Topics in Dynamic Macroeconomic Theory: On the Causes and Consequences of Income Growth and Uncertainty

DISSERTATION

zur Erlangung des Grades eines Doktors der Wirtschaftswissenschaft

eingereicht an der
Wirtschaftswissenschaftlichen Fakultät
der Universität Regensburg

vorgelegt von
Christian Bauer

Berichterstatter:
Prof. Dr. Lutz Arnold (Universität Regensburg)
Prof. Dr. Alfred Maußner (Universität Augsburg)

Tag der Disputation: 30.10.2008

Overview

Contents

List of Figures

Part I

Trade with Imperfect Competition and Heterogeneous Firms

Chapter 1

The "Almost Static" Melitz (2003) Trade Model with Fixed Export Costs

In a series of influential papers, Clerides, Lach and Tybout (1998), Bernard and Jensen (1999,2004), Bernard, Eaton, Jenson, and Kortum (2003), Eaton, Kortum, and Kramarz (2004), among others, substantiate the existence of large and persistent productivity differences between firms in narrowly defined industries. They show, inter alia, that even in net export sectors, the fraction of exporting firms is small, and that the propensity of a firm to export is largely driven by its productivity level.[1] While exporting does not feed back to exporters' productivity, a more pronounced exposure to trade reallocates resources towards the more productive firms and forces the least productive ones to exit. Recent theoretical advances by Melitz (2003), Helpman, Melitz, and Yeaple (2004), Melitz and Ottaviano (2005), Yeaple (2005), and Bernard, Redding, and Schott (2007), have confronted this evidence by explicitly modeling costly trade in environments where firms have heterogeneous marginal costs and face market entry costs. With a focus on reallocation and firm-selection, this "new new trade" theory takes technologies as given, and analyzes trade liberalization in settings with zero steady-state productivity growth. If productivity growth is positive and endogenous, trade-induced increases in productivity may come at the expense of profit-driven product innovation (Baldwin and Robert-Nicoud, 2007). In knowledge-driven growth models, the aggregate effect on welfare then depends on the strength of knowledge spillovers in R&D (Gustafsson and Segerstrom, 2007). Similarly, Atkeson and Burstein (2007) show that the exposure to trade fosters process innovation if

[1]See Bernard, Jensen, Redding, and Schott (2007) for a compendium on firms in international trade.

only a subset of firms exports, but comes at the expense of a large decline in product innovation.[2] We contribute to this literature by generalizing the "almost static" Melitz (2003) trade model to include semi-endogenous variety growth, production using physical capital, and capital accumulation. To begin with, we introduce the Melitz (2003) model.

1.1 The Melitz (2003) Model

Melitz (2003) has become the workhorse model in international trade theory with productivity differences at the firm level.[3] Five years after being published in Econometrica, Google Scholar reports almost 1,000 citations for "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity".[4] In a nutshell, Melitz (2003) adds firm heterogeneity and fixed costs of exporting to Krugman's (1980) model of intra-industry trade under monopolistic competition and variable trading costs. Adjusted to imperfect competition and embedded in a general equilibrium framework, Melitz' model employs the Hopenhayn (1992) mechanism of firm development based on productivity differences to explain the markedly heterogeneous impact of international trade on individual firms and its potential to shift market shares from small and less productive firms to big and more productive firms within narrowly defined industries. The beauty of Melitz's (2003) model is that, while remaining analytically tractable, it takes several of the stylized facts about firms in international trade into account, for which there is no scope within the framework of the new trade theory.[5] In particular, the model provides for pronounced intra-industry productivity differences across firms, small fractions of

[2] Atkeson and Burstein (2007) conclude from their quantitative results that excluding process innovation generates very similar dynamic welfare gains, even if process innovation reacts very elastically to trade liberalization.

[3] Other important contributions in international trade with heterogeneous firms include Bernard, Eaton, and Kortum (2003), who study heterogeneous firms under Bertrand competition and no fixed costs, Helpman, Melitz, and Yeaple (2004), who include the possibility of FDI, Melitz and Ottaviano (2005), who dispense with the constant elasticity of substitution assumption between horizontally differentiated goods, and Yeaple (2005), who explains the heterogeneity of a priori identical firms by different technology choices and workers with heterogenous skills. See Bernard, Jensen, Redding, and Schott (2007) for a survey.

[4] According to the same source, Krugman's (1980) paper "Scale Economies, Product Differentiation, and the Pattern of Trade" received about 1330 citations in the past 28 years.

[5] The neoclassical (or "traditional") trade theory prior to Krugman (1979a,b) employs aggregate production functions and constant returns to scale technologies, which leave the firm size indeterminate and thus also has no scope for the impact of trade at the firm level.

exporting firms in all industries, substantially higher productivity among exporting firms, high fixed costs of exporting, and no productivity gains from learning by exporting.

The main results of Melitz (2003) are that trade liberalization i) forces the least productive firms to shut down, ii) shifts resources from less productive firms to more productive firms, iii) allows some firms to start exporting, and iv) increases profits only for the most productive exporters. Opening to trade therefore generates substantial turmoil among firms. Melitz (2003) predicts, however, that the firm selection and resource reallocation induced by the exposure to international trade unambiguously increases aggregate productivity and welfare.

By accounting for intra-industry firm heterogeneity and an extensive margin of trade liberalization, the model considerably extends the canonical Helpman-Krugman framework and has proven extremely fruitful for research into firms in international trade (cf. Bernard, Jensen, Redding, and Schott, 2007, for a survey of this strand of literature).

We next introduce the Melitz (2003) model. In deriving the main results on trade liberalization, we pay special attention to the mechanism at work. To conclude, we specify the fairly general distribution function for firms' productivity levels in Melitz (2003) and improve our understanding by deriving a closed form solution in Section 2.

1.2 Autarky

Following Melitz (2003), we start by describing the model environment in autarky.

1.2.1 Model Setup

The model is described by preferences, firm-specific production technologies, and assumptions about the production structure in the economy. We consider each in turn. The model is dynamic in nature, but the analysis is confined to "almost static" stationary equilibria so that there is no need for a time index. The notion of this "stationary equilibrium" was introduced by Hopenhayn (1992) and will be explained in more detail below.

Demand

The model is populated by households whose actions can be summarized in the behavior of a representative agent.[6] Her preferences exhibit love of variety and are given by

$$U = \left[\int_{j \in J} x(j)^{\alpha} \, dj \right]^{1/\alpha}, \quad 0 < \alpha < 1, \tag{1.1}$$

where J is the set of available products and $x(j)$ denotes the quantity consumed of good $j \in J$. All available goods are equally valuable substitutes with a constant elasticity of substitution between any two goods equal to $\varepsilon \equiv 1/(1-\alpha) \in (1, \infty)$. If more products become available, e.g. because of an increasing number of domestic producers or due to international trade, the newly available goods similarly substitute imperfectly for previously available goods.

Let the aggregate consumption expenditures of the representative consumer be equal to E,

$$\int_{j \in J} p(j) \, x(j) \, dj = E. \tag{1.2}$$

Utility is then maximized by choosing consumption bundles so that the marginal rate of substitution between any pair of goods equals the relative price of the two goods,

$$\frac{\partial U / \partial x(j)}{\partial U / \partial x(j')} = \frac{p(j)}{p(j')}, \tag{1.3}$$

and the marginal utility per unit of expenditures satisfies

$$\frac{\partial U / \partial x(j)}{p(j)} = \lambda, \quad \forall j \in J, \tag{1.4}$$

where λ is the Lagrange multiplier associated with the constraint in (1.2). From (1.1) we have

$$\frac{\partial U}{\partial x(j)} = U^{\frac{1}{\alpha}-1} x(j)^{\alpha - 1}, \ \forall j \in J. \tag{1.5}$$

As an aside, note that, from (1.5), the marginal utility on the left hand side of (1.4) can be expressed as $U^{\frac{1}{\alpha}-1} x(j)^{\alpha-1} = \lambda p(j)$, which implies that

$$x(j) \, p(j)^{\varepsilon} = U^{\frac{1}{\alpha}} \lambda^{-\varepsilon} \tag{1.6}$$

[6] Instead of working with a representative consumer, we could equivalently assume a mass of identical consumers. In this case, the individual demand functions derived from (1.1) for each consumer can be aggregated across all consumers and yield demand functions identical to the ones above, with E then referring to the aggregate consumption expenditures of all consumers. For an introduction on the existence of a representative agent see, e.g., Huang and Litzenberger (1988), Chapter 5.

is equal across all j's. Using (1.5) in (1.3), optimality requires

$$\left[\frac{x(j)}{x(j')}\right]^{\alpha-1} = \frac{p(j)}{p(j')}, \ \forall j, j' \in J,$$

or, after solving for $x(j)$,

$$x(j) = \left[\frac{p(j')}{p(j)}\right]^{\varepsilon} x(j').$$

Substituting for $x(j)$ in the budget constraint in (1.2) with this expression yields

$$E = \int_{j\in J} p(j)\, x(j)\, dj = \int_{j\in J} p(j)^{1-\varepsilon} x(j')\, p(j')^{\varepsilon} dj. \tag{1.7}$$

We can pull $x(j')\, p(j')^{\varepsilon}$ out of the integral in (1.7) and get ($\forall j \in J$)

$$E = x(j)\, p(j)^{\varepsilon} \int_{j\in J} p(j)^{1-\varepsilon} dj. \tag{1.8}$$

Solving for $x(j)$ yields the optimal demand for each available good:

$$x(j) = \frac{Ep(j)^{-\varepsilon}}{\int_{j\in J} p(j)^{1-\varepsilon} dj}, \ \ \forall j \in J.$$

The denominator thereby inversely reflects the aggregate price level ($\varepsilon > 1$). Defining the aggregate price index as

$$P \equiv \left[\int_{j\in J} p(j)^{1-\varepsilon} dj\right]^{\frac{1}{1-\varepsilon}}, \tag{1.9}$$

the demand functions for available products are given by

$$x(j) = \frac{P^{\varepsilon-1}E}{p(j)^{\varepsilon}} = \left[\frac{P}{p(j)}\right]^{\varepsilon} \frac{E}{P}, \ \ \forall j \in J. \tag{1.10}$$

A direct consequence of the love of variety embodied in (1.1) is that the demand for each good is positive as long as its price is finite. The quantity demanded of each good is c.p. decreasing in its own price and increasing in the expenditure level and the price index ($\varepsilon > 1$) since the other available goods are competing substitutes.[7] The demands in (1.10) are iso-elastic and the price elasticity of demand is the same for all available goods,

$$-\frac{\partial x(j)}{\partial p(j)} \frac{p(j)}{x(j)} = \varepsilon x(j)\, p(j)^{-1} \frac{p(j)}{x(j)} = \varepsilon. \tag{1.11}$$

[7] An increase in the price index comes both with an income and a substitution effect for the quantity demanded of each product. First, high overall prices lower the quantity demanded of all goods (the P in the denominator in (1.10)). Second, the quantity demanded from relatively cheap products is larger (the P in the nominator in (1.10)). The latter effect dominates as $\varepsilon > 1$.

A 1% increase in in the total amount spent on consumption c.p. raises the quantity demanded of each available variety by 1%,

$$\frac{\partial x(j)}{\partial E}\frac{E}{x(j)} = 1,$$

i.e., demands are homothetic. Using (1.10), the spending on each product equals

$$e(j) \equiv p(j)\,x(j) = \left[\frac{P}{p(j)}\right]^{\varepsilon-1} E, \quad \forall j \in J. \tag{1.12}$$

The consumer spends more on a given variety the cheaper this product is relative to the aggregate price index. If prices scatter over a wide range, there is lots of variation in the equilibrium expenditure and demand profile across different products if the consumer heavily substitutes relatively expensive goods against cheaper goods, i.e. if $\varepsilon \to \infty$. Evidently, there is only little variation in expenditures and demands even if prices scatter widely if ε is close to 1.

Production

The economy is endowed with a single non-durable and inelastically supplied primary factor, "labor", that serves as the numéraire. Firms operate under monopolistic competition and produce one product variety each (j thus indexes both the firm and its product). Entry into the industry is costly (all costs are wage payments for hiring the required quantity of labor). Upon paying a uniform sunk cost f_e, i.e. "building a firm", the entrants have access to a constant marginal costs technology. In addition to the variable costs, the production of output also incurs (quasi-) fixed "overhead" costs f, so that production occurs under increasing returns to scale. The production structure is therefore closely akin to Krugman (1980) with the only difference that $f_e > 0$. Crucially, however, Melitz (2003) adds two further ingredients. First, firms are heterogeneous with respect to the marginal productivity of their technologies.[8] Second, firms do not know their productivity before entering the market. Newcomers thus face uncertainty about their market value before entry. Immediately after paying the entry costs, each firm learns its technology, which is given by the costs $c(j,x)$ of producing x units of good j,

$$c(j,x) = \frac{x}{\varphi(j)} + f, \tag{1.13}$$

where $\varphi(j)$ denotes the firm specific marginal productivity of labor. Formally, each $\varphi(j)$ is randomly drawn from a common and commonly known distribution $G(\varphi)$ defined over $\mathbb{R}^+$, $G'(\varphi) \equiv g(\varphi) > 0$ and $\int_{\mathbb{R}^+} \varphi^{\varepsilon-1} dG(\varphi) < \infty$.[9] Firms with a high marginal productivity of labor, i.e. a high φ, are able

[8] A firm's productivity can equivalently be interpreted as the quality of its product.

[9] The latter regularity condition ensures that, in equilibrium, the average productivity level is finite.

to produce a given amount of output at a lower wage cost than less productive firms. The overhead costs f arise in each period and are identical across all producers.

Firm Entry and Exit

When a firm has learned about its production technology, it decides whether to shut down or to start production using the realized productivity level. If the firm starts production, it engages in monopolistic competition with other producers until it is hit by a deadly productivity shock which then forces instantaneous exit. This idiosyncratic death shock originates from a memoryless stochastic (Poisson) process that exogenously hits any producer with a common probability $0 < \delta < 1$ in each period. Hence, a fraction δ of all producers is forced to exit in every period. Following Melitz (2003), our attention is confined to stationary equilibria where there is continuous entry and exit of firms, but the aggregate productivity distribution of producers remains constant over time. In these equilibria, aggregate output, the number of producers, and each producer's profit is constant over time (until the firm exits). While the unconditional exogenous exit of firms is not particularly realistic, it is an easy way to enable the transition between different stationary distributions of productivity levels after an exogenous change in the environment. It also implies that the distribution of productivity levels in the stationary equilibrium is determined by the distribution of productivity levels of new entrants (which evidently must be stationary itself).[10]

For simplicity, the interest rate is set to zero so that there is no discounting other than forming expectations over a producers' lifetime (we show in Appendix 5.A how a Poisson shock with arrival rate δ translates into the usual discounting with discount factor $e^{-\delta t}$).

A Remark on the Capital Market

Every new firm must raise the entry costs f_e. From the point of view of the households, this investment is uncertain in two dimensions. First, with probability $1 - G(\varphi^*)$ it yields a positive return $\pi(\varphi)$ where $\varphi \geq \varphi^*$. The expected value of this return is uncertain as well. An increase in the cutoff productivity raises the expected return conditional on an investment that pays at all, but lowers the fraction of investments that do so. With probability $G(\varphi^*)$, the financed firm fails to draw a productivity $\varphi \geq \varphi^*$ and the return accruing to the investors is zero. The aggregate income must

[10] In Hopenhayn's (1992) article, firms' death rates differ so that the distribution of entrants' productivities does not coincide with the stationary equilibrium distribution of producers' productivity levels.

thus not be equal to the aggregate wage payments at all times. Note, however, that this will be the case in the stationary equilibrium. To see this, recall that Melitz (2003) simplifies his model by assuming zero time discounting (the only discounting is with respect to firm's market values due to the productivity death shock, cf. Appendix 5.A). In the stationary equilibrium, therefore, the sum of dividend payments is equal to the sum of wage payments to the entry workers, i.e. the aggregate investment in new firms. Accordingly, there is no net income from investing in new firms, the income of entry workers and payments of investors cancel (cf. Section 1.2.2, Footnote 17, and Section 1.2.3 below). The present value of these income flows would be different, however, if there was positive time discounting. As noted by Melitz (2003, Footnote 16), the absence of a positive net investment income is not directly related to the aggregation of heterogenous firms but rather due to the stationary equilibrium assumption (and zero discounting). A final remark: While Melitz (2003) explicitly chooses to ignore intertemporal assessments, the assumption of zero time discounting requires Melitz (2003) to assume an instantaneous utility function equal to the Dixit-Stiglitz index. Households are thus risk-neutral. We show in a dynamic model of growth and trade in Chapter 3 that risk aversion dues not alter Melitz' (2003) results.

Optimal Firm Behavior

A firm's decision to start production is based on the prospects of its future profits. Since the chance of dying is the only source of uncertainty once $\varphi(j)$ is revealed, each firm can simply calculate the expected return from production using its profit maximizing sequence of output quantities and compare the resulting returns to the necessary fixed costs.

Profit maximization.

Denote by $\pi(j,x)$ the period profit of a firm that produces x units of output,

$$\pi(j,x) = p(x(j))\,x(j) - c(j,x)$$

where $p(x(j))$ is the inverse demand for good j implied by (1.10). Given the consumers' demand curve, each producer chooses the profit maximizing output quantity, i.e. chooses its output so as to equate marginal revenues to (the firm specific) marginal costs. From (1.10) and (1.13), optimality requires

$$\frac{\partial p(j)}{\partial x(j)} x(j) + p(j) = \frac{1}{\varphi(j)}.$$

Using $p(x(j))$ as implied by (1.10), we get the optimal price as a function of the constant price elasticity of demand and the marginal productivity of labor employed in firm j,[11]

$$1-\frac{1}{p(j)\,\varphi(j)}=-\frac{1}{\frac{\partial x(j)}{\partial p(j)}\frac{p(j)}{x(j)}}=\frac{1}{\varepsilon}.$$

Solving for $p(j)$, the profit maximizing price given the iso-elastic demand (cf. (1.11)) is the usual markup $1/\alpha$, but over firm specific marginal costs $1/\varphi(j)$:

$$p(j)=\frac{1}{\alpha\varphi(j)}. \tag{1.14}$$

Firms with a higher marginal productivity of labor charge lower prices and, from (1.11), sell higher quantities and thus have larger market shares. We derive the firm's equilibrium output as a function of the total consumption expenditures $E\equiv\int_{j\in J}e(j)\,dj$. Substituting for $p(j)$ in (1.10) with (1.14), the quantity supplied by firm j is

$$x(j)=[\alpha\varphi(j)]^{\varepsilon}P^{\varepsilon-1}E. \tag{1.15}$$

Given that all firms with identical productivity levels charge the same price and vice versa, $p(j)=p(j')\Leftrightarrow\varphi=\varphi'$ see (1.14), we can state the equilibrium prices and quantities in (1.10) as functions of the firm's productivity levels,

$$p(\varphi)\equiv\frac{1}{\alpha\varphi}, \tag{1.16}$$
$$x(\varphi,P,E)\equiv(\alpha\varphi)^{\varepsilon}P^{\varepsilon-1}E. \tag{1.17}$$

Using these functions, the indirect/maximized revenue and profit functions for a firm with productivity φ are given by

$$r(\varphi,P,E)\equiv p(\varphi)\,x(\varphi,P,E)=\left[\frac{p(\varphi)}{P}\right]^{1-\varepsilon}E=(\alpha\varphi P)^{\varepsilon-1}E, \tag{1.18}$$
$$\pi(\varphi,P,E)=r(\varphi)-\frac{r(\varphi)}{p(\varphi)\,\varphi}-f=(1-\alpha)\,r(\varphi)-f=\frac{(\alpha\varphi P)^{\varepsilon-1}E}{\varepsilon}-f. \tag{1.19}$$

Profits are the usual fraction $1-\alpha$ of revenues minus the fixed overhead costs. Note that revenues and profits vary across firms with different levels of productivity. Profits are strictly increasing in φ, linear in $\varphi^{\varepsilon-1}$, and $\pi(0,P,E)=-f<0$.

To summarize, more productive firms (with a higher marginal productivity of labor φ) c.p. produce more output, charge lower prices, and earn higher revenues and profits than less productive firms.

[11] $\frac{\partial p(j)}{\partial x(j)}=\frac{1}{\frac{\partial x(j)}{\partial p(j)}}$ since $p(x)$ is the (differentiable) inverse of $x(p)$ in (1.10).

Evidently, more productive firms are larger, i.e. have higher market shares and employ more labor in equilibrium.

Profitable production.

When a firm has learned its productivity, it decides whether it can produce profitably given the realized technology. Since the entry cost f_e is sunk, the decision to start production boils down to wether or not a firm is able to cover its fixed overhead costs. The firm thus makes a simple forward looking decision: if the present value of operating profits exceeds the present value of fixed cost of production, it will start to produce. Otherwise, it will immediately shut down and exit the market. In a stationary equilibrium, where the distribution of productivity levels and also P and E are constant, the present value of profits of a firm with productivity φ at time $t = 0$ equals[12]

$$\begin{aligned} v(\varphi, P, E) &= \max\left\{0, \sum_{t=0}^{\infty} (1-\delta)^t \pi(\varphi, P, E)\right\} = \max\left\{0, \pi(\varphi, P, E) \sum_{t=0}^{\infty} (1-\delta)^t\right\} = \\ &= \max\left\{0, \frac{\pi(\varphi, P, E)}{\delta}\right\}. \end{aligned} \tag{1.20}$$

Each firm treats the aggregates P and E as given. We indicate the firm's view using a semicolon and write $\pi(\varphi; P, E)$. Since the equilibrium profits are c.p. strictly increasing in φ and $\pi(0, P, E) < 0$ (cf. 1.19), (1.20) implies that there is a strictly positive unique cutoff productivity level φ^*, below which firms decide to exit immediately:[13]

$$\varphi^* \equiv \inf\{\varphi : \pi(\varphi; P, E) > 0\} \quad (> 0). \tag{1.21}$$

[12]The last equation in (1.20) follows from an infinite geometric series. For $0 < \bar{\delta} \equiv 1 - \delta < 1$,

$$\begin{aligned} \sum_{t=0}^{\infty} \bar{\delta}^t &= 1 + \bar{\delta} + \bar{\delta}^2 + \ldots, \\ \bar{\delta} \sum_{t=0}^{\infty} \bar{\delta}^t &= \bar{\delta} + \bar{\delta}^2 + \ldots. \end{aligned}$$

Subtracting the second equation from the first gives

$$\left(1 - \bar{\delta}\right) \sum_{t=0}^{\infty} \bar{\delta}^t = 1,$$

so that, using the definition of $\bar{\delta}$,

$$\sum_{t=0}^{\infty} (1-\delta)^t = \frac{1}{\delta}.$$

[13]The infimum is used as firms with productivity φ^* earn zero profits (and thus are indifferent between production and exit), so that the cutoff φ^* is not part of the set of $\varphi' s$ that permit $\pi(\varphi) > 0$.

Only firms with productivity larger than φ^* are able to operate profitably and thus start production in the first place. Accordingly, the distribution of active firms' productivity levels μ is the distribution of the population of productivity levels, $G(\varphi)$, conditional on a sufficiently high productivity level that permits profitable entry, i.e. $\varphi \geq \varphi^*$ (there are no producers with productivity $\varphi < \varphi^*$ and the probability for entry into manufacturing is $1 - G(\varphi^*) > 0$):

$$\mu(\varphi, \varphi^*) \equiv \begin{cases} \frac{G(\varphi)}{1-G(\varphi^*)} & \text{for } \varphi \geq \varphi^* \\ 0 & \text{else} \end{cases} . \tag{1.22}$$

The equilibrium distribution is endogenously determined from the exogenous distribution via the upper bound of its support. Evidently, the underlying distribution thus determines the characteristics of the equilibrium distribution. A notable special case is the Pareto distribution, which, if truncated, again yields a Pareto distribution with the same shape parameter (cf. Chapter 2).

While the overall equilibrium distribution of $\varphi' s$ remains exogenous, its support $[\varphi^*, \infty)$ and thereby its first moment are endogenously determined in equilibrium.

1.2.2 The Autarky Equilibrium

Following Melitz (2003), we solve the model for a stationary equilibrium where the distribution of productivities $\mu(\varphi)$ is stationary, the number of firms is constant and, as usual, prices and quantities maximize firms' profits and households' utility and all markets clear.

The cutoff productivity is determined together with the average profit of producers by the endogenous entry and exit decisions of the firms. In particular, we pin down the threshold productivity φ^* by two equations that naturally emerge from (i) free entry into production and (ii) the minimum productivity requirement necessary for profitable production (cf. (1.21)).

Free Entry

On the one hand, there is potentially unbounded entry of firms if outsiders expect positive profits in the market. On the other hand, no firm is willing to enter if it expects negative profits from production in equilibrium so that the firm deaths would lead to a continuous decline in the mass of producers. Free entry therefore requires that firms expect zero profits from entering the market, i.e. the ex ante expected value of a firm must match the entry costs,[14]

$$\int_{\varphi\in\mathbb{R}+} v(\varphi, P, E)\, dG(\varphi) = f_e.$$

Since $v(\varphi) = 0$ for all $\varphi < \varphi^*$ from (1.20) and (1.21), we can equivalently express this equilibrium condition as

$$\int_{\varphi^*}^{\infty} \frac{\pi(\varphi, P, E)}{\delta} g(\varphi)\, d\varphi = f_e.$$

Inserting the density $g(\varphi)$ of equilibrium productivity levels from (1.22),

$$g(\varphi) = [1 - G(\varphi^*)]\, \mu'(\varphi), \tag{1.23}$$

we then have

$$\int_{\varphi^*}^{\infty} \pi(\varphi)\, [1 - G(\varphi^*)]\, \mu'(\varphi)\, d\varphi = \delta f_e.$$

Dividing by $1 - G(\varphi^*)$, we find that free entry equates the expected per period profit of a producer (for whom the productivity distribution is given by $\mu(\varphi)$) to the "annuity payment" of the expected entry costs (on average, it takes $1/[1 - G(\varphi^*)]$ draws for a sufficiently productive technology with $\varphi \geq \varphi^*$):

$$\int_{\varphi^*}^{\infty} \pi(\varphi)\, d\mu(\varphi) = \frac{\delta f_e}{1 - G(\varphi^*)}. \tag{1.24}$$

Hence, with free entry, the average profit of a producer, i.e. the average profit of a firm conditional on starting production after market entry, [15]

$$\bar{\pi} \equiv \int_0^{\infty} \pi(\varphi, P, E)\, d\mu(\varphi) = \int_{\varphi^*}^{\infty} \pi(\varphi, P, E)\, d\mu(\varphi), \tag{1.25}$$

depends on the cutoff productivity, P, and E, $\bar{\pi} = \bar{\pi}(\varphi^*, P, E)$. Taken together, (1.24) and the definition in (1.25) yield a first relation between the average profits of a producer and the cutoff productivity, namely

$$\int_{\varphi\in\mathbb{R}+} v(\varphi)\, dG(\varphi) - f_e \leq 0 \;\Leftrightarrow\; \bar{\pi} = \frac{\delta f_e}{1 - G(\varphi^*)}. \tag{FE_a}$$

We refer to this equation as free entry condition, FE for short, and use the subscript 'a' to indicate autarky. Put differently, given E, the cutoff must be such that P and E drop out from the profit

[14] We show below that $v(\varphi, P, E)$ is decreasing in the number of producers (like $v(j)$ in Krugman, 1980), so that the above argument for zero expected profits in fact holds. Note, however, that off the stationary equilibrium considered in the main text, the expected value of entry may well be strictly negative for some time during the transition to the stationary equilibrium.

[15] Since there are no producers with productivities below the cutoff, $\mu(\varphi) = \mu'(\varphi) = 0$ for $\varphi < \varphi^*$, cf. (1.22).

function in (1.19) in a way that (FE_a) holds. The FE curve is upward sloping in $(\varphi^*, \bar{\pi})$-space as $g(\varphi) \geq 0$ (see (FE_a)).

With (FE_a), we have derived a first equation in $\bar{\pi}$ and φ^*. A second equation in the same variables is readily obtained from the determination of the cutoff itself.

The Zero Cutoff Profit Condition

Using (1.19), the cutoff in (1.21) defines a relation between the profits of firms operating with productivity φ^* and the cutoff productivity level itself. Evidently, $\pi(\varphi^*)$ is zero (cf. (1.19), (1.20), and (1.21)):[16]

$$\pi(\varphi^*, P, E) = (1-\alpha)\, r(\varphi^*, P, E) - f = 0. \tag{1.26}$$

In view of (FE_a), all we have to do to obtain a system of two equations in the same variables is to express the left hand side of (1.26) in terms of the average profits of producers, $\bar{\pi}$. In a stationary equilibrium, this "translation" is easily achieved since the cutoff is a "sufficient statistic" in that it contains all relevant information about the equilibrium distribution of productivity types, see (1.22). Focussing on the stationary equilibrium, we achieve the transition from $\pi(\varphi^*)$ to $\bar{\pi}$ in two steps. First, we explicitly derive the productivity level associated with $\bar{\pi}$, i.e. an average productivity. Second, we employ this average productivity to relate $\bar{\pi}$ to $\pi(\varphi^*)$, using the functional form in (1.19) for the average and the cutoff productivity level.

The average productivity level. Together with (1.25), the profit function in (1.19) defines an average productivity level $\tilde{\varphi}$ by $\bar{\pi} \equiv \pi(\tilde{\varphi}, P, E)$, i.e. $\tilde{\varphi}$ is the productivity of a firm whose profit equals $\bar{\pi}$. Substituting for $\pi(\varphi, P, E)$ with (1.19) in (1.25) and yields

$$\pi(\tilde{\varphi}, P, E) = \int_0^\infty \left[(1-\alpha)(\alpha\varphi P)^{\varepsilon-1} E - f\right] d\mu(\varphi) = (1-\alpha)(\alpha P)^{\varepsilon-1} E \int_0^\infty \varphi^{\varepsilon-1} d\mu(\varphi) - f.$$

The second equality follows because the price index can be pulled out of the integral (from (1.6) and (1.8), $P^{1-\varepsilon} = \lambda^\varepsilon E U^{-1/\alpha}$ is the same for all φ) and $\int_0^\infty d\mu(\varphi) = 1 - \mu(0) = 1$. Solving for $\tilde{\varphi}$ using

[16] Since $\pi(\varphi^*) = 0$, equation (1.19) implies that φ^* is a decreasing function of the price index and the aggregate expenditures,

$$\varphi^* = \frac{(\varepsilon f)^{\frac{1}{\varepsilon-1}}}{\alpha E^{\frac{1}{\varepsilon-1}} P} = \frac{(\varepsilon f)^{\frac{1}{\varepsilon-1}}}{\alpha \left(\frac{E}{P}\right)^{\frac{1}{\varepsilon-1}} P^{\frac{1}{\alpha}}}.$$

Intuitively, a decline in real consumption requires a higher efficiency at the firm level. Note, however, that we yet have to solve for P since given the number of producers, the price index also reflects an average productivity, cf. (1.9) and (1.16).

(1.19) yields the average productivity:

$$\tilde{\varphi} = \left[\int_0^{\infty} \varphi^{\varepsilon - 1} d\mu(\varphi) \right]^{\frac{1}{\varepsilon - 1}}. \tag{1.27}$$

Evidently, as $G(\varphi)$ is exogenous and the support of μ is determined by the cutoff, the average productivity depends only on φ^*. To get an explicit expression for $\tilde{\varphi}$ as a function of the cutoff, we can substitute for $d\mu(\varphi) = \mu'(\varphi)\, d\varphi$ using (1.22):

$$\begin{aligned} \tilde{\varphi}(\varphi^*) &= \left[\int_0^{\infty} \varphi^{\varepsilon - 1} \mu'(\varphi)\, d\varphi \right]^{\frac{1}{\varepsilon - 1}} = \left[\int_0^{\varphi^*} \varphi^{\varepsilon - 1} \mu'(\varphi)\, d\varphi + \int_{\varphi^*}^{\infty} \varphi^{\varepsilon - 1} \mu'(\varphi)\, d\varphi \right]^{\frac{1}{\varepsilon - 1}} = \\ &= \left[\frac{1}{1 - G(\varphi^*)} \int_{\varphi^*}^{\infty} \varphi^{\varepsilon - 1} dG(\varphi) \right]^{\frac{1}{\varepsilon - 1}}. \end{aligned} \tag{1.28}$$

Hence, we are now in the position to relate $\bar{\pi} = \pi(\tilde{\varphi}, P, E)$ to $\pi(\varphi^*, P, E)$, so that (1.26) and (FE_a) can be solved for the two unknowns $\bar{\pi}$ and φ^*.

The cutoff productivity level and the average profits. To substitute for $\pi(\varphi^*, P, E)$ in (1.26), note that all producers' relative revenues depend only on their relative productivities. Dividing (1.18) for productivity levels φ and φ' verifies

$$\frac{r(\varphi, P, E)}{r(\varphi', P, E)} = \left(\frac{\varphi}{\varphi'} \right)^{\varepsilon - 1} \quad \forall \varphi \geq \varphi^*. \tag{1.29}$$

In particular, using the average and the cutoff productivity, this yields

$$\frac{r(\tilde{\varphi}, P, E)}{r(\varphi^*, P, E)} = \left(\frac{\tilde{\varphi}}{\varphi^*} \right)^{\varepsilon - 1} \quad \forall \varphi \geq \varphi^*. \tag{1.30}$$

From (1.19), $r(\varphi, P, E) = \varepsilon [\pi(\varphi, P, E) + f]$ so that the average profits relate to $\pi(\varphi^*, P, E)$ $(= 0)$ by

$$\pi(\tilde{\varphi}, P, E) + f = \left(\frac{\tilde{\varphi}}{\varphi^*} \right)^{\varepsilon - 1} [\pi(\varphi^*, P, E) + f].$$

Solving for $\pi(\tilde{\varphi}, P, E) = \bar{\pi}$ and noting (1.28), we arrive at

$$\pi(\varphi^*, P, E) = 0 \;\Leftrightarrow\; \bar{\pi} = \left\{ \left[\frac{\tilde{\varphi}(\varphi^*)}{\varphi^*} \right]^{\varepsilon - 1} - 1 \right\} f, \tag{ZCP_a}$$

where $\tilde{\varphi}(\varphi^*)$ is explicitly given in (1.28). Melitz (2003) coins this equation the zero cutoff profit condition (ZCP for short), since it simply rephrases that firms with productivity levels below φ^* would incur losses while firms with $\varphi > \varphi^*$ earn positive profits and $\pi(\varphi^*) = 0$.

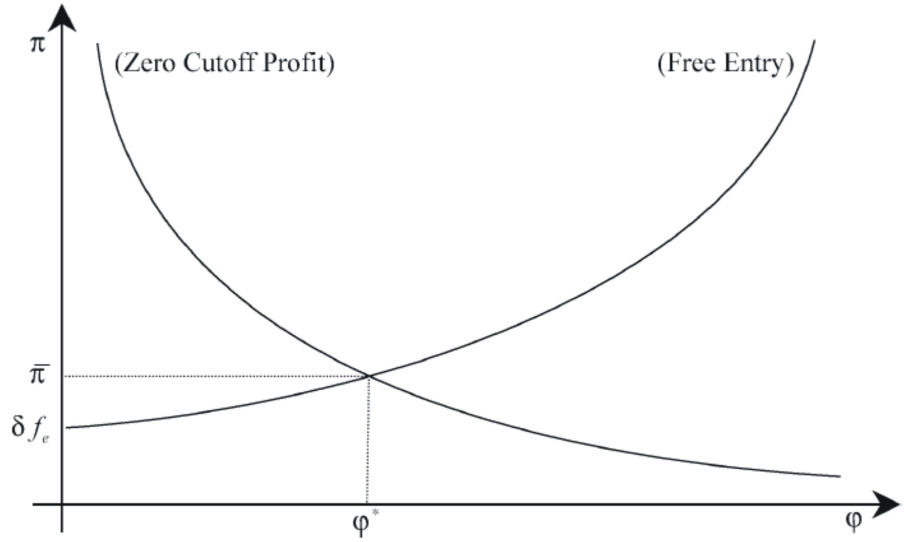


Figure 1.1: Determination of the Autarky Cutoff (Adopted from Melitz, 2003, p. 1704)

Characterization of the Autarky Equilibrium

Equations (FE_a) and (ZCP_a) can now be solved for the equilibrium values of $\bar{\pi}$ and φ^*. Given these values, we can then characterize the entire equilibrium outcome and its welfare properties. Note that the cutoff productivity level and the average profits are determined by fixed production and entry costs independently of the endogenous variables P and E (and also do not depend on the country size). Graphically, Figure 1.1 depicts the FE and the ZCP curves in the $(\varphi^*, \bar{\pi})$-space. From (FE_a), $\bar{\pi}$ is increasing in φ^* ($g(\varphi) > 0$ for all φ so $G'(\varphi^*) > 0$) from $\bar{\pi}(0) = \delta f_e$ ($\varphi \in \mathbb{R}^+$ thus $G(0) = 0$) to $\lim_{\varphi^* \to \infty} \bar{\pi}(\varphi^*) = \infty$. The average profits increase in the cutoff productivity because the average productivity is increasing in the cutoff and profits are strictly increasing in productivity. In Appendix 1.B, we present Melitz' (2003) proof that the ZCP cuts the FE curve exactly once (viz., from above), so that the solution to (FE_a) and (ZCP_a) in fact exists and is unique.

Given the equilibrium values for $\bar{\pi}$ and φ^*, we can solve for the number of producers/available varieties. From (1.19), the average profit of producers equals

$$\bar{\pi} = (1-\alpha)(\alpha\tilde{\varphi}P)^{\varepsilon-1} E - f.$$

Let M denote the (constant) number of producers in equilibrium. Since $\mu(\varphi) = 0$ for $\varphi < \varphi^*$, $P^{1-\varepsilon}$ can be rewritten as (see the detailed derivation in Appendix 1.A)

$$P^{1-\varepsilon} = \int_0^M p(j)^{1-\varepsilon} dj = \int_{\varphi^*}^{\infty} (\alpha\varphi)^{\varepsilon-1} M\mu'(\varphi) d\varphi = \alpha^{\varepsilon-1} M \int_0^{\infty} \varphi^{\varepsilon-1} d\mu(\varphi) = M(\alpha\tilde{\varphi})^{\varepsilon-1}. \quad (1.31)$$

As an aside, note that the aggregate price index simply is the price index with homogenous firms and a common marginal productivity of $\tilde{\varphi}$, $P = M^{1/(1-\varepsilon)}/(\alpha\tilde{\varphi}) = p(\tilde{\varphi}) M^{1/(1-\varepsilon)}$, cf. (1.16). From (1.31), $(\alpha\tilde{\varphi}P)^{\varepsilon-1} = 1/M$ so that with $1-\alpha = 1/\varepsilon$

$$\bar{\pi} = \frac{E}{\varepsilon M} - f. \tag{1.32}$$

We are thus left to find a second equation to determine M and E. Melitz’ (2003) model is closed by the assumption of market clearing. With full employment, total consumption expenditures equal L. To see this explicitly, start with $L = L_e + L_p$ where L_e and L_p is labor employed in entry and in production, respectively, and L is the total stock of labor available. If we denote by M_e the mass of firms that incur the entry costs in each period and resort to a law of large numbers, the resulting mass of new producers is $M_e [1 - G(\varphi^*)]$. In the stationary equilibrium, the mass of entering producers must equal the mass of dying firms, δM. Hence, the labor market clearing condition becomes

$$L = M_e f_e + L_p = \frac{\delta f_e}{1 - G(\varphi^*)} M + L_P.$$

Recognizing that $\delta f_e / [1 - G(\varphi^*)] = \bar{\pi}$ from free entry into production, see (1.24), we verify that total income (i.e. profit plus labor income) equals the total consumption expenditures,

$$L = \bar{\pi} M + L_P = E. \tag{1.33}$$

The entry worker’s income cancels since it equals the investment in new firms that cannot be used for consumption.[17] Returning to the determination of M, we can simply plug $E = L$ in (1.32) and find the equilibrium number of firms

$$M = \frac{L}{\varepsilon(\bar{\pi} + f)}. \tag{1.34}$$

From (FE_a) and (ZCP_a), the average profit is independent of L. The number of firms is thus higher in countries with a large labor endowment relative to countries with little labor resources (“large” countries have a more diversified product portfolio than “small” countries). Equation (1.34) shows nicely the tension between the number of available products and the productivity with which they are produced. From (1.19), (FE_a), and (1.28), an increase in the cutoff translates into a higher average productivity and higher average profits, drives down the number of producers and decreases utility due to the fact that households value variety.

Note, that the increase in the aggregate level of productivity decreases the mass of firms (and hence the mass of available products) only if the average profit in fact reacts to changes in φ^*. If, however,

[17] $\bar{\pi} + wL = E + (wL - wL_P)$.

the average profit remains constant for all possible cutoff levels, then the number of firms is constant and there is no trade-off between product diversity and productivity (we explore this feature further in Chapter 2 below).

Given that a tradeoff between product variety and productivity exists, we are interested to see in equilibrium whether the decrease in utility due to a reduced number of available products is overcompensated by the impact of lower prices due to more productive producers. To answer this question, following Melitz (2003), we calculate the period utility of the representative individual (i.e., aggregate welfare) in the stationary equilibrium. Substituting for the firm index j in the utility function with productivities φ, utility can be rewritten as (cf. Appendix 1.A)[18]

$$U = \left[\int_{j \in J} x(j)^{\alpha} \, dj\right]^{\frac{1}{\alpha}} = \left[\int_{\varphi^*}^{\infty} x(\varphi, P, E)^{\alpha} \, M d\mu(\varphi)\right]^{\frac{1}{\alpha}}.$$

Inserting $x(\varphi, P, E)$ from (1.17) and using $E = L$ from (1.33), U equivalently reads

$$U = \alpha^{\varepsilon} M^{\frac{1}{\alpha}} P^{\varepsilon - 1} L \left[\int_{\varphi^*}^{\infty} \varphi^{\varepsilon\alpha} d\mu(\varphi)\right]^{\frac{1}{\alpha}}.$$

Replacing $P^{\varepsilon-1} = (\alpha\tilde{\varphi})^{1-\varepsilon}/M$ as implied by (1.31) and noting that $\alpha\varepsilon = \varepsilon - 1$, we have

$$U = \alpha M^{\frac{1-\alpha}{\alpha}} L \tilde{\varphi}^{1-\varepsilon} \left[\int_{0}^{\infty} \varphi^{\varepsilon - 1} d\mu(\varphi)\right]^{1/\alpha}.$$

Since the term in squared brackets equals $\tilde{\varphi}^{(\varepsilon-1)/\alpha}$ from the definition of $\tilde{\varphi}$ in (1.27), $1-\varepsilon+(\varepsilon-1)/\alpha = 1$, and $(1-\alpha)/\alpha = 1/(\varepsilon - 1)$, the (per period) welfare in the stationary equilibrium equals

$$U = \alpha M^{1/(\varepsilon-1)} L \tilde{\varphi}(\varphi^*), \tag{1.35}$$

where $\tilde{\varphi}(\varphi^*)$ is given in (1.28). Both an increase in the cutoff/average productivity (evidently $\tilde{\varphi}(\varphi^*)' > 0$, cf. (1.28)) and an increase in the mass of available products raise utility. To arrive at a conclusive result, substitute for M using (1.34):

$$U = \alpha \left[\frac{L^{\varepsilon}}{\varepsilon(\bar{\pi} + f)}\right]^{\frac{1}{\varepsilon-1}} \tilde{\varphi}(\varphi^*).$$

An increase in the cutoff unambiguously raises utility if $\bar{\pi}$ is independent of φ^*. In the general case, we know from (ZCP_a) that $\bar{\pi} + f = (\tilde{\varphi}/\varphi^*)^{\varepsilon-1} f$, whereby

$$U = \alpha \left(\frac{L^{\varepsilon}}{\varepsilon f}\right)^{\frac{1}{\varepsilon-1}} \varphi^*. \tag{1.36}$$

[18] Melitz (2003) measures welfare W as utility per worker in the steady state equilibrium and simply assigns each worker one unit of labor. Hence, $W = U/L$. Cf. the notes on steady state welfare measures and left-out transitional dynamics in the open economy case in Footnote 24 and Footnote 4 in Chapter 2 below.

We thus find that an increase in the cutoff productivity level unambiguously increases aggregate welfare, even if it comes at the cost of a lower mass of available products.

Let us briefly recapitulate the determination and properties of the equilibrium cutoff, mass of firms, average profits, and welfare. We solved the model for an equilibrium where the distribution of productivity level remains stationary. The cutoff productivity is a sufficient statistic for this distribution. To determine the cutoff, we proceeded as follows. The endogenous exit of newcomers with low levels of realized productivity implies a zero cutoff profit condition, which relates the average profits earned in the market to the cutoff productivity level. Free entry implies a second equation in the same variables and imposes that the average profit $\bar{\pi}$ depends only on φ^*. Hence, we can solve for the equilibrium φ^* and $\bar{\pi}$ independently of other endogenous variables. Labor market clearing closes the model. It implies that the aggregate spending on consumption equals the total of wages earned ($E = L$).

In the autarky equilibrium, the cutoff productivity φ^* and hence, from (1.28) and (ZCP_a), the distribution of productivity levels and the average profits are independent of the size of the country as measured by L. The number of producers M is proportional to L and typically decreases with the average productivity of producers (we provide a counter-example in Section 2). Aggregate welfare, however, is unambiguously increasing in the average productivity, which is itself increasing in the cutoff. Welfare is also increasing in L because of the larger mass of available varieties implied by abundant labor resources/consumption expenditures. To summarize, the equilibrium with firm heterogeneity in Melitz (2003) can be expressed in terms of a representative firm and as such resembles the autarky equilibrium in Krugman (1980). Put differently, if the distribution of productivity levels in Melitz (2003) is "degenerate" with all probability mass at one productivity level, the equilibrium outcome boils down to that in Krugman (1980).

1.2.3 Who Pays for Entry?

Baldwin (2005) points out that producers in this environment are "luck rentiers": they earn pure profits for being lucky enough to draw high productivity levels. Of course, these earnings are necessary to allow for investments in new firms to break even on average across all productivity types. In fact, in the stationary equilibrium, the positive profits of incumbent producers exactly cover the fixed entry investment in each period. To see this, recall that the mass of firms that exits equals the mass of entrants in the stationary equilibrium. If we denote by M^e the mass of entrants, the mass of producers equals the sum of surviving firms from "the previous period", $(1 - \delta)\, M$, and the mass of new producers (i.e., assuming a law of large numbers, a fraction $1 - G(\varphi^*)$ of entrants): $M = (1-\delta) M + [1 - G(\varphi^*)] M^e$. Solving for the mass of entrants yields

$$M^e = \frac{\delta M}{1 - G(\varphi^*)}.$$

Now, multiplying (FE_a) by M and using the definition of $\bar{\pi}$ in (1.25), we get

$$f_e \frac{\delta M}{1 - G(\varphi^*)} = M \int_{\varphi^*}^{\infty} \pi(\varphi)\, d\mu(\varphi) = \int_0^M \pi(j)\, dj.$$

The investment in new firms is profitable if firms want to enter and the dividends from the incumbent producers in the stationary equilibrium exactly match the financing needs of the entrants. Accordingly, the aggregate dividend payments are completely used to cover the entry costs of unlucky entrepreneurs (with productivity draws $\varphi < \varphi^*$). However, we assent to Baldwin's (2005) view that the pure profits in equilibrium deserve closer attention.

1.3 Open Economy

Consider a world economy where international trade costlessly increases the product market for domestic firms. Suppose furthermore that trade costs do not alter the elasticity of demand and competition from foreign exporters does not affect the markup charged by domestic firms. Then, all domestic producers export and the average productivity level is the same as in autarky (as an example, see Krugman, 1980). Put differently, the usual "replication argument" according to which the multi-country economy behaves exactly identically to the hypothetical integrated economy that occurs in the absence of national borders applies in the case without trade costs. Crucially, therefore, Melitz (2003) adds to this environment fixed costs of exporting to prevent the least productive firms from exporting – an obstacle strongly supported by empirical evidence. The addition of fixed export costs leads to substantial new insights on the intra-industry reallocation of resources and the productivity effects of trade liberalization typically found in firm level trade data. We study these effects after presenting the open economy model in the next section.

1.3.1 Additional Assumptions

The world now consists of $n + 1$ identical countries of the type described in the autarky section.[19] In particular, each country is inhabited by L individuals. Hence, the wage rate equalizes in all countries.[20] International trade is a simple mutual exchange of consumption goods (j). Two trade frictions hamper the international flow of goods. First, there are variable per-unit iceberg costs, so that $\tau \geq 1$ units must be shipped for one unit to arrive. Second, and crucially for Melitz' (2003) advancements, there are initial fixed costs f_X of exporting, again denominated in units of labor (with exporting, the fixed production costs accrue only in the local market; see Helpman, Melitz, and Yeaple, 2004, for a similar model with FDI). Firms decide whether or not to start exporting and incur the fixed cost of exporting after their productivity level $\varphi(j)$ is realized. This timing structure generates an endogenous selection of only the most productive firms into exporting, once the fixed costs of exporting prevent some firms from exporting (recall that profits are c.p. strictly increasing in the firm's productivity, so that only sufficiently productive firms can afford to enter into the export market). To simplify the exposition, we stick to the symmetric case where trade frictions are uniform across all foreign destinations. For the model to match the stylized fact that only a fraction of all producers exports, we impose the following parameter restriction (for reasons that will become clear below):

$$f < \tau^{\varepsilon-1}\delta f_X. \tag{PA}$$

Intuitively, this condition ensures that it is more costly to sell a good to a market for a foreign company than it is for a local company (even if $\tau = 1$).

1.3.2 Equilibrium

The firms' profit maximization problem is separable in the different destinations of output. In particular, the optimal price for domestically sold units is not affected by the possibility of exporting. We

[19] Melitz (2003) briefly considers the case of asymmetric labor endowments (in which free trade of a homogenous good ensures factor price equalization). He finds that the average productivity, steady state welfare, and wages are higher in large countries. Falvey, Greenaway, and Yu (2006) additionally include asymmetries in production technologies. In the case of costly intra-industry trade between a technologically leading country and a laggard country, they find that reallocation of resources towards the most productive firms is more pronounced in the leading country. If the technological lead is sufficiently strong, or if a country is sufficiently larger than its trading partner, the superior country will run a trade surplus in the differentiated goods sector and induce the inferior country to stop production in this sector.

[20] Note that this does not restrict the size of a country relative to the rest of the world.

only add to (1.16) a subscript d to indicate domestic sales (exports will be labeled with subscript x),

$$p_d(\varphi) = \frac{1}{\alpha\varphi}. \tag{1.37}$$

Accordingly, the revenue and profit functions can be adopted from (1.18) and (1.19). For $\varphi > \varphi^*$,

$$r_d(\varphi, P, E) = (\alpha\varphi P)^{\varepsilon-1} E, \tag{1.38}$$

$$\pi_d(\varphi, P, E) = \frac{r_d}{\varepsilon} - f. \tag{1.39}$$

Turning to the exports, iceberg costs increase the marginal costs for exported units without affecting the elasticity of demand, and fixed exporting costs do not affect the output decision. We can thus simply treat the production of exported units as if it occurred with productivity $\varphi(j)/\tau$ at the firm level. Hence, exporters charge a constant markup over the effective marginal production costs (variable production plus trading costs) for each unit sold abroad, cf. the derivation of (1.16):

$$p_x(\varphi) = \frac{\tau}{\alpha\varphi} = \tau p_d(\varphi). \tag{1.40}$$

Similarly, the resulting equilibrium revenues are readily inferred from (1.38), accounting for the reduced productivity $\varphi(j)/\tau$ for foreign sales:

$$r_x(\varphi, P_i, E_i) = \left(\alpha\frac{\varphi}{\tau}P_i\right)^{\varepsilon-1} E_i = \tau^{1-\varepsilon}(\alpha\varphi P_i)^{\varepsilon-1} E_i = \tau^{1-\varepsilon} r_d(\varphi, P, E). \tag{1.41}$$

The subindex i refers to the target country $1 \leq i \leq n$, but since all countries are inhabited by the same number of people, the aggregate price indices and consumption expenditures are also the same in all destinations ($P_i = P$ and $E_i = E$). Exporting firms earn additional revenues r_x and incur the one time entry cost f_X in each foreign market. Including the periodized amortization payment f_x, the period profit from exporting to a single country amounts to[21]

$$\pi_x(\varphi, P, E) = \frac{r_d(\varphi, P, E)}{\tau^{\varepsilon-1}\varepsilon} - f_x. \tag{1.42}$$

Both exporting costs reduce the profits from foreign sales. In contrast to the Krugman (1980) model, however, the fixed export costs imply that not all domestic firms export. From (1.38) and (1.41), it follows that $\pi_x < 0$ for firms with a sufficiently low productivity (in particular, $\pi_x(0, P, E) < 0$). These firms will not engage in exporting and not sink f_X in the first place. If a firm decides to export,

[21] In the stationary equilibrium with an interest rate equal to zero and no uncertainty other than the death shock, exporters can borrow the upfront entry payment f_X from competitive bankers and pay it back in equal amounts of $f_x \equiv \delta f_X$ per period (so that the lenders make zero profits on average).

however, it exports in all periods throughout its lifetime and, due to symmetric exporting costs across all destinations, to all foreign countries. Accordingly, the total profits of a firm with productivity level φ are

$$\pi(\varphi, P, E) = \max\langle 0, \pi_d(\varphi, P, E) + n \max\{0, \pi_x(\varphi, P, E)\}\rangle. \tag{1.43}$$

In the stationary equilibrium, the value of each firm is again given by

$$v(\varphi, P, E) = \max\left\{0, \frac{\pi(\varphi, P, E)}{\delta}\right\}. \tag{1.44}$$

Just as in autarky, a firm only starts to produce at all if its productivity level exceeds the domestic cutoff productivity φ^*, which is again given by

$$\varphi^* = \inf\{\varphi : \pi_d(\varphi^*) > 0\}. \tag{1.45}$$

Following the same logic, there is now an additional cutoff productivity level for exporting conditional on producing:

$$\varphi_x^* = \inf\{\varphi : \varphi \geq \varphi^* \text{ and } \pi_x(\varphi) > 0\}. \tag{1.46}$$

Exporting naturally requires production (i.e., $\varphi \geq \varphi^*$) and φ_x^* defines the lowest productivity level that allows a firm to export profitably (profits are strictly increasing in productivity and $\pi_x(\varphi_x^*) = 0$). Empirically, φ_x^* is binding for the majority of producers. Bernard et al. (2003) find that in 1992, only 21% of U.S. plants report to export at all (and among exporters, only a third sells more than 10% of its output abroad).[22] The parameter assumption in (PA) accounts for this finding by imposing that $\varphi_x^* > \varphi^*$ holds in equilibrium, so that only a fraction of firms, made up of the most productive firms, exports (and all exporters also sell domestically).[23] Accordingly, there are three types of firms: first, firms with productivity levels below φ^*, which exit immediately upon recognizing their productivity; second, firms with $\varphi^* \leq \varphi < \varphi^*$, which sell only locally; third, the most productive firms with $\varphi \geq \varphi^*$, which sell both locally and export to all countries in the world.

Note that (1.41) allows us to relate the two cutoff productivity levels independently of endogenous variables. From $\pi_x(\varphi_x^*) = 0$, we have

$$\frac{r_d(\varphi_x^*, P, E)}{\varepsilon \tau^{\varepsilon - 1}} = f_x,$$

[22] We are not aware of similar carefully crafted studies with more recent data/other countries. In particular, it would be interesting to see if similar characteristics hold in Germany, the largest exporter in the world, as well (see, e.g., The Economist Intelligence Unit's Country Viewswire, 2008).

[23] We prove this implication of the parameter assumption in (PA) below.

which, from (1.38), implies

$$\left(\alpha\varphi_x^* P\right)^{\varepsilon-1} E = \varepsilon\tau^{\varepsilon-1} f_x. \tag{1.47}$$

Equivalently, $\pi_d(\varphi^*) = 0$ implies

$$\frac{r_d(\varphi^*, P, E)}{\varepsilon} = f,$$

which, again using (1.38), can be expressed as

$$\left(\alpha\varphi^* P\right)^{\varepsilon-1} E = f. \tag{1.48}$$

Dividing (1.47) by (1.48), P and E drop out due to symmetry and we find

$$\varphi_x^* = \tau\left(\frac{f_x}{f}\right)^{\frac{1}{\varepsilon-1}} \varphi^*. \tag{1.49}$$

The threshold productivity for exporting is distorted away from the domestic cutoff by both kinds of trading costs. Under (PA), $\tau f_x^{1/(\varepsilon-1)}/f^{1/(\varepsilon-1)} > 1$. This proves that, in equilibrium, only the most productive firms export (i.e., $\varphi_x^* > \varphi^*$). The decision to produce is based on the same reasoning as in autarky, so the equilibrium distribution of productivity levels for all producers is again given by

$$\mu(\varphi) = G(\varphi) / \left[1 - G(\varphi^*)\right]. \tag{1.50}$$

The domestic cutoff is a "sufficient" statistic for the impact of trade on the equilibrium distribution of productivity levels.

The ex ante probability of producing at the time of entry, i.e. before incurring the entry costs f_e, equals

$$\Lambda(\varphi^*) \equiv 1 - G(\varphi^*). \tag{1.51}$$

The ex ante probability of exporting in turn is $1 - G(\varphi_x^*)$. Conditional on starting production (i.e. conditional on $\varphi \geq \varphi^*$), a producer becomes an exporter with probability $\left[1 - G(\varphi_x^*)\right]/\Lambda(\varphi^*)$. In view of (1.49), $\varphi_x^* = \varphi_x^*(\varphi^*)$ so that the conditional probability of exporting is given by

$$\Lambda_x(\varphi^*) \equiv \frac{1 - G(\varphi_x^*(\varphi^*))}{1 - G(\varphi^*)} \quad < 1. \tag{1.52}$$

The inequality follows directly from the assumption that only a fraction of firms exports ($\varphi_x^* \geq \varphi^*$ implies $G(\varphi_x^*) > G(\varphi^*)$, cf. (PA) and (1.49)). To solve for the equilibrium level of the domestic cutoff φ^*, we follow the same steps as before and derive the open economy analogue to the free entry and zero cutoff profit conditions. Starting with free entry, the expected operating profit must again be equal to the entry cost f_e. By definition of v and π, both variables are zero for productivity levels that imply losses so the free entry condition remains unchanged:

$$\int_0^\infty v(\varphi)\, dG(\varphi) = f_e.$$

Ex ante, firms anticipate that they either exit immediately (if $\varphi < \varphi^*$), or earn positive profits from domestic sales ($\varphi^* \leq \varphi < \varphi_x^*$), or from both domestic sales and exports ($\varphi_x^* \leq \varphi$). Hence, (1.43) and (1.44) allow us to rewrite the free entry condition as

$$\int_{\varphi^*}^\infty \frac{\pi_d(\varphi, P, E)}{\delta} dG(\varphi) + n \int_{\varphi^*}^\infty \frac{\pi_x(\varphi, P, E)}{\delta} dG(\varphi) = f_e.$$

After multiplying by δ and inserting π_d from (1.39) and π_x from (1.42), we find

$$\int_{\varphi^*}^\infty \left[\frac{r_d}{\varepsilon} - f\right] dG(\varphi) + n \int_{\varphi^*}^\infty \left[\frac{r_d(\varphi; P, E)}{\tau^{\varepsilon-1}\varepsilon} - f_x\right] dG(\varphi) = \delta f_e.$$

Dividing by $1 - G(\varphi^*)$, inserting $\varphi_x^* = \varphi_x^*(\varphi^*)$ from (1.49), and recognizing $dG(\varphi) = [1 - G(\varphi^*)]\, d\mu(\varphi)$ from (1.50),

$$\int_{\varphi^*}^\infty \left[\frac{r_d}{\varepsilon} - f\right] d\mu(\varphi) + n \int_{\varphi_x^*(\varphi^*)}^\infty \left[\frac{r_d(\varphi; P, E)}{\tau^{\varepsilon-1}\varepsilon} - f_x\right] d\mu(\varphi) = \frac{\delta f_e}{1 - G(\varphi^*)} \tag{1.53}$$

where $\varphi_x^*(\varphi^*)$ is explicitly given in (1.49). Analogously to the closed economy, the left hand side is the average/expected profit earned by a producer, which now includes profits from domestic sales and also, potentially, additional exporting profits. We can again use the free entry condition to define an average productivity of producers and label the associated profits, i.e. the left hand side of (1.53), $\bar{\pi}$. With this definition, the free entry condition is formally identical to the FE condition in autarky:

$$\bar{\pi} = \frac{\delta f_e}{1 - G(\varphi^*)}. \tag{FE}$$

We continue to use the free entry condition to derive the average productivity of domestic producers, which then can be used to state the zero cutoff profit condition in the open economy in terms of $\bar{\pi}$ and φ^*. The average productivity is again a function of the local market cutoff productivity only. After drawing f and f_x out of the integrals in (1.53), using (1.51), and rearranging we have

$$\int_{\varphi^*}^\infty \frac{r_d}{\varepsilon} d\mu(\varphi) + n\left[\tau^{1-\varepsilon} \int_{\varphi_x^*(\varphi^*)}^\infty \frac{r_d(\varphi; P, E)}{\varepsilon} d\mu(\varphi)\right] = \frac{\delta f_e}{\Lambda(\varphi^*)} + f + n f_x \Lambda_x(\varphi^*). \tag{1.54}$$

Finally, using the specific form for revenues from (1.38) and noting (FE),

$$\frac{(\alpha P)^{\varepsilon-1} E}{\varepsilon} \int_{\varphi^*}^\infty \varphi^{\varepsilon-1} d\mu(\varphi) + \Lambda_x(\varphi^*)\, n \left[\tau^{1-\varepsilon} \frac{(\alpha P)^{\varepsilon-1} E}{\varepsilon} \int_{\varphi_x^*(\varphi^*)}^\infty \frac{\varphi^{\varepsilon-1}}{\Lambda_x(\varphi^*)} d\mu(\varphi)\right] = \bar{\pi} + f + n f_x \Lambda_x(\varphi^*).$$

In view of (1.38), we can define the average productivity of all domestic producers in the home market,

$$\tilde{\varphi}\left(\varphi^{*}\right) \equiv \left[\int_{\varphi^{*}}^{\infty} \varphi^{\varepsilon-1} d\mu\left(\varphi\right)\right]^{\frac{1}{\varepsilon-1}} = \left[\frac{1}{1-G\left(\varphi^{*}\right)} \int_{\varphi^{*}}^{\infty} \varphi^{\varepsilon-1} dG\left(\varphi\right)\right]^{\frac{1}{\varepsilon-1}}, \tag{1.55}$$

as well as the average productivity of exporting producers excluding the variable trading costs,

$$\tilde{\varphi}_{x}\left(\varphi^{*}\right) \equiv \left[\int_{\varphi_{x}^{*}\left(\varphi^{*}\right)}^{\infty} \frac{\varphi^{\varepsilon-1}}{\Lambda_{x}\left(\varphi^{*}\right)} d\mu\left(\varphi\right)\right]^{\frac{1}{\varepsilon-1}} = \left[\frac{1}{1-G\left(\varphi_{x}^{*}\right)} \int_{\varphi_{x}^{*}\left(\varphi^{*}\right)}^{\infty} \varphi^{\varepsilon-1} dG\left(\varphi\right)\right]^{\frac{1}{\varepsilon-1}}. \tag{1.56}$$

These definitions allow us to state the average profit of a producer as

$$\bar{\pi} = \pi_{d}\left(\tilde{\varphi}, P, E\right) + \Lambda_{x}\left(\varphi^{*}\right) n \pi_{x}\left(\tilde{\varphi}_{x}, P, E\right). \tag{1.57}$$

This formulation makes explicit that $\tilde{\varphi}$ is the average productivity of domestic producers as measured only in the domestic market. Evidently, the average productivity of domestic firms is even higher if take account of the fact that only the most productive firms export and thereby gain further market shares relative to the less productive producers who only sell locally. Observe also that the trading costs τ, which are modeled as iceberg costs that do not generate any income, are not accounted for in $\tilde{\varphi}_{x}$. Accordingly, we assess aggregate productivity with $\tilde{\varphi}$ and $\tilde{\varphi}_{x}$.

Returning to the derivation of the ZCP, the next step is to relate $\pi_{d}\left(\tilde{\varphi}, P, E\right)$ and $\pi_{x}\left(\tilde{\varphi}_{x}, P, E\right)$ to φ^{*}. This is easily achieved in analogy to the autarky case and separately for the domestic and the foreign market as follows. Using (1.39), $\pi_{d}\left(\tilde{\varphi}, P, E\right)$ explicitly reads

$$\pi_{d}\left(\tilde{\varphi}_{d}, P, E\right) = \frac{r_{d}\left(\tilde{\varphi}_{d}\right)}{\varepsilon} - f = \frac{\left(\alpha \tilde{\varphi}_{d} P\right)^{\varepsilon-1} E}{\varepsilon} - f.$$

Expressing the relative domestic revenues of producers with productivity levels $\tilde{\varphi}$ and φ^{*} in terms of their profits, respectively, yields

$$\frac{r_{d}\left(\tilde{\varphi}_{d}\right)}{r_{d}\left(\varphi^{*}\right)} = \frac{\pi_{d}\left(\tilde{\varphi}_{d}\right) + f}{\pi_{d}\left(\varphi^{*}\right) + f}.$$

Hence, as $\pi_{d}\left(\varphi^{*}\right) = 0$,

$$\pi_{d}\left(\tilde{\varphi}_{d}\right) = f\left[\left(\frac{\tilde{\varphi}_{d}\left(\varphi^{*}\right)}{\varphi^{*}}\right)^{\varepsilon-1} - 1\right]. \tag{1.58}$$

Similarly, from (1.42) and (1.49), we have

$$\pi_{x}\left(\tilde{\varphi}_{x}, P, E\right) = \frac{r_{d}\left(\tilde{\varphi}_{x}; P, E\right)}{\tau^{\varepsilon-1}\varepsilon} - f_{x} = \tau^{1-\varepsilon}\left[\frac{r_{d}\left(\tilde{\varphi}_{x}; P, E\right)}{\varepsilon} - f\right] + \tau^{1-\varepsilon} f - f_{x}.$$

The relative revenues from exporting for firms with productivity levels $\tilde{\varphi}_{x}$ and φ_{x}^{*}, respectively, thus can be expressed as

$$\frac{r_{x}\left(\tilde{\varphi}_{x}\right)}{r_{x}\left(\varphi_{x}^{*}\right)} = \frac{\pi_{x}\left(\tilde{\varphi}_{x}\right) + f_{x}}{\pi\left(\varphi_{x}^{*}\right) + f_{x}},$$

where, using $\pi\left(\varphi_x^*, P, E\right)=0$ and (1.41),

$$\pi_x\left(\tilde{\varphi}_x, P, E\right)=\left[\left(\frac{r_x\left(\tilde{\varphi}_x\right)}{r_x\left(\varphi_x^*\right)}\right)-1\right] f_x=\left[\left(\frac{\tilde{\varphi}_x\left(\varphi^*\right)}{\varphi_x^*}\right)^{\varepsilon-1}-1\right] f_x . \tag{1.59}$$

Finally, inserting (1.58) and (1.59) in (1.57), we arrive at the zero cutoff profit condition in the open economy:

$$\bar{\pi}=f\left[\left(\frac{\tilde{\varphi}_d\left(\varphi^*\right)}{\varphi^*}\right)^{\varepsilon-1}-1\right]+\Lambda_x\left(\varphi^*\right) n f_x\left[\left(\frac{\tilde{\varphi}_x\left(\varphi^*\right)}{\varphi_x^*}\right)^{\varepsilon-1}-1\right] . \tag{ZCP}$$

Recall that the closed economy ZCP is given by a similar expression without the second summand ($\tilde{\varphi}_d$ with trade corresponds to $\tilde{\varphi}$ in autarky). With (ZCP) and (FE), we again have two equations in two unknowns which can be solved for the average profits $\bar{\pi}$ and the local cutoff productivity level φ^*. Inserting the domestic cutoff in (1.49) then directly gives the cutoff productivity level for profitable exporting. In Appendix (1.B), we present Melitz' (2003) proof for (ZCP) to cut the FE curve once from above, so that a unique equilibrium exists in the open economy as well.

We are left to determine the mass of local producers, M, and the mass of exporters, M_x. The assumption of identical countries simplifies this solution. In particular, balanced trade requires that the spending on imported varieties of local consumers equals the (gross i.e. before accounting for the iceberg costs) revenues of domestic exporters in the foreign countries, so that the revenues of all domestic firms equal the consumption expenditures of local consumers on all available products. Denoting the average revenues of domestic producers from local and export sales E/M by $\bar{r}$ (expenditures on imported goods equal the aggregate revenues of domestic producers from exports), and using $E=L$ from (1.33), which is equally valid in the open economy, this gives

$$M \bar{r}=E=L . \tag{1.60}$$

We can thereby directly infer the average revenue

$$\bar{r}=\int_0^{\infty} r(\varphi) d \mu(\varphi)=\int_{\varphi^*}^{\infty} r_d(\varphi) d \mu(\varphi)+\int_{\varphi_x^*\left(\varphi^*\right)}^{\infty} r_x(\varphi) d \mu(\varphi)$$

from the average profits in (1.54) using (FE):

$$\bar{r}=\varepsilon\left[\bar{\pi}+f+n \Lambda_x\left(\varphi^*\right) f_x\right] . \tag{1.61}$$

After substituting for $\bar{r}$ in (1.60), the equilibrium number of local producers is

$$M=\frac{L}{\varepsilon\left[\bar{\pi}+f+n \Lambda_x\left(\varphi^*\right) f_x\right]} . \tag{1.62}$$

Resorting to a law of large numbers, the mass of exporters is $M_x = \Lambda_x(\varphi^*) M$ $(< M)$. Since the equilibrium distributions of productivity levels are identical throughout the world, the mass of imported products is $nM_x = n\Lambda_x(\varphi^*) M$. Hence, the mass of available varieties in the domestic market is $M_t \equiv M + nM_x = [1 + n\Lambda_x(\varphi^*)] M$. Alternatively, using $r_d(\tilde{\varphi}_t, P, E) = L/M_t$ (M_t firms with average productivity $\tilde{\varphi}_t$ receive a total spending of $E = L$ and revenues can be described by (1.38)) and $r_d(\varphi^*, P, E) = \varepsilon f$ together with (1.38) for $\varphi = \varphi^*$ and $\varphi = \tilde{\varphi}_t$, we have

$$\frac{L}{\varepsilon f M_t} = \frac{r_{d(\tilde{\varphi}_t,P,E)}}{r_{d(\varphi^*,P,E)}} = \left(\frac{\tilde{\varphi}_t}{\varphi^*}\right)^{\varepsilon-1},$$

and hence

$$M_t = \left(\frac{\varphi^*}{\tilde{\varphi}_t}\right)^{\varepsilon-1} \frac{L}{\varepsilon f}. \tag{1.63}$$

We can also solve for the price index which now additionally includes the prices of imported products:

$$P^{1-\varepsilon} \equiv \int_{j\in J} p(j)^{1-\varepsilon} dj = \int_0^M p_d(j)^{1-\varepsilon} dj + \int_0^{nM_x} p_x(j)^{1-\varepsilon} dj,$$

where we use in the second summand that the firm index is a number and that firms can always be rearranged along the real line so that j in $[0, nM_x]$ indicates the imported products from the n countries. Analogously to the derivation of (1.31), the $j's$ can be substituted against $\varphi's$ using integration by substitution (cf. Appendix 1.A):

$$P^{1-\varepsilon} = \int_0^M p_d(j)^{1-\varepsilon} dj + \int_0^{nM_x} p_x(j)^{1-\varepsilon} dj = \int_{\varphi^*}^{\infty} p_d(\varphi)^{1-\varepsilon} M\mu'(\varphi)\, d\varphi + \int_{\varphi_x^*}^{\infty} p_x(\varphi)^{1-\varepsilon} nM_x\mu'(\varphi)\, d\varphi.$$

As for the last term, we used the fact that the domestic markets' cut-off productivity level is the same in all countries, so that μ is the distribution of productivity levels in each country. Inserting the equilibrium pricing rules from (1.37) and (1.40), we obtain

$$\begin{aligned} P^{1-\varepsilon} &= \int_{\varphi^*}^{\infty} (\alpha\varphi)^{\varepsilon-1} M d\mu(\varphi) + \int_{\varphi_x^*}^{\infty} (\alpha\varphi\tau)^{\varepsilon-1} nM_x d\mu(\varphi) \\ &= \alpha^{\varepsilon-1}\left[M \int_{\varphi^*}^{\infty} \varphi^{\varepsilon-1} d\mu(\varphi) + \tau^{\varepsilon-1} nM_x \int_{\varphi_x^*}^{\infty} \varphi^{\varepsilon-1} d\mu(\varphi)\right]. \end{aligned}$$

Using the definitions of $\tilde{\varphi}_d$ and $\tilde{\varphi}_x$, (1.55) and (1.56), the price index further becomes

$$P^{1-\varepsilon} = \alpha^{\varepsilon-1}\left[M \int_{\varphi^*}^{\infty} \varphi^{\varepsilon-1} d\mu(\varphi) + \tau^{\varepsilon-1} nM_x \int_{\varphi_x^*}^{\infty} \varphi^{\varepsilon-1} d\mu(\varphi)\right].$$

If we finally define the average productivity level of all producers competing in a country as

$$\tilde{\varphi}_t(\varphi^*) \equiv \left\{\frac{1}{M_t}\left[M \int_{\varphi^*}^{\infty} \varphi^{\varepsilon-1} d\mu(\varphi) + \tau^{1-\varepsilon} nM_x \int_{\varphi_x^*(\varphi^*)}^{\infty} \varphi^{\varepsilon-1} d\mu(\varphi)\right]\right\}^{\frac{1}{\varepsilon-1}}, \tag{1.64}$$

the price index boils down to

$$P = \frac{1}{\alpha \tilde{\varphi}_t M_t^{\frac{\epsilon}{\epsilon-1}}}, \tag{1.65}$$

which is analogous to the autarky case. In contrast to the definitions of $\tilde{\varphi}$ or $\tilde{\varphi}_x$, that refer to the actual productivity of firms, $\tilde{\varphi}_t$ also takes into account that productive exporters gain additional total market shares due to the fact that some, less productive producers only sell locally and the decline in aggregate output due to the ”melting” of shipped units.

Finally, consider aggregate welfare in the stationary equilibrium.[24] Households derive utility from consuming domestically produced and imported products. Substituting for firms j with productivity levels φ (cf. Appendix 1.A) and recognizing (1.49) gives

$$U = \left[\int_{j\in J} x(j)^{\alpha}\, dj\right]^{\frac{1}{\alpha}} = \left[M \int_{\varphi^*}^{\infty} x_d(\varphi, P, E)^{\alpha}\, d\mu(\varphi) + nM_x \int_{\varphi_x^*(\varphi^*)}^{\infty} x_x(\varphi, P, E)^{\alpha}\, d\mu(\varphi)\right]^{\frac{1}{\alpha}}.$$

Inserting $x_d(\varphi, P, E) = r_d(\varphi, P, E)/p_d(\varphi) = (\alpha\varphi)^{\varepsilon} P^{\varepsilon-1}E$, $x_x(\varphi, P, E) = r_x(\varphi, P, E)/p_x(\varphi) = (\alpha\varphi)^{\varepsilon} \tau^{-\varepsilon} P^{\varepsilon-1}E$, and $E = L$ from (1.37), (1.38), (1.40), (1.41), and (1.33), U can be rewritten as (note that $\varepsilon\alpha = \varepsilon - 1$)

$$U = \alpha^{\varepsilon} P^{\varepsilon-1} L \left[M \int_{\varphi^*}^{\infty} \varphi^{\varepsilon-1} d\mu(\varphi) + \tau^{1-\varepsilon} nM_x \int_{\varphi_x^*(\varphi^*)}^{\infty} \varphi^{\varepsilon-1} d\mu(\varphi)\, d\mu(\varphi)\right]^{\frac{1}{\alpha}}.$$

Now, inserting $P = (\alpha\tilde{\varphi}_t)^{1-\varepsilon}/M_t$ and replacing the expression in squared brackets with $\tilde{\varphi}_t^{\varepsilon-1} M_t$, see (1.64), we find

$$U = \alpha L M_t^{1/(\varepsilon-1)} \tilde{\varphi}_t(\varphi^*), \tag{1.66}$$

similar as in autarky, but with the average productivity and the number of firms referring to the total of domestic products and imports. To derive aggregate welfare as a function of the domestic cutoff productivity φ^*, substitute for $M_t^{1/(\varepsilon-1)}$ in U with M_t from (1.63):

$$U = \alpha \left(\frac{L^{\varepsilon}}{\varepsilon f}\right)^{\frac{1}{\varepsilon-1}} \varphi^*. \tag{1.67}$$

Note that this expression is identical to (1.36) under autarky. The impact of trade on aggregate welfare thus becomes apparent from the impact of trade on the domestic cutoff.

[24] Melitz (2003) assesses the impact of trade on aggregate welfare by a comparison of welfare levels in the steady state equilibria of different trade regimes, thus ignoring welfare along the transitional path. Alessandria and Choi (2007) take the transitional dynamics into account and find that, if calibrated to U.S. data, the stationary comparison is ill-suited for the purpose of measuring welfare gains of trade policies, cf. footnote 4 in Chapter 2 below.

1.3.3 The Impact of Trade

In this model, producers self-select into the domestic and the export market according to their level of productivity by the assumption on exporting costs in (PA). The main message from Melitz' (2003) model is thus *not* the partitioning of producers into exporters and non-exporters. Its explanatory power rather lies in the ability to generate resource reallocations along two margins. First, opening up to trade allows some firms to start exporting and forces the least productive firms to exit the market. Second, international trade reallocates resources among surviving producers and shifts market shares from less productive domestic firms and exporters to the most productive exporters.

To assess these effects, we compare the open economy equilibrium to the equilibrium in autarky and first analyze how the opening up to trade affects the domestic productivity cutoff, the average producers' profits, the mass of available products, and aggregate welfare. Subsequently, we study how the exposure to trade has the potential to reallocate resources within industries.

The market share shifting and exit of the least productive producers equivalently occurs if open economies decrease barriers to trade further (see Melitz, 2003, Appendix E for a proof). Instead of studying the effects of a larger number of trading partners and smooth declines in the iceberg and export market entry costs in the general model, we consider, as an example, a specified distribution of productivity levels in Chapter 2. This allows us to derive a closed form solution and provide an easier understanding of the mechanisms at play.

The Firm Distribution, Aggregate Productivity, and Welfare

To begin with, we compare the domestic cutoff productivity level under autarky to the respective figure in the open economy. Under both regimes, φ^* is determined together with $\bar{\pi}$ independently of other endogenous variables by the FE and the ZCP condition. We re-state the open economy variants of both conditions for convenience:

$$\bar{\pi} = \frac{\delta f_e}{1 - G\left(\varphi^*_{(a)}\right)} \tag{FE}$$

$$\bar{\pi} = f\left\{\left[\frac{\tilde{\varphi}\left(\varphi^*_{(a)}\right)}{\varphi^*_{(a)}}\right]^{\varepsilon-1} - 1\right\} + \Lambda_x\left(\varphi^*\right) n f_x \left\{\left[\frac{\tilde{\varphi}_x\left(\varphi^*\right)}{\varphi^*_x\left(\varphi^*\right)}\right]^{\varepsilon-1} - 1\right\} \tag{ZCP}$$

The FE curve is identical under both regimes. Relative to the ZCP curve in autarky, the ZCP curve in the open economy has an additional summand (which is positive whenever there is some trade so that

$\Lambda_x(\varphi) > 0$ since $\tilde{\varphi}_x(\varphi^*) \geq \varphi_x^*(\varphi^*)$). That is, the ZCP curve in the open economy is the ZCP curve in autarky shifted upwards. It follows that the cutoff productivity level φ^* in the domestic market is increasing. Evidently, the increasing minimum productivity raises the average productivity based on domestic market shares $\tilde{\varphi}$, cf. (1.55). This productivity increase is accompanied by an increase in the average profit from domestic sales and added export sales, $\bar{\pi}$. That is, while the possibility for international trade allows firms with productivity levels $\varphi \geq \varphi_x^*$ to serve the export markets, at the same time, it forces the least productive firms to cease production.

The opening to trade thereby leads to a decline in the mass of domestic producers: as $\bar{\pi}$ increases and $n\Lambda_x f_x > 0$, (1.62) implies that M is unambiguously lower than under autarky. Typically, however, the total mass of available products increases, since the mass of different imported products exceeds the mass of firms that exit due to the exposure to trade. Interestingly, however, M_t $(= (1 + n\Lambda_x f_x) M)$ may also be smaller than the mass of available products under autarky ($n\Lambda_x f_x > 0$ but M decreases). In this case, we infer from (1.35) and (1.66) that $\varphi^*/\tilde{\varphi}$ in autarky is strictly greater than $\varphi^*/\tilde{\varphi}_t$ in the open economy with international trade. Since the domestic cutoff unambiguously increases if the economy opens up to international trade, it has to be that $\tilde{\varphi}_t$ strictly exceeds $\tilde{\varphi}$ in autarky (this property need not hold in general since $\tilde{\varphi}_t$ includes the decline in output due to iceberg costs).[25] No matter whether both the mass of available products and aggregate productivity increase or only aggregate productivity increases, trade liberalization unambiguously increases aggregate welfare since the gain in aggregate productivity outweighs the loss of variety. Trade exposure increases φ^* which, from (1.67), unambiguously increases welfare.

Intra-Industry Resource Reallocation

To begin with, opening up to international trade allows sufficiently productive firms to start exporting. At the same time, the least productive firms are driven out of the market. This observation suggests that trade liberalization benefits the most productive and harms the least productive producers. In fact, the surviving, non-exporting firms become smaller and earn lower revenues than in autarky. To see this, denote autarky variables by a subscript ’a’ and consider (1.29) noting that $r(\varphi^*, P, L) = \varepsilon f$

[25] In general, $\tilde{\varphi}_t$ must thus not exceed the average productivity as defined by $\tilde{\varphi}$ under autarky. Melitz (2003, Appendix D.3), however, shows that any revenue weighted average productivity, measured “at the factory gate” (Melitz, 2003, p. 1721), i.e. before the iceberg costs occur, must unambiguously increase with trade openness.

both in the open economy and in autarky. This gives

$$r_a(\varphi, P_a, L) = \left(\frac{\varphi}{\varphi_a^*}\right)^{\varepsilon-1} \varepsilon f, \tag{1.68}$$

$$r_d(\varphi, P, L) = \left(\frac{\varphi}{\varphi^*}\right)^{\varepsilon-1} \varepsilon f, \tag{1.69}$$

so that $r_a(\varphi, P_a, L) > r_d(\varphi, P, L)$ since $\varphi^* > \varphi_a^*$. Melitz (2003, Appendices D.2 and E.2), however, shows that producers who export earn higher combined revenues from domestic sales and exports than in autarky, $r_a(\varphi, P_a, L) < r_d(\varphi, P, L) + nr_x(\varphi, P, L)$. Including the export market entry costs, it can be shown that only the most productive exporters gain from trade liberalization, i.e. that there is a threshold productivity level that separates the most productive exporters who gain from trade from the remainder of producers.[26] Evidently, the exposure to trade thus raises the inequality in profit levels across all producers.

The model thus accounts for two major empirical facts. First, it is consistent with the findings by Bernard and Jensen (1999), Aw, Chung, and Roberts (2000), and Clerides, Lach, and Tybout (1998), that opening up to trade drives the least productive firms out of the market. Second, as surviving non-exporters shrink while exporters grow, the equilibrium characteristics are in line with the evidence uncovered by Pavcnik (2002) and Bernard, Jensen, and Schott (2006), namely that trade induces a large scale reallocation of resources from the less productive, surviving firms to the most productive, exporting firms.

The bottom line of the preceding analysis is that trade liberalization unambiguously increases aggregate productivity (which drives the increase in welfare). We now take a closer look at the mechanism behind the productivity gain.

Deciphering the Productivity Effects from Wages and Competition

In principle, two forces can account for the exit of the least productive firms. First, following trade liberalization, domestic producers compete with an increased number of imports produced by the most productive foreign manufacturers. Second, relative to autarky, the most productive firms aim at earning additional profits from exports and thus increase their labor demand to conduct export

[26]The proof for the last two statements in the general case with an unspecified utility function involves somewhat more cumbersome comparative statics of the combined revenues of exporters with respect to τ, so that we do not reproduce them and refer to Melitz (2003, Appendices D.2 and E.2). Instead, we demonstrate these claims in the case of Pareto-distributed productivities below (cf. Chapter 2).

entry and expanded production. In addition, the increase in the average profits induces more firms to enter, again increasing the aggregate demand for labor for the conduct of entry and the subsequent production. Since the total stock of labor is fixed, the corresponding increase in the real wage lowers profits and thereby forces the least productive firms to exit. Melitz (2003) asserts that the monopolistic competition model with CES preferences restrictively shuts down the “increasing competition” channel, since the price elasticity of demand is independent of the number of competitors and their productivity. Note, however, that the “increasing competition” effect works through two forces. On the one hand, the average productivity of competitors increases because only the most productive firms export. This increase in competition indeed does not affect the profits of incumbent producers due to the CES assumption and monopolistic competition. On the other hand, however, the mere fact that imports increase the number of different products available in the domestic market drives down the (Dixit-Stiglitz) market share of each domestic incumbent. The constant elasticity of substitution implies that the decline in revenues is proportional across all varieties, but the increased horizontal competition also has the potential to force the least productive firms to exit. Solving for the transitional dynamics following trade liberalization, Chaney (2005) proves that immediately after trade liberalization occurs, the least productive domestic producers exit because of the market entry of foreign firms. Subsequently, resources reallocate towards the most productive firms and the aggregate productivity gain is due to both an increase in the real wage *and* horizontal competition.

Accounting for the Extensive Margin of Trade Liberalization

Note that the addition of fixed costs for exporting together with heterogeneous firms substantially extends earlier findings on the effects of trade liberalization as e.g. obtained from Krugman’s (1980) model. In Melitz (2003), there is an endogenous self-selection of firms into the export markets implies an extensive margin of trade. That is, while a lowering in (variable) trading costs in Krugman (1980) simply induces all firms to export more, lowering both fixed and variable export costs not only increase the trading quantity of all previously active exporters, but also induces some firms to export which did not export before trade liberalization occurred. In Krugman (1980) all firms export to begin with and trade liberalization only has an intensive margin. Hence there is no possibility for some firms to lose from trade.

1.3.4 Measured Productivity

In this model, firm-level productivity is defined as the marginal productivity of labor in a firm, φ. When looking at the data, however, productivity at the firm level is often measured as output per worker, i.e. by the average productivity of labor in each firm. In the model, such a measure is ill suited if it excludes fixed costs. In fact, if we exclude fixed costs and output is measured by revenues, "output per worker" is the same for all firms: from (1.16), we have

$$\frac{r(j)}{x(j)/\varphi(j)} = \varphi(j)\, p(j) = \frac{1}{\alpha},$$

where we substituted for j with φ in the output, price, and revenue functions to define $x(j)$, $p(j)$, and $r(j)$ (we do the same for the domestic and export variants below without further remark). Consider, however, a similar productivity measure including the fixed overhead costs (using (1.17) and (1.18)):

$$\mathcal{P}(j) \equiv \frac{r(j)}{\frac{x(j)}{\varphi(j)} + f} = \frac{[\alpha\varphi(j)\, P]^{\varepsilon-1}}{\alpha^{\varepsilon}[\varphi(j)\, P]^{\varepsilon-1} + f} = \frac{1}{\alpha + \frac{f}{[\alpha\varphi(j)P]^{\varepsilon-1}}}.$$

$\mathcal{P}(j)$ is c.p. strictly increasing in $\varphi(j)$. Analogously, a theoretical counterpart of the "measured productivity" of an exporting firm and a non-exporting firm are

$$\mathcal{P}^{x}(j) \equiv \frac{r_d(j) + n r_x(j)}{\frac{x_d(j)}{\varphi(j)} + n\frac{\tau x_x(j)}{\varphi(j)} + f + n f_x} \text{ and } \mathcal{P}^{d}(j) \equiv \frac{r_d(j)}{\frac{x_d(j)}{\varphi(j)} + f},$$

respectively. $x_x(\varphi)$ in the denominator of $\mathcal{P}^x$ is thereby the quantity sold in the export market for which τx_x units must be shipped (and hence produced) for x_x units to arrive. $\mathcal{P}^x$ can be simplified by substituting for $r_x = \tau^{1-\varepsilon} r_d$ from (1.41) and $x_x = r_x/p_x = \tau^{1-\varepsilon} r_d/(\tau p_d) = \tau^{-\varepsilon} x_d$ from (1.40) and (1.41):

$$\mathcal{P}^{x} \equiv \frac{\left(1+n\tau^{1-\varepsilon}\right) r_d(j)}{\left(1+n\tau^{1-\varepsilon}\right)\frac{x_d(j)}{\varphi(j)} + f + n f_x} = \mathcal{P}^{d}\frac{\left(1+n\tau^{1-\varepsilon}\right)\left[\frac{x_d(j)}{\varphi(j)} + f\right]}{\left(1+n\tau^{1-\varepsilon}\right)\frac{x_d(j)}{\varphi(j)} + f + n f_x} = \mathcal{P}_d \frac{\frac{x_d(j)}{\varphi(j)} + f}{\frac{x_{d(j)}}{\varphi(j)} + \frac{f + n f_x}{1+n\tau^{1-\varepsilon}}}.$$

Hence, $\mathcal{P}^x > \mathcal{P}^d$ if and only if

$$f > \frac{f + n f_x}{1 + n\tau^{1-\varepsilon}} \quad \Leftrightarrow \quad f > \tau^{\varepsilon-1} f_x = \tau^{\varepsilon-1}\delta f_X,$$

i.e. under the previous parameter assumption (PA). Hence, the model is in fact in line with the evidence on heterogeneity in measured productivity if fixed costs are included, with exporters outperforming the non-exporters.

As an aside, note that the logic of the Melitz model imposes a clear direction of causality for the positive correlation between the measured productivity and a firm's export status: firms export because they are more productive than their non-exporting competitors, and not because they anticipate a technological improvement by exporting.[27]

1.4 A Remark on the ZCP Condition

Melitz (2003) proves that the ZCP curve cuts the FE curve exactly once (viz., from above). The slope of the ZCP curve is thus not essential for the result that opening up to trade, i.e. an upward shift of the ZCP curve, increases the domestic cutoff. Given that the ZCP pins down the average profit of producers, however, we should be aware of the fact that the slope of the ZCP curve is determined by the assumptions on $G(\varphi)$, which exogenously fix $\tilde{\varphi}(\varphi^*)/\varphi^*$. In general, $\bar{\pi}$ can be upward or downward sloping in $(\bar{\pi}, \varphi^*)$-space depending on whether $\tilde{\varphi}(\varphi^*)/\varphi^*$ is in- or decreasing in φ^*. Melitz states that a sufficient condition for the ZCP to decrease monotonically from infinity to zero is that $\Gamma(\varphi)$ is increasing to infinity where

$$\Gamma(\varphi) \equiv \frac{\varphi g(\varphi)}{1 - G(\varphi)}$$

To see this, note first that $\lim_{\varphi^* \to 0} \bar{\pi}(\varphi^*) = \infty$ (if the ZCP curve depends on φ^*, i.e. $\tilde{\varphi}(\varphi^*)$ is not linear in φ^*). The ZCP is decreasing to zero in the limit as $\varphi^* \to \infty$ if

$$\omega(\varphi) \equiv \left[\frac{\tilde{\varphi}(\varphi)}{\varphi}\right]^{\varepsilon - 1} - 1$$

is strictly monotonically decreasing in φ since $\omega(\varphi) > 0$ ($\tilde{\varphi}(\varphi) > \varphi$ follows from the definition in (1.28) for any non-degenerated distribution). In Appendix 1.B, Equation (A.2) we derive

$$\omega'(\varphi) = \frac{g(\varphi)}{1 - G(\varphi)}\omega(\varphi) - \frac{(\varepsilon - 1)[\omega(\varphi) + 1]}{\varphi}.$$

Hence, $\omega'(\varphi) < 0$ if and only if

$$\Gamma(\varphi) < (\varepsilon - 1)\left[1 + \frac{1}{\omega(\varphi)}\right],$$

which holds if $\Gamma(\varphi)$ is increasing to infinity. This property holds for most common distributions (restricted to a positive support). Note, however, that $\tilde{\varphi}(\varphi^*)/\varphi^*$ may also be independent of the cutoff productivity, namely if $\tilde{\varphi}(\varphi^*)$ is linear in φ^* (a case not considered in Melitz, 2003). The ZCP

[27] Also, “productivity” in the model is the only determinant of a firm’s profit and hence for the decision to export.

is horizontal in $(\bar{\pi}, \varphi^*)$-space if $\tilde{\varphi}(\varphi^*)$ is linear in φ^* so that $\tilde{\varphi}(\varphi^*)/\varphi^*$ is independent of φ^*.[28] In this case, the average profit is independent of the cutoff. To grasp this feature, consider again the ZCP condition (repeated here for convenience):

$$\bar{\pi} = \left\{ \left[\frac{\tilde{\varphi}(\varphi^*)}{\varphi^*} \right]^{\varepsilon-1} - 1 \right\} f.$$

$\bar{\pi}$ is increasing in φ^* if $\varphi^* \tilde{\varphi}'(\varphi^*) - \tilde{\varphi}(\varphi^*) > 0$ or, equivalently,

$$\frac{\partial \tilde{\varphi}(\varphi^*)}{\partial \varphi^*} \frac{\varphi^*}{\tilde{\varphi}(\varphi^*)} \equiv \varepsilon_{\tilde{\varphi},\varphi^*} > 1.$$

Similarly, an increase in the cutoff implies a decrease in $\bar{\pi}$ if $\varepsilon_{\tilde{\varphi},\varphi^*} < 1$. If $\tilde{\varphi}(\varphi^*)$ is linear in φ^*, $\varepsilon_{\tilde{\varphi},\varphi^*} = 1$ and $\bar{\pi} = const. \times f$. We can readily assess the economic effects behind $\varepsilon_{\tilde{\varphi},\varphi^*}$. We derived (ZCP_a) using (1.30), whereby

$$\left[\frac{\tilde{\varphi}(\varphi^*)}{\varphi^*} \right]^{\varepsilon-1} = \frac{r(\tilde{\varphi}, P, E)}{r(\varphi^*, P, E)} = \frac{r(\tilde{\varphi}, P, E)}{f}.$$

We just showed that this term is increasing if $\varepsilon_{\tilde{\varphi},\varphi^*} > 1$, decreasing if $\varepsilon_{\tilde{\varphi},\varphi^*} < 1$, and constant if $\varepsilon_{\tilde{\varphi},\varphi^*} = 1$. From (1.18) and $E = L$ we hence know that $\tilde{\varphi}(\varphi^*) P$ is increasing in φ^* if $\varepsilon_{\tilde{\varphi},\varphi^*} > 1$, decreasing if $\varepsilon_{\tilde{\varphi},\varphi^*} < 1$, and constant if $\varepsilon_{\tilde{\varphi},\varphi^*} = 1$. That is, $\varepsilon_{\tilde{\varphi},\varphi^*}$ captures two opposing effects on $r(\tilde{\varphi})$: on the one hand, the increase in φ^* raises $\tilde{\varphi}$ and hence directly the average profits, cf.(1.19) for a given P. On the other hand, the average profit is decreasing in the aggregate productivity of producers, which shows up as a low aggregate price level. In equilibrium, from (1.31), $\tilde{\varphi}(\varphi^*) P$ equals $\tilde{\varphi}(\varphi^*) P = M^{1/(\varepsilon-1)}/\alpha$. Hence, the direct and indirect effects of $\tilde{\varphi}(\varphi^*)$ and P on $\bar{\pi}$ determine the reaction of the number of available products as φ^* changes. In particular,

$$\varepsilon_{\tilde{\varphi},\varphi^*} \gtreqless 1 \Rightarrow \frac{\partial M}{\partial \varphi^*} \gtreqless 0.$$

This implies that comparative statics of aggregates that are indirectly influenced by φ^* via $\bar{\pi}$ with respect to φ^* crucially depend on our choice of $G(\varphi)$. This is particularly important in the theoretical

[28] Evidently, $\tilde{\varphi}$ increases with φ^* (as long as the underlying distribution is not a degenerated one-point distribution). The marginal change in the average productivity is given by

$$\begin{aligned}
\tilde{\varphi}'(\varphi^*) &= \frac{[1-G(\varphi^*)]\,\tilde{\varphi}_a}{(\varepsilon-1)\int_{\varphi^*}^{\infty} \varphi^{\varepsilon-1} dG(\varphi)} \left\{ \frac{g(\varphi^*)}{[1-G(\varphi^*)]^2} \int_{\varphi^*}^{\infty} \varphi^{\varepsilon-1} dG(\varphi) - \frac{g(\varphi^*)}{1-G(\varphi^*)} (\varphi^*)^{\varepsilon-1} \right\} \\
&= \frac{g(\varphi^*)\,\tilde{\varphi}}{(\varepsilon-1)\int_{\varphi^*}^{\infty} \varphi^{\varepsilon-1} dG(\varphi)} \left[\frac{1}{1-G(\varphi^*)} \int_{\varphi^*}^{\infty} \varphi^{\varepsilon-1} dG(\varphi) - (\varphi^*)^{\varepsilon-1} \right] \\
&= \frac{g(\varphi^*)\,(\tilde{\varphi})^{2-\varepsilon}}{(\varepsilon-1)[1-G(\varphi^*)]} \left[(\tilde{\varphi})^{\varepsilon-1} - (\varphi^*)^{\varepsilon-1} \right] > 0.
\end{aligned}$$

literature, which often resorts to the Pareto distribution when a closed form solution is needed (see, amongst others, Baldwin and Robert-Nicoud, 2008, and Gustafson and Segerstrom, 2007). We will see in Chapter 2 that the ZCP curve is flat in the $(\bar{\pi}, \varphi^*)$-space under Pareto-distributed productivities since $\varepsilon_{\tilde{\varphi},\varphi_a^*} = 1$. In this case the direct effect via $\tilde{\varphi}(\varphi^*)$ and the indirect effect via P on $\bar{\pi}$ exactly offset each other, so that $\bar{\pi}$ is independent of φ^* which implies that the number of producers is also independent of the cutoff.

Appendix

This appendix contains the rule for integrating by substitution which is used throughout the previous sections to substitute for firms with index j by productivity levels φ and Melitz' (2003) proof for the existence and uniqueness of the equilibrium in his seminal model.

Appendix 1.A Substituting for Firms with Productivity Levels

Consider the previously defined price index

$$P^{1-\varepsilon} \equiv \int_{j\in J} p(j)^{1-\varepsilon}\, dj = \int_0^M p(j)^{1-\varepsilon}\, dj$$

and the equilibrium price

$$p(j) = \frac{1}{\alpha\varphi}.$$

We aim at integrating over φ's instead of j's in the price index, since the price of all goods is a function of their productivity levels, $p(j) = p(j(\varphi))$. We thus apply the substitution rule from integral calculus. Intuitively speaking (under some condition, which are pointed out below), we may apply a substitution rule to the integrand and the inverse of the substitution rule to the limits of integration and integrate over φ instead of integrating over j. In the price index example, we thus need to find a substitution rule s for j, namely $s : [\varphi^*, \infty] \to [0, M]$. Since the firm index is just a number which can always be rearranged so that it indicates increasing levels of productivity, we define

$$s(\varphi) \equiv M\mu(\varphi).$$

$s(\varphi)$ is increasing in φ, with $s(\varphi^*) = M\mu(\varphi^*) = 0$ and $\lim_{\varphi\to\infty} s(\varphi) = \lim_{\varphi\to\infty} M\mu(\varphi) = M$. To perform the substitution, we replace j by $s(\varphi)$ and adjust the limits of integration by applying s^{-1}. Since each firm's price depends only on the firm's productivity level, substituting for j in the integrand simply means that we regard $p(j(\varphi))^{1-\varepsilon}$ as $p(\varphi)^{1-\varepsilon}$. From the definition of s, $dj = ds(\varphi) = M\mu'(\varphi)\, d\varphi$. The new limits of integration are also readily given: from $s(\varphi^*) = 0$ we have $s^{-1}(0) = \varphi^*$ and $\lim_{\varphi\to\infty} s(\varphi) = M$ gives $\lim_{M\to\infty} s^{-1}(M) = \infty$. Taken together, we have

$$P^{1-\varepsilon} = \int_0^M p(j)^{1-\varepsilon}\, dj = \int_{s^{-1}(0)}^{s^{-1}(M)} p(\varphi)^{1-\varepsilon} M\mu'(\varphi)\, d\varphi = M\int_{\varphi^*}^{\infty} p(\varphi)^{1-\varepsilon}\, d\mu(\varphi).$$

In what follows, we state the substitution rule somewhat more precisely and provide a short proof. Here is the rule for integration by substitution of the variable $x \equiv s(t)$ for a definite integral:

If I is a real valued integral $I \subseteq \mathbb{R}$, $s : [\gamma_1, \gamma_2] \rightarrow I$ is a continuous, differentiable function, and $f : I \rightarrow \mathbb{R}$ is a continuous function, then

$$\int_{s(\gamma_1)}^{s(\gamma_2)} f(x)\, dx = \int_{\gamma_1}^{\gamma_2} f(s(t))\, s'(t)\, dt. \tag{A.1}$$

A short proof is given as follows. Suppose we are given functions f, s, and s' that satisfy the required assumptions on continuity and differentiability. Continuity implies that $F' \equiv f$ exists so that $F(s(t))$ is defined. Taking the derivative with respect to t, we have

$$F(s(t))'(t) = F'(s(t))\, s'(t) = f(s(t))\, s'(t)$$

over the domain $t \in [\gamma_1, \gamma_2]$. Integrating this expression yields the desired result, i.e. equation (A.1):

$$\int_{\gamma_1}^{\gamma_2} f(s(t))\, s'(t)\, dt = F(s(\gamma_2)) - F(s(\gamma_1)) = \int_{s(\gamma_1)}^{s(\gamma_2)} f(x)\, dx.$$

In our case, I corresponds to the interval $[0, M]$, s is $s(\varphi)$ from above, γ_1 and γ_2 are given by φ^* and ∞, respectively, whereby $\lim_{\varphi\rightarrow\infty} s(\varphi) = M$ and $s(\varphi^*) = 0$, and $f(x)$ is given by $p(j)^{1-\varepsilon}$. To verify our result from above, we simply plug these expressions into (A.1), recognizing that $f(s(t))$ is simply $p^{1-\varepsilon}(j(\varphi)) = p^{1-\varepsilon}(\varphi)$, $ds(\varphi) = M\mu'(\varphi)$, and $dt = d\varphi$:

$$\int_0^M p(j)^{1-\varepsilon}\, dj = \int_{\varphi^*}^{\infty} p(\varphi)^{1-\varepsilon} M\mu'(\varphi)\, d\varphi.$$

Appendix 1.B Existence and Uniqueness of the Equilibrium

In this appendix we reproduce Melitz’ (2003) proof for the ZCP to cut the FE curve once from above in $(\varphi, \bar{\pi})$ space so that a unique equilibrium exists. We start with the autarky case and subsequently consider the equilibrium with international trade.

1.B.1 Existence and Uniqueness of the Autarky Equilibrium

If a solution exists, from (FE_a) and (ZCP_a), it satisfies

$$f\left\{\left[\frac{\tilde{\varphi}(\varphi)}{\varphi}\right]^{\varepsilon-1} - 1\right\} = \frac{\delta f_e}{1 - G(\varphi)} \Leftrightarrow \frac{\delta f_e}{f} = \Phi(\varphi) \equiv \omega(\varphi)\left[1 - G(\varphi)\right].$$

$\tilde{\varphi}(\varphi)$ is given in (1.28) (with φ^* replaced by φ, so that the index of integration must be changed to some other variable to avoid confusion), so that $\omega(\varphi) > 0$ as long as the distribution is non-degenerate. Thus, if $\Phi(\varphi)$ is strictly monotonically decreasing from ∞ to 0 on $(0, \infty)$, a unique intersection with

$\delta f_e/f$ exists so that there is a unique solution to $\bar{\pi}$ and φ^* in from (FE_a) and (ZCP_a); the same is true if $\Phi(\varphi)$ were strictly monotonically increasing from 0 to ∞, but we directly see that $\lim_{\varphi^* \to 0} \Phi(\varphi) = \infty$ as $\lim_{\varphi^* \to 0} \omega(\varphi) = \infty$. Hence, we are left to show that $d\Phi(\varphi)/d\varphi < 0$ and that $\Phi(\varphi)$ decreases to (at least) zero. To see this, recall from footnote 28 that

$$\tilde{\varphi}'(\varphi^*) = \frac{g(\varphi^*)(\tilde{\varphi})^{2-\varepsilon}}{(\varepsilon-1)[1-G(\varphi^*)]}\left[(\tilde{\varphi})^{\varepsilon-1} - (\varphi^*)^{\varepsilon-1}\right] > 0.$$

Accordingly,

$$\begin{aligned}
\omega'(\varphi) &= (\varepsilon-1)\left\{\frac{\tilde{\varphi}^{\varepsilon-2}g(\varphi)\tilde{\varphi}^{2-\varepsilon}}{\varphi^{\varepsilon-1}(\varepsilon-1)[1-G(\varphi)]}\left[\tilde{\varphi}^{\varepsilon-1}-\varphi^{\varepsilon-1}\right] - \frac{\tilde{\varphi}^{\varepsilon-1}}{\varphi^{\varepsilon}}\right\} = && \text{(A.2)}\\
&= (\varepsilon-1)\left[\frac{g(\varphi)}{(\varepsilon-1)[1-G(\varphi)]}\left(\frac{\tilde{\varphi}}{\varphi}\right)^{\varepsilon-1} - \frac{g(\varphi)}{(\varepsilon-1)[1-G(\varphi)]} - \frac{1}{\varphi}\left(\frac{\tilde{\varphi}}{\varphi}\right)^{\varepsilon-1}\right] = \\
&= \frac{g(\varphi)}{1-G(\varphi)}\omega(\varphi) - \frac{(\varepsilon-1)[\omega(\varphi)+1]}{\varphi}. && \text{(A.3)}
\end{aligned}$$

Therefore, $\Phi'(\varphi)$ is easily calculated:

$$\Phi'(\varphi) = \omega'(\varphi)[1-G(\varphi)] - \omega(\varphi)g(\varphi) = -\frac{(\varepsilon-1)[\omega(\varphi)+1][1-G(\varphi)]}{\varphi} < 0.$$

Moreover, the elasticity of $\Phi(\varphi)$ is strictly negative as $\omega(\varphi) > 0$,

$$\frac{\partial\Phi(\varphi)}{\partial\varphi}\frac{\varphi}{\Phi} = -(\varepsilon-1)\left[1+\frac{1}{\omega(\varphi)}\right] < -(\varepsilon-1) < 0$$

so that $\Phi(\varphi)$ is decreasing to zero in the limit as $\varphi \to \infty$ ($\Phi'(\varphi) < 0$ and $\Phi > 0$ from the last inequality), which completes the proof.

1.B.2 Existence and Uniqueness of the Equilibrium with International Trade

The equilibrium in the open economy is determined by the FE condition and the open economy adjusted ZCP condition. The solution satisfies (FE) and (ZCP)

$$f\left\{\left[\frac{\tilde{\varphi}(\varphi^*)}{\varphi^*}\right]^{\varepsilon-1} - 1\right\} + \Lambda_x(\varphi^*)nf_x\left\{\left[\frac{\tilde{\varphi}_x(\varphi^*)}{\varphi_x^*(\varphi^*)}\right]^{\varepsilon-1} - 1\right\} = \frac{\delta fe}{1-G(\varphi^*)},$$

or, equivalently using the definitions of $\omega(\varphi)$, $\Phi(\varphi)$, and $\Lambda_x(\varphi^*)$, and recognizing (1.49),

$$f\omega(\varphi^*) + \Lambda_x(\varphi^*)n\omega(\varphi_x^*)f_x = \frac{\delta fe}{1-G(\varphi^*)} \quad\Leftrightarrow\quad f\Phi(\varphi^*) + f\Phi(\varphi_x^*(\varphi^*)) = \delta f_e.$$

$\Phi(\varphi)$ was shown above to decrease from infinity to zero on $(0,\infty)$. Since the left hand side is a non-negative linear combination of $\Phi's$ while the right hand side is horizontal, a unique intersection exists.

Chapter 2

Melitz (2003) with Pareto-Distributed Productivities

To improve our understanding of the equilibrium characteristics, we consider a particular specification for $G(\varphi)$ and assume that productivity levels are drawn from a Pareto distribution:

$$\tilde{G}(\varphi) \equiv \begin{cases} 1 - \left(\frac{\varphi_0}{\varphi}\right)^k & \text{if } \varphi \geq \varphi_0 > 0 \\ 0 & \text{else} \end{cases}, \tag{2.1}$$

where $k > \varepsilon - 1$. The corresponding density is

$$\tilde{g}(\varphi) \equiv \tilde{G}'(\varphi) = \begin{cases} k\varphi_0^k \varphi^{-(k+1)} & \text{if } \varphi \geq \varphi_0 > 0 \\ 0 & \text{else} \end{cases}.$$

This assumption was made e.g. by Helpman, Melitz, and Yeaple (2004), Baldwin and Robert-Nicoud (2007), Gustafsson and Segerstrom (2007) and others. Figure 2.1 illustrates $\tilde{G}(\varphi)$, $\tilde{g}(\varphi)$ together with the resulting distribution of active firm's productivities $\mu(\varphi)$ and $\mu'(\varphi)$ for an arbitrary cutoff. We again start by deriving the equilibrium in autarky.

2.1 Autarky

2.1.1 Equilibrium Conditions

Under the Pareto assumption, the average productivity $\tilde{\varphi}$ is linear in the cutoff so that φ^* drops out from (ZCP_a). The ZCP curve is therefore horizontal in the $(\bar{\pi}, \varphi^*)$-space so that the average profits are exogenously determined (by k, ϵ, and f).

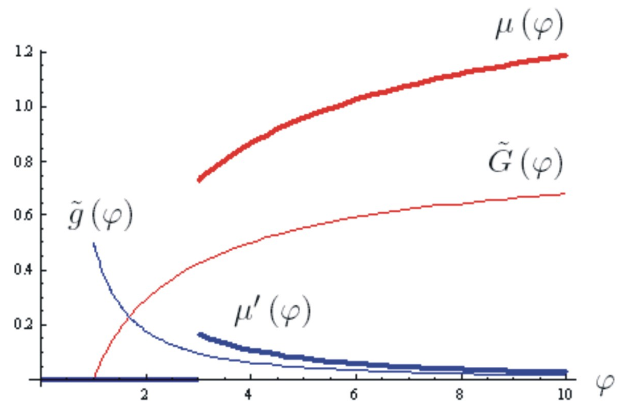


Figure 2.1: The Pareto Specification: Illustration of the Underlying and Equilibrium Productivity Distributions

To begin with, using $\tilde{G}(\varphi)$ in (1.22), the distribution of productivity levels of producers is given by

$$\mu(\varphi,\varphi^*) = \begin{cases} \frac{1-\left(\frac{\varphi_0}{\varphi}\right)^k}{\left(\frac{\varphi_0}{\varphi^*}\right)^k} & \text{for } \varphi \geq \varphi^* \\ 0 & \text{else} \end{cases}.$$

Hence, the density is

$$\mu'(\varphi,\varphi^*) = \begin{cases} k(\varphi^*)^k \varphi^{-(k+1)} & \text{if } \varphi \geq \varphi^* \\ 0 & \text{else} \end{cases}. \tag{2.2}$$

Substituting for $\mu'(\varphi)$ in (1.27) with (2.2), using $\varepsilon - 1 - k < 0$ and $\beta \equiv k/(k-\varepsilon+1) > 1$, the average productivity level of a producer in equilibrium equals

$$\begin{aligned} \tilde{\varphi}(\varphi^*) &= \left[\int_{\varphi^*}^{\infty} \varphi^{\varepsilon-1}\mu'(\varphi)\,d\varphi\right]^{\frac{1}{\varepsilon-1}} = \left[(\varphi^*)^k k \int_{\varphi^*}^{\infty} \varphi^{\varepsilon-k-2} d\varphi\right]^{\frac{1}{\varepsilon-1}} \\ &= \left\{(\varphi^*)^k \frac{k}{\varepsilon-1-k}\left[\varphi^{\varepsilon-1-k}\right]_{\varphi^*}^{\infty}\right\}^{\frac{1}{\varepsilon-1}} = \beta^{\frac{1}{\varepsilon-1}}\varphi^*. \end{aligned} \tag{2.3}$$

Note, that the average productivity is linear in the cutoff productivity level. This proves the earlier remark on the slope of the ZCP condition, namely that a Pareto distribution implies

$$\varepsilon_{\tilde{\varphi},\varphi^*} = \frac{\partial\tilde{\varphi}(\varphi^*)}{\partial\varphi^*}\frac{\varphi^*}{\tilde{\varphi}(\varphi^*)} = \beta^{\frac{1}{\varepsilon-1}}\frac{\varphi^*}{\beta^{\frac{1}{\varepsilon-1}}\varphi^*} = 1.$$

A one percent increase in the equilibrium cutoff raises the average productivity by exactly one percent. Using (2.3) in (ZCP_a), we find the ZCP curve to be horizontal,

$$\bar{\pi} = (\beta - 1) f. \tag{$\widetilde{ZCP_a}$}$$

Using $\tilde{G}(\varphi)$ in (FE_a) directly, the FE condition reads

$$\bar{\pi} = \delta f_e \left(\frac{\varphi^*}{\varphi_0} \right)^k . \qquad (\widetilde{FE}_a)$$

The fact that the ZCP curve is flat and strictly positive in $(\bar{\pi}, \varphi^*)$-space while the FE curve starts at the origin and increases monotonically to infinity proves that a unique steady state exists.

2.1.2 The Autarky Equilibrium under the Pareto Distribution

From $(\widetilde{ZCP}_a)$ and $(\widetilde{FE}_a)$, the equilibrium cutoff productivity equals

$$\varphi^* = \left[\frac{(\beta - 1) f}{\delta f_e} \right]^{\frac{1}{k}} \varphi_0 . \qquad (2.4)$$

The minimum productivity requirement is increasing in the overhead costs, but decreasing in the market entry costs. The equilibrium average profit is directly given in $(\widetilde{ZCP}_a)$. Under the Pareto assumption, it depends only on the overhead costs. Since the average profit is independent of φ^* under Pareto-distributed productivities, the number of producers $M = \varepsilon / \bar{r}$ ($\bar{r} = \bar{\pi} + f$) is also independent of φ^* (in particular, it is independent of the upfront costs f_e) and, using $(\widetilde{ZCP}_a)$ in (1.34), equals

$$M = \frac{L}{\varepsilon \beta f} . \qquad (2.5)$$

The mass of producers is increasing in the size of the country and decreasing in the overhead costs. Aggregate profits, however are independent of f, $M\bar{\pi} = (1 - 1/\beta) L / \varepsilon$. Note, however, that f affects the average productivity positively and thereby generates a trade-off between the mass of available products and the productivity in production:[1]

$$\tilde{\varphi} = \beta^{\frac{1}{\varepsilon - 1}} \left[\frac{(\beta - 1) f}{\delta f_e} \right]^{\frac{1}{k}} \varphi_0 . \qquad (2.6)$$

Interestingly, low entry costs imply a high average productivity. Since decreasing f_e increases the average productivity without altering the average profits, entry costs do not affect the mass of producers either. Since M is independent of the entry costs, the mass of workers hired for market entry in the stationary equilibrium must be independent of f_e, too. Using (2.4) and (2.5) verifies

$$f_e M^e = \frac{\delta M}{\left(\frac{\varphi_0}{\varphi^*} \right)^k} = \frac{\delta L (\beta - 1)}{\varepsilon \beta \delta} .$$

[1] This is given the distribution of productivities. One way of capturing technological progress is an outward shift of the support of $G(\varphi)$. Another one is to endogenize the availability of new products, cf. Chapter 3.

Evidently, this is a direct implication of the ZCP curve being horizontal under the Pareto distribution. As a caveat, suppose that we had actual discounting in addition to δ. The interest rate would then simply add to δ in the denominator in (2.4), suggesting a co-movement of interest rates and average productivity. In Section 3, we explicitly include physical capital as a factor in production in a variant of the Melitz (2003) model with positive growth and show that in this environment, φ^* and hence $\tilde{\varphi}$ are not affected by the interest rate.

Finally, using φ^* from (2.4) in (1.36), (instantaneous) aggregate welfare amounts to

$$U = \alpha \left(\frac{L^\varepsilon}{\varepsilon f}\right)^{\frac{1}{\varepsilon-1}} \left[\frac{(\beta-1) f}{\delta f_e}\right]^{\frac{1}{k}} \varphi_0$$

or, after collecting parameters,

$$U = \frac{\varepsilon - 1}{f^{\left(\frac{1}{\varepsilon-1}-\frac{1}{k}\right)}} \left(\frac{L}{\varepsilon}\right)^{\frac{1}{\alpha}} \left(\frac{\beta-1}{\delta f_e}\right)^{\frac{1}{k}} \varphi_0.$$

Welfare is increasing in L (and ε) and decreasing in δ, f, and f_e. From (2.5), larger countries, i.e. economies with abundant labor supply, generate a high product diversity, which per se is valuable to the consumer, see (1.35). With respect to welfare, the productivity-raising impact of increasing overhead costs is dominated by the accompanied decrease in product variety. Similarly, barriers to entry, as measured by f_e, unambiguously lower welfare – even with M and $f_e M^e$ unaffected by f_e. Since $P = M^{1/(1-\varepsilon)}/(\alpha\tilde{\varphi})$ as implied by (1.31) and $\tilde{\varphi}$ is independent of L (due to the independence of φ^* from L, see (2.4)), the high product diversity in labor abundant economies also implies a lower aggregate price level. We can explicitly solve for the price index using (2.3) and (2.5):

$$P = \left[M^{\frac{1}{\varepsilon-1}} \alpha \tilde{\varphi}\left(\varphi^*\right)\right]^{-1} = \frac{\left(\frac{L}{\varepsilon\beta f}\right)^{\frac{1}{1-\varepsilon}}}{\alpha \beta^{\frac{1}{\varepsilon-1}} \varphi^*} = \frac{\left(\frac{\varepsilon f}{L}\right)^{\frac{1}{\varepsilon-1}}}{\alpha \left[\frac{(\beta-1) f}{\delta f_e}\right]^{\frac{1}{k}} \varphi_0} = \left(\frac{\varepsilon}{L}\right)^{\frac{1}{\varepsilon-1}} \left(\frac{\delta f^{\frac{1}{\beta-1}} f_e}{\beta-1}\right)^{\frac{1}{k}} \left(\alpha\varphi_0\right)^{-1}.$$

The aggregate price index is decreasing in L and increasing in δ, f, and f_e. It reflects the scarcity of labor as indicated by the real wage $1/P$.

2.1.3 On the Impact of Fixed Production and Entry Costs on the Cutoff

The preceding analysis revealed that the cutoff is increasing in the overhead costs f and decreasing in the entry costs f_e (see (2.4)). Here we take a closer look at the effects exerted by fixed production and entry costs and how they affect the average productivity. Our point of departure is the price index from above, where we substitute for $\tilde{\varphi}$ from (2.6):

$$\alpha \tilde{\varphi}\left(\varphi^*\right) = \alpha \beta^{\frac{1}{\varepsilon-1}} \varphi^* = \frac{\frac{1}{P}}{M^{\frac{1}{\varepsilon-1}}}. \tag{2.7}$$

C.p., there are two channels through which the overhead costs affect the average productivity level: the real wage ($1/P$)-channel and the M-channel. The channels have countervailing effects. First, if we hold M constant, potential entrants compete for scarce labor and the most productive firms "win" since they can afford to pay high wages. Therefore, holding the mass of producers constant, the average productivity of entrants exceeds the average productivity of incumbent producers, leading to a rise in $\tilde{\varphi}$. Second, however, if we fix the real wage, entrants must have a lower average productivity than incumbents, since less productive firms demand less labor in equilibrium. Thereby labor demand remains at a level compatible with the fixed real wage $1/P$. C.p. this decreases the average productivity. A decline in both f and f_e raises the real wage. In the case of f_e under $\tilde{G}(\varphi)$, where the ZCP is horizontal in $(\varphi^*, \bar{\pi})$-space, M remains unaffected, so that the decline in entry costs raises the average productivity. In fact, the increase in the average productivity is true in any case, i.e. even if a decline in f_e induces additional entry. We can infer from Figure 1.1 that the impact via the $1/P$ channel dominates the impact via the M channel (since φ^* goes up): if the ZCP is downward sloping, lowering f_e implies a decline in $\bar{\pi}$ and hence an increase in the number of producers. The decline in M absorbs the increase in the minimum productivity due to the increase in the real wage only to some extent. If the ZCP curve is upward sloping, this means that a decrease in f_e goes together with an increase in $\bar{\pi}$ and hence a decreasing number of producers. In this case, the decline in M reinforces the exit pressure on the least productive firms exerted by the increase in the real wage. Note that utility is increasing both in productivity and in product variety. Given this tension, a closer closer look at the strength of the gains from specialization implicitly determined in the consumers' preferences is an interesting task for future work (cf. footnote 7).[2] in this direction. Interestingly, with Pareto-distributed productivities, there is no effect on M as $\varphi^*/\tilde{\varphi}(\varphi^*)$ changes due to f_e. A decline in f_e then only affects utility via its positive impact on aggregate productivity.

The overhead costs affect productivity through both channels. In the case of f, however, the decline in the cutoff due to a larger mass of available products outweighs the productivity gains from rising wages (otherwise, the cutoff would actually increase with $\tilde{\varphi}$, see (2.4) or, equivalently, from the stationary condition $\delta M = M^e\left[1 - G(\varphi^*)\right]$ with M fixed, an increase in M^e must be accompanied by an increase in φ^*). Accordingly, a decline in f induces less productive firms to enter and their products

[2]The framework with specialized capital goods in Section 3 further provides the point of departure for disentangling the degree of substitutability from the degree of market power along the lines of Alvarez-Pelaez and Groth (2005).

overcompensate the consumer for her utility loss from higher prices.

Dying producers at first relax the resource constraint (so that labor becomes less scarce and $1/P$ falls as δ increases). Likewise, an increase in δ therefore lowers the real wage and allows less productive firms to start production: φ^* falls.

To summarize, we have seen that the impact of fixed costs works through the labor market. The overhead costs and the market entry costs thereby work through two channels, one emphasizing product variety, the other emphasizing low prices/high productivity. Overhead costs thereby have a stronger impact via their effect on product variety relative to productivity. In the case of entry costs, this relation is reversed and the productivity channel dominates. These observations also hold when we dispense with the Pareto assumption. We have seen, however, that the Pareto assumption implies $\varepsilon_{\tilde{\varphi},\varphi^*} = 1$, i.e. a flat ZCP curve. The only effect of f_e is then on productivity as M remains unaffected by entry costs. In general, falling entry costs raise the cutoff by more, the higher the slope of the ZCP curve. In case of the Pareto distribution, it is zero.

We next turn to the open economy.

2.2 Open Economy

2.2.1 Equilibrium Conditions

The cutoff productivity is again determined by the FE and the ZCP conditions. Given the definition of $\bar{\pi}$, international trade does not affect the FE curve, see (FE), hence $(\widetilde{FE}_a)$ holds in the open economy also:

$$\bar{\pi} = \delta f_e \left(\frac{\varphi^*}{\varphi_0}\right)^k . \qquad (\widetilde{FE})$$

The first summand of the ZCP condition (repeated here for convenience),

$$\bar{\pi} = f\left\{\left[\frac{\tilde{\varphi}_d(\varphi^*)}{\varphi^*}\right]^{\varepsilon-1} - 1\right\} + \Lambda_x(\varphi^*)\, n f_x \left\{\left[\frac{\tilde{\varphi}_x(\varphi^*)}{\varphi_x^*}\right]^{\varepsilon-1} - 1\right\},$$

is the same as in autarky,

$$f\left\{\left[\frac{\tilde{\varphi}_d(\varphi^*)}{\varphi^*}\right]^{\varepsilon-1} - 1\right\} = f(\beta - 1).$$

We are thus left to calculate the second term. From (2.3), $\tilde{\varphi}_x/\varphi_x^*$ equals $\tilde{\varphi}/\varphi^* = \beta^{1/(\varepsilon-1)}$. Due to the fact that the average productivity is linear in the cutoff under the Pareto distribution, the cutoff drops out from this ratio. Applying the Pareto specification to (1.51) and (1.52), the probability of starting production is

$$\Lambda(\varphi^*) = 1 - \tilde{G}(\varphi^*) = \left(\frac{\varphi_0}{\varphi^*}\right)^k$$

so that, from (1.49), the probability of exporting equals

$$\Lambda_x(\varphi^*) = \frac{\left(\frac{\varphi_0}{\varphi_x^*}\right)^k}{\Lambda(\varphi^*)} = \frac{\left(\frac{\varphi_0}{\varphi_x^*}\right)^k}{\left(\frac{\varphi_0}{\varphi^*}\right)^k} = \left(\frac{\varphi^*}{\varphi_x^*}\right)^k = \left(\frac{f}{\tau^{\varepsilon-1} f_x}\right)^{\frac{\beta}{\beta-1}}, \tag{2.8}$$

where we used $k/(\varepsilon-1) = \beta/(\beta-1)$. Substituting for $\Lambda_x(\varphi^*)$ and $\tilde{\varphi}_x/\varphi_x^*$ from above, the expected additional profits from exporting, i.e. the additional term in the open economy ZCP condition is

$$\Lambda_x(\varphi^*)\, n f_x \left\{\left[\frac{\tilde{\varphi}_x(\varphi^*)}{\varphi_x^*}\right]^{\varepsilon-1} - 1\right\} = \left(\frac{f}{\tau^{\varepsilon-1} f_x^{\frac{1}{\beta}}}\right)^{\frac{\beta}{\beta-1}} n(\beta-1).$$

Define further

$$0 < \sigma \equiv \tau^{-k}\left(\frac{f}{f_x}\right)^{\frac{1}{\beta-1}} < \frac{f_x}{f}.$$

The second inequality thereby follows from (PA).[3] Taken together, the ZCP condition is given by

$$\bar{\pi} = (\beta-1) f + n(\beta-1)\left(\frac{f}{\tau^{\varepsilon-1} f_x^{\frac{1}{\beta}}}\right)^{\frac{\beta}{\beta-1}} = (\beta-1) f (1 + n\sigma). \tag{$\widetilde{ZCP}$}$$

This verifies the earlier assessment that ($\widetilde{ZCP}$) is the ZCP curve from autarky shifted upwards, accounting for the additional profit opportunities in the export markets for producers with productivity levels above φ_x^*. Given the definition of $\bar{\pi}$ based on $\tilde{\varphi}$ and $\tilde{\varphi}_x$, the average profit is independent of the cutoff productivity. Including the possibility of exporting raises $\bar{\pi}$ above the average profit in autarky as long as there is some trade, i.e. as long as τ and f_x are finite.

2.2.2 Equilibrium with International Trade

Substituting for $\bar{\pi}$ in ($\widetilde{ZCP}$) with ($\widetilde{FE}$), the cutoff associated with profitable production equals

$$\varphi^* = \left[\frac{(\beta-1) f}{\delta f_e}(1 + n\sigma)\right]^{\frac{1}{k}} \varphi_0. \tag{2.9}$$

[3] The parameter assumption that guarantees partitioning into exporters and non-exporters can be rewritten as

$$\frac{f_x}{f}\tau^{\varepsilon-1} > 1 \quad \Leftrightarrow \quad \left(\frac{f}{f_x}\right)^{\frac{\beta}{\beta-1}} \tau^{-k} < 1$$

so that

$$0 < \tau^{-k}\left(\frac{f}{f_x}\right)^{\frac{1}{\beta-1}} < \frac{f_x}{f}.$$

Inserting the definition of σ yields the expression above.

Relative to the autarky cutoff, the only difference is the additional summand in the squared brackets, i.e. $\varphi^* = (1+n\sigma)^{1/k}\,\varphi^*$. Using the definition of σ in (1.49), we have

$$\tilde{\varphi} = \sigma^{-\frac{1}{k}} \left(\frac{f}{f_x}\right)^{-\frac{1}{k}} \varphi^*. \tag{2.10}$$

Substituting for φ^* with (2.9), the cutoff productivity level for exporting is given by

$$\varphi_x^* = \sigma^{-\frac{1}{k}} \left(\frac{f}{f_x}\right)^{-\frac{1}{k}} \left[\frac{(\beta-1)\,f}{\delta f_e}\,(1+n\sigma)\right]^{\frac{1}{k}} \varphi_0 = \left[\frac{(\beta-1)\,nf_x}{\delta f_e}\left(1+\frac{1}{n\sigma}\right)\right]^{\frac{1}{k}} \varphi_0.$$

Switching from autarky to some limited level of trade openness discretely raises the domestic cutoff φ^* and forces the least productive firms to shut down. The intuition for this is, contrary to the claim in Melitz (2003), twofold. First, as argued by Melitz, the additional production of the exporting firms bids up the wage rate, thereby leading to a decline in all firms' domestic revenues and shifting market shares from non-exporters to exporters. Second, all local producers' market shares shrink proportionally as foreign products become available (cf. the discussion in Section 1.3.3 above).

By inspection of (2.9), the domestic cutoff productivity level is monotonic in n, f_x, and τ and converges to the autarky cutoff as barriers to trade become prohibitive ($\tau \to \infty$ and/or $f_x \to \infty$ so that $\sigma \to 0$) or $n \to 0$. Hence, a smooth change in the exposure to trade has a similar impact as the switch from autarky to some limited level of trade openness. In particular, taking the derivative of φ^* with respect to n, τ, and f_x shows that the least productive firms exit as the barriers to trade, τ or f_x, fall. At the same time, a reduction in either τ or f_x allows some non-exporting producers to start exporting (φ_x^* is decreasing as τ or f_x fall). An increase in the number of trading partners induces the least productive exporters to cease exporting (φ_x^* rises with n).

From (2.10) and the definition of σ, $\Lambda_x = (\varphi^*/\varphi_x^*)^k = \sigma f/f_x$. Using the average profit $\bar{\pi}$ from $(\widetilde{ZCP})$ and substituting for $\Lambda_x f_x = \sigma f$, the average revenue as derived in (1.61) reads

$$\bar{r} = \varepsilon\left[\bar{\pi} + f + n\Lambda_x f_x\right] = \varepsilon\left[(\beta-1)\,f\,(1+n\sigma) + f\,(1+n\sigma)\right] = \varepsilon\beta f\,(1+n\sigma)\,.$$

Therefore, the mass of domestic producers is given by

$$M = \frac{L}{\bar{r}} = \frac{1}{\varepsilon\,(1+n\sigma)}\left(\frac{L}{\beta f}\right) \;<\; \frac{L}{\beta f},$$

which is lower than in autarky, see (2.5). Corresponding to our assessment above, everything that raises revenues from exports, like a decline in τ or f_x so that σ increases, or an increase in the number of trading partners, reduces the mass of domestic producers. A reduction in the barriers to trade (σ increases) similarly allows more producers to become exporters, increasing the number of trading partners again reduces the mass of exporters (M_x is increasing in σ and falls in n):

$$M_x = M\Lambda_x(\varphi^*) = \frac{\sigma\frac{f}{f_x}}{\varepsilon(1+n\sigma)}\left(\frac{L}{\beta f}\right) = \frac{\sigma}{\varepsilon(1+n\sigma)}\left(\frac{L}{\beta f_x}\right).$$

Accordingly, there are

$$M_t = [1+n\Lambda_x(\varphi^*)]\,M = \frac{1+n\sigma\frac{f}{f_x}}{\varepsilon(1+n\sigma)}\left(\frac{L}{\beta f}\right)$$

goods available in each economy, so that, under $\tilde{G}(\varphi)$, the consumer has more goods available with trade if and only if $f \geq f_x$ (recall that M was $L/(\beta f)$ in autarky). Under (PA), $\tau^{\varepsilon-1} f_x > f$ so that both $f > f_x$ and $f < f_x$ is possible if $\tau > 1$.

To calculate the average productivity of all producers, $\tilde{\varphi}_t$, note that from the definition of $\tilde{\varphi}$ and $\tilde{\varphi}_x$ and their expressions under the Pareto assumption above

$$\begin{aligned}\int_{\varphi^*}^{\infty}\varphi^{\varepsilon-1}d\tilde{G}(\varphi) &= \left[1-\tilde{G}(\varphi^*)\right]\int_{\varphi^*}^{\infty}\varphi^{\varepsilon-1}d\mu \;\Leftrightarrow\; \int_{\varphi^*}^{\infty}\varphi^{\varepsilon-1}d\mu = \beta(\varphi^*)^{\varepsilon-1},\\ \int_{\varphi_x^*}^{\infty}\varphi^{\varepsilon-1}d\mu &= \beta(\varphi_x^*)^{\varepsilon-1} = \beta\tau^{\varepsilon-1}\left(\frac{f_x}{f}\right)(\varphi^*)^{\varepsilon-1}.\end{aligned}$$

Substituting for these expressions in $\tilde{\varphi}_t$ from (1.64) gives

$$\tilde{\varphi}_t = \left\{\frac{1}{M_t}\left[M\beta(\varphi^*)^{\varepsilon-1} + \tau^{1-\varepsilon}nM_x\beta\tau^{\varepsilon-1}\frac{f_x}{f}(\varphi^*)^{\varepsilon-1}\right]\right\}^{\frac{1}{\varepsilon-1}} = \beta^{\frac{1}{\varepsilon-1}}\varphi^*\left(\frac{M}{M_t} + \frac{M_x}{M_t}n\frac{f_x}{f}\right)^{\frac{1}{\varepsilon-1}}.$$

Inserting the shares of domestically produced and exported goods,

$$\frac{M}{M_t} = \frac{1}{1+n\sigma\frac{f}{f_x}},$$

and $nM_x/M_t = 1 - M/M_t$, we find

$$\tilde{\varphi}_t = \left(\frac{1+n\sigma}{1+n\sigma\frac{f}{f_x}}\right)^{\frac{1}{\varepsilon-1}}\beta^{\frac{1}{\varepsilon-1}}\varphi^*.$$

Recall that $\beta^{1/(\varepsilon-1)}\varphi^*$ was the average productivity in autarky. The average productivity, not measured "at the factory gate" but via $\tilde{\varphi}_t$, i.e. including the output shrinkage from iceberg costs, may be larger or smaller than under autarky, depending on the relative size of f and f_x. If the fixed exports costs are sufficiently high so that $f_x > f$, the term in brackets is less than 1 so that the increase in φ^* due to international trade may not be sufficient for $\tilde{\varphi}_t$ to exceed $\tilde{\varphi}$ from autarky. In the opposite case where $f \geq f_x$, $\tilde{\varphi}_t$, the average productivity in the open economy exceeds the average productivity in autarky. In what follows, we show that even if $\tilde{\phi}_t$ falls, aggregate welfare will increase with trade openness in the stationary equilibrium.

We firstly calculate the equilibrium price index by inserting M_t and $\tilde{\varphi}_t$ in (1.65):

$$P = \frac{M_t^{\frac{1}{1-\varepsilon}}}{\alpha\tilde{\varphi}_t} = \frac{\left(\frac{1+n\sigma\frac{f}{f_x}}{\varepsilon(1+n\sigma)}\right)^{\frac{1}{1-\varepsilon}}\left(\frac{L}{\beta f}\right)^{\frac{1}{1-\varepsilon}}}{\alpha\left(\frac{1+n\sigma}{1+n\sigma\frac{f}{f_x}}\right)^{\frac{1}{\varepsilon-1}}\beta^{\frac{1}{\varepsilon-1}}\left[\frac{(\beta-1)f}{\delta f_e}\left(1+n\sigma\right)\right]^{\frac{1}{k}}\varphi_0} = \frac{\left(\frac{L}{\varepsilon f}\right)^{\frac{1}{1-\varepsilon}}}{\alpha\left[\frac{(\beta-1)f}{\delta f_e}\left(1+n\sigma\right)\right]^{\frac{1}{k}}\varphi_0}.$$

For comparison, the price index under autarky was

$$P_a = \frac{\left(\frac{L}{\varepsilon f}\right)^{\frac{1}{1-\varepsilon}}}{\alpha\left[\frac{(\beta-1)f}{\delta f_e}\right]^{\frac{1}{k}}\varphi_0}.$$

International trade triggers an expansion of the production of the most productive firms, which translates into an increase in labor demand and therefore an increase in the real wage (the inverse of the price index).

Using the expression for P, the instantaneous utility flow in the stationary equilibrium equals

$$U = \alpha\tilde{\varphi}_t M_t^{1/(\varepsilon-1)}L = \frac{L}{P} = \alpha\left(\frac{L^{\varepsilon}}{\varepsilon f}\right)^{\frac{1}{\varepsilon-1}}\left[\frac{(\beta-1)f}{\delta f_e}\left(1+n\sigma\right)\right]^{\frac{1}{k}}\varphi_0.$$

Again for comparison, the corresponding figure under autarky is

$$U_a = \alpha\left(\frac{L^{\varepsilon}}{\varepsilon f}\right)^{\frac{1}{\varepsilon-1}}\left[\frac{(\beta-1)f}{\delta f_e}\right]^{\frac{1}{k}}\varphi_0.$$

In this environment, international trade unambiguously raises the level of welfare in the stationary equilibrium ($U > U_a$ whenever there is some international trade, i.e., if σ is bounded away from zero). Note, however, that we do not evaluate the welfare effects along the transition path from the stationary distribution under autarky to the new stationary distribution.[4]

[4] Calibrating a variant of the Melitz (2003) model to match the U.S. employment size distribution of manufacturing establishments, Alessandria and Choi (2007) find that steady state consumption is a notoriously bad measure of welfare (especially in the presence of sunk costs to exporting). They find that, in models with fixed costs of exporting, comparing steady state consumption levels after trade liberalization understates welfare gains and overstates welfare gains in models without fixed costs. Cf. Chaney (2005) on the transitional dynamics of trade reforms in a Melitz-type model.

2.2.3 Who Wins from Trade Liberalization?

Finally, we use our example to demonstrate the reallocation effects among surviving producers exerted by international trade (evidently, producers that are forced to shut down lose).[5] We already know from the discussion in Section 1.3.3, that non-exporters unambiguously lose from trade, since their domestic revenues decline: $r_{d,a}(\varphi, P_a, L) < r_d(\varphi, P, L)$. Since exporting comes with fixed costs and domestic revenues also decline for exporters, there exists a productivity level $\bar{\varphi} > \varphi_x^*$ such that all producers with $\varphi < \bar{\varphi}$ lose from trade and producers with $\varphi \geq \bar{\varphi}$ gain from trade. The threshold is implicitly defined by

$$\pi_a(\bar{\varphi}, P_a, L) \equiv \pi_d(\bar{\varphi}, P, L) + n\pi_x(\bar{\varphi}, P, L). \tag{2.11}$$

Solving for $\bar{\varphi}$, we find[6]

$$\varphi_x^* \; < \; \bar{\varphi} \; = \; \left[\frac{n\frac{f_x}{f}}{(1+n\tau^{1-\varepsilon}) - (1+n\sigma)^{\frac{\beta-1}{\beta}}}\right]^{\frac{1}{\varepsilon-1}} \varphi^* = \left[\frac{(n\sigma)^{\frac{\beta-1}{\beta}}\left(n\frac{f_x}{f}\right)^{\frac{1}{\beta}}}{(1+n\tau^{1-\varepsilon}) - (1+n\sigma)^{\frac{\beta-1}{\beta}}}\right]^{\frac{1}{\varepsilon-1}} \varphi_x^*.$$

---

[5]This subsection presents the discussion in Melitz (2003) for a Pareto-specified distribution of productivity levels.

[6]Inserting the expressions for profits, (2.11) becomes

$$\frac{r_a(\bar{\varphi}, P_a, L)}{\varepsilon} - f = \left(1 + n\tau^{1-\varepsilon}\right)\frac{r_d(\bar{\varphi}, P, L)}{\varepsilon} - f - nf_x.$$

Using (1.68) and (1.69) to replace

$$\frac{r_a(\bar{\varphi}, P_a, L)}{\varepsilon} = \left(\frac{\bar{\varphi}}{\varphi_a^*}\right)^{\varepsilon-1} f \text{ and } \frac{r_d(\bar{\varphi}, P, L)}{\varepsilon} = \left(\frac{\bar{\varphi}}{\varphi^*}\right)^{\varepsilon-1} f,$$

and rearranging terms yields

$$\bar{\varphi}^{\varepsilon-1}\left[\left(1+n\tau^{1-\varepsilon}\right)(\varphi^*)^{1-\varepsilon} - (\varphi_a^*)^{1-\varepsilon}\right] = n\frac{f_x}{f}.$$

Inserting $\varphi^*/(1+n\sigma)^{1/k} = \varphi_a^*$ we find

$$\left(\frac{\bar{\varphi}}{\varphi^*}\right)^{\varepsilon-1}\left[\left(1+n\tau^{1-\varepsilon}\right) - (1+n\sigma)^{\frac{\beta-1}{\beta}}\right] = n\frac{f_x}{f}.$$

Solving for $\bar{\phi}$ yields the first expression above, substituting for φ^* using (2.9) gives the second expression.

Formally, the inequality sign follows from $\sigma > 0$, $\beta > 1$, and $n \geq 2$.[7] The impact of trade on the industry equilibrium is illustrated in Figure 2.2. As argued above, opening up to trade raises the cutoff productivity from its autarky level φ_a^* to φ^*. This increase comes with a decline in revenues for surviving firms that do not find it profitable to engage in exporting. Another increase in the exposure to trade raises the domestic cutoff further and lowers the cutoff for exporting profitably. The formally least productive exporters loose profits. Sufficiently productive firms, those with $\varphi \geq \bar{\varphi}$, gain additional profits.

[7]This is easily verified as follows. For $\bar{\varphi} > \phi_x^*$ to hold, it has to be that

$$(n\sigma)^{\frac{\beta-1}{\beta}} \left(n\frac{f_x}{f}\right)^{\frac{1}{\beta}} > \left(1 + n\tau^{1-\varepsilon}\right) - (1+n\sigma)^{\frac{\beta-1}{\beta}}.$$

Inserting the definition of σ on the left hand side, collecting terms, and rearranging, it equivalently has to hold that

$$(1+n\sigma)^{\frac{\beta-1}{\beta}} > 1.$$

$\sigma > 0$, $\beta > 1$, and $n \geq 2$ imply that this is true.

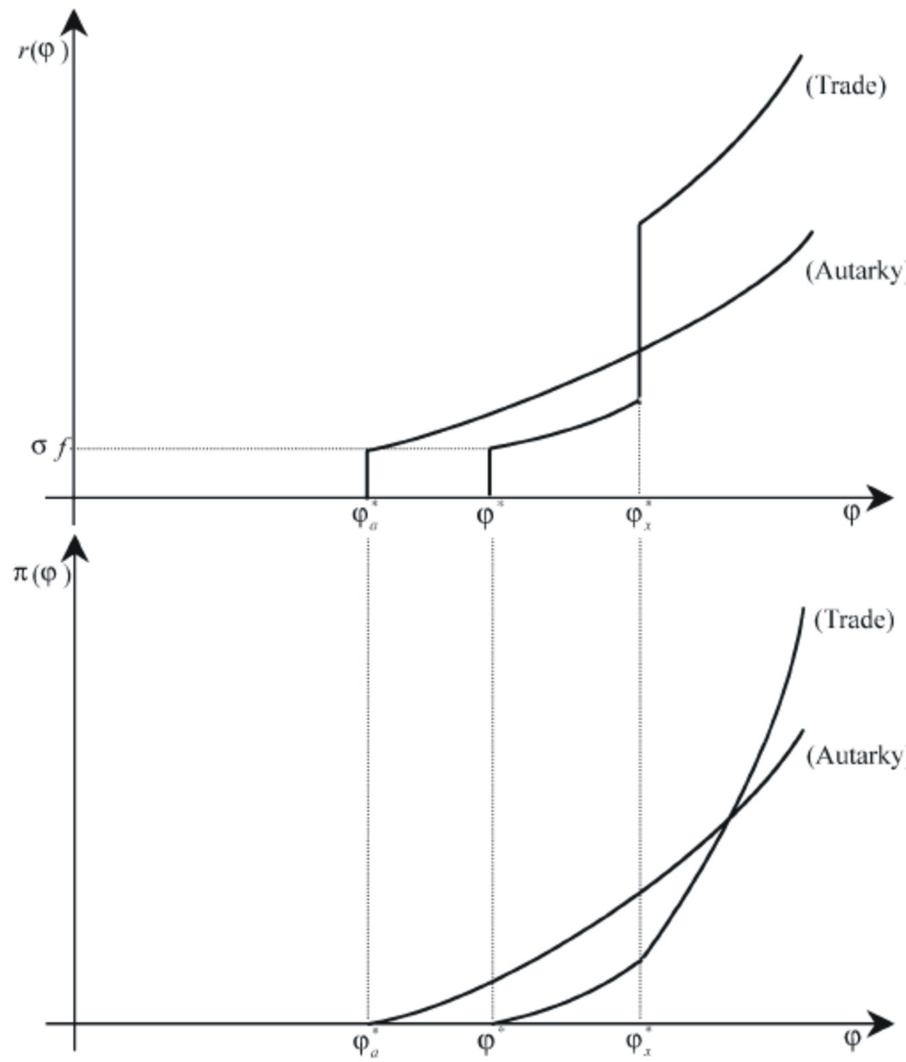


Figure 2.2: Reallocation of Market Shares and Profits (Adopted from Melitz, 2003, p. 1715)

Appendix 2.A The Distribution of Input Coefficients

In Chapter 3 below we work with the distribution of input coefficients $b \equiv 1/\varphi$ instead of productivity levels. We can simply infer their density $h(b)$ from $G(\varphi)$:[8]

$$h(b) = g\left(\frac{1}{b}\right) \cdot \left| \left(\frac{1}{\varphi}\right)' \right|_{\varphi=b}.$$

In particular, using a Pareto distribution for $G(\varphi)$, i.e. the specification of $\tilde{G}$ in (2.1), we find

$$h(b) = \tilde{g}\left(\frac{1}{b}\right) \cdot \left| \left(\frac{1}{\varphi}\right)' \right|_{\varphi=b} = k\varphi_0^k b^{k+1} \cdot \left| -b^{-2} \right| = k\varphi_0^k b^{k-1}.$$

Hence, after integrating, the distribution function of input coefficients that implies Pareto-distributed productivity levels is

$$H(b) = \left(\frac{b}{b_0}\right)^k, \; b \in (0, b_0]$$

where $b_0 \equiv 1/\varphi_0$.

[8]The expression for $h(b)$ above follows from the definition of $b \equiv 1/\varphi$ as a monotonically decreasing and differentiable function for all $\varphi > 0$. More generally, if $f(x)$ is the (value of the) density of a continuous random variable x (at "x"), then, if $y = u(x)$ with u a continuous and monotonic function over all x for which $f(x) \neq 0$ (so that the inverse of u exists) the density of y is $h(y) = f\left(u^{-1}(y)\right) \cdot |\left(u^{-1}\right)'(y)|$ given that $u'(x) \neq 0$ (elsewhere, $g(y) = 0$). See e.g. Freund and Walpole (1980).

Chapter 3

A Dynamic Trade Model with Heterogeneous Firms and Semi-Endogenous Growth[1]

3.1 Abstract

We investigate the impact of incremental trade liberalization in a dynamic model of endogenous growth with heterogeneous firms and costly trade. Growth originates from horizontal specialization and the steady state productivity growth rate is positive. Innovations require costly R&D and are conducted by profit-seeking researchers. Including physical capital as a factor of production, we find that after appropriate adjustments in the production structure, previous results on the reallocation of resources and the selection of firms following trade liberalization continue to hold. We show, however, that unlike in the Melitz (2003) model, the reallocation effect does not work through increases in the factor price in production.

3.2 Introduction

The relation between trade and growth remains unfinished business. On the one hand, recent empirical research convincingly argues that commonly used measures of "trade openness" are either poor measures of barriers to trade or otherwise are highly correlated with important determinants

[1]This section is a slightly modified version of Bauer (2008a).

of growth (cf. Rodriguez and Rodrik, 2001). Theoretical investigations, on the other hand, highlight various specific mechanisms by which trade liberalization may affect growth and/or productivity, but this literature suffers from clear-cut results and hardly produces testable predictions. For example, trade liberalization lowers the real gross domestic product in a typical Heckscher-Ohlin model, but increases the real gross domestic product in models of monopolistic competition. Unfortunately, the key variables in competing models often correspond to different empirical measures of real income or are not observable in the data, thus making it hard to substantiate the findings. Moreover, most recent theoretical papers abstract from consumer durables and capital goods, which account for 32% and 30% of non-energy imports and 16% and 45% of non-energy exports in the U.S., respectively (Erceg, Guerrieri, and Gust, 2008).

In this section, we lay out a specific environment to study how trade affects endogenous R&D in a dynamic model with heterogeneous firms and costly trade. In particular, we set up a model in which growth originates from horizontal specialization and the steady state productivity growth rate is positive. Innovations require costly R&D and are conducted by profit-seeking researchers. This feature is the main difference from the canonical Melitz (2003) model.

Our model accounts for typical characteristics of both growth and trade. First, growth is semi-endogenous and thus does not display a strong scale effect. That is, the steady state productivity growth rate is exogenous, but policy makers may well exert level effects and influence the growth rate along a transition path to the steady state. Second, we account for various firm-level facts uncovered by the empirical trade literature. Most importantly, the distribution of firms' productivities is highly skewed and only the most productive firms export in equilibrium (cf. Aw, Chung, and Roberts, 2000, Bernard and Jensen, 1999, Clerides, Lach, and Tybout, 1998, Pavcnik, 2002, and Tybout, 2003, for a survey). Trade liberalization implies a reallocation of resources towards the more productive firms (cf. Melitz, 2003). Further, there is no feedback effect from exporting to a firm's productivity (Bernard and Jensen, 1999, and Bernard, Jensen and Schott, 2006). The environment laid out below, is suited to allow for both trade in final goods and trade in durables. In this chapter, however, we focus on trade in intermediate goods which are produced from durable physical capital. The production of output uses specialized capital inputs and labor. Traded goods are used to produce both consumption and investment goods. Intermediate firms face endogenous fixed costs for R&D and discover production technologies with heterogenous productivities. When successful, firms enter the local product market at a cost and decide wether or not to export their goods to a foreign market. Technical barriers to trade imply that only the the most productive firms export. International trade is hampered by both variable trade costs and fixed market entry costs. Accounting for the different natures of both types of barriers to trade, we model transportation costs as capital costs and fixed trade costs as labor costs.

The reduced form of the autarky economy resembles the Jones (1995) model. Crucially, however, the productivity in R&D is not exogenous in the absence of knowledge spillovers. In this model with firm heterogeneity and market entry costs, the productivity in R&D is endogenously determined by the amount of labor necessary for market entry and the average R&D cost in the face of a minimum productivity requirement for firms.

In the open economy, we show that including trade in intermediate goods as well as production using physical capital does not alter previous findings on the reallocation of resources and the selection of firms. Similarly, modeling labor intensive technical barriers and capital intensive marginal trading costs is not essential in the baseline specification. In search of the specific mechanisms implied by the monopolistic competition heterogeneous firms models, including physical capital is an informative exercise. In Melitz (2003), trade offers additional profit opportunities only for the most productive firms. With a constant returns to scale technology, the implied market expansion effect increases the scarcity of labor, which is the only factor in production. The increase in the wage rate drives the least productive firms out of the market. In our model, the factor price for intermediate goods producing firms is independent of the exposure to trade. Furthermore, including a factor that can be accumulated potentially allows for a more pronounced impact of trade openness. The model builds on two strands of the literature, namely research on costly trade with heterogeneous firms and non-scale variety growth. We essentially include firms with heterogeneous marginal productivities and costly trade in Jones' (1995) non-scale variety growth model to account for the firm selection effect of trade openness (Bernard, Eaton, Jensen and Kortum 2003, Melitz 2003). Compared to the seminal contribution of Melitz (2003), we model endogenous entry cost and positive long-run productivity growth. Baldwin and Robert-Nicoud (2007) study these two extensions in a fully endogenous growth framework (with scale effects) and with labor as the only factor in production. They find that depending on the specification of the engine of growth, trade is likely to depress the rate of growth because with endogenous R&D, the average R&D costs are likely to increase with the necessary productivity for firms to produce profitably. Using a non-scale R&D technology, Gustafsson and Segerstrom (2007) challenge this view because of the strong knowledge spillovers implicitly assumed in Baldwin and Robert-Nicoud's (2007) analysis. Using a semi-endogenous growth model, they argue that trade only has level effects. In contrast to the Baldwin and Robert-Nicoud (2007) model, more trade makes consumers better off as long as the knowledge spillovers in R&D are not too strong. Both Baldwin and Robert-Nicoud (2007) and Gustafsson and Segerstrom (2007) focus on the effect of trade liberalization on productivity and firm selection, and thus use one factor models and perishable output. A common shortcoming is the lack of a thorough welfare analysis which is due to the complexity of the models' dynamics.

The remainder of the chapter is organized as follows. We first present the closed economy model. After discussing its production structure, we characterize the autarky equilibrium. Section 3.4 introduces international trade. Some qualitative effects of trade liberalization are discussed in Section 3.5. Section 3.6 concludes.

3.3 Model

Following Rivera-Batiz and Romer (1991), the world consists of two identical economies. International trade occurs only in the form of exchanges of intermediate goods. The production structure in each economy is adapted from Jones (1995), where we include heterogeneous firms and market entry costs in the spirit of Hopenhayn (1992a,b) and Melitz (2003).

3.3.1 Overview

Production structure. We explicitly distinguish between three sectors in each economy. The R&D sector invents blueprints for intermediate goods and conducts their market launch. Two manufacturing sectors produce intermediate goods and aggregate output, respectively.[2] Output includes consumption and investment goods.[3] There are three factors in production: labor, raw capital, and knowledge. Raw capital is the investment good, measured in terms of forgone output. The R&D technology requires labor as the only private input, and the existing stock of knowledge can have an external effect on its productivity. Aggregate output is produced from labor and a variety of imperfectly substitutable intermediate goods with additive-separable effects on output. The production of every intermediate good takes a blueprint and raw capital and is conducted by a single intermediate firm.[4] Each blueprint

[2] In what follows, we use the terms "output" and "final good" interchangeably.

[3] We ignore government purchases and there will be no international trade in the final good in the open economy.

[4] We simply take firms to produce exactly one variety and equate firms with their products (i.e. good j is produced by firm j and vice versa). The boundary of intermediate firms is only essential in that we require each firm to have measure zero so that each firm takes the price index of intermediate goods as given.

implies a specific level of productivity that remains constant over time.

Market entry costs. When entering the market, intermediate firms must bear a uniform entry or "beachhead" cost. Market entry is conducted using labor only, hence the entry costs take the form of a wage payment. Newly born firms make a forward looking entry decision based on their productivity. Firms which are sufficiently productive earn sufficiently high profits to cover the fixed entry cost. They therefore actually launch production in the first place and become profitable producers. Less productive firms, however, perceive that the sunk costs exceed their discounted future profits and exit right upon recognizing their productivity.

Costly trade. Each variety faces a positive demand in every country, but international trade is costly. It involves marginal trading costs as well as fixed export costs. The fixed export costs capture the additional costs a foreign company faces when selling to the local market. Importantly, country specific regulations, standards, and similar "technical" obstacles make it more costly for foreign firms to enter the home market then it is for local firms.[5] The key implication of the existence of technical barriers to trade (TBTs for short) is that only the most productive firms self-select into the foreign market and earn additional profits from exporting.

Endogenous growth. Upon investing the entry costs, intermediate firms operate under monopolistic competition and earn positive profits. The prospect of these rents stimulates researchers to invent specialized inputs for the production of output.[6] Introducing new intermediate goods continuously increases the total factor productivity (TFP) and causes growth.

Before we describe the model in greater detail, we briefly contrast the present environment with the Jones (1995) model with homogenous firms, discuss its production structure in the open economy with variable trade costs, and explain how firms with heterogeneous productivities arise from newly discovered blueprints.

3.3.2 Heterogeneous Firms, Trade, and the Jones (1995) Model

Homogeneous firms, durable intermediates. The production structure of the Jones (1995) model is taken from Romer (1990). In Romer (1990), the capital stock comprises a continuum of durable capital goods, which imperfectly substitute in the production of output, with additively separable

[5] See Baldwin (2001) for an illustrative introduction to technical barriers to trade.

[6] Monopolistic competition was introduced to growth theory by Romer (1987).

effects.[7] The capital goods are assembled by intermediate firms. Using $k(j)$ units of the investment good, firm j assembles $x(j) = k(j)$ units of the specialized capital good j. The investment good, "raw capital", is produced from labor and existing durable goods. It is convenient and common practice to assume identical production technologies for the consumption good and the investment good so that the output from both sectors can be summarized as aggregate output which can either be used for investment or for consumption. Romer (1990) already noted that the one-to-one production of intermediate goods from raw capital is merely assumed to keep the model simple. Similarly, uniform production technologies across intermediate firms are typically used only for analytical convenience.

Heterogenous firms. In this research, intermediate firms are heterogeneous with respect to their productivity. We thereby incrementally extend two workhorse models. First, relative to the Jones (1995) model, the average "efficiency" of intermediate firms contributes as a second, "vertical" dimension of productivity to the level of TFP.[8] The range, and along with it the average of firms' productivities in production, is endogenously determined by the degree of trade openness as measured by trade costs. In contrast to growth models with both horizontal and vertical innovations, only the number of varieties increases continuously over time (R&D with heterogeneous firms is addressed in detail in the next but one paragraph). Second, relative to the existing literature on growth and trade with heterogeneous firms, intermediate goods are not only used for consumption, but also for investment. This extension opens up the possibility of a more pronounced impact of trade. Accounting for the accumulation of physical capital, we further add a second factor in production.

Marginal trade costs and the allocation of capital. The presence of marginal trade costs requires a careful modeling of the spatial allocation of physical capital. The production structure of the Jones (1995) model in principle allows two equitable interpretations. The first, classical interpretation (used by Romer, 1990 and Jones, 1995) is that intermediate goods are durable inputs in the production of output. Intermediate good producing firms assemble the durables from raw capital and pass the processed capital on to output producing firms. In this case, capital accumulates at the location of the final good production. In the second interpretation, raw capital is a durable good in the production

[7] Breaking up the capital stock in a continuum of imperfectly substitutable goods allows for positive market rents, which are necessary to cover the innovation costs when production technologies are not strictly convex (see, among others, Romer, 1990).

[8] Li (2000), Young (1998), and Kornprobst (2008, Ch. 9) present models with two R&D sectors and both horizontal and vertical innovations. Sorger (2007) considers quality improving horizontal innovations in a one-sector R&D model, where researchers can influence the quality of their innovations at the cost of a reduced quantity of innovations.

of intermediate goods. Intermediate firms accumulate physical capital to produce perishable inputs for the production of aggregate output. In this case, the capital stock is located at the origin of the intermediate good production.

No trade in durable commodities. In the closed economy, both interpretations are equivalent. As long as there are no variable transportation cost both interpretations are equivalent in the open economy as well. To simplify matters, in what follows, we focus on perishable inputs in the production of durable investment and consumption goods (we stick to the second interpretation above). Since we also rule out trade in aggregate output, there is no accumulation of physical capital by imports.[9] From an empirical point of view, neglecting trade in durable/capital goods appears as a severe shortcut. Erceg, Guerrieri, and Gust (2008) find for the U.S. that consumer durables and capital goods amount to 32% and 30% of non-energy imports, and 16% and 45% of non-energy exports, respectively. In their data, consumer non-durables represent about one-fourth of non-energy imports and exports. The remainder is non-energy industrial supplies used in the production of durables.

Variety expanding R&D and heterogeneous firms. The discovery of blueprints for new intermediate goods is at the heart of our model of growth and trade. A crucial question is how labor and knowledge are transformed into blueprints with heterogeneous productivities. We adapt the modeling in Baldwin and Robert-Nicoud (2007), but use a non-scale technology like Gustafsson and Segerstrom (2007). Following Melitz (2003), the productivity types of blueprints are drawn from a given stationary distribution. The resources necessary to produce a sufficiently valuable blueprint, however, are endogenously determined.

Stochastic productivity draws. While researchers can be certain about finding a new blueprint, its inherent productivity is random. Every research attempt is a costly draw. Due to the entry costs, only blueprints with a sufficiently high productivity (and hence a sufficiently high market value) sell at a positive price. For the sake of clarity, we formally treat R&D and manufacturing as performed in separate sectors. As regards content, we may equivalently combine the two activities for a given variety in "the firm". With a slight abuse of terms, we then also call costly developed blueprints which do not make it into the product market "firms". This gives us a theoretical counterpart to those very low productivity type firms for which the empirical trade literature has identified a high death rate. In the model, these "firms" exit immediately upon recognizing their productivity.

Costly aggregate productivity gains. One of the contributions of Baldwin and Robert-Nicoud (2007) is

[9] The "trade in intermediate goods only" approach follows Rivera-Batiz and Romer (1991).

to incorporate the idea that increasing the productivity of innovations is costly, in the sense that R&D (c.p. and on average) requires more resources if its outcome is to be more productive. In modeling this notion, Baldwin and Robert-Nicoud (2007) look at R&D from an aggregate point of view and consider the average costs associated with the discovery of a marketable blueprint. A potential drawback of this "aggregate R&D" approach is the lack of intentional investments in more productive capital goods. In fact, individual researchers cannot influence the productivities of their innovations. From the individual researcher's perspective, conditional on being usable, high productivity type blueprints are "lucky draws" and as such, they come for free: every draw is equally costly. As will be discussed in more detail in Section 3.3.5, free entry into R&D does not remove the windfall gains associated with high productivities because researchers must break even across usable and unusable innovations in expectations.

A productivity frontier in R&D. As a final remark, note that there is a close analogy between the "aggregate R&D" approach and a productivity-quantity frontier in R&D. That is, an increase in the quality of products will c.p. come at the cost of fewer innovations.[10] Increasing the minimum productivity requirement (again c.p.) forces researchers to move along the technologically given productivity-quantity frontier towards more productive blueprints and fewer innovations. Trade liberalization, as measured by a decrease in the foreign market entry costs, actually raises the minimum productivity requirement, thereby increasing the average productivity of intermediate firms. This productivity gain however is not "manna from heaven" but takes costly resources and implies that the set of intermediate goods at least temporarily expands at a lower rate. Via this channel, the exposure to trade has the potential to slow down productivity gains from specialization. Hence, trade liberalization may at least temporarily depress growth and at the same time have ambiguous effects on TFP.

To begin with, we show how endogenous horizontal innovation and TFP is affected by a minimum productivity requirement in autarky. We then turn to the open economy with international trade in Section 3.4.

[10] Sorger (2007) explicitly includes such a frontier in R&D in a closed economy, free entry model of variety growth. In his model, researchers choose the quality of their innovations optimally, recognizing that higher qualities imply fewer R&D output (cf. footnote 8).

3.3.3 Autarky

The economy is characterized by preferences, endowments, technologies, and a specific institutional environment. As in Romer (1990), Jones (1995), or Rivera-Batiz and Romer (1991), the specific environment laid out below allows a concise exposition and is only one example of an environment that supports the decentralization. The model is set in continuous time and final output is used as the numéraire. We omit the time argument, t, wherever it is not confusing, and occasionally abbreviate variables in the argument of functions by a centered dot (" $\cdot$ ").

Households

The economy is populated by a continuum of mass one of identical households. Every household consists of L homogenous members, who inelastically supply one unit of labor each (there is no disutility from work). The population grows at an exogenously given, constant rate $\dot{L}/L \equiv n \geq 0$, and $L(0) > 0$.[11] The households are infinitely-lived Barrovian (1974) dynasties, where each generation cares about the well-being of all its future offsprings. Every household member consumes an equal amount c of aggregate output Y. The consumption behavior is therefore appropriately summarized by the optimal decision of one household. Preferences are given by a standard intertemporal utility function with constant intertemporal elasticity of substitution in consumption equal to $1/\sigma$ (≥ 0):[12]

$$U = \int_0^\infty e^{-\rho t} u(c(t))dt, \quad u(c(t)) = \frac{c(t)^{1-\sigma} - 1}{1 - \sigma}.$$

ρ (> 0) is the subjective discount rate.

Every household earns income from working and returns on assets and purchases consumption goods and assets. The flow budget constraint is $\dot{\zeta} = wL + r\zeta - cL$, where wL and cL denote the household's labor income and consumption, respectively, and $r\zeta$ is the return on asset holdings ζ at interest rate r. Assets comprise ownership claims on physical and financial capital (loans and debts between households cancel in the representative households' budget constraint). Subsequent assumptions on the observability of firm types and the capital market ensure that physical capital and all types of equity are perfect substitutes as vehicles of savings. They all pay a common rate of return r.

[11] Arnold (1998) replaces population growth with human capital accumulation in a Grossman-Helpman (1991, Ch. 3) framework (without physical capital) and thereby shows explicitly that L can be interpreted more broadly as the effective labor force.

[12] The elasticity of marginal utility is also constant and equals $-\sigma$.

Ponzi-games, where some households borrow infinitely to "repay" consumption loans (and in fact never actually repay their credit), are ruled out by a borrowing constraint imposed in the capital market. Bankers will not lend out more than the present value of a household's income. Hence the present value of consumption expenditures is bounded above by the present value of income. As usual, the appropriate condition is that the present value of assets is asymptotically non-negative, $\lim_{t\to\infty}\left\{\zeta(t)\exp\left[-\int_0^t r(s)ds+nt\right]\right\}\geq 0$.[13]

Technology in Manufacturing

Output. Aggregate output Y is produced using a set of measure A of vertically differentiated intermediate goods j in quantities $x(j)$ and labor L_Y:

$$Y = L_Y^{1-\alpha}\int_0^A x(j)^{\alpha}dj, \quad 0<\alpha<1. \tag{3.1}$$

Output is manufactured by a large number of identical firms (the number of firms is indeterminate because of constant returns to scale for a given level of A).[14] Labor and intermediate goods are complements ($\partial^2 Y/(\partial x \partial L_Y)>0$). The elasticity of substitution between any pair of intermediates is $(1<)\ \epsilon \equiv 1/(1-\alpha)\ (<\infty)$. Given the parameter restriction implicit in (3.1), the intermediate goods have an additively separable effect on output ($\partial^2 Y/[\partial x(j)\partial x(j')]=0$).[15] As usual, the parameter α jointly determines the returns to horizontal specialization in the production of output, the elasticity of substitution between intermediate goods (which indicates the degree of market power of intermediate producers), the price elasticity of demand, and also pins down constant shares of factor incomes in equilibrium.[16]

Intermediates. Every intermediate good is produced from raw capital by an intermediate firm that exclusively owns its blueprint. Each blueprint implies a constant level of productivity in production which

[13]Non-negativity constraints on consumption can be ignored as the instantaneous utility function $u(c)$ satisfies $u'(c)\to\infty$ as $c\to 0$.

[14]The production function in (3.1) of course displays increasing returns in L_Y, all $x(j)$, and A jointly.

[15]Alvarez-Pelaez and Groth (2005) introduce a more general production function $Y = A^{\gamma}X^{\beta}L_Y^{1-\beta}$, $X = A\left[\frac{1}{A}\int_0^A x(j)^{\alpha}dj\right]^{\frac{1}{\alpha}}$ where intermediates can be substitutes ($\alpha>\beta$) or complements ($\alpha<\beta$). We implicitly impose $\gamma=1-\beta$ and $\alpha=\beta$ for simplicity.

[16] It is possible to disentangle the elasticity of output with respect to (horizontal) specialization and the substitutability of capital goods, see Benassy (1998). Alvarez-Pelaez and Groth (2005) also disentangle the degree of substitutability from the capital share, see footnote 15.

carries over to its producer. The firm-level differences in productivities are captured by heterogenous per unit input coefficients $b(j)$:

$$x(j) = \frac{k(j)}{b(j)}, \quad b(j) \in (0, b_0]. \tag{3.2}$$

More productive firms, i.e. firms with low $b(j)$, require less raw capital $k(j)$ to produce one unit of their intermediate good. Unlike in the original Romer model (1990), we treat raw capital as a durable good in the production of perishable intermediate goods. The production and export of the intermediates implies a permanent flow of production and transport costs and simplifies the solution of the open economy model in Section 3.4 below.

Technology in R&D

The presence of entry costs implies that forward looking, profit-driven firms only launch production with blueprints that yield a positive operating profit. Firms' profits are obviously increasing in productivity, which implies that the lowest productivity-type blueprints will be discarded due to the entry costs. If this minimum productivity requirement is binding, the number of intermediate goods (A) is lower than the total number of discovered blueprints (B). In Romer (1990) and Jones (1995), there are no barriers to entry and every discovered blueprint is used to produce a new variety ($A = B$). To tackle this issue, we may think of the R&D technology conceptually as involving two parts. "Research" comprises the process of discovering a previously unknown blueprint. "Development" involves the productivity in production inherent in each blueprint. We consider both parts in turn.

Discovery of blueprints. Researchers deterministically invent new blueprints $\dot{B}$ using Jones' (1995) R&D technology:

$$\dot{B} = \frac{L_B A^{1-\chi}}{F_B}, \quad \chi > 0, \; F_B > 0. \tag{3.3}$$

L_B is the number of people searching for new blueprints, and F_B inversely measures their productivity. Following the common practice in endogenous growth theory, innovation displays constant returns to scale in its only private input, labor. Previous research efforts can have external effects on the magnitude of labor required for innovation, and we follow Baldwin and Robert-Nicoud (2007) in choosing the existing number of intermediate goods (A) to represent the relevant knowledge stock.[17] The exponent $1 - \chi$ accounts for the strength and the sign of the knowledge spillovers. Researchers

[17] Without intentional investments in qualities, B seems equally appropriate as A to represent past innovation efforts.

may either "stand on the shoulders of giants" and benefit from past innovations ($\chi < 1$) or face the "fishing out" of ideas ($\chi > 1$). If $\chi = 1$, there are no spillovers. In this case,

$$\frac{\dot{B}}{B} = \frac{L_B}{BF_B}$$

so that the growth rate of B declines if B increases and L_B is held constant. It then takes positive growth of the labor input to maintain positive long-run growth (which balances the growth of B in the denominator). At $t = 0$, the economy is endowed with a mass $B(0) = B_0$ of blueprints with distribution $G(b)$.

Jones' (1995) R&D technology is intended to eliminate the strong scale effect, i.e. the dependence of the productivity growth rate on the level of labor engaged in R&D in the long run. In doing so, his specification "exogenizes" long-run growth. Suppose $A = B$ and $\dot{L}_B/L_B = n$ (as is the case along a balanced growth path in Jones' model). Then, a constant growth rate of the number of blueprints requires $n - \chi\dot{B}/B = 0$, or $\dot{B}/B = n/\chi$.[18] Thus, growth is semi-endogenous (in that the long run growth rate cannot be influenced by policy) and trade liberalization can "only" exert level effects.

Having described the discovery process, we now turn to the productivity in production that comes along with each blueprint.

Stochastic assignment of productivities. The level of productivity is given by variety-specific input coefficients, which are randomly assigned to each blueprint and revealed after the R&D investment is made (i.e. at the time a blueprint is discovered).[19] The input coefficients are drawn from a distribution which has many low productivity types, fewer intermediate productivity types, and only a few types of very high productivity. To be specific, the input coefficients are drawn from the "mirrored" Pareto distribution

$$G(b) = (b/b_0)^{\theta}, \ b \in [0, b_0], \tag{3.4}$$

where the parameters b_0 (> 0) and $\theta > \max\{\epsilon - 1, 1\}$ govern the width of the support and the shape of the cumulative distribution function, respectively.[20] Figure 3.1 depicts the cumulative distribution function and the density $g(b)$ of input coefficients for $\theta = 2$ (red) and $\theta = 8$ (blue) with $b_0 = 1$.

[18] In the aforementioned Sorger (2007) model with intentional investment in quality, growth also does not display a strong scale effect. In his model, however, policy makers can affect the growth rate if they are able to design quality contingent subsidies (see also Howitt, 1998).

[19] This is analogous to the costly ("black box") draw of productivities in Melitz (2003).

[20] We explicitly deduced this distribution from Pareto-distributed productivity levels in Appendix 2.A.

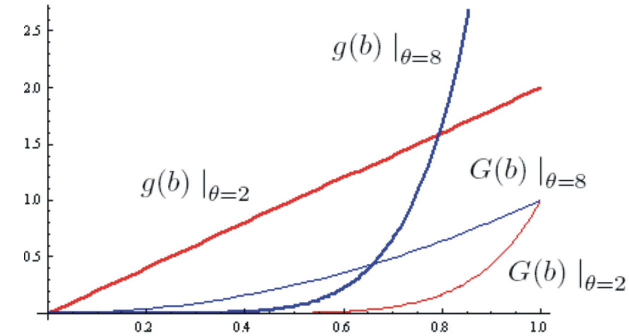


Figure 3.1: Pareto-Distributed Input Coefficients

Imposing a lower bound on θ serves two purposes. First, as will become clear below, $\theta > \epsilon - 1$ ensures that the input coefficient of the least productive firm is strictly positive (so that there is a non-degenerated distribution of firms). Second, it preserves the intended skewness towards low productivity types in case of $\alpha < 0.5$ (in this case, $\theta > \epsilon - 1$ does not imply $\theta > 1$).[21] θ measures the steepness or "dispersion" of the distribution and can therefore be interpreted as the inherent likelihood (or "difficulty") of inventing high productivity types. Increasing θ gives first-order stochastically dominated distributions, i.e. distributions that are more skewed towards high input coefficients ($\theta = 0$ is the uniform distribution and $\theta \to \infty$ yields a degenerate distribution at b_0, in which case $G(b) \to 0$ for all $b < b_0$).[22]

From blueprints to firms. The distribution underlying the productivity types of newly discovered blueprints directly translates into the productivity distribution of firms. This is because the Pareto distribution has the property of scale invariance: truncating a Pareto distribution yields another Pareto distribution with the same shape parameter.[23] As an example, suppose that the cumulative distribution function $G(b)$ is truncated at some minimum productivity $1/b_{trunc}$. The resulting distribution of input coefficients is

$$G(b|b \leq b_{trunc}) = \frac{G(b)}{G(b_{trunc})} = \frac{\left(\frac{b}{b_0}\right)^{\theta}}{\left(\frac{b_{trunc}}{b_0}\right)^{\theta}} = \left(\frac{b}{b_{trunc}}\right)^{\theta}$$

[21] Both Baldwin and Robert-Nicoud (2007) and Gustafsson and Segerstrom (2007) do not impose the second parameter restriction which, however, is only important for the interpretation.

[22] The expected value and variance are $E(b) = \int_0^{b_0} b dG(b) = \int_0^{b_0} \frac{\theta}{b_0^{\theta}} b^{\theta} db = \frac{\theta}{b_0^{\theta}} \left[\frac{b^{\theta+1}}{1+\theta}\right]_0^{b_0} = \frac{\theta}{1+\theta} b_0$ (increasing in θ) and $Var(b) = E(b^2) - [E(b)]^2 = \int_0^{b_0} b^2 dG(b) - \left(\frac{\theta}{\theta+2} b_0\right)^2 = \frac{\theta}{(2+\theta)(1+\theta)^2} b_0^2$ (decreasing in θ).

[23] More generally, the Pareto distribution belongs to the class of power law distributions, which are characterized by the scale invariance property (θ is then consistently called the scaling parameter).

for $b \in [0, b_{trunc}]$. Thus, if some blueprints are not used due to the minimum productivity requirement, the distribution of firms productivities will still remain Pareto, and θ will equivalently reflect the dispersion in the truncated distribution (the support simply shrinks from $[0, b_0]$ to $[0, b_{trunc}]$). Given the shape of the underlying productivity distribution, the distribution of firms' productivities matches the empirical regularity that the fraction of less productive firms is large.

Justifying the Pareto distribution. Like Baldwin and Robert-Nicoud (2007) and Gustafsson and Segerstrom (2007), we specify a functional form to obtain a closed form solution. The Pareto distribution is attractive for two reasons. Firstly, it receives strong empirical support when it comes to matching the observable distribution of productivities, see e.g. Cabral and Mata (2003) and Corcos, Del Gatto, Mion, and Ottaviano (2007). Secondly, as pointed out above, it allows a simple analytical exposition of the distribution of firm types because truncating a Pareto distribution yields another Pareto distribution with the same shape parameter (it is scale invariant).

Markets

The markets for labor, the final good, and financial capital are all perfectly competitive. Producers of capital goods hold infinitely-lived, fully enforced patents. All markets clear. Ownership claims on physical capital and financial wealth are perfect substitutes and pay the same rate of return, r.

Fundamental evaluation. Once a firm's input coefficient is revealed (upon discovery of its blueprint), it immediately becomes common knowledge. We denote by $\pi(j)$ the instantaneous profits of firm j and let

$$v(j) \equiv \int_t^\infty e^{-\bar{r}(s-t)} \pi(j) ds, \tag{3.5}$$

where $\bar{r} \equiv \int_t^s r(\varsigma) d\varsigma$ is the cumulative interest rate up to time $s \geq t$. In the absence of bubbles, and due to the sunk nature of both innovation and entry costs, $v(j)$ is the market value of firm j (with input coefficient $b(j)$). Differentiating (3.5) with respect to time t reveals that, given the definition of $v(j)$ as fundamental value, the returns from investing in any productivity-type of firm, i.e. the dividend payments plus capital gains, have to equal the common return on either asset:

$$\pi(j) + \dot{v}(j) = rv(j) \quad \forall j \in [0, A]. \tag{3.6}$$

Market clearing. Labor market clearing requires that the sum of labor in innovation, market entry, and production is equal to the labor force,

$$L = L_B + L_E + L_Y. \tag{3.7}$$

We further denote by $L_A \equiv L_B + L_E$ the total labor force engaged in the process of R&D and market entry, which we henceforth refer to as R&E (a mnemonic for R&D plus entry).

The stock of raw capital is

$$K \equiv \int_0^A k(j)dj. \tag{3.8}$$

Capital does not depreciate. In Jones (1995), where $b(j) = b = 1$, the sum of intermediate goods equals the amount of accumulated forgone consumption, i.e., the stock of raw capital. Here, with heterogeneously productive firms, the sum of intermediate outputs is proportional to the stock of raw capital and the factor of proportionality equals the output weighted average input coefficient.[24] From (3.2) and (3.8)

$$K = \int_0^A b(j)x(j)dj. \tag{3.9}$$

If intermediate firms become more productive on average, an increased amount of intermediate goods can be obtained from forgoing a given amount of consumption.[25]

Finally, the resource constraint defined over economy-wide aggregates is

$$Y = cL + \dot{K}. \tag{3.10}$$

Market Entry

Launching the production of a newly discovered intermediate good is equally costly to all entrants. To keep the analytical exposition simple, we follow Baldwin and Robert-Nicoud (2007) and assume identical production functions (and thereby "factor intensities") in R&D and the conduct of entry. The productivity in the entry process thereby indicates the markets' "openness". Strictly speaking, the entrant is required to hire $A^{\chi-1}F_L$ workers and pay the associated wage bill $wA^{\chi-1}F_L$. F_L measures the strength of the barriers to entry.[26] To ensure that the input coefficient of the least productive firm

[24]Given mark-up pricing in the intermediate good sector, the output weighted average productivity is closely related to the CES price index. In his one-factor zero-growth model, Melitz (2003, footnote 9) uses such a output-weighted average to measure overall productivity.

[25]As pointed out in the model introduction, assuming that capital can be accumulated as forgone output implies that raw capital is produced with the same technology as the final good. "Forgone consumption" in the above interpretation is thus not actually produced in the first place, but the respective resources are used to produce, i.e. accumulate, raw capital instead.

[26]In Baldwin and Robert-Nicoud (2007) and Gustafsson and Segerstrom (2007), the interpretation of the innovation and entry process is that researchers have to accumulate F_B units of knowledge for inventing a new blueprint and F_L units of knowledge to cope with market entry. Note that this is the Melitz (2003) setting. The one-time initial entry

in equilibrium is strictly smaller than the upper bound of the underlying distribution, b_0 (i.e. that the minimum productivity requirement introduced by the entry cost is binding in equilibrium), we impose a lower bound on F_L:

$$F_B < (\phi - 1) b_0^{\theta} F_L. \tag{PA1}$$

At any point in time, the economy-wide amount of labor devoted to preparing entry is

$$L_E = \dot{A} A^{\chi - 1} F_L. \tag{3.11}$$

Since all productivities are immediately revealed and become common knowledge when the blueprint is discovered, the entry decision involves no uncertainty.[27]

Justifying the entry specification. Four remarks on the specification of entry costs are in order. First, the scaling of entry costs by $A^{\chi-1}$ enables a balanced growth equilibrium with a constant ratio of entry costs and the market value of a new capital good (which, by construction, lies between zero and one). Without resorting to (completely) arbitrary scaling factors, we could alternatively employ $F_L K/A$ or $F_L Y/A$ (and include the use of resources in the respective market clearing/resource condition). Second, identical production functions in R&D and entry turn out to be particularly convenient because they allow a manageable analytical treatment of the free entry into R&D condition. Third, exploiting the block-recursive structure of the Jones (1995) model, identical "factor intensities" in R&D and entry allow simple aggregations of both processes. Fourth, in the open economy, trade is restricted by marginal costs and TBTs. Modeling variable trade costs as iceberg costs implies that they are capital costs. With respect to the nature of TBTs, we assume that overcoming technical obstacles is more labor intensive, and take the extreme standpoint that fixed barriers to trade imply only labor costs.

Having described the environment, we now derive optimality conditions, define the equilibrium, and aggregate over the different types of firms. The subsequent section then characterizes the equilibrium

costs and the periodically occurring fixed overhead costs (summarized in present value terms), correspond to the R&D outlays and the entry costs, respectively. The present interpretation with entry costs, however, simplifies the exposition in the presence of growing firm values.

[27]This timing structure emphasizes the importance of entry cost. If researchers individually knew the productivity of their future innovations, the sunk innovation cost would obviously be sufficient to prevent low productivity types from being invented in the first place.

balanced growth path.

3.3.4 Optimality Conditions

Households and firms maximize their utility and profits, respectively. We consider their decisions in turn.

Households

Optimal behavior of households boils down to choosing a path for consumption. Given a measure $B \geq B_0$ of firms, households are able to pool the risk of investing firms whose type is a priori unknown. Hence, optimal consumption is not affected by the actually prevailing productivity distribution of firms in a household's portfolio or in the economy. Maximizing intertemporal utility subject to the flow budget constraint and the no Ponzi game condition (or, equivalently, to an intertemporal budget constraint that limits the present value of consumption spending to the present value of total income) yields the well-known Euler equation[28]

$$\frac{\dot{c}}{c} = \frac{r - \rho - n}{\sigma} \tag{3.12}$$

[28] If households maximize utility in per capita terms, the present value Hamiltonian is

$$H = e^{-\rho t} u(c) + \lambda (wL + r\zeta - cL),$$

where λ denotes the shadow price of wealth. H is concave in c and ζ, so that the following first-order conditions are sufficient for optimality:

$$\begin{aligned}
\frac{\partial H}{\partial c} &= e^{-\rho t} c^{-\sigma} - \lambda L \stackrel{!}{=} 0, \\
\frac{\partial H}{\partial \zeta} &= r\lambda \stackrel{!}{=} -\dot{\lambda}, \\
\lim_{t \to \infty} \zeta \lambda &= 0.
\end{aligned}$$

Inserting $\lambda = e^{-\rho t} c^{-\sigma}/L$ from the first condition and its time derivative,

$$\begin{aligned}
\dot{\lambda} &= \frac{L\left(-\rho e^{-\rho t} c^{-\sigma} - \sigma e^{-\rho t} c^{-\sigma-1} \dot{c}\right) - e^{-\rho t} c^{-\sigma} \dot{L}}{L^2} \\
&= \frac{e^{-\rho t} c^{-\sigma}}{L} \left(-\rho - n - \sigma \frac{\dot{c}}{c}\right),
\end{aligned}$$

in the second optimality condition yields (3.12). Substituting $\lambda = e^{-\rho t} c^{-\sigma}/L$ and $u'(c) = c^{-\sigma}$ in the third optimality condition, the transversality condition requires that households must not get any utility out assets as $t \to \infty$,

$$\lim_{t \to \infty} \frac{\zeta e^{-\rho t} c^{-\sigma}}{L} = \frac{e^{-\rho t} u'(c)}{L} = 0.$$

and a transversality condition. As usual, the Euler equation gives the rate of consumption growth that optimally relates the subjective discount rate (including household growth) and the market interest rate.

Firms

Profit maximization and competition in the output producing sector imply that the aggregate demand for production workers L_Y and intermediate goods $x(j)$, $j \in [0, A]$, satisfy

$$\begin{aligned} L_Y &= \frac{(1-\alpha)Y}{w}, & (3.13)\\ x(j) &= \left[\frac{\alpha}{p(j)}\right]^{\epsilon} L_Y. & (3.14)\end{aligned}$$

As mentioned earlier, the price elasticity of demand is

$$\frac{\partial x\left(j\right) p\left(j\right)}{\partial p\left(j\right) x\left(j\right)} = -\epsilon.$$

Given the demand function in (3.14), every intermediate goods producer producing firm maximizes its profit $\pi(j)$ by charging a price equal to a constant mark-up over the firm-specific marginal cost (irrespective of the time of invention):

$$p(j) = \frac{rb(j)}{\alpha}, \quad \forall j \in [0, A]. \tag{3.15}$$

Using (3.15) in (3.14), the equilibrium demand and revenues $R\left(j\right) \equiv p\left(j\right) x\left(j\right)$ of firm j with input coefficient $b(j)$ are

$$\begin{aligned} x(j) &= \alpha^{2\epsilon} \left[rb(j)\right]^{-\epsilon} L_Y, & (3.16)\\ R(j) &= \alpha^{2\epsilon-1} \left[rb(j)\right]^{1-\epsilon} L_Y, \quad \forall j \in [0, A]. & (3.17)\end{aligned}$$

From (3.15), profits amount to

$$\pi(j) = (1-\alpha)R(j), \quad \forall j \in [0, A]. \tag{3.18}$$

Obviously, profits are increasing in productivity $1/b$. From (3.18),

$$\frac{\partial \pi\left(b, \cdot\right)}{\partial \left(\frac{1}{b}\right)} = \left(1-\alpha\right) \alpha^{2\epsilon-1} \left(\epsilon - 1\right) \left(\frac{1}{b}\right)^{\epsilon-2} \left(\frac{1}{r}\right)^{\epsilon-1} L_Y,$$

which implies that profits are convex (concave) in productivity if $\epsilon > 2$ ($\epsilon < 2$), i.e. $\alpha >$ ($<$) $1/2$.

Gains from increasing degrees of specialization with imperfectly substitutable intermediate goods limits a complete allocation of resources towards the most productive firms. In particular, the market entry costs are the necessary ingredient to prevent the least productive firms from operating: There is always a positive demand for any variety as long as any output is produced ($L_Y > 0$), and mark-up pricing guarantees positive operating profits for firms of all productivity types. In the absence of barriers to entry ($F_L = 0$), all firms launch production, $b_L^* = b_0$, so that $A = B$.

No durable goods monopoly problem. Note that our interpretation of the production structure with durable goods in the intermediate rather than the final good sector naturally avoids the usual "durable goods monopoly problem". When monopolists actually sell durable goods, tomorrow's demand is a close substitute to today's demand, and firms with market power account for the fact that today's sales come at the expense of tomorrow's sales. Tirole (1988, Section 1.5) shows that monopolists then have an incentive to increase today's quantities at the expense of tomorrow's demand and do so in the absence of commitment to output quantities. Romer (1990) points out that in his model environment, selling durable goods to the final good sector potentially results in a more complicated pricing problem than the "static" program stated above. To avoid this complication, Romer (1990, in a closed economy) and Rivera-Batiz and Romer (1991, in an open economy) formally assume that the durable goods are rented. In our interpretation, the problem is resolved since there is no monopolistic supplier of the investment good.

From goods to productivities. In this environment, the intermediate firms' prices, quantities, profits, and firm values differ only due to heterogeneous productivities. As of this point, it is thus reasonable to drop the firm index j and phrase the equilibrium expressions in terms of productivity types b, i.e. from (3.15) and (3.16),

$$p(b,\cdot) = \frac{rb}{\alpha}, \quad x(b,\cdot) = \alpha^{2\epsilon}(rb)^{-\epsilon}L_Y, \tag{3.19}$$

and from (3.18) and the definition of the firm value in (3.5),

$$\pi(b,\cdot) = (1-\alpha)\alpha^{2\epsilon-1}(rb)^{1-\epsilon}L_Y, \quad v(b,\cdot) = \int_t^\infty e^{-\bar{r}(s-t)}\pi(b,\cdot)ds. \tag{3.20}$$

Similarly, the time derivatives of the firm values in (3.6) simplify to

$$rv(b,\cdot) = \pi(b,\cdot) + \dot{v}(b,\cdot). \tag{3.21}$$

Understanding firm heterogeneity. To improve our understanding of firm heterogeneity in this production environment, consider a firm with input coefficient b that is more efficient than another firm with input coefficient $b' \geq b$. From (3.19), we find that relative output is

$$\frac{x(b,\cdot)}{x(b',\cdot)} = \left(\frac{b}{b'}\right)^{-\epsilon} = \left(\frac{b'}{b}\right)^{\epsilon} \quad (\geq 1).$$

Similarly, the relative input requirement in production is

$$\frac{bx(b,\cdot)}{b'x(b',\cdot)} = \left(\frac{b}{b'}\right)^{1-\epsilon} = \left(\frac{b'}{b}\right)^{\epsilon-1} \quad (\geq 1).$$

The relative output and input quantities are thus independent of endogenous variables, and the only parameter that has an impact at all is α. Finally,

$$\frac{v(b,\cdot)}{v(b',\cdot)} = \frac{\pi(b,\cdot)}{\pi(b',\cdot)} = \frac{R(b,\cdot)}{R(b',\cdot)} = \left(\frac{b}{b'}\right)^{1-\epsilon} = \left(\frac{b'}{b}\right)^{\epsilon-1} \quad (\geq 1).$$

The second equality holds because of our assumptions on the fundamental evaluation at the capital market above. Since the input coefficients are constant over time, the profits of firms of all productivity types and hence their market values grow at equal rates (b cancels from the last term because it can be pulled out of the integral):

$$\frac{\dot{v}\,(b,\cdot)}{v(b,\cdot)} = r - \frac{\pi(b,\cdot)}{v(b,\cdot)} = r - \frac{\pi(b,\cdot)}{\int_t^\infty e^{-\bar{r}(s-t)}\pi(b,\cdot)ds}, \tag{3.22}$$

Hence, $\hat{v}\,(b,\cdot) = \hat{v}\,(j) = \hat{v}$ so that the dividend ratio is identical across firms of all productivity types. In equilibrium, firms with a higher productivity sell higher quantities, demand more raw capital (as the lower input coefficient is offset by the rise in total demand), receive higher profits, and have a higher market value. An increase in α amplifies the differences. Figure 3.2 depicts a firm's profit and its market value as a function of its productivity for $\epsilon < 2$.[29] We summarize these findings in

Result 3.1 (**Productivity and firm size**). *In equilibrium, more efficient firms are larger: they produce more output and use more raw capital than less efficient firms. Profits and firm values are increasing and concave (convex) in the firm's productivity if* $\epsilon < (>) 2$.

Obviously, higher input prices (an increase in r), and less demand from the final good sector (a decline in L_Y) c.p. imply smaller profits. Clearly also, the profits of more efficient firms react stronger to such changes in absolute terms (here exemplarily for r):

$$\frac{\partial \pi(j)/\partial r}{\partial \pi(j')/\partial r} = \left(\frac{b'(j)}{b(j)}\right)^{\epsilon-1} \quad (> 1).$$

[29] Differentiating equilibrium profits with respect to b yields $\frac{\partial \pi(b,\cdot)}{\partial b} = \epsilon(\epsilon-1)\frac{\pi(b,\cdot)}{b^2} > 0$. This immediately gives a firm value function that is of the same shape, since the dividend ratio is the same for all productivity type firms which implies that the ratio of the slopes of v and π is identical across productivity types.

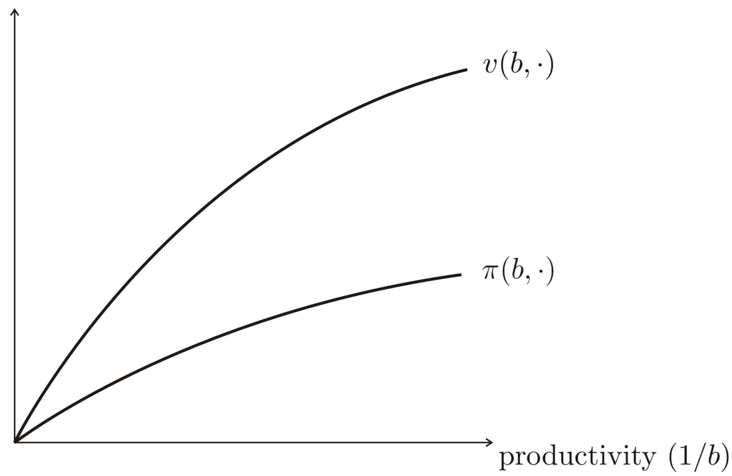


Figure 3.2: The Firm's Profit and Market Value as a Function of Productivity ($\epsilon < 2$)

Profits and α. The relation between firms' profits and the parameter α deserves a short comment. As pointed out above, changing α has multiple implications, and it also captures opposing effects on intermediate firms' profits. On the one hand, like in the canonical trade models with love of variety preferences, a low degree of substitutability between the differentiated final good inputs (a low α) allows the monopolists to charge a high mark-up $1/\alpha$, and (as demand is inelastic) earn high revenues and high profits. On the other hand, α also measures the capital share in the production of final output. Hence, a small α also presumes less demand for capital goods from producers of output goods. Using standard parameters, the latter effect prevails and profits are increasing in α.[30]

Entry

Let us return to the entry decision of the firm. The imposed upper bound on F_B restricts the analysis to the case where F_L (or b_0) is sufficiently "large" so that the entry costs exceed the market value of the least productive firms (i.e. $v(b_0, \cdot) < wA^{1-\chi}F_L$ holds in equilibrium by assumption). Thus, only sufficiently productive firms are willing to bear the entry cost. Given market prices, the cutoff productivity associated with profitable entry, $1/b_L$, is determined by

$$v(b_L, \cdot) \equiv wA^{\chi-1}F_L. \tag{3.23}$$

Equation (3.23), the zero cutoff profit condition, is illustrated in Figure 3.3.[31] Firms with a productivity below $1/b_L$ will not incur the entry costs and "die" instantaneously. More productive firms incur

[30] In fact, the sign of the net effect actually depends on the size of the input coefficient. From (3.20),

$$\ln \pi (b, \cdot) = \ln (1-\alpha) + (2\epsilon - 1) \ln \alpha + (1-\epsilon) \ln (rb) + \ln L_Y,$$

and hence

$$\frac{\partial \ln \pi (b, \cdot)}{\partial \alpha} = \frac{1}{\alpha - 1} + \frac{2\epsilon - 1}{\alpha} + 2\frac{\partial \epsilon}{\partial \alpha} \ln \alpha - \frac{\partial \epsilon}{\partial \alpha} \ln (rb).$$

After collecting terms and inserting $\partial \epsilon / \partial \alpha = \epsilon^2$,

$$\frac{\partial \ln \pi (b, \cdot)}{\partial \alpha} = \frac{\epsilon}{\alpha} + \epsilon^2 \left[2 \ln \alpha - \ln (rb)\right].$$

Hence, profits are increasing in α if

$$\ln \frac{\alpha^2}{rb} > -\frac{1}{\alpha\epsilon},$$

or, using $-\alpha\epsilon = -\alpha/(\alpha - 1)$,

$$\frac{\alpha^2}{rb} > e^{-\frac{1-\alpha}{\alpha}}.$$

Increasing α raises profits if $b < \bar{b}$, and lowers profits if $b > \bar{b}$, where

$$\bar{b} = \frac{\alpha^2 e^{\frac{1-\alpha}{\alpha}}}{r}.$$

For more productive firms (with $b < \bar{b}$), profits increase in α, since for them the positive effect of a high final good demand outweighs the negative effect due to a low mark-up. For less productive firms (those with input coefficient $b > \bar{b}$), the increase in demand is not sufficiently strong to outweigh the profit decreasing effect of a lower mark up. Since α captures opposing effects, comparative statics with respect to α are not unambiguous.

[31] In Melitz (2003), $w = 1$ and $\chi = 1$. We verify below that $\hat{w} + (\chi - 1)\hat{A} = \hat{v}$ in the equilibrium along the balanced growth path so that the snapshot above in fact illustrates a stationary cutoff.

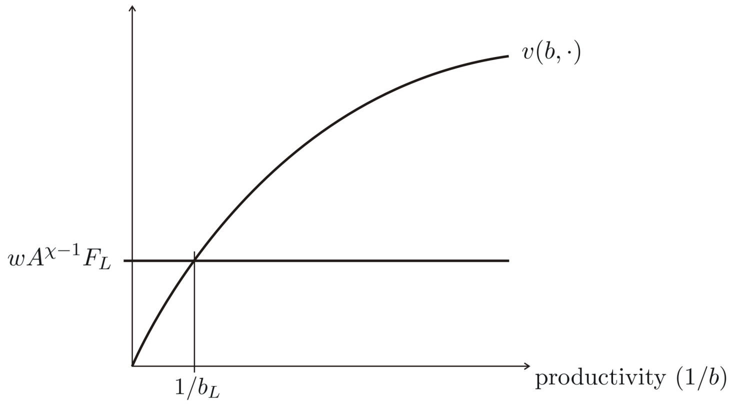


Figure 3.3: The Cutoff Productivity in Autarky

the costs and launch production. Due to the scale invariant nature of the Pareto distribution, whereby truncating the distribution maintains both the type of the distribution and its shape parameter, all information about the equilibrium distribution of firms' productivities is contained in the cutoff productivity (for example, b_L easily translates into the output weighted average productivity). We explore this convenient feature further in the next section.

A law of motion for A. A binding cutoff ($b_L < b_0$) implies that researchers can only sell sufficiently productive blueprints to profit-seeking manufacturers. Given a continuum of newly discovered blueprints at any point in time, we rely on a law of large numbers and conclude that the fraction of profitable blueprints is $G(b_L)$. Hence, the evolution of A given that $\dot{b}_L = 0$ is governed by

$$\dot{A} = G(b_L)\dot{B}. \tag{3.24}$$

Since only a fraction $G(b_L) < 1$ of newly discovered blueprint will actually go into production (and increase the specialization in the production of aggregate output), an increase in the minimum productivity requirement c.p. depresses the dynamic gains from horizontal specialization.

Labor allocation in R&E. In view of (3.24), let us clarify the allocation of labor between R&D and market entry. By construction, the ratio of labor in R&D to labor in entry is fixed for a given cutoff. From (3.3), (3.11), and (3.24),

$$\frac{L_B}{L_E} = \frac{F_B}{G(b_L)F_L}. \tag{3.25}$$

Every newly invented intermediate good requires F_L (times $A^{\chi-1}$) workers to realize its market entry and, on average, it takes $F_B/G(b_L)$ (times $A^{\chi-1}$) workers to discover a producible blueprint. Labor

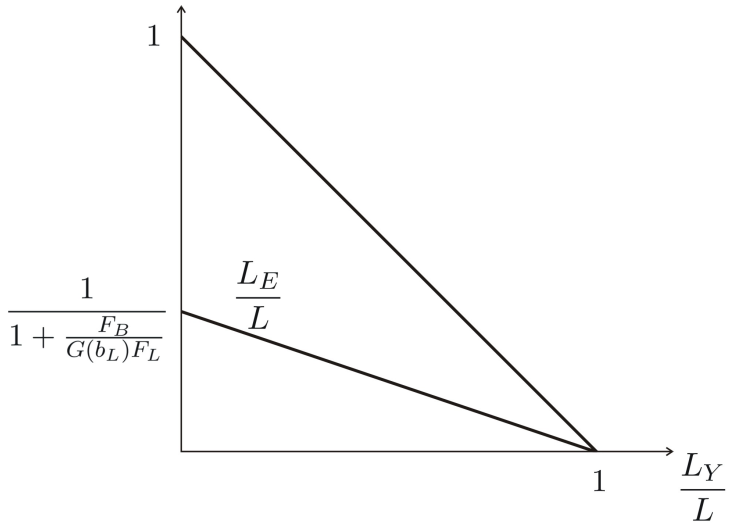


Figure 3.4: Labor Shares in R&D and Entry Against the Labor Share in Production for a Given Cutoff

market clearing requires that the labor shares in entry, R&D, and production sum up to unity. Using this relation to replace L_B, and solving for the share of labor in the conduct of entry yields

$$\frac{L_E}{L} = \frac{1}{1 + \frac{F_B}{G(b_L)F_L}} \left(1 - \frac{L_Y}{L}\right). \tag{3.26}$$

Figure 3.4 shows the labor shares in R&E as a function of the labor share in the production of output for a given cutoff productivity $1/b_L$. The upper line depicts the labor market clearing condition as a function of the share of labor in production,

$$\frac{L_A}{L} = 1 - \frac{L_Y}{L}.$$

The lower line corresponds to the allocation of labor between entry and R&D, i.e. to equation (3.26). Of course, the horizontal distance between the two lines is the share of labor in R&D, L_B/L, since the labor market clearing line has slope -1. Suppose that the share of labor in production is not affected by the productivity distribution of intermediate firms (which we shall prove later on in Corollary 3.8). Then, for a given cutoff, $L_E/L(L_Y/L)$ simply centers around $L_Y/L = 1$ as F_L changes. We will return to this property after having characterized the equilibrium cutoff.

Free entry into R&D. In an equilibrium with free entry into R&D, the expected operating value net of market entry costs must at most outweigh the innovation cost. If $\dot{A} > 0$, we thus have

$$\int_0^{b_L} \left[v(b, \cdot) - wA^{\chi-1}F_L\right] dG(b) = wA^{\chi-1}F_B. \tag{3.27}$$

If the expected net return to R&D (the left hand side), i.e. the market value of a capital good net of the entry cost (the term in squared brackets on the left hand side), exceeds the R&D cost (the right hand side), more researchers enter and discover a higher number of blueprints, thereby driving down the value of innovations. Similarly, if the expected net returns to R&D are not sufficient to cover the R&D cost, researchers leave and become production workers, thereby reducing the number of innovations and increasing the market value of innovations. Hence, the expected return to R&D must equal the total innovation costs; from (3.27),

$$\int_0^{b_L} v\left(b,\cdot\right) dG\left(b\right) = wA^{\chi-1}\left[F_B + G\left(b_L\right) F_L\right]. \tag{3.28}$$

In the absence of knowledge spillovers ($\chi = 0$), this free entry condition is again identical to Melitz (2003). Finally, we define an equilibrium in this economy.

Definition 3.1 (Equilibrium). *An equilibrium is a path of quantities* $c, L_A, L_E, L_Y, Y, K, A, B, \{x(j), k(j)\}_{j\in[0,A]}$, *prices* $r, w, \{p(j), \pi(j), v(j)\}_{j\in[0,A]}$, *and the cutoff productivity* b_L *that satisfies technologies (3.1), (3.2), (3.3), (3.11), and (3.24), the entry conditions (3.23) and (3.27), the optimality conditions (3.12), (3.13), (3.14), and (3.15), the resource constraints (3.7) and (3.10), as well as the definitions of* π, v, *and* K.[32]

3.3.5 Aggregation for a Given Cutoff

We derive the equilibrium outcome in aggregate terms in two steps. First, we aggregate over all productivity type firms for a given level of the cutoff productivity. In a second step, we solve for the cutoff and characterize the equilibrium.

Suppose for the time being that the cutoff productivity $1/b_L$ is initially given and constant. Since all entrants are required to pay the entry costs, the productivity distribution in the product market,

[32] As usual in general equilibrium theory, the households' budget constraint is another, but dependent, equation in the same variables.

denoted by $\mu(b; b_L)$, is the productivity distribution of blueprints, $G(b)$, conditional on entry:[33]

$$\mu(b; b_L) \equiv \frac{G(b)}{G(b_L)} = \left(\frac{b}{b_L}\right)^{\theta}, \ b \in [0, b_L]. \tag{3.29}$$

Using $\mu(b)$, it is an easy task to aggregate over all active firm types.[34] Intuitively speaking, the probability density function $\mu'(b)$ gives the mass of firms for each level of productivity, relative to the total mass of active firms, A. The "number" of firms with the same level of productivity hence equals $A\mu'(b)$ for each productivity level b. Taking into account that only firms with productivities above the cutoff productivity incur the entry cost, integrating over all active productivity levels $b \leq b_L$ then gives the aggregate intermediate outcome. To begin with, consider the capital stock in (3.9). Instead of aggregating over the raw capital inputs $k(j) = b(j)\,x(j)$ of all firms $j \in [0, A]$, we equivalently aggregate over all active productivity types $b \in [0, b_L]$, taking into account that there is a mass $A\mu'(b)$ of firms per level of productivity:

$$K = \int_0^A b(j)x(j)dj = \int_0^{b_L} bx(b, \cdot)\, A\mu'(b)\, db.$$

Now, using the conventional notation $d\mu(b) = \mu'(b)\, db$ and the equilibrium quantities from (3.16),

$$K = A \int_0^{b_L} b\alpha^{2\epsilon}(rb)^{-\epsilon} L_Y d\mu(b).$$

After inserting

$$d\mu(b) = \frac{\theta b^{\theta-1}}{b_L^{\theta}} db \tag{3.30}$$

[33] In a setup with a random positive death rate for active firms and a large pool of potential entrants, Melitz (2003) shows that the long-run equilibrium distribution of active firms is $\tilde{G}(b)/\tilde{G}(b_L)$ if the universe of productivities is described by a more general class of probability distributions $\tilde{G}(b)$. The random death of firms of all productivity types is needed for the distribution of active productivity types to converge back to $\tilde{G}(b)/\tilde{G}(b_L)$ after a shock to b_L. In our environment, where A grows at a positive rate, the transition between two distributions of active firms' productivity types with different cutoffs is naturally achieved as the share of those productivities that are no longer introduced goes to zero in finite time. This is equivalent to randomized firm death, which steadily brings the productivity distribution of active firms back to the productivity distribution of newcomers whenever this distribution remains constant over time. To simplify the exposition, we drop the dependency of the active firms' productivity distribution's support on the cutoff whenever doing so does not lead to confusion.

[34] When aggregating over all firm types, we choose to express the outcome of the aggregation by equilibrium quantities of the cutoff productivity type firm. Following Melitz (2003), Baldwin and Robert-Nicoud (2007), and Gustafsson and Segerstrom (2007) we could alternatively apply an output-weighted average productivity type firm. Our choice, which of course is as good as any other productivity type, is motivated by the fact that the aggregate outcome in terms of the cutoff productivity makes the basic mechanism of the model visible quite well.

from (3.29) and integrating, the capital stock equals

$$K = \frac{A\alpha^{2\epsilon} r^{-\epsilon} L_Y \theta}{b_L^{\theta}} \int_0^{b_L} b^{\theta-\epsilon} db = \frac{A\alpha^{2\epsilon} r^{-\epsilon} L_Y \theta}{b_L^{\theta}} \left[\frac{b^{\theta-\epsilon+1}}{\theta-\epsilon+1}\right]_0^{b_L} = A\alpha^{2\epsilon} r^{-\epsilon} L_Y \phi b_L^{1-\epsilon} \tag{3.31}$$

where $\phi \equiv \theta/(\theta - \epsilon + 1)$ (> 1).[35] To ease the exposition, use (3.16) again:

$$K = \phi A b_L x(b_L, \cdot). \tag{3.32}$$

The average productivity. The average output weighted productivity $\bar{b}$ is defined by

$$K = \bar{b} \int_0^A x(j) dj = \bar{b} A \int_0^{b_L} x(b, \cdot) d\mu(b).$$

Applying (3.30), inserting $x(b, \cdot)$ from (3.19), and integrating we have

$$K = \frac{\bar{b} A\alpha^{2\epsilon} \theta r^{-\epsilon}}{b_L^{\theta}} \int_0^{b_L} b^{\theta-1-\epsilon} db = \frac{\bar{b}\theta A\alpha^{2\epsilon} (r b_L)^{-\epsilon} L_Y}{\theta - \epsilon}.$$

Accordingly, using (3.19) and (3.32),

$$K = \frac{\theta A \bar{b} x(b_L, \cdot)}{\theta - \epsilon} = \frac{\theta A b_L x(b_L, \cdot)}{\theta - \epsilon + 1}, \tag{3.33}$$

so that

$$\bar{b} = \frac{b_L}{1 + \frac{1}{\theta - \epsilon}}. \tag{3.34}$$

For a given amount of accumulated savings, the output of intermediate firms is obviously larger, the more efficiently resources are transformed into intermediate goods, i.e. the smaller $\bar{b}$. Of course, with a Pareto distribution, the output-weighted average input coefficient increases with the input coefficient of the least productive firm. Comparing the output-weighted average, $\bar{b} = (\theta - \epsilon)/(\theta + 1 - \epsilon) b_L$, to the unweighted average which corresponds to symmetric varieties,

$$\int_0^{b_L} b d\mu = \frac{\theta \int_0^{b_L} b^{\theta} db}{b_L^{\theta}} = \frac{b_L}{1 + \frac{1}{\theta}},$$

confirms the intuition that the difference in firms' output is more pronounced, the lower the degree of substitutability between intermediate goods (i.e. as ϵ is increasing). That is, competition in the product market (measured by the degree of substitutability between intermediate goods) works against the variance reducing effect of fixed cost.[36]

[35] $\phi = \frac{\theta}{\theta - (\epsilon - 1)} > 1$ since $\theta > \epsilon - 1$ by (PA1) and $\epsilon > 1$.

[36] Note that this conclusion is again ambiguous due to the fact that α also measures the share of capital income and the gains from specialization.

Aggregate profits and firm values. Turning to firm's market values, we first derive the aggregate intermediate producers' profits. Using (3.15) and (3.31),

$$\int_0^A \pi(j)\,dj = (1-\alpha)\int_0^A p(j)\,x(j)\,dj = \frac{(1-\alpha)\,r}{\alpha}\int_0^A b(j)\,x(j)\,dj = (1-\alpha)\,\alpha^{2\epsilon-1}\phi A\,(rb_L)^{1-\epsilon}\,L_Y.$$

From (3.17), we have

$$\int_0^A \pi(j)dj = (1-\alpha)A\phi R(b_L,\cdot) = A\phi\pi(b_L,\cdot). \tag{3.35}$$

The average profit is thus ϕ times the profit of firms operating with the cutoff productivity, $\phi\pi(b_L) = \int_0^A \pi(j)dj/A$. Using (3.6), the same is true for the cutoff productivity type firm value. From $\pi(j) = v(j)(r-\hat{v})$ and (3.35), we find

$$\int_0^A v(j)dj = A\phi v(b_L,\cdot). \tag{3.36}$$

The difference in the market value of firms with the cutoff productivity and the average productivity is larger, the larger ϕ. ϕ accounts for the characteristics of the underlying distribution of productivities (as summarized by θ and b_0) and includes α as an indicator of the value of productivity.[37] Consistent with the previous observation on relative profits, the value of average productivity type firms is low relative to the value of firms operating with the cutoff productivity if α is small (ϕ is larger, the larger α). A large α implies a high level of all firms' values[38], and more unevenly distributed profits. Put differently, the dispersion in the values of firms with different productivity levels depends positively on α ($\alpha \to 0$ implies $\phi \to 1$ and $v(b,\cdot) \to v(b_L,\cdot)$).[39]

Next, aggregating over the intermediate firm's outputs in (3.1) using the equilibrium quantities from (3.19) and $d\mu(b)$ from (3.30), the production function for aggregate output can be rewritten as

$$Y = L_Y^{1-\alpha}\int_0^A x(j)^\alpha dj = L_Y^{1-\alpha}A\alpha^{2\epsilon\alpha}r^{1-\epsilon}L_Y^\alpha\int_0^{b_L} b^{1-\epsilon}d\mu(b) = L_Y^{1-\alpha}\phi A[\alpha^{2\epsilon}(rb_L)^{-\epsilon}L_Y]^\alpha,$$

or, using (3.16) again,

$$Y = AL_Y^{1-\alpha}\phi x(b_L,\cdot)^\alpha. \tag{3.37}$$

[37] Of course, a higher average productivity implies a higher distance of the average to the cutoff, so ϕ increases with the breadth and dispersion of the underlying distribution, b_0 and θ.

[38] This is because profits are higher the larger α, see the paragraph below equation (3.17).

[39] The latter observation is easily verified by looking at relative firm values. For $b < b_L$,

$$\frac{d\left[\frac{v(b,\cdot)}{v(b_L,\cdot)}\right]}{d\alpha} = \frac{d\left[\frac{\pi(b,\cdot)}{\pi(b_L,\cdot)}\right]}{d\alpha} = \frac{d\left(\frac{b_L}{b}\right)^{\frac{\alpha}{1-\alpha}}}{d\alpha} = \left[\frac{1}{1-\alpha}+\frac{\alpha}{(1-\alpha)^2}\right]\left(\frac{b_L}{b}\right)^{\frac{\alpha}{1-\alpha}}\log\left(\frac{b_L}{b}\right) > 0.$$

Replacing $x(b_L)$ using (3.32) we find

$$Y = (\phi A L_Y)^{1-\alpha} \left(\frac{K}{b_L}\right)^{\alpha} = \phi (A L_Y)^{1-\alpha} \left(\frac{K}{\phi b_L}\right)^{\alpha}. \tag{3.38}$$

This expression demonstrates quite clearly the close analogy to a representative firm model, where $Y = (AL_Y)^{1-\alpha} K^{\alpha}$. In the present environment with heterogeneous firms, the (endogenously increasing) degree of specialization is complemented by the (static) average input coefficient that describes how efficient capital goods can be manufactured to produce output. A degenerate one point distribution at $b = b_L$ ($\theta \to \infty$) implies $\phi \to 1$.[40]

The output-capital ratio. We already know from the household's optimal consumption decision, that the savings behavior is not affected by the productivity distribution of intermediate firms. Hence, the output-capital ratio should be independent of the distribution of firm productivities. From (3.38),

$$\frac{Y}{K} = \frac{(\phi A L_Y)^{1-\alpha} K^{\alpha-1}}{b_L^{\alpha}},$$

and using (3.31),

$$K^{\alpha-1} = A^{\alpha-1} \alpha^{2\epsilon(\alpha-1)} r^{-\epsilon(\alpha-1)} L_Y^{\alpha-1} \phi^{\alpha-1} b_L^{(1-\epsilon)(\alpha-1)}.$$

By definition, $\epsilon(\alpha - 1) = -1$ and $(1-\epsilon)(\alpha-1) = \alpha$ so that

$$\frac{Y}{K} = \frac{r}{\alpha^2}. \tag{3.39}$$

We thus note:

Result 3.2 (Output-capital ratio). *The output-capital ratio depends positively on the interest rate. Unless entry costs have an impact on the interest rate, the output-capital ratio is independent of barriers to entry.*

The evolution of A for a given cutoff. Given b_L, a compact law of motion for A is readily obtained by combining (3.24), (3.3), and (3.11). From the first two equations,

$$\dot{A} = \frac{L_B A^{1-\chi} G(b_L)}{F_B} = \frac{(L_A - L_E) A^{1-\chi} G(b_L)}{F_B}.$$

Inserting L_E from (3.11) yields

$$\dot{A} = \frac{(L_A - \dot{A} A^{\chi-1} F_L) A^{1-\chi} G(b_L)}{F_B}.$$

[40] For $\theta \to \infty$, $\mu(b_L) = 0$ for all $b < b_L$ and $\mu(b_L) = 1$ for $b = b_L$. At the same time, $\lim_{\theta\to\infty} \frac{1}{1-\frac{\epsilon-1}{\theta}} = 1$.

Solving for $\dot{A}$, the R&E process is described by a standard Jones (1995) R&D technology:

$$\dot{A} = \frac{L_A A^{1-\chi}}{F^a(b_L)}, \quad F^a(b_L) \equiv F_L + \frac{F_B}{G(b_L)}. \tag{3.40}$$

Compared to Jones (1995), the innovation technology is augmented in two aspects. Firstly, without entry costs, all research attempts are successful. Here, the R&E productivity $(1/F^a(b_L))$ decreases endogenously with b_L because some innovations must be discarded. $F^a(b_L) = F_L + F_B/G(b_L)$ captures the effect of entry costs and the implied minimum productivity requirement on the R&D productivity in terms of output quantities. Given A, discovering and launching production for a new intermediate good is obviously less labor intensive if the minimum productivity requirement is low, or easy to meet (because θ is high so that $G(b_L)$ is high), and if few workers are necessary to conduct market entry (i.e. if F_L is low). Note that without entry cost (more precisely, with F_L violating (PA1)), there is no need to dispense with low productivity types. Here, in contrast, it takes $1/G(b_L)$ times more resources on average to discover a usable blueprint. Secondly, entry is modeled in such a way that it takes workers away from R&D. This further increases the labor requirement necessary for usable blueprints.

Free entry in R&D for a given cutoff. Diving the free entry into innovation condition in (3.28) by $G(b_L)$ gives

$$\int_0^{b_L} v(b,\cdot)\frac{dG(b)}{G(b_L)} = wA^{\chi-1}F^a(b_L).$$

After recognizing

$$\frac{dG(b)}{G(b_L)} = \frac{G'(b)\,db}{G(b_L)} = \mu'(b)\,db = d\mu(b),$$

and substituting

$$v(b,\cdot) = \frac{\pi(b,\cdot)}{\left[r - \frac{\dot{v}(b,\cdot)}{v(b,\cdot)}\right]} \tag{3.41}$$

from (3.6) in terms of input coefficients (see (3.20)) we get

$$\int_0^{b_L}\left[\frac{\pi(b,\cdot)}{r - \frac{\dot{v}(b,\cdot)}{v(b,\cdot)}}\right]d\mu(b) = wA^{\chi-1}F^a(b_L).$$

As shown before, $\hat{v}(b,\cdot) = \hat{v}$, hence $r - \hat{v}$ is independent of b and can be pulled out of the integral. Moreover, since

$$\int_0^{b_L}\pi(b,\cdot)d\mu(b) = \frac{\int_0^A \pi(j)dj}{A},$$

we can replace the remaining integral term, the average profits, with the expression implied by (3.35), i.e. $\phi\pi(b_L,\cdot)$. Finally, using (3.41), the free entry into R&D condition becomes

$$\phi v(b_L,\cdot) = wA^{\chi-1}F^a(b_L). \tag{3.42}$$

The right hand side equals the average development costs of an actually producible durable good: a newly discovered blueprint requires F_L times $A^{\chi-1}$ workers to conduct its market entry and it takes $F_B/G(b_L)$ times $A^{\chi-1}$ workers on average to discover a producible blueprint in the first place (researchers on average must "draw" $1/G(b_L)$ times to find a sufficiently productive type). In endogenenous growth models with free entry into innovation and costless entry into the product market, the R&D costs of every undertaken research project must equal its costs in an equilibrium with positive growth. In the present environment, however, researchers face uncertainty about the productivity and thus the market value of their innovations. In particular, blueprints with a productivity below the cut-off will not earn their R&D costs. Hence, in equilibrium, sucessful innovations must earn excess rents. Inserting the definition of F^a from (3.40) in (3.42) and solving for the R&D costs of a single discovery shows this most clearly:

$$wA^{\chi-1}F_B = G(b_L)\left[\phi v\left(b_L,\cdot\right) - wA^{\chi-1}F_L\right] < \phi v\left(b_L,\cdot\right) - wA^{\chi-1}F_L.$$

Given that the cutoff is binding, i.e. $G(b_L) < 1$, the average net value of entry (the right hand side of the inequality) exceeds the actual R&D costs of a single innovation to ensures that research investments break even across all undertaken projects. More generally, if there is a positive probability that research projects fail, the ex post return on sucessful projects must exceed one, to ensure free entry ex ante. Since this feature is the main difference between the heterogeneous firms and entry costs models and the canonical growth models, we explicitly state it in

Result 3.3 (**Excess rents for innovators**). *The average net value of entry exceeds the innovation cost.*

Return on investment in R&D. To avoid confusion, we explicitly state that ex ante zero profits free entry in R&D imply that the ratio of the average firm value to the average R&D costs for a producible blueprint is independent from the entry costs. From (3.42),

$$\frac{\phi v(b_L,\cdot)}{wA^{\chi-1}F^a(b_L)} = 1.$$

Whenever the R&D costs should increase as a consequence of increasing entry costs, the average returns to successful R&D would increase by the same factor.

Recap. Let us recapitulate briefly. Melitz (2003) showed that dealing with firm heterogeneity is easy when consumers have love of variety preferences à la Dixit-Stiglitz because these preferences still allow us to work with a single, representative firm. The same is true if we follow Ethier (1982) and use a Dixit-Stiglitz aggregator in the production of output. To reveal the basic mechanics of the model, we choose to express the aggregate firm outcome in terms of the cutoff productivity type firms. Given the cutoff, R&E is conducted with a standard Jones (1995) R&D technology. In fact, the closed economy model with heterogeneous firms and entry costs boils down to the Jones (1995) model. The two additional ingredients, costly entry and firm heterogeneity, are included as follows. In Jones' model, the R&D technology is

$$\dot{A} = \frac{A^{1-\chi} L_A}{a},$$

where a is an exogenously given productivity parameter. Including entry costs, we can interpret the productivity as being endogenous. In our formulation, it incorporates the labor requirement necessary for market entry and to find a sufficiently productive blueprint (a in Jones' model can take any value so we set $a \equiv F^a(b_L)$).[41]

The aggregate equilibrium outcome with firm heterogeneity can conveniently be expressed as the outcome with a representative firm. Market entry costs introduce a minimum productivity requirement that increases the average productivity of firms. The net effect of entry costs on the level of TFP, however, is ambiguous, since an increase in productivity in production comes at the cost of an increase in the average labor requirement necessary to invent a new variety. If the share of labor in R&D remains fixed, this increase translates into a lower rate at which new intermediate goods are introduced to the output sector. Note, however, that we cannot assert that the R&D costs actually increase until we know more about the effects of the entry costs on the wage rate and the level of A which governs the spillover effects.

Returning to the derivation of an equilibrium, it remains for us to solve for the lowest productivity level that allows firms to earn the entry costs (the cutoff productivity).

3.3.6 The Equilibrium Cutoff Productivity

Our choice of expressing the aggregate intermediate firm outcome in terms of the cutoff productivity type firms shows clearly that solving for the cutoff productivity requires only the free entry into

[41] L_A in our formulation also includes entry workers (which do not exist in Jones' model).

R&D condition and the condition for profitable market entry. To see this, recall that the free entry condition requires that the average productivity type firms' values net of entry costs are equal to their R&D costs. The value of firms with the average productivity in turn is closely linked to the cutoff productivity firms' values, see (3.36) for a given A. By definition, the cutoff productivity type firms' net/market value in turn is zero, i.e. their operating value equals the entry costs. The equilibrium cutoff productivity must therefore imply an average firm value that exactly meets the average R&D costs of finding a usable blueprint. Combining (3.23) and (3.42) yields

$$F^a(b_L^*) = \phi F_L, \tag{3.43}$$

or from (3.40),

$$G(b_L^*) = \frac{F_B}{F_L(\phi - 1)}. \tag{3.44}$$

The wage rate and the scaling factor $A^{\chi-1}$ drop out because of identical technologies in R&D and market entry. Hence, free entry into R&D requires the average net value of a profitably usable blueprint $F_L(\phi-1)wA^{\chi-1} = (\phi F_L - F_L)\, wA^{\chi-1} = \phi v_L wA^{\chi-1} - F_L wA^{\chi-1}$ times the fraction of usable blueprints $G(b_L^*)$, to equal the discovery costs $F_B wA^{\chi-1}$. Put differently, researchers may expect a usable blueprint after $1/G(b_L^*) = F_L(\phi-1)/F_B$ draws on average. If successful, the return on the research investment equals $F_L(\phi-1)/F_B$, so that in expectation, researchers exactly break even on average. Since blueprints with productivities below the cutoff have zero value, the share of usable discoveries is equal to the inverse of the return on investment in R&D for any usable blueprint.

From the definition of $G(b)$ in (3.4), $b_L^* = [G(b_L)^*]^{\frac{1}{\theta}}\, b_0$. Inserting (3.44) yields the equilibrium cutoff input coefficient:

$$b_L^* = \left[\frac{F_B}{F_L(\phi - 1)}\right]^{\frac{1}{\theta}} b_0 \quad (> 0). \tag{3.45}$$

The separation of firms into profitable producers and firms that exit comes solely from the entry costs. As mentioned earlier, $b_L^* \to \infty$ as $F_L \to 0$ (i.e. the cutoff is not binding since b_0 is finite, $b_L^* = b_0$). The minimum productivity requirement, $1/b_L^*$, is obviously higher, the higher F_L.[42] Interestingly, the efficiency with which researchers operate to find new blueprints ($1/F_B$) has a negative impact on

[42]The closed economy model in this chapter merely serves as a starting point for the analysis of marginal changes in the openness of foreign markets (as indicated by the foreign market entry costs). While the above comparative static is helpful to understand the model's mechanics, we do not want to take the productivity-increasing effect of local market entry costs too serious. This is because sunk entry costs (like, e.g., costly regulation) in general deter the creation of new firms, a feature broadly supported by the data, see Alesina et al. (2005), Nicoletti and Scarpetta (2003), and Klapper et al. (forthcoming). Loosely speaking, the average firm in Greece, where entry costs are about US$ 6900, is hardly believed

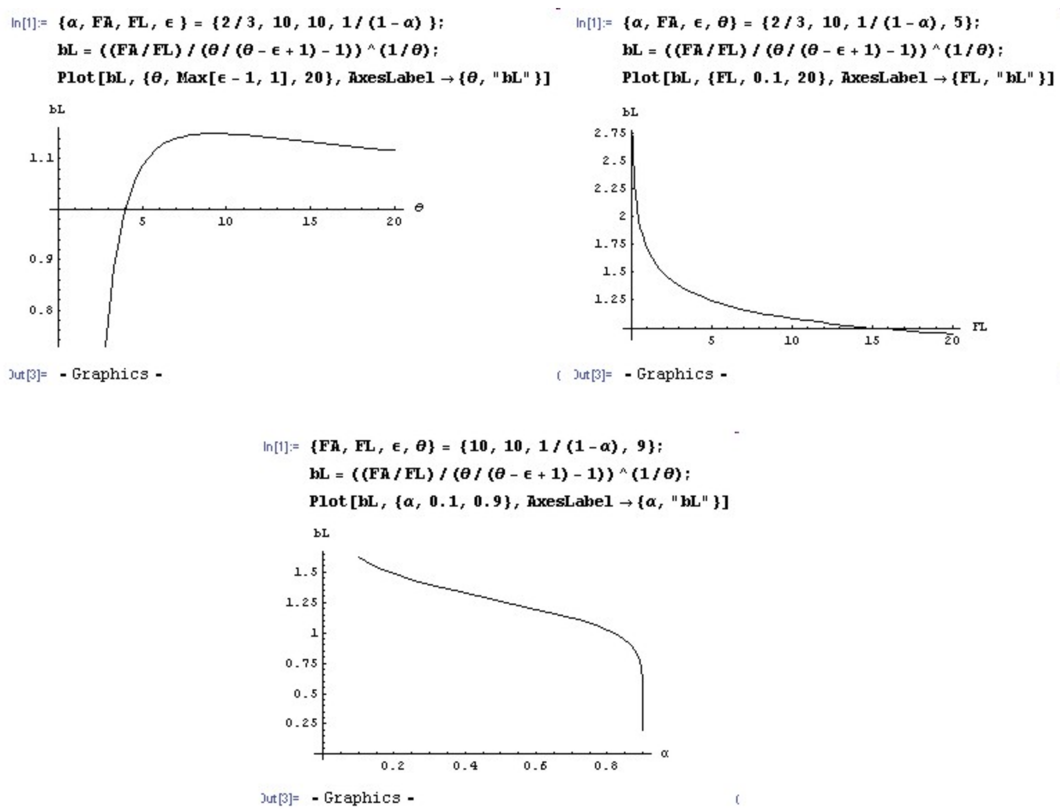


Figure 3.5: b_L^* as a Function of θ (Upper Left Panel), F_L (Upper Right Panel), and α (Lower Panel)

the highest admissible input coefficient: the cutoff input coefficient is smaller, the more efficient the development of new blueprints occurs (i.e. the smaller F_B). In other words, an R&D sector with a low productivity allows intermediate firms to be less efficient in production (horizontal and vertical productivity are complements). This is intuitive because a less productive R&D sector implies less pronounced horizontal competition from new entrants (and higher profits for incumbent firms). Figure 3.5 depicts the cutoff input coefficient as a function of θ (upper left panel), F_L (upper right panel), and α (lower panel). We summarize these findings in

Result 3.4 (**Minimum productivity requirement**). *Entry barriers introduce a minimum productivity requirement for intermediate goods firms. This requirement is higher when the capital share in the production of output is large (when α is large) and when researchers are productive in the discovery*

to be more productive than the average firm in Canada, where entry cost are much lower (US$ 280 according to Büttner, 2006).

of blueprints (when F_B is small).[43]

Note that since the factor prices drop out in the determination of the cutoff, see (3.43), the production structure is not essential for the determination of the cutoff when the production functions in R&D and entry are identical.

3.3.7 Properties of the Autarky Equilibrium

We are now in a position to characterize the equilibrium labor allocation and the evolution of horizontal specialization.

Irrelevance of entry costs for the allocation of labor in R&E. Starting with the labor allocation, we find that the relative inputs of labor in R&D and entry are not affected by barriers to entry as measured by F_L. This at first glance astonishing feature is concealed in existing models, where there is no explicit distinction between entry workers and researchers. Baldwin and Robert-Nicoud (2007) and Gustafsson and Segerstrom (2007) take a short cut by assuming that the discovery of new intermediates takes a certain amount of knowledge and that it takes an additional amount of knowledge to enter a market subsequently. As a consequence, they directly employ an R&E-knowledge production function like (3.40). Clearly, we we do not alter the modeling substantially given that we maintain the mechanical link between entry workers and researchers implied by identical production functions. Exploring the relation between entry workers and researchers, however, reveals how restrictive this assumption actually is: in equilibrium, the allocation of labor between market entry and R&D is not affected by the entry barriers.

To see this, plug

$$G(b_L^*)F_L = \frac{F_B}{\phi - 1}$$

from (3.44) into (3.25):

$$\frac{L_B}{L_E} = \phi - 1. \tag{3.46}$$

From $L_B + L_E \equiv L_A$, $L_B = (\phi - 1)(L_A - L_B)$ such that

$$\frac{L_B}{L_A} = \frac{\phi - 1}{\phi}, \tag{3.47}$$

$$\frac{L_E}{L_A} = \frac{1}{\phi}. \tag{3.48}$$

[43] Given the different meanings of α, the result with respect to the capital share is again ambiguous.

Hence, the existence of a binding cutoff is sufficient to fix the relative labor shares in R&D and entry, irrespective of the level of the barriers to entry. At second glance, of course, the increase in the labor requirement per usable blueprint is only one side of the coin. Turning the minimum productivity requirement up side down, barriers to entry reduce the share of newly invented, usable blueprints. In equilibrium, a change in the barriers to entry induces a change in the share of usable blueprints that exactly offsets the change in the labor requirement for developing a usable blueprint. Formally, from (3.44), $G(b_L^*)F_L$ is fixed independent of F_L.

More explicitly, lowering F_L has two opposing effects. On the one hand, a reduction in the barriers to entry frees labor from the conduct of entry for a given number of innovations (i.e. L_A/L_E rotates counter-clockwise around $L_Y/L = 1$ in Figure 3.4). On the other hand, the number of newly invented intermediate goods increases for a given number of researchers because the reduction in entry costs relaxes the minimum productivity requirement (and thereby reduces the average labor requirement necessary to discover a usable blueprint). Hence, there are more usable innovations for which the free entry workers conduct market entry. If there is a change in the share of researchers as a fraction of the labor force, it is accompanied by an equally sized change in the share of entry workers.

Result 3.5 (**Labor allocation between R&D and entry**). *In equilibrium, the relative use of labor in R&E, L_E/L_B, is independent of the level of the entry costs.*

Law of motion for A. Using $F^a(b_L^*)$ from (3.43) in (3.40) gives the equilibrium evolution of A:

$$\dot{A} = \frac{A^{1-\chi}L_A}{\phi F_L}. \tag{3.49}$$

Barriers to entry decrease the rate at which new blueprints are introduced to the production of output, and their impact is stronger the easier it is for the producers of output to replace inputs of less efficient firms by (cheaper) inputs of more efficient firms (as ϕ is increasing in ϵ).

The underlying R&D productivity ($1/F_B$) drops out because the productivity of blueprints used in active firms is conditional on exceeding the cutoff. On average, the discovery of these blueprints requires $F_B/G(b_L^*)$ (times $A^{\chi-1}$) workers. Since the probability of drawing a productivity of at least $1/b_L^*$ in equilibrium always takes the same effort ($G(b_L^*) = F_B/[F_L(\phi-1)]$ is linear in F_B), conditioning on $b \leq b_L^*$ removes F_B from the law of motion for A.

Result 3.6 (**Irrelevance of $\mathbf{F_B}$ for $\dot{\mathbf{A}}$**). *In equilibrium, the law of motion for A is pinned down by the entry costs irrespective of the productivity in R&D.*

Equation (3.40) indicates that the reduced form of our model yields the Jones (1995) model. We will verify this conjecture explicitly in the following section. For now, note that we are free to choose the units of measurement in the production of output, so that the production function of output in (3.1) can equivalently be stated as

$$\tilde{Y} = \delta L_Y^{1-\alpha} \int_0^A x(j)^\alpha dj, \ \delta > 0, \ \dot{\delta} = 0. \tag{3.50}$$

Inserting the equilibrium cutoff from (3.45) in the reduced form production function of output in (3.38) gives

$$Y = \phi (A L_Y)^{1-\alpha} \left\{ \frac{K}{\phi \left[\frac{F_B}{F_L(\phi-1)}\right]^{\frac{1}{\theta}} b_0} \right\}^\alpha = \left[\frac{F_L(\phi-1)}{F_B}\right]^{\frac{\alpha}{\theta}} \frac{\phi^{1-\alpha}}{b_0^\alpha} (A L_Y)^{1-\alpha} K^\alpha.$$

Without loss of generality for a given F_L let

$$\delta \equiv \left[\frac{F_B}{F_L(\phi-1)}\right]^{\frac{\alpha}{\theta}} \frac{b_0^\alpha}{\phi^{1-\alpha}}$$

so that we get a standard Cobb-Douglas production function for output:

$$\tilde{Y} = (A L_Y)^{1-\alpha} K^\alpha. \tag{3.51}$$

By definition, more productive firms produce more output out of a given amount of input. Hence, δ increases as F_L decreases since the average productivity depends positively on the monotonically increasing minimum productivity requirement introduced by entry costs.

3.3.8 Balanced Growth Path

Given that the cutoff is determined by instantaneous optimality conditions, the dynamics of the model are identical to the dynamics of the Jones (1995) model (which are analyzed in Arnold, 2006). For the sake of completeness, we adapt the analysis in Arnold (2006) to the present environment where intermediate firms have heterogeneous input coefficients, but can be summarized by a representative firm. We then use the laws of motions of key variables to solve for the equilibrium allocation along a balanced growth path. Following Arnold (2006), define the stationary variables $\tilde{l} \equiv L/A^\chi$, $z \equiv Y/K$, $\gamma \equiv cL/K$, and $\nu \equiv (1-\alpha)Y/[\int_0^A v(j)dj] = (1-\alpha)Y/[A\phi v(b_L^*)$. We use $\tilde{l}$ instead of $l \equiv L/\left[\phi F_L A^\chi\right] = \tilde{l}/\left(\phi F_L\right)$ to trace the R&E productivity in the law of motion for A. Of course, we can alternatively derive the balanced growth path without these definitions.[44] Deriving the entire dynamic system, however, increases our understanding of the model's mechanics and gives us an idea of the interdependencies off the balanced growth path.

Definition 3.2 (**Steady state**). *A steady state is an equilibrium with l, z, γ and ν constant.*

Before we consider the dynamic system, we first report the steady state growth rates of all endogenous variables.

Steady State Growth Rates

By definition of a steady state, all variables grow at constant rates. Imposing a constant growth rate on consumption requires, via the Euler equation (3.12), that the interest rate is a constant as well. Accordingly, from (3.15), the prices of the intermediate goods are constant. From labor market clearing, labor in all sectors must grow at the population growth rate so as to ensure that the labor shares are constant. Profit maximization in the production of output then implies that the input quantities of intermediate goods also grow at rate n, see (3.14). As prices and the interest rate are constant, profits and firm values also grow at rate n, see (3.18) and (3.22). Then, using the results of the aggregation in (3.32) and (3.51), the capital stock and final output grow at rate $n + n/\chi$. For the growth rate of blueprints to be a constant, B must grow at rate n/χ.[45] As b_L^* is a constant (see (3.45)), (3.24) demands that A is a constant fraction of B. Hence, A grows at the same rate as B. Finally, the profit maximizing labor demand in the production of output in (3.13) implies that the wage rate grows at rate n. To summarize, $\hat{A} = \hat{B} = \hat{w} = \hat{c} = n/\chi$, $\hat{K} = \hat{Y} = n + n/\chi$, $\hat{L}_Y = \hat{L}_E = \hat{L}_B = \hat{x} = \hat{\pi} = \hat{v} = n$, and $\hat{r} = \hat{p} = \hat{b}_L^* = 0$.

In a steady state, the growth rate of consumption per capita, $\hat{c} = n/\chi$, and the Euler equation in (3.12) pin down the interest rate irrespective of the barriers to entry:

$$r^* = \sigma \frac{n}{\chi} + n + \rho. \tag{3.52}$$

That is, the long-run interest rate is such that it removes the dissaving motives from population

[44] As an example, consider ν. ν^* is simply derived from the long-run values of r and $\hat{v}^*$. $\nu = \frac{(1-\alpha)Y}{A\phi v(b_L,\cdot)} = \frac{\phi A \pi_L}{\alpha \phi A v_L}$ since $A\phi\pi_L = (1-\alpha)\alpha Y$. From (3.6) in a steady state, $r^* - \hat{v} = \pi(j)/v(j) = \pi(b_L,\cdot)/v(b_L,\cdot)$. Using this equation to substitute for the dividend ratio in the expression for ν^* delivers ν^*.

[45] In (3.40), $\dot{A} = \frac{A^{1-\chi}L_A}{\phi F_L}$, so that $\hat{\dot{A}} = 0$ requires $\hat{A} = \frac{n}{\chi}$. Hence, (3.24) and (3.44) yield $\hat{B} = \hat{A} = \frac{n}{\chi}$.

growth, subjective discounting, and growth (which is larger the smaller the elasticity of intertemporal substitution).[46]

Result 3.7 (Optimal consumption determines steady state interest rate). *In a steady state, the interest rate, i.e. the factor price for intermediate goods firms, is independent of barriers to entry.*

Impact channels of entry costs. While this relation is generally well known for the Jones (1995) model, it is an important observation with respect to the impact of trade. Intuitively speaking, a reduction in the barriers to trade exerts an production-expanding effect (via additional entry and production for foreign markets by exporters) which is expected to bid up the factor price. In fact, in the Melitz (2003) model, the trade-induced increase in the real wage rate is the only channel through which trade openness drives the least productive firms out of the domestic market (by increasing the local market's minimum productivity requirement). In particular, the constant price elasticity of demand implied by the Dixit-Stiglitz index severely limits the possible impact of trade on factor price effects. Changing the number of competitors or their productivity leaves the elasticity of demand unaffected (see Melitz, 2003, p. 1715). In the present model, Result 3.7 implies that the steady state factor price is independent of the minimum productivity requirement implied by the entry costs. In looking for the impact of trade in a monopolistic competition model under CES production, this observation hints at looking for decreases in the prices for the other production factors, labor and knowledge.

In what follows, we deduce the laws of motions and the steady state values for the transformed variables. If one is less interested in this rather technical derivation, one can skip this paragraph. The stationary values of key variables are identical to those in the Jones (1995) model since the minimum productivity requirement only enters through the productivity in R&E.

Laws of Motions for Key Variables

To economize on notation, let $\eta(b_L^*) = \eta_L$, $\pi(b_L^*, \cdot) = \pi_L$ and $v(b_L^*, \cdot) = v_L$. Log-differentiating the definition of γ, we have

$$\dot{\gamma} = \gamma \left(\frac{\dot{c}}{c} + n - \frac{\dot{K}}{K} \right).$$

[46] From the point of view of the market for the investment good, the determinants of r equivalently reflect the scarcity of raw capital.

To substitute for $\dot{c}/c$ and $\dot{K}/K$, note from the output-capital ratio in (3.39) that

$$\alpha^2 z = r.$$

Dividing by K, the resource constraint in (3.10) implies $\dot{K}/K = z-\gamma$. Using these expressions together with the Euler equation yields the law of motion for γ:

$$\dot{\gamma} = \gamma\left[\left(\frac{\alpha^2}{\sigma}-1\right)z+\gamma+\left(1-\frac{1}{\sigma}\right)n-\frac{\rho}{\sigma}\right]. \tag{3.53}$$

The aggregate firm value is equal to $\int_0^A v(j)dj = \phi A v_L$ (see (3.36)). As $v_L(r-\hat{v}) = \pi_L$ from (3.22) and (3.20), the denominator of $\nu = (1-\alpha)Y/(\phi A v_L)$ is

$$\phi A v_L = \frac{\phi A \pi_L}{r-\hat{v}}. \tag{3.54}$$

Aggregate profits from (3.35), rewritten as

$$\phi A\pi(b_L) = \int_0^A \pi(j)dj = (1-\alpha)\alpha^{2\epsilon-1}\int_0^A [rb(j)]^{1-\epsilon}L_Y dj = (1-\alpha)\alpha L_Y^{1-\alpha}\int_0^A \alpha^{2\alpha\epsilon}[rb(j)]^{1-\epsilon}L_Y^{\alpha}dj,$$

can be expressed as a constant fraction of aggregate output. Recognizing $-\alpha\epsilon = 1-\epsilon$ and using (3.16),

$$A\phi\pi_L = (1-\alpha)\alpha\int_0^A x(j)^{\alpha}dj = \alpha(1-\alpha)Y. \tag{3.55}$$

Taken together, (3.54) and (3.55) imply $r-\hat{v} = \alpha\nu$. Replacing $r = \alpha^2 z$ gives the law of motion for the intermediate firms' values:

$$\hat{v} = \alpha(\alpha z - \nu). \tag{3.56}$$

The equilibrium law of motion for A from (3.49) can be rewritten using the labor market clearing condition (3.7), the equilibrium labor demand from the aggregate output sector (3.13), and the free entry condition (3.42), which equals the free entry into the product market condition in equilibrium:

$$\frac{\dot{A}}{A} = \frac{A^{-\chi}L_A}{\phi F_L} = \frac{L-L_Y}{A^{\chi}\phi F_L} = \frac{\tilde{l}}{\phi F_L} - \frac{(1-\alpha)YF_L}{A^{\chi}\phi F_L v_L A^{1-\chi}} = \frac{\tilde{l}}{\phi F_L} - \nu, \tag{3.57}$$

or simply $\dot{A}/A = l - \nu$.

Now, combining (3.32) and (3.37), we have $Y/K = \phi\varphi^{-\alpha}(AL_Y)^{1-\alpha}K^{\alpha-1}$, or

$$z = \frac{\phi}{\varphi^{\alpha}}\left(\frac{AL_Y}{K}\right)^{1-\alpha}. \tag{3.58}$$

After substituting for L_Y using (3.13), (3.58) becomes

$$z = \frac{\phi}{\varphi^{\alpha}}\left[\frac{(1-\alpha)AY}{wK}\right]^{1-\alpha}.$$

Solving for $z = Y/K$ yields

$$z = \frac{\phi^{\frac{1}{\alpha}}}{\varphi}\left[\frac{(1-\alpha)A}{w}\right]^{\frac{1-\alpha}{\alpha}}.$$

We can replace the wage rate in the above expression from the free entry condition (which in equilibrium coincides with the entry into the product market condition), $w = v_L A^{1-\chi}/F_L$, and get

$$z = \frac{\phi^{\frac{1}{\alpha}}}{\varphi}\left[\frac{(1-\alpha)A^{\chi}F_L}{v_L}\right]^{\frac{1-\alpha}{\alpha}}. \tag{3.59}$$

Log-differentiating yields

$$\dot{z} = z\frac{1-\alpha}{\alpha}\left(\chi\frac{\dot{A}}{A} - \frac{\dot{v}}{v}\right),$$

or, replacing $\dot{A}/A$ from (3.57) and $\dot{v}/v$ from (3.56),

$$\dot{z} = z\left(\frac{1-\alpha}{\alpha}\right)\left[\frac{\chi}{\phi F_L}\tilde{l} + (\alpha-\chi)\nu - \alpha^2 z\right]. \tag{3.60}$$

Turning to $\tilde{l}$, $\dot{\tilde{l}} = \tilde{l}(g - \chi\dot{A}/A)$, and inserting $\dot{A}/A$ from (3.49), one obtains

$$\dot{\tilde{l}} = \tilde{l}\left[n - \chi\left(\frac{\tilde{l}}{\phi F_L} - \nu\right)\right]. \tag{3.61}$$

Finally, a differential equation for ν is obtained as follows. Dividing (3.10) by K, we get the law of motion for the capital stock,

$$\frac{\dot{K}}{K} = (z - \gamma). \tag{3.62}$$

Using the definition of z, $\dot{Y}/Y = \dot{z}/z + \dot{K}/K$, and after replacing $\dot{K}/K$ from (3.62) and $\dot{z}/z$ from (3.60),

$$\frac{\dot{Y}}{Y} = \left(\frac{1-\alpha}{\alpha}\right)\left[\frac{\chi}{\phi F_L}\tilde{l} + (\alpha-\chi)\nu - \alpha^2 z\right] + z - \gamma. \tag{3.63}$$

Now plugging this expression together with the laws of motion for $v(j)$ and A in (3.56) and (3.57) in the log-differentiated definition of ν, $\dot{\nu} = \nu(\hat{Y} - \hat{A} - \hat{v})$,

$$\frac{\dot{\nu}}{\nu} = \left(\frac{1-\alpha}{\alpha}\right)\left[\frac{\chi}{\phi F_L}\tilde{l} + (\alpha-\chi)\nu - \alpha^2 z\right] + z - \gamma - \frac{\tilde{l}}{\phi F_L} + \nu - \alpha(\alpha z - \nu).$$

After collecting terms, the law of motion for ν equals

$$\dot{\nu} = \nu\left[\left(\frac{1-\alpha}{\alpha}\chi - 1\right)\frac{\tilde{l}}{\phi F_L} + \left(2 - \frac{1-\alpha}{\alpha}\chi\right)\nu + (1-\alpha)z - \gamma\right]. \tag{3.64}$$

Steady State

In a steady state, where $\dot{\gamma} = \dot{z} = \dot{\tilde{l}} = \dot{\nu} = 0$, (3.53), (3.60), (3.61), and (3.64) imply

$$\begin{aligned}
\left[\left(\frac{\alpha^2}{\sigma} - 1\right) z^* + \gamma^*\right] + \left(1 - \frac{1}{\sigma}\right) n &= \frac{\rho}{\sigma} &(3.65)\\
\frac{\chi}{\phi F_L}\tilde{l}^* + (\alpha - \chi)\nu^* &= \alpha^2 z^* &(3.66)\\
\frac{\chi}{\phi F_L}\tilde{l}^* &= n + \chi\nu^* &(3.67)\\
\left(\frac{1-\alpha}{\alpha} - \frac{1}{\chi}\right)\frac{\chi}{\phi F_L}\tilde{l}^* &= \gamma^* - \left(2 - \frac{1-\alpha}{\alpha}\chi\right)\nu^* - (1-\alpha)z^*. &(3.68)
\end{aligned}$$

These four equations are readily solved for a unique steady state. Eliminating $\chi/(\phi F_L)\tilde{l}^*$ from (3.66) and (3.67) and (3.67) and (3.68), respectively gives:

$$\begin{aligned}
\frac{n}{\alpha} + \nu^* &= \alpha z^* &(3.69)\\
\left(\frac{1-\alpha}{\alpha} - \frac{1}{\chi}\right) n + \nu^* &= \gamma^* - (1-\alpha)z^*. &(3.70)
\end{aligned}$$

Equation (3.65) can be used to eliminate γ^* from (3.70):

$$\left(\frac{1}{\alpha} - \frac{1}{\sigma} - \frac{1}{\chi}\right) n + \nu^* + \alpha\left(\frac{\alpha}{\sigma} - 1\right) z^* = \frac{\rho}{\sigma}. \tag{3.71}$$

Let $\Delta = (\sigma - 1)n/\chi + \rho$. Combining (3.69) and (3.71) to derive ν^* and z^*, and using (3.67) and (3.68) gives (see the detailed derivation in Appendix 3.A):

$$\begin{aligned}
\nu^* &= \frac{1}{\alpha}\left(\Delta + \frac{1}{\chi}n\right) &(3.72)\\
z^* &= \frac{1}{\alpha^2}\left(\Delta + \frac{1+\chi}{\chi}n\right) &(3.73)\\
\gamma^* &= \frac{1}{\alpha^2}\left[\Delta + \frac{1+\chi}{\chi}(1-\alpha^2)n\right] &(3.74)\\
\tilde{l}^* &= \frac{\phi F_L}{\alpha}\left[\Delta + \frac{1+\alpha}{\chi}n\right]. &(3.75)
\end{aligned}$$

Replacing $\tilde{l}^* = l^*\phi F_L$ we also get,

$$l^* = \frac{1}{\alpha}\left(\Delta + \frac{1+\alpha}{\chi}n\right). \tag{3.76}$$

The steady state in (3.72)-(3.74) and (3.76) is the steady state in the Jones (1995) model with $a \equiv \phi F_L$ (where "a" in Jones, 1995, is the inverse of the R&D productivity). The present modeling of firm heterogeneity and R&D implies that, in the long-run, entry costs and firm heterogeneity exclusively affect the productivity of R&E, and hence $\tilde{l} = L/A^{\chi}$.

Arnold (2006)'s findings on the dynamics of the Jones (1995) model are thus robust to our extensions. In particular, $\Delta > 0$ is sufficient to ensure that all stationary variables are positive, utility is bounded ($(1-\sigma)\dot{c}/c - \rho < 0$ since $\dot{c}/c = n/\chi$), and that the transversality condition holds, so that a steady state exists if and only if $\Delta > 0$.

Using the steady state values we find that this "independence" result carries over to the allocation of labor.

Result 3.8 (BGP labor shares). *In steady state, the allocation of labor between R&E and production is independent of F_L.*

This observation is easily verified by combining the condition for optimal labor in production in (3.16), the free entry condition in aggregate terms in (3.42), and ν^*:

$$\nu^* = \frac{(1-\alpha)Y}{A\phi v_L} = \frac{wL_Y}{wA^\chi \phi F_L} = \frac{L_Y}{A^\chi \phi F_L} = \frac{L_Y}{L} l^*,$$

and therefore,[47]

$$\frac{L_Y}{L} = \frac{\nu^*}{l^*} = \frac{\Delta + \frac{n}{\chi}}{\Delta + \frac{1+\alpha}{\chi} n}. \tag{3.77}$$

As $\alpha > 0$, and $\Delta + n/\chi > 0$, we have $0 < L_Y/L < 1$. From this observation and the labor market clearing condition (3.7), we directly infer that $0 < L_A/L < 1$ and $L_A/L = 1 - \frac{\nu^*}{l^*}$ is given irrespective of the level of F_L.[48]

Similarly, using (3.3), (3.11), (3.24), and $G(b_L^*) = F_B/[(\phi-1)F_L]$, in steady state,

$$\left(\frac{L_B}{L}\right)^* = \frac{n}{\chi}\frac{A^\chi}{L}(\phi - 1)F_L \tag{3.78}$$

$$\left(\frac{L_E}{L}\right)^* = \frac{n}{\chi}\frac{A^\chi}{L}F_L. \tag{3.79}$$

Since the labor shares are independent of the barriers to entry, the total labor income is also independent of the entry costs. Due to the Cobb-Douglas production of output we have $wL_Y = (1-\alpha)Y$, see (3.13), and hence

$$\frac{wL}{Y} = (1-\alpha)\frac{L}{L_Y} = (1-\alpha)\frac{\nu^*}{l^*}.$$

No reallocation between factor incomes. Do entry costs affect the aggregate distribution of wage and capital income in the steady state? Using the definition of z in $wL_Y = (1-\alpha)Y$ from the last paragraph, $wL_Y = (1-\alpha)r^*K/\alpha^2$ and hence

$$\frac{wL}{rK} = \frac{(1-\alpha)l^*}{\alpha^2 \nu^*}$$

[47] Substituting for Δ, (3.77) equivalently states $\frac{L_Y}{L} = \frac{\sigma n + \chi\rho}{(\sigma+\alpha)n + \chi\rho}$.

[48] $\frac{L_A}{L} = \frac{\alpha n}{[(\sigma+\alpha)n + \rho\chi}$.

is independent of barriers to entry.

Having characterized the equilibrium in the closed economy, we now turn to the open economy.

3.4 Trade

We include characteristic features of international trade in the simplest possible way. Consider a world of two economies, each one as described in the previous section. The two economies have identical preferences, technologies, production structures, and identical capital and labor endowments. Only intermediate goods are traded internationally.[49] The free flow of intermediates between the two countries is hampered by marginal trading costs and TBTs.

TBTs. Empirically, TBTs remain important obstacles between developed countries despite various rounds of free trade negotiations. Importantly, TBTs are pure trading costs, and as such should be interpreted distinctly from the local market entry costs. TBTs are fixed costs associated with the entry of firms into the export market and account for country specific product/production standards/regulations, additional certification procedures, or additional bureaucratic burdens that make it harder for foreign firms to supply the domestic market than for their local competitors.[50] To capture the relative disadvantage for foreign firms, we follow Melitz (2003) and assume that foreign firms face higher entry costs when entering the export market than local firms that enter that same market. Due to symmetry, exporting thus comes at higher fixed costs than producing for the local market from a domestic firm's point of view. Hence, there is another cutoff productivity, $1/b_E$, for exporting. In the presence of TBTs, b_E is lower than b_L, so that the equilibrium productivity pattern in the local and the export market matches the empirical regularity that the bulk of firms sells only locally and only the most productive firms export. As an aside, the empirical trade literature has also clarified that there are no feedback effects from exporting to a firm's productivity (Bernard and Jensen, 1999, and Bernard, Jensen, and Schott, 2006). The fact that input coefficients remain constant in the model when a firm starts to export is thus in line with the empirical evidence. Returning to the model, we account for the different sources of TBTs and iceberg costs by using different "factor intensities" for the two types of trade barriers. With respect to the fixed export costs, we adapt the modeling in the literature and assume that fixed export costs are wage costs.

[49] We abstract from including trade in the final good since this would allow imports of new physical capital.

[50] Roberts and Tybout (1997) validate empirically that the sunk cost associated with exporting are of substantial magnitude for exporting firms. See also the evidence in Bernard and Jensen (2004).

Iceberg costs. The variable trading costs are modeled as Samuelson-type iceberg costs and as such decrease the productivity in the production of exported units. Since intermediate goods are manufactured using physical capital, variable trading costs constitute capital costs to the firms.

In what follows, we describe the additional assumptions for the open economy and then deduce the adjusted optimality conditions.

3.4.1 Open Economy

Technologies

Output. We distinguish local and export market variables of intermediate firms by a subscript i, $i \in \{L, E\}$. To avoid additional complexity, we focus on the case where there is no overlap between local and foreign varieties at the point in time when trade liberalization occurs.[51] That is, if foreign firms decide to export, local output producers are able to employ a larger number of intermediate inputs and explore a higher degree of specialization. We denote the worldwide "number" of different existing intermediate goods by $\tilde{A} > A_L$. Without loss of generality, let the intermediate goods index be such that $j \in [0, A_L]$ indicates locally produced intermediate goods, while $j \in (A_L, \tilde{A}]$ refers to imported varieties. Aggregate output in each economy is given by

$$Y = L_Y^{1-\alpha} \int_0^{\tilde{A}} x_i(j)^\alpha dj, \tag{3.80}$$

where $x_i(j) = x_L(j)$ for $j \leq A_L$ represents the input quantity of a locally produced good and $x_i(j) = x_E(j)$ for $j > A_L$ is the input quantity of an imported intermediate good.

Intermediates. When exported, $\tau \geq 1$ units of an intermediate good must be shipped for every unit that arrives (exporters get paid for the arriving units). From an exporting firm's perspective, producing one actually sellable unit for the foreign market thus ties up τ times more resources than producing the same unit for the local market:

$$x_i(j) = \frac{k_i(j)}{\tau_i b(j)}, \tag{3.81}$$

where here and in what follows $\tau_i = \tau$ if goods are exported ($i = E$), and $\tau_i = 1$ if goods are sold locally ($i = L$). Iceberg costs therefore imply that the productivity of a firm depends on the destination of the manufactured output. The level of productivity in the production of exports thereby decreases linearly in the iceberg costs. To simplify the exposition, we treat the marginal trade

[51] Cf. Tang and Wälde (2001) for a model with initial overlap.

costs as a technology such that they do not yield any income.[52]

Markets

Barriers to trade. Launching an export business with a newly discovered intermediate good requires the entrant to hire $A_L^{\chi-1}F_E$ "entry workers", where $F_E \geq F_L$ so that $T \equiv F_E/F_L \geq 1$. T (a mnemonic for TBTs) measures how much harder it is for a foreign firm to enter the local market compared to a domestic firm. If $T > 1$, additional profit opportunities from exporting accrue only to the most productive firms for which it is profitable to sink the foreign market entry costs. Incremental trade liberalization is modeled as a decrease in either transportation costs, τ, or TBTs, T. In the case of TBTs, we formally consider the comparative statics with respect to F_E (evaluated at $F_E \geq F_L$).

Profitable market entry. Firms base the decision to export on the same forward-looking investment calculus as the decision to enter the domestic market in autarky. Given its productivity, each firm knows its future profits in either market and decides to enter a market only if the present value of its profits in that market exceeds the entry costs to that market. We denote by $1/b_L$ and $1/b_E$ the lowest productivity levels that allow firms to operate profitably in the local and the export market, respectively. The presence of TBTs (i.e., $F_E > F_L$) implies that $b_E < b_L$ so that only the most productive firms export. Hence, $x_E(j)$ in (3.80) is zero for some $j > A_L$.[53] While entering the domestic market is less involved for local firms, it still takes costly resources and hence there is a minimum productivity requirement for active firms.[54] The least productive firms, which do not meet this minimum productivity requirement, exit immediately. Firms with input coefficients $b_L \geq b > b_E$ sell exclusively in their home market and the most productive firms with $b \leq b_E$ sell both in the local market and export.

Market entry and research again use the same production technology and we define its productivity so as to already include the effects of international knowledge spillovers.[55] For further reference define

[52] See Matsuyama (2007) for a theory of factor biased trading costs.

[53] By the definition of A_L as the number of actually active firms, all $x_L(j)$ are positive for $j \leq A_L$.

[54] The lower bound on F_L in (PA1) ensures that $b_L < b_0$ in equilibrium so that the minimum productivity requirement is binding for some firms.

[55] Consider the Jones (1995) R&D production function from the closed economy, augmented by international knowledge spillovers:

$$\dot{B} = \frac{(A_L + \sigma A_F)^{1-\chi} L_A}{\tilde{F}_B},$$

where A_F is the knowledge stock in the foreign country, $\sigma \geq 0$ measures the intenisty of the across-the-border spillovers, and $\tilde{F}_B$ is an exogenous productivity parameter. With symmetric countries it holds that $A_F = A_L$ and hence $A_L + \sigma A_F =$

the present value of operating profits $\pi_i\,(j)$ of firm j in market i:

$$v_i(j) \equiv \int_t^\infty e^{-\bar{r}(s-t)}\pi_i(j)ds. \tag{3.82}$$

Since firms differ only in terms of their productivities, profit maximization together with (3.82) again implies that $v_i(j;b\,(j)) = v_i(j';b\,(j'))$ if and only if $b\,(j) = b\,(j')$. With a slight abuse of notation, we can therefore drop the firm index and equivalently state the present value of profits of firm j in market i as a function of the firm's input coefficient (and other variables), i.e.

$$v_i\,(b,\cdot) \equiv \int_t^\infty e^{-\bar{r}(s-t)}\pi_i(b,\cdot)ds, \tag{3.83}$$

where $\pi_i\,(b,\cdot)$ is the operating profit of a firm with input coefficient b in market i. Differentiating (3.83) with respect to time t, the definition of $v_i\,(b,\cdot)$ implies

$$rv_i(b,\cdot) = \pi_i(b,\cdot) + \dot{v}_i(b,\cdot). \tag{3.84}$$

Entry into R&D. Including the additional profit opportunities net of entry costs for productivity types $b \leq b_E$, free entry into R&D in the open economy requires

$$\int_0^{b_0} \max\left[v_L(b,\cdot) - A_L^{\chi-1}wF_L, 0\right] dG(b) + \int_0^{b_0} \max\left[v_E(b,\cdot) - A_L^{\chi-1}wF_E, 0\right] dG(b) = A_L^{\chi-1}wF_B \tag{3.85}$$

whenever $\dot{B_L} > 0$. In equilibrium, the innovation costs have to equal the expected market value of a newly discovered blueprint, i.e. the sum of expected operating values net of entry costs in both the local and the foreign market. Expectations are taken with respect to productivity, accounting for the fact that only sufficiently productive blueprints sell at all and that only blueprints with a high productivity allow for additional profits in the export market. We already know that market values decrease strictly monotonically in b so that the cutoffs b_L and b_E, i.e. the productivities that yield zeros in the squared brackets in (3.85), are unique.[56] In particular, the cutoff associated with exporting, b_E, is determined via

$$v_E(b_E,\cdot) = wA_L^{\chi-1}F_E, \tag{3.86}$$

$(1+\sigma)A_L$. We let $F_B \equiv \tilde{F}_B/(1+\sigma)^{1-\chi}$ to maintain

$$\dot{A}_L = \frac{A_L^{1-\chi}L_A}{F_B}$$

as in the closed economy.

[56]Under our parameter assumptions, both cutoffs also exist within the support of the equilibrium distribution of active firms' productivities.

whereas the minimum productivity requirement for active firms, $1/b_L$, is implicitly given by

$$v_L(b_L,\cdot) = wA_L^{\chi-1}F_L. \tag{3.87}$$

In view of (3.86) and (3.87), the free entry condition in (3.85) equivalently reads

$$\int_0^{b_L}\left[v_L(b,\cdot) - A_L^{\chi-1}wF_L\right]dG(b) + \int_0^{b_E}\left[v_E(b,\cdot) - A_L^{\chi-1}wF_E\right]dG(b) = A_L^{\chi-1}wF_B. \tag{3.88}$$

No imitation or "footloose" production. Two simplifying assumptions ensure that firms have no means to avoid the trading costs. First, firms are not able to form multinational companies or issue production licenses, i.e. there is no "footloose" production. Second, transportation costs are lower than the cost of patent infringement, so that imitation and limit pricing by foreign firms is not profitable.

3.4.2 Equilibrium

We proceed by deriving the equilibrium for given cutoff productivity levels, then aggregate across firms, and use the results to determine the cutoffs.

Optimality Conditions

Households and firms. Optimal consumption is not affected by the degree of trade openness in the model. The demand for intermediates from the final good sector in (3.14) now applies to all $j \in [0,\tilde{A}]$ and the profit-maximizing monopoly price in (3.15) refers to the price of selling locally, $p_L(j) = rb(j)/\alpha$. When exporting, the monopolists charge the profit maximizing price $p_E(j) = \tau rb(j)/\alpha = \tau p_L(j)$. Using these pricing rules, the demand for variety j in market i obeys

$$x_i(j) = \alpha^{2\epsilon}\left[r\tau_i b(j)\right]^{-\epsilon} L_Y. \tag{3.89}$$

If firm j is active in market i, it receives equilibrium revenues

$$R_i(j) = \alpha^{2\epsilon-1}\left[r\tau_i b(j)\right]^{1-\epsilon} L_Y, \tag{3.90}$$

and earns profits $\pi_i(j) = (1-\alpha)R_i(j)$ in that market.Using the fact that $p_i(j) = p_i(j')$ and $x_i(j) = x_i(j')$ if and only if $b(j) = b(j')$, we can again rewrite the equilibrium prices and quantities in terms of productivities:

$$p_i(b,\cdot) = \frac{\tau_i rb}{\alpha},\ x_i(b,\cdot) = \alpha^{2\epsilon}\left[r\tau_i b\right]^{-\epsilon} L_Y.$$

The same is true for revenue and profits,

$$R_i(b) = \alpha^{2\epsilon-1} \left[r\tau_i b \right]^{1-\epsilon} L_Y, \;\; \pi_i(b) = (1-\alpha) R_i(b),$$

which was already used in the definition of the present value of profits in (3.83) above.

A simple relation between the cutoffs. Due to the symmetry assumption, the relation of the two cutoff productivities is easily derived as follows. Combining the local and the foreign market entry condition, (3.86) and (3.87), gives $v_E(b_E, \cdot)/v_L(b_L, \cdot) = T$. For a given b, $v_E(b, \cdot)$ and $v_L(b, \cdot)$ differ only because of the marginal trading costs. Since $\hat{v}_i(b, \cdot) = \hat{v}(\cdot)$ from (3.84), $\pi_E(b, \cdot) = \pi_L(\tau b, \cdot)$ implies $v_E(b, \cdot) = v_L(\tau b, \cdot)$. We also know from the closed economy that the ratio of any two firms' market values only depends on their input coefficients (and ϵ, cf. (3.22)). In this setup, the two cutoff input coefficients are therefore exclusively related by variable and fixed barriers to trade:

$$\frac{b_E}{b_L} = \psi, \quad 0 > \psi \equiv \tau^{-1} T^{-\frac{1}{\epsilon-1}} \geq 1. \tag{3.91}$$

ψ is an inverse measure of the real barriers to trade and measures the economies' openness. Under free trade, $\tau = T = 1$ and $\psi = 1$. The more restricted trade is (i.e. the larger τ and T), the smaller is ψ. Autarky corresponds to $\psi \to 0$.

We verify from (3.91) that both marginal trading costs and TBTs drive a wedge between the minimum productivity requirement for the local and the export market. The impact of TBTs is thereby more severe if ϵ is large. This is because a high elasticity of substitution in the production of output depresses prices, profits, and hence market values in the intermediate sector so that firms must be more productive to cover the entry costs.[57]

Aggregation

Output. Making use of the convention that goods $j \in [0, A_L]$ refer to locally produced intermediates while $j \in [A_L, \tilde{A}]$ indicate imported varieties, aggregate output can be rewritten as

$$Y = L_Y^{1-\alpha} \int_0^{\tilde{A}} x(j)^\alpha dj = L_Y \left[\int_0^{A_L} x_L(j)^\alpha dj + \int_{A_L}^{\tilde{A}} x_E(j)^\alpha dj \right].$$

Using integration by substitution to switch from goods j to productivities b, the term in squared brackets becomes $A_L \int_0^{b_L} x_L(b)^\alpha d\mu(b) + A_L^\diamondsuit \int_0^{\psi b_L^\diamondsuit} x_E(b)^\alpha d\mu^\diamondsuit(b)$, where the diamonds indicate foreign

[57] Recall, however, that α also measures the capital intensity in the final good production and the gains from specialization.

values and $b_E^\Diamond = \psi b_L^\Diamond$ was used (see (3.91)). We again solve the model in terms of the local cutoff productivity b_L. Due to symmetry and identical initial conditions, $A_L = A_L^\Diamond$, $b_L = b_L^\Diamond$ (such that $\mu^\Diamond(b) = \mu(b)$) and using (3.89) and (3.91), we obtain

$$Y = L_Y^{1-\alpha}\left[A_L \alpha^{2\alpha\epsilon} r^{-\alpha\epsilon} L_Y^\alpha \left(\int_0^{b_L} b^{-\alpha\epsilon} d\mu(b) + \tau^{-\alpha\epsilon}\int_0^{\psi b_L} b^{-\alpha\epsilon} d\mu(b)\right)\right].$$

Applying the Pareto specification and integrating yields

$$\begin{aligned} Y &= L_Y^{1-\alpha}\left[\frac{A_L \alpha^{2\alpha\epsilon} r^{-\alpha\epsilon} L_Y^\alpha \theta}{b_L^\theta}\left(\int_0^{b_L} b^{\theta-\epsilon} db + \tau^{1-\epsilon}\int_0^{\psi b_L} b^{\theta-\epsilon} db\right)\right] \\ &= L_Y^{1-\alpha}\left[\phi A_L \alpha^{2\alpha\epsilon} r^{-\alpha\epsilon} L_Y^\alpha \left(b_L^{1-\epsilon} + \tau^{1-\epsilon}\psi^{\theta-\epsilon+1} b_L^{1-\epsilon}\right)\right]. \end{aligned}$$

Using (3.89) again and recognizing $-\alpha\epsilon = 1-\epsilon$, aggregate output can be rewritten as

$$Y = \phi A_L L_Y^{1-\alpha} x(b_L, \cdot)^\alpha \left(1 + \tau^{1-\epsilon}\psi^{\theta-\epsilon+1}\right).$$

After collecting the parameters in the last term, we get

$$Y = \phi \Psi A_L L_Y^{1-\alpha} x(b_L, \cdot)^\alpha, \tag{3.92}$$

where

$$1 \leq \Psi \equiv 1 + \psi^\theta T \leq 2, \quad \partial\Psi/\partial\tau < 0, \ \partial\Psi/\partial T < 0.$$

For a given cutoff b_L, the expression for Y converges to the closed economy counterpart $Y = \phi A_L L_Y^{1-\alpha} x(b_L, \cdot)^\alpha$ as trade costs become prohibitively high ($\psi \to 0$ such that $\Psi \to 1$), (cf. (3.37)). Under free trade, (where $T = \tau = \psi = 1$) $\Psi = 2$ and output were twice the amount in the closed economy if the autarky cutoff were to prevail in the open economy also (the mass of domestic firms also would remain unchanged).

Capital. The capital stock is analogously derived as follows. From

$$K = \int_0^{A_L} k(j)dj = \int_0^{A_L} b(j)x(j)dj,$$

where $b(j)$ and $x(j)$ refer to the total output of an intermediate firm (i.e. output sold in the local market and, if applicable, output sold in the export market) and the destination dependent productivity, respectively. In terms of productivities, the capital stock reads

$$K = A_L \int_0^{b_L} b x_L(b) d\mu(b) + A_L \int_0^{b_E} b\tau x_E(b) d\mu(b).$$

After substituting with the output functions $x_i(b)$ and integrating over the distribution of firms' productivity types we get

$$\begin{aligned}
K &= A_L\alpha^{2\epsilon}r^{-\epsilon}L_Y\left[\int_0^{b_L} b^{1-\epsilon}d\mu(b) + \int_0^{b_E}(\tau b)^{1-\epsilon}d\mu(b)\right] \\
&= \frac{A_L\alpha^{2\epsilon}r^{-\epsilon}L_Y\theta}{b_L^\theta}\left[\int_0^{b_L} b^{\theta-\epsilon}db + \int_0^{\psi b_L}\tau^{1-\epsilon}b^{\theta-\epsilon}db\right] \\
&= \phi A_L\alpha^{2\epsilon}(rb_L)^{-\epsilon}L_Y b_L(1+\tau^{1-\epsilon}\psi^{\theta-\epsilon+1}).
\end{aligned}$$

The last term is again $1+\tau^{1-\epsilon}\tau^{\epsilon-1-\theta}T^{-\frac{\theta}{\epsilon-1}+1} = \Psi$ such that

$$K = \phi\Psi A_L b_L x(b_L,\cdot). \tag{3.93}$$

After replacing $x(b_L,\cdot)^\alpha = K^\alpha(\phi\Psi A_L b_L)^{-\alpha}$, aggregate output in the open economy with costly trade obeys

$$Y = (\phi\Psi A_L L_Y)^{1-\alpha}\left(\frac{K}{b_L}\right)^\alpha. \tag{3.94}$$

Given b_L, the only difference to the closed economy is again the additional parameter Ψ (cf. (3.38)).

Aggregate revenues, profits, and market values. From (3.80) and (3.90), the total payments for intermediate goods in each economy amount to

$$\int_0^{\tilde{A}} R(j)dj = \int_0^{A_L} R_L(j) + \int_{A_L}^{\tilde{A}} R_E(j)dj.$$

Trade balance gives that the revenues of foreign producers (the second term) equal the revenues of domestic firms from exporting,

$$\int_{\tilde{A}-A_L}^{\tilde{A}} R_E(j)dj = \int_0^{A_L} R_E(j)dj,$$

where $R_E(j) = 0$ for all firms with $b(j) > b_E$. Denoting the total revenues accruing to firm j, i.e. $R_L(b,\cdot)$ for $b_L \geq b > b_E$, and $R_L(b,\cdot) + R_E(b,\cdot)$ for $b \leq b_E$ by $R(j)$, it thus has to hold that

$$\int_0^{\tilde{A}} R(j)dj = \int_0^{A_L} R(j)dj.$$

Rewriting this relation in terms of productivity levels we obtain

$$\int_0^{A_L} R(j)dj = A_L\int_0^{b_L} p_L(b,\cdot)x_L(b,\cdot)d\mu(b) + A_L\int_0^{b_E} p_E(b,\cdot)x_E(b,\cdot)d\mu(b).$$

Using the equilibrium pricing rules and demands, the revenues of domestic producers become[58]

$$\int_0^{A_L} R(j)dj = A_L\alpha^{2\epsilon-1}r^{1-\epsilon}L_Y\left[\int_0^{b_L} b^{1-\epsilon}d\mu(b) + \int_0^{b_E}(\tau b)^{1-\epsilon}d\mu(b)\right].$$

Applying the Pareto specification and integrating again, we equivalently have

$$\int_0^{A_L} R(j)dj = A_L\alpha^{2\epsilon-1}(rb_L)^{1-\epsilon}L_Y\phi\left(1+\tau^{1-\epsilon}\psi^{\theta-\epsilon+1}\right).$$

The last term again equals Ψ. Using the equilibrium profits in (3.90), aggregate revenues can be expressed as

$$\int_0^{A_L} R(j)dj = \Psi\phi A_L R_L(b_L,\cdot).$$

Since profits are a fraction $1-\alpha$ of revenues, we also have

$$\int_0^{A_L} \pi(j)dj = \Psi\phi A_L\pi_L(b_L,\cdot),$$

so that, along a balanced growth path, (3.84) implies

$$\int_0^{A_L} v(j)dj = \Psi\phi A_L v_L(b_L,\cdot).$$

Given the cutoff, accounting for costly international trade simply adds the factor $\Psi \leq 2$ here as well. The average firm value is $\int_0^{A_L} v(j)dj/A_L = \Psi\phi v_L(b_L,\cdot)$. In the absence of trade (so that $\Psi = 1$) we are back in the closed economy where ϕ relates the value of firms with the cutoff productivity to the average value of firms in the local market. Under free trade, the average firm value equals $\int_0^{A_L} v(j)dj/A_L = 2\phi v_L(b_L,\cdot)$. The average firm value of producers that are only selling to the local market as a function of b_L was derived in the closed economy and equals $A_L\int_0^{b_L} v_L(b,\cdot)d\mu(b) = \phi v_L(b_L,\cdot)$. The domestic cutoff in the open economy, however, will differ from the cutoff in autarky.

[58] We could take the following shortcut here: replacing the term in squared brackets with the expression from the derivation of aggregate output, the revenues of local producers amount to

$$\int_0^{A_L} R(j)dj = YL_Y^{\alpha-1}A_L^{-1}\alpha^{-2\alpha\epsilon}r^{\alpha\epsilon}L_Y^{-\alpha}A_L\alpha^{2\epsilon-1}r^{1-\epsilon}L_Y = \alpha Y.$$

Since aggregate profits are a fraction $1-\alpha$ of aggregate revenues, aggregate profits are

$$\int_0^{A_L} \pi(j)dj = \alpha(1-\alpha)Y.$$

In an equilibrium with balanced growth, using (3.84), we have

$$\int_0^{A_L} v(j)dj = \alpha(1-\alpha)Y/(r-\hat{v}).$$

This directly yields the open economy $\nu\nu = (1-\alpha)Y/\int_0^{A_L} v_i(b,\cdot)d\mu(b_L^*) = (1-\alpha)Y/[\Psi\phi A_L v_L(b_L^*)]$.

Entry into R&D for given cutoffs. Equation (3.91) is the key to solve the open economy model in the same block recursive manner as the autarky model. It again allows us to derive the FE and the ZCP conditions as a function of the domestic cutoff only. To begin with, we rewrite the FE condition in (3.88) in terms of productivity levels

$$\int_0^{b_L}\left[v_L(b,\cdot)-A_L^{\chi-1}wF_L\right]dG(b)+\int_0^{b_E}\left[v_E(b,\cdot)-A_L^{\chi-1}wF_E\right]dG(b)=A_L^{\chi-1}wF_B. \tag{3.95}$$

Splitting up the integral terms, integrating out the terms without productivity levels, and rearranging, the FE condition equivalently reads

$$\int_0^{b_L}v_L(b,\cdot)dG(b)+\int_0^{b_E}v_E(b,\cdot)dG(b)=A_L^{\chi-1}w\left[F_B+G\left(b_L\right)F_L+G(b_E)F_E\right].$$

After dividing by $G(b_L)$ and recognizing from the definition of $\mu\left(b\right)$ in (3.29) that $dG\left(b\right)/G(b_L)=\mu'\left(b\right)db=d\mu\left(b\right)$, we obtain

$$\int_0^{b_L}v_L(b,\cdot)d\mu(b)+\int_0^{b_E}v_E(b,\cdot)d\mu(b)=A^{\chi-1}w\left[\frac{F_B}{G(b_L)}+F_L+F_E\frac{G(b_E)}{G(b_L)}\right]. \tag{3.96}$$

The term in squared brackets on the right hand side is, like $F^a(b_L)$ in the closed economy, the quantity of labor necessary to invent and market a new intermediate good in the absence of knowledge spillovers ($\chi=1$). Using (3.91), it can be expressed as a function of the local cutoff only. Substituting for b_E with (3.91) and applying the Pareto specification yields $G(b_E)/G(b_L)=\psi^\theta$. Hence,

$$\frac{F_B}{G(b_L)}+F_L+F_E\frac{G(b_E)}{G(b_L)}=\frac{F_B}{G(b_L)}+F_L+\theta^{-\theta}F_E^{1-\frac{\theta}{\epsilon-1}}F_L^{\frac{\theta}{\epsilon-1}}=\frac{F_B}{G(b_L)}+F_L\left[1+\left(\frac{F_E}{F_L}\right)^{1-\frac{\theta}{\epsilon-1}}\tau^{-\theta}\right].$$

To avoid cumbersome expressions, we keep in mind that T is contained in ψ also, but collect parameters so that the last term on the right hand side becomes

$$\frac{F_B}{G(b_L)}+F_L\left[1+\left(\frac{F_E}{F_L}\right)^{1-\frac{\theta}{\epsilon-1}}\tau^{-\theta}\right]=\frac{F_B}{G(b_L)}+F_L\left(1+\psi^\theta T\right)=\frac{F_B}{G(b_L)+\Psi F_L}\equiv F\left(b_L\right). \tag{3.97}$$

The left hand side of (3.96) can be expressed in similar terms. Using $\hat{v}_i(b,\cdot)=\hat{v}(\cdot)$, $d\mu=\theta b^{\theta-1}/b_L^\theta db$, and (3.91), the average value of a usable blueprint becomes

$$\frac{(1-\alpha)\alpha^{2\epsilon-1}r^{1-\epsilon}L_Y\theta}{(r-\hat{v})b_L^\theta}\left(\int_0^{b_L}b^{\theta-\epsilon}db+\tau^{1-\epsilon}\int_0^{\psi b_L}b^{\theta-\epsilon}db\right).$$

Integrating and rearranging terms yields

$$\frac{(1-\alpha)\alpha^{2\epsilon-1}r^{1-\epsilon}L_Y\theta}{(r-\hat{v})b_L^\theta}\left[\frac{b_L^{\theta-\epsilon+1}}{\theta-\epsilon+1}+\tau^{1-\epsilon}\frac{(\psi b_L)^{\theta-\epsilon+1}}{\theta-\epsilon+1}\right]=\frac{\phi(1-\alpha)\alpha^{2\epsilon-1}(rb_L)^{1-\epsilon}L_Y\left(1+\tau^{1-\epsilon}\psi^{\theta-\epsilon+1}\right)}{r-\hat{v}}.$$

From (3.89), this equals $\phi v_L + \psi^{\theta-(\epsilon-1)}\phi v_E$. Costly trade thus lowers the value of exporting relative to local sales ($\psi \in (0,1]$ and $\theta > \epsilon - 1$). Like in the closed economy, ϕ relates the average market value of exclusively locally selling firms to the market value of firms with the local market cutoff productivity. In order to obtain an aggregation in terms of the local market cutoff, we rewrite the average value of a usable blueprint as

$$\phi v_L \left(1 + \tau^{1-\epsilon}\psi^{\theta-\epsilon+1}\right) = \phi v_L \left[1 + \tau^{-\theta} T^{-\left(\frac{\theta}{\epsilon-1}-1\right)}\right] = \Psi \phi v_L.$$

Again for expositional convenience, we collect parameters in the last equation to get $\phi v_L \left(1 + \psi^\theta T\right)$. Inserting this expression together with (3.97) in the free entry condition, we arrive at

$$\Psi \phi v_L = w A_L^{\chi-1} F(b_L). \tag{3.98}$$

Given the cutoff, all we have to take into account when including international trade between identical countries is that firms now expected present additional profits in the foreign market. Discounted to the present, these profits amount to ψ times the average value of a domestic producer.

We are now in the position to solve the model for its steady-state equilibrium. Upon deriving the steady state, we provide further economic intuition and discuss the reallocation and incentive effects induced by international trade.

Equilibrium

In view of the close and block-recursive relation of b_L and b_E in (3.91), the definition of an equilibrium in the closed economy carries over to the open economy (including b_E, the additional equation is (3.91)). The steady state growth rates are the same as in the closed economy, and $\hat{b}_E = 0$ follows from (3.91). The cutoffs are again instantaneously fixed.

Equilibrium cutoffs. The entry conditions for the local product market and for R&D in (3.87) and (3.98) again determines the local productivity cutoff b_L^* and via (3.91) also b_E^*. In equilibrium, the share of sufficiently productive blueprints is

$$G(b_L^*) = \frac{F_B}{(\phi - 1) F_L \Psi}, \tag{3.99}$$

and the implied cutoffs are

$$\begin{aligned} b_L^* &= \left[\frac{F_B}{(\phi - 1) F_L \Psi}\right]^{\frac{1}{\theta}} b_0, \\ b_E^* &= \psi \left[\frac{F_B}{(\phi - 1) F_L \Psi}\right]^{\frac{1}{\theta}} b_0 \quad (\geq b_L^*). \end{aligned}$$

Inserting the definition of Ψ, the local market cutoff explicitly reads

$$b_L^* = \left[\frac{F_B}{F_L(\phi-1)}\right]^{\frac{1}{\theta}} \frac{b_0}{\left(1+\tau^{-\theta}T^{1-\frac{\theta}{\epsilon-1}}\right)^{-\frac{1}{\theta}}}. \tag{3.100}$$

Similarly, the equilibrium cutoff associated with exporting is given by

$$b_E^* = \left[\frac{F_B}{F_L(\phi-1)}\right]^{\frac{1}{\theta}} \frac{b_0\tau^{-1}T^{-\frac{1}{\epsilon-1}}}{\left(1+\tau^{-\theta}T^{1-\frac{\theta}{\epsilon-1}}\right)^{\frac{1}{\theta}}}. \tag{3.101}$$

Using (3.99) together with these expressions, the average labor requirement for a usable blueprints is found to equal[59]

$$A^{\chi-1}F(b_L^*) = A^{\chi-1}\frac{F_B}{G(b_L^*)} + \Psi F_L = A^{\chi-1}\Psi\phi F_L. \tag{3.102}$$

In the absence of knowledge spillovers, the average labor requirement in R&D is determined by Ψ and the parameters relating average values to the domestic cutoff values (ψ) alone. In this case, the increase in the required productivity of intermediates raises the average labor requirement in R&D. If $\chi = 1$ and there is free trade, the average labor requirement per usable blueprint increases by 100 percent. Barriers to trade relax this necessity by lowering the required minimum productivity.

Dynamic Equilibrium

Laws of motions and steady state. Turning to the dynamic equilibrium, the same transformed variables as in the closed economy apply. The integral in the denominator of ν now goes from 0 to A_L, and $\nu = (1-\alpha)Y/\int_0^{A_L} v_i(b,\cdot)d\mu(b_L^*) = (1-\alpha)Y/[\Psi\phi A_L v_L(b_L^*)]$. Just like in the closed economy, $rK/\alpha{=}\alpha Y$, or $r = \alpha^2 z$. Also, using the adjusted definition of ν, $\dot{K}/K = z - \gamma$ and $\dot{v}/v = \alpha(\alpha z - \nu)$ continue to hold. The innovation costs change from $F^a(b_L^*)wA^{\chi-1} = \phi F_L w A^{\chi-1}$ to $F(b_L^*)wA^{\chi-1} = \Psi\phi F_L w A_L^{\chi-1}$, and this only affects the law of motion of A_L in terms of $\tilde{l}$: $\dot{A}_L/A_L = \tilde{l}/(\Psi\phi F_L) - \nu$. Accordingly, the laws of motion for γ and z are still given by (3.53) and (3.60), where ν in (3.60) refers to the adjusted definition from above. The law of motion for $\tilde{l}$ then becomes

$$\dot{\tilde{l}} = \tilde{l}\left[n - \chi\left(\frac{\tilde{l}}{\Psi\phi F_L} - \nu\right)\right].$$

[59]For free entry into R&D to be in line with the local market entry condition, the impact of trade on the average firm value (relative to the local cutoff productivity type firm value) must be equal to the impact of trade on the average development costs (relative to the entry costs), i.e. Ψ must drop out in the free entry into R&D condition, compare (3.98) using (3.102).

Updating the law of motion for ν with $\dot{A}_L/A_L$ in the open economy, we obtain

$$\dot{\nu} = \nu\left[\left(\frac{1-\alpha}{\alpha}\chi - 1\right)\frac{\tilde{l}}{\Psi\phi F_L} + \left(2 - \frac{1-\alpha}{\alpha}\chi\right)\nu + (1-\alpha)z - \gamma\right].$$

Accordingly, the steady state in the open economy is the steady state in the closed economy in terms of ν^*, z^*, γ^*, and l^*. Ψ only affects the resources in R&E and

$$\tilde{l}^* = \frac{\Psi\phi F_L}{\chi}\left[\Delta + \frac{1+\alpha}{\chi}n\right].$$

3.5 Trade Liberalization

To ease the exposition, suppose without loss of generality in the following analysis that $b_0 = 1$ and $F_B = F_L(\phi - 1)$. Then, the cutoffs simplify to

$$b_L^* = \frac{1}{\left(1+\tau^{-\theta}T^{1-\frac{\theta}{\epsilon-1}}\right)^{\frac{1}{\theta}}}, \tag{3.103}$$

$$b_E^* = \frac{\tau^{-1}T^{-\frac{1}{\epsilon-1}}}{\left(1+\tau^{-\theta}T^{1-\frac{\theta}{\epsilon-1}}\right)^{\frac{1}{\theta}}}. \tag{3.104}$$

With free trade, $\tau = T = 1$, and all firms are exporting ($b_E = b_L$). In this case, output in both economies is produced using $\tilde{A} = 2A_L$ intermediate goods at any point in time.

We now consider policy induced changes in the barriers to trade and show the presence of a Melitz (2003)-type reallocation towards the more productive firms.

3.5.1 Cutoffs and Industry Reallocation

Differentiating (3.104) exemplarily with respect to τ yields

$$\frac{\partial b_E^*}{\partial \tau} = \frac{-\left(1+\tau^{-\theta}T^{1-\frac{\theta}{\epsilon-1}}\right)^{\frac{1}{\theta}}\tau^{-2}T^{-\frac{1}{\epsilon-1}} + \tau^{-1}T^{-\frac{1}{\epsilon-1}}\left(1+\tau^{-\theta}T^{1-\frac{\theta}{\epsilon-1}}\right)^{\frac{1}{\theta}-1}\tau^{-\theta-1}T^{1-\frac{\theta}{\epsilon-1}}}{[.]^2} < 0.$$

The derivative with respect to T has the same sign.[60] They are negative, implying that a decrease in both types of trade costs increases b_E^* and hence lower the minimum productivity requirement necessary

[60] More precisely, to analyze the effect of a decrease in TBTs, we take the derivative with respect to F_E and evaluate it at $F_E \geq F_L$. The sign of the derivative with respect to T is $sgn\left[-\tau^{-2}T^{-\frac{1}{\epsilon-1}} + \tau^{-1}T^{-\frac{1}{\epsilon-1}}\left(1+\tau^{-\theta}T^{1-\frac{\theta}{\epsilon-1}}\right)^{-1}\tau^{-\theta-1}T^{-1-\frac{\theta}{\epsilon-1}}\right] = sgn\left(\frac{\tau^{-\theta}T^{1-\frac{\theta}{\epsilon-1}}}{1+\tau^{-\theta}T^{1-\frac{\theta}{\epsilon-1}}} - 1\right) = -1.$

for profitable exporting. A simple inspection of (3.103) shows that a reduction in the barriers to trade has the opposite effect on the local cutoff:

$$\frac{\partial b_L^*}{\partial \tau} > 0, \quad \frac{\partial b_L^*}{\partial T} > 0.$$

A decrease in either τ or T raises the minimum productivity requirement for all firms. Taken together, trade liberalization allows more firms to export profitably (and implies an increase in the intensive margin), but at the same time requires all newcomers to be more productive.

Result 3.9 (**Reallocation**). *Trade liberalization lowers the productivity requirement for exporting, but increases the minimum productivity requirement for newcomers.*

The implied reallocation of resources from less productive firms towards more productive firms is the same as in Melitz (without productivity growth) and Baldwin and Robert-Nicoud (2007) (with fully endogenous steady state growth and scale effects). Including capital in production, factor prices are irrelevant for the determination of the productivity cutoffs as long as R&D and entry are conducted with identical production functions. Therefore, Gustafsson and Segerstrom (2007) find exactly the same cutoffs.

3.5.2 Labor Shares

A direct implication of identical steady state values in autarky and in the open economy is that the allocation of labor between production and R&E is not affected by the exposure to trade. Moreover, for the same reasons as argued in Section 3.3.7), the allocation of labor between R&D and entry is not affected by a reduction of trade costs.

Result 3.10 (**No impact on labor shares**). *The allocation of labor in R&D, market entry, and production is not affected by the degree of trade exposure.*

3.5.3 Trade Liberalization and the Incentives to Innovate

We have seen three channels by which trade openness affects the incentives to innovate. First, trade liberalization increases the minimum productivity requirement for all firms and thereby raises the average discovery costs of newly invented varieties. Second, at the same time, the expected value of a usable blueprint increases because a reduction of TBTs lowers the entry costs for the export market. Third, this innovation enhancing effect from an increase in the returns to successful R&D is reinforced by the fact that more blueprints can be used to launch a profitable export business.

Growth effects under fully endogenous growth. Baldwin and Robert-Nicoud (2007) study the growth effects of trade liberalization in a fully endogenous growth framework (with scale effects). They find that openness to trade is growth enhancing if and only if the expected sunk cost of R&E decrease (their Result 1, p. 10) as a result of trade liberalization.

The sunk costs of R&E consist of the quantity of workers necessary to conduct R&E for marketable blueprints, magnified by the impact of spillovers, and the associated wage. Baldwin and Robert-Nicoud (2007) state that the actual labor requirement unambiguously increases if a country opens up to trade incrementally. The impact on the price for R&E depends on the exact specification of the engine of growth, but is likely to be positive. In the Grossman-Helpman specification (Grossman and Helpman, 1991b, Ch.3), the net-effect permanently depresses growth.

Incentive effects under semi-endogenous growth. In our formulation with Jones (1995) technology as the engine of growth, the impact of trade exposure on the labor requirement is similar. The innovation enhancing reduction in the labor requirement for entry is offset by the increase in the labor requirement due to a lower local cutoff:[61]

$$F(b_L^*) = \Psi \phi F_L = \frac{\phi}{\phi - 1} \Psi F_B$$

increases as T and/or τ decrease (which increases Ψ). If trade where free, $F(b_L^*) = 2\phi F_L$.

Gustafsson and Segerstrom (2007) study how trade affects the level of total productivity (which in their model coincides with per capita consumption), i.e. variety growth and the productivity in production. To do so, they compare per capita consumption along the steady-state path of two economies which exclusively differ in terms of the trade costs. Since the ratio is constant over time, they conclude that trade increases productivity if and only if the path associated with lower barriers to trade has higher per capita consumption. This conclusion depends on the strength of knowledge spillovers in R&D. If spillovers are sufficiently strong, trade liberalization retards productivity growth in the short run (and makes consumers worse off in the long-run). This assertion is true in our model also because physical capital in the production of intermediates does not alter the effects of trade on the incentives to innovate. This is because the assumption of identical technologies in R&D and entry "exogenizes" the determination of the cutoff.

[61] The second equality follows by replacing F_L with its normalized value in terms of F_B.

3.6 Conclusion

Recap. We described a specific environment to investigate how trade affects endogenous R&D in a dynamic model with heterogeneous firms and costly trade. Focusing on trade in non-durable intermediate goods, we highlighted important features of the production structure when firms use physical capital and showed that trade in intermediate goods and a careful introduction of capital in production does not alter previous findings on the reallocation of resources and the selection of firms. We further clarified that, albeit convenient, the assumption of identical production technologies in R&D and entry is restrictive. In particular, the labor shares between innovation and market entry are fixed independently of the level of barriers to entry. Technically speaking, the cutoff productivity levels are determined independently of other endogenous variables. This disentangles the cutoffs from the particular production structure.

Avenues for future research. The present chapter provides a framework for various robustness checks. In particular, including trade in the final good and trade in durables is a straightforward extension. More importantly, however, the environment described in this chapter allows us to include physical capital and/or units of output as an input in the entry process. A second extension concerns the average R&D costs approach. Recall that inventing a blueprint is costly, but the productivity implied by the discovered blueprint was a random draw. Hence, there is no intentional investment in productivity, and high productivity types come "for free". Exploring the trade-off between high productivities and the number of usable blueprints at the level of an individual researcher and thereby allowing for purposive investment in productivity is a promising task. Finally, stripping down the model to its essential ingredients is an important step in building a model that is both in line with empirical evidence and also amenable to a thorough welfare analysis. Such a model is necessary to assess the suitability of trade and welfare measures in empirical work. We leave these important challenges for future work.

Appendix 3.A Derivation of ν^*, z^*, l^*, and γ^*

Combining (3.69) and (3.71) yields

$$\begin{aligned}
\left(\frac{1}{\alpha}-\frac{1}{\sigma}-\frac{1}{\chi}-\frac{1}{\alpha}\right)n+\alpha\left(\frac{\alpha}{\sigma}-1\right)z^*+\alpha z^* &= \frac{\rho}{\sigma} \\
-\left(\frac{1}{\sigma}+\frac{1}{\chi}\right)n+\left[\alpha\left(\frac{\alpha}{\sigma}-1\right)+\alpha\right]z^* &= \frac{\rho}{\sigma} \\
\frac{\rho}{\sigma}+\left(\frac{1}{\sigma}+\frac{1}{\chi}\right)n &= z^*\frac{\alpha^2}{\sigma} \\
\rho+\left(1+\frac{\sigma}{\chi}\right) &= \alpha^2 z^*,
\end{aligned}$$

and hence

$$\begin{aligned}
z^* &= \frac{1}{\alpha^2}\left[\rho+n+\frac{\sigma}{\chi}n+\frac{(\sigma-1)n}{\chi}-\frac{(\sigma-1)n}{\chi}\right] \\
&= \frac{1}{\alpha^2}\left[\Delta+\frac{(\chi+\sigma)n-(\sigma-1)n}{\chi}\right] \\
&= \frac{1}{\alpha^2}\left[\Delta+\frac{1+\chi}{\chi}\right].
\end{aligned}$$

Using this expression in (3.69) yields ν^*.

$$\begin{aligned}
\frac{n}{\alpha}+\nu^* &= \frac{1}{\alpha}\left[\Delta+\frac{1-\chi}{\chi}n\right] \\
\nu^* &= \frac{1}{\alpha}\left[\Delta+\frac{1+\chi}{\chi}n-n\right] \\
\nu^* &= \frac{1}{\alpha}\left[\Delta+\frac{n}{\chi}\right].
\end{aligned}$$

This gives $\frac{\chi\tilde{l}}{\phi F_L}=n+\frac{\chi}{\alpha}\left[\Delta+\frac{n}{\chi}\right]$. Hence,

$$\begin{aligned}
\tilde{l}^* &= \frac{\phi F_L}{\chi}\left[n+\frac{\chi\Delta}{\alpha}+\frac{n}{\alpha}\right] \\
&= \phi F_L\left[\frac{n}{\chi}\left(1+\frac{1}{\alpha}\right)+\frac{\Delta}{\alpha}\right] \\
&= \frac{\phi F_L}{\alpha}\left[n\frac{1+\alpha}{\chi}+\Delta\right].
\end{aligned}$$

Finally, from (3.65),

$$\left[\frac{\rho}{\sigma}-\left(1-\frac{1}{\sigma}n\right)\right]-\left(\frac{\alpha^2}{\sigma}-1\right)\frac{1}{\alpha^2}\left(\frac{1+\chi}{\chi}n+\Delta\right)=\gamma^*.$$

After rearranging and canceling terms, we have

$$
\begin{aligned}
\gamma^* &= \frac{\rho}{\sigma} - n + \frac{n}{\sigma} - \left(\frac{1}{\sigma} - \alpha^2\right)\left(\frac{1+\chi}{\chi}n + \frac{\sigma-1}{\chi}n + \rho\right) \\
&= \frac{\rho}{\sigma} - n + \frac{n}{\sigma} - \frac{1+\chi}{\chi}\frac{n}{\sigma} - \frac{\sigma-1}{\chi}\frac{n}{\sigma} - \frac{\rho}{\sigma} + \frac{1+\chi}{\chi}\alpha^2 n + \frac{\sigma-1}{\chi}\alpha^2 n + \alpha^2\rho \\
&= \frac{1}{\alpha^2}\left[\alpha^2\left(\frac{\rho}{\sigma} - n + \frac{n}{\sigma} - \frac{1+\chi}{\chi}\frac{n}{\sigma} - \frac{\sigma-1}{\chi}\frac{n}{\chi} - \frac{\rho}{\sigma}\right) + \left(\frac{1+\chi}{\chi} + \frac{\sigma-1}{\chi}\right)n + \rho\right] \\
&= \frac{1}{\alpha^2}\left[\alpha^2 n\left(\frac{1}{\sigma} - 1 - \frac{1+\chi}{\chi}\frac{1}{\sigma} - \frac{\sigma-1}{\chi\sigma}\right) + \frac{\chi+\sigma}{\chi}n + \rho\right] \\
&= \frac{1}{\alpha^2}\left[\alpha^2 n\left(\frac{\chi - \chi\sigma - (1+\chi) - \sigma + 1}{\chi\sigma}\right) + \frac{\chi+\sigma}{\chi}n + \rho\right] \\
&= \frac{1}{\alpha^2}\left[-\alpha^2 n\left(\frac{1+\chi}{\chi}\right) + \frac{\chi+\sigma}{\chi}n + \rho\right] \\
&= \frac{1}{\alpha^2}\left\{\frac{n}{\chi}\left[-\alpha^2(1+\chi) + \chi + \sigma\right] + \rho\right\} \\
&= \frac{1}{\alpha^2}\left\{\frac{n}{\chi}\left[(1-\alpha^2)\chi - \alpha^2 + \sigma\right] + \rho\right\} \\
&= \frac{1}{\alpha^2}\left\{\frac{n}{\chi}\sigma + n\left[(1-\alpha^2) - \frac{\alpha^2}{\chi}\right] + \rho\right\} \\
&= \frac{1}{\alpha^2}\left\{\frac{n}{\chi}\sigma - \frac{n}{\chi} + \rho + \frac{n}{\chi} + n\left[(1-\alpha^2) - \frac{\alpha^2}{\chi}\right]\right\} \\
&= \frac{1}{\alpha^2}\left\{\Delta + \frac{n}{\chi}\left[1 + \chi(1-\alpha^2) - \alpha^2\right]\right\} \\
&= \frac{1}{\alpha^2}\left[\Delta + (1-\alpha^2)n\frac{1+\chi}{\chi}\right].
\end{aligned}
$$

Part II

On the Growth and Welfare Effects of Monopolistic Competition

Chapter 4

Monopolistic Competition and Endogenous Growth

4.1 Introduction

In this section, we present the canonical Grossman-Helpman (1991b, Ch. 3) increasing variety endogenous growth model. It was first published as a building block in Grossman and Helpman's book "Innovation and Growth in the Global Economy" in 1991. While the book, which according to Google Scholar received about 4350 citations, was designed to analyze the linkages between innovation, growth, and trade, the variety growth model in Chapter 3 serves as a useful device to answer various questions related to profit driven growth. Today, more than 15 years later, the model remains a helpful modeling tool, mainly because of its marked flexibility and great analytical tractability.

We proceed as follows. Section 4.2 provides a quick overview over the production structure and presents the model. The equilibrium is derived and characterized in Section 4.3. We then turn to its welfare properties in Section 4.4. Section 4.5 assesses the effectiveness of R&D policies in the model. Finally, we recognize that the model has come under criticism for its counterfactual "strong scale" effect implication and conclude this section with some supportive comments.

4.2 The Grossman and Helpman (1991b, Ch. 3) Model

4.2.1 Overview

Consider a closed economy endowed with a single primary factor, labor, in fixed supply. There are two sectors: In R&D, workers invent blueprints for the production of differentiated consumption goods. In manufacturing, workers employ the blueprints and produce consumption goods. Households consume the manufacturing output and value product variety. Holding the total physical amount of inputs constant, the consumers' utility is increasing in the number of different products consumed. R&D is conducted to invent new products and thereby has the potential to raise households' utility, even if the amount of resources and the technologies employed in manufacturing remain constant. Industrial innovation, however, is costly. Romer (1990) has clarified that endogenous technological change necessitates either imperfect product markets or strictly convex production functions to cover the upfront R&D costs. Following Romer (1990), Grossman and Helpman (1991b, Ch. 3) allow for market power in manufacturing (and employ constant returns to scale technologies): Researchers sell newly discovered blueprints exclusively to one manufacturing firm, which remains the monopolistic supplier forever, and different products substitute only imperfectly in consumption. Manufacturing firms thus operate under monopolistic competition, and their rents provide the incentive for R&D.[1] It is thus the intentional, profit-driven innovation that leads to an increasing mass of available blueprints, and hence to an increase in welfare.

Increasing product variety, however, exerts a destructive effect on the incentives for R&D, which hinge crucially on the profits in manufacturing. As more and more products become available, an increasing number of firms competes over limited resources, and profits per firm decline. More severe horizontal competition implies diminishing returns to R&D as forward looking manufacturers lower their bids on newly discovered blueprints if they expect declining returns from selling the good.

Accordingly, sustainable R&D driven growth requires an opposing force to offset the "profit destruction effect" in the long-run. This force is found in the non-rival and non-excludable nature of (at least some forms of technological) knowledge.[2] Following Romer (1990), Grossman and Helpman (1991b, Ch. 3)

---

[1] In this model, competition has an unambiguously negative impact on growth. See quality upgrading models with step-by-step innovation for settings where competition spurs growth in imperfectly competitive markets (Aghion, Harris, Vickers, 1997, and Aghion, Harris, Howitt, and Vickers, 2001). See also Hellwig and Irmen (2001) and Funk (2008) for models with endogenous growth under perfect competition.

[2] For an empirical assessment on R&D spillovers see Griliches (1992) and Coe and Helpman (1995) for the open

include publicly available knowledge as an input in the production of blueprints. In conducting R&D, researchers "stand on the shoulders of giants" and employ technological knowledge uncovered by previous research. As a consequence of the imperfect appropriability of knowledge, the "knowledge capital" accumulates as a by-product of innovation. It is freely assessable by current researchers and increases the marginal productivity of labor in the production of blueprints.[3] Accordingly, if many innovations have been uncovered in the past, less labor is needed to invent a new blueprint so that the R&D cost per blueprint decreases as the number of products blueprint rises over time.

A long run equilibrium is found when the "knowledge spillover effect" exactly balances the "profit destruction effect" and the incentive for R&D remains unchanged. The central trade off is then the allocation of labor between R&D, where it increases the variety of available products, and manufacturing, where it increases the output of existing products. Below, the market economy is shown to provide inefficiently little research and hence too little growth relative to the social optimum.

The model is set in continuous time (as usual, we suppress the time index t whenever this causes no confusion). It is deterministic and solved for a perfect foresight equilibrium. The model environment is described by behavioral assumptions about households and firms, technologies, and a specific institutional structure. Beginning with households, we consider each in turn.

4.2.2 Households

Households supply labor, receive wages and capital income, consume, and save by accumulating financial wealth. The population is exogenously given and constant over time. For simplicity, all households are identical (with respect to preferences, productivity, and initial asset holdings), and we use a representative consumer to describe their behavior.[4] She aims at maximizing the present value of utility flows,

$$U_t = \int_t^\infty e^{-\rho(\tau - t)} \log D(\tau)\, d\tau, \tag{4.1}$$

economy.

[3]Evidently, intertemporal knowledge spillovers from past R&D on current innovation may well be negative so that there is a "fishing out" of ideas. In this case, it takes different sources to sustain growth.

[4]See Caselli and Ventura (2000) for consumer heterogeneity in one-sector representative consumer growth models. Inter alia, they show that agents with homothetic preferences and heterogeneous levels of asset holdings save the same fraction of total wealth independently of the individual level of asset holdings.

where ρ denotes the constant subjective discount rate and $\log D$ is the instantaneous utility function[5] where

$$D(t) \equiv \left[\int_0^{n(t)} x(j,t)^{\alpha} \, dj \right]^{\frac{1}{\alpha}}. \tag{4.2}$$

The Dixit-Stiglitz (1977) index $D(t)$ generates positive demands $x(j)$ for differentiated products $j \in [0, \infty)$ from which only goods $j \in [0, n(t)]$, $n(0) \equiv n_0 > 0$, are available at time t, (the product space is assumed to be continuous). As in earlier sections, $\varepsilon \equiv 1/(1-\alpha)$ measures the constant elasticity of substitution between each pair of available products. As pointed out earlier, consumers with Dixit-Stiglitz (1977) preferences value product diversity.[6] Importantly for the analysis of endogenous growth, (4.2) equivalently implies that the utility derived from a given stock of resources increases with the mass of available products. To see this, suppose that different goods are consumed in equal amounts so that $x(j) = x$. Suppose further that all goods are produced one-to-one from labor. Then, $D(t) = n^{1/\alpha - 1} nx$ or $D/(nx) = n^{1/(\varepsilon-1)}$ is increasing in n even if the resources employed in manufacturing, nx, remain unchanged. In particular, if n increases by 1%, D goes up by $1/\alpha - 1\% > 0\%$.[7]

The range of products grows endogenously due to intentional investment in R&D (and never diminishes). Newly invented products substitute imperfectly for existing products. Ongoing variety growth increases diversity in consumption and thus accommodates consumers' demand for variety. An important characteristic of this type of variety growth is that new products are in no way superior to

[5] Log-utility is used for convenience and can be replaced by a constant intertemporal elasticity of substitution (CIES) utility function. This generalization, however, does not provide additional insights in the present environment and requires further parameter restrictions for the existence of a steady state and boundedness of the utility integral in (4.1), see Kornprobst (2008, Chapter 4).

[6] Following Ethier (1982), D may alternatively be interpreted as a production function for a unique final good, which is assembled (i.e. without further inputs) from specialized inputs $x(j)$. In this interpretation, an increase in $n(t)$ indicates a more specialized production process. Both interpretations are identical. There is an important difference with respect to welfare, however, if D is interpreted as a production function and competitively priced labor is used in the production of the final good as well. Unlike in the love-of-variety/production-by-assembling variant, this alteration implies that if D is produced by a mass of competitive firms from specialized inputs and labor, and labor is priced at the competitive wage rate, there are welfare losses due to monopoly power in manufacturing. This is because mark-ups then differ between the final good and the intermediate good sector. In the love-of-variety/production-by-assembling variant, all relative prices are unbiased so that the allocation is Pareto efficient if we exclude knowledge spillovers (cf. Section 4.4). In what follows, we stick to the somewhat simpler interpretation with love of variety preferences (so that there is no welfare loss from monopoly, cf. Lerner, 1934, or Samuelson, 1965, pp. 239–240, and the discussion in Section 4.4 below).

[7] The elasticity of D with respect to product variety is again fixed by the parameter assumption implicit in (4.2), cf. the reference to Benassy (1996) in footnote 16.

existing products. This is very different from quality ladder growth models, where technologically superior goods replace existing products in the same product line so that innovation implies Schumpeter's notion of "creative destruction". Note, however, that there is some creative destruction in terms of market shares in the increasing variety model. Though there is no product replacement like in quality upgrading growth models, product innovation increases the horizontal competition between existing firms and diminishes their market shares proportionally.

Anticipating the welfare properties of the model, we have thus already uncovered three potential sources of market failure: first, producers have market power; second, newly introduced products increase consumers' utility by more than is reflected in the earnings of the innovator (the "consumer surplus" effect); third, by introducing new products, innovators decrease the incumbent firms' profits (the "profit destruction" effect). These externalities suggests that the market equilibrium is not efficient. We will see below, however, that the particular production structure in this model implies that the allocation is Pareto optimal as long as no intertemporal knowledge spillovers occur. We will also see that knowledge spillovers are the source of sustained growth in this model so that the market equilibrium with growth is inefficient.

In maximizing (4.1), the representative consumer is subject to the flow budget constraint

$$\dot{(V^{-1})} = r\left(V^{-1}\right) + wL - E, \tag{4.3}$$

where r, w, and E denote the market interest rate, the wage rate, and consumption expenditures, respectively, and V^{-1} are the representative household's asset holdings (or, equivalently, the stock market value). A "no-Ponzi game" condition additionally limits "lifetime" consumption to "lifetime" income:

$$\int_t^{\infty} e^{-\left[\int_0^{\tau} r(s)ds - \int_0^{t} r(s)ds\right]} E\left(\tau\right) d\tau \leq \int_t^{\infty} e^{-\left[\int_0^{\tau} r(s)ds - \int_0^{t} r(s)ds\right]} w\left(\tau\right) L d\tau + V^{-1}\left(t\right). \tag{4.4}$$

Each household supplies labor inelastically (leisure does not enter the utility function), and we suppose that the aggregate labor supply is constant and equal to L at each point in time.

Formally, households solve a similar program as in Chapter 3, with the slight difference that L is constant over time (i.e., $n = 0$), aggregate consumption expenditures are denoted E (instead of Lc), and $\sigma \to 1$, i.e. log-utility applies for simplicity (cf. footnote 28 below).

4.2.3 Producers

There are two types of firms: research labs and manufacturing plants.[8] Workers employed in the labs conduct R&D to discover blueprints for the production of new consumer goods. At the aggregate level, this process is treated as certain and described by a deterministic production function. The R&D technology is a "bottleneck technology" in the sense that only a finite number of innovations are made at each instant. There is free entry into R&D, but due to this limiting characteristic, no matter how many researchers pursue the development of new products, the mass of blueprints will never jump discretely but rather increase smoothly over time. Specifically, $n(t)$ increases proportionally with the aggregate amount of labor employed in R&D, L_n:

$$\dot{n} = \frac{L_n}{a}. \tag{4.5}$$

Innovation has a constant marginal product in the labor input; a is a productivity parameter. We will see below that growth under (4.5) comes to a halt in finite time. In Section 4.3.2, we add existing knowledge as a factor in production, which then allows for sustained growth.

Upon discovery, the blueprint for a new good is exclusively sold to one (atomistic) manufacturing firm. Since the knowledge embodied in the blueprint is essential for production, the buyer becomes the only producer of the newly invented good. Further down the road, imitation might become possible at a cost, but we assume that upon imitation, the incumbent and the imitating firm engage in Bertrand competition with identical marginal costs.[9] This leaves the imitating firm with losses so that no costly imitation will ever occur in a perfect foresight equilibrium (for a model with imitation cf. Chapter 5).[10] For simplicity, we also suppose that each firm produces exactly one good/owns one blueprint. Accordingly, we refer to the firm that owns the blueprint for/produces good j as firm j. In contrast

[8] R&D can equivalently be conducted either by independent research labs or in-house, i.e. within the prospective manufacturing firm (see Romer, 1990, p. 82). The distinction matters in the case of quality upgrading growth models since with external R&D, the quality leader has a larger willingness to pay for new blueprints than outsiders (see Kornprobst, 2008, p. 83).

[9] In the so-called North-South growth models, imitation occurs by foreign firms. If the wage rate in the North exceeds the wage rate in the South and identical manufacturing technologies, the imitating firms captures the entire product market. The equilibrium is then characterized by product cycle trade patterns. See, e.g. Grossman and Helpman (1991a), Segerstrom (1991), or, more recently, Dinopoulos and Segerstrom (2007) and Dinopoulos and Kottaridi (2008).

[10] Alternatively, the bulk of the endogenous growth literature assumes perpetual, fully enforced patents to ensure that the initial buyer remains a monopolistic supplier forever.

to Melitz (2003), firms are homogenous in production. All producers use constant returns to scale technologies with an identical level of productivity. By choice of units, we set the labor input coefficient equal to 1. The aggregate labor demand in manufacturing therefore amounts to

$$L_p = nx. \tag{4.6}$$

The instantaneous profits of firm $j \in [0, n]$, which are paid out as dividends to its shareholders, amount to

$$\pi(j) = p(j)\,x(j) - wx(j), \tag{4.7}$$

where $p(j)$ and $x(j)$ denote the price charged and the quantity sold by firm j, respectively. Given the households' demand schedule, whereby $x(j) = x(p(j))$, all active firms $j \in [0, n]$ engage in Chamberlinian monopolistic competition in the product market.

4.2.4 Equilibrium Conditions

Households

Households maximize (4.1) subject to (4.3) and the no-Ponzi game condition. Let $E(t)$ denote the aggregate spending consistent with market clearing on the markets for available consumer goods at time t,

$$E(t) \equiv \int_0^{n(t)} p(j,t)\,x(j,t)\,dj. \tag{4.8}$$

The problem of solving for the optimal demands $x(j)$, $j \in [0, n]$ is the same as in Section 1.2.1. There, we found that given the aggregate consumption expenditures E and the ideal price index $P = \left[\int_0^n p(j)^{1-\varepsilon}\,dj\right]^{1/(1-\varepsilon)}$, D is maximized by choosing

$$x(j) = \frac{Ep(j)^{-\varepsilon}}{P^{1-\varepsilon}}. \tag{4.9}$$

Recall that these demands feature a uniform elasticity of substitution across goods (ε), a constant price elasticity of demand across all quantities demanded (ε), and a constant expenditure elasticity (equal to 1).[11] Note, in particular, that the price elasticity of demand for any variety is independent of the mass of available products (i.e. there is "large group" monopolistic competition in the sense that a single firm has a sufficiently small market share).

[11] As mentioned earlier, the demand functions $x(j)$ can be derived separately for each consumer. Aggregating over all consumers then yields the demands given above.

Since all firms have identical production technologies and the demand functions are the same for all products, the equilibrium is symmetric with regard to firms:

$$p(j) = p, \quad x(j) = x, \quad \pi(j) = \pi.$$

Accordingly,

$$D = n^{\frac{1}{\alpha}} x, \quad P = n^{-\frac{1}{\varepsilon-1}} p = n^{-\frac{1-\alpha}{\alpha}} p, \quad E = npx = DP.$$

Substituting for $D = E/P$ in (4.1), the households' utility obeys

$$U_t = \int_t^\infty e^{-\rho(\tau-t)} \log E(\tau)\, d\tau - \int_t^\infty e^{-\rho(\tau-t)} \log P(\tau)\, d\tau.$$

Hence, $\partial U/\partial E$ is independent of P. The intertemporal maximization problem can therefore be split into two separate steps. First, given E and P, optimality requires that households choose quantities $x(j)$ such that D is maximized, i.e. according to (4.9). Second, the path of expenditures $\{E\}_{\tau=0}^\infty$ can be chosen given P.[12] A similar problem was discussed in Chapter 3. Its solution yields the standard Euler equation for expenditure growth, $\dot{E}/E = (r-\rho)/\sigma$ (as well as a transversality condition to pin down the solution to this differential equation and equality in (4.4)).[13] With log-utility ($\sigma \to 1$), the optimal expenditure path boils down to $\dot{E}/E = r - \rho$. The model is most easily solved by choosing E as the numéraire. With $E(t) \equiv 1$, the optimal expenditure path pins down the interest rate. From $\dot{E} = 0$ in the Euler equation,

$$r(t) = \rho. \tag{4.10}$$

In anticipation of the dynamic equilibrium, note that $E(t) = 1$ also implies that $\hat{w} = \hat{p} = 0$ and $\hat{\pi} = -\hat{x}$ if the sectoral allocation of labor between R&D and manufacturing remains constant (since then $\hat{n} = -\hat{x}$ from $\hat{L} = 0$). Moreover, in this case, $\hat{P} = -\hat{D} = -1/(\varepsilon-1)\,\hat{n}$.

Production

Manufacturing. Manufacturing firms maximize the profits (4.7) subject to the consumers' demands in (4.9). We solved this problem earlier for the more general case of input coefficients in manufacturing

[12] This is a special property of the log-utility in (4.1). In the more general CIES case with $u(D) = D^{1-\sigma}/(1-\sigma)$, $u'(D) = E^{-\sigma}/P^{1-\sigma}$, the necessary condition for optimal consumption is that $u'(D)$ grows at the same rate as the shadow price of wealth, $\rho - r$. Accordingly, $-\sigma\hat{E} - (1-\sigma)\hat{P} = \rho - r$ or

$$\hat{E} = \frac{r-\rho}{\sigma} - \frac{1-\sigma}{\sigma}\hat{P}.$$

With log-utility ($\sigma \to 1$), expenditure growth is independent of $\hat{P}$ and $\hat{E} = r - \rho$.

[13] Cf. footnote 28 with $n = 0$ and $Lc = E$.

equal to $1/\varphi$, (cf. Chapter 1). We found that firms charge a constant mark-up over marginal cost, so for $\varphi = 1$ we have

$$p = \frac{w}{\alpha}. \tag{4.11}$$

In equilibrium and with expenditures as numéraire, the quantity demanded of each good is thus $x = 1/(np)$ or, using (4.11),

$$x = \frac{\alpha}{nw}.$$

Substituting for w with (4.11) in (4.7) reveals that profits are a fraction $1-\alpha$ of revenues, $\pi = (1-\alpha)\, px$. Since $px = 1/n$, the instantaneous profit of each producer equals

$$\pi = \frac{1-\alpha}{n}. \tag{4.12}$$

Hence, the instantaneous aggregate profits or, equivalently, dividend income $n\pi$ is a constant fraction $1-\alpha$ of consumption expenditures ($E = 1$).

Equation (4.12) demonstrates the "profit destruction effect": product innovations increase horizontal competition and decrease the profits of incumbent firms. Profits are lower if the elasticity of demand, ε, is high (so that α is high).

Let $v(t) \equiv \int_t^{\infty} e^{-\rho(\tau-t)} \pi(t)\, d\tau$ denote the present value of a firm's profits. In the absence of speculative bubbles and uncertainty, the capital market prices each firm with value v.[14] Since manufacturing firms are willing to pay up to v for each newly discovered blueprint, v is also the equilibrium revenue for the innovating firms, i.e. the market price for blueprints. Taking the derivative with respect to time t, v evolves according to

$$\dot{v} = \rho v - \pi. \tag{4.13}$$

The value of blueprints declines over time if the attainable profits from its use in manufacturing exceed the annuity equivalent of all future profits/dividend payments. In view of (4.12) and the definition of v, $\dot{v} < 0$ if n increases over time, i.e. if innovation occurs. Since (4.13) equivalently states that bonds and equity are indeed perfect substitutes (the return on an investment of v in the bond, $r = \rho$, equals the equivalently certain return of an investment of v in any firm's equity, i.e. the sum of capital gains $\dot{v}$ and dividend payments π relative to v, $(\dot{v} + \pi)/v = \rho$), it is often named "capital market equilibrium condition". This wording is somewhat misleading since any equity price that equates supply and demand implements an equilibrium in the capital market.

[14] With infinitively lived agents, equity prices do not include bubbles, see Grossman and Yanagawa. For a similar result cf. Blanchard and Fisher (1989, Ch. 5).

R&D. Free entry into R&D and profit maximization in the research labs imply that labor demand is unbounded if the (instantaneous) return on newly discovered blueprints exceeds the invention costs, i.e. if $v\dot{n}\ dt > wL_n\ dt$. Since $\dot{n}$ blueprints require $L_n = \dot{n}a$ units of labor, see (4.5), we know that $v \leq wa$ holds in equilibrium. If $v < wa$, however, no researcher would rationally incur the R&D cost and $L_n = \dot{n} = 0$. Summarizing, in equilibrium,

$$wa \begin{cases} = v & \text{if } \dot{n} > 0 \\ \geq v & \text{if } \dot{n} = 0 \end{cases} . \tag{4.14}$$

Due to the normalization of expenditures, wages must decline with v as long as ongoing innovation decreases the return on R&D.

Resource Constraint

Finally, labor market clearing (i.e. the aggregate resource constraint) requires labor in R&D and labor in production, $L_p = nx$, to sum up to the aggregate labor endowment, $L_n + L_p = L$. After substituting for L_n using (4.5), the resource constraint reads

$$L = a\dot{n} + L_p. \tag{4.15}$$

4.2.5 Model Structure

The model involves two state variables, n and v. While n is given by historical innovations (and the exogenous initial value $n(0) = n_0$) at each instant, v continuously adjusts according to what agents believe about the future rates of return to R&D. These "beliefs" involve no uncertainty, so v can be calculated by the agents in the model. Given n and v, we can solve for the equilibrium labor allocation and prices at each instant, and (4.5) and (4.13) connect the resulting "instantaneous equilibria" through time.

Note that given our assumptions on technologies, $v > 0$ provides the only incentive for R&D (the sufficient condition is that n is sufficiently small).[15] That is, in equilibrium, the "if" in (4.14) goes in both directions: $v \to 0$ implies $\dot{n} \to 0$.

[15] Cf. footnote 2 on necessary conditions for innovation-driven growth.

4.3 Equilibrium

An equilibrium in this economy is a sequence of quantities $\{L_p, L_n, x, n\}_t^\infty$ and prices $\{p, w, r, v, \pi\}_t^\infty$ that satisfies[16]

- agents' optimality conditions
  - expenditure is chosen optimally, $r(t) = \rho$
  - prices maximize profits, $p = w/\alpha$
- technologies
  - in R&D, $\dot{n} = L_n/a$
  - in production, $nx = L_p$
- the absence of bubbles (to rule out divergent paths), $\rho v = \dot{v} + \pi$ where profits are given by $\pi = (1-\alpha)/n$
- bounded labor demand in R&D with free entry: $wa \geq v$ if $\dot{n} = 0$ and $wa = v$ if $\dot{n} > 0$
- Market clearing
  - at the labor market, $L = L_p + L_n$.
  - at the markets for consumption goods, $nxp = 1$

We begin to characterize the equilibrium by deriving the necessary prerequisites for profit driven growth in this environment.

4.3.1 No Endogenous Growth in the Long-Run

The incumbent manufacturers' profits in (4.12), $\pi = (1-\alpha)/n$, are decreasing in the number of competitors. This profit destruction effect of R&D hints at where the economy is headed in the long-run. Given that the present value of future profits provides the incentive for researchers to conduct R&D, innovation must come to a halt as $n \to \infty$ since then $\pi \to 0$ and $v \to 0$ along with it.[17] In fact,

[16] As usual, the budget constraint is another, but dependent, equation in the same variables.

[17] This is the Grossman and Helpman (1989) model.

we can easily verify that the threshold number of competitors compatible with profit-driven growth is finite. Let $\bar{n}(t')$ denote the number of incumbents/blueprints at the point in time where innovation stops, i.e. suppose that $n(t'+\tau)=\bar{n}$ for all $\tau \geq 0$. From (4.5), this implies that $L_n = 0$ as of time t', and $1 = npx = pL_p$ from (4.15) and (4.11) give $L = L_p = 1/p = \alpha/w$. In the stationary environment, the wage rate thus equals

$$w|_{L_n=0} = \frac{\alpha}{L}. \tag{4.16}$$

It is increasing in ε (α), reflecting the fact that a high elasticity of substitution/demand implies that firms charge low prices, sell large quantities, and thus demand much labor.

If the number of producers is constant at n, so are the incumbent firms' profits. Since $r = \rho$, the market value of each firm amounts to $v = (1-\alpha)/n \int_t^\infty e^{-\rho(\tau-t)} d\tau$, or,[18]

$$v = \frac{1-\alpha}{\rho n}. \tag{4.17}$$

Finally, without ongoing innovation, the free entry into R&D condition in (4.14) is given by $v \leq wa$. Inserting the wage rate from (4.16) and the market value from (4.17) yields $\alpha a/L \geq (1-\alpha)/(\rho n)$. Solving for n we find that $\dot{n} = 0$ occurs if and only if $n \geq \bar{n}$, where

$$\bar{n} = \frac{(1-\alpha)L}{\alpha a \rho}. \tag{4.18}$$

The economy is able to generate profit-driven innovation as long as $n < \bar{n}$. The threshold value of producers is increasing in the labor endowment and the productivity in R&D ($1/a$). A low elasticity of substitution between the goods as well as a high discount rate are detrimental to $\bar{n}$.

If $n_0 \geq \bar{n}$ to begin with, the economy immediately finds itself in a unique stationary equilibrium (without growth): no innovation occurs ($L_n = 0$) and n remains constant at n_0. Labor market clearing, $L = L_p = \alpha/w$, determines the equilibrium wage rate (see (4.16)). Substituting for w in the optimal output decision of the firms with (4.16), we find that each of the n_0 manufacturers produces $x = 1/(n_0 p) = \alpha/(n_0 w) = \alpha L/(n_0 \alpha)$ units of output and charges a price equal to $p = 1/L$. Letting $\bar{n} = n_0$ in (4.17), the value of manufacturers equals $v = (1-\alpha)/(\rho n_0)$.

"No growth" equivalently obtains if a growing economy reaches the number $\bar{n}$ of producers. The intuition for this is simple. As more firms enter the product market with new goods, households allocate their expenditures across an increasing number of products, market shares and profits per product diminish and along with it the incentives for R&D. Eventually, profits (and thereby the value

[18] $\int_t^\infty e^{-\rho(\tau-t)} d\tau = \frac{\left[e^{-\rho(\tau-t)}\right]_t^\infty}{\rho} = \frac{1}{\rho}$.

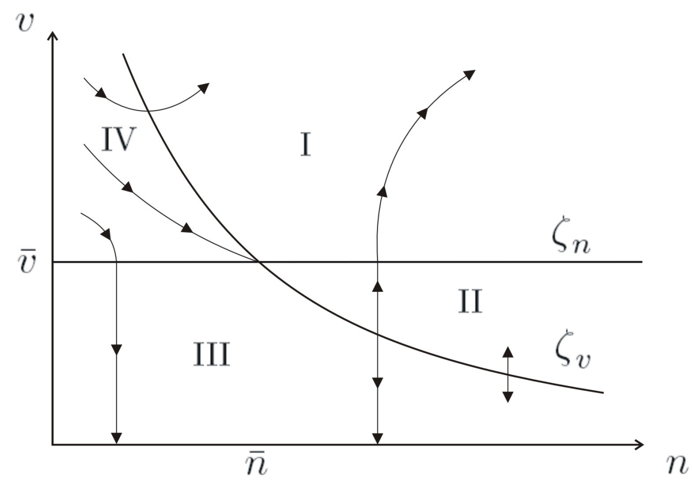


Figure 4.1: Global Dynamics (Adapted from Grossman and Helpman, 1991b, p. 56)

of blueprints) are so low that no research is conducted and all labor is employed in manufacturing. This occurs at the point in time when n reaches $\bar{n}$ or, using (4.18) with equality in (4.17)), if v reaches $v = \alpha a/L$.[19]

Put differently, growth of n is positive if and only if $n < \bar{n}$ or, equivalently, if and only if

$$v > \frac{\alpha a}{L} \equiv \bar{v}. \tag{4.19}$$

The dynamics of the two-dimensional system are captured by two differential equations in n and v. First, if (4.19) holds, the mass of innovations at each point in time is

$$\dot{n} = \frac{L - \frac{1}{p}}{a} = \frac{L}{a} - \frac{\alpha}{aw} = \frac{L}{a} - \frac{\alpha}{v} \quad (> 0)\,. \tag{4.20}$$

Second, using (4.12) in (4.13), the market value of blueprints changes by

$$\dot{v} = \rho v - \frac{1-\alpha}{n}. \tag{4.21}$$

Given n_0, the two ordinary differential equations in (4.20) and (4.21) completely describe the dynamic behavior of the system and are amenable to a simple phase diagram analysis (see Figure 4.1).

From (4.21), v is constant on

$$\zeta_v\,(n) \equiv \frac{1-\alpha}{\rho n},$$

[19]Equivalently, if $\dot{n} > 0$, free entry into R&D requires $v = wa = \alpha a p$ and, since $L_n > 0$, $v/\,(\alpha a) = p > 1/L$ from labor market clearing. Hence $\dot{n} > 0$ requires $v > \alpha a/L \equiv \bar{v}$.

and only if we impose non-negativity, $\dot{v} = 0$ at $v = 0$ (otherwise $\dot{v}(0) < 0$). ζ_v is downward sloping from ∞ to zero on $(0,\infty)$ in (n, v)-space. v increases above ζ_v and decreases below. With respect to n, we have seen above that no innovation occurs if $v \leq \bar{v}$, see (4.19). The border of this region,

$$\zeta_n \equiv \bar{v},$$

is horizontal in (n, v)-space so that ζ_v and ζ_n uniquely intersect in the first quadrant at $(\bar{n}, \bar{v})$. If $v > \bar{v}$, the rewards to R&D spur innovation.

ζ_v and ζ_n separate (the first quadrant of) the (n, v)-space in four regions. In the North-East region (region I), all trajectories diverge and imply $v \to \infty$, $n \to \infty$. In the South-East region (region II), all trajectories are vertical and eventually enter region I. Below both ζ_n and ζ_v, in the South-West region (region III) all trajectories are also vertical but point to the South. The dynamics in this region eventually implies $v \to 0$ at a constant n. Finally, in the North-West region (region IV), trajectories point to the South-East. Accordingly, a saddle path leading to $(\bar{n}, \bar{v})$ exists. Trajectories above and below this path eventually enter region I and region III, respectively. We show below that perfect foresight allows us to sort out all paths where $v \to \infty$ or $v \to 0$, so that we can dismiss regions I and III and all paths leading to these regions.

If the economy starts with $n_0 \geq \bar{n}$, we have already seen that prices immediately adjust so that the equilibrium without growth is reached where $n = n_0$ and $v = \zeta_v(n_0)$ forever. If $n < n_0$ and given that we can rule out explosive paths, we know from the phase diagram analysis that v jumps to its value on the saddle path, innovation occurs, and the economy converges gradually along the saddle path towards the steady state $(\bar{n}, \bar{v})$ where finally $\dot{n} = 0$.

The main message of this section is the following. No matter how well endowed the economy is with labor, in the long-run, diminishing returns to R&D bring profit-driven growth to a halt. In particular, growth cannot be sustained because the incentives for R&D diminish as innovations decrease the future profits of manufacturing. With $\dot{n} > 0$, π/v converges towards $\pi/\bar{v} = \pi\rho/\bar{\pi}$ ($\dot{v} = -\pi > \dot{\pi}$) and eventually, as $n \to \bar{n}$, the rate of return to R&D converges to $r = \rho$.

4.3.2 Sustainable Growth with Knowledge Spillovers

Up to now, labor was the only input in the production of blueprints. In particular, technical knowledge was implicitly assumed to be perfectly appropriable: each researcher starts from scratch and sells all the knowledge he gathers in the form of a blueprint. Romer (1990) argues that R&D not only produces this type of excludable knowledge. R&D rather generates lots of techniques and insights that not only apply to the particular research project at hand. Put differently, R&D involves a lot of existing knowledge and "standing on the shoulders of giants" (i.e. researchers benefiting from the achievements of others). Following Romer, Grossman and Helpman assert that R&D uses a substantial amount of "public knowledge" that is created automatically along with the private, patentable knowledge that leads to new varieties. Somewhat more accurate, "public knowledge" refers to available knowledge that did not originate from the given R&D project, but is freely available to the researchers.

In what follows, we include non-excludable knowledge as an input in the production of blueprints (Grossman and Helpman, 1991b, Sec. 3.2). The accumulated, non-excludable part of R&D output is thereby referred to as knowledge capital. Its flows arise steadily as a by-product of purposive R&D. The stock of available knowledge capital serves as a freely accessible input in the production of blueprints. Knowledge capital – at first similar to physical capital – increases the marginal productivity of labor. The crucial difference, however, is the non-rivalry of knowledge (Romer, 1990). If Mr. Smith uses the particle accelerator to speed up electrons, Mr. Jones will have to wait to smash atoms himself, but both Smith and Jones may use Newton's law of motions at the same time.[20]

Introducing Knowledge Capital in R&D

In their baseline model, Grossman and Helpman (1991b, Sec. 3.2) assume strong positive externalities from past R&D on current product development. Abstracting from dissemination lags and non-linearities, knowledge capital is simply assumed to be directly proportional to the number of previous research projects, n. By appropriate choice of units, the factor of proportionality is set to one so that n equivalently denotes the knowledge available. The knowledge capital enters the blueprint production function linearly, so that the labor requirement per blueprint declines exponentially if n grows

[20]This example, however, immediately points to a severe shortcoming of the pure externality approach to sustained growth. In the model, knowledge is a freely available input, but in reality it surely takes costly resources to allow researchers to apply the body of knowledge uncovered by others (cf. Scotchmer, 1991). Moreover, in theory, we can work with positive knowledge spillover effects as well as with negative "fishing out" effects whereby research into new products becomes increasingly harder the more innovations have been made in the past (cf. Jones, 1995a, Segerstrom, 1998, and the empirical evidence in Caballero and Jaffe, 1993, and Kortum, 1993). We can further think of various dissemination effects that generate all kinds of growth patterns. While some functional forms can be excluded on theoretical grounds, more detailed empirical research on the production function is required to further open the black box of R&D. In particular, the literature lacks the micro-foundations of a R&D production function when it comes to the trade-off between the number and the quality of innovations.

exponentially. In particular, (4.5) is replaced by[21]

$$\dot{n} = \frac{nL_n}{a}. \tag{4.22}$$

This modification of the R&D production function leaves firms' and households' decisions unaffected and only affects the resource constraint and the free entry into R&D condition. Given (4.22), a/n units of labor are sufficient to produce one blueprint. Accordingly, replacing $L_n = a\dot{n}$ by $a\dot{n}/n$ from (4.22), the labor market clearing condition in (4.15) becomes

$$L = a\frac{\dot{n}}{n} + \frac{1}{p}. \tag{4.23}$$

Correspondingly, the R&D cost per blueprint reduces to wa/n so that free entry into R&D now says

$$wa \begin{cases} = vn & \text{if } \dot{n} > 0 \\ \geq vn & \text{if } \dot{n} = 0 \end{cases}. \tag{4.24}$$

Following Grossman and Helpman (1991b, Sec. 3.2), we characterize the equilibrium in terms of $\dot{n}/n \equiv g$ and the auxiliary variable $V \equiv (nv)^{-1}$. The definition of V implies that $\hat{V} = -g - \hat{v}$. Substituting for $\hat{v}$ from (4.21) in terms of V, i.e. $\dot{v}/v = \rho - \pi/v = \rho - (1-\alpha)V$, V evolves according to

$$\frac{\dot{V}}{V} = -(g+\rho) + (1-\alpha)V. \tag{4.25}$$

$V = 0$ implies $\dot{V} = 0$. In analogy to the model without knowledge spillovers, $\dot{n} = ng > 0$ requires v to be sufficiently large (π and hence v continue to decline as n increases). In terms of V, we have from (4.24) with $\dot{n} > 0$ and (4.11) that

$$V = \frac{1}{wa} = \frac{1}{\alpha ap}. \tag{4.26}$$

Again, with ongoing innovation, some resources must be employed in the R&D sector so that $L_p = 1/p > L$. Inserting p from (4.26) we find that $\dot{n} = gn > 0$ if and only if

$$V < \frac{L}{\alpha a} \equiv \bar{V}.$$

Evidently, not all combinations of g and V are feasible ("growth" and "manufacturing" compete for scarce labor). Using the R&D technology (4.22), labor market clearing (4.23), the pricing equation (4.11), and free entry into R&D with $\dot{n} > 0$, we can express the resource constraint as

$$L = L_p + L_n = \frac{1}{p} + a\frac{\dot{n}}{n} = \frac{\alpha}{w} + a\frac{\dot{n}}{n} = a\frac{\alpha}{nv} + a\frac{\dot{n}}{n}.$$

[21] Evidently, the restriction $n_0 > 0$ allows for the possibility of growth.

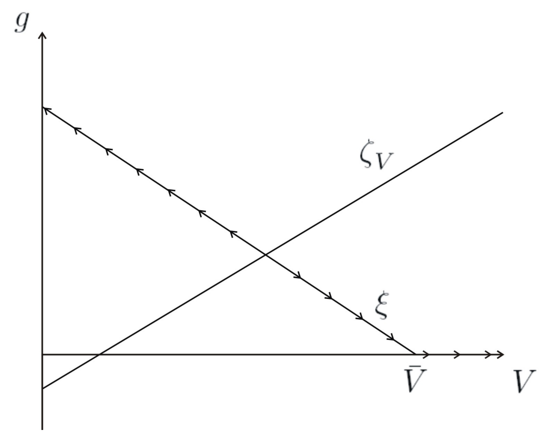


Figure 4.2: Dynamics if the Parameter Assumption (PA) Holds (Adapted from Grossman and Helpman, 1991b, p. 60)

This constraint can be rewritten in terms of V and g:

$$\xi\left(V\right)=\begin{cases}\frac{L}{a}-\alpha V & \text{if } V<\bar{V}\\ 0 & \text{if } V\geq\bar{V}\end{cases}. \tag{4.27}$$

The dynamics of the system is entirely characterized by the evolution of V as described in (4.25). A feasible steady state is found where V remains constant and (4.27) holds. If $V\neq 0$ is constant, (4.25) implies $g=-\rho+(1-\alpha)V$. Similarly, if $V=0$, $\dot{V}=0$ for all g. Hence, from (4.25), V is constant on

$$\zeta_V\left(V\right)=\begin{cases}-\rho+(1-\alpha)V & \text{if } V>0\\ 0 & \text{if } V=0\end{cases}.$$

Depending on parameter values, there are two possibilities. Consider first the case where $\zeta_V\left(0\right)=\rho/\left(1-\alpha\right)<\bar{V}=L/\left(\alpha a\right)$, i.e.

$$\frac{\rho}{1-\alpha}<\frac{L}{\alpha a}. \tag{PA}$$

In this case, ζ_V intersects ξ once from below in the first quadrant so that a unique steady state with positive growth exists, see Figure 4.2. Note that V is decreasing to the left of ζ_V and increasing to the right. If the economy starts not at the intersection of ζ_V and ξ, it converges along ξ to either $V=0$ (if $V(0)$ is less than the intersection value) or $V\to\infty$ (if $V(0)$ exceeds the intersection value). Since $0<n<\infty$ (as $g\to 0$ as $V\to\bar{V}<\infty$), we can sort out both trajectories based on the arguments put forward in Appendix 4.A so that the economy must jump into its steady state initially.

Consider next the case where (PA) does not hold. In this case, ξ lies above ζ_V for all $V<\bar{V}$ (cf. Figure 4.3). If the economy starts out to the left of the intersection of ζ_V and ξ, the system eventually

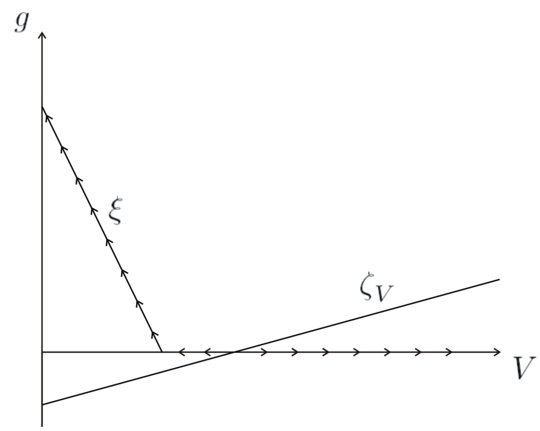


Figure 4.3: Dynamics if (PA) is Violated (Adapted from Grossman and Helpman, 1991b, p. 64)

converges to $V = 0$. If it starts to the right of the intersection, $\xi > \zeta_V$ and $V \rightarrow \infty$. By familiar arguments (see above), both types of paths can be ruled out. We conclude that the economy directly jumps into its steady state where $g = 0$, $V = \bar{V}$ (that is to say, given n_0, v adjusts such that $v = \alpha a / (n_0 L)$).

Let us briefly summarize. The parameter assumption in (PA) distinguishes two cases. If the resource base is sufficiently low so that (PA) is violated, then no growth will ever occur and $(g = 0, V = \bar{V})$ instantaneously constitutes the unique (steady state) equilibrium. If, however, (PA) is satisfied, the economy jumps into a unique steady state with positive growth, and from $\zeta_V (L/a + \rho) = \xi (L/a + \rho)$, the growth rate equals

$$g = (1 - \alpha) \frac{L}{a} - \alpha\rho \quad (> 0) . \tag{4.28}$$

Growth can be sustained in the long-run since the falling returns to R&D due to ongoing innovation are accompanied by a continuous decline in the R&D cost per blueprint. In the steady state equilibrium, the two forces exactly offset each other and the incentives for R&D remain constant. Due to the normalization of expenditures, wages (and thus prices) are constant and equal $w = (L + a\rho)^{-1}$. From g and (4.5), we infer $L_p = (1 - \alpha) L - \alpha a \rho$ and from (4.15), $L_n = \alpha L + \alpha a \rho$.

We now turn to the welfare properties of the equilibria with and without knowledge spillovers.

4.4 Welfare

Recall that we encountered three possible sources of inefficiencies in the model. First, consumers get extra utility from every newly invented good since they value product variety per se. This "consumer surplus effect" of innovations does not accrue to the profit-seeking behavior of firms. Second, newly entering manufacturers decrease the profits of incumbent firms (the "profit destruction effect"). Third, producers have market power and set prices above marginal costs. Absent competitive pricing, however, all mark-ups in the economy are of equal size so that relative prices are unbiased. Hence, as claimed already by Lerner (1934), no distortions arise from the monopolistic price setting (cf. Samuelson, 1965, pp. 239-240).

In the model with knowledge spillovers, there is a fourth reason to expect the equilibrium to be inefficient: current researchers do not account for the positive side effect on future innovations.

To get a grip on the welfare properties of the models with and without knowledge spillovers, we follow Grossman and Helpman (1991b, Ch. 3) and break the labor allocation problem down into a static and a dynamic part. That is, in a first step, we take the sectoral allocation of labor as given and ask how production labor has to be split up among the different products. In a second step, we then ask how labor should be allocated efficiently between R&D and manufacturing over time.

This section is organized as follows. We begin by considering R&D without knowledge spillovers (e.g. perfectly appropriable knowledge). In this environment, the decentral market outcome will be shown to be optimal. We discuss this somewhat surprising result right after its derivation below. Next, we include knowledge spillovers and show that, in this case, our earlier intuition is confirmed: with public knowledge, the private value of innovations falls short of its social return and the equilibrium allocation generates too little growth compared to the social optimum.

4.4.1 Static Efficiency

Welfare is given by the intertemporal utility U_t. To begin with, note that any allocation that maximizes utility over time must necessarily achieve efficiency in a static sense. That is, at any point in time labor must be allocated across all available products so that the quantities produced of all goods maximize D in (4.2).[22] With the marginal gains from $x(j)$ falling and identical production technologies, this evidently requires symmetry across all products: $x(j) = x$ or, equivalently, $x = L_p/n$.

[22]Evidently, the log is a strictly increasing function and attains its highest value if D is maximized.

The market equilibrium is efficient in this static sense. Both with and without knowledge spillovers, $npx = E = 1$ and hence $x = p^{-\varepsilon}/\left(np^{1-\varepsilon}\right) = 1/\left(np\right) = L_p/n$. We are thus left to check whether the allocation of labor between R&D and manufacturing is efficient across time.

4.4.2 Dynamic Allocation with Perfectly Appropriable Knowledge

We next derive the socially optimal paths for the allocation of labor in production and R&D when R&D uses labor as the only factor in production.

Given static efficiency and one-to-one production from labor, the Dixit-Stiglitz index equals $D = n^{1/\alpha}x = n^{(1-\alpha)/\alpha}L_p$. Substituting this expression in the households utility functional in (4.1), the social planners' objective function is given by

$$U_t = \int_t^{\infty} e^{-\rho(\tau-t)} \left[\frac{1-\alpha}{\alpha}\log n + \log L_p\right] d\tau.$$

Due to the log preferences, U_t is bounded away from infinity since growth of n is at most exponential so growth of $\log n$ is linear. In the social planner problem, n is the only state variable and L_p is used as the control. A feasibility constraint was derived in (4.15), taking into account the resource constraint and technologies. Solving (4.15) for $\dot{n}$ gives the transition function for n :

$$\dot{n} = \frac{L - L_p}{a}. \tag{4.29}$$

The current value Hamiltonian for the planner's program is[23]

$$\mathcal{H} = \frac{1-\alpha}{\alpha}\log n + \log L_p + \lambda\left(\frac{L - L_p}{a}\right),$$

where λ denotes the co-state variable. $\mathcal{H}$ measures the flow value of consuming L_p units of n different available products and investing $L - L_P$ units of labor in "new firms" in terms of current utility. By construction, the Hamiltonian also puts a price on the foregone consumption wrapped up in the R&D investment, namely the incremental change in (future) lifetime utility due to a change in n, λ (recall from Bellman's principle that λ is the time derivative of the corresponding value function and that the flow from the optimal choice, i.e. the maximized Hamiltonian, has a return equal to ρ). λ can thus be interpreted as the shadow price of n, which, as will become clear below, is closely linked to the firm's market value in equilibrium.

[23] For a primer on optimal control theory see e.g. Kamien and Schwartz (1991).

Log-utility implies that $\mathcal{H}$ is concave in both its state and control with its state being non-negative. Accordingly, the following necessary conditions are sufficient for a maximum:

$$\begin{aligned}
\frac{\partial \mathcal{H}}{\partial L_p} &= \frac{1}{L_p} - \frac{\lambda}{a} \leq 0, \ =, \text{ if } L_p > 0 \quad \Leftrightarrow \quad L_p \geq \frac{a}{\lambda} \ =, \text{ if } L_p > 0 && (4.30) \\
\frac{\partial \mathcal{H}}{\partial n} &= \frac{1-\alpha}{\alpha n} = -\dot{\lambda} + \rho\lambda \quad \Leftrightarrow \quad \dot{\lambda} = \rho\lambda - \frac{1-\alpha}{\alpha n} && (4.31) \\
\lim_{\tau\to\infty} \lambda(\tau)\, n(\tau) &= 0. && (4.32)
\end{aligned}$$

The first condition, (4.30), requires the marginal utility loss from foregoing one unit of consumption of an existing good (i.e. holding n constant, $1/L_p$) to equal the marginal utility gain from an additional unit of labor employed in R&D (reducing output by one unit allows one manufacturing worker to become a researcher, invent $1/a$ new blueprints, and thereby raise U_t by λ/a). The second condition, (4.31), requires the rate of return on n to equal the discount factor:

$$\rho = \frac{\frac{1-\alpha}{\alpha n} + \dot{\lambda}}{\lambda} = \frac{\frac{\partial \log D}{\partial n} + \dot{\lambda}}{\lambda}. \tag{4.33}$$

As usual, the rate of return consists of a dividend and a capital gain (in terms of utility): increasing n directly raises utility (the first term) and the value of doing so, λ, may change over time (the second term).[24] The last condition (4.32) imposes that the present value (in utility terms) of blueprints must converge to zero in the long run so that households cannot raise utility by inducing labor to reallocate into manufacturing.

In the decentral equilibrium, one unit of consumption costs $p = w/\alpha$. Foregoing one unit of consumption thus allows the household to finance the upfront blueprint costs for $1/\alpha$ firms (p equals $1/\alpha-$th of the value of $1/a$ blueprints, see (4.14) for $\dot{n} > 0$). If v/α were equal to the shadow price of n in the socially optimal allocation (i.e. equal to λ), the market would provide the socially optimal incentive for R&D.

In what follows, we prove that this is in fact the case. To do so, we show that a sequence of asset prices $\{v \equiv \alpha\lambda\}_t^\infty$ satisfies the market equilibrium conditions not yet accounted for in the social planner's problem (namely, the evolution of asset prices and free entry into R&D), so that the solution to the social planner's problem also solves the market equilibrium conditions. Since the market equilibrium is unique, this proves that the decentral outcome is efficient.

[24] Including public knowledge as a factor in R&D, we will find that since the transition function (i.e. the R&D technology in terms of L_P) is no longer independent of n, there will be an additional indirect dividend term from each invented blueprint, see Section 4.4.3 below.

First, by definition, $\hat{\lambda} = \hat{v}$ so that (using (4.7)), (4.31) coincides with the evolution of asset prices in (4.13). Second, substituting for λ with $v = \alpha\lambda$, (4.30) becomes $L_p = (\alpha a)/v$. After replacing L_p using (4.29) we find that (4.30) yields

$$\frac{\dot{n}}{n} = \frac{L}{a} - \frac{L_p}{a} = \frac{L}{a} - \frac{\alpha}{v},$$

(i.e. equation (4.20)). Accordingly, $L_p/a = \alpha/(wa) = \alpha v$ or $wa = v$, which satisfies (4.14).

Finally, let us ask under what conditions the social planner would implement zero growth. A simple way to answer this question is to impose $L = L_p$ in the first order condition for L_p in the planner's solution. Then, $\lambda = a/L$ and $\dot{\lambda} = 0$ from (4.30) so that (4.31) implies

$$n = \frac{1-\alpha}{\alpha a \rho} L = \bar{n}.$$

The same argument yields that optimally $L_p^* < L$ (so that $g > 0$) if $n < \bar{n}$. In the market equilibrium, $n \geq \bar{n}$ was shown to imply $L_n = g = 0$. This, however, is exactly the outcome in the market equilibrium. The socially optimal allocation is thus an equilibrium and uniqueness implies that the equilibrium path is socially optimal.

Since the growth rate associated with the share of labor in R&D implies the efficient growth rate, from the social point of view, we conclude that the tendency to overinvest in innovation (due to the nonobservance of profit destruction) exactly offsets the tendency to underinvest in innovation (due to the consumer surplus effect; see Grossman and Helpman, 1991b, Appendix A3.3 for a formal proof).

We next consider the economy with knowledge spillovers in R&D and show that adding this externality induces the inefficiency of the market outcome.

4.4.3 Dynamic Allocation with Knowledge Externalities

Including knowledge spillovers, we only have to use the modified R&D technology $\dot{n} = L_n n/a$ and replace the transition function for n in (4.29) with

$$\dot{n} = \frac{(L - L_p)\, n}{a}. \tag{4.34}$$

The Hamiltonian in this case becomes

$$\mathcal{H} = \frac{1-\alpha}{\alpha} \log n + \log L_p + \lambda \left(\frac{L - L_p}{a} \right),$$

and the necessary and sufficient conditions for an optimum are

$$\frac{\partial \mathcal{H}}{\partial L_p} = \frac{1}{L_p} - \frac{\lambda n}{a} = 0 \;\Leftrightarrow\; L_p = \frac{a}{\lambda n} \tag{4.35}$$

$$\frac{\partial \mathcal{H}}{\partial n} = \frac{1-\alpha}{\alpha n} + \frac{\lambda \left(L - L_p\right)}{a} = -\dot{\lambda} + \rho\lambda \;\Leftrightarrow\; \dot{\lambda} = \rho\lambda - \frac{1-\alpha}{\alpha n} - \frac{\lambda \left(L - L_p\right)}{a} \tag{4.36}$$

$$\lim_{\tau\to\infty} e^{-\rho\tau}\lambda\left(\tau\right) n\left(\tau\right) = 0. \tag{4.37}$$

Comparing (4.31) to (4.36), we see that the latter now implies an additional "social dividend" term for the rate of return on n. In addition to the direct increase in utility (which was the only reward in the economy without knowledge spillovers), every additional blueprint now increases the productivity of researchers and c.p. boosts innovation in the future. In utility terms, this additional reward is worth $\lambda\dot{n}$.

We assess the properties of the optimal allocation in two steps. First, we show that the optimal path involves no transitional dynamics so that the economy finds itself directly in a steady state allocation as of time 0. We then proceed to show that along the optimal path, the share of labor employed in R&D exceeds the share of labor in R&D in the equilibrium allocation. Hence the equilibrium growth rate is inefficiently low whenever the optimal allocation implies a positive growth rate. Otherwise, the equilibrium is efficient, just like in the case without knowledge spillovers above (cf. Section 4.4.2).

With $M \equiv n\lambda$ and (4.34), (4.36) can be rewritten as

$$\frac{\dot{\lambda}}{\lambda} + \frac{\dot{n}}{n} = \rho - \frac{1-\alpha}{\alpha M} = \hat{M}\,,$$

or, equivalently, $\dot{M} = \rho M - \left(1-\alpha\right)/\alpha$. Intuitively speaking, since this differential equation is unstable ($\rho > 0$) and neither $M \to 0$ nor $M \to \infty$ can be part of a feasible optimal path, $\dot{M} = 0$ and L_p is immediately pinned down at its steady state value by (4.35). To prove this statement, we go ahead and solve the (linear, first order) differential equation for its general solution:[25]

$$M = \frac{1-\alpha}{\alpha\rho} - \left(\gamma_0 + \gamma_1 \frac{1-\alpha}{\alpha}\right) e^{\rho t},$$

where the $\gamma' s$ are arbitrary constants. The transversality condition pins down the particular solution:

[25] Multiplying $\dot{M} - \rho M = 1 - 1/\alpha$ with $e^{-\rho t}$ and integrating, we have

$$\int \dot{M} e^{-\rho t} - \rho M e^{-\rho t} dt = \left(1 - \frac{1}{\alpha}\right) \int e^{-\rho t} dt.$$

Recognizing that the term in the integral on the left hand side equals the time derivative of $e^{-\rho t} M$ and solving for M yields the solution above.

from (4.37), $\lim_{\tau\to\infty} e^{-\rho\tau}M = 0$ so that the term in brackets is zero and $M = (1-\alpha)/(\alpha\rho)$.[26] Accordingly, (4.35) imposes that L_p directly jumps to its steady state value,

$$L_p^* = \frac{a}{M} = \frac{a\alpha\rho}{1-\alpha}.$$

The growth rate along the optimal path is obtained from (4.34):

$$g^* = \frac{L - L_p^*}{a} = \frac{L}{a} - \frac{\alpha\rho}{1-\alpha} = \frac{1}{1-\alpha}\left[(1-\alpha)\frac{L}{a} - \alpha\rho\right] = \varepsilon g > g. \tag{4.38}$$

Observe that labor employed in production is independent of the size of the labor force. Along the optimal path, economies with $L \leq L^*$ allocate all labor to manufacturing. If $L > L_p^*$, $L - L_p^*$ resources are available for the conduct of R&D and growth is positive. In two economies A and B with $L^A > L^B > L_p^*$, the "larger" economy A grows faster along the optimal path, $\left(g^A\right)^* > \left(g^B\right)^*$, but the manufacturing sectors are of equal size: $n^A x^A = n^B x^B$.

The inefficiency of the laissez-faire outcome in the economy that permits sustainable growth points to the need of government intervention. We consider two instruments that naturally come to mind: (i) a sales subsidy and (ii) a subsidy on R&D costs. We conclude the section by deriving the optimal R&D subsidy (which implements the Pareto efficient allocation).

4.5 R&D Policy

Following Grossman and Helpman (1991b, Ch. 3.3), we analyze the impact of subsidies in a diagram that highlights the very basic allocation trade-off between growth and manufacturing. Recall that the economy with knowledge spillovers jumps into its steady state so that there are no transitional dynamics to be considered. In (g, L_p)-space, the steady state equilibrium can be exemplified by the intersection of the resource constraint and a "participation constraint", i.e. combinations of L_p and g where the return to manufacturing equals the real interest rate. These conditions are readily derived. Rearranging (4.23), we redefine the resource constraint (i.e. the economy's production possibility frontier) as

$$\Pi_L(g) \equiv L - ag. \tag{4.39}$$

[26] In some problems, the transversality condition employed above must not be necessary for optimality ($\lim_{\tau\to\infty}\mathcal{H}(\tau) = 0$ is). Since we employ the transversality condition to pick the particular solution for M, let us point out that in the planning problem at hand, $\lim_{\tau\to\infty} e^{-\rho\tau}\lambda n = 0$ is in fact the appropriate condition since $\rho > 0$ and U_t converges. See Benveniste and Scheinkman (1982) for a proof.

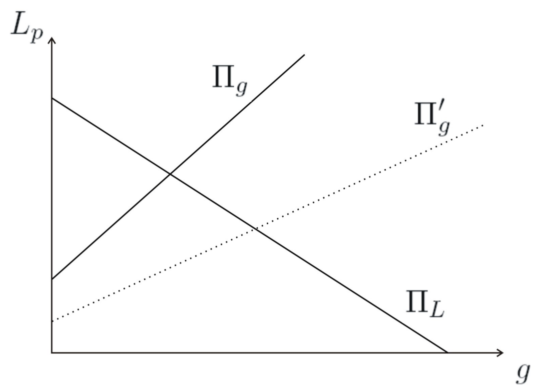


Figure 4.4: Illustration of the Equilibrium: Resource and "Participation" Constraint (Adapted from Grossman and Helpman, 1991b, p. 65)

The negative slope of Π_L says that labor employed in R&D is not available for production, and thus a high growth rate goes hand in hand with little goods production. A useful feature of our analysis of growth policies is that this equilibrium condition is independent of all policy interventions.

The second condition is readily derived from (4.13), which says that the rate of return on R&D satisfies $r = \dot{v}/v + \pi/v$. In equilibrium (in steady state), v grows at rate $-g$ (since V is constant), $r = \rho$, and, using the free entry into R&D condition and the pricing equation, the dividend ratio equals $\pi/v = (1-\alpha)/(nv) = (1-a)/(aw) = (1-\alpha)/(\alpha ap)$. Inserting $1/p = L_p$ from the normalization of expenditures and the manufacturing technology and solving for L_p, we get the "participation constraint":

$$\Pi_g(g) \equiv \frac{\alpha a\rho}{1-\alpha} + \frac{\alpha a}{1-\alpha} g.$$

Given (PA), the upward sloping Π_g curve cuts Π_L uniquely from below, see Figure 4.4. We now use this diagram to examine how a subsidy to R&D affects the equilibrium allocation.

4.5.1 R&D Subsidy

Suppose that the government aims at increasing the incentives for R&D via a subsidy. In particular, suppose that a fraction ϕ of the R&D expenses per newly invented blueprint is covered by the subsidy. While this intervention leaves the "production possibilities frontier" Π_L unaffected, the slope of Π_g and its intersection with the ordinate decline. That is to say, the profit rate in manufacturing must be higher for any given allocation to offset the increase in real interest rate due to the subsidy. Starting from the free entry into R&D condition, $V^{-1} = wa(1-\phi)$, or, using $w = \alpha p$ from the pricing equation and $L_p^{-1} = p$ from $nxp = 1$, $V^{-1} = \alpha a (1-\phi) L_p^{-1}$. Solving for the resources employed in manufacturing L_p yields the modified Π_g curve:

$$\Pi_g'(g) \equiv \frac{\alpha a (1-\phi)\rho}{1-\alpha} + \frac{\alpha a (1-\phi)}{1-\alpha} g. \tag{4.40}$$

Equating $\Pi_g'(g)$ and Π_L generalizes the expression for the equilibrium growth rate to include the subsidy,[27]

$$g^{\phi} \equiv \frac{(1-\alpha)\frac{L}{a} - (1-\phi)\alpha\rho}{1-\alpha\phi}. \tag{4.41}$$

Evidently, $g^{\phi=0} = g$ in (4.28). Comparing (4.41) to the optimal growth rate in (4.38), we find that the efficient allocation can be implemented by a subsidy equal to

$$\phi^* = \frac{\frac{L}{a} - (\varepsilon - 1)\rho}{\frac{L}{a} - (\varepsilon-1)\rho + \rho} = \frac{g^*}{g^* + \rho} \quad (< 1).$$

Qualitatively, the optimal subsidy is increasing in L and $1/a$ and decreasing in ε and ρ. Perhaps surprisingly, ϕ^* is increasing in L since, optimally, all resources in excess of $a\alpha\rho/(1-\alpha)$ should be allocated to R&D. It takes ever stronger subsidization to achieve this target, the larger L (as $L \to \infty$, $\phi^* \to 1$).

4.5.2 Inefficacy of Sales Subsidies

Now suppose that instead of paying a fraction ϕ of R&D expenses, the government subsidizes the sales of manufactured goods so as to increase the value of blueprints and thereby spur innovation via the product market. Following Grossman and Helpman, we show that such a subsidy leaves the sectoral allocation of labor, and hence the growth rate, unaffected.

A sales subsidy affects the price setting behavior of manufacturing firms. Denoting the subsidy by ϕ_x (> 0), the profit of a manufacturer reads

$$\pi(j) = p(j)(1+\phi_x)x(j) - wx(j),$$

[27] Equating (4.39) and (4.40) yields

$$\begin{aligned} L - ag^{\phi} &= \frac{\alpha a (1-\phi)}{1-\alpha} g^{\phi} + \frac{\alpha a (1-\phi)}{1-\alpha}\rho \quad \Leftrightarrow \\ \frac{1-\alpha}{\alpha a (1-\phi) L} - \rho &= g^{\phi}\left[1 + \frac{1-\alpha}{\alpha(1-\phi)}\right]. \end{aligned}$$

Solving for g^{ϕ} gives (4.41) above.

where $x(j) = p(j)^{-\varepsilon} P^{\varepsilon-1}$, see (4.9). Quantities are chosen to maximize profits so that $\alpha p(j)(1+\phi_x) = w$. C.p., the subsidy lowers the equilibrium prices and increases the quantity supplied,

$$p(j) = \frac{w}{\bar{\alpha}}, \quad x(j) = \frac{1}{np} = \frac{\bar{\alpha}}{nw},$$

where $\bar{\alpha} \equiv (1+\phi_x)\alpha > \alpha$. The equilibrium profit amounts to

$$\pi = \frac{\bar{\alpha}}{\alpha} px - wx = \left(\frac{p}{w}\frac{\bar{\alpha}}{\alpha} - 1\right) wx = \left(\frac{1}{\alpha} - 1\right)\frac{\bar{\alpha}}{n} = \frac{1+\phi_x-\bar{\alpha}}{n} = \frac{(1+\phi_x)(1-\alpha)}{n}.$$

Note, however, that the wage rate is endogenous. In fact, since $wa = V^{-1}$, the sales subsidy does not change the rate of return on R&D (the R&D costs increase with the value of the blueprints). To prove this statement, insert the equilibrium price in the free entry into R&D condition, $V = (wa)^{-1} = (\bar{\alpha} a p)^{-1} = L_p / (\bar{\alpha} a)$ (since $L_p = p^{-1}$ from $nxp = 1$), and calculate the rate of return to R&D:

$$r = \frac{\pi}{v} - g = (1+\phi_x)(1-\alpha) V - g = \frac{(1+\phi_x)(1-\alpha) L_p}{(1+\phi_x)\alpha a} - g = \frac{(1-\alpha) L_p}{\alpha a} - g.$$

Accordingly, Π_g remains unchanged. A sales subsidy does not alter the allocation of labor between R&D and manufacturing and hence leaves the growth rate unaffected. While the R&D subsidy increases the rate of return on R&D without affecting the rate of return from manufacturing, the sales subsidy at the same time increases the return to manufacturing, thereby leaving the relative rate of return unchanged.[28]

4.6 Scale Effects in the Grossman-Helpman (1991b, Ch. 3) Model

A major drawback of the Grossman-Helpman model, as well as of other endogenous growth models of the early 1990s, is that it suffers from the strong scale effect: the long-run growth rate depends positively on the labor endowment (i.e. the "size") of the economy.[29] Jones (1995a) prominently argued that this feature is at odds with empirical evidence.[30] Using U.S. data from 1950-1990, he shows that while the number of scientists and engineers grew by factor 5, the growth rates of total

[28]This result is more general. The crucial assumption is that there is only one factor in manufacturing.

[29]Jones (2005) distinguishes the strong scale effect, i.e. the dependency of the growth rate on the "scale" of the economy, and the dependency of the level of GDP on the "scale" of the economy, the weak scale effect.

[30]Similar concerns are raised by Kremer (1993) and Young (1998) among others. The first "search for the scale effects" in growth (and trade) seems to be conducted by Backus, Kehoe, and Kehoe (1992), who found little empirical evidence for the strong scale effect, but uncovered that the growth of output per worker in manufacturing is significantly affected by the scale of manufacturing.

factor productivity (TFP) continued to stagger around 2%. Growth models marked by the strong scale effect, however, predict a rise of the growth rate as the number of researchers increases.

The invariability of growth rates further casts doubts on the efficacy of growth policy, and, more generally, on the influenceability of innovation rates, another central hypothesis of the Grossman-Helpman model.

To cure these shortcomings, Jones (1995b) and various other authors proposed slight modifications of the R&D technology among other things. Important contributions in this strand of literature include Segerstrom (1998), Eicher and Turnovsky (1999), and Arnold (1998). Basically, these models impose that the labor requirement per newly discovered blueprint increases over time. Sustainable industrial growth in these environments is then driven by e.g. population growth (Jones, 1995b) or growth of human capital (Arnold, 1998). At first glance, the modifications in the R&D production functions of these so-called "second generation" models are rather small. In fact, however, they lead to tremendous differences in the steady state comparative statics. In particular, the long-run growth rate is no longer endogenous and thus also invariant to policy interventions.[31] Importantly, however, policy interventions still affect the transitional growth path as the economy converges to its long-run growth path. The level effects exerted by growth policies or other shocks are of great importance for consumers' welfare. Their evaluation, however, more often than not, requires careful calibration. Generally, second generation growth models are less flexible and generally harder to work with than first generation growth models (at least, they involve an additional state variable).

Two further arguments support the continued use of first generation growth models like Grossman and Helpman (1991b, Ch. 3). First, Baldwin and Forslid (1999) warn against taking the Jones critique as too substantial an argument against the Grossman-Helpman model. Their point is that Jones measures "innovation" by the growth rate of TFP which, as previously claimed by Nelson (1996), is a bad measure of innovation. In particular, Baldwin and Forslid (1999, p. 802) argue that since TFP figures crucially depend on aggregate prices, the TFP growth rate will constantly underestimate innovation if the involved price indices are not continuously adapted to the increase in product variety. In Grossman and Helpman's (1991b, Ch. 3) growth model, it is precisely the continuous decline in the ideal price index[32] induced by the increasing product variety that drives growth. Baldwin and Forslid point to an extreme case: if the empirical price index is constructed as consumption expenditures over

[31] See, however, Howitt (1998).

[32] P is called the "ideal" price index because it follows with perfect competition from cost minimization in the manufacturing sector (which yields unit production costs equal to $\left[\int_0^n p(j)^{1-\varepsilon}\, dj\right]^{\varepsilon/(1-\varepsilon)} / (p(j)^{\varepsilon}$ for each j).

average prices, a figure that remains constant in the model, there will be no correlation between labor employed in R&D and growth of TFP as measured by the empirical price index.

Second, Lingens (2005) calibrates both a version with and without scale effects of Jones' (1999) endogenous growth model and finds that the impact of a marginal change in the amount of labor devoted to R&D leads to similar reactions of the growth rate right after the shock in both models. That is, the first generation model and the second generation model behave similarly if we look at the initial impact in the second generation framework. Lingens (2005, p. 5) concludes that "'first generation' models, hence, offer a convenient approximation of the transitional behavior of 'second generation' models". He further claims that first generation models even provide a good quantitative approximation of the transitional effects of R&D policies in second generation models. If this assertion is robust, we may well exploit the flexibility of the Grossman-Helpman in further policy experiments.

A couple of years after the initial "Jones shock", some authors put their feet back on the grounds of first generation models. In doing so, they are very careful in interpreting the results and the mechanisms at work. Recent publications, like e.g. Baldwin and Robert-Nicoud (2007), suggest that a careful assessment is acceptable by the profession. Evidently, however, building on first generation frameworks, one is inevitably subject to criticism by sceptical readers. The development of sufficiently simple and flexible frameworks for endogenous growth without scale effects therefore still remains an important task. Sorger (2007) is a first step in this direction, although his "quantity–quality frontier in R&D" approach has not yet received adequate attention.

Appendix 4.A Rational Expectations and Diverging Trajectories

In this appendix, we show that neither $nv \to \infty$ $(V \to 0)$ nor $nv = 0$ $(V \to \infty)$ is compatible with rational expectations so that the divergent paths in both the economy with and without knowledge spillovers can be sorted out (Grossman and Helpman, 1991b, p. 61).

To begin with, consider the the case $nv \to \infty$ $(V \to 0)$. The most general argument against this case follows from the no-Ponzi game condition in (4.4). If the consumption path is chosen optimally, (4.4) must hold with equality and $r = \rho$ from $E(t) \equiv 1$. Accordingly, along the optimal consumption path, (4.4) obeys

$$\frac{1}{\rho} = \int_t^\infty e^{-\rho(\tau - t)} wL d\tau + vn.$$

Recall that $wa \geq vn$ and $wa \geq v$ in the economy with and without knowledge spillovers, respectively. Hence, if both n and v grow without bound in the economy without knowledge spillovers, the consumption path cannot be chosen optimal. The same is true for unbounded growth of nv in the economy with knowledge spillovers. If the stock market value blows up to infinity, the firm values must be inconsistent with rational expectations and optimal consumer behavior.

Consider next the case where $nv \to 0$ $(V \to \infty)$. In both economies with and without knowledge spillovers, this case implies $\dot{n} = 0$ so that we are actually looking at $v \to 0$. Since the economy is in a stationary environment with a bounded number of firms and each incumbent earns strictly positive profits at each moment in time, households cannot rationally expect $v \to 0$. Actually, in the economy without knowledge spillovers, trajectories that lead to $v \to 0$ imply $v < 0$ if we do not impose $v \geq 0$ exogenously. Evidently, this must violate rational expectations.

Chapter 5

Some Second Thoughts on Monopolistic Distortions and Endogenous Growth[1]

5.1 Abstract

The most fundamental proposition about growth and competition is that there is a tradeoff between static welfare and long-term growth. This chapter reconsiders this basic proposition in an expanding variety endogenous growth model with competitive markets for "old" innovative products and for a traditional good. We shed light on some implications of monopolistic distortions which tend to be ignored by standard models. First, no growth may be better than some growth, since modest positive growth potentially requires sizeable static welfare losses. Second, the economy may converge to a steady state with zero growth, even though another locally saddle-point stable steady state with positive growth exists if the initial share of "cheap" competitive markets is sufficiently high, as this implies a relatively low demand for "expensive" innovative goods. Third, such a "no-growth trap" may happen in a world economy made up of several countries engaged in free trade with each other. The policy implications are that growth-enhancing policies may be misguided and that quick deregulation as well as quick trade liberalization can lead to stagnation in the long term.

[1]This chapter is joint work with Lutz Arnold. It presents a slightly modified version of Arnold and Bauer (2008). An earlier working paper contains detailed derivations of Propositions 5.1 and 5.2 (Bauer, 2006). The model in this chapter builds on Arnold (1995).

5.2 Introduction

The most fundamental proposition about growth and competition, taught in introductory economics courses, is that there is a tradeoff between static welfare and long-term growth: perfect competition brings about static efficiency but undermines the incentives to invest in the innovation of new goods, services, and processes (see, e.g., Blanchard, 2006, p. 256).[2] This chapter highlights several important macroeconomic implications of this basic proposition. To do so, we consider the standard Grossman-Helpman (1991, Ch. 3) expanding variety endogenous growth model augmented to include erosion of monopoly power due to (exogenous) imitation and a non-innovative traditional sector. Neither of these two extensions is novel. Textbook expositions can be found, for instance, in Barro and Sala-i-Martin (2004, Section 6.2, pp. 305 ff.) and Grossman and Helpman (1991b, Section 5.3, pp. 130 ff.), respectively.[3] However, the implications of the resulting monopolistic distortions for model dynamics and welfare are not fully worked out. We prove three results on the model's dynamics and welfare properties and derive corollaries which characterize second-best competition and patent policies.

The first result is directly concerned with the tradeoff between static welfare and the incentives to innovate. In the Grossman-Helpman (1991, Ch. 3) model (i.e., without competitive markets), the equilibrium growth rate is lower than optimal, and no growth cannot be better than some growth. We show that, in our model, it can be. This is because the static welfare loss due to monopoly pricing in the innovative sector is non-infinitesimal. So if strict enforcement of IPRs brings about only a modest growth rate, it is preferable to dispense with growth altogether and implement static efficiency instead.

The second result says that the economy may get stuck in a "no-growth trap" (poverty trap): the unique perfect-foresight equilibrium possibly entails convergence to a steady state with zero growth, even though another (locally) saddle-point stable steady state with positive growth exists. This will happen if the initial share of competitive markets is sufficiently high.[4] In that case, a potential innovator would compete with many relatively cheap products, so that it does not pay to innovate. This result

---

[2] Hellwig and Irmen (2001) point out that perfect competition per se does not rule out innovation-driven growth: positive profit and costly innovation are compatible with perfect competition in the presence of diminishing returns to scale and inframarginal rents (see also the discussion in Romer, 1990, pp. S75-S77).

[3] For models with exogenous imitation, see also Rustichini and Schmitz (1991), and Pelka (2005, Chapter 7). In Segerstrom (1991) and Walz (1995), imitation is endogenous. Perez-Sebastian (2000) shows that "growth miracles" can be explained in a model that interprets the process of imitation as the costly adaption of knowledge created abroad.

[4] The possibility of a no-growth trap is ignored in the papers with exogenous imitation mentioned in Footnote 2 except in Pelka (2005, Chapter 7).

is related to the literature on poverty traps, surveyed by Azariadis and Stachurski (2005).[5] As for economic policy, it implies that quick deregulation, which turns many monopolistic markets (e.g., state monopolies) into perfectly competitive markets simultaneously may be very detrimental to long-term growth, since innovating becomes unattractive as the incumbent competitive producers attract the major part of the goods demand.

The second result is reminiscent of Tang and Wälde's (2001) finding that a two-country world economy may find itself in a no-growth trap if there are sufficiently many competitive markets due to a large initial overlap of products, invented before trade is opened up between the countries. Our third result is concerned with the open-economy version of our model and relates our model to Tang and Wälde's (2001). Adapting the analysis in Arnold (2007) appropriately, we prove that, under certain conditions, the world economy made up of several (identical) countries replicates the equilibrium of the hypothetical integrated economy (that would prevail if national borders did not exist). Together with the second result, it follows immediately that if the model parameters are such that the no-growth trap occurs in the integrated economy, then the no-growth trap is also an equilibrium of the world economy if there are sufficiently many competitive markets due to a large initial overlap of products. From a policy point of view, it follows that, like quick deregulation in a closed economy, quick trade liberalization can lead to stagnation in the long term: the opening up of free trade at a point in time when the overlap exceeds the threshold number of competitive markets, above which the (world) economy is stuck in a no-growth trap, leads to long-term stagnation.

Several related recent papers investigate the growth and welfare effects of changes in IPRs. These papers challenge Judd's (1985) case for infinitely-lived patents. Kwan and Lai (2003) and Iwaisako and Futagami (2003) analyze imperfect protection of IPRs in expanding variety growth models. Kwan and Lai (2003) work out the welfare effects of a marginal shock to IPR protection (including transitional dynamics) and show that there is a finite optimal patent length, which probably exceeds the status quo for the U.S. Iwaisako and Futagami (2003) draw a similar conclusion from a comparison of steady-state intertemporal utility levels. Horii and Iwaisako (2007) show, in a quality upgrading model, that

[5]The result most closely related to our finding of a no-growth trap is due to Ciccone and Matsuyama (1996, Section 7). Allowing final goods producers to substitute labor for a composite good made up of differentiated intermediate goods, they show that a low initial number of innovative goods may lead to a no-growth trap if the elasticity of substitution is sufficiently high. The intuition is that a low number of existing innovative goods implies high costs of the composite input. Together with the possibility to substitute for innovative goods with labor, the resulting demand for innovative goods is too low to induce innovation and get growth started.

the growth rate effect of strengthening IPRs is not necessarily positive if competitive sectors are more innovative than monopolies. Similarly, Furakawa (2007), using an expanding variety model, points out that tightening IPRs may be detrimental to growth if the ensuing reduction in market size reduces the scope for productivity gains due to learning by doing. Our chapter's main contribution to this strand of the literature is two-fold. For one thing, to investigate the possibility of a no-growth trap, we analyze our model's *global* dynamic behavior. For another, we present results for the open as well as for the closed economy.

A different strand of the literature calibrates dynamic general equilibrium models in order to challenge Harberger's (1954, p. 87) classical presumption that "monopoly does not seem to affect aggregate welfare very seriously through its effect on resource allocation", in view of his calculation that the gains from eliminating the distortions in the manufacturing sector (in the U.S. in the 1920s) are about one-tenth of a percentage point of GNP. Matheron (2002) shows that in a human capital growth model these gains are equivalent to a permanent 2.5 percent increase in consumption. The corresponding figure is 3.6 percent in Jonson's (2007) model and grows more than three-fold if one takes account of distortions caused by the tax system. These models lend support to the view that the monopolistic distortions analyzed in the present chapter are of non-negligible magnitude.

As the above-mentioned papers constitute only a small fraction of the vast literature on growth and competition, it might be helpful to point out what the present chapter does not do.

First, a huge and growing literature addresses the issue of competition between several firms in given markets and the relation between the intensity of competition and the pace of equilibrium growth. For instance, Aghion et al. (2001) demonstrate that more intense competition may spur growth in a model with innovation by both technological leaders and laggards, as it induces firms to try to escape fierce competition. Such effects are not present in our model, in which markets are either monopolies or perfectly competitive. Our motivation is that this is the easiest way of approaching the question of how distortions which stem from the fact that some markets are less competitive than others affect model dynamics.

Second, in their influential "Case Against Intellectual Property", Boldrin and Levine argue that "the case against monopoly rests less upon the [Harberger] welfare triangle from monopoly pricing than upon the rent-seeking activity used to get and keep a monopoly" (Boldrin and Levine, 2002, p. 211). This de-emphasis of monopoly distortions is at least debatable. At any rate, the two effects are complementary, and including the rent seeking problems emphasized by Boldrin and Levine would strengthen our conclusions.

Third, O'Donoghue and Zweimüller (2004, p. 87) argue that, not patent length, but "[T]he patentability requirements and patent breadth reflect the two main tasks that confront the patent authorities" and analyze these dimensions of patent policy in depth. We do not address these issues.

Fourth, the terms deregulation and liberalization can be given different meanings. We call a change from monopoly to perfect competition deregulation (since our model is an expanding variety model, the market then remains competitive indefinitely). As an example of a different definition, see Büttner (2006). She considers a quality upgrading model in which some goods are publicly provided (without being upgraded) at monopoly prices and defines deregulation as a decrease in the number of such monopolies. Deregulation is then unambiguously conducive to growth. In the open economy version of our model, liberalization means a switch from no international trade at all to unrestricted free trade. A more general formulation would allow for finite iceberg costs and define trade liberalization as a decrease in the iceberg costs (see, e.g., Baldwin and Robert-Nicoud, 2007, or Gustafsson and Segerstrom, 2007) or introduce tariffs explicitly (as, e.g., in Dinopoulos and Segerstrom, 1999, or Baldwin and Forslid, 1999, 2000).

Fifth, it is well known that in models with quality upgrading as the source of growth, contrary to the standard expanding variety models, growth can be too fast and zero growth can be preferable when equilibrium growth is positive (see, e.g., Grossman and Helpman, 1991b, Ch. 4, pp. 103-106). To highlight that it is the monopolistic distortions which are responsible for our first result, we choose as our point of departure the Grossman-Helpman (1991, Ch. 3) expanding variety growth model, so that growth cannot be too fast in the absence of competitive markets.[6]

Sixth, the Grossman-Helpman (1991, Ch. 3) model is a first-generation R&D model, which displays scale effects. Time series observations pose a great challenge to such models (see Jones, 1995a) and have led to the development of non-scale growth models, such as Jones (1995b), Young (1998), or Arnold (1998).[7] A relatively general lesson of these models is that growth rates are much less responsive to

[6]In standard expanding variety models, one parameter pins down both the elasticity of substitution between any two varieties of a differentiated good and the magnitude of the gains from specialization. Bénassy (1998) shows that growth can be too fast if one disentangles these two variables.

[7]Young (1998) emphasizes that an increase in the labor force may be absorbed by a sector of the economy that does not spur long-term growth. Jones (1995b) assumes diminishing returns to knowledge in the creation of new knowledge, in which case population growth is required to sustain long-term growth. Arnold (1998) replaces population growth with human capital accumulation.

changes in the model parameters than models with scale effects indicate.[8] Our motivation for using a first-generation model is that this limits the number of state variables in such a way that we can carry out the (phase diagram) analysis of the model's global dynamic behavior which is necessary in order to identify a no-growth trap. This appears acceptable in view of the fact that the presence of a no-growth trap is a property of the model's qualitative dynamic behavior, which should not relate to the responsiveness of the steady-state growth rate to changes in model parameters.

The remainder of the chapter is organized as follows. Section 5.3 introduces the model. The growth equilibrium is derived in Section 5.4. Section 5.5 proves our main results on growth and competition. Section 5.6 concludes.

5.3 Model

There is a continuum of mass one of identical households. Each household inelastically supplies L units of labor, the only primary factor of production. Their intertemporal utility is $U = \int_0^\infty e^{-\rho t}[\sigma \ln X + (1-\sigma)\ln Y]dt$, where X and Y are the quantities consumed of two homogeneous goods, x and y (and $\rho > 0$, $0 < \sigma < 1$).[9] Good x is produced using a set of intermediates, j, according to the production function $X = [\int_0^n x(j)^\alpha dj]^{1/\alpha}$, where $x(j)$ is the input of intermediate j, n is the "number" of producible intermediates, and $0 < \alpha < 1$. Each producible intermediate, j, is obtained one-to-one from labor. The "traditional" good y is also obtained one-to-one from labor: $Y = L_Y$, where L_Y is labor employed in the production of y (one may think of services with less scope for innovation than in manufacturing). Blueprints for new intermediates are invented in R&D according to $\dot{n} = nL_R/a$ (with $a > 0$), where L_R is employment in R&D (there are scale effects). As for market structure, we assume that all markets are perfectly competitive except for the markets for "new" intermediates. Immediately after the development of a new variety, the innovator is a monopolist (due to either patent protection or the fact that other agents are not yet able technologically to imitate the intermediate). Subsequently, in any short time interval dt the innovator loses his monopoly with probability $\psi\, dt$ (due to the loss of patent protection or of technological leadership), in which case the market becomes perfectly competitive. ψ (≥ 0) is called the rate of imitation. Consequently, letting n_m and n_c denote the "numbers" of monopolistic and competitive markets for intermediate goods, respectively, we have

$$\dot{n_c} = \psi n_m, \quad n = n_c + n_m. \tag{5.1}$$

[8] See, however, Howitt (1999).

[9] The time argument is suppressed here and in what follows whenever this does not cause confusion.

As mentioned in the Introduction, the presence of competitive markets (for "old" innovative goods and for the traditional good) is the only difference to Grossman and Helpman (1991b, Ch. 3). Our main results, explained in the Introduction, go through for $\psi = 0$. The purpose of including $\psi > 0$, at the cost of some additional complexity, is two-fold. First, with $\psi = 0$, obviously, the number of competitive innovative goods markets converges to zero, which runs counter our focus on the role of competitive versus monopolistic markets. Second, by allowing for positive values, we can use ψ as a measure of the strength of intellectual property rights, which will be convenient in the policy experiments we consider.

5.4 Equilibrium

Using aggregate expenditure as the numéraire, utility maximization yields

$$p_X X = \sigma,\ p_Y Y = 1 - \sigma,\ r = \rho, \tag{5.2}$$

where p_X and p_Y are the prices of goods x and y, respectively, and r is the interest rate. Cost minimization in the x-sector yields the input coefficient $a(j) = p(j)^{-\epsilon}[\int_0^n p(j')^{1-\epsilon} dj']^{\epsilon/(1-\epsilon)}$ for good j, where $p(j)$ is the price of intermediate j and $\epsilon \equiv 1/(1-\alpha)$. Consequently, the unit production cost and, because of perfect competition, the price of good x is $p_X = [\int_0^n p(j)^{1-\epsilon} dj]^{1/(1-\epsilon)}$. The x-sector's demand for intermediate j is $x(j) = a(j)X$. The price elasticity of demand is ϵ ($< \infty$). Monopolists in the intermediate goods sector maximize profit, π, given these demand functions. Letting p_m and p_c denote the prices in monopolistic and competitive intermediate goods markets, respectively, and x_m the output of monopolistically supplied intermediates, we obtain the familiar pricing rules and ensuing profits:

$$p_m = \frac{w}{\alpha},\ p_c = w,\ \pi = (1-\alpha) p_m x_m. \tag{5.3}$$

Substituting the pricing rules into the expression for the input coefficients, $a(j)$, the demands $x(j) = a(j)X$ can be rewritten as

$$x_m = \alpha^{\epsilon} \left(n_c + \alpha^{\epsilon-1} n_m\right)^{-\frac{1}{\alpha}} X,\ x_c = \left(n_c + \alpha^{\epsilon-1} n_m\right)^{-\frac{1}{\alpha}} X, \tag{5.4}$$

where x_c denotes the output of competitively supplied intermediates. Moreover, substituting the pricing rules in (5.3) into $p_X = [\int_0^n p(j)^{1-\epsilon} dj]^{1/(1-\epsilon)}$ and using the fact that good y is obtained one-to-one from labor, we get the final goods' prices:

$$p_X = \left(n_c + \alpha^{\epsilon-1} n_m\right)^{-\frac{1}{\epsilon-1}} w,\ p_Y = w. \tag{5.5}$$

Using the fact that, as of time t, a monopolist's probability of still being a monopolist at τ ($\geq t$) is $e^{-\psi(\tau-t)}$, the value of a monopoly is

$$v(t) \equiv \int_t^\infty \exp\left\{-\int_t^\tau [r(s)+\psi]ds\right\}\pi(\tau)d\tau. \tag{5.6}$$

Imitation acts like additional discounting.[10] Free entry into R&D requires

$$wa \geq nv, \text{ with equality if } \dot{n} > 0. \tag{5.7}$$

Finally, the labor market clearing condition reads:

$$L = a\frac{\dot{n}}{n} + n_c x_c + n_m x_m + Y. \tag{5.8}$$

Equations (5.1)-(5.8) comprise a system of 15 equations in 15 unknowns: n_c, n_m, n, p_X, X, p_Y, Y, r, p_m, w, p_c, π, x_m, x_c, and v.[11] A vector of these 15 variables which solves (5.1)-(5.8) for all t is an equilibrium.

Let $\theta \equiv n_c/n$ and $g = \dot{n}/n$ denote the proportion of intermediate goods markets which are competitive and the growth rate of the "number" of intermediates, respectively. Further, let $V \equiv 1/(nv)$.[12] Using $Y = L_Y$, (5.2), (5.4), (5.5), (5.7), and these definitions, the labor market clearing condition (5.8) can be rewritten as

$$g = \max\left\{0,\ \frac{L}{a} - \sigma V\left[\frac{\theta(1-\alpha^\epsilon)+\alpha^\epsilon}{\theta(1-\alpha^{\epsilon-1})+\alpha^{\epsilon-1}} + \frac{1-\sigma}{\sigma}\right]\right\}. \tag{5.9}$$

Differentiating the definition of θ and using (5.1) yields

$$\dot{\theta} = (1-\theta)\psi - \theta g. \tag{5.10}$$

Imitation tends to increase the proportion of competitive markets, growth tends to reduce it. From (5.2)-(5.5), we have

$$\pi = \frac{\sigma(1-\alpha)}{\left[1-\theta(1-\alpha^{1-\epsilon})\right]n}.$$

Differentiating the definition of V and (5.6) with respect to time, eliminating $\dot{v}$, and using (5.2) and the equation for monopoly profit above, we obtain

$$\frac{\dot{V}}{V} = \frac{(1-\alpha)\sigma}{1-\theta(1-\alpha^{1-\epsilon})}V - (\rho+\psi+g). \tag{5.11}$$

[10] See Appendix 5.A below.

[11] The budget constraint $d(n_m v)/dt = rn_m v + wL - 1$ represents another equation in the same variables but, as usual in general equilibrium theory, one of the 16 equations can be obtained from the other 15 so that we have as many equations as unknowns.

[12] Note that this is not the inverse of the aggregate firm value.

Given (5.9), equations (5.10) and (5.11) comprise an autonomous system of ordinary differential equations in θ and V. In the present section, we analyze this system of equations. The findings will be used in Section 5.5 to bring forth our main results. As mentioned above, it is possible to focus on the (easier) case $\psi = 0$.

From (5.9), $g > 0$ if, and only if,

$$V < \frac{L}{a} \frac{1-\theta(1-\alpha^{1-\epsilon})}{1-\theta[1-\alpha^{1-\epsilon}-(1-\alpha)\sigma]-(1-\alpha)\sigma} \equiv \tilde{V}(\theta) \tag{5.12}$$

and $g = 0$ otherwise. $\tilde{V}(\theta)$ is continuous and strictly decreasing for $\theta \in [0,1]$, with $\tilde{V}(0) = (L/a)/[1-(1-\alpha)\sigma]$ and $\tilde{V}(1) = L/a$.

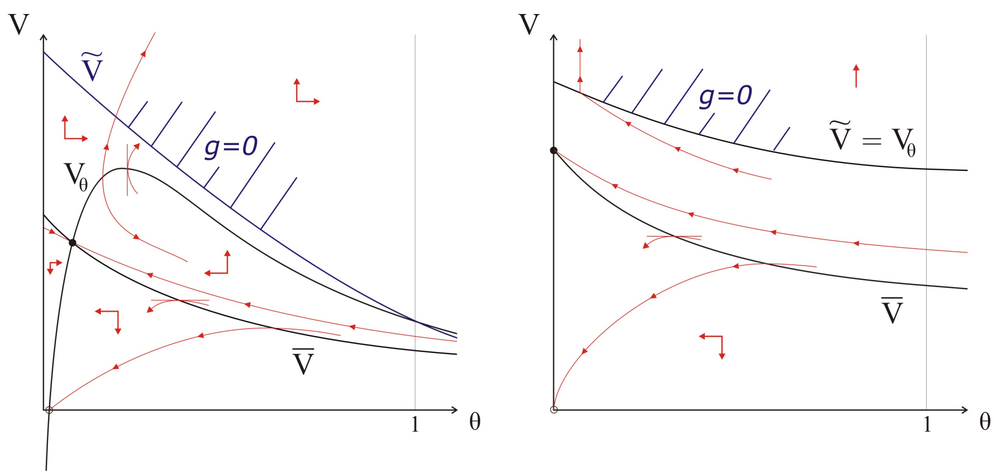


Figure 5.1: Dynamics in Case 1 (Left Panel: $\psi > 0$, Right Panel: $\psi = 0$)

Consider first the $g = 0$-region. According to (5.11), V is constant for $V = 0$ and for

$$V = (\rho+\psi)\frac{1-\theta(1-\alpha^{1-\epsilon})}{(1-\alpha)\sigma} \equiv \bar{V}_0(\theta), \tag{5.13}$$

where $\bar{V}_0(0) = (\rho+\psi)/[(1-\alpha)\sigma]$, $\bar{V}_0(1) = (\rho+\psi)/[(1-\alpha)\alpha^{\epsilon-1}\sigma]$, and $\bar{V}_0'(\theta) > 0$. $\dot{V}$ is positive for $V > \bar{V}_0(\theta)$ and negative for $V < \bar{V}_0(\theta)$. For $\psi > 0$, from (5.10) and $g = 0$, θ is constant or increases depending on whether $\theta = 1$ or $\theta < 1$, respectively. $\psi = 0$, together with $g = 0$, implies $\dot{\theta} = 0$.

Here and in what follows, we distinguish three cases:

Case 1:

$$\frac{(1-\alpha)\sigma\frac{L}{a}}{\rho+\psi} > \alpha^{1-\epsilon}. \tag{5.14}$$

In this case, $\bar{V}_0(1) < \tilde{V}(1)$. As $\bar{V}_0'(\theta) > 0 > \tilde{V}'(\theta)$ for all $\theta \in [0,1]$, it follows that $\bar{V}_0(\theta) < \tilde{V}(\theta)$ and, hence, $\dot{V}/V > 0$ for all $\theta \in [0,1]$. That is, trajectories in the $g = 0$-region point to the north-east for $\psi > 0$ (see the left panel of Figure 5.1) and to the north for $\psi = 0$ (see the right panel of Figure 5.1).

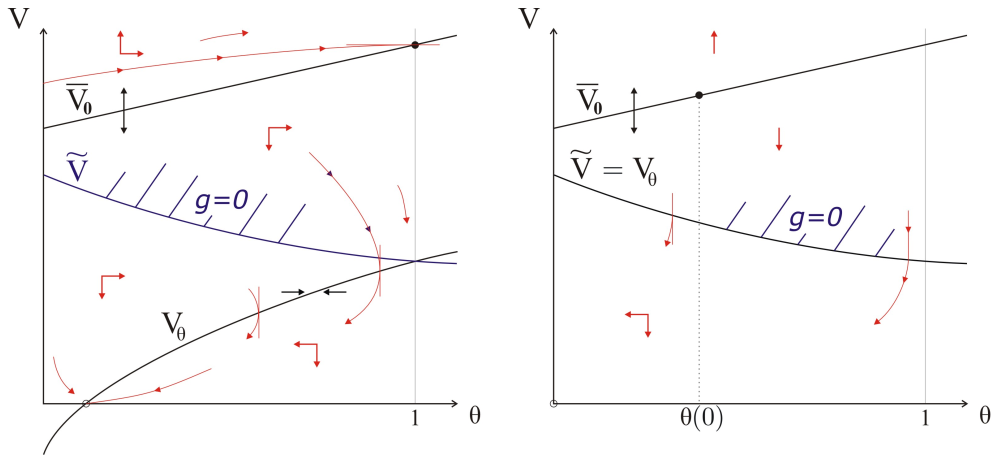


Figure 5.2: Dynamics in Case 2 (Left Panel: $\psi > 0$, Right Panel: $\psi = 0$)

Case 2:

$$\frac{(1-\alpha)\sigma\frac{L}{a}}{\rho+\psi} < 1-(1-\alpha)\sigma. \tag{5.15}$$

Here, $\bar{V}_0(0) > \tilde{V}(0)$ so that the curve $\bar{V}_0(\theta)$ is located above the curve $\tilde{V}(\theta)$. V rises above and falls below $\bar{V}_0(\theta)$. Suppose $\psi > 0$. Then, the point $(\theta, V) = (1, \bar{V}_0(1))$ is a steady state. As can be seen from the left panel of Figure 5.4, for each $\theta \in [0,1]$, there exists a unique path converging to this steady state. For $\psi = 0$, we have $(\dot{\theta}, \dot{V}) = (0,0)$ for all $(\theta, \bar{V}_0(\theta))$. Hence, for any initial proportion of competitive markets, $\theta(0)$, there is a steady state $(\theta, V) = (\theta(0), \bar{V}_0(\theta(0)))$ (see the right panel of Figure 5.4).

Case 3:

$$1-(1-\alpha)\sigma \le \frac{(1-\alpha)\sigma\frac{L}{a}}{\rho+\psi} \le \alpha^{1-\epsilon}. \tag{5.16}$$

In this, intermediate, case the curves $\bar{V}_0(\theta)$ and $\tilde{V}(\theta)$ intersect for some $\theta \in [0,1]$.[13] As in case 2, V rises above $\bar{V}_0(\theta)$ and falls below the curve (see Figure 5.3).

Next, consider the region with positive growth (i.e., $g > 0$). From (5.9) and (5.11), V is constant if $V = 0$ or if

$$V = \left(\rho+\psi+\frac{L}{a}\right)\frac{1-\theta(1-\alpha^{1-\epsilon})}{1-\theta[1-\alpha^{1-\epsilon}-(1-\alpha)\sigma]} \equiv \bar{V}(\theta). \tag{5.17}$$

$\bar{V}(0) = \rho+\psi+L/a$, $\bar{V}(1) = (\rho+\psi+L/a)/[1+(1-\alpha)\alpha^{\epsilon-1}\sigma]$, and $\bar{V}'(\theta) < 0$ for all $\theta \in [0,1]$. V rises

[13]Notice that the terminology is somewhat loose with regard to a common point at one of the boundaries, $\theta = 0$ or $\theta = 1$ (i.e., when one equality in (5.16) is strict).

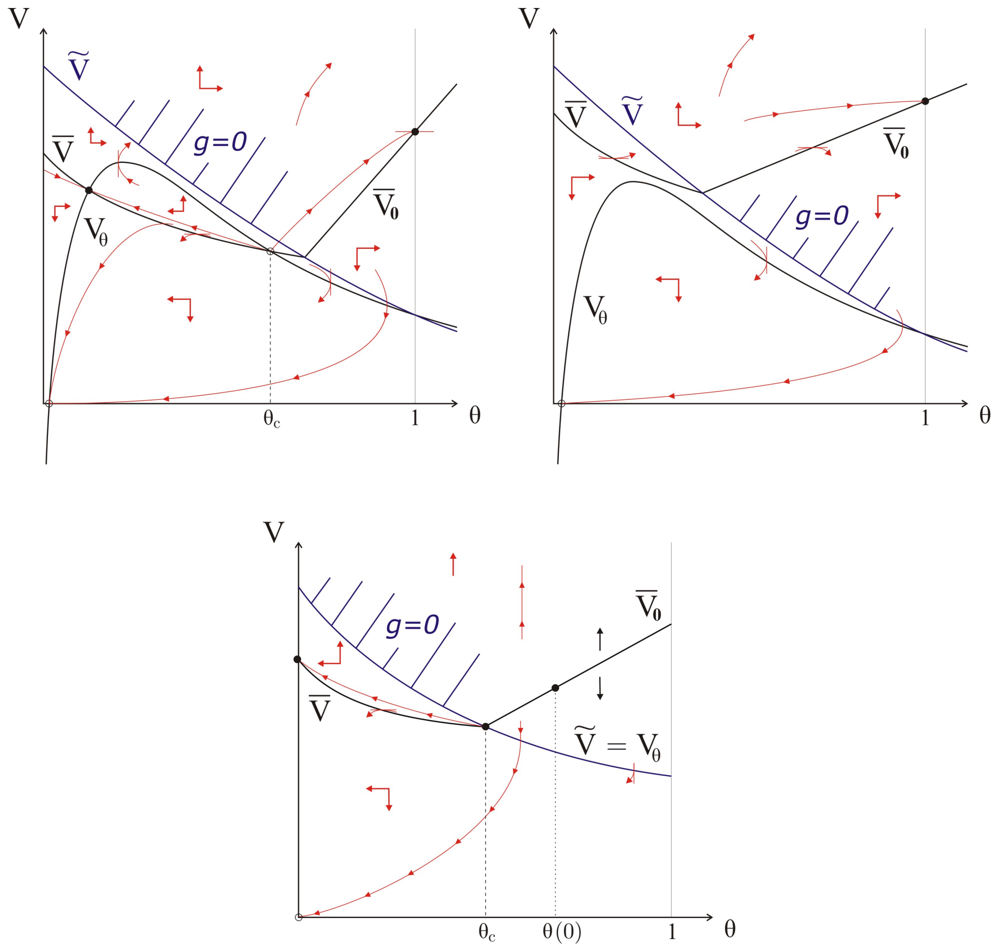


Figure 5.3: Dynamics in Case 3 (Upper Panels: $\psi > 0$, Lower Panel: $\psi = 0$)

above the stationary locus and falls below. From (5.10), $\dot{\theta} = 0$ for

$$V = \left(\frac{L}{a} - \frac{1-\theta}{\theta}\psi\right)\frac{1-\theta(1-\alpha^{1-\epsilon})}{1-\theta\left[1-\alpha^{1-\epsilon}-(1-\alpha)\sigma\right]-(1-\alpha)\sigma} \equiv V_\theta(\theta). \tag{5.18}$$

Using (5.12), we can rewrite $V_\theta(\theta)$ as

$$V_\theta(\theta) = \tilde{V}(\theta) - \frac{1-\theta}{\theta}\psi\frac{1-\theta(1-\alpha^{1-\epsilon})}{1-\theta\left[1-\alpha^{1-\epsilon}-(1-\alpha)\sigma\right]-(1-\alpha)\sigma}. \tag{5.19}$$

For $\psi > 0$, from (5.18) and (5.19), we have $V_\theta(\theta) = 0$ for $(0 <)\ \theta = \psi/(\psi + L/a)\ (< 1)$ and for $\theta = 1/(1-\alpha^{1-\epsilon})\ (< 0)$. Moreover, $V_\theta(\theta) \to -\infty$ as $\theta \to 0$ from above, V_θ is continuous on $\theta \in (0, 1]$, $V_\theta'(1) = \psi - (L/a)\alpha^{\epsilon-1}(1-\alpha)\sigma$, and $V_\theta(\theta) < \tilde{V}(\theta)$ for all $\theta \in (0, 1)$. For $\psi = 0$, on the other hand, $V_\theta(\theta) = \tilde{V}(\theta)$ for all θ and $V_\theta(1) = \tilde{V}(1) = L/a$.

Case 1: Suppose $\psi > 0$. Then, the case distinction (5.14) implies $\bar{V}(\theta) < \tilde{V}(\theta)$ for all $\theta \in [0, 1]$ (since $\bar{V}(0) < \tilde{V}(0)$ and $\bar{V}(\theta) = \tilde{V}(\theta)$ for some $\theta \in (0, 1]$ contradicts the case distinction). Moreover,

$V_\theta(\theta) < \tilde{V}(\theta)$ for all $\theta \in (0,1)$, and $V_\theta(1) = \tilde{V}(1)$ (i.e., $V_\theta(1) > \bar{V}(1)$). It follows that the stationary loci $\bar{V}(\theta)$ and $V_\theta(\theta)$ intersect an odd number of times on $(0,1)$. The fact that $V'_\theta(1) < 0$ implies that $V_\theta(\theta)$ has an interior local maximum on $(0,1]$. From (5.17) and (5.18), $\bar{V}(\theta) = V_\theta(\theta)$ for $\theta = 1/(1-\alpha^{1-\epsilon})$ (< 0) and for those θ's which satisfy the equality $\bar{V}(\theta)/[1-\theta(1-\alpha^{1-\epsilon})] = V_\theta(\theta)/[1-\theta(1-\alpha^{1-\epsilon})]$. This is a quadratic equation, with an even number of real-valued solutions. It follows that $\bar{V}(\theta)$ and $V_\theta(\theta)$ intersect exactly once in the interval $[0,1]$, which proves that a unique steady state exists in the $g > 0$-region. As can be seen from the left panel of Figure 5.1, the steady state is a saddle point. For $\psi = 0$, we have $\dot{\theta}/\theta = -g < 0$. $(\theta, V) = (0, \bar{V}(0))$ is the unique steady state in the $g > 0$-region and is a saddle point (see the right panel of Figure 5.1). For each $\theta \in [0,1]$, there exists a unique trajectory converging to the steady state both for $\psi > 0$ and for $\psi = 0$. Divergent paths can be ruled out adapting the arguments put forward by Grossman and Helpman (1991b, p. 61): paths starting above the saddle path yield $V \to \infty$ and $\theta \to \theta' > 0$, where $\theta' = 1$ if $\psi > 0$ (see Figure 5.1). However, once the economy is in the $g = 0$-region, $\pi n = \sigma(1-\alpha)/[1-\theta(1-\alpha^{1-\epsilon})] \geq \sigma(1-\alpha)\alpha^{\epsilon-1}$ and, from (5.6), $vn \geq \sigma(1-\alpha)\alpha^{\epsilon-1}/(\rho+\psi)$. This contradicts $V \to \infty$. Paths starting below the saddle path converge to $(\theta, V) = (\psi/(\psi + L/a), 0)$ (see Figure 5.1). As $\pi \leq \sigma(1-\alpha)/n$, we have $nv \leq \sigma(1-\alpha)/(\rho+\psi)$, which contradicts $V \to 0$.

Case 2: In the $g = 0$-region, $\dot{V}/V < 0$ below the curve $\bar{V}_0(\theta)$. A fortiori, from (5.11), $\dot{V}/V < 0$ in the $g > 0$-region. As in case 1, if $\psi > 0$, the $\dot{\theta} = 0$-locus diverges to $-\infty$ as $\theta \to 0$ from above and satisfies $V_\theta(1) = \tilde{V}(\theta)$. As can be seen from the left panel in Figure 5.4, all paths except the one converging to $(\theta, V) = (1, \bar{V}_0(1))$ violate perfect foresight analogously to the divergent paths in case 1. For $\psi = 0$, given a starting value $\theta(0)$, the only trajectory consistent with perfect foresight entails that the economy jumps to the steady state $(\theta(0), V_0(\theta(0)))$.

Case 3: In this intermediate case, the curves $\bar{V}_0(\theta)$ and $\tilde{V}(\theta)$ intersect for some $\theta \in [0,1]$. From (5.11), the stationary locus for V is continuous on the border between positive and zero growth, $\tilde{V}(\theta)$. When the first inequality in (5.16) is strict, we have $\bar{V}(0) = \tilde{V}(0)$. For $\psi > 0$, by the arguments put forward in case 2, the number of intersections of $\bar{V}(\theta)$ and $V_\theta(\theta)$ in the $g > 0$-region is two (see the upper left panel of Figure 5.3) or zero (see the upper right panel of Figure 5.3).[14] In the former subcase (two intersections), let θ_c denote the abscissa value of the south-eastern intersection. Then, for each $\theta < \theta_c$, the unique trajectory consistent with perfect foresight converges to the north-western steady state, and

[14]The dynamics are similar in all respects important for our purposes if $V_\theta(\theta)$ is montonically increasing, rather than taking on a maximum on $\theta \in (0,1]$.

for each $\theta > \theta_c$, the unique trajectory consistent with perfect foresight converges to $(\theta, V) = (1, \bar{V}_0(1))$. Similarly, in case of $\psi = 0$, let θ_c denote the θ-value at which the $\dot{V} = 0$-locus intersects the $g = 0$-boundary, $\tilde{V}(\theta)$. For $\theta(0) < \theta_c$, the economy converges to $(0, \bar{V}(0))$. By contrast, for $\theta(0) > \theta_c$, the economy jumps to $(\theta(0), \bar{V}_\theta(\theta(0)))$ (see the lower panel of Figure 5.3).

5.5 Results

In this section, we use our findings about the model dynamics to prove three propositions about growth and welfare and derive corollaries addressing the issue of second-best competition policies mentioned in the Introduction.

The first result states that no growth may be better than some growth. To illustrate this, suppose it is possible to effectively protect monopolies indefinitely so that $\psi = 0$. Suppose L is sufficiently large so that (5.14) or (5.16) holds for $\psi = 0$, i.e. case 1 or case 3 applies, and there is a steady state with positive growth. We know from Section 5.4 that if $\theta(0) = 0$, the economy settles down at a steady state with $\theta(t) = 0$ for all t (see the right panel of Figure 5.1 and the lower panel of Figure 5.3, respectively). From (5.9) and (5.11) with $\dot{\theta} = 0$, it follows that

$$g = (1-\alpha)\sigma\frac{L}{a} - [1-(1-\alpha)\sigma]\rho. \tag{5.20}$$

Let L_+ denote the value for L such that $g = 0$ for $L \leq L_+$ and $g > 0$ for $L > L_+$ ($L = L_+$ implies that the first weak inequality in (5.16) holds with equality):

$$L_+ \equiv a\rho\left[\frac{1}{(1-\alpha)\sigma} - 1\right].$$

Proposition 5.1 ("benefits of no growth"). *Suppose (5.14) or (5.16) holds for $\psi = 0$ (i.e., $L \geq L_+$). Then, there exists L_c ($> L_+$) such that for $L \in (L_+, L_c)$, intertemporal utility with competitive prices, full employment, $\theta(t) = 1$, and $g(t) = 0$ for all $t \geq 0$ is higher than in the equilibrium with $\theta(t) = 0$ and $g(t) > 0$ given by (5.20) for all $t \geq 0$.*[15]

Proof. Let θ and g be constant. Furthermore, assume prices are set competitively in the y-sector and in n_c intermediate goods markets, while monopoly prices are charged in n_m intermediate goods markets. Then intertemporal utility, U, is

$$\rho U - \frac{1-\alpha}{\alpha}\sigma \ln n(0) = \frac{\sigma}{\alpha}\ln\left[(n_c x_c)^\alpha \theta^{1-\alpha} + (n_m x_m)^\alpha (1-\theta)^{1-\alpha}\right] + (1-\sigma)\ln Y + \frac{1-\alpha}{\alpha}\frac{\sigma}{\rho}g. \tag{5.21}$$

[15]In Grossman and Helpman's (1991b, Ch. 4) quality upgrading model, equilibrium growth is positive although zero growth would be preferable if $\rho a/L$ lies in the interval $(\log\lambda, \lambda - 1)$, where λ (> 1) is the size of a quality jump.

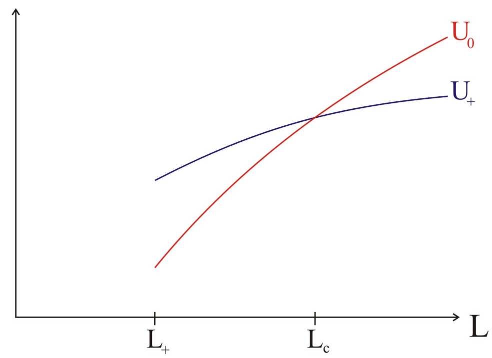


Figure 5.4: Comparing Growth and No-Growth Equilibria

The important point to notice is that monopoly pricing in the intermediate goods sector causes the usual static welfare loss. To see this, notice that an allocation of labor across the intermediates and good y that maximizes static welfare requires symmetry across the intermediates (i.e., $x(j) = x$ for all $j \in [0, n]$) and $nx/Y = \sigma/(1-\sigma)$. For $\theta(t) = 1$, as $n_c(t) = n(t)$, the symmetry condition is satisfied. And from zero profit in x- and y-production (i.e., $n_c x_c = \sigma/w$ and $Y = (1-\sigma)/w$, respectively), the allocation of labor is efficient. Next, consider the allocation with $\theta(t) = 0$ and $g(t) > 0$. Equations (5.2)-(5.5) and $\theta = 0$ imply $nx/Y = \alpha\sigma/(1-\sigma)$. That is, markup pricing leads to too low a level of x-production relative to y.

Let U_0 denote the intertemporal utility level with $\theta(t) = 1$ and $g(t) = 0$, and U_+ the utility level obtained in the steady-state equilibrium with $\theta(t) = 0$ and $g(t) > 0$ given by (5.20). Let $L \to L_+$ from above. By the definition of L_+, the growth rate, g, converges to zero. Given that this implies that the combined labor input in x- and y-production converges to L, the static welfare loss is of non-infinitesimal magnitude. Consequently, $U_0 > U_+$ as $L \to L_+$ from above.[16] The right hand side of (5.21) is concave in $n_c x_c$ and in $n_m x_m$ but linear in g (which is itself linear in L, see (5.20)). So there is an $L_c > L_+$ such that $U_+ \geq U_0$ for $L \geq L_c$ (see Figure 5.5). □

[16] It is possible (albeit not necessary) to calculate the difference in welfare levels explicitly. Simple manipulations show that $\rho U_0 - [(1-\alpha)/\alpha]\sigma \ln n(0) = \ln L + \sigma \ln \sigma + (1-\sigma)\ln(1-\sigma)$ and $\rho U_+ - [(1-\alpha)/\alpha]\sigma \ln n(0) = \ln L + \sigma \ln \sigma + (1-\sigma)\ln(1-\sigma) + \sigma \ln \alpha - \ln[1-(1-\alpha)\sigma]$ as g goes to zero. So $\rho(U_0 - U_+) = \ln[1-(1-\alpha)\sigma] - \sigma \ln \alpha$ as g goes to zero. That $U_0 - U_+$ is strictly positive follows from the fact that it equals zero for $\sigma = 0$ and $\sigma = 1$ and is strictly concave. The fact that $U_0 - U_+ = 0$ for $\sigma = 1$ highlights that the presence of a traditional sector is essential for our argument. Monopolistic price setting per se does not cause a static distortion; distortions obtain when different goods have different markups (see Grossman and Helpman, 1991b, p. 70).

A direct corollary of Proposition 5.1 is that no patent protection may be preferable to very strict patent protection. Suppose by giving up patent protection, the policymaker can raise the imitation rate from $\psi = 0$ to $\psi = \bar{\psi}$ (> 0). Consider the extreme case in which imitators learn instantaneously how to copy new innovations, so that $\bar{\psi} \to \infty$.

Corollary 5.1 ("benefits of preventing growth"). *For L slightly greater than L_+, $\theta(0) = 0$, and $\bar{\psi} \to \infty$, intertemporal utility is higher with $\psi = \bar{\psi}$ than with $\psi = 0$.*

Proof. As explained above, with $\psi = 0$, the economy settles down at a steady state with positive growth, $\theta = 0$, and, hence, with intertemporal utility U_+ (the first inequality in (5.16) is strict). As ψ rises sufficiently far, (5.15) becomes valid and case 2 applies. As illustrated by the left panel of Figure 5.4, zero growth prevails, and the economy converges towards $(1, \bar{V}_0(1))$. For $\bar{\psi} \to \infty$, the convergence process becomes infinitely short, so that θ quickly goes to unity and intertemporal utility is close to U_0. Since, by Proposition 5.1, $U_0 > U_+$ for L slightly greater than L_+, giving up patent protection raises welfare. □

Numerical results, admittedly, suggest that some growth is better than no growth. However, it is possible to construct counterexamples with parameter values that should not be deemed unrealistic a priori. For instance, let $\rho = 0.04$, $\alpha = 5/8$ (which gives rise to the standard 60% markup), $a = 1$, and $n(0) = 1$. We let $\sigma = 0.999$ or $\sigma = 0.5$ and choose L such that without imitation (i.e., if $\psi = 0$) $g = 0.5\%$, which implies 0.3% growth in the x-sector's real output (i.e., $L = 0.0801$ or $L = 0.2$, respectively). For $\sigma = 0.999$, we get $U_0 = -63.3097 < -63.0515 = U_+$. For $\sigma = 0.5$, on the other hand, $U_0 = -57.5646 > -57.9446 = U_+$. So the economy with $\sigma = 0.5$ (but not the economy with σ close to unity) would be willing to give up long-term manufacturing output growth of 0.3% in exchange for static efficiency. That is, raising the imitation rate from zero to infinity is beneficial to this economy's representative consumer.

Our second proposition states that the economy may get stuck in a "no-growth" trap (even though there exists a steady state with positive growth) due to too much competition initially.

Proposition 5.2 ("no-growth trap"). *Suppose (5.16) holds (case 3). Suppose further that either $\psi > 0$ and $\bar{V}(\theta)$ and $V_\theta(\theta)$ intersect twice in the $g > 0$-region, or else $\psi = 0$. Then, $g(t) > 0$ for all $t \geq 0$ if $\theta(0) < \theta_c$, while there is $t_c \geq 0$ such that $g(t) = 0$ for all $t \geq t_c$ if $\theta(0) > \theta_c$.*

Proof. This simply rephrases the results of the analysis of case 3 in Section 5.4. If $\psi > 0$, t_c (> 0) is the point in time at which the trajectory converging to $(1, \bar{V}_0(1))$ crosses the $g = 0$-boundary. $t_c = 0$

for $\psi = 0$ (see the upper left and the lower panels of Figure 5.3, respectively). $\square$

The intuition for Proposition 5.2 is: the lower its competitors' prices, the lower the share of aggregate demand that accrues to a potential innovator. If too many competitors supply at competitive prices, it does not pay to innovate, even though it would pay if the competitors' products were more expensive. As an example, let $\rho = 0.02$, $\psi = 0.01$, $\alpha = 0.6$, $a = 1$, $\sigma = 1$, and $L = 0.15$. This gives rise to case 3, and the stationary loci for V and θ intersect twice in the $g > 0$-region, at $\theta = 0.4352$ and $\theta = 0.7404 \equiv \theta_c$. The growth rate corresponding to the steady state with $\theta = 0.4352$ is $g = 1.30\%$, which implies 0.87% growth of the x-sector. So this economy fails to reach a steady state with 0.87% manufacturing output growth if the initial proportion of competitive markets exceeds 74.04%.

The implication of Proposition 5.2 for competition policy is that quick deregulation of monopolies may do more harm than good, as it makes it harder for a potential innovator to compete with incumbent producers. To illustrate this, consider an emerging economy with state monopolies in $n_m = \theta_m n$ intermediate goods markets initially. Assume $\theta_m < \theta_c$. Without deregulation, $\theta(0) = 1 - \theta_m$, and the economy converges to the saddle-point stable steady state, in which growth is positive. Suppose, by contrast, the government deregulates some of the monopolies, in which case they instantaneously become perfectly competitive. That is, the government determines a starting value $\theta(0)$ in the interval $[\theta_m, 1]$ ($\ni \theta_c$). From Proposition 5.2, we have:

Corollary 5.2 ("perils of quick liberalization"). *Let the conditions of Proposition 5.2 be satisfied. If markets are deregulated such that $\theta(0) > \theta_c$, then there is a $t_c > 0$ such that $g(t) = 0$ for all $t \geq t_c$, while $g(t) > 0$ for all $t \geq 0$ without deregulation.*

In the example above, if $\theta_m = 74\%$, then deregulating a further 0.1% of the markets means giving up 0.87% long-term manufacturing growth.

Tang and Wälde (2001) show that a no-growth trap is possible in the two-country open economy version of our model with $\sigma = 1$ and $\psi = 0$. Proposition 5.2 is strongly reminiscent of their finding. We now turn to the m-country open economy version of our model. We show that under certain conditions the m-country economy behaves exactly identically to the hypothetical integrated economy that occurs in the absence of national borders (i.e., the restrictions on labor movements they imply). To do so, we generalize the analysis in Arnold (2007) (which assumes $\psi = 0$). The Tang-Wälde (2001) result is then obtained as a corollary to this replication theorem.[17] Consider a world economy

[17] Our result also generalizes Tang and Wälde (2001), to the cases $\sigma > 0$ and $m > 2$.

made up of m (≥ 2) countries of the type introduced in Section 5.3 (i.e., with identical parameter values everywhere). Variables referring to individual countries, i, are distinguished by a superscript i ($= 1, \ldots, m$). Variables without the superscript are world aggregates. We assume that knowledge spillovers in R&D are international in scope, so that the R&D technologies become $\dot{n}^i = nL_R^i/a$. An important issue is which products can be produced where. We start with the assumptions least conducive to the possibility of replication: non-imitated goods have to be produced where they were invented, and imitation is also "local", in that in short time intervals, dt, a fraction $(L^i/L)\psi\, dt$ of the n_m goods not yet imitated before becomes producible in country i and is also produced in country i: $\dot{n}_c^i = (L^i/L)\psi n_m$. We say that replication of the equilibrium of the integrated economy (ignoring national borders that inhibit movements of labor across borders) is feasible if this allocation is an equilibrium of the world economy (with national borders) as well.

Proposition 5.3 (**"replication"**). *If*

$$L^i - n_c^i(t)x_c(t) - n_m^i(t)x_m(t) \geq 0, \qquad \textit{for all } i = 1, \ldots, m,\ t \geq 0, \tag{5.22}$$

then replication is feasible.

Proof. Ignoring national borders, the equilibrium obeys equations (5.1)-(5.8) in Section 5.4 (where $L \equiv \sum_{i=1}^m L^i$). We have to show that this set of equations is also satisfied in the world economy with national borders. Equation (5.1) follows from adding up $\dot{n}_c^i = (L^i/L)\psi n_m$ for all $i = 1, \ldots, m$. The conditions for utility maximization in (5.2) are unaffected by the presence of national borders. Since the cost minimization problem is also unchanged, so are the input coefficients, $a(j)$, the zero profit condition $p_X = [\int_0^n p(j)^{1-\epsilon}dj]^{1/(1-\epsilon)}$, the demands for intermediates $x(j) = a(j)X$, and, therefore, the pricing rules and ensuing profit in (5.3) and the expressions for x_m and x_c in (5.4) (where X is the world production of good x, x_m is the output of any monopolistically supplied intermediate, and x_c is the output level of any competitively supplied intermediate). Evidently, the equations for pricing of the final goods, the value of an innovation, and free entry into R&D, (i.e., (5.5)-(5.7), respectively) hold true in equilibrium. Finally, labor market clearing in country i requires

$$L^i = a\frac{\dot{n}^i}{n} + n_c^i x_c + n_m^i x_m + Y^i. \tag{5.23}$$

Assumption (5.22) ensures that for each country, i, given n_c^i and n_m^i, there exist $\dot{n}^i \geq 0$ and $Y^i \geq 0$ such that (5.23) is satisfied. Adding up (5.23) for all $i = 1, \ldots, m$ yields (5.8). $\square$

Propositions 5.2 and 5.3 can be jointly used to prove a generalized version of the Tang-Wälde (2001) theorem on the existence of a no-growth trap due to the opening up of international trade between several countries. To do so, assume that at time $t = 0$, m countries with free international flows of knowledge between them start to engage in trade with each other. Suppose that, while still in autarky (i.e., before time $t = 0$), the producers do not take into account the possibility of future trade liberalization, so that they do not have an incentive to avoid the invention of identical intermediates in different countries ("duplication"). Let n denote the total "number" of different intermediates producible somewhere in the world economy and n_d the "number" of duplicated intermediates at the point in time when trade is liberalized. From Propositions 5.2 and 5.3, we obtain:

Corollary 5.3 (**generalization of Tang and Wälde, 2001**). *Let the conditions of Proposition 5.2 be satisfied, and (5.22) holds. Then the no-growth trap described in Proposition 5.2 is an equilibrium of the world economy if* $n_d/n > \theta_c$.

From a policy point of view, this corollary implies that, like quick deregulation in a closed economy, quick trade liberalization can lead to stagnation in the long term: if countries decide to liberalize trade at a point in time when $n^d/n \geq \theta_c$, growth will come to a halt.[18]

Under the maintained assumptions, replication may fail due to the fact that a country, i, starts out with a disproportionately large number of blueprints. If, for instance, $n^i_m x_m > L^i$, then replication is not feasible, as country i does not have enough resources to manufacture the integrated equilibrium outputs of the intermediates with a domestic monopolist. On the other hand, as $L^i - n^i_c(t)x_c(t) - n^i_m(t)x_m(t) = L^i_R + Y_i$, (5.22) is satisfied with strict equality in a steady state with L^i_R and Y_i positive, it follows that if the world economy is close to its steady state initially, then (5.22) will hold. Moreover, the problem vanishes altogether under assumptions more conducive to the possibility of replication. To see this, assume that intermediates invented in one country can be manufactured in a different country, i, either within multinational firms or via international patent licensing. Assume further that once imitation is possible in one country, i, it is possible in each country, $i = 1, \ldots, m$ ("simultaneous imitation"). Then, all production activities are "footloose". Equation (5.23) is satisfied, for instance, for $\dot{n}^i/\dot{n} = n^i_c/n_c = n^i_m/n_m = Y^i/Y = L^i/L$ (and for many other allocations of productive activities across countries as well). This proves:

[18] To maintain growth, one has to wait for a point in time where $n_d/n < \theta_c$, or one has to choose a non-stationary path of the imitation rate, ψ, which prevents too much competition. A thorough analysis would require announcement effects and/or explicitly making ψ non-stationary.

Corollary 5.4. *Let the conditions of Proposition 5.2 be satisfied. Then, with simultaneous imitation and either multinational firms or international patent licensing, the no-growth trap described in Proposition 5.2 is an equilibrium of the world economy if* $n_d/n > \theta_c$.

5.6 Conclusion

This chapter is concerned with the question of how competition with "cheap", i.e., old or traditional, goods affects the incentives to enter markets with new, innovative products. It shows that no growth may be better than some growth and that both a closed economy and a world economy made up of several countries engaged in free trade with each other may get stuck in a no-growth trap. As a result, growth-enhancing policies may be misguided, and quick deregulation as well as quick trade liberalization possibly lead to avoidable stagnation in the long term.

Appendix 5.A Imitation Acts Like Additional Discounting

In this appendix, we show that imitation as defined in the main text acts like additional discounting.[19] Albeit intuitive, this is readily proven as follows. Let $\Psi(t, \tau + \Delta)$ denote the probability that no imitation occurs in the time span $(t, \tau + \Delta)$. Since the Poisson process lacks memory, $\Psi(t, \tau + \Delta) = \Psi(t, \tau)\Psi(\tau, \tau + \Delta)$ holds. If Δ refers to a short period of time, the latter factor equals $1 - \psi\Delta$. Now, let Δ go to zero,

$$\lim_{\Delta \to 0} \frac{\Psi(t, \tau + \Delta) - \Psi(t, \tau)}{\Delta} = \frac{\partial \Psi(t, \tau)}{\partial \tau} = -\Psi(t, \tau)\psi.$$

Solving this first order variable coefficient differential equation gives the probability of still being a monopolist τ periods after t. Since $\Psi(t, t) = 1$, we have $\Psi(t, \tau) = \exp\left[-\int_t^\tau \psi(\xi) d\xi\right]$.

[19] Cf. the Poisson imitation model by Grossman and Helpman (1991a).

Part III

Idiosyncratic Income Risk and the Labor Market

Chapter 6

Labor Market Matching with Savings

More often than not, individual's decisions are subject to risk and uncertainty.[1] While some risk is due to systematic, aggregate shocks, other risk is less systematic or even characteristic of a single asset or agent. These idiosyncratic risks are diversifiable to a large extent if the underlying shock is publicly observable. There are then, typically, markets where individuals trade state-contingent claims to insure against the financial consequences of risks such as credit or liability fraud, accidents, or, in some cases, disability or health problems, to name just a few. As indicated by the latter example, however, some, often substantial, idiosyncratic risks lack a market for insurance. This is in particular true if the shocks are private information. Informational frictions, through moral hazard and adverse selection, among other frictions may prevent markets from being complete. For example, as love is private information, there is no market for a spouse to insure against the potential loss of wealth caused by a divorce due to reported loss of love. Similarly, there is no market for individual workers to insure against unemployment, since again, key determinants of unemployment like productivity or effort are often unobservable. In certain cases where no market exists, governments can improve allocations by enforcing transfers among individuals, effectively avoiding adverse selection through statutory insurance, or by disentangling the insurance costs from the benefits. Governments may further be able to collect information that is inaccessible in a market or otherwise extremely costly.[2] Evidently, any normative analysis of such policies requires a thorough understanding of how individuals

[1]The distinction between risk and uncertainty dates back to Knight (1921). Hubbard (2007, p. 46) refers to uncertainty as the existence of more than one possible outcome and defines risk as a state of uncertainty where some of the possibilities involve a loss for the decision maker.

[2]Cf. Acemoglu, Golosov, and Tsyvinski (2008) for a comparison of government-regulated allocations to market allocations.

behave optimally in the presence of uninsured idiosyncratic risks.

In what follows, we employ dynamic partial equilibrium heterogeneous agents models in continuous time to show how risk averse individuals react optimally to idiosyncratic fluctuations in their labor incomes. There is no insurance market for this risk, but since we abstract from aggregate shocks, individuals may use a riskless asset with a deterministic return to smooth their consumption to some extent.

We start off by briefly reviewing some of the influential work on risk-averse individuals and equilibrium unemployment with incomplete markets. The following section introduces a dynamic partial equilibrium framework that serves as a baseline model for the analysis of labor market search and matching when individual workers are able to invest in a riskless asset. Upon describing the environment, we characterize the individual's optimal behavior and consumption dynamics under CRRA preferences. The following section is devoted to the individual's consumption Euler equation under labor income uncertainty. We proceed by reviewing the recent model by Shimer and Werning (2008), who characterize the optimal unemployment policy in a dynamic search model with CARA preferences. Building on their analysis, we then show how imposing a borrowing constraint and nonnegative consumption accounts for history dependence in the decision to accept a job, even under CARA. We characterize the optimal savings behavior and derive the equilibrium reservation wage as a function of wealth. Finally, we conclude this part by showing how a change in the attitude towards risk and uncertainty affects the optimal amount of savings in the standard two period model of optimal consumption under income uncertainty.

6.1 Introduction

Research into risk averse agents, savings, and labor market search in economies with incomplete markets has a longstanding tradition. Somewhat roughly, we can subdivide the literature into three strands.

The first strand is on labor market search and business cycles. It involves a merger of the canonical Kydland-Prescott (1982) real business cycle (RBC) model with search frictions of the Mortensen-Pissarides (1994) type of equilibrium unemployment. Seminal contributions include Andolfatto (1996), Merz (1995), and den Haan, Ramey, and Watson (2000). The starting point of this literature is an unsatisfying propagation of shocks in early RBC models. Including costly job search in a Kydland-Prescott (1982) RBC model, Andolfatto (1996) reduces the contemporaneous correlation of hours worked and productivity relative to the standard RBC model, matches the observed larger fluctuation in hours worked compared to the fluctuation of wages, and generates output dynamics that no longer simply replicate the impulse dynamics.[3] Den Haan, Ramey, and Watson (2000) further increase the persistence of shocks by endogenizing the process of job destruction. The fluctuations of the job destruction rate over the cycle, moreover, magnify the output effects of shocks.[4] The model matches well the empirically observed cyclical patterns of job creation and destruction. Typical for this strand of literature, however, den Haan, Ramey, and Watson (2000) assume income pooling in large families and thus abstract from idiosyncratic consumption risk due to job separation and unemployment.

In pioneering work, a second strand of literature dispenses with the assumption of income pooling and takes uninsured income fluctuations into account. Here, substantial contributions are by İmrohoroğlu (1989), Castañeda, Díaz-Giménez, and Ríos-Rull (1998, 2003), and Heer (2001).[5] İmrohoroğlu (1989) computes the welfare costs of business cycle fluctuations holding fixed the capital stock and the interest rate. She finds large utility costs of cycles, in the range of 0.3 and 1.5% of aggregate consumption. In a complete markets model by Lucas (1987), the corresponding figure is substantially lower, about 0.1% of consumption. Allowing agents to react to income fluctuations by building up precautionary savings so that asset holdings differ between the benchmark cases with and without income fluctuations,

[3]In a similar RBC model with search unemployment, Merz (1995) generates dynamic patterns in line with labor market data: real wages deviate from productivity, productivity is leading unemployment over the cycle, and the contemporary correlation between vacancies and the unemployment rate is negative.

[4]In their model, the standard deviation of output is 2.5 times the standard deviation of shocks; in the standard RBC model, it is less than two.

[5]Heer and Maußner (2005, Chapters 5 and 6) provide a compendium of the numerical methods necessary to compute the stationary equilibrium and the evolution of the wealth distribution of heterogeneous agent economies. Cf. Ríos-Rull (1995, 1999). Heer and Maußner (2005, Ch. 5.3 and 6.5) also provide various applications and point to further important research in the field (İmrohoroğlu, 1992, on the welfare cost of inflation; Aiyagari, 1994, on an endogenous borrowing constraint that ensures repayment with probability one in all possible states; Hubbard, Skinner, and Zeldes, 1995, on the implications of various idiosyncratic shocks for the accumulation of wealth in a life-cycle model; Huggett, 1996, on the wealth distribution in life-cycle models when the earnings distribution is calibrated to real world data; Heckman, Lochner, and Taber, 1998, on dynamic general equilibrium models with endogenous heterogeneous human capital accumulation; Ventura, 1999, on the quantitative consequences of a revenue-neutral flat tax reform for capital accumulation, labor supply, and the distribution of earnings, income, and wealth; Huggett and Ventura, 2000, on the impact of a social security reform on the distribution of welfare, consumption, and leisure across households; Caucutt, İmrohoroğlu, and Kumar, 2003, on the implications of increasing subsidies to higher education for inequality, welfare, and efficiency; Heer and Trede, 2003, on the quantitative effects of revenue-neutral flat tax and consumption tax reforms).

Heer (2001) suggests that the welfare costs of cycles obtained by İmrohoroğlu (1989) provide a lower bound. Castañeda, Díaz-Giménez, and Ríos-Rull (1998) employ a neoclassical growth model with heterogeneous agents and aggregate shocks to investigate the dynamics of the income distribution over the cycle. Inter alia, they show that uninsured unemployment spells, not cyclical moving factor shares, account for much of the dynamics of the income distribution when partitioning the set of households in five groups according to differences in permanent earnings.[6] Castañeda, Díaz-Giménez, and Ríos-Rull (2003) substantiate the outstanding performance of a model with uninsured idiosyncratic efficiency labor endowment shocks and ex ante identical households. Calibrated to the Lorenz curve of earnings and wealth from 1992 U.S. consumer survey data, their model matches the observed U.S. earnings and wealth inequality almost exactly.

More recently, a third strand of literature emerged from an extension of the Mortensen-Pissarides (1994) equilibrium search model to include incomplete markets in the spirit of Bewley (undated), Huggett (1993), and Aiyagari (1994). This literature, initiated by Krussel, Mukoyama, and Şahin (2007) and Nakajima (2007), offers new mechanisms to address the ongoing rise of within-group heterogeneity. For example, in a model with aggregate certainty and period-per-period wage negotiations, Krussel, Mukoyama, and Şahin (2007) show that financial wealth acts as an outside option in wage negotiations and raises the worker's wage. Further, including aggregate shocks, consumption smoothing permits capital from jumping to its new long-run level and thereby generates transitional dynamics for the tightness of the labor market (measured by the ratio of vacancies to unemployed). In models with linear utility, labor market tightness is a jump variable. However, Krussel, Mukoyama, and Şahin (2007) assert that both mechanisms matter little for aggregate dynamics. In most calibrations, there is also little dispersion of wages due to different levels of wealth.[7] While there is some difference in key variables between linear and concave utility in Hagedorn and Manovskii's (2008) calibration, risk aversion does not add anything if the model is calibrated to labor market flows like in Shimer (2005) and Hall (2005). In a related paper, Nakajima (2007) replaces individual wage bargaining by the bargain between a representative firm and a representative worker, who is only constructed for that single purpose. He finds that risk aversion and the lack of insurance strongly amplify shocks, independently

[6] The differences in permanent earnings are measured by an index constructed from longitudinal U.S. Panel Study of Income Dynamics data.

[7] This may be due to the assumption of exogenous job separation. Using estimates on French panel data, Algan, Chéron, Hairault, and Langot (2003) substantiate a significant positive effect of short-term liquid asset holdings on job quits.

of including leisure in the utility function or not. Nakajima's (2007) model, however, generates little inequality among workers as well. The only heterogeneity is due to different (un)employment histories.

Like Krussel, Mukoyama, and Şahin (2007), most research into uninsured idiosyncratic labor income risk builds on an individual sequential search problem or uses insights from such a model in reduced form. Recent work by Shimer and Werning (2008, cf. Chapter 7) presents analytical results on the optimal unemployment insurance in a partial equilibrium job search model with CARA preferences. In their model, individuals have unlimited access to liquidity and employment is an absorbing state. In this case, constant unemployment benefits and employment taxes are optimal in that they provide the unemployed individual with a given level of utility at the lowest possible cost. Using numerical methods to solve the model under CRRA preferences, they show that the cost of a constant benefit and tax policy relative to an optimal, time-varying policy is minuscule, so that the optimal policy obtained for CARA provides a good benchmark for CRRA as well. In what follows, we focus on the optimal behavior of the individual. Independently of Shimer and Werning's research, we advanced and clarified a similar analysis of optimal consumption in a standard Pissarides matching model due to Sennewald (2007b). In this model, individuals have CRRA preferences, switch back and forth between employment and unemployment according to a two-state Poisson process, and face interest rates that generally differ from their subjective discount rates. In the simplest version, we abstract from heterogenous job offers and thus ignore reservation wages. In an extension, we include job offers drawn from a distribution and derive the equilibrium reservation wage. The baseline model serves well as a starting point for the analysis of matching and savings for two reasons. First, it allows us to derive analytical expressions for the evolution of optimal consumption with CRRA preferences, where an individual's degree of risk aversion changes if her level of wealth changes, and account also for the level effect implied by consumption smoothing in cases where the interest rate differs from the subjective discount rate. In passing, we derive the law of motion of optimal consumption with CARA preferences as well, which will be useful in the following applications. Second, exploring individual consumption Euler equations, it is an easy task to assess the dynamics of consumption and wealth and to characterize the economic determinants of threshold wealth levels that were first uncovered in computable general equilibrium models (see, e.g., Heer and Maußner, 2005, Ch. 5).

In what follows, we describe the baseline model environment. It contains as a special case (with zero job separation, an interest rate equal to the discount rate, and a degenerated one-point distribution of wage offers) the unemployed's problem in Shimer-Werning (2008) under CRRA preferences. After solving the model, we characterize the dynamics of consumption and wealth. The appendix comprises several derivations and technical conditions for the model in the main text. It also extends the baseline model to include wage offers from a distribution and a wealth dependent reservation wage.

A cautionary note is in order. The following section is intended to introduce the basic modeling tools in continuous time matching with savings models and to present some immediate findings. The appendices to this chapter go somewhat beyond this goal and present material that is relevant for a reader interested in working with the models. We briefly repeat essential steps when needed, so that the main text is sufficient to prepare the grounds for the following applications.

6.2 A Baseline Model of Matching with Savings[8]

6.2.1 An Individual

Consider an individual who faces idiosyncratic, uninsured income risk from job separation and search frictions at the labor market. As in standard matching models in continuous time (cf. Pissarides, 2000), an individual working on a job loses her position with Poisson separation rate s and, when unemployed, finds a job with Poisson rate μ (a mnemonic for 'match'). Over the course of her lifetime, the individual switches back and forth between working on a job and being unemployed. The only uncertainty is about the duration of the employment and the unemployment spells. The wage rate is given by some constant w and constant unemployment benefits $b \leq w$ are paid as long as the individual is unemployed. Consumption is the numéraire. In a slight abuse of the word, we refer to $z(t) \in \{w, b\}$ as an individual's "labor" income irrespective of the actual employment state. We further economize on notation and denote by $z(t)$ equivalently an individual's employment state, i.e. $z(t) = w$ for 'employment' and $z(t) = b$ for 'unemployment'.

While there is no insurance against the labor income shocks, individuals can borrow and save in a riskless asset. Let a denote the stock of an individual's asset holdings. The individual's dynamic (or flow) budget constraint reads[9]

$$da(t) = \{ra(t) + z(t) - c(t)\}\, dt. \tag{6.1}$$

[8]Sections 6.2 and 6.4 are joint work with Klaus Wälde.

[9]With CRRA preferences, nonnegative consumption implies what Aiyagari (1994) calls the natural debt limit. That is, even if consumption was zero, the individual would not be able to repay more than a certain level of debt with probability one. The asset space is thus bounded below, cf. Section 6.3.

Per unit of time dt, financial wealth $a(t)$ increases (or decreases) if capital income $ra(t)$ plus labor income $z(t)$ is larger (or smaller) than consumption $c(t)$. Labor income $z(t)$ is given by w when employed and b when unemployed. All variables with a time argument can change over time, all others, like the interest rate r, are constant. Dividing the budget constraint by dt and using $\dot{a}(t) \equiv da(t)/dt$ would yield a more standard expression, $\dot{a}(t) = ra(t) + z(t) - c(t)$. As $a(t)$ is not differentiable with respect to time at moments where an individual jumps between employment and unemployment or vice versa, we prefer the above representation. The latter is also more consistent with the subsequent stochastic differential equations.

The dynamics of $z(t)$, the individual's employment status and labor income, can be described by a stochastic differential equation,

$$dz(t) = \Delta dq_\mu - \Delta dq_s, \qquad \Delta \equiv w - b. \tag{6.2}$$

The Poisson process q_s counts how often the individual moves from employment into unemployment. Its arrival rate is given by $s(z(t))$. The Poisson process related to finding a job is denoted by q_μ with arrival rate $\mu(z(t))$. It counts an individual's transitions out of unemployment, i.e. how often she finds a job.[10] As an individual cannot loose her job when she is unemployed and as searching for a job makes no sense for someone who has a job since all jobs are equally valuable, both arrival rates are state dependent (see table 6.2.1). If an individual is employed, for example, $\mu(w) = 0$, whereas when she is unemployed, $s(b) = 0$.

$z(t)$	w	b
$\mu(z(t))$	0	μ
$s(z(t))$	s	0

Table State dependent arrival rates.

Suppose the individual is employed: $z(t) = w$. The equation for the employment status then simplifies to $dz = -(w-b)\,dq_s$. Whenever the process q_s jumps, i.e. when the individual loses her job and $dq_s = 1$, the change in labor income is given by $-w + b$ and, given that the individual earns w before

[10] Alternatively, we could use a single process that jumps back and forth between two values indicating 'employment' and 'unemployment', respectively (cf. Bayer, 2007) instead of two alternating Poisson processes. Both formulations are equivalent in the present environment.

losing the job, earns $w - w + b = b$ afterwards. Similarly, when unemployed, the employment status follows $dz = (w - b)\, dq_\mu$ and finding a job, i.e. $dq_\mu = 1$, means that labor income raises from b to w. The description of the dynamics of the job status in (6.2) formalizes by the tool of a stochastic differential equation what has been used for a long time in many discrete-time models with two-state Markov processes.

The individual maximizes a standard expected intertemporal (present value) utility function,

$$U(t) = E_t \int_t^\infty e^{-\rho[\tau - t]} u(c(\tau))\, d\tau.$$

Expectations need to be formed due to the uncertainty of labor income which in turn makes consumption $c(\tau)$ uncertain. The planning horizon starts in t (as today) and is infinite. The time preference rate is $\rho > 0$ and instantaneous utility is of the CRRA type,

$$u(c(\tau)) = \frac{c(\tau)^{1-\sigma} - 1}{1 - \sigma}, \quad \sigma > 0. \tag{6.3}$$

We now ask how individuals behave optimally in such a framework.

6.2.2 Optimal Behavior

The Solution to the Maximization Problem

The individual maximizes her objective function by choosing a path $\{c(\tau)\}$ of consumption subject to the budget constraint (6.1) and the stochastic evolution of her employment status (6.2). We approach this problem by continuous-time stochastic dynamic programming. While this approach has a longstanding tradition in almost all fields of economics (see, e.g., Obstfeld, 1994, Turnovsky, 1997, Dinopoulos and Thompson, 1998, , Wälde, 1999, Stein, 2006, among many others), a rigorous proof for its applicability to standard economic problems was only recently provided by Sennewald (2007a). At least, this is true for the subset of problems where the instantaneous objective function is unbounded like in our case (see (6.3)).[11] Sennewald (2007a) proves that dynamic programming in fact yields the necessary and sufficient conditions for an optimum with this type of instantaneous objective function. To the best of our knowledge, however, there exists no rigorous proof of the differentiability of value functions in continuous-time stochastic dynamic programming when uncertainty stems from

[11]More precisely, Sennewald (2007a) proves that the stochastic dynamic programming approach can be used as a necessary criterion for optimality if the utility function and the coefficients are linearly bounded. He further proves sufficiency without any boundedness requirements.

Poisson processes.[12] Sennewald (2007a, p. 1126) notes this shortcoming: "Nevertheless, we required (...) the value function to be once continuously differentiable with bounded first derivatives. Relaxing these issues is left for further research." This lack of rigorous foundation is hardly believable given the long lasting and widespread use of continuous-time stochastic dynamic programming. Evidently, the issue is naturally resolved ex post in the case of problems that allow closed-form solutions (see, e.g., Dinopoulos and Thompson, 1998, or Wälde, 1999). As noted by Wälde (2008, p. 226), however, many standard models do not permit a closed form solution. Our general model shares this property. In what follows, we thus pursue a two-fold strategy. On the one hand, we take as given that the value function is twice continuously differentiable in the level of wealth whenever the employment state does not jump and proceed analytically. This approach offers important insights into the dynamics of optimal consumption and wealth that are in line with economic reasoning on consumption smoothing and precautionary savings. In view of closed-form solutions in special cases and rigorous proofs for analogous problems in discrete time, we conjecture that differentiability holds more generally as well.[13] On the other hand, we adopt a numerical solution and show qualitatively how the solution looks like. This strategy of course cannot replace a rigorous analytical treatment. Like our predecessors, we leave this important work for future research.

Returning to the model, an individual's state is described by her current labor income $z(t)$ and her current wealth $a(t)$. We thus define the value function as $V(a(t), z(t)) = \max_{\{c(\tau)\}} U(t)$ such that it solves the Hamilton-Jacobi-Bellman (HJB) equation (for the math on stochastic differential equations see Øksendahl, 2003; for the derivation of the HJB equation in the present setup cf. Sennewald, 2007a,b, Sennewald and Wälde, 2006, and the compendium in Wälde, 2008)

$$\rho V(z(t), a(t)) = \max_{c(t)} \left\{ u(c(\tau)) + \frac{1}{dt} E_t dV(z(t), a(t)) \right\}. \tag{6.4}$$

[12] See Stockey and Lucas (with Prescott) (1989) for a rigorous treatment in discrete time. In continuous-time under certainty (and also in discrete time under certainty), Beneviste and Scheinkmann (1979) prove the differentiability of the value function.

[13] It will become clear below, however, that the issue is not easily dismissable. For example, we find that consumption jumps at points in time where the individual changes her employment state, so that the instantaneous utility functional is not continuously differentiable in t. Note, however, that we only require differentiability of the value function with respect to the endogenous state variable at points in time where the Poisson process does not jump. Seierstad and Sydsæter (1987) provide theorems for the Maximum principle and sufficient conditions in problems where the optimizing decision involves jumps in the state variable (similar to the problem of an unemployed with endogenous reservation wage in Appendix 6.C below).

To compute the differential $dV\left(a\left(t\right),z\left(t\right)\right)$ we need the rule for differentiating functions of stochastic processes governed by Poisson processes, the Change of Variable Formula (CVF, see Wälde, 2008, Lemma 5). Let $F(t,Q)$ be a function of a stochastic process Q that evolves according to $dQ(t)=\gamma_1\left(\cdot\right)dt+\gamma_2\left(\cdot\right)dq\left(t\right)$, where $\gamma_1\left(\cdot\right)$ and $\gamma_2\left(\cdot\right)$ stand for the deterministic part and the stochastic part of the stochastic differential equation, respectively, and $dq\left(t\right)$ denotes the increment of the Poisson process (i.e., 1 with probability $\lambda_q dt$ and 0 with probability $1-\lambda_q dt$ where λ_q is the arrival rate of q). The CVF then says that the differential of F is given by the sum of the "deterministic derivatives" and heuristic terms that indicate the jumps of F due to jumps of the Poisson process q:

$$dF\left(t,Q\right)=F_t dt+F_Q\gamma_1\left(\cdot\right)dt+\left[F\left(t,Q+\gamma_2\left(\cdot\right)\right)-F\left(t,Q\right)\right]dq.$$

F_t and F_Q thereby denote the partial derivative of F with respect to t and Q, respectively.

Applying the CVF to $V\left(z,a\right)$, the differential $dV(z,a)$ is

$$dV(z,a)=V_a(z,a)da+\left[V(w,a)-V(b,a)\right]dq_\mu(z)+\left[V(b,a)-V(w,a)\right]dq_s(z),$$

where the change in the employment status is state dependent with $dq_\mu(z)=0$ in case of employment and $dq_\mu(z)=dq_\mu$ when the worker is unemployed. Similarly, $dq_s(z)=0$ in case of unemployment and $dq_s(z)=dq_s$ if the worker has a job.

Forming expectations, noting that $E_t(q_\lambda(\tau)-q_\lambda(t))=\lambda(\tau-t)$ for both Poisson arrival rates $\lambda\in\{s,\mu\}$, yields

$$E_t dV(z,a)=V_a(z,a)da+\mu(z)\left[V(w,a)-V(b,a)\right]+s(z)\left[V(b,a)-V(w,a)\right],$$

where again $s(z)=0$ if $z=u$ and $\mu(z)=0$ if $z=e$ (and s and μ otherwise, respectively).

After replacing the evolution of assets from (6.1) in $E_t dV(z,a)$, and inserting the resulting expression in (8.5), the HJB equation becomes

$$\begin{aligned}\rho V\left(z,a\right) &= \max_c\left\{u\left(c\right)+\left[ra+z-c\right]V_a\left(z,a\right)\right.\\ &\quad \left.+s\left(z\right)\left[V\left(b,a\right)-V\left(w,a\right)\right]+\mu\left(z\right)\left[V\left(w,a\right)-V\left(b,a\right)\right]\right\},\end{aligned} \tag{6.5}$$

where $V_a\left(z,a\right)$ denotes the partial derivative of V with respect to a. We suppressed the time arguments as this should not lead to confusion as of this point. Note that this Bellman equation holds in both employment states as the arrival rates are state dependent.

The first order condition for optimal consumption defines $c = c(z, a)$. It requires marginal utility of consumption to equal the shadow price of wealth,

$$u'(c(z,a)) = V_a(z,a). \tag{6.6}$$

We know from the budget constraint (6.1) that one unit of consumption costs one unit of wealth. Hence, in the optimum, the instantaneous increase in utility due to consuming marginally more is identical to the present value increase in overall utility due to an additional unit of wealth.

Consumption is a Normal Good

Normality of consumption is implied by the strict concavity of the value function of the unemployed (see, e.g. Sennewald, 2007b, Appendix A.3 to Chapter 3, or in a related setting with Brownian motion Chang, 2004, Section 4.3.1). To see this, differentiate the first order condition for consumption (6.6) with respect to wealth. Since the utility function (6.3) is strictly concave, we find that

$$\frac{\partial c(z,a)}{\partial a} = \frac{V_{aa}(z,a)}{u''(c(z,a))} > 0$$

if and only if $V(z,a)$ is strictly concave in a ($V_{aa}(z,a)$ thereby denotes the partial derivative of $V_a(z,a)$ with respect to a).[14] We are thus left to prove that $V(z,a)$ is in fact strictly concave in a, i.e. that $V(z,a_\lambda) > \lambda V(z,a_1) + (1-\lambda) V(z,a_2)$, where $a_\lambda \equiv \lambda a_1 + (1-\lambda) a_2$ and $\lambda \in [0,1]$. To this end, denote by $c_1(z,a_1)$ and $c_2(z,a_2)$ the optimal consumption levels associated with a_1 and a_2, respectively, and let $c_\lambda \equiv \lambda c_1 + (1-\lambda) c_2$. While not changing her employment status, the wealth levels evolve according to (6.1), i.e. $da_1/dt = \dot{a}_1 = ra_1 + z - c_1$ and $da_2/dt = \dot{a}_2 = ra_2 + z - c_2$, respectively. The consumption levels are thereby the optimal choices given a_1 and a_2. Note that the evolution of a_λ follows in fact from consumption level c_λ: $da_\lambda/dt = \dot{a}_\lambda = \lambda \dot{a}_1 + (1-\lambda)\dot{a}_2 = \lambda [ra_1 + z - c_1] + (1-\lambda)[ra_1 + z - c_1] = ra_\lambda + z - c_\lambda$.

Inserting the definition of the value function in the inequality above, consumption is normal if and only if

$$\begin{aligned} V(z,a_\lambda) > \lambda V(z,a_1) + (1-\lambda) V(z,a_2) \quad &= \\ E_t \int_t^\infty e^{-\rho(\tau-t)} \lambda u(c(z,a_1))\, d\tau \quad &+ \quad E_t \int_t^\infty e^{-\rho(\tau-t)} (1-\lambda) u(c(z,a_2))\, d\tau. \end{aligned}$$

[14]See Stockey and Lucas (with Prescott) (1989) Theorem 9.8 for the concavity of the value function in the time-discrete case.

Since $u''(c) < 0$, we know that $\lambda u(c_1) + (1-\lambda) u(c_2) < u(c_\lambda)$. Accordingly, we actually have

$$\begin{aligned} \lambda V(z, a_1) + (1-\lambda) V(z, a_2) &= \\ = E_t \int_t^\infty e^{-\rho(\tau - t)} \left[\lambda u(c_1) + (1-\lambda) u(c_2) \, d\tau\right] &< E_t \int_t^\infty e^{-\rho(\tau - t)} u(c_\lambda) \, d\tau \leq V(z, a_\lambda). \end{aligned}$$

Since c_λ is affordable and something else is better, we have shown

Corollary 6.1 ("normality", Sennewald, 2007b). *The consumption good is normal. The consumption function is strictly increasing in wealth,* $\frac{\partial c(z,a)}{\partial a} > 0$.

The Reduced Form

Let us collect the equations we need in order to be able to determine optimal behavior. We can insert the first order condition (6.6), keeping (6.3) in mind, into the Bellman equation (6.5) and take the state dependence into account. This gives (see Appendix 6.A)

$$\rho V(w, a) = \frac{\sigma}{1-\sigma} \left[(V_a(w, a))^{(\sigma-1)/\sigma} - \frac{1}{\sigma} \right] + [ra + w] V_a(w, a) + s [V(b, a) - V(w, a)], \tag{6.7}$$

$$\rho V(b, a) = \frac{\sigma}{1-\sigma} \left[(V_a(b, a))^{(\sigma-1)/\sigma} - \frac{1}{\sigma} \right] + [ra + b] V_a(b, a) + \mu [V(w, a) - V(b, a)], \tag{6.8}$$

i.e. the maximized Bellman equation for employed workers in (A.20) and for unemployed workers in (A.21). These two equations, formally speaking, are implicit differential equations (as one can not solve explicitly for the derivative $V_a(w, a)$ given the utility term $\sigma/(1-\sigma) V_a(b, a)^{(\sigma-1)/\sigma}$ and form a differential algebraic system. They are also non-autonomous due to the $ra + w$ and $ra + b$ terms. In principle, the two differential equations can be solved for value functions $V(b, a)$ and $V(w, a)$ which are sufficient for an optimum.

From the solutions to the value functions we can then, given the first order condition (6.6), compute consumption levels as a function of the state variables. This in turn allows us to compute the evolution of wealth of unemployed and employed workers using the budget constraint (6.1).

In Appendix 6.D.1 we show analytically that there is a simple solution for the special case where $w = b = 0$ (so that there is only one implicit differential equation and no $zV_a(z, a)$ term). It is given by $V(w, a) = V(b, a) = \kappa a^{1-\sigma}$ with κ some constant. We also prove that this solution is the boundary condition for the two differential equations in the general case with $w > b \geq 0$, i.e. $\lim_{a\to\infty} V(w, a) = \lim_{a\to\infty} V(b, a) = \kappa a^{1-\sigma}$ (see Appendix 6.D.1). Accordingly, from (6.6) and (6.3), consumption becomes approximately linear in wealth if wealth grows sufficiently large (cf. Ljungqvist and Sargent, 2004, Ch. 17.13).

As analytical solutions to the above system have not been found so far (the structure is related to the so-called Clairaut equation), we employ a numerical approximation to graphically demonstrate the ensuing policy functions below. In what follows, we obtain further insights from the evolution of optimal consumption over the (un)employment spell. Before continuing with the analysis of the optimal consumption behavior, we thus derive the consumption Euler equations in the two employment states.[15]

6.3 The Euler Equation in Continuous-Time Matching and Saving Problems

The first step towards the consumption Euler equation is to deduce the evolution of the costate variable (cf. Sennewald and Wälde, 2006, and Sennewald, 2007a and 2007b, Ch. 3 Appendix 2). The evolution of marginal utility then follows from the first order condition. Finally, we translate the evolution of marginal utility into the consumption Euler equation.

6.3.1 Deriving the Evolution of the Costate Variable

Differentiating the maximized HJB,

$$\begin{aligned}\rho V(z,a) &= u(c(z,a)) + V_a(z,a)\left[ra+z-c(z,a)\right]+\\ &\quad +\mu(z)\left[V(w,a)-V(b,a)\right]+s(z)\left[V(b,a)-V(w,a)\right],\end{aligned}$$

with respect to a yields

$$\begin{aligned}\rho V_a(z,a) &= u'(c(z,a))c_a(z,a) + V_{aa}(z,a)\left[ra+z-c(z,a)\right]+\left[r-c_a(z,a)\right]V_a(z,a)+\\ &\quad +\mu(z)\left[V_a(w,a)-V_a(b,a)\right]+s(z)\left[V_a(b,a)-V_a(w,a)\right].\end{aligned}$$

where $c_a(z,a)$ denotes the partial derivative of $c(z,a)$ with respect to a and $V_{aa}(z,a)$ stands for the partial derivative of $V_a(z,a)$ with respect to a. Using the first order condition (6.6), the change in consumption cancels (the "envelope theorem"):

$$\begin{aligned}\rho V_a(z,a) &= rV_a(z,a) + \left[ra+z-c(z,a)\right]V_{aa}(z,a)+\\ &\quad +\mu(z)\left[V_a(w,a)-V_a(b,a)\right]+s(z)\left[V_a(b,a)-V_a(w,a)\right].\end{aligned} \tag{6.9}$$

[15]Section 6.3 follows the derivation of the evolution of marginal utility and optimal consumption under CRRA in Sennewald (2007b).

An expression for $V_{aa}(z,a)$ is readily derived by applying the CVF to $V_a(z,a)$:

$$dV_a(z,a) = V_{aa}(z,a)da + \left[V_a(w,a) - V_a(b,a)\right] dq_\mu(z) + \left[V_a\left(b,a\right) - V_a(w,a)\right] dq_s(z).$$

Replacing the evolution of assets gives

$$\begin{aligned} dV_a(z,a) &= \left[ra + z - c(z,a)\right] V_{aa}(z,a)dt + \\ &\quad + \left[V_a(w,a) - V_a(b,a)\right] dq_\mu(z) + \left[V_a\left(b,a\right) - V_a(w,a)\right] dq_s(z). \end{aligned} \tag{6.10}$$

Using equation (6.10) to substitute for $\left[ra + z - c^z(a)\right] V^{z\prime\prime}(a)$ in (6.9), we find that $V^{z\prime}(a)$ evolves according to

$$\begin{aligned} dV_a(z,a) &= \left\{(\rho - r)V_a(z,a) - \mu(z)(V_a(w,a) - V_a(b,a)) - s(z)(V_a(b,a) - V_a(w,a))\right\} dt + \\ &\quad + \left[V_a(w,a) - V_a(b,a)\right] dq_\mu(z) + \left[V_a\left(b,a\right) - V_a(w,a)\right] dq_s(z). \end{aligned}$$

Note that we did not use the functional form of the utility function yet. To obtain the evolution of optimal consumption under fairly general preferences, we substitute $dV_a(z,a) = du'(c(z,a))$ from the first order condition (without using the CRRA specification). This gives the evolution of marginal utility:

$$\begin{aligned} \frac{du'(c(z,a))}{u'(c(z,a))} &= \left\{\rho - r - \left[\mu(z) - s(z)\right] \frac{u'(c(w,a)) - u'(c(b,a))}{u'(c(z,a))}\right\} dt + \\ &\quad + \frac{u'(c(w,a)) - u'(c(b,a))}{u'(c(z,a))} dq_\mu(z) + \frac{u'(c(b,a)) - u'(c(w,a))}{u'(c(z,a))} dq_s(z) \end{aligned} \tag{6.11}$$

6.3.2 Deriving the General Consumption Euler Equation

We can use the first order condition to translate the evolution of marginal utility into the evolution of consumption and derive a consumption Euler equation. To do so, define $f(\cdot) \equiv (u'(c))^{-1}$ (such that $f(u'(c)) = c$). Applying the CVF to f, we get

$$\begin{aligned} df((u'(c)) &= f'(u'(c))du'(c(z,a))dt + \\ &\quad + \left[f(u'(c\left(b,a\right)) - u'(c\left(w,a\right))\right] dq_\mu(z) + \left[f(u'(c\left(w,a\right)) - u'(c\left(b,a\right))\right] dq_s(z), \end{aligned}$$

where $du'(c(z,a))$ is given in (6.11). By construction, $df = dc$ and

$$f'(\cdot) = \frac{df}{du'(c(z,a))} u'(c(z,a)) = \frac{1}{u''(c(z,a)}.$$

Hence (6.11) becomes

$$\begin{aligned} dc(z,a) &= \left\{(\rho-r)\frac{u'(c(z,a))}{u''(c(z,a))} - \mu(z)\frac{(u'(c(w,a)) - u'(c(b,a))}{u''(c(z,a))} - s(z)\frac{u'(c(b,a)) - u'(c(w,a))}{u''(c(z,a))}\right\} dt + \\ &\quad + (c(w,a) - c(b,a))dq_\mu(z) + (c(b,a) - c(w,a))dq_s(z). \end{aligned}$$

Rearranging and taking the state-dependency into account, we finally find

$$-dc(b,a)\frac{u''(c(b,a))}{u'(c(b,a))} = \left\{r - \rho - \mu\left[1 - \frac{u'(c(w,a))}{u'(c(b,a))}\right]\right\} dt - \frac{u''(c(b,a))}{u'(c(b,a))}\left[c(w,a) - c(b,a)\right] dq_\mu, \quad (6.12)$$

$$-dc(w,a)\frac{u''(c(w,a))}{u'(c(w,a))} = \left\{r - \rho + s\left[\frac{u'(c(b,a))}{u'(c(w,a))} - 1\right]\right\} dt - \frac{u''(c(w,a))}{u'(c(w,a))}\left[c(b,a) - c(w,a)\right] dq_s. \quad (6.13)$$

As usual, the change of consumption is decreasing in the Arrow-Pratt measure of risk aversion $-u''(c)/u'(c)$.

The Consumption Euler Equation under CRRA Utility

From (6.3), $u''(c)/u'(c) = -\sigma/c$ and $u'(c^u)/u'(c^e) = (c^e/c^u)^\sigma$. Using these expressions, and letting $c^u = c(b,a)$ and $c^e = c(w,a)$, we get

$$\begin{aligned} \frac{dc^u}{c^u} &= \left\{\frac{r-\rho}{\sigma} - \frac{\mu}{\sigma}\left[1 - \left(\frac{c^u}{c^e}\right)^\sigma\right]\right\} dt + \left[\frac{c^u}{c^e} - 1\right] dq_\mu, \\ \frac{dc^e}{c^e} &= \left\{\frac{r-\rho}{\sigma} + \frac{s}{\sigma}\left[\left(\frac{c^e}{c^u}\right)^\sigma - 1\right]\right\} dt - \left[1 - \frac{c^e}{c^u}\right] dq_s. \end{aligned} \quad (6.14)$$

If we concentrate on the effects of income uncertainty and let $r = \rho$, consumption decreases (increases) deterministically for the unemployed (employed), and jumps to its new level whenever the employment status changes:

$$\begin{aligned} \frac{dc^u}{c^u} &= -\frac{\mu}{\sigma}\left[1 - \left(\frac{c^u}{c^e}\right)^\sigma\right] dt + \left[\frac{c^u}{c^e} - 1\right] dq_\mu, \\ \frac{dc^e}{c^e} &= \frac{s}{\sigma}\left[\left(\frac{c^e}{c^u}\right)^\sigma - 1\right] dt - \left[1 - \frac{c^e}{c^u}\right] dq_s. \end{aligned}$$

Note that, with CRRA preferences, $c^u = 0$ and $c^e = 0$ implies $dc^u = 0$ and $dc^e = 0$, respectively.

The Consumption Euler Equation under CARA Utility

Alternatively, suppose that preferences are of the CARA type,

$$u(c) = -e^{-\gamma c}. \quad (6.15)$$

In this case, $u''(c)/u'(c) = -\gamma$ is independent of c and $u'(c^u)/u'(c^e) = e^{-\gamma(c^u - c^e)}$. Accordingly, from (6.12) and (6.13),

$$\begin{aligned} dc^u &= \left\{ \frac{r-\rho}{\gamma} - \frac{\mu}{\gamma}\left[1 - e^{-\gamma(c^e - c^u)}\right] \right\} dt + (c^e - c^u) dq_\mu, \\ dc^e &= \left\{ \frac{r-\rho}{\gamma} + \frac{s}{\gamma}\left[e^{-\gamma(c^u - c^e)} - 1\right] \right\} dt - (c^e - c^u) dq_s. \end{aligned}$$

Letting again $r = \rho$, optimal consumption follows

$$\begin{aligned} dc^u &= -\frac{\mu}{\gamma}\left[1 - e^{-\gamma(c^e - c^u)}\right] dt + (c^e - c^u) dq_\mu, \\ dc^e &= \frac{s}{\gamma}\left[e^{-\gamma(c^u - c^e)} - 1\right] dt - (c^e - c^u) dq_s. \end{aligned}$$

We now return to the analysis of the optimal consumption in our baseline matching with savings model.

6.4 Consumption and Wealth Dynamics under CRRA

The last section provided the consumption Euler equation in the case of CRRA in the present environment where the labor income is uncertain due to matching frictions. We found that the consumption of an employed individual obeys

$$\frac{dc(w,a)}{c} = \left\{ \frac{r-\rho}{\sigma} + \frac{s}{\sigma}\left\{ \left[\frac{c(w,a)}{c(b,a)}\right]^\sigma - 1 \right\} \right\} dt - \left[1 - \frac{c(b,a)}{c(w,a)}\right] dq_s, \tag{6.16}$$

while her wealth evolves according to (6.1) with $z = w$, i.e. $da = [ra + w - c(w,a)]dt$. Analogously, the evolution of the optimal consumption of an unemployed individual is given by

$$\frac{dc(b,a)}{c} = \left\{ \frac{r-\rho}{\sigma} - \frac{\mu}{\sigma}\left\{ 1 - \left[\frac{c(b,a)}{c(w,a)}\right]^\sigma \right\} \right\} dt + \left[\frac{c(w,a)}{c(b,a)} - 1\right] dq_\mu, \tag{6.17}$$

and her wealth follows $da = [ra + b - c(b,a)]dt$.

Without uncertainty about future labor income, i.e. with $s = \mu = dq_s = dq_m = 0$, the above consumption Euler equations reduce to the classical consumption rule in deterministic models, $\dot{c}/c = (r-\rho)/\sigma$. The additional $s\{\cdot\}$ term in (6.16) suggests that, when exposed to the risk of losing her job, an employed individual will c.p. want to sacrifice more consumption today and accumulate additional wealth to "insure" herself against sudden drops in labor income.[16] Accumulating wealth allows her to smooth

[16] It is known from the partial equilibrium analysis in Leland (1968) and Sandmo (1970) that utility functions with convex marginal utility generate a positive precautionary demand for savings. The CRRA specification in (6.3) satisfies $u'''(c) > 0$. Cf. Kimball (1990) and Chapter 9.

consumption over the unemployment spell to some extent. Due to these precautionary savings, the risk of unemployment potentially increases the relative change in consumption of employed individuals. In the case of unemployment, as indicated by the $s\{\cdot\}$ term in (6.17), the possibility to find a new job induces unemployed individuals to expand their current consumption. Relative to a situation in which unemployment is an absorbing state (i.e., in the absence of job separation so that $s = 0$), the prospect of a higher labor income in the future reduces the individual's willingness to give up today's consumption. Precautionary savings are less important if the unemployed can expect to return to a higher labor income after having found a job. Relaxing the anxiety of having to rely on a low unemployment benefit thus allows a more optimistic spending and has the potential to reduce consumption growth for unemployed individuals.

The last terms in (6.16) and (6.17) (tautologically) represent the discrete jumps in the level of consumption whenever the employment status changes.

We are now in a position to explore the individuals savings decisions when her income is to some extent uncertain due to matching at the labor market.

6.4.1 Consumption and Wealth

Figure 6.1 shows the relationship between consumption and wealth of an employed and of an unemployed worker. This figure was obtained by numerically solving for the value functions $V(b,a)$ and $V(w,a)$ and then computing the optimal consumption levels $c(w,a)$ and $c(b,a)$ from the first order condition (6.6).[17] The following parameter values were used to generate Figure 6.1:

ρ	s	μ	r	σ	w	b
.05	.01	.09	.05	1.5	.5	$w/2$

The parameter values were chosen without a particular calibration goal. The separation and matching rate imply an unemployment rate equal to $s/(\mu+s) = 10\%$.

Financial wealth a is plotted on the horizontal, consumption levels $c(w,a)$ and $c(b,a)$ on the vertical axis. Not surprisingly, an employed worker consumes strictly more for a given wealth level than an

[17] We thank Burkhard Heer for numerically approximating the value functions and providing the following figures via Klaus Wälde. Some figures were modified to include threshold values for financial wealth and to indicate the dynamics of consumption and wealth, see below.

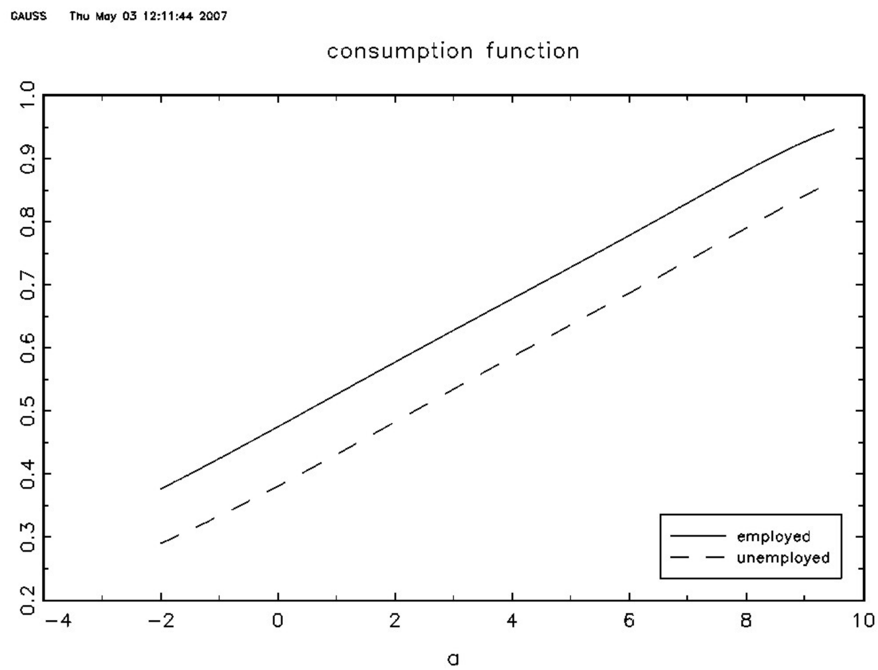


Figure 6.1: A Qualitative Example of the Consumption Function

unemployed worker,[18]

$$c\,(w,a) > c\,(b,a)\,, \quad \forall a. \tag{6.18}$$

Intuitively, consumption of the unemployed is lower at any given level of wealth since a currently employed worker expects a strictly higher lifetime income than a currently unemployed individual. One can also find (although hardly visible) that the relative consumption of employed to unemployed workers decreases in the level of wealth and approaches unity (see Appendix 6.D.1, equation (A.19) for a proof of the latter result),

$$\frac{d}{da}\frac{c\,(w,a)}{c\,(b,a)} < 0, \quad \lim_{a\to\infty}\frac{c\,(w,a)}{c\,(b,a)} = 1. \tag{6.19}$$

[18] Lentz and Tranæs (2005) provide a rigorous proof of this property for the time-discrete case. Let $f(z,a) \equiv ra + z - c(z,a)$. In the case where $f(w,a) \leq f(b,a)$, $w > b$ directly implies $c(w,a) > c(b,a)$. In the opposite case, Lentz and Tranæs (2005) use the contraction mapping property of the mapping defined by the system of maximized value functions for the two employment states to prove the existence of a fixed point $(V(w,a), V(b,a))$. It is then shown that this fixed point lies in the set of continuous functions that satisfy the analogue of $V_a(w,a) < V_a(b,a)$, so that, from the first-order condition, $c(w,a) > c(b,a)$ also in the case where $f(w,a) > f(b,a)$ (cf. Lentz and Tranæs, 2005, Lemma 3).

6.4.2 Dynamics

Let us understand the evolution of consumption in both employment states. As it turns out, saving and wealth dynamics crucially depend on the level of the interest rate relative to the subjective discount rate and the Poisson arrival rates. We therefore subdivide our discussion into three parts.

Very High Interest Rates ($\mathbf{r} \geq \boldsymbol{\rho} + \boldsymbol{\mu}$)

To begin with, consider an employed worker and her consumption Euler equation (6.16). Suppose she does not lose her job while we analyze her consumption behavior. Then, optimal consumption is given by

$$\sigma \frac{\dot{c}(w, a)}{c(w, a)} = r - \rho + s\left\{\left[\frac{c(w,a)}{c(b,a)}\right]^{\sigma} - 1\right\}. \tag{6.20}$$

From (6.18), $c(b, a)/c(w, a) < 1$ so that the term in curly brackets is positive. This gives us part (i) in the following

Proposition 6.1. *If $r > \rho + \mu$ (sufficient condition),*
(i) consumption of employed workers always increases through time.
(ii) consumption of unemployed workers also always increases through time.
For both groups, wealth rises as well.

When we look at an unemployed worker who does not find a job, her consumption follows

$$\sigma \frac{\dot{c}(b, a)}{c(b, a)} = r - \rho - \mu\left\{1 - \left[\frac{c(b,a)}{c(w,a)}\right]^{\sigma}\right\}. \tag{6.21}$$

In a situation with a high interest rate, i.e. for $r > \rho + \mu$, the right hand side is always positive and consumption always rises. Normality requires wealth to rise as well. This gives us the second part of Proposition 6.1.

We illustrate this proposition in Figure 6.2. Rising consumption levels are indicated by arrows on the consumption curves. With very high interest rates, an individual continuously increases her wealth and consumption over her lifetime. Consumption jumps at points in time when she changes her job situation: when finding a job, consumption jumps up, when losing a job, consumption jumps down. The opportunity costs of consumption and the desire to save for unemployment spells work in the same directions. During unemployment spells, the high interest rate boosts consumption growth by reducing the incentives to borrow against a higher expected future labor income. Since $da = ra + z - c(z, a) > 0$, we have that $c(z, a) < ra + z$ at all times.

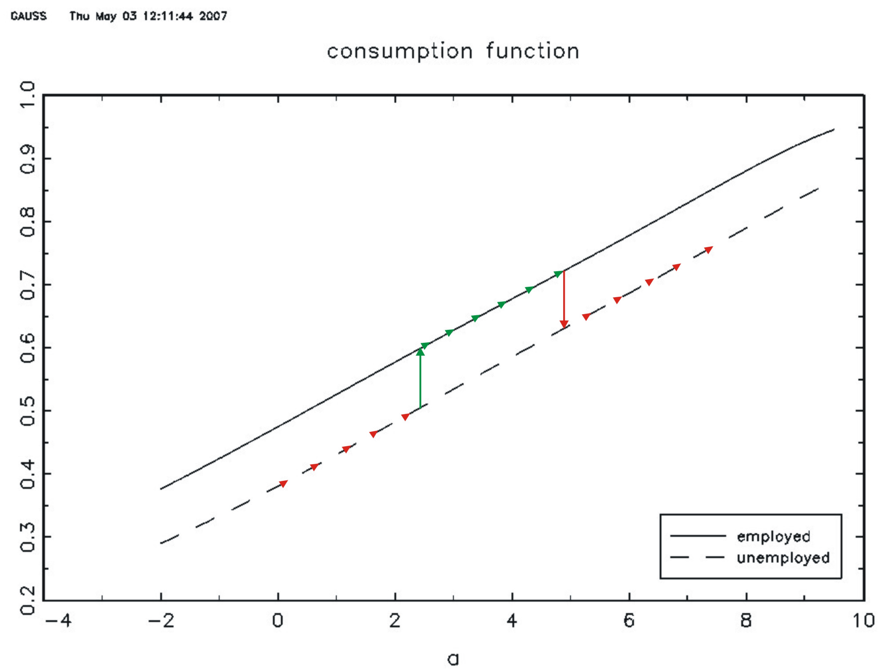


Figure 6.2: Consumption Dynamics in the Case of High Interest Rates (Qualitative Example)

Intermediate Level Interest Rates ($\boldsymbol{\rho < r < \rho + \mu}$)

Consider again an employed worker who keeps her job. As the right hand side of her deterministic consumption rule in (6.20) is positive, her consumption still increases unambiguously. She accumulates wealth at a positive rate (i.e., she consumes less than her contemporary income $ra + w$) as long as she keeps her job and is only forced to decrease her consumption level discretely when she becomes unemployed.

Let us now turn to an unemployed individual and suppose that she does not find a new job. For her, there typically exists a critical wealth level a_b^* implicitly defined by

$$\frac{c(b, a_b^*)}{c(w, a_b^*)} \equiv \left[1 - \frac{r - \rho}{\mu}\right]^{\frac{1}{\sigma}}, \tag{6.22}$$

above which consumption and wealth increase (see Appendix 6.B for a proof of the existence of a_b^* given the existence of an optimal control). In this case, $c\,(b, a) < ra + b$. If endowed with lower wealth, i.e. $a < a_b^*$, returns on assets and the unemployment benefit add up to an income level for which a continuously growing consumption is not optimal. Instead, she will eat up her savings by consuming more than her contemporary income and decrease her consumption over time. To see this formally,

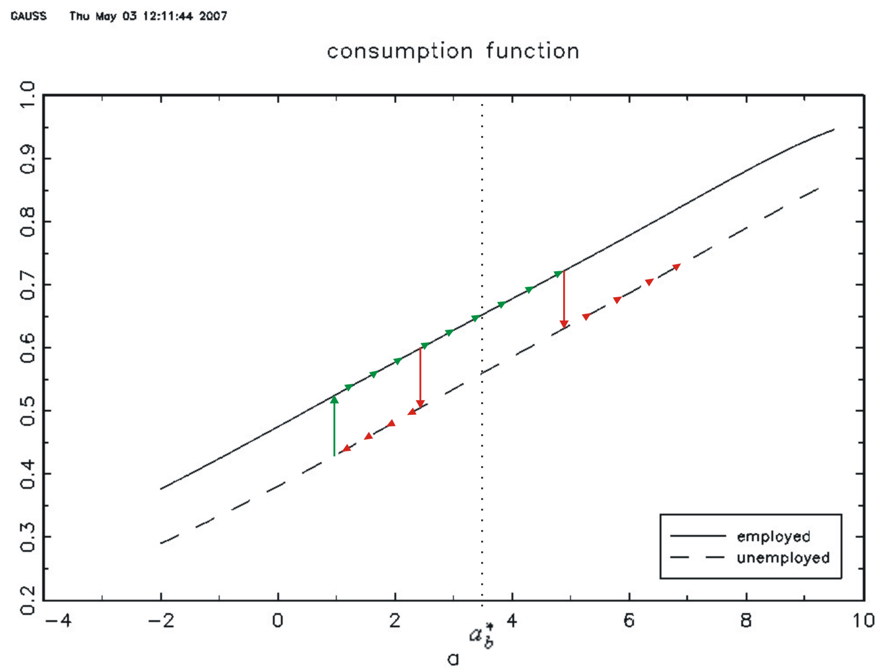


Figure 6.3: Consumption Dynamics in the Case of Intermediate Interest Rates (Qualitative Example)

consider the deterministic consumption rule for the unemployed in (6.21). The right hand side is positive if and only if $r - \rho > \mu[1 - [c(b,a)/c(w,a)^{\sigma}]$ or, using the definition of a_b^* in (6.22), if and only if $a \geq a_b^*$. Accordingly, sufficiently large levels of wealth ($a \geq a_b^*$) ensure a continuously increasing level of consumption as well as wealth. We qualitatively describe the resulting dynamics by arrows in Figure 6.3.

Next, consider an employed worker who has accumulated wealth $a < a_b^*$. If she loses her job, the level of consumption jumps downward and continues to decrease continuously along $c(b, a)$. Only finding a new job can reverse the decrease in consumption and wealth. Suppose she actually finds a job (so that her consumption jumps upward) and that she is lucky enough to stay employed long enough to accumulate wealth $a \geq a_b^*$. The next employment shock will again reduce her level of consumption discretely ($c(w, a)$ drops to $c(b, a)$), but this time unemployment will not reverse the growth of consumption and wealth: she will continue to accumulate assets, increase consumption, and finally reach the level of consumption she had before becoming unemployed. Put differently, all individuals with wealth $a \geq a_b^*$ optimally choose consumption levels above $c(b, a_b^*)$. More poorly endowed unemployed individuals are forced to optimally decrease their consumption and run down their assets further, potentially incurring debt. We show below that consumption and wealth decline and consumption converges towards zero, implying an endogenous lower bound on wealth. We have thus established

Proposition 6.2. *If $\rho < r < \rho + \mu$ (sufficient condition),*
(i) consumption and wealth of employed workers always increases.
(ii) consumption and wealth of unemployed workers increases if $a \geq a_b^*$ where a_b^* is defined in (6.22) and decreases otherwise.

Low Level of the Interest Rate ($\mathbf{r} \leq \boldsymbol{\rho}$)

Finally, consider low levels of the interest rate where $r \leq \rho$. In deterministic settings, such an interest rate typically leads to a decrease in the level of consumption as the individual does not receive a sufficiently high compensation for her impatience. This classical result must hold true for unemployed workers in this model as well since the "optimistic spending" effect further exaggerates the usual dissaving incentives of a low interest rate (compare the previous paragraph). Consumption exceeds the contemporary income. Formally, inspection of the right hand side of (6.21) shows that unemployed workers indeed always reduce wealth and consumption if $r \leq \rho$.

For employed workers, the "precautionary savings" motive counteracts the dissaving incentives from the low interest rate (i.e., consumption smoothing).[19] Since the threat of sudden drops in consumption becomes less severe the higher the agents' wealth, only employed workers with a low level of accumulated savings will continue to save and build up assets. Reducing "today's" level of consumption then implies positive consumption growth if "tomorrow's" consumption is not decreased by more than current consumption. Since the "precautionary savings" motive becomes weaker as wealth increases, "tomorrow's" decrease is lower than "today's" decrease as long as there are positive savings. Accordingly, consumption and wealth go up for less endowed working individuals, see Figure 6.4.
A formal argument is readily derived from (6.20). When $r \leq \rho$, the right hand side can only be positive if $c(w,a)/c(b,a)$ is sufficiently large or, more precisely, if and only if $a \leq a_w^*$ where a_w^* is implicitly defined by

$$\frac{c(w,a_w^*)}{c(b,a_w^*)} \equiv \left[1 - \frac{r-\rho}{s}\right]^{\frac{1}{\sigma}}. \tag{6.23}$$

[19]Cf. Chapter 9.

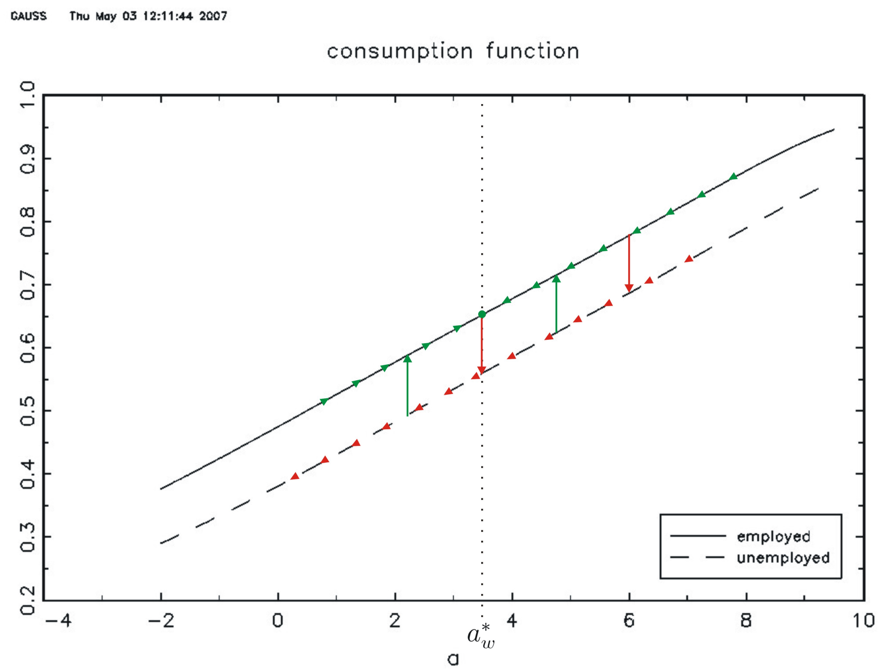


Figure 6.4: Consumption Dynamics in the Case of Low Interest Rates (Qualitative Example)

See Appendix 6.B for a proof of the existence of a_w^* given the existence of an optimal control. The threshold again separates wealth levels that imply positive savings ($c(w, a) < ra + w$) from those that imply dissaving ($c(w, a) > ra + w$). We summarize the above results in

Proposition 6.3. *If $r \leq \rho$ (sufficient condition),*
(i) consumption and wealth of unemployed workers always decrease.
(ii) consumption and wealth of employed workers increase if $a \leq a_w^*$ where a_w^* is defined in (6.22) and decrease otherwise.

6.5 Conclusion

We have studied a baseline model of labor market matching with savings. In the model, individuals move back and forth between employment and unemployment and thus face idiosyncratic labor income shocks. Under CRRA preferences, individuals have a desire to build up precautionary savings while working on a job and like to spend more during the unemployment spell when having the prospect of finding a job. Consumption smoothing may either work in the same direction or run counter the precautionary savings motive (we further study this feature in a two period model in Chapter 9).

Depending on the constellation of employment transition parameters, the interest-, and the discount rate, we encountered threshold wealth levels in the dynamics of consumption and wealth. If the interest rate is in an intermediate range, poor unemployed individuals may find it hard to escape wealth levels below the threshold. If the interest rate is low, consumption of the employed will converge to a stationary target level. Sufficiently well endowed, employed individuals decrease consumption and eat up their wealth, while less endowed, employed individuals save and increase their consumption (until they reach the stationary level of consumption and wealth). Unemployed individuals, however, always dissave.

Appendix

The following appendices contain several derivations and technical conditions for the model in the main text, proves the existence and uniqueness of the threshold levels, and introduces reservation wages.

Appendix 6.A Derivation of the System of Partial Differential Equations for $\mathbf{V}(\mathbf{z},\mathbf{a})$

Inserting the first order condition for optimal consumption (6.6) in (6.5) gives the maximized HJB equation

$$\rho V(z,a) = u(c(z,a)) + [ra + z - c(z,a)]\, V_a(z,a) + s(z)\,[V(b,a) - V(w,a)] + \mu(z)\,[V(w,a) - V(b,a)]\,.$$

After substituting for $c(z,a)$ with $V_a(z,a)^{-1/\sigma}$ from (6.6) and

$$u(c(z,a)) = \frac{c(z,a)^{-\sigma}\, c(z,a) - 1}{1-\sigma} = \frac{V_a(z,a)^{\frac{\sigma-1}{\sigma}} - 1}{1-\sigma} \tag{A.1}$$

from (6.6) and the utility functional (6.3) and taking the state dependency into account, we get

$$\begin{aligned}
\rho V(w,a) &= \frac{V_a(w,a)^{\frac{\sigma-1}{\sigma}}}{1-\sigma} - \frac{1}{1-\sigma} + \left[ra + w - V_a(w,a)^{-\frac{1}{\sigma}}\right] V_a(w,a) + s\,[V(b,a) - V(w,a)]\,,\\
\rho V(b,a) &= \frac{V_a(b,a)^{\frac{\sigma-1}{\sigma}}}{1-\sigma} - \frac{1}{1-\sigma} + \left[ra + b - V_a(b,a)^{-\frac{1}{\sigma}}\right] V_a(b,a) + \mu\,[V(w,a) - V(b,a)]\,.
\end{aligned}$$

Rearranging and collecting terms yields the system in the main text. Note that the notation $V(z,a)$ indicates different value functions since, in general, not only $w \neq b$ but also $s \neq \mu$.

Appendix 6.B Existence and Uniqueness of the Wealth Thresholds

We aim at providing conditions under which the threshold wealth levels a_b^* and a_w^* exist (taking as given the existence of an optimal control). To begin with, consider the unemployed in the case of intermediate level interest rates. By definition of $c(a_b^*, b)$, we have that $c(a_b^*, b) = ra_b^* + b$ as $\dot{c}(a_b^*) = 0$ and hence $da_b^* = 0$ from normality of the consumption good. Using this expression in (6.22) and rearranging gives

$$\frac{ra_b^* + b}{\left[1 - \frac{r-\rho}{\mu}\right]^{\frac{1}{\sigma}}} = c(w, a_b^*)\,.$$

The left hand side is zero at $a_b^* = -r/b$ and, in the case of intermediate level interest rates, increases with slope $r/\left(1-\frac{r-\rho}{\mu}\right)^{\frac{1}{\sigma}} > r$. The right hand side is smaller than ra_b^*+w as the unemployed individual chooses $da < 0$ at all levels of wealth in this intermediate case. It is, however larger than $ra_b^* + b$ as $c(w,a) > c(b,a)$. By normality, $c(w,a_b^*)$ is monotonically increasing in a_b^*. Hence, a_b^* exists uniquely without further parameter restrictions.

Consider next the employed individual in the case of low interest rates. For her, by the same argument as above, we have $c(a_w^*, w) = ra_w^* + w$. Inserting this expression in (6.23) and rearranging, we get

$$\frac{ra_w^* + w}{\left[1-\frac{r-\rho}{s}\right]^{\frac{1}{\sigma}}} = c(b, a_w^*).$$

The left hand is zero at $a_w^* = -w/r$ and increases with slope $r/\left[1-\frac{r-\rho}{s}\right]^{\frac{1}{\sigma}} < r$ since $r < \rho$. At the same time, $c(b,a_w^*) > ra_w^* + b$ as wealth declines for the unemployed individual. Again, consumption is normal so $c(b,a_w^*)$ increases monotonically in a_w^*. Accordingly, a unique intersection exists without further restrictions.

Appendix 6.C Endogenous Reservation Wages

In this appendix, we show how a distribution of wage offers affects the unemployed individual's optimal behavior. An immediate implication of including a wage offer distribution is that the unemployed individual will only accept wages above some threshold, i.e. a reservation wage exists.[20] In what follows, we derive the HJB equation including a wage offer distribution and an additional equation that characterizes the optimal reservation wage.

6.C.1 Derivation of the HJB Equation with Reservation Wage

Suppose the wage offers are sampled from a continuous distribution $F(w)$ with finite expectation and $F(w) < 1$ for some $w > 0$. The draw of the wage is assumed to be independent of the stochastic process that governs the transition between the two employment states. "Search on the job" is ruled out by assumption (cf. the state dependent arrival rates in Section 6.2.1). The unemployed individual thus faces a random sequence of independently and identically distributed wage offers. Put differently, we consider a Poisson differential equation with random amplitudes of the upward jumps of z. Let $\bar{w}$

[20]For an introduction to reservation wages, see Ljungqvist and Sargent (2004, Ch. 5) in discrete time or Mortensen (1986, Ch. 15), see further Chapters 7 and 8.

denote the individual's reservation wage, i.e., the lowest wage offer accepted by the individual. Such a reservation value rule is well-known from the analysis of an optimal stopping rule in sequential search models with costly wage draws (see, e.g. DeGroot, 2004). By definition, the reservation wage satisfies $V(b,a) = V(\bar{w},a)$, so that it is potentially a function of wealth, $\bar{w} = \bar{w}(a)$. As will become clear in Chapter 8 below, the reservation wage need not always depend on an individual's level of wealth.[21]

Since the stochastic processes underlying wage offers and employment transitions are independent, the arrival rate of a job for an unemployed conditional on $w \geq \bar{w}(a)$ is given by $\tilde{\mu}(a) \equiv \mu[1 - F(\bar{w}(a))]$. Hence, z is now governed by

$$dz = (w-b)dq_{\tilde{\mu}(a)} + (b-w)dq_s. \tag{A.2}$$

where $dq_{\tilde{\mu}(a)}$ is the increment of the Poisson process resulting from $\tilde{\mu}(a)$ and $z \in \{b, w\}$, $w \in [\bar{w}, \infty]$. Wealth still evolves according to (6.1), i.e. $da = ra + z - c(z,a)$, where $z = w$ in the case of employment is now the realization of the wage draw conditional on $w \geq \bar{w}(a)$. In addition to the uncertain duration of the employment spells, an unemployed individual now also faces uncertainty about the future wage income.

Let $V(z,a)$ again denote the unique solution to this problem's Bellman equation. In principle, $V(z,a)$ looks just like the value function in (6.4), except that we need to take account of the fact that the individual only accepts jobs that make her better off relative to remaining unemployed. After all, working on a poorly paid job forecloses the possibility of finding a better job. To compute the differential $dV(a(t), z(t))$, we first calculate the differential for a given wage draw, $d\bar{V}$, and then form expectations over w and the evolution of the employment state to obtain dV. Using the CVF and taking the constraints (6.1) and (A.2) into account gives

$$d\bar{V}(z,a) = V_a[ra + z - c]dt + \max[0, V(w,a) - V(z,a)]dq_\mu + [V(b,a) - V(w,a)]dq_s.$$

Taking as given that $V(\bar{w}(a), a) = V(b,a)$ yields a unique reservation wage $\bar{w}(a)$ and using that

[21] We show in Chapter 8 that if preferences are CARA, consumption is allowed to be real-valued, debt is unbounded, and employment is an absorbing state (i.e. $s = 0$), the reservation wage is in fact independent of the unemployed individual's wealth. Chapter 8 demonstrates that including a borrowing constraint in this environment introduces reservation wages that are increasing in wealth.

$V(z,a)$ is increasing in z, we can equivalently write[22]

$$d\bar{V}(z,a) = V_a[ra+z-c]dt + [\bar{V}(\tilde{w},a) - V(b,a)]dq_\mu + [V(b,a) - V(w,a)]dq_s,$$

where $\bar{V}(w,a)$ stands for $V(w,a)$ if $w \geq \bar{w}(a)$ and $V(b,a)$ otherwise. For $z=b$ we have

$$d\bar{V}(b,a) = V_a[ra+b-c]dt + [\bar{V}(\tilde{w},a) - V(b,a)]dq_\mu.$$

The first term on the right hand side contains only the state and prices, which are known in t. Forming expectations about both wages and states yields

$$\begin{aligned}
&\int_0^\infty \int_0^\infty d\bar{V}(b,a)dG(dq)dF(w) = \\
&= V_a(ra+b-c)dt + \int_0^\infty \left[\bar{V}(w,a) - V(b,a)\right] \int_0^\infty dq_\mu dG(dq_\mu)dF(w) = \\
&= V_a(ra+b-c)dt + \mu \int_0^\infty \left[\bar{V}(w,a) - V(b,a)\right] dF(w)dt = \\
&= V_a(ra+b-c)dt + \mu\left[1 - F(\bar{w}(a))\right] \left[\int_{\bar{w}(a)}^\infty V(w,a)\frac{dF(w)}{1-F(\bar{w})} - V(b,a)\right] dt. \qquad \text{(A.3)}
\end{aligned}$$

The expected change of $V(b,a)$ has a deterministic and a stochastic component. First, wealth a evolves deterministically according to (6.1) as long as the unemployed individual does not accept a job. This change itself is known in t. Second, the labor income jumps upward at the point in time where the unemployed individual accepts a job. Her expected wage is drawn from the truncated wage distribution with support $[\bar{w}(a), \infty)$, i.e. from the cumulative distribution function $F(w)/[1-F(\bar{w}(a))]$.
Using $dV(b,a)$ and making explicit the choice between remaining unemployed and accepting a job, an unemployed's situation may be expressed via the following partial differential equation, denoting by $\breve{w}$ some minimum wage:

$$\begin{aligned}
\rho V(b,a) = \max \Big\{ &\rho V(w,a), \max_c \langle u(c) + [ra+b-c] V_a(b,a) + \\
&+\mu \left[\int_{\breve{w}}^\infty V(w,a) - V(b,a) dF(w)\right] \rangle \Big\}. \qquad \text{(A.4)}
\end{aligned}$$

If an unemployed accepts a wage offer, her "overall happiness" is given by the value function $V(w,a)$ where w is the offered wage (the first part of the above Bellman equation). If she rejects the job, she

[22] Note that the indirect marginal effect of wealth on V via the reservation wage is zero since, by definition of the reservation wage and the independency of the stochastic processes for employment transition and the wage draw, the change in $[\bar{V}(\tilde{w},a) - V(b,a)]dq_\mu$ due to change of wealth via the reservation wage is zero, since $[V(a,\bar{w}) - V(a,b)] = 0$, cf. Chapter 8.

finds herself in the same situation as if no job had been offered at all (this is the second part of this Bellman equation). The reservation wage $\bar{w}(a)$ is the wage offer that makes the individual indifferent between continuing to search and accepting the job. If $\bar{w}(a)$ is chosen, the Bellman equation in (A.4) can simply be written as

$$\rho V(b,a) = \max_c \left\{ u(c) + [ra + b - c] V_a(b,a) + \mu \int_{\bar{w}(a)}^{\infty} [V(w,a) - V(b,a)] \, dF(w) \right\} \tag{A.5}$$

The first order condition for optimal consumption thus again equates the shadow price of wealth to the marginal utility of consumption: $u'(c(b,a)) = V_a(b,a)$.

Now, insert the first order conditions in both states into the Bellman equations and equate the maximized Bellman equations in both employment states. This gives an expression for the reservation wage in terms of optimal consumption:

$$\begin{aligned} &u(c(a,\bar{w}(a))) + [ra + \bar{w}(a) - c(a,\bar{w}(a))] V_a(\bar{w}(a),a) + s[V(b,a) - V(\bar{w}(a),a)] = \\ &\qquad u(c(a,b)) + [ra + b - c(a,b)] V_a(b,a) + \mu \left[\int_{\bar{w}(a)}^{\infty} V(w,a) - V(b,a) \, dF(w) \right]. \end{aligned} \tag{A.6}$$

6.C.2 The Reduced Form

After substituting for $u(c(z,a))$ with (A.1), we obtain an equation for the reservation wage in terms of $V(z,a)$ and $V_a(z,a)$:

$$\begin{aligned} &\frac{\sigma}{1-\sigma} V_a(\bar{w}(a),a)^{(\sigma-1)/\sigma} + [ra + \bar{w}(a)] V_a(\bar{w}(a),a) + s[V(b,a) - V(\bar{w}(a),a)] = \\ &\frac{\sigma}{1-\sigma} V_a(b,a)^{(\sigma-1)/\sigma} + [ra + b] V_a(b,a) + \mu [1 - F(\bar{w}(a))] \left[\int_{\bar{w}(a)}^{\infty} V(w,a) \frac{dF(w)}{1-F(w)} - V(b,a) \right] \end{aligned} \tag{A.7}$$

By means of the identity $V(\bar{w}(a),a) = V(b,a)$, we further have

$$V_a(\bar{w}(a),a) \left[1 + \frac{\partial \bar{w}(a)}{\partial a} \right] = V_a(b,a).$$

After substituting for $V_a(b,a)$ on the left hand side, using $V(b,a) = V(\bar{w}(a),a)$, and rearranging, equation (A.7) simplifies to

$$\begin{aligned} &\frac{\sigma}{\sigma-1} \left[V_a(\bar{w}(a)) \frac{\partial \bar{w}(a)}{\partial a} \right]^{\frac{\sigma-1}{\sigma}} + (ra - b) V_a(\bar{w}(a),a) \frac{\partial \bar{w}(a)}{\partial a} + (\bar{w}(a) - b) V_a(\bar{w}(a),a) = \\ &\mu \int_{\bar{w}(a)}^{\infty} [V(w,a) - V(b,a)] \, dF(w). \end{aligned} \tag{A.8}$$

In the example in Chapter 7 below, where the reservation wage is independent of wealth, the reservation wage is determined by (letting $\bar{w}'(a) = 0$)

$$(\bar{w} - b)V_a(\bar{w}, a) = \mu \int_{\bar{w}}^{\infty} [V(w, a) - V(b, a)] \, dF(w). \tag{A.9}$$

Equation (A.8) is an additional equation to the system of partial differential equations for $V(w, a)$ and $V(b, a)$ considered in the main text, augmented to include the expected wage in the case of employment where w is drawn from the cumulative distribution function $F(w)/[1 - F(\bar{w})]$. A solution to the augmented system of three equations in three unknowns provides us with $V(z, a)$, $V_a(z, a)$, and $\bar{w}(a)$ (see Chapter 8 for an example). Using this solution, the policy functions for consumption follow from the first order conditions above.

Appendix 6.D Technical Appendix

6.D.1 On the Boundary Conditions of the General Model

In this final technical appendix, we derive two boundary conditions for the system of differential equations in the value functions as the level of wealth grows large.

In the main text, we derived the reduced form of the model from the maximized Bellman equation using the first order condition to replace consumption with $V_a(z, a)$ with the help of the specified CRRA utility function $u(c) = c^{1-\sigma}/(1 - \sigma) - 1/(1 - \sigma)$. Including the constant $-1/(1 - \sigma)$ is usually convenient because it implies that $u(c)$ approaches the log-utility function as $\sigma \to 1$ (this can easily be verified using l'Hôpital's rule). It turns out that omitting the term $-1/(1 - \sigma)$ allows us to derive solutions to the system of partial differential equations for the value functions in special cases. This is helpful because it allows us to check the validity of numerical solutions for parameter values close to the special cases. In what follows, we therefore replace (6.3) with

$$u(c(\tau)) = \frac{c(\tau)^{1-\sigma}}{1 - \sigma}, \quad \sigma > 0, \quad \sigma \neq 1. \tag{A.10}$$

Now, instantaneous utility no longer converges to $\log(c)$ as $\sigma \to 1$, but the elasticity of marginal utility remains $u''(c)c/u'(c) = -\sigma$ and the individual's consumption choice evidently remains unaffected by this transformation of the utility function since, as usual, we rely on a limited form of cardinal utility (see Koopmans, 1965).

With this change, the constant drops out from the system of differential equations: [23]

$$\rho V(w,a) = \frac{\sigma}{1-\sigma} V_a(w,a)^{\frac{\sigma-1}{\sigma}} + (ra+w)\, V_a(w,a) + s\,(V(b,a) - V(w,a))\,, \tag{A.11}$$

$$\rho V(b,a) = \frac{\sigma}{1-\sigma} V_a(b,a)^{\frac{\sigma-1}{\sigma}} + (ra+b)\, V_a(b,a) + \mu\,(V(w,a) - V(b,a))\,. \tag{A.12}$$

In what follows, we first describe a solution to the special case where $w = b = 0$ (so that there is no income uncertainty and no $zV_a(z,a)$ term in the remaining differential equation) and then use this information to derive a boundary condition for large levels of wealth. Intuitively, as an individual becomes richer and richer, optimal consumption should depend less and less on current labor income.

A Special Case: $\mathbf{w} = \mathbf{b} = \mathbf{0}$

Consider an unemployed individual that holds the same level of wealth as another, employed, individual. If the unemployment benefit and the wage rate are equal ($w = b$), there is no economic reason for the maximized value functions of the two individuals to differ. The only difference is due to heterogeneous initial employment states, but after all, the labor income is all that matters for the individual's lifetime utility and it is the same in both employment states. In this case, the system of equations (A.11) and (A.12) boils down to a single equation ($V(w,a) = V(b,a)$),

$$\rho V(w,a) = \frac{\sigma}{1-\sigma} V_a(w,a)^{\frac{\sigma-1}{\sigma}} + (ra+w)\, V_a(w,a). \tag{A.13}$$

If, in addition, the wage rate is zero, we arrive at an implicit, non-autonomous differential equation that is amenable to an analytical solution:

$$\rho V(0,a) = \frac{\sigma}{1-\sigma} V_a(0,a)^{\frac{\sigma-1}{\sigma}} + raV_a(0,a). \tag{A.14}$$

A solution to (A.14) is

$$V(0,a) = a^{1-\sigma}\kappa, \quad \kappa = \frac{\sigma^\sigma}{(1-\sigma)(\rho - r(1-\sigma))^\sigma}. \tag{A.15}$$

Proof. The solution can easily be verified by plugging (A.15) in the right hand side of (A.14). From the solution in (A.15) we get

$$V(0,a)^{\frac{\sigma-1}{\sigma}} = a^{1-\sigma}\left(\frac{\sigma}{\rho - r(1-\sigma)}\right)^{\sigma-1} = \frac{1-\sigma}{\sigma} a^{1-\sigma}\kappa\,(\rho - r(1-\sigma))\,, \tag{A.16}$$

[23] From (A.10) and the first order condition (6.6) it holds that $u(c) = 1/(1-\sigma)V_a^{(\sigma-1)/\sigma}$ and $cV_a = V_a^{(\sigma-1)/\sigma}$. Plugging these expressions into the Bellman equation (6.5) separately for each state gives the system under modified utility functional (A.10) .

and

$$aV_a(0,a) = a^{1-\sigma}\left(\frac{\sigma}{\rho - r(1-\sigma)}\right)^{\sigma} = a^{1-\sigma}\kappa\,(1-\sigma)\,. \tag{A.17}$$

Using both expressions in (A.14) yields $a^{1-\sigma}\kappa\rho = \rho V(0,a)$. □

For κ not to become a complex number, impose $\rho > (1-\sigma)r$.

Taking the Limit $\mathbf{a} \to \infty$

Ultimately, we are interested in the characteristics of the solution if wages exceed positive unemployment benefits. The special case where $w = b = 0$, however, has an important implication for the general case where $w \geq b \geq 0$. This is because the effect of current labor income on the individual's consumption choice vanishes if wealth holdings grow sufficiently large. In the limit as a goes to infinity, the impact of z on the value function vanishes.

These conjectures are validated as follows. Suppose the solution to the special case where $w = b = 0$, $V(z,a) = a^{1-\sigma}\kappa$, also solves the general system. Using $V(z,a) = a^{1-\sigma}\kappa$ in equations (A.11) and (A.12) gives[24]

$$\rho a^{1-\sigma}\kappa = \frac{\sigma}{1-\sigma}\left((1-\sigma)\,a^{-\sigma}\kappa\right)^{\frac{\sigma-1}{\sigma}} + (ra+z)\,(1-\sigma)\,a^{-\sigma}\kappa,$$

where z equals w for (A.11) and b for (A.12). Dividing by a gives

$$\begin{aligned}
\rho a^{-\sigma}\kappa &= \frac{\sigma}{1-\sigma}\left((1-\sigma)\,\kappa\right)^{\frac{\sigma-1}{\sigma}} a^{-\sigma} + \left(r + \frac{z}{a}\right)(1-\sigma)\,a^{-\sigma}\kappa \Leftrightarrow \\
\rho - \left(r + \frac{z}{a}\right)(1-\sigma) &= \sigma\,(1-\sigma)^{-\frac{1}{\sigma}}\,\kappa^{-\frac{1}{\sigma}}.
\end{aligned}$$

As $a \to \infty$, there is a constant κ which solves this equation. In fact, it is the same κ as in (A.15). The true solution runs below or above $a^{1-\sigma}\kappa$, depending on wether $\sigma < 1$ or $\sigma > 1$, respectively. For any given level of wealth, the difference between the true value function and $a^{1-\sigma}\kappa$ is smaller, the closer σ is to unity (from either side) and the smaller the wage rate.

We have thus found two "boundary conditions" for the general system (A.11) and (A.12): $V(z,a)$ converges to $a^{1-\sigma}\kappa$ as a grows large. Taking the limit of $V(z,a) = a^{1-\sigma}\kappa$ as $a \to \infty$, we further find that

$$\lim_{a\to\infty} V(z,a)\,|_{\sigma>1} = 0, \quad \lim_{a\to\infty} V(z,a)\,|_{\sigma<1} = +\infty, \tag{A.18}$$

[24] To calculate the right hand sides, we can directly use the expressions in equations (A.16) and (A.17) for both $z = w$ and $z = b$, $V(z,a)^{\frac{\sigma-1}{\sigma}} = \frac{1-\sigma}{\sigma}a^{1-\sigma}\kappa\,(\rho - r(1-\sigma))$ and $aV_a(z,a) = a^{1-\sigma}\kappa\,(1-\sigma)$. Taken together, the right hand sides obey $a^{1-\sigma}\kappa(\rho + (1-\sigma)z/a)$, while the left hand sides read $a^{1-\sigma}\kappa\rho$. Dividing both sides by $a^{1-\sigma}\kappa$ and taking the limit $a \to \infty$ verifies that the solution to the general system converges to $a^{1-\sigma}\kappa$ as a becomes "large" ($\sigma \neq 1$).

i.e. the value functions approach zero or plus infinity, depending on the size of σ. We also obtain $\lim_{a\to\infty} V_a(z,a) = 0$. Using the first order condition (6.6) to relate $V_a(z,a)$ to optimal consumption, this proves equation (6.19) in the main text, namely that

$$\lim_{a\to\infty} \frac{c(w,a)}{c(b,a)} = 1. \tag{A.19}$$

6.D.2 No "Typical" Closed-Form Solution in the Baseline Model

A solution to the individual's optimal consumption problem is a set of value functions $V(z,a)$ that satisfy the reduced system, (A.11) and (A.12), the first order conditions, and the boundary conditions. The following guess and verify approach shows the difficulty of the problem at hand. Suppose the solution is of the "typical" form: $c_g(w,a) = \gamma_0 a + \gamma_1$, $V_g(w,a)(a) = \gamma_2 u(c_g(w,a))$, and $c_g(b,a) = \delta_0 a + \delta_1$, $V_g(b,a) = \delta_2 u(c_g(b,a))$ where all γ_i, δ_i, $i \in \{0,1\}$ are constants. If these guesses are correct, they must solve the system of differential equations (re-stated here for convenience),

$$\rho V(w,a)(a) = \frac{\sigma}{1-\sigma}\left[V_a(w,a)(a)^{\frac{\sigma-1}{\sigma}} - \frac{1}{\sigma}\right] + [ra+w]\,V_a(w,a)(a) + s\,[V(b,a) - V(w,a)], \tag{A.20}$$

$$\rho V(b,a) = \frac{\sigma}{1-\sigma}\left[V_a(b,a)^{\frac{\sigma-1}{\sigma}} - \frac{1}{\sigma}\right] + [ra+b]\,V_a(b,a)(a) + \mu\,[V(w,a) - V(b,a)]. \tag{A.21}$$

Applying CRRA utility, the guesses become

$$V_g^e(a) = \gamma_2 \frac{(\gamma_0 a + \gamma_1)^{1-\sigma}}{1-\sigma}, \quad V_g^u(a) = \delta_2 \frac{(\delta_0 a + \delta_1)^{1-\sigma}}{1-\sigma}.$$

Accordingly, $V_g^{e\prime}(a) = \gamma_0\gamma_2(\gamma_0 a + \gamma_1)^{-\sigma}$, $V_g^{u\prime}(a) = \delta_0\delta_2(\gamma_0 a + \gamma_1)^{-\sigma}$ and $u'(c_g^e) = (\gamma_0 a + \gamma_1)^{-\sigma}$ as well as $u'(c_g^u) = (\delta_0 a + \delta_1)^{-\sigma}$. Note that this derivation explicitly uses $d\gamma_i/da = d\delta_i/da = 0$. For the guess to solve the individual's problem, it also has to solve the f.o.c.'s , namely equate marginal utility to the shadow price of wealth. After substituting the linear guesses, we thus have $(\gamma_0 a + \gamma_1)^{-\sigma} = \gamma_0\gamma_2(\gamma_0 a + \gamma_1)^{-\sigma}$ and $(\delta_0 a + \delta_1)^{-\sigma} = \delta_0\delta_2(\delta_0 a + \delta_1)^{-\sigma}$, so that $\gamma_0\gamma_2 = \delta_0\delta_2 = 1$. Inserting this finding and the above expressions for $V_g^e(a)$, $V_g^u(a)$, $V_g^{e\prime}(a)$, and $V_g^{u\prime}(a)$ into the system of differential equations above, we get

$$\begin{aligned}
\rho\frac{(\gamma_0 a + \gamma_1)^{1-\sigma}}{(1-\sigma)\gamma_0} &= \frac{1}{\sigma-1} + \frac{\sigma}{1-\sigma}(\gamma_0 a + \gamma_1)^{1-\sigma} + [ra+w](\gamma_0 a + \gamma_1)^{-\sigma} + \\
&\quad + s\left[\frac{(\delta_0 a + \delta_1)^{1-\sigma}}{(1-\sigma)\delta_0} - \frac{(\gamma_0 a + \gamma_1)^{1-\sigma}}{(1-\sigma)\gamma_0}\right], \\
\rho\frac{(\delta_0 a + \delta_1)^{1-\sigma}}{(1-\sigma)\delta_0} &= \frac{1}{\sigma-1} + \frac{\sigma}{1-\sigma}(\delta_0 a + \delta_1)^{1-\sigma} + [ra+b](\delta_0 a + \delta_1)^{-\sigma} + \\
&\quad + \alpha\left[\frac{(\gamma_0 a + \gamma_1)^{1-\sigma}}{(1-\sigma)\gamma_0} - \frac{(\delta_0 a + \delta_1)^{1-\sigma}}{(1-\sigma)\delta_0}\right].
\end{aligned}$$

Now, we cannot make any reasonable assumption about the constants to get rid of a. We will see below that $\gamma_0 = \delta_0 = r$ is a solution to a similar system with CARA preferences and $s = 0$. Here in contrast, we are always left with two equations that depend on the level of wealth, a contradiction. If one is interested in the implications of a closed form solution, a possible way out is to assume that wages in equilibrium depend on assets and equal $w = \gamma^e a^\sigma$ and $b = \gamma^u a^\sigma$ (or, $w = \gamma^e a$ and $b = \gamma^u a$), but that the individual does not anticipate these relations when making her consumption decision. In general, however, it seems a good idea to think about numerical solutions.

6.D.3 Matching with Savings under CARA: A Boundary Condition

Finally, we briefly consider the same model with CARA preferences and provide a boundary condition for a numerical solution. After substituting for CRRA utility with CARA utility and performing analogous steps as in the main text, we obtain the two dimensional system of value functions describing the optimal behavior.

$$rV(w,a) = V_a(w,a)\left[ra + w - \frac{1}{\gamma}(1+\log\gamma) + \frac{1}{\gamma}\log V_a(w,a)\right] + s\left[V(b,a) - V(w,a)\right], \quad \text{(A.22)}$$

$$rV(b,a) = V_a(b,a)\left[ra + b - \frac{1}{\gamma}(1+\log\gamma) + \frac{1}{\gamma}\log V_a(b,a)\right] + \alpha\left[V(w,a) - V(b,a)\right]. \quad \text{(A.23)}$$

Upon solving for V^z and differentiating with respect to a, the first order condition for optimal consumption yields $c^e(a)$ and $c^u(a)$.

Lemma 6.1. *The value function*

$$V(w,a) = V(b,a) \equiv V(a) = \frac{-e^{-\gamma r a}}{r}$$

solves the system (A.22)-(A.23) as $a \to \infty$.

Proof. Inserting $V(a)$ and its derivative $V'(a) = \gamma e^{-\gamma r a}$ into the general system yields

$$-e^{-\gamma r a} = \gamma e^{-\gamma r a}\left[ra + w - \frac{1}{\gamma}(1+\log\gamma) + \frac{1}{\gamma}\log\left[\gamma e^{-\gamma r a}\right]\right],$$

$$-e^{-\gamma r a} = \gamma e^{-\gamma r a}\left[ra + b - \frac{1}{\gamma}(1+\log\gamma) + \frac{1}{\gamma}\log\left[\gamma e^{-\gamma r a}\right]\right],$$

which is equivalent to

$$-\frac{1}{\gamma} = ra + w - \frac{1}{\gamma}(1+\log\gamma) + \frac{1}{\gamma}\log\gamma - ra,$$

$$-\frac{1}{\gamma} = ra + b - \frac{1}{\gamma}(1+\log\gamma) + \frac{1}{\gamma}\log\gamma - ra,$$

and to $ra = ra + w$ and $ra = ra + b$. Dividing by a yields $r = r + w/a$ and $r = r + b/a$ which holds for $a \to \infty$. $\square$

This lemma implies that for large a, the solution of the general system must approach the solution $V(a)$. In other words, $V(a)$ provides a boundary condition for the system of differential equations. As wealth approaches infinity, consumption converges to $c^u = c^e = ra$.

Chapter 7

Optimal Unemployment Policy with Savings: Shimer and Werning (2008)

7.1 Introduction

After 30 years of intensive research, there is still a lively debate about the optimal design of unemployment insurance.[1] The main challenge tackled in this literature is to provide income insurance for risk-averse individuals with the incentive scheme most conducive to efficient outcomes (unemployment benefits offer insurance but at the same time induce individuals to reduce their search effort, increase their reservation wages, and exert less effort at work). On the normative side, the optimal timing of the unemployment benefits is of prominent interest. In seminal work on optimal unemployment insurance, Shavell and Weiss (1979) show that in the absence of asset holdings and moral hazard, the optimal benefit sequence is constant over time. Allowing for the incentives to search for a job to vary with unemployment benefits, they demonstrate that the inherent moral hazard problem requires benefits to decline over the unemployment spell. Including a constant but potentially history dependent wage tax/subsidy, Hopenhayn and Nicolini (1997) show that the declining benefits remain characteristic of an optimal unemployment policy (the wage tax should increase over the unemployment spell to further increase the cost of unemployment and spur job search).[2] While Shavell and Weiss (1979)

[1]See Atkinson and Micklewright (1992) and Holmlund (1998) for surveys of the "classical" literature on unemployment insurance.

[2]This finding is the prevalent justification underlying many countries' unemployment policies in the EU where, similar to the U.S., unemployment benefits are typically paid for about one year (usually followed by lower welfare payments).

account for wealth holdings at the beginning of the unemployment spell in some cases, Hopenhayn and Nicolini (1997) explicitly rule out savings throughout their analysis.[3] Allowing workers to save but taking the probability of finding a job as exogenous to the unemployed (i.e. abstracting from reservation wages), Shavell and Weiss (1979) find that benefits should initially be zero and then jump to a constant positive level over the unemployment spell.[4]

In their forthcoming AER article, Shimer and Werning (2008) employ an optimal contracting model in the spirit of Hopenhayn and Nicolini (1997) to address some of the previously unanswered questions on optimal unemployment benefits when workers have access to liquidity (i.e., can borrow and save in a riskless asset). Inter alia, they show that if individuals have constant absolute risk aversion (CARA) preferences, constant net benefits (equal to unemployment benefits net of employment taxes) are optimal. Using numerical simulations, they demonstrate that the optimal policy under constant relative risk aversion (CRRA) preferences involves "nearly constant" benefits and taxes so that the CARA case provides a good benchmark for CRRA preferences also. Consequently, Shimer and Werning (2008) advocate simple constant unemployment schemes. Accounting for individual savings, their paper presents a simple, analytically tractable model that allows a distinction between the insurance against the uncertain duration of unemployment spells on the one hand and consumption smoothing by means of liquidity on the other hand, a feature typically found in calibrated models on optimal unemployment insurance.[5][6] Put differently, the timing of consumption can be distinguished from the

A noticeable exception is Belgium, where unemployment benefits are paid for the entire duration of unemployment.

[3]In both papers, employment is an absorbing state; the job separation rate is zero.

[4]Shavell and Weiss (1979) demonstrate in a two period model that if the probability of finding a job can be influenced by the worker, optimal benefits may be hump-shaped. They were not able to derive a solution to the general model, however.

[5]The seminal contribution in the class of calibrated models is by Hansen and İmrohoroğlu (1992), who analyze the optimal level of unemployment benefits in an environment with a interest rate on savings equal to zero and without borrowing. Optimal unemployment policies in numerical models include Abdulkadiroglu, Kuruscu, and Şahin (2002) and Wang and Williamson (2002), who account for saving but rule out borrowing, and find that the optimal benefit sequence is u-shaped. Unfortunately, as pointed out by Shimer and Werning (2008, p. 4), it is hard to identify the source of the conflicting outcomes in the calibrated models. For a model with exogenous duration of unemployment and both saving and borrowing up to an exogenous constraint see Heer and Maußner (2005, Ch. 5).

[6]See Bloemen and Stancanelli (2005) for an empirical validation of the consumption smoothing effect of unemployment benefits. Lentz (2008) estimates optimal unemployment insurance in a job search model with savings using Danish micro data.

timing of benefits.

In what follows, we review the model with CARA preferences of Shimer and Werning (2008). Section 7.2 briefly introduces the model (the model is a special case of the general environment introduced in Section 6, with $s = 0$, $r = \rho$, and including a wage offer distribution). Section 7.3 considers the optimal behavior of the agents and establishes the optimality of a constant policy scheme if individuals have access to liquidity. In the main text, we concentrate on central arguments and relegate more cumbersome derivations to the appendix.

After presenting the model, we argue that the closed form solution obtained by Shimer and Werning (2008) relies on a government that, with positive probability, provides more debt to an unemployed than she can ever pay back, even if she commits to zero consumption forever after accepting a job. That is, a Aiyagari (1994)-type natural debt limit is violated with positive probability (cf. Aiyagari, 1994).[7] This equivalently implies negative consumption levels in finite time with positive probability. In the next section, we propose an alternative solution that accounts for a natural borrowing limit.

7.2 The Shimer and Werning (2008) Model

Consider a partial equilibrium sequential job search model with risk-averse individuals (McCall, 1970). Each unemployed individual exogenously receives job offers with a constant Poisson arrival rate $\alpha > 0$. Jobs pay a constant wage w and each offer is an independent draw from a known wage distribution $F(w)$ where $F(w) < 1$ for some $w > 0$. If the individual accepts a job, she keeps it forever; there is no "on the job" search and the separation rate is zero (in the limit as t goes to infinity, the unemployment rate goes to zero).[8] The individual receives benefits $b\,(t)$ when unemployed and pays an employment tax $\tau\,(t)$ when working on a job. Utility is derived from a single consumption good. Crucially, individuals have no means to pool their incomes so that they cannot insure against the uncertain duration of

[7] A debt limit is usually called "natural debt limit" if it emerges from the requirement of risk-free government bonds, i.e. if repayment must occur almost surely under optimal taxation. A more stringent debt limit is called an ad hoc limit (see Aiyagari, 1994, or Ljungqvist and Sargent, 2004, p. 577; cf. Chamberlain and Wilson, 2000). Aiyagari (1994) calls the debt limit "natural" that emerges from solving the dynamic budget constraint and imposing $c \geq 0$ and repayment with probability one. Following Aiyagari (1994), we take seriously the assumption that the interest rate is risk-free.

[8] Following the tradition in the optimal unemployment insurance literature, employment is an absorbing state. Chapter 6 provided a more general environment which includes temporary unemployment spells so that income shocks are transitory.

unemployment.

An unemployment agency aims at providing each individual with a given level of intertemporal utility at the lowest possible cost. In doing so, the agency observes only the individual's employment status, but not the realized wage once the individual accepts a job. Restricting the agency's information to the employment status forecloses the possibility of taxing every worker's income by 100% and redistributing the proceeds lump-sum whereby the first best attainable outcome would be feasible.

Shimer and Werning (2008) consider two policies. The first involves constant unemployment benefits and employment taxes and allows individuals to borrow and lend at the risk-free market interest rate r (UIP I). The second policy is the benchmark case of optimal unemployment insurance where the unemployment agency dictates the consumption path for individuals (UIP II). Under this policy scenario, the agency again cannot observe the wage offers, and the unemployed cannot borrow or save, i.e. they live from (the agency's) hand to mouth. In Appendix A of their paper, Shimer and Werning use the revelation principle to prove that UIP II is the best available deterministic mechanism under non-increasing risk aversion preferences given that the agency cannot observe the wage offers (which is used to prevent the first best) in that it provides a given amount of utility at the lowest possible cost.[9]

Throughout the paper, the interest rate r is equal to the subjective discount rate ρ. Following Shimer and Werning (2008) we proceed to show that the cost of providing a certain level of utility is the same under both policies, so that UIP I is in fact optimal (since UIP II yields the information-constrained efficient allocation).

7.3 Optimal Behavior

7.3.1 Optimal Behavior under UIP I

Under UIP I, the unemployment agency sets constant unemployment benefits $\bar{b}$ and employment taxes $\bar{\tau}$. Workers have access to liquidity in terms of a riskless asset and choose optimal consumption given $\bar{b}$ and $\bar{\tau}$.

[9]The main insight from this proof is that neither reports on the received wage offers nor the possibility to vary taxes over the employment spell reduce the costs of providing a given level of utility relative to what can be achieved using a policy that only accounts for the duration of unemployment.

The Agency's Program

The unemployment agency aims at maximizing the utility of each unemployed worker subject to its budget and the initial wealth level a_0. That is, the agency pays unemployment benefits when the individual is unemployed and continuously receives the tax proceeds as of the time she starts working on a job. Shimer and Werning (2008) solve this program via its dual, i.e. by minimizing the expected cost of the unemployment benefits and initial assets net of employment taxes given a certain expected level of intertemporal utility v_0 (expectations need to be formed about the duration of the unemployment spell). With E_0 denoting the expectations operator and $z(t) \in \{\bar{b}, -\bar{\tau}\}$, the total expected costs obey

$$E_0 \int_0^\infty e^{-rt} z(t)\, dt + a_0.$$

We show in Appendix 7.A that this objective can be rewritten as

$$\int_0^\infty e^{-\int_0^t (r+\alpha(1-F(\bar{w}^*(s))))ds} \left(\bar{b} - \alpha\left(1 - F\left(\bar{w}^*(t)\right)\right)\frac{\bar{\tau}}{r}\right) dt + a_0,$$

where $\bar{w}^*(t) \equiv \bar{w}(a(t), \bar{b}, \bar{\tau})$ denotes the reservation wage implied by the optimal behavior of an unemployed individual.

Denote by $V^u\left(a_0, \bar{b}, \bar{\tau}\right)$ the maximized Bellman equation of the unemployed, i.e. her equilibrium lifetime utility given initial assets a_0 and the sequence of constant unemployment benefits and taxes $\{\bar{b}, \bar{\tau}\}$. Taken together, the dual problem is

$$C^c(v_0, a_0) = \min_{\bar{b}, \bar{\tau}} \int_0^\infty e^{-\int_0^t (r+\alpha(1-F(\bar{w}^*(s))))ds} \left(\bar{b} - \alpha\left(1 - F\left(\bar{w}^*(t)\right)\right)\frac{\bar{\tau}}{r}\right) dt + a_0 \tag{7.1}$$

$$\text{subject to } v_0 = V^u\left(a_0, \bar{b}, \bar{\tau}\right). \tag{7.2}$$

Evidently, the constraint in (7.2) ensures that the policy endows the individual with utility v_0 in equilibrium, where the unemployed chooses her optimal consumption path and reservation wage optimally given $\{\bar{b}, \bar{\tau}\}$.

Note that in this environment with $r = \rho$, rational individuals are indifferent between taxes tomorrow distributed as lump-sum income today so that any initial income transfer x/r given the agency's budget leaves the unemployed's utility unaffected, $V^u\left(a; \bar{b}, \bar{\tau}\right) = V^u\left(a + x/r; \bar{b} - x, \bar{\tau} + x\right)$; Ricardian equivalence applies. Suppose $\left\{\bar{b}, \bar{\tau}\right\}^*$ is optimal. Since the $\left\{\bar{b}, \bar{\tau}\right\}^*$ and $\left\{\bar{b} - x, \bar{\tau} + x\right\}$ give rise to the same allocations (the individual is indifferent between both policies), the optimal policy leaves a_0 indeterminate. Following Shimer and Werning (2008), we thus suppress a_0 and simplify the notation by letting $C^c(v_0, a_0) = C^c(v_0)$.

We proceed to derive the closed-form solution for optimal consumption obtained by Shimer and Werning (2008). Under their solution, the optimal consumption path implies a constant reservation wage $\bar{w}$ that leaves the individual indifferent between accepting a job and remaining unemployed. The fact that this reservation wage is independent of the level of wealth is a direct implication of the optimal consumption rule being linear in the level of assets (see Chapter 8 for a nonlinear solution).

Workers

The worker's state is described completely by her employment status and current level of wealth. In the program considered by Shimer and Werning (2008), the asset space is unbounded so that the individual state space is given by the set $\{e, u\} \times (-\infty, \infty)$.[10]

As long as an unemployed individual does not find a job, her assets evolve according to $\dot{a}(t) = ra(t) + \bar{b} - c^u(t)$. After accepting a job, the worker's assets follow $\dot{a}(t) = ra(t) + w - \bar{\tau} - c^e(t)$ where $w \geq \bar{w}$.

An unemployed individual chooses consumption c^u to maximize her intertemporal utility,

$$\begin{aligned} rV^u(a) &= \max_{c^u} \left\{ u(c^u) + V^{u\prime}(a) \left[ra + \bar{b} - c^u \right] \right\} + \\ &\quad + \alpha \int_0^\infty \max \left[V^e(w, a) - V^u(a), 0 \right] dF(w). \end{aligned}$$

We solve this program for $c^u(a)$, $\bar{w}$, and the maximized value function by going through a "verification theorem" in Section 8 (cf. Merton, 1969 and 1971).[11] For CARA preferences, where $u(c) = -e^{-\gamma c}$, the Shimer-Werning (2008) solution obeys the following linear consumption rule (see the proof of Proposition 8.3 in Chapter 8):

$$c^u(a) = ra + \bar{w}. \tag{7.3}$$

The reservation wage $\bar{w}$ is implicitly given by

$$(\bar{w} - b)\gamma = \frac{\alpha}{r} \int_{\bar{w}}^\infty \left[1 + u(w - \bar{w}) \right] dF(w), \tag{7.4}$$

[10] In Chapter 8, we restrict the asset space to $[a_{\min}, \infty)$ where $a_{\min}$ is a limit on borrowing.

[11] The only constraint on this maximization problem is a no-Ponzi game condition of the form $\lim_{t\to\infty} e^{-rt} a(t) \geq 0$. Consumption is allowed to become negative as $c \in \mathbb{R}$. For a numerical solution of a general equilibrium model with an exogenous lower bound on wealth and CRRA preferences see Heer and Maußner, 2005, Ch. 5. Cf. the discussion in Chapter 8 below.

and the maximized value function satisfies

$$V^u = \frac{u\,(ra + \bar{w})}{r}. \tag{7.5}$$

The existence and uniqueness of $\bar{w}$ is easily verified. Rearranging (7.4) gives

$$\frac{\gamma r\,(\bar{w} - b)}{\alpha} = \int_{\bar{w}}^{\infty} \left[1 + u\,(w - \bar{w})\right] dF\,(w)\,.$$

The left hand side is equal to $-\gamma rb/\alpha < 0$ if $\bar{w} = 0$ and linearly increasing in $\bar{w}$. The right hand side is positive at $\bar{w} = 0$ (since $e^{\gamma(w-\bar{w})} \geq 1$ for $w \geq \bar{w}$) and strictly decreasing in $\bar{w}$.[12] Hence, a unique $\bar{w}$ exists.

When working on a job, the individual lives in a stationary world, and $r = \rho$ ensures that $\dot{a} = 0$. Hence,

$$c^e\,(w, a) = ra + w - \bar{\tau} \;\text{ and }\; V^e = \frac{u\,(c^e)}{r}. \tag{7.6}$$

In both employment states, optimal consumption is linear in the asset level. In the case of unemployment, this implies that the reservation wage is independent of the level of an individual's wealth, see (7.4) (which does not involve a). If unemployed, the individual consumes an amount $ra + \bar{w} > b$ and runs down her assets ($\dot{a}\,(t) = \bar{b} - \bar{w} < 0$, cf. the discussion in Chapter 8). If she accepts a job, $\dot{a} = 0$ and consumption equals the constant net income.

7.3.2 Optimal Behavior under UIP II

In the second scenario, UIP II, the unemployment agency sets time-dependent benefits $b(t)$ and employment taxes $\tau\,(t)$, and individuals have no means to borrow or save. The insurance agency thus dictates the level of consumption during the unemployment spell and workers consume their after-tax income. If unemployed, however, the individual can decide which job offer to accept. The agency therefore accounts for the unemployed's optimal reservation wage decision, but cannot control the reservation wage $\bar{w}\,(t)$ directly.

The objective of the unemployment agency is to provide the individual with a given amount of utility v_0 at the lowest possible cost. Since now the individuals can neither borrow nor save, the unemployed's reservation wage under UIP II is found by simple backward induction. At the second stage, the unemployed individual decides about her reservation wage given the unemployment policy $\{b\,(t)\,, \tau\,(t)\}$. At

[12] $\frac{\partial \int_{\bar{w}}^{\infty} [1+u(w-\bar{w})]dF(w)}{\partial \bar{w}} = \gamma \int_{\bar{w}}^{\infty} \left[u\,(w - \bar{w})\right] dF\,(w) < 0.$

the first stage, the agency sets the sequence of benefits and taxes, taking into account the unemployed's optimizing behavior. The agency thereby indirectly pins down the sequence of reservation wages.

To state the ensuing program, denote by $U(t', \{w(t), b(t), \tau(t)\})$ an unemployed individual's expected lifetime utility as of time t'. The agency then solves

$$\min_{\{b(t),\tau(t),\bar{w}(t)\}} \int_0^\infty e^{-\int_0^t (r+\alpha(1-F(\bar{w}(s))))ds} \left(b(t) - \alpha(1-F(\bar{w}(t))) \frac{\tau(t)}{r} \right) dt \tag{7.7}$$

such that the recommended sequence of reservation wages $\{\bar{w}'(t)\}$ (i) provides the worker with utility v_0,

$$v_0 = U(0, \{\bar{w}(t), b(t), \tau(t)\}), \tag{7.8}$$

and (ii), given $\{b(t), \tau(t)\}$, $\{\bar{w}(t)\}$ is at least as good as any other sequence of reservation wages $\{\bar{w}'(t)\}$,

$$U(0, \{\bar{w}(t), b(t), \tau(t)\}) \geq U\left(0, \{\bar{w}'(t), b(t), \tau(t)\}\right). \tag{7.9}$$

The unemployed's utility mirrors the costs of the unemployment agency. Its derivation is therefore analogous to the derivation of the operand in (7.1) deduced in Appendix 7.A. We thus simply adjust (7.7) to yield the present value utility of an unemployed individual. As long as she is looking for a job, the unemployed worker must consume whatever benefit the agency provides. While the agency pays out $b(t)$, the unemployed individual receives instantaneous utility $u(b(t))$. Once the individual accepts a job, she consumes her after-tax income. The worker then receives utility $u(w-\tau(t))$ where $w \geq \bar{w}(t)$, while the agency earns the proceeds from the employment tax (cf. (7.7)).
The distribution of the acceptable wages is the underlying distribution of wages conditional on $w \geq \bar{w}(t)$, i.e. $F(w)/[1-F(\bar{w})]$. Accordingly, an unemployed's expected lifetime utility as of the time she accepts a job obeys

$$\tilde{u}(\bar{w}(t)) = \int_{\bar{w}}^\infty \frac{u(w-\tau(t))}{r} \frac{dF(w)}{1-F(\bar{w}(t))}.$$

Sufficiently well paying jobs are offered with Poisson arrival rate $\alpha(1-F(\bar{w}(t)))$. Hence, replacing $b(t)$ by $u(b(t))$, $-\tau(t)/r$ by $\tilde{u}(\bar{w}(t))$, and adjusting the limits of integration in the objective in (7.7), the present value utility of an unemployed individual as of time t' equals

$$U\left(t', \bar{w}, \{b(t), \tau(t)\}\right) = \int_{t'}^\infty e^{-\int_{t'}^t (r+\alpha(1-F(\bar{w}(s))))ds} \left(u(b(t)) + \alpha \int_{\bar{w}}^\infty \frac{u(w-\tau(t))}{r} dF(w) \right) dt.$$

Recursive Formulation

The agency's program in (7.7), (7.8), and (7.9) can be expressed recursively, whereby utility $v(t') = U(t', \bar{w}, \{b(t), \tau(t)\})$ serves as the only state variable (Werning, 2002). Taking the derivative of $U(t', \bar{w}, \{b(t), \tau(t)\})$ in (7.10) with respect to time and substituting for the definition of $v(t)$ yields the evolution of utility given the reservation wage (see Appendix 7.B for the derivation):

$$\dot{v}(t) = rv(t) - u(b(t)) - \alpha \int_{\bar{w}(t)}^{\infty} \left[\frac{u(w - \tau(t))}{r} - v(t) \right] dF(w). \tag{7.10}$$

For the recommended reservation wage to be compatible with the unemployed's optimal choice, it must maximize the unemployed's intertemporal utility, see (7.9). Hence, $\bar{w}(t)$ must maximize $v(t)$ at any point in time, and Appendix 7.C shows that the constraint in (7.9) can equivalently be stated as

$$v(t) = \frac{u(\bar{w}(t) - \tau(t))}{r}. \tag{7.11}$$

If $C^*(v_0)$ is the minimum attainable cost implied by the solution of the agency's problem, $C^*(v_0)$ must solve the Hamilton-Jacobi-Bellman (HJB) equation

$$\begin{aligned} rC^*(v) = \min_{\bar{w},b,\tau} \Bigg(& b + (C^*)'(v) \left[rv(t) - u(b(t)) - \alpha \int_{\bar{w}(t)}^{\infty} \frac{u(w - \tau(t))}{r} - v(t)\, dF(w) \right] \\ & - \alpha (1 - F(\bar{w})) \left(\frac{\tau}{r} + C^*(v) \right) \Bigg) \end{aligned} \tag{7.12}$$

subject to (7.11). The sequence $\{\bar{w}(t), b(t), \tau(t)\}$ then follows from the first order conditions of (7.12).

7.4 Characterization of the Equilibrium under UIP I

Under UIP I, the equilibrium unemployment policy is given by the solution of (7.1) subject to (7.2) and the unemployed individual's optimality conditions (7.3)–(7.6), which enter via the agency's objective utility level v_0. In what follows, we substitute for v_0 and $\bar{\tau}$ in the agency's program, so that the equilibrium under UIP I is fully characterized by the reservation wage.

7.4.1 Exploring the Linear Solution

Two properties of the optimal behavior of the unemployed are used: first, the constancy of the reservation wage (see (7.4)) and, second, the closed form solution to the value function (see (7.5)). Inserting $\bar{\tau}$ and the "net benefit" $\bar{B} \equiv \bar{\tau} + \bar{b}$ as a function of $\bar{w}$ from the unemployed individual's optimality conditions for utility level v_0 and the unemployed's value function, we show in Appendix 7.D that the agency's objective can be rewritten as

$$C^c(v_0) = \max_{\bar{w}} \Phi(\bar{w})$$

where

$$\Phi(\bar{w}) = \frac{\int_{\bar{w}}^{\infty} 1 + \gamma \bar{w} + u(w - \bar{w})\, dF(w)}{r + \alpha(1 - F(\bar{w}))}. \tag{7.13}$$

It follows that the optimal reservation wage $\bar{w}^*$ is given by

$$\bar{w}^* \in \arg\max_{\bar{w}} \Phi(\bar{w}).$$

As demonstrated in Appendix 7.D, the implied insurance costs under UIP I with CARA preferences and a linear consumption rule amount to

$$C^c(v_0) = \frac{u^{-1}(rv_0)}{r} - \frac{\alpha \Phi(\bar{w}^*)}{\gamma r}. \tag{7.14}$$

7.5 Characterization of the Equilibrium under UIP II

The optimal unemployment policy under UIP II follows from the policy functions of the HJB equation in (7.12). The optimal behavior of the unemployed enters through the incentive compatibility constraint, which was expressed in recursive terms in (7.9).

The policy outcome is again characterized by the minimum expected cost associated with utility level v. Let us directly look at the solution:

$$C^*(v) = \frac{u^{-1}(rv)}{r} - \frac{\alpha \Phi(\bar{w}^*)}{\gamma r}. \tag{7.15}$$

where Φ is defined in (7.13). Accordingly, for $v = v_0$, the optimal policy under UIP II implies the same cost as the constant benefits and taxes under UIP I (cf. (7.14)). This is Shimer and Werning's main result: constant benefits and taxes exert constrained optimal incentives given that the worker has CARA preferences and is free to borrow and save in a riskless asset.

We are left to prove the validity of the cost function in (7.15). In doing so, we make a point of studying how the optimal decision of the unemployed affects the minimum costs. Following Shimer and Werning (2008), the proof is organized in two steps. The first step is to show that the minimum cost function for any initial level of promised utility v must be of the following form:

$$C^*(v) = \frac{u^{-1}(rv)}{r} + C_0, \tag{7.16}$$

where C_0 is a constant. To see this, we make use of the following helpful fact (see Appendix 7.E for a detailed proof): CARA preferences imply that a simultaneous shift in the non–capital income in both employment states does not alter the reservation wage and affects intertemporal utility linearly,

$$U\left(t', \{\bar{w}(t), b(t), \tau(t)\}\right) = -u(x)\, U\left(t', \{\bar{w}(t), b(t) + x, \tau(t) - x\}\right)$$

for all x. Therefore, if a policy $\{\bar{w}(t)^*, b(t)^*, \tau(t)^*\}$ is optimal for utility v_0, then the alternative policy $\{\bar{w}(t)^*, b(t)^* + x, \tau(t)^* - x\}$ is also accepted by the unemployed, and thus feasible. This alternative policy delivers utility $-u(x)\, v_0$. Since a lump-sum income transfer, i.e. a transfer that does not alter the decision to accept a job offer, gives the highest possible utility, it is also the agency's optimal choice to deliver utility $-u(x)\, v_0$. This is our argument for optimality. Shimer and Werning (2008) "establish" optimality of $\{\bar{w}(t)^*, b(t)^* + x, \tau(t)^* - x\}$ for utility $-u(x)\, v_0$ by arguing that since $\{\bar{w}(t)^*, b(t)^*, \tau(t)^*\}$ is revealed as good as any other feasible policy to deliver v_0, and $\{\bar{w}(t)^*, b(t)^* + x, \tau(t)^* - x\}$ is feasible, $\{\bar{w}(t)^*, b(t)^* + x, \tau(t)^* - x\}$ is optimal to deliver $-u(x)\, v_0$ by a "standard revealed preference argument".

Suppose that the agency wants to ensure zero consumption forever and thus deliver utility $v_0 = u(0)/r = -1/r$. Given the optimal sequence $\{\bar{w}(t)^*, b(t)^*, \tau(t)^*\}$, denote by $C_0 \equiv C^*(u(0)/r)$ the associated expected minimum attainable cost. Then, the alternative optimal policy $\{\bar{w}(t)^*, b(t)^* + x, \tau(t)^* - x\}$, which implies utility $-u(x)\, v_0 = u(x)/r$, yields costs $C_0 + x/r$ since the agency now pays out an additional amount x at any point in time.

Taken together, we can solve for the income shift x that yields a given initial utility v and the associated costs $C^*(v) = C_0 + x/r$. From $v = u(x)/r$, we have $x = u^{-1}(rv)$ so that the minimum attainable cost to deliver v is $C_0 + u^{-1}(rv)/r$, i.e. equation (7.16).

We are left with the second step, which is to determine the constant C_0. Here is where the optimal decision of the unemployed enters.

By definition, C_0 is the cost of delivering utility level $v = u(0)/r$. Since we know that the minimum costs $C^*(v)$ must satisfy the HJB equation (7.12), its first order conditions, and the incentive constraint (7.11), we can simply insert all these formulas in the HJB equation in (7.12) and solve for C_0. We again re-state the HJB equation from above for convenience:

$$\begin{aligned} rC^*(v) \quad = \quad & \min_{\bar{w}, b, \tau} \Bigg(b + (C^*)'(v) \left[rv(t) - u(b(t)) - \alpha \int_{\bar{w}(t)}^{\infty} \frac{u(w - \tau(t))}{r} - v(t)\, dF(w) \right] \\ & -\alpha\left(1 - F(\bar{w})\right)\left(\frac{\tau}{r} + C^*(v)\right)\Bigg). \end{aligned}$$

To begin with, note that $u^{-1}(rv) = 0$ and $C^*(v) = C_0$ if $v = u(0)/r$. The first order condition with respect to b is

$$1 - C^*(v)' u'(b) = 0. \tag{7.17}$$

Since

$$C^*(v)' = \frac{\left[u^{-1}(rv)\right]'}{r} = \frac{r}{ru'\left[u^{-1}(rv)\right]} = \frac{1}{u'\left[u^{-1}(rv)\right]}$$

from (7.16), $v = u(0)/r$ implies

$$C^*(v)' = \frac{1}{u'(0)} = \frac{1}{\gamma}.$$

Substituting for $C^*(v)$ in (7.17), optimality requires

$$1 - \frac{u'(b)}{u'(0)} = 0,$$

or equivalently $b = 0$. Moreover, evaluated at $v = u(0)/r$, the incentive constraint in (7.11) implies $\bar{w} - \tau = 0$, or $\bar{w} = \tau$. Taken together, if $v = u(0)/r = -1/r$, $C^*(v) = C_0$, $C^*(v)' = 1/\gamma$, $rv = u(0)$, $\tau = \bar{w}$, and $b = 0$, the HJB boils down to

$$rC_0 = \min_{\bar{w}} -\frac{\alpha}{\gamma} \int_{\bar{w}(t)}^{\infty} \frac{u(w - \tau(t)) + 1}{r} dF(w) - \alpha(1 - F(\bar{w})) \left(\frac{\bar{w}}{r} + C_0\right).$$

After collecting terms using $1 - F(\bar{w}) = \int_{\bar{w}}^{\infty} dF(w)$, the HJB equivalently reads

$$rC_0 = \min_{\bar{w}} -\frac{\alpha}{\gamma r} \int_{\bar{w}(t)}^{\infty} \left[u(w - \tau(t)) + 1 + \gamma\bar{w}\right] dF(w) - \alpha(1 - F(\bar{w})) C_0,$$

or, solving for C_0 and using the definition of $\Phi(\bar{w})$ in (7.13),

$$C_0 = \min_{\bar{w}} \frac{-\frac{\alpha}{\gamma r} \int_{\bar{w}(t)}^{\infty} \left[u(w - \tau(t)) + 1 + \gamma\bar{w}\right] dF(w)}{r + \alpha(1 - F(\bar{w}))} = -\frac{\alpha}{\gamma r} \max_{\bar{w}} \Phi(\bar{w}).$$

After inserting this constant in (7.16), we arrive at (7.15), which establishes optimality of the constant unemployment policy under CARA, if there is no separation, $r = \rho$, and consumption is allowed to take any (i.e. possibly negative) real number.

Appendix

The following appendices contain several detailed derivations of formulas used in the main text.

Appendix 7.A Derivation of the Expected Insurance Costs

In this appendix, we show that the expected net payment to an individual who is unemployed at $t = 0$ plus initial assets, $E_0 \int_0^\infty e^{-rt} z(t)dt + a_0$ with $z(t) \in \{\bar{b}, -\bar{\tau}\}$, can be rewritten as the operand in (7.1):

$$\int_0^\infty e^{-\int_0^t (r+\alpha(1-F(\bar{w}^*(s))))ds} \left(\bar{b} - \alpha\left(1 - F\left(\bar{w}^*\left(t\right)\right)\right)\frac{\bar{\tau}}{r}\right) dt + a_0.$$

To see this, consider

$$C_0 = E_0 \int_0^\infty e^{-rt} z(t)dt + a_0.$$

Expectations are formed only about the duration of unemployment, or, more precisely, about the future employment state, given that the individual is unemployed at $t = 0$. Exploring the Markovian evolution of the employment status and the fact that there is no separation after the individual has found a job, the conditional probability of being unemployed at $t \geq 0$ is simply given by

$$p_{uu}(t) = e^{-\int_0^t \alpha(1-F(w^*(s)))ds} = e^{-\bar{\alpha}(t)t},$$

where $\bar{\alpha}\left(t\right) \equiv 1/t \int_0^t \left(\alpha\left(1 - F\left(w^*\left(s\right)\right)\right)\right) ds$. Correspondingly, the probability of working on a job at $t \geq 0$, given unemployment at $t = 0$ is

$$1 - p_{uu}(t) = 1 - e^{-\bar{\alpha}(t)t}.$$

Taken together, the present value of the total expected resource costs amounts to

$$C_0 = \int_0^\infty e^{-rt} e^{-\int_0^t \alpha(1-F(\bar{w}^*(s)))ds}\bar{b} - e^{-rt}\left(1 - e^{-\bar{\alpha}(t)t}\right)\bar{\tau}dt + a_0.$$

Since $r = \rho$, the interest rate is constant so that $-rt = -\int_0^t rds$. We thus have

$$C_0 = \int_0^\infty e^{-\int_0^t r+\alpha(1-F(\bar{w}^*(s)))ds}\bar{b}dt - \int_0^\infty e^{-rt}\left(1 - e^{-\bar{\alpha}(t)t}\right)\bar{\tau}dt + a_0.$$

In order to arrive at the operand in (7.1), we need to show that

$$\int_0^\infty e^{-rt}\left(1 - e^{-\bar{\alpha}(t)t}\right)dt = \int_0^\infty e^{-rt} e^{-\int_0^t \bar{\alpha}(s)ds}\frac{\bar{\alpha}\left(t\right)}{r}dt.$$

To do so, let $f(t) \equiv 1 - e^{-\int_0^t \bar{\alpha}(s)ds}$ and $g(t) \equiv -e^{-rt}/r$. Using these definitions and $g'(t) = e^{-rt}$, the left hand side equals

$$\int_0^\infty e^{-rt}\left(1 - e^{-\bar{\alpha}(t)t}\right)dt = \int_0^\infty g'(t) f(t)\,dt.$$

After integrating by parts, whereby

$$\int_0^\infty g'(t) f(t)\,dt = [f(t)\,g(t)]_0^\infty - \int_0^\infty f'(t)\,g(t)\,dt,$$

we get, upon recognizing $f'(t) = \bar{\alpha}(t)\,e^{-\int_0^t \bar{\alpha}(s)ds}$,

$$\begin{aligned}\int_0^\infty e^{-rt}\left(1 - e^{-\bar{\alpha}(t)t}\right)dt &= \left[\left(1 - e^{-\int_0^t \bar{\alpha}(s)ds}\right)\left(-\frac{e^{-rt}}{r}\right)\right]_0^\infty + \\ &\quad + \int_0^\infty e^{-rt} e^{-\int_0^t \bar{\alpha}(s)ds}\frac{\bar{\alpha}(t)}{r}dt\end{aligned}$$

We are left to show that the term in squared brackets is zero. The first factor, $1 - e^{-\int_0^t \bar{\alpha}(s)ds}$, is again the conditional probability of being employed at $t \geq 0$. As $t \to \infty$, this probability goes to unity. Evaluated at $t \to \infty$, however, the second factor, $-e^{-rt}/r$, goes to zero, so that the product of the two factors goes to zero as $t \to \infty$. If $t = 0$, the second factor, $-e^{-rt}/r$, equals $-1/r$. Evaluated at $t = 0$, $1 - e^{-\int_0^t \bar{\alpha}(s)ds} = 0$ and again the product of both expressions is zero. Hence, the term in squared brackets is actually zero so that

$$E_0 \int_0^T e^{-rt} z(t)dt = \int_0^\infty e^{-\int_0^t (r+\alpha(1-F(\bar{w}^*(s))))ds}\bar{b} - e^{-rt}e^{-\int_0^t \bar{\alpha}(s)ds}\frac{\bar{\alpha}(t)}{r}\bar{\tau}dt.$$

After replacing $\bar{\alpha}(s)$ and $rt = \int_0^t r ds$, we arrive at

$$C_0 = \int_0^\infty e^{-\int_0^t (r+\alpha(1-F(\bar{w}^*(s))))ds}\left(\bar{b} - \alpha\left(1 - F(\bar{w}^*(t))\frac{\bar{\tau}}{r}\right)\right) + a_0.$$

Appendix 7.B Derivation of the Evolution of Utility

Equation (7.10) describes the lifetime utility of an unemployed as of time t' (repeated here for convenience):

$$U\left(t', \bar{w}, \{b(t), \tau(t)\}\right) = \int_{t'}^\infty e^{-\int_{t'}^t (r+\alpha(1-F(\bar{w}(s))))ds}\left(u(b(t)) + \alpha \int_{\bar{w}}^\infty \frac{u(w - \tau(t))}{r}dF(w)\right)dt.$$

Taking the derivative with respect to t', we get

$$\begin{aligned}\frac{\partial U(\cdot)}{\partial t'} &= -u\left(b\left(t'\right)\right) - \alpha \int_{\bar{w}(t')}^\infty \frac{u(w - \tau(t'))}{r}dF(w) + \\ &\quad + \int_{t'}^\infty e^{-\int_{t'}^t (r+\alpha(1-F(\bar{w}(s))))ds}\left[r + \alpha\left(1 - F\left(\bar{w}\left(t'\right)\right)\right)\right] \times \\ &\quad \left(u(b(t)) + \alpha \int_{\bar{w}}^\infty \frac{u(w - \tau(t))}{r}dF(w)\right)dt.\end{aligned}$$

Using $v(t') = U(t', \bar{w}, \{b(t), \tau(t)\})$ and changing the time argument from t' to t,

$$\dot{v}(t) = [r + \alpha(1 - F(\bar{w}(t)))]\, v(t) - u(b(t)) - \alpha \int_{\bar{w}(t)}^{\infty} \frac{u(w - \tau(t))}{r} dF(w).$$

Recognizing $1 - F(\bar{w}(t')) = \int_{-\infty}^{\infty} dF(w) - \int_{-\infty}^{\bar{w}(t')} dF(w) = \int_{\bar{w}}^{\infty} dF(w)$, we arrive at the evolution of utility in equation (7.10):

$$\dot{v}(t) = rv(t) - u(b(t)) - \alpha \int_{\bar{w}(t)}^{\infty} \left[\frac{u(w - \tau(t))}{r} - v(t)\right] dF(w).$$

Appendix 7.C Recursive Formulation of the Participation Constraint

The worker only uses the agency's recommended reservation wage sequence if it is at least as good as any other reservation wage sequence. In recursive terms, this is to say that $\bar{w}(t)$ must maximize $v(t)$. Finding the respective necessary condition is most easily achieved by using (7.10) to define the implicit function

$$\Gamma(v(t), \bar{w}(t)) \equiv \dot{v}(t) - rv(t) + u(b(t)) + \alpha \int_{\bar{w}(t)}^{\infty} \left[\frac{u(w - \tau(t))}{r} - v(t)\right] dF(w) = 0. \tag{A.1}$$

From the implicit function theorem,

$$\frac{\partial v(t)}{\partial \bar{w}(t)} = -\frac{\partial \Gamma(v(t), \bar{w}(t)) / \partial \bar{w}(t)}{\partial \Gamma(v(t), \bar{w}(t)) / \partial v(t)},$$

where we remember that $v(t') = U(t', \{\bar{w}, b(t), \tau(t)\})$. For $\bar{w}(t)$ to maximize $v(t)$, we require

$$\frac{\partial \Gamma(v(t), \bar{w}(t))}{\partial \bar{w}(t)} = -\frac{u(\bar{w}(t) - \tau(t))}{r} + v(t) = 0.$$

Rearranging gives $U(0, \bar{w}, \{b(t), \tau(t)\}) \geq U(0, \{\bar{w}'(t), b(t), \tau(t)\})$ equivalently as

$$v(t) = \frac{u(\bar{w}(t) - \tau(t))}{r}.$$

Appendix 7.D Derivation of the Insurance Costs under UIP I and CARA

The unemployment agency's objective was given in (7.1), and is repeated here for convenience:

$$\min_{\bar{b}, \bar{\tau}} \int_0^{\infty} e^{-\int_0^t (r + \alpha(1 - F(\bar{w}(s))))ds} \left(\bar{b} - \alpha(1 - F(\bar{w}(t))) \frac{\bar{\tau}}{r}\right) dt + a_0.$$

Given that $\bar{w}$ is constant (see (7.4)),

$$e^{-\int_0^t (r+\alpha(1-F(\bar{w}(s))))ds} = e^{-(r+\alpha(1-F(\bar{w})))t},$$

and $\bar{b} - \alpha\left(1 - F\left(\bar{w}\left(t\right)\right)\right)\bar{\tau}/r = \bar{b} - \alpha\left(1 - F\left(\bar{w}\right)\right)\bar{\tau}/r$ can be pulled out of the integral:

$$\min_{\bar{b},\bar{\tau}} \left(\bar{b} - \alpha\left(1 - F\left(\bar{w}\right)\right)\frac{\bar{\tau}}{r}\right)\int_0^\infty e^{-(r+\alpha(1-F(\bar{w})))t}dt + a_0.$$

After integrating, we have

$$\min_{\bar{b},\bar{\tau}} \frac{\bar{b} - \alpha\left(1 - F\left(\bar{w}\left(t\right)\right)\right)\frac{\bar{\tau}}{r}}{r + \alpha\left(1 - F\left(\bar{w}\right)\right)} + a_0,$$

or, substituting $\bar{b} + \bar{\tau} \equiv \bar{B}$,

$$\min_{\bar{b},\bar{\tau}} \frac{\bar{B} - \tau - \alpha\left(1 - F\left(\bar{w}\left(t\right)\right)\right)\frac{\bar{\tau}}{r}}{r + \alpha\left(1 - F\left(\bar{w}\right)\right)} + a_0.$$

Now, after collecting terms,

$$\min_{\bar{b},\bar{\tau}} \frac{\bar{B}}{r + \alpha\left(1 - F\left(\bar{w}\right)\right)} - \bar{\tau}\frac{\frac{r+\alpha(1-F(\bar{w}(t)))}{r}}{r + \alpha\left(1 - F\left(\bar{w}\right)\right)} + a_0,$$

and canceling $r + \alpha\left(1 - F\left(\bar{w}\right)\right)$,

$$\min_{\bar{b},\bar{\tau}} \frac{\bar{B}}{r + \alpha\left(1 - F\left(\bar{w}\right)\right)} - \frac{\bar{\tau}}{r} + a_0. \tag{A.2}$$

The next step is to substitute for $\bar{\tau}$ using the unemployed's optimal consumption decision (7.3), her value function (7.5), and the objective utility, $v_0 \equiv V^u\left(a_0; \bar{b}, \bar{\tau}\right)$. After inserting the value function V^u from (7.5), we get

$$v_0 = \frac{u\left(c^u\right)}{r},$$

where $c^u = ra - \bar{\tau} + \bar{w}$ is the optimal consumption rule when unemployed. To solve this equation for $\bar{\tau}$, multiply by r, apply u^{-1} on both sides, and insert c^u at $t = 0$ to obtain $u^{-1}\left(rv_0\right) = ra_0 - \tau + \bar{w}$. Accordingly,

$$\tau = ra_0 + \bar{w} - u^{-1}\left(rv_0\right) \tag{A.3}$$

so that (A.2) becomes

$$\min_{\bar{B},\bar{w}} \frac{\bar{B}}{r + \alpha\left(1 - F\left(\bar{w}\right)\right)} - \frac{ra_0 + \bar{w} - u^{-1}\left(rv_0\right)}{r} + a_0.$$

The initial asset level a_0 drops out:

$$\min_{\bar{B},\bar{w}} \frac{\bar{B}}{r + \alpha\left(1 - F\left(\bar{w}\right)\right)} + \frac{u^{-1}\left(rv_0\right) - \bar{w}}{r}, \tag{A.4}$$

and the term $u^{-1}(rv_0)/r$ is a constant in the cost minimization. We are left to substitute for $\bar{B}$. To do so, we solve the unemployed's optimality condition that determines the reservation wage (again repeated for convenience),

$$\gamma\left(\bar{w}-\bar{B}\right)=\frac{\alpha}{r}\int_{\bar{w}}^{\infty}\left[1+u\left(w-\bar{w}\right)\right]dF\left(w\right),$$

for $\bar{B}$ and get

$$\bar{B}=\bar{w}-\frac{\alpha}{\gamma r}\int_{\bar{w}}^{\infty}\left[1+u\left(w-\bar{w}\right)\right]dF\left(w\right). \tag{A.5}$$

After substituting for $\bar{B}$ and dropping the constant $u^{-1}(rv_0)/r$, (A.4) becomes

$$\min_{\bar{B},\bar{w}}\frac{\bar{w}-\frac{\alpha}{\gamma r}\int_{\bar{w}}^{\infty}1+u\left(w-\bar{w}\right)dF\left(w\right)}{r+\alpha\left(1-F\left(\bar{w}\right)\right)}-\frac{\bar{w}}{r}.$$

Collecting the reservation wage terms, we have

$$\min_{\bar{B},\bar{w}}\frac{\bar{w}\left(1-\frac{r+\alpha(1-F(\bar{w}))}{r}\right)}{r+\alpha\left(1-F\left(\bar{w}\right)\right)}-\frac{\frac{\alpha}{\gamma r}\int_{\bar{w}}^{\infty}\left[1+u\left(w-\bar{w}\right)\right]dF\left(w\right)}{r+\alpha\left(1-F\left(\bar{w}\right)\right)}$$

so that, after canceling r from the numerator of the first term, we get

$$\min_{\bar{B},\bar{w}}-\frac{\frac{\alpha}{r}\bar{w}\left(1-F\left(\bar{w}\right)\right)}{r+\alpha\left(1-F\left(\bar{w}\right)\right)}-\frac{\frac{\alpha}{\gamma r}\int_{\bar{w}}^{\infty}\left[1+u\left(w-\bar{w}\right)\right]dF\left(w\right)}{r+\alpha\left(1-F\left(\bar{w}\right)\right)}.$$

Dropping the common factor α/r from both terms, and recognizing $1-F(\bar{w})=\int_{\bar{w}}^{\infty}dF(w)$, we can combine the two terms to yield

$$\min_{\bar{B},\bar{w}}-\frac{\frac{1}{\gamma}\int_{\bar{w}}^{\infty}\left[1+\gamma\bar{w}+u\left(w-\bar{w}\right)\right]dF\left(w\right)}{r+\alpha\left(1-F\left(\bar{w}\right)\right)}.$$

Finally, let

$$\Phi\left(\bar{w}\right)=\frac{\int_{\bar{w}}^{\infty}\left[1+\gamma\bar{w}+u\left(w-\bar{w}\right)\right]dF\left(w\right)}{r+\alpha\left(1-F\left(\bar{w}\right)\right)}$$

and drop $-1/\gamma$ to arrive at

$$C^{c}\left(v_0\right)=\max_{\bar{w}}\Phi\left(\bar{w}\right).$$

Hence, the reservation wage $\bar{w}^*$ is given by

$$\bar{w}^*\in\arg\max_{\bar{w}}\Phi\left(\bar{w}\right).$$

The unemployment policy $\left(\bar{b}^*,\bar{\tau}^*\right)$ follows from (A.3) and (A.5) using the definition of $\bar{B}$. Adding again the dropped factor $\alpha/(\gamma r)$ and the summand $u^{-1}(rv_0)/r$, the insurance costs are

$$C^{c}\left(v_0\right)=\frac{u^{-1}\left(rv_0\right)}{r}-\frac{\alpha\Phi\left(\bar{w}\right)}{\gamma r}.$$

Appendix 7.E A Useful Property of CARA

We show in this appendix that a simultaneous shift in the "labor" income in both states does not affect the reservation wage under CARA preferences. Level effects in income simply affect intertemporal utility linearly:

$$U\left(t', \{\bar{w}(t), b(t), \tau(t)\}\right) = -u(x)\, U\left(t', \{\bar{w}(t), b(t) + x, \tau(t) - x\}\right) \tag{A.6}$$

for all x. Moreover, if a sequence $\{\bar{w}(t)^*, b(t)^*, \tau(t)^*\}$ is optimal to implement utility $u(0)/r$, then $\{\bar{w}(t)^*, b(t)^* + u^{-1}(rv), \tau(t)^* - u^{-1}(rv)\}$ is optimal for utility level v_0 (cf. the discussion in Section 7.5 and Lemma 1 in Shimer and Werning, 2008). An unemployed's utility as of time t' was derived in (7.10), and is repeated here for convenience:

$$U\left(t', \bar{w}, \{b(t), \tau(t)\}\right) \equiv \int_{t'}^{\infty} e^{-\int_{t'}^{t}(r+\alpha(1-F(\bar{w}(s))))ds} \left(u(b(t)) + \alpha \int_{\bar{w}}^{\infty} \frac{u(w - \tau(t))}{r} dF(w)\right) dt.$$

Raising the "labor" income in both states by x, utility becomes

$$\begin{aligned} U\left(t', \bar{w}, \{b(t) + x, \tau(t) - x\}\right) = \int_{t'}^{\infty} & e^{-\int_{t'}^{t}(r+\alpha(1-F(\bar{w}(s))))ds} \big(u(b(t) + x) + \\ & + \alpha \int_{\bar{w}}^{\infty} \frac{u(w - \tau(t) + x)}{r} dF(w)\bigg) dt. \end{aligned}$$

Since

$$u(c + x) = -e^{-\gamma(c+x)} = -(-e)^{-\gamma c}\left(-e^{-\gamma x}\right) = -u(c)\, u(x),$$

we have

$$\begin{aligned} U\left(t', \bar{w}, \{b(t) + x, \tau(t) - x\}\right) &= -u(x) \int_{t'}^{\infty} e^{-\int_{t'}^{t}(r+\alpha(1-F(\bar{w}(s))))ds} \big(u(b(t)) + \\ & \alpha \int_{\bar{w}}^{\infty} \frac{u(w - \tau(t))}{r} dF(w)\bigg) dt, \end{aligned}$$

i.e. equation (A.6). Hence, utility is simply multiplied by $-u(x)$, and this carries over to the evolution of utility. Accordingly, changing the level of income by the same amount in both states leaves Γ in (A.1) unaffected and hence does not alter the sequence of reservation wages that maximizes the unemployed's intertemporal utility.

Chapter 8

Job Search with Borrowing Constraint under CARA

8.1 Introduction

In a standard job search model with employment as an absorbing state and without consumption smoothing motives from the interest rate, Shimer and Werning (2008, cf. Chapter 7) have shown that consumption is linear in wealth and reservation wages are independent of wealth if individuals have CARA preferences. These results rest on the assumption that borrowing is unlimited and only a no-Ponzi game condition needs to be taken into account.

Shimer and Werning (2008) characterize the optimal unemployment policy in a sequential search environment where agents can borrow and save in a riskless asset. A closed form solution for consumption allows an analytical treatment of the optimal contract, and unlimited access to liquidity is a pragmatic way to obtain a closed-form solution. Intuitively speaking, unbounded debt allows a linear consumption function to solve the necessary conditions for optimality.[1]

In this section, we introduce a borrowing constraint in the spirit of Aiyagari (1994) into the optimal savings problem considered by Shimer and Werning (2008) under CARA preferences.[2] Our motivation is twofold. First, accounting for a borrowing constraint that ensures repayment with probability one

[1]A direct implication is that the optimal decision to accept a job is independent of the unemployed's history because (i) the probability of finding a job is constant and (ii) unlimited borrowing guarantees that consumption smoothing is equally feasible at all levels of wealth.

[2]This chapter originates from joint work with Klaus Wälde. It presents a modified version of Bauer (2008b).

and nonnegative consumption is intrinsically reasonable.[3] Ruling out the possibility of default takes serious the fact that the bond interest rate is risk-free.

Second, empirical research has identified a significant positive impact of wealth on the reservation wage (Algan, Chèron, Hairault, and Langot, 2003, Bloemen and Stancanelli, 2001, Stancanelli 1999, among others)[4] on the one hand, and strict borrowing limits for a large fraction of the unemployed on the other hand (cf. Browning and Crossley, 2001, Chetty, 2005, Sullivian, 2002 and 2008, and Rendon, 2006). Imposing repayment with probability one naturally generates wealth dependent reservation wages in line with the data even under CARA preferences. We contribute to the understanding of the empirical findings by differentiating between the effects of liquidity, insurance, and risk aversion[5].

Introducing what Aiyagari (1994) calls a "natural borrowing limit", i.e. imposing nonnegative consumption and repayment with probability one, we show that the optimal consumption of the unemployed is no longer linear in wealth. The intuition is that the borrowing constraint rules out debt levels of unemployed workers which are implied by a linear solution while leaving consumption of wealthy individuals unaffected. We prove that a linear solution requires an unemployed worker to borrow more (with a strictly positive probability) than the present value of her lifetime income, even if she commits to a zero consumption level after having found a job. We present a solution where the intertemporal budget constraint (IBC for short) is satisfied at each point and clarify the role of borrowing constraints for the duration dependence of hazard rates. The solution implies that unlike in the case of unlimited borrowing, the distribution of wages below the reservation wage affects the individual's optimal behavior.

The cost of restricting consumption to nonnegative levels is that no closed-form solution as in Shimer

[3]Under CRRA preferences, nonnegativity of consumption naturally implies a lower bound on debt that allows repayment with probability one. As shown in a discrete time model by Aiyagari (1994), nonnegative consumption together with the no-Ponzi game condition $\lim_{t\to\infty} a(t)\, e^{-rt} \geq 0$ is equivalent to a constraint on debt that allows repayment with probability one.

[4]Addison, Centeno, and Portugal (2008) find little evidence for declining reservation wages across the employment spell. It is not evident from their study, however, if individual asset holdings are taken into account.

[5]Allowing for negative consumption, the reservation wage in an environment with constant net benefits is independent of an individual's wealth under CARA, see Shimer and Werning (2007 and 2008). In the numerical solution to the CRRA case with optimal unemployment insurance in Shimer and Werning (2008), the reservation wage is nearly constant, reflecting a balance of decreasing reservation wages as a response to increasing risk aversion as the unemployed becomes poorer and an increase in the moral hazard from raising unemployment benefits, which is a feature of the optimal unemployment policy scheme.

and Werning (2008) can be found. In deriving our results, however, we characterize the optimal behavior by exploring analytically an expression for the optimal evolution of consumption in a simple phase diagram.

The idea that limited asset endowments mitigate the moral hazard problem of prolonging unemployment spells has a long tradition. It dates back at least to Danforth (1979), who show that risk averse workers choose a lower reservation wage if they hold little wealth. Similarly, Mortensen (1986) establishes that a liquidity constraint imposes an upper bound on the search time an individual is able to incur so that the reservation wage is decreasing in the worker's wealth. To isolate the impact of the borrowing constraint, we consider an individual with CARA preferences and impose an interest rate equal to the subjective discount rate.

The paper most closely related to our model is by Lentz and Tranæs (2005), who consider a discrete-time sequential search model where a risk averse individual switches back and forth between employment and unemployment according to a two-state Markov process. In pioneering work, Lentz and Tranæs (2005) demonstrate that the expected length of unemployment originates endogenously from the optimal consumption behavior under risk aversion if individuals have access to liquidity. Over the unemployment spell, unemployed individuals reduce their level of financial wealth and increase their search efforts. Our research clarifies that bounded debt is a sufficient condition to obtain this pattern.

In related work, Rendon (2006) estimates a labor market search model with on-the-job search and longitudinal survey data. He finds finds that larger initial wealth and free access to liquidity increase the expected duration of unemployment as well as future wage incomes. He also confirms the earlier assessment of tight borrowing constraints. According to Rendon's (2006) estimates, borrowing is typically limited to 14% of an individual's present value of risk-free income and thus far tighter than assumed in this section (we impose a limit equal to 100% of the risk-free income).[6] Rendon (2006) derives his results from a discrete-time job search model with finite planning horizon in the spirit of Danforth (1979). As usual in the literature, he obtains policy rules numerically by discretizing the state variables, applying constant relative risk aversion preferences and a specified (truncated lognormal) wage offer distribution among other functional forms.

In what follows, we employ a standard job search model in continuous time where the transition between unemployment and employment is governed by a Poisson process. Abstracting from changing

[6] Most of the arguments in the analysis go through as long as any borrowing constraint exists, independently of its level.

risk aversion as wealth changes using CARA preferences and imposing an interest rate equal to the subjective discount rate, our analysis provides further insights into how borrowing constraints, risk aversion, and insurance affect the reservation wage as a function of wealth as well as equilibrium consumption. Inter alia, we prove that a borrowing limit naturally generates a reservation wage that increases monotonically in an individual's wealth even under CARA preferences and assumptions most conducive towards a constant reservation wage (the interest rate equals the subjective discount rate and is the same for borrowing and for saving, the job arrival rate is exogenous, and the degree of risk aversion is constant). We obtain our results using analytical expressions for optimal consumption in the presence of idiosyncratic labor income risk and assess the impact of borrowing constraints on optimal consumption and the reservation wage via phase diagrams. To the best of our knowledge, there are no other studies who link nonnegativity constraints on consumption to wealth dependent reservation wages under constant absolute risk aversion in a similar environment of sequential job search with individual savings.

The remainder of this section is organized as follows. Section 8.2 briefly recapitulates the relevant parts of the Shimer-Werning (2008) model. Section 8.3 demonstrates our claim in a "pure matching setup", i.e. there is no wage distribution and therefore no reservation wage. We show that a linear consumption rule violates the IBC and characterize the optimal consumption path in the case with a borrowing constraint/nonnegative consumption. A wage distribution and wealth dependent reservation wages are introduced in Section 8.4. Section 8.5 concludes. The appendix contains numerical illustrations and considers the liquidity effect under CARA.

8.2 The Model

Consider the sequential search model employed by Shimer and Werning (2008, cf. Chapter 7). At the heart of the model is an individual with infinite planning horizon and preferences given by

$$U(t) = \int_t^\infty e^{-\rho(\tau - t)} u\left(c(\tau)\right) d\tau,$$

where ρ denotes the subjective discount rate and t is the time index. Instantaneous utility is of the CARA type

$$u\left(c\right) = -e^{-\gamma c}. \tag{8.1}$$

Like in Shimer and Werning (2008), we impose an interest rate r equal to the discount rate ρ, $r = \rho$, to abstract from impatience effects. We concentrate on the case of constant unemployment benefits and employment taxes and set the employment tax to zero for simplicity ($\bar{\tau} = 0$ in Shimer and Werning's notation).[7] With a denoting an individual's wealth, the dynamic budget constraint of an employed worker obeys $\dot{a}^e = ra + w - c^e$, where w is the wage rate and c^e the optimal level of consumption during the employment spell. Similarly, while unemployed, the flow budget constraint is given by

$$\dot{a}^u = ra^u + \bar{b} - c^u. \tag{8.2}$$

If an unemployed individual accepts a job, her labor income jumps from constant unemployment benefits $\bar{b}$ to w ($> \bar{b}$) and remains there indefinitely (i.e., employment is an absorbing state, the job separation rate is zero). The Poisson arrival rate of job offers is given by α.[8]

8.3 The Case without a Wage Offer Distribution

We first look at the "pure matching case", i.e. a setup without wage distribution, and derive a closed-form solution. We show that the closed-form solution violates the IBC and demonstrate what the optimal consumption path looks like under the IBC.

8.3.1 A Closed-Form Solution without a Borrowing Constraint

The Consumption Rule

We derive a closed-form solution by going through a verification theorem. This approach is in the tradition of the "educated guess" (cf. Merton, 1969, 1971, among others).

Proposition 8.1 (**Shimer and Werning, 2008, without a wage distribution/reservation wage**). *Optimal consumption for an employed worker and her value function are given by*

$$\begin{aligned} V^e(a) &= \frac{u(ra+w)}{r}, && (8.3) \\ c^e(a) &= ra + w. && (8.4) \end{aligned}$$

Proof. The value function for an employed worker is

$$rV^e(a) = \max_c \left\{ u(c) + [ra + w - c]\, V^{e\prime}(a) \right\},$$

[7] Since Ricardian equivalence holds, doing so comes without loss of generality.

[8] Under CARA and including a wage offer distribution, it is not optimal to get back to an earlier job offer so that recall can safely be ignored.

and the first order condition reads $u'(c) = V^{e\prime}(a)$. Now make an "educated guess" and use as candidate solutions (g for guess and γ_i are constants)

$$c_g^e = \gamma_0 a + \gamma_1,\ V_g^e(a) = \gamma_2 u(c_g^e).$$

Inserting them into the first order condition reveals $\gamma_0\gamma_2 = 1$. Inserting them into the Bellman equation gives

$$\frac{r}{\gamma_0} = 1 - ((r-\gamma_0)a + w - \gamma_1)\gamma.$$

As we see, the wealth level drops out for $\gamma_0 = r$. The constants are therefore $\gamma_0 = r,\ \gamma_1 = w,\ \gamma_2 = 1/r$. Accordingly, the optimal consumption level and value function are

$$c^e = ra + w,\ V^e = \frac{u(c^e)}{r}.$$

□

Now we perform the same procedure for the unemployed worker. The optimal behavior of the unemployed individual can be described by a Hamilton-Jacobi-Bellman (HJB) equation. With $r = \rho$ and without a wage distribution to start with, the HJB equation obeys

$$rV^u(a) = \max_{c(t)} \left\{ u(c(\tau)) + V^{u\prime}(a)\left(ra + \bar{b} - c\right) \right\} + \alpha\left\{V^e(a) - V^u(a)\right\}. \tag{8.5}$$

The first-order condition is

$$u'(c) = V^{u\prime}(a). \tag{8.6}$$

Proposition 8.2. *Optimal behavior of an unemployed individual is described by*

$$c^u = ra + \delta_1,\ V^u(a) = \frac{u(c^u)}{r}. \tag{8.7}$$

where δ_1 strictly lies between $\bar{b}$ and w, being determined by $\left(\bar{b} - \delta_1\right)\frac{\gamma r}{\alpha} = e^{-\gamma(w-\delta_1)} - 1$.

Proof. Assume that optimal behavior is given by

$$c_g^u = \delta_0 a + \delta_1,\ V_g^u(a) = \delta_2 u(c_g^u).$$

Analogously to the above, inserting c_g^u and $V_g^u(a)$ into the first order condition requires $\delta_0\delta_2 = 1$. Inserting the guess into the Bellman equation (8.5) yields

$$\begin{aligned} -r\delta_2 e^{-\gamma(\delta_0 a+\delta_1)} &= -e^{-\gamma(\delta_0 a+\delta_1)} + \gamma\delta_0\delta_2 e^{-\gamma(\delta_0 a+\delta_1)}(ra + \bar{b} - \delta_0 a - \delta_1) + \\ &\quad +\alpha\left[-\gamma_2 e^{-\gamma(\gamma_0 a+\gamma_1)} + \delta_2 e^{-\gamma(\delta_0 a+\delta_1)}\right]. \end{aligned}$$

After canceling $e^{-\gamma(\delta_0 a+\delta_1)}$, inserting $\gamma_0 = r$, $\gamma_1 = w$, $\gamma_2 = 1/r$, noting that $\delta_2 = 1/\delta_1$, and rearranging terms, we get

$$\frac{r}{\delta_0} = 1 - ((r-\delta_0)a + \bar{b} - \delta_1)\gamma + \alpha\left[\frac{e^{-\gamma((r-\delta_0)a+w-\delta_1)}}{r} - \frac{1}{\delta_0}\right].$$

The wealth level drops out for $\delta_0 = r$. Then,

$$\left(\bar{b} - \delta_1\right)\frac{\gamma r}{\alpha} = e^{-\gamma(w-\delta_1)} - 1. \tag{8.8}$$

The left hand side is a linearly decreasing function in δ_1, intersecting the horizontal axis at $\delta_1 = \bar{b}$ from above. The right hand side is monotonically increasing in δ_1 and equals $exp(-\gamma w)$ at δ_1. It intersects the horizontal axis at $\delta_1 = w$ from below. Hence, equation (8.8) determines a unique $\delta_1 \in \left]\bar{b}, w\right[$. $\square$

The optimal consumption path is thus linear in wealth.

The Dynamics of Savings

We now analyze how savings of the unemployed evolve. Given the budget constraint of the unemployed worker and optimal consumption from (8.7),

$$\dot{a}^u = ra + \bar{b} - (ra + \delta_1) = \bar{b} - \delta_1. \tag{8.9}$$

As $\delta_1 > \bar{b}$, wealth falls. As both δ_1 and $\bar{b}$ are constant, it seems that wealth drops without bound. At the wealth level

$$a^u_{neg} = -\delta_1/r, \tag{8.10}$$

however, consumption c^u becomes negative, see (8.7). If we assume (in an ad hoc fashion) that consumption is zero from there on, wealth continues to follow (8.9) with $c^u = 0$, i.e. $\dot{a}^u = ra^u + \bar{b}$. Using (8.10), we find $\dot{a}^u = -\delta_1 + \bar{b}$. Following the argument from before, as δ_1 is strictly larger than $\bar{b}$, we conclude that assets continue to fall even if consumption remains constant at zero. Income of the unemployed is not enough to pay back interest on debt. This cannot be a feasible optimal behavior if the lender does not want to take risk from the unemployed (i.e., if the bond is risk-free).

8.3.2 The Intertemporal Budget Constraint

Given this conclusion, we now impose a more restrictive intertemporal constraint on the individual's maximization program, which also implies a no-Ponzi game condition.[9]

[9] The typical wording "dynamic budget constraint" for the law of motion of the state variable, i.e. the evolution of wealth in our case, is generally misleading since it does not really constrain the individual by itself. If the evolution of

What does such an intertemporal constraint look like? If we formulate a constraint for an unemployed worker in realizations (i.e., suppose we knew the realization of the state variables at any instant), it reads

$$\int_t^\infty e^{-\rho[\tau-t]}c\left(\tau\right)d\tau = a_t + \int_t^\infty e^{-\rho[\tau-t]}z\left(\tau\right)d\tau, \tag{8.11}$$

where $z\left(\tau\right)$ denotes the wage income in the case of employment and the unemployment benefit in the case of unemployment, respectively. In a slight abuse of the word, we call z the "labor" income. As we do not know this future "labor" income and therefore consumption levels, such a constraint is relatively useless. We should rather think of an IBC in expected terms. As expectations need only be formed about the point in time T when the individual accepts a job, this yields

$$E^T\int_t^T e^{-\rho[\tau-t]}c^u\left(\tau\right)d\tau + E^T\int_T^\infty e^{-\rho[\tau-t]}c^e d\tau = a_t + E^T\int_t^T e^{-\rho[\tau-t]}b d\tau + E^T\int_T^\infty e^{-\rho[\tau-t]}w d\tau. \tag{8.12}$$

Clearly, we do not have sufficient information about $c^u\left(\tau\right)$ to compute the first integral.[10] We can consider the IBC in the limit, however, where consumption of the unemployed is zero. In this case, after some steps which are described in Appendix 8.A,

$$\frac{\alpha}{r}c^e = a^{\min}(\alpha + r) + b + \frac{\alpha}{r}w. \tag{8.13}$$

where $a^{\min}$ is the lowest possible wealth level, i.e. $-(a^{\min})$ is the highest debt level an unemployed worker is ever able to borrow. Solving for $a_{\min}$,

$$a^{\min} = \frac{\alpha}{r}\frac{c^e - w}{\alpha + r} - \frac{b}{\alpha + r}, \tag{8.14}$$

and inserting the optimal consumption of the employed worker, $c^e = ra + w$ at $a = a^{\min}$, we get

$$a^{\min} = -\frac{b}{r}. \tag{8.15}$$

The present value of an infinite stream of unemployment benefits is the largest level of debt that can be repaid with probability one. This implies that the maximum debt which is allowed under an IBC at $c^u = 0$, $a^{\min}$, is smaller than the debt level (8.10) under the linear consumption rule, $-b/r > -\delta_1/r$. Hence, under a linear consumption rule, wealth of the unemployed with positive probability becomes

the state where the only "constraint" on the maximization problem, consumption (in our case) should simply be set to infinity (the typical no-Ponzi game condition in fact limits the growth rate of debt to r). Imposing a transversality condition to rule out Ponzi schemes and solving the dynamic budget constraint gives the intertemporal constraint.

[10] This is simple in setups where we have a closed form solution for consumption. See, e.g., Blanchard (1985).

so small (debt grows so high) that the IBC is violated. The individual should never have been able to get so much debt in the first place.

As a thought experiment, suppose the unemployed individual were able to write a binding contract that limits her consumption to zero, even when she finds a job. Intuitively, she should be able to borrow funds equal to the present value of unemployment benefits for the expected duration of unemployment and wage income thereafter. Using $c^e = 0$ in equation (8.14), where c^e is constant due to $r = \rho$ and $s = 0$, confirms

$$a^{\min,0} = -\left(\frac{b}{\alpha + r} + \frac{\alpha}{\alpha + r}\frac{w}{r}\right). \tag{8.16}$$

As $-\delta_1/r$ is not necessarily smaller than $a^{\min,0}$, the linear solution may not violate the IBC and the unemployed could actually end up in a situation with zero consumption. As this occurs at a wealth level lower than $-b/r$ where $\dot{a} < 0$, the unemployed individual, however, will continue to run into debt, and eventually hit $a^{\min,0}$. We therefore know that in either case, the IBC is eventually violated: $-a$ exceeds any real number with positive probability.

8.3.3 Optimal Consumption

We now analyze the evolution of consumption and wealth generally without imposing a linear solution. We first compute the evolution of optimal consumption. The resulting consumption Euler equation, which is sometimes called "Keynes-Ramsey-rule" (KRR), was derived in Section 6.3. Here, we apply the result for the CARA case (cf. Section 6.3.2):

$$dc^u = \left\{\frac{r - \rho}{\gamma} - \frac{\alpha}{\gamma}\left[1 - e^{-\gamma(c^e - c^u)}\right]\right\} dt + (c^e - c^u) dq_\alpha, \tag{8.17}$$

where dq_α is the increment of the Poisson process that counts the switch from unemployment to employment, i.e. from 0 where the individual is unemployed to 1 when she finds a job. Chapter 6 provided a detailed interpretation of this consumption Euler equation so that we restrict ourselves to a short description here. As long as the unemployed individual does not change her employment status, $dq_\alpha = 0$, consumption follows a deterministic rule (and we can write $dc^u/dt \equiv \dot{c}^u$):

$$\dot{c}^u = \frac{r - \rho - \alpha\left[1 - e^{-\gamma(c^e - c^u)}\right]}{\gamma}.$$

The unemployed individual spends more than she would in the absence of the chance of finding a job so that consumption grows more slowly than in the standard deterministic setting (where $\alpha = 0$). If the individual accepts a job, her consumption jumps discretely to c^e.

Returning to the model, we note that the (stationary) consumption level of the employed remains given by (8.4), i.e. $c^e(a) = ra^e + w$ (neither the first-order condition nor the value function of the employed changed due to the lower bound on wealth, and we impose $a \geq -b/r$ to ensure repayment with probability one so that $c^u > 0$ as $w > b$). Hence, with $r = \rho$ and while not receiving a job offer, (8.17) can be written as

$$\dot{c}^u = -\frac{\alpha}{\gamma}\left[1 - e^{-\gamma(ra+w-c^u)}\right],$$

As finding a job can occur at each instant, the wealth level in the employed's consumption $c^e(a) = ra + w$ used in the unemployed's Euler equation above is the current wealth level as described in (8.2). We therefore characterize the optimal behavior of an unemployed individual by a two-dimensional differential equation system,

$$\begin{aligned} \dot{c}^u &= -\frac{\alpha}{\gamma}\left[1 - e^{-\gamma(ra+w-c^u)}\right], \\ \dot{a} &= ra + \bar{b} - c^u. \end{aligned}$$

Identifying the equilibrium trajectory in a phase diagram requires two boundary values. The first is given by the initial wealth level a_0^u. The second is initial consumption, c_0^u, which remains to be identified.

Properties of the System

This system can easily be analyzed by a phase diagram. The zero motion line for optimal consumption during the unemployment spell is given by

$$\zeta_c(a) \equiv ra + w.$$

It has a constant positive slope $\zeta_c' = r = \rho$, $\zeta_c(0) = w$, and $\zeta_c(-w/r) = 0$. Similarly, a is constant on

$$\zeta_a(a) \equiv ra + \bar{b},$$

which is parallel to ζ_c with $\zeta_a(0) = \bar{b}$ $(< w)$. The dynamics of the system are illustrated in Figure 8.1. Above $\zeta_c(a)$, consumption goes to infinity while wealth converges to minus infinity. Laterally reversed, consumption goes to minus infinity while a approaches infinity below $\zeta_a(a)$. Evidently, neither case can be part of a feasible optimal behavior so that the optimal path has to be included between both zero motion lines. This verifies Corollary 6.1 (which also holds under CARA preferences and without job separation), viz. that the consumption good is normal.

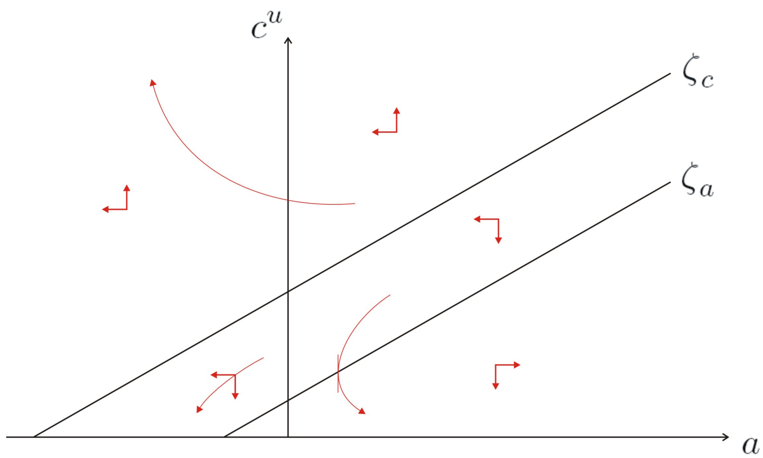


Figure 8.1: Wealth and Consumption Dynamics during Unemployment

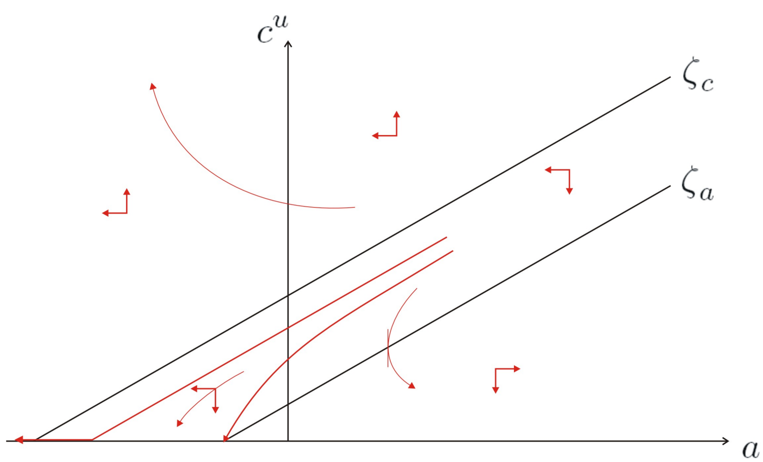


Figure 8.2: Linear and IBC Consistent Consumption Paths

Figure 8.2 adds the Shimer-Werning (2008) solution. The linear consumption rule hits the horizontal axis at $-\delta_1/r$. If we impose nonnegative consumption, c^u stays zero from then on and wealth falls further. We know from (8.15) that the minimum wealth level an individual is allowed to have (if IBC is to hold) is $-b/r$. We thus conclude that the optimal consumption path consistent with the IBC must be as drawn. Consumption starts on this path from whatever wealth level the unemployed has, approaches a zero-consumption level (in finite time if she does not leave unemployment) and stays at $(c, a) = (0, -b/r)$ until the unemployed finds a job. The optimal path can easily be solved numerically.

8.4 Wage Distribution and Reservation Wage

With this background, we now include a wage distribution. This section analyzes the model as presented by Shimer and Werning (2008, cf. Chapter 7) and provides the benchmark for the solution with a borrowing constraint.[11]

8.4.1 The Closed-Form Solution without a Borrowing Constraint

The Consumption Rule

Allowing for wage offers from $[0, \infty)$, the optimal program involves a reservation wage $\bar{w}$, which is defined by the unemployed individual being indifferent between accepting a job that pays the reservation wage $\bar{w}$ or remaining unemployed:[12]

$$V^e(a, \bar{w}) = V^u(a). \tag{8.18}$$

Hence, $\bar{w}$ is potentially a function of the assets if V^e and V^u have different slopes. The value function for the unemployed reads

$$rV^u(a) = \max_c \left\{ u(c) + V^{u\prime}(a) \left[ra + \bar{b} - c \right] + \alpha \int_0^\infty \max \left[V^e(a, w) - V^u(a), 0 \right] dF(w) \right\}. \tag{8.19}$$

As usual, $dF(w) = f(w)\,dw$. A new job arrives with Poisson arrival rate α, and only jobs with wages that yield higher discounted utility, $V^e(a, w) > V^u(a)$, are accepted. We now provide a similar proposition to the ones above without wage distribution, namely the solution obtained in Shimer and Werning (2008).

Proposition 8.3 (**Shimer and Werning, 2008**). *Optimal consumption levels and value functions for employed and unemployed workers with CARA preferences are given by*

$$c^e(a, w) = ra + w, \ V^e(a) = \frac{u(ra + w)}{r}. \tag{8.20}$$

$$c^u(a) = ra + \bar{w}, \ V^u(a) = \frac{u(ra + \bar{w})}{r}. \tag{8.21}$$

The reservation wage is determined by

$$[\bar{w} - b]\, \gamma = \frac{\alpha}{r} \int_{\bar{w}}^\infty [1 + u(w - \bar{w})] dF(w). \tag{8.22}$$

[11]This section proves the statements on the optimal consumption rule and the value function of the unemployed in Chapter 7.

[12]We verify from the resulting solution for the value functions that $V^e(a, \bar{w})$ and $V^u(a)$ define a unique reservation wage.

Proof. The Bellman equation and first order condition of an employed individual do not change when we include a wage distribution. This proves the first part of the proposition. We adopt the Shimer-Werning guess for the unemployed:

$$c_g^u(a) = \kappa_0 a + \kappa_1,\ V_g^u(a) = \kappa_2 u(c_g^u(a)),$$

where the κ_i's are constants. The first order condition of the unemployed is not affected by her reservation wage choice either: $u'(c) = V^{u\prime}(a)$, from (8.19). We have three equations to determine the three constants κ_i. The first order condition requires again $\kappa_0\kappa_2 = 1$. From the definition of the reservation wage and the maximized HJB we get, directly using $\kappa_0\kappa_2 = 1$,

$$-\frac{e^{-\gamma(ra+\bar{w})}}{r} = -\kappa_2 e^{-\gamma(\kappa_0 a+\kappa_1)}, \tag{8.23}$$

and

$$\begin{aligned} -r\kappa_2 e^{-\gamma(\kappa_0 a+\kappa_1)} &= -e^{-\gamma(\kappa_0 a+\kappa_1)} + \gamma e^{-\gamma(\kappa_0 a+\kappa_1)}\left[ra + b - \kappa_0 a - \kappa_1\right] + \\ &\quad +\alpha \int_0^\infty \max\left[-\frac{e^{-\gamma(ra+w)}}{r} + \kappa_2 e^{-\gamma(\kappa_0 a+\kappa_1)}, 0\right] dF(w). \end{aligned} \tag{8.24}$$

Following Shimer and Werning (2008), suppose $\kappa_2 = 1/r$. Thus, $\kappa_0 = r$ and from (8.23) $\kappa_1 = \bar{w}$. Using these expressions in (8.24), we find

$$\left[\bar{w} - b\right]\gamma e^{-\gamma(ra+\bar{w})} = \frac{\alpha}{r}\int_0^\infty \max\left[e^{-\gamma(ra+\bar{w})} - e^{-\gamma(ra+w)}, 0\right] dF(w).$$

Note that integrating from 0 to ∞ using the max operator and an arrival rate α is identical to integrating from $\bar{w}$ to ∞, dropping the max operator and integrating over the truncated (conditional on $w \geq \bar{w}$) distribution $F(w)/[1-F(\bar{w})]$ while using the effective job arrival rate $\alpha[1-F(\bar{w})]$. Dividing by $u(c_g^u) = -e^{-\gamma(ra+\bar{w})}$ we arrive at (7.4) (equation (9) in Shimer and Werning, 2008):

$$\left[\bar{w} - b\right]\gamma = \frac{\alpha}{r}\int_{\bar{w}}^\infty \left[1 + u(w - \bar{w})\right] dF(w). \tag{8.25}$$

□

Equation (8.25) shows the beauty of the Shimer-Werning result. What seemed to be a more or less technical condition in (8.8) which determined δ_1, i.e. the part out of labor income which is used for consumption, is here an equation that determines the reservation wage. In fact, multiplying (8.25) by -1 and inserting the utility function shows that δ_1 in (8.8) corresponds to the reservation wage $\bar{w}$ in (8.25). Note that the reservation wage is determined independently of the level of financial wealth. This is the standard result obtained for CARA preferences (cf. Acemoglu and Shimer, 1999).

The Dynamics of Savings

We now analyze again how savings of the unemployed evolve. The budget constraint of the unemployed worker does not change compared to a situation without wage distribution. Current income is b in both cases. Optimal consumption, however, is now given by (8.21) and inserting this into the budget constraint (8.2) yields

$$\dot{a}^u = ra^u + \bar{b} - c^u = \bar{b} - \bar{w}. \tag{8.26}$$

With a wage distribution, the reservation wage is higher than unemployment benefits, $\bar{w} > \bar{b}$, and since both are constant, wealth decreases linearly. In finite time, consumption of the unemployed becomes negative as well, as soon as the individual's wealth hits

$$a_{neg}^{u,\, distri} = -\bar{w}/r. \tag{8.27}$$

Wealth continues to fall according to (8.26). Repayment cannot occur with probability one.

Let us see whether an ad hoc restriction on consumption solves the problem. That is, let us impose nonnegativity of consumption. In this case, consumption is zero when the debt level hits $a_{neg}^{u,\, distri}$ as defined in (8.27). Following the same steps as before, we find that wealth continues to fall according to (8.26) (with consumption remaining constant at zero). Hence, the linear solution again implies credit default with positive probability.

8.4.2 The Intertemporal Budget Constraint

The present value of unemployment benefits is the highest debt level that an unemployed can repay with probability one. We prove this statement following the same steps as in Section 8.3.2 and derive the intertemporal budget constraint. We then check whether the linear solution violates the IBC. Finally, we analyze the general dynamics again and show what the optimal consumption path looks like.

The intertemporal budget constraint in realizations with a wage distribution is identical to (8.11). Forming expectations, however, now requires forming expectations about w as well, i.e. in contrast to (8.12) we have an additional expectations operator $E^{T,w}$:

$$E^T \int_t^T e^{-\rho[\tau-t]} c^u(\tau)\, d\tau + E^{T,w} \int_T^\infty e^{-\rho[\tau-t]} c^e d\tau = a_t + E^T \int_t^T e^{-\rho[\tau-t]} b d\tau + E^{T,w} \int_T^\infty e^{-\rho[\tau-t]} w d\tau.$$

Note that we assume that the draw from the wage distribution is independent of the draw of the length of unemployment in the derivation above. Put differently, independently of when the worker accepts a job, she always faces the same wage distribution. This allows us to rewrite $E^{T,w}\int_T^\infty e^{-\rho[\tau-t]}wd\tau$ as $E^T\int_T^\infty e^{-\rho[\tau-t]}E^w wd\tau$. We also need to do the same for consumption of the employed worker, i.e. $E^{T,w}\int_T^\infty e^{-\rho[\tau-t]}c^e d\tau = E^T\int_T^\infty e^{-\rho[\tau-t]}E^w c^e d\tau$. With these expressions at hand, we proceed just like in the previous paragraph, compare (8.13) and Appendix 8.A, and find

$$\frac{\alpha}{r}c^e = a^{\min}[\alpha + r] + b + \frac{\alpha}{r}E^w w. \tag{8.28}$$

Inserting the consumption level of the employed worker from (8.4), $c^e(a) = ra^{\min} + E^w w$, yields

$$a^{\min} = -\frac{b}{r}.$$

8.4.3 Optimal Consumption

We now analyze the evolution of consumption and wealth generally without imposing a linear solution. We already know that accounting for the borrowing constraint rules out a linear solution. Also, we know from the flow budget constraint for the unemployed that $a^{\min}$ with $\dot{a} = 0$ corresponds to zero consumption. To characterize the optimal path, we begin by computing the evolution of optimal consumption including a distribution of wage offers.

Properties of the Reservation Wage

Consider again the HJB equation for the unemployed individual when she faces wage draws from a distribution:

$$\rho V^u(a) = \max_{c^u}\left\{u(c^u) + [ra + b - c^u]V^{u\prime}(a) + \alpha\int_0^\infty \max\{0, V^e(a,w) - V^u(a)\}dF(w)\right\}.$$

In this program, wealth has no direct effect on the expected value of $V^e - V^u$ via the reservation wage. To see this, note that for a given level of wealth, the third summand in this HJB equation can be rewritten as[13]

$$\alpha\int_0^\infty \max\{0, V^e(a,w) - V^u(a)\}dFw = \alpha\int_{\bar{w}(a)}^\infty [V^e(a,w) - V^u(a)]dF(w). \tag{8.29}$$

Intuitively, reservation wages on the one hand prolong the expected duration of unemployment by lowering the arrival rate of an acceptable job from α (without a binding reservation wage) to

[13] $\alpha\int_0^\infty \max\{0, V^e(a,w) - V^u(a)\}dF(w) =$
$\alpha\left[\int_0^{\bar{w}(a)} \max\{0, V^e(a,w) - V^u(a)\}dF(w) + \int_{\bar{w}(a)}^\infty \max\{0, V^e(a,w) - V^u(a)\}dF(w)\right]$ and using the definition of $\bar{w}$ yields the expression above.

$\alpha[1-F(\bar{w}(a))]$. On the other hand, the density of acceptable wages adjusts from $f(w)$ (without a binding reservation wage) to $f(w)/[1-F(\bar{w}(a))]$, implying a higher expected wage. The change in the distribution of wages above $\bar{w}$ and the lower arrival rate cancel. Accordingly, at the margin, a change in a has no direct effect via the reservation wage. Taking the derivative with respect to a verifies that the direct impact via $\bar{w}(a)$ drops out.[14]

As argued above, the optimal reservation wage is potentially a function of the level of wealth. We say "potentially" since we already know that if the consumption function for the unemployed is linear in wealth and no borrowing constraint is taken into account, both value functions are linear in wealth and a cancels so that $\bar{w}$ is in fact independent of a (cf. Chapter 7 or (8.22) above).

In the following, the situation of the employed remains unchanged so that her consumption is constant but we look for a solution that satisfies the borrowing constraint in the case of unemployment (which is non-linear). Returning to the reservation wage and using the closed-form solution for the maximized value function of the employed, $V^e = u(ra+w)/r$, the optimal reservation wage is characterized by

$$rV^u(a) \equiv u(c^e(a,\bar{w}(a))), \tag{8.30}$$

where we made explicit the (potential) dependency of the reservation wage on wealth. Taking the derivative with respect to a reveals

$$rV^{u\prime}(a) = u'(ra+\bar{w}(a))\left[r+\bar{w}'(a)\right].$$

The terms in squared brackets on the right hand side reflect the direct impact of wealth on consumption and the indirect effect via the change in the reservation wage, $\partial c/\partial a + (\partial c/\partial \bar{w})(\partial \bar{w}/\partial a) = r + w'(a)$. In the optimum, it holds that $V^{u\prime}(a) = u'(c^u)$ from the first-order condition for optimal consumption of the unemployed. Substituting for $V^{u\prime}(a)$ and rearranging proves the following

Lemma 8.1 ("optimal reservation wage"). *The optimal reservation wage satisifies*

$$\frac{u'(c^u(a))}{u'(c^e(a,\bar{w}(a)))} = 1 + \frac{\bar{w}'(a)}{r}, \tag{8.31}$$

where $c^e(a)$ is the optimal consumption of the employed, $c^e = ra + w$ with $w \geq \bar{w}(a)$, and $c^u(a)$ is the optimal consumption during the unemployment spell. Hence, $c^u(a) <,=,> c^e(a,\bar{w}(a))$ implies $\bar{w}'(a) >,=,< 0$.

[14] $\frac{\partial\left\{\alpha\int_{\bar{w}(a)}^{\infty}[V^e(a,w)-V^u(a)]dF(w)\right\}}{\partial a} = -\alpha[V^e(a,\bar{w}(a))-V^u(a)]\frac{\partial \bar{w}(a)}{\partial a} + \alpha\int_{\bar{w}(a)}^{\infty}[V^{e\prime}(a,w)-V^{u\prime}(a)]dF(w) = \alpha\int_{\bar{w}(a)}^{\infty}[V^{e\prime}(a,w)-V^{u\prime}(a)]dF(w)$.

We infer from Lemma 8.1 that imposing a linear consumption function (that abstracts from borrowing constraints/nonnegativity restrictions on consumption) rules out the wealth dependency of the reservation wage along the optimal path (cf. (8.25)). Imposing a linear solution implies that consumption of the unemployed equates marginal utility during the unemployment spell and the marginal utility obtained when working on a job that pays the (wealth-independent) reservation wage from (8.25). In what follows, we use stars (*) to indicate variables obtained under the linear consumption rule. Equation (8.31) shows again that a constant reservation wage is in fact optimal if $c^{u\star}(a) = ra + \bar{w}^\star = c^e(\bar{w}^\star, a)$. It follows that consumption smoothing is not perfect at the point in time where the individual accepts a job that pays strictly more than her reservation wage. Under the solution obtained by Shimer and Werning (2008), consumption jumps upward from $c^u(a)^\star = ra + \bar{w}^\star$ to $c^e(a) = ra + w$ with $w \geq \bar{w}^\star$. If the reservation wage is indeed a function of wealth, the change in $\bar{w}(a)$ drives a wedge between the relative marginal utilities as wealth changes.

Evolution of Optimal Consumption with Reservation Wage

Let us begin by deriving the evolution of optimal consumption. Using (8.29), the unemployed's Bellman equation can be written as

$$\rho V^u(a) = \max_{c^u} \left\{ u(c^u) + [ra + b - c^u] V^{u\prime}(a) + \alpha \int_{\bar{w}(a)}^{\infty} [V^e(a,w) - V^u(a)]\, dF(w) \right\}.$$

Taking the derivative with respect to c^u on the right hand side, the first-order condition for optimal consumption is given by $u'(c^u) = V^{u\prime}(a)$ so that again $c^u = c^u(a)$. Plugging this solution into the HJB, the maximized Bellman equation reads

$$\rho V^u(a) = u(c^u(a)) + [ra + b - c^u(a)] V^{u\prime}(a) + \alpha \int_{\bar{w}(a)}^{\infty} [V^e(a,w) - V^u(a)]\, dF(w). \tag{8.32}$$

Now, taking the derivative with respect to a, using the first-order condition (i.e., the envelope theorem) and noting that $-\alpha [V^e(a, \bar{w}(a)) - V^u(a)] \bar{w}'(a) = 0$ (see (8.18)), we find

$$\rho V^{u\prime}(a) = rV^{u\prime}(a) + [ra + b - c^u(a)] V^{u\prime\prime}(a) + \alpha \int_{\bar{w}(a)}^{\infty} \left[V^{e\prime}(a,w) - V^{u\prime}(a)\right] dF(w). \tag{8.33}$$

We have to find an expression for the evolution of the costate variable to substitute for $V^{u\prime\prime}(a)$. To get there, we "calculate" $dV^{u\prime}$ using the Change of Variable Formula (the CVF was introduced in Section 6.2.2; cf. Øksendahl, 2003, Sennewald, 2007a,b, and Wälde, 2008):

$$dV^{u\prime}(a) = V^{u\prime\prime}(a)\, da^u + \left[\int_{\bar{w}(a)}^{\infty} \left[V^{e\prime}(a,w) - V^{u\prime}(a)\right] dF(w) \right] dq_\alpha,$$

where dq_α is the increment of the Poisson process and again the arrival rate of acceptable jobs and the conditioning of $f(w)$ on $w \geq \bar{w}(a)$ drop out.[15] Therefore, inserting $da^u = [ra + b - c^u]\, dt$ and rearranging, we have

$$[ra + b - c^u]\, V^{u\prime\prime}(a)\, dt = dV^{u\prime}(a) - \left[\int_{\bar{w}(a)}^{\infty} \left[V^{e\prime}(a, w) - V^{u\prime}(a)\right] dF(w)\right] dq_\alpha.$$

This expression can be used to substitute for $[ra + b - c^u]\, V^{u\prime\prime}(a)$ in (8.33):

$$dV^{u\prime}(a) =$$
$$\left[(\rho - r)\, V^{u\prime}(a) - \alpha \textstyle\int_{\bar{w}(a)}^{\infty} [V^{e\prime}(a, w) - V^{u\prime}(a)]\, dF(w)\right] dt + \left[\textstyle\int_{\bar{w}(a)}^{\infty} [V^{e\prime}(a, w) - V^{u\prime}(a)]\, dF(w)\right] dq_\alpha.$$

If the individual does not change her employment state, and $r = \rho$ like in Shimer and Werning (2008), we find

$$dV^{u\prime}(a) = -\alpha \int_{\bar{w}(a)}^{\infty} \left[V^{e\prime}(a, w) - V^{u\prime}(a)\right] dF(w)\, dt.$$

While the reservation wage depends on the individual's wealth, relative to the case without a wage distribution/reservation wage, only the expected wage as a worker needs to be taken into account. Accordingly, we can substitute for $V^{e\prime}$ and $V^{u\prime}$ from the first-order conditions in both employment states to derive the evolution of marginal utility (where we omit all arguments for notational convenience):

$$du'(c^u) = -\alpha \int_{\bar{w}}^{\infty} \left[u'(c^e) - u'(c^u)\right] dF(w)\, dt. \tag{8.34}$$

To arrive at the evolution of optimal consumption, let $\upsilon \equiv (u'(c^u))^{-1}$ such that $d\upsilon = dc^u$ and $\upsilon'(\cdot) = u''(c^u(a))^{-1}$. Applying the CVF to υ gives

$$d\upsilon((u'(c)) = \upsilon'(u'(c))du'(c^z(a))dt + \left[\upsilon\left(u'(c^u)\right) - \upsilon\left(u'(c^e)\right)\right] dq_\alpha.$$

After substituting for $d\upsilon = dc^u$, $\upsilon' = u''(c^u(a))^{-1}$, and $du'(c^u)$ with (8.34), the change in consumption during an ongoing unemployment spell (so that $dq_\alpha = 0$) obeys

$$\dot{c}^u = -\frac{\alpha}{u''(c^u(a))} \int_{\bar{w}(a)}^{\infty} \left[u'(c^e) - u'(c^u)\right] dF(w).$$

[15] As noted above, if one feels more comfortable adjusting the arrival rate to the actual change in the new employment state to $\alpha(1 - F(\bar{w}(a)))$, the change in the distribution from $f(w)$ to $f(w)/(1 - F(\bar{w}(a)))\, dw$ has to be taken into account.

Under CARA, $u''(c) = -\gamma u'(c)$ so that along the optimal path, consumption evolves according to[16]

$$\dot{c}^u = \frac{\alpha}{\gamma} \int_{\bar{w}(a)}^{\infty} \left[\frac{u'(c^e) - u'(c^u)}{u'(c^u)} \right] dF(w) . \tag{8.35}$$

Using $u'(c^e)/u'(c^u) = -u(c^e - c^u)$ we have thus shown

Lemma 8.2 ("evolution of optimal consumption with reservation wage"). *Under CARA, $r = \rho$, and no job separation, the optimal consumption of the unemployed satisfies*

$$\dot{c}^u = -\frac{\alpha}{\gamma} \int_{\bar{w}(a)}^{\infty} [u(c^e(a,w) - c^u(a)) + 1] \, dF(w) , \tag{8.36}$$

where $\bar{w}$ is characterized in (8.31).

Note that allowing for a wage distribution and a reservation wage does not add another state variable. We thus again derive a phase diagram in (c^u, a)-space to characterize optimal consumption and assess the reservation by via the optimality conditions.

Properties of the Dynamic System

According to Lemma 8.2, consumption is constant if and only if

$$\int_{\bar{w}(a)}^{\infty} \left[\frac{u'(c^e(a,w))}{u'(c_0^u(a))} - 1 \right] dF(w) = 0,$$

where $c_0^u(a)$ denotes the zero growth locus of c^u. Pulling the constant out of the integral and rearranging yields

$$u'(c_0^u(a)) = \frac{\int_{\bar{w}(a)}^{\infty} u'(c^e(a,w)) \, dF(w)}{1 - F(\bar{w}(a))} . \tag{8.37}$$

We proceed in two steps. First, we allow for wage offers that are drawn from a (non-degenerated) distribution, but impose an exogenous minimum wage that has to be accepted by the unemployed. We set this threshold equal to the reservation wage $\bar{w}^{\star}$ under Shimer and Werning's (2008) solution. In the second step, we consider the optimal consumption path and sequence of reservation wages if the individual is free to choose her reservation wage.

[16] Using CARA explicitly, i.e. applying $u'(c) = \gamma e^{-\gamma c}$ and $u''(c) = \gamma^2 e^{-\gamma c}$, this equals

$$\dot{c}^u = -\frac{\alpha}{\gamma e^{-\gamma c^u}} \int_{\bar{w}(a)}^{\infty} \left[e^{-\gamma c^u} - e^{-\gamma c^e} \right] dF(w) .$$

Optimal Consumption Given a Minimum Wage $\bar{\mathbf{w}}^{\star}$

Replacing $\bar{w}(a)$ with $\bar{w}^{\star}$, consumption of the unemployed is constant along $c_u^{0\star}$ which is determined by

$$u'\left(c_0^{u\star}(a) - ra\right) = \frac{\int_{\bar{w}^{\star}}^{\infty} u'(w)\, dF(w)}{1 - F(\bar{w}^{\star})}.$$

The right hand side is a constant, and we denote it by $\omega^{\star}(\bar{w}^{\star})$. Accordingly, solving for $c_0^{u\star}$ yields

$$c_0^{u\star}(a) = ra + (u')^{-1}(\omega^{\star}).$$

This lies above $c^{u\star}$ $(w^{\star} < u'^{-1}(\omega^{\star}))$:

$$u'\left(c_0^{u\star}(a) - ra\right) = \frac{\int_{\bar{w}^{\star}}^{\infty} u'(w)\, dF(w)}{1 - F(\bar{w}^{\star})} < \frac{\int_{\bar{w}^{\star}}^{\infty} u'(\bar{w}^{\star})\, dF(w)}{1 - F(\bar{w}^{\star})} = u'(\bar{w}^{\star}), \tag{8.38}$$

and $u''(\cdot) < 0$ implies $c_0^{u\star}(a) > ra + \bar{w}^{\star}$. Consumption is falling below $c_0^{u\star}$ and increasing above $c_0^{u\star}$. We thus find the same global dynamics as in the case without a wage distribution, see Figure 8.3. We also include again the linear Shimer-Werning solution (recall that a falls above $ra + b$ and increases below). As was shown earlier, the linear solution hits the wealth axis at a debt level $\bar{w}^{\star}/r$ to the left of $a_{\min}$, thereby violating the IBC. In the presence of a borrowing constraint at $a_{\min}$ and a fixed reservation wage $\bar{w}^{\star}$, the best the individual can do is to pick the consumption path that goes through $c^u = 0$ at $a = a_{\min}$, just like in the case without a wage distribution. The Shimer-Werning solution would satisfy an optimal reservation wage choice (cf. (8.31)), but eventually violate the IBC.

Since optimal consumption satisfies $c^u(a) < c^e(\bar{w}(a))$, Lemma 8.1 implies that it is optimal to choose $\bar{w}'(a) > 0$. Keeping in mind that the unemployed would like to decrease the minimum wage along the non-linear consumption path as wealth declines, we turn to the general solution where the individual is free to chose her reservation wage.

The General Solution

Roughly speaking, earlier research (see e.g. the authoritative work of Acemoglu and Shimer, 1999) has concluded that an individual's wealth does not prolong an individual's unemployment spell if under CARA. In what follows, we show that this conclusion is incomplete.

Suppose that $\bar{w}'(a) = 0$ for all levels of wealth (by all levels of wealth, here and in what follows, we refer to all admissible levels of wealth $a \geq a_{\min}$). Then, (8.31) implies that consumption is linear,

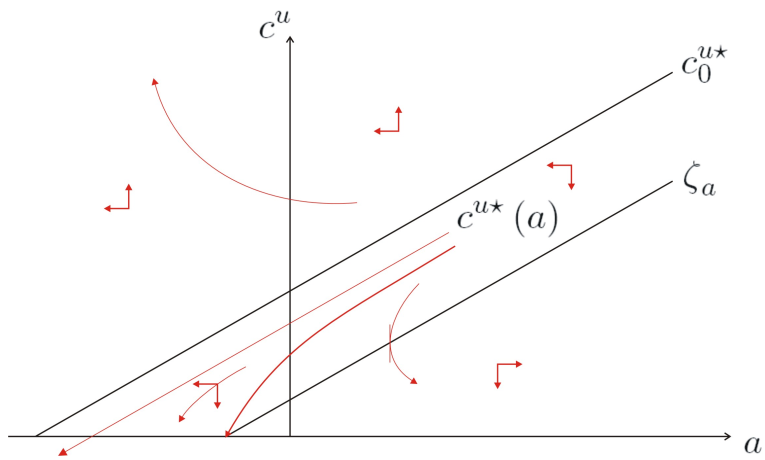


Figure 8.3: Global Dynamics with a Fixed Minimum Wage Equal to $\bar{w}^{\star}(a)$

i.e. $c^u = ra + \bar{w}^{\diamond}$ for some constant reservation wage $\bar{w}^{\diamond}$. With the IBC, from $a \geq a_{\min} = -b/r$ and $c^u(-b/r) = 0$, the constant reservation wage would have to satisfy $\bar{w}^{\diamond} = b$. If $c^u = ra + b$, however, $\dot{a} = 0$ and $\dot{c}^u = 0$ as well. Then, (8.37) implies

$$u'(b) < \frac{\int_b^{\infty} u'(w)\, dF(w)}{1 - F(b)} \tag{8.39}$$

which, at $a > a_{\min}$, violates the differential equation for optimal consumption in Lemma 8.2 (which requires equality in (8.39) if c^u is to be constant). Hence, the only globally constant reservation wage consistent with the IBC violates optimality.

Consider briefly a second thought experiment. Suppose $\bar{w}'(a) < 0$ for all levels of wealth. An immediate implication is that, from (8.31), $c^u > ra + \bar{w}(a)$. This implies that at $a_{\min}$ and zero consumption, $\bar{w}(a_{\min}) < b$. Let alone optimality, by accepting a job with reservation wage $\bar{w} < b$ the worker would not even be able to repay her loan.

Intuitively, if unlimited debt is possible, consumption is set to the highest level that maintains $\dot{c}^u < 0$ if $\dot{a} < 0$ (so that consumption is normal), viz. the Shimer-Werning solution $c^{\star}(u) = ra + \bar{w}^{\star}$.

We proceed to show that if a borrowing constraint is taken into account, wealth has a positive impact on the reservation wage even if preferences exhibit CARA. We achieve this in several steps. First, we show that $c(a_{\min}) = 0$ and $c_0^u(a_{\min}) > 0$ whereas $\zeta_a(a_{\min}) = 0$. This gives, by continuity of c_0^u and from the instability of the $\dot{a} = 0$ and the $\dot{c} = 0$ loci, that $c(a) > ra + b$ and $\dot{c} < 0$ for wealth levels slightly above $a_{\min}$. We then infer from (8.39) that $c^u(a) \neq ra + b$ at any level of wealth $a > a_{\min}$, since this contradicts the law of motion for optimal consumption (8.36). Accordingly, $c_0^u(a) > ra+b = \zeta_a(a)$ for all levels of wealth (and $c_0^u > ra+b$ for all a). Finally, we show that $c^u(a) < c^e(a, \bar{w}(a))$ such that Lemma 8.1 implies $\bar{w}'(a) > 0$ for all levels of wealth $a > a_{\min}$. In passing, we describe the evolution of consumption during the unemployment spell.

Lemma 8.3. *For wealth levels slightly above $a_{\min}$, $c^u(a) > ra + b$, $\dot{c}^u(a) < 0$, and $\dot{a} < 0$ since $c^u(a_{\min}) > \zeta_a(a_{\min}) = 0$.*

Proof. $c^u(a_{\min}) = 0$ since $c^u(a) \geq 0$ and $c^u(a_{\min}) > 0$ implies $\dot{a} < 0$ (since $c^u(a_{\min} > 0$ lies to the left of $\zeta_a(a_{\min}) = 0$), a violation of the IBC. Using $c^u(a_{\min}) = 0$ and Lemma 8.1 yields

$$e^{\gamma(\bar{w}(a_{\min})-b)} = 1 + \frac{\bar{w}'(a_{\min})}{r},$$

so that $\bar{w}(a_{\min}) >,=,< b$ if and only if $\bar{w}'(a_{\min}) >,=,< 0$. $\bar{w}(a_{\min}) < b$ can be sorted out since it violates the IBC (if an individual accepts this low reservation wage, she cannot repay her debt). Hence, $\bar{w}(a_{\min}) \geq b$. By assumption, $f(w) > 0$ for at least one $w > \bar{w}(a_{\min})$, hence strict concavity of $u(c)$ and (8.37) imply

$$\frac{\int_{\bar{w}(a_{\min})}^{\infty} u'(\bar{w}(a_{\min}) - b)\, dF(w)}{1 - F(\bar{w}(a_{\min}))} > \frac{\int_{\bar{w}(a_{\min})}^{\infty} u'(w - b)\, dF(w)}{1 - F(\bar{w}(a_{\min}))} = u'(c_0^u(a_{\min})).$$

Accordingly, $u'(\bar{w}(a_{\min}) - b) > u'(c_0^u(a_{\min}))$ so that, from $u''(c) < 0$ and our earlier result $\bar{w}(a_{\min} \geq b$, $c_0^u(a_{\min}) > \bar{w}(a_{\min}) - b \geq 0$. Hence, $c_0^u(a_{\min}) > \zeta_a(a_{\min})$. Continuity of $c_0^u(a)$ and the instability of $c_0^u(a)$ and $\zeta_a(a)$ complete the proof. □

We have shown in (8.39) that $c^u(a) \neq ra + b$ for $a > a_{\min}$. Together with Lemma 8.3, it follows that $c_0^u(a) > \zeta_a(a)$ for all levels of wealth and hence $\dot{c}^u(a) < 0$ for all $a > a_{\min}$ together with $\dot{a} < 0$. We summarize this finding in the following

Proposition 8.4. *Under CARA with no job separation, $r = \rho$, Poisson job offers from a wage distribution, and a borrowing constraint, optimal consumption and wealth decline strictly over the unemployment spell. If the unemployed does not find a job, consumption and wealth continue to decline until the stationary point $(c, a) = (0, a_{\min})$ is reached.*

We proceed to show that $c^u(a)$ is less than what an individual would consume when she accepts a job that pays exactly the reservation wage $\bar{w}(a)$ if consumption of the unemployed exceeds her instantaneous income. Hence, consumption always jumps upward when an individual with wealth $a \geq a_{\min}$ accepts a job.

Lemma 8.4. *If $c^u(a) > ra + b$, we have*

$$c^e(a, \bar{w}(a)) > c^u(a). \tag{8.40}$$

Proof. Since the $\dot{c}^u = 0$ locus is unstable, and $da < 0$ is implied by $c^u(a) > ra + b$, we have $c^u(a) < c_0^u(a)$. Concavity of the utility function thus yields

$$u'(c^u(a)) > u'(c_0^u(a)) = \frac{\int_{\bar{w}(a)}^{\infty} u'(c^e(a, w))\, dF(w)}{1 - F(\bar{w}(a))}.$$

This implies

$$u'(c^u(a))\,[1 - F(\bar{w}(a))] > \int_{\bar{w}(a)}^{\infty} u'(c^e(a, w))\, dF(w) > \int_{\bar{w}(a)}^{\infty} u'(c^e(\bar{w}(a), a))\, dF(w),$$

or, equivalently, $u'(c^u(a)) > u'(c^e(\bar{w}(a), a))$. Hence, $c^u(a) < c^e(a, \bar{w}(a))$. □

Taken together, Lemma 8.1 and Lemma 8.4 prove

Proposition 8.5. *Under CARA with no job separation, $r = \rho$, Poisson job offers from a wage distribution, and a borrowing constraint, the reservation wage is strictly increasing in wealth, $\bar{w}'(a) > 0$ for $a > a_{\min}$.*

A direct implication of Proposition 8.5 is the following

Corollary 8.1. *Including the borrowing constraint, the reservation wage remains strictly larger than the unemployment benefits at $a > a_{\min}$.*

Proof. The corollary follows from the proof of Lemma 8.3 ($\bar{w}(a_{\min} \geq b)$) and Proposition 8.5. □

In fact, Corollary 8.1 must apply to $a_{\min}$ also. Intuitively, when the unemployed receives benefits b, she cannot be indifferent to accepting a job that pays b and her current situation as long as there is some chance of receiving a better offer than b.

Lemma 8.5. *The reservation wage exceeds the unemployment benefits at all levels of wealth.*

Proof. $\bar{w}'(a) > 0$ for $a > a_{\min}$ is shown in Corollary 8.1. In the case of $a_{\min}$, we know that $\bar{w}(a_{\min}) \geq b$ from the proof of Lemma 8.3. $\bar{w}(a_{\min}) > b$ is shown by contradiction. Suppose $\bar{w}(a_{\min}) = b$. Then, (8.30) implies $rV^u(a_{\min}) = u(0)$. Evaluating the maximized Bellman equation (8.32) at $a = a_{\min}$ using $c^u(a_{\min}) = 0$ (see the proof of Lemma 8.3) and $\rho = r$, we obtain

$$0 = \int_b^{\infty} [V^e(w, a_{\min}) - V^u(a_{\min})]\, dF(w).$$

Substituting for $V^u(a_{\min})$ and $V^e(w, a_{\min})$ using (8.20) yields

$$0 = \int_b^\infty [1 + u(w - b)]\, dF(w),$$

a contradiction if $f(w) > 0$ for any $w > b$ since $u(0) = -1$. □

We conclude that wealth prolongs the duration of unemployment if the borrowing limit is taken into account ($\bar{w}'(a) > 0$ for all a). As the unemployed runs down her assets over the unemployment spell, however, optimal consumption behavior mitigates the moral hazard problem and raises the hazard rate even if risk aversion remains constant. We have thus disentangled the (reinforcing) effects of risk aversion and borrowing limits on the reservation wage.[17]

Let us briefly summarize our findings on the optimal consumption path of the unemployed. The Euler equation implies that smooth changes in consumption are optimal so that the individual avoids jumps within an employment state. From the dynamics of the system and the location of the zero growth loci for consumption and wealth, we deduced that consumption of the unemployed is strictly larger than her instantaneous income if $a > a_{\min}$. Thus, the optimal path involves decreasing wealth and, as implied by the location and instability of the zero growth loci, decreasing consumption as well. From the dynamics of the system, we conclude that the policy function $c^u(a)$ is strictly increasing and concave. The borrowing constraint provides a stationary state at zero consumption and assets $-b/r$. Along the linear consumption path obtained by Shimer and Werning (2008), $\bar{w}'(a)$ is constant (and equal to $\bar{w}^\star$). Accordingly, $c^{u\star}(a)$ gives the consumption choice that leaves the reservation wage unaffected. If consumption is more than linearly decreasing over the unemployment spell, from (8.31) and (8.40), $c^u(a) < c^{u\star}(a)$ equivalently states $w'(a) > 0$. Along the optimal path, $\dot{a} < 0$ and (since $\bar{w}'(a) > 0$) the reservation wage decreases. In contrast to free borrowing, if individuals face a borrowing constraint, the moral hazard problem is mitigated even with CARA preferences.

These findings suggest that the optimality result of constant unemployment benefits and taxes in Shimer and Werning (2008) crucially hinges on the assumption of linear consumption. We leave the analysis of an optimal unemployment policy with nonnegative consumption for future research. Note, however, that the solution in Shimer and Werning (2008) matches well the outcome with optimal unemployment benefits and taxes under CRRA, where the reservation wage is almost constant due to the counteracting effects of falling risk aversion and increasing optimal benefits.

[17] Evidently, this observation crucially depends on an exogenous and homogenous job arrival rate.

8.5 Conclusion

We have studied the optimal consumption behavior of an unemployed individual with CARA preferences in a standard job search model where the only history dependence comes from an individual's level of financial wealth. We included a natural borrowing constraint in the optimal savings problem considered by Shimer and Werning (2008) to maintain nonnegative consumption and repayment with probability one. In doing so, we provided a general solution for the optimal evolution of consumption in the form of a stochastic differential equation. In contrast to Shimer and Werning (2008), we pick a specific solution to this differential equation by imposing a borrowing constraint instead of a linear consumption rule. As a result, the value function no longer obeys a closed form solution. Using simple phase diagrams, however, we establish that the policy function in the case with a borrowing constraint approaches the Shimer-Werning solution from below in (a, c^u)-space, with zero consumption at the maximum debt level. The ensuing optimal path features a declining level of wealth, falling consumption, and a sequence of declining reservation wages. The novel argument is that this occurs even under CARA preferences if consumption is non-linear in wealth. The intuition is that the borrowing constraint forecloses levels of debt implied by a linear solution, but does not affect consumption at high levels of wealth. Accordingly, consumption is non-linear and the reservation wage optimally declines as wealth decreases so as to offset the increase in marginal utility. Our analysis thus disentangled the effects of risk aversion and the borrowing constraint on the reservation wage. Dissecting the effects allows us a better understanding of the main result in Shimer and Werning (2008), namely that the CARA case without borrowing constraints provides a good benchmark for optimal unemployment policy.

Appendix

This appendix contains the derivation of the IBC, numerical solutions to the consumption and wealth paths in the job search model without a reservation wage, and the derivation of the expected cost of a constant benefits and taxes policy if an individual moves back and forth between employment and unemployment.

Appendix 8.A Deriving an Intertemporal Budget Constraint

We can write the IBC in realizations as

$$\int_t^T e^{-\rho(\tau-t)} c^u(\tau)\, d\tau + \int_T^\infty e^{-\rho(\tau-t)} c^e d\tau = a_t + \int_t^T e^{-\rho(\tau-t)} b d\tau + \int_T^\infty e^{-\rho(\tau-t)} w d\tau,$$

where T is the (unknown) point in time where the individual switches from unemployment to employment. If we assume that there is some constant c^u before the switch (in the main text, we consider the limiting case where $c^u = 0$), forming expectations about T gives equation (8.12) in the main text,

$$E^T \int_t^T e^{-\rho(\tau-t)} c^u(\tau)\, d\tau + E^T \int_T^\infty e^{-\rho(\tau-t)} c^e d\tau = a_t + E^T \int_t^T e^{-\rho(\tau-t)} b d\tau + E^T \int_T^\infty e^{-\rho(\tau-t)} w d\tau.$$

Collecting terms, we equivalently have

$$E^T \int_t^T e^{-\rho(\tau-t)} \left[c^u(\tau) - b\right] d\tau + E^T \int_T^\infty e^{-\rho(\tau-t)} \left[c^e - w\right] d\tau = a_t.$$

As we know that T is exponentially distributed, the density is given by $f(T) = \alpha e^{-\alpha[T-t]}$. The expected value is thus given by $E^T g(t,T) = \int_t^\infty g(t,T)\, f(T)\, dT$. If $c^u(\tau)$ is constant, we obtain

$$\begin{aligned} E^T \int_t^T e^{-\rho(\tau-t)} [c^u - b] d\tau &= \int_t^\infty \int_t^T e^{-\rho(\tau-t)} [c^u - b] d\tau f(T) dT \\ &= (c^u - b) \int_t^\infty \left[\int_t^T \alpha e^{-\rho(\tau-t)} d\tau\right] e^{-\alpha[T-t]} dT. \end{aligned}$$

Inserting the solution for the integral in squared brackets, $\alpha\left[\frac{-e^{-\rho(\tau-t)}}{\rho}\right]_{\tau=t}^{\tau=T} = \frac{\alpha}{\rho}\left[-e^{\rho(T-t)}+1\right]$, the right hand side becomes

$$
\begin{aligned}
(c^u-b)\int_t^\infty \frac{\alpha}{\rho}\left[-e^{\rho(T-t)}+1\right]e^{-\alpha[T-t]}dT &= \\
\frac{\alpha(c^u-b)}{\rho}\int_t^\infty \left[1-e^{\rho(T-t)}\right]e^{-\alpha[T-t]}dT &= \\
\frac{\alpha(c^u-b)}{\rho}\left[\int_t^\infty e^{-\alpha(T-t)}dT-\int_t^\infty e^{-(\alpha+\rho)(T-t)}dT\right] &= \\
\frac{\alpha(c^u-b)}{\rho}\left\{\left[\frac{-e^{-\alpha(T-t)}}{\alpha}\right]_{T=t}^{T=\infty}-\left[\frac{-e^{-(\alpha+\rho)(T-t)}}{\alpha+\rho}\right]_{T=t}^{T=\infty}\right\} &= \\
\frac{\alpha(c^u-b)}{\rho}\left\{\frac{1}{\alpha}-\frac{1}{\alpha+\rho}\right\} &= \\
\frac{c^u-b}{\alpha+\rho}&.
\end{aligned}
$$

Similarly, as the wage rate and c^e are constant (the employed's world is stationary and we assume $r=\rho$),

$$
\begin{aligned}
E^T\int_T^\infty e^{-\rho[\tau-t]}(c^e-w)d\tau &= \\
(c^e-w)\int_t^\infty\int_T^\infty e^{-\rho[\tau-t]}d\tau f(T)dT &= \\
(c^e-w)\int_t^\infty\int_T^\infty e^{-\rho[\tau-t]}d\tau\alpha e^{-\alpha[T-t]}dT &= \\
(c^e-w)\frac{\alpha}{\rho}\int_t^\infty e^{-\rho[T-t]}e^{-\alpha[T-t]}dT &= \\
\frac{\alpha}{\rho}\left(\frac{c^e-w}{\alpha+\rho}\right)&.
\end{aligned}
$$

Taken together, the IBC in expectations in the limiting case where c^u is constant, is given by

$$
\frac{c^u-b}{\alpha+\rho}+\frac{\alpha}{\rho}\left(\frac{c^e-w}{\alpha+\rho}\right)=a_t,
$$

or, using $r=\rho$ and rearranging terms,

$$
c^u+\frac{\alpha}{r}c^e=a_t(\alpha+r)+b+\frac{\alpha}{r}w.
$$

Appendix 8.B Numerical Solution

In this appendix, we compute numerically the solution to the pure job search model with savings (the model with $s=0$ and uncertainty only about the duration of the unemployment spell). The objective of this exercise is to illustrate and compare the consumption path obtained by Shimer and Werning (2008) for the case of CARA on the one hand, and the consumption path under optimal IBC consistent behavior on the other hand. Technically speaking, the optimality condition for consumption is the general solution to the two-dimensional system of differential equations for consumption and wealth. We change the boundary condition from what implies linear consumption as a specific solution to the natural borrowing limit. For illustration purposes, we also solve the CRRA version of the model under the simplifying assumption that $r = \rho$ so that consumption of the employed remains stationary. Finally, we briefly consider the case of simultaneously changing risk aversion and prudence (relative prudence in the case of CRRA), i.e. calculate and compare the optimal consumption paths for different values of γ (CARA) and σ (CRRA), respectively.

We use the computer algebra system `Mathematica 6.0.1` and describe all calculations in detail, so that they can easily be reproduced without prior computing knowledge.[18]

8.B.1 Consumption and Wealth Paths in the Job Search Model

Recall that an individual optimally runs down her assets during the course of unemployment and consumes a constant amount when working on a job. In what follows, we focus on the optimal behavior during the unemployment spell. We describe the situation of an unemployed who does not find a job and keep in mind that consumption jumps upward when she finds a job (and subsequently earns w forever).

CARA Preferences

As shown in the main text, if $r = \rho$ and the individual has CARA preferences, optimal consumption during the unemployment spell is completely described by the following two differential equations:

$$\begin{aligned} \dot{c}^u &= -\frac{\alpha}{\gamma}\left[1 - e^{-\gamma(ra+w-c^u)}\right] \\ \dot{a} &= ra + b - c^u. \end{aligned}$$

The linear policy function $c^u = ra + \delta_1$ obtained by Shimer and Werning (2008) hits the a axis at $-\delta_1/r$. Their solution thus implicitly imposes the boundary condition $(a, c) = (-\delta_1/r, 0)$ at the point

[18] `Mathematica` is a registered trademark of Wolfram Research, Inc. It is protected by copyright law and international treaties. © 1988-2007 Wolfram Research, Inc.

in time where consumption becomes negative.[19] We use this boundary condition to explicitly calculate the linear solution. The policy function to the IBC consistent solution is then found by replacing $-\delta_1/r$ with $-b/r$, i.e. using the borrowing limit as boundary condition.

In what follows, we solve the system of differential equations numerically and plot the policy functions in (a,c^u)-space. To do so, we firstly reverse the direction of time in the differential equations and calculate the linear solution using $c = 0$ at $a = -\delta_1/r$ as initial condition (Step 1). Upon reversing the direction of time again, we draw the solution, which is given by numerical function objects for $c(t)$ and $a(t)$, by combining consumption and wealth pairs in (a, c^u)-space. Including the $\dot{a} = 0$ and the $\dot{c}^u = 0$ loci, this gives a figure similar to a phase diagram.

Secondly, we solve the system again, using the same procedure, but this time imposing zero consumption at $a = -b/r$ to find the specific solution corresponding to the optimal IBC consistent behavior (Step 2). The resulting policy function is then drawn in the same (a, c^u)-diagram as the linear solution. Note, however, that we cannot compare the time paths for consumption and wealth obtained in this step since, in general, the underlying wealth level will be different.

To compare the optimal paths for consumption and wealth, we must use identical levels of initial wealth. We therefore pick an arbitrary initial asset level and solve the time-reversed system again, using as terminal conditions the amount of consumption obtained in Step 1 and 2 for the specified level of initial wealth (Step 3).

Step 1 (Calculating the Linear Policy Function)

To begin with, define the parameters.

```
Clear[α, γ, r, w, b];
{α, γ, r, w, b} = {0.09, 1.5, .05, 1, w / 4};
```

All parameter values are chosen arbitrarily. With this specification, the expected duration of unemployment is $1/\alpha = 11.1111$ periods and unemployment benefits amount to 25% of the labor income during employment. The implied valued for δ_1 is 0.692924.[20] The unemployed individual thus uses about 2.8 times her unemployment benefits in addition to her capital income to finance instantaneous

[19]Recall that δ_1 replaces the reservation wage in the original Shimer-Werning (2008) setting. It determines the consumption component out of labor income before the employment shock and is defined by $(b - \delta_1)\,\gamma r/\alpha = \exp[-\gamma(w - \delta_1)] - 1$, see Proposition 8.2.

[20]δ_1 is readily calculated from $(b - \delta_1)\,\gamma r/\alpha = \exp[-\gamma(w - \delta_1)] - 1$ using `Mathematica's` equation solver `Solve`.

consumption.

Next, define the system of differential equations for consumption and asset holdings.

```
Clear[dec, dea];
dec = c'[t] == α / γ * (1 - Exp[-γ * (r * a[t] + w - c[t])]);
dea = a'[t] == -(r * a[t] + b - c[t]);
```

Note that we multiplied the right hand sides by -1. This is because we solve the system backwards, starting from the actual terminal condition. This is easily achieved by reversing the direction of time. We obtain the solution, namely the time paths for consumption and wealth, using `Mathematica`'s numerical solver `NDSolve`.[21]

```
Clear[solSW, cbSW, abSW, cfSW, afSW];
solSW = NDSolve[ {dec, dea, c[0] == 0, a[0] == -δ1 / r},
  {c[t], a[t]}, {t, 0, 100}]
```

```
{{c[t] → InterpolatingFunction[{{0., 100.}}, <>][t],
  a[t] → InterpolatingFunction[{{0., 100.}}, <>][t]}}
```

The abbreviations `cbSW` and `abSW` stand for "consumption backward" and "assets backward" under the linear solution obtained by Shimer and Werning (2008). Analogously, `cfSW` and `afSW` denote "consumption forward" and "assets forward" under the linear solution, respectively. Upon extracting `cbSW` and `abSW`, `cfSW` and `afSW` are obtained by reversing the direction of time again.

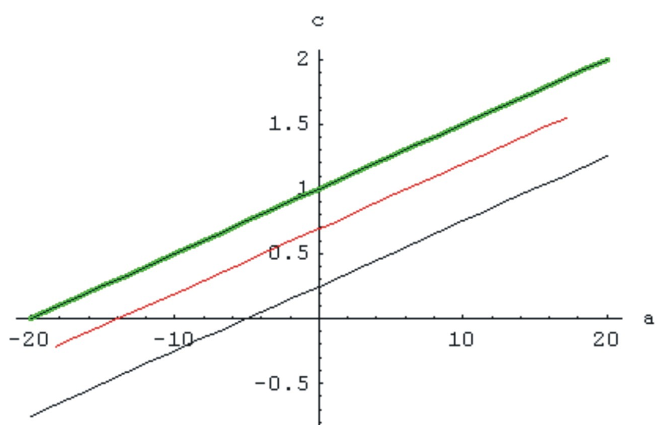
[21] `NDSolve` applies a variety of algorithms to obtain a solution. These include the forward Euler method, the midpoint rule method, the smoothed midpoint rule method (sometimes called midpoint rule with Gragg smoothing), the linearly implicit Euler method, the linearly implicit smoothed midpoint rule method (sometimes called linearly implicit midpoint rule with Bader smoothing), and the numerical evaluation of parts of systems that are locally amenable to an analytical solution. In principle, all but the last method take a sequence of steps in the dependent variable. `NDSolve` then estimates the error in the solution and compares it to pre-specified tolerances. If not satisfied, `NDSolve` adapts the step size in the numerical integration scheme or uses an alternative method to find a more accurate solution. In the problem above, accuracy does not seems to be an issue. Cf. http://documents.wolfram.com/mathematica/functions/NDSolve for an introduction to `NDSolve`.

```
cbSW[t_] := Evaluate[solSW[[1, 1, 2]]]
abSW[t_] := Evaluate[solSW[[1, 2, 2]]]
```

```
Clear[tend];
tend = 50;
cfSW[t_] := cbSW[tend - t]
afSW[t_] := abSW[tend - t]
```

`tend` is the largest time value used in this calculation and corresponds to the point in time where an unemployed individual arrives at the borrowing limit. Our choice of `tend=50` (< 100, the upper limit in the calculation) is arbitrary at this point. We verify below that it corresponds to an initial amount of wealth of about 12.7. By choosing different time values for `tend`, we obtain the consumption and wealth paths for all possible levels of initial wealth.

The policy function $c^u = c^u(a)$ is simply the pairwise combination of the solutions for $a(t)$ and $c(t)$. Using a sufficiently large value for `tend`, we obtain the policy function for all levels of wealth (the red line).[22]



It runs between the $\dot{a} = 0$ (lower black line) and the $\dot{c}^u = 0$ (upper black line) locus. The green line (identical to the $\dot{c}^u = 0$ locus) corresponds to the stationary consumption level after finding a job, given the momentary level of wealth.

Step 2 (Calculating the Policy Function to the IBC Consistent Solution)

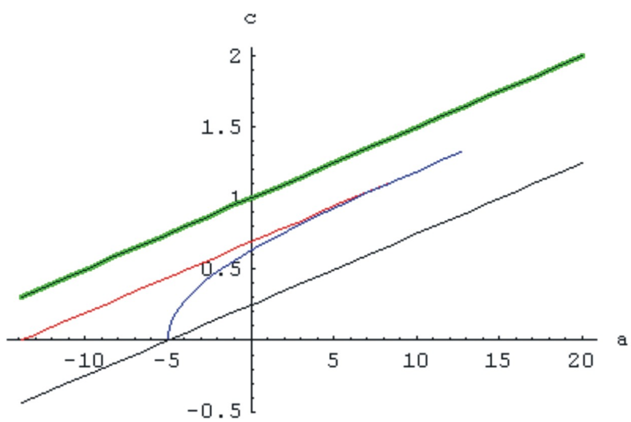
The policy function to the solution with borrowing constraint is obtained by imposing the boundary condition $(c, a) = (0, -b/r)$ when solving the system of differential equations:

[22] The figure above was generated using `tend=100` (a[0]=34.7908) to include a broader range of asset levels. Note that $a = w/r = 20$ corresponds to the present value of a workers' "lifetime" income.

```
Clear[solBC, cbBC, abBC, cfBC, afBC];
solBC = NDSolve[ {dec, dea, c[0] == 0, a[0] == -b / r},
   {c[t], a[t]}, {t, 0, 100}]
```

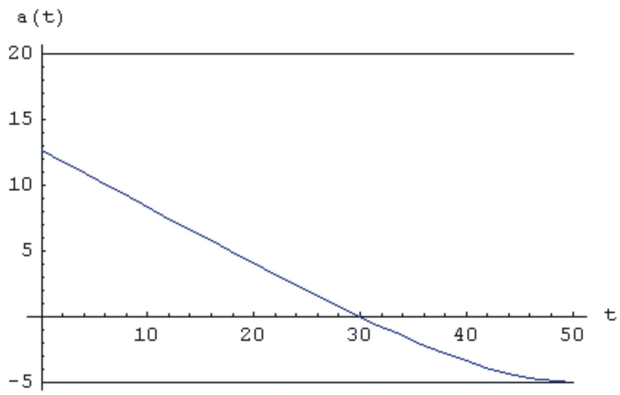
```
{{c[t] → InterpolatingFunction[{{0., 100.}}, <>][t],
  a[t] → InterpolatingFunction[{{0., 100.}}, <>][t]}}
```

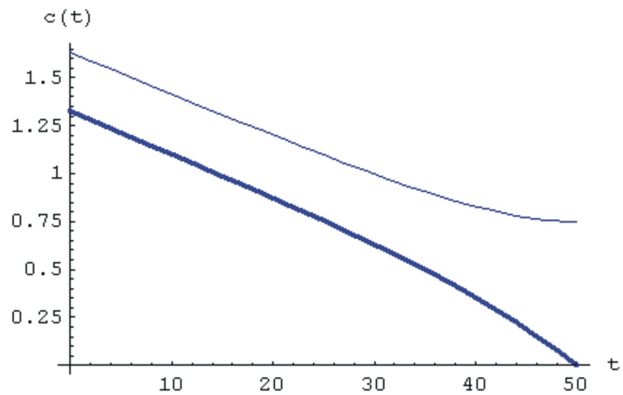
Upon reversing the direction of time, we construct the policy function from the solutions for $a(t)$ and $c(t)$ and add it to the figure obtained in Step 1 above.



Starting at the borrowing limit and moving to larger levels of wealth, the policy function to the IBC consistent solution approaches the linear policy function from below. If individuals are sufficiently rich, the presence of a borrowing constraint barely affects their consumption choice. At lower levels of wealth, however, the borrowing constraint forces the unemployed to decrease consumption by more than what is implied by the linear solution. In the figure, there is a visible reduction in consumption up to about $a = 6.7$, which corresponds to about one third of the present value of a workers' "lifetime" income ($= w/r = 20$). Using a careful calibration and/or estimation, future research may provide important insights into the quantitative impact of borrowing limits and the observed heterogeneity in consumption of the unemployed. In this assessment, the CARA case is an important assumption to distinguish between the impact of borrowing constraints and changing risk aversion.

Given our exemplary choice of `tend=50`, `a[0]` is equal to 12.6681. If an unemployed individual starts out with this level of wealth and does not find a job for 50 periods, she ends up with the maximum attainable debt and zero consumption. Here are the time paths for wealth and consumption under the IBC consistent solution for this unlucky individual:





Step 3 (Comparing Consumption and Wealth Under the Two Solutions)

Since we are free to choose an arbitrary level of wealth, we simply use `afSW[0]=8.28773` from the linear solution and the same asset level under the IBC consistent solution (which turns out to correspond to `t=9.99128` in the initial run):

```
afSWNull = afSW[0]
cfSWNull = cfSW[0]
```

```
8.28773
```

```
1.10731
```

```
afBCNull = afBC[9.99128]
cfBCNull = cfBC[9.99128]
```

```
8.28773
```

```
1.09999
```

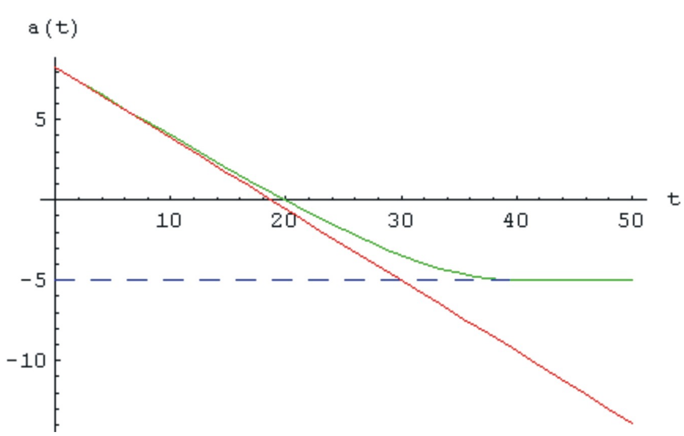
The next step is to solve the time-reversed system again, using as terminal conditions the identical asset levels (`a[100]=8.28773`) and the corresponding consumption levels on the respective optimal paths (`c[100]=1.10731` and `c[100]=1.09999`, respectively):

```
Clear[solBCa, solSWa, cbBCa, cbSWa, abSWa, abBCa,
  cfBCa, cfSWa, afBCa, afSWa];
solBCa =
 NDSolve[ {dec, dea, c[100] == cfBCNull, a[100] == afBCNull},
  {c[t], a[t]}, {t, 0, 100}]
solSWa =
 NDSolve[ {dec, dea, c[100] == cfSWNull, a[100] == afSWNull},
  {c[t], a[t]}, {t, 0, 100}]
```

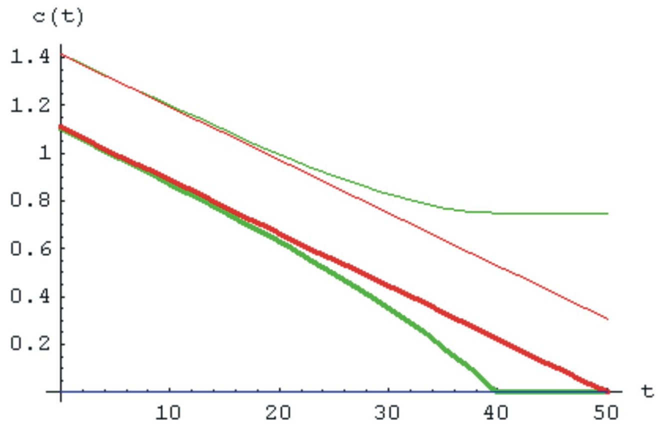
```
{{c[t] → InterpolatingFunction[{{0., 100.}}, <>][t],
  a[t] → InterpolatingFunction[{{0., 100.}}, <>][t]}}
```

```
{{c[t] → InterpolatingFunction[{{0., 100.}}, <>][t],
  a[t] → InterpolatingFunction[{{0., 100.}}, <>][t]}}
```

Upon extracting the solution and reversing time, we obtain the optimal time paths for consumption and wealth under the two solutions. Looking first at the evolution of wealth, we find that starting at a level of wealth equal to 8.28773 or about 41% of the present value of working on a job, there is little impact of the borrowing constraint on the evolution of wealth. After some time, say, somewhat longer than half of the average duration of an unemployment spell, the borrowing constraint starts to affect the consumption behavior. The unemployed reduces her assets slower relative to the case of unbounded borrowing. While the level of asset holdings under the linear solution continues to decline until the individual finds a job, the IBC consistent solution reaches the maximum level of attainable debt after 3.6 times the average duration of unemployment, and remains constant thereafter.

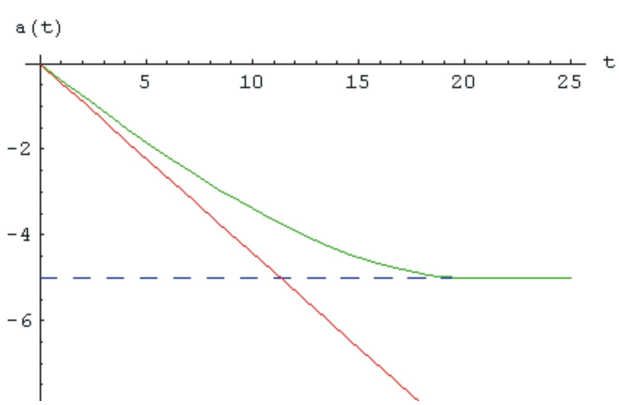


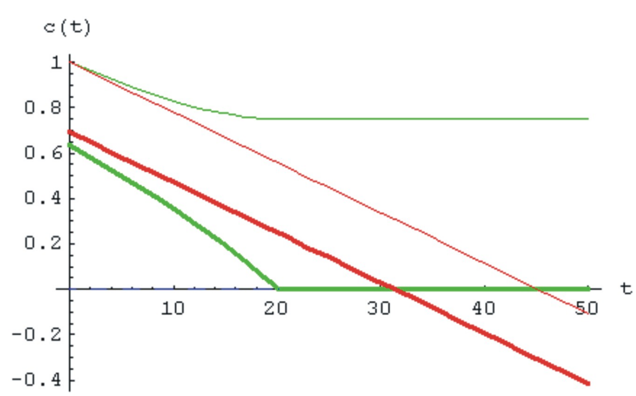
The time path of consumption reflects the evolution of wealth. Initially, the individual is sufficiently rich so that consumption differs only slightly in the cases with or without borrowing constraint. While consumption decreases linearly without the borrowing constraint, the IBC consistent consumption path declines faster, thereby slowing down the decline in wealth and, eventually, the increase in debt. Including the borrowing constraint, consumption hits zero faster, and then remains constant until the unemployed finds a job. In the linear case, consumption becomes zero at time 50, but continues to fall to negative levels.



The thin lines depict the stationary consumption levels after finding a job, given the momentary level of asset holdings. Note an important difference between the two solutions. In the case of linear consumption, the increment in consumption when finding a job remains constant across time and, as can be verified from the policy function above, also across asset levels. This feature causes the reservation wage in the original Shimer-Werning (2008) model to be independent of an individual's financial wealth, reflecting CARA. Under the IBC consistent behavior, the difference between consumption in the two employment states increases steadily, starting (with a larger initial difference) at day one of the unemployment spell. This is what generates declining reservation wages in the model with a wage offer distribution in the main text – even under CARA.

Evidently, the quantitative effects crucially hinge on the exact parametrization and the initial level of wealth. Consider, for example, an unemployed individual with zero financial wealth initially.





She considerably retards the increase in debt, both by cutting initial consumption and by decreasing consumption faster over time relative to the solution with unlimited borrowing. Measured by the difference in consumption between the two employment states, the individual facing a borrowing constraint should have a significantly stronger incentive to work.

In Shimer and Werning (2008), imperfect information on behalf of the unemployment agency generates the well-known Samaritan's dilemma: the reservation wage (the probability of finding a job) is increasing (decreasing) in the net subsidy (i.e. benefits net of employment taxes which we set to zero) received by an individual. Yet, we have just seen that unlimited access to liquidity per se prevents individuals from becoming less selective as wealth declines, and this finding is independent from the current degree of risk aversion. With unlimited access to liquidity, the opportunity costs of consumption under CARA are such that a constant reservation wage is optimal. Including the IBC, the costs of consumption are higher, leading to larger levels of wealth so that lower expected wages are sufficient to finance the target consumption level after finding a job.

Under CRRA preferences and the optimal insurance scheme, Shimer and Werning (2008) identify two determinants of the reservation wage: first, as consumption falls, the unemployed's risk aversion increases, leading to a decline in her reservation wage. Second, the optimal unemployment policy accompanies the increase in risk aversion and raises benefits over time, reinforcing the moral hazard problem. In their calibration, the net impact of both effects is less, implying an almost constant reservation wage.[23]

After shutting down both the decline in risk aversion and the increase in benefits, we have seen that a constraint on borrowing induces optimal consumption to decline faster relative to the case where debt is potentially unlimited. From the point of view of the unemployed, this implies that the expected level of consumption after finding a job declines slower over the course of unemployment, because the reduced levels of consumption during the unemployment spell retard the decline in wealth. Higher levels of wealth c.p. enable higher levels of consumption when working on a job. This induces the unemployed to accept lower wages (consumption of the unemployed does not change if the decline in labor income is compensated by a higher capital income so that $dw = -rda$). Since imposing a no-Ponzi game condition is equivalent to the natural borrowing constraint under CRRA, this mechanism plays

[23] An immediate implication is that imposing constant benefits comes at little additional cost.

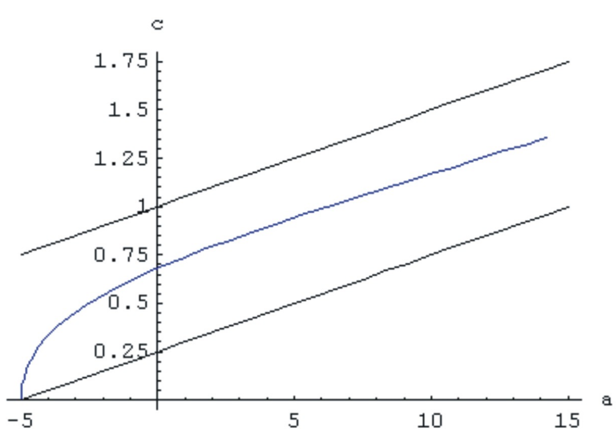
a role in the determination of the optimal level of constant unemployment benefits also. As argued by Shimer and Werning (2008), the unemployment benefit under CRRA has the dual role of relaxing the borrowing constraint and providing insurance. Interestingly, Shimer and Werning (2008) find that providing extra liquidity by means of a welfare program that generates little incentive effects *reduces* insurance costs. In particular, the solution to the optimal unemployment insurance implies notable slowly adjusting job finding rates that in fact *decline* over time. Optimality in the sense of providing the individual with a given level of utility at the lowest possible cost thus does not imply a policy that puts the individual to work as fast as possible. It instead allows her to wait for a sufficiently well paying job, thereby accepting lower capital income and potentially lower consumption in the future. The interplay of risk aversion, liquidity, and optimal consumption in generating these findings provides an interesting area for future research.

CRRA Preferences

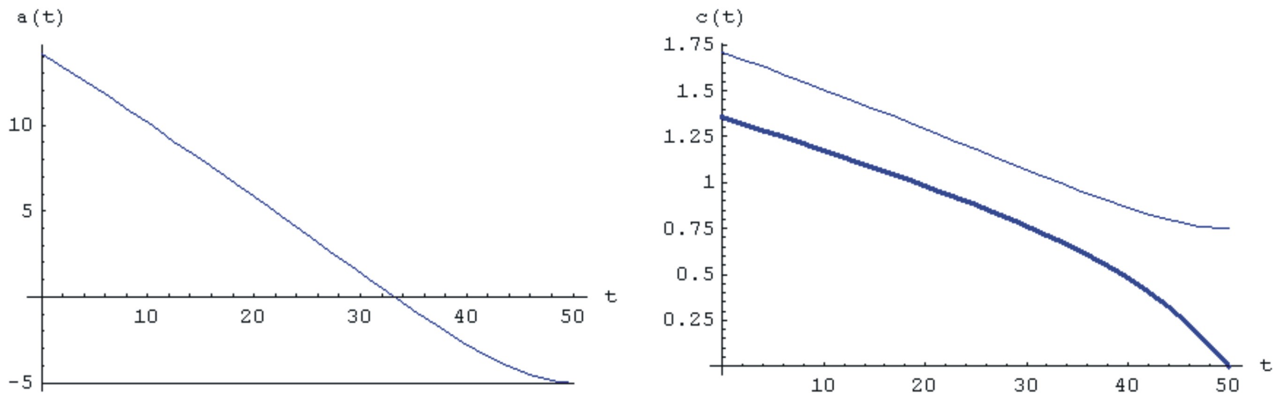
Let us briefly consider the case of CRRA preferences in an economy with $r = \rho$ explicitly. We proceed just as in the previous section, but with the law of motion for $\dot{c}^u$ given by

$$\dot{c}^u = -\frac{\alpha}{\sigma}\left[1 - \left(\frac{c^u}{ra + w}\right)^{\sigma}\right].$$

Following Shimer and Werning (2008), we set $\sigma = 1.5$. Performing analogous steps as in the CARA case again gives a phase diagram-like figure of the policy rule:



The choice of `tend=50` in this case implies an initial level of wealth equal to 14.1885. The time paths of wealth and consumption of an individual with this initial level of wealth look similar to the respective paths under CARA.



Note, in particular, that the increment of consumption between the two employment states only increases after quite a long duration of unemployment in this example.

If we relax the assumption of $r = \rho$, we must take the change in the optimal consumption of the employed worker ($ra + w$ in the case of $r = \rho$) into account. Since there is no job separation, the employed individual lives in a deterministic world after finding a job. Her consumption Euler equation is then given by the well-known expression $\dot{c}^e = (r - \rho)/\sigma$ (see equation 6.14 in Section 6.3 with $s = dq_\mu = 0$). After integrating, we get $c^e(t) = \exp[(r-\rho)t/\sigma]c^e(0)$, where $c^e(0)$ follows from the IBC in the case of employment (cf. Barro and Sala-i-Martin, 2004, pp. 93).[24] Together with the evolution of wealth, $\dot{a} = ra + w - c^e$, we can construct a policy function $c^e(a)$ and then calculate $c^u(a)$ just like before.

---

[24] If we denote by $\bar{a}$ the level of wealth at the point in time where the unemployed finds a job, and further denote this point in time by $t = 0$, the IBC is given by

$$\int_0^\infty c^e(t)\,dt = \bar{a} + \frac{w}{r}.$$

Using $c(t) = \exp[(r-\rho)t/\sigma]c^e(0)$, this gives

$$c^e(0) = \frac{\bar{a} + \frac{w}{r}}{\int_0^\infty e^{\frac{r(1-\sigma)-\rho}{\sigma}t}dt}.$$

If $r(1-\sigma) < \rho$, the exponent in the denominator is negative and

$$c^e(0) = \frac{\left(\bar{a} + \frac{w}{r}\right)(\rho - r(1-\sigma))}{\sigma}.$$

In the case where $r = \rho$, the consumption rule boils down to the linear solution in Shimer and Werning (2008) for the case of employment,

$$c^e(t) = c(0) = r\bar{a} + w$$

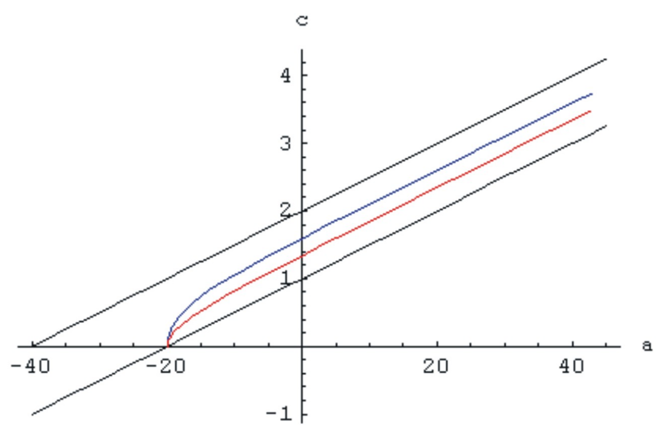
since $\dot{a}^e = 0$.

8.B.2 Changing the Attitude Towards Risk

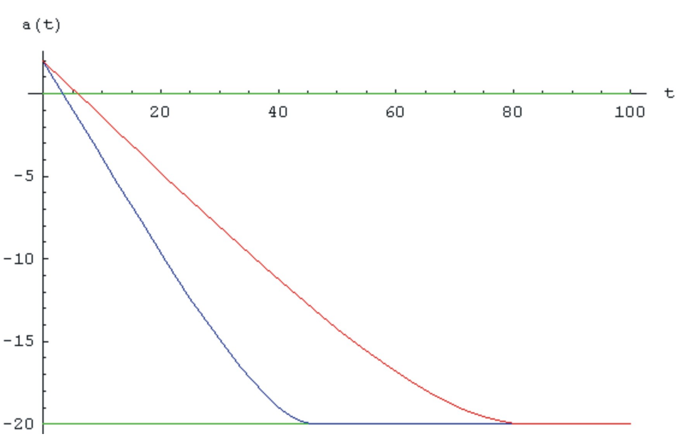
Finally, we consider the case of different degrees of risk aversion and prudence by varying γ and σ. This section provides an introduction to the section to come, where we study the impact of changing risk aversion and prudence on the optimal amount of savings in the standard two-period model of savings under uncertainty.

CARA

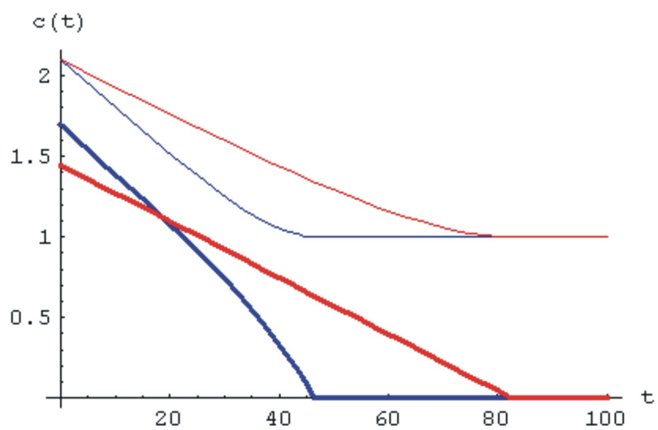
Performing the by now familiar steps, we obtain the phase diagram for two different degrees of risk aversion/prudence. We plot the policy function for $\gamma = 1$ (blue line) and $\gamma = 5$ (red line).



At any given level of wealth, the individual with higher degree of prudence/risk aversion chooses a smaller amount of consumption than the less risk averse/prudent individual. To compare both policy functions in terms of the optimal path for consumption and wealth over the course of unemployment, we again evaluate the two solutions at identical initial levels of asset holdings. We pick `a=2` and obtain the following figure for the evolution of wealth:

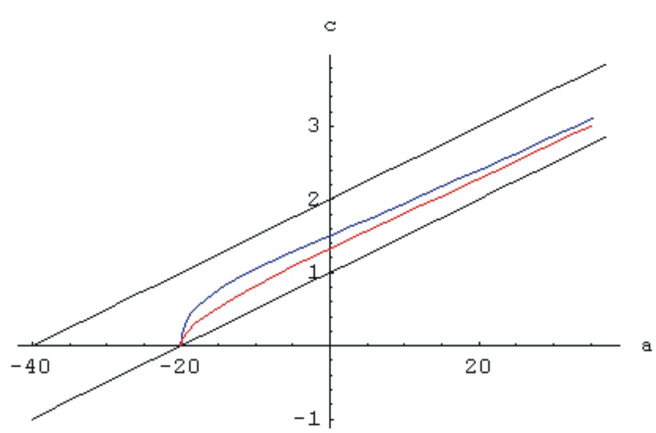


Corresponding to the lower level of consumption, the more risk averse individual runs down her assets slower than the less risk averse individual.[25] If the unemployment spells last for longer than approximately 1.7 times the average duration of unemployment, the faster decline in the less risk averse individual's wealth implies that her consumption falls short of the consumption of the more risk averse individual for the remainder of the unemployment spell. The conservative consumption behavior of the more risk averse individual implies that her consumption level after finding a job always exceeds the consumption level of the less risk averse individual, who has eaten up more assets and thus earns less capital income.



CRRA

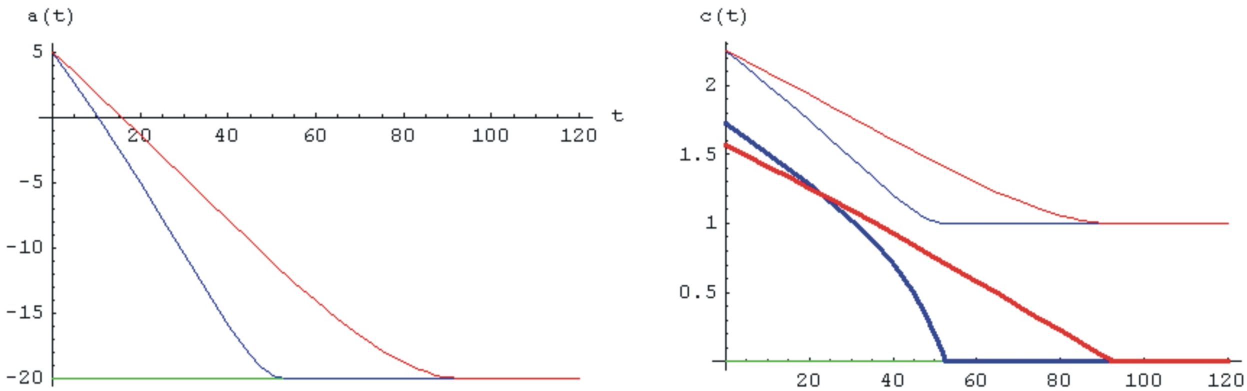
We calculate and plot the phase diagram for $\sigma = 1$ (blue line) and $\sigma = 5$ (red line).



To compare both paths, we evaluate the optimal wealth and consumption path for both individuals

[25]With CARA and CRRA preferences, a variation in the global degree of risk aversion ($= -u''(c)/u'(c)$) implies a variation of global prudence ($= -u'''(c)/u''(c)$) in the same direction as well. Cf. Chapter 9.

at an initial level of financial wealth equal to `a=5`. The resulting patterns of wealth and consumption are similar to the CARA case.



Appendix 8.C On the Agency's Problem if Unemployment is Temporary

To the best of our knowledge, no one has studied the path of an optimal unemployment policy in a matching model with savings where individuals jump back and forth between employment and unemployment, i.e. with stochastic durations of un-/employment spells. In this appendix, we take a first step and derive the expected costs of a constant benefits and taxes policy to insure an individual over the course of her lifetime. Given the lack of closed form solutions for consumption, we expect that solving the unemployment agency's dual problem is easier to do. In what follows, we assume that the agency is committed to provide an actuarially fair insurance.

With positive job separation and matching rates, and constant benefits and taxes, the expected cost of insuring an unemployed worker for T periods is

$$Z_{0,T} = E_0 \int_0^T e^{-rt} z(t) dt,$$

where $z(t) \in \{b, -\tau\}$ denotes the agency's payment of benefit and revenue from the employment tax, respectively. Expectations are formed about future employment states, given that the individual is unemployed at $t = 0$. Given the Markovian evolution of employment, the conditional probability for a currently unemployed worker to be employed at some future point in time $t \geq 0$ only depends on the elapsed time:

$$p_{uu}(t) = \frac{s}{\alpha + s} + \frac{\alpha}{\alpha + s} e^{-(\alpha+s)t}.$$

As t increases, p_{uu} converges to the fraction of time the individual will be unemployed over her lifetime, i.e. the unemployment rate $s/(\alpha+s)$.

From today's point of view ($t=0$), the insurance company pays out the unemployment benefit with probability $p_{uu}(t)$ and receives the employment tax with probability $1-p_{uu}(t)$ at time t. The expected cost of running the insurance agency thus is

$$\begin{aligned}
Z_{0,T} &= \int_0^T e^{-rt}\left\{p_{uu}(t)b-[1-p_{uu}(t)]\,\tau\right\}dt \\
&= \int_0^T e^{-rt}\left\{\left[\frac{s}{\alpha+s}+\frac{\alpha}{\alpha+s}e^{-(\alpha+s)t}\right]b-\left[\frac{\alpha}{\alpha+s}-\frac{\alpha}{\alpha+s}e^{-(\alpha+s)t}\right]\tau\right\}dt \\
&= \frac{sb-\alpha\tau}{\alpha+s}\int_0^T e^{-rt}dt+\frac{\alpha(b+\tau)}{\alpha+s}\int_0^T e^{-(r+\alpha+s)t}dt \\
&= \frac{sb-\alpha\tau}{\alpha+s}\left(\frac{1-e^{-rT}}{r}\right)+\frac{\alpha(b+\tau)}{\alpha+s}\left[\frac{1-e^{-(r+\alpha+s)T}}{r+\alpha+s}\right].
\end{aligned}$$

Taking the limit as $T\to\infty$, the expected cost of insuring an individual throughout her life amounts to

$$\begin{aligned}
Z_{0,\infty} &= \frac{sb-\alpha\tau}{(\alpha+s)r}+\frac{\alpha(b+\tau)}{(\alpha+s)\,(r+\alpha+s)} \\
&= \left(\frac{s}{\alpha+s}\right)\frac{b}{r}-\left(\frac{\alpha}{\alpha+s}\right)\frac{\tau}{r}+\left(\frac{\alpha}{\alpha+s}\right)\frac{b+\tau}{r+\alpha+s}. \qquad \text{(A.1)}
\end{aligned}$$

Suppose for the moment that once the unemployed finds a job, she can keep it forever so that $s=0$. The expected insurance cost for T periods then becomes

$$Z_{0,T}^{s=0}=-\tau\left(\frac{1-e^{-rT}}{r}\right)+\frac{\alpha(b+\tau)}{\alpha}\left[\frac{1-e^{-(r+\alpha)T}}{r+\alpha}\right],$$

and over an individuals lifetime ($T\to\infty$), this boils down to

$$Z_{0,\infty}^{s=0}=\frac{b+\tau}{r+\alpha}-\frac{\tau}{r}.$$

Without separation, the insurance agency simply pays out a continuous stream of unemployment benefits (that ends stochastically with probability α) until the unemployed finds a job. The present value of this cost is $b/(r+\alpha)$. When employed, the worker pays the employment tax forever, hence the agency receives $\tau/r-\tau/(r+\alpha)$. In the case with separation, equation (A.1) is analogous, accounting for the fact that the individual stochastically changes her employment status.

The insurance agency is limited to choose combinations of b and τ to provide an actuarially fair insurance. Running the agency therefore comes at zero expected cost or gain, $Z_{0,\infty}=0$. The insurance agency is thus free to choose τ, but must pay the benefit $b(\tau)$ that is consistent with a fair insurance. Solving $Z_{0,\infty} = 0$ for b gives

$$b_f(\tau) = \frac{\alpha(\alpha + s)\tau}{\alpha r + (r + \alpha + s)s}.$$

In the no-separation case, $b_f^{s=0}(\tau) = \alpha\tau/r$. If we look at an individual with a positive reservation wage, the arrival rate of a job is $\alpha[1 - F(\bar{w})]$. Hence, in the no-separation case, $b_f^{s=0}(\tau) = \alpha[1 - F(\bar{w})]/r$.

Chapter 9

How Changing Prudence and Risk Aversion Affects Optimal Saving[1]

9.1 Abstract

In this chapter we show how optimal saving in a two-period model is affected when prudence and risk aversion of the underlying utility function change. Increasing prudence alone will induce higher savings only if, for certain combinations of the interest rate and the pure time discount rate, there is distributional neutrality between the two periods. Otherwise, changes of risk aversion that affect the distribution between the periods must also be taken into account.

9.2 Introduction

A famous result in expected utility theory states that a mean preserving spread of risky exogenous future wealth leads to higher savings if the third derivative of the investor's von Neumann-Morgenstern utility function is positive (see Leland, 1968, Sandmo, 1970, and Drèze and Modigliani, 1972). Utility functions with this property thus reflect a specific precautionary savings motive and accordingly have been coined as "prudent" (Kimball, 1990). Just like different utility functions may show different degrees of risk aversion indicated by the Arrow-Pratt measure, they may in a quite analogous way also show different degrees of absolute and relative prudence (see also Kimball, 1990, and the exposition

[1]This section is joint work with Wolfgang Buchholz. It presents a slightly modified version of Bauer and Buchholz (2008).

in Gollier, 2001). Whereas in some cases a globally higher degree of prudence will increase savings, this assertion is not generally true (see, e.g., Menegatti, 2001, 2007, and Hau, 2002). In this chapter we further explore, in the framework of the standard two period model with identical utility functions in both periods, how a higher degree of prudence affects the optimal level of savings. The findings of our analysis are ambiguous: If, through adequate combination of the exogenous interest rate and the pure time discount rate, some equal treatment of the two periods is ensured, higher prudence will induce higher savings. In other cases, the replacement of the utility functions typically has impacts on the distribution of consumption over time such that, in addition, changes of risk aversion have to be taken into account. If risk is low or the interest rate is high, the partial effect brought about by a change of risk aversion will dominate, and the change of prudence becomes irrelevant. Moreover, it can be shown that in the more general case with different utility functions in both periods, it cannot a priori be expected that criteria based only on changes in prudence and risk aversion will generate clear-cut effects on savings behavior.

9.3 The Model

Consider the standard optimal savings model under uncertainty when there are two periods, which we synonymously interpret as two subsequent generations.[2] We first assume that the utility function is the same in the two periods 0 and 1 such that the objective function, i.e. the social welfare function in the intergenerational case, is

$$u\left(w_0 - s\right) + \beta E u\left(\tilde{w}_1 + \rho s\right). \tag{9.1}$$

Here, w_0 denotes the given certain wealth in the first period, $\tilde{w}_1$ is the uncertain wealth in the second period and s is the endogenous amount of savings such that the (safe) consumption in the earlier period is $c_0 = w_0 - s$ and (risky) consumption in the latter period is $c_1 = \tilde{w}_1 + \rho s$. The von Neumann-Morgenstern utility function $u\left(c_i\right)$ (with $i = 1, 2$) is assumed to be defined on $\mathbb{R}^+$ and to be three times continuously differentiable with $u'\left(c_i\right) > 0$, $u''\left(c_i\right) < 0$, and $u'''\left(c_i\right) > 0$, i.e. it is strictly monotonically increasing in consumption c, strictly concave, and prudent.

The marginal rate of transformation ρ between consumption in period 0 and 1 and the pure time discount factor β are exogenously given by $\rho = 1 + r$ and $\beta = 1/\left(1 + \delta\right)$, where r is the interest rate

[2] With this interpretation, our results also have some relevance for the problem of intergenerational distribution which is an important issue, e.g., in the current debate on global warming (see, e.g., Stern, 2006).

and δ is the pure rate of time discount. We assume that maximizing (9.1) with respect to s yields an interior solution s_u^*, which is characterized by the first order condition

$$u'(w_0 - s_u^*) = \beta\rho E u'(\tilde{w}_1 + \rho s_u^*). \tag{9.2}$$

An important role in our analysis is played by the "precautionary equivalent wealth level" $\hat{w}_1 = \hat{w}_1(\rho s_u^*, u, \tilde{w}_1)$, which is defined as the certainty-equivalent of the wealth distribution under optimal savings s_u^* in period 1 when $-u'(c)$ is taken to be the utility function. Thus,

$$u'(\hat{w}_1) = E u'(\tilde{w}_1 + \rho s_u^*). \tag{9.3}$$

In general, the precautionary equivalent wealth level $\hat{w}_1$ is related to the well-known precautionary equivalent premium φ via $\hat{w}_1 = \rho s_u^* - \varphi(\rho s_u^*, u, \tilde{w}_1)$ (see Kimball, 1990, and Gollier, 2001, esp. pp. 237–238).

The relation between $\hat{w}_1$ and consumption $w_0 - s_u^*$ in period 0 then crucially depends on the size of $\beta\rho$.

Lemma 9.1. *If* $\beta\rho < 1$ $(= 1,\ > 1)$ *we have*

$$\frac{\hat{w}_1}{w_0 - s_u^*} < 1 \quad (= 1,\ > 1). \tag{9.4}$$

Proof. The assertion follows as (9.2) and (9.3) imply

$$u'(\hat{w}_1) = \frac{u'(w_0 - s_u^*)}{\beta\rho} \tag{9.5}$$

and $u(c)$ is strictly concave. □

We now analyze how optimal savings will change if the utility function $u(c)$ is substituted by another utility function $v(c)$.

9.4 The Results

We assume that the new utility function $v(c)$ has the same properties as the original utility function $u(c)$, i.e. that it is three times differentiable with $v'(c) > 0$, $v''(c) < 0$, and $v'''(c) > 0$. Furthermore, $v(c)$ is supposed to be more prudent than $u(c)$ according to the definition of Kimball (1990), i.e.

$$-\frac{v'''(c)}{v''(v)} > -\frac{u'''(c)}{u''(c)} \tag{9.6}$$

holds for all consumption levels $c > 0$. Then, we get the following result:

Lemma 9.2. *If $v(c)$ is more prudent than $u(c)$ according to (9.6), then*

$$v'(\hat{w}_1) < Ev'(\tilde{w}_1 + \rho s_u^*). \tag{9.7}$$

Proof. Condition (9.6) implies that the utility function $-v'(c)$ is more risk averse than the utility function $-u'(c)$. The assertion then follows from the identity (9.3) by applying a standard result concerning changes of Arrow-Pratt risk aversion (see, e.g., Gollier, 2001, p. 21). □

This result can now be used to show that in specific cases higher prudence will induce higher savings.

Proposition 9.1. *If $\beta\rho$ is sufficiently close to 1, more prudence implies higher savings.*

Proof. We first consider the case $\beta\rho = 1$. Then, $\hat{w}_1 = w_0 - s_u^*$ from Lemma 9.1 such that Lemma 9.2 gives

$$v'(w_0 - s_u^*) < Ev'(\tilde{w}_1 + \rho s_u^*). \tag{9.8}$$

Starting from (9.8) with $s = s_u^*$, it is a straightforward implication of the concavity of $v(c)$ that s has to be increased to restore equality, i.e. to get

$$v'(w_0 - s_v^*) = Ev'(\tilde{w}_1 + \rho s_v^*) \tag{9.9}$$

as the first order condition for optimal savings s_v^* with the new utility function $v(c)$. Therefore, $s_v^* > s_u^*$ holds in the case $\beta\rho = 1$ and then, from continuity, also if $\beta\rho$ is sufficiently close to 1. □

In general, however, higher prudence alone is not sufficient to provide unambiguous results on an increase in optimal savings. Rather, additional assumptions on an accompanying change of risk aversion are required. For the proof of these results, we need a lemma that is some kind of folk theorem in the theory of risk aversion.

Lemma 9.3. *The utility function $v(c)$ is more (less) risk averse than the utility function $u(c)$, if and only if the ratio of marginal utilities $v'(c)/u'(c)$ is decreasing (increasing) in c.*

Proof. By means of a short calculation, it is shown that the following equivalence holds

$$\frac{d\left(\frac{v'(c)}{u'(c)}\right)}{dc} < 0 \quad (> 0) \quad \Leftrightarrow \quad -\frac{v''(c)}{v'(c)} > -\frac{u''(c)}{u'(c)} \quad \left(< \frac{u''(c)}{u'(c)}\right), \tag{9.10}$$

which means that $v(c)$ is more (less) risk averse than $u(c)$ according to Arrow-Pratt's standard definition. □

We then have two results on the change of optimal savings depending on whether $\beta\rho < 1$ or $\beta\rho > 1$.

Proposition 9.2. *If $\beta\rho < 1$, higher prudence combined with higher risk aversion implies higher savings.*

Proof. Since, in the case $\beta\rho < 1$, Lemma 9.1 gives $\hat{w}_1 < w_0 - s_u^*$, it follows from Lemma 9.3, i.e. higher risk aversion of $v(c)$, that

$$\frac{v'(w_0 - s_u^*)}{v'(\hat{w}_1)} < \frac{u'(w_0 - s_u^*)}{u'(\hat{w}_1)} = \beta\rho. \tag{9.11}$$

From (9.11) and Lemma 9.2, i.e. higher prudence of $v(c)$, we get

$$v'(w_0 - s_u^*) < \beta\rho E v'(\tilde{w}_1 + \rho s_u^*). \tag{9.12}$$

A similar reasoning as at the end of the proof of Proposition 9.1 then shows $s_v^* > s_u^*$. □

Quite analogously, a result for the case $\beta\rho > 1$ can be obtained.

Proposition 9.3. *If $\beta\rho > 1$, higher prudence combined with lower risk aversion implies higher savings.*

Proof. Since, in the case $\beta\rho > 1$, Lemma 9.1 gives $\hat{w}_1 > w_0 - s_u^*$, it follows from Lemma 9.3, i.e. lower risk aversion of $v(c)$, that condition (9.11) again holds. The proof then continues just like in the case of Proposition 9.2. □

We now want to provide some intuitive explanation for these results, which should make it more transparent why savings behavior depends both on prudence and on risk aversion.

9.5 The Interaction of Changes in Prudence and Risk Aversion: An Interpretation

For an interpretation of the results derived in the previous section, we start with the case $\beta\rho = 1$ in which β and ρ balance each other. Under the standard assumption that the economy is productive, i.e. $\rho > 1$ holds, this advantage for the later generation is compensated by a positive pure time discount rate $\delta > 0$, i.e. $\beta < 1$, so as to avoid an unequal outcome and thus to ensure distributional neutrality. This is a classical justification for pure time preference that dates back to Böhm-Bawerk (1883)(see also e.g. Arrow, 1999, and – clearly expressed but quite unnoticed – Rawls, 1972, pp. 297-298). How smoothing of consumption across the two generations is brought about by $\beta\rho = 1$ is particularly obvious in the special case when there is no wealth risk in the later period, i.e. if $\tilde{w}_1$ is non random. In this situation, $\beta\rho = 1$ implies equal consumption levels for both generations. In the case where $\tilde{w}_1$ is a random variable, the distributional balance between the two generations manifests itself in the identity between consumption in period 0 and the size of the precautionary equivalent wealth level. Then, as described by Proposition 9.1, the savings level is only affected by changes in prudence since effects on intergenerational distribution are canceled out.

If, however, $\beta\rho \neq 1$, things look quite different because in this case, a change of the utility function not only exerts an influence on precautionary savings, but also on the distribution of consumption across generations. First, consider the case $\beta\rho < 1$ in which the future generation is disadvantaged through a discount rate δ that is higher than the interest rate r, i.e. β is smaller than ρ. In the benchmark case without wealth risk, the future generation then would have a lower level of consumption than the present generation. With uncertainty in wealth $\tilde{w}_1$ in period 1, the precautionary equivalent wealth level is lower than consumption in period 1, i.e. $\hat{w}_1 < w_0 - \rho s_u^*$. Now, higher prudence still induces higher saving via the precautionary motive (as in the case $\beta\rho = 1$) but, in addition, the effects on the intergenerational distribution that are implied by the replacement of the utility function have to be taken into account, too. Since higher saving corresponds to a more equal intergenerational distribution in the case $\beta\rho < 1$, the new utility function $v(c)$ must be more risk averse in order to ensure a higher level of optimal saving (see Proposition 9.2).[3] In the other case with $\beta\rho > 1$, it is the future generation that is privileged by the underlying combination of β and ρ which is reflected through $\hat{w}_1 > w_0 - \rho s_u^*$. To generate higher savings in this situation, the intergenerational distribution has to become less equal such that higher prudence must be combined with less risk aversion (see Proposition 9.3).

Considering general risk averse utility functions, there is no systematic relationship between changes of prudence and changes of risk aversion,[4] which makes our results substantial. For specific classes of utility functions, however, increased prudence goes along with increased risk aversion such that there are opposing effects. Consider, as an example, the important case of iso-elastic utility functions for which the constant elasticity of marginal utility is denoted by η. Further assume that the economy is productive, i.e. $\rho > 1$, and that there is no pure time discount such that utility in both periods is given

[3] For some hints at the importance of risk aversion in this context see Ventura (2007).

[4] See Eeckhoudt and Schlesinger (1994) for examples of the independence and an analysis of some existing relationship between changes of prudence and risk aversion. Additional results on this are in Maggi, Magnani, and Menegatti (2006).

equal weight, i.e. $\beta = 1$. An increase in risk aversion η now leads to an increase in the degree of relative prudence which is $\eta + 1$. Therefore, the negative impact on savings that then results from higher risk aversion via the consumption smoothing effect over time is counteracted by the precautionary effect that stems from higher prudence. This ambiguity has clearly been noted by Dasgupta (2008) in his comment on Stern (2006).

If future wealth is certain, i.e. $\tilde{w}_1 = w_1$, only changes of risk aversion matter. Therefore, by continuity, for any given $u(c)$, β, and ρ with $\beta\rho < 1$ and any utility function $v(c)$ that is more risk averse than $u(c)$, there always exists, irrespective of the prudence of $v(c)$, a random wealth distribution $\tilde{w}_1$ with $E\tilde{w}_1 = w_1$ such that $s_v^* > s_u^*$. If $\beta\rho > 1$, the analogous result hold for utility functions $v(c)$ that are less risk averse than $u(c)$. In this case, more saving is also compatible with lower prudence if future wealth is uncertain.

Concerning changes of prudence, another irrelevance result is obtained when, for given u, β, and ρ, the condition

$$w_0 - s_u^* \leq \underline{w}_1 + \rho s_u^* \tag{9.13}$$

holds for s_u^* and $\underline{w}_1 := \min \tilde{w}_1$. Then, with optimal savings, wealth in period 1 in all states of the world is at least as high as wealth in period 0. This clearly requires $\beta\rho > 1$, and it is typically possible to generate the situation described in (9.13) by only decreasing ρ strongly enough.[5] Now, assume that $u(c)$ is replaced by any utility function $v(c)$ that is less risk averse than $u(c)$. Then, $h(c) := v'(c)/u'(c)$ is increasing in c by Lemma 9.3, such that we get

$$\begin{aligned} v'(w_0 - s_u^*) &= h(w_0 - s_u^*)\, u'(w_0 - s_u^*) = Eh(w_0 - s_u^*)\, u'(\tilde{w}_1 + \rho s_u^*) \\ &< Eh(\tilde{w}_1 + \rho s_u^*)\, u'(\tilde{w}_1 + \rho s_u^*) = Ev'(w_0 + \rho s_u^*)\,. \end{aligned} \tag{9.14}$$

By the standard argument already applied in the proofs of Propositions 9.1, 9.2, and 9.3 it then follows that $s_v^* > s_u^*$, independently of any assumption on the change in prudence. As we have started with a general utility function $u(c)$, these considerations also show that the potential irrelevance of changes in prudence for changes in savings is not a remote possibility, but rather a generic phenomenon.

---

[5]Let $u'(c) > 0$ for all $c > 0$. Now assume that $\rho s_u^* < M < \infty$ for all $\rho > 0$. Then, from concavity $Eu'(\tilde{w}_1 + \rho s_u^*) > Eu'(\tilde{w}_1 + M) > 0$ for all ρ such that, for any $\beta > 0$, $\lim_{\rho\to\infty} \beta\rho Eu'(\tilde{w}_1 + \rho s_u^*) = \infty$. The supposed boundedness of ρs_u^*, however, implies $\lim_{\rho\to\infty} s_u^* = 0$, such that $\lim_{\rho\to\infty} u'(w_0 - s_u^*) = u'(w_0) < \infty$, which is not compatible with the first order condition (9.2). Thus, $\lim_{\rho\to\infty} (\underline{w}_1 + \rho s_u^*) = \lim_{\rho\to\infty} \rho s_u^* = \infty$. This implies that there must exist a $\tilde{\rho}$ such that $\underline{w}_1 + \rho s_u^* > w_0 > w_0 - s_u^*$ for all $\rho > \tilde{\rho}$.

9.6 An Impossibility Result

We finally consider the general case where the utility functions in both periods are different. By $u_0(c_0)$ we denote the utility function in the earlier, and by $u_1(c_1)$ that in the later period. Under otherwise unchanged assumptions, the objective function then becomes

$$u_0(w_0 - s) + \beta E u_1(\tilde{w}_1 + \rho s). \tag{9.15}$$

We now show that, given $u_0(c_0)$, β, and ρ, it is not possible to characterize the class of period 1 utility functions $v_1(c_1)$ that induce higher savings than the original utility function $u_1(c_1)$ only by referring to their (absolute) degrees of risk aversion and prudence. This impossibility result follows from the following proposition.

Proposition 9.4. *Let $u_1(c_1)$ be replaced by some other utility function $v_1(c_1)$. Then, there always exists a utility function $\tilde{v}_1(c_1)$ which has the same degree of risk aversion and prudence as $v_1(c_1)$ everywhere, but induces a lower amount of savings than $u_1(c_1)$.*

Proof. Define $\tilde{v}_1(c_1)$ as $\tilde{v}_1(c_1) := \gamma v_1(c_1)$ for some constant $\gamma > 0$. Thus, $\tilde{v}_1(c_1)$ clearly has the same degrees of risk aversion and prudence as $v_1(c_1)$. Now, choose γ small enough such that

$$u_0'\left(w_1 - s^*_{u_0,u_1}\right) > \beta E \gamma v_1'\left(\tilde{w}_1 + \rho s^*_{u_0,u_1}\right) = \beta E \tilde{v}_1'\left(\tilde{w}_1 + \rho s^*_u\right), \tag{9.16}$$

where $s^*_{u_0,u_1}$ denotes optimal savings under the original combination $(u_0(c_0), u_1(c_1))$ of utility functions. Then, again by the standard argument described in the proof of Proposition 9.1, savings must decrease when $u_1(c_1)$ is substituted by $\tilde{v}_1(c_1)$. □

So we see that, because of a level effect, it cannot be expected in the general case that changes of risk aversion and/or prudence will provide sensible results on changes of savings behavior.

9.7 Conclusion

This chapter has confirmed that only in rather limited cases, changes in the degree of prudence of utility functions have unambiguous effects on optimal savings in the standard two period model. Only if there are identical utility functions in both periods and the underlying combination of the interest rate and the pure discount rate approximately give rise to distributional neutrality across the two periods, is it ensured that higher prudence induces higher savings. Otherwise, additional properties of the utility functions also play an important role. With identical utility functions in both periods, changes of risk aversion are also relevant when the intergenerational distribution is not balanced. Then, distributional effects that are not grasped by changing prudence but instead by changing risk aversion as a separate determinant become relevant for the savings decision. In general it is, depending on the given interest and pure time discount rate, well possible that the precautionary effect and the consumption smoothing effect over time resulting from a change of the utility function either support or work against each other.

Bibliography

Abdulkadiroğlu, A., B. Kuruşçu, and A. Şahin (2002): "Unemployment Insurance and the Role of Self-Insurance," *Review of Economic Dynamics*, 5(3), 681–703.

Acemoglu, D., M. Golosov, and A. Tsyvinski (2008): "Markets versus Governments," *Journal of Monetary Economics*, 55(1), 159–189.

Acemoglu, D., and R. Shimer (1999): "Efficient Unemployment Insurance," *Journal of Political Economy*, 107(5), 893–928.

Addison, J. T., M. Centeno, and P. Portugal (2008): "Do Reservation Wages Really Decline? Some International Evidence on the Determinants of Reservation Wages," *IZA Discussion Paper*, 3289.

Aghion, P., C. Harris, P. Howitt, and J. Vickers (2001): "Competition, Imitation and Growth with Step-by-Step Innovation," *The Review of Economic Studies*, 68(3), 467–492.

Aghion, P., C. Harris, and J. Vickers (1997): "Competition and Growth with Step-by-Step Innovation: An Example," *European Economic Review*, 41(3-5), 771–782.

Aiyagari, S. R. (1994): "Uninsured Idiosyncratic Risk and Aggregate Saving," *Quarterly Journal of Economics*, 109(3), 659–684.

Alesina, A., S. Ardagna, G. Nicoletti, and F. Schiantarelli (2005): "Regulation And Investment," *Journal of the European Economic Association*, 3(4), 791–825.

Alessandria, G., and H. Choi (2007): "Establishment Heterogeneity, Exporter Dynamics, and the Effects of Trade Liberalization," *Federal Reserve Bank of Philadelphia Working Paper*, 07–17.

Algan, Y., A. Chéron, J. Hairault, and F. Langot (2003): "Wealth Effect on Labor Market Transitions," *Review of Economic Dynamics*, 6(1), 156–178.

Alvarez-Pelaez, M. J., and C. Groth (2005): “Too Little or Too Much R&D?,” *European Economic Review*, 49(2), 437–456.

Andolfatto, D. (1996): “Business Cycles and Labor-Market Search,” *American Economic Review*, 86, 112–132.

Arnold, L. G. (1995): “Imitation, Intellectual Property Rights, and Growth,” *Universität Dortmund Discussion Papers in Econommics*, 95–02.

——— (1998): “Growth, Welfare, and Trade in an Integrated Model of Human-Capital Accumulation and Research,” *Journal of Macroeconomics*, 20(1), 81–105.

——— (2006): “The Dynamics of the Jones R & D Growth Model,” *Review of Economic Dynamics*, 9(1), 143–152.

——— (2007): “A Generalized Multi-Country Endogenous Growth Model,” *International Economics and Economic Policy*, 4(1), 61–100.

Arnold, L. G., and C. Bauer (2008): “Some Second Thoughts on Monopolistic Distortions and Endogenous Growth,” *BGPE Discussion Paper*, 41.

Arrow, K. (1999): “Discounting and Intergenerational Equity,” in *Discounting, Morality, and Gaming*, ed. by P. R. Portney, and J. P. Weyant. Resources for the Future, Washington, DC, 13–21.

Atkeson, A., and A. T. Burstein (2007): “Innovation, Firm Dynamics, and International Trade,” *NBER Working Paper*, W13326.

Atkinson, A. B., and J. Micklewright (1991): “Unemployment Compensation and Labor Market Transitions: A Critical Review,” *Journal of Economic Literature*, 29(4), 1679–1727.

Aw, B. Y., S. Chung, and M. J. Roberts (2000): “Productivity and Turnover in the Export Market: Micro-Level Evidence from the Republic of Korea and Taiwan (China),” *The World Bank Economic Review*, 14(1), 65–90.

Azariadis, C., and J. Stachurski (2005): “Poverty Traps,” in *Handbook of Economic Growth*, ed. by P. Aghion, and S. Durlauf. Elsevier, Amsterdam, 295–384.

Backus, D. K., P. J. Kehoe, and T. J. Kehoe (1992): “In Search of Scale Effects in Trade and Growth,” *Journal of Economic Theory*, 58(2), 377–409.

Baldwin, R. E. (2001): "Regulatory Protectionism, Developing Nations and a Two-Tier World Trade System," in *Quantifying the Impact of Technical Barriers to Trade: Can It Be Done?*, ed. by K. E. Maskus, and J. S. Wilson. University of Michigan Press.

——— (2005): "Heterogeneous Firms and Trade: Testable and Untestable Properties of the Melitz Model," *NBER Working Paper*, W11471.

Baldwin, R. E., and R. Forslid (1999): "Incremental Trade Policy and Endogenous Growth: A q-Theory Approach," *Journal of Economic Dynamics and Control*, 23(5-6), 797–822.

——— (2000): "Trade Liberalisation and Endogenous Growth A q-Theory Approach," *Journal of International Economics*, 50(2), 497–517.

Baldwin, R. E., and F. Robert-Nicoud (2008): "Trade and Growth with Heterogeneous Firms," *Journal of International Economics*, 74(1), 21–34.

Barro, R. J. (1974): "Are Government Bonds Net Wealth?," *Journal of Political Economy*, 82(6), 1095.

Barro, R. J., and X. Sala-i Martin (2004): *Economic Growth*. MIT Press.

Bauer, C. (2006): "Competition in an Increasing Variety Growth Model," *BGPE Discussion Paper*, 1.

——— (2008a): "Job Search with Savings under CARA and Nonnegative Consumption," *Manuscript*, University of Regensburg.

——— (2008b): "Trade, Productivity, and Semi-Endogenous Growth," *BGPE Discussion Paper*, 56.

Bauer, C., and W. Buchholz (2008): "How Changing Prudence and Risk Aversion Affect Optimal Saving," *Manuscript*, University of Regensburg.

Bayer, C. (2007): "Finding a Density," *Manuscript*, Vienna University of Technology.

Benassy, J. P. (1998): "Is There Always Too Little Research in Endogenous Growth with Expanding Product Variety?," *European Economic Review*, 42(1), 61–69.

Benveniste, L. M., and J. A. Scheinkman (1979): "On the Differentiability of the Value Function in Dynamic Models of Economics," *Econometrica*, 47(3), 727–732.

——— (1982): "Duality Theory for Dynamic Optimization Models of Economics: The Continuous Time Case," *Journal of Economic Theory*, 27(1), 1–19.

Bernard, A. B., J. Eaton, J. B. Jensen, and S. S. Kortum (2003): "Plants and Productivity in International Trade," *American Economic Review*, 93(4), 1268–1290.

Bernard, A. B., and J. B. Jensen (1999): "Exceptional Exporter Performance: Cause, Effect, or Both?," *Journal of International Economics*, 47(1), 1–25.

——— (2004): "Entry, Expansion, and Intensity in the US Export Boom, 1987-1992," *Review of International Economics*, 12(4), 662–675.

Bernard, A. B., J. B. Jensen, S. Redding, and P. K. Schott (2007): "Firms in International Trade," *Journal of Economic Perspectives*, 21(3), 105–130.

Bernard, A. B., J. B. Jensen, and P. K. Schott (2006a): "Survival of the Best Fit: Exposure to Low-Wage Countries and the (Uneven) Growth of US Manufacturing Plants," *Journal of International Economics*, 68(1), 219–237.

——— (2006b): "Trade Costs, Firms and Productivity," *Journal of Monetary Economics*, 53(5), 917–937.

Bernard, A. B., S. J. Redding, and P. K. Schott (2007): "Comparative Advantage and Heterogeneous Firms," *Review of Economic Studies*, 74(1), 31–66.

Bewley, T. F. (undated): "Interest Bearing Money and the Equilibrium Stock of Capital," *Manuscript*, Yale University.

Blanchard, O. J. (1985): "Debt, Deficits, and Finite Horizons," *The Journal of Political Economy*, 93(2), 223.

——— (2006): *Macroeconomics*. New Jersey, Pearson.

Blanchard, O. J., and S. Fischer (1989): *Lectures on Macroeconomics*. MIT Press.

Bloemen, H. G., and E. G. F. Stancanelli (2001): "Individual Wealth, Reservation Wages, and Transitions into Employment," *Journal of Labor Economics*, 19(2), 400–439.

——— (2005): "Financial Wealth, Consumption Smoothing and Income Shocks Arising from Job Loss," *Economica*, 72(287), 431–452.

Böhm-Bawerk, E. v. (1891): *The Positive Theory of Capital.* Translated by W. Smart, New York, GE Stechert (originally published in German in 1888).

Boldrin, M., and D. K. Levine (2002): "The Case against Intellectual Property," *American Economic Review*, 92, 209–212.

Browning, M., and T. F. Crossley (2001): "Unemployment Insurance Benefit Levels and Consumption Changes," *Journal of Public Economics*, 80(1), 1–23.

Büttner, B. (2006): "Entry Barriers and Growth," *Economics Letters*, 93(1), 150–155.

Caballero, R., and A. Jaffe (1993): "How High are the Giants' Shoulders: An Empirical Assessment of Knowledge Spillovers and Creative Destruction in a Model of Economic Growth," in *NBER Macroeconomics Annual 1993*, vol. 8. MIT Press, Cambridge MA, 15–86 (Reprinted in "Economic Growth: Theory and Evidence", ed. by G. M. Grossman, 1997, and in "Technology and Growth", ed. by P.M. Romer and M. Trajtenberg, 1997).

Cabral, L. M. B., and J. Mata (2003): "On the Evolution of the Firm Size Distribution: Facts and Theory," *The American Economic Review*, 93(4), 1075–1090.

Caselli, F., and J. Ventura (2000): "A Representative Consumer Theory of Distribution," *American Economic Review*, 90(4), 909–926.

Castañeda, A., J. Díaz-Giménez, and J.-V. Ríos-Rull (1998): "Exploring the Income Distribution Business Cycle Dynamics," *Journal of Monetary Economics*, 42(1), 93–130.

——— (2003): "Accounting for the US Earnings and Wealth Inequality," *Journal of Political Economy*, 111(4), 818–857.

Caucutt, E. M., A. İmrohoroğlu, and K. B. Kumar (2003): "Growth and Welfare Analysis of Tax Progressivity in a Heterogeneous-Agent Model," *Review of Economic Dynamics*, 6(3), 546–577.

Chamberlain, G., and C. A. Wilson (2000): "Optimal Intertemporal Consumption under Uncertainty," *Review of Economic Dynamics*, 3(3), 365–395.

Chaney, T. (2005): "The Dynamic Impact of Trade Opening: Productivity Overshooting with Heterogeneous Firms," *Manuscript*, University of Chicago.

Chetty, R. (2005): "Why Do Unemployment Benefits Raise Unemployment Durations? The Role of Borrowing Constraints and Income Effects," *NBER Working Paper*, W11760.

Ciccone, A., and K. Matsuyama (1996): "Start-Up Costs and Pecuniary Externalities as Barriers to Economic Development," *Journal of Development Economics*, 49, 33–59.

Clerides, S. K., S. Lach, and J. R. Tybout (1998): "Is Learning by Exporting Important? Micro-Dynamic Evidence from Colombia, Mexico, and Morocco," *Quarterly Journal of Economics*, 113(3), 903–947.

Coe, D. T., and E. Helpman (1995): "International R&D Spillovers," *European Economic Review*, 39(5), 859–887.

Danforth, J. P. (1979): "On the Role of Consumption and Decreasing Absolute Risk Aversion in the Theory of Job Search," *Studies in the Economics of Search*, 123, 109–131.

Dasgupta, P. (2008): "Discounting Climate Change," *Journal of Risk and Uncertainty*, forthcoming.

DeGroot, M. H. (2004): *Optimal Statistical Decisions*. Wiley-Interscience.

den Haan, W. J., G. Ramey, and J. C. Watson (2000): "Job Destruction and the Propagation of Shocks," *American Economic Review*, 90, 482–98.

Dinopoulos, E., and C. Kottaridi (2008): "The Growth Effects of National Patent Policies," *Review of International Economics*, 16(3), 499–515.

Dinopoulos, E., and P. S. Segerstrom (1999): "A Schumpeterian Model of Protection and Relative Wages," *The American Economic Review*, 89(3), 450–472.

——— (2007): "North-South Trade and Economic Growth," *Manuscript*, University of Florida and Stockholm School of Economics.

Dinopoulos, E., and P. Thompson (1998): "Schumpeterian Growth Without Scale Effects," *Journal of Economic Growth*, 3(4), 313–336.

Dixit, A. K., and J. E. Stiglitz (1977): "Monopolistic Competition and Optimum Product Diversity," *American Economic Review*, 67(3), 297–308.

Dreze, J. H., and F. Modigliani (1972): "Consumption Decisions under Uncertainty," *Journal of Economic Theory*, 5(3), 308–335.

Eaton, J., and S. S. Kortum (2001): "Trade in Capital Goods," *European Economic Review*, 45(7), 1195–1235.

Eaton, J. K., S. S. Kortum, and F. Kramarz (2004): "Dissecting Trade: Firms, Industries, and Export Destinations," *American Economic Review (Papers and Proceedings)*, 94, 150–154.

Eeckhoudt, L., and H. Schlesinger (1994): "Increases in Prudence and Increases in Risk Aversion," *Economics Letters*, 45, 51–53.

Eicher, T. S., and S. J. Turnovsky (1999): "Non-Scale Models of Economic Growth," *The Economic Journal*, 109(457), 394–415.

Erceg, C. J., L. Guerrieri, and C. Gust (2008): "Trade Adjustment and the Composition of Trade," *Journal of Economic Dynamics and Control*, 32(8), 2622–2650.

Ethier, W. J. (1982): "National and International Returns to Scale in the Modern Theory of International Trade," *American Economic Review*, 72(3), 389–405.

Falvey, R., D. Greenaway, and Z. Yu (2006): "Extending the Melitz Model to Asymmetric Countries," *University of Nottingham Research Paper*, 2006/07.

Freund, J. E., and R. E. Walpole (1980): *Mathematical Statistics*. Prentice-Hall Englewood Cliffs, NJ.

Funk, P. (2008): "Entry and Growth in a Perfectly Competitive Vintage Model," *Journal of Economic Theory*, 138(1), 211–236.

Furukawa, Y. (2007): "The Protection of Intellectual Property Rights and Endogenous Growth: Is Stronger Always Better?," *Journal of Economic Dynamics and Control*, 31(11), 3644–3670.

Gollier, C. (2001): *The Economics of Risk and Time*. MIT Press, Cambridge, MA.

Griliches, Z. (1992): "The Search for R&D Spillovers," *Scandinavian Journal of Economics*, 94(Supplement), 29–47.

Grossman, G. M., and E. Helpman (1989): "Product Development and International Trade," *Journal of Political Economy*, 97(6), 1261–1283.

——— (1991a): "Endogenous Product Cycles," *The Economic Journal*, 101(408), 1214–1229.

——— (1991b): *Innovation and Growth in the Global Economy*. MIT Press.

——— (1991c): "Quality Ladders in the Theory of Growth," *Review of Economic Studies*, 58(1), 43–61.

Grossman, G. M., and N. Yanagawa (1993): "Asset Bubbles and Endogenous Growth," *Journal of Monetary Economics*, 31(1), 3–19.

Gustafsson, P., and P. S. Segerstrom (2007): "Trade Liberalization and Productivity Growth," *CEPR Discussion Paper*, 5894.

Hagedorn, M., and I. Manovskii (2008): "The Cyclical Behavior of Equilibrium Unemployment and Vacancies Revisited," *American Economic Review*, forthcoming.

Hall, R. E. (2005): "Employment Fluctuations with Equilibrium Wage Stickiness," *American Economic Review*, 95(1), 50–65.

Hansen, G. D., and A. İmrohoroğlu (1992): "The Role of Unemployment Insurance in an Economy with Liquidity Constraints and Moral Hazard," *Journal of Political Economy*, 100(1), 118.

Harberger, A. C. (1954): "Monopoly and Resource Allocation," *The American Economic Review (Papers and Proceedings)*, 44(2), 77–87.

Hau, A. (2002): "Prudence, Intertemporal Substitution, and Savings under Uncertainty," *Journal of Economics*, 77(1), 35–52.

Heckman, J. J., L. Lochner, and C. Taber (1998): "Explaining Rising Wage Inequality: Explorations with a Dynamic General Equilibrium Model of Labor Earnings with Heterogeneous Agents," *Review of Economic Dynamics*, 1(1), 1–58.

Heer, B. (2001): "On the Welfare Gain from Stabilizing Cyclical Fluctuations," *Applied Economics Letters*, 8(5), 331–334.

Heer, B., and A. Maussner (2005): *Dynamic General Equilibrium Modelling: Computational Methods and Applications*. Springer.

Heer, B., and M. Trede (2003): "Efficiency and Distribution Effects of a Revenue-Neutral Income Tax Reform," *Journal of Macroeconomics*, 25(1), 87–107.

Hellwig, M., and A. Irmen (2001): "Endogenous Technical Change in a Competitive Economy," *Journal of Economic Theory*, 101(1), 1–39.

Helpman, E., M. J. Melitz, and S. R. Yeaple (2004): "Export versus FDI with Heterogeneous Firms," *American Economic Review*, 94(1), 300–316.

Holmlund, B. (1998): "Unemployment Insurance in Theory and Practice," *Scandinavian Journal of Economics*, 100(1), 113–141.

Hopenhayn, H. A. (1992a): "Entry, Exit, and Firm Dynamics in Long Run Equilibrium," *Econometrica*, 60(5), 1127–1150.

——— (1992b): "Exit, Selection, and the Value of Firms," *Journal of Economic Dynamics and Control*, 16(1), 621–653.

Hopenhayn, H. A., and J. Nicolini (1997): "Optimal Unemployment Insurance," *Journal of Political Economy*, 105(2), 412–438.

Horii, R., and T. Iwaisako (2007): "Economic Growth with Imperfect Protection of Intellectual Property Rights," *Journal of Economics*, 90(1), 45–85.

Howitt, P. (1999): "Steady Endogenous Growth with Population and R&D Inputs Growing," *Journal of Political Economy*, 107(4), 715–730.

Huang, C., and R. H. Litzenberger (1988): *Foundations for Financial Economics*. Prentice Hall.

Hubbard, D. W. (2007): *How to Measure Anything: Finding the Value of 'Intangibles' in Business*. John Wiley & Sons.

Hubbard, R. G., J. Skinner, and S. P. Zeldes (1995): "Precautionary Saving and Social Insurance," *Journal of Political Economy*, 103(2), 360.

Huggett, M. (1993): "The Risk-Free Rate in Heterogeneous-Agent Incomplete-Insurance Economies," *Journal of Economic Dynamics and Control*, 17(5/6), 953–970.

——— (1996): "Wealth Distribution in Life-Cycle Economies," *Journal of Monetary Economics*, 38(3), 469–494.

Huggett, M., and G. Ventura (2000): "Understanding Why High Income Households Save More Than Low Income Households," *Journal of Monetary Economics*, 45(2), 361–397.

İmrohoroğlu, A. (1989): "Cost of Business Cycles with Indivisibilities and Liquidity Constraints," *The Journal of Political Economy*, 97(6), 1364.

——— (1992): "The Welfare Cost of Inflation under Imperfect Insurance," *Journal of Economic Dynamics and Control*, 16(1), 79–91.

Iwaisako, T., and K. Futagami (2003): "Patent Policy in an Endogenous Growth Model," *Journal of Economics*, 78(3), 239–258.

Jones, C. I. (1995a): "R & D-Based Models of Economic Growth," *Journal of Political Economy*, 103(4), 759–784.

——— (1995b): "Time Series Tests of Endogenous Growth Models," *Quarterly Journal of Economics*, 110(2), 495–525.

——— (2005): "Growth and Ideas," in *Handbook of Economic Growth*, ed. by P. Aghion, and S. Durlauf. Elsevier, Amsterdam, 295–384.

Jonsson, M. (2007): "The Welfare Cost of Imperfect Competition and Distortionary Taxation," *Review of Economic Dynamics*, 10(4), 576–594.

Judd, K. L. (1985): "On the Performance of Patents," *Econometrica*, 53(3), 567–585.

Kamien, M. I., and N. L. Schwartz (1991): *Dynamic Optimization*. Elsevier Science New York, NY.

KIMBALL, M. S. (1990): "Precautionary Saving in the Large and in the Small," *Econometrica*, 58, 53–73.

KNIGHT, F. H. (1921): *Risk, Uncertainty, and Profit.* Houghton Mifflin Company, Boston, MA.

KOOPMANS, T. C. (1965): *On the Concept of Optimal Economic Growth.* Cowles Foundation for Research in Economics at Yale University.

KORNPROBST, W. (2008): *Schumpeterian Growth Theory.* Dissertation, University of Regensburg, Germany.

KORTUM, S. S. (1993): "Equilibrium R&D and the Patent-R&D Ratio: US Evidence," *American Economic Review*, 83, 450–450.

KREMER, M. (1993): "Population Growth and Technological Change: One Million BC to 1990," *Quarterly Journal of Economics*, 108(3), 681–716.

KRUGMAN, P. (1979a): "A Model of Innovation, Technology Transfer, and the World Distribution of Income," *The Journal of Political Economy*, 87(2), 253.

——— (1979b): "Increasing Returns, Monopolistic Competition, and International Trade," *Journal of International Economics*, 9(4), 469–479.

——— (1980): "Scale Economies, Product Differentiation, and the Pattern of Trade," *American Economic Review*, 70(5), 950–959.

KRUSELL, P., T. MUKOYAMA, AND A. ŞAHIN (2007): "Labor-Market Matching with Precautionary Savings and Aggregate Fluctuations," *Manuscript*, Princeton University.

KWAN, Y. K., AND E. L. C. LAI (2003): "Intellectual Property Rights Protection and Endogenous Economic Growth," *Journal of Economic Dynamics and Control*, 27(5), 853–873.

KYDLAND, F. E., AND E. C. PRESCOTT (1982): "Time to Build and Aggregate Fluctuations," *Econometrica*, 50(6), 1345–1370.

LELAND, H. E. (1968): "Saving and Uncertainty: The Precautionary Demand for Saving," *Quarterly Journal of Economics*, 82(3), 465–473.

Lentz, R. (2008): "Optimal Unemployment Insurance in an Estimated Job Search Model with Savings," *Review of Economic Dynamics*, forthcoming.

Lentz, R., and T. Tranæs (2005): "Job Search and Savings: Wealth Effects and Duration Dependence," *Journal of Labor Economics*, 23(3), 467–489.

Lerner, A. P. (1934): "The Concept of Monopoly and the Measurement of Monopoly Power," *Review of Economic Studies*, 1(3), 157–175.

Li, C. W. (2000): "Endogenous vs. Semi-Endogenous Growth in a Two-R&D-Sector Model," *The Economic Journal*, 110(462), 109–122.

Lingens, J. (2005): "Policy Implications of Endogenous Growth Models: A Note," *Economics Bulletin*, 5(13), 1–7.

Ljungqvist, L., and T. J. Sargent (2004): *Recursive Macroeconomic Theory*. MIT Press.

Lucas, R. E. (1987): *Models of Business Cycles*. Blackwell, New York.

Maggi, M. A., U. Magnani, and M. Menegatti (2006): "On the Relationship Between Absolute Prudence and Absolute Risk Aversion," *Decisions in Economics and Finance*, 29(2), 155–160.

Matheron, J. (2002): "The Welfare Cost of Monopolistic Competition Revisited," *Economics Letters*, 75(1), 129–133.

Matsuyama, K. (2007): "Beyond Icebergs: Towards a Theory of Biased Globalization," *Review of Economic Studies*, 74(1), 237–253.

McCall, J. J. (1970): "Economics of Information and Job Search," *Quarterly Journal of Economics*, 84(1), 113–126.

Melitz, M. J. (2003): "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, 71(6), 1695–1725.

Melitz, M. J., and G. I. P. Ottaviano (2008): "Market Size, Trade, and Productivity," *Review of Economic Studies*, 75(1), 295–316.

Menegatti, M. (2001): "On the Conditions for Precautionary Saving," *Journal of Economic Theory*, 98(1), 189–193.

——— (2007): "A New Interpretation for the Precautionary Saving Motive: A Note," *Journal of Economics*, 92(3), 275–280.

Merton, R. C. (1969): "Lifetime Portfolio Selection under Uncertainty: The Continuous-Time Case," *Review of Economics and Statistics*, 51(3), 247–257.

——— (1971): "Optimum Consumption and Portfolio Rules in a Continuous-Time Model," *Journal of Economic Theory*, 3(4), 373–413.

Merz, M. (1995): "Search in the Labor Market and the Real Business Cycle," *Journal of Monetary Economics*, 36(2), 269–300.

Mortensen, D. T. (1986): "Job Search and Labor Market Analysis," in *Handbook of Labor Economics*, ed. by O. Ashenfelter, and R. Layard, vol. 2. North-Holland, 849–919.

Mortensen, D. T., and C. Pissarides (1994): "Job Creation and Destruction in the Theory of Unemployment," *Review of Economic Studies*, 61(3), 397–415.

Nakajima, M. (2007): "Business Cycles in the Equilibrium Model of Labor Market Search and Self-Insurance," *Manuscript*, University of Illinois at Urbana-Champaign and University of Pennsylvania.

Nelson, R. R. (1996): *The Sources of Economic Growth*. Harvard University Press Cambridge, MA.

Nicoletti, G., and S. Scarpetta (2003): "Regulation, Productivity and Growth: OECD Evidence," *Economic Policy*, 18(36), 9–72.

Obstfeld, M. (1994): "Risk-Taking, Global Diversification, and Growth," *American Economic Review*, 84(5), 1310–1329.

O'Donoghue, T., and J. Zweimüller (2004): "Patents in a Model of Endogenous Growth," *Journal of Economic Growth*, 9(1), 81–123.

Øksendal, B. (2003): *Stochastic Differential Equations: An Introduction with Applications*. Springer.

Pavcnik, N. (2002): "Trade Liberalization, Exit, and Productivity Improvements: Evidence from Chilean Plants," *The Review of Economic Studies*, 69(1), 245–276.

Pelka, G. J. (2005): *Wachstum und Strukturwandel: Eine theoretische Analyse der Zusammenhänge*. Marburg, Metropolis.

Perez-Sebastian, F. (2000): "Transitional Dynamics in an R&D-Based Growth Model with Imitation: Comparing its Predictions to the Data," *Journal of Monetary Economics*, 45(2), 437–461.

Pissarides, C. A. (2000): *Equilibrium Unemployment Theory*. MIT Press.

Rawls, J. (1972): *A Theory of Justice*. Oxford University Press, Oxford, UK.

Rendon, S. (2006): "Job Search and Asset Accumulation under Borrowing Constraints," *International Economic Review*, 47(1), 233–263.

Rivera-Batiz, L. A., and P. M. Romer (1991): "Economic Integration and Endogenous Growth," *Quarterly Journal of Economics*, 106(2), 531–555.

Roberts, M., and J. R. Tybout (1997): "The Decision to Export in Colombia: An Empirical Model of Entry with Sunk Costs," *American Economic Review*, 87(4), 545–564.

Romer, P. M. (1987): "Growth Based on Increasing Returns Due to Specialization," *American Economic Review*, 77(2), 56–62.

——— (1990): "Endogenous Technological Change," *Journal of Political Economy*, 98(5), S71–102.

Ríos-Rull, J.-V. (1995): "Models with Heterogeneous Agents," in *Frontier of Real Business Cycle Research*, ed. by T. Cooley. Princeton University Press.

——— (1999): "Computation of Equilibria in Heterogeneous Agent Models," in *Computational Methods for the Study of Dynamic Economies: An Introduction*, ed. by R. Marimón, and A. Scott. Oxford University Press.

Rustichini, A., and J. Schmitz (1991): "Research and Imitation in Long-Run Growth," *Journal of Monetary Economics*, 27(2), 271–292.

Samuelson, P. A. (1954): "The Transfer Problem and Transport Costs, II: Analysis of Effects of Trade Impediments," *Economic Journal*, 64(254), 264–289.

——— (1965): *Foundations of Economic Analysis*. Cambridge, MA.

Sandmo, A. (1970): "The Effect of Uncertainty on Saving Decisions," *Review of Economic Studies*, 37(3), 353–360.

Scotchmer, S. (1991): "Standing on the Shoulders of Giants: Cumulative Research and the Patent Law," *Journal of Economic Perspectives*, 5(1), 29–41.

Segerstrom, P. S. (1991): "Innovation, Imitation, and Economic Growth," *Journal of Political Economy*, 99(4), 807–827.

——— (1998): "Endogenous Growth Without Scale Effects," *American Economic Review*, 88(5), 1290–1310.

Seierstad, A., and K. Sydsæter (1986): *Optimal Control Theory with Economic Applications*. Elsevier.

Sennewald, K. (2007a): "Controlled Stochastic Differential Equations under Poisson Uncertainty and with Unbounded Utility," *Journal of Economic Dynamics and Control*, 31(4), 1106–1131.

——— (2007b): *Stochastic Control, Optimal Saving, and Job Search in Continuous Time*. Dissertation, TU Dresden, Germany.

Sennewald, K., and K. Wälde (2006): "'Itô's Lemma' and the Bellman Equation for Poisson Processes: An Applied View," *Journal of Economics*, 89(1), 1–36.

Shavell, S., and L. Weiss (1979): "The Optimal Payment of Unemployment Insurance Benefits over Time," *The Journal of Political Economy*, 87(6), 1347–62.

Shimer, R. (2005): "The Cyclical Behavior of Equilibrium Unemployment, Vacancies, and Wages: Evidence and Theory," *American Economic Review*, 95(1), 25–49.

Shimer, R., and I. Werning (2007): "Reservation Wages and Unemployment Insurance," *The Quarterly Journal of Economics*, 122(3), 1145–1185.

——— (2008): "Liquidity and Insurance for the Unemployed," *American Economic Review*, forthcoming.

Sorger, G. (2007): "Quality-Improving Horizontal Innovations," *Working Paper, University of Vienna*.

Stancanelli, E. G. F. (1999a): "Do the Rich Stay Unemployed Longer? An Empirical Study for the UK," *Oxford Bulletin of Economics and Statistics*, 61, 295–314.

——— (1999b): "Do the Rich Stay Unemployed Longer? An Empirical Study for the UK," *Oxford Bulletin of Economics and Statistics*, 61, 295–314.

Stein, J. L. (2006): *Stochastic Optimal Control, International Finance, and Debt Crises*. Oxford University Press.

Stern, N. H. (2006): *Stern Review on the Economics of Climate Change*. Available online at http://www.hm-treasury.gov.uk. (30.09.2008).

Stockey, N. L. and Lucas, R. E. Jr. with E. C. Prescott (1989): *Recursive Methods in Economic Dynamics*. Harvard University Press.

Sullivan, J. X. (2002): *Essays on the Consumption, Saving, and Borrowing Behavior of Poor Households*. Dissertation, Northwestern University.

——— (2008): "Borrowing during Unemployment: Unsecured Debt as a Safety Net," *Journal of Human Resources*, 43(2), 383–412.

Tang, P. J. G., and K. Wälde (2001): "International Competition, Growth and Welfare," *European Economic Review*, 45(8), 1439–1459.

The Economist Intelligence Unit's Country Viewswire (2008): "Country ViewsWire Germany," available at http://viewswire.eiu.com/ (30.09.2008).

Tirole, J. (1988): *The Theory of Industrial Organization*. MIT Press.

Turnovsky, S. J. (1997): *International Macroeconomic Dynamics*. MIT Press.

Tybout, J. R. (2003): "Plant- and Firm-Level Evidence on 'New' Trade Theories," in *Handbook of International Trade*, ed. by E. K. Choi, and J. Harrigan. Wiley.

Ventura, G. (1999): "Flat Tax Reform: A Quantitative Exploration," *Journal of Economic Dynamics and Control*, 23(9-10), 1425–1458.

Ventura, L. (2007): "A Note on the Relevance of Prudence in Precautionary Saving," *Economics Bulletin*, 4(23), 1–11.

Walz, U. (1996): "Endogenous Innovation and Imitation in a Model of Equilibrium Growth," *European Journal of Political Economy*, 11(4), 709–723.

Wang, C., and S. D. Williamson (2002): "Moral Hazard, Optimal Unemployment Insurance, and Experience Rating," *Journal of Monetary Economics*, 49(7), 1337–1371.

Werning, I. (2002): "Optimal Unemployment Insurance with Unobservable Savings," *Manuscript*, University of Chicago and UTDT.

Wälde, K. (1999): "Optimal Saving under Poisson Uncertainty," *Journal of Economic Theory*, 87(1), 194–217.

——— (2008): *Applied Intertemporal Optimization*. Lecture Notes, University of Glasgow, Available at www.waelde.com/aio (30.09.2008).

Wolfram Research Inc. (2008): *Mathematica 6.0.1 Documentation*. Available at http://reference.wolfram.com/mathematica (30.09.2008).

Yeaple, S. R. (2005): "A Simple Model of Firm Heterogeneity, International Trade, and Wages," *Journal of International Economics*, 65(1), 1–20.

Young, A. (1998): "Growth without Scale Effects," *Journal of Political Economy*, 106(1), 41–63.