On threshold estimation in threshold vector error correction models

Friederike Greb, Tatyana Krivobokova, Stephan von Cramon-Taubadel, Axel Munk
University of Göttingen
Courant Research Centre „Poverty, Equity and Growth in Developing Countries"
Wilhelm-Weber-Str. 2, D-37073 Göttingen, Germany
fgreb@uni-goettingen.de

Paper prepared for presentation at the EAAE 2011 Congress
Change and Uncertainty
Challenges for Agriculture,
Food and Natural Resources

August 30 to September 2, 2011
ETH Zurich, Zurich, Switzerland

1 Introduction

When assessing the integration of spatially separated markets, agricultural economists typically analyze the transmission of price shocks between these markets. The law of one price suggests that the same assets traded in markets in different locations should have the same price. Otherwise traders would buy the good in the cheapest and sell it in the dearest market to profit from arbitrage. Hence, prices in the two different places would be drawn back together. However, this mechanism obviously only sets in if the price differential exceeds the transaction costs incurred by moving goods between markets. Threshold vector error correction models (TVECMs) have proven particularly adequate to model these dynamics. They became widely popular with Balke and Fomby's article on threshold cointegration (Balke and Fomby, 1997). A TVECM differs from a vector error correction model in that the coefficients are not fixed over time unless observations fall into the same regime. Regimes are determined by relating the price differential (or, more precisely, the error correction term) as the transition variable to the threshold parameters. These lend themselves to be interpreted as transaction costs. Consequently, in the context of modeling spatial arbitrage outlined above, three different regimes are defined by (i) the value of the transition variable being smaller than the lower threshold or assets being moved from the first to the second market, (ii) the transition variable taking a value in between the two thresholds or no trade occurring and (iii) the value of the transition variable exceeding the upper threshold or assets being moved from the second to the first market.

Estimation of the threshold parameters is complicated by the fact that they are not the only unknown model parameters. Of course, estimation in the presence of so-called nuisance parameters is a situation commonly encountered. However, for TVECMs these additional unknowns give rise to identification problems: On the one hand, if the difference between model parameters for two adjoining regimes equals zero, the threshold which separates them is not identified. On the other hand, if one of the threshold parameters is located at the boundary of the space in which the transition variable takes its values, the model parameters of the outer regime remain unidentified. The simplest approach to handle nuisance parameters is to maximize them out for fixed values of the parameters of interest – the thresholds – and base estimation on the profile likelihood function. The profile likelihood estimator is indeed the prevalent threshold estimator in the econometrics literature, see for example Hansen and Seo (2002). However, in certain situations this estimator performs poorly, especially if there is only a small number of observations available whereas the number of nuisance parameters is high. One would suspect threshold estimation to become increasingly difficult the closer the true parameters get to a situation in which identification problems occur; Balcombe et al. (2007) acknowledge this. As a matter of fact, simulations reveal that the profile likelihood estimator tends to fail when there is little difference in model parameters between regimes or thresholds divide the data into sets which are quite unequal in size. This becomes a very sensitive issue when modeling arbitrage. Clearly, the concept of arbitrage implies that the price differential will revert to the middle regime. Especially when integration is strong, so that deviations from the long-run equilibrium are corrected quickly, there will be few observations in the outer regimes, making estimation of the threshold and the error correction parameters in these regimes unreliable.

In this article, we suggest an alternative threshold estimator based on an empirical Bayes approach. The idea is to both account for nuisance parameter variability and regularize their estimates if necessary. Simulation studies confirm that this new estimator yields good results even in settings where the profile likelihood estimator is highly susceptible to faults.

The rest of this article is organized as follows. In the second section, we specify the model, discuss existing threshold estimators together with their deficiencies and present our modified estimator. Its performance is illustrated by a simulation study in the third section. The fourth section reviews an application of TVECMs in the analysis of spatial market integration for the case of four corn and soybean markets in North Carolina detailed in a seminal paper by Goodwin and Piggott (2001). We apply our empirical Bayes estimator to their data in order to contrast old and new estimation results.

2 Theory

2.1 The Model

Observations $p_t = (p_{t,1}, p_{t,2})^T$ of a two-dimensional time series generated by a TVECM with two thresholds or three different regimes, which are characterized by parameters $\rho_k, \theta_k \in \mathbb{R}^2$ and $\Theta_{km} \in \mathbb{R}^{2\times 2}$ for $k = 1, 2, 3$ and $m = 1, \ldots, M$, can be written as

$$\Delta p_t = \begin{cases} \rho_1 \gamma^T p_{t-1} + \theta_1 + \sum\limits_{m=1}^{M} \Theta_{1m} \Delta p_{t-m} + \varepsilon_t & , \quad \gamma^T p_{t-1} \leq \psi_1 \\ \rho_2 \gamma^T p_{t-1} + \theta_2 + \sum\limits_{m=1}^{M} \Theta_{2m} \Delta p_{t-m} + \varepsilon_t & , \quad \psi_1 < \gamma^T p_{t-1} \leq \psi_2 \\ \rho_3 \gamma^T p_{t-1} + \theta_3 + \sum\limits_{m=1}^{M} \Theta_{3m} \Delta p_{t-m} + \varepsilon_t & , \quad \psi_2 < \gamma^T p_{t-1} \end{cases} \tag{1}$$

p_t is assumed to form an $I(1)$ time series with cointegrating vector $\gamma \in \mathbb{R}^2$ and error-correction term $\gamma^T p_t$. The errors denoted by ε_t are taken to have expected value $\mathrm{E}(\varepsilon_t) = 0$ and covariance matrix $\mathrm{Cov}(\varepsilon_t) = \Omega \in (\mathbb{R}^+)^{2\times 2}$. $\psi_1, \psi_2 \in \mathbb{R}$ are called the threshold parameters. Note that although all of the regression coefficients may change between regimes, it is also possible for some of them to stay constant.

To express the model in matrix notation, we define vectors $\Delta P^{(i)}$, $\varepsilon^{(i)}$ by stacking the ith component of Δp_t and ε_t and $\mathbf{I}_{\{\gamma^T P \leq \psi_1\}}$,$\mathbf{I}_{\{\psi_1 < \gamma^T P \leq \psi_2\}}$, $\mathbf{I}_{\{\psi_2 < \gamma^T P\}}$ by stacking $\mathbb{I}_{\{\gamma^T p_{t-1} \leq \psi_1\}}$,$\mathbb{I}_{\{\psi_1 < \gamma^T p_{t-1} \leq \psi_2\}}$, and $\mathbb{I}_{\{\psi_2 < \gamma^T p_{t-1}\}}$, respectively. $\mathbb{I}_{\{\cdot\}}$ denotes the indicator function. For observations at N time points, an $N \times p$ matrix X is constructed by stacking rows $x_t^T = (\gamma^T p_{t-1}, 1, \Delta p_{t-1}^T, \ldots, \Delta p_{t-M}^T)$ of length $p = 2M + 2$, while $\beta_k^{(i)}$ is given as the ith column of the matrix $(\rho_k, \theta_k, \Theta_{k1}, \ldots, \Theta_{kM})^T$, $i = 1, 2$ and $k = 1, 2, 3$. With $\mathrm{diag}\left(\mathbf{I}_{\{\cdot\}}\right)$ defined as the diagonal matrix with entries $\mathbf{I}_{\{\cdot\}}$ in the diagonal and $E_2 \in \mathbb{R}^{2\times 2}$ the identity matrix, we are now able to write

$$\begin{aligned} \Delta P^{(i)} &= \mathrm{diag}\left(\mathbf{I}_{\{\gamma^T P \leq \psi_1\}}\right) X \beta_1^{(i)} + \mathrm{diag}\left(\mathbf{I}_{\{\psi_1 < \gamma^T P \leq \psi_2\}}\right) X \beta_2^{(i)} + \mathrm{diag}\left(\mathbf{I}_{\{\psi_2 < \gamma^T P\}}\right) X \beta_3^{(i)} + \varepsilon^{(i)} \\ &= X_1 \beta_1^{(i)} + X_2 \beta_2^{(i)} + X_3 \beta_3^{(i)} + \varepsilon^{(i)} \end{aligned}$$

to obtain the following representation of model (1),

$$\begin{aligned} \Delta P = \begin{pmatrix} \Delta P^{(1)} \\ \Delta P^{(2)} \end{pmatrix} = & \begin{pmatrix} X_1 \beta_1^{(1)} + X_2 \beta_2^{(1)} + X_3 \beta_3^{(1)} + \varepsilon^{(1)} \\ X_1 \beta_1^{(2)} + X_2 \beta_2^{(2)} + X_3 \beta_3^{(2)} + \varepsilon^{(2)} \end{pmatrix} \\ = & (E_2 \otimes X_1)\beta_1 + (E_2 \otimes X_2)\beta_2 + (E_2 \otimes X_3)\beta_3 + \varepsilon \\ = & X_\psi \beta + \varepsilon. \end{aligned}$$

A variety of modifications and restrictions of this specification of the TVECM are used in price transmission studies to represent different dynamics. A popular choice is the so-called continuous BAND-TVECM. Neglecting lags, this can be written as

$$\Delta p_t = \begin{cases} \rho_1 \left(\gamma^T p_{t-1} - \psi_1\right) + \varepsilon_t & , \quad \gamma^T p_t \leq \psi_1 \\ \varepsilon_t & , \quad \psi_1 < \gamma^T p_t \leq \psi_2 \\ \rho_3 \left(\gamma^T p_{t-1} - \psi_2\right) + \varepsilon_t & , \quad \psi_2 < \gamma^T p_t \end{cases} \tag{2}$$

In the middle regime, prices follow a random walk without drift while their difference (assuming that $\gamma^T p_t = p_{t,1} - p_{t,2}$) reverts to the threshold values in the outer regimes. Lo and Zivot (2001) provide details on various threshold cointegration models which appear in the context of modeling arbitrage.

We investigate both models of type (1) and (2). Instead of considering a q-dimensional TVECM with r thresholds, we focus on a two-dimensional TVECM with two thresholds. This setting is most frequently encountered in the analysis of market integration, and it allows to keep the exposition simple. We expect analogous results to hold for the more general model.

2.2 Commonly used threshold estimators

2.2.1 Profile likelihood estimator

As noted in the introduction, the most frequently encountered threshold estimator in the econometrics literature is the profile likelihood estimator (which is identical to the least squares estimator in case of independent Gaussian errors). Splitting all model parameters into parameters of interest ψ and nuisance parameters λ, the profile likelihood function $\mathcal{L}_p$ is constructed by maximizing out λ for given values of ψ, that is, $\mathcal{L}_p(\psi) = \mathcal{L}(\psi, \hat{\lambda}_\psi)$ with $\hat{\lambda}_\psi = \underset{\lambda \in \Lambda}{\operatorname{argmax}}\, \mathcal{L}(\psi, \lambda)$. In turn, the profile likelihood estimator $\hat{\psi}$ is defined as the argument maximizing $\mathcal{L}_p$, $\hat{\psi} = \underset{\psi \in \Psi}{\operatorname{argmax}}\, \mathcal{L}_p(\psi)$. Here, $\mathcal{L}$ constitutes the likelihood function of the model, Λ the nuisance parameter space and Ψ the space in which the parameter of interest takes its values. In our case, the thresholds form the parameter of interest, $\psi = (\psi_1, \psi_2)$, whereas the remaining unknowns in the model make up the nuisance parameter λ. Hence, Ψ will typically equal the space of all $(\psi_1, \psi_2) \in \mathbb{R}^2$ which satisfy $\min_{1 \leq t \leq N} (\gamma^T p_t) < \psi_1 < 0 < \psi_2 < \max_{1 \leq t \leq N} (\gamma^T p_t)$ whereas $\Lambda = \mathbb{R}^{6p} \times (\mathbb{R}^+)^{2 \times 2}$.

As ψ determines the distribution of observations into regimes, estimates $\hat{\lambda}_\psi$ are only well-defined for those ψ which allow for a sufficient number of observations to fall into each regime. Consequently, to guarantee adequate nuisance parameter estimates $\hat{\lambda}_\psi$, most authors simply require an arbitrary minimum proportion (or number) of observations per regime. Typically, a share of about 15% is imposed, see for example Andrews (1993). Clearly, this is inappropriate if the true fraction is less. The importance of taking such situation into account when modeling arbitrage has been emphasized above: If markets are well-integrated it is not unlikely that only 10% of the data is left in each of the outer regimes. In this case, an estimator based on the prerequisite of a minimum of 15% of the observations to be associated with each regime cannot be consistent as the parameter space Ψ is restricted to exclude the true thresholds. Moreover, in this case, the profile likelihood function $\mathcal{L}_p$ does not necessarily take its maximum at the boundary of the parameter space, since $\mathcal{L}_p$ is typically not unimodal, but jagged with a number of local maxima. For a sample run of model (2) with true lower threshold $\psi_1 = -5$, Figure 1 illustrates how the minimum condition influences the estimates of ψ_1. ψ_2 is kept at a fixed value to simplify visualization. For values of ψ_1 that leave less than the required share of observations in one of the regimes, the line is dotted; these parts are not considered when searching for the maximum of $\mathcal{L}_p$. While in this example, criteria of 5% or 10% result in identical estimates $\hat{\psi}_1 = -3.89$, which are relatively close to the true value -5, the estimate changes dramatically when at least 15% of the observations are forced to be present in each regime. This results in $\hat{\psi}_1 = 6$.

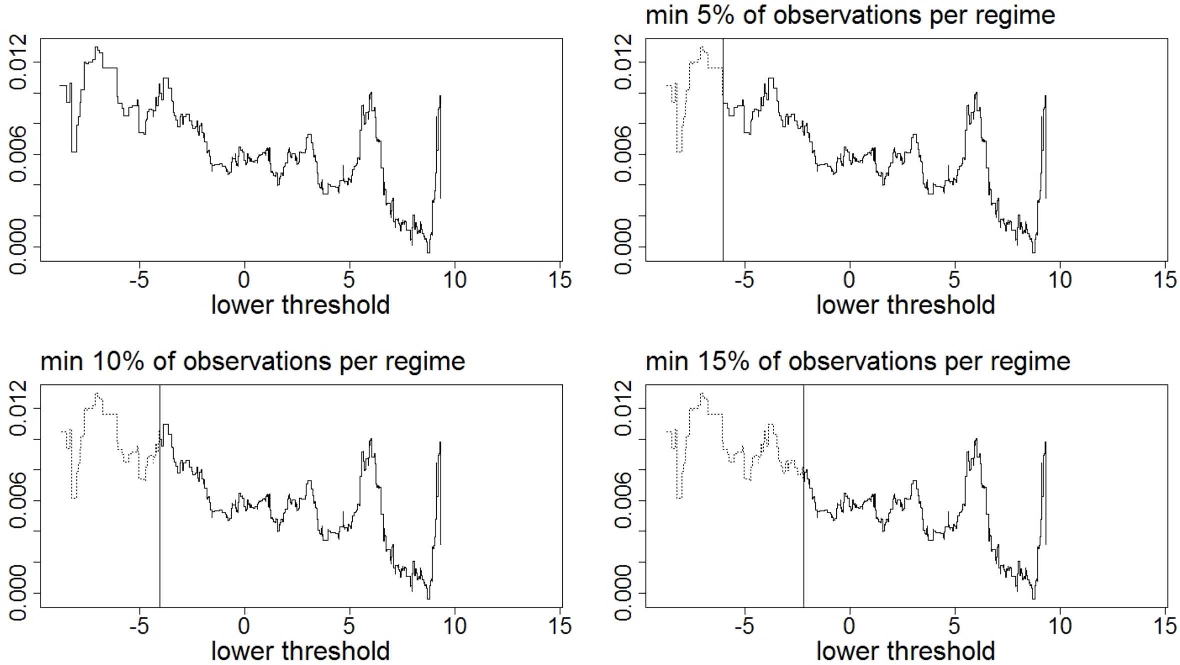

Figure 1: Effect of the minimum requirement on the profile likelihood function

However, it is not only the arbitrary constraint on the division of observations into regimes which might distort estimation results. Profile likelihood estimation does not account for the variability of the nuisance parameters. It is known that maximization of the profile likelihood function produces biased estimates as, in general, the profile likelihood function is not score unbiased. In addition, there are even examples of inconsistent profile likelihood estimates – the "many normal means" problem (McCullagh and Tibshirani, 1990) to mention just one. In case of the TVECM, treating estimated nuisance parameters as known and ignoring a loss of degrees of freedom can lead to biased threshold estimates in small samples.

2.2.2 Bayesian estimator

In contrast to profiling, which involves maximization, estimation based on integrating out nuisance parameters naturally accounts for their variability. This is the usual way of treating nuisance parameters in a Bayesian setting. Hence, Bayesian techniques might be particularly suitable for our problem. In fact, Bayesian estimators (BEs) have been employed in some price transmission studies: for example, Balcombe et al. (2007) and Balcombe and Rapsomanikis (2008) use a BE in the context of a TVECM. In the related setting of modeling financial arbitrage, Bayesian threshold estimators can be found in Forbes et al. (1999) and Balcombe (2006). However, none of these papers motivates the choice of a BE by its superior performance in comparison to the profile likelihood estimator. In contrast, Balcombe (2006) explicitly states that there is nothing which prohibits the model to be estimated using classical methods, but points to computational convenience as an argument in favor of a BE. All of these applications base their BEs on noninformative priors[1] and a quadratic loss function (which is equivalent to calculating the estimator as the posterior mean). Yet, it turns out that given a limited number of observations, the performance of the BE crucially depends on the priors selected; noninformative priors distort estimates.

For a TVECM $\Delta P = X_\psi \beta + \varepsilon$ with normal errors $\varepsilon \sim \mathcal{N}(0, \sigma^2 E_{2N})$ a noninformative (improper) prior $p(\beta, \sigma^2) \propto \sigma^{-2}$ is associated with a posterior

$$p(\psi|\Delta P, X) \propto \left(\Delta P - X_\psi \hat{\beta}_\psi\right)^T \left(\Delta P - X_\psi \hat{\beta}_\psi\right)^{-(N-2p)} \det\left(X_\psi^T X_\psi\right)^{-1/2}$$

for a uniform prior on the threshold parameters, $\psi \sim U(\Psi)$, and $\hat{\beta}_\psi = \left(X_\psi^T X_\psi\right)^{-1} X_\psi^T \Delta P$ (Lubrano, 2000). As in case of the profile likelihood function, we see that calculation of the posterior density requires that a sufficient number of observations be present in each regime in order to assure that $\left(X_\psi^T X_\psi\right)^{-1}$ and, hence, $\hat{\beta}_\psi$, are well-defined. Again, the suggestion found in the literature is to require a certain share of observations per regime. Balcombe et al. (2007), for example, specify that at least 20% must fall into each regime. Comparing the posterior density $p(\psi|\Delta P, X)$ to the profile likelihood function

$$\mathcal{L}_p(\psi) \propto \left(\Delta P - X_\psi \hat{\beta}_\psi\right)^T \left(\Delta P - X_\psi \hat{\beta}_\psi\right)^{-N}$$

it is easily seen that they differ by the term

$$\left(\Delta P - X_\psi \hat{\beta}_\psi\right)^T \left(\Delta P - X_\psi \hat{\beta}_\psi\right)^{2p} \det\left(X_\psi^T X_\psi\right)^{-1/2}.$$

In case ψ leaves only few observations in one of the regimes, the determinant $\det\left(X_\psi^T X_\psi\right)$ goes towards zero. Consequently, $p(\psi|\Delta P, X)$ becomes very large in comparison to $\mathcal{L}_p(\psi)$. Figure 2 illustrates this behavior for a sample run of model (2) with the true lower threshold value of -5. For ψ_2 fixed at its profile likelihood estimate $\hat{\psi}_2$, $\mathcal{L}_p(\psi_1, \hat{\psi}_2)$ is shown on the left. In this particular sample run, the profile likelihood estimator performs well. The middle plot depicts the posterior $p(\psi_1, \hat{\psi}_2|\Delta P, X)$ for the same run. Clearly, it increases drastically approaching the boundary of the parameter space. It takes its largest values exactly for values of ψ which are arbitrarily included or excluded from Ψ by varying the minimum requirement of observations per regime. As this is where most of the mass of the integral is acquired, the minimum criterion strongly affects its value, the threshold estimate calculated as the posterior mean, $\displaystyle\int \psi p(\psi|\Delta P, X) d\psi$.

2.3 Empirical Bayes Estimator

When rethinking the threshold estimator, there are good arguments to stick to a Bayesian framework. On the one hand, BEs naturally incorporate the variability of nuisance parameters. On the other hand, there is reason to believe that the thresholds in a TVECM

[1]For Forbes et al. (1999), this is only partially correct. In their specific application, they have at least some prior knowledge about the location of the thresholds which they include.

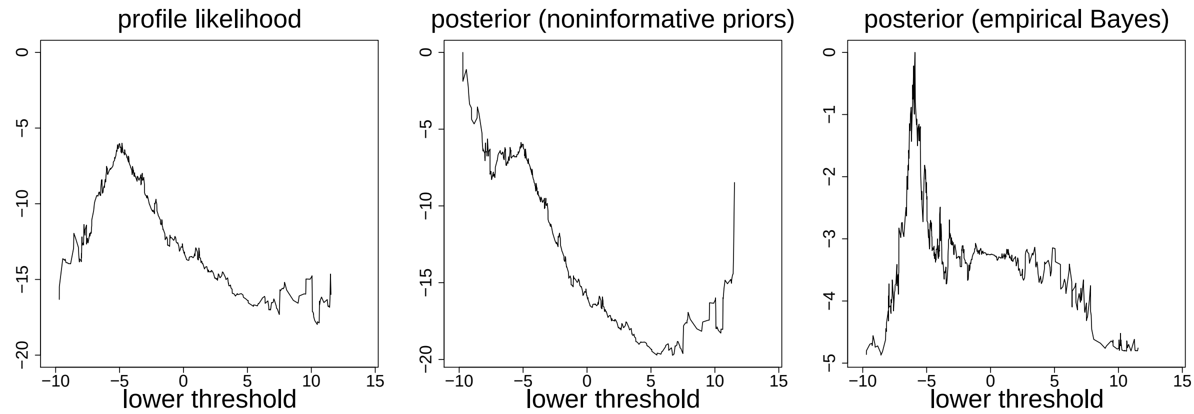

Figure 2: Comparison of different estimating functions for a sample run

are estimated more efficiently (at least asymptotically) by a BE. While Bayesian and maximum likelihood estimators (MLEs)[2] are asymptotically equivalent under certain regularity conditions (Van der Vaart, 2000, Chapter 10), this is not necessarily the case for nonregular problems as ours - it is not clear that the MLE is asymptotically efficient. The performance of Bayesian threshold estimators in TVECMs has not been studied yet. However, there exist results for some less intricate threshold models. Early assessments of the efficiency of a Bayesian threshold estimator for related situations include Ibragimov and Has' minskii's analysis of the estimation of a discontinuous signal in Gaussian white noise. They proved that the MLE is asymptotically less efficient than BEs with respect to a quadratic loss function (Ibragimov et al., 1981, Chapter 7). Yu (2007) first investigates the efficiency of the threshold estimator in a simple threshold regression model with i.i.d. observations. He shows that BEs are asymptotically efficient among all estimators in the locally asymptotically minimax sense, and strictly more efficient than the MLE. In a related paper, Chan and Kutoyants (2010) examine asymptotic properties of BEs for the threshold autoregression model. While they do not particularly focus on efficiency, they note that the limit variance of the BE is smaller than that of the MLE and comment on how to establish asymptotic efficiency for the BE.

While asymptotically the choice of priors does not have an impact, we have seen that it can critically influence estimation results in small samples. In particular, an estimator based on noninformative priors can perform poorly. Yet, when no prior knowledge is available, what are the other options? Roughly speaking, our idea is to exploit understanding of when estimation fails to regularize the posterior $p(\psi|\Delta P, X)$. We have observed that trouble frequently occurs in case of diminutive differences between model parameters in adjoining regimes. To formalize the notion of conceiving the model parameters in terms of their discrepancies between regimes, we reparametrize the model. Using the notation introduced in Section 2.1,

$$\begin{aligned}\Delta P &= (E_2 \otimes X_1)\beta_1 + (E_2 \otimes X_2)\beta_2 + (E_2 \otimes X_3)\beta_3 + \varepsilon \\ &= (E_2 \otimes X_1)(\beta_1 - \beta_2) + (E_2 \otimes X)\beta_2 + (E_2 \otimes X_3)(\beta_3 - \beta_2) + \varepsilon \\ &= (E_2 \otimes X_1)\delta_1 + (E_2 \otimes X)\beta_2 + (E_2 \otimes X_3)\delta_3 + \varepsilon.\end{aligned}$$

While holding on to a noninformative constant prior for β_2, we pick normal priors for δ_i,

[2]Since the profile likelihood estimator is identical to the MLE of the parameter of interest, we use these terms synonymously; the former expression stresses the presence of nuisance parameters.

$\delta_i \sim \mathcal{N}(0, \sigma^2_{\delta_i} E_{2p})$, $i = 1, 3$. For $\sigma^2_{\delta_i}$ tending towards infinity, these priors become noninformative, too. However, for small values $\sigma^2_{\delta_i}$, we introduce prior knowledge suggesting that δ_i takes values close to zero. This becomes manifest when looking at the predictor $\hat{\delta}$ for $\delta = (\delta_1^T, \delta_3^T)^T$ obtained as the posterior mode under the assumption of normal errors $\varepsilon \sim \mathcal{N}(0, \sigma^2 E_{2N})$,

$$\begin{aligned}\hat{\delta} &= \arg\max_{\delta} p(\delta|\Delta P, X) \\ &= \arg\min_{\delta} (\Delta P - W\tilde{\beta} - Z\delta)^T(\Delta P - W\tilde{\beta} - Z\delta) + \frac{\sigma^2}{\sigma^2_{\delta_1}}\delta_1^T\delta_1 + \frac{\sigma^2}{\sigma^2_{\delta_3}}\delta_3^T\delta_3\end{aligned}$$

where $W = E_2 \otimes X$, $\tilde{\beta} = (W^TW)^{-1}W^T\Delta P$ and $Z = (Z_1, Z_3) = (E_2 \otimes X_1, E_2 \otimes X_3)$. For large values of $\sigma^2_{\delta_i}$, the first term will drive estimation results $\hat{\delta}_i$, for small values it is the latter. Note that rather than arbitrarily restricting the parameter space Ψ to avoid the ill-posed problem of estimating m parameters from $n < m$ data points, which comes up near the boundary of Ψ, our choice of priors leads to the strategy of turning an ill-posed into a well-posed problem adopted by Tikhonov (Tikhonov et al., 1977). While a variety of techniques how to adequately choose the regularization parameters $\rho_i = \sigma^2/\sigma^2_{\delta_i}$ have been developed in the context of Tikhonov regularization, they naturally emerge in an empirical Bayesian setting. If the hyperparameters (the parameters of the prior distribution) in a Bayesian modeling framework are unknown, one option is to introduce another hierarchical level, that is, prior distributions for the hyperparameters; the other option is to estimate the hyperparameters. This is the empirical Bayes approach. We estimate $\sigma^2$and $\sigma^2_{\delta_i}$ for each ψ by maximizing the log-likelihood function $\ell_B(\psi, \sigma^2, \sigma^2_{\delta_1}, \sigma^2_{\delta_3}) = \log \mathcal{L}_B(\psi, \sigma^2, \sigma^2_{\delta_1}, \sigma^2_{\delta_3}) = \log p(\Delta P|X, \psi, \sigma^2, \sigma^2_{\delta_1}, \sigma^2_{\delta_3})$ of the model, [3]

$$\begin{aligned}\ell_B(\psi, \sigma^2, \sigma^2_{\delta_1}, \sigma^2_{\delta_3}) \propto -\frac{1}{2}\Bigg(&(2N - 2p)\log\sigma^2 + \log|V| + \log|W^TV^{-1}W| \\ &+ \frac{1}{\sigma^2}(\Delta P - W\hat{\beta})^TV^{-1}(\Delta P - W\hat{\beta})\Bigg)\end{aligned} \tag{3}$$

Here, $\hat{\beta} = (W^TV^{-1}W)^{-1}W^TV^{-1}\Delta P$ and $V = E_{2N} + \sigma^2_{\delta_1}/\sigma^2 Z_1Z_1^T + \sigma^2_{\delta_3}/\sigma^2 Z_3Z_3^T$. This leaves us with estimates $\hat{\sigma}^2$ and $\hat{\sigma}^2_{\delta_i}$, which are subsequently plugged into the posterior density

$$p(\psi|\Delta P, X, \hat{\sigma}^2, \hat{\sigma}^2_{\delta_1}, \hat{\sigma}^2_{\delta_3}) \propto p(\Delta P|X, \psi, \hat{\sigma}^2, \hat{\sigma}^2_{\delta_1}, \hat{\sigma}^2_{\delta_3})p(\psi|X) \tag{4}$$

to finally calculate the threshold estimator $\hat{\psi}$ as the posterior mean

$$\hat{\psi} = \int_{\Psi} \psi p(\Delta P|X, \psi, \hat{\sigma}^2, \hat{\sigma}^2_{\delta_1}, \hat{\sigma}^2_{\delta_3})d\psi$$

assuming a noninformative prior $\psi \sim U(\Psi)$ for ψ.

Looking at this posterior density $p(\psi|\Delta P, X)$ in comparison with that of a Bayesian model with exclusively noninformative priors for the sample run in Figure 2, it is evident that irregular behavior near the boundary has been corrected for. Hence, we are able to account for the variability of nuisance parameters, yet avoid problems at the boundary of the parameter space. We do not need to artificially constrain the parameter space Ψ by requiring a minimum number of observations to fall into each regime in order for our estimating function to be well-defined.

[3] Strictly speaking, σ^2 is not a hyperparameter. Nevertheless, treating it as such by estimating it simultaneously with $\sigma^2_{\delta_i}$ greatly facilitates computation.

3 Simulations

In this section we investigate the performance of our new threshold estimator and compare it with that of the profile likelihood estimator. Beforehand, a comment on computational issues. At least for model (1), any threshold parameters $\psi^A \neq \psi^B$ which produce the same distributions of observations into regimes result in identical values of the profile and posterior likelihood functions, $\mathcal{L}_p(\psi^A) = \mathcal{L}_p(\psi^B)$ and $\mathcal{L}_B(\psi^A) = \mathcal{L}_B(\psi^B)$. Hence, $\mathcal{L}_p$ and $\mathcal{L}_B$ are step functions which take only a finite number of different values. This means that for an appropriately specified grid (namely, consisting of the values observed for the transition variable), a grid search yields an exact maximum and the integral as well can be easily calculated accurately. In case of model (2), ψ enters the specification of the different regimes which complicates the matter. We will exemplify the effect of different choices for the grid in the next section. It is noteworthy that it is possible to explicitly calculate the posterior density as (4) in our Bayesian setup. This leaves only the challenge of computing estimates $\hat{\sigma}^2$, $\hat{\sigma}^2_{\delta_1}$ and $\hat{\sigma}^2_{\delta_3}$ in order to calculate the estimator. However, these are readily available using the R-package for mixed models, nlme. More precisely, the function lme.r with the default setting method="REML" yields the arguments $\hat{\sigma}^2$, $\hat{\sigma}^2_{\delta_1}$ and $\hat{\sigma}^2_{\delta_3}$ maximizing (3).

In our simulation study, we consider the following BAND-TVECM,

$$\Delta p_t = \begin{cases} \binom{-0.25}{0}\left(\gamma^T p_{t-1} + 5\right) + \begin{pmatrix} 0.2 & 0.2 \\ 0 & 0 \end{pmatrix} \Delta p_{t-1} + \varepsilon_t & , \quad \gamma^T p_t \leq -5 \\ \varepsilon_t & , \quad -5 < \gamma^T p_t \leq 10 \\ \binom{-0.25}{0}\left(\gamma^T p_{t-1} - 10\right) + \begin{pmatrix} 0.2 & 0.2 \\ 0 & 0 \end{pmatrix} \Delta p_{t-1} + \varepsilon_t & , \quad 10 < \gamma^T p_t \end{cases}$$

Errors are normally distributed, $\varepsilon \sim \mathcal{N}(0, \sigma^2 E_2)$ with $\sigma^2 = 1$. The cointegrating vector γ is assumed to be known to equal $\gamma = (1, -1)^T$. Results are summarized in Figure 3 and Table 1.

	Profile likelihood estimator				**Empirical Bayes estimator**			
	lower threshold		upper threshold		lower threshold		upper threshold	
		restr.		restr.		restr.		restr.
true value	-5	-5	10	10	-5	-5	10	10
mean	-0.932	-4.603	5.847	8.609	-2.709	-4.629	7.572	8.525
variance	41.166	3.921	39.604	10.269	5.573	1.444	8.445	4.594
MSE	57.663	4.073	56.802	12.19	10.812	1.58	14.327	6.764

Table 1: Comparison of profile likelihood and empirical Bayes estimator

The left-hand half of Figure 3 shows histograms for the lower threshold, while those for the upper threshold are arranged on the right-hand half. The upper row displays histograms for the profile likelihood estimator, and those for the empirical Bayes estimator are found below. The difference between histograms in the first/third columns and those in the second/fourth columns is that the latter deal with an estimator which presupposes knowledge that $\psi_1 < 0$ and $\psi_2 > 0$. In Table 1, this is indicated by "restr.". The

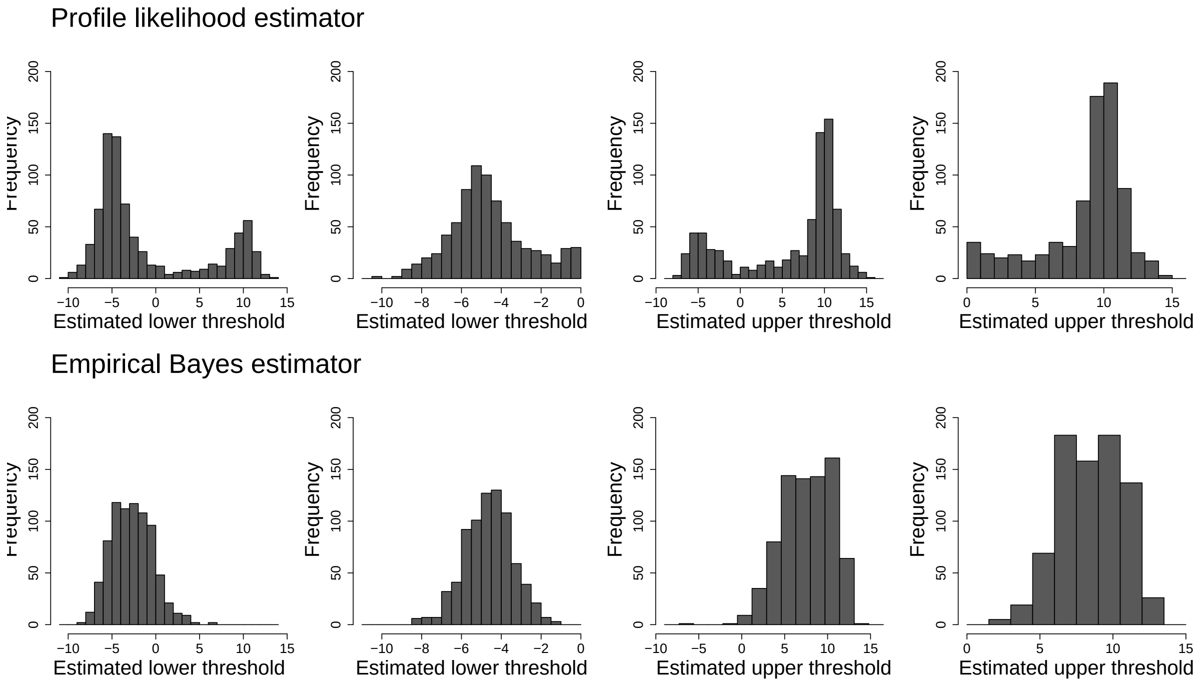

Figure 3: Histograms

former estimates are based on the parameters space of all $(\psi_1, \psi_2) \in \mathbb{R}^2$ which satisfy $\min\limits_{1 \le t \le N}(\gamma^T p_t) < \psi_1 < \psi_2 < \max\limits_{1 \le t \le N}(\gamma^T p_t)$, the latter on that of all those $(\psi_1, \psi_2) \in \mathbb{R}^2$ which additionally fulfill $\psi_1 < 0 < \psi_2$. It is clearly visible that the profile likelihood estimator overestimates the lower and underestimates upper threshold. The empirical Bayes estimates, in contrast, are less biased. Their variance and MSE are reduced drastically in comparison to the profile likelihood estimator.

4 Empirical Application

We next apply our empirical Bayes estimator to the data analyzed by Goodwin and Piggott (2001). They explore daily corn and soybean prices at important North Carolina terminal markets. These include Williamston, Candor, Cofield and Kinston for corn; and Fayetteville, Raleigh, Greenville and Kinston for soybeans. Observations range from 2 January 1992 until 4 March 1999. For each commodity, they evaluate pairs consisting of the central market - Williamston for corn and Fayetteville for soybeans - and each one of the other markets. They estimate the general TVECM (1) for logarithmic prices by minimizing the sum of squared errors (which is equivalent to maximizing the profile likelihood assuming independent Gaussian errors). Tables 2 and 3 compare threshold estimates obtained by maximizing the profile likelihood function with empirical Bayes estimates in terms of the thresholds as well as the resulting distribution of observations into regimes. It is evident that relative to the profile likelihood estimates, the empirical Bayes estimates for both lower and upper threshold tend to be drawn away from zero. This suggests that here the profile likelihood estimator might indeed yield biased results and markets might be even more tightly integrated than concluded by Goodwin and Piggott (2001).

	Profile likelihood estimator		**Empirical Bayes estimator**	
	lower threshold	upper threshold	lower threshold	upper threshold
CORN				
Candor-Williamston	-0.0269	0.0033	-0.0668	0.0287
Cofield-Williamston	-0.0036	0.0679	-0.0635	0.0377
Kinston-Williamston	-0.0125	0.0192	-0.0201	0.0192
SOYBEANS				
Raleigh-Fayettesville	-0.0016	0.0105	-0.0214	0.0142
Greenville-Fayettesville	-0.0075	0.0124	-0.0213	0.0117
Kinston-Fayettesville	-0.0059	0.0264	-0.0876	0.0167

Table 2: Estimated thresholds

	Profile likelihood estimator			**Empirical Bayes estimator**		
	1st reg	2nd reg	3rd reg	1st reg	2nd reg	3rd reg
CORN						
Candor-Williamston	290 (295)	678 (755)	797 (715)	13	1525	227
Cofield-Williamston	748 (69)	1011 (1661)	6 (35)	60	1463	242
Kinston-Williamston	244 (248)	1514 (1462)	7 (55)	7	1751	7
SOYBEANS						
Raleigh-Fayettesville	409 (162)	1309 (1556)	47 (47)	12	1714	39
Greenville-Fayettesville	495 (410)	965 (1285)	305 (70)	95	1335	335
Kinston-Fayettesville	545 (543)	1208 (506)	12 (716)	9	1600	156

Table 3: Estimated number of observations per regime

Even though we have found the profile likelihood estimates by maximizing the same function as Goodwin and Piggott (2001), the results presented in Tables 2 and 3 differ from those reported in their paper. This is due to differences in the grid used to search for the maximum. Goodwin and Piggott (2001) evaluate the estimating function at 100 equally spaced grid points for each threshold; while these grids cover the range between the smallest observed error correction term and zero for the lower threshold, they run from zero to the largest of them for the upper threshold. This procedure might not yield the correct maximum. Revisiting Figure 1 it is easy to understand that the true maximum might not be approximated well by specifying a grid which is too coarse. However, as mentioned before, the function to be maximized is a step function. Hence, evaluating it at each step (by specifying a complete grid of all observations for the error correction term), we will necessarily end up with the correct maximum. The impact of the proper grid is shown in Table 3. In the column "Profile likelihood estimator", estimates for Goodwin and Piggott's grid are reported in parentheses behind the estimates obtained for the complete grid. The case of Cofield and Williamston clearly illustrates that this effect of the type of grid used cannot be neglected. Caution has to be taken not only in choosing the estimator, but also in its practical implementation.

5 Conclusions

We discuss threshold estimators in TVCEMs. We point out weaknesses of commonly used estimators and emphasize the relevance of these problems for price transmission studies. We suggest a new estimator based on an empirical Bayes approach and demonstrate its superior performance in a simulation study. Revisiting an empirical application dealing with corn and soya prices, we find that using the new estimator changes results.

References

Andrews, D. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica*, 61(4):821–856.

Balcombe, K. (2006). Bayesian estimation of cointegrating thresholds in the term structure of interest rates. *Empirical Economics*, 31(2):277–289.

Balcombe, K., Bailey, A., and Brooks, J. (2007). Threshold effects in price transmission: the case of Brazilian wheat, maize, and soya prices. *American Journal of Agricultural Economics*, 89(2):308–323.

Balcombe, K. and Rapsomanikis, G. (2008). Bayesian estimation and selection of nonlinear vector error correction models: the case of the sugar-ethanol-oil nexus in Brazil. *American Journal of Agricultural Economics*, 90(3):658–668.

Balke, N. and Fomby, T. (1997). Threshold cointegration. *International Economic Review*, 38(3):627–645.

Chan, N. and Kutoyants, Y. (2010). On Parameter Estimation of Threshold Autoregressive Models. *Arxiv preprint arXiv:1003.3800.*

Forbes, C., Kalb, G., and Kofman, P. (1999). Bayesian arbitrage threshold analysis. *Journal of Business & Economic Statistics*, 17(3):364–372.

Goodwin, B. and Piggott, N. (2001). Spatial market integration in the presence of threshold effects. *American Journal of Agricultural Economics*, 83(2):302–317.

Hansen, B. and Seo, B. (2002). Testing for two-regime threshold cointegration in vector error-correction models. *Journal of Econometrics*, 110(2):293–318.

Ibragimov, I., Has' minskii, R., and Kotz, S. (1981). *Statistical estimation: asymptotic theory.* Springer-Verlag New York.

Lo, M. and Zivot, E. (2001). Threshold cointegration and nonlinear adjustment to the law of one price. *Macroeconomic Dynamics*, 5(04):533–576.

Lubrano, M. (2000). Bayesian analysis of nonlinear time series models with a threshold. In *Nonlinear Econometric Modeling in Time Series: Proceedings of the Eleventh International Symposium in Economic Theory.*

McCullagh, P. and Tibshirani, R. (1990). A simple method for the adjustment of profile likelihoods. *Journal of the Royal Statistical Society. Series B (Methodological)*, 52(2):325–344.

Tikhonov, A., Arsenin, V., and John, F. (1977). *Solutions of ill-posed problems.* Vh Winston Washington, DC.

Van der Vaart, A. (2000). *Asymptotic statistics.* Cambridge Univ Pr.

Yu, P. (2007). Likelihood Estimation and Inference in Threshold Regression.