

# Estimating unknown arbitrage costs: Evidence from a 3-regime threshold vector error correction model\*

Kristyna Ters

*University of Basel, Basel, Switzerland.*

Jörg Urban<sup>1</sup>

*University of Basel, Basel, Switzerland.*

---

## Abstract

We present a methodology for estimating a multivariate 3-regime threshold vector error correction model (TVECM) with an unknown cointegrating vector based on a new dynamic grid evaluation. This model is particularly useful for estimating deviations from parity conditions, such as unknown arbitrage costs in markets with a persistent non-zero basis between two similar financial market instruments traded in the spot and derivative markets. Our proposed 3-regime TVECM can estimate the area where arbitrageurs have no incentives to trade. It is only when the basis exceeds a critical threshold, that is when the potential gain from the basis trade exceeds the overall transaction costs, that we expect arbitrageurs to step in and carry out the respective trade. This leads to non-linear adjustment dynamics and regimes with different characteristics. Our methodology allows overall transaction costs to be quantified in markets where trading costs are opaque or unknown.

*Keywords:* Arbitrage, error correction, regime switch, threshold, and transaction cost.

*JEL:* G12, G14, and G15.

---

---

\* This is a preprint of an article published in the Journal of Financial Markets, Volume 47, January 2020, 100503. Copyright © 2019 Elsevier B.V. All rights reserved. doi: 10.1016/j.finmar.2019.07.002.

The authors thank Darrell Duffie, Erik Theissen, Myunghwan Seo and OTC markets seminar participants at the SNB Gerzensee Study Center, the Bank for International Settlements, and the EFMA 2016 conference for useful discussions and comments. Jörg Urban developed this paper while was working at the BIS. The research leading to these results received funding from the Swiss National Science Foundation under career grant agreement no. PMPDPI-158369.

E-mail addresses: kristyna.ters@unibas.ch (K. Ters), joerg.urban@unibas.ch (J. Urban)

<sup>1</sup> Corresponding author, University of Basel, Peter Merian-Weg 6, 4002 Basel, Switzerland.

## 1. Introduction

We present a methodology for estimating multivariate 3-regime vector error correction models with two thresholds and an unknown cointegrating vector. Our methodology is particularly useful for modelling arbitrage in markets with frictions and estimating unknown transaction costs.

The theoretical no-arbitrage condition between two similar financial market instruments traded in the spot and derivative markets requires pricing to be equal in both markets. Otherwise, any pricing discrepancy would provide investors with an arbitrage opportunity that will rapidly disappear, as arbitrageurs exploit any mispricing. This mispricing is measured by the so-called basis. The no-arbitrage condition holds only in perfect and frictionless markets. In practice, however, frictions and imperfections (e.g., due to limited and time-varying liquidity across market segments) often make such arbitrage trades difficult and costly. As a result, the basis is not zero and can become sizable and persistent.

A persistent non-zero basis is likely to reflect the unwillingness of arbitrageurs to try to exploit it, unless the pricing mismatch is greater than the cost of carrying out the arbitrage trade. Empirically, therefore, we would expect arbitrage forces to intensify as the size of the basis exceeds a level that reflects the costs that traders face in the market. This suggests a non-linear adjustment process towards the long-run equilibrium.

With modern information and telecommunication techniques, the access to markets has become less costly and one might expect that transaction costs or friction do not play a significant role. However, both the recent financial crisis and the European sovereign debt crisis have shown that the contrary is true. In crisis periods, where liquidity dries out or risk perception changes, transaction costs can become sizable. Supervisory scrutiny in such crisis periods requires effective tools for policy makers and central banks to understand how policies or market measures affect the basis and transaction costs (e.g., Gyntelberg et al., 2018).

Martens et al. (1998) employ the cost-of-carry model to describe the pricing relationship between futures and spot prices and then use a univariate threshold autoregressive model. In other words, they ex ante assume a specific structure of the error correction term. Stevens

(2015) finds that transaction costs increase the persistence of the basis in the market for crude oil. He explains the non-zero basis by the absence of arbitrage. Forbes et al. (1999) find significant transaction costs in index futures arbitrage for the S&P 500 Index that prevent arbitrage and a clear indication of arbitrage trading if the basis goes beyond a threshold. Forbes et al. (1999) and Stevens (2015) employ a univariate structural change test to the error correction term based on Tsay (1989). However, this approach is only valid when the error correction term is known a priori.

Balke and Fomby (1997) analyze the relation between error correction models and cointegration and introduce threshold cointegration as a feasible means to combine regime switches (non-linearities) and cointegration. Theissen (2012) employs a 1-threshold error correction model for DAX and FDAX price quotes. However, he assumes that the theoretical cointegration relationship spot minus futures prices holds (based on the cost-of-carry-model). Theissen's innovation is to allow for time-varying transaction costs by making the threshold dependent on bid-ask spreads in the two markets. While this seems an economically sound assumption, it does not hold for all markets and periods. For example, Gyntelberg et al. (2017) report that the bid-ask spread in the CDS and bond market between 2010 and 2011 remained relatively stable, while the basis drastically widened. The model proposed by Theissen (2012) is partly an extension and partly nested in the model provided by Hansen and Seo (2002), who introduced a methodology for estimating a 2-regime TVECM with an unknown cointegrating relationship. Our extension to a 3-regime TVECM with two thresholds is important from an economic viewpoint, as it allows testing for the existence of transaction costs in positive and negative basis trading strategies. We alter the model setup proposed by Hansen and Seo (2002) by introducing a constant in the cointegration relationship, which is especially suitable for financial markets, where often a persistent deviation of the long-term equilibrium from the theoretical parity relationship exists.

Assuming the existence of two time-independent thresholds, one for a positive and the other for a negative basis trade, is an approximation, because traders are confronted with different transaction costs and are also heterogeneous in terms of information. An alternative model to use in situations with heterogeneous traders is the smooth error correction model. Anderson

(1997) compares and contrasts adjustments with homogeneous and heterogeneous transaction costs. Granger and Teräsvirta (1993) and Teräsvirta (1994) develop a smooth transition error correction model. This model requires the specification of a transition function. Panos et al. (1997) employ the smooth transition autoregressive model of Granger and Teräsvirta (1993) in the real exchange market by assuming an exponential transition function. Dumas (1992) also derives a smooth adjustment process. Gaul and Theissen (2015) also employ smooth models arguing that the transition function must be exogenously specified and that there is no theory to guide such modeling. They propose a partially linear error correction model, but, unlike us, they approximate the cointegration relationship to the theoretical no-arbitrage relationship.

Despite the rich literature, we think that our “abrupt” threshold model is suitable to understand how policy changes or shocks affect the market, without assuming any exogenous transition modeling. Even though the “true” threshold is blurred, our 2-threshold VECM is sufficient for event studies, because only a few parameters need to be estimated and no specification of a transition function is required. Policy makers have therefore a tool at hand to understand how shocks affect market friction or transaction costs. An additional advantage of TVECM models compared to smooth transition models is that the estimation error of the long-run terms does not affect the estimation of the short-run terms (Seo, 2011).

We contribute to the literature by developing an estimation procedure for multivariate TVECMs with three regimes (two thresholds) and an unknown cointegrating vector. The estimation of an unknown cointegrating vector is particularly important for distorted parity relationships such as in financial markets and economic applications that exhibit a significant and persistent non-zero deviation from the theoretical parity relationship. Our model allows for non-linear adjustment of prices, in derivative and spot markets, towards the long-run equilibrium. Consequently, we can estimate the region where arbitrageurs enter the market as the trading opportunity is “sufficiently profitable”. The overall transaction costs can be derived from the no-arbitrage regime. We do not aim to model different types of transaction costs. In line with this, we do not assume that the positive and negative thresholds are symmetric around zero, because different transaction costs in the negative and positive arbitrage regime may exist (e.g., Berkman et al., 2005; Gaul and Theissen, 2015).

Regulators can employ our non-linear model to estimate unknown transaction costs, to study which market moves first or to monitor how fast markets return to the long-run equilibrium after an exogenous shock. Such an analysis is especially interesting, after regulatory changes were introduced, in order to study the impact of the implemented changes on market participants' behavior. In markets, where high frequency data are available (e.g., index-trading), a pre- and post-event analysis can be performed to evaluate the effectiveness of policy actions.

The rest of the paper is structured as follows: In Section 2, we discuss the setup and estimation of the TVECM, as well as present a simulation exercise. In Section 3, we provide an empirical application for the illustration of our methodology. Concluding remarks are in Section 4.

## 2. 3-regime threshold vector error correction model

We start section with an introduction of the multi-regime TVEVM and describe how the model parameters can be estimated. Then we describe how to test for the significance of the estimated threshold(s). A detailed simulation is presented to validate the proposed approach.

### 2.1. The model setup

The TVECM approach extends the VECM by allowing the behavior of price quotes for spots  $S_t$  and derivatives  $D_t$  for a specific reference entity or underlying to depend on the state of the system. In case the vector of price quotes  $y_t = (S_t \ D_t)^\top$  is non-stationary, cointegrated, and  $\Delta y_t$  is  $I(0)$ , we can formulate a  $l$ -regime TVECM as follows:

$$\mathcal{H}_l : \quad \Delta y_t = \sum_{j=1}^l [\lambda_l^j \text{ec}_{t-1}(\beta_0, \beta_1) + \Gamma_l^j(L) \Delta y_t] d_t(\beta_0, \beta_1, \theta^{j-1}, \theta^j) + \varepsilon_t, \quad (1)$$

with  $\Gamma_l^j(L) = \sum_{k=1}^m \alpha^{j,k} L^k$ ,  $L$  as lag operator,  $m$  as the number of VAR lags,<sup>1</sup> and a stationary error correction term  $\text{ec}_{t-1}(\beta_0, \beta_1) = (S - \beta_1 D - \beta_0)_{t-1}$ . The index  $j$  denotes the different regimes. For later convenience, we introduce labels  $l$  at the speed of adjustment  $\lambda_l^j$  and  $\Gamma_l^j(L)$  indicating to which model class they belong.  $\varepsilon_t = (\varepsilon_t^S \ \varepsilon_t^D)^\top$  is a vector of i.i.d. shocks. All

---

<sup>1</sup> We use the Schwarz (Bayesian) information criterion (SIC) as Lütkepohl (2006) states that, in large samples for multivariate models when  $T \rightarrow \infty$ , only the SIC criterion is consistent.

thresholds  $\theta^j$  are ordered and  $d_t(\beta_0, \beta_1, \theta^{j-1}, \theta^j) = 1$  if the error correction term  $ec_{t-1}(\beta_0, \beta_1)$  is in the interval  $[\theta^{j-1}, \theta^j)$  and otherwise 0. Further, by definition, the threshold  $\theta^0$  is  $-\infty$  and  $\theta^l$  is  $\infty$ .

The  $2 \times 1$  vector  $\lambda_t^j$  and the lagged VAR terms are regime-dependent conditioned on the state of the error correction term. The speed of adjustment parameters characterizes to what extent the price changes in  $\Delta y_t$  react to deviations from the long-term equilibrium.

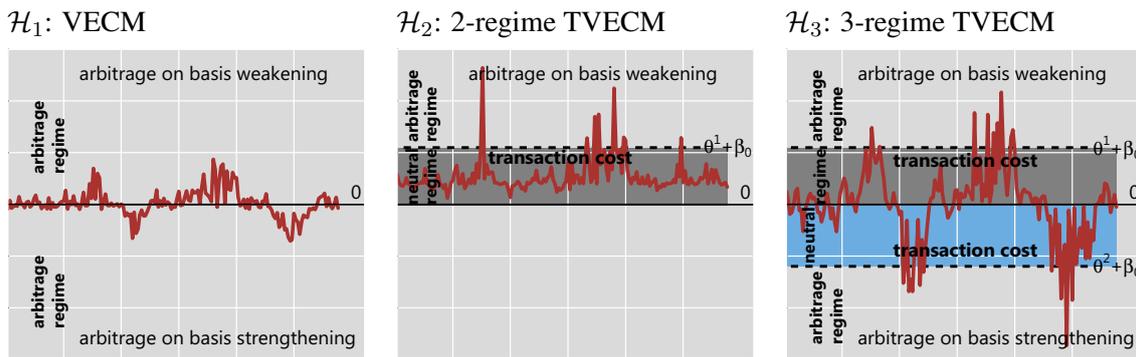
The error correction term represents the long-term equilibrium of the two time series, which has to be an AR(1) process by construction (Johansen, 1988). Therefore, we cannot fix  $\beta_0$  and  $\beta_1$  to 0 and 1, respectively, as market friction might generate persistent and large deviations from these theoretical values. The VAR term represents the short-run dynamics coming from market imperfections (Baillie et al., 2002).

As previously discussed, contrary to the 1-threshold VECM proposed by Hansen and Seo (2002), we introduce an intercept  $\beta_0$  in the error correction term that denotes the deviation from the theoretical parity relationship (long-term equilibrium). This is motivated by our no-arbitrage discussion in Section 1 as the local constant  $\beta_0$  represents the persistent non-zero basis. The Hansen and Seo (2002) model is more general than our model at the cost of another parameter needed to get fit. The alteration leading to our model is motivated by economic reasoning.

The transaction costs for a basis trade prevent a complete adjustment towards a zero basis. As such, in markets with frictions there may be a neutral band between the derivative and the spot market, in which the error correction term in equation (1) may fluctuate without incentives for market participants to switch funds between the spot and derivative market. Outside of that neutral band, however, there might be strong incentives for arbitrage trades, which results in an adjustment towards the long-term equilibrium. We expect larger magnitudes of the speed of adjustment parameters as soon as the basis exceeds a threshold. In a market with a positive basis ( $S_t > D_t$ ), arbitrageurs bet on a declining basis and will go short in the spot market and go long in the derivative market. In case of a negative basis ( $S_t < D_t$ ), arbitrageurs bet on an increasing basis while carrying out the reverse trade.

According to arbitrage theory, we would in general expect to find a 3-regime TVECM with

two thresholds when the basis fluctuates between positive and negative figures with sizable and persistent deviations from zero. The lower regime is defined as  $ec_{t-1}(\beta_0, \beta_1) < \theta^1$ , the middle regime as  $\theta^1 \leq ec_{t-1}(\beta_0, \beta_1) < \theta^2$ , and the upper regime as  $\theta^2 \leq ec_{t-1}(\beta_0, \beta_1)$ . The middle regime is the neutral band, while the outer regimes are the arbitrage regimes. There may also be certain markets or time periods with only a persistent positive or negative basis. In that case, we expect to find at most two regimes ( $l = 2$ ) with only one threshold  $\theta^1$  (for a simple representation see Figure 1).



**Figure 1:** Model classes  $\mathcal{H}_l$

The VECM  $\mathcal{H}_1$  represents markets or periods, where the basis only marginally deviates from zero. Model classes  $\mathcal{H}_2$  and  $\mathcal{H}_3$  are multi-regime models, where outside the neutral regime arbitrageurs engage in basis trades. A model belonging to class  $\mathcal{H}_2$  can also have a negative basis and consequently a negative threshold. The graphs represent the time evolution of the basis.

Therefore, we discuss three classes of nested models with unknown cointegrating vector: model class  $\mathcal{H}_1$  is a 1-regime VECM, with no statistical significant threshold (efficient market). Model class  $\mathcal{H}_2$  is a 2-regime TVECM and model class  $\mathcal{H}_3$  is a 3-regime TVECM.

The threshold  $\theta^j$  is computed relative to the estimated basis  $S_t - \beta_1 D_t - \beta_0$ . Transaction costs are defined relative to the observed/real basis  $S_t - D_t$ . Hence, in the case of  $\theta^j + \beta_0 < 0$ , we have transaction costs  $|\theta^j + \beta_0|$  for an arbitrage trade on basis strengthening and in the case of  $\theta^j + \beta_0 > 0$ , we have transaction costs  $(\theta^j + \beta_0)$  for an arbitrage trade on basis weakening. We cannot assume symmetry of the two thresholds  $\theta^1$  and  $\theta^2$ , because in most markets transaction costs will be different for a basis weakening and basis strengthening trade, among others due to different liquidity.

Following Hansen and Seo (2002), we estimate the model classes by imposing the following

additional constraint for each regime:

$$\pi_0 \leq P(\theta^{j-1} \leq \text{ec}_{t-1}(\beta_0, \beta_1) < \theta^j), \quad (2)$$

where  $\pi_0 > 0$  is a trimming parameter and  $P$  is the share of observations in each regime. This constraint allows us to identify a threshold effect only if the share of observations in each regime is large enough (i.e., is greater than  $\pi_0$ ). If this condition is not met, the model  $\mathcal{H}_l$  reduces to  $\mathcal{H}_{l-1}$  for  $l > 2$ . Andrews (1993) argues that setting  $\pi_0$  between 0.05 and 0.15 is typically a good choice. We choose as a baseline setup  $\pi_0 = 0.1$ , but also perform robustness tests for  $\pi_0 = 0.05$  and  $\pi_0 = 0.15$ .

## 2.2. Estimating the model

The most important statistical issue for threshold models is estimating the unknown threshold(s) and testing for their significance. We implement a maximum likelihood estimation for  $l$  regimes, assuming i.i.d. Gaussian error terms (Hansen and Seo, 2002):

$$\mathcal{L}_n(\lambda_l^1, \dots, \lambda_l^l, \Gamma_l^1, \dots, \Gamma_l^l, \beta_0, \beta_1, \theta^1, \dots, \theta^{l-1}, \Sigma) = -\frac{n}{2} \ln |\Sigma| - \sum_{t=1}^n \frac{1}{2} \varepsilon_t^\top \Sigma^{-1} \varepsilon_t, \quad (3)$$

with  $\Sigma = E(\varepsilon_t \varepsilon_t^\top)$  and sample size  $n$ .  $\varepsilon_t$  and  $\Sigma$  are functions of  $\lambda_l^i, \Gamma_l^i, \beta_0, \beta_1$ , and  $\theta^j$ .

Hansen and Seo (2002) suggest that it is computationally convenient to hold  $\beta_0$  and  $\beta_1$  as well as  $\theta^j$  fixed and compute the concentrated maximum likelihood function. Due to the linearity of the model, this is an OLS regression, which results in:

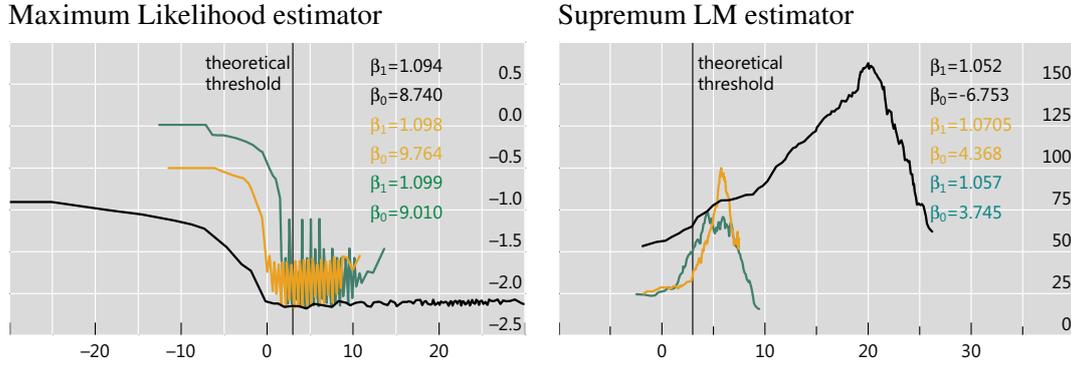
$$\begin{aligned} \mathcal{L}_n(\beta_0, \beta_1, \theta^j) &= \mathcal{L}_n(\hat{\lambda}_l^i(\beta_0, \beta_1, \theta^j), \hat{\Gamma}_l^i(\beta_0, \beta_1, \theta^j), \hat{\Sigma}(\beta_0, \beta_1, \theta^j), \beta_0, \beta_1, \theta^j) \\ &= -\frac{n}{2} \ln |\hat{\Sigma}(\beta_0, \beta_1, \theta^j)| - \frac{n \cdot p}{2}, \end{aligned} \quad (4)$$

with  $p = 2$  in the bivariate case.

All variables with a hat are OLS estimators. The remaining task of finding the maximum likelihood estimation of  $\beta_0, \beta_1$ , and the thresholds is to minimize  $\ln |\hat{\Sigma}(\beta_0, \beta_1, \theta^j)|$ , subject to the constraint in equation (2).

The function in equation (4) is not smooth (see the left-hand panel of Figure 2), hence conventional hill-climbing algorithms cannot be used to find the minimum. Hansen and Seo

(2002) suggest a joint grid search. We present evidence in Subsection 2.4 for the outstanding performance of the grid search even for three regimes.



**Figure 2:** Maximum likelihood estimator versus supremum LM estimator

The graphs show the maximum likelihood estimation and the supremum LM estimation for three sample simulations.  $\beta_0$ ,  $\beta_1$ , and the threshold have been set to 10, 1.1, and 3, in the data-generating process. Each graph contains the estimators for each simulation. The left-hand side shows the  $\beta_i$  values estimated under the alternative hypothesis (model class  $\mathcal{H}_2$ ). The minimum is very close to the theoretically expected threshold. The right-hand side shows  $\beta_i$  obtained from the VECM estimation (null hypothesis), which are far off from the values of the data-generating process, and so is the maximum of the supremum LM estimator. The horizontal axis represents part of the grid used for the threshold search.

Our model requires a search over a three-dimensional grid ( $\beta_0, \beta_1, \theta^1$ ) in the 2-regime case and a four-dimensional grid search ( $\beta_0, \beta_1, \theta^1, \theta^2$ ) in the 3-regime case. In order to keep the computation feasible, we suggest a sequential threshold search as it is proven to be consistent by Bai (1997) and Bai and Perron (1998). The superconsistency is also thoroughly discussed by (Seo, 2011). Using 100 grid points for each parameter, a sequential search would require  $2 \times 100^3$  grid point evaluations ( $100^3$  for  $\beta_0, \beta_1$ , and  $\theta^1$  during the first grid search and another  $100^3$  grid point evaluations for  $\beta_0, \beta_1$ , and  $\theta^2$  during the second search). We discuss and justify in a comprehensive simulation exercise in Subsection 2.4 that it is efficient to fix  $\beta_1$  in the second threshold search to the value found in the first threshold search and hence reduce the second search to a two-dimensional space. This further reduces the computational burden. In the same subsection, we also show that we cannot fix  $\beta_0$  in the threshold search for  $\theta^2$ , as the  $\beta_0$  estimate suffers from a large uncertainty.

The remaining issue to be addressed is the setup of the sufficient search area for each parameter. The search region  $[\theta_L, \theta_U]$  for the thresholds is straightforward as it must be identical to the interval  $[\min(\text{ec}_t(\beta_0, \beta_1)), \max(\text{ec}_t(\beta_0, \beta_1))]$ . The region for the  $\beta_0$  and  $\beta_1$  parameters

can be calibrated based on the estimates of the linear VECM and the theoretical values  $\beta_1 = 1$  and  $\beta_0 = 0$ . It is important to keep the search area for  $\beta_0$  and  $\beta_1$  large enough to include the minimum, but not too large to reduce the precision of the grid search. The grid search for  $\theta^2$  will be reduced by the constraint  $\theta^2 + \beta_0 > 0$  if the first search resulted in  $\theta^1 + \beta_0 < 0$  and vice versa if  $\theta^1 + \beta_0 > 0$ . This is purely based on our previous arbitrage discussion, where we expect to find at most two transaction costs.

We show in our simulation exercise that the precision of the estimated parameters will depend on the distance between two neighboring grid points, despite the complicated non-differentiable structure of the maximum likelihood function.

### 2.3. Testing for a threshold

The next step is to determine whether the estimated thresholds  $\hat{\theta}^j$  are statistically significant. As the model class  $\mathcal{H}_1$  is nested in  $\mathcal{H}_2$ , and  $\mathcal{H}_2$  is nested in class  $\mathcal{H}_3$ , we start by discussing the 1-threshold model  $\mathcal{H}_2$ . In that case, under the null hypothesis, there is no threshold, so the model reduces to a conventional linear VECM  $\mathcal{H}_1$ , where  $\lambda_2^1 = \lambda_2^2 = \lambda_1^1$  and  $\Gamma_2^1(L) = \Gamma_2^2(L) = \Gamma_1^1(L)$ . The 1-threshold VECM  $\mathcal{H}_2$  is detected under the alternative hypothesis with  $\lambda_2^1 \neq \lambda_2^2$  and  $\Gamma_2^1(L) \neq \Gamma_2^2(L)$ . As the models are linear, a regular LM test with an asymptotic  $\chi^2(N)$ -distribution can be calculated from a linear regression on the model  $\mathcal{H}_2$ . However, the LM test statistic can only be applied if  $\beta_i$  and the threshold variable  $\theta^1$  are known a priori (Hansen and Seo, 2002). While the point estimates of  $\beta_0$  and  $\beta_1$  under the null hypothesis are  $\tilde{\beta}_0$  and  $\tilde{\beta}_1$  from the linear model, there is no estimate of  $\theta^1$  under the null hypothesis. This implies that there is no distribution theory for the parameter estimates and no conventionally defined LM-like statistic.

Hansen and Seo (2002) derived the LM-like statistic for the one threshold case:

$$\text{LM}(\beta_0, \beta_1, \theta^1) = \text{vec}(\hat{A}_2^1 - \hat{A}_2^2)^\top (\hat{V}_2^1 + \hat{V}_2^2)^{-1} \text{vec}(\hat{A}_2^1 - \hat{A}_2^2), \quad (5)$$

with the OLS estimator  $\hat{A}_2^i$  ( $i \in \{1, 2\}$ ):

$$\hat{A}_2^i(\beta_0, \beta_1, \theta^1) = \left( \sum_{t=1}^n Y_{t-1}(\beta_0, \beta_1) Y_{t-1}(\beta_0, \beta_1)^\top d_t(\beta_0, \beta_1, \theta^{i-1}, \theta^i) \right)^{-1} \times \left( \sum_{t=1}^n Y_{t-1}(\beta_0, \beta_1) \Delta y_t^\top d_t(\beta_0, \beta_1, \theta^{i-1}, \theta^i) \right), \quad (6)$$

where  $n$  is the length of the time series.  $\hat{A}_2^i$  are  $(1 + pm) \times p$  matrices, with  $m$  denoting the number of VAR lags and  $p = 2$  for the bi-variate case.  $\hat{V}_2^i$  is defined via the moment functionals  $M_2^i$  and  $\Omega_2^i$ :

$$\hat{V}_2^i(\beta_0, \beta_1, \theta^1) = M_2^i(\beta_0, \beta_1, \theta^1)^{-1} \Omega_2^i(\beta_0, \beta_1, \theta^1) M_2^i(\beta_0, \beta_1, \theta^1)^{-1}. \quad (7)$$

The moment functionals are given as:

$$M_2^i(\beta_0, \beta_1, \theta^1) = \mathbb{1}_p \otimes Y^i(\beta_0, \beta_1, \theta^1)^\top Y^i(\beta_0, \beta_1, \theta^1), \quad (8)$$

$$\Omega_2^i(\beta_0, \beta_1, \theta^1) = \xi^i(\beta_0, \beta_1, \theta^1)^\top \xi^i(\beta_0, \beta_1, \theta^1), \quad (9)$$

which are both  $p \cdot (1 + p \cdot m) \times p \cdot (1 + p \cdot m)$  matrices.  $Y^i$  is a short form of the matrices of the stacked rows  $Y_{t-1}(\beta_0, \beta_1) \circ d_t(\beta_0, \beta_1, \theta^{i-1}, \theta^i)$ , with

$$Y_{t-1}(\beta_0, \beta_1) = \begin{pmatrix} \text{ec}_{t-1}(\beta_0, \beta_1) \\ \Delta y_{t-1} \\ \vdots \\ \Delta y_{t-m} \end{pmatrix}. \quad (10)$$

Hence,  $Y^i$  is a  $t \times (1 + p \cdot m)$  matrix of the following form:

$$Y^i(\beta_0, \beta_1, \theta^1) = \begin{pmatrix} \vdots \\ \text{ec}_{t-1}(\beta_0, \beta_1) & \Delta y_{t-1}^\top & \dots & \Delta y_{t-m}^\top \\ \vdots \end{pmatrix} \circ d_t(\beta_0, \beta_1, \theta^{i-1}, \theta^i), \quad (11)$$

where  $\circ$  denotes element-wise multiplication.  $Y^i$  contains only non-zero entries if  $\theta^{i-1} \leq \text{ec}_{t-1} < \theta^i$ , whereby we have  $\theta^0 = -\infty$  and  $\theta^2 = \infty$  for the 2-regime TVECM  $\mathcal{H}_2$ .  $\xi^i$  is defined as  $\tilde{\varepsilon}_t \otimes Y_{t-1}(\beta_0, \beta_1) \circ d_t(\beta_0, \beta_1, \theta^{i-1}, \theta^i)$ , with  $\tilde{\varepsilon}_t$  being the OLS estimate of the

residual vector from the linear model and  $\otimes$  is the Kronecker product.

As there is no estimate of  $\theta^1$  for model class  $\mathcal{H}_1$  (null hypothesis), we employ the union-intersection principle proposed by Davies (1987):

$$\text{SupLM} = \sup_{\theta_L^1 \leq \theta^1 \leq \theta_U^1} \text{LM}(\tilde{\beta}_0, \tilde{\beta}_1, \theta^1). \quad (12)$$

The search region  $[\theta_L^1, \theta_U^1]$  is defined such that the constraint in equation (2) holds. This grid evaluation over  $[\theta_L^1, \theta_U^1]$  is necessary to implement the maximization defined in equation (12) as the function  $\text{LM}(\tilde{\beta}_0, \tilde{\beta}_1, \theta^1)$  is non-differentiable in  $\theta^1$  and hence conventional hill-climbing algorithms cannot be used to find the extrema.

The value of  $\theta^1$  that maximizes equation (12) is different from the MLE  $\hat{\theta}^1$  in Subsection 2.2, as equation (12) is a LM test based on parameter estimates obtained under the null hypothesis (model class  $\mathcal{H}_1$ ). Also, the test statistic is calculated with HAC-consistent covariance matrix estimates, which leads to differing estimates compared to the estimate in Subsection 2.2 (see the discussion in Hansen and Seo (2002) and the illustration in Figure 2, where the difference can be seen by comparing the left-hand with the right-hand panel). The supremum LM test is performed to compute the critical value, which is later used to decide if the computed threshold is statistically significant. The supremum LM test, unlike the maximum likelihood method, is not used to estimate the model parameters.

Just like in the one threshold case ( $\mathcal{H}_2$ ), the next step is to determine whether the estimated 2-threshold VECM ( $\mathcal{H}_3$ ) is statistically significant. Under the null hypothesis, there is one threshold  $\theta^1$ , so the model reduces to a TVECM with two regimes ( $\mathcal{H}_2$ ) described in Subsection 2.2 with  $\lambda_3^1 \neq \lambda_3^2 = \lambda_3^3$ . The 2-threshold VECM is detected under the alternative hypothesis with  $\lambda_3^2 \neq \lambda_3^3$ , after we already know that  $\lambda_3^1 \neq \lambda_3^2$ . We do not explicitly test if the lower and the upper regime (in a 3-regime model) differ in their adjustment speeds. However, the cointegration requirement ensures that the speed of adjustments in the upper and lower regime must at least have a different sign by construction (in order for the time series to revert to the long-term equilibrium).

Following the same steps and arguments discussed above, we can define the following LM-

like statistics:

$$\text{LM}(\beta_0, \beta_1, \theta^1, \theta^2) = \text{vec}(\hat{A}_3^2 - \hat{A}_3^3)^\top (\hat{V}_3^2 + \hat{V}_3^3)^{-1} \text{vec}(\hat{A}_3^2 - \hat{A}_3^3). \quad (13)$$

Following the proposition by Davies (1987), the union-intersection principle now reads:

$$\text{SupLM} = \sup_{\theta_L^2 \leq \theta^2 \leq \theta_U^2} \text{LM}(\tilde{\beta}_0, \tilde{\beta}_1, \tilde{\theta}^1, \theta^2), \quad (14)$$

with the point estimates  $\tilde{\beta}_i$  and  $\tilde{\theta}^1$  obtained under the null (1-threshold VECM). We perform again a grid search over the search region for  $\theta^2$  subject to the constraint in equation (2). Furthermore, based on our arbitrage discussion in Section 1 and the assumption that we have two transaction costs, one for a positive and one for a negative basis trade, we further impose for the grid search of  $\theta^2$  and  $\beta_0$  the constraints that  $\theta^2 + \beta_0 > 0$  if the first search resulted in transaction costs  $\theta^1 + \beta_0 < 0$  and  $\theta^2 + \beta_0 < 0$  if the first search resulted in transaction costs  $\theta^1 + \beta_0 > 0$ .

As there is no formal distribution theory in the case under discussion, we follow Hansen and Seo (2002) (see also Hansen (2000) and Hansen (1996)) and perform two different bootstrap methodologies to estimate the asymptotic distribution of equation (1).

We implement a non-parametric bootstrap of the residuals, called the "fixed regressor bootstrap", which uses Monte Carlo resampling of the residuals from the estimated linear VECM or 1-threshold VECM, in the case of the threshold search for  $\theta^1$  or  $\theta^2$ , respectively.

We take estimates under the null hypothesis of  $\beta_i$ , denoted as  $\tilde{\beta}_i$ , and define  $\tilde{e}_{t-1} \equiv e_{t-1}(\tilde{\beta}_i)$  and  $\tilde{Y}_{t-1} \equiv Y_{t-1}(\tilde{\beta}_i)$ , whereby  $Y_{t-1}$  is defined in equation (10). Further,  $\tilde{\varepsilon}_t$  are the residuals of the null. The name "fixed regressor bootstrap" conveys the message that  $\tilde{\beta}_i$ ,  $\tilde{\varepsilon}_t$ ,  $\tilde{e}_{t-1}$ , and  $\tilde{Y}_{t-1}$  are kept fixed to their sample values.

Next, we compute a large number of times (e.g., 1,000)  $y_{bt} = \tilde{\varepsilon}_t e_{bt}$ , whereby  $e_{bt}$  is i.i.d.  $N(0, 1)$  and in each draw  $e_{bt}$  is independently chosen. For each draw (identified by the index  $b$ ), we perform an LM test.  $\tilde{\varepsilon}_{bt}$  is computed by regressing  $y_{bt}$  on  $\tilde{Y}_{t-1}$ .  $\hat{A}_2^j(\tilde{\beta}_i, \theta^1)_b$  (see equation (6) for the 1-threshold case)<sup>2</sup> and  $\hat{\varepsilon}_{bt}(\tilde{\beta}_i, \theta^1)$  are computed by regressing  $y_{bt}$  on  $\tilde{Y}_{t-1} d_{1t}(\tilde{\beta}_i, \theta^1)$  and  $\tilde{Y}_{t-1} d_{2t}(\tilde{\beta}_i, \theta^1)$ , whereby  $\tilde{\beta}_i$  is kept fixed. Further, for each draw  $b$  we compute  $\hat{V}_2^j(\tilde{\beta}_i, \theta^1)_b$  (see

<sup>2</sup> For the 2-threshold VECM, we need to compute  $\hat{A}_3^j(\tilde{\beta}_i, \tilde{\theta}^1, \theta^2)_b$ , whereby the values of the null  $\tilde{\beta}_i$  and  $\tilde{\theta}^1$  are kept fixed.

equation (7) for the 1-threshold VECM)<sup>3</sup>, whereby  $\tilde{\beta}_i$  is kept fixed again. Similar to equations (5) and (12),<sup>4</sup> we compute for each draw  $b$ :

$$\text{SupLM}^* = \sup_{\theta_L^1 \leq \theta^1 < \theta_U^1} \text{vec}(\hat{A}_{2b}^1 - \hat{A}_{2b}^2)^\top (\hat{V}_{2b}^1 + \hat{V}_{2b}^2)^{-1} \text{vec}(\hat{A}_{2b}^1 - \hat{A}_{2b}^2), \quad (15)$$

where  $\hat{A}_{2b}^i$  and  $\hat{V}_{2b}^i$  are functions of the fixed  $\tilde{\beta}_i$  and  $\theta^1$ .

Hansen (1996) shows that  $\text{SupLM}^*$  is a valid first-order approximation to the asymptotic null distribution of  $\text{SupLM}$ . Despite having the computational cost of a bootstrap, it only approximates the asymptotic distribution.

The  $p$ -value is calculated as the percentage of  $\text{SupLM}^*$  values that exceed the actual  $\text{SupLM}$  value of the original time series.

The second method, named the "residual bootstrap", is fully parametric with respect to the data-generating process, which means for the one threshold case we use the complete specification for the null as given by  $\mathcal{H}_1$  (VECM vs. 1-threshold VECM as alternative) and for the 2-threshold VECM we use the complete specification for the null as given by  $\mathcal{H}_2$  (1-threshold VECM vs. 2-threshold VECM as alternative). We further assume  $\varepsilon_t$  to be i.i.d. from an unknown distribution and fixed initial conditions. To be specific, random draws are made from  $\tilde{\varepsilon}_t$ , which are the residuals under the null. Using the given initial conditions from the data, and the parameters estimated under the null ( $\tilde{\lambda}^i$ ,  $\tilde{\beta}_0$ ,  $\tilde{\beta}_1$ , and  $\tilde{\Gamma}^i$ ), we recursively generate the bivariate vector series  $x_{bt}$  for the given model ( $\mathcal{H}_1$  or  $\mathcal{H}_2$ ). For each draw the  $\text{SupLM}^*$  value is computed and then again the percentage of the  $\text{SupLM}^*$  values that exceed the actual  $\text{SupLM}$  value (computed from the original time series) gives the  $p$ -value.

Hansen and Seo (2002) conjecture that this bootstrap method gives better finite sample performance at the computational cost of being fully parametric with respect to the data-generating process.

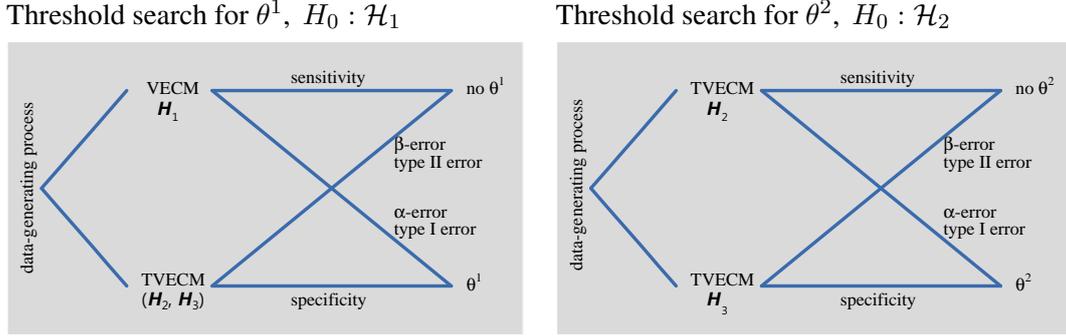
#### 2.4. Simulation

We perform numerous simulations to estimate the model parameters and to test the power of the proposed LM tests using different data-generating processes for the model specifications

<sup>3</sup> For the 2-threshold VECM, we need to compute  $\hat{V}_3^j(\tilde{\beta}_i, \tilde{\theta}^1, \theta^2)_b$ , whereby the values of the null  $\tilde{\beta}_i$  and  $\tilde{\theta}^1$  are kept fixed.

<sup>4</sup> In the 2-threshold VECM, we need equation (14).

from equation (1). Figure 3 presents the sequential search for the two thresholds.



**Figure 3:** Null versus alternative hypothesis in the sequential search for  $\theta^1$  and  $\theta^2$

We test for the existence of thresholds in the data-generating process in equation (1) using the fixed regressor and the residual bootstrap, as well as a sequential threshold search. In both bootstrap tests, we assume that there is no threshold in the search for  $\theta^1$ , i.e., the null hypothesis is a VECM ( $\mathcal{H}_1$ ). For the threshold search for  $\theta^2$ , the null hypothesis is a 2-regime TVECM ( $\mathcal{H}_2$ ).

In this subsection, we use the following notation:  $\beta_i$  and  $\theta^i$  are the parameters fixed in the data-generating process, while  $\hat{\beta}_i$  and  $\hat{\theta}^i$  are the estimates. For each set of fixed parameters, we perform 1,000 simulations. The means of the 1,000 estimators  $\hat{\beta}_i$  and  $\hat{\theta}^i$  are compared to  $\beta_i$  and  $\theta^i$ . The standard deviation is used as a measure of the precision of the estimators.

We introduce additional restrictions regarding the generated time series: first, we expect that the data-generating process produces time series that are I(1) and a basis  $ec_{t-1}(\beta_0, \beta_1)$  that is I(0) at the 90% confidence level using the augmented Dickey-Fuller test. Second, we do not allow the time series to become large in absolute terms.<sup>5</sup> We perform several robustness checks, such as different time series lengths and different equidistant grid sizes of 10, 50, and 100, in addition to a dynamic grid setting. The dynamic grid setting is determined by a minimum distance between two individual grid points. The dynamic setting is of course potentially very expensive, as it may lead to a large number of grid points. In our simulation exercise, we use 0.01 as the distance between two grid points for the  $\beta_1$  grid and 0.5 for the  $\beta_0$  and  $\theta^1$  grid. In the restricted case, where we have the spot and derivative times series confined in absolute terms to 500, the basis can vary between -1,000 and 1,000 in extreme situations.<sup>6</sup> In this case, we can

<sup>5</sup> The maximum value is set to 500 in our simulations. Behind this assumption lays the reason that economic time series do not grow arbitrarily large. For cross-checking purposes, we have also tested the unrestricted case and found similar results, but the standard deviation is larger, as expected.

<sup>6</sup> This is strictly correct only if  $\beta_1$  is around 1.

expect on the order of  $2,000/0.5$  grid points for the  $\beta_0$  and for the  $\theta^1$  grid. The number of grid points for the  $\beta_1$  grid does usually not reach such extreme values. The grid is built around the theoretical value  $\beta_1 = 1$  (observed basis) and the value found in the initial VECM estimation. Assuming that the VECM estimation is 4, then the grid is gauged from around 0.8 and 4.2, to include the VECM estimate and the theoretical value of 1. This leads to  $(4.2-0.8)/0.01=340$  grid points. The advantage of the dynamic grid setting is that the precision of the estimation process is predefined and not the number of grid points.

All parameters in the data-generating process are created with a random number generator, except for  $\beta_0$ ,  $\beta_1$ , and the thresholds  $\theta^j$  (in case of a threshold model). We fix the VAR lag in the data-generating process to one. Relaxing this restriction is straight-forward and yields robust results, however with lower precision due to the larger number of parameters to be estimated.

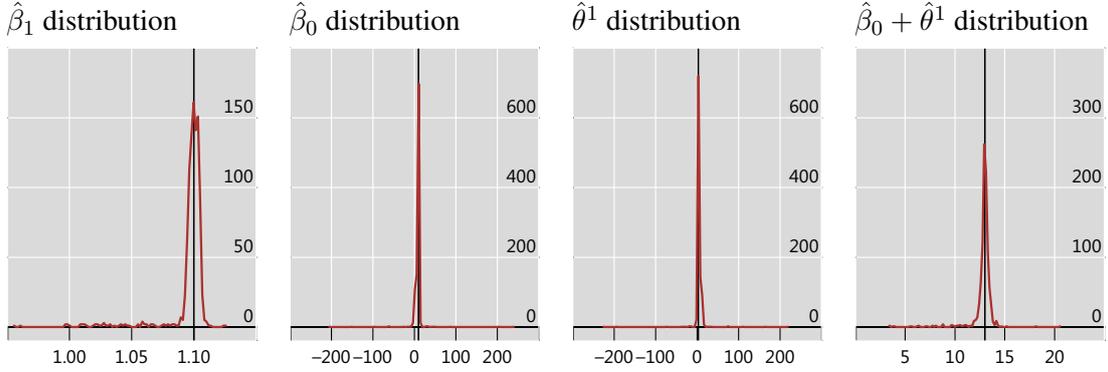
#### 2.4.1. Model class $\mathcal{H}_2$ : 2-regime TVECM

We begin our simulation with the test for a 2-regime TVECM  $\mathcal{H}_2$ . We choose as an exemplification  $\beta_1 = 1.10$ ,  $\beta_0 = 10$ , and  $\theta^1 = 3$  for our data-generating process. In this subsection we estimated the parameters and test the significance of the single threshold using the bootstrap methodologies.

The results of the simulation for the first threshold search are presented in Table 1, where two observations can be made: the mean of the 1,000 simulations for  $\hat{\beta}_1$  is practically independent of the grid setup and the outcomes of the bootstrap methodologies are also independent. The fixed regressor and residual bootstrap have very small  $\beta$ -errors, below 2%. As expected, the standard deviation of  $\hat{\beta}_1$  is getting larger for coarser grid settings. The dynamic grid setting yields the best performance. The mean values of  $\hat{\beta}_0$  and  $\hat{\theta}^1$  from our 1,000 simulations are far-off from their theoretical values, fixed in the data-generating process. No relationship between the grid setup and the precision can be inferred.

However and most importantly, if we compute the sum of  $\hat{\beta}_0$  and  $\hat{\theta}^1$  for each simulation and therefrom the average and standard deviation, we find that the point estimate of the sum  $\hat{\beta}_0 + \hat{\theta}^1$  (transaction costs) has a high estimation precision, which improves with finer grids. This finding is not surprising because, in our model setup,  $\beta_0$  is determined outside the process of determining the threshold and then the threshold  $\theta$  is computed for a given  $\beta_0$ . Hence,  $\beta_0$

and  $\theta$  cannot be identified simultaneously, however the sum can. These results (also shown in Figure 4) are very promising, because for arbitrageurs the sum  $\hat{\beta}_0 + \hat{\theta}^1$  is more important than the individual parameter estimates.



**Figure 4:** 2-regime TVECM  $\mathcal{H}_2$  - distribution of parameter estimates

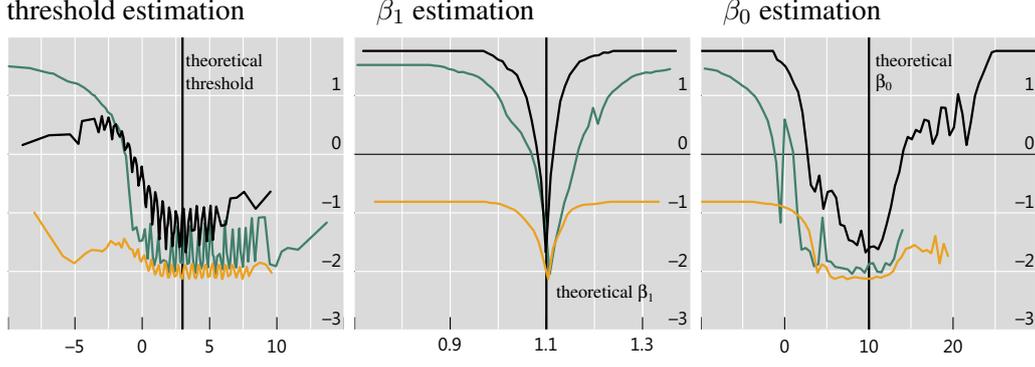
The graphs show the empirical distribution resulting from 1,000 simulations. The horizontal axis shows the values of the estimates and the vertical axis shows their frequency.  $\hat{\beta}_1$  and  $\hat{\beta}_0 + \hat{\theta}^1$  are narrowly distributed, whereas  $\hat{\beta}_0$  and  $\hat{\theta}^1$  have many and far away outliers.  $\beta_0$ ,  $\beta_1$ , and  $\theta^1$  are fixed to 10, 1.1, and 3, respectively.

The graphs in Figure 4 contain, despite their similarity, several interesting features. As shown in Table 1, the means generated from our 1,000 simulations are close to the theoretically expected value. The range of the horizontal axis is determined by the chosen grid. The plot for the estimator of  $\beta_1$  is shifted to the right, because we calibrated the grid to include the theoretical value  $\hat{\beta}_1 = 1$ . For  $\hat{\beta}_0$  and  $\hat{\theta}^1$ , we find that the weight of the distribution is heavily centered around the values of the data-generating process. However, we find several outliers far away from the theoretical values. The most notable finding is, that if we look at the very right graph in Figure 4, namely the sum of  $\hat{\beta}_0 + \hat{\theta}^1$ , we see that the dispersion is dramatically lower compared to its individual parameter estimates (two graphs in the middle).

The negative log-likelihood functions in Figure 5, which need to be minimized, show a non-smooth behavior, hence the proposed grid search is necessary. The graphs show that the estimation of  $\beta_1$  has a higher precision than  $\beta_0$  and  $\theta^1$ .

#### 2.4.2. Testing for a second threshold in model class $\mathcal{H}_2$

As a next and important step, we generalize the simulation to search and test for two thresholds. Based on our no-arbitrage discussion in Section 1, we assume that we have at most one positive and/or one negative threshold. The previous analysis shows that the dynamic grid



**Figure 5:** 2-regime TVECM  $\mathcal{H}_2$  - negative log-likelihood functions

The figures show the non-differentiable nature of the log-likelihood function for three simulations. The horizontal axis represents the grid search area. It is obvious from the shape of the curves that the precision for  $\beta_1$  is high and poor for the other parameter estimates.  $\beta_0$ ,  $\beta_1$ , and  $\theta^1$  are fixed to 10, 1.1, and 3, respectively.

point setting is the most precise methodology to estimate the transaction costs (see Table 1).

Therefore, from now on we use this grid setting only.

**Table 1:** 2-regime TVECM  $\mathcal{H}_2$  - estimation precision

The table shows the means and standard deviations (in parentheses) of 1,000 parameter estimates and their theoretical values (second row). In the columns underneath *fixed regressor* and *residual*, we report how often the null hypothesis (VECM) is incorrectly not rejected at the 95% and 99% confidence levels (e.g., a value 0.004 below 95% means that in 4 out of 1,000 simulation the null hypothesis is not rejected at the 95% confidence level). “dyn” denotes the dynamic grid setting, while the other grids have a fixed number of grid points.

	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\theta}^1$	$\hat{\beta}_0 + \hat{\theta}^1$	fixed regressor		residual	
grid	1.1	10	3	13	95%	99%	95%	99%
10	1.10 (0.29)	7.49 (50.56)	4.79 (47.48)	12.28 (4.37)	0.004	0.007	0.005	0.005
100	1.10 (0.04)	5.42 (16.63)	7.49 (16.00)	12.91 (1.33)	0.005	0.007	0.005	0.012
dyn	1.10 (0.02)	8.39 (32.99)	4.46 (33.00)	12.85 (1.17)	0.002	0.007	0.005	0.012

As in the previous subsection, we use the same model parameters and the same model, a 1-threshold VECM  $\mathcal{H}_2$ , however, we test whether the residual and fixed regressor bootstrap methodologies reject the existence of a second threshold.

We fix  $\beta_1$  during the second threshold search to the precisely estimated parameter value found in the first threshold search and hence reduce the dimensionality of the grid space to two dimensions ( $\beta_0$  and  $\theta^2$ ). We performed simulations with  $\beta_1$  not fixed during the second threshold search and observed a poor computational and statistical performance.

In order to convince the reader about the reliability of the methodology for the second threshold search, we also test different time series lengths. The results are presented in Table 2.

**Table 2:** 2-regime TVECM  $\mathcal{H}_2$  - estimation precision (fixed  $\beta_1$ )

The table shows the parameter estimates and standard deviations (in parentheses) for the two threshold searches using a dynamic grid setup and  $\beta_1$  is fixed in the second search. The theoretical parameter values are shown in the second row. The data-generating process has no second threshold, hence n/a.

	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\theta}^1$	$\hat{\beta}_0 + \hat{\theta}^1$	$\hat{\beta}_0$	$\hat{\theta}^2$
periods	1.1	10	3	13	10	n/a
1,000	1.10 (0.02)	8.39 (32.99)	4.46 (33.00)	12.85 (1.17)	8.64 (19.90)	-10.38 (19.82)
2,000	1.10 (0.01)	9.35 (7.93)	3.57 (7.91)	12.93 (0.72)	8.09 (13.83)	-9.51 (13.72)

As in the previous simulations, the estimators  $\hat{\beta}_1$  as well as  $\hat{\theta}^1 + \hat{\beta}_0$  are very precise. The precision of the estimation increases with the number of observations (i.e., with the time series length). The results presented here are for the restricted case; however, the use of the dynamic grid setting yields similar results in the unrestricted case, but at increased computational cost, due to the larger grid space.

The results of the bootstrap algorithms for both threshold searches are given in Table 3. Again, we find that the fixed regressor and the residual bootstrap produce consistent and acceptable  $\alpha$ - and  $\beta$ -errors.

**Table 3:** 2-regime TVECM  $\mathcal{H}_2$  - bootstrap precision

The table shows the  $\beta$ -errors for the threshold search for  $\theta^1$  and the  $\alpha$ -errors of the  $\theta^2$  search. For the first threshold bootstrap we report similar to Table 1 how often the null hypothesis (VECM) is incorrectly not rejected at the 95% and 99% confidence levels. For the second threshold bootstrap we report how often the null hypothesis (this time a 2-regime TVECM) is incorrectly rejected at the 95% and 99% confidence levels.

periods	first threshold ( $\theta^1$ ) bootstrap				second threshold ( $\theta^2$ ) bootstrap			
	fixed regressor		residual		fixed regressor		residual	
	95%	99%	95%	99%	95%	99%	95%	99%
1,000	0.002	0.007	0.005	0.012	0.15	0.07	0.15	0.06
2,000	0.003	0.006	0.004	0.014	0.30	0.23	0.30	0.18

### 2.4.3. Model class $\mathcal{H}_3$ : 3-regime TVECM

Finally, we present the results for a data-generating process with two thresholds. Without loss of generality, we show the results for the following choice of parameters:  $\beta_1 = 1.1$  and  $\beta_0 = 1$ , as well as two thresholds  $\theta^1 = -4$  and  $\theta^2 = 6$ . We again use the dynamic grid point method and keep  $\beta_1$  for the second threshold search fixed as we could show that this yields the best results. We perform a variety of additional robustness checks. For example, we extend the time series length to 5,000 and we vary the precision parameter settings in the dynamic grid search. As previously, we use the standard setting  $\Delta\beta_0 = 0.5$ ,  $\Delta\beta_1 = 0.01$ , and  $\Delta\theta^j = 0.5$ , but also consider cases where  $\Delta\beta_0 = 0.1$ ,  $\Delta\beta_1 = 0.005$ , and  $\Delta\theta^j = 0.1$ .<sup>7</sup>

We perform a sequential threshold search. Depending on the time series, either of the two thresholds may be found in the first round. Table 4 shows that the estimation of  $\beta_1$  is again very precise, whereas the estimate of  $\beta_0$  is poor. For the estimation of  $\beta_1$ , we find a clear improvement of precision with increasing time series length and smaller grid spacing. The estimate of  $\beta_0$  shows a large standard deviation, which seems independent from the time series length and the grid spacing. Further, the threshold estimate is poor, for two reasons: First, we know from the one threshold case that the estimation of the threshold is imprecise and second, in the 2-threshold case the estimation procedure may find either the negative or the positive threshold during the first round.

**Table 4:** 3-regime TVECM  $\mathcal{H}_3$  - first threshold search estimation precision

The table shows the parameter estimates and standard deviations (in parentheses) for the first threshold search using a dynamic grid setup. The theoretical parameter values are shown in the second row.

		$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\theta}$	$\hat{\beta}_0 + \hat{\theta}^1$	$\hat{\beta}_0 + \hat{\theta}^2$
periods	precision	1.1	1	-4 or 6	-3	7
1,000	0.01/0.5	1.10 (0.02)	0.08 (6.15)	2.43 (8.69)	-2.98 (0.65)	6.94 (1.59)
1,000	0.005/0.1	1.10 (0.01)	-0.24 (5.89)	2.69 (8.43)	-2.94 (0.34)	6.86 (0.89)
5,000	0.01/0.5	1.10 (0.01)	-0.10 (21.03)	2.65 (21.90)	-3.00 (0.89)	6.75 (1.33)

Again, the most important finding is that the sum of  $\hat{\beta}_0$  and the threshold is for the negative

<sup>7</sup> The unrestricted time series setup yields robust results, because of the use of the dynamic grid point setting. These results and further robustness checks can be found in Ters and Urban (2018).

values very close to  $\hat{\beta}_0 + \hat{\theta}^1 = -3$  and for the positive values very close to  $\hat{\beta}_0 + \hat{\theta}^2 = 7$ . We find a clear dependence of the quality of the estimator on the time series length and the grid spacing. The  $\beta$ -errors of the two bootstrap methods are below 0.002 at the 95% and 99% confidence levels, showing the high power of the tests.

The results for the second threshold search, presented in Table 5, are extremely promising, because the two transaction costs are well estimated. In Table 5, we find a similar picture to the first threshold search, however, as expected, with slightly bigger standard deviations. Again, the sum of  $\hat{\beta}_0$  and the threshold are for the negative values very close to  $\hat{\beta}_0 + \hat{\theta}^1 = -3$  and for the positive values very close to  $\hat{\beta}_0 + \hat{\theta}^2 = 7$ . We observe a higher precision for longer time series length and for smaller grid spacing.

**Table 5:** 3-regime TVECM  $\mathcal{H}_3$  - second threshold search estimation precision

The table shows for different time series length (periods) and precision parameters the estimates and standard deviations (in parentheses) for the second search using a dynamic grid setup.  $\beta_1$  is fixed to the value found during the first search, hence n/a. The theoretical parameter values are given in the second row.

		$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\theta}$	$\hat{\beta}_0 + \hat{\theta}^1$	$\hat{\beta}_0 + \hat{\theta}^2$
periods	precision	1.1	1	-4 or 6	-3	7
1,000	0.01/0.5	n/a	0.84 (4.28)	0.43 (7.56)	-2.91 (1.80)	6.28 (2.22)
1,000	0.005/0.1	n/a	0.99 (6.43)	0.25 (9.25)	-2.96 (2.48)	6.42 (2.04)
5,000	0.01/0.5	n/a	2.36 (19.75)	-1.07 (20.90)	-2.63 (1.84)	6.34 (2.10)

The bootstrap precision for the second threshold search is satisfying, even though, as expected, not as good as in the search for the first threshold. The power of the tests increases with the time series length.

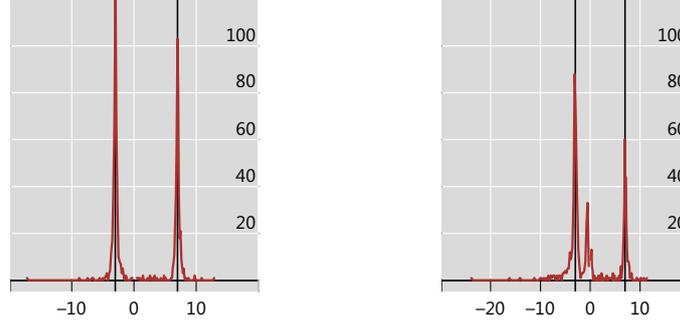
The negative log-likelihood function shows a non-differentiable behavior similar to Figure 5 for the 1-threshold case.

The distribution of the estimators of the transaction costs for the first and second threshold search are presented in Figure 6.

#### 2.4.4. Conclusion from the simulation exercise

The proposed methodology leads to accurate and robust results for the 1- and 2-threshold VECM. The transaction costs  $\theta^i + \beta_0$  as well as  $\beta_1$  can be estimated with high accuracy. The

First search:  $\hat{\beta}_0 + \hat{\theta}^i$  distribution    Second search:  $\hat{\beta}_0 + \hat{\theta}^i$  distribution



**Figure 6:** 3-regime TVECM  $\mathcal{H}_3$  - distribution of transaction cost estimates

The graphs show the estimation of  $\hat{\beta}_0 + \hat{\theta}^i$ , obtained from the two threshold searches, based on 1,000 MC simulations. The horizontal axis shows the  $\hat{\beta}_0 + \hat{\theta}^i$  values and the vertical axis shows their frequency. The distribution shows only a few outliers closely located to the theoretical value. The behavior of  $\hat{\beta}_1$ ,  $\hat{\beta}_0$ , and  $\hat{\theta}^i$ , not shown here, is very similar to Figure 4. The theoretical values for  $\beta_0 + \theta^i$  are -3 and 7.

dynamic grid search strategy has turned out to be superior. Despite the non-differentiable structure of the maximum likelihood function, we find that the accuracy of the estimation improves with smaller precision parameters and also with an increasing number of observations in the time series.  $\beta_1$  can be fixed in the second threshold search to the value found in the first search.

### 3. Application: Commodity arbitrage

As a simple illustration, we apply our model to commodity trading for palladium.

We are interested in estimating the thresholds for basis trading, which represent the unknown transaction costs. Further, the speeds of adjustments in the different regimes provide an insight on how fast the different prices revert back to the long-run equilibrium. From the speed of adjustments in equation (1), we can compute the Hasbrouck (HAS) and Gonzalo-Granger (GG) measures of price discovery (e.g., Hasbrouck, 1995; Gonzalo and Granger, 1995). HAS and GG are VECM-based measures. Hence, the error correction term must be  $I(0)$ . Further, they are applicable to a situation with a general cointegration vector (e.g., Figuerola-Ferretti and Gonzalo, 2010; Man et al., 2013). The different speeds of adjustments give an intuitive picture, which market moves first and faster towards the long run equilibrium.

We define HAS as the average of the two independent measures:  $HAS_1$  and  $HAS_2$  (e.g., Man et al., 2013). The two GG measures,  $GG^{\text{spot}}$  and  $GG^{\text{futures}}$ , are dependent and add up to 1. We use  $GG^{\text{spot}}$  and suppress the superscript. Further, for simplicity we also suppress the label

indicating the regime dependence of the HAS and GG measures.

HAS and GG measures greater than 0.5 imply that more than 50% of the price discovery occurs in the spot market. When the measures are close to 0.5, both markets contribute to price discovery without evidence on which market is dominant.<sup>8</sup>

The spot and futures time series in the bivariate VECM share a common stochastic trend, therefore the impulse response function for the cointegrating residual can be used to determine the speed of adjustment to the long-run equilibrium (Zivot and Wang, 2006). The vector error correction mechanism directly links the speed of adjustments to the cointegration residual  $u_t$ , which follows an implied AR(1) process:

$$u_t = (1 + \lambda_1 - \beta_1 \lambda_2)u_{t-1} + \varepsilon_t^{\text{spot}} - \beta_1 \varepsilon_t^{\text{futures}} \equiv \phi u_{t-1} + \varepsilon_t^{\text{spot}} - \beta_1 \varepsilon_t^{\text{futures}}. \quad (16)$$

The half-life of a shock can now be calculated from the AR(1) coefficient  $\phi$  as  $n = \ln(0.5)/\ln(\phi)$ .

We test whether all involved daily time series are I(1) and that the spot and future prices are cointegrated. Seo (2006) provides cointegration tests for the TVECM. Further, we test whether the regression residual sample means are zero to ensure the suitability of the model. The close relationship (cointegration) between the two time series ensures that the sample mean in our model is zero for our applications (basis trading and arbitrage).

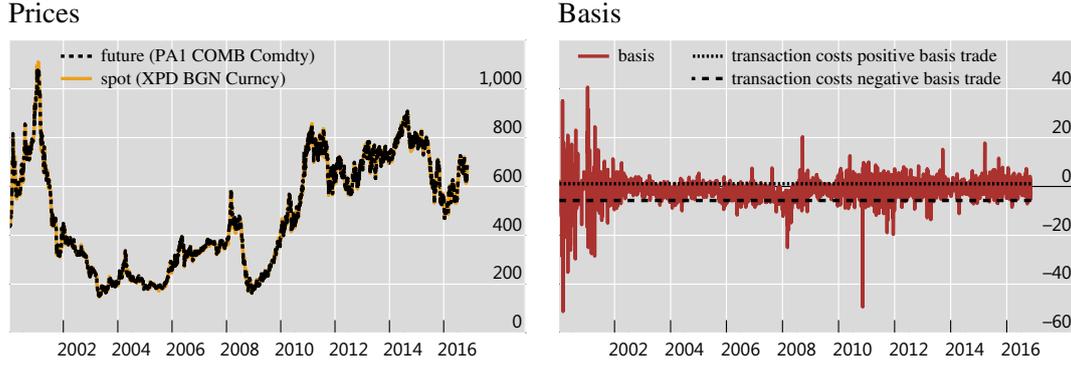
Palladium is a precious metal similar to platinum. It is used as an investment and industrial commodity. Its main industrial use is in catalytic converters. The time series of the spot, futures as well as the basis from 2000 to 2017 are presented in Figure 7.

We find evidence for two thresholds, resulting in two transaction costs, for a negative basis trade  $\hat{\beta}_0 + \hat{\theta}^1 = -5.7\text{USD}$  and a positive basis trade  $\hat{\beta}_0 + \hat{\theta}^2 = 1.1\text{USD}$  per contract size of a 100 troy ounce for a model with one VAR lag. The transaction costs for the negative basis trade have to be interpreted in absolute terms. The two transactions costs are shown as horizontal lines in the right-hand panel of Figure 7.

The estimation procedure shows that the model fit based on SIC is not straight-forward because the value of the SIC is declining with a weak slope, reaching a plateau at around 7

---

<sup>8</sup> Unlike the HAS measure, the GG measure is not mathematically confined to the closed interval [0,1], which seems to make an interpretation similar to the Hasbrouck measure difficult. GG measures below 0 and above 1 should be interpreted as 0 and 1, respectively (see Gyntelberg et al. (2013) for more discussion).



**Figure 7:** Palladium spot, future prices, and spot-future basis

The figure on the left-hand side shows the palladium spot and future prices in USD for a contract size of 100 troy ounces. The figure on the right-hand side shows the basis, as well as the two transaction costs (horizontal lines). Source: Bloomberg, authors' calculations

lags before declining further with an increasing number of lags. The most parsimonious model suggests the existence of two thresholds with high significance (see Table 6). Apart from lag=2, the transaction costs for a negative basis trade are stable. The lower regime is an arbitrage regime, which as expected contains only 11% of all observations. We also find that the upper regime appears to be an arbitrage regime because it contains 12% of the overall observations.

**Table 6:** Threshold search - Palladium

The table shows the results for the threshold searches for  $\theta^1$  and  $\theta^2$  for the period from April 2000 to end December 2016. The value of SIC is declining until it reaches a local minimum at lag =7 (results are provided upon request). The most parsimonious model (lag=1) has a significant threshold. \*\* and \*\*\* denote the 95% and the 99% confidence levels, respectively.  $\lambda_{iL/U}^j$  are the regime dependent speed of adjustments ( $L/U$  denote the lower/upper regime).

First threshold search:

lag	SIC	Bootstrap		$\hat{\beta}_1$	$\hat{\beta}_0 + \hat{\theta}^1$	$\lambda_{1L}^1$	$\lambda_{2L}^1$	$\lambda_{1U}^1$	$\lambda_{2U}^1$	lower regime
		fixed	residual							
1	7.22	0.06	0.08	1.002	-5.75	-0.14	0.60***	-0.37***	0.28***	11%
2	7.19	0.44	0.52	1.003	0.67	-0.18	0.39***	-0.29***	0.39***	88%
3	7.16	0.67	0.74	1.003	-6.36	0.12	0.50**	-0.19***	0.31***	10%
4	7.08	0.28	0.34	1.002	-5.66	0.20	0.51**	-0.17	0.25**	10%

Second threshold search (relevant first lag only):

lag	SIC	Bootstrap		$\hat{\beta}_1$	$\hat{\beta}_0 + \hat{\theta}^2$	$\lambda_{1L}^2$	$\lambda_{2L}^2$	$\lambda_{1U}^2$	$\lambda_{2U}^2$	upper regime
		fixed	residual							
1	7.23	0.02	0.03	1.002	1.14	-0.28***	0.44***	-0.49***	0.30**	12%

In addition to the estimated transaction cost, we compute measures of price discovery for the relevant lag=1, which are presented in Table 7.

**Table 7:** Measures of price discovery for one lag - Palladium  
 Results of the neutral regime are averages from the first and second threshold search.

regime	HAS	GG	leading market	half life (in days)
upper	0.48	0.38	futures	0.44
neutral	0.46	0.56	-	0.60
lower	0.56	0.81	spot	0.51

In the upper regime, we find some indication that the futures market leads the price discovery process, whereas in the lower regime there is leadership of the spot market. Both markets seem to absorb one-unit shocks rather quickly. The neutral regime has as expected a longer half life, however the difference is not very large. Moreover, as expected, we have no clear indication of which market leads the price finding in the neutral regime.

#### 4. Conclusion

The identification of unknown thresholds in non-linear vector correction models is of rather complex origin with several unresolved problems. In this paper, we present the solution for one unresolved issue, which is the estimation of a 3-regime multivariate TVECM with an unknown cointegrating vector. Our model extends the 2-regime TVECM as proposed by Hansen and Seo (2002). In contrast to Hansen and Seo (2002), we also introduce an intercept  $\beta_0$  in the cointegrating relation  $(S - \beta_1 D - \beta_0)_t$  to account for a distorted parity relationship (e.g., the non-zero basis). Using a sequential grid search for the first and second threshold search, we estimate the cointegration relationship, as well as the two thresholds ( $\theta^1$  and  $\theta^2$ ) by employing a maximum likelihood approach. As there are practically no empirical studies for TVECMs with an unknown cointegrating vector even for the 1-threshold case, we present a comprehensive simulation study to investigate the reliability of our model. Hansen and Seo (2002) have suggested a grid point search to estimate the variables  $\beta_i$  and  $\theta^i$  with a fixed number of grid points. We show that a dynamic grid point setting, where the distance between two grid points is set via a precision parameter, instead of defining a fixed number of grid points, yields the best results, albeit at potentially high computational costs. The slope  $\beta_1$  in the cointegrating relationship is estimated at a very high degree of precision. We show that the value  $\hat{\beta}_1$  found in the first threshold search can be fixed in the second search. This lowers the dimension of

the grid space and hence reduces the computing time in the second grid search. The thresholds  $\theta^i$  and the shift in the cointegration relation  $\beta_0$  are poorly estimated, but the sum  $\theta^i + \beta_0$  is estimated with a very high degree of precision.

Our model is particularly appealing for the analysis of distorted parity relationships in economics, such as a persistent non-zero basis, which points towards the presence of transaction costs. In our applications, we show that our 3-regime TVECM can estimate unknown transaction costs on arbitrage trades (sum of  $\theta^i + \beta_0$ ) with a very high degree of precision. In addition, we computed measures of price discovery, as well as half-lives of shocks. These measures could help regulators evaluate which markets are more dynamic, and hence provide them with a better picture of market activity.

## References

- Anderson, H. M. (1997). Transaction costs and nonlinear adjustment towards equilibrium in the us treasury bill market, *Oxford Bulletin of Economics and Statistics* **59**(4): 465–484.
- Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point, *Econometrica* **61**: 821–1414.
- Bai, J. (1997). Estimating multiple breaks one at a time, *Econometric Theory* **13**(3): 315–352.
- Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes, *Econometrica* **66**(1): 47–78.
- Baillie, R. T., Booth, G. G., Tse, Y. and Zobotina, T. (2002). Price discovery and common factor models, *Journal of Financial Markets* **5**: 309–321.
- Balke, N. S. and Fomby, T. B. (1997). Threshold cointegration, *International Economic Review* **38**(3): 627–645.
- Berkman, H., Brailsford, T. and Frino, A. (2005). A note on execution costs for stock index futures: Information versus liquidity effects, *Journal of Banking and Finance* **29**: 565–577.
- Davies, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative, *Biometrika* **74**(1): 33–43.
- Dumas, B. (1992). Dynamic equilibrium and the real exchange rate in a spatially separated world, *Review of Financial Studies* **5**(2): 153–180.
- Figuerola-Ferretti, I. and Gonzalo, J. (2010). Modelling and measuring price discovery in commodity markets, *Journal of Econometrics* **158**: 95–107.
- Forbes, C. S., Kalb, G. R. J. and Kofhian, P. (1999). Bayesian arbitrage threshold analysis, *Journal of Business & Economic Statistics* **17**: 364–372.
- Gaul, J. and Theissen, E. (2015). A partially linear approach to modeling the dynamics of spot and futures prices, *The Journal of Futures Markets* **35**(4): 371–384.
- Gonzalo, J. and Granger, C. W. J. (1995). Estimation of common long-memory components in cointegrated systems, *Journal of Business & Economic Statistics* **13**(1): 27–36.
- Granger, C. W. J. and Teräsvirta, T. (1993). *Modelling Economic Nonlinear Relationships*, Oxford University Press.
- Gyntelberg, J., Hoerdahl, P., Ters, K. and Urban, J. (2013). Intraday dynamics of euro area sovereign cds and bonds, *BIS Working Papers 423*, Bank for International Settlements.
- Gyntelberg, J., Hoerdahl, P., Ters, K. and Urban, J. (2017). Arbitrage costs and the persistent non-zero cds-bond basis: Evidence from intraday euro area sovereign debt markets, *BIS Working Papers 631*, Bank for International Settlements.
- Gyntelberg, J., Hoerdahl, P., Ters, K. and Urban, J. (2018). Price discovery in euro area sovereign credit markets and the ban on naked cds, *Journal of Banking and Finance* **96**: 106–125.
- Hansen, B. E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis, *Econometrica* **64**: 413–430.

- Hansen, B. E. (2000). Testing for structural change in conditional models, *Journal of Econometrics* **97**: 93–115.
- Hansen, B. E. and Seo, B. (2002). Testing for two-regime threshold cointegration in vector error-correction models, *Journal of Econometrics* **110**(2): 293–318.
- Hasbrouck, J. (1995). One security, many markets: Determining the contributions to price discovery, *Journal of Finance* **50**(4): 1175–1199.
- Johansen, S. (1988). Statistical analysis of cointegrated vectors, *Journal of Economic Dynamics and Control* **12**: 231–254.
- Lütkepohl, H. (2006). *New Introduction to Multiple Time Series Analysis*, Springer.
- Man, K., Wang, J. and Wu, C. (2013). Price discovery in the u.s. treasury market: Automation versus intermediation, *Management Science* **59**: 695–714.
- Martens, M., Kofman, P. and Vorst, T. (1998). A threshold error-correction model for intraday futures and index returns, *Journal of Applied Econometrics* **13**: 245–263.
- Panos, M., Nobay, A. R. and Peel, D. A. (1997). Transactions costs and nonlinear adjustment in real exchange rates: An empirical investigation, *Journal of Political Economy* **105**(5): 862–879.
- Seo, M. H. (2006). Bootstrap testing for the null of no cointegration in a threshold vector error correction model, *Journal of Econometrics* **134**: 129–150.
- Seo, M. H. (2011). Estimation of nonlinear error correction models, *Econometric Theory* **27**: 201–234.
- Stevens, J. (2015). Do transaction costs prevent arbitrage in the market for crude oil? Evidence from a threshold autoregression, *Applied Economic Letters* **22**: 169–172.
- Teräsvirta, T. (1994). Specification, estimation and evaluation of smooth transition autoregression models, *Journal of the American Statistician* **89**: 208–218.
- Ters, K. and Urban, J. (2018). Estimating unknown arbitrage costs: evidence from a three-regime threshold vector error correction model, *BIS Working Paper 689*, Bank for International Settlements.
- Theissen, E. (2012). Price discovery in spot and futures markets: a reconsideration, *The European Journal of Finance* **18**(10): 969–987.
- Tsay, R. S. (1989). Testing and modeling threshold autoregressive processes, *Journal of the American Statistical Association* **84**(405): 231–240.
- Zivot, E. and Wang, J. (2006). *Modeling Financial Time Series with S-PLUS*, Springer-Verlag New York, Inc.