

**Regensburger  
DISKUSSIONSBEITRÄGE  
zur Wirtschaftswissenschaft**

University of Regensburg Working Papers in Business,  
Economics and Management Information Systems

**Updating the Option Implied Probability of Default Methodology**

Johannes Vilsmeier\*

October 2011

Nr. 462

*JEL Classification: C51, C52, C61, G12, G24, G32*

*Key Words: Option Implied Probability of Default, Risk Neutral Density, Cross Entropy*

---

\* Johannes Vilsmeier is a research assistant at the Department of Economics, Faculty of Business, Economics and Management Information Systems at the University of Regensburg, 93040 Regensburg, Germany  
Phone: +49-941-943-5619, E-mail: johannes.vilsmeier@wiwi.uni-regensburg.de

# Updating the Option Implied Probability of Default Methodology

Johannes Vilsmeier

## Abstract

In this paper we ‘update’ the option implied probability of default (option iPoD) approach recently suggested in the literature. First, a numerically more stable objective function for the estimation of the risk neutral density is derived whose integrals can be solved analytically. Second, it is reasoned that the originally proposed approach for the estimation of the PoD has some serious drawbacks and hence an alternative procedure is suggested that is based on the Lagrange multipliers. Carrying out numerical evaluations and a practical application we find that the framework provides very promising results.

*Keywords:* Option Implied Probability of Default, Risk Neutral Density, Cross Entropy

*JEL classification:* C51, C52, C61, G12, G24, G32

# 1 Introduction

The extraction of information from option prices is a very appealing way to construct meaningful indicators for financial fragility. Risk neutral densities (RNDs) derived from option data observed at the market can give us an image of the investors' expectation regarding the future evolution of a bank's equity. Hence, one directly obtains a forward looking measure for the bank's soundness without any need to construct comprehensive forecast models and especially avoiding the drawbacks of traditional forecast models based on backward looking accounting data.<sup>1</sup> Particularly during crisis periods it is a well known fact that traditional models almost exclusively fail to provide reasonable signals in advance of crisis periods. The reasons are manifold, ranging from problems to identify potential risk factors and transmission channels, to correctly modeling the highly nonlinear character of risk transmission in advance of crisis periods, and last but not least to obtaining the necessary data.

Facing these problems, we stress the proper use of information from investors through the option market. Due to high potential losses and the fact that banks and institutional investors use options to hedge their risk positions the market participants should be at least partly 'well informed' and changing risks should hence be displayed to a certain degree in the observed option prices. By estimating a RND one derives an entire probability distribution from the option prices and accordingly observes the market consensus (the expected value) for the future value of equity as well as the uncertainty prevailing among investors regarding the consensus. Any changes in the distribution will be associated with changes in some investors' expectation and hence may provide essential information. Consequently, one should try to impose as few a priori restrictions as possible on the shape of the density when estimating the RND as otherwise one may bias the information contained in higher statistical moments of the density.

An appealing methodology that is able to estimate RNDs of almost arbitrary form was suggested by Capuano (2008). The flexibility of the framework allows to incorporate the idea of the structural approach of Merton (1974) in a risk neutral pricing framework in order to define a default mechanism for a bank. In this way one can derive a (risk neutral) option implied probability of default (option iPoD) for a bank from the estimated RND. Contrary to other approaches based on information in market instruments the methodology requires no assumptions for the recovery rate to determine the PoD. The PoD estimates provide information regarding the expectation of the most pessimistic investors at the market, i.e.

---

<sup>1</sup>See e.g. Knaup (2011) for a discussion on the drawbacks and information content of accounting data.

the fraction of investors that expects the bank to default. One may expect the fraction of pessimists to increase in advance of crisis periods even though the market consensus may stay unchanged. The option iPoD hence is a promising addition to the usual set of option based financial indicators consisting of the statistical moments characterizing the RND.

Despite of offering an attractive tool for the construction of financial indicators the framework of Capuano (2008) faces some serious problems regarding its numerical stability and, as shown in the numerical evaluations in this paper, also regarding the accuracy of its estimates.

We suggest in this paper some technical modifications to the original framework which solve these problems to a large degree and hence considerably improve the general applicability of the option iPoD approach. The first modification concerns the estimation procedure for the cross entropy distribution as the approach applied in Capuano (2008) is generally known to be highly unstable due to singularity problems in wide ranges of its parameter space. Following Alhassid et al. (1978), the search for the roots of a highly non-linear system of equations is transformed into a stable and computationally efficient minimization problem for a strictly convex scalar function. This is done by defining a lower bound on the value of the cross entropy function for the minimum cross entropy density compatible with the available information. Further, we provide an analytical solution to the integrals of the objective function, such that no numerical methods are necessary.

The second modification concerns the determination of the optimal PoD. Numerical evaluations of the framework in this paper show that the originally suggested procedure for the estimation of the PoD is unreliable and crucially depends on the choice of arbitrary model parameters. We suggest an alternative procedure which is based on the evolution of the Lagrange multipliers when estimating the RND for different default barriers. Despite of the ad hoc nature of our approach the numerical evaluations show clearly its accuracy. In an application to option data for the Bank of America the framework indicates an elevated level of risk way before the actual downgrading by Moody's took place.

The remainder of the paper is organized as follows. In the next section the general idea of the framework suggested in Capuano (2008) is explained. Section 3 presents the updated methodology for the estimation of the cross entropy density. This comprises a detailed derivation of a new objective function and an analytical solution to the involved integrals. Section 4 provides an analysis of the mechanism that allows for the estimation of the option iPoD followed by an discussion on how to determine the optimal PoD. In this context an

ad hoc procedure is suggested whose accuracy is comprehensively evaluated in section 5. The evaluation comprises numerical examples as well as an application to real option data. Section 6 offers some conclusions.

## 2 The General Idea of the Option iPoD Framework

The basic idea of the option iPoD framework can be described as follows. One aims for a procedure that allows to estimate a ‘mass point’ in the RND that indicates the probability that the underlying of a stock option will have value zero at time of maturity of the option. This mass point can be interpreted as a probability of default as it corresponds to the fraction of investors’ that expects events leading to a stock price of zero and consequently to the default of the issuing firm. In order to be able to estimate a potential ‘jump’ in the RND (if a PoD exists) at a stock price of zero within a continuous estimation framework a ‘trick’ is applied in Capuano (2008). This trick is to define an interval of negative equity values in the RND such that density assigned to this interval in the estimation process can be interpreted as a PoD. Formally, this is achieved by integrating the structural approach of Merton (1974) in an entropy based RND estimation framework. In the following we sketch the basic building blocks of the framework.

The framework of Capuano (2008) starts by applying the structural approach of Merton (1974) to a bank’s balance sheet. The structural approach assumes that a company goes bankrupt if the value of its asset is lower than the value of its debt, implying only two claims on firm’s asset, namely equity and debt. Hence, the value of a stock  $S$ , which is a claim on firm’s equity  $E$ , can be expressed as the value of asset  $V$  minus debt  $D$ , i.e.  $S = E = (V - D)$ . Regarding an option written on a stock, it follows that the inner value of a call with strike price  $K_i$  at time of maturity  $T$  must be equal to:  $C_T^{K_i} = \max(V_T - D - K_i; 0)$ .<sup>2</sup>

Using the theory of risk neutral pricing one can obtain an estimator of the RND of  $V_T$ ,  $f(V_T)$ , based on the information contained in today’s prices for calls (or equally puts) with different strikes  $K_i$ . Based on  $f(V_T)$  one obtains for arbitrary  $D$  the following definition for the firm’s PoD:

$$PoD(D) = \int_0^D f(V_T) dV_T. \quad (1)$$

---

<sup>2</sup>The reason  $D$  is not time dependent is that it is not the investors’ expectation regarding the value of debt in  $T$  but for a value that triggers default until time of maturity.

To extract  $f(V_T)$  from the available option prices the concept of cross entropy is applied, originally introduced by Kullback and Leibler (1951). The term entropy originally refers to a concept in thermodynamics but Shannon (1948) has shown that it can as well be interpreted as a measure of the average uncertainty in a random variable. Using a weak law of large numbers and Stirling's approximation, the entropy function  $H[f(x)] = -\int_0^\infty f(x) \log f(x) dx$  can be directly derived from the multinomial coefficient in which the relative frequencies of the different outcomes are replaced by probabilities (see e.g. Jaynes (1968)). Each vector of probabilities (assignable to a given domain) entails a certain amount of possible outcomes, whereat the degree of uncertainty in a random variable increases with the number of possible outcomes. The maximum entropy distribution for a range of possible outcomes (domain) will hence be that distribution that provides the most uncertainty regarding a future outcome and it is therefore also said to be the least informative distribution. On a closed interval this will be the uniform distribution, on a unbounded positive real valued domain (for a given mean) the exponential distribution and on a unbounded real valued interval (given a mean and a variance) the normal distribution.

The related cross-entropy-function  $CE[f(V_T), f^0(V_T)]$  (also known as relative entropy) is defined as:

$$CE[f(V_T), f^0(V_T)] = \int_0^\infty f(V_T) \log \frac{f(V_T)}{f^0(V_T)} dV_T, \quad (2)$$

and can be interpreted as an entropic measure of the discrepancy between the two probability distributions  $f(x)$  and  $f^0(x)$ . In the following the latter can be thought of as a prior distribution. As suggested by Jaynes (1957), both the entropy and the cross entropy function can be used for the estimation of probability distributions when there is only partial information available, i.e. the estimation problem is under-identified as there are many distributions compatible with the information at hand. Using the measure defined in equation (2), the so called principle of minimum cross entropy allows the estimation of probability densities when there are a prior distribution and several constraints in form of expected values (moment restrictions) available. The principle chooses among all distributions that are consistent with the imposed moment conditions, the one that minimizes (2), that is the distribution that has the closest entropic discrepancy to the (known) prior distribution. If the chosen prior distribution is of maximum entropy on its defined domain, the resulting (posterior) distribution will be identical to the density that would be obtained if one maximizes the entropy function using the same moment conditions. Consequently, among

all distributions, consistent with the available information (restrictions), the one with the highest degree of uncertainty regarding a future outcome will be identified as optimal. The reason for minimizing equation (2) rather than maximizing the entropy function itself lies in the necessity of the prior to get an estimation for the PoD, as will be seen in section 4.

Subsequently we present the mathematical methodology corresponding to the above described framework.

### 3 Estimation of the RND

#### 3.1 Basic Setup

Using the cross entropy principle, the option implied RND is obtained by minimizing equation (2) under several moment constraints given by the theory of risk neutral pricing and the observed option prices. The theory of risk neutral pricing postulates that the expectation over all inner values of an option for given  $K_i$  and maturity  $T$  (measured in years), discounted with the (annual) risk free rate  $r$ , should be equal to the current option price observed at the market. Formally, one gets the following constraints:

$$C_0^{K_i} = e^{-rT} \int_{V_T=D+K_i}^{\infty} (V_T - D - K_i) f(V_T) dV_T, \quad i = 1 \dots B, \quad (3)$$

with  $B$  denoting the number of observable option prices  $C_0^{K_i}$  whereat the current stock price  $S_0$  is included as an option with strike  $K_1 = 0$ .

Applying the Lagrange multiplier technique, and taking into account the additivity constraint  $\int_0^{\infty} f(V_T) dV_T = 1$  to ensure that the density integrates to one, the Lagrangian reads as:

$$L = \int_{V_T=0}^{\infty} f(V_T) \left[ \log \frac{f(V_T)}{f^0(V_T)} \right] dV_T + \lambda_0 \left[ 1 - \int_{V_T=0}^{\infty} f(V_T) dV_T \right] + \sum_{i=1}^B \lambda_i \left[ C_0^{K_i} - e^{-rT} \int_{V_T=D+K_i}^{\infty} (V_T - D - K_i) f(V_T) dV_T \right], \quad (4)$$

where  $f^0(V_T)$  is the distribution of maximum entropy on the defined domain and  $\lambda_0, \dots, \lambda_B$  are the Lagrange multipliers. To obtain the first order conditions for  $f(V_T)$  one needs the

Fréchet derivative of the Lagrange function with respect to the density (see e.g. Cover and Thomas (2006)). The Fréchet derivative corresponds to the total differential generalized to infinite-dimensional function spaces and can be expressed by the Jacobi matrix  $J_f$  whose elements are the partial derivatives  $\partial L(f(V_T = v_i))/\partial f(V_T = v_i)$ . Setting  $J_f$  equal to zero, one obtains:

$$f^*(V_T) = f^0(V_T) \exp \left[ \lambda_0 - 1 + \sum_{i=1}^B \lambda_i e^{-rT} \mathbf{1}_{V_T > D + K_i} (V_T - D - K_i) \right], \quad (5)$$

where  $f^*(V_T) = (f^*(V_T = v_1) \cdots f^*(V_T = v_N))'$  is an infinite-dimensional vector and  $\mathbf{1}$  is an indicator function that is one if the condition is true and zero otherwise.

By inserting equation (5) in the additivity constraint,  $\exp[\lambda_0 - 1]$  can be expressed as function of the residual  $\lambda_i$  such that  $f^*(V_T)$  can be rewritten as:

$$f^*(V_T) = \frac{1}{\mu(\lambda)} f^0(V_T) \exp \left[ \sum_{i=1}^B \lambda_i e^{-rT} \mathbf{1}_{V_T > D + K_i} (V_T - D - K_i) \right], \quad (6)$$

with

$$\mu(\lambda) = \exp(1 - \lambda_0) = \exp(-\lambda'_0) = \int_{V_T=0}^{\infty} f^0(V_T) \exp \left[ \sum_{i=1}^B \lambda_i e^{-rT} \mathbf{1}_{V_T > D + K_i} (V_T - D - K_i) \right] dV_T. \quad (7)$$

It turns out that for a given value of debt  $D$  the optimization for  $f(V_T)$  results in the necessity to determine the optimal set of  $\lambda_i$ 's in (6). This can be achieved by inserting equation (6) in equation (4), deriving the resulting function with respect to the remaining  $\lambda_i$ 's and setting the latter to zero. One obtains the following nonlinear system of equations:

$$\frac{\partial L}{\partial \lambda_i} = e^{-rt} \int_{V_T=0}^{\infty} \mathbf{1}_{V_T > D + K_i} (V_T - D - K_i) f^*(V_T) dV_T - C_0^{K_i} \stackrel{!}{=} 0, \quad i = 1 \dots B. \quad (8)$$

The optimal set is usually calculated by solving (8) with a multivariate Newton-Raphson algorithm (see e.g. Zellner and Highfield (1988)), that is by linearizing the system with a first order Taylor approximation. Unfortunately the search for the roots of the system is unfeasible in many applications for various reasons. First, the Jacobi matrix resulting from the Taylor approximation has near singularities in large regions of the  $\lambda$ -space which

makes the required inversion of the Jacobi matrix impossible in most cases. Matters are further complicated by the fact that the iterative procedure used by the Newton-Raphson algorithm is very vulnerable to inaccuracies in the numerical solution of the integrals involved in equation (8). As a result, the search for the roots is unstable and converges only for a small number of constraints and when the initial values for  $\lambda$  are set near the final solution (see e.g. Ormoneit and White (1999), Maasoumi (1993)). Hence, following Alhassid et al. (1978), we subsequently suggest a robust and computationally efficient algorithm to calculate the optimal set of  $\lambda$  in equation (6).

### 3.2 Derivation of the New Objective Function

Alhassid et al. (1978) showed that a function can be defined such that for any trial set of parameters  $\lambda_1^T \dots \lambda_B^T$  it provides a theoretical upper bound to the entropy of the maximum entropy density that satisfies the imposed moment conditions. Equivalently, we will derive a lower bound to the cross entropy of the corresponding minimum cross entropy. The resulting function can also be derived directly from the Lagrange function (4) using a so called Legendre transform. In physics this kind of transformation is used to convert the fundamental equation of thermodynamics in so called thermodynamical potentials which are scalar functions and embody the same information as the original equation. Therefore, subsequently we will refer to the new objective function as a potential.

To derive the potential function we start by denoting every density that satisfies the moment constraints given by equations (3) with  $f(V_T)$  and the particular  $f(V_T)$  that is of minimum cross entropy with  $f^*(V_T)$ . Further we define  $f^{Tr}(V_T)$  as any (trial) distribution of minimum cross entropy, that is a distribution of form (6) with parameters  $\lambda_1^{Tr} \dots \lambda_B^{Tr}$ . Subsequently we will show that a strictly convex function  $W$  of  $\lambda_1^{Tr} \dots \lambda_B^{Tr}$  exists which has a minimum at that set of  $\lambda_i^{Tr} = \lambda_i^*$  that satisfies the system of equations (8) and therefore provides us with  $f^*(V_T)$ .

In order to obtain  $W$  we use the non-negativity characteristic (Cover and Thomas (2006), p. 28) of the cross entropy function, that is

$$CE[f(V_T), f^{Tr}(V_T)] = \int_{V_T=0}^{\infty} f(V_T) \log \frac{f(V_T)}{f^{Tr}(V_T)} dV_T \geq 0. \quad (9)$$

Adding and subtracting  $\int_{V_T=0}^{\infty} f(V_T) \log f^0(V_T) dV_T$  on the LHS of (9) and rearranging terms

yields:

$$CE[f(V_T), f^0(V_T)] = \int_{V_T=0}^{\infty} f(V_T) \log \frac{f(V_T)}{f^0(V_T)} dV_T \geq \int_{V_T=0}^{\infty} f(V_T) \log \frac{f^{Tr}(V_T)}{f^0(V_T)} dV_T, \quad (10)$$

with equality if and only if  $f(V_T) = f^{Tr}(V_T)$ . Next we insert equation (6) for  $f^{Tr}(V_T)$  and get for the RHS of (10):

$$\int_{V_T=0}^{\infty} f(V_T) \left[ \lambda_0^{Tr'} + \sum_{i=1}^B \lambda_i^{Tr} e^{-rT} \mathbf{1}_{V_T > D+K_i} (V_T - D - K_i) \right] dV_T, \quad (11)$$

where  $\lambda_0^{Tr'} = (\lambda_0^{Tr} - 1)$ .

As it holds that  $\int_{V_T=0}^{\infty} f(V_T) \lambda_0^{Tr'} dV_T = \lambda_0^{Tr'}$  and  $\int_{V_T=0}^{\infty} f(V_T) e^{-rT} \mathbf{1}_{V_T > D+K_i} (V_T - D - K_i) dV_T = C_0^{K_i}$  one finally obtains:

$$CE[f(V_T), f^0(V_T)] \geq CE[f^*(V_T), f^0(V_T)] \geq \lambda_0^{Tr'} + \sum_{i=1}^B \lambda_i^{Tr'} C_0^{K_i}, \quad (12)$$

whereat the first inequality holds because  $f^*(V_T)$  is just a particular  $f(V_T)$  such that the RHS also applies to  $CE[f^*(V_T), f^0(V_T)]$ . Therefore equation (12) provides a lower bound on the entropy of the distribution of minimum cross entropy, with equality if and only if  $f^{Tr}(V_T) = f^*(V_T)$ , implying  $\lambda_1^{Tr} \dots \lambda_B^{Tr} = \lambda_1^* \dots \lambda_B^*$ .

Rewriting (12) we get our working function  $W$ :

$$W = \left( CE[f^*(V_T), f^0(V_T)] - (\lambda_0^{Tr'} + \sum_{i=1}^B \lambda_i^{Tr'} C_0^{K_i}) \right) \geq 0, \quad (13)$$

which can be interpreted as a 'goodness of fit' measure of  $f^{Tr}(V_T)$  regarding  $f^*(V_T)$  and is therefore minimized.

The FOCs for  $W$  are given by the conditions:

$$\partial W / \partial \lambda_i^{Tr} \stackrel{!}{=} 0 \quad \text{or} \quad -\partial \lambda_0^{Tr'} / \partial \lambda_i^{Tr} \stackrel{!}{=} C_0^{K_i}, \quad i = 1 \dots B. \quad (14)$$

In the Appendix it is shown that  $W$  is a strictly convex function for any set of  $\lambda_i^{Tr}$  imply-

ing a unique minimum. Consequently one faces a simple minimization problem for a scalar function in  $B$  variables which, given that there is a solution<sup>3</sup>, will yield convergence for any starting values  $\lambda_{i,0}^{Tr}$ .

In practice we will minimize the potential  $F = -\lambda_0^{Tr'} - \sum_{i=1}^B \lambda_i C_0^{K_i}$  rather than  $W$  as the two functions differ only by the constant  $CE[f^*(V_T), f^0(V_T)]$  and hence  $F$  is also strictly convex and has a unique minimum. Further, following Agmon et al. (1979), we can calculate  $F$  in a computationally more efficient way by multiplying equation (5) with  $\exp(\sum_{i=1}^B \lambda_i C_0^{K_i} - \sum_{i=1}^B \lambda_i C_0^{K_i})$ , yielding:

$$\exp(-\lambda_0^{Tr''}) = \exp(-\lambda_0^{Tr'} + \sum_{i=1}^B \lambda_i^{Tr} C_0^{K_i}), \quad (15)$$

and

$$F = -\lambda_0^{Tr''} = \log \left\{ \int_{V_T=0}^{\infty} f^0(V_T) \exp \left[ \sum_{i=1}^B \lambda_i^{Tr} (e^{-rT} \mathbf{1}_{V_T > D+K_i} (V_T - D - K_i) - C_0^{K_i}) \right] dV_T \right\}, \quad (16)$$

which is the function that we minimize in our applications. Subsequently we show that we can carry out the integration implied by (16) analytically such that no numerical quadrature methods are necessary.

### 3.3 Analytical Solution of the Integrals

In order to derive an analytical solution for the integration we assume a finite domain for  $V_T$  with lower bound  $V_{min}$  and upper bound  $V_{max}$ . Further we define an uniform prior, i.e.  $f^0(V_T) = \frac{1}{V_{max}-V_{min}}$ . Then we split up the integral in (16) such that we can rewrite the

---

<sup>3</sup>This requires the derivative to change sign from positive to negative as the set  $\lambda_i^{Tr}$  varies from  $-\infty$  to  $+\infty$  (see Alhassid et al. (1978)).

objective function  $F$  without the indicator function:

$$\begin{aligned}
F = & \log \left( \frac{1}{V_{max} - V_{min}} \right) + \log \left\{ \int_{V_{min}}^D \exp \left( - \sum_{i=1}^B \lambda_i C_0^{K_i} \right) dV_T \right. \\
& + \sum_{i=1}^{B-1} \int_{D+K_i}^{D+K_{i+1}} \exp \left( \sum_{j=1}^i \lambda_j (e^{-rT}(V_T - D - K_j) - C_0^{K_j}) - \sum_{k=i+1}^B \lambda_k C_0^{K_k} \right) dV_T \\
& \left. + \int_{D+K_B}^{V_{max}} \exp \left( \sum_{j=1}^B \lambda_j (e^{-rT}(V_T - D - K_j) - C_0^{K_j}) \right) dV_T \right\}, \quad (17)
\end{aligned}$$

For this form of  $F$  the implied integrals can be solved in a straightforward way, leading to:

$$\begin{aligned}
F = & \log \left( \frac{1}{V_{max} - V_{min}} \right) + \log \left\{ \exp \left( \sum_{i=1}^B \lambda_i C_0^{K_i} \right) (D - V_{min}) \right. \\
& - \sum_{i=1}^{B-1} \left[ \frac{\exp \left( \sum_{j=1}^i \lambda_j (e^{-rT} K_j - C_0^{K_j}) - \sum_{k=i+1}^B \lambda_k C_0^{K_k} \right)}{e^{-rT} (\sum_{j=1}^i \lambda_j)} \right. \\
& \quad \left. \left. - \frac{\exp \left( \sum_{j=1}^i \lambda_j (e^{-rT} (K_{j+1} - K_j) - C_0^{K_j}) - \sum_{k=i+1}^B \lambda_k C_0^{K_k} \right)}{e^{-rT} (\sum_{j=1}^i \lambda_j)} \right) \right] \\
& \left. - \left[ \frac{\exp \left( \sum_{j=1}^B \lambda_j (e^{-rT} (K_B - K_j) - C_0^{K_j}) \right) - \exp \left( \sum_{j=1}^B \lambda_j (e^{-rT} (V_{max} - D - K_j) - C_0^{K_j}) \right)}{e^{-rT} (\sum_{j=1}^B \lambda_j)} \right] \right\} \quad (18)
\end{aligned}$$

## 4 Estimation of the option iPoD

So far we focused on the estimation of the optimal set of  $\lambda$  where we had to assume that the default barrier  $D$  is known. In this section we turn to the estimation of the optimal  $D$  and the related determination of the option implied PoD. In 4.1 we provide deeper insight to the PoD estimation mechanism and show that the accuracy of the PoD estimates depends on the choice for the length of the interval  $[V_{min}, D]$ . In 4.2 we discuss how we can identify this optimal interval length.

## 4.1 The Mechanism

We start by taking a closer look at the mechanism that enables us to estimate the option iPoD. The basic idea is to allow for a mass point for the value zero of the stock price at time of maturity of the option. In order to estimate that mass point we define a uniform prior  $f^0(V_T) = \frac{1}{(V_{max}-V_{min})}$  for an interval  $[V_{min}, V_{max}]$ , with  $V_{min} < D$ . The exact values of  $V_{min}$  and  $V_{max}$  can be freely chosen, but only the choice of  $V_{min}$  will essentially influence the results of the estimation.<sup>4</sup> The estimation procedure will use the moment constraints (3) to modify the prior to the posterior density whereat all restrictions are equal to zero for any values  $V_T \leq D$  except of the additivity constraint. The additivity constraint will assign equal density according to  $f^*(V_T) = \frac{1}{(V_{max}-V_{min})} \frac{\exp(-\sum_{i=1}^B C_0^{K_i} \lambda_i)}{\exp(F)}$  to all values  $V_T \in [V_{min}, D]$ . These values of  $V_T$  do not contribute to the value of the option as the inner values are all zero for  $V_T \leq D$  but they contribute in that way that the density above  $D$  might be more consistent with the constraints if the density for  $V_T > D$  does not integrate to one. It's important to note that all  $V_T \in [V_{min}, D]$  imply a future stock price of zero such that the integral over the assigned density for this interval can be interpreted as a PoD. Accordingly, we can define the following PoD function:

$$PoD(\lambda, D) = \frac{1}{(V_{max} - V_{min})} \frac{\exp(-\sum_{i=1}^B C_0^{K_i} \lambda_i)}{\exp(F(\lambda, D))} (D - V_{min}). \quad (19)$$

As can be seen from (19) the PoD depends crucially on the length of the interval  $[V_{min}, D]$  as a longer interval implies more values  $V_T < D$  for which a density has to be assigned. On the other hand one has to take into account that the nonparametric estimation procedure that we use will find a density that provides a similar good fit to the observed option prices for any interval length. As we will see these two features lead to a trade-off problem when searching for an optimal interval length.

If we assume that there is a mass point and hence a PoD in the true RND then there is just one interval length that provides a density that fits the prices and simultaneously exhibits the correct PoD. This will be that density that has an interval  $[V_{min}, D]$  which leads, after apportioning the true PoD equally to the values within this interval, to a density at  $V_T = D$  that is equal to the density of the true RND at  $V_T = D + \epsilon$ , with  $\epsilon$  as an infinitesimal increment of  $V_T$ . This is required by the continuity restriction of the estimation framework, and otherwise the estimated RND can not coincide with the true density.

---

<sup>4</sup>Note that not the absolute level of  $V_{min}$  does influence the results but the value of  $V_{min}$  relative to  $D$ . Further, the 'location' of the interval  $[V_{min}, V_{max}]$  does not influence the results as the inner values of the options depend on  $V_T - D$ , and  $D$  will increase by the same amount as  $V_T$  if we move the interval.

If we choose the interval too short, apportioning the true PoD to the available values in  $[V_{min}, D]$  will lead to a density at  $V_T = D$  that is too high for a smooth transition to the true density at  $V_T = D + \epsilon$  and hence the estimation procedure has to carry out some adjustments to the true density form in order to meet the observed option prices. In order to fit the prices the approach will adjust the true density such that very small values for  $V_T \in [D, V_{max}]$  will get higher density and larger values accordingly lower density. The estimated RND will exhibit a too low PoD and a too high density for small values for  $V_T \in [D, V_{max}]$ . In contrast, an interval that is too long will lead after apportioning the true PoD to a density at  $V_T = D$  that is too low for a smooth transition to the true density at  $V_T = D + \epsilon$  and hence the estimated PoD will be too large and the density for  $V_T > D$  too low for small values of  $V_T > D$ .

The Tables 1 and 2 as well as the Figures 1.(a)-1.(d) illustrate the mentioned problems using numerical evaluation examples for our estimation procedure. For the numerical evaluation user specified ('true') densities were created (see Figure 1.(a)) and the theoretically implied option prices for different strike prices were calculated from these densities. The theoretical option prices were subsequently used as input data in the cross entropy estimations for different interval lengths (see Figures 1.(b)-1.(d)). To get a realistic impression of the reliability of the estimates in practice we used about ten equally spaced different strike prices within a range of  $[0.7 \times S_0, 1.3 \times S_0]$  in order to estimate the respective densities. This corresponds to a quite realistic set of strikes for options with time to maturity of 3 months.

$D - V_{min}$	1	2	3	4	5	<b>6</b>	10	15	20
$PoD$	0.010	0.017	0.022	0.027	0.027	0.032	0.037	0.042	0.045

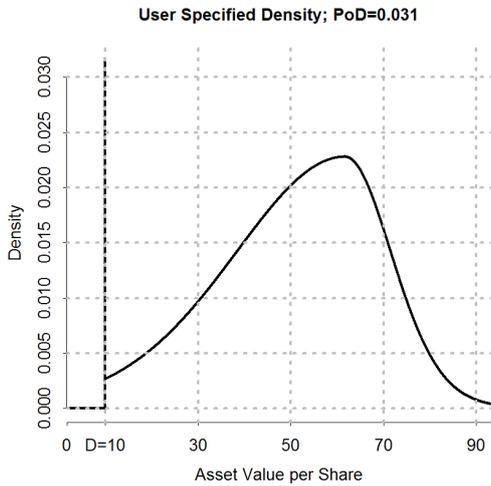
**Table 1:** Estimated PoDs for different interval lengths (Specified PoD: 0.031).

$D - V_{min}$	1	2	3	4	5	<b>7</b>	10	15	20
$PoD$	0.0004	0.0009	0.0012	0.0016	0.0020	0.00279	0.0037	0.0055	0.0065

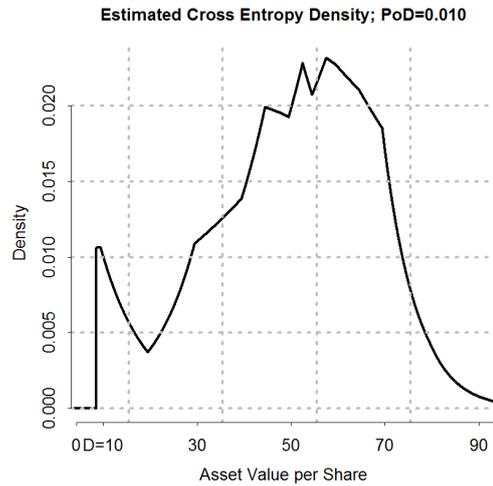
**Table 2:** Estimated PoDs for different interval lengths (Specified PoD: 0.0028).

The Tables 1 and 2 show the PoD estimates for differing interval lengths  $D - V_{min}$  for two different user specified densities. As expected the estimated PoDs clearly increase with the length of the intervals, but also can be seen that the estimation procedure roughly gets the

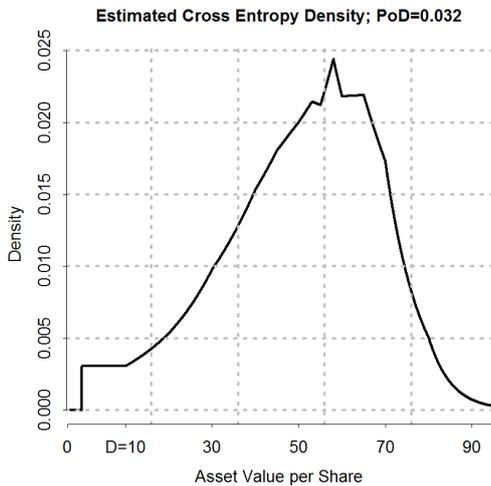
magnitude of the PoD right even for too long intervals. Especially notable is the fact, as will be illustrated in section 4.2, that the estimation procedure can clearly distinguish between a RND with a PoD and a RND that does not exhibit a PoD.



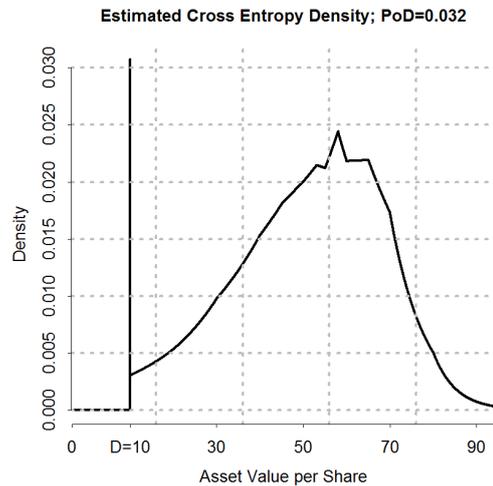
**Figure 1.(a):** User-Specified density with PoD= 0.032.



**Figure 1.(b):** Cross-Entropy density estimate with  $V_{min} = 9$  and  $D = 10$ . Estimated PoD=0.010.



**Figure 1.(c):** Cross-Entropy density estimate with  $V_{min} = 4$  and  $D = 10$ . Estimated PoD=0.32.



**Figure 1.(d):** Cross-Entropy density estimate with  $V_{min} = 4$  and  $D = 10$ . PoD assigned to  $V_T = D$ .

## 4.2 Determination of the optimal $D$

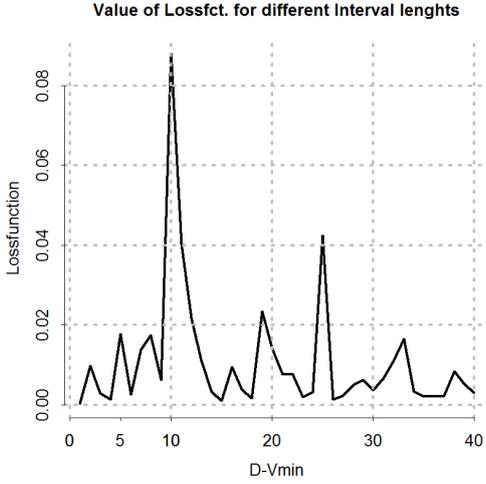
A closer look at the results in Table 1 and 2 reveals that we should aim for more than just rough estimates for the PoD. The flexibility of the framework allows us to obtain very accurate estimates for any PoD level and RND form if we are able to identify the correct interval length. An obvious approach is to evaluate our objective function (18) for different  $D$  and leaving  $V_{min}$  fixed (this is the approach suggested by Capuano (2008)). We can derive a formula for the optimal  $D$  for a given set of  $\lambda$  from the objective function in (18). Solving  $\frac{\partial F}{\partial D} = 0$  for  $D$  yields  $D^* = V_{max} - \frac{\sum_{j=1}^B \lambda_j K_j}{\sum_{j=1}^B \lambda_j}$ . Inserting this formula for  $D$  in the objective function and optimizing will give us the optimal interval length.

However, the flexibility of our approach is a curse here. To show this, a Loss-Function which measures the quadratic distance between the observed prices and the prices that are implied by the respective RND estimate for differing interval length  $D - V_{min}$  is defined. In Figure 2.(a) the Loss-Function for the numerical example of 1 is displayed, and one notes that for arbitrary interval lengths a good fit to the data can be achieved. Hence, we have no moment restrictions to identify the optimal  $D$  as any  $D$  provides a good data fit, and we will only minimize the cross entropy function if we optimize (18). Consequently we will identify that density as optimal that is the closest to the uniform distribution.

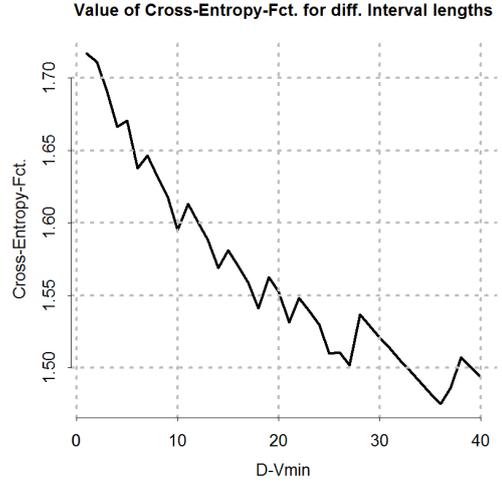
We remember that we uniformly apportion the PoD to values  $V_T \in [V_{min}, D]$  such that the 'uniform nature' of our posterior will increase with increasing  $D$ . As a consequence one will identify that  $D$  as optimal that is as large as possible and simultaneously still can provide a good fit to the data. The fit will decrease if  $D$  is so close to  $V_{max}$  that there are not enough inner values left to guarantee a good fit. As a consequence the optimal PoD will depend on how large we set  $V_{max}$  relative to  $V_{min}$ . If we choose a large interval  $[V_{min}, V_{max}]$  we will estimate a large PoD as for the optimal default barrier  $D - V_{min}$  will be large. Figure 2.(b) illustrates these detections by showing the value of the cross entropy function of estimated densities for different  $D$  using the same numerical example as in Table 1.

Noting the arbitrariness of the results using the above described approach, we suggest an alternative approach. The approach is still quite ad hoc and requires further research but, as will be seen in section 5, provides very promising results. It is based on the evolution of the PoD function (19) and of the Lagrange multipliers when estimating the optimal density for different  $D$ . We start by looking at a numerical example in which we define a RND with no PoD and estimate this RND for different  $D$ . The Figure 3.(a) shows the evolution of the

PoD and Figure 3.(b) the 'aggregated' evolution of the estimated Lagrange multipliers.



**Figure 2.(a):** Quadratic Loss-Function for estimates with different interval lengths.



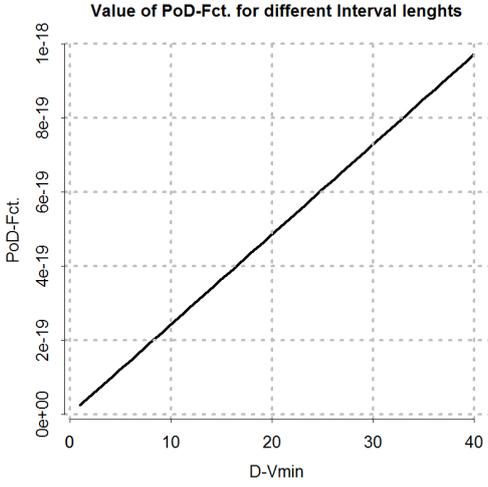
**Figure 2.(b):** Value of cross entropy function for different interval lengths.

One sees that the estimated Lagrange multipliers stay the same for all chosen  $D$  and consequently the PoD function increases linearly with growing interval length. The reason is that there is no PoD to assign to the interval  $[V_{min}, D]$  and hence no shape modifications have to be carried out for the RND for  $V_T > D$  (which would be displayed in changing  $\lambda$ s) for changing  $D$ .

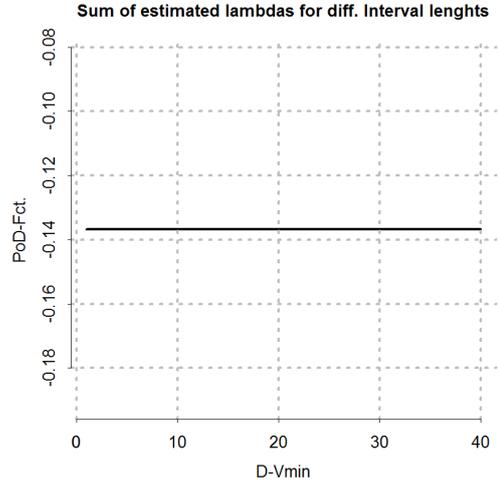
Now we look in contrast at the evolution of the PoD and the Lagrange multipliers if we define the same RND as for Table 1 (see Figure 4.(a) and 4.(b)). The evolution of the Lagrange multipliers illustrates the shape adjustments that are necessary for increasing  $D$  in order to get a good fit to the data in each case. The PoD function displays clearly a concave form and hence the slope of the function decreases with growing  $D$ . Empirically we found the PoD to be more concave if the PoD is high. Looking at the evolution of the  $\lambda$ s one detects strong fluctuations. These fluctuations will be the stronger the higher the PoD is. But equal to the PoD function (which is governed by the  $\lambda$ ) the  $\lambda$ -function is flattening with increasing  $D$ . This characteristic would clearly be more striking if the function would be smoothed.

The exact nature of the evolution of the PoD and the  $\lambda$ -Function are matter of current research. It seems that the exact determination of the optimal interval length has to be based on the second derivative of the functions as the slope of the functions is the steepest before reaching the optimal interval length of six (see Table 1) and is flattening afterwards.

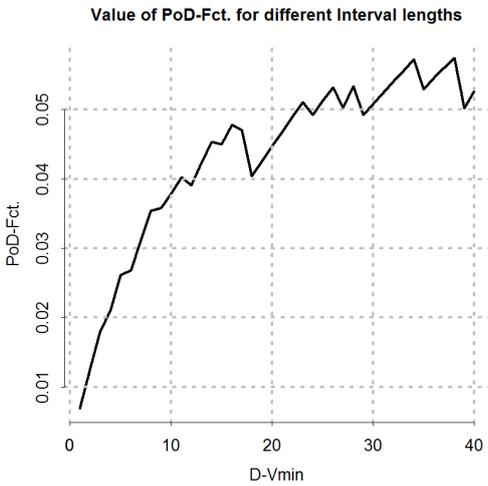
This fast convergence to the true PoD was found in all of our empirical evaluations. So far though, it is not clear what the exact decision rule should be as the degree of flattening depends on the level of the PoD.



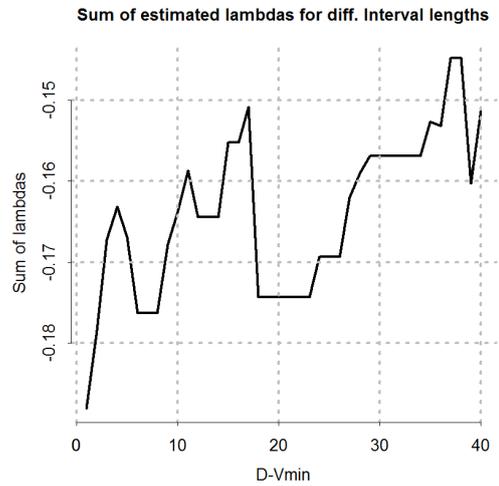
**Figure 3.(a):** Evolution of the PoD for different D; Specified PoD=0.



**Figure 3.(b):** Sum of estimated Lagrange multipliers ( $\sum_{i=1}^B \lambda_i$ ); Specified PoD=0.



**Figure 4.(a):** Evolution of the PoD for different D; Specified PoD=0.031



**Figure 4.(b):** Sum of estimated Lagrange multipliers ( $\sum_{i=1}^B \lambda_i$ ); Specified PoD=0.031

Until now we suggest the following ad hoc procedure which led to convincing results in our numerical experiments (see section 5). As we do not know the exact decision rule we decide

to average estimated PoDs over several interval lengths after deciding for an upper bound to the interval length. Empirically we found that setting  $V_{min}$  equal to zero and choosing 20 as our maximal value for  $D$  provides accurate results for arbitrary PoD levels and RND forms. The decision for  $D_{max} = 20$  is backed by the finding that the PoD function is quite flat for this value of  $D$  for any PoD level that we specified in our numerical experiments. This strongly indicates that in practical applications the true PoD will also be within an interval of length  $D - V_{min} = 20$  and averaging over estimates for lengths close to the optimal one will provide good results. To identify the optimal  $D$  and RND one chooses that interval length as optimal that provides a PoD estimate that is the closest to the 'average PoD'.

In the next section we examine the accuracy and reliability of the suggested framework regarding shape and PoD estimates for the RND.

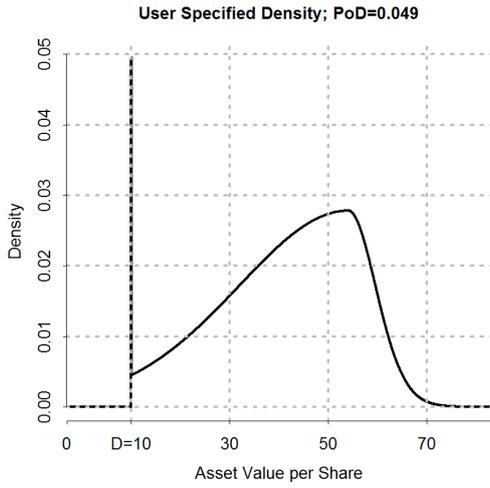
## 5 Evaluation

As addressed in the previous section, we evaluate the estimation procedure by defining different densities from which we generate our option data.

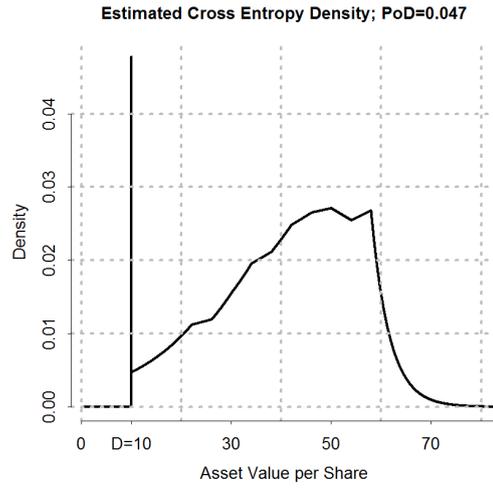
Figure 5.(a) and 6.(a) show two user-specified densities, Table 3 their respective statistical characteristics. The density in 5.(a) exhibits the typical shape of RNDs often found in empirical studies (negative skewness and positive excess kurtosis ('fat tails')) except that the entire density below the default barrier is assigned to  $D$  leading in this case to a negative (excess) kurtosis. In contrast, in Figure 6.(a) a density of quite unusual form is specified such that the great flexibility of the estimation procedure can be demonstrated. In Figures 5.(b), 6.(b) and Table 3 the respective results of the estimation are shown.

	PoD	Mean	Variance	Skewness	Kurtosis
Figure 5.(a)	0.0496	40.6024	213.3029	-0.4614	-0.6215
Figure 5.(b)	0.0478	40.6205	214.1831	-0.4392	-0.6058
Figure 6.(a)	0.0027	85.8212	680.5172	-0.1146	-0.6574
Figure 6.(b)	0.0032	86.2269	676.8333	-0.0584	-0.4193

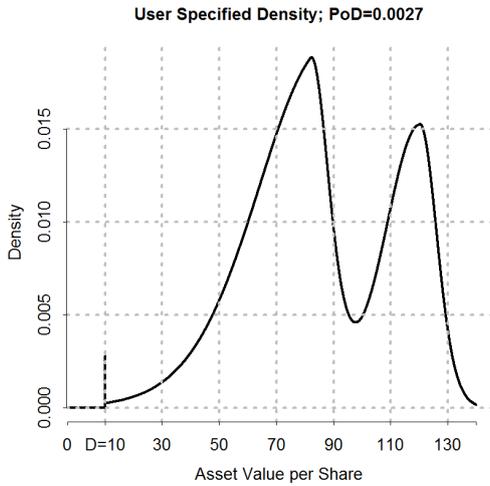
**Table 3:** PoDs and moments of specified densities and their corresponding cross-entropy density estimates.



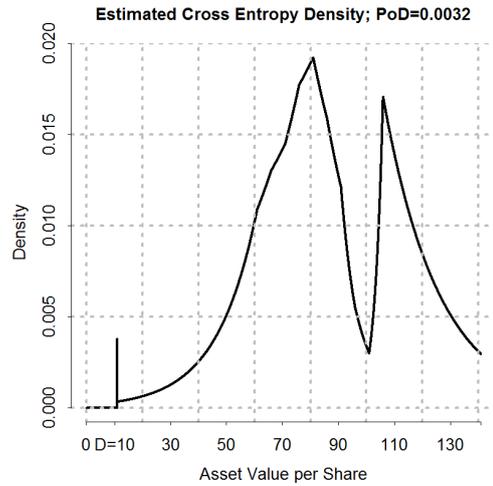
**Figure 5.(a):** User-Specified density with PoD= 0.049.



**Figure 5.(b):** Cross-Entropy density for the density in Figure 5.(a), with vector of strikes  $K = (0, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39)$ .



**Figure 6.(a):** User-Specified density with PoD= 0.0027.



**Figure 6.(b):** Cross-Entropy density for the density in Figure 6.(a), with vector of strikes  $K = (0, 55, 60, 65, 70, 75, 80, 85, 90, 95)$ .

One can easily verify the accuracy of the estimates regarding shape and PoD of the RNDs. In each case the optimal PoD was determined as the mean of the PoDs estimated for differing interval lengths ( $D_{max} = 20$ ). The optimal density was then identified as the density that exhibits a PoD that is the closest to the average PoD.

In Table 4 one can see the results of density estimates for further density specifications with PoDs ranging from very high ( $\approx 20\%$ ) to very low ( $0.0078\%$ ) and differing statistical moments. Also for these numerical experiments the framework provided very reliable results. Especially remarkable is that our easy ad hoc procedure regarding the determination of the PoD is able to obtain accurate estimates for any levels of the true PoD. A clear feature of the framework seems that lower PoDs can be estimated more accurate than high PoDs. This is due to the strong shape modifications that have to be carried out when estimating RNDs with high PoDs for different interval lengths. Consequently the PoD estimates for different  $D$  vary more for high PoDs than for low PoDs and the averaging approach is less accurate.

	PoD	Mean	Variance	Skewness	Kurtosis
Specified	0.1977	45.8692	179.3605	0.5490	-0.0811
Estimated	0.2044	45.0393	185.9304	0.7241	0.5073
Specified	0.0838	70.6151	9.5013	0.6569	4.5840
Estimated	0.0873	70.6192	8.8162	0.3001	1.4496
Specified	0.0159	23.0837	23.1187	-0.5768	-0.0459
Estimated	0.0188	23.1054	23.5051	-0.5481	0.0833
Specified	0.0010	70.0067	8.8962	-0.6949	0.4190
Estimated	0.0040	70.0634	8.9593	-0.7501	0.8354
Specified	0.000078	90.6685	51.5272	-3.6873	24.6349
Estimated	0.000121	90.6784	51.3571	-3.6397	19.8385
Specified	—	155.1997	24.9505	-1.0545	2.4086
Estimated	$10^{-23}$	155.0426	25.4788	-0.9470	1.9518

**Table 4:** PoDs and Moments of specified densities and their corresponding (optimal) Cross-Entropy density estimates.

Finally, we want to show a short application of our framework to real option prices of banks. The true shapes and PoDs of banks are of course not known, hence we evaluate our approach by contrasting the estimation results with real events.

On 9/21/2011 the Bank of America (BoA) was downgraded two levels by Moody's. This implies that BoA should clearly exhibit a higher level of PoD than e.g. JP Morgan Chase & Co. (JPM), whose top rating stayed unchanged. We now want to test if our framework is able to identify this elevated risk for BoA relative to JPM months before the ac-

tual downgrade did happen. We look at two data sets, each available at Yahoo!Finance (<http://finance.yahoo.com>). One for 4/25/2011, and one for 8/30/2011. The used options are 3 month contracts with maturity at 07/15/2011 and 11/19/2011 respectively.<sup>5</sup>

	4/25/2011	PoD	Mean	Variance	Skewness	Kurtosis
JP Morgan Chase & Co. (JPM)	$5.8 \times 10^{-9}$	54.8529	19.9191	0.0382	0.6768	
Bank of America (BoA)	0.00152	22.4115	3.5336	-0.3701	6.6342	

**Table 5:** PoD and moments of RNDs for JP Morgan Chase & Co. and Bank of America based on Options from 4/25/2011.

	8/30/2011	PoD	Mean	Variance	Skewness	Kurtosis
JP Morgan Chase & Co. (JPM)	$2.4 \times 10^{-14}$	47.6295	42.8599	1.3031	0.9717	
Bank of America (BoA)	0.04568	17.4347	8.2326	-0.2795	2.6762	

**Table 6:** PoD and moments of RNDs for JP Morgan Chase & Co. and Bank of America based on Options from 8/30/2011.

We clearly see that the framework indicates an elevated PoD of BoA relative to JPM for both dates. In addition, the PoD for BoA increases sharply getting closer to the date of the downgrading, which seems highly plausible. Finally, one notices an increasing variance for both banks, JPM and BoA, which might indicate a general increase in risk perception in the market regarding financial institutions in the last months.

## 6 Conclusion

We presented in this paper some technical modifications to the framework proposed in Capuano (2008) to derive an option implied probability of default (option iPoD). The first modification concerns the optimization algorithm to calculate the cross entropy density associated with the option prices observed at the market. We derived an objective function whose minimization is stable and yields unique solutions for the Lagrange parameters which determine the optimal density. Further, we show how the integrals of the objective function can be solved analytically. Another modification was proposed regarding the determination

<sup>5</sup>Note that both data sets have the same time to maturity such that the densities for both dates can be compared, as the maturity dependence is equal in both cases.

of the optimal PoD. After reasoning that a pure entropic approach to determining the optimal default barrier leads to quite arbitrary results, we suggested an easy to implement algorithm for the calculation of the optimal PoD based on the characteristics of the PoD function. Both modifications to the framework increase the general practical applicability of the option iPoD framework.

In section 5 we comprehensively tested our approach by applying it to user-specified data sets as well as to real option data observed at the market. The results are very convincing as the estimation procedure was shown to be highly accurate regarding the estimation of the moments and the PoD of the true density. Especially remarkable is the ability of the framework to estimate densities with very low probability of defaults which is essential for practical applications as in these one will mostly deal with low-PoD densities. In our application to real option data the framework is able to anticipate the downgrading decision taken by Moody's regarding the Bank of America.

We conclude that the suggested framework can be very useful in the derivation of market based stability indicators, e.g. for the financial system. Compared to other estimation procedures which try to extract RNDs from market data our framework is extremely flexible regarding the underlying shape of the density, assuming the least information regarding the future evolution of the firm beyond what is known from the data. In contrast to other approaches estimating PoDs from market instruments (e.g. using Credit Default Swaps) our framework has the great advantage that we do not have to assume a recovery rate for the case that the firm defaults.

Further research is required regarding the exact nature of the PoD function and the evolution of the Lagrange multipliers when the default barrier  $D$  is changed in order to obtain an exact decision rule for the determination of the option iPoD.

## Acknowledgments

The author thanks the FAZIT-STIFTUNG Gemeinnützige Verlagsgesellschaft mbH and the Bavarian Graduates Program of Economics (BGPE) for financial support during the writing process. I also thank Professor Rolf Tschernig for insightful comments and valuable support. Any errors, misrepresentations, and omissions are my own.

## References

- Agmon, N., Y. Alhassid, and R. Levine (1979). An algorithm for finding the distribution of maximal entropy. *Journal of Computational Physics* 30, 250–258.
- Alhassid, Y., N. Agmon, and R. D. Levine (1978). An upper bound for the entropy and its applications to the maximal entropy problem. *Chemical Physics Letters* 53, 22–26.
- Brockwell, P. J. and R. A. Davis (1991). *Time Series: Theory and Methods* (2nd ed.). Springer.
- Capuano, C. (2008). The option-ipod. the probability of default implied by option prices based on entropy. *IMF Working Paper* 08(194).
- Cover, T. M. and J. A. Thomas (2006). *Elements of Information Theory* (2nd ed.). John Wiley & Sons.
- Jaynes, E. T. (1957). Information theory and statistical mechanics. *Physical Review* 106, 620–630.
- Jaynes, E. T. (1968). Prior probabilities. *IEEE Transactions on Systems Science and Cybernetics* 4(3), 227–241.
- Knaup, M. (2011). *Market-Based Measures of Bank Risk and Bank Aggressiveness*. Ph. D. thesis, Tilburg University.
- Kullback, S. and R. Leibler (1951). On information and sufficiency. *Annals of Mathematical Statistics* 22, 79–86.
- Maasoumi, E. (1993). A compendium to information theory in economics and econometrics. *Econometric Reviews* 12, 137–182.
- Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. *Journal of Finance* 29, 449–470.
- Ormoneit, D. and H. White (1999). An efficient algorithm to compute maximum entropy densities. *Econometric Reviews* 18(2), 127–140.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal* 27, 379–423.
- Zellner, A. and R. A. Highfield (1988). Calculation of maximum entropy distributions and approximation of marginal posterior distributions. *Journal of Econometrics* 37, 195–209.

## Appendix

### Proof A: The strict convexity of $W$

The proof follows strongly Alhassid et al. (1978) and it is shown that the Hessian matrix for the function  $W$  is positive definite. We start by deriving the negative definiteness and hence strict concavity for the function  $F'' = -F = \lambda_0^{Tr'} + \sum_{i=1}^B \lambda_i^{Tr} C_0^{K_i}$ .

From our definition for  $W$  in equation (13) it follows that  $F''$  is smaller than  $CE[f^*(V_T), f^0(V_T)]$  for arbitrary  $\lambda_1^{Tr} \dots \lambda_B^{Tr}$ , except if  $\lambda_1^{Tr} \dots \lambda_B^{Tr} = \lambda_1^* \dots \lambda_B^*$  for which both terms are equal. Therefore,  $F''$  is a concave function with a unique maximum at  $F'' = CE[f^*(V_T), f^0(V_T)]$  if its Hessian matrix is negative definite for arbitrary  $\lambda_1^{Tr} \dots \lambda_B^{Tr}$ .

We define the following shorthand notations:

$$\phi_i(V_T) = e^{-rT} \mathbf{1}_{V_T > D + K_i} (V_T - D - K_i) \quad (20)$$

and

$$\begin{aligned} C_0^{K_i, Tr} &= \int_0^{V_{max}} \phi_i(V_T) f^{Tr}(V_T) dV_T \\ &= \int_0^{V_{max}} \phi_i(V_T) \frac{1}{\mu(\lambda)} f^0(V_T) \exp \left[ \sum_{i=1}^B \lambda_i^{Tr} \phi_i(V_T) \right] dV_T = -\partial \lambda_0^{Tr'} / \partial \lambda_i^{Tr}. \end{aligned} \quad (21)$$

Hence, we get for the first derivative of  $F''$  with respect to  $\lambda_i^{Tr}$ :

$$\partial F'' / \partial \lambda_i^{Tr} = \partial \lambda_0^{Tr'} / \partial \lambda_i^{Tr} + C_0^{K_i} = C_0^{K_i} - C_0^{K_i, Tr}, \quad (22)$$

and the Hessian of  $F''$  is given by:

$$\begin{aligned} \partial F''^2 / \partial \lambda_i^{Tr} \partial \lambda_j^{Tr} &= -\partial C_0^{K_i, Tr} / \partial \lambda_j^{Tr} = -\partial C_0^{K_j, Tr} / \partial \lambda_i^{Tr} \\ &= (C_0^{K_i, Tr} C_0^{K_j, Tr}) - \int_0^{V_{max}} \phi_i(V_T) \phi_j(V_T) f^{Tr}(V_T) dV_T \\ &= - \int_0^{V_{max}} f^{Tr}(V_T) \left[ \phi_i(V_T) - C_0^{K_i, Tr} \right] \left[ \phi_j(V_T) - C_0^{K_j, Tr} \right] dV_T, \end{aligned} \quad (23)$$

where we use the expansion  $\pm C_0^{K_i,Tr} C_0^{K_j,Tr}$  and the fact that  $C_0^{K_i,Tr} C_0^{K_j,Tr}$  can be rewritten as  $C_0^{K_i,Tr} \int_0^{V_{max}} f^{Tr}(V_T) \phi_j(V_T) dV_T$  to get from the second to the third line.

The Hessian can be interpreted as a covariance matrix of  $\phi_i(V_T)$  and  $\phi_j(V_T)$  where  $C_0^{K_i,Tr}$  and  $C_0^{K_j,Tr}$  are the respective expected values. To formally show that the Hessian is positive definite we have to show that for linearly independent constraints the matrix  $M$  with

$$M_{i,j} = \partial F''^2 / \partial \lambda_i^{Tr} \partial \lambda_j^{Tr} \quad (24)$$

satisfies for any column vector  $x$  ( $x \neq 0$ ),  $x^t M x > 0$ . To do so we note that we can write  $M_{i,j}$  as a (weighted) scalar-product of the two constraints  $B_i(V_T) = \phi_i(V_T) - C_0^{K_i,Tr}$  and  $B_j(V_T) = \phi_j(V_T) - C_0^{K_j,Tr}$  (see e.g. Brockwell and Davis (1991) for the axioms that define a scalar product), i.e.

$$M_{i,j} = - \int_0^{V_{max}} f^{Tr}(V_T) B_i(V_T) B_j(V_T) dV_T = - \langle B_i, B_j \rangle^{f^{Tr}}, \quad (25)$$

where  $f^{Tr}(V_T)$  is a strictly positive weighting function. Using the properties of a scalar product we can further write  $x^t M x$  as:

$$x^t M x = - \sum_{i,j=1}^B x_i \langle B_i, B_j \rangle^{f^{Tr}} x_j = - \langle \sum_{i=1}^B x_i B_i, \sum_{j=1}^B x_j B_j \rangle^{f^{Tr}} = - \langle C, C \rangle^{f^{Tr}}. \quad (26)$$

An alternative way to write  $\langle C, C \rangle$  is:

$$\langle C, C \rangle = (x_1 B_1(V_1) + \dots + x_B B_B(V_1))^2 + \dots + (x_1 B_1(V_{V_{max}}) + \dots + x_B B_B(V_{V_{max}}))^2, \quad (27)$$

so that that it holds that  $x^t M x < 0$  and hence  $M$  is negative definite for arbitrary sets of  $\lambda_i^{Tr}$  if  $C \neq 0$ . One obtains  $C = 0$  if and only if for every  $V_T$  holds:

$$\sum_{i=1}^B x_i B_i = \sum_{i=1}^B x_i \phi_i(V_T) - \sum_{i=1}^B x_i C_0^{K_i,Tr} = 0 \quad (28)$$

or

$$\sum_{i=0}^B x_i \phi_i(V_T) = 0, \quad (29)$$

with  $\phi_0 = 1$  and  $x_0$  equal to the second term in (28). Equation (29) will only be satisfied if one or more constraints are linearly dependent. If this is the case the cross entropy distribution will be still unique but the magnitude of the  $\lambda_i^{Tr}$  are not identified uniquely. In practice one can always eliminate one or more constraints in order to obtain a linearly independent set of moment conditions (constraints).

The derivation of the positive definiteness for our working function  $W$  is straightforward as we defined  $F'' = -F$  and  $W = CE[f^*(V_T), f^0(V_T)] + F$  where  $CE[f^*(V_T), f^0(V_T)]$  is just a constant.