Stochastic model specification in Markov switching vector error correction models

NIKO HAUZENBERGER[1,2], FLORIAN HUBER[1], MICHAEL PFARRHOFER[1,2]
and THOMAS O. ZÖRNER[2*]

[1]*Paris Lodron University of Salzburg*
[2]*Vienna University of Economics and Business*

This paper proposes a hierarchical modeling approach to perform stochastic model specification in Markov switching vector error correction models. We assume that a common distribution gives rise to the regime-specific regression coefficients. The mean as well as the variances of this distribution are treated as fully stochastic and suitable shrinkage priors are used. These shrinkage priors enable to assess which coefficients differ across regimes in a flexible manner. In the case of similar coefficients, our model pushes the respective regions of the parameter space towards the common distribution. This allows for selecting a parsimonious model while still maintaining sufficient flexibility to control for sudden shifts in the parameters, if necessary. We apply our modeling approach to real-time Euro area data and assume transition probabilities between expansionary and recessionary regimes to be driven by the cointegration errors. The results suggest that the regime allocation is governed by a subset of short-run adjustment coefficients and regime-specific variance-covariance matrices. These findings are complemented by an out-of-sample forecast exercise, illustrating the advantages of the model for predicting Euro area inflation in real time.

JEL: C11, C32, E31, E32, E44
KEYWORDS: Nonlinear vector error correction model, hierarchical modeling, inflation forecasting, Euro area

1. INTRODUCTION

In this paper we propose a hierarchical multivariate regime switching model that allows for detecting which parameters differ across regimes. The model, a Markov switching vector error correction model (MS-VECM), explicitly discriminates between short- and long-run dynamics and potentially allows for time-varying transition probabilities that depend on the cointegration errors. We assess whether parameters differ across regimes by using novel shrinkage priors that have recently been utilized in finite mixture modeling (see Yau and Holmes, 2011; Malsiner-Walli *et al.*, 2016).

The literature on Bayesian estimation of Markov switching (MS) models is voluminous (see, among many others, Chib, 1996; Kim and Nelson, 1998; Kaufmann, 2000; Sims *et al.*, 2008; Kaufmann, 2015; Droumaguet *et al.*, 2017; Bognanni and Herbst, 2018). By contrast, contributions that explicitly deal with the Bayesian estimation of MS-VECMs are comparatively sparse (Martin, 2000; Paap and Van Dijk, 2003; Jochmann and Koop, 2015). Most of these contributions use Bayesian shrinkage priors to enable reliable and efficient estimation. These priors are typically specified in the spirit of standard Minnesota priors (Doan *et al.*, 1984; Sims and Zha, 1998a) and symmetric across regimes. Here, one implication is that coefficients are pushed towards a stylized prior model (like a multivariate random walk), irrespective of the regime and what the (regime-specific) likelihood suggests. For instance, a regression parameter may be zero in one regime, but different in another. Such a situation is effectively ruled out as both coefficients are pushed towards zero.

In this paper, we circumvent such issues by proposing a novel hierarchical modeling approach that has originally been proposed in the literature on finite mixture models (see Yau and Holmes, 2011; Malsiner-Walli *et al.*, 2016). We estimate an MS-VECM assuming that the regime-specific coefficients arise from a common distribution. The mean of this common distribution is treated as unknown and estimated from the data. Using Normal-Gamma shrinkage priors (Griffin and Brown, 2010) on the variance-covariance matrix of the common distribution enables us to gain an understanding on what covariates drive the corresponding regime allocation. When compared to the existing literature, our approach allows for flexibly testing which coefficients (or sets of them) should differ across regimes while pushing similar coefficients towards a common mean. In addition, we follow the literature on MS models with time-varying transition probabilities (see, among others, Filardo, 1994; Kim and Nelson, 1998; Kaufmann, 2015) and assume that the transition probability matrix is time-varying and depends on the lagged cointegration errors. This setup implies that we test whether deviations of macroeconomic fundamentals from their long-run equilibrium values impact the transition probabilities between business cycle stages.

For an empirical illustration, we estimate a medium-scale Euro area business cycle model that discriminates between business cycle expansions and contractions based on real-time data (Giannone *et al.*, 2012). The results suggest that our model successfully replicates Euro area business cycle behavior and performs well in forecasting inflation relative to a set of competing specifications. We find that deviations of inflation, unemployment and expected inflation (measured by the 2-year inflation projections of the ECBs survey of professional forecasters) from their long-run fundamentals have predictive power for the transition probabilities. When considering differences in parameters across regimes, we find that some short-run adjustment coefficients differ markedly across regimes. Moreover, a subset of autoregressive coefficients displays variation across regimes, while others exhibit no structural breaks. Besides allowing for time variation in the autoregressive coefficients, it appears crucial to consider heteroscedastic shocks by allowing for a regime-specific variance-covariance matrix.

The remainder of the paper is organized as follows. Section 2 outlines the proposed idea by means of a simple switching regression model while Section 3 presents the MS-VECM model as well as the prior setup. Section 4 first gives an overview of the dataset used and subsequently shows the empirical results and the forecasting performance over time. The last section summarizes and concludes the paper.

2. A SIMPLE HIERARCHICAL MODEL

In this section we outline the main idea on how to determine whether coefficients differ across regimes within a simple regression model. Subsequently, we generalize the stylized example to a multivariate nonlinear error correction model.

We set the stage by assuming that a scalar time series $\{y\}_{t=1}^{T}$ follows a switching regression model (Goldfeld and Quandt, 1973),[1]

$$y_t = \beta_{1S_t} x_{1t} + \beta_{2S_t} x_{2t} + \sigma_{S_t} \eta_t, \quad \eta_t \sim \mathcal{N}(0, 1), \tag{1}$$

with x_{jt} being exogenous covariates and β_{jS_t} (for $j = 1, 2$) the associated regression coefficients while $\sigma_{S_t}^2$ denotes the error variance. We assume that S_t represents a latent quantity that takes values $0, \ldots, R$ and follows a first-order Markov process.[2]

The Bayesian literature typically uses Gaussian priors on β_{jS_t},

$$\beta_{jS_t} \sim \mathcal{N}(0, \tau_j), \tag{2}$$

where τ_j denotes a fixed prior scaling for coefficient j. Notice that τ_j is regime invariant and the prior is centered on zero. In time series applications and for non-stationary data, it is common practice that the prior on the first lag of y_t is centered on unity, while for higher lag orders the prior mean is set equal to zero.

The symmetric prior in Eq. (2) translates into a prior on the standardized distance between coefficients. For illustration, we compute the prior distance between β_{jk} and β_{jl} for $k \neq l$,

$$\frac{\beta_{jk} - \beta_{jl}}{\sqrt{2}} \sim \mathcal{N}(0, \tau_j). \tag{3}$$

If τ_j is close to zero, the corresponding coefficients do not differ significantly across regimes. However, they are simultaneously strongly shrunk to zero. Thus, while such a prior is able to push selected coefficients towards homogeneity, it is not capable of handling cases where coefficients are non-zero but at the same time appear to be the same across regimes.

As a solution, we follow Yau and Holmes (2011) and assume that the regime-specific coefficients arise from a common distribution,

$$\beta_{jS_t} \sim \mathcal{N}(\beta_j, \tau_j). \tag{4}$$

The common mean β_j is treated as unknown and estimated from the data. It is noteworthy that if τ_j is close to zero, β_{jS_t} is pushed to β_j across all regimes. Computing the standardized distance between β_{jk} and β_{jl} yields Eq. (3) and the same intuition applies. However, instead of pulling β_{jS_t} to zero for all regimes, this specification shrinks towards parameter homogeneity and Eq. (4) can be interpreted as a Gaussian hierarchical prior on β_{jS_t}.

In what follows, we use shrinkage priors on τ_j to test the existence of significant differences across regimes. Specifically, we assume that τ_j arises from a Gamma distribution,

$$\tau_j \sim \mathcal{G}(d_0, d_1), \tag{5}$$

with d_0 and d_1 being hyperparameters. This specification has been introduced by Griffin and Brown (2010) within a regression context and subsequently adopted by Malsiner-Walli et al. (2016) to determine variable relevance in finite mixture models. Using a Gamma prior allows for shrinking τ_j to zero if necessary, and thus permits selecting whether parameters differ across regimes.

---

[1] For textbook introductions, see Kim and Nelson (1999) and Frühwirth-Schnatter (2006).
[2] The arguments we provide below hold for any law of motion of S_t.

3. ECONOMETRIC FRAMEWORK

In Section 3.1, we first discuss the multivariate MS error correction specification while Section 3.2 discusses the prior choice and Section 3.3 briefly summarizes the main steps necessary to perform posterior inference.

3.1. *A nonlinear vector error correction model*

The simple model outlined in the previous section is now generalized to a MS-VECM with two regimes. We opt for two regimes since we are interested in developing a model of the Euro area that encompasses several stylized facts about business cycles, namely pronounced co-movement of macroeconomic quantities over the business cycle as well as the distinction between expansionary and recessionary stages (Burns and Mitchell, 1946). Moreover, our proposed framework explicitly aims at discriminating between short- and long-run dynamics.

We assume that the first difference of an m-dimensional vector of time series $\{\boldsymbol{y}_t\}_{t=1}^T$ follows an MS-VECM with P lags,

$$\Delta \boldsymbol{y}_t = \boldsymbol{\lambda}_{S_t} \boldsymbol{b}' \boldsymbol{y}_{t-1} + \sum_{p=1}^{P} \boldsymbol{B}_{pS_t} \Delta \boldsymbol{y}_{t-p} + \boldsymbol{H}_{S_t} \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_m). \tag{6}$$

Here we let $\boldsymbol{\lambda}_{S_t}$ be an $m \times r$ matrix of short-run adjustment coefficients, $\boldsymbol{b}$ is an $m \times r$ matrix of long-run relations and $\boldsymbol{B}_{pS_t}$ are $m \times m$ coefficient matrices associated with the pth lag of $\Delta \boldsymbol{y}_t$. Furthermore, $\boldsymbol{H}_{S_t}$ is the lower Cholesky factor of a regime-specific variance-covariance matrix $\boldsymbol{\Sigma}_{S_t} = \boldsymbol{H}_{S_t} \boldsymbol{H}'_{S_t}$ and S_t is a discrete Markov process that takes values zero or unity detailed below.[3] In what follows we rewrite Eq. (6) as

$$\Delta \boldsymbol{y}_t = \boldsymbol{A}_{S_t} \boldsymbol{x}_t + \boldsymbol{H}_{S_t} \boldsymbol{\eta}_t, \tag{7}$$

whereby $\boldsymbol{A}_{S_t} = (\boldsymbol{\lambda}_{S_t}, \boldsymbol{B}_{1S_t}, \dots, \boldsymbol{B}_{PS_t})$ is an $m \times K$ matrix of stacked coefficients with $K = r + mP$. Moreover, $\boldsymbol{x}_t = (\boldsymbol{w}'_t, \Delta \boldsymbol{y}'_{t-1}, \dots, \Delta \boldsymbol{y}'_{t-P})'$ and $\boldsymbol{w}_t = \boldsymbol{b}' \boldsymbol{y}_{t-1}$ denotes the r cointegration errors.[4] Notice that Eq. (7) is a standard multivariate regression model conditional on $\boldsymbol{b}$. We explicitly rule out the possibility of breaks in $\boldsymbol{b}$ since they are assumed to be long-run fundamental relations and thus not subject to abrupt changes. However, there exist studies which allow for nonlinearities in the cointegration relationship (see, for instance, Martin, 2000; Bec and Rahbek, 2004; Koop *et al.*, 2011; Jochmann and Koop, 2015).

The transition probabilities of S_t, $\text{Pr}(S_t = j | S_{t-1} = i, \boldsymbol{\gamma}, \boldsymbol{w}_t) = p_{ij,t}$, are specified to be time-varying and depend on the (lagged) cointegration errors $\boldsymbol{w}_t$ through a set of regression coefficients in $\boldsymbol{\gamma}$. More precisely, the matrix of transition probabilities is given by

$$\boldsymbol{P}_t = \begin{pmatrix} p_{00,t} & p_{01,t} \\ p_{10,t} & p_{11,t} \end{pmatrix}. \tag{8}$$

The rows of Eq. (8) sum to unity for all i and t. Following the literature on early warning Markov switching models (Filardo, 1994; Kim and Nelson, 1998; Amisano and Fagan, 2013; Huber and Fischer, 2018), we assume that the transition probabilities are parameterized using a probit specification,[5]

$$p_{ij,t} = \Phi(c_{0i} + \boldsymbol{\gamma}' \boldsymbol{w}_t), \tag{9}$$

with c_{0i} denoting a regime-specific intercept term and Φ denotes the cumulative distribution function of the standard normal distribution. The jth element in $\boldsymbol{\gamma}$ measures the sensitivity of the transition probabilities with respect to the jth cointegration error, w_{jt}. Following Amisano and Fagan (2013), we

---

[3]Note that it is also possible to estimate different chains for the coefficients and volatilities, as in Hubrich and Tetlow (2015).

[4]For notational simplicity we suppress the dependence of $\boldsymbol{w}_t$ on $\boldsymbol{b}$.

[5]In the case of more than two regimes, a potential alternative would be a logit specification (Kaufmann, 2015; Billio *et al.*, 2016).

postulate that γ is time-invariant while the intercept term depends on the prevailing regime. Using the latent variable representation of the probit model yields

$$z_t^* = c_{0i} + \gamma' w_t + u_t, \quad u_t \sim \mathcal{N}(0, 1). \tag{10}$$

For identification purposes, the error variance is set equal to unity.

Before proceeding to the prior specification it is worth noting that λ and b are not identified since they enter Eq. (6) as a product. We achieve identification by using the linear normalization $b = (I_r, \Xi')'$ with Ξ being a $(m - r) \times r$ matrix of coefficients. This choice is clearly not invariant to the ordering of the elements y_t but ensures that the model is exactly identified.[6]

3.2. *Prior specification*

The model outlined in the previous section is heavily parameterized and we thus adopt a Bayesian approach to estimation and inference. Consistent with the discussion in Section 2 we assume that the common distribution that gives rise to $a_{S_t} = \text{vec}(A_{S_t})$ follows a multivariate Gaussian distribution,

$$a_{S_t} \sim \mathcal{N}(a, \Omega), \tag{11}$$

where $\Omega = \text{diag}(\tau_1, \ldots, \tau_k)$ is a diagonal variance-covariance matrix with variances τ_j for $k = Km$. Analogous to Section 2, τ_j determines the similarity between elements in a_0 and a_1. For instance, if only the short-run adjustment coefficients differ across regimes, the corresponding elements in Ω will be rather large whereas for the remaining coefficients, the associated τ_js will be close to zero.

Following Malsiner-Walli *et al.* (2016) we specify a Gaussian prior on a and the Gamma prior outlined in Eq. (5) for all j,

$$a \sim \mathcal{N}(\underline{a}, \underline{\Omega}), \tag{12}$$
$$\tau_j \sim \mathcal{G}(d_0, d_1). \tag{13}$$

Hereby, we let $\underline{a}$ denote the k-dimensional prior mean vector and $\underline{\Omega}$ is a $k \times k$ prior variance-covariance matrix. In what follows, we set $\underline{a} = 0$ and $\underline{\Omega}^{-1} = 0 \times I_k$, leading to an improper prior on a. For τ_j, it is worth emphasizing that if $d_0 = 1$, we obtain the Bayesian Lasso (Park and Casella, 2008) used in Yau and Holmes (2011) while if $d_0 < 1$ increasing prior mass is placed on zero while the tails of the marginal prior become heavier (Malsiner-Walli *et al.*, 2016). In our empirical application, we set $d_0 = d_1 = 0.1$ to strongly center the prior on zero and allow for heavy tails.

The prior setup described in Eqs. (11)–(13) effectively allows for detecting which elements in a_{S_t} differ across regimes and which of them appear to be homogeneous over distinct business cycle stages. Especially in light of a moderate to large number of time series in y_t as well as a moderate number of lags P, the number of parameters per regime is large relative to the length of a typical dataset. In the presence of multiple regimes, however, this problem is even more severe and shrinkage is necessary to obtain reliable parameter estimates.

It is worth mentioning that the proposed setup suffers from the well-known label switching problem (see Frühwirth-Schnatter, 2006). This issue arises due to the labeling of the states not affecting the likelihood function. In such cases, one may either rely on random permutation techniques, or achieve identification by employing economic reasoning. In the empirical application, for instance, we assume that the conditional mean of the equation for unemployment is lower for $S_t = 0$, and achieve this by implementing a rejection step in the MCMC algorithm.

We proceed with the prior on the free elements of the long-run adjustment coefficients. For $\xi = \text{vec}(\Xi)$, the $v = (m - r)r$ free elements of b, we use a Gaussian prior,

$$\xi \sim \mathcal{N}(0, \zeta \times I_v), \tag{14}$$

---

[6]Another potential choice would be to identify the space spanned by the cointegrating vectors and introduce a restriction on this space (Strachan, 2003; Koop *et al.*, 2009).

where ζ is a prior hyperparameter that controls the tightness of the prior. We set $\zeta = 1$, which is a fairly uninformative choice given the scale of our data. In principle, it would also be possible to elicit a prior directly on the cointegrating space (Strachan, 2003). Here, we follow the traditional approach since we are interested in directly interpreting the corresponding cointegration error.

Following Frühwirth-Schnatter (2006), Malsiner-Walli *et al.* (2016) and Huber and Zoerner (2019), we use a hierarchical Wishart prior on $\boldsymbol{\Sigma}_{S_t}^{-1}$,

$$\boldsymbol{\Sigma}_{S_t}^{-1} \sim \mathcal{W}(\boldsymbol{S}, s), \tag{15}$$

$$\boldsymbol{S} \sim \mathcal{W}(\boldsymbol{Q}, q). \tag{16}$$

The prior hyperparameters are specified such that (see Frühwirth-Schnatter, 2006; Malsiner-Walli *et al.*, 2016)

$$s = 2.5 + \frac{m-1}{2}, \tag{17}$$

$$q = 0.5 + \frac{m-1}{2}, \tag{18}$$

$$\boldsymbol{Q} = \frac{100s}{\underline{s}} \operatorname{diag}(\hat{\sigma}_1^2, \ldots, \hat{\sigma}_m^2). \tag{19}$$

Here we let $\hat{\sigma}_j^2$ denote the OLS variance obtained by running m univariate autoregressive models of order P. This choice implies that the all of the regime-specific variance-covariance matrices stem from a common distribution, similar to the case for $\boldsymbol{a}_{S_t}$ outlined above.

Finally, we use Gaussian priors on $\boldsymbol{\Gamma} = (c_{0i}, \boldsymbol{\gamma}')'$,

$$\boldsymbol{\Gamma} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{V}), \tag{20}$$

with $\boldsymbol{V} = 10 \times \boldsymbol{I}_{r+1}$ to introduce relatively little prior information on $\boldsymbol{\Gamma}$. Decreasing the prior variance would lead to a situation where the researcher suspects that transition probabilities do not depend on $\boldsymbol{w}_t$ and tend to be time-invariant.

3.3. *Full conditional posterior simulation*

The model outlined in the previous sections features a joint posterior density that is intractable. Fortunately, all full conditional posterior distributions take a well known form and are thus amenable to perform Gibbs sampling. In this section, we briefly summarize the steps involved in order to obtain a valid draw from the joint posterior, focusing attention on the non-standard parts while providing references for the full conditionals that are standard.

Conditional on a suitable set of starting values, our Markov chain Monte Carlo (MCMC) algorithm cycles through the following steps:

1. Sample $\boldsymbol{a}_j$ ($j = 0, 1$) conditional on the remaining parameters and the states from a k-dimensional multivariate Gaussian distribution. The corresponding moments take a standard form (Zellner, 1973).

2. Simulate $\boldsymbol{\xi}$ from a Gaussian posterior distribution conditional on the remaining parameters, states and a set of identifying assumptions. The precise formulas can be found in Villani (2001) and Huber and Zoerner (2019).

3. The common mean a is simulated conditional on a_0, a_1 and $\underline{\Omega}$ from a Normal distribution, with $\odot$ indicating point-wise multiplication,

$$a|a_0, a_1, \Omega \sim \mathcal{N}(\overline{a}, \overline{\Omega})$$
$$\overline{\Omega} = \frac{1}{2}\Omega,$$
$$\overline{a} = \frac{1}{2}\left(a_0 + a_1\right).$$

4. Draws from the conditional posterior of τ_j are obtained by noting that $p(\tau_j|a_0, a_1)$ follows a generalized inverted Gaussian (GIG) distribution,

$$\tau_j|a, a_0, a_1 \sim \mathcal{GIG}\left(d_0 - 1, \sum_{j=0}^{1}(a_j - a)^2, 2d_1\right). \tag{21}$$

5. Draw Σ_j^{-1} (for $j = 0, 1$) from a Wishart conditional posterior distribution given by

$$\Sigma_j^{-1}|a_0, a_1, S^T, S, \mathcal{D} \sim \mathcal{W}(\overline{S}_j, \overline{s}_j)$$
$$\overline{S}_j = S + \frac{1}{2}\sum_{t:S_t=j}(\Delta y_t - A_j x_t')(\Delta y_t - A_j x_t')',$$
$$\overline{s}_j = s + N_j/2$$

whereby $S^T = (S_1, \ldots, S_T)'$ denotes the full history of the states, $\mathcal{D}$ the data and $N_j = \#(t : S_t = j)$, that is, the number of observations in regime j.

6. The common scaling matrix for the Wishart prior S is simulated from its Wishart distributed conditional distribution,

$$S|a_0, a_1 \sim \mathcal{W}(\overline{S}, \overline{q})$$
$$\overline{S} = Q + \sum_{j=0}^{1}\Sigma_j^{-1},$$
$$\overline{q} = q + 2s.$$

7. Simulate the full history of the states S^T as well as the transition probabilities using the algorithm outlined in Kim and Nelson (1998) and adopted in Amisano and Fagan (2013).

8. Estimate the full history of z_t^* and Γ using the algorithm proposed in Albert and Chib (1993).

We repeat this algorithm 85,000 times where the first 50,000 draws are discarded as burn-in. Convergence is assessed using standard trace plots as well as inefficiency factors and the Raftery and Lewis (1992) diagnostic. All measures point toward rapid convergence for most parts of the parameter space under scrutiny.[7]

4. EMPIRICAL APPLICATION

In this section, we first present our dataset and discuss details on model specification. We proceed by illustrating the features of the modeling approach by providing key in-sample empirical findings. The section is complemented by a real-time out-of-sample forecasting exercise.

[7]Detailed convergence diagnostics can be obtained from the authors upon request.

4.1. *Data overview and model specification*

For the empirical application we use the real-time database of Giannone *et al.* (2012), covering quarterly Euro area time series ranging from 1999:Q1 to 2017:Q4. Based on our focus of modeling inflation dynamics in the forecasting exercise, we select five quantities of interest from the full dataset and include two-year inflation projections of the ECB's survey of professional forecasters to measure inflation expectations (see, for instance, Koop and Onorante, 2012; Jarociński and Lenza, 2018).

The information set comprises the growth rate of the harmonized index of consumer prices (excluding energy and unprocessed food, labeled HICPEX), the unemployment rate (UNEMP), inflation expectations (HICP-EXPEC), Euro area capacity utilization in manufacturing (UTILIZ) in logs, oil prices (OIL) in logs, and the three-month Euribor (I3M) as a short-term interest rate. Due to the normalization assumption on $\boldsymbol{b}$, the ordering of the variables in $\boldsymbol{y}_t$ plays a crucial role. Here, we order

$$\boldsymbol{y}_t = (HICPXE, UNEMP, HICP\text{-}EXPEC, UTILIZ, OIL, I3M)'. \tag{22}$$

With this ordering we assume that the equilibrium realized inflation depends on unemployment, the inflation expectations, the capacity utilization, oil prices and the interest rate, resembling a variant of the expectation augmented Phillips curve. As is standard in the literature for quarterly data, we choose $P = 4$ lags in the empirical specification and include an intercept term. This translates to five lags of $\boldsymbol{y}_t$ in terms of a vector autoregression (VAR) in levels.

4.2. *Posterior distribution of the cointegrating errors*

A crucial modeling decision in the context of VECMs is to select the appropriate cointegration rank r. Our specific choice of r is based on using log predictive likelihoods, reported in Section 4.5. The following discussion is based on setting $r = 3$, which is a choice that makes sense from a theoretical perspective but also yields strong density predictions (see the discussion in Section 4.5).

The posterior distributions of the three cointegration errors that are subsequently used to inform the transition distributions are depicted in Fig. 1. Notice that the first error can be interpreted as the deviation of inflation from its long-run equilibrium value determined by the remaining elements in $\boldsymbol{y}_t$. Consequently, the second error can be considered as the deviation of unemployment from its long-run equilibrium value, related to the non-accelerating inflation rate of unemployment (NAIRU). By contrast, the third error can be viewed as the deviation of inflation expectations from the target level.

The first cointegration error in Fig. 1(a) indicates that inflation is below its equilibrium value during the first two years of the sample. From 2002 onward, it quickly reverts towards its long-run fundamental value and overshoots until mid 2005. The figure suggests that the difference between inflation and its long-run equilibrium value increased until the onset of the European sovereign debt crisis. From this point on, inflation is below its implied equilibrium value in most quarters, except for the period between 2012 and 2013. Turning to the second cointegration error, Fig. 1(b) indicates that unemployment is roughly at its long-run equilibrium until the beginning of 2006, while negative errors occur subsequently until 2009. In the build-up to the Great Recession, we observe large negative cointegration errors for unemployment, mirroring conventional measures of the Euro area output gap. A sharp increase occurs in early 2010, and the Double-Dip recession of the Euro area is clearly visible. By the end of 2015, the model suggests a return of unemployment to its long-run equilibrium value. The third cointegration error in Fig. 1(c) refers to the deviation of inflation expectations from trend inflation. Three distinct periods are worth emphasizing: First, a period of approximate equilibrium that harshly ended during the financial crisis with strong negative deviations. Second, two pronounced positive spikes are observable between 2011 and 2014. Finally, we observe negative cointegration errors in 2014 and 2015 similar in magnitude to the dip during the Great Recession, while inflation expectations return to their equilibrium level towards the end of the sample period.
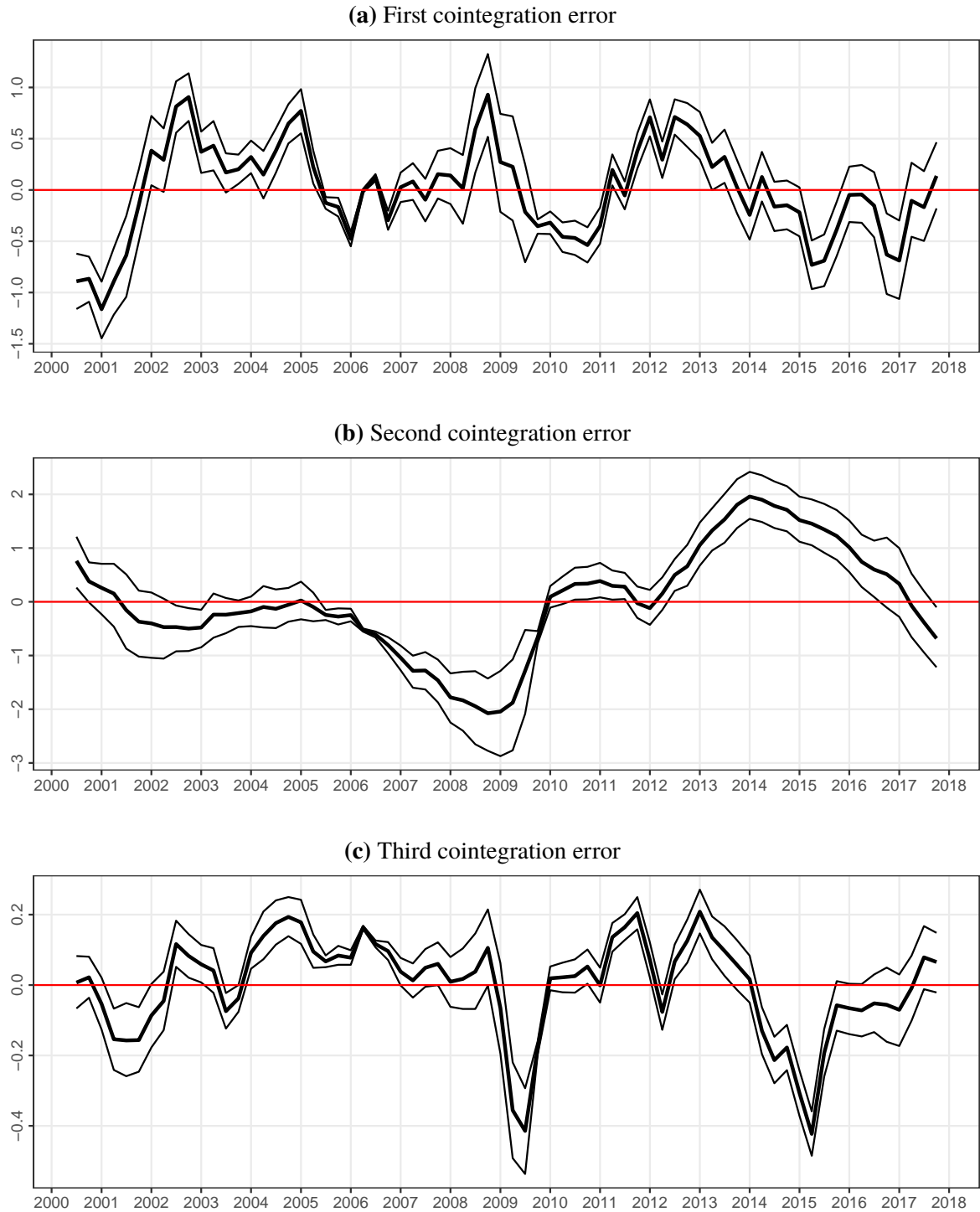
(a) First cointegration error

(b) Second cointegration error

(c) Third cointegration error

Fig. 1: Posterior distribution of the cointegration errors.

Notes: The thick black line is the posterior median of the cointegration errors, while the thin black lines refer to the 16th and 84th percentiles. The red line marks zero.

4.3. *Regime allocation and time-varying transition distributions*

In this section, we provide evidence for time variation in the posterior mean of the transition probabilities. Time-varying transition probabilities alongside posterior regime probabilities are depicted in Fig. 2, with gray shaded areas referring to the smoothed probability of being in the recession state. The solid and
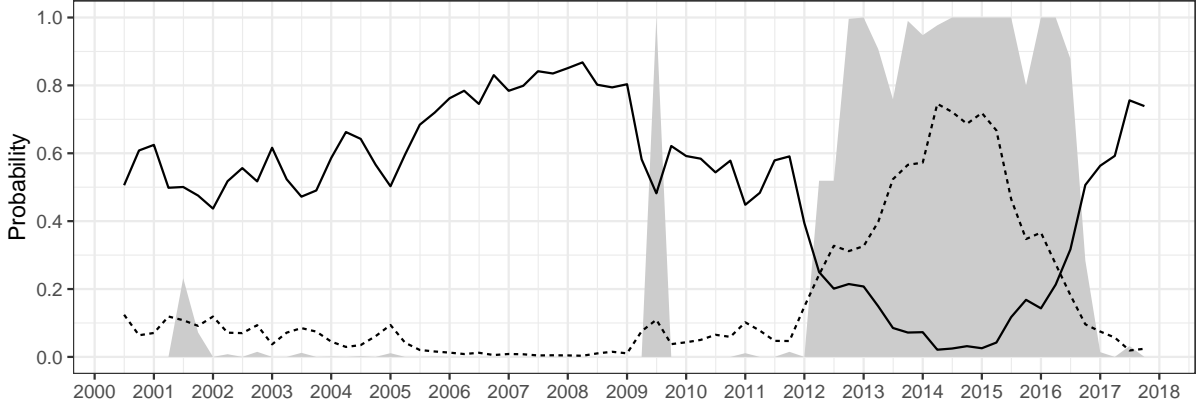
Fig. 2: Posterior mean of transition probabilities and filtered probabilities of being in the expansion or recession state.

Notes: The solid line in the upper part indicates $\Pr(S_t = 0|S_{t-1} = 1)$, the dashed line in the lower part denotes $\Pr(S_t = 1|S_{t-1} = 0)$. The gray shaded area indicates the posterior mean probability of the recession state.

dashed black lines are the posterior median of the off-diagonal elements in the time-varying transition probability matrix, indicate the probability of entering a recession at time t when being the expansionary state in $t-1$ (that is, $\Pr(S_t = 1|S_{t-1} = 0)$), and vice versa.

Our model tracks several periods of economic stress, with three recessionary episodes: First, the burst of the Dotcom bubble and the 9/11 terrorist attacks are visible. Second, the period surrounding the global financial crisis in 2008/2009 is surrounded by increases in the likelihood of moving into a recession. Finally, the European sovereign debt crisis is captured by our model with a prolonged period of elevated probabilities for $S_t = 1$ between 2012 and 2016. After the first crisis in our sample, a period of relative stability emerges, evidenced by a decline in the probability of moving into a recession when being in an expansion until the beginning of 2008. This period ends in 2008 with slightly higher probabilities of entering a recessionary phase, reflecting the beginning of the global financial crisis. After elevated probabilities of recession between 2012 and 2016, the period until the end of our sample is characterized by a steady decline of the probability of switching to the recessionary regime.

4.4. *Do parameters differ across regimes?*

The key novelty of our proposed approach is that it allows for flexible testing which coefficients differ across regimes. In order to assess differences in parameters, we rely on two visualizations that enable us to assess how much shrinkage is introduced, and to quantify the size of the deviations of $\boldsymbol{a}_{S_t}$ from the common mean $\boldsymbol{a}$. First, we consider the posterior median of the log of τ_j. Figure 3 presents the scaling parameters associated with the short-run adjustment coefficients in λ_{S_t} across equations.

Heterogeneity both across cointegration errors and equations is observable. This suggests that some elements in λ_{S_t} are clearly regime dependent, while others are pushed towards homogeneity. For example, the respective coefficients of *UTILIZ* are pushed towards the common mean and thus a linear specification (large negative values on the log-scale), while the coefficients for *HICPXE* and *I3M* differ across regimes. A key takeaway of Fig. 3 is that economic adjustment towards long-run equilibrium values appear to be dependent on the prevailing state of the economy.

We proceed by investigating whether the coefficients associated with the lagged endogenous variables differ for expansions and recessions. Fig. 4 shows the log scaling parameters for all autoregressive coefficients and the intercept term per equation. A few points are worth noting here. First, an apparent observable pattern is that shrinkage for the autoregressive coefficients tends to be similar across the lags
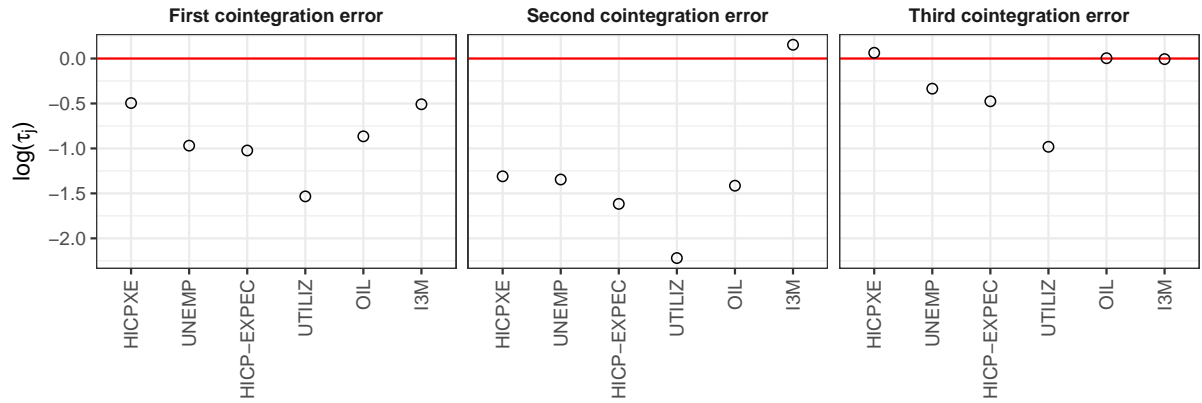
Fig. 3: Posterior mean of the log of the scaling parameters across equations: Adjustment coefficients associated with the cointegration cointegration errors.

Notes: Variables included are the harmonized index of consumer prices (HICPEX), the unemployment rate (UNEMP), inflation expectations (HICP-EXPEC), Euro area capacity utilization in manufacturing (UTILIZ), oil prices (OIL), and the three-month Euribor (I3M).

Fig. 4: Posterior mean of the log of the scaling parameters across equations: autoregressive coefficients.

Notes: Variables included are the harmonized index of consumer prices (HICPEX), the unemployment rate (UNEMP), inflation expectations (HICP-EXPEC), Euro area capacity utilization in manufacturing (UTILIZ), oil prices (OIL), and the three-month Euribor (I3M).
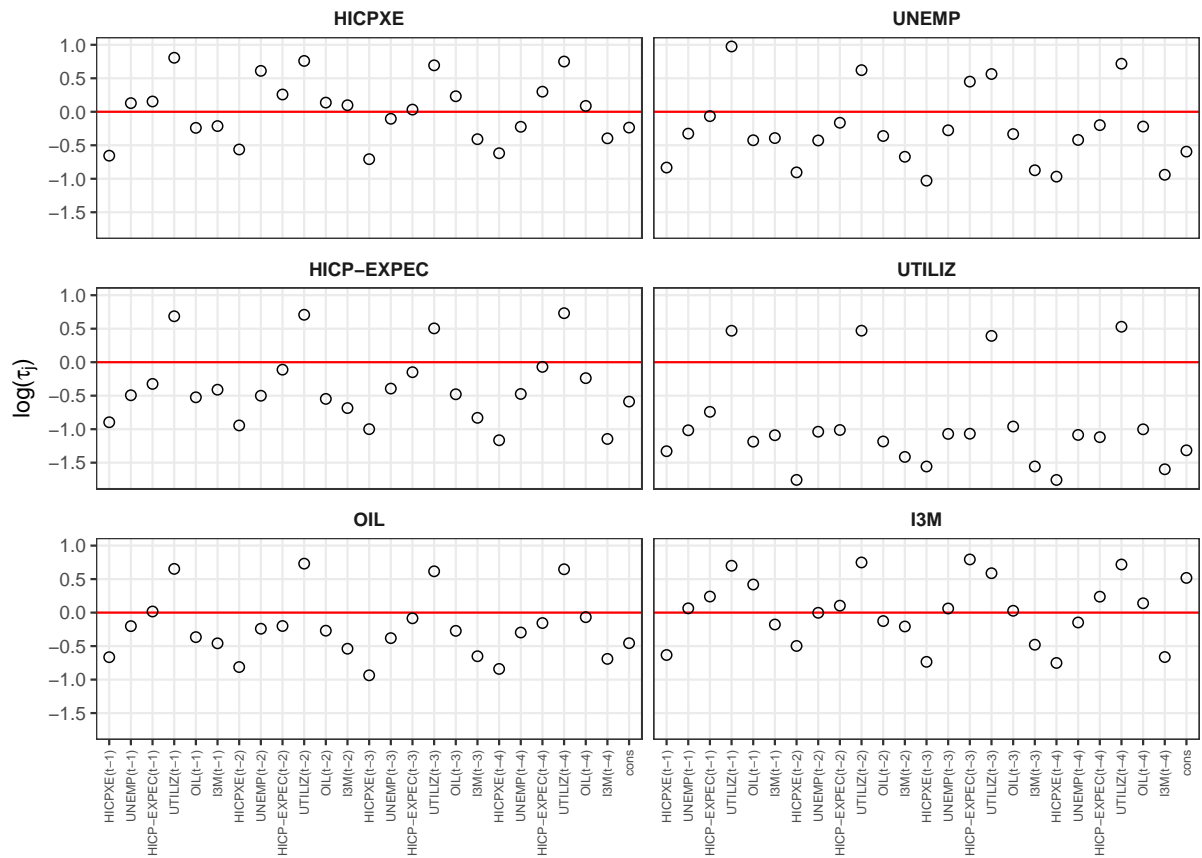
of specific quantities for all equations. In other words, if heavy shrinkage for the first lag of a variable is introduced, remaining lags are also strongly pushed towards the common mean.

Second, the amount of shrinkage towards linearity is equation specific. The coefficients in the equation for *UTILIZ* in Fig. 4, for instance, display only little deviations from the common mean and can thus be treated as time-invariant. This is indicated by large negative values of $\log \tau_j$ for most coefficients. By contrast, *I3M*, and to a slightly lesser degree *HICPEXE*, appear to require larger breaks in the model coefficients.

Third, independent of the respective equation, coefficients associated with the lags of *UTILIZ* seem to feature the largest differences between regimes. The lags of *I3M* and *HICPEXE*, on the other hand, are heavily pushed towards the common mean across all equations. Interestingly, this translates to similar patterns of shrinkage on autoregressive coefficients for all variables, even though the amount of shrinkage towards homogeneity differs. Finally, variation in the intercept across regimes is observable for all equations.

One shortcoming of the analysis presented in Figs. 3 and 4 is that it is not invariant with respect to the scaling in $\mathbf{y}_t$ (and its changes). To provide some evidence on the quantitative differences in the autoregressive coefficients, we compute the posterior mean of the distance between regime-specific coefficients and the underlying common distribution. The results are reported in Fig. 5. We find comparatively large differences for the equations associated with *HICPEXE* and *I3M* in both regimes, in line with the previous discussion of Fig. 4. Minor deviations are also apparent in the case of *OIL* and *HICP-EXPEC*, while *UTILIZ* shows only minor differences across regimes except for its own lag. Note that the two states are approximately symmetric, that is, positive deviations in the expansionary regime are typically accompanied by negative differences in the recessionary state. Interestingly, for unemployment *UNEMP* the differences do not mirror each other.
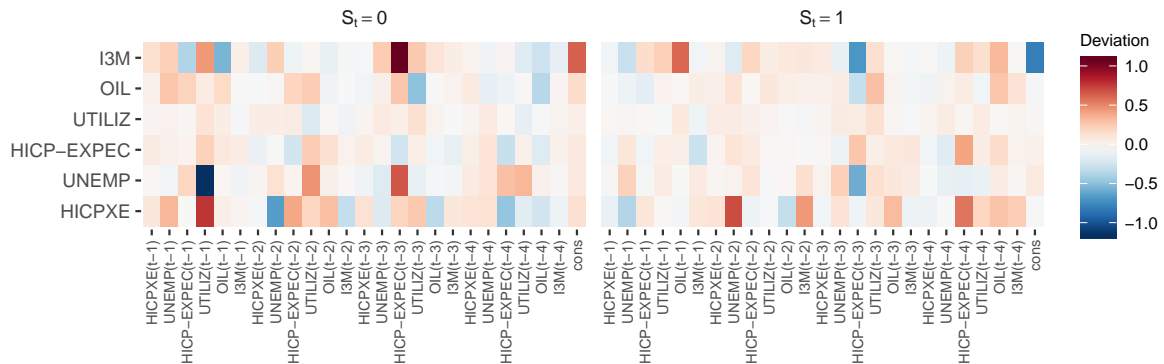

Fig. 5: Difference of posterior means between state-specific coefficients and the common distribution.

Notes: Variables included are the harmonized index of consumer prices (HICPEX), the unemployment rate (UNEMP), inflation expectations (HICP-EXPEC), Euro area capacity utilization in manufacturing (UTILIZ), oil prices (OIL), and the three-month Euribor (I3M).

Finally, we consider whether the variance-covariance matrices differ across regimes. To this end, Fig. 6 presents a boxplot of the marginal posterior distributions of the log determinant, the log trace and the log maximum eigenvalue of the variance-covariance matrix for the recessionary as well as for the expansionary regime. The plots report the median (bold line), together with the first and third quartile and the maximum and minimum value of the posterior distribution of the respective statistics. Two findings are worth emphasizing. First, considering the posterior median highlights that all considered measures of Σ_1 are larger, pointing towards larger error variances in recessions. This is consistent with other findings in the literature that highlight the necessity to allow for heteroscedasticity if recessionary episodes are included in the sample (for forecasting evidence, see, Clark, 2011). Second, and finally, we find that
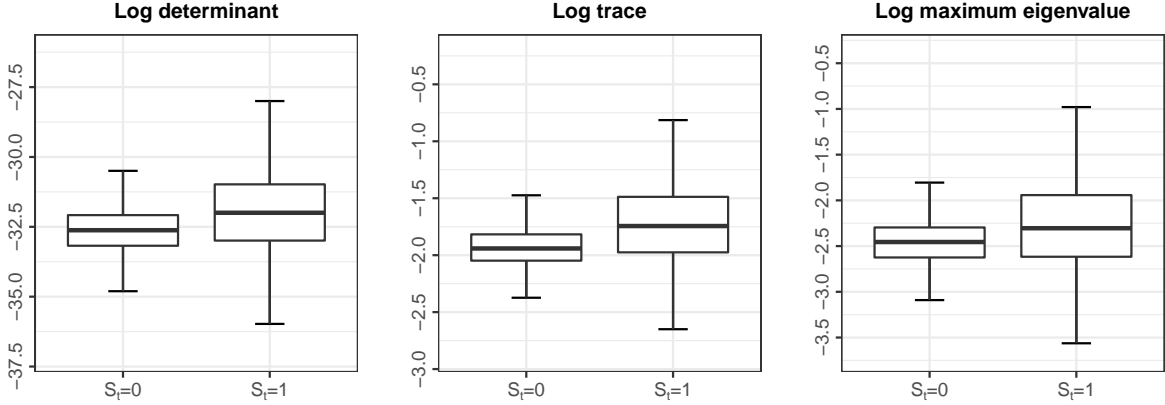
Fig. 6: Posterior distribution of the log determinant, log trace and the log maximum eigenvalue of the regime-specific variance-covariance matrices

Notes: The thick black line marks the posterior median, the white box covers the 25th and 75th percentile of the posterior distribution. The whiskers refer to the largest (smallest) value within 1.5 times the interquartile range above the 75th percentile (below the 25th percentile).

posterior uncertainty increases quite sharply during recessionary episodes. This stems from the fact that the recessionary regime features considerably less observations.

To sum up, we find that a subset of the regression coefficients in a_0 and a_1 are different from zero, and we provide strong evidence that error variance-covariance matrices differ across regimes. Our modeling approach stochastically selects a model where only selected VAR coefficients differ while some of short-run adjustment coefficients appear to be regime invariant. Moreover, our findings indicate that the regime allocation is also driven by differences in the variance-covariance matrices and this, to some extent, corroborates findings presented in Sims and Zha (2006). Before proceeding to the forecasting exercise, it is worth emphasizing that our model could be relaxed by introducing separate regime indicators that drive the regime allocation across different regions of the parameter space. However, our analysis essentially indicates that the bulk of time variation in the parameters of the model stem from the error variance-covariance matrices while the remaining coefficients are strongly pushed towards homogeneity. Introducing separate Markov chains to drive the transition of different blocks of the parameter space would be possible but probably not change much since, in terms of the VAR coefficients, we find comparatively little evidence that parameters tend to change (as opposed to rather strong evidence that the error variances change). This implies that in the recursions used to obtain a draw from the regime indicators, the regime allocation is dominated by the differences in the error variance-covariance matrices. Using different Markov chains to drive the switching behavior of the model would be necessary if we also find that regression coefficients differ sharply across regimes, and thus using a single regime indicator would translate into a miss-specified model. In addition, when it comes to our empirical application we feel that using additional state indicators would render illustrating the key features of the proposed prior much more difficult. We have included an additional discussion in the text where we discuss differences between regimes.

4.5. *Forecasting exercise*

We illustrate the merits of our model in a real time out-of-sample forecast exercise. We use the period from 1999:Q1 to 2005:Q4 as an initial estimation period while leaving out the period from 2006Q1 to 2017:Q4 to serve as a verification sample. We then proceed by estimating the model, taking into account different data vintages, and then compute the one-step-ahead predictive distribution. The predictive density is then used to evaluate model performance by computing log predictive scores (LPSs, see Geweke and Amisano,

2010) relative to a six variable Bayesian VAR model (BVAR) with five lags in Tab. 1 and a conjugate Minnesota prior in the spirit of Sims and Zha (1998b). Afterwards, we add an additional observation from the holdout sample and re-estimate the model.

Table 1: Out-of-sample forecasting results: 2006:Q1 to 2017:Q4.

	$r = 1$	$r = 2$	$r = 3$	$r = 4$	$r = 5$	r = NA
MS-VECM-TVP	14.09	21.69	**39.42**	28.76	26.59	
MS-VECM-FTP	28.22	29.42	21.71	21.67	29.8	
VECM	22.16	10.02	−66.76	−75.38	−117.13	
RW						35.89
AR(1)						34.58

Notes: The table presents relative log predictive scores for inflation against the BVAR. Large numbers signal better predictive performance. TVP refers to time varying transition probabilities while FTP denote fixed transition probabilities. $r = 1$ to $r = 5$ refers to the cointegration rank of the VECMs, VECM is a linear VECM, RW and AR(1) denote univariate random walk and autoregressive models of order one.

To investigate the merits of our approach, we include additional nested benchmark models. The set of competing models includes different variants of our model approach, differing in the number of cointegration relations ($r = 1, 2, \ldots, 5$), constant transition probabilities (labeled MS-VECM-FTP), and whether the coefficients feature time-variation (in this case, a linear VECM is adopted). Moreover, to illustrate whether using a multivariate approach pays off in forecasting terms, we include an AR(1) model and the simple random walk (RW) benchmark.

The cumulative LPSs in Tab. 1 clearly indicate MS-VECM with time varying transition probabilities and $r = 3$ as the best performing model in terms of predictive densities. Our highly flexible model does not only outperform the simple BVAR benchmark, but also the simpler random walk specification. In comparison, the linear specification does not perform very well, as it neglects various characteristics the nonlinear model explicitly takes into account.

To assess whether differences in forecast performance occur systematically, we assess the evolution of cumulative LPSs relative to the BVAR over time in Fig. 7. While the linear VECM specification underperforms consistently against the nonlinear MS-VECM models throughout the sample, the picture for the remaining models provides a rather homogeneous picture until 2008. Largest relative gains in predictive accuracy for MS-VECM-TVP $r = 3$ are observable in 2010 and 2011, and in late 2014.

5. CLOSING REMARKS

In this paper, we propose a hierarchical nonlinear model that allows for stochastically selecting coefficients that differ across regimes. The proposed MS-VECM discriminates between short- and long-run dynamics and assumes that the transition probability matrix of the underlying Markov process is time-varying. We assume that the autoregressive coefficients, the error variance-covariance matrices, as well as the short-run adjustment coefficients differ across regimes and arise from a common distribution. Moreover, another novel feature of our model is that the transition distributions are parameterized using a simple binary probit model with the (lagged) cointegration errors included as covariates.

We apply the proposed framework to a medium-scale dataset for the Euro area. Our empirical model discriminates between expansionary and recessionary business cycle stages and allows for assessing whether transition probabilities vary over time. Considering the posterior estimates for the filtered probabilities indicates that our model succeeds in replicating business cycle features of the Euro area. In addition, we investigate which coefficients differ across regimes and how this impacts regime allocation. In-sample evidence is complemented by a real time out-of-sample forecast exercise featuring various benchmark specifications.
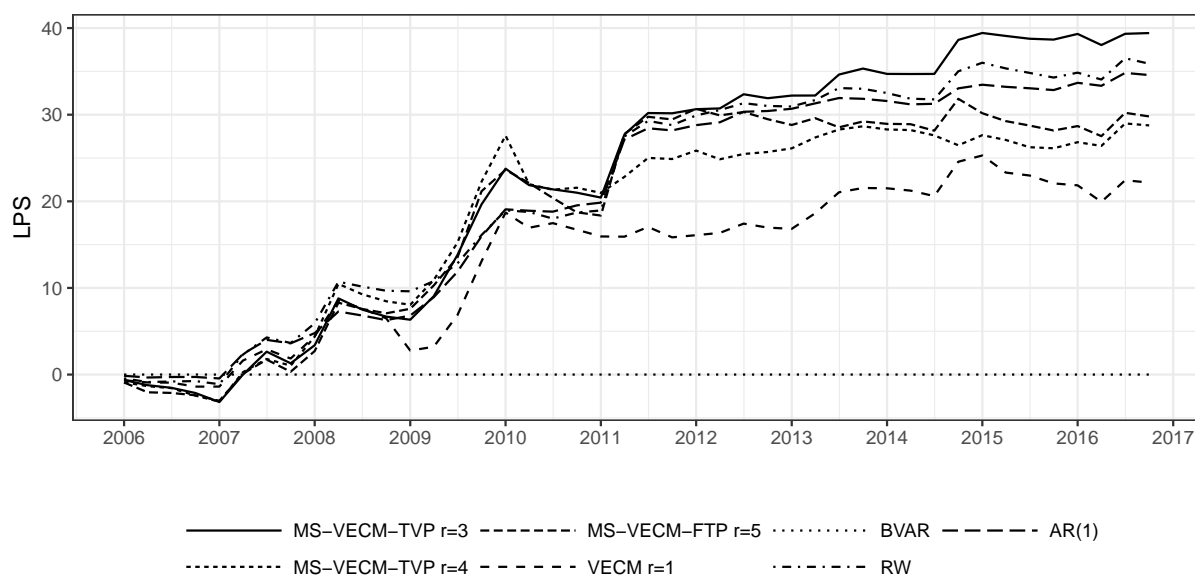
Fig. 7: Evolution of log predictive scores for inflation relative to the benchmark over the hold-out period: 2006:Q1 to 2017:Q4.

Notes: Large numbers signal better predictive density performance. TVP refers to time varying transition probabilities while FTP denote fixed transition probabilities. $r = 1$ to $r = 5$ refers to the cointegration rank of the VECMs, RW and AR(1) denote univariate random walk and autoregressive models of order one.

Our findings suggest that the proposed approach succeeds in reproducing Euro area recessionary episodes and, additionally, shows that deviations of inflation, unemployment and inflation expectations from their long-run fundamentals drive transitions between regimes. The proposed hierarchical model moreover suggests that only a subset of short-run adjustment coefficients can be assumed to be regime-invariant, and selected autoregressive coefficients differ markedly across expansion and recession periods. A further key modeling aspect is that volatilities of the shocks feature sharp differences across regimes, playing a substantial role in determining regime allocation. The forecast exercise shows that the proposed model is competitive in terms of forecasting Euro area inflation.

REFERENCES

ALBERT JH, AND CHIB S (1993), "Bayesian analysis of binary and polychotomous response data," *Journal of the American Statistical Association* **88**(422), 669–679.

AMISANO G, AND FAGAN G (2013), "Money growth and inflation: A regime switching approach," *Journal of International Money and Finance* **33**(3), 118–145.

BEC F, AND RAHBEK A (2004), "Vector equilibrium correction models with nonlinear discontinuous adjustments," *The Econometrics Journal* **7**(2), 628–651.

BILLIO M, CASARIN R, RAVAZZOLO F, AND VAN DIJK HK (2016), "Interconnections between eurozone and US booms and busts using a Bayesian panel Markov-switching VAR model," *Journal of Applied Econometrics* **31**(7), 1352–1370.

BOGNANNI M, AND HERBST E (2018), "A sequential Monte Carlo approach to inference in multiple-equation Markov-switching models," *Journal of Applied Econometrics* **33**(1), 126–140.

BURNS AF, AND MITCHELL WC (1946), *Measuring Business Cycles*, NBER Book Series Studies in Business Cycles, National Bureau of Economic Research, Cambridge (MA).

CHIB S (1996), "Calculating posterior distributions and modal estimates in Markov mixture models," *Journal of Econometrics* **75**(1), 79–97.

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

DOAN T, LITTERMAN R, AND SIMS C (1984), "Forecasting and conditional projection using realistic prior distributions," *Econometric Reviews* **3**(1), 1–100.

DROUMAGUET M, WARNE A, AND WOŹNIAK T (2017), "Granger Causality and Regime Inference in Markov Switching VAR Models with Bayesian Methods," *Journal of Applied Econometrics* **32**(4), 802–818.

FILARDO AJ (1994), "Business-cycle phases and their transitional dynamics," *Journal of Business & Economic Statistics* **12**(3), 299–308.

FRÜHWIRTH-SCHNATTER S (2006), *Finite Mixture and Markov Switching Models*, Springer, Berlin.

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

GIANNONE D, HENRY J, LALIK M, AND MODUGNO M (2012), "An area-wide real-time database for the euro area," *Review of Economics and Statistics* **94**(4), 1000–1013.

GOLDFELD SM, AND QUANDT RE (1973), "A Markov model for switching regressions," *Journal of econometrics* **1**(1), 3–15.

GRIFFIN JE, AND BROWN PJ (2010), "Inference with normal-gamma prior distributions in regression problems," *Bayesian Analysis* **5**(1), 171–188.

HUBER F, AND FISCHER MM (2018), "A Markov Switching Factor-Augmented VAR Model for Analyzing US Business Cycles and Monetary Policy," *Oxford Bulletin of Economics and Statistics* **80**(3), 575–604.

HUBER F, AND ZOERNER TO (2019), "Threshold cointegration in international exchange rates: A Bayesian approach," *International Journal of Forecasting* **35**(2), 458 – 473.

HUBRICH K, AND TETLOW RJ (2015), "Financial stress and economic dynamics: The transmission of crises," *Journal of Monetary Economics* **70**, 100 – 115.

JAROCIŃSKI M, AND LENZA M (2018), "An inflation-predicting measure of the output gap in the euro area," *Journal of Money, Credit and Banking* **50**(6), 1189–1224.

JOCHMANN M, AND KOOP G (2015), "Regime-switching cointegration," *Studies in Nonlinear Dynamics & Econometrics* **19**(1), 35–48.

KAUFMANN S (2000), "Measuring business cycles with a dynamic Markov switching factor model: An assessment using Bayesian simulation methods," *Econometrics Journal* **3**(1), 39–65.

——— (2015), "K-state switching models with time-varying transition distributions – Does loan growth signal stronger effects of variables on inflation?" *Journal of Econometrics* **187**(1), 82–94.

KIM CJ, AND NELSON CR (1998), "Business cycle turning points, a new coincident index, and tests of duration dependence based on a dynamic factor model with regime switching," *Review of Economics and Statistics* **80**(2), 188–201.

——— (1999), *State-space models with regime switching: Classical and Gibbs-sampling approaches with applications*, MIT Press, Cambridge (MA) and London (England).

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.

KOOP G, LEON-GONZALEZ R, AND STRACHAN RW (2011), "Bayesian inference in a time varying cointegration model," *Journal of Econometrics* **165**(2), 210–220.

KOOP G, AND ONORANTE L (2012), "Estimating Phillips curves in turbulent times using the ECB's survey of professional forecasters," *ECB Working Paper No. 1422* .

MALSINER-WALLI G, FRÜHWIRTH-SCHNATTER S, AND GRÜN B (2016), "Model-based clustering based on sparse finite Gaussian mixtures," *Statistics and computing* **26**(1-2), 303–324.

MARTIN GM (2000), "US deficit sustainability: a new approach based on multiple endogenous breaks," *Journal of Applied Econometrics* 83–105.

PAAP R, AND VAN DIJK HK (2003), "Bayes estimates of Markov trends in possibly cointegrated series: An application to US consumption and income," *Journal of Business & Economic Statistics* **21**(4), 547–563.

PARK T, AND CASELLA G (2008), "The Bayesian lasso," *Journal of the American Statistical Association* **103**(482), 681–686.

RAFTERY AE, AND LEWIS S (1992), "How many iterations in the Gibbs sampler?" in JM BERNARDO, JO BERGER, AP DAWID, AND AFM SMITH (eds.) "Bayesian Statistics 4," pp. 763–773, Oxford University Press, Oxford.

SIMS CA, WAGGONER DF, AND ZHA T (2008), "Methods for inference in large multiple-equation Markov-switching models," *Journal of Econometrics* **146**(2), 255–274.

SIMS CA, AND ZHA T (1998a), "Bayesian methods for dynamic multivariate models," *International Economic Review* 949–968.

——— (1998b), "Bayesian methods for dynamic multivariate models," *International Economic Review* **39**(4), 949–968.

——— (2006), "Were there regime switches in US monetary policy?" *American Economic Review* **96**(1), 54–81.

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.

YAU C, AND HOLMES C (2011), "Hierarchical Bayesian nonparametric mixture models for clustering with variable relevance determination," *Bayesian Analysis* **6**(2), 329.

ZELLNER A (1973), *An Introduction to Bayesian Inference in Econometrics*, Wiley, New York.