Regensburger
DISKUSSIONSBEITRÄGE
zur Wirtschaftswissenschaft

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

Fractionally Integrated VAR Models with a Fractional Lag Operator and Deterministic Trends: Finite Sample Identification and Two-step Estimation

Rolf Tschernig[*], Enzo Weber[**], Roland Weigand[***]

January 2013

Nr. 471

JEL Classification: C32, C51

Key Words: fractional integration, long memory, maximum likelihood estimation, fractional lag operator

[*] Rolf Tschernig holds the chair of Econometrics at the Department of Economics and Econometrics at the University of Regensburg, 93040 Regensburg, Germany.
Phone: +49-941-943-2737, E-mail: rolf.tschernig[at]wiwi.uni-regensburg.de

[**] Enzo Weber holds the chair of Empirical Economics, especially Macroeconometrics and Labour Markets at the Department of Economics and Econometrics at the University of Regensburg, 93040 Regensburg, Germany, and is head of the research department Forecasts and Structural Analysis of the Institute for Employment Research (IAB), 90478 Nuremberg, Germany.
Phone: +49 -911-179-7643, E-mail: enzo.weber[at]iab.de

[***] Roland Weigand is a researcher at the research department Forecasts and Structural Analysis of the Institute for Employment Research (IAB), 90478 Nuremberg, Germany.
Phone: +49 -911-179-3291, E-mail: roland.weigand[at]iab.de

Fractionally Integrated VAR Models with a Fractional Lag Operator and Deterministic Trends: Finite Sample Identification and Two-step Estimation

Rolf Tschernig[a,*], Enzo Weber[a,b], and Roland Weigand[b]

[a]University of Regensburg, Department of Economics, D-93040 Regensburg
[b]Institute for Employment Research (IAB), D-90478 Nuremberg

January 2013

Abstract Fractionally integrated vector autoregressive models allow to capture persistence in time series data in a very flexible way. Additional flexibility for the short memory properties of the model can be attained by using the fractional lag operator of Johansen (2008) in the vector autoregressive polynomial. However, it also makes maximum likelihood estimation more difficult. In this paper we first identify parameter settings for univariate and bivariate models that suffer from poor identification in finite samples and may therefore lead to estimation problems. Second, we propose to investigate the extent of poor identification by using expected log-likelihoods and variations thereof which are faster to simulate than multivariate finite sample distributions of parameter estimates. Third, we provide a line of reasoning that explains the finding from several univariate and bivariate simulation examples that the two-step estimator suggested by Tschernig et al. (2010) can be more robust with respect to estimating the deterministic components than the maximum likelihood estimator.

Keywords fractional integration, long memory, maximum likelihood estimation, fractional lag operator.

*Corresponding author. Email: Rolf.Tschernig@wiwi.uni-regensburg.de. Phone: (+49) 941 943-2737

A former version was titled: Conditional versus approximate conditional maximum likelihood estimation in fractionally integrated vector autoregressive models with a fractional lag operator and deterministic trends.

1 Introduction

Fractionally integrated vector autoregressive (VAR) models have become a valuable extension of VAR models with integer orders of integration. Recently, Tschernig et al. (2010) introduced the fractional lag operator (see Johansen, 2008) into standard fractionally integrated VAR models in order to avoid certain shortcomings in impulse response analysis under long-run identification restrictions. The additional modeling flexibility due to the fractional lag operator, however, makes maximum likelihood estimation more difficult, in particular if deterministic components are included.

In this paper we first identify parameter settings for univariate and bivariate model versions that suffer from poor identification in finite samples and may therefore lead to estimation problems. Second, we propose to investigate the extent of poor identification by use of expected log-likelihoods and variations thereof which are faster to simulate than multivariate finite sample distributions of parameter estimates. Third, we provide a line of reasoning that explains the finding from several univariate and bivariate simulation examples that the two-step estimator suggested by Tschernig et al. (2010) can be more robust with respect to estimating the deterministic components than the maximum likelihood estimator. Within the maximum likelihood approach the estimator of the deterministic components and its properties depend on the simultaneously estimated fractional parameters. If the latter are subject to poor finite sample identification, the deterministic components may be poorly estimated which adds to the difficulties of estimating the fractional parameters. We therefore suggest to apply the two-step estimator in practice.

Section 2 briefly presents fractionally integrated VAR models with a fractional lag operator while Section 3 treats the maximum likelihood estimator for these models. In Section 4 we discuss prominent cases of univariate data generating processes that may be prone to poor identification. Section 5 explains how the expected log-likelihood and variations thereof allow to visualize the magnitude of poor identification for a given data generating process. In Section 6 we show how the problem of poor identification worsens once deterministic components have to be estimated as well. Finally, Section 7 extends the analysis to bivariate fractionally integrated VAR processes with a fractional lag operator.

2 Fractionally integrated VAR$_b$ models

In this section we consider fractional (vector) autoregressive processes with a fractional lag operator and deterministic terms. Using the fractional difference operator

$$(1-L)^d = \sum_{j=0}^{\infty} \frac{\Gamma(j-d)}{\Gamma(-d)\Gamma(j+1)} L^j,$$

where $\Gamma(\cdot)$ denotes the gamma function, the fractional lag operator L_b is defined (see Johansen, 2008) as

$$L_b = 1-(1-L)^b = c_1 L + c_2 L^2 + \cdots \quad \text{with } b > 0.$$

The degree of fractional integration b is required to be positive in order to guarantee that applying the fractional lag operator does not change the degree of integration. For $b = 1$ one obtains the standard lag operator L. A fractionally integrated k-variate vector autoregressive process with fractional lag operator (FIVAR$_b$) process for $\boldsymbol{x}_t$ is given by

$$\boldsymbol{A}(L_b)\boldsymbol{\Delta}(L,\boldsymbol{d})\boldsymbol{x}_t = \boldsymbol{u}_t, \quad \boldsymbol{u}_t \sim WN(\boldsymbol{0},\boldsymbol{\Sigma}), \quad t = 1,2,\ldots \tag{1a}$$

$$\boldsymbol{\Delta}(L,\boldsymbol{d}) := \operatorname{diag}\left((1-L)^{d_1},(1-L)^{d_2},\ldots,(1-L)^{d_k}\right). \tag{1b}$$

Here the errors $\boldsymbol{u}_t$ are multivariate white noise (WN) with mean zero and homoscedastic covariance matrix $\boldsymbol{\Sigma}$. For the pth-order vector autoregressive lag polynomial $\boldsymbol{A}(z) = \boldsymbol{I} - \boldsymbol{A}_1 z - \cdots - \boldsymbol{A}_p z^p$ we require the stability condition of Johansen (2008, Corollary 6) to hold. It provides a condition such that each element in the vector process $\boldsymbol{\eta}_t$ given by

$$\boldsymbol{A}(L_b)\boldsymbol{\eta}_t = \boldsymbol{u}_t, \quad t = \ldots,-2,-1,0,1,2,\ldots, \tag{2}$$

is $I(0)$. The roots of $|\boldsymbol{A}(z)| = 0$ have to be outside $\mathbb{C}_b$, which is the image of the unit circle under the mapping $f : z \mapsto 1-(1-z)^b$. This condition depends both on $\boldsymbol{A}(\cdot)$ and on b and can easily be checked once parameter values are given. Since $z = 1$ lies on $\mathbb{C}_b$ regardless of the value of b, stability of the $\boldsymbol{A}(L_b)$ polynomial excludes the unit root case and also implies nonsingularity of $\boldsymbol{A}(1)$.

Under the stability condition the parameter b adds some flexibility to the *short-run* properties of the process rather than having influence on the integration orders. For $b = 1$ one obtains a standard fractionally integrated VAR process (e.g. Nielsen, 2004a). In the sequel denote the parameter vector with all parameters of the model by $\boldsymbol{\theta} \in \boldsymbol{\Theta}_S$ where $\boldsymbol{\Theta}_S$ is the parameter space that contains all parameters that fulfil the stability condition.

2.1 Treatment of presample values

In order to obtain a solution to the process (1) for nonstationary $\boldsymbol{x}_t$ we make use of the truncated operator notation (Johansen, 2008, Appendices A.4, A.5). Let $\boldsymbol{\Pi}(L)$ denote an infinite matrix polynomial and $\mathbb{I}(\cdot)$ the indicator function. Then $\boldsymbol{\Pi}_+(L)\boldsymbol{x}_t = \mathbb{I}(t \geq 1)\sum_{i=0}^{t-1} \boldsymbol{\Pi}_i \boldsymbol{x}_{t-i}$ and $\boldsymbol{\Pi}_-(L)\boldsymbol{x}_t = \boldsymbol{\Pi}(L)\boldsymbol{x}_t - \boldsymbol{\Pi}_+(L)\boldsymbol{x}_t$.[1] Note also that $\boldsymbol{\Pi}_+(L)$ can always be inverted by expanding $\boldsymbol{\Pi}(z)^{-1}$ around zero and taking the first t terms.

The presample values $\boldsymbol{x}_t$, $t = 0, -1, \ldots$, are not modeled although they are allowed to be stochastic. Under mild conditions on the presample values, see e.g. Johansen and Nielsen (2012a), an equivalent representation of the FIVAR$_b$ process (1) is given by

$$\begin{aligned} \boldsymbol{A}_+(L_b)\boldsymbol{\Delta}_+(L,\boldsymbol{d})\boldsymbol{x}_t &= \boldsymbol{u}_t + \boldsymbol{m}_t, \quad t = 1, 2, \ldots, \\ \boldsymbol{m}_t &= \left[\boldsymbol{A}_+(L_b)\boldsymbol{\Delta}_-(L,\boldsymbol{d}) + \boldsymbol{A}_-(L_b)\boldsymbol{\Delta}_+(L,\boldsymbol{d}) + \boldsymbol{A}_-(L_b)\boldsymbol{\Delta}_-(L,\boldsymbol{d})\right]\boldsymbol{x}_t \end{aligned} \tag{3}$$

with its solution given by

$$\boldsymbol{x}_t = \boldsymbol{A}_+(L_b)^{-1}\boldsymbol{\Delta}_+(L,\boldsymbol{d})^{-1}\boldsymbol{u}_t + \boldsymbol{\mu}_t, \quad \boldsymbol{\mu}_t = \boldsymbol{A}_+(L_b)^{-1}\boldsymbol{\Delta}_+(L,\boldsymbol{d})^{-1}\boldsymbol{m}_t, \tag{4}$$

where $\boldsymbol{\mu}_t$ captures the impact of the presample values. Note that the forecast error impulse responses $\boldsymbol{\Psi}_h$ for horizon h can be computed from the truncated lag polynomial $\boldsymbol{\Psi}_+(L) = \boldsymbol{A}_+(L_b)^{-1}\boldsymbol{\Delta}_+(L,\boldsymbol{d})^{-1}$, replacing in (4) t by $t+h$.

2.2 Deterministic linear trends

The modeling of linear deterministic time trends is restricted to the sample of modeled data by assuming that

$$\boldsymbol{y}_t = \begin{cases} \boldsymbol{\nu}_0 + \boldsymbol{\nu}_1 t + \boldsymbol{x}_t, & \text{if } t \geq 1, \\ \boldsymbol{x}_t & \text{if } t \leq 0. \end{cases} \tag{5}$$

Therefore, the FIVAR$_b$ model with deterministic trends is given by

$$\boldsymbol{A}(L_b)\Delta(L,\boldsymbol{d})\left(\boldsymbol{y}_t - \boldsymbol{\nu}_0 - \boldsymbol{\nu}_1 t\right) = \boldsymbol{u}_t, \quad t = 1, 2, \ldots. \tag{6}$$

[1] Note that for a product $\boldsymbol{\Pi}(L) = \boldsymbol{\Phi}(L)\boldsymbol{\Theta}(L)$ of VAR polynomials $\boldsymbol{\Phi}(L)$, $\boldsymbol{\Theta}(L)$ one has

$$\begin{aligned} \boldsymbol{\Pi}_+(L) &= \boldsymbol{\Phi}_+(L)\boldsymbol{\Theta}_+(L), \\ \boldsymbol{\Pi}_-(L) &= \boldsymbol{\Phi}_+(L)\boldsymbol{\Theta}_-(L) + \boldsymbol{\Phi}_-(L)\boldsymbol{\Theta}_+(L) + \boldsymbol{\Phi}_-(L)\boldsymbol{\Theta}_-(L). \end{aligned}$$

By explicitly stating the dependence on presample values one obtains the representations

$$\boldsymbol{A}_+(L_b)\boldsymbol{\Delta}_+(L,\boldsymbol{d})\boldsymbol{y}_t = \boldsymbol{u}_t + \boldsymbol{A}_+(L_b)\boldsymbol{\Delta}_+(L,\boldsymbol{d})\,(\boldsymbol{\nu}_0+\boldsymbol{\nu}_1 t) + \boldsymbol{m}_t, \quad t=1,2,\ldots, \tag{7}$$

$$\boldsymbol{y}_t = \boldsymbol{A}_+(L_b)^{-1}\boldsymbol{\Delta}_+(L,\boldsymbol{d})^{-1}\boldsymbol{u}_t + (\boldsymbol{\nu}_0+\boldsymbol{\nu}_1 t) + \boldsymbol{\mu}_t, \quad t=1,2,\ldots. \tag{8}$$

If it is assumed that all presample values are zero, $\boldsymbol{x}_t = 0$, $t \leq 0$, then $\boldsymbol{m}_t = \boldsymbol{\mu}_t = \boldsymbol{0}$, $t = 1, 2, \ldots$.

3 Maximum likelihood estimation

In the following we state the conditional maximum likelihood estimator for given presample values $\boldsymbol{x}_t$, $t = 0, -1, \ldots, T_p$. For brevity it is called maximum likelihood estimator throughout the paper. Let $\boldsymbol{\alpha} = \operatorname{vec}(\boldsymbol{A}_1, \ldots, \boldsymbol{A}_p)$ denote the vector of all VAR coefficients and $\boldsymbol{Y} = \left(\boldsymbol{y}_{-T_p}, \ldots, \boldsymbol{y}_{-1}, \boldsymbol{y}_0, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_T\right)$ the vector of observable presample and sample values. For an observed time series, the maximum likelihood estimators for the general model (6) allowing for deterministic trends is given by

$$\left(\hat{\boldsymbol{d}}, \hat{b}, \hat{\boldsymbol{\nu}}_0, \hat{\boldsymbol{\nu}}_1, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\Sigma}}\right) = \arg \max_{\boldsymbol{d}, b, \boldsymbol{\nu}_0, \boldsymbol{\nu}_1, \boldsymbol{\alpha}, \boldsymbol{\Sigma}} L(\boldsymbol{d}, b, \boldsymbol{\nu}_0, \boldsymbol{\nu}_1, \boldsymbol{\alpha}, \boldsymbol{\Sigma}; \boldsymbol{Y}), \tag{9}$$

where the maximization is carried out over an appropriate parameter space. Further, we assume normally distributed errors in order to derive the log-likelihood function explicitly:

$$\begin{aligned} &L(\boldsymbol{d}, b, \boldsymbol{\nu}_0, \boldsymbol{\nu}_1, \boldsymbol{\alpha}, \boldsymbol{\Sigma}; \boldsymbol{Y}) = \\ &-\frac{Tk}{2}\log 2\pi - \frac{T}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\sum_{t=1}^{T} \boldsymbol{u}_t(\boldsymbol{d}, b, \boldsymbol{\nu}_0, \boldsymbol{\nu}_1, \boldsymbol{\alpha})' \boldsymbol{\Sigma}^{-1} \boldsymbol{u}_t(\boldsymbol{d}, b, \boldsymbol{\nu}_0, \boldsymbol{\nu}_1, \boldsymbol{\alpha}), \end{aligned} \tag{10}$$

where $\boldsymbol{u}_t(\boldsymbol{d}, b, \boldsymbol{\nu}_0, \boldsymbol{\nu}_1, \boldsymbol{\alpha})$ is obtained by rearranging (6) as

$$\begin{aligned} \boldsymbol{u}_t(\boldsymbol{d}, b, \boldsymbol{\nu}_0, \boldsymbol{\nu}_1, \boldsymbol{\alpha}) = &\underbrace{\Delta(L,\boldsymbol{d})\,(\boldsymbol{y}_t - (\boldsymbol{\nu}_0+\boldsymbol{\nu}_1 t))}_{\boldsymbol{z}_t(\boldsymbol{d},\boldsymbol{\nu}_0,\boldsymbol{\nu}_1)} - \boldsymbol{A}_1 \underbrace{L_b\Delta(L,\boldsymbol{d})\,(\boldsymbol{y}_t - (\boldsymbol{\nu}_0+\boldsymbol{\nu}_1 t))}_{\boldsymbol{z}_{t-1}(\boldsymbol{d},b,\boldsymbol{\nu}_0,\boldsymbol{\nu}_1)} \\ &- \cdots - \boldsymbol{A}_p \underbrace{L_b^p\Delta(L,\boldsymbol{d})\,(\boldsymbol{y}_t - (\boldsymbol{\nu}_0+\boldsymbol{\nu}_1 t))}_{\boldsymbol{z}_{t-p}(\boldsymbol{d},b,\boldsymbol{\nu}_0,\boldsymbol{\nu}_1)}, \quad t=1,2,\ldots. \end{aligned} \tag{11}$$

Since $\boldsymbol{z}_t(\boldsymbol{d}, \boldsymbol{\nu}_0, \boldsymbol{\nu}_1)$ does not depend on $\boldsymbol{\alpha}$ it follows from (11) that for given $\boldsymbol{d}$, b, $\boldsymbol{\nu}_0$, $\boldsymbol{\nu}_1$ the VAR coefficient matrices $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_p$, if unrestricted, are obtained by least squares, regressing $\boldsymbol{z}_t(\boldsymbol{d}, \boldsymbol{\nu}_0, \boldsymbol{\nu}_1)$ on $\boldsymbol{z}_{t-1}(\boldsymbol{d}, b, \boldsymbol{\nu}_0, \boldsymbol{\nu}_1), \ldots, \boldsymbol{z}_{t-p}(\boldsymbol{d}, b, \boldsymbol{\nu}_0, \boldsymbol{\nu}_1)$. Together with concentrating out $\boldsymbol{\Sigma}$ this simplifies the maximization considerably and leads to the concentrated log-likelihood

$$L^c(\boldsymbol{d}, b, \boldsymbol{\nu}_0, \boldsymbol{\nu}_1; \boldsymbol{Y}) = -\frac{Tk}{2}(\log 2\pi + 1) - \frac{1}{2}\log\left|\sum_{t=1}^{T} \boldsymbol{u}_t(\boldsymbol{d}, b, \boldsymbol{\nu}_0, \boldsymbol{\nu}_1)\boldsymbol{u}_t(\boldsymbol{d}, b, \boldsymbol{\nu}_0, \boldsymbol{\nu}_1)'\right|, \tag{12}$$

that can be maximized instead of (10). For example, in the bivariate case the maximization of this concentrated log-likelihood function only requires a nonlinear optimization on seven parameters independently of the VAR order p.

Available related asymptotic results are derived by Johansen and Nielsen (2012a) in a framework of fractional cointegration under the assumption of no deterministic trends. While in their setting identical d_is are assumed, this does not imply that the individual fractional orders of integration are identical due to the possibility of trivial cointegration with $\beta = (1, 0)'$. However, we exclude the possibility of fractional cointegration by the stability restriction on the VAR polynomial $\boldsymbol{A}(L_b)$. For the standard FIVAR model under the stability condition with $b = 1$ asymptotic results are derived by Nielsen (2004a) and implied by Hualde and Robinson (2011). It remains to be checked if Hualde and Robinson (2011) covers the FIVAR$_b$ model for $b \neq 1$ and zero presample values.

With respect to univariate processes the asymptotic behavior of the maximum likelihood estimator (9) is investigated by Nielsen (2004b), who requires $b = 1$, a stable AR polynomial, and zero presample values but allows for a deterministic trend, and by Johansen and Nielsen (2010), who allow for nonzero presample values, $b \neq 1$ as well as for unit roots in the AR polynomial. They exclude deterministic trends, however. Johansen and Nielsen (2012b) derive the asymptotic second-order bias due to presample values for a pure univariate fractional process $\Delta^d x_t = u_t$. With the exception of Johansen and Nielsen (2010) and Johansen and Nielsen (2012a) all mentioned results require the $\boldsymbol{A}(L_b)$ polynomial to be stable.

4 Poor finite sample identification in case of univariate processes

It is well known that parameter estimation may be more difficult if the parameter values of the data generating process are close to the boundary of the parameter space where all or some parameters are not identified. To give an example, let the data generating process be a smooth transition autoregressive process that is very close to a linear autoregressive process. Since the parameters of a smooth transition autoregressive model are not identified if the data generating process is in fact linear, a weak nonlinear structure may not be detectable in many samples such that estimation takes place *as if* parameters are not identified.

Another well known example are common roots in ARMA(p, q) models. If the true orders

are p_0 and q_0, then estimating an ARMA(p_0+1, q_0+1) model suffers from common roots which causes the parameters of the model to be not identified. In contrast to the previous example where a weak nonlinear structure in the data generating process causes estimation problems, it is too large a model order in the latter example.

When estimating FIVAR$_b$ processes similar scenarios may occur where the order of the autoregressive polynomial is too large, $p > p_0$:

a) $p_0 = 0$ and $p = 1$: Assume that the data generating process is a univariate white noise but a univariate FIVAR$_b$ model of order one is fitted,

$$(1 - a_1 L_b)\Delta^d x_t = u_t, \quad u_t \sim WN(0, \sigma^2), \quad t = 1, 2, \dots . \tag{13}$$

This model will henceforth be called a FAR$_b$ model of order one.

i) Inserting the true parameter values $a_{1,0} = 0$ and d_0 into the lag polynomials delivers $(1 - 0\, L_b)\Delta^{d_0} = \Delta^{d_0}$ so that b can take any positive value. Thus, b is not identified but d is.

ii) However, one may also insert $a_{1,0} = 1$. Then $(1 - L_b)\Delta^d = \Delta^{b+d}$ which has to be equal to Δ^{d_0}. Then there is a continuum of combinations for b and d for which $b + d = d_0$ holds.

b) $p_0 = 1$ and $p = 2$: Since $(1 - a_1 L_b - a_2 L_b^2)\Delta^d = (1 - \lambda_1 L_b)(1 - \lambda_2 L_b)\Delta^d$, inserting the true parameters $0 < a_{1,0} < 1$, $b_0 > 0$, and $d_0 > 0$ into a FAR$_b$ model with order 2 delivers two representations using $\lambda_1 = a_{1,0}$:

i) $\lambda_2 = 0$ and thus $a_2 = 0$, being equivalent to a FAR$_b$ model of order one.

ii) $\lambda_2 = 1$. Then

$$\begin{aligned}(1 - a_{1,0} L_{b_0})\Delta^{d_0} &= (1 - a_{1,0} L_{b_0})\Delta^{b_0}\Delta^{d_0 - b_0} = (1 - a_{1,0} L_{b_0})(1 - 1 L_{b_0})\Delta^{d_0 - b_0} \\ &= (1 - a_1 L_{b_0} - a_2 L_{b_0}^2)\Delta^{d_0 - b_0}\end{aligned}$$

where $a_1 = a_{1,0} + 1$, $a_2 = -a_{1,0}$.

Note that in ii) the stability condition is violated since $\lambda_2 = 1$. In contrast to Case a), the parameters for each scenario are locally identified but not globally. Thus, one may expect the log-likelihood to be bimodal.

Even if the parameters are identified, they may be close to the boundary of (partial) non-identification. Assuming $p_0 = 1$ and $0 < a_{1,0} < 1$ such that the stability condition holds, it can still happen that the sample information is not sufficient to keep the estimate of a_1 reasonably far away from zero so that an identification problem may result in finite samples, resembling case a) i) above. Similarly, if $a_{1,0}$ is smaller than unity but not distinguishable from one in a finite sample, then

$$d = d_0 + b_0 - b \tag{14}$$

gives the approximate locations of estimated d and b. Only $d + b$ is appropriately identified. We call these scenarios *poor finite sample identification*.

For higher order FAR_b models, $p \geq 1$, poor finite sample identification may also result from the possibility that an estimate of b is close to zero. Then one has

$$a_j L_{b_0}^j = a_j \left(1 - (1 - L)^{b_0}\right) \approx 0 \tag{15}$$

independently of the value of a_j. Therefore, if a DGP that is white noise is modeled by a FAR_b process with order p, the AR parameters a_j, $j = 1, 2, \ldots, p$ are poorly identified for values of b close to zero. Hence also for this reason it is important to aim at using correct lag orders.

Depending on the purpose of the model, poor finite sample identification may or may not be harmful. If one is interested in the long-run dynamics implied by the degree of integration, then it can be very problematic. As an example, compare the degree of integration implied by a process with a_1 close to one to the case $a_1 = 1$ for identical d_0 and b_0: while the former is an $I(d_0)$ process, the latter is an $I(d_0 + b_0)$ process.

Whether poor finite sample identification is an issue for a given sample may be checked by investigating the log-likelihood function on the relevant range of the parameters of interest. For the FAR_b model (13) of order one, one may plot the concentrated log-likelihood on a grid for d and b and visually check whether there are two peaks or mountain ridges that indicate poor finite sample identification.

In case one wants to check the potential of poor finite sample identification prior to sampling, one may simply "average" the log-likelihood function (10) over possible samples by taking expectations of the log-likelihood function at the true parameter vector. This delivers the expected log-likelihood function (16) further described in the next section.

Finally, if the true parameters are *far enough away* from the boundary of (partial) non-identification, poor finite sample identification should not be a major issue. Thus, in case of the

FAR$_b$ model (13), one may expect reasonable finite sample identification in case of $a_{1,0} = 0.6$.

5 Visualizing the expected log-likelihood

In the previous section it was argued that poor finite sample identification may be checked without referring to any sample by investigating the expected log-likelihood function. This suggestion differs from the commonly used method for investigating finite sample estimation properties by simulating the finite sample distribution of $\hat{\boldsymbol{\theta}}$. The expected likelihood computations take into account additional information about the shape of possible likelihood functions away from their maxima. Additionally, when jointly considering more than two parameters, the expected likelihood can be simulated faster than the joint density of the parameter estimator.

5.1 The expected log-likelihood

Let $E_0[\cdot]$ indicate that the expectation is taken with respect to the data generating process. Then the expected log-likelihood is given by

$$E_0\left[L(\boldsymbol{\theta};\boldsymbol{Y})\right] = \int \log f\left(\boldsymbol{Y};\boldsymbol{\theta}\right) f\left(\boldsymbol{Y};\boldsymbol{\theta}_0\right) d\boldsymbol{Y}. \tag{16}$$

Note that $\boldsymbol{\theta}_0$ may not be unique without the restriction $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}_S$.

Plotting contour lines or surfaces of the expected log-likelihood (16) is only possible if $\boldsymbol{\theta}$ is of length $l = 2$. One way to deal with the case $l > 2$ is to split the $(l \times 1)$ vector $\boldsymbol{\theta}$ into a (2×1) vector $\boldsymbol{\theta}_I$ that contains the two parameters of interest and a $((l-2) \times 1)$ vector $\boldsymbol{\theta}_{II}$ of all other parameters and then maximize (16) with respect to $\boldsymbol{\theta}_{II}$:

$$E_0\left[L(\boldsymbol{\theta}_I, \boldsymbol{\theta}_{II}^m(\boldsymbol{\theta}_I);\boldsymbol{Y})\right], \quad \text{where} \quad \boldsymbol{\theta}_{II}^m(\boldsymbol{\theta}_I) = \arg\max_{\boldsymbol{\theta}_{II}} E_0\left[L(\boldsymbol{\theta}_I, \boldsymbol{\theta}_{II};\boldsymbol{Y})\right]. \tag{17}$$

For the univariate FAR$_b$ process (13) one may define $\boldsymbol{\theta}_I = (d_1, b)'$ and $\boldsymbol{\theta}_{II} = \left(d_2, a_1, \sigma^2\right)'$.

If a concentrated log-likelihood is available, it may be preferable to consider the expected concentrated log-likelihood. If $\boldsymbol{\theta}_{II}$ can be concentrated out completely, the *expected concentrated log-likelihood* is defined by

$$E_0\left[L\left(\boldsymbol{\theta}_I, \hat{\boldsymbol{\theta}}_{II}(\boldsymbol{\theta}_I;\boldsymbol{Y});\boldsymbol{Y}\right)\right] \tag{18}$$

through concentrating

$$\hat{\boldsymbol{\theta}}_{II}(\boldsymbol{\theta}_I;\boldsymbol{Y}) = \arg\max_{\boldsymbol{\theta}_{II}} L\left(\boldsymbol{\theta}_I, \boldsymbol{\theta}_{II};\boldsymbol{Y}\right). \tag{19}$$

Considering the expected concentrated log-likelihood allows to study the indirect effects that estimating $\boldsymbol{\theta}_{II}$ has on the estimation of $\boldsymbol{\theta}_I$, while such indirect effects are ignored in the maximization approach (17). Further, one may use (18) to compare various estimators for $\hat{\boldsymbol{\theta}}_{II}(\boldsymbol{Y})$ with respect to their influence on the estimation problem of $\boldsymbol{\theta}_I$. If an alternative estimator to (19) is used, then one obtains different objective functions for the estimation of $\boldsymbol{\theta}_I$ which can be compared.

Finally one may combine both approaches by concentrating out some parameters of $\boldsymbol{\theta}_{II}$ while taking the maximum with respect to the remaining ones. As an example with respect to (10) one may define $\boldsymbol{\theta}_I = (d_1, b)'$, $\boldsymbol{\theta}_{II} = d_2$, and $\boldsymbol{\theta}_{III} = (\mathrm{vec}(\boldsymbol{A})', \boldsymbol{\nu}_0', \boldsymbol{\nu}_1', \mathrm{vech}(\boldsymbol{\Sigma})')'$. Then the expected concentrated log-likelihood depending on $\boldsymbol{\theta}_I$ is given by

$$E_0\left[L\left(\boldsymbol{\theta}_I, \boldsymbol{\theta}_{II}^m(\boldsymbol{\theta}_I), \hat{\boldsymbol{\theta}}_{III}(\boldsymbol{\theta}_I, \boldsymbol{\theta}_{II}^m(\boldsymbol{\theta}_I; \boldsymbol{Y}); \boldsymbol{Y}\right)\right] \tag{20}$$

through concentrating

$$\hat{\boldsymbol{\theta}}_{III}(\boldsymbol{\theta}_I, \boldsymbol{\theta}_{II}; \boldsymbol{Y}) = \arg\max_{\boldsymbol{\theta}_{III}} L\left(\boldsymbol{\theta}_I, \boldsymbol{\theta}_{II}, \boldsymbol{\theta}_{III}; \boldsymbol{Y}\right)$$

and 'optimizing out'

$$\boldsymbol{\theta}_{II}^m(\boldsymbol{\theta}_I) = \arg\max_{\boldsymbol{\theta}_{II}} E_0\left[L\left(\boldsymbol{\theta}_I, \boldsymbol{\theta}_{II}, \hat{\boldsymbol{\theta}}_{III}(\boldsymbol{\theta}_I, \boldsymbol{\theta}_{II}; \boldsymbol{Y}); \boldsymbol{Y}\right)\right].$$

5.2 Expected concentrated log-likelihoods for FAR$_b$ processes of order one

In this section we use simulations to compute the expected concentrated log-likelihoods for various FAR$_b$ processes (13) of order one. We choose $d_0 = b_0 = 0.8$ and consider $a_{1,0} = 0.1, 0.6, 0.9$. For given $\boldsymbol{\theta}_I = (d, b)'$ and $\boldsymbol{\theta}_{II} = (a_1, \sigma^2)'$ the concentrated log-likelihood is easily computed as described in Section 3. In order to approximate the expected concentrated log-likelihood (18) we draw 100 realizations for given d and b. For obtaining contour plots we vary the parameters $d \in [-1, 1.5]$ and $b = [0.02, 1.5]$ using grids with step size of 0.02.

The magnitude of poor finite sample identification of d and b is visualized by the shape and size of the area with the largest values of the expected concentrated log-likelihood, which can well be seen from a plot with contour lines. For sample size $T = 250$ and $a_{1,0} = 0.9$ the contour lines of the expected concentrated log-likelihood are shown in the top panel of Figure 1. If $a_{1,0}$ were exactly 1, one would expect from (14) that the location of the largest values of

the expected concentrated log-likelihood is described by $\hat{d} \approx d_0 + b_0 - \hat{b} = 1.6 - \hat{b}$. From the top panel of Figure 1 it is seen that by letting $a_{1,0}$ deviate slightly from one, this location is shifted somewhat to $\hat{d} \approx 1.8 - \hat{b}$.

As argued in Section 4, poor identification in small samples is less of a problem if $a_{1,0}$ is neither close to zero nor to one, say $a_{1,0} = 0.6$, as can be seen from the middle panel of Figure 1. The lower panel of Figure 1 shows that the poor finite sample identification issue is again prominent if $a_{1,0} = 0.1$ and thus close to zero. From Case a) ii) in Section 4 it follows that if a_1 were exactly zero, $\hat{b}$ can float arbitrarily while $\hat{d} \approx 0.8$. This explains the upper ridge if a_1 is estimated close to zero. The lower ridge is explained by estimates $a_1 \approx 1$ leading to a negative trade-off between d and b.

6 Dealing with deterministic trends

In this section we investigate why the problem of poor finite sample identification worsens if deterministic trends are allowed in the model. In order to estimate the parameters of the FIVAR$_b$ model (6) with linear trends one may use the maximum likelihood estimator based on (10) in Section 3. In the following we provide arguments and a few simulation results that a two-step estimation is more robust.

6.1 Pitfalls in the maximum likelihood estimation

One may rewrite (6) so that $\boldsymbol{\nu}_0$ and $\boldsymbol{\nu}_1$ can be estimated by least squares *if* all the other parameters $\boldsymbol{d}$, b, and $\boldsymbol{\alpha}$ are given:

$$\boldsymbol{A}(L_b)\boldsymbol{\Delta}(L,\boldsymbol{d})\boldsymbol{y}_t = \boldsymbol{A}_+(L_b)\boldsymbol{\Delta}_+(L,\boldsymbol{d})\,1\;\boldsymbol{\nu}_0 + \boldsymbol{A}_+(L_b)\boldsymbol{\Delta}_+(L,\boldsymbol{d})\,t\;\boldsymbol{\nu}_1 + \boldsymbol{u}_t, \quad t = 1,2,\ldots,T. \tag{21}$$

Therefore the estimates of the deterministic components are influenced by both the long memory parameters $\boldsymbol{d}$ and the parameters $\boldsymbol{\alpha}$ and b determining the $I(0)$ dynamics. As a consequence, if one of the cases investigated in Sections 4 and 5.2 occurs where for a given sample size and DGP only $d+b$ is well identified while d and b are not, b may be estimated too large and d too small. To see the possible implications, consider a univariate DGP with $d_0 = 1$, $\nu_{0,0} \neq 0$, $\nu_{1,0} = 0$, and $a_{1,0}$ close to one. Suppose $\hat{d}$ takes the true value $d_0 = 1$, then the regression for estimating ν_0 and ν_1 corresponding to (21) is

$$(1 - a_1 L_b)\Delta y_t = (1 - a_1 L_b)\mathbb{I}(t = 1)\,\nu_0 + (1 - a_1 L_b)\mathbb{I}(t \geq 1)\,\nu_1 + u_t, \quad t = 1,2,\ldots,T. \tag{22}$$

Then ν_0 is estimated from only one observation, $t = 1$, and $\mathrm{Var}(\hat{\nu}_0) < \infty$ for $T \to \infty$. This is not problematic as the impact of the estimated ν_0 vanishes with growing sample size.

Due to poor finite sample information a_1 may be estimated close to one. Since for $a_{1,0} = 1$, only $d + b$ is identified by (14), $\hat{d}$ may be close to zero and $\hat{b}$ close to $1 + b_0$. Then, setting $d = 0$, no (fractional) differences are taken and the errors in

$$(1 - a_1 L_b) y_t = (1 - a_1 L_b) \mathbb{I}(t \geq 1)\ \nu_0 + (1 - a_1 L_b) t \mathbb{I}(t \geq 1)\ \nu_1 + u_t, \quad t = 1, 2, \ldots, T \tag{23}$$

exhibit a unit root. In this case it can be shown that $\mathrm{Var}(\hat{\nu}_0)$ increases with sample size T. Then, the estimate of ν_0 can be expected to remain influential. Thus, the implicit estimation properties for ν_0 and its impact on the other estimates crucially depend on the d estimate. It can be shown that this also holds for the ν_1 estimate.

Such difficulties due to a grossly wrong d estimates can be avoided if the poor finite sample identification problem is circumvented when estimating $\boldsymbol{\nu}_0$ and $\boldsymbol{\nu}_1$. To achieve this, we invert $\boldsymbol{A}(L_b)$ in (21), which is always possible if the stability condition holds, and consider the regression

$$\boldsymbol{\Delta}(L, \boldsymbol{d}) \boldsymbol{y}_t = \boldsymbol{\Delta}_+(L, \boldsymbol{d})\, 1\ \boldsymbol{\nu}_0 + \boldsymbol{\Delta}_+(L, \boldsymbol{d}) t\ \boldsymbol{\nu}_1 + \boldsymbol{\varepsilon}_t, \quad t = 1, 2, \ldots, T, \tag{24}$$

where the errors $\boldsymbol{\varepsilon}_t = \boldsymbol{A}(L_b)^{-1} \boldsymbol{u}_t$ are autocorrelated but $I(0)$. In this regression b does neither enter the regressand nor the regressor. Using the regression (24) works since the regressors are deterministic and therefore the autocorrelated errors do not matter much for estimating $\boldsymbol{\nu}_0$ and $\boldsymbol{\nu}_1$. Therefore, estimating the parameters of the deterministic terms on basis of (24) only requires knowledge of $\boldsymbol{d}$ which can be estimated by some semiparametric estimator that does not suffer from the finite sample identification problems. This leads directly to the two-step estimator described next.

6.2 Two-step estimation

Within the two-step estimation procedure, the deterministic components are estimated in the first step. In the second step the log-likelihood function is maximized after replacing the deterministic components by their estimates from the first step.

The first step involves running the regression (24). This amounts to computing the least

squares estimator from

$$(1-L)^{d_s} y_{s,t} = (1-L)_+^{d_s} 1\, \nu_{0,s} + (1-L)_+^{d_s}\, t\ \nu_{1,s} + \varepsilon_{s,t}, \qquad s = 1, 2, \ldots, k, \quad t = 1, 2, \ldots, T. \tag{25}$$

We obtain the following two-step estimator suggested by Tschernig et al. (2010):

First step:

- For each series $s = 1, 2, \ldots, k$ estimate the memory parameter d_s with the semiparametric exact local Whittle estimator of Shimotsu (2010) that allows for deterministic trends.
- In order to obtain $\tilde{\nu}_{0,s}$ and $\tilde{\nu}_{1,s}$, run the regression (25) after taking fractional differences based on the $\tilde{d}_s$ estimate. Do this for $s = 1, 2, \ldots, k$.

Second step: Maximize the log-likelihood function (10) with $\boldsymbol{\nu}_0 = \tilde{\boldsymbol{\nu}}_0$ and $\boldsymbol{\nu}_1 = \tilde{\boldsymbol{\nu}}_1$.

Note that in the bivariate case the nonlinear optimization in the second step only includes the three parameters d_1, d_2, b.

6.3 Expected concentrated log-likelihoods for FAR$_b$ processes of order one with deterministic trends

We consider the same three DGPs as in Section 5.2 but estimate a FAR$_b$ model allowing for a deterministic trend. Figures 2, 4, and 6 display the contour lines of the expected concentrated log-likelihood where the maximum likelihood estimator is used for concentrating out $\boldsymbol{\nu}_0$ and $\boldsymbol{\nu}_1$. Figures 3, 5, and 7 display the contour lines of the expected concentrated log-likelihood where the two-step estimator of Section 6.2 is used for estimating the deterministic terms. Comparing the contour lines of both estimators for $a_{1,0} = 0.9$ in Figures 2 and 3 shows that the region of highest expected concentrated log-likelihood is closer to the true values for the two-step estimator than for the maximum likelihood estimator as conjectured in the previous subsection. For the other FAR$_b$ with $a_{1,0} = 0.1$ whose estimation may also suffer from poor finite sample identification (close to Case a) in Section 4), the maximum likelihood estimator in Figure 6 is even more off in comparison to the two-step estimator in Figure 7. Even for the case of $a_{1,0} = 0.6$ which shows the least finite sample identification problems in case of no deterministic component the two-step procedure seems superior as can be seen from comparing Figures 4 and 5. When comparing the contour lines from the two-step procedure

with the results in case of known deterministic terms one observes that estimating deterministic components is costly since the regions of highest expected concentrated log-likelihood no longer include the true parameters.

In order to check whether the identification problem really depends on the sample size we also computed the expected concentrated log-likelihoods for the FAR$_b$ with $a_{1,0} = 0.9$ for sample size $T = 1000$. Now both estimation procedures work much better as can be seen from Figures 8 and 9 although the two-step concentrated likelihood still shows a higher curvature. In sum, the two-step estimator works better as was expected from the reasoning in Sections 6.1 and 6.2. It remains to be investigated how these findings carry over to the bivariate FIVAR$_b$ model (6) which is done next.

7 Poor finite sample identification in bivariate FIVAR$_b$ processes

In this section we consider the FIVAR$_b$ processes (1) and (6).

7.1 Processes with poor finite sample identification

If the bivariate DGP has a diagonal VAR coefficient matrix

$$\left(\boldsymbol{I} - \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} L_b\right) \begin{pmatrix} (1-L)^{d_1} & 0 \\ 0 & (1-L)^{d_2} \end{pmatrix} \boldsymbol{x}_t = \boldsymbol{u}_t, \quad t = 1, 2, \ldots, \tag{26}$$

then maximizing the likelihood can be expected to be close although not identical to maximizing the likelihood of each univariate series. The latter generally deviates from the joint estimation since b is estimated for each series separately. If $\boldsymbol{d}$ and b are known, both estimators coincide. However, since b is identical across the individual series, one eigenvalue neither being close to one nor zero should help to estimate b sufficiently well so that the other eigenvalue does no longer cause problems even if it is close to one or zero.

For this reason, we expect estimation problems from poor finite sample identification if neither individual process helps to determine b and we expect that such a FIVAR$_b$ process inherits the poor finite sample identification problems from the individual FAR$_b$ processes. In the following we will investigate processes (26) with $\lambda = \lambda_1 = \lambda_2$.

Since λ is the eigenvalue (with multiplicity 2) of the autoregressive parameter matrix $\boldsymbol{A}$, it is interesting to investigate whether the poor finite sample identification problems diminish once

dependence between the two processes through the autoregressive polynomial is introduced while keeping the eigenvalue λ constant. In order to obtain a matrix $\boldsymbol{A}$ that exhibits the same eigenvalue but nonzero off-diagonal elements we apply the Jordan decomposition for real matrices (Lütkepohl, 1996, Section 6.2.1 (2)). Using a nonsingular $\boldsymbol{J} = \left(\begin{smallmatrix} a & c \\ b & d \end{smallmatrix}\right)$, a (2×2) matrix $\boldsymbol{A}$ with one eigenvalue λ can be written as

$$\boldsymbol{A} = \boldsymbol{J} \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix} \boldsymbol{J}^{-1} = \begin{pmatrix} \lambda - \frac{ac}{ad-bc} & \frac{a^2}{ad-bc} \\ -\frac{c^2}{ad-bc} & \lambda + \frac{ac}{ad-bc} \end{pmatrix}. \tag{27}$$

We conjecture that the poor finite sample identification problems diminish for a common eigenvalue λ once the dependence between the two processes is increased through off-diagonal elements in $\boldsymbol{A}$. We also conjecture that the impact of these problems matters more for the maximum likelihood estimator than for the two-step estimator like in the univariate case. In order to check these claims we investigate the expected concentrated log-likelihood functions for two DGPs next.

7.2 Expected concentrated log-likelihoods for FIVAR$_b$ processes of order one

All DGPs considered have normally distributed errors with $\boldsymbol{\Sigma}_0 = \left(\begin{smallmatrix} 1 & 0.5 \\ 0.5 & 1 \end{smallmatrix}\right)$, no deterministic trend, $\boldsymbol{\nu}_{0,0} = \boldsymbol{\nu}_{1,0} = \boldsymbol{0}$, and zero presample values. To check the claim that identical eigenvalues close to one without autoregressive dependence pose problems we consider:

DGPdiag the diagonal process (26) with $d_{1,0} = b_0 = 0.8$ and $d_{2,0} = 1.8$. The fractional parameter values resemble values that were estimated by Tschernig et al. (2010) for US real GDP and price data.

In order to investigate the effect of autoregressive dependence, we choose $\boldsymbol{J}$ in (27) such that the correlation between y_{1t} and y_{2t} is about -0.5 when $\boldsymbol{\Sigma} = \boldsymbol{I}$. This is the case for $\boldsymbol{J}_0 = \left(\begin{smallmatrix} 1 & -1 \\ 1 & 1 \end{smallmatrix}\right)$. This delivers the following DGP:

DGPdep

$$\left(\boldsymbol{I} - \begin{pmatrix} \lambda + \frac{1}{2} & +\frac{1}{2} \\ -\frac{1}{2} & \lambda - \frac{1}{2} \end{pmatrix} L_{0.8} \right) \begin{pmatrix} (1-L)^{0.8} & 0 \\ 0 & (1-L)^{1.8} \end{pmatrix} \boldsymbol{x}_t = \boldsymbol{u}_t, \quad t = 1, 2, \ldots \tag{28}$$

Suppose that there are now three parameters of interest: d_1, d_2 and b. For computing the expected log-likelihoods we first concentrate out all parameters except those three and then

plot contour lines where we take the maximum of the expected concentrated log-likelihood with respect to the third parameter. As an example choose in (20) $\boldsymbol{\theta}_I = (d_1, b)'$, $\boldsymbol{\theta}_{II} = d_2$, and $\boldsymbol{\theta}_{III} = (\boldsymbol{\alpha}', \boldsymbol{\nu}_0', \boldsymbol{\nu}_1', \text{vech}(\boldsymbol{\Sigma})')'$.

We first discuss the results of the diagonal **DGPdiag**. If no deterministic components are estimated, Figure 10 shows that the true parameters are close to the point of highest expected concentrated log-likelihood. Once deterministic components are allowed for in the estimation model, this is no longer the case for the maximum likelihood estimator as can be seen from the left column in Figure 11. In contrast, the two-step estimator still delivers reasonable results. This finding supports the reasoning of Section 7.1.

If dynamic spill-overs between the series are present as in the **DGPdep**, then it turns out from inspecting Figures 12 and 13 (for estimation without and with deterministic components, respectively) that the weak finite sample identification problem is less pronounced than in the diagonal case although b seems to be estimable less precisely than d_1 and d_2. Notably different effects of the estimators for the deterministic terms are not present in Figure 13.

8 Conclusion

We discussed finite sample estimation properties of fractionally integrated VAR models where high flexibility is introduced through the fractional lag operator and deterministic trends. We identify situations where identification may be poor in finite samples and verify these claims by plotting expected (concentrated) likelihoods. Deterministic trends aggravate the problems. A two-step estimator helps to circumvent at least part of the flaws which are faced if the maximum likelihood estimator is used. Subsequent work may be concerned with the asymptotic properties of the estimators. As a general recommendation, future empirical results using fractionally integrated time series techniques should be checked with respect to finite sample identification issues both to assess robustness of the results and to support an appropriate estimator choice.

References

Hualde, J. and Robinson, P. M. (2011), "Gaussian Pseudo-Maximum Likelihood Estimation of Fractional Time Series Models," *The Annals of Statistics*, 39, 3152–3181.

Johansen, S. (2008), "A Representation Theory for a Class of Vector Autoregressive Models for Fractional Processes," *Econometric Theory*, 24, 651–676.

Johansen, S. and Nielsen, M. Ø. (2010), "Likelihood Inference for a Nonstationary Fractional Autoregressive Model," *Jounal of Econometrics*, 158, 51–66.

— (2012a), "Likelihood Inference for a Fractionally Cointegrated Vector Autoregressive Model," *Econometrica*, 80, 2667–2732.

— (2012b), "The role of initial values in nonstationary fractional time series models," Working Paper 1300, Queen's Economics Department.

Lütkepohl, H. (1996), *Handbook of Matrices*, Wiley & Sons.

Nielsen, M. Ø. (2004a), "Efficient Inference in Multivariate Fractionally Integrated Time Series Models," *Econometrics Journal*, 7, 63–97.

— (2004b), "Efficient Likelihood Inference in Nonstationary Univariate Models," *Econometric Theory*, 20, 116–146.

Shimotsu, K. (2010), "Exact Local Whittle Estimation of Fractional Integration with Unknown Mean and Time Trend," *Econometric Theory*, 26, 501–540.

Tschernig, R., Weber, E., and Weigand, R. (2010), "Long-run Identification in a Fractionally Integrated System," Regensburger Diskussionsbeiträge zur Wirtschaftswissenschaft 447, University of Regensburg.

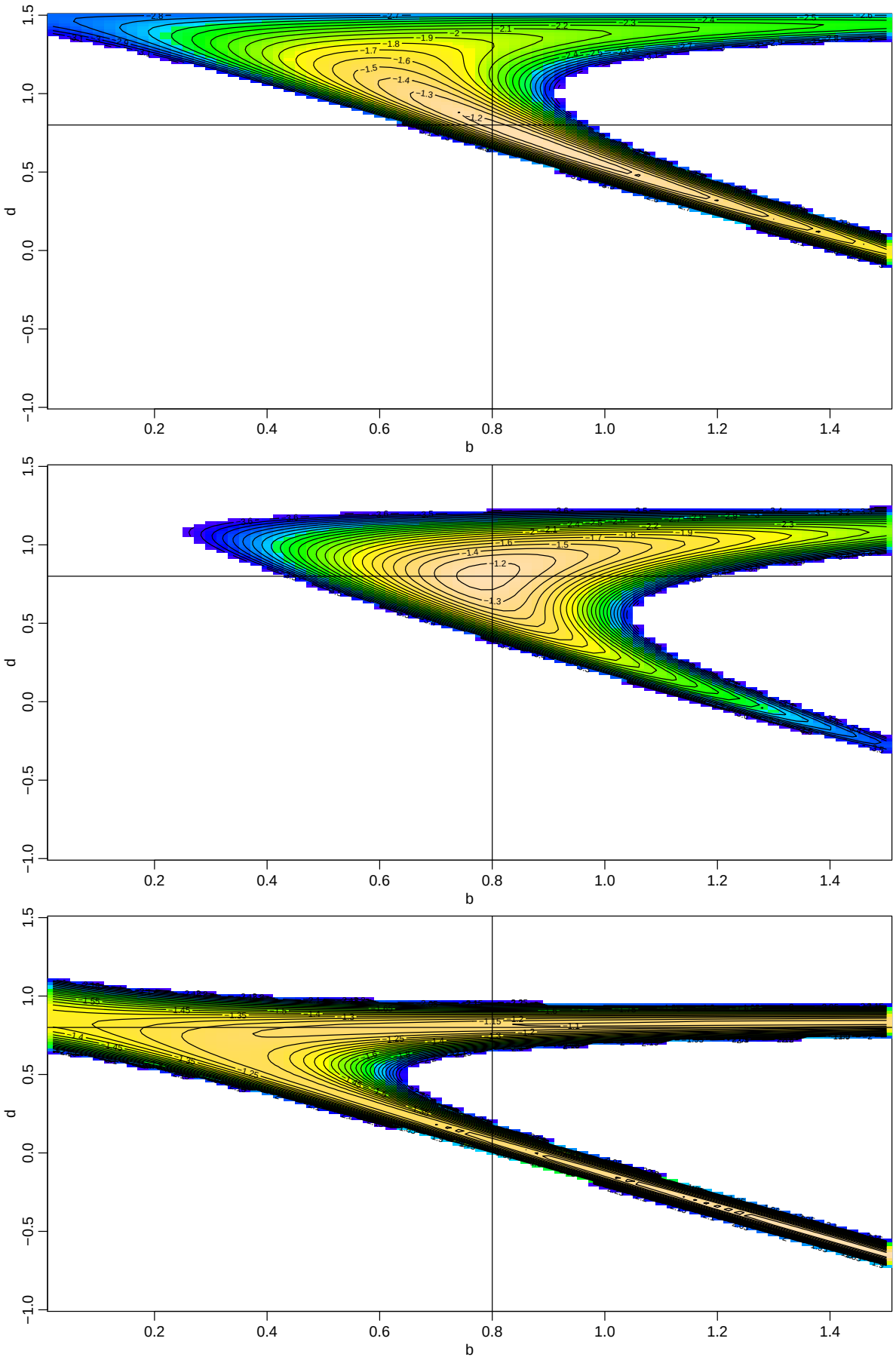

Figure 1: The expected concentrated log-likelihood function for a $\mathrm{FAR}_{0.8}$ process (13) with 250 observations, $d = 0.8$ and $a_1 = 0.9$ (above), $a_1 = 0.6$ (middle) and $a_1 = 0.1$ (below)

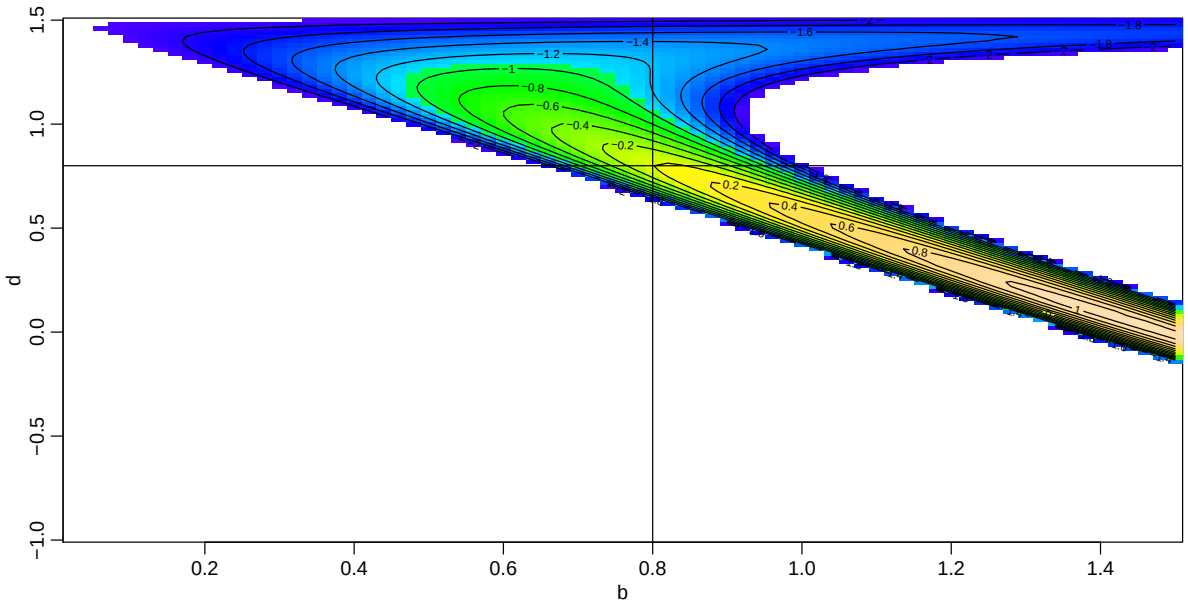

Figure 2: The expected concentrated log-likelihood function with estimated deterministic trend for a $\text{FAR}_{0.8}$ process (13) with $d = 0.8$ and $a_1 = 0.9$ with 250 observations.

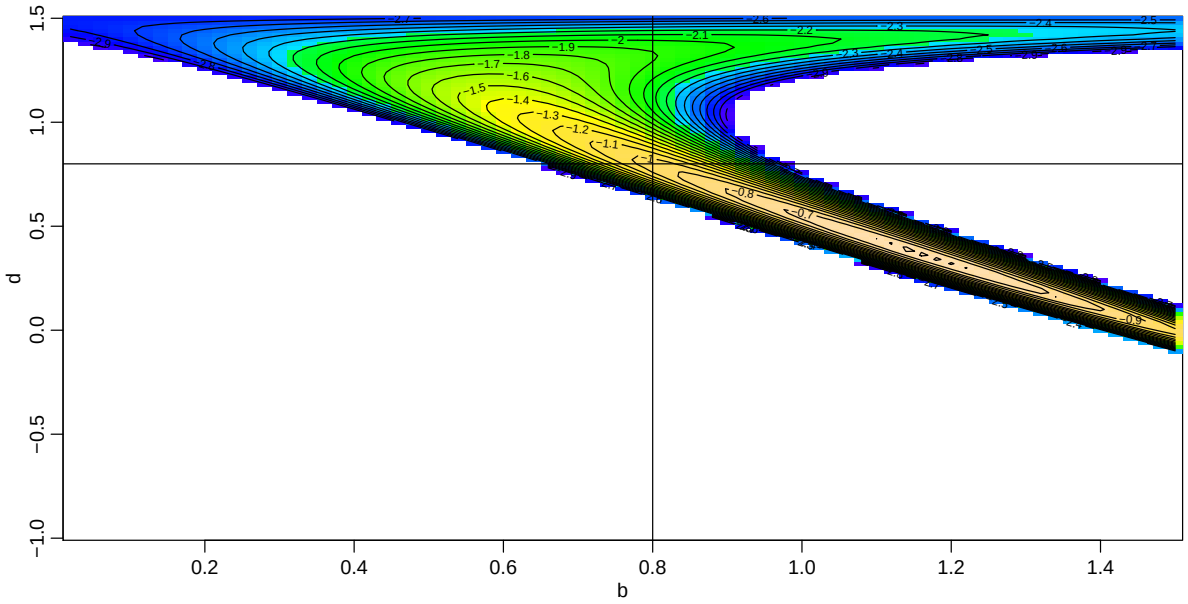

Figure 3: The expected concentrated **two-step approximate** log-likelihood function with estimated deterministic trend for a $\text{FAR}_{0.8}$ process (13) with $d = 0.8$ and $a_1 = 0.9$ with 250 observations.

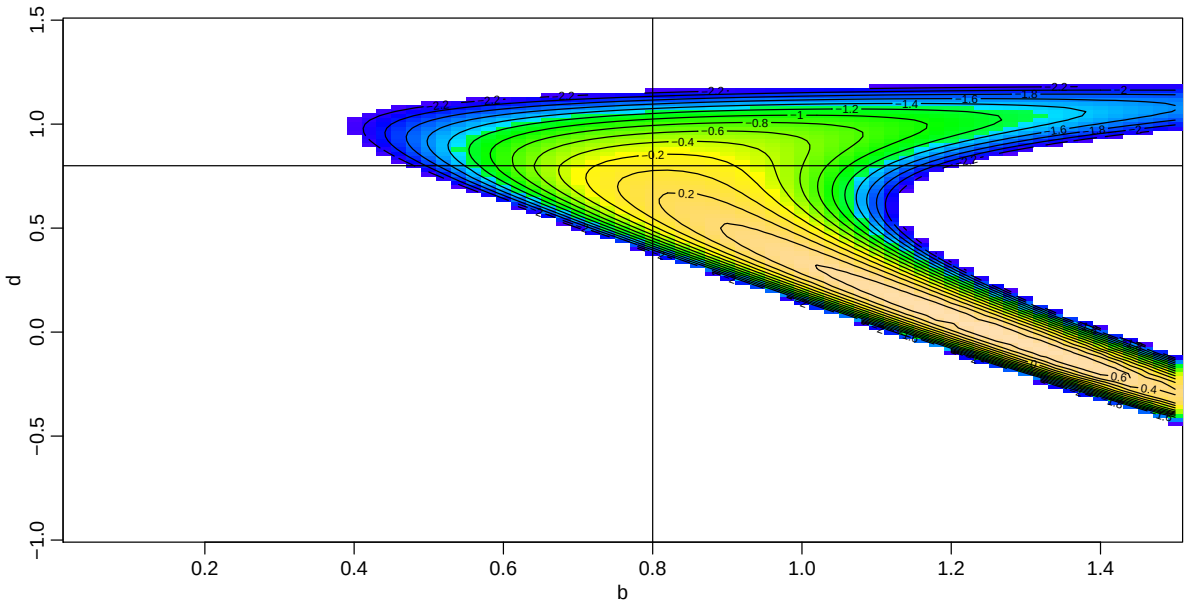

Figure 4: The expected concentrated log-likelihood function with estimated deterministic trend for a $\text{FAR}_{0.8}$ process (13) with $d = 0.8$ and $a_1 = 0.6$ with 250 observations.

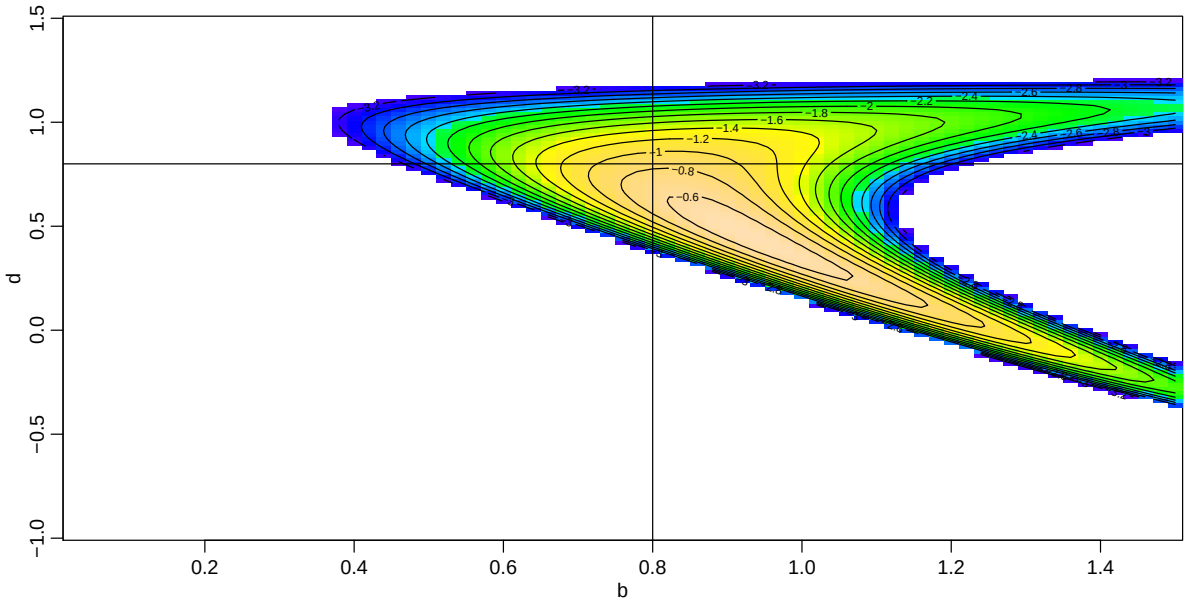

Figure 5: The expected concentrated **two-step approximate** log-likelihood function with estimated deterministic trend for a $\text{FAR}_{0.8}$ process (13) with $d = 0.8$ and $a_1 = 0.6$ with 250 observations.

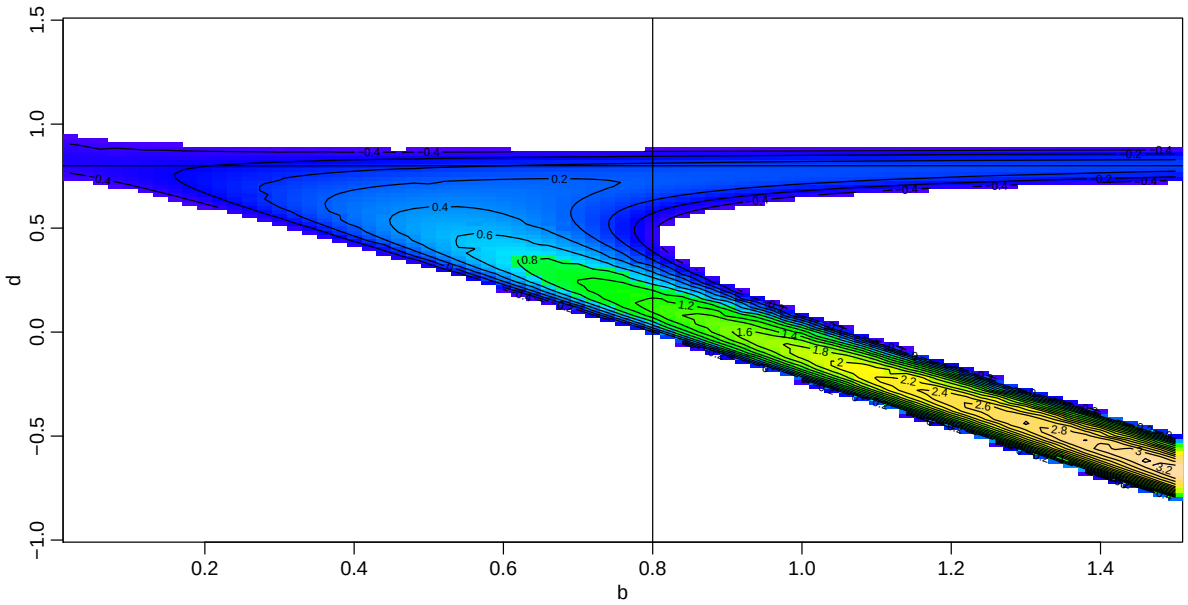

Figure 6: The expected concentrated log-likelihood function with estimated deterministic trend for a $\text{FAR}_{0.8}$ process (13) with $d = 0.8$ and $a_1 = 0.1$ with 250 observations.

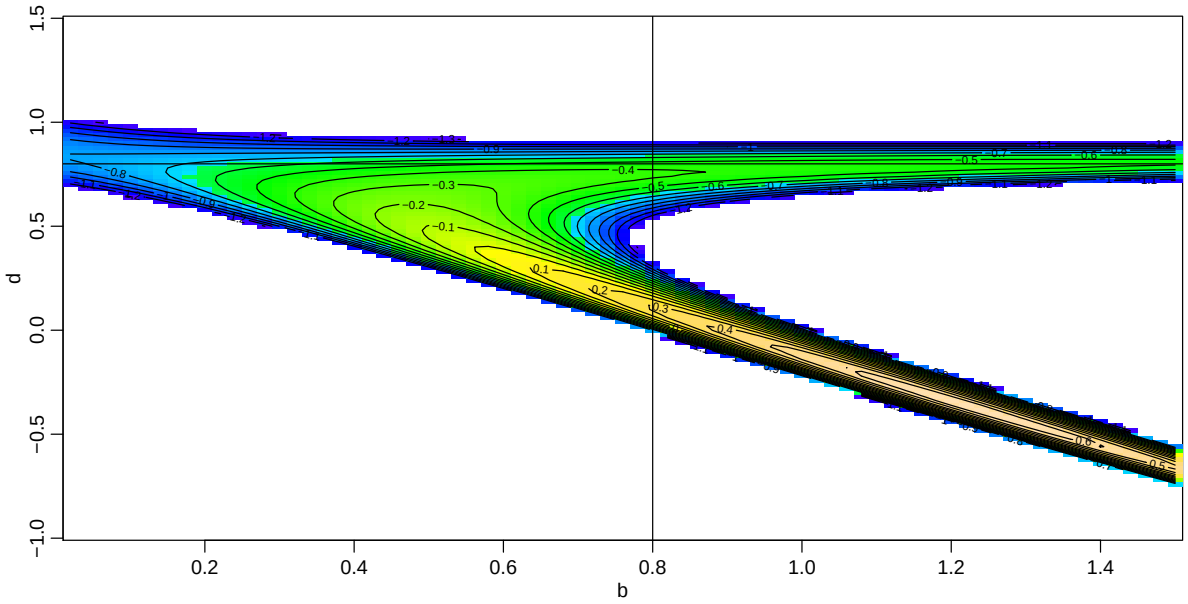

Figure 7: The expected concentrated **two-step approximate** log-likelihood function with estimated deterministic trend for a $\text{FAR}_{0.8}$ process (13) with $d = 0.8$ and $a_1 = 0.1$ with 250 observations.

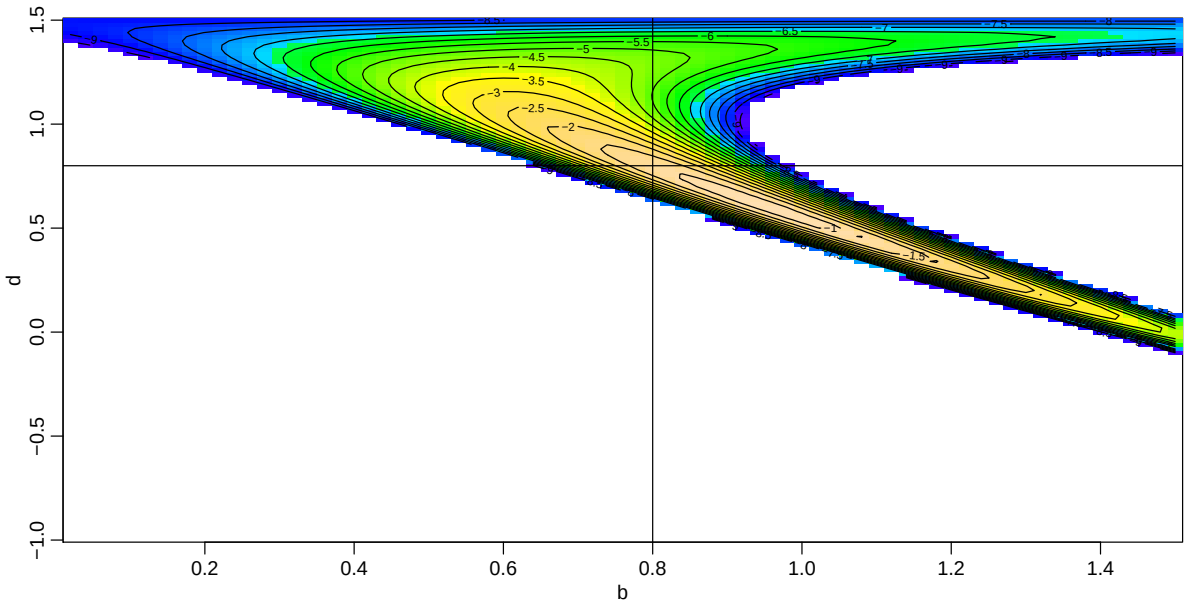

Figure 8: The expected concentrated log-likelihood function with estimated deterministic trend for a $\text{FAR}_{0.8}$ process (13) with $d = 0.8$ and $a_1 = 0.9$ with 1000 observations.

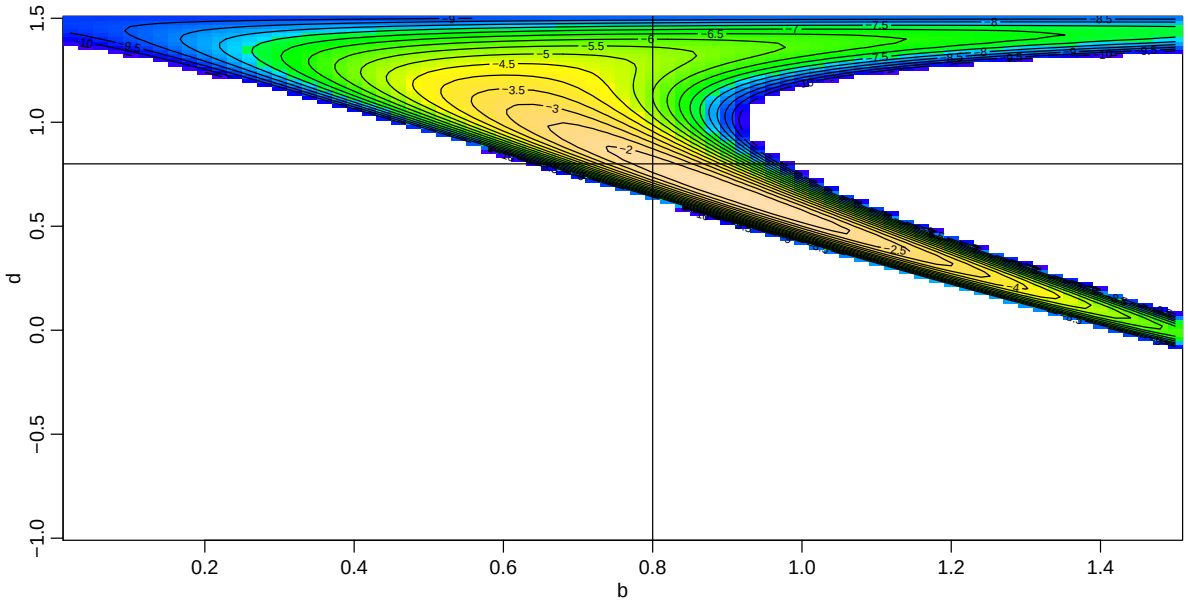

Figure 9: The expected concentrated **two-step approximate** log-likelihood function with estimated deterministic trend for a $\text{FAR}_{0.8}$ process (13) with $d = 0.8$ and $a_1 = 0.9$ with 1000 observations.

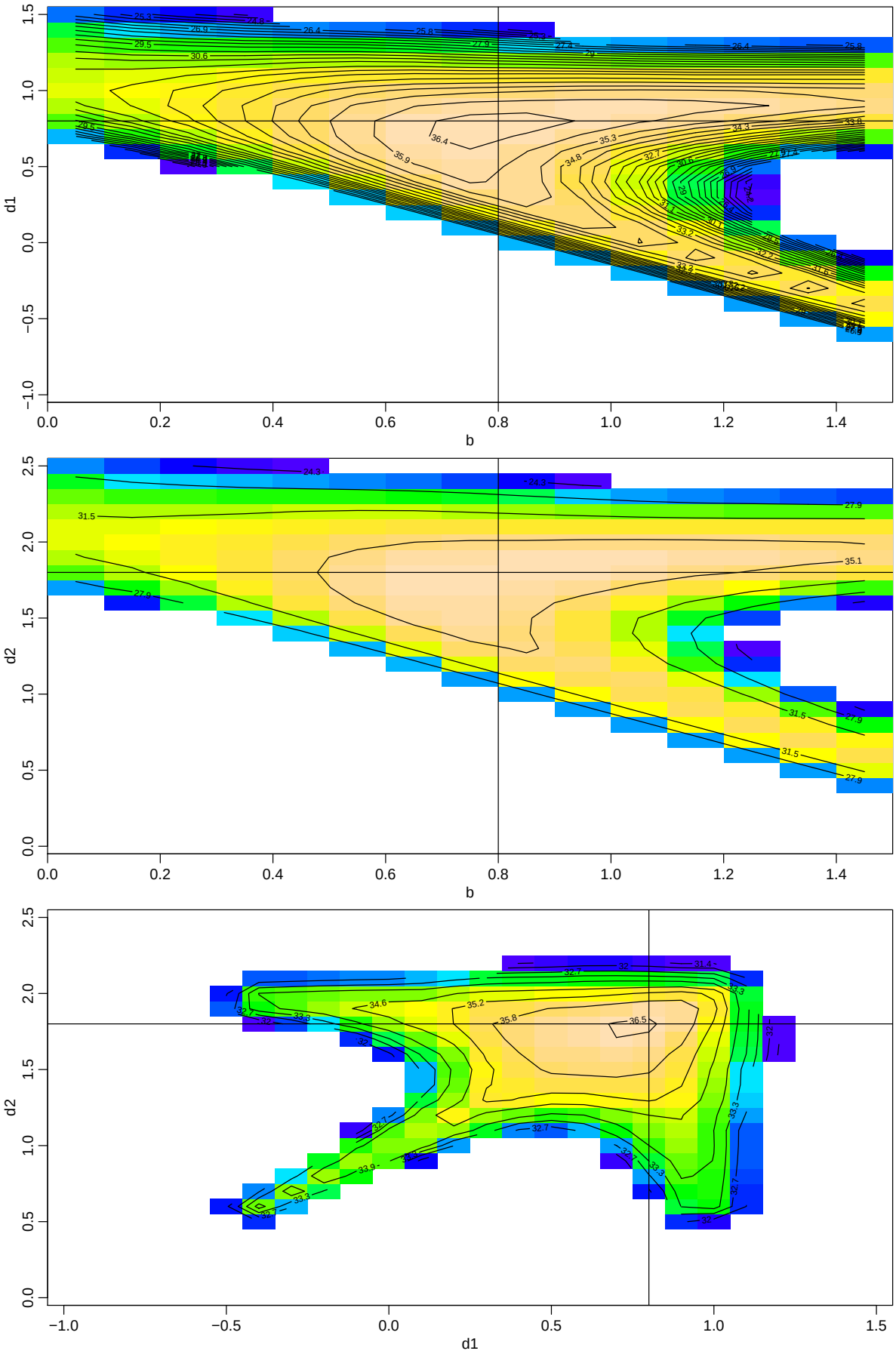

Figure 10: The expected concentrated log-likelihood function for the $\text{FIVAR}_{0.8}$ process **DG-Pdiag** (26) with $d_1 = 0.8$, $d_2 = 1.8$ and $a_1 = 0.8$ with 250 observations.

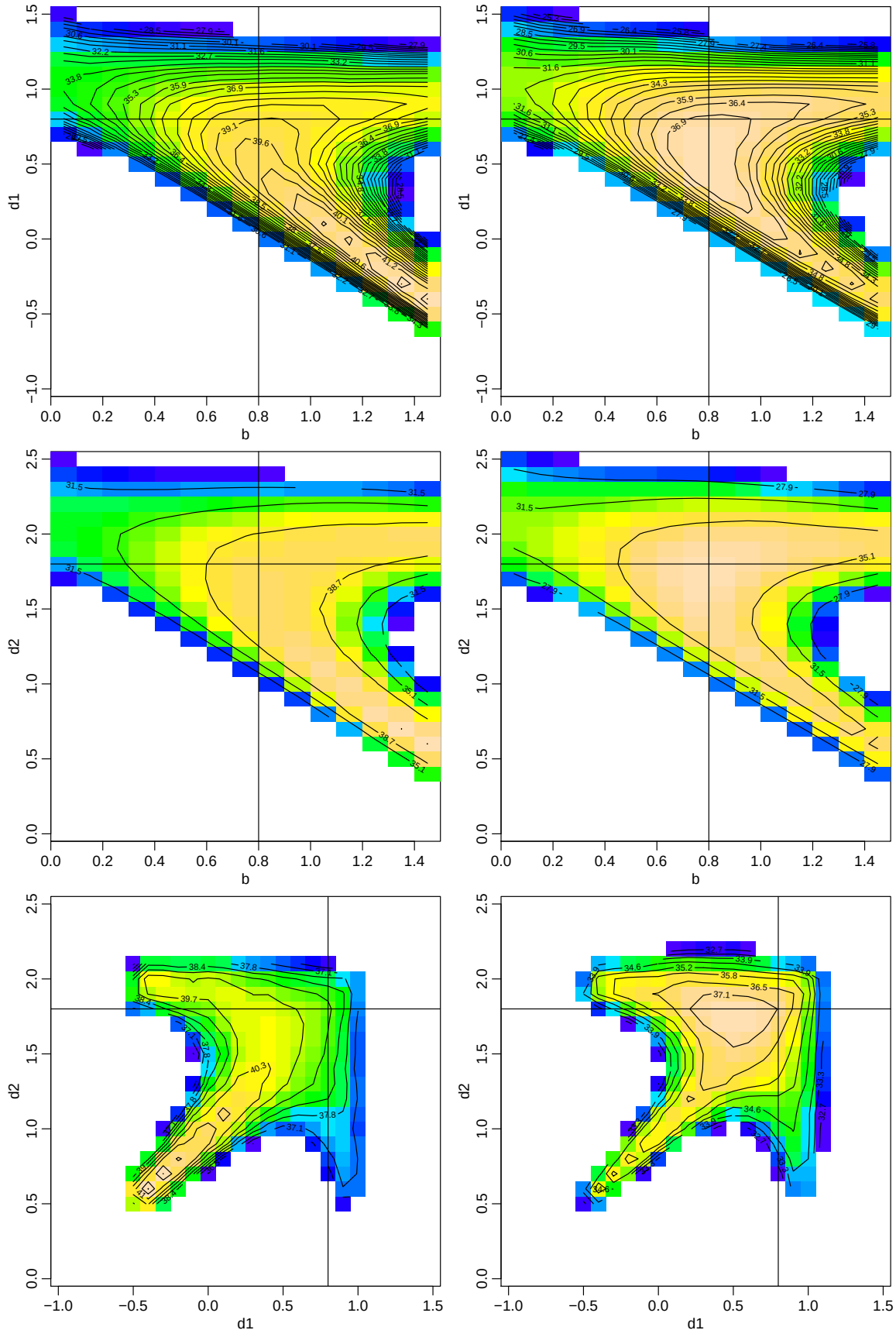

Figure 11: **Left column:** The expected concentrated log-likelihood function with estimated deterministic trends. **Right column:** expected concentrated **two-step approximate** log-likelihood function with estimated deterministic trends. Each for the $\text{FIVAR}_{0.8}$ process **DG-Pdiag** (26) with $d_1 = 0.8$, $d_2 = 1.8$ and $a_1 = 0.8$ with 250 observations.

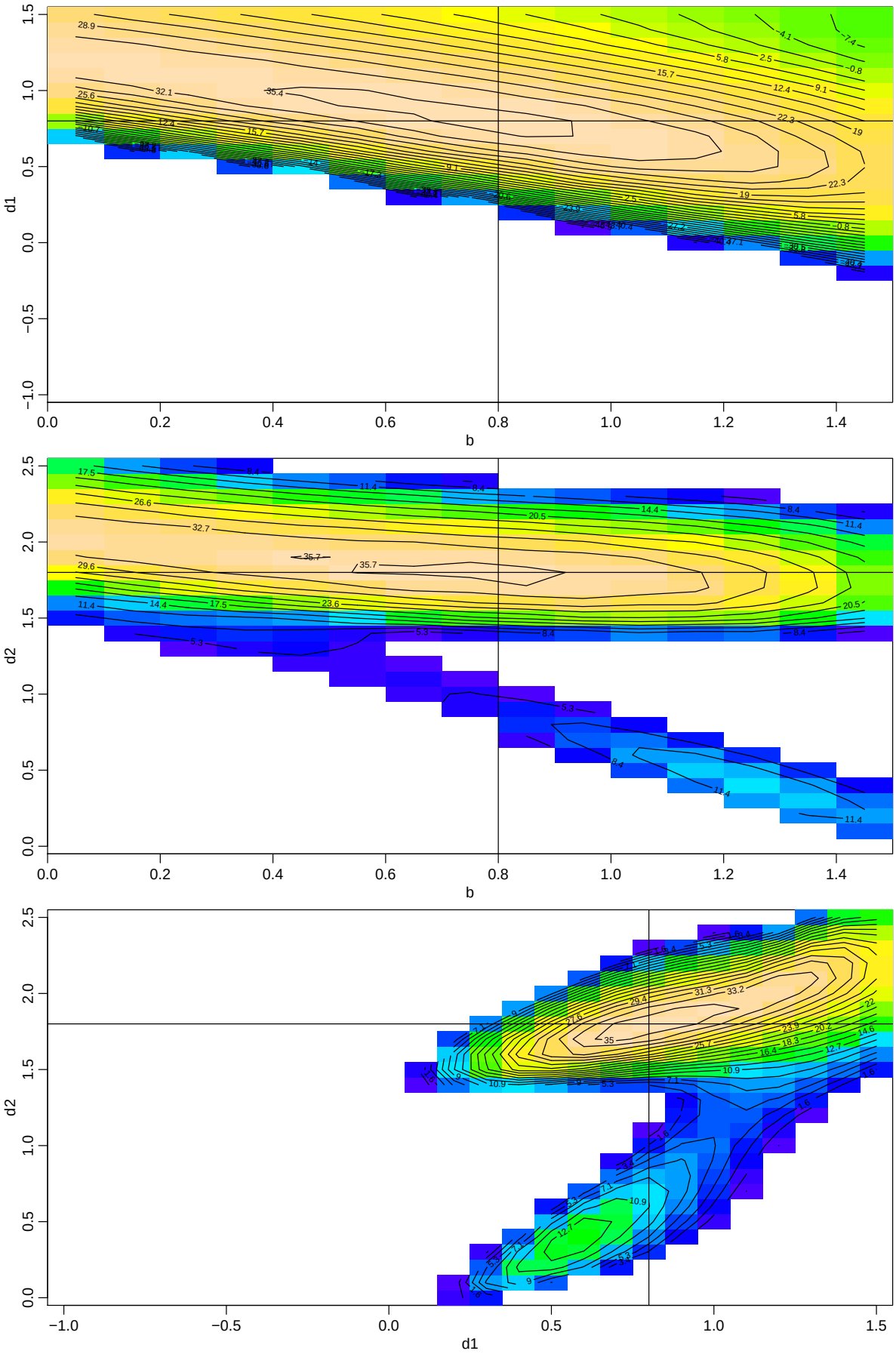

Figure 12: The expected concentrated log-likelihood function for the $\text{FIVAR}_{0.8}$ process **DG-Pdep** (28) with $d_1 = 0.8$, $d_2 = 1.8$ and $\lambda \underset{25}{} 0.8$ but autoregressive dependence with 250 observations.

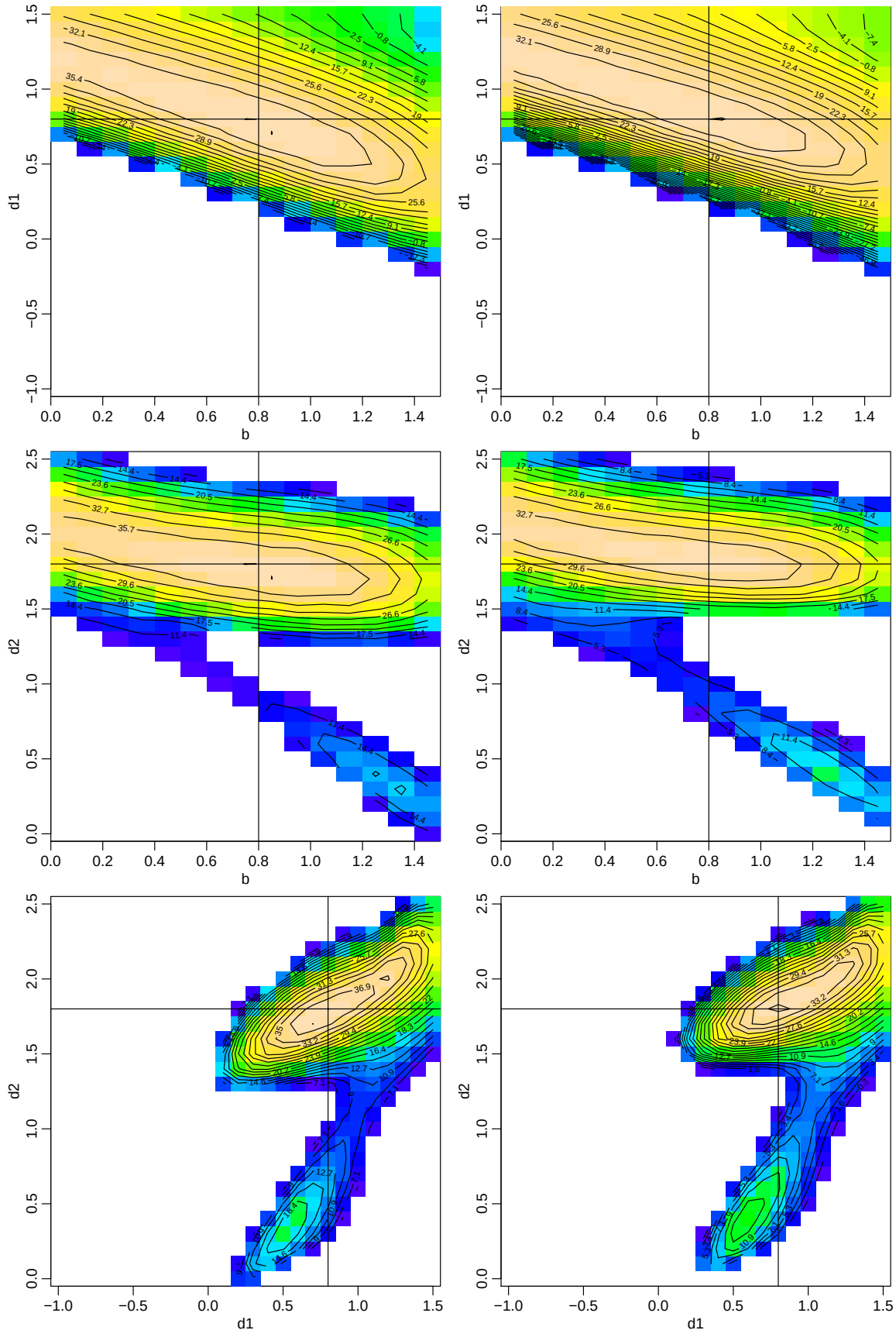

Figure 13: **Left column:** The expected concentrated log-likelihood function with estimated deterministic trends. **Right column:** expected concentrated **two-step approximate** log-likelihood function with estimated deterministic trends. Each for the FIVAR$_{0.8}$ process **DG-Pdep** (28) with $d_1 = 0.8$, $d_2 = 1.8$ and λ 0.8 but autoregressive dependence with 250 observations.