Department of Economics
Working Paper No. 250

Threshold Cointegration and Adaptive Shrinkage

Florian Huber
Thomas O. Zörner

June 2017

WIRTSCHAFTS
UNIVERSITÄT
WIEN VIENNA
UNIVERSITY OF
ECONOMICS
AND BUSINESS

EFMD EQUIS ACCREDITED
AACSB ACCREDITED
ASSOCIATION OF MBAS ACCREDITED

Threshold cointegration and adaptive shrinkage

Florian Huber* and Thomas O. Zörner

Vienna University of Economics and Business

Abstract

This paper considers Bayesian estimation of the threshold vector error correction (TVECM) model in moderate to large dimensions. Using the lagged cointegrating error as a threshold variable gives rise to additional difficulties that are typically solved by relying on large sample approximations. Relying on Markov chain Monte Carlo methods we circumvent these issues by avoiding computationally prohibitive estimation strategies like the grid search. Due to the proliferation of parameters we use novel global-local shrinkage priors in the spirit of Griffin and Brown (2010). We illustrate the merits of our approach in an application to five exchange rates vis-á-vis the US dollar and assess whether a given currency is over or undervalued. Moreover, we perform a forecasting comparison to investigate whether it pays off to adopt a non-linear modeling approach relative to a set of simpler benchmark models.

Keywords: non-linear modeling, shrinkage priors, multivariate cointegration, exchange rate modeling.

JEL Codes: C11, C32, C53, F31, F47.

*Corresponding author: Florian Huber, Vienna University of Economics and Business. E-mail: fhuber@wu.ac.at. The second author gratefully acknowledges financial support from the Austrian National Bank, Jubilaeumsfond grant no. 16748.

1 Introduction

Economists employed in central banks and governmental institutions have been relying on small-scale linear models that are easily estimated and proved to work well during the run-up to the last financial crisis that hit the world economy in 2008. In light of the most pronounced economic downturn since the great depression, policy makers switched to extraordinary fiscal and monetary policy strategies that pushed interest rates to zero rapidly. In such an environment, linear models suddenly failed to work since the underlying transmission channels changed markedly, calling for non-standard econometric methods that are flexible enough to incorporate such shifts in the modeling framework.

Another important regularity commonly observed in macroeconomic data is cointegration between several key quantities of interest. Since standard estimation tools for vector error correction models (VECMs) heavily rely on pre-testing and the interpretation of the estimation results becomes somewhat more involved, most practitioners stick to standard vector autoregressive (VAR) models estimated either in differences or in log-levels. While the Bayesian literature (Sims et al., 1990; Sims and Uhlig, 1991) suggests that in general, estimation of the underlying multivariate time series models in log-levels does not lead to a miss-specified model because long-run relations are implicitly taken into account, common practice is to transform the data to be approximately stationary and thus rule out long-run relationships. Nevertheless, estimation of VECMs provides additional inferential possibilities like the explicit discrimination between short- and long-run dynamics.

The literature on non-linear error correction models takes both points mentioned seriously. The seminal contribution by Balke and Fomby (1997), for instance, assumes that any non-linearities stem from a thresholding mechanisms that explicitly depends on the magnitude of the cointegrating error. This implies that if deviations from the long-run equilibrium are large, adjustment mechanics change, capturing the notion that economic agents change their behavior when faced with a situation characterized by sustained disequilibrium. Another prominent example where this sort of framework applies is the modeling of financial market arbitrage, where participants only engage in trading activities if the deviation of an asset price from its fundamental or long-run value is sufficiently large (Martens et al., 1998; Forbes et al., 1999).

While threshold VECM (TVECM) models provide a great deal of flexibility and can capture salient features of the time series under scrutiny, they are also challenging to estimate. Early contributions have been restricted to fairly small dimensional settings, including only a single cointegration vector and two endogenous variables and regimes, rendering estimation of the underlying long-run relations and the thresholds feasible (Lo and Zivot, 2001; Hansen and Seo, 2002). Specifically, for such small models, numerical estimation procedures such as a two dimensional grid search are employed that become prohibitively slow when the dimensionality of the problem is increased. Another strand of the literature uses a plug-in estimator for the cointegration matrix based on a linear VECM or assumes that the cointegration matrix is known, leading to a relatively standard multivariate threshold model where the threshold variable is observed as opposed to latent.

In this paper, we contribute to the literature along three important dimensions, with two contributions being related to the Bayesian estimation of non-linear VECM models (Forbes et al., 1999; Balcombe, 2006; Balcombe and Rapsomanikis, 2008; Gefang, 2012; Jochmann and Koop, 2015) and a third one which is related to the literature on exchange rate modeling (Mark, 1995; Mark and Sul, 2001; Rapach and Wohar, 2002; 2004; Molodtsova and Papell, 2009; Huber, 2016; 2017). First, we propose a straightforward Markov chain Monte Carlo (MCMC) algorithm that enables us to estimate medium to large sized models in a computationally efficient manner. Moreover, our algorithm also allows us to estimate models that feature more than two regimes and more than a single cointegrating vector, i.e. a model with cointegration rank greater than unity. Second, since multivariate threshold time series models are heavily parameterized, we impose a set of global-local shrinkage priors in the spirit of Polson and Scott (2010) and Griffin and Brown (2010) on all coefficients of the model. These priors have been recently introduced in the VAR framework by Huber and Feldkircher (2016) and possess convenient statistical features that provides enough flexibility to allow for non-zero regression coefficients in the presence of heavy global shrinkage.

Finally, our last contribution relates to exchange rate modeling. As an empirical application we apply a three regime TVECM model to five exchange rate pairs relative to the US dollar. Inspection of several key features of our model reveals that we are able to infer periods of over/undervaluation of a given currency. In addition, a forecasting exercise provides some evidence that our model performs well against simpler linear models in terms of density predictions.

The remainder of the paper is structured as follows. Section 2 introduces the econometric model along with the prior specification adopted. Section 3 discusses the corresponding conditional posterior distributions and the MCMC algorithm while Section 4 presents the findings our empirical application. Finally, the last section summarizes and concludes the paper.

2 Econometric framework

2.1 The threshold vector error correction model

Let us assume that our object of interest is a M-dimensional vector $\boldsymbol{y}_t$ of $I(1)$ time series which are cointegrated with a $M \times c$-dimensional cointegration matrix $\boldsymbol{\beta}$. We let $\boldsymbol{w}_t(\boldsymbol{\beta}) = \boldsymbol{\beta}'\boldsymbol{y}_t$ denote a c-dimensional vector of error correction terms with the jth element denoted by $w_{jt}(\boldsymbol{\beta})$. Consistent with Balke and Fomby (1997) and Hansen and Seo (2002) we furthermore assume that adjustment to the long-run equilibrium happens non-linearly and depends on the magnitude of w_{jt}.

We let $\mathcal{R}_r$ $(r = 1, \ldots, R)$ denote the rth regime that is defined as,

$$\mathcal{R}_r = \{w_{jt-1}(\boldsymbol{\beta}) : \gamma_{r-1} \leq w_{jt-1}(\boldsymbol{\beta}) < \gamma_r\},$$

with $\gamma_0 = -\infty$ and $\gamma_R = \infty$. If $w_{jt-1}(\boldsymbol{\beta}) \in \mathcal{R}_r$, then the regime-specific VECM is given by

$$\Delta \boldsymbol{y}_t = \boldsymbol{\alpha}_r \boldsymbol{\beta}' \boldsymbol{y}_{t-1} + \boldsymbol{B}_{r1} \Delta \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{B}_{rp} \Delta \boldsymbol{y}_{t-p} + \boldsymbol{\eta}_t. \tag{2.1}$$

Within each regime r we let $\boldsymbol{\alpha}_r$ denote a $M \times c$ matrix of short-run adjustment coefficients, $\boldsymbol{B}_{rj}$ $(j = 1, \ldots, p)$ a set of regime-specific autoregressive coefficient matrices of dimension $M \times M$ and $\boldsymbol{\eta}_t$ is a normally distributed vector white noise error with regime-specific $M \times M$ variance-covariance matrix $\boldsymbol{\Sigma}_r$, i.e.

$$\boldsymbol{\eta}_t \sim \mathcal{N}(\mathbf{0}_M, \boldsymbol{\Sigma}_r). \tag{2.2}$$

This model separates the M-dimensional Euclidean space into R regimes where within each regime a linear VECM is adopted.

The specific model structure has interesting implications on the dynamic adjustment mechanisms back towards equilibrium. For instance, if the jth equilibrium error gets large, a regime transition takes place and most coefficients of the model change. Thus, while $\boldsymbol{\alpha}_t$ determines the short-run adjustment back to equilibrium, the matrices $\boldsymbol{B}_{rj}$ determine short-run dynamics within each regime. It is also noteworthy that our model is heteroscedastic since we allow the variance-covariance matrix to change, capturing the notion that specific regimes might be characterized by higher uncertainty.

The only set of coefficients that is assumed to remain constant are the r cointegrating vectors in $\boldsymbol{\beta}$. Since these measure long-run relations we rule out structural breaks and assume that the underlying coefficients stay constant over time. Allowing for $\boldsymbol{\beta}$ to be regime-dependent, however, would be straightforward and the corresponding MCMC algorithm outlined in Section 2.4 still applies with minor modifications.

Before proceeding, a few words on the challenges involved in the estimation of the model given by Eq. (2.1) are in order. First, the presence of the thresholding mechanism leads to a ragged likelihood function that calls for numerical methods to estimate the model. Since the cointegration matrix is a parameter to be estimated, the proposed threshold variable is latent. This gives rise to additional difficulties which are generally tackled by either relying on approximations (i.e. estimation of a linear VECM and using the corresponding estimate of $\boldsymbol{\beta}, \hat{\boldsymbol{\beta}}$, as a plug-in estimator in the second step) or utilizing a grid search over the thresholds and the cointegration matrix. The latter procedure is computationally intensive and restricts the researcher to stick to small and simple models. Relying on approximations that are justified based on asymptotics is also generally problematic if $\boldsymbol{y}_t$ contains quarterly macroeconomic data with relatively few observations.

A second difficulty arises since the number of parameters to estimate becomes large. As is well known from the standard VAR literature (Doan et al., 1984; Sims and Zha, 1998; Bańbura et al., 2010; Giannone et al., 2015; Huber and Feldkircher, 2016), VAR models feature a proliferation of parameters which in turn leads to the curse of dimensionality. In our model this problem is vastly intensified since the number of parameters we have to estimate is multiplied by R. In typical macroeconomic applications, the number of available observations is rather limited and M might be moderate or large. This calls for shrinkage methods that softly shrink the corresponding coefficients towards a fixed value (often to zero).

Finally, the model in Eq. (2.1) is not identified and we thus have to impose identifying assumptions on $\boldsymbol{\alpha}_r$ or $\boldsymbol{\beta}$ since they appear in product form, implying that $\boldsymbol{\alpha}_r\boldsymbol{\beta}' = \boldsymbol{\alpha}_r\boldsymbol{U}^{-1}\boldsymbol{U}\boldsymbol{\beta}'$ for any invertible matrix $\boldsymbol{U}$. The traditional approach is to impose the normalization $\boldsymbol{\beta} = (\boldsymbol{I}_c, \boldsymbol{\xi}')'$. This implies that we impose c^2 restrictions on $\boldsymbol{\beta}$ and the model is thus just identified. From a practical perspective, this identification scheme is sensitive with respect to permutations of the elements in $\boldsymbol{y}_t$, rendering the ordering of the time series an important modeling decision. Moreover, it is well known that sufficiently informative priors need to be imposed on $\boldsymbol{\alpha}_r$ and $\boldsymbol{\xi}$ since under weakly informative priors the posterior of $\boldsymbol{\xi}$ is heavy tailed with the possibility that no moments exist (Kleibergen and Van Dijk, 1994; Geweke, 1996).

A more recent approach to identification is to restrict the cointegration space to a semi-orthogonal subspace by imposing that $\boldsymbol{\beta}'\boldsymbol{\beta} = \boldsymbol{I}_r'$. This choice avoids some of the issues mentioned above and permits prior elicitation not in terms of the cointegrating vectors but in terms of the space spanned by the cointegrating vectors (Strachan, 2003; Koop et al., 2009). In the present paper, we adopt the first identification strategy since we specify informative priors on the free elements in $\boldsymbol{\beta}$ in the empirical application and due to the fact that we are interested in using w_{jt} as a threshold variable. Note, however, that with minor modifications we could also adopt an identification strategy based on restricting the cointegration space.

2.2 Prior specification

As mentioned previously, the model outlined in Section 2.1 is heavily parameterized. We thus opt for a Bayesian approach which circumvents issues related to irregularly behaved likelihood functions and allows for flexible shrinkage possibilities to reduce estimation uncertainty and select appropriate models in a flexible manner.

This is achieved by specifying a Normal-Gamma (NG) prior in the spirit of Griffin and Brown (2010) on the parameters of the model in Eq. (2.1). Specifically, we impose a Gaussian prior on the elements $\boldsymbol{b}_r = \text{vec}(\boldsymbol{B}_r) = [\boldsymbol{B}_{r1}, \dots, \boldsymbol{B}_{rp}]'$, denoted as

b_{rj} for $j = 1, \dots, K$ with $K = M(Mp)$, i.e.

$$b_{rj}|\tau_{rj}^2 \sim \mathcal{N}(0, \tau_{rj}^2), \ \tau_{rj}^2 \sim \mathcal{G}(\varphi_{br}, \varphi_{br}\lambda_{br}/2). \tag{2.3}$$

Here we let τ_{rj}^2 be a prior scaling factor which features its own Gamma distributed prior where φ_{br} and λ_{br} are prior hyperparameters. As Griffin and Brown (2010) note, λ_{br} controls the overall level of shrinkage, acting as a global shrinkage parameter that pulls all elements in $\boldsymbol{b}_r$ towards zero. By contrast, φ_{br} controls the excess kurtosis of the underlying marginal prior obtained by integrating out the τ_{rj}^2's. This implies that small levels of φ_{br} lead to a fat tailed marginal prior on b_{rj} that allows for non-zero regression coefficients even in the presence of a large global shrinkage parameter λ_{br}. For implementation, we stack all local scaling parameters in a K-dimensional vector $\boldsymbol{\tau}_r = (\tau_{1r}^2, \dots, \tau_{Kr}^2)'$.

For the global shrinkage parameter λ_{br}, we also adopt a Gamma prior,

$$\lambda_{br} \sim \mathcal{G}(d_b, e_b), \tag{2.4}$$

where $d_b = e_b = 0.01$ are prior hyperparameters with low values inducing heavy shrinkage on $\boldsymbol{b}_r$.

Similarly to the autoregressive coefficients, we adopt a NG prior on the elements of $\boldsymbol{a}_r = \text{vec}(\boldsymbol{\alpha}_r)$, denoted as a_{rj} $(j = 1, \dots, cM)$, i.e.

$$a_{rj}|\zeta_{rj}^2 \sim \mathcal{N}(0, \zeta_{rj}^2), \ \zeta_{rj}^2 \sim \mathcal{G}(\varphi_{ar}, \varphi_{ar}\lambda_{ar}/2). \tag{2.5}$$

Hereby, φ_{ar} and λ_{ar} denote again prior hyperparameters that control the overall degree of shrinkage and the excess kurtosis of the marginal prior. The local scaling parameters ζ_{rj}^2 are again stored in a cM-dimensional vector $\boldsymbol{\zeta}_r = (\zeta_{r1}^2, \dots, \zeta_{r,cN}^2)$. The shrinkage prior on the elements of $\boldsymbol{a}_r$ allows us to flexibly infer whether cointegration is present in regime r. Note that even in the case of heavy shrinkage (i.e. λ_{ar} being large) we still allow for non-zero adjustment coefficients through the properties of our global-local shrinkage prior specification. Again, we place a Gamma prior on $\lambda_{ar} \sim \mathcal{G}(d_a, e_a)$ and set $d_a = e_a = 0.01$. As Kleibergen and Van Dijk (1994) note, if the prior is specified to be too informative (i.e. in the limiting case $\lambda_{ar} = \infty$), the posterior distribution of $\boldsymbol{a}_r$ will have point mass on zero and this leads to local non-identification of $\boldsymbol{\beta}$. We avoid this issue by monitoring the performance of our MCMC algorithm.

For $\boldsymbol{\beta}$, we also impose a NG prior on each of the elements[1] in vec($\boldsymbol{\xi}$), denoted as ξ_j,

$$\xi_j|\phi_j^2 \sim \mathcal{N}(0, \phi_j^2), \ \phi_j^2 \sim \mathcal{G}(\varphi_\xi, \varphi_\xi\lambda_\xi) \text{ for } j = 1, \dots, c(M-c), \tag{2.6}$$

[1]Note that we specify the prior directly on the r columns of $\boldsymbol{\beta}$. As mentioned previously another feasible approach would be to elicit the prior directly on the space spanned by $\boldsymbol{\beta}$ (see Strachan, 2003).

with φ_ξ and λ_ξ denoting hyperparameters similarly to the ones outlined above. On the global shrinkage parameter we impose yet another Gamma prior, $\lambda_\xi \sim \mathcal{G}(d_\xi, e_\xi)$ with $d_\xi = e_\xi = 0.01$.

Since the hyperparameters φ_{ir} for $i \in \{b, a\}$ and φ_ξ play a vital role we an exponentially distributed prior to infer φ_{ir} from the data. Specifically, we set

$$\varphi_{ir} \sim Exp(\underline{\varphi}_i) \tag{2.7}$$
$$\varphi_\xi \sim Exp(\underline{\varphi}_\xi) \tag{2.8}$$

with $\underline{\varphi}_i = \underline{\varphi}_\xi = 0.1$ specified to place significant prior mass on low values of φ_{ir} and φ_ξ. This choice is based on recent findings in Huber and Feldkircher (2016) who report an empirical estimate of φ_{ir} around 0.1 for US data.

On each threshold γ_j we impose independent normally distributed priors given by

$$\gamma_j \sim \mathcal{N}(0, \psi_j^2), \text{ for } j = 1, \dots, r, \tag{2.9}$$

where we let ψ_j^2 be a prior scaling factor set to a rather large value (in our empirical application we specify $\psi_j^2 = 10^2$). In principle, we try to avoid introducing significant prior information on the specific threshold value γ_j. However, note that it would be straightforward to introduce other priors that place more prior mass on interesting regions of γ_j, if this information is available a priori.

Finally, we use an inverted Wishart prior on $\boldsymbol{\Sigma}_r$,

$$\boldsymbol{\Sigma}_r \sim \mathcal{IW}(v_r, \boldsymbol{S}_r). \tag{2.10}$$

The prior degrees of freedom are denoted by $v_r = M + 1$ and the $M \times M$-dimensional prior scaling matrix is given by $\boldsymbol{S}_r = \frac{1}{100}\boldsymbol{I}_M$.

2.3 Posterior distributions

After specifying a suitable set of prior distributions we apply Bayes theorem and combine the prior with the likelihood function. For most parameters this leads to relatively simple conditional posterior distributions. This enables Gibbs updating while for the threshold parameters we do not obtain a convenient conditional posterior distribution and thus have to rely on alternative methods to sample from the conditional posterior.

Conditional on $\boldsymbol{\gamma}$, $\boldsymbol{\beta}$, $\boldsymbol{\Sigma}_r$, $\boldsymbol{\tau}_r$, $\boldsymbol{\zeta}_r$ and the available data $\mathcal{D}$ the conditional posterior distribution of $(\boldsymbol{a}_r', \boldsymbol{b}_r')'$ takes a standard form (Zellner, 1986) and is given by

$$\begin{pmatrix} \boldsymbol{a}_r \\ \boldsymbol{b}_r \end{pmatrix} \Bigg| \; \boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\Sigma}_r, \boldsymbol{\tau}_r, \boldsymbol{\zeta}_r, \mathcal{D} \sim \mathcal{N}(\boldsymbol{\mu}_r, \boldsymbol{V}_r). \tag{2.11}$$

The posterior variance is given by

$$\boldsymbol{V}_r = (\boldsymbol{\Sigma}_r^{-1} \otimes \boldsymbol{Z}_r'\boldsymbol{Z}_r + \underline{\boldsymbol{V}}_r^{-1})^{-1}, \tag{2.12}$$

with $\boldsymbol{Z}_r'$ being a $T_r \times Mp+c$ matrix with typical row t given by $(\boldsymbol{y}_{t-1}'\boldsymbol{\beta}, \Delta\boldsymbol{y}_{t-1}', \dots, \Delta\boldsymbol{y}_{t-p}')$, T_r equals the number of observations in regime r and $\underline{\boldsymbol{V}}_r$ is a diagonal matrix with $\text{diag}(\underline{\boldsymbol{V}}_r) = (\boldsymbol{\zeta}_r', \boldsymbol{\tau}_r')'$.

The posterior mean is given by

$$\boldsymbol{\mu}_r = \boldsymbol{V}_r \left\{ (\boldsymbol{I}_M \otimes \boldsymbol{Z}_r'\boldsymbol{Y}_r)\text{vec}(\boldsymbol{\Sigma}_r) \right\}. \tag{2.13}$$

Hereby $\boldsymbol{Y}_r$ is constructed analogously to $\boldsymbol{Z}_r$ with typical t'th row given by $\Delta\boldsymbol{y}_t'$.

Turning to the conditional posterior distribution of $\boldsymbol{\beta}$ it is noteworthy that conditional on $\boldsymbol{a} = (\boldsymbol{a}_1', \dots, \boldsymbol{a}_R')', \boldsymbol{b} = (\boldsymbol{b}_1', \dots, \boldsymbol{b}_R')'$ and $\boldsymbol{\Sigma} = \{\boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_r\}$ it is possible to rewrite Eq. (2.1) as a standard linear regression model with regression coefficients $\boldsymbol{\beta}$. However, the presence of the identifying assumptions introduces additional difficulties (see Geweke, 1996). Fortunately, Villani (2001) demonstrates that conditional on the identifying restrictions it is possible to show that the posterior of $\text{vec}(\boldsymbol{\xi})$ follows a multivariate Gaussian distribution,

$$\text{vec}(\boldsymbol{\xi})|\boldsymbol{\gamma}, \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{\Sigma}, \mathcal{D} \sim \mathcal{N}(\boldsymbol{\delta}, \boldsymbol{\Lambda}). \tag{2.14}$$

Posterior variance and mean are given by

$$\boldsymbol{\Lambda} = [\boldsymbol{H}' \sum_{r=1}^{R} \left\{ (\boldsymbol{\alpha}_r'\boldsymbol{\Sigma}_r^{-1}\boldsymbol{\alpha}_r) \otimes \boldsymbol{Y}_r'\boldsymbol{Y}_r \right\} \boldsymbol{H} + \boldsymbol{\Omega}^{-1}]^{-1}, \tag{2.15}$$

$$\boldsymbol{\delta} = \boldsymbol{\Lambda}(\boldsymbol{H}' \sum_{r=1}^{R} \left\{ (\boldsymbol{\alpha}_r'\boldsymbol{\Sigma}_r^{-1} \otimes \boldsymbol{Y}_r')\text{vec}(\hat{\boldsymbol{y}}_r) \right\}). \tag{2.16}$$

We let $\boldsymbol{H}$ be a $cM \times c(M-c)$ matrix that imposes the linear identifying restrictions on $\boldsymbol{\beta}$, $\boldsymbol{Y}_r$ a $T_r \times M$ matrix with typical row given by $\boldsymbol{y}_{t-1}'$, $\boldsymbol{\Omega} = \text{diag}(\phi_1^2, \dots, \phi_{c(M-c)}^2)$ and $\hat{\boldsymbol{y}}_r = \text{vec}(\boldsymbol{Y}_r - \boldsymbol{X}_r\boldsymbol{B}_r) - (\boldsymbol{\alpha}_r \otimes \boldsymbol{Y}_r)\boldsymbol{s}$. Furthermore, $\boldsymbol{X}_r$ is a $T_r \times M$ matrix with typical row $(\Delta\boldsymbol{y}_{t-1}', \dots, \Delta\boldsymbol{y}_{t-p}')$ and $\boldsymbol{s}$ is a cM-dimensional vector obtained by selecting and vectorizing the first c columns of a $M \times M$-dimensional identity matrix.

The regime-specific variance-covariance matrices are simulated from their inverted Wishart distributed conditional posterior distributions,

$$\boldsymbol{\Sigma}_r|\boldsymbol{\gamma}, \boldsymbol{a}_r, \boldsymbol{b}_r, \boldsymbol{\beta}, \mathcal{D} \sim \mathcal{IW}(\nu_r, \boldsymbol{\Gamma}_r), \tag{2.17}$$

with $\boldsymbol{\Gamma}_r = (\boldsymbol{Y}_r - \boldsymbol{Z}_r\boldsymbol{\theta}_r)'(\boldsymbol{Y}_r - \boldsymbol{Z}_r\boldsymbol{\theta}_r) + \boldsymbol{S}_r$ and $\boldsymbol{\theta}_r = (\boldsymbol{\alpha}_r, \boldsymbol{B}_{r1}, \dots, \boldsymbol{B}_{rp})'$. The posterior degrees of freedom are given by $\nu_r = T_r + v_r$.

For the shrinkage parameters associated with the NG prior we derive the full conditional posterior distributions by applying Bayes theorem,

$$p(\tau_{rj}^2|b_{rj},\lambda_{br}) \propto p(b_{rj}|\tau_{rj}^2) \times p(\varphi_{br},\varphi_{br}\lambda_{br}/2). \tag{2.18}$$

After straightforward algebra it is possible to show that $p(\tau_{rj}^2|b_{rj},\lambda_{br})$, for $r=1,\dots,R; j=1,\dots,K$, follows a generalized inverted Gaussian (GIG) distribution,

$$\tau_{rj}^2|b_{rj},\lambda_{br} \sim \mathcal{GIG}(\varphi_{br}-1/2, b_{rj}^2, \varphi_{br}\lambda_{br}). \tag{2.19}$$

Note that the hierarchical nature of the model implies that the conditional posterior of τ_{rj}^2 is independent of the data. In addition, the global shrinkage parameter λ_{br} introduces dependence between the elements in $\boldsymbol{b}_r$.

Similarly to the local scaling parameters of the VAR coefficients, the conditional posterior of ζ_{rj}^2 with $r=1,\dots,R; j=1,\dots,cM$, and ϕ_j^2 $(j=1,\dots,c(M-c))$ is,

$$\zeta_{rj}^2|a_{rj},\lambda_{ar} \sim \mathcal{GIG}(\varphi_a-1/2, a_{rj}^2, \varphi_a\lambda_{ar}), \tag{2.20}$$
$$\phi_j^2|\xi_j,\varphi_\xi,\lambda_\xi \sim \mathcal{GIG}(\varphi_\xi-1/2, \xi_j^2, \varphi_\xi\lambda_\xi). \tag{2.21}$$

We now turn to the full conditional posterior distributions of the global scaling parameters. Again, after applying Bayes theorem to the Gamma likelihood induced by the prior on the local scaling parameters and the Gamma prior on the global scaling parameters the resulting conditional posterior distributions for $\lambda_{ar},\lambda_{br}$ and λ_ξ are Gamma distributed,

$$\lambda_{ar}|\boldsymbol{\zeta}_r \sim \mathcal{G}(d_a+\varphi_{ar}cM, e_a+\frac{\varphi_{ar}}{2}\sum_{i=1}^{cM}\zeta_{ir}^2), \tag{2.22}$$

$$\lambda_{br}|\boldsymbol{\tau}_r \sim \mathcal{G}(d_b+\varphi_{br}K, e_b+\frac{\varphi_{br}}{2}\sum_{i=1}^{K}\tau_{ir}^2), \tag{2.23}$$

$$\lambda_\xi|\boldsymbol{\Omega} \sim \mathcal{G}(d_\xi+\varphi_\xi c(M-c), e_\xi+\frac{\varphi_\xi}{2}\sum_{i=1}^{c(M-c)}\phi_i^2). \tag{2.24}$$

Up to this point all relevant conditional posterior distributions possess a well-known form. This implies that to simulate from the relevant joint posterior it suffices to simulate from the conditionals described above.

Unfortunately, the conditional posterior distribution of φ_{ir}, for $i \in \{a,b\}$, and φ_ξ is of no well-known form. We thus follow Griffin and Brown (2010) and adopt a random walk Metropolis Hastings (RWMH) algorithm with proposal distribution given by

$$\varphi_{ir}^{(prop)} = \varphi_{ir}^{(J-1)}\exp(\mathfrak{s}_i z_i), \tag{2.25}$$
$$\varphi_\xi^{(prop)} = \varphi_\xi^{(J-1)}\exp(\mathfrak{s}_\xi z_\xi), \tag{2.26}$$

We let $\varphi_{ir}^{(prop)}$ and $\varphi_{\xi}^{(prop)}$ be proposed values of φ_{ir} and φ_{ξ}, $\varphi_{ir}^{(J-1)}$, $\varphi_{\xi}^{(J-1)}$ the last accepted draws, $\mathfrak{s}_i, \mathfrak{s}_{\xi}$ scaling parameters specified such that the acceptance rate is between 20 and 40 percent and z_i, z_{ξ} standard normally distributed white noise increments.

Similarly to φ_{ir}, the conditional posterior distribution of $\boldsymbol{\gamma}$ is of no well known form. Typically, sampling from the relevant conditional posterior of a threshold parameter is achieved by employing a RWMH step (Geweke and Terui, 1993; Chen and Lee, 1995; Alessandri and Mumtaz, 2017). However, the latent nature of our threshold variable and the rather high correlation between $\boldsymbol{\gamma}$ and $\boldsymbol{\xi}$ renders a standard RWMH step infeasible. Given the fact that the support of γ_j is a bounded conditional on γ_{j-1} and γ_{j+1} (except for γ_0 and γ_R), we can utilize the Griddy Gibbs sampler (see Ritter and Tanner, 1992) and approximate the true cumulative distribution function (CDF) of the full conditional posterior distribution by a piecewise linear function and then use the resulting approximation to obtain draws by applying inverse transform sampling.

The Griddy Gibbs sampler evaluates the conditional posterior $p(\gamma_j|\boldsymbol{\gamma}_{-j}, \boldsymbol{\Sigma}, \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{\beta}, \mathcal{D})$, with $\boldsymbol{\gamma}_{-j}$ being the vector $\boldsymbol{\gamma}$ with the jth element excluded, at a set of candidate points $\tilde{\gamma}_j^{(1)}, \dots, \tilde{\gamma}_j^{(Q)}$ and computes

$$\mathfrak{w}_{ij} = \frac{p(\tilde{\gamma}_j^{(i)}|\boldsymbol{\gamma}_{-j}, \boldsymbol{\Sigma}, \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{\beta}, \mathcal{D})}{\sum_{i=1}^{Q} p(\tilde{\gamma}_j^{(i)}|\boldsymbol{\gamma}_{-j}, \boldsymbol{\Sigma}, \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{\beta}, \mathcal{D})} \quad \text{for } i = 1, \dots, Q. \tag{2.27}$$

These weights are then used to perform inverse transform sampling to obtain draws from $p(\gamma_j|\boldsymbol{\gamma}_{-j}, \boldsymbol{\Sigma}, \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{\beta}, \mathcal{D})$. In our empirical applications we set $Q = 50$ and apply linear interpolation to obtain a sequence of weights that accurately approximate the actual CDF. Note that the likelihood function of the model is flat with respect to specific values of γ_j if no change in the corresponding regime allocation is induced. Finally, it is worth emphasizing that we also introduce the restriction that a certain fractions of observations (unless otherwise noted five percent of T) have to remain within each regime.

2.4 Full conditional posterior simulation

Posterior inference is carried out by iteratively drawing from the relevant conditional posterior distributions outlined above and discarding the first set of draws as burn-in. More precisely our MCMC algorithm cycles between the following steps:

Regime-specific coefficients for $r = 1, \dots, R$

1. Obtain draws of $(\boldsymbol{a}_r', \boldsymbol{b}_r')'$ from $\mathcal{N}(\boldsymbol{\mu}_r, \boldsymbol{V}_r)$
2. Draw $\boldsymbol{\Sigma}_r$ from $\mathcal{IW}(\nu_r, \boldsymbol{\Gamma}_r)$
3. For $j = 1, \dots, K$ sample τ_{rj}^2 from $\mathcal{GIG}(\varphi_{br} - 1/2, b_{rj}^2, \varphi_{br}\lambda_{br})$
4. For $j = 1, \dots, cM$ draw ζ_{rj}^2 from $\mathcal{GIG}(\varphi_{ar} - 1/2, a_{rj}^2, \varphi_{ar}\lambda_{ar})$
5. Simulate λ_{br} from $\mathcal{G}(d_b + \varphi_{br}K, e_b + \frac{\varphi_{br}}{2}\sum_{i=1}^{K}\tau_{ir}^2)$
6. Sample λ_{ar} from $\mathcal{G}(d_a + \varphi_{ar}cM, e_a + \frac{\varphi_{ar}}{2}\sum_{i=1}^{cM}\zeta_{ir}^2)$
7. Draw φ_{ir} for $i \in \{b, a\}$ using the RWMH algorithm

Regime-independent coefficients

8. Simulate φ_ξ using the RWMH algorithm
9. Draw $\boldsymbol{\xi}$ from $\mathcal{N}(\boldsymbol{\delta}, \boldsymbol{\Lambda})$
10. Obtain draws of γ_j using the Griddy Gibbs sampler
11. For $j = 1, \dots, c(M - c)$ draw ϕ_j^2 from $\mathcal{GIG}(\varphi_\xi - 1/2, \xi_j^2, \varphi_\xi\lambda_\xi)$

The proposed algorithm scales quite well in high dimensions with the main computational hurdle stemming from Eq. (2.11), which involves inverting a $K \times K$ variance-covariance matrix per regime and iteration. This step, however, may be speed up considerably by resorting to recent advances in the estimation of high dimensional VAR models (Carriero et al., 2015; Kastner and Huber, 2017).

3 Empirical application: Threshold cointegration in international exchange rates

In this section we take on an empirically relevant question, namely whether a given currency is under/overvalued relative to the US dollar. This is achieved by adopting the modeling approach outlined in the previous section. After outlining the theoretical framework that drives our choice of covariates in the next section we conduct an extensive forecasting exercise where we evaluate the merits of our approach relative to a set of linear models. Moreover, we select the number of cointegrating relations using the predictive likelihood. Finally, the posterior distribution of the regime allocation and the cointegrating error is discussed.

3.1 Theoretical framework, model specification and data overview

Our possible set of endogenous variables included in $\{\boldsymbol{y}_t\}$ is given by

$$\boldsymbol{y}_t = (e_t, i_t, i_t^*, p_t, p_t^*, y_t, y_t^*, m_t, m_t^*)', \tag{3.1}$$

where e_t denotes the logarithm of the nominal exchange rate in terms of the number of foreign currency per unit of domestic currency (the US dollar in what follows), i_t is the short-term interest rate (approximated through the three-months money market rates), p_t is the consumer price index, y_t measures output by including data on industrial production and m_t denotes the money supply measured through the M3 money stock for each country. Asterisks indicate foreign counterparts of the aforementioned variables. Our monthly sample starts in $t_0 = 1983 : M01$ and ends in $T = 2014 : M12$.

This choice of variables is motivated by combining information sets from two successfully used exchange rate models, namely a Taylor rule-based exchange rate model Molodtsova et al. (2008); Molodtsova and Papell (2009); Molodtsova et al. (2011); Huber (2017) and the long-run monetary model (Mark and Sul, 2001; Rapach and Wohar, 2002; 2004). More importantly, however, is the fact that we rely on the monetary model to identify our TVECM and, in addition, specify the prior on the first column of $\boldsymbol{\xi}$.

Specifically, in the case that $c = 1$ the cointegrating error is given by

$$\boldsymbol{\beta}'\boldsymbol{y}_t = e_t - [\xi_2 i_t + \xi_3 i_t^* + \xi_4 p_t + \xi_5 p_t^* + \xi_6 y_t + \xi_7 y_t^* + \xi_8 m_t + \xi_9 m_t^*]. \tag{3.2}$$

From Eq. (3.2) we directly observe that the typical identification scheme described in Section 2 can be applied in light of the ordering given in Eq. (3.1) and still be consistent with the long-run monetary model (that sets $\xi_2 = \xi_3 = \xi_4 = \xi_5 = 0$) as long as $c \leq 5$.

We softly introduce the restrictions introduced by the monetary model by specifying a normally distributed prior on the first column of $\boldsymbol{\xi}$ with prior mean centered on the coefficients suggested by the long-run monetary model. This implies that the right-hand side of Eq. (3.2) reduces to

$$e_t - [(y_t - y_t^*) + (-m_t + m_t^*)]. \tag{3.3}$$

For the variances we specify the NG prior described in Section 2.2. One implication is that we assess in a data driven way whether we should force the elements of the first column of $\boldsymbol{\xi}$ towards the implied coefficients of the long-run monetary model. In the case that $c = 5$ our identification strategy can be interpreted as a dogmatic prior that imposes the restriction that uncovered interest and purchasing power parity hold.

One additional important choice is the number of regimes R. We assume that three regimes sufficiently control for non-linearities in the behavior of the exchange rate. Since our threshold variable is the deviation of the log exchange rate from its long-run fundamental value, three regimes capture the notion that if deviations from the underlying fundamental model become large, a given currency is either over or undervalued. In addition, the range between both thresholds defines a certain range of inaction, where economic agents perceive the deviation as being too small to actively engage in foreign exchange trading.

The model specification is completed by specifying the lag length and the prior hyperparameters. The remaining priors are specified as described in Section 2.2. In addition, we set $p = 1$ which implies that the underlying threshold VAR specification features two lags. This choice is based on the fact that for $p = 13$ computation becomes excessively slow and we run into severe issues associated with regimes that feature only a small number of observations. In addition, we have some evidence that our shrinkage prior strongly pushes coefficients associated with higher lag orders towards a zero matrix.

3.2 Model validation: Out of sample forecasting performance

We assess the merits of our modeling approach by evaluating the forecasting performance in terms of one-step-ahead predictive likelihoods (LPS) computed over a hold-out sample that ranges from 1998:M06 to 2014:M12 (200 monthly observations). As competing models we include a VAR estimated in first differences (labeled VAR (Δ)), a linear VECM with $r = 5$ for all countries[2] and a VAR in levels. All models feature the shrinkage prior outlined in Section 2. These models enable us to assess what features of our proposed model improve exchange rate predictability.

We use an expanding forecasting window that uses the first 181 observations as an training sample to compute the one-step-ahead predictive density for the first period in the hold-out sample. In the next step we expand the initial forecasting window by a single observation and repeat this procedure until we reach the end of the sample. To assess forecast accuracy we rely on the one-step-ahead predictive likelihood that can be interpreted as a training sample marginal likelihood and thus provides a natural measure of model adequacy Geweke and Amisano (2010).

Table 1 displays the sum of marginal log predictive scores for the five exchange rate pairs considered. Before we proceed to comparing our proposed specification to different competing models, we select the cointegration rank by assessing the marginal log predictive likelihood of a given model conditional on the rank $c \in \{1, 2, 3, 4, 5\}$. Since we are interested exclusively in the predictive accuracy of our

[2]Our results for the non-linear model suggests that the presence of our shrinkage prior is able to alleviate issues associated with the increased number of parameters that comes from choosing r to be too large and to some extent allows for stochastic model specification.

	DE	CA	JP	SE	UK
$c = 1$	417.32	487.21	451.39	442.34	490.99
$c = 2$	415.56	500.22	449.77	447.00	**494.10**
$c = 3$	414.78	488.25	446.40	448.72	489.20
$c = 4$	417.80	**502.48**	449.70	448.72	488.67
$c = 5$	**417.88**	491.50	451.03	446.30	489.13
VAR (Δ)	415.28	476.73	451.19	450.51	489.28
VECM	413.66	474.64	**452.03**	449.74	488.82
VAR	414.04	472.71	450.89	**451.01**	490.19

Notes: The table presents the cumulative sum of log predictive likelihoods (LPS) over the hold-out period. $c = 1$ to $c = 5$ refers to the cointegration rank of the TVECM, VAR (Δ) is a VAR in first differences, VECM a linear VECM, and VAR denotes a VAR estimated in levels.

Table 1: Sum of marginal log predictive likelihoods over the hold-out period: 1998:M06 to 2014:M12.

model in terms of exchange rate prediction, this seems to be a natural choice to discriminate between differing cointegration ranks.

Notice that we select $c = 5$ for Germany. In this case, accuracy differences between allowing for a single cointegrating relation and four or five appear to be rather small. The differences across ranks suggests almost no predictive differences, suggesting that we could simply use the simplest specification (i.e. $c = 1$). However, our proposed method also permits reliable estimation of larger models by relying on a suitable set of shrinkage priors that also enable, to some extent, model selection.

Turning to the results for Canada reveals that the optimal number of c is four. Here we see larger differences between the competing numbers of cointegrating relations, with the simplest specification (i.e. $c = 1$) displaying the weakest performance. This could be traced back to the fact that the remaining cointegration errors embody important information for exchange rate movements which are neglected by relying on a single cointegration relationship. Put differently, it could also be the case that other cointegration relations appear to be important to model the dynamics of the remaining $M - 1$ variables in the system. For Japan and the United Kingdom we select $c = 1$ and $c = 2$, respectively. In what follows, we label the specification that yields the highest LPS as TVECM(r^*).

Comparing the predictive performance of the TVECM(r^*) with a VAR in first differences reveals that for Germany and Canada we find that allowing for both, cointegration and non-linearities, improves predictions. Especially for Canada, the improvement is particularly pronounced. To quantitatively assess the forecast gain obtained by introducing a non-linear model specification we compare the LPS between the TVECM(r^*) and the VECM. Note that in this case, the difference is even more pronounced as in the case of a VAR in first differences for Germany and Canada.

This finding, however, does not carry over to Japan. There we find that the linear VECM outperforms all competing model specifications slightly. We conjecture that this finding is linked with the estimated regime allocation (see Section 3.3) as well as the generally higher number of parameters we have to infer from the data. Moreover, Japan is the only country in our sample where short-term interest rates have been constrained by the zero lower bound (ZLB) for an extended period of time. While this holds also true for some of the remaining currencies we would like to stress that interest rates reached the ZLB already in the early 2000s in Japan whereas for the remaining countries the ZLB is reached only after the great financial crisis (i.e. after 2008).

For Sweden, we find that a VAR in levels also outperforms our proposed model specification. Sims et al. (1990) show that a VAR in levels, which proves to be a rather parsimonious specification relative to our TVECM(r^*), is able to control for the presence of cointegrating relations while at the same time it does not suffer from issues associated with selecting an appropriate rank of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$.

Finally, considering the results for the United Kingdom reveals that the TVECM(r^*) outperforms all competing models over the hold-out period. The VAR estimated in levels ranks second while the VAR in first differences ranks third. We again find that having a VECM does not pay off in terms of predictive capabilities, at least in the short-run.

Since the discussion above is solely related to the sum of the LPS over the hold-out period it proves to be interesting to investigate whether a given model tends to work well during certain points in time. A typical result in the forecasting literature (Clark, 2011; Groen et al., 2013; Clark and Ravazzolo, 2015; Byrne et al., 2016; Huber and Feldkircher, 2016) is that especially during economic crises it pays off to allow for non-linearities in the model to capture sudden shifts in macroeconomic volatility. Our TVECM(r^*) is able to control for this since error variances are allowed to change between regimes. Moreover, if the underlying structural relationships depend on whether a given currency is under/overvalued we might gain an important advantage relative to a linear specification in forecasting terms.

[Fig. 1 about here.]

Fig. 1 depicts the evolution of the log predictive Bayes factor relative to the VAR(Δ) over time. A few findings are worth emphasizing. First, note that for some currencies the predictive performance of the non-linear specifications improves sharply during the recent global financial crisis. This holds true for Germany, Sweden and Canada. For Germany and Canada, we observe that non-linear models also outperform the linear competitors prior to the financial crisis. For instance, we improve upon linear models in terms of predicting the DM/USD exchange rate[3] up to the first half of the 2000s. For the Canadian dollar a similar picture arises, with the predictive Bayes factor turning positive for the first time in 2003 for the best performing TVECM.

Second, for the British pound we observe consistent improvements relative to the VAR(Δ). To some extent, this finding also carries over to the DM/USD with the main exception that predictive accuracy further increases during the financial crisis. For Sweden, the forecasting performance of our model is consistently worse than the forecasting performance of the linear VAR model in differences, except during the crisis period. This finding again highlights the additional forecasting accuracy premia obtained by estimating a non-linear model specification that permits time variation in the underlying autoregressive parameters and the error variances.

Finally, the results for Japan in Fig. 1 (e) indicate that our non-linear model worked well up to the beginning of the financial crisis. During that period, however, we observe a decline in predictive accuracy that might be caused by the regime allocation induced by our model. We conjecture that this can be traced back to the economic interpretation behind the different regimes and the specific nature of the Yen. Since the Yen also serves as a funding currency for carry trades and thus tends to appreciate in times of increased economic uncertainty, the prevailing regime might signal a rather fair valuation relative to the US dollar which might be characterized by a generally low level of macroeconomic volatility. However, this behavior of our model might thus turn out to be detrimental for predictions.

Summing up our results, we find that allowing for non-linearities in our model improves exchange rate forecasts for three out of five currencies considered. The two exceptions, Japan and Sweden, prove to be countries that differ strongly in terms of their macroeconomic fundamentals when compared to the remaining economies in our panel. This could point towards specification issues associated with the underlying structural model in the previous section. Thus, to improve predictive accuracy for the remaining two currency pairs we might have to resort to other structural models or perform forecast combination to improve upon the existing results.

3.3 Regime allocation and equilibrium exchange rates

After providing some evidence that our models tend to fit the data well and provide more accurate predictions as simpler benchmark models we now turn to describing some of the features of our flexible econometric specification.

To provide a comprehensive picture of the estimated regime allocation, Fig. 2 shows the regime probabilities obtained by integrating out the remaining model parameters. Our threshold specification implies that, conditional on β and γ, movements between regimes happen in a deterministic manner. However, after integrat-

[3] After the introduction of the Euro, this series is constructed by using the EUR/USD exchange rate from 1999 onwards.

ing out $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ the underlying regime indicators follow a Bernoulli distribution. This enables us to infer the probability of a given currency being over/undervalued for a given point in time.

[Fig. 2 about here.]

Looking at Fig. 2 yields several interesting insights. First, note that the upper regime can be interpreted as being closely related to an undervaluation of a given currency vis-a-vis the US dollar. Put differently, if the posterior distribution of w_{1t-1} is located above the upper threshold we can speak of an overvaluation of the US dollar relative to the currency of interest. During the beginning of the 1980's we observe that for the United Kingdom, Sweden, Canada, and Japan our model points towards an overvaluation of the US dollar to the respective currencies. This, to some extent, coincides with the dollar bubble that was driven by the restrictive monetary policy conducted under Fed Chairman Paul Volcker and a general de-regulation of the US banking system.

Second, from the late 1980s to the early 2000s our model generally predicts a rather fair valuation of the US dollar for most currencies. While we tend to observe transient shifts in the underlying regime probabilities especially for European currencies during the crisis of the European Exchange Rate Mechanisms in 1992/1993, the remainder of this period was characterized by rather small deviations of the exchange rate from the model implied long-run equilibrium value.

Finally, in the run-up to the recent global financial crisis starting in late 2008, our model points towards an overvaluation of most currencies under scrutiny relative to the dollar. The bankruptcy of Lehman Brothers, however, led to a sharp appreciation of the US dollar. This was caused by a major shift in financial market participants' risk preferences. The pronounced increase in uncertainty led to portfolio rebalancing associated with carry trading strategies that seek to exploit interest rate differentials between low and high-yielding currencies. If uncertainty increases, the expected return on such strategies falls, rendering traditional carry trading strategies unfeasible.

4 Conclusive remarks

In this paper we propose a flexible econometric approach that permits estimation of moderately to large-sized threshold vector error correction models (TVECM). Typical issues encountered in the literature are solved by relying on a Markov chain Monte Carlo (MCMC) algorithm that avoids optimization of an ill-behaved likelihood function in the presence of a threshold process that drives the transition between regimes. One important feature of our model is that the proposed threshold variable is a latent quantity and depends on the cointegration vector. From a frequentist perspective, this model is typically estimated by relying on a grid search or by resorting to large-sample approximation. This quickly turns computationally prohibitive if the number of included variables or the cointegrating rank becomes large. One additional issue is related to overparameterization if the number of endogenous quantities is moderate to large. We solve this issue through Bayesian shrinkage priors that enable us to restrict the parameter space and shrink the corresponding posterior distribution towards a simpler prior model.

From an empirical perspective, we illustrate the merits of our approach using a dataset consisting of five currencies vis-á-vis the US dollar. In a forecasting application we highlight that allowing for threshold cointegration pays off in terms of predictive performance, especially during turbulent times such as the recent global financial crisis. Moreover, our flexible framework enables us to investigate whether a given currency is under or overvalued against the US dollar. Looking at the regime probabilities suggests that most currencies have been undervalued against the dollar during the first half of the 1980s and have been rather fair valued during the 1990s, a trend which quickly reversed after the Lehman event. In the run-up to the global financial crisis, however, most currencies have been overvalued.

We would like to stress that our econometric framework enables researchers to comprehensively investigate issues commonly encountered in macroeconomics and finance. From an econometric point of view an interesting avenue of further research could be a combination between the conditionally deterministic regime-switching behavior of our model with Markov switching models in the spirit of Kaufmann (2015).

References

Alessandri P and Mumtaz H (2017) Financial conditions and density forecasts for US output and inflation. *Review of Economic Dynamics* 24, 66–78

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

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

Bańbura M, Giannone D and Reichlin L (2010) Large Bayesian vector auto regressions. *Journal of Applied Econometrics* 25(1), 71–92

Byrne JP, Korobilis D and Ribeiro PJ (2016) Exchange rate predictability in a changing world. *Journal of International Money and Finance* 62, 1–24

Carriero A, Clark TE and Marcellino M (2015) Large vector autoregressions with asymmetric priors. Technical report, Working Paper, School of Economics and Finance, Queen Mary University of London

Chen CW and Lee JC (1995) Bayesian inference of threshold autoregressive models. *Journal of Time Series Analysis* 16(5), 483–492

Clark TE (2011) Real-time density forecasts from Bayesian vector autoregressions with stochastic volatility. *Journal of Business & Economic Statistics* 29(3), 327–341

Clark TE and Ravazzolo F (2015) Macroeconomic Forecasting Performance under Alternative Specifications of Time-Varying Volatility. *Journal of Applied Econometrics* 30(4), 551–575

Doan T, Litterman R and Sims C (1984) Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews* 3(1), 1–100

Forbes CS, Kalb GR and Kofhian P (1999) Bayesian arbitrage threshold analysis. *Journal of Business & Economic Statistics* 17(3), 364–372

Gefang D (2012) Money-output Causality Revisited–A Bayesian Logistic Smooth Transition VECM Perspective. *Oxford Bulletin of Economics and Statistics* 74(1), 131–151

Geweke J (1996) Bayesian reduced rank regression in econometrics. *Journal of Econometrics* 75(1), 121–146

Geweke J and Amisano G (2010) Comparing and evaluating Bayesian predictive distributions of asset returns. *International Journal of Forecasting* 26(2), 216–230

Geweke J and Terui N (1993) Bayesian threshold autoregressive models for nonlinear time series. *Journal of Time Series Analysis* 14(5), 441–454

Giannone D, Lenza M and Primiceri GE (2015) Prior selection for vector autoregressions. *Review of Economics and Statistics* 97(2), 436–451

Griffin JE and Brown PJ (2010) Inference with normal-gamma prior distributions in regression problems. *Bayesian Analysis* 5(1), 171–188

Groen JJ, Paap R and Ravazzolo F (2013) Real-time inflation forecasting in a changing world. *Journal of Business & Economic Statistics* 31(1), 29–44

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

Huber F (2016) Forecasting exchange rates using multivariate threshold models. *The BE Journal of Macroeconomics* 16(1), 193–210

Huber F (2017) Structural breaks in Taylor rule based exchange rate models - Evidence from threshold time varying parameter models. *Economics Letters* 150, 48–52

Huber F and Feldkircher M (2016) Adaptive Shrinkage in Bayesian vector autoregressive models. *Journal of Business & Economic Statistics* , forthcoming

Jochmann M and Koop G (2015) Regime-switching cointegration. *Studies in Nonlinear Dynamics & Econometrics* 19(1), 35–48

Kastner G and Huber F (2017) Sparse Bayesian vector autoregressions in huge dimensions. *arXiv preprint arXiv:1704.03239*

Kaufmann S (2015) K-state switching models with time-varying transition distributionsDoes loan growth signal stronger effects of variables on inflation? *Journal of Econometrics* 187(1), 82–94

Kleibergen F and Van Dijk HK (1994) On the shape of the likelihood/posterior in cointegration models. *Econometric Theory* 10(3-4), 514–551

Koop G, León-González R and Strachan RW (2009) Efficient posterior simulation for cointegrated models with priors on the cointegration space. *Econometric Reviews* 29(2), 224–242

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

Mark NC (1995) Exchange rates and fundamentals: Evidence on long-horizon predictability. *The American Economic Review* , 201–218

Mark NC and Sul D (2001) Nominal exchange rates and monetary fundamentals: evidence from a small post-Bretton Woods panel. *Journal of International Economics* 53(1), 29–52

Martens M, Kofman P and Vorst TC (1998) A threshold error-correction model for intraday futures and index returns. *Journal of Applied Econometrics* , 245–263

Molodtsova T, Nikolsko-Rzhevskyy A and Papell DH (2008) Taylor rules with real-time data: A tale of two countries and one exchange rate. *Journal of Monetary Economics* 55, S63–S79

Molodtsova T, Nikolsko-Rzhevskyy A and Papell DH (2011) Taylor rules and the euro. *Journal of Money, Credit and Banking* 43(2-3), 535–552

Molodtsova T and Papell DH (2009) Out-of-sample exchange rate predictability with Taylor rule fundamentals. *Journal of International Economics* 77(2), 167–180

Polson NG and Scott JG (2010) Shrink globally, act locally: Sparse Bayesian regularization and prediction. *Bayesian Statistics* 9, 501–538

Rapach DE and Wohar ME (2002) Testing the monetary model of exchange rate determination: new evidence from a century of data. *Journal of International Economics* 58(2), 359–385

Rapach DE and Wohar ME (2004) Testing the monetary model of exchange rate determination: a closer look at panels. *Journal of International Money and Finance* 23(6), 867–895

Ritter C and Tanner MA (1992) Facilitating the Gibbs sampler: the Gibbs stopper and the griddy-Gibbs sampler. *Journal of the American Statistical Association* 87(419), 861–868

Sims CA, Stock JH and Watson MW (1990) Inference in linear time series models with some unit roots. *Econometrica: Journal of the Econometric Society* , 113–144

Sims CA and Uhlig H (1991) Understanding unit rooters: A helicopter tour. *Econometrica: Journal of the Econometric Society* , 1591–1599

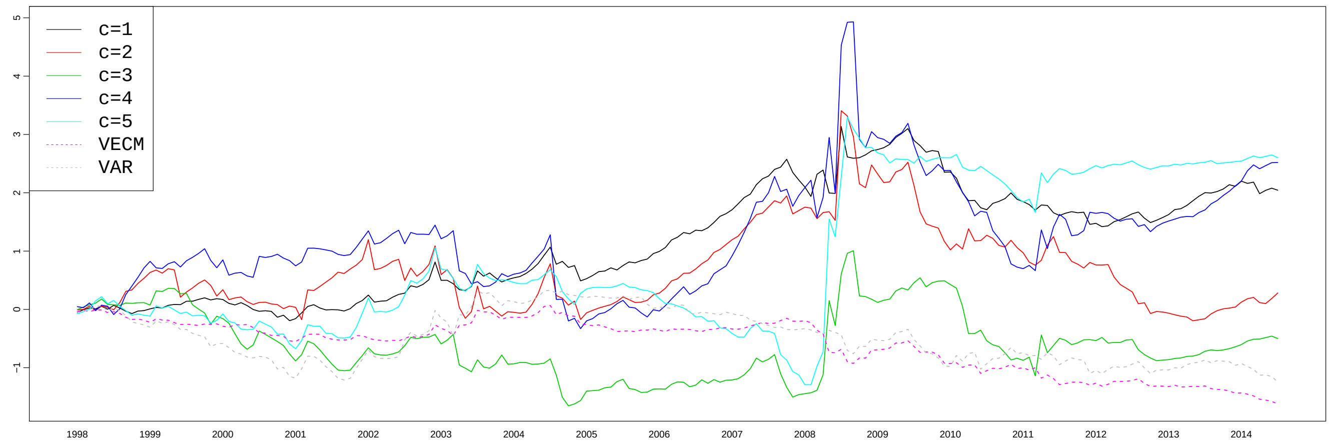
Sims CA and Zha T (1998) Bayesian methods for dynamic multivariate models. *International Economic Review* , 949–968

Strachan RW (2003) Valid Bayesian estimation of the cointegrating error correction model. *Journal of Business & Economic Statistics* 21(1), 185–195

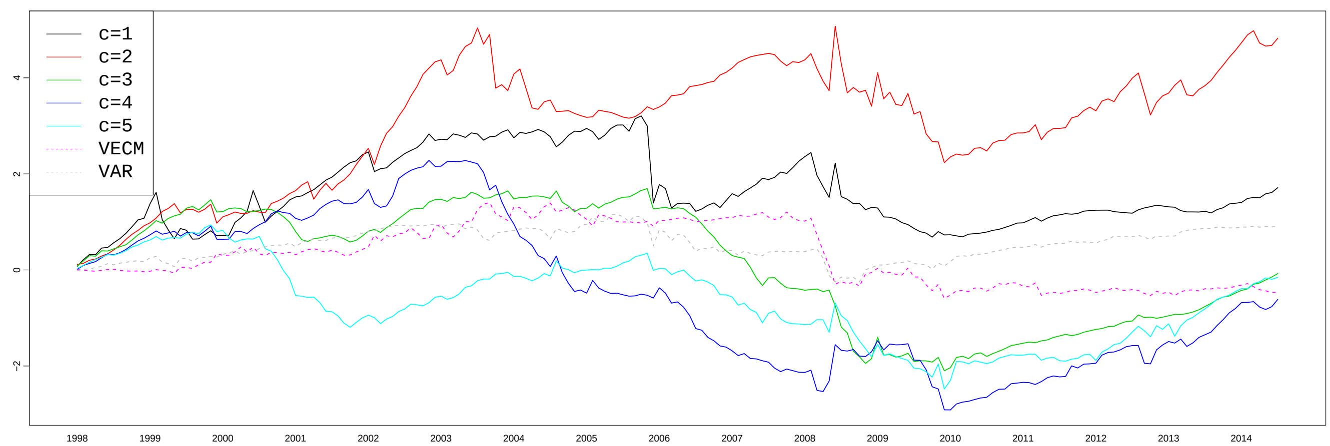
Villani M (2001) Bayesian prediction with cointegrated vector autoregressions. *International Journal of Forecasting* 17(4), 585–605

Zellner A (1986) *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti*, chapter On Assessing Prior Distributions and Bayesian Regression Analysis with g-Prior Distributions. North-Holland: Amsterdam

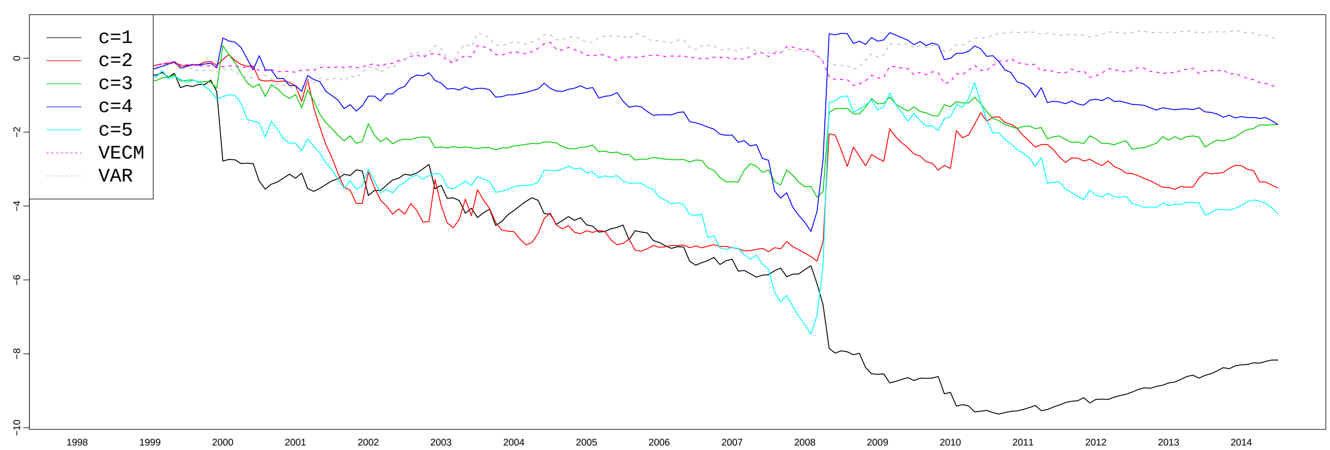
(a) Germany



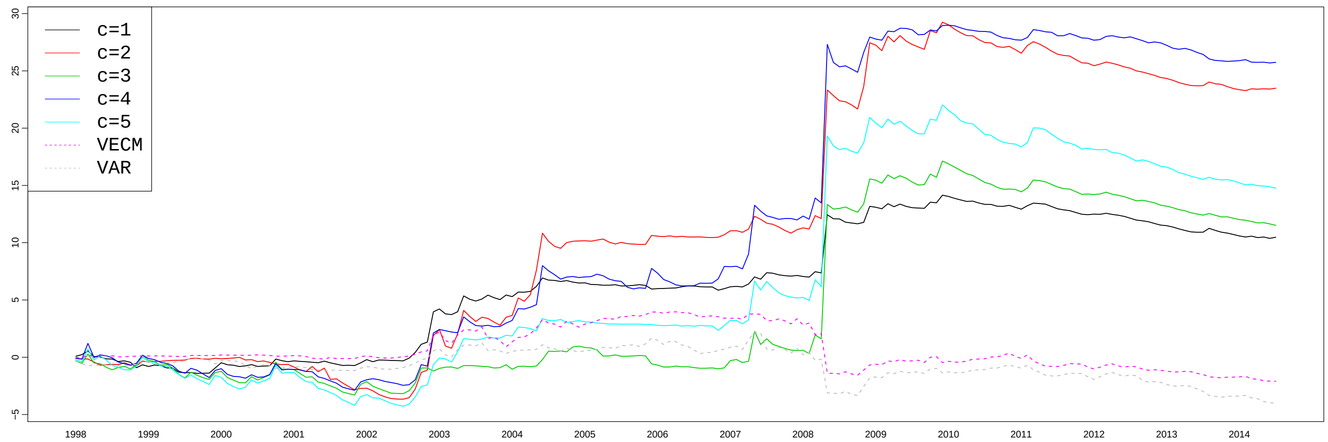
(b) United Kingdom



(c) Sweden



(d) Canada



(e) Japan

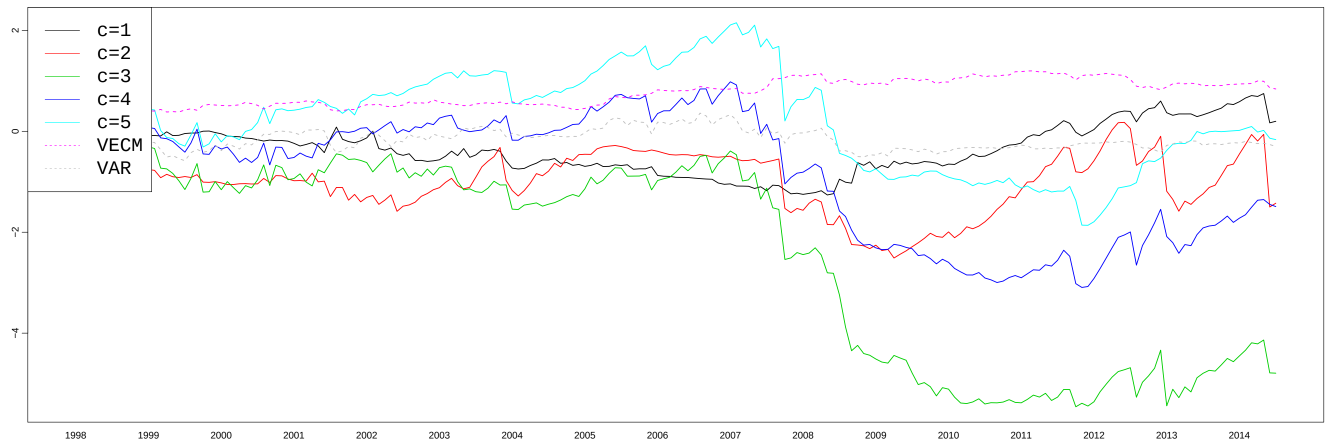
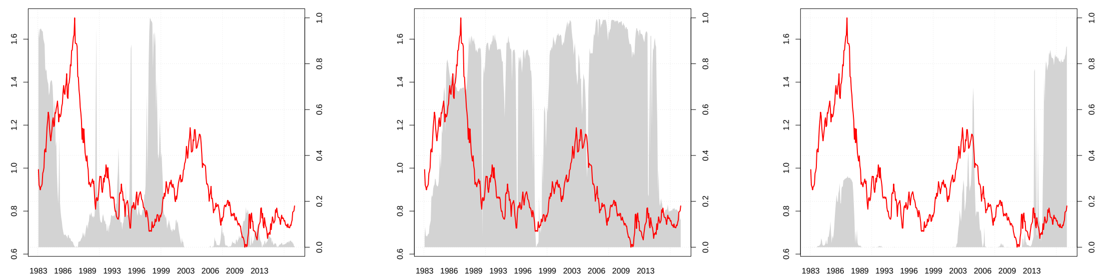


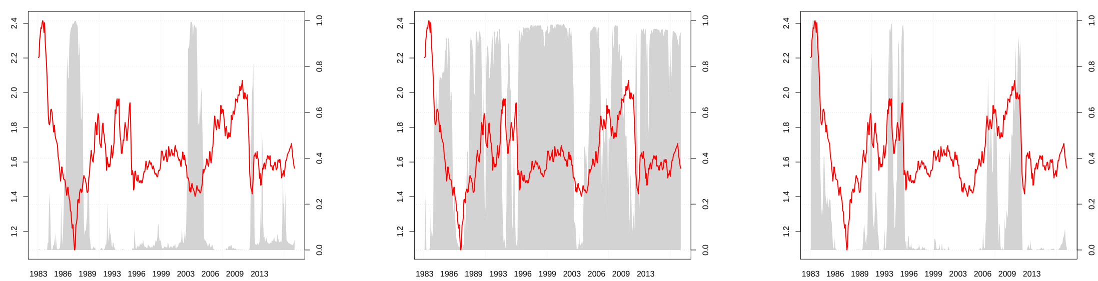
Fig. 1: Evolution of log predictive Bayes factors over the hold-out period: 1998:M06 to 2014:M12.

Germany

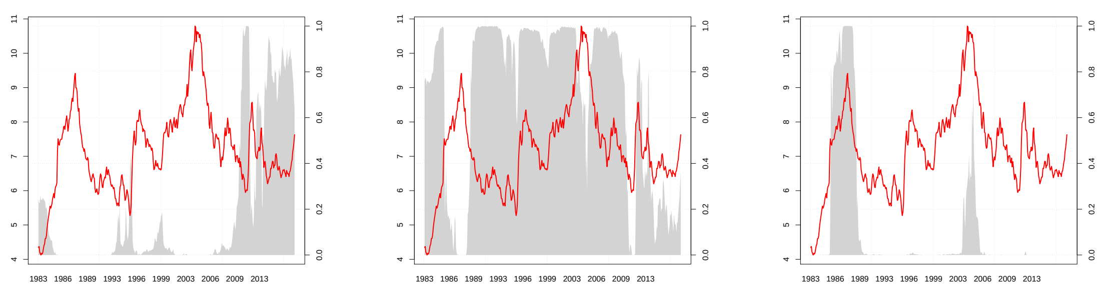
(a) Lower regime **(b)** Middle regime **(c)** Upper regime



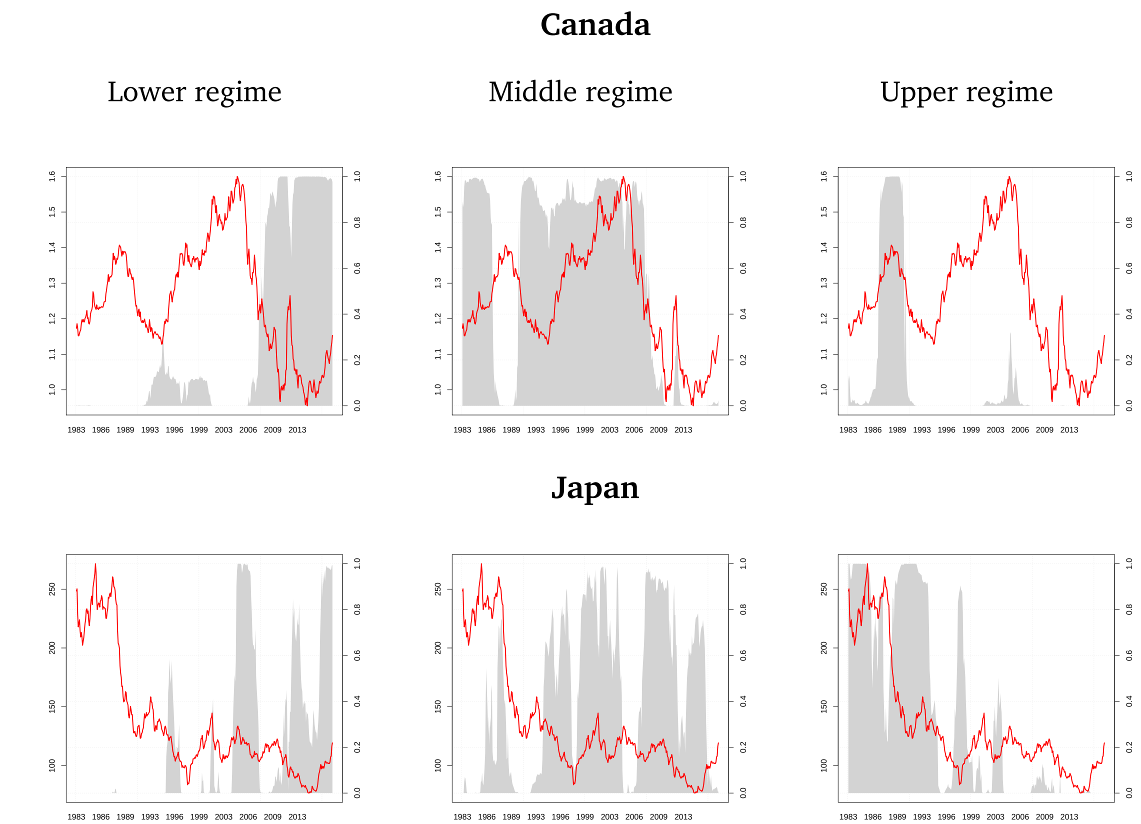
United Kingdom



Sweden



Notes: Posterior mean of the regime probabilities (shaded gray areas) along with the actual exchange rate (in red). Results are based on 35,000 posterior draws.



Notes: Posterior mean of the regime probabilities (shaded gray areas) along with the actual exchange rate (in red). Results are based on 35,000 posterior draws.

Fig. 2: Posterior probabilities of being within a given regime over time