K-state switching models with time-varying transition distributions – Does credit growth signal stronger effects of variables on inflation?

Sylvia Kaufmann

Working Paper 14.04

K-state switching models with time-varying transition distributions – Does credit growth signal stronger effects of variables on inflation?

Sylvia Kaufmann*†

February 2014

Abstract

Two Bayesian sampling schemes are outlined to estimate a K-state Markov switching model with time-varying transition probabilities. Data augmentation for the multinomial logit model of the transition probabilities is alternatively based on a random utility and a difference in random utility extension. We propose a definition to determine a relevant threshold level of the covariate determining the transition distribution, at which the transition distributions are balanced across states. Identification issues are addressed with random permutation sampling. In terms of efficiency, the extension to the difference in random utility specification in combination with random permutation sampling performs best. We apply the method to estimate a regime dependent two-pillar Phillips curve for the euro area, in which lagged credit growth determines the transition distribution of the states.

JEL classification: C11,C22,E31,E52

Key words: Bayesian analysis, credit, M3 growth, Markov switching, Phillips curve, permutation sampling, threshold level, time-varying probabilities.

*Study Center Gerzensee, P.O. Box 21, CH-3115 Gerzensee, sylvia.kaufmann@szgerzensee.ch

†This working paper is a revised version of *K-state switching models with endogenous transition distributions* Swiss National Bank, Working Paper No. 2011-13, November 2011. I thank for comments obtained by an Associate Editor and two anonymous referees. They helped improve the paper a lot. I assume the responsibility for remaining errors or omissions.

1 Introduction

Bayesian estimation of Markov regime switching models is by now well developed in the literature (Chib 1996, Frühwirth-Schnatter 2006, Sims et al. 2008) and many applications have proved the model to be useful in the analysis of economic data. Among many others, see the multivariate approaches of Kim and Nelson (1998), Paap and van Dijk (2003), Hamilton and Owyang (2012), Kaufmann (2010). Generally, the transition probabilities are assumed to be exogenous, which represents a major critique addressed to Markov switching models (and to models with exogenous break dates in general), as they lack an explicit interpretation of the driving variables behind the switching process. Extensions to time-varying probabilities have usually focussed on the restriction to two states and have been parameterized using a probit specification (see Filardo 1994, Filardo and Gordon 1998). A multinomial logit specification is adopted in Meligkotsidou and Dellaportas (2011) who use recent derivations of auxiliary samplers for multinomial logistic models (Holmes and Held 2006) to estimate hidden Markov models.

In the present paper, time-varying probabilities are also parameterized using a multinomial logit function which provides a mean to extend Bayesian estimation to a K-state switching model in a straightforward way. Two Markov chain Monte Carlo (MCMC) samplers are proposed to estimate the model, both of which are based on data augmentation. The first one uses the extension of the multinomial logit model to the random utility representation and the second one the extension to the difference in random utility representation (Frühwirth-Schnatter and Frühwirth 2010). The advantage of introducing the additional layers is that draws from the posterior distribution of all parameters, including those driving the time-varying transition probabilities, are obtained from full conditional posterior distributions. Hence, we can rely on the Gibbs sampler while the alternative sampler of Holmes and Held (2006) involves rejection sampling in the random utility representation of the logit regression model. While parameter inference with both auxiliary samplers is straightforward and easy, it turns out that the extension to the difference in random utility representation is more efficient than the extension to the random utility model. Finally, note that although the samplers are presented within a univariate framework here, the schemes can be readily integrated in multivariate time series or panel data approaches like those mentioned before.

The posterior inference of the model allows to discriminate the Markov switching model against nested alternatives. A Markov switching model with constant, exogenous transition distribution is obtained if the parameters on the covariate are restricted to zero. If the parameters governing the transition distribution do not depend on the previous state, we obtain a mixture model with time-varying weights. In analogy to smooth transition models (STAR, Teräsvirta and Anderson 1992) in which the threshold level is the level of the covariate at which the state probability equals 0.5, we define a threshold level for the covariate, at which the state transition distributions are what we call balanced across states. We achieve this by defining the threshold level as the one at which the divergence between the persistence probabilities of states is minimized.

Another issue that we also address is identification, which is important to obtain an unbiased estimate of the identified model, (Hamilton et al. 2007). Regime switching models are not identified unless an ordering of the states is provided. Finding a uniquely state-identifying restriction is often driven by the investigation at hand. Nevertheless, there often are cases, in particular in models including an increasing number of parameters to estimate, where it is unclear a priori which coefficient may be used to uniquely identify the states. The issue is addressed by using the random permutation sampler (Frühwirth-Schnatter 2001) to first obtain an estimate of the unconstrained posterior distribution, which also yields an inference about the presence of Markov switching. Then the sample from the unconstrained posterior is postprocessed to infer a uniquely state-identifying restriction. Meligkotsidou and Dellaportas (2011) argue that identification is not an issue if the purpose of investigation is forecasting. Nevertheless, one might be interested in obtaining state-dependent forecasts, if e.g. the states would represent different macroeconomic scenarios, each of which would imply a state-specific policy response. In that case, model identification would again be a prerequisite.

Additional literature most directly related to the present paper includes Hamilton and Owyang (2012), who estimate US state-level recession clusters. They model cluster association of US state-level employment growth rates using a multinomial logit specification with four covariates. There is no path-dependence in cluster association, however. Another approach to model endogenous transition probabilities is presented in Billio and Casarin (2011), who specify regime-specific Beta autoregressive transition distributions depending on covariates like duration or past transition probabilities. A specification of time-varying transition probabilities depending on the strength of a latent, continuous state variable has been proposed in Chib and Dueker (2004). In Billio and Casarin (2010), endogenous transition probabilities are latent Beta random variables. Change-point models (Chib 1998) with a fixed number of regimes are nested in Markov switching models. Setting the appropriate zero restrictions in the transition matrix yields a process with switches to non-recurrent states. While Chib (1998) and Pesaran et al. (2007) assume constant transition probabilities, Koop and Potter (2007) render the approach more flexible by introducing a hierarchical prior for state duration which induces duration dependent transition probabilities. Moreover, the setup they pursue does not restrict the number of breaks to a predetermined value. Most recently, Geweke and Jiang (2011) present a multiple-break model in which the unknown number of break dates are indicated by a latent Bernoulli variable, with exogenous probability distribution, however. A logit specification of the break probability including explanatory covariates, as pursued in the present paper, could also be integrated in their approach.

We apply the model to the two-pillar Phillips curve for the euro area investigated in Assenmacher-Wesche and Gerlach (2008). They regress the quarterly inflation rate on the low-frequency components of M3 growth, real GDP growth and the change in the government bond yield, and on the high-frequency component of the output gap. They find that the coefficient on the low-frequency components of M3 growth and real GDP growth are not significantly different from 1 and -1, respectively. The low-frequency component of the change in the government bond yield looses its significance when the frequency band is shifted towards longer frequencies. The high-frequency component of the output gap remains significant in all frequency bands considered. This analysis confirmed the importance of M3 growth as an indicator for inflation prospects. However, it turns out that the relationship breaks down if the Phillips curve is estimated with shorter and more recent data series running from 1983 to 2010, and renders inflation nearly unpredictable – a pattern also present in US data (Stock and Watson 2007, Sargent and Surico 2011). Extending the setup to a Markov switching framework recovers the relevance of the variables for inflation. Lagged credit growth rate above a threshold level of 1.9% quarterly growth rate is estimated to be indicative of switches to the state in which is the low-frequency component of M3 growth and economic variables are significant for inflation. This state is characterized by an increased volatility in quarterly inflation, matched with, besides above trend M3 growth of credit, a trend M3 growth rate staying at a relatively high level, being on an increasing or strongly decreasing path.

The next section outlines the econometric model and discusses the parametrization of the transition distribution. Section 3 presents the MCMC sampling scheme. In section 4 the estimation method is illustrated with simulated data and contains the efficiency evaluation of the RUM and dRUM auxiliary samplers, each of which is implemented within the random and alternatively the constrained permutation sampler. The application to the two-pillar Phillips curve for the euro area is presented in section 5. Section 6 concludes. The detailed RUM and dRUM auxiliary sampler for the parameters of the transition distribution are found in appendix A. Appendix B describes the permutation steps used in model identification.

2 The econometric model

2.1 Specification

The usual representation of a regime-switching model for a time series y_t, $t = 1, \ldots, T$, is

$$y_t = X_t' \beta_{S_t} + \varepsilon_t \tag{1}$$

$$\varepsilon_t \sim \text{i.i.d } N(0, \sigma^2) \tag{2}$$

where X_t is a $p \times 1$ vector of explanatory variables which may include lagged observations of y_t if autoregressive dynamics are taken into account. The parameter vector β is state-dependent, $\beta_{S_t} = \beta_k$ if $S_t = k,\ k = 1, \ldots, K$. In the general case, the variance of the error terms may also be subject to regime changes, $\sigma^2_{S_t} = \sigma^2_k$ if $S_t = k$. The variance may even be driven by a state variable that is independent of the state variable governing the parameter vector β. For expositional convenience, we drop this assumption. The estimation of the model extended to state-dependent variances is straightforward. For completeness, we will discuss it in section 3, which outlines the sampling scheme.

The state indicator $S_t = k,\ k = 1, \ldots, K$, follows a first-order Markov process. A usual critique to Markov switching models with exogenous transition probabilities, in particular in macroeconomic applications, is the lack of an explicit inclusion/interpretation of the driving variable(s) behind the switching process. The usual procedure is then to correlate the estimated state probabilities to business cycle measures or to variables expected to influence the regimes. One can also compute moments of the variables like the state-dependent means and/or variances to characterize the estimated regimes. Another avenue has been to set up a model for the transition probabilities and to include explicitly the variables expected to influence them, which yields a model with time-varying transition probabilities. A covariate $\tilde{Z}_t$ affecting the transition distribution of the state variable then, through β_{S_t}, indirectly influences the effect of a variable in X_t.

In the present paper, we will parameterize the time-varying transition probabilities in what we call a *centered* way:

$$P(S_t = k|S_{t-1} = l, Z_t, \gamma) = \xi_{lk,t} = \frac{\exp\left(Z_t\gamma^z_{lk} + \gamma_{lk}\right)}{\sum_{j=1}^{K} \exp\left(Z_t\gamma^z_{lj} + \gamma_{lj}\right)}, \; k = 1, \ldots, K, \tag{3}$$

where the influence of the covariate is decomposed into two components. Namely, the time-varying component $\left(\tilde{Z}_t - \bar{Z}\right)\gamma^z_{lk}$, capturing the effect of deviations from the mean in the first term and the mean effect $\bar{Z}\gamma^z_{lk}$ entering the second term $\gamma_{lk} = \tilde{\gamma}_{lk} + \bar{Z}\gamma^z_{lk}$, which ultimately affects the time-invariant average state persistence.[1] The prior on γ_{lk} can then be specified jointly on all time-invariant components of the transition distribution, those coming from the truly exogenous part and those coming from mean effects of covariates.

For identification purposes, the parameters governing the transition to the "reference" state k_0, $k_0 \in \mathcal{K} = \{1, \ldots, K\}$, are assumed to be zero, $\left(\gamma^z_{lk_0}, \; \gamma_{lk_0}\right) = 0$. This yields

$$P(S_t = k_0|S_{t-1} = l, Z_t) = \frac{1}{1 + \sum\limits_{j \in \mathcal{K}_{-k_0}} \exp\left(Z_t\gamma^z_{lj} + \gamma_{lj}\right)} \tag{4}$$

where $\mathcal{K} = \{1, \ldots, K\}$ is the set of all states and $\mathcal{K}_{-k_0}$ means all states but the reference transition to state k_0.

The reasons why we explicitly use the centered parametrization (3) are twofold. First, it defines the average $\bar{Z}$ as an (initial arbitrary) threshold level. This is not restrictive, as we show below how the posterior estimate of the model can be used to define a threshold level which would differ from the average. Second, in the *uncentered* specification

$$\xi_{lk,t} = \frac{\exp\left(\tilde{Z}_t\gamma^z_{lk} + \tilde{\gamma}_{lk}\right)}{\sum_{j=1}^{K} \exp\left(\tilde{Z}_t\gamma^z_{lj} + \tilde{\gamma}_{lj}\right)} = \frac{\exp\left(\tilde{Z}_t\gamma^z_{lk} + \left(\gamma_{lk} - \bar{Z}\gamma^z_{lk}\right)\right)}{\sum_{j=1}^{K} \exp\left(\tilde{Z}_t\gamma^z_{lj} + \tilde{\gamma}_{lj}\right)} \tag{5}$$

[1] The model can be generalized to include more than one covariate to influence the transition probabilities. In that case Z_t and γ^z_{lk} would be $m \times 1$ vectors of variables and of parameters, respectively. The product in the numerator and denominator would then read $Z_t'\gamma^z_{lk}$.

the time-invariant part of the transition probabilities $\tilde{\gamma}_{lk}$ would reflect the time-invariant part net of the mean effect of $\tilde{Z}_t$. Formulating a prior on $\tilde{\gamma}_{lk}$ is then not scale invariant with respect to $\tilde{Z}_t$. In fact, only diffuse priors might be appropriate in this parametrization given that $\tilde{\gamma}_{lk}$ might be a large negative or positive number, depending on the sign of $\tilde{Z}_t$ (think of survey indices which may take on only positive values). As already mentioned, using the centered specification, we circumvent the problem in that we formulate a prior on γ_{lk} which contains all time-invariant parts of the transition probabilities.

Although we do not put any restrictions on γ^z_{lk}, after estimation they should reflect a property that we may think of as being reasonable in a Markov switching process (see also the examples in subsection 2.3). When deviating from zero (or another non-trivial threshold), the covariate Z_t should increase the dispersion in the persistence of the states, by e.g. increasing the switching probability from state 1 to state 2 (decreasing the persistence of state 1) and increasing the persistence of state 2. Thus, when $K = 2$, parameters in γ^z_k considerably shifted away from zero should be so in the same direction. When $K > 2$, this property should at least be present between parameters relating to two (past) states.

At first sight, the choice of using the logit specification in (3) instead of the probit specification mostly used otherwise in the literature is purely a matter of personal preferences, and comes even at the cost of working with a non-linear, non-normal model for the parameters. In this view, the probit model seems more attractive because the latent utility specification yields a linear normal model. This advantage against the logit specification, besides the pro of the possibility to resolve the issue of independence of irrelevant alternatives (IIA), is readily exploitable in models in which the state or classification indicator is known (Hausman and Wise 1978, Albert and Chib 1993, Nobile 1998, McCulloch et al. 2000), because the outcome probabilities do not need to be evaluated to obtain an inference on the parameters. However, when the state indicator is not known and is to be estimated, it turns out that, after taking a second, a more general look (in particular for $K > 2$), the probit specification is less trivial than expected at first sight. In order to make an inference on the state indicator, the transition probabilities have to be evaluated. In a K-state Markov switching model for a time series of length T, this amounts to evaluate $TK(K-1)$ $K-1$-dimensional multivariate normal integrals (Nobile 1998). This is numerically feasible, but computationally intensive and based on numerical methods introducing approximation errors which propagate through time because of state dependence and directly bias likelihood evaluation. I view this computational aspect as a major disadvantage of the probit specification for latent, in particular $K > 2$-state processes.[2] Besides, the independent probit specification used so far in applications of latent Markov switching models (see also Lu and Berliner 1999 for an application to other than economic data) yields approximately the same odds structure among states as the logit specification (Hausman and Wise 1978). Therefore, given the computational advantage in evaluating the time-varying transition distributions and the recent advances in Bayesian methods (Frühwirth-Schnatter and Frühwirth 2007 and 2010) which render the estimation of multinomial logit models amenable to Gibbs sampling, the logit specification overtakes the at-first-sight attractiveness of the probit specification.

Nevertheless, the use of a fully specified probit specification for Markov switching transition distributions remains a very interesting avenue for future research, see e.g. Imai and van Dyk (2005). It is obviously conceivable that one would like to introduce correlation among the probability of states which are perceived to have similar characteristics. In such a way, their common odds ratio against all other states would be the same as if only one of the two states would be included in the set of possible alternatives. Note that the latent ordered probit specification proposed in Kim, Piger, and Startz (2008) might also provide a basis to design a posterior sampler. Ordered probit reduces the computational burden to the evaluation of $TK(K-1)$ integrals of univariate rather than multivariate normals. However, using ordered probit for Markov switching models implies a structure on the probit parameters which is not obvious to implement.

2.2 Nested alternatives and a digression: Defining a threshold

In the literature implementing time-varying transition probabilities (Filardo 1994, Amisano and Fagan (2010)) it is sometimes assumed that the effect of the covariate is independent of the past state, which restricts $\gamma_{lk}^z = \gamma_k^z$. The Markov dependence is then only governed by the time-invariant part γ_{lk}. If the effect of the covariate is irrelevant, $\gamma_{lk}^z = 0,\ \forall l, k$, we obtain a K-state

[2] So far, applications using the probit specification involving more than 2 states are rare, if existent at all. On the other hand, some recent working papers (Billio et al. 2013 and Gaggl and Kaufmann 2014) refer to the method proposed in this paper to latent $K > 2$-state switching processes.

Markov switching model with constant transition probabilities.

If both $\gamma_{lk}^z = \gamma_k^z$ and $\gamma_{lk} = \gamma_k$, $\forall k$, the time-varying state probabilities are independent of the lagged prevailing state. The regime probability is then a monotone function of Z_t only:

$$P(S_t = k|Z_t, \gamma) = \xi_{kt} = \frac{\exp\left(Z_t \gamma_k^z + \gamma_k\right)}{\sum_{j=1}^{K} \exp\left(Z_t \gamma_j^z + \gamma_j\right)}$$

and we obtain a mixture model with time-varying weights.

Usually, threshold levels of covariates are not an issue in Markov switching models. However, in addition to know about the sensitivity of the state probabilities to changes in the covariate, it might equally well be of interest at which level of the covariate the mass of the transition distributions is critically shift towards one of the states. For instance, Amisano and Fagan (2010) use the money growth as an indicator for the probability to switch to a regime of high inflation. Knowing about the level of money growth at which the persistence of or the probability to switch to the high-inflation states increases critically, would be a useful information for policy makers. In smooth transition models (Teräsvirta and Anderson 1992), the threshold level is defined as the level of Z_t at which the regime probability, define it φ_t here, equals 0.5. We can re-parameterize the well-known transition function to our notation ($\gamma^z \neq 0$ holds throughout):

$$\varphi_t = \frac{1}{1 + \exp\left(-\gamma^z \left(Z_t - c\right)\right)} \quad \text{or} \quad \varphi_t = \frac{1}{1 + \exp\left(-\gamma^z Z_t + \gamma^c\right)} \tag{6}$$

where γ^z represents the curvature or the steepness of the transition function, c is the threshold, and $\gamma^c = \gamma^z c$. From an estimate of the right-hand parametrization, we can recover the threshold by solving $(-\hat{\gamma}^z Z_t + \hat{\gamma}^c) = 0$ for Z_t: The solution is $Z_t = \hat{\gamma}^c / \hat{\gamma}^z$. If Z_t is mean-adjusted, as defined so far, and γ^c is zero, the threshold is equal to the mean $\bar{Z}$. If $\gamma^c \neq 0$, the threshold is $\bar{Z} + \gamma^c/\gamma^z$. In a sidestep, note that the mean of a covariate might serve as a reasonable starting value for the threshold. The true one can then be recovered from an estimate of the transition function as parameterized on the right-hand side in (6).

From these considerations, we will now define a threshold for a covariate $\tilde{Z}_t$ determining the transition distribution in a Markov switching model, where states are characterized by their persistence and their switching probabilities. Our definition is:

Definition 1: The relevant threshold level of a covariate $\tilde{Z}_t$ which affects the transition distribution of a Markov switching state variable is the level of $\tilde{Z}_t$ at which the divergence between the persistence probabilities of states is minimized.

To motivate the definition, assume $K = 2$. According to the definition, the threshold level would be the level of $\tilde{Z}_t$, at which the persistence of both states is equalized. This represents a balanced situation where the probability to remain in the prevailing state and the probability to switch to the other state would be equal across the states. A deviation of $\tilde{Z}_t$ from that level renders the transition distribution asymmetric across states and shifts probability mass towards one of the states. In case $K > 2$, and depending on the parameter constellations, it is highly probable that there is no level of $\tilde{Z}_t$ which will equalize the persistence probabilities across states. In that case, we define the threshold as the level of $\tilde{Z}_t$ at which the divergence across the persistence of states is minimized. Finally, if we work with the mean-adjusted covariate Z_t, we apply Definition 1 to an estimate of (3). The obtained threshold level is then added to the mean $\bar{Z}$ to obtain the relevant threshold level of $\tilde{Z}_t$.

To sum up, having obtained an inference on the posterior distribution of the parameters governing the transition probabilities in (3), we may assess whether the model could be restricted to one of the discussed alternative parametrization. We may also deduce a threshold level for $\tilde{Z}_t$, at which the transition distribution is balanced and as symmetric as possible across states.

2.3 Some examples

To illustrate the various effects of the covariate on the transition distribution, let us assume three scenarios for Z_t, $Z_t \in \{0, 0.3, -0.3\}$. Assume two states for S_t, $S_t = 1, 2$ and state 1 to be the reference transition state. The model (3) can be written as

$$\xi_{l2,t} = \frac{\exp\left(\mathbf{Z}_t' \gamma_2\right)}{1 + \exp\left(\mathbf{Z}_t' \gamma_2\right)} \tag{7}$$

where $\mathbf{Z}_t' = \left(Z_t D_{t-1}^{(1)}, Z_t D_{t-1}^{(2)}, D_{t-1}^{(1)}, D_{t-1}^{(2)}\right)$, with $D_t^{(j)} = 1$ if $S_t = j$ and 0 otherwise, $j = 1, 2$. The parameter γ_2 has four elements, $\gamma_2 = (\gamma_{12}^z, \gamma_{22}^z, \gamma_{12}, \gamma_{22})'$. The first two elements determine the state-dependent effect of the covariate on the transition distribution to state 2. The last two elements, $(\gamma_{12}, \gamma_{22})$, are the parameters governing the time-invariant transition distributions from, respectively, state 1 and 2 in period $t - 1$ to state 2 in period t. Four different settings for γ_2 are assumed. In the first three, $\gamma_2 = (4, g, -2, 2)'$, $g \in \{0, 1, 4\}$, which yields an average persistence of 0.88 for each state. The influence of the various settings on ξ_t is depicted in table 1. In the first row where $\gamma_2 = (4, 0, -2, 2)'$, we observe that Z_t influences only the transition distribution of state 1. When Z_t is positive, the probability to switch to state 2 increases from 0.12 to 0.31. Conversely, as soon as Z_t would decrease, the persistence of state 1 would increase. In the second row where $\gamma_2 = (4, 1, -2, 2)'$, we observe that now an increase (a decrease) in Z_t also increases (decreases) the persistence of state 2. The two settings thus illustrate the property that the dispersion between state persistence is positively related to deviations of the covariate from its mean (or threshold). In the third row, $\gamma_2 = (4, 4, -2, 2)'$, the effect of Z_t is independent of the past prevailing state and the transition probabilities are a monotone function of Z_t only. The changes in the persistence probabilities are then symmetric for deviations of Z_t from zero. The second last row contains the effects when $\gamma_2 = (4, 4, 2, 2)'$, which represents the setting where the state probabilities are a monotone function of Z_t only, without dependence on the past prevailing state. For completeness, we add a parameter setting, in which the effect of Z_t goes into opposite directions for the state transition distributions. We observe that this case would capture situations in which positive (negative) deviations of the covariate from its mean would render an economic system more labile (inert), reflected in a decrease (an increase) in both state persistence probabilities.

From these examples, we would argue that in macroeconomic investigations the first three settings would be the most expected ones for Markov sitching models with significantly time-varying transition probabilities. A relevant covariate, in our view, would shift the mass of all (or most) transition distributions towards the same state.

Figure 1 illustrates the nonlinear effect of the covariate on the persistence probabilities of the states for the second, second last and last parameter settings of table 1, respectively. Panel (a) depicts the effect on the state persistence probabilities in the case we think is the most expected one in macroeconomic analysis. Panel (a) and (c) illustrate that in all settings of table 1 except for the second last one, the relevant threshold level according to our definition would be zero. At that level, the persistence probabilities are equal. They diverge, as Z_t deviates from zero. In the second last setting, and in fact also in the last one for equal parameters of opposite sign in γ^z (in which case the lines in panel (c) would overlap), given our definition the parameters would imply a threshold level of respectively $Z_t = -0.5$ and

Table 1: Time varying transition probabilities. Some examples for $\xi_t = P(S_t|S_{t-1}, Z_t, \gamma)$, $\gamma' = (\gamma^z_{12}, \gamma^z_{22}, \gamma_{12}, \gamma_{22})$

$\gamma' =$	$Z_t = 0$	$Z_t = 0.3$	$Z_t = -0.3$
(4,0,-2,2)	$\begin{bmatrix} 0.88 & 0.12 \\ 0.12 & 0.88 \end{bmatrix}$	$\begin{bmatrix} 0.69 & 0.31 \\ 0.12 & 0.88 \end{bmatrix}$	$\begin{bmatrix} 0.96 & 0.04 \\ 0.12 & 0.88 \end{bmatrix}$
(4,1,-2,2)	$\begin{bmatrix} 0.88 & 0.12 \\ 0.12 & 0.88 \end{bmatrix}$	$\begin{bmatrix} 0.69 & 0.31 \\ 0.09 & 0.91 \end{bmatrix}$	$\begin{bmatrix} 0.96 & 0.04 \\ 0.15 & 0.85 \end{bmatrix}$
(4,4,-2,2)	$\begin{bmatrix} 0.88 & 0.12 \\ 0.12 & 0.88 \end{bmatrix}$	$\begin{bmatrix} 0.69 & 0.31 \\ 0.04 & 0.96 \end{bmatrix}$	$\begin{bmatrix} 0.96 & 0.04 \\ 0.31 & 0.69 \end{bmatrix}$
(4,4,2,2)	$\begin{bmatrix} 0.12 & 0.88 \\ 0.12 & 0.88 \end{bmatrix}$	$\begin{bmatrix} 0.04 & 0.96 \\ 0.04 & 0.96 \end{bmatrix}$	$\begin{bmatrix} 0.31 & 0.69 \\ 0.31 & 0.69 \end{bmatrix}$
(4,-2,-2,2)	$\begin{bmatrix} 0.88 & 0.12 \\ 0.12 & 0.88 \end{bmatrix}$	$\begin{bmatrix} 0.69 & 0.31 \\ 0.20 & 0.80 \end{bmatrix}$	$\begin{bmatrix} 0.96 & 0.04 \\ 0.07 & 0.93 \end{bmatrix}$

$Z_t = 0.5$, yielding a state probability of $\varphi_t = 0.5$.

3 MCMC Estimation

3.1 The likelihood and prior specification

To outline the estimation of model (1), we introduce the following notation. With the time subscript t we indicate observations as of period t, while with the time superscript we indicate the entire history of observations up to time t, i.e. $y^t = (y_t, y_{t-1}, \ldots, y_1)$, and similarly for X^t, Z^t, S^t. The regression parameters are gathered into the parameter vector $\beta = (\beta_1, \ldots, \beta_K)$, where $\beta_k = (\beta_{1,k}, \ldots, \beta_{p,k})$, for $k = 1, \ldots, K$. Finally, the parameters governing the transition probabilities are denoted by $\gamma = \{\gamma_k | k \in \mathcal{K}_{-k_0}\}$ with $\gamma_k = (\gamma^z_{1k}, \ldots, \gamma^z_{Kk}, \gamma_{1k}, \ldots, \gamma_{Kk})$. All model parameters are contained in $\theta = (\beta, \gamma, \sigma^2)$, and the extended parameter vector $\psi = (\theta, S^T)$ gathers the model parameters and the unobservable state vector S^T.

Conditional on the state vector S^T, the complete data likelihood of the regression model (1) is

$$L\left(y^T | X^T, S^T, \theta\right) = \prod_{t=1}^{T} f\left(y_t | X_t, S_t, \theta\right) \tag{8}$$

with a normally distributed observation density

$$f\left(y_t | X_t, S_t, \theta\right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}\left(y_t - X_t'\beta_{S_t}\right)^2\right\} \tag{9}$$

Conditional on γ and Z_t, the prior density of the state vector factorizes

$$\pi\left(S^T | Z^T, \gamma\right) = \prod_{t=1}^{T} \pi\left(S_t | Z_t, S_{t-1}, \gamma\right) \pi\left(S_0\right) \tag{10}$$

in which the prior distribution of the initial state $\pi\left(S_0\right)$ is assumed to be uniform over the number of states: $P\left(S_0 = k\right) = 1/K$.

To complete the setup, the prior distribution of the regression parameters, the error variance and of the parameters governing the transition distribution are assumed to be independent

$$\pi\left(\theta\right) = \pi(\beta)\pi(\sigma^2)\pi(\gamma) \tag{11}$$

Conditional on the state, we face a traditional piecewise linear regression model and therefore, we may specify the usual normal-inverse Gamma prior distributions for β and σ^2, respectively:

$$\pi(\beta) = \prod_{k=1}^{K} \pi(\beta_k) = \prod_{k=1}^{K} N(b_0, B_0) \tag{12}$$

$$\pi(\sigma^2) = IG(w_0, W_0) \tag{13}$$

The prior specification in (12) assumes the state-dependent regression parameters to be independent of each other, and to follow a state-independent prior distribution. This renders the prior invariant to state permutations and allows to explore the full unconditional posterior distribution of the parameters using the random permutation sampler (Frühwirth-Schnatter 2001).

The logit specification for the transition probabilities in (3)-(4) allows to assume a normal prior distribution for the parameter γ:

$$\pi(\gamma) = \prod_{k \in \mathcal{K}_{-k_0}} \pi(\gamma_k) = \prod_{k \in \mathcal{K}_{-k_0}} N(g_0, G_0) \tag{14}$$

where also here, the equal specification across states renders the prior invariant to state permutations.

Depending on the purpose of investigation, the researcher might render the prior more informative with respect to the states. For example, there might be situations where one would like to specify coefficients which might be state-dependently negative or positive, i.e. to have different means b_{0k} a priori. Or one would expect to have states with different persistence and transition distributions, reflected in state-dependent means g_{0k}. On the other hand, one might incorporate this information into the prior specification, without explicitly assigning the information to a specific state. In that case, an invariant prior might be obtained by construction:

$$\pi(\beta) = \frac{1}{K!} \sum_{j=1}^{K!} \prod_{k=1}^{K} N\left(b_{0\rho_j(k)}, B_{0\rho_j(k)}\right) \prod_{k=1}^{K} N\left(g_{0\rho_j(k)}, G_{0\rho_j(k)}\right) \tag{15}$$

where ρ_j, $j = 1, \ldots, K!$, represent all possible $K!$ permutations of $\{1, \ldots, K\}$. In this case, the sampler described below would include a Metropolis-Hastings step to sample the coefficients. A straightforward proposal would be the state-dependent normal priors with appropriately permuted moments. In the present paper, we work with state-independent prior specifications, so we do not pursue this avenue here. The interested reader finds an application in Kaufmann and Frühwirth-Schnatter (2002).

There might also be situations where the researcher explicitly wants to assign state-dependent information to a specific state, e.g. in outlier detection analysis or when working with financial data typically having infrequent states of high volatility and low persistence, see e.g. also (Wasserman 2000). In that case, the prior will not be state-invariant any more. However, the random permutation sampler might still be applied if well-separating state-dependent information is not included for all coefficients. But then, the state-dependent prior hyperparameters have to be permuted along with the state-dependent parameters at the end of each iteration of the sampler.

3.2 The sampling scheme

The posterior distribution $\pi\left(\psi|y^T, X^T, Z^T\right)$ is obtained by combining the prior with the likelihood

$$\pi\left(\psi|y^T, X^T, Z^T\right) \propto f\left(y^T|X^T, S^T, \theta\right)\pi\left(S^T|Z^T, \gamma\right)\pi\left(\theta\right) \tag{16}$$

To obtain a sample from (16), we iterate over the following Markov chain Monte Carlo sampling steps:

(i) Sample the state indicator from $\pi\left(S^T|y^T, X^T, Z^T, \theta\right)$ by multi-move sampling

(ii) Sample the parameters governing the transition probabilities from $\pi(\gamma|S^T, Z^T)$ based on data augmentation (Frühwirth-Schnatter and Frühwirth 2010), taking into account the path-dependence implied by the Markov structure in the logit transition distribution.
Compute ξ_t, the matrices of time-varying transition probabilities which determine the posterior in (i)

(iii) Sample the remaining parameters $p(\theta_{-\gamma}|S^T, y^T, X^T)$

(iv) Permutation step: Either randomly permute all state-dependent parameters to obtain a sample from the unconditional distribution, or permute the state-dependent parameters according to a uniquely state-identifying restriction.

Step (i) is by now standard in Bayesian MCMC methods. The way we proceed is to adjust the multi-move sampler described in Chib (1996) to the time-varying Markov structure in the transition probabilities, see e.g. Frühwirth-Schnatter (2006), Algorithm 11.1 and 11.2.

Step (ii) is based on data augmentation procedures proposed in Frühwirth-Schnatter and Frühwirth (2010), the advantage of which are that, by conditioning on two auxiliary latent variables, namely the utilities (or the utility differences) and the mixture component indicators, the full conditional posterior distribution of γ can be derived and drawn from in a Gibbs step. In a first step, extending the model to the random utility model (RUM, McFadden 1974) yields a non-normal model for so-called state-dependent latent utilities,

$$\begin{aligned} S^u_{kt} &= \mathbf{Z}'_t \gamma_k + \nu_{kt}, \ \forall k \in \mathcal{K}_{-k_0} &\quad (17) \\ S^u_{k_0,t} &= \nu_{k_0,t}, \ \text{for the identification restriction } \gamma_{k_0} = 0, &\quad (18) \end{aligned}$$

where $\mathbf{Z}'_t = \left(Z_t D^{(1)}_{t-1}, Z_t D^{(2)}_{t-1}, \ldots, Z_t D^{(K)}_{t-1}, D^{(1)}_{t-1}, D^{(2)}_{t-1}, \ldots, D^{(K)}_{t-1} \right)$. The errors ν_{kt} are i.i.d over k and t, and follow a Type I extreme value distribution. Conditional on S^u_{kt}, $\forall k, t$, we could sample γ from the posterior distribution applying a Metropolis-Hastings algorithm and using a multivariate normal proposal (Scott 2011). Frühwirth-Schnatter and Frühwirth (2007) introduce an additional layer to approximate the density of ν_{kt} by a mixture of M normal components (see Frühwirth-Schnatter and Frühwirth 2007, table 1). Conditional on the components R_{kt} and the utilities S^u_{kt}, the non-normal model becomes conditionally linear

$$S^u_{kt} = \mathbf{Z}'_t \gamma_k + m_{R_{kt}} + s_{R_{kt}} \upsilon_{kt}, \quad \upsilon_{kt} \text{ i.i.d. } N(0,1), \text{ over } k, t \qquad (19)$$

Assuming a normal prior for γ_k, the conditional posterior is also normal $\gamma_k \sim N(g_k, G_k)$, with

$$\begin{aligned} G_k &= \left(\sum_{t=1}^{T} \mathbf{Z}_t \mathbf{Z}'_t / s^2_{R_{kt}} + G_0^{-1} \right)^{-1} &\quad (20) \\ g_k &= G_k \left(\sum_{t=1}^{T} \mathbf{Z}_t \left(S^u_{kt} - m_{R_{kt}} \right) / s^2_{R_{kt}} + G_0^{-1} g_0 \right) &\quad (21) \end{aligned}$$

A second approach uses the extension to a difference in random utility model (dRUM), i.e. expresses the differences in the latent utilities

$$s_{kt} = \mathbf{Z}'_t \gamma_k + \epsilon_{kt}, \ \epsilon_{kt} \sim \text{Logistic}, \ \forall k \in \mathcal{K}_{-k_0} \qquad (22)$$

where $s_{kt} = S^u_{kt} - S^u_{k_0,t}$ and $\epsilon_{kt} = \nu_{kt} - \nu_{k_0,t}$. Therefore, the errors ϵ_{kt} are not independent over k any more. Given that the parameters of the reference transition are zero, $\gamma_{k_0} = 0$, γ_k is the same as in (17). The model can further be condensed to obtain the partial dRUM representation:

$$\begin{aligned} \omega_{kt} &= S^u_{kt} - S^u_{-k,t}, \; D_t^{(k)} = I\{\omega_{kt} > 0\} \quad (23) \\ &= \mathbf{Z}_t' \gamma_k - \log(\lambda_{-k,t}) + \underbrace{\nu_{kt} - \nu_{-k,t}}_{=\epsilon_{kt}} \quad (24) \end{aligned}$$

where $S^u_{-k,t}$ indicates the maximum value of all utilities excluding $S^u_{k,t}$, $S^u_{-k,t} = \max_{j \in \mathcal{K}_{-k}} S^u_{jt}$, and the constant $\lambda_{-k,t} = \sum_{j \in \mathcal{K}_{-k}} \exp(\mathbf{Z}_t' \gamma_j)$. Given that the constant $-\log(\lambda_{-k,t})$ is independent of the coefficient γ_k, we obtain a linear regression γ_k with logistic errors. The logistic error distribution can again be approximated by a mixture of mean zero normal distributions with M components, and conditional on the component R_{kt}, the non-normal model becomes normal (see Frühwirth-Schnatter and Frühwirth 2010, table 1):

$$\tilde{\omega}_{kt} = \omega_{kt} + \log(\lambda_{-k,t}) = \mathbf{Z}_t' \gamma_k + \epsilon_{kt}, \quad \epsilon_{kt} | R_{kt} \sim N\left(0, s^2_{R_{kt}}\right) \quad (25)$$

Again, assuming a normal prior for γ_k, the posterior is normal $\gamma_k \sim N(g_k, G_k)$, with

$$\begin{aligned} G_k &= \left(\sum_{t=1}^{T} \mathbf{Z}_t \mathbf{Z}_t' / s^2_{R_{kt}} + G_0^{-1} \right)^{-1} \quad (26) \\ g_k &= G_k \left(\sum_{t=1}^{T} \mathbf{Z}_t \tilde{\omega}_{kt} / s^2_{R_{kt}} + G_0^{-1} g_0 \right) \quad (27) \end{aligned}$$

The interested reader finds a detailed derivation of the sampling scheme in appendix A.

In step (iii), we further block the parameter vector into the regression vectors $\beta = \text{vec}(\beta_1, \ldots, \beta_K)$ and σ^2. Conditional on data and S^T, the posterior of β is normal,

$$\begin{aligned} \pi(\beta) &\sim N(b, B) \\ B &= \left(\frac{1}{\sigma^2} \tilde{X}' \tilde{X} + B_0^{-1} \right)^{-1} \\ b &= B^{-1} \left(\frac{1}{\sigma^2} \tilde{X}' y + B_0^{-1} b_0 \right) \end{aligned}$$

where the rows of $\tilde{X}$ are $\tilde{X}_t = \left(X_t D_t^{(1)}, X_t D_t^{(2)}, \ldots, X_t D_t^{(K)} \right)$. If some parameters were not switching, we gather the variables with non-switching parameters in X_{1t} and the ones with switching parameters in X_{2t}. The corresponding rows of $\tilde{X}$ are then $\tilde{X}_t = \left(X_{1t}, X_{2t}D_t^{(1)}, X_{2t}D_t^{(2)}, \ldots, X_{2t}D_t^{(K)}\right)$.

The posterior of σ^2 is inverse Gamma, $IG\,(w, W)$ with $w = w_0 + 0.5T$ and $W = W_0 + 0.5\sum_{t=1}^{T}\left(y_t - \tilde{X}_t\beta\right)^2$. In case of state-dependent variances the posterior would also be inverse Gamma $IG\,(w_k, W_k)$ with $w_k = w_0 + 0.5T_k$, $T_k = \sum_{t=1}^{T} D_t^{(k)}$ and $W = W_0 + 0.5\sum_{t=1}^{T} D_t^{(k)}\,(y_t - X_t'\beta_k)^2$

To motivate step (iv), note that the model (1) is not identified with respect to the states. So, the likelihood in (8), $L\left(y^T|X^T, S^T, \theta\right)$, remains unchanged with respect to any state permutation. Under an invariant prior, the posterior will also be invariant to any state permutation $\rho = (\rho_1, \ldots, \rho_K)$

$$\pi\left(\theta, S^T|y^T, X^T, Z^T\right) = \pi\left(\rho(\theta), \rho(S^T)|y^T, X^T, Z^T\right)$$

The investigator may choose one of two options to estimate an identified model. The one most often pursued is to define a state-identifying restriction based on one of the state-dependent coefficients. In the present case, one could set a restriction on the regression coefficients or on the parameters governing the transition distribution:

$$\beta_{j,1} < \cdots < \beta_{j,K} \text{ or } \gamma_{l1} < \cdots < \gamma_{lK}, \quad j \in \{1, \ldots, p\},\ l \in \mathcal{K}_{-k_0} \tag{28}$$

Obviously, in case $K > 2$, one could also choose a combination of restrictions

$$\beta_{j,1} < \min\left(\beta_{j,2}, \ldots \beta_{j,K}\right) \text{ and } \gamma_{l2} < \cdots < \gamma_{lK} \tag{29}$$

In this case, each iteration would be terminated by re-ordering the state-dependent parameters and the states to fulfill the restriction (constrained permutation sampling) and by re-normalizing the parameters of the transition distribution to $\gamma_{k_0} = 0$.

If the investigator does not know a priori which parameter yields a unique state-identifying restriction, she may sample from the unconditional posterior by forcing the sampler to visit all posterior modes (random permutation sampling, Frühwirth-Schnatter 2001). State-identification is then obtained by post-processing the MCMC output. A unique state-identifying restriction might be obtained from inspecting the multimodal marginal posteriors of the state-dependent coefficients or by applying k-means clustering to all MCMC iterations, see Frühwirth-Schnatter (2011). A detailed description of the permutation steps is found in appendix B.

4 Illustration and evaluation

4.1 Model estimation

To illustrate the usefulness of the sampling procedures outlined in the previous section, we first use simulated data. We assume an autoregressive process y_t to depend on two exogenous variables

$$\begin{aligned} y_t &= \beta_{1S_t} x_{1t} + \beta_{2S_t} x_{2t} + \varepsilon_t \\ &\quad \varepsilon_t \sim N\left(0, \sigma^2_{S_t}\right) \end{aligned} \tag{30}$$

in which the state-dependent regression parameters are set to $\beta_1 = \{0,\ 0.8\}$ and $\beta_2 = \{0.2,\ 0.2\}$, and the state-dependent variances of the error terms to $\sigma^2 = \{0.05,\ 0.1\}$

The exogenous variables are drawn from independent normal distributions, and the Markov switching process S_t is modelled to depend on a covariate Z_t generated by an autoregressive process:

$$x_{1t}, x_{2t} \text{ i.i.d. } N(0,1) \tag{31}$$

$$Z_t = 0.8 Z_{t-1} + \eta_t,\ \ \eta_t \text{ i.i.d. } N(0, 0.5) \tag{32}$$

where the relative strong autoregressive process for Z_t is chosen to induce some persistence in the simulated Markov variable S_t. We shift Z_t to induce a threshold level different from zero. Assuming two states, $K = 2$, and $k_0 = 1$, the transition distribution writes:

$$\xi_{k2,t} = \frac{\exp\left(\gamma^z_{k2}\left(Z_t - 0.5\right) + \gamma_{k2}\right)}{1 + \exp\left(\gamma^z_{k2}\left(Z_t - 0.5\right) + \gamma_{k2}\right)} \tag{33}$$

For the parameters we assume $\gamma^z_{12} = 4$, $\gamma^z_{22} = 1$ and $\gamma_{12} = -2$, $\gamma_{22} = 2$. The specification reflects the property we think of being most intuitive in macroeconomic applications of Markov switching models (see section 2.3). When Z_t is at its threshold, the values for γ_{k2} correspond to a transition probability matrix (see also table (1))

$$\xi = \begin{bmatrix} 0.88 & 0.12 \\ 0.12 & 0.88 \end{bmatrix}$$

We simulate 400 observations, $T = 400$, and use the last 200 to estimate the model. In a first round, we work with Z_t as covariate, given that its mean is zero. We estimate the model assuming all parameters to be state-dependent under quite uninformative prior specifications. We specify for $\beta_{j,k}$ $\pi\left(\beta_{j,k}\right) \sim N(0,1/4)$, for σ_k $\pi\left(\sigma_k\right) \sim IG(2,0.25)$ and for γ_2 $\pi\left(\gamma_2\right) \sim N\left(\left[0,\ 0,\ -1,\ 1\right]', 6.25 \cdot I_4\right)$. We estimate the model using alternatively random and constrained permutation sampling. In each case, the parameters of the transition distribution are sampled using both alternatives of the auxiliary sampling schemes. The last 20,000 iterations out of a total of 50,000 under the RUM specification and out of 30,000 under the dRUM specification are used to evaluate the posterior distribution. Fewer iterations are needed in the latter case due to faster convergence and higher efficiency (see below).

Before comparing the various estimation methods, we discuss the results of the ultimately preferred procedure in terms of efficiency: Random permutation with dRUM auxiliary sampling of the transition distribution parameters. Given that the sampler is forced to visit both modes of the posterior, the marginal posterior densities of the state-dependent parameter ($\beta_{1,k}$ and σ_k^2) in figure 2, panel (a), are bi-modal and overlap for $k = 1, 2$. For γ_{k2}^z, random permutation brings about bi-modality with one mode at zero as the base category also switches between the two states. The scatter plots in figure 3, which plot the simulated regression parameters against the simulated constant transition parameters γ_{k2}, convey the same information. Obviously, $\beta_{1,k}$ is switching between states, while $\beta_{2,k}$ is apparently state-independent.

To obtain state-identification, we may re-order the simulated values according to the state-identifying restriction $\beta_{1,1} < \beta_{1,2}$ and normalize the parameters of the transition distribution choosing $k_0 = 1$ (see permutation scheme (58) in appendix B). The result of the identification step for the marginal posterior distributions is plotted in figure 2, panel (b). The re-ordered draws for $\beta_{1,k}$ are plotted in figure 3 in the right panel. The panel plots all draws inclusive of the burn-in and we observe that the sampler converges quite quickly.

The plots in figure 4 show different views on the threshold value determined by Definition 1 using the sampled values of the transition distribution parameters. The boxplot in the left panel reports a median estimate of 0.37 with a 95% highest posterior density interval ranging between 0.18 and 0.60. The right panel plots $\tilde{Z}_t$ against $\xi_{11,t}^{(m)}$ and $\xi_{22,t}^{(m)}$ implied by the simulated values for $\gamma_2^{(m)}$. The green dots plot the implied threshold level against the corresponding persistence probability of state one $\xi_{11,t}^{(m)}$.

4.2 Efficiency evaluation

The various sampling designs are compared in evaluating their inefficiency in sampling the parameters of the transition distribution, γ. The inefficiency measure (Geweke 1992) relates the variance of a hypothetical i.i.d. sampler to the sampling variance. We estimate the ratio by dividing the squared numerical standard error (an estimate of the sampling variance at frequency zero) by the posterior sampling variance of γ, $\hat{\sigma}^2_\gamma$. The square of the numerical standard error is estimated taking into account serial dependence in the sampled values:

$$\hat{S}(0) = \Omega_0 + 2\sum_{j=1}^{J}\left(1 - \frac{j}{J+1}\right)\Omega_j$$

where Ω_j is the autocovariance for lag j. For the measures summarized in table 2, we set $J = 2000$. Moreover, the measures are scaled by the number of retained iterations. We either retain all of the last 20,000 iterations or retain every 4th iteration to remove some of the autocorrelation, which leaves us with 5,000 iterations in that case. For expositional convenience, the inefficiency factors reported in table 2 are multiplied by 100.

We observe that random permutation with auxiliary sampling based on the dRUM specification shows the best performance (last two columns, top two panels). The output of the random permutation sampler shows virtually the same inefficiency irrespective of whether we use all iterations or only every 4th one. Working with every 4th iteration in the identified model, removes considerably autocorrelation in the simulated values (see figure 5), the inefficiency is roughly halved. This is not the case for the constrained permutation sampler, where inefficiency does markedly decrease only for two parameters if we retain every 4th observation. Auxiliary sampling based on the dRUM strongly outperforms auxiliary sampling based on the RUM (see also Frühwirth-Schnatter and Frühwirth (2010) for further comparisons). The inefficiency factor increases by a factor of (at least) 10 if we use the random permutation sampler. The increase in inefficiency is not as strong if we use the constrained permutation sampler. Nevertheless, the differences are considerable.

The results about the inefficiency factors are mirrored in the autocorrelation functions (ACF) of the sampled values for γ. Figure 5 plots the ACFs for the various MCMC outputs. The pictures document again the superiority of the random permutation sampler with dRUM auxiliary sampling. The autocorrelation function drops very quickly to zero for all parameters in the randomly permuted MCMC output. Retaining only every 4th iteration in the identified model also removes considerable autocorrelation in the simulated values. The same applies to constrained permutation sampling. The considerable inefficiency of RUM auxiliary sampling is revealed in the high and very slowly decreasing autocorrelation functions. In the case of constrained permutation, the posterior sample has to be thinned out considerably to remove correlation.

5 Application: The two-pillar Phillips curve

We apply the model to the same setting as in Assenmacher-Wesche and Gerlach (2008), who estimate an empirical, so-called two-pillar Phillips curve for the euro area:

$$\begin{aligned}\pi_t &= \beta_{0,S_t} + \beta_{1,S_t}\Delta\tilde{m}_t + \beta_{2,S_t}\Delta\tilde{R}_t + \beta_{3,S_t}\Delta\tilde{y}_t + \beta_{4,S_t}\hat{y}_t + \sum_{j=1}^{p}\phi_{j,S_t}\pi_{t-j} + \varepsilon_t \\ &\quad \varepsilon_t \sim \text{i.i.d } N(0,\sigma^2_{S_t}) \qquad (34)\end{aligned}$$

where π_t represents the quarterly rate of inflation, Δm_t, ΔR_t and Δy are M3 growth, the change in the government bond yield, and GDP growth, respectively. The tilde indicates that the long-run component of the respective variables enters the regression, while the hat on the output gap indicates its cyclical component. Up to p autoregressive terms are included to take into account dynamics.

Assenmacher-Wesche and Gerlach (2008) extended the traditional Phillips curve with money growth as explanatory variable to motivate its importance in determining the inflation rate, in particular the long-run prospects of inflation. The specification of the empirical Phillips curve captures the notion that long-run or low-frequency components of inflation are determined by low-frequency nominal (trend M3 growth, trend change in bond yield) and real components (trend GDP growth), its high-frequency components are determined by the high-frequency (cyclical) component of the output gap.

In a companion paper, Assenmacher-Wesche and Gerlach (2007) estimate this empirical two-pillar Phillips curve for the US, the UK and Japan. They find significant evidence for the frequency-components determination of the inflation rate in all countries, with the only exception for the low-frequency component of output growth in the UK. Both papers contributed to the discussion surrounding the first evaluation of the European Central Bank's two-pillar monetary policy strategy, which lead to a re-ordering of the first and second pillar, the so-called, respectively, monetary pillar and the economic pillar (ECB 2004). Although repeatedly, investigations based on either cross-section data (De Grauwe and Polan 2005) or country-specific data (De Santis et al. 2013, Sargent and Surico 2011) documented instabilities in money demand functions, the role of money in monetary policy and it use in inflation forecasting has remained highly debated in the euro area between the profession and academia (ECB 2008). Revisiting the two-pillar Phillips curve adds to the discussion, as it turns out that the significant link between the low frequency components of money growth and inflation evidenced in Assenmacher-Wesche and Gerlach (2008) deteriorates if shorter data series – starting in the early 80s rather than back in the 70s – are used to estimate equation (34). This evidence fits the results presented in Sargent and Surico (2011), who show that the link between US money growth and inflation is insignificant since 1984, a period during which inflation has stabilized at low levels. However, with the use of a structural model, they also show that the link might reappear any time if monetary policy were to depart from its response rule which puts a relatively large weight on inflation.

The following results also put into perspective the deterioration of the euro area two-pillar Phillips curve estimates obtained with a shorter data sample. We present three estimations of equation (34). The first estimate reproduces the results of Assenmacher-Wesche and Gerlach (2008), obtained using the long observation sample covering the period 1970-2010. The second estimate uses the data sample beginning in 1983, and yields evidence that the significance of all variables but trend M3 growth deteriorates. Nevertheless, also the importance of M3 growth apparently diminishes. In the third estimate, we allow for regime-specific parameters. It turns out that the results for one regime are very similar to those obtained for the linear specification. The results for the other regime recover, besides a considerable effect of trend M3 growth, the significance of trend GDP growth and the cyclical output gap. These periods are often lead by periods during which loan growth is persistently above trend money growth (see figure 11 below). This, and because loans are one counterpart of money, motivates why we include lagged credit growth as a covariate in the transition distribution between states. We obtain evidence that in particular the probability to switch to state 2 increases considerably when the quarterly credit growth rate is above 1.9%.

We first describe how the data are compiled, before discussing the results.

5.1 Data

Most data are retrieved from the statistical website of the European Central Bank. To obtain longer data series where necessary, we use published data on the euro area wide model and chain time series backwards by growth rates. Proceeding this way, we obtain long quarterly data series for real GDP, the harmonized index of consumer prices (HICP), and the government bond yield. They cover the period from the first quarter of 1970 to the first quarter of 2010. The historical loan series starts only in 1983. Therefore, the model estimated with time-varying transition probabilities will use data from 1983 onwards. This can also be seen as an advantage, as we can assess whether the estimate of the two-pillar Phillips curve for long time series is robust when only more recent data are available.

To obtain the low- and high-frequency components of time series, we use the HP-filter rather than extraction by frequency bands. One advantage is that no observations are lost, in particular at the end of the sample, which may be of interest if the model is used for forecasting. Moreover, comparing the extracted HP-trend with the component extracting frequencies longer than 6 years, reveals no large differences between the series. As an example, see figure 6 in which the low-frequency and the HP-trend of M3 growth are depicted. The HP-trend shows less volatility, but basically, both time series feature the same dynamics.

5.2 Results

All estimations are based on 150,000 draws from the posterior distribution, discarding the first 25,000 ones to remove dependence on initial conditions. To remove some of the autocorrelation in the draws, we retain every 10th one for posterior inference. To estimate the switching specification, we use the random permutation sampler described in section 3, and base data augmentation on the dRUM extension. State identification is obtained by post-processing the MCMC output, by re-ordering the sampled values according to a state-identifying restriction.

We work with rather uninformative prior specifications, setting $b_0 = 0$, $B_0 = 0.25$ for the regressors' coefficients and $B_0 = 0.09$ for the autoregressive coefficients. The latter specification determines the 95% interval for the autoregressive coefficients to cover the interval $(-0.6, 0.6)$. We might also work with a more uninformative prior and truncating it to the stationarity region. Note however, that with the present data, even when working with a more uninformative prior, sampling from the posterior of the autoregressive coefficient has not to be restricted as the sampled value lie within the unit circle.

The prior for the error variances is specified with $w_0 = 5$ and $W_0 = 0.25$. Finally, to obtain an intuition how to specify the prior for the coefficients of the transition distribution, we first estimate the model without prior information. The resulting posterior distributions are rather vague, but indicating that the posteriors are shifted away from zero by around 5 units. Therefore, for the final estimate, we specify $g_0 = (0,\ 0,\ -1,\ 1)'$, $G_0 = \mathrm{diag}(25,\ 25,\ 1,\ 1)$ for $\gamma = (\gamma_{12}^z,\ \gamma_{22}^z,\ \gamma_{12},\ \gamma_{22})'$.

5.2.1 Baseline estimation

The results of the baseline estimation are depicted in the first column of table 3. In parentheses, we report the 95% highest posterior density interval (HPDI) on the first line and the one-sided P-value of zero on the second line. Basically, we can reproduce the results of Assenmacher-Wesche and Gerlach (2008), even using HP- rather than frequency filtered data. In particular, trend M3 growth and the cyclical output gap are significantly positive. Taking into account the dynamics, the mean long-run effects of the variables amount to 0.75 and 0.35 for trend M3 growth and the cyclical output gap, respectively. A unit long-run effect of trend M3 growth lies in the 95% HPDI, which corresponds to the estimates presented in Assenmacher-Wesche and Gerlach (2008). In contrast to Assenmacher-Wesche and Gerlach (2008) however, we do not find a significant coefficient on trend GDP growth, and the estimate on the trend in the change of the government bond yield is positive (0.48).

When the estimation sample is restricted to begin in 1983, all the coefficients but the one on trend M3 growth turn insignificant. Trend M3 growth looses some of its long-run importance, given that the 95% HPDI does not include a unit coefficient anymore. GDP looses its significance at all, given that the effect of the cyclical output gap looses becomes insignificant. The trend change in the bond yield now obtains the expected sign, but looses its significance, too. Without going deeper into the analysis to find a structural reason for the changes in the parameter estimates, we nevertheless note that the results match those reported on decreased inflation predictability in the US. Stock and Watson (2007) show that the time series properties of inflation have changed after the mid 80s. According to their results, the variance share attributed to the permanent component in inflation has considerably decreased after 1983, the period when the Fed's policy became increasingly forward-looking and strongly reactive to inflation developments (see also Benati and Surico 2008).

5.2.2 Regime switching with time-varying transition distribution

In the switching specification, we assume all variables and the error variance to be state-dependent. The output of the random permutation sampler is depicted in figure 7. All coefficients, but the one on the government bond yield appear to be state-dependent.[3] We continue working with this specification and identify the states by k-means clustering, using the simulated values fulfilling all of the following restrictions (see the permutation steps in (57)):

$$\beta_{11} < \beta_{12},\ \beta_{21} < \beta_{22},\ \beta_{41} < \beta_{42},\ \phi_{11} > \phi_{12} \tag{35}$$

which defines state 2 as the one with a stronger effect of trend M3 growth, trend GDP growth and of the cyclical output gap on inflation. At the same time, state 2 has a lower autoregressive coefficient, in fact changes from positive to negative. The marginal posterior distributions of the state-identified parameters are depicted in the figures 8 and 9. The right-hand plot in figure 9 shows that the posterior of γ^z is shifted considerably away from zero, taking into account the diffuse prior we work with. Lagged credit growth thus affects the transition distribution of the state variable. We also see this in the first panel of table 4 where the one-sided P-value of 0 indicates significance at the 10% level of lagged credit growth in the transition distribution. The upper left panel in figure 10 presents boxplots of the parameter estimates in the transition distribution. In particular for γ^z, both 99% intervals include zero and the interval for γ^z_{22} encompasses the one of γ^z_{12}. This motivates a closer evaluation of these parameters.

The second panel in table 4 reports the log Bayes factor for various restriction sets on these parameters. Given that the restrictions are nested in the general

[3]The extension to three states, the results of which are available upon request, revealed that only two modes characterize the posterior distributions of the regression parameters and that the mean posterior state probabilities of one of the three states were lower than 0.5 over the whole observation period. This evidence confirms the two state specification.

model, we obtain these by evaluating the Savage-Dickey density ratio:

$$\log BF\left(M_0|M\right) = \log \frac{\pi\left(\gamma^z|y^T, X^T, Z^T\right)|_{\gamma^z=0}}{\pi\left(\gamma^z\right)|_{\gamma^z=0}}$$

Because a relatively diffuse prior advantages the posterior odds of zero coefficients, the table reports the log Bayes factor obtained for models estimated with increasingly informative priors on γ^z. The three remaining panels in figure 10 depict the effects that an increasingly informative prior has on the posterior update for γ^z. From table 4 we observe that increased prior information deteriorates the posterior odds of zero coefficients, both for the single and the joint evaluation. Although the numbers do not clearly favour a specification, they nevertheless turn negative, the most negative for the joint zero restriction. Interestingly, the log Bayes factor for equal coefficients, $\gamma^z_{12} = \gamma^z_{22}$, derived from an evaluation implied by the joint prior and posterior distributions, remains positive, without clearly favouring a specification, either.

The posterior inference on the state-dependent parameters and the posterior state probabilities does not change significantly if we restrict the effect of lagged credit growth to be state-independent in the transition distribution. Therefore, in the following we report the results obtained for the state-dependent covariate effects. The posterior inference on the state-identified parameters is summarized in the last two columns of table 3. Comparing the results with the ones of the linear specification for the shorter data sample, we observe that results for regime 1 are very similar to the ones obtained with the linear specification. As already mentioned, these results match those for the US on decreased inflation predictability since the mid 80s. Trend M3 growth is the only marginally relevant variable for inflation, although the effect is considerably lower than estimated for the longer sample covering the years since 1970. This fits the evidence presented for the US in Sargent and Surico (2011). In regime 2 however, trend GDP growth and the cyclical output gap are significant determinants of inflation. On impact, in this regime, trend M3 growth also has a stronger effect on inflation than in regime 1. Even the unit coefficient is included in the 95% HPDI.

Figure 11 depicts the mean posterior probabilities of state 2, $P\left(S_t = 2|y^T, X^T, Z^T\right)$. Preceding each longer lasting episode during which state 2 has been relevant – from 1988 third quarter, 2000 second quarter and 2007 fourth quarter onwards –, quarterly loan growth increased to levels above 1.9% – the threshold level composed of an average loan growth rate of 1.7% and of 0.2% inferred according to Definition 1 of subsection 2.2. In all these periods, inflation gets more volatile (see figure 9, left panel) and reaches peak levels during the first two periods and a historical trough in the last period. This increased volatility in the inflation rate might be the reason why the close link between trend M3 growth and inflation is unmasked during these periods. Indeed, when state 2 prevails, trend M3 growth is either at a relatively high level (first period), is on an increasing path (second period) or is sharply decreasing (third period). On the other hand, when state 1 is prevailing, corresponding to periods during which loan growth remained subdued relatively to trend M3 growth, inflation displays a low volatility and is either decreasing or remaining at a low levels. These are typically the periods in which the relationship between trend M3 growth and inflation might be masked and, moreover, render inflation less predictable (Benati and Surico 2008).

The importance of loan growth as a signal for regime changes is also reflected in figure 12. The figure plots the median posterior transition probabilities, and we observe that the persistence of state 1 decreases below 0.8 and the persistence of state 2 increases nearly to unity in particular ahead of state 2 periods mentioned above.

6 Conclusion

The present paper proposes to use a multinomial logit model to parameterize a K-state regime switching process with time-varying transition distribution. To derive a Bayesian sampling scheme, the multinomial logit model is extended to a random utility and a difference in random utility model. In a second layer, the non-normal but linear models are approximated by mixture of normals to derive the full conditional posterior distributions of the coefficients governing the transition distributions. Identification issues are addressed with the random permutation sampler, which, in combination with the model extension to the difference in utility model, performs best in terms of efficiency.

The model estimate can be used to discriminate the Markov switching specification with time-varying transition probabilities against related alternatives like Markov switching models with constant transition distribution or mixture models with time-varying weights. We give a definition to determine a relevant threshold of the covariate influencing the transition distribution.

The method is applied to estimate the empirical two-pillar Phillips curve for the euro area (Assenmacher-Wesche and Gerlach 2008), in which the trend components of M3 growth, real GDP growth and of the government bond yield change, and the cyclical component of the output gap are the explanatory variables for headline inflation. Using the nonlinear specification for quarterly data covering the period 1983 to 2010, we are able to recover first evidence provided for data series going back to the 1970s, which would not be the case using the original linear specification.

Although the sampling scheme is derived within the univariate framework, it readily can be included in multivariate approaches like vector autoregressive systems or panel data analysis.

References

Albert, J. H. and S. Chib (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association 88*, 669–679.

Amisano, G. and G. Fagan (2010). Money growth and inflation: A regime switching approach. Working Paper 1207, ECB.

Assenmacher-Wesche, K. and S. Gerlach (2007). Money at low frequencies. *Journal of the European Economic Association 5*, 534–542.

Assenmacher-Wesche, K. and S. Gerlach (2008). Interpreting euro area inflation at high and low frequencies. *European Economic Review 52*, 964–986.

Benati, L. and P. Surico (2008). Evolving U.S. monetary policy and the decline of inflation predictability. *Journal of the European Economic Association 6*, 634–646.

Billio, M. and R. Casarin (2011). Beta autoregressive transition Markov-switching models for business cycle analysis. *Studies in Nonlinear Dynamics & Econometrics 15*, 1–32.

Billio, M., R. Casarin, F. Ravazzolo, and H. K. van Dijk (2013). Interactions between eurozone and US booms and busts: A Bayesian panel Markov-switching VAR model. Working Paper 2013/20, Norges Bank.

Billio, N. and R. Casarin (2010). Identifying business cycle turning points with sequential monte carlo: An online and real-time application to the euro area. *Journal of Forecasting 29*, 145–167.

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

Chib, S. (1998). Estimation and comparison of multiple change-point models. *Journal of Econometrics 86*, 221–241.

Chib, S. and M. Dueker (2004). Non-Markovian regime switching with endogenous states and time-varying state strengths. Working Paper 2004-030A, The Federal Reserve Bank of St. Louis.

De Grauwe, P. and M. Polan (2005). Is inflation always and everywhere a monetary phenomenon? *Scandinavian Journal of Economics 107*, 239–259.

De Santis, R., C. A. Favero, and B. Roffia (2013). Euro area money demand and international portfolio allocation: A contribution to assessing risks to price stability. *Journal of International Money and Finance 32*.

ECB (2004). *The Monetary Policy of the ECB* (second ed.). European Central Bank: Frankfurt a. Main.

ECB (2008). *The Role of Money – Money and Monetary Policy in the Twenty-First Century*. ECB: Frankfurt a. Main.

Filardo, A. J. (1994). Business-cycle phases and their transitional dynamics. *Journal of Business & Economic Statistics 12*, 299–308.

Filardo, A. J. and S. F. Gordon (1998). Business cycle durations. *Journal of Econometrics 85*, 99–123.

Frühwirth-Schnatter, S. (2001). MCMC estimation of classical and dynamic switching and mixture models. *Journal of the American Statistical Association 96*, 194–209.

Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*. Springer.

Frühwirth-Schnatter, S. (2011). Dealing with label switching under model uncertainty. In C. P. R. Kerrie L. Mengersen and D. M. Titterington (Eds.), *Mixtures: Estimation and Applications*, pp. 213–240. The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom: John Wiley & Sons Ltd.

Frühwirth-Schnatter, S. and R. Frühwirth (2007). Auxiliary mixture sampling with applications to logistic models. *Computational Statistics and Data Analysis 51*, 3509–3528.

Frühwirth-Schnatter, S. and R. Frühwirth (2010). Data augmentation and MCMC for binary and multinomial logit models. In T. Kneib and G. Tutz (Eds.), *Statistical Modelling and Regression Structures – Festschrift in Honour of Ludwig Fahrmeir*, pp. 111–132. Heidelberg: Physica-Verlag.

Frühwirth-Schnatter, S., C. Pamminger, A. Weber, and R. Winter-Ebmer (2012). Labor market entry and earnings dynamics: Bayesian inference using mixture-of-experts Markov chain clustering. *Journal of Applied Econometrics 27*, 1116–1137.

Gaggl, P. and S. Kaufmann (2014, January). The cyclical component of labor market polarization and jobless recoveries in the U.S. mimeo.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In J. Bernardo, J. Berger, A. Dawid, and A. Smith (Eds.), *Bayesian Statistics 4*, pp. 169–193. Oxford: Oxford University Press.

Geweke, J. and Y. Jiang (2011). Inference and prediction in a multiple-structural-break model. *Journal of Econometrics 163*, 172–185.

Hamilton, J. D. and M. T. Owyang (2012). The propagation of regional recessions. *Review of Economics & Statistics 94*, 935–947.

Hamilton, J. D., D. F. Waggoner, and T. Zah (2007). Normalization in econometrics. *Econometric Reviews 26*, 221–252.

Hausman, J. A. and D. A. Wise (1978). A conditional probit model for qualitative choice: Discrete decisions recognizing interdependence and heterogeneous preferences. *Econometrica 46*, 403–426.

Holmes, C. C. and L. Held (2006). Bayesian auxiliary variable models for binary and multinomial regression. *Bayesian Analysis 1*, 145–168.

Imai, K. and D. A. van Dyk (2005). A Bayesian analysis of the multinomial probit model using marginal data augmentation. *Journal of Econometrics 124*, 311–334.

Kaufmann, S. (2010). Dating and forecasting turning points by Bayesian clustering with dynamic structure: A suggestion with an application to Austrian data. *Journal of Applied Econometrics 25*, 309–344.

Kaufmann, S. and S. Frühwirth-Schnatter (2002). Bayesian analysis of switching ARCH models. *Journal of Time Series Analysis 23*, 425–458.

Kim, C.-J. and C. R. Nelson (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 & Statistics 80*, 188–201.

Kim, C.-J., J. Piger, and R. Startz (2008). Estimation of Markov regime switching regression models with endogenous switching. *Journal of Econometrics 143*, 263–273.

Koop, G. and S. M. Potter (2007). Estimation and forecasting in models with multiple breaks. *Review of Economics & Statistics 74*, 763–789.

Lu, Z.-Q. and L. M. Berliner (1999). Markov switching time series models with application to a daily runoff series. *Water Resources Research 35*, 523–534.

McCulloch, R. E., N. G. Polson, and P. E. Rossi (2000). A Bayesian analysis of the multinomial probit model with fully identified parameters. *Journal of Econometrics 99*, 173–193.

Meligkotsidou, L. and P. Dellaportas (2011). Forecasting with non-homogenous hidden Markov models. *Statistics and Computing 21*, 439–449.

Nobile, A. (1998). A hybrid Markov chain for the Bayesian analysis of the multinomial probit model. *Statistics and Computing 8*, 229–242.

Paap, R. and H. K. van Dijk (2003). Bayes estimates of Markov trends in possibly cointegrated series: An application to US consumption and income. *Journal of Business & Economic Statistics 21*, 547–563.

Pesaran, M. H., D. Pettenuzzo, and A. Timmerman (2007). Forecasting time series subject to multiple structural breaks. *Review of Economics & Statistics 74*, 763–789.

Sargent, T. J. and P. Surico (2011). Two illustrations of the quantity theory of money: Breakdowns and revivals. *The American Economic Review 101*, 109–128.

Scott, S. L. (2011). Data augmentation, frequentistic estimation, and the Bayesian analysis of multinomial logit models. *Statistical Papers 52*, 87–109.

Sims, C. A., D. F. Waggoner, and T. Zha (2008). Methods for inference in large multiple-equation markov-switching models. *Journal of Econometrics 146*, 255–274.

Stock, J. H. and M. W. Watson (2007). Why has U.S. inflation become harder to forecast? *Journal of Money Credit and Banking 39*, 3–33.

Teräsvirta, T. and H. M. Anderson (1992). Characterizing nonlinearities in business cycles using smooth transition autoregressive models. *Journal of Applied Econometrics 7*, S119–S136.

Wasserman, L. (2000). Asymptotic inference for mixture models using data-dependent priors. *Journal of the Royal Statistical Society. Series B (Statistical Methodology) 62*, 159–180.

A Auxiliary mixture sampling of γ

Given that so far regime switching models with time varying probabilities usually have been parameterized using the probit distribution (Filardo 1994, Filardo and Gordon 1998), we derive in detail the two sampling schemes for the logit model (3)-(4). Basically, step (ii) of the sampling scheme outlined in section 3 consists of three sub-steps. The derivations follow (Frühwirth-Schnatter and Frühwirth 2010) and are adjusted to the path-dependent structure of the Markov specification.

A.1 Data augmentation for RUM

In the following, the three sampling steps leading to a draw from $\pi\left(\gamma|S^T, Z^T\right)$, step (ii) in section 3.2, are described in more detail.

Step (ii.a): Sample the utilities S^u_{kt} *from* $\pi\left(S^{u,KT}|S^T,\gamma\right) = \prod_{t=1}^{T}\pi\left(S^u_{1t},\ldots,S^u_{Kt}|S^T,\gamma\right)$
To sample the utilities

$$\begin{aligned} S^u_{kt} &= \mathbf{Z}_t'\gamma_k + \nu_{kt},\ \forall k \in \mathcal{K}_{-k_0} \\ S^u_{k_0 t} &= \nu_{k_0 t},\ \text{implied by the identification restriction } \gamma_{k_0} = 0, \end{aligned} \tag{36}$$

conditional on the state variable S^T, we first note that the maximal utility should obtain for the observed state,

$$S^u_{jt} = \max_{k\in\mathcal{K}} S^u_{kt},\ \text{if } S_t = j$$

Therefore, $\exp\left(-S^u_{jt}\right)$ is the minimum value among all values $\exp\left(-S^u_{kt}\right)$ and

$$\exp\left(-S^u_{jt}\right) \sim \mathcal{E}\left(\sum_{k=1}^{K}\lambda_{kt}\right) \tag{37}$$

where $\mathcal{E}$ represents the exponential distribution and $\lambda_{kt} = \exp\left(\mathbf{Z}_t'\gamma_k\right)$.[4]

[4]The exponential distribution is implied by the Type I extreme value distribution of ν_{kt} and from the fact that the minimum of exponentially distributed variables follows again an exponential distribution:

$$\begin{aligned} \exp\left(-S^u_{kt}\right) &\sim \mathcal{E}\left(\lambda_{kt}\right), \\ \min_{k\in\mathcal{K}}\exp\left(-S^u_{kt}\right) &\sim \mathcal{E}\left(\sum_{k=1}^{K}\lambda_{kt}\right), \end{aligned}$$

Given the minimum, all other utilities are conditionally independent and the posterior factorizes:

$$\pi\left(S^u_{1t},\ldots,S^u_{Kt}|S_t=j,\gamma\right) = \pi\left(S^u_{jt}|S_t=j,\gamma\right)\prod_{k\in\mathcal{K}_{-j}}\pi\left(S^u_{kt}|S_t=j,\gamma\right). \tag{38}$$

The distribution $\pi\left(S^u_{jt}|S_t=j,\gamma\right)$ is given by (37) and implies

$$\exp\left(-S^u_{kt}\right) = \exp\left(-S^u_{jt}\right)+\chi_{kt}, \qquad \chi_{kt}\sim\mathcal{E}\left(\lambda_{kt}\right),\ \forall k\in\mathcal{K}_{-j} \tag{39}$$

for $\pi\left(S^u_{kt}|S_t=j,k\neq j,\gamma\right)$. To sample S^u_{kt} for each $t=1,\ldots,T$, we sample K independent uniform random numbers W_t and $V_{2t},\ldots,V_{Kt}$ and obtain:

$$S^u_{kt} = -\log\left(-\frac{\log\left(W_t\right)}{\sum_{l=1}^K\lambda_{lt}}-\frac{\log\left(V_{kt}\right)}{\lambda_{kt}}I_{\{S_t\neq k\}}\right) \tag{40}$$

Step (ii.b): Sample the components R_{kt} from $\pi\left(R^{KT}|S^{u,KT},\gamma\right)$
Conditional on S^u_{kt}, the component indicator R_{kt} is sampled from:

$$P\left(R_{kt}=r|S^u_{kt},\gamma_k\right)\propto\frac{w_r}{s_r}\exp\left\{-\frac{1}{2}\left(\frac{S^u_{kt}-\mathbf{Z}_t'\gamma_k-m_r}{s_r}\right)^2\right\},\ k\in\mathcal{K}_{-k_0} \tag{41}$$

where $r=1,\ldots,10$, and the respective component's mean m_r, standard deviation s_r and weight w_r, are taken from Frühwirth-Schnatter and Frühwirth (2007), Table 1.

Step (ii.c): Sample γ from $\pi\left(\gamma|S^{u,KT},R^{KT}\right)$
Finally, given all utilities $S^{u,KT}=\left(S^u_{11},\ldots,S^u_{K1},\ldots,S^u_{KT}\right)$ and all component indicators $R^{KT}=\left(R_{11},\ldots,R_{K1},\ldots,R_{KT}\right)$, we obtain a linear regression model for the parameters governing the transition probabilities to each state $k,\ \ k\in\mathcal{K}_{-k_0}$:

$$S^u_{kt}=\mathbf{Z}_t'\gamma_k+m_{R_{kt}}+s_{R_{kt}}\upsilon_{kt}, \quad \upsilon_{kt}\sim N(0,1) \tag{42}$$

Assuming a normal prior for γ_k, $\pi\left(\gamma_k\right)=N(g_0,G_0)$, conditional on $S^{u,KT}$ and R^{KT} the posterior is normal, too:

$$\pi\left(\gamma_k|S^{u,T}_k,R^T_k\right) = N\left(g_k,G_k\right),\ \forall k\in\mathcal{K}_{-k_0} \tag{43}$$

$$G_k=\left(\sum_{t=1}^T\mathbf{Z}_t\mathbf{Z}_t'/s^2_{R_{kt}}+G_0^{-1}\right)^{-1} \tag{44}$$

$$g_k=G_k\left(\sum_{t=1}^T\mathbf{Z}_t\left(S^u_{kt}-m_{R_{kt}}\right)/s^2_{R_{kt}}+G_0^{-1}g_0\right) \tag{45}$$

A.2 Data augmentation for the dRUM

The steps described in the following yield a draw from $\pi\left(\gamma|S^T, Z^T\right)$ based on the dRUM representation.

Step (ii.a): Sample the utility differences from
$\pi\left(\omega^{KT}|S^T,\gamma\right) = \prod_{k\in\mathcal{K}_{-k_0}} \pi\left(\omega_{k1},\ldots,\omega_{kT}|S^T,\gamma\right)$
The dRUM extension expresses the multinomial logit model as differences in the latent utilities (36)

$$s_{kt} \;=\; \mathbf{Z}_t'\gamma_k + \epsilon_{kt},\;\; \epsilon_{kt} \sim \text{Logistic},\;\; \forall k \in \mathcal{K}_{-k_0} \tag{46}$$

where $s_{kt} = S^u_{kt} - S^u_{k_0 t}$ and $\epsilon_{kt} = \nu_{kt} - \nu_{k_0 t}$. Given that the parameters of the reference transition are zero, $\gamma_{k_0} = 0$, γ_k is the same as in (36). Working with this representation would be quite involving because, in contrast to the error terms ν_{kt} in (36), the error terms ϵ_{kt} in (46) are not independent any more across states. Therefore, Frühwirth-Schnatter and Frühwirth (2010) consider a partial representation of the dRUM model, which relies on the observation that

$$S_t = k \Leftrightarrow S^u_{kt} > S^u_{-k,t},\;\; S^u_{-k,t} = \max_{j\in\mathcal{K}_{-k}} S^u_{jt} \tag{47}$$

i.e. that state k is observed if S^u_{kt} is larger than the maximum of all other utilities. For all states but the reference state we define the latent difference utilities ω_{kt} and the binary observation $D_t^{(k)}$:

$$\omega_{kt} = S^u_{kt} - S^u_{-k,t},\;\; D_t^{(k)} = I\{S_t = k\},\;\; \forall k \in \mathcal{K}_{-k_0} \tag{48}$$

Given the multinomial logit model for S_t, ω_{kt} has an explicit distributional form. Recall that (see footnote 4)

$$\exp\left(-S^u_{-k,t}\right) \sim \mathcal{E}\left(\sum_{j\in\mathcal{K}_{-k}} \lambda_{jt}\right) \tag{49}$$

where $\lambda_{jt} = \exp\left(\mathbf{Z}_t'\gamma_j\right)$ and define $\lambda_{-k,t} = \sum_{j\in\mathcal{K}_{-k}} \lambda_{jt}$. We then can write $S^u_{-k,t} = \log\left(\lambda_{-k,t}\right) + \nu_{-k,t}$, where $\nu_{-k,t}$ follows an EV distribution. Thus, the multinomial logit model has the partial dRUM representation

$$\begin{aligned} \omega_{kt} \;&=\; S^u_{kt} - S^u_{-k,t} = \mathbf{Z}_t'\gamma_k - \log\left(\lambda_{-k,t}\right) + \nu_{k,t} - \nu_{-k,t} \\ &=\; \mathbf{Z}_t'\gamma_k - \log\left(\lambda_{-k,t}\right) + \epsilon_{k,t},\;\; D_t^{(k)} = I\{S_t = k\} \end{aligned} \tag{50}$$

where $\nu_{k,t}$ and $\nu_{-k,t}$ are i.i.d. and follow an EV distribution, and $\epsilon_{k,t}$ follows a logistic distribution. The constant $-\log\left(\lambda_{-k,t}\right)$ in (50) depends only on the parameters γ_{-k}. Therefore, given $\omega_k^T = (\omega_{k1}, \ldots, \omega_{kT})$ and γ_{-k}, we obtain a linear regression with parameter γ_k and logistic errors.

The sub-sampling steps can now be outlined explicitly. For each state k, we first sample the latent utility differences ω_k^T from logistic distributions.[5] Across k, we sample independently T values W_{kt} from a uniform distribution $W_{kt} \sim U[0,\ 1]$ and obtain

$$\omega_{kt} = \mathbf{Z}_t' \gamma_k - \log(\lambda_{-k,t}) + F_\epsilon^{-1}\left(D_t^{(k)} + W_{kt}\left(1 - D_t^{(k)} - \pi_{kt}\right)\right) \tag{51}$$

where $\pi_{kt} = P\left(D_t^{(k)} = 1|\gamma\right) = 1 - F_\epsilon\left(-\mathbf{Z}_t'\gamma_k + \log(\lambda_{-k,t})\right) \propto \lambda_{kt}/\lambda_{-k,t}$; $F_\epsilon(p)$ represents the cumulative distribution function of the logistic distribution, and $F_\epsilon^{-1}(p) = \log(p) - \log(1-p)$ its inverse.

Step (ii.b) Sample the components R^{KT} from $\pi\left(R^{KT}|\omega^{KT}, \gamma\right)$
Given ω^{KT}, the posterior of γ_k is derived based on (50), approximating the logistic distribution of the errors ϵ_{kt} by a mixture of normal distributions with M components. The components R_{kt} are drawn from a multinomial distribution

$$P(R_{kt} = r|\omega_{kt}, \gamma_k) \propto \frac{w_r}{s_r} \exp\left\{-\frac{1}{2}\left(\frac{\omega_{kt} + \log(\lambda_{-k,t}) - \mathbf{Z}_t'\gamma_k}{s_r}\right)^2\right\} \tag{52}$$

where $r = 1, \ldots, 6$, and the respective component's standard deviation s_r and weight w_r, are taken from Frühwirth-Schnatter and Frühwirth (2010), Table 1.

Step (ii.c): Sample γ from $\pi\left(\gamma|\omega^{KT}, R^{KT}\right)$
Conditional on the components R_k^T, model (50) becomes normal in γ_k:

$$\tilde{\omega}_{kt} = \omega_{kt} + \log(\lambda_{-k,t}) \quad = \quad \mathbf{Z}_t'\gamma_k + \epsilon_{kt}, \quad \epsilon_{kt}|R_{kt} \sim N\left(0, s_{R_{kt}}^2\right) \tag{53}$$

Assuming a normal prior for γ_k, $\pi(\gamma_k) = N(g_0, G_0)$, conditional on ω_k^T and R_k^T the posterior is normal, too:

$$\pi\left(\gamma_k|\omega_k^T, R_k^T\right) \quad = \quad N(g_k, G_k) \tag{54}$$

$$G_k = \left(\sum_{t=1}^T \mathbf{Z}_t\mathbf{Z}_t'/s_{R_{kt}}^2 + G_0^{-1}\right)^{-1} \tag{55}$$

$$g_k = G_k\left(\sum_{t=1}^T \mathbf{Z}_t\tilde{\omega}_{kt}/s_{R_{kt}}^2 + G_0^{-1}g_0\right) \tag{56}$$

[5] $\omega_{kt}|S^T, \gamma_k$ follows a logistic distribution truncated to $[0,\ \infty)$ if $S_t = k$, and truncated to $(-\infty,\ 0]$ if $S_t \neq k$.

B Model identification

A more detailed description of the permutation step (iv) in the sampling scheme outlined in section 3.2 is given here, because the multinomial logit specification of the transition probabilities has a path-dependent structure, i.e. depends not only on the current state but also on the past state. Identification in mixture of experts models has also been considered in Frühwirth-Schnatter et al. (2012), without path dependence in the logit classification distribution, however.

Using the constrained permutation sampler, the sampled values are re-ordered after each iteration according to the state-identifying restriction

$$
\begin{aligned}
&\text{for the state-dependent parameters and the states} \\
&\beta_k^{(m)} := \beta_{\rho(k)}^{(m)},\ S^{T,(m)} := \rho(S^{T,(m)}) \\
&\text{for the state-dependent transition parameters} \\
&\tilde{\gamma}_k^{(m)} := \left(\gamma_{\rho(k),\rho(k)}^{z(m)} \gamma_{\rho(k),\rho(k)}^{(m)}\right),\ \text{with } \gamma_1 = 0 \\
&\gamma_k^{(m)} := \tilde{\gamma}_k^{(m)} - \tilde{\gamma}_1^{(m)}
\end{aligned}
\tag{57}
$$

For example, in case $K = 2$, for γ this would amount to:

$$
\gamma = \begin{bmatrix} 0 & \gamma_{12}^z \\ 0 & \gamma_{22}^z \\ 0 & \gamma_{12} \\ 0 & \gamma_{22} \end{bmatrix}, \tilde{\gamma} := \begin{bmatrix} \gamma_{22}^z & 0 \\ \gamma_{12}^z & 0 \\ \gamma_{22} & 0 \\ \gamma_{12} & 0 \end{bmatrix}, \gamma := \begin{bmatrix} 0 & -\gamma_{22}^z \\ 0 & -\gamma_{12}^z \\ 0 & -\gamma_{22} \\ 0 & -\gamma_{12} \end{bmatrix}
$$

Note that the normalization $\gamma_k := \tilde{\gamma}_k - \tilde{\gamma}_1$ is important here to keep the same reference state across simulations. If random permutation sampling is chosen to visit all modes of the posterior, the states, the state-dependent parameters are randomly permuted in step (iv) of the sampler. For a given permutation ρ at iteration m, we permute:

$$
\begin{aligned}
&\text{the state-dependent parameters and priors, states} \\
&\beta_k^{(m)} := \beta_{\rho(k)}^{(m)},\ S^{T,(m)} := \rho(S^{T,(m)}) \\
&\text{state-dependent transition parameters} \\
&\gamma_k^{(m)} := \left(\gamma_{\rho(k),\rho(k)}^{z(m)} \gamma_{\rho(k),\rho(k)}^{(m)}\right),\ \text{with } \gamma_{k_0} = 0
\end{aligned}
\tag{58}
$$

In this case, the normalization takes place after post-processing the MCMC output, i.e. after re-ordering the sampled values according to a restriction:

$$
\gamma_k^{(m)} := \gamma_k^{(m)} - \gamma_{k_0}^{(m)},\ \forall k
$$

Note that, if state-dependent priors are specified, the hyperparameters have to be permuted accordingly.

C Tables

Table 2: Simulated data. Inefficiency factors for γ. Scaled by the number of retained iterations, and multiplied by 100 for expositional convenience. The autocovariance at zero frequency is estimated taking into account 2,000 autocovariances.

		Auxiliary sampling based on			
		RUM		dRUM	
		Iterations retained		Iterations retained	
Random permutation:		all[(a)]	every 4th	all[(b)]	every 4th
– unidentified model	γ^z_{12}	0.24	0.16	0.02	0.01
	γ^z_{22}	0.51	0.56	0.04	0.01
	γ_{12}	0.22	0.15	0.02	0.01
	γ_{22}	0.28	0.37	0.03	0.01
– identified model	γ^z_{12}	6.18	7.31	0.62	0.36
	γ^z_{22}	2.48	1.61	0.20	0.12
	γ_{12}	6.49	9.03	0.58	0.32
	γ_{22}	1.11	0.63	0.11	0.05
Constrained permutation		all	every 4th	all	every 4th
– identified model	γ^z_{12}	3.22	4.51	0.97	0.57
	γ^z_{22}	2.40	1.31	0.33	0.33
	γ_{12}	3.93	5.93	0.96	0.53
	γ_{22}	1.21	1.79	0.18	0.19

(a) The last 20,000 of a total of 50,000 iterations.
(b) The last 20,000 of a total of 30,000 iterations.

Table 3: Two-pillar Phillips curve. Posterior mean effects. 95% highest posterior density interval and P-value of 0 (one-sided) in parentheses.

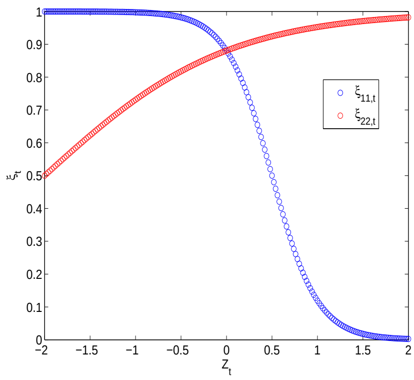
	No switching		Regime switching	
	1970Q2-2010Q1	1983Q1-2010Q1	1983Q1-2010Q1	
	3 AR lags	1 AR lag	1 AR lag	
Impact effect			Regime 1	Regime 2
trend M3 growth	0.17	0.25	0.14	0.66
	(0.02 – 0.33)	(0.08 – 0.41)	(-0.02 – 0.30)	(0.27 – 1.01)
	(0.02)	(0.00)	(0.04)	(0)
trend GDP growth	0.01	-0.05	-0.10	0.40
	(-0.20 – 0.26)	(-0.27 – 0.18)	(-0.33 – 0.13)	(0.01 – 0.82)
	(0.45)	(0.35)	(0.21)	(0.02)
trend change in gov. bond yield	0.48	-0.36	-0.40	0.08
	(-0.06 – 1.03)	(-0.95 – 0.28)	(-1.11 – 0.33)	(-0.75 – 0.96)
	(0.04)	(0.13)	(0.14)	(0.41)
cyclical output gap	0.07	0.02	-0.02	0.26
	(0.03 – 0.12)	(-0.03 – 0.08)	(-0.09 – 0.05)	(0.10 – 0.39)
	(0.00)	(0.20)	(0.28)	(0.00)
Long run effects				
trend M3 growth	0.75	0.53	0.43	0.53
	(0.19 – 1.34)	(0.21 – 0.83)	(-0.10 – 0.89)	(0.26 – 0.77)
	(0.02)	(0.00)	(0.04)	(0)
trend GDP growth	0.13	-0.09	-0.30	0.33
	(-1.00 – 1.41)	(-0.59 – 0.39)	(-1.08 – 0.48)	(-0.01 – 0.70)
	(0.45)	(0.35)	(0.20)	(0.02)
trend change in gov. bond yield	2.37	-0.75	-1.13	0.07
	(-0.42 – 5.92)	(-1.98 – 0.63)	(-3.43 – 1.04)	(-0.67 – 0.83)
	(0.04)	(0.13)	(0.14)	(0.41)
cyclical output gap	0.35	0.05	-0.07	0.20
	(0.10 – 0.66)	(-0.08 – 0.17)	(-0.34 – 0.16)	(0.10 – 0.31)
	(0.00)	(0.20)	(0.28)	(0.00)

Table 4: Posterior evaluation of γ_2, for $G^{-1}_{0,\gamma^z} = 0.04$ in the first panel and for increasingly informative prior specification in the second panel.

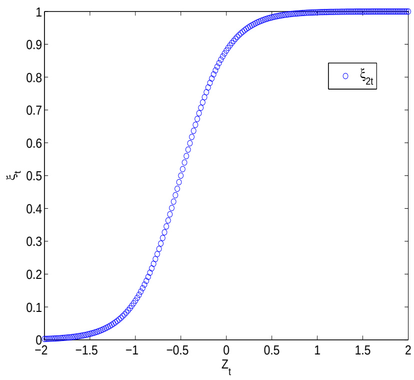
	γ^z_{12}	γ^z_{22}	γ_{12}	γ_{22}
Mean	0.97	3.23	-1.96	1.05
P-value of zero (one sided)	(0.10)	(0.10)	(0.00)	(0.09)
	$\log BF(M_0 \vert M)$, $M_0 =$			
G^{-1}_{0,γ^z}	$\gamma^z_{12} = 0$	$\gamma^z_{22} = 0$	$\gamma^z_{12} = \gamma^z_{22} = 0$	$\gamma^z_{12} = \gamma^z_{22}$
0.04	2.49	0.29	1.24	1.49
0.16	0.86	-0.07	-0.41	1.11
1.00	-0.66	-0.25	-1.44	0.47

D Figures

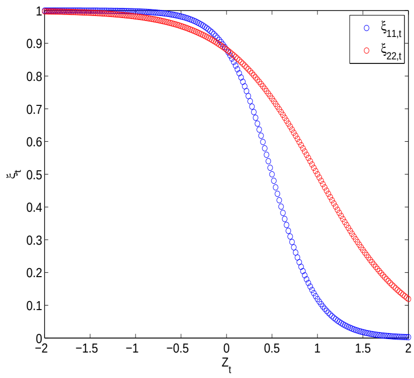
Figure 1: Some examples: Nonlinear effect of the covariate on the state persistence



(a) $\gamma = (4, 1, -2, 2)$



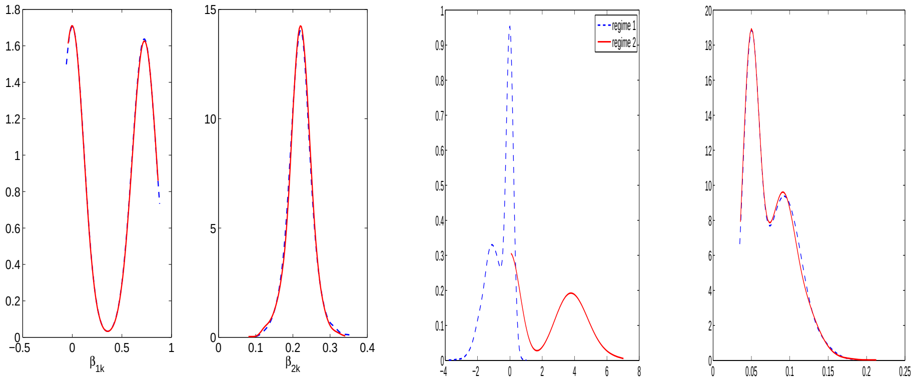
(b) $\gamma = (4, 4, 2, 2)$



(c) $\gamma = (4, -2, -2, 2)$

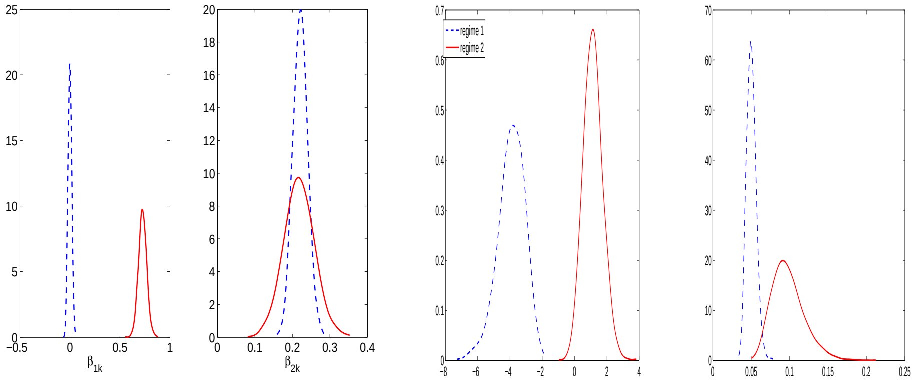
Figure 2: Random permutation with dRUM auxiliary sampling for the transition distribution. Marginal distribution of selected parameters.

(a) Simulated values obtained from the random permutation sampler



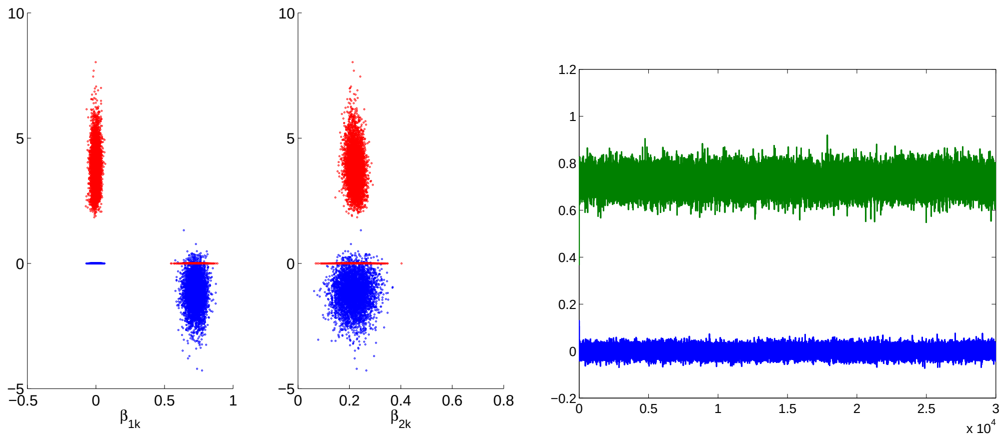
$\beta_{1,k}$ $\beta_{2,k}$ γ_{k2} σ_k^2

(b) Simulated values re-ordered according to $\beta_{1,1} < \beta_{1,2}$



$\beta_{1,k}$ $\beta_{2,k}$ γ_{k2} σ_k^2

Figure 3: Random permutation with dRUM auxiliary sampling for the transition distribution. Simulated values obtained from the random permutation sampler, scatter plots of regressions parameters against constant transition parameters γ_{k2}, $k = 1, 2$.



$\beta_{1,k}$ $\beta_{2,k}$ re-ordered $\beta_{1,k}$

Figure 4: Recovering the threshold (0.5 in simulated data). Boxplot of values implied by sampled transition parameters, marginal distribution of the threshold level (left and middle panel). Scatter plots of Z_t against $\xi_{11,t}^{(m)}$ (blue), $\xi_{22,t}^{(m)}$ (red) implied by the mth simulated parameter value γ_2. The filled dots plot the threshold level against $\xi_{11,t}^{(m)}$ (green)

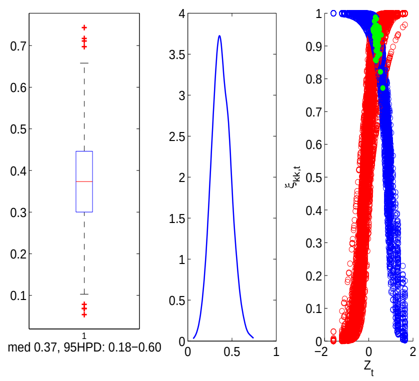
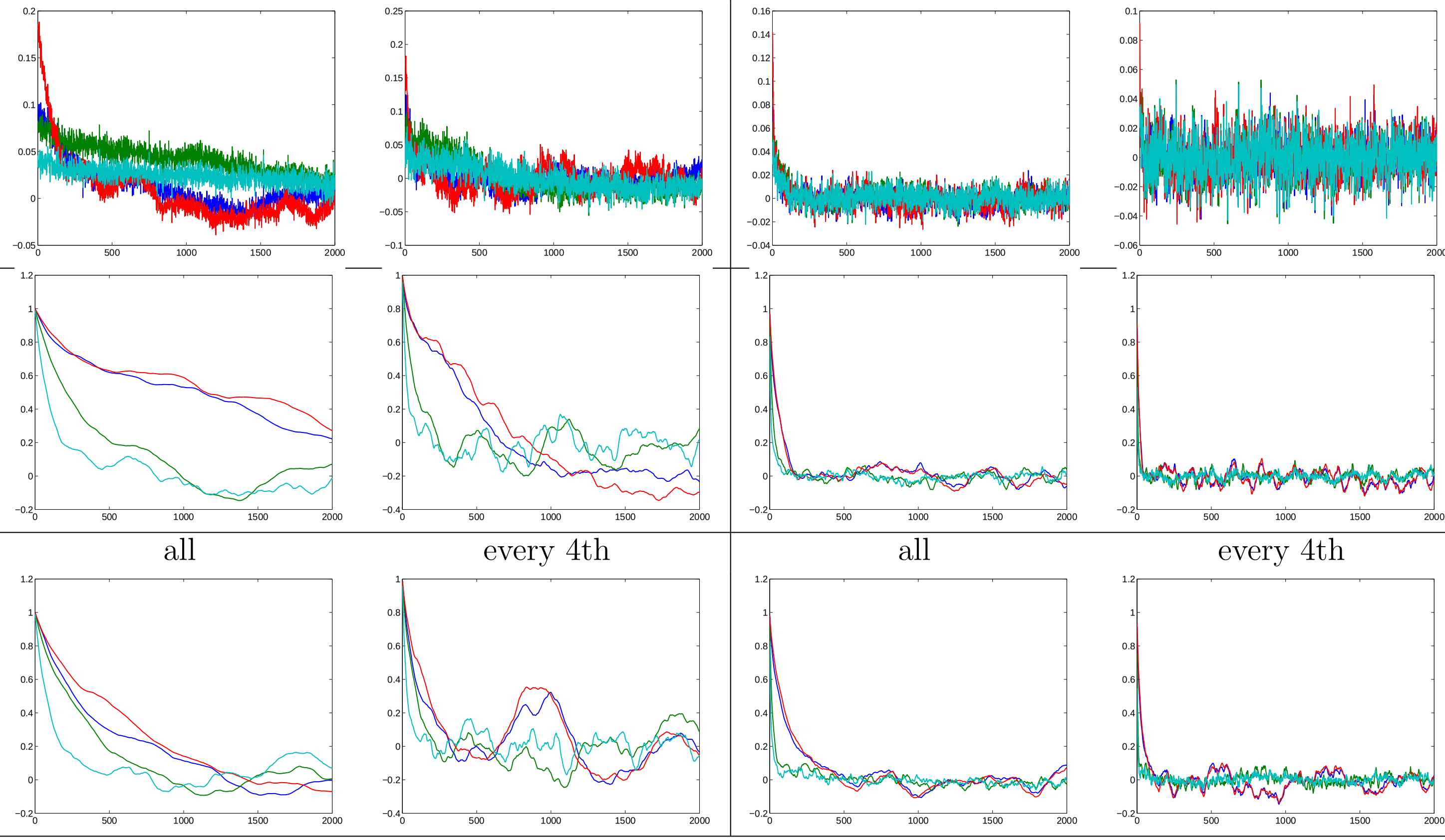


Figure 5: Simulated data. Autocorrelation function of sampled values γ.

	Auxiliary sampling based on RUM		dRUM	
	Iterations retained		Iterations retained	
Random permutation:	all[(a)]	every 4th	all	every 4th
– unidentified model γ^z_{12} γ^z_{22} γ_{12} γ_{22}				
– identified model γ^z_{12} γ^z_{22} γ_{12} γ_{22}				
Constrained permutation	all	every 4th	all	every 4th
– identified model γ^z_{12} γ^z_{22} γ_{12} γ_{22}				



(a) The last 20,000 of a total of 50,000 iterations.

Figure 6: M3 growth, HP-trend and low-frequency component (> 6 years).

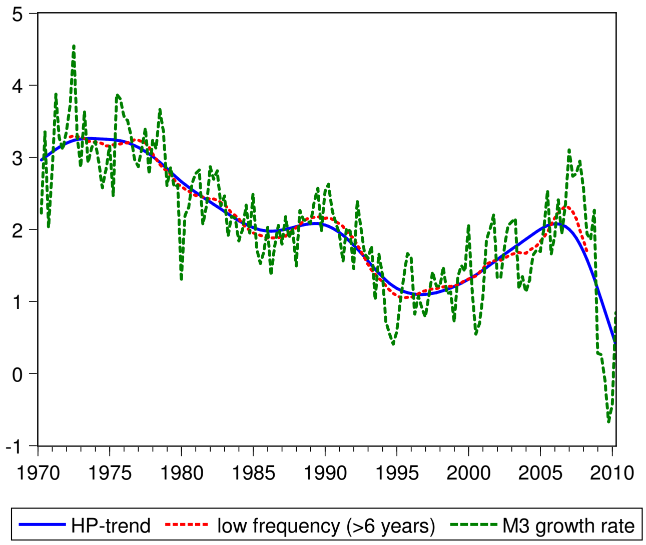


Figure 7: Scatter plot of sampled regression parameter against constant transition effect.

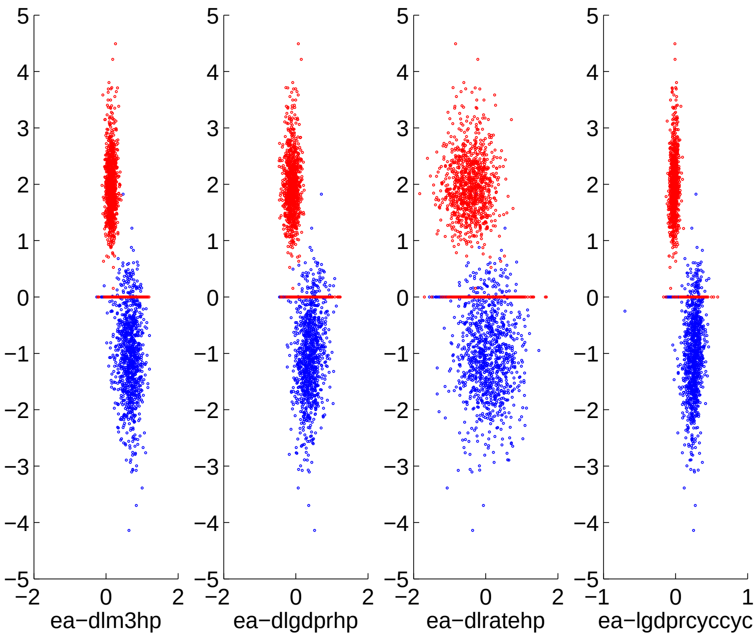


Figure 8: Marginal posterior distribution of state-identified regression coefficients (solid line, regime 2).

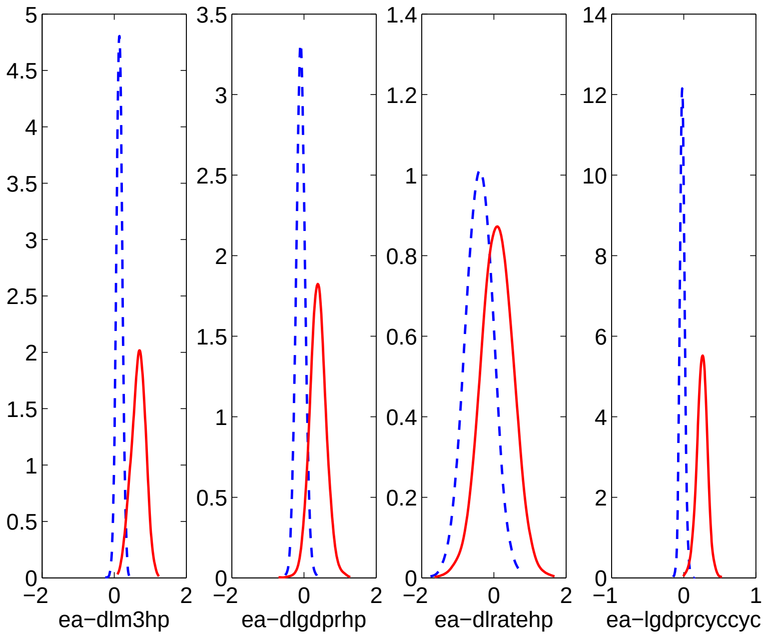


Figure 9: Marginal posterior distribution of error variance and state-identified covariate effects on the transition probability (solid line, regime 2).

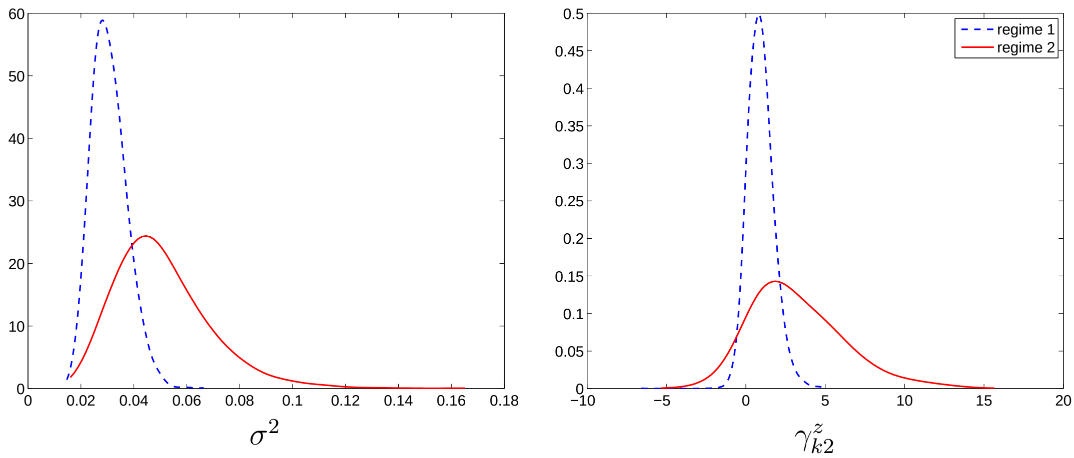


Figure 10: Boxplot of the sampled $\gamma_{\cdot 2} = (\gamma^z_{12}, \gamma^z_{22}, \gamma_{12}, \gamma_{22})$ and marginal posterior of the covariate effects for increasingly informative prior distributions

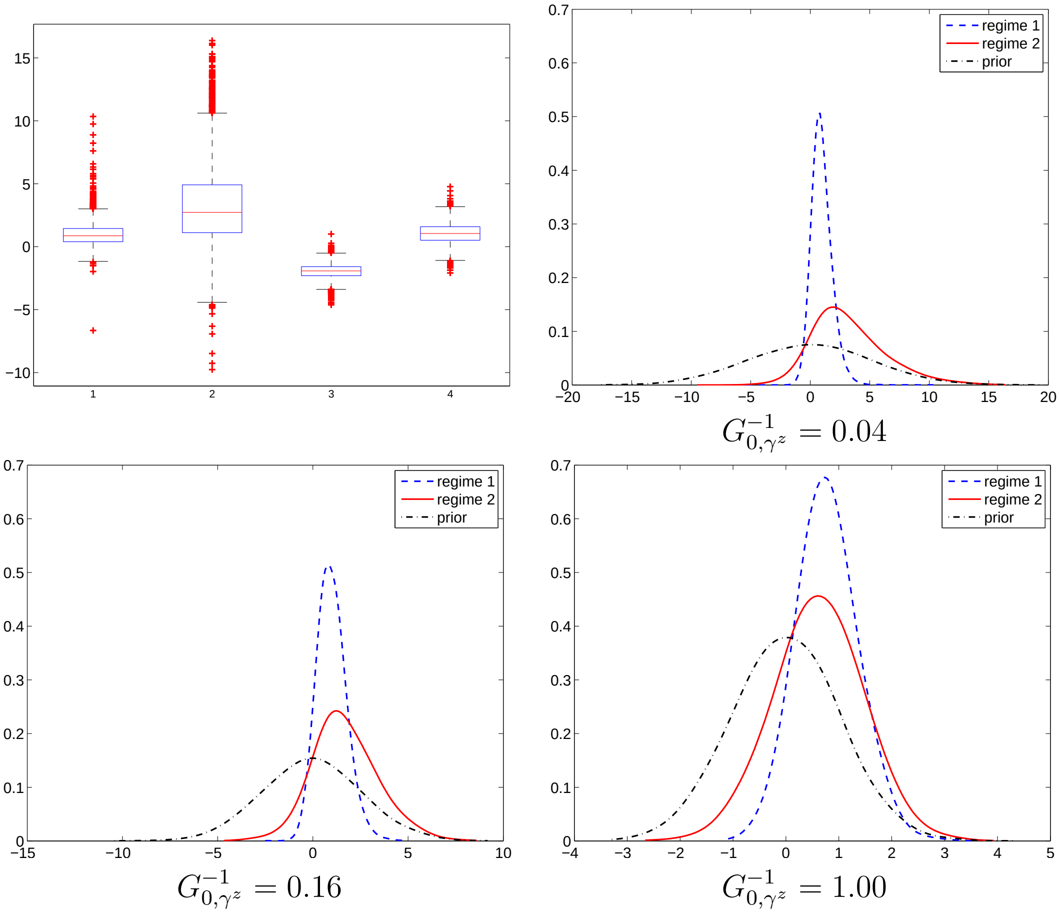


Figure 11: Posterior mean state probabilities (cyan) along with HCIP inflation, loan growth and trend M3 growth. The horizontal line corresponds to a threshold level of 1.9% quarterly credit growth rate, composed from an average of 1.7% growth rate and an inferred 0.2% according to Definition 1 (see section 2.2). The yellow bars correspond to periods where loan growth is continuously above trend M3 growth.

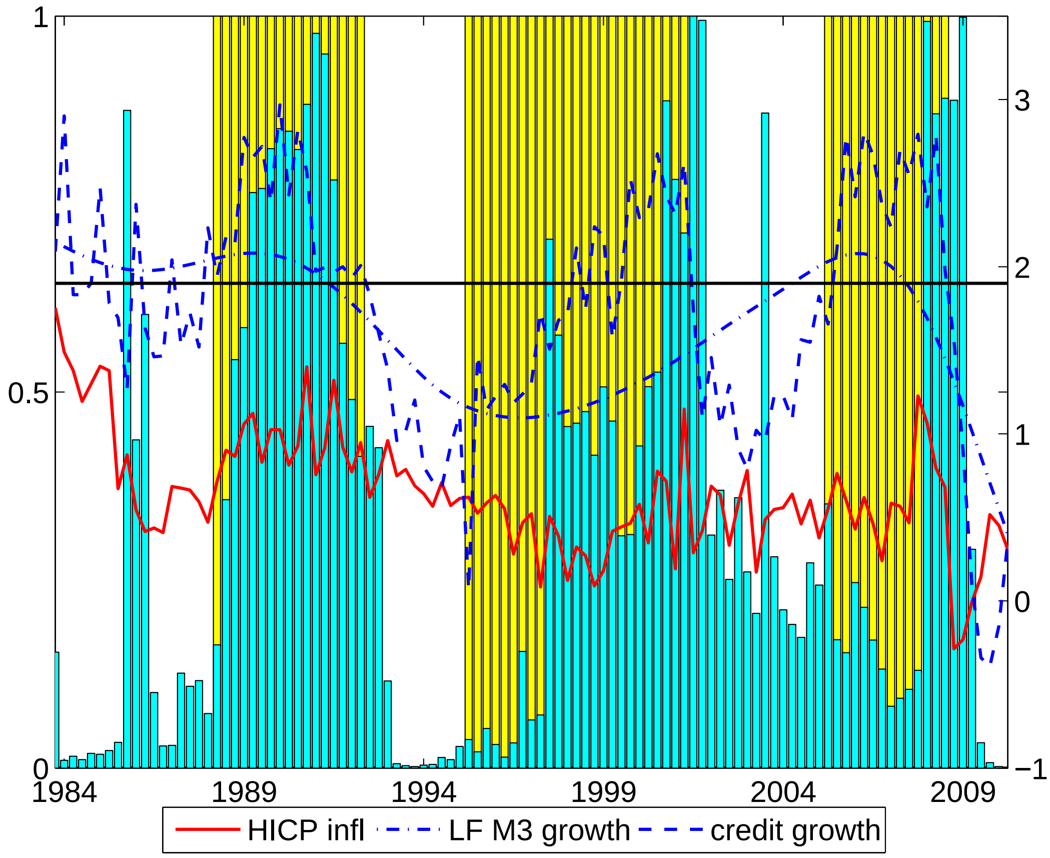


Figure 12: Median posterior transition probabilities.