K-state switching models with endogenous transition distributions

Sylvia Kaufmann

SCHWEIZERISCHE NATIONALBANK
BANQUE NATIONALE SUISSE
BANCA NAZIONALE SVIZZERA
BANCA NAZIUNALA SVIZRA
SWISS NATIONAL BANK

2011-13

Swiss National Bank Working Papers

The views expressed in this paper are those of the author(s) and do not necessarily represent those of the Swiss National Bank. Working Papers describe research in progress. Their aim is to elicit comments and to further debate.

ISSN 1660-7716 (printed version)
ISSN 1660-7724 (online version)

K-state switching models with endogenous transition distributions

Sylvia Kaufmann[*†]

November 2011

Abstract

Two Bayesian sampling schemes are outlined to estimate a K-state Markov switching model with time-varying transition probabilities. The multinomial logit model for the transition probabilities is alternatively expressed as a random utility model and as a difference random utility model. The estimation uses data augmentation and both sampling schemes can be based on Gibbs sampling. Based on the model estimate, we are able to discriminate the model against a smooth transition model, in which the state probability may be influenced by a variable, but without depending on the past prevailing state. Formulating a definition allows to determine the relevant threshold level of the covariate influencing the transition distribution without resorting to the usual grid search. Identification issues are addressed with random permutation sampling. In terms of efficiency the extension to difference random utility in combination with random permutation sampling performs best. To illustrate the method, we estimate a two-pillar Phillips curve for the euro area, in which the inflation rate depends 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. Using recent data series, the effect of the low-frequency component of M3 growth depends on regimes determined by lagged credit growth.

JEL classification: C11,C22,E31,E52

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

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 (2009), Kaufmann (2010). Generally, the transition probabilities

[*]Swiss National Bank, Börsenstrasse 15, CH-8022 Zürich, sylvia.kaufmann@snb.ch

[†]The paper contains the views of the author and not necessarily those of the SNB.

Thanks for comments go to an anonymous referee. The author assumes the responsibility for errors or omissions.

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 smooth transition model (STAR, Teräsvirta and Anderson 1992). The latter models usually include a threshold to be estimated. Using a so-called centered parametrization which leaves the threshold inherently unidentified allows to estimate the time-varying influence of the covariate irrespectively of the threshold. Nevertheless, we show that a threshold different from the mean of the covariate can be recovered by exploiting the role that the covariates play in the time-varying transition distribution. In short, after model estimation, the threshold level is defined as the level at which the divergence between the persistence probabilities of states is minimized.

Another issue that is also addressed 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 (2009), 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 (2009), who specify the moments of the Beta distribution governing a two-state switching process to depend on covariates like duration or past transition probabilities. 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. It turns out that these results are not reproducible if the empirical Phillips curve is estimated for shorter and more recent data series running from 1983 to 2010. Extending the setup to a Markov switching framework recovers a state-specific long-run unity coefficient for M3 growth. Lagged credit growth rate above a threshold level of 2% quarterly growth rate is estimated to be indicative of switches to the state in which M3 growth is significant for inflation.

The next section outlines the econometric model and discusses the parametrization of the transition distribution. Section 3 presents the MCMC sampling scheme. The interested reader finds the detailed derivations of the posterior distributions in appendices A and B. 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.

2 The econometric model

2.1 Specification

The usual representation of a regime-switching model for a time series y_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 taking into account all time-invariant parts simultaneously, 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

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

[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 $Z_t'\gamma^z_{lk}$.

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

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

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 simultaneously on all time-invariant parts of the transition probabilities.

Although we do not put any restrictions on γ_{lk}^z, 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 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.

Finally, the parametrization is quite general and nests some interesting alternatives, which are discussed in the following subsection.

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 would restrict $\gamma_{lk}^z = \gamma_k^z$. 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 Markov switching model with constant transition probabilities.

If on the other hand $\gamma_{lk}^z = \gamma_k^z$ 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 multi-state analogue to the logistic smooth transition model of Teräsvirta and Anderson (1992):

$$\begin{aligned} y_t &= (1-\xi_t)\,\beta_1 X_t + \xi_t \beta_2 X_t + \varepsilon_t && (6)\\ &= \beta_1 X_t + \xi_t\,(\beta_2-\beta_1)\,X_t + \varepsilon_t && (7)\\ \xi_t &= \frac{1}{1+\exp\left(\gamma^z\,(Z_t-c)\right)} \end{aligned}$$

where γ^z represents the curvature and c the threshold. The parametrization we adopt inherently leaves the threshold unidentified, given that any level (also different from the mean) may be recovered from a posterior estimate of (3):

$$Z_t\gamma^z_{jk} + \gamma_{jk} = (Z_t - c)\gamma^z_{jk} + \tilde{\gamma}_{jk}$$

where $\gamma_{jk} = -c\gamma^z_{jk} + \tilde{\gamma}_{jk}$.

We may nevertheless define a relevant threshold level:

Definition 1: The relevant threshold level is the level of $\tilde{Z}_t$ at which the divergence between the persistence probabilities of states is minimized.

According to this definition, in case $K = 2$, in the Markov switching model the level of $\tilde{Z}_t$ would be the level at which the persistence probabilities of states is equalized, $\xi_{11,t} = \xi_{22,t}$.[2] Using the centered specification (3), the threshold level is composed of two components: the average level $\bar{Z}$ and the level c, which can be determined after model estimation using Definition 1 applied to Z_t instead of $\tilde{Z}_t$. The obvious advantage of the procedure is that the inference about the threshold level is done without having to resort to a grid search, which represents the common approach in estimating transition models.

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, the smooth transition model or the constant transition Markov model.

2.3 Some examples

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

$$\xi_{k2,t} = \frac{\exp\left(\mathbf{Z}_t'\gamma_2\right)}{1+\exp\left(\mathbf{Z}_t'\gamma_2\right)} \tag{8}$$

where $\mathbf{Z}_t = \left(Z_t D^{(1)}_{t-1}, Z_t D^{(2)}_{t-1}, D^{(1)}_{t-1}, D^{(2)}_{t-1}\right)'$, with $D^{(j)}_t = 1$ if $S_t = j$ and 0 otherwise, $j = 1, 2$. The parameter γ_2 has four elements, $\gamma_2 = (\gamma^z_{12}, \gamma^z_{22}, \gamma_{12}, \gamma_{22})$. The first two elements determine the time-varying effect of the covariate on the transition probability to state 2, which depends on the state prevailing in period $t-1$. The last two elements, $(\gamma_{12}, \gamma_{22})$, are the parameters governing the time-invariant transition probability from state 1 in period $t-1$ to state 2 in period t and to the persistence of remaining in state 2, respectively. Four

[2]For the smooth transition model, the analogously defined threshold level is the level of $\tilde{Z}_t$ at which the absolute difference between the state probabilities is minimized, i.e. the level of $\tilde{Z}_t$ at which the state probability is equal to 0.5, $\xi_t = 0.5$.

different settings for γ_2 are assumed. In the first three, $\gamma_2 = (4, g, -2, 2)$, $g = 0, 1, 4$, which yields an average persistence of 0.88 for each state. The influence of the combinations 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 ill ustrates 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 $Z_t = 0.5$, yielding a state probability of 0.5, $\xi_t = 0.5$.

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. $y^t = (y_t, y_{t-1}, \dots, y_1)$, and similarly for X^t, Z^t, S^t. The regression parameters are gathered into the parameter vector $\beta = (\beta_1, \dots, \beta_K)$, where $\beta_k = (\beta_{1,k}, \dots, \beta_{p,k})$, for $k = 1, \dots, K$. Finally, the parameters governing the transition probabilities are denoted by $\gamma = \{\gamma_j | j \in \mathcal{K}_{-k_0}\}$ with $\gamma_j = (\gamma^z_{1j}, \dots, \gamma^z_{Kj}, \gamma_{1j}, \dots, \gamma_{K,j})$. All model parameters are contained in $\theta = (\beta, \gamma, \sigma^2)$,

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)	$\begin{bmatrix} 0.88 & 0.12 \\ 0.12 & 0.88 \end{bmatrix}$	$\begin{bmatrix} 0.69 & 0.31 \\ 0.20 & 0.80 \end{bmatrix}$	$\begin{bmatrix} 0.96 & 0.04 \\ 0.07 & 0.93 \end{bmatrix}$

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

$$L\left(y^T|X^T, S^T, \theta\right) = \prod_{t=1}^{T} f\left(y_t|X_t, S_t, \theta\right) \tag{9}$$

with a normally distributed observation density

$$f\left(y_t|X_t, S_t, \theta\right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}\left(y_t - X_t'\beta_{S_t}\right)^2\right\} \tag{10}$$

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

$$\pi\left(S^T|Z^T, \gamma\right) = \prod_{t=1}^{T} \pi\left(S_t|Z_t, S_{t-1}, \gamma\right)\pi\left(S_0\right) \tag{11}$$

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

$$\pi\left(\theta\right) = \pi(\beta)\pi(\sigma^2)\pi(\gamma) \tag{12}$$

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:[3]

$$\begin{aligned}
\pi\left(\beta\right) &= \prod_{k=1}^{K} \pi\left(\beta_k\right) = \prod_{k=1}^{K} N\left(b_0, B_0\right) && (13)\\
\pi\left(\sigma^2\right) &= IG\left(w_0, W_0\right) && (14)
\end{aligned}$$

The prior specification in (13) additionally assumes the state-dependent regression parameters to be independent of each other, and to follow a state-independent prior distribution. The specification can be generalized to include state-dependent prior hyperparameters, $\pi\left(\beta_k\right) = N\left(b_{0k}, B_{0k}\right)$. The logit specification for the transition probabilities in (3)-(4) allows to assume a normal prior distribution for the parameter γ:

$$\pi\left(\gamma\right) = \prod_{k \in \mathcal{K}_{-k_0}} \pi\left(\gamma_k\right) = \prod_{k \in \mathcal{K}_{-k_0}} N\left(g_{0k}, G_{0k}\right) \tag{15}$$

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

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

[3] In case of state-specific error variances we would specify the prior $\pi\left(\sigma_1^2, \ldots, \sigma_K^2\right) = \prod_{k=1}^{K} IG\left(w_{0k}, W_{0k}\right)$.

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-dependent structure in the present logit model. 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 specification of the transition probabilities. The interested reader finds the derivation of the posterior sampling densities in appendix A.

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,

$$S^u_{kt} = \mathbf{Z}_t'\gamma_k + \nu_{kt},\ \forall k \in \mathcal{K}_{-k_0} \tag{17}$$

$$S^u_{k_0,t} = \nu_{k_0,t},\ \text{for the identification restriction } \gamma_{k_0} = 0, \tag{18}$$

where $\mathbf{Z}_t = \left(Z_t D^{(1)}_{t-1}, Z_t D^{(2)}_{t-1}, \ldots, Z_t D^{(K)}_{t-1}, D^{(1)}_{t-1}, D^{(2)}_{t-1}, \ldots, D^{(K)}_{t-1}\right)'$. If $\nu_{kt},\ k = 1, \ldots, K$, follow a Type I extreme value distribution, the marginal distribution of S_t will be the multinomial logit model as in (3)-(4). 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 2006). 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

$$S^u_{kt} = \mathbf{Z}_t'\gamma_k + m_{R_{kt}} + s_{R_{kt}}\upsilon_{kt}, \quad \upsilon_{kt} \sim N(0, 1). \tag{19}$$

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

$$G_k = \left(\sum_{t=1}^{T} \mathbf{Z}_t \mathbf{Z}_t' / s^2_{R_{kt}} + G^{-1}_{0k}\right)^{-1} \tag{20}$$

$$g_k = G_k \left(\sum_{t=1}^{T} \mathbf{Z}_t \left(S^u_{kt} - m_{R_{kt}}\right) / s^2_{R_{kt}} + G^{-1}_{0k} g_{0k}\right) \tag{21}$$

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

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

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 (17). The model can further be condensed to obtain the partial dRUM representation:

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

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 $\lambda_{-k,t} = \sum_{j\in\mathcal{K}_{-k}} \exp\left(\mathbf{Z}_t'\gamma_j\right)$. Given that the constant $-\log\left(\lambda_{-k,t}\right)$ 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):

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

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

$$\begin{aligned}G_k &= \left(\sum_{t=1}^{T} \mathbf{Z}_t\mathbf{Z}_t'/s^2_{R_{kt}} + G_{0k}^{-1}\right)^{-1} &\quad (26)\\ g_k &= G_k\left(\sum_{t=1}^{T} \mathbf{Z}_t\tilde{\omega}_{kt}/s^2_{R_{kt}} + G_{0k}^{-1}g_{0k}\right) &\quad (27)\end{aligned}$$

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

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,

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

where the rows of $\tilde{X}$, $\tilde{X}_t = \left(X_tD_t^{(1)}, X_tD_t^{(2)}, \ldots, X_tD_t^{(K)}\right)$. The posterior of σ^2 is inverse Gamma, $IG\left(w, W\right)$ 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 $IG\left(w_k, W_k\right)$ with $w_k = w_0 + 0.5T_k$, $T_k = \sum_{t=1}^{T} D_t^{(k)}$ and $W = W_0 + 0.5\sum_{t=1}^{T} D_t^{(k)}\left(y_t - X_t'\beta_k\right)^2$

To motivate step (iv), note that the model (1) is not identified with respect to the states. The likelihood (9), $L\left(y^T|X^T, S^T, \theta\right)$ and hence the posterior remain unchanged with respect to any state permutation $\rho = \left(\rho_1, \ldots, \rho_K\right)$[4]

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

[4] For example in case $K = 2$, a permutation $\rho = (2, 1)$ would reorder the states and the state-dependent parameter such that state 2 would become state 1.

The investigator may choose one of two options. 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:

$$\beta_{j1} < \cdots < \beta_{jK} \text{ or } \gamma_{j1} < \cdots < \gamma_{jK} \tag{28}$$

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

$$\beta_{j1} < \min\left(\beta_{j2}, \ldots \beta_{jK}\right) \text{ and } \gamma_{j2} < \cdots < \gamma_{jK} \tag{29}$$

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 $k_0 = 0$. In this case, the specification of the hyperparameters should not be at odds with the state-identifying restrictions.

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. At the end of each sweep, the states and the state-dependent parameters are permuted randomly. The multimodal posteriors can then be used to find a state-identifying restriction, according to which the sampled values of the states and the state-dependent parameters are re-ordered to obtain the posterior inference on the identified model. A detailed description of the permutation steps is found in appendix C.

4 Illustration and evaluation

4.1 Model estimation

To illustrate the usefulness of the random 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 Markov switching process S_t is modelled to depend on one covariate Z_t. Assuming two states $K = 2$ and $k_0 = 1$, we obtain:

$$\xi_{k2,t} = \frac{\exp\left(\mathbf{Z}_t' \gamma_2\right)}{1 + \exp\left(\mathbf{Z}_t' \gamma_2\right)} \tag{31}$$

where the parameter $\gamma_2 = (4,\ 1,\ -2,\ 2)$ reflects the property we think of being most intuitive in macroeconomic applications of Markov switching models. The values $(-2,\ 2)$

correspond, when $\tilde{Z}_t$ is at its threshold, to a transition probability matrix (see also table (1))

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

The exogenous variables and the covariate are drawn from independent normal distributions:

$$x_{1t}, x_{2t} \text{ i.i.d. } N(0,1) \qquad \tilde{Z}_t = 0.8\tilde{Z}_{t-1} + \eta_t,\ \eta_t \text{ i.i.d. } N(0,0.5) \tag{32}$$

$$Z_t = \tilde{Z}_t - 0.5 \tag{33}$$

where the relative strong autoregressive process for $\tilde{Z}_t$ is chosen to induce some persistence in the simulated Markov variable S_t. The subtraction of 0.5 from $\tilde{Z}_t$ is introduced to illustrate the possibility of recovering the threshold level from the model estimate using Definition 1. We simulate 400 observations, $T = 400$, and use the last 200 to estimate the model. Figure 2 plots the simulated state variable along with the covariate in the top panel and the time series y_t in the bottom panel. The influence of the covariate is nicely observable. If $\tilde{Z}_t$ is above the threshold of 0.5, the indicator S_t switches to state 2.

In a first round, we work with $\tilde{Z}_t$ 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 β_{jk} $\pi(\beta_{jk}) \sim N(0, 1/4)$, for σ_k $\pi(\sigma_k) \sim IG(2, 0.25)$ and for γ_2 $\pi(\gamma_2) \sim N\left([4,\ 0,\ 0,\ 0]', \text{diag}(1,\ 1,\ 4,\ 4)\right)$. We iterate 50'000 times over the sampler outlined in subsection 3.2 and estimate the model using alternatively random and constrained permutation sampling. The parameters of the transition distribution are sampled using both alternatives of the auxiliary sampling schemes. In both cases, we apply random and alternatively constrained permutation sampling.

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. The simulated values for β_{1k} (switching parameter) and β_{2k} (not switching parameter) are plotted in figure 3, panel (a). The sampler converges quickly. Given that the sampler is forced to visit both modes of the posterior, the simulation paths for β_{1k} and β_{2k}, $k = 1, 2$, overlap. The scatter plots in figure 4 plot the simulated regression parameters against the simulated constant transition parameters γ_{k2} (every 4th of the last 20,000 iterations). These obviously reveal that β_{1k} is switching between states, while β_{2k} apparently not. The bimodality of state-dependent parameters is reflected in the marginal posterior densities depicted in figure 5, panel (a). At first sight, the state-dependency of the error variance is not obvious.

To obtain state-identification, we may re-order the simulated values according to the state-identifying restriction $\beta_{11} < \beta_{22}$ and normalize the parameters of the transition distribution choosing $k_0 = 1$ (see permutation scheme (61) in appendix C). The result of the identification step is plotted in figure 3, panel (b), for the regression coefficients β_{1k} and β_{2k}. Obviously, the restriction is able to uniquely identify the two modes. This is also reflected in the marginal distributions depicted in figure 5, panel (b).

To illustrate the importance of appropriate identification, figure 6 depicts the marginal posteriors obtained when imposing an inappropriate state-identifying restriction while sampling, namely $\beta_{21} < \beta_{22}$. Remember that β_{2k} in (30) is truly not state-dependent. While the state-specific marginal posteriors of β_{2k} are unimodal, the restriction fails to invoke unimodality in the posterior of the truly state-dependent parameters, β_{1k} and γ^z_{k2}.

Based on these we would clearly obtain a biased inference on first and second moments of the marginal posterior distributions of the model parameters.

Figure 7 shows that applying Definition 1 to recover the threshold would yield a median estimate of 0.48 with an interquartile range of 0.17. 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 points plot the threshold level $\tilde{Z}_t^{(m)}$ determined according to Definition 1 against the implied 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 can 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, $\hat{\sigma}_\gamma^2$. 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 of a total of 50,000 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 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 8), 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 only 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 nearly every parameter, the inefficiency more than triples, irrespectively of whether we retain all iterations or retain only every 4th one. The increase in inefficiency is even more stronger, by a factor of at least 4 to one of 10, when comparing the factors for the identified models. Finally, constrained permutation with auxiliary sampling based on RUM leads to the most inefficient sampled MCMC output, the inefficiency factor is not quite reduced by thinning out the MCMC sample. The inefficiency using the RUM extension is larger by a factor of at least 6 up to a factor of 40 (for γ_{22}) when compared to the dRUM extension.

The results about the inefficiency factors are mirrored in the autocorrelation functions (ACF) of the sampled values for γ. Figure 8 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 permutated 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:

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

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 (extracted by the HP-filter) of the respective variables is thought to affect the long-run frequency component of the inflation rate, while its high frequency component is thought to be affected by the cyclical component of the output gap, indicated by a hat. To take into account dynamics, we also include up to p lagged values of the inflation rate. As a result of a first investigation, the autocorrelation coefficients and the error variance turned out to be state-independent, therefore we omit a state-dependent specification in equation (34).

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 9 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

We present three estimations of the two-pillar Phillips curve (34). The first estimate will reproduce the results of Assenmacher-Wesche and Gerlach (2008), in which the coefficients will not be subject to regime switching. We will thus work with the whole available observation sample, covering the period 1970-2010. To account for dynamics, we also include three lagged values of the inflation rate, the fourth being insignificant in a preliminary estimation. In the second estimate the sample is restricted to begin in 1983. Thus, the results yield evidence about the robustness of the estimates when the investigation concentrates only on more recent data. Last, we present results for the estimation where the coefficients are regime switching. We additionally estimate whether the transition distribution of the regime indicator is endogenous and depends on lagged credit growth. The model specifying the transition distribution is the one in (3)-(4), where Z_t is lagged credit growth adjusted by its mean of 1.7% quarterly growth rate.

All estimations are based on 75,000 iterations of the MCMC sampler described in section 3, discarding the first 35,000 and retaining only every 4th for posterior inference. Based on the efficiency evaluation presented in section 4, we sample out of the unconstrained posterior, i.e. the MCMC sample is obtained by applying the random permutation sampler. We base auxiliary sampling of the transition distribution parameters on the dRUM extension. State identification is then obtained by post-processing the MCMC output by re-ordering the sampled values according to a state-identifying restriction.

5.2.1 Baseline estimation

The results of the baseline estimation are depicted in the first column of table 3. Basically, we can reproduce the results of Assenmacher-Wesche and Gerlach (2008), also 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 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% highest posterior density interval (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 marginally positive (0.48).

When the estimation sample is restricted to begin in 1983, the results basically remain robust, although the long-run importance of trend M3 growth is estimated to have decreased. The 95% HPDI does not include a unit coefficient anymore. The cyclical output gap remains marginally significant for inflation dynamics. Its effect has also decreased, however.

5.2.2 Regime switching with time-varying transition distribution

Given that the effect of trend M3 growth on inflation has apparently decreased over time, it is interesting to assess whether the effect depends on regimes which characterize specific macroeconomic conditions. We will investigate whether the lagged loan growth rate might be one determinant of the regime transitions.

In a first round, all variables and the error variance are assumed to be state-dependent. The output of the random permutation sampler is depicted in figure 10. The coefficients on trend M3 growth and on the trend in GDP growth are the most obvious ones to be state-dependent. The scatter plot for the error variance (not displayed) reveals that this parameter is not state-dependent, either. Based on this first inference, the final estimate will restrict the coefficient on the cyclical output gap and the error variances to be state-independent.[5] The states are identified by re-ordering the sampled values according to $\beta_{11} < \beta_{12}$ (see the permutation steps in (60)), i.e. state two is the one with a stronger effect of the long-run component of M3 growth. The marginal posterior distributions of the state-identified parameters are depicted in the figures 11 and 12). The right-hand plot in figure 12 shows that lagged credit growth affects the transition distribution of state 1 but not significantly the one of state 2.

The posterior inference on the state-identified parameters is summarized in table 4. Because there is some overlap in the posterior distributions, we report the 95% and the 90% HPDI in brackets on the first and second line, respectively, below the mean estimate of the coefficients. Regime 2 now recovers the expected influence of the variables. In particular trend M3 growth has a strong positive effect on inflation, the short-run 95% and the long-run HPDI intervals cover the unit coefficient. The negative effects of trend GDP growth and the trend in the change of the government bond yield are marginally significant, zero is excluded from the 90% HPDI. In the first regime, mainly real variables determine inflation. Trend GDP growth and the cyclical output gap (the latter in both states) have a marginally positive effect on inflation in the short-run and in the long-run as well.

Figure 13 depicts the posterior probabilities of state 2, $P\left(S_t = 2|y^T, X^T, Z^T\right)$. At the beginning of each episode during which state 2 has been relevant, loan growth was initially high and inflation was at above-average levels. Moreover, these episodes are mainly characterized by trend M3 and loan growth moving in parallel. The median posterior transition probabilities are plotted in figure 14. We observe the effect of lagged loan growth on the transition distribution of state 1. In particular in 1989 and after 2005, the persistence of state 1 decreases from nearly unity to below 0.8, indicating the switches to state 2 in figure 13. The horizontal line in figure 13 corresponds to a threshold level of 2.0% quarterly credit growth rate composed of an average growth rate of 1.7% and of 0.3% inferred according to Definition 1 of subsection 2.2. The latter corresponds to the median (across the MCMC output) of the corresponding level of Z_t at which the divergence between the state persistence probabilities is minimized.

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

[5] 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.

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, in particular against a smooth transition model and a Markov model with time-invariant transition probabilities. We give a definition to determine a relevant threshold of the covariate influencing the transition distribution. The advantage of the procedure is to obtain an inference on the threshold without resorting to a grid search, the procedure usually pursued to estimate smooth transition models.

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

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

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

Billio, M. and R. Casarin (2009). Bayesian estimation of stochastic-transition Markov-switching models for business cycle analysis. mimeo, University of Venice and Brescia.

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.

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. 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. IFAS Research Paper Series 2010-48, Department of Applied Statistics, Johannes Kepler University Linz.

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 (2009). The propagation of regional recessions. Working Paper 13, revised version October 2010, Federal Reserve Bank of St. Louis.

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

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

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*, 309344.

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.

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

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

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.

Scott, S. L. (2006). Data augmentation, frequentistic estimation, and the Bayesian analysis of multinomial logit models. mimeo, The Marshall School of Business, University of Southern California, Los Angeles, CA 90089-1421.

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.

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

A Sampling out of $\pi\left(S^T|y^T,X^T,Z^T,\theta\right)$

To derive the sampling scheme for S^T, we define the time-varying matrix ξ_t with element $\xi_{lk,t}$, $l,k=1,\ldots,K$ representing the transition probability from state l in $t-1$ to state k in t. For $k=2,\ldots,K$:

$$\xi_{lk,t}=P(S_t=k|S_{t-1}=l,Z_t,\gamma)=\frac{\exp\left(Z_t\gamma^z_{lk}+\gamma_{lk}\right)}{1+\sum_{k=2}^K\exp\left(Z_t\gamma^z_{lk}+\gamma_{lk}\right)}$$

where γ^z_{lk} and and γ_{lk} represent the state-dependent effect of the covariate Z_t (here: lagged credit growth) and the average state-dependent effect, respectively. The transition to state 1 defining the first column of ξ_t, $\xi_{l1,t}$ is the reference transition, and thus is independent of the covariate:

$$\xi_{l1,t}=P(S_t=1|S_{t-1}=l,Z_t,\gamma)=\frac{1}{1+\sum_{k=2}^K\exp\left(Z_t\gamma^z_{lk}+\gamma_{lk}\right)}$$

We express the posterior $\pi\left(S^T|y^T,X^T,Z^T,\theta\right)$ as $\pi\left(S^T|y^T,X^T,\xi^T,\theta_{-\gamma}\right)$ and factorize it

$$\pi\left(S^T|y^T,X^T,\xi^T,\theta\right)=\pi\left(S_T|y_T,X_T,\xi_T,\theta_{-\gamma}\right)\prod_{t=1}^{T-1}\pi\left(S_t|y_t,X_t,\xi_t,\theta_{-\gamma}\right)\pi\left(S_{t+1}|S_t,\xi_{t+1}\right)$$

The filter density $\pi\left(S_t|y_t,X_t,\xi_t,\theta_{-\gamma}\right)$ is obtained by iterating forward through $t=1,\ldots,T$

$$\begin{aligned}\pi\left(S_t|y_t,X_t,\xi_t,\theta_{-\gamma}\right) &\propto f\left(y_t|X_t,S_t,\theta_{-\gamma}\right)\pi\left(S_t|y_{t-1},X_{t-1},\xi_t,\theta_{-\gamma}\right)\\ \pi\left(S_t|y_{t-1},X_{t-1},\xi_t,\theta_{-\gamma}\right) &= \xi_t'\pi\left(S_{t-1}|y_{t-1},X_{t-1},\xi_{t-1},\theta_{-\gamma}\right)\end{aligned}$$

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

State S_T is sampled out of $\pi\left(S_T|y_T,X_T,\xi_T,\theta_{-\gamma}\right)$. We proceed backwards $t=T-1,\ldots,0$ and draw from the posterior sampling density

$$\pi\left(S_t|y_t,X_t,S_{t+1},\xi_t,\theta_{-\gamma}\right)\propto\pi\left(S_t|y_t,X_t,\xi_t,\theta_{-\gamma}\right)\xi_{S_tS_{t+1},t+1}$$

where $\xi_{S_tS_{t+1},t+1}$ extracts the column S_{t+1} of the matrix ξ_{t+1}.

B 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, which are described for each model extension in the following sub-sections.

B.1 Data augmentation for RUM

The three following sampling steps form step (ii) in a sweep of the whole sampling scheme (see section 3).

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

(ii.b) Sample the components R_{kt} from $\pi\left(R^{KT}|S^{u,KT},\gamma\right)$

(ii.c) Sample γ from $\pi\left(\gamma|S^{u,KT},R^{KT}\right)$

To sample the utilities

$$\begin{aligned} S^u_{kt} &= \mathbf{Z}'_t\gamma_k + \nu_{kt},\ \forall k\in\mathcal{K}_{-k_0} \\ S^u_{k_0t} &= \nu_{k_0t},\ \text{from the identification restriction } \gamma_{k_0}=0, \end{aligned} \tag{35}$$

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

$$\exp\left(-S^u_{jt}\right) \sim \mathcal{E}\left(\sum_{k=1}^K \lambda_{kt}\right) \tag{36}$$

where $\lambda_{kt} = \exp\left(\mathbf{Z}'_t\gamma_k\right)$.[6]

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

$$\pi\left(S^u_{1t},\ldots,S^u_{Kt}|S_t=j,\gamma\right) = \pi\left(S^u_{jt}|S_t=j,\gamma\right)\prod_{k\in\mathcal{K}_{-j}}\pi\left(S^u_{kt}|S_t=j,\gamma\right). \tag{37}$$

The distribution $\pi\left(S^u_{jt}|S_t=j,\gamma\right)$ is given by 36 and implies

$$\exp\left(-S^u_{kt}\right) = \exp\left(-S^u_{jt}\right) + \xi_{kt}, \qquad \xi_{kt}\sim\mathcal{E}\left(\lambda_{kt}\right),\ \forall k\in\mathcal{K}_{-j} \tag{38}$$

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:

$$S^u_{kt} = -\log\left(-\frac{\log\left(W_t\right)}{\sum_{l=1}^K\lambda_{lt}} - \frac{\log\left(V_{kt}\right)}{\lambda_{kt}} I_{\{S_t\neq k\}}\right) \tag{39}$$

Conditional on S^u_{kt}, the component indicator R_{kt} (step ii.b) is sampled from:

$$P\left(R_{kt}=r|S^u_{kt},\gamma_k\right) \propto \frac{w_r}{s_r}\exp\left\{-\frac{1}{2}\left(\frac{S^u_{kt}-\mathbf{Z}'_t\gamma_k-m_r}{s_r}\right)^2\right\},\ k\in\mathcal{K}_{-k_0} \tag{40}$$

[6]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:

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

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.

Finally, given all utilities $S^{u,KT}$ and all component indicators R^{KT}, we obtain a linear regression model for the parameters governing the transition probabilities to each state $k, \ k \in \mathcal{K}_{-k_0}$:

$$S^u_{kt} = \mathbf{Z}'_t \gamma_k + m_{R_{kt}} + s_{R_{kt}} \upsilon_{kt}, \quad \upsilon_{kt} \sim N(0,1) \tag{41}$$

Assuming a normal prior for γ_k, $\pi(\gamma_k) = N(g_{0k}, G_{0k})$, conditional on $S^{u,KT}$ and R^{KT} the posterior is normal, too:

$$\pi\left(\gamma_k | S^{u,T}_k, R^T_k\right) = N(g_k, G_k), \ \forall k \in \mathcal{K}_{-k_0} \tag{42}$$

$$G_k = \left(\sum_{t=1}^T \mathbf{Z}_t \mathbf{Z}'_t / s^2_{R_{kt}} + G^{-1}_{0k}\right)^{-1} \tag{43}$$

$$g_k = G_k \left(\sum_{t=1}^T \mathbf{Z}_t \left(S^u_{kt} - m_{R_{kt}}\right) / s^2_{R_{kt}} + G^{-1}_{0k} g_{0k}\right) \tag{44}$$

B.2 Data augmentation for the dRUM

The three sub-steps of step (ii) for the dRUM consist of:

(ii.a) Sample the utility differences ω^{KT} from $\pi\left(\omega^{KT}|S^T, \gamma\right) = \prod_{k \in \mathcal{K}_{-k_0}} \pi\left(\omega_{k1}, \ldots, \omega_{kT}|S^T, \gamma\right)$

(ii.b) Sample the components R^{KT} from $\pi\left(R^{KT}|\omega^{KT}, \gamma\right)$

(ii.c) Sample γ from $\pi\left(\gamma|\omega^{KT}, R^{KT}\right)$

The dRUM extension expresses the multinomial logit model as differences in the latent utilities (35)

$$s_{kt} = \mathbf{Z}'_t \gamma_k + \epsilon_{kt}, \ \epsilon_{kt} \sim \text{Logistic}, \ \forall k \in \mathcal{K}_{-k_0} \tag{45}$$

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 (35). Working with this representation would be quite involving because, in contrast to the error terms ν_{kt} in (35), the error terms ϵ_{kt} in (45) 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

$$S_t = k \Leftrightarrow S^u_{kt} > S^u_{-k,t}, \ S^u_{-k,t} = \max_{j \in \mathcal{K}_{-k}} S^u_{jt} \tag{46}$$

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 $D^{(k)}_t$:

$$\omega_{kt} = S^u_{kt} - S^u_{-k,t}, \ D^{(k)}_t = I\{\omega_{kt} > 0\}, \ \forall k \in \mathcal{K}_{-k_0} \tag{47}$$

Given the multinomial logit model for S_t, ω_{kt} has an explicit distributional form. Recall that (see footnote 6)

$$\exp\left(-S^u_{-k,t}\right) \sim \mathcal{E}\left(\sum_{j \in \mathcal{K}_{-k}} \lambda_{jt}\right) \tag{48}$$

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

$$\begin{aligned} \omega_{kt} &= S^u_{kt} - S^u_{-k,t} = \mathbf{Z}_t'\gamma_k - \log(\lambda_{-k,t}) + \nu_{k,t} - \nu_{-k,t} \\ &= \mathbf{Z}_t'\gamma_k - \log(\lambda_{-k,t}) + \epsilon_{k,t},\ D_t^{(k)} = I\{\omega_{kt} > 0\} \end{aligned} \tag{49}$$

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 $-\log(\lambda_{-k,t})$ in (49) depends only on the parameters γ_{-k}. Therefore, given $\omega_k^T = (\omega_{k1}, \dots, \omega_{kT})$ 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 (step (ii.a)) from logistic distributions.[7] Across k, we sample independently T values W_{kt} from a uniform distribution $W_{kt} \sim U[0,\ 1]$ and obtain

$$\omega_{kt} = \mathbf{Z}_t'\gamma_k - \log(\lambda_{-k,t}) + F_\epsilon^{-1}\left(D_t^{(k)} + W_{kt}\left(1 - D_t^{(k)} - \pi_{kt}\right)\right) \tag{50}$$

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.

Given ω^{KT}, the posterior of γ_k is derived based on (49), approximating the logistic distribution of the errors ϵ_{kt} by a mixture of normal distributions with M components. The components R_{kt} (step (ii.b)) are drawn from a multinomial distribution

$$P(R_{kt} = r|\omega_{kt}, \gamma_k) \propto \frac{w_r}{s_r} \exp\left\{-\frac{1}{2}\left(\frac{\omega_{kt} + \log(\lambda_{-k,t}) - \mathbf{Z}_t'\gamma_k}{s_r}\right)^2\right\} \tag{51}$$

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

Conditional on the components R_k^T, model (49) becomes normal in γ_k:

$$\tilde{\omega}_{kt} = \omega_{kt} + \log(\lambda_{-k,t}) \quad = \quad \mathbf{Z}_t'\gamma_k + \epsilon_{kt}, \quad \epsilon_{kt}|R_{kt} \sim N\left(0, s^2_{R_{kt}}\right) \tag{52}$$

Assuming a normal prior for γ_k, $\pi(\gamma_k) = N(g_{0k}, G_{0k})$, conditional on ω_k^T and R_k^T the posterior is normal, too:

$$\pi\left(\gamma_k|\omega_k^T, R_k^T\right) = N(g_k, G_k) \tag{53}$$

$$G_k = \left(\sum_{t=1}^T \mathbf{Z}_t\mathbf{Z}_t'/s^2_{R_{kt}} + G_{0k}^{-1}\right)^{-1} \tag{54}$$

$$g_k = G_k\left(\sum_{t=1}^T \mathbf{Z}_t\tilde{\omega}_{kt}/s^2_{R_{kt}} + G_{0k}^{-1}g_{0k}\right) \tag{55}$$

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

C Model identification

A more detailed description of the permutation step (iv) in the sampling scheme outlined in section 3.2 is given here, given that 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. Recall that the model (1)-(3) needs a restriction to identify the states. Given that the likelihood is invariant for a given state permutation $\rho = (\rho_1, \ldots, \rho_K)$, the same holds for the posterior:

$$\pi\left(\theta, S^T | y^T, X^T, Z^T\right) = \pi\left(\rho(\theta), \rho(S^T) | y^T, X^T, Z^T\right) \tag{56}$$

Thus the unconstrained posterior has $K!$ modes. Usually, the model is estimated by assuming a state-identifying restriction. In the sampling scheme outlined in section 3.2, this would amount to complete each iteration by re-ordering the state-dependent parameters and the states according to a restriction, e.g.

$$\beta_{j1} < \cdots < \beta_{jK} \text{ or } \gamma_{j1} < \cdots < \gamma_{jK} \tag{57}$$

for any j indicating a state-dependent parameter or one of the state-dependent parameter governing the transition distribution. This would be termed constrained permutation sampling. In this case, the specification of the hyperparameters should not be at odds with the state-identifying restrictions.

Another approach would be to sample from the unconstrained posterior, i.e. to force the sampler to visit all modes of the posterior (16) by randomly permuting the states and the state-dependent parameters at the end of each iteration (random permutation sampling). A state-identifying restriction may then be found by post-processing the MCMC output. For instance, looking at the marginal posterior distributions or scatter plots of state-dependent parameters may reveal adequate uniquely state-identifying restrictions. This procedure is useful, if the researcher has no information on which parameter(s) are significantly different across regimes or on whether there is regime-switching at all (see application sections below).

In any case, the permutation of the state-dependent parameters in the multinomial logit specification (3) needs special attention. It is best introduced by considering the example given in section 2.3. Assuming two states, $S_t \in \{1, 2\}$ and a scalar covariate determining the transition distribution, the transition probabilities are written as

$$\xi_{lk,t} = \frac{\exp\left(\mathbf{Z}_t' \gamma_k\right)}{\sum_{j=1}^{2} \exp\left(Z_t \gamma_{lj}^z + \gamma_{lj}\right)}, \; l, k = 1, 2 \tag{58}$$

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, $j = 1, 2$. Each parameter γ_k has four elements, $\gamma_k = (\gamma_{1k}^z, \gamma_{2k}^z, \gamma_{1k}, \gamma_{2k})$. For identification reasons, one of the γ_k would equal zero, $\gamma_{k_0} = 0$. If $k_0 = 1$, then

$$\gamma = \begin{bmatrix} 0 & \gamma_{12}^z \\ 0 & \gamma_{22}^z \\ 0 & \gamma_{12} \\ 0 & \gamma_{22} \end{bmatrix} \tag{59}$$

Assume that at iteration m, the sampled value of the regression coefficients would violate the pre-defined state-identifying condition $\beta_{11} < \beta_{12}$. This would imply re-ordering the states and the state-dependent parameters according to $\rho = (2\ 1)$. The constrained permutation step (iv) consists in:

$$\begin{aligned}
&\text{for the state-dependent parameters and the states}\\
&\beta_k^{(m)} := \beta_{\rho(k)}^{(m)},\ S^{T,(m)} := \rho(S^{T,(m)})\\
&\text{for the state-dependent transition parameters}\\
&\tilde{\gamma}_k^{(m)} := \left(\gamma_{\rho(k),\rho(k)}^{z(m)} \gamma_{\rho(k),\rho(k)}^{(m)}\right),\ \text{with } \gamma_1 = 0\\
&\gamma_k^{(m)} := \tilde{\gamma}_k^{(m)} - \tilde{\gamma}_1^{(m)}
\end{aligned} \tag{60}$$

For γ in (59) 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 and hyperparameters are randomly permuted in step (iv) of the sampler. For a given permutation ρ at iteration m, we permute:

$$\begin{aligned}
&\text{the state-dependent parameters and priors, states}\\
&\beta_k^{(m)} := \beta_{\rho(k)}^{(m)},\ b_{0k}, B_{0k} := b_{0\rho(k)}, B_{0\rho(k)} \qquad (61)\\
&S^{T,(m)} := \rho(S^{T,(m)})\\
&\text{state-dependent transition parameters and priors} \qquad (62)\\
&\gamma_k^{(m)} := \left(\gamma_{\rho(k),\rho(k)}^{z(m)} \gamma_{\rho(k),\rho(k)}^{(m)}\right),\ \text{with } \gamma_{k_0} = 0\\
&g_{0k} := \left(g_{0,\rho(k)\rho(k)}^z g_{0,\rho(k)0\rho(k)}\right)\\
&G_{0k} := G_{0,\rho(k)\rho(k)}
\end{aligned}$$

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:

$$\gamma_k^{(m)} := \gamma_k^{(m)} - \gamma_{k_0}^{(m)},\ \forall k \in \mathcal{K}_{-k_0},\ \text{for a chosen } k_0.$$

D 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	every 4th
– unidentified model	γ^z_{12}	0.09	0.07	0.03	0.02
	γ^z_{22}	0.16	0.03	0.01	0.01
	γ_{12}	0.13	0.07	0.03	0.02
	γ_{22}	0.07	0.04	0.01	0.01
– identified model	γ^z_{12}	3.10	1.66	0.59	0.25
	γ^z_{22}	0.95	0.25	0.10	0.04
	γ_{12}	2.29	0.65	0.33	0.16
	γ_{22}	1.14	0.53	0.12	0.07
Constrained permutation		all	every 4th	all	every 4th
– identified model	γ^z_{12}	2.99	2.21	0.46	0.37
	γ^z_{22}	1.72	0.59	0.12	0.03
	γ_{12}	1.76	1.42	0.27	0.24
	γ_{22}	1.60	1.65	0.12	0.04

[a] The last 20,000 of a total of 50,000 iterations.

Table 3: Two-pillar Phillips curve. No switching. 95% (first line) and 90% (second line) highest posterior density interval in parentheses.

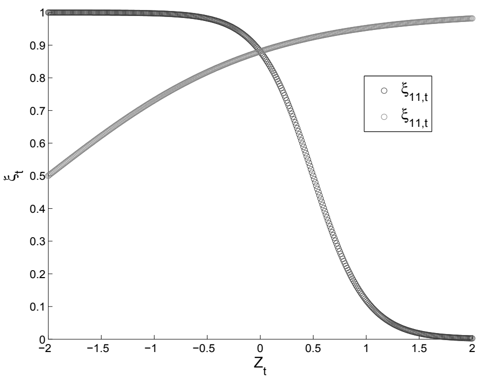
HP-filter	1970Q2-2010Q1 3 AR lags	1983Q1-2010Q1 1 AR lag
trend M3 growth	0.17	0.25
	(0.02 0.33)	(0.08 0.41)
	(0.03 0.30)	(0.11 0.39)
trend GDP growth	0.01	-0.05
	(-0.20 0.26)	(-0.27 0.18)
	(-0.18 0.21)	(-0.23 0.14)
trend change in gov. bond yield	0.48	-0.36
	(-0.06 1.03)	(-0.95 0.28)
	(0.02 0.94)	(-0.90 0.16)
cyclical output gap	0.07	0.02
	(0.03 0.12)	(-0.03 0.08)
	(0.03 0.11)	(-0.02 0.07)
	Long run effects	
trend M3 growth	0.75	0.53
	(0.19 1.34)	(0.21 0.83)
	(0.24 1.19)	(0.29 0.79)
trend GDP growth	0.13	-0.09
	(-1.00 1.41)	(-0.59 0.39)
	(-0.77 1.22)	(-0.48 0.31)
trend change in gov. bond yield	2.37	-0.75
	(-0.42 5.92)	(-1.98 0.63)
	(0.10 5.28)	(-1.88 0.33)
cyclical output gap	0.35	0.05
	(0.10 0.66)	(-0.08 0.17)
	(0.15 0.61)	(-0.05 0.15)

Table 4: Two-pillar Phillips curve. Switching in effects of trend variables. 95% (first line) and 90% (second line) highest posterior density interval in parentheses.

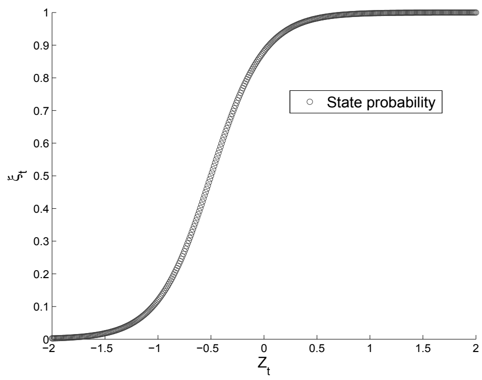
HP-filter	1983Q1-2010Q1 1 AR lag	
	Regime 1	Regime 2
trend M3 growth	0.09	0.73
	(-0.21 0.35)	(0.44 1.01)
	(-0.15 0.32)	(0.50 0.96)
trend GDP growth	0.29	-0.40
	(-0.11 0.67)	(-0.86 0.16)
	(-0.04 0.61)	(-0.82 -0.02)
trend change in gov. bond yield	0.12	-0.63
	(-0.71 0.93)	(-1.35 0.09)
	(-0.56 0.80)	(-1.21 -0.01)
cyclical output gap	0.02	
	(-0.04 0.08)	
	(-0.03 0.07)	
	Long run effects	
trend M3 growth	0.10	0.83
	(-0.23 0.41)	(0.57 1.05)
	(-0.17 0.37)	(0.64 1.03)
trend GDP growth	0.33	-0.45
	(-0.14 0.76)	(-0.99 0.23)
	(-0.05 0.69)	(-0.93 -0.00)
trend change in gov. bond yield	0.14	-0.71
	(-0.80 0.76)	(-1.54 0.08)
	(-0.67 0.91)	(-1.37 -0.04)
cyclical output gap	0.02	
	(-0.05 0.10)	
	(-0.04 0.08)	

E Figures

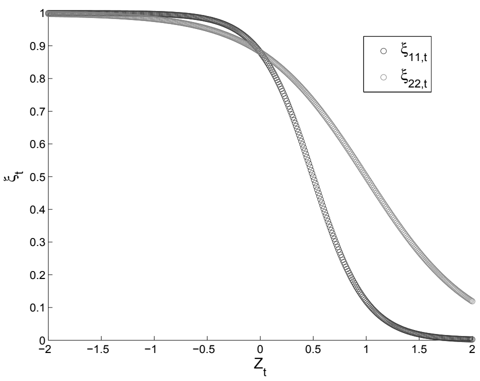
Figure 1: Some examples: Nonlinear effect of the covariate on the state persistence



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



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



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

Figure 2: Simulated data, the covariate Z_t along with state 2 (top panel) and the time series y_t (bottom panel)

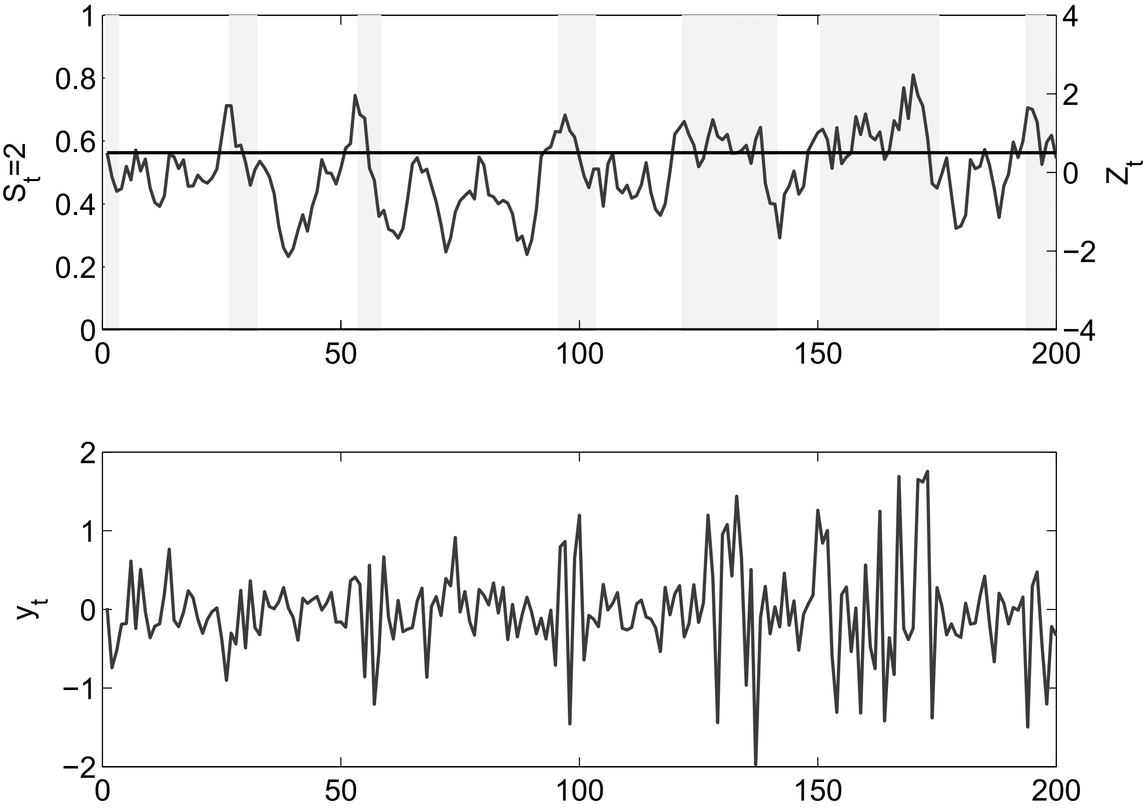
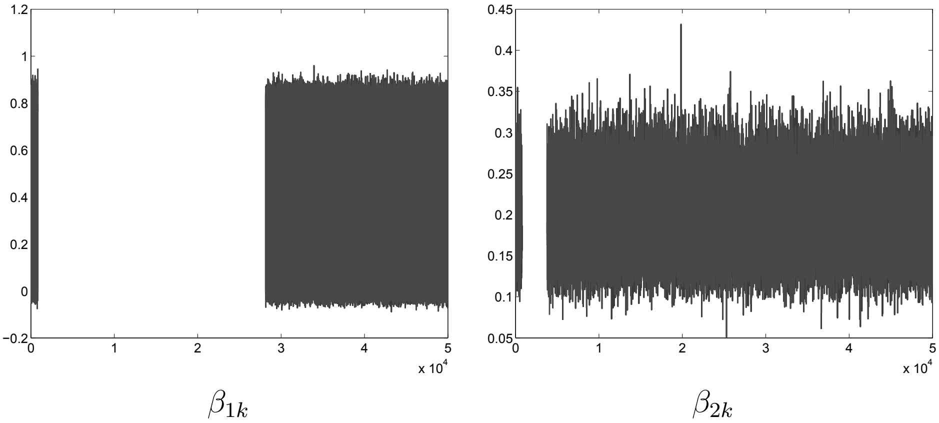


Figure 3: Random permutation with dRUM auxiliary sampling for the transition distribution. Simulated values of the regression parameters obtained from the random permutation sampler (panel (a)). State-identified simulated values (panel (b)).

(a) Random permutation sampling



(b) Identified model according to $\beta_{11} < \beta_{12}$

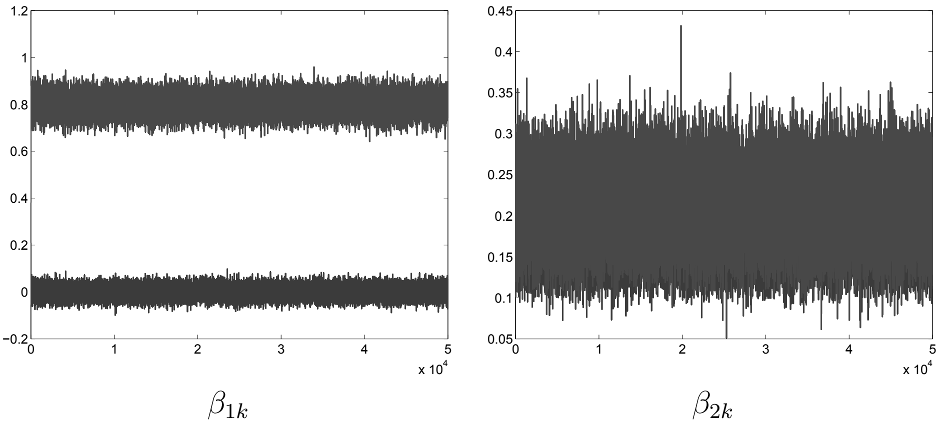


Figure 4: Random permutation with dRUM auxiliary sampling for the transition distribution. Simulated values obtained from the random permutation sampler, scatter plots of regressions parameters against constant transition parameters γ_{k2}, $k = 1, 2$.

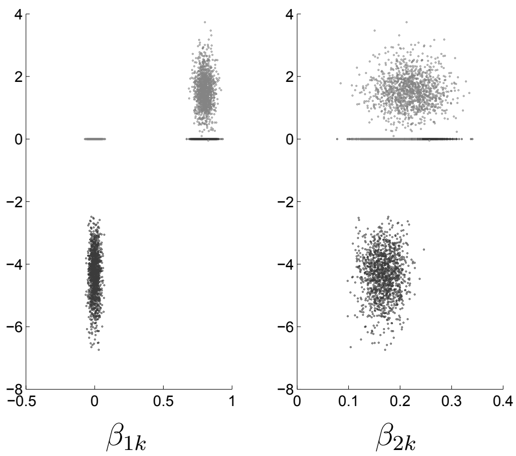
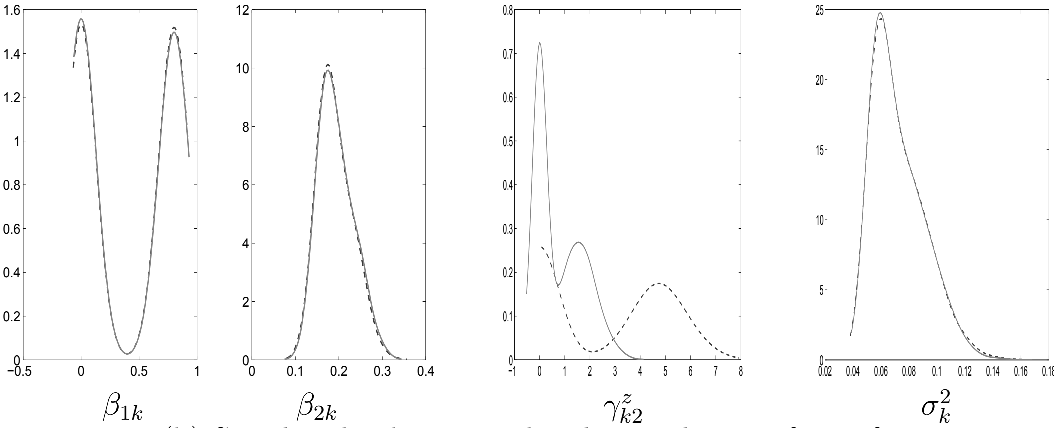


Figure 5: Random permutation with dRUM auxiliary sampling for the transition distribution. Marginal distribution of selected parameters.

(a) Simulated values obtained from the random permutation sampler



(b) Simulated values re-ordered according to $\beta_{11} < \beta_{12}$

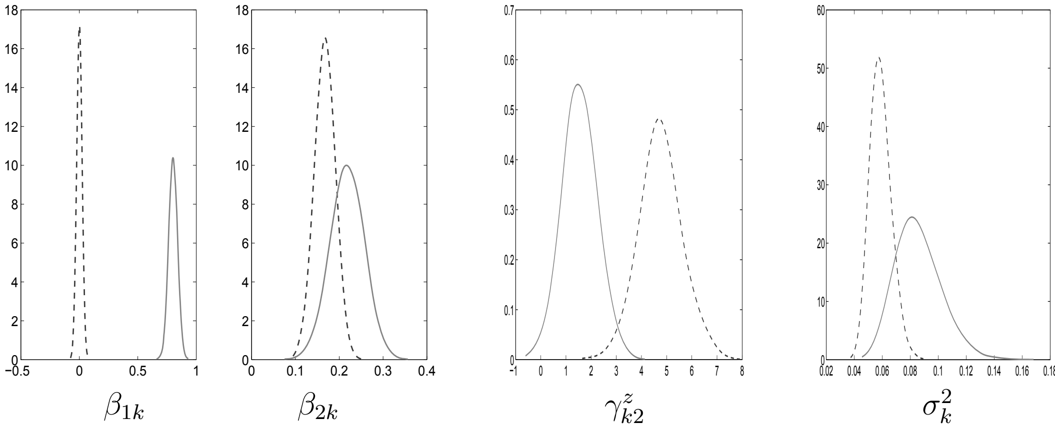


Figure 6: Marginal distribution of simulated values obtained from constrained permutation with dRUM auxiliary sampling for the transition distribution. Based on inappropriate restriction $\beta_{21} < \beta_{22}$.

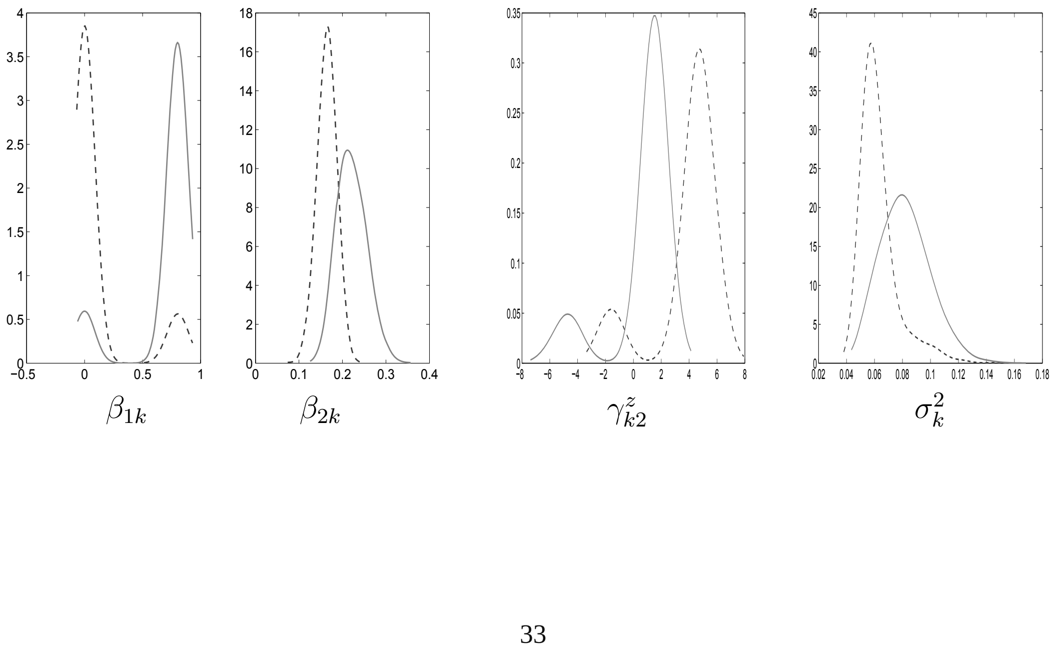


Figure 7: Recovering the threshold (0.5 in simulated data). Values and marginal distribution of the threshold level (left panels). Scatter plots of Z_t against $\xi_{11,t}^{(m)}$ (blue), $\xi_{22,t}^{(m)}$ (red) implied by the mth simulated parameter value γ_2, and of the threshold level against $\xi_{11,t}^{(m)}$ (green)

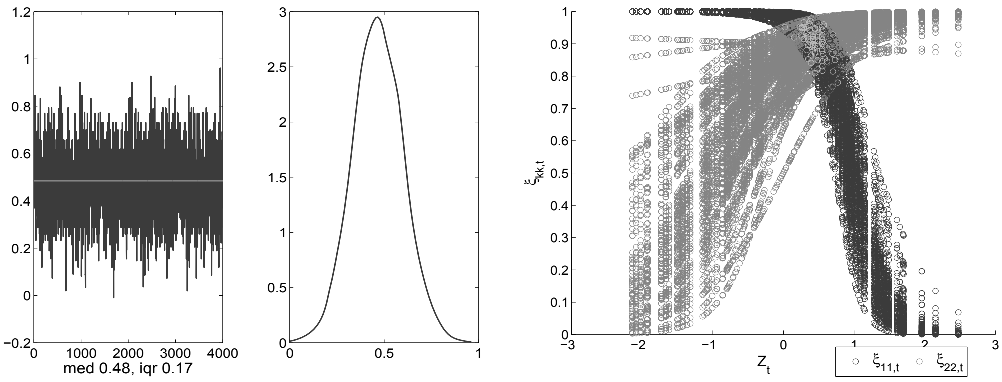
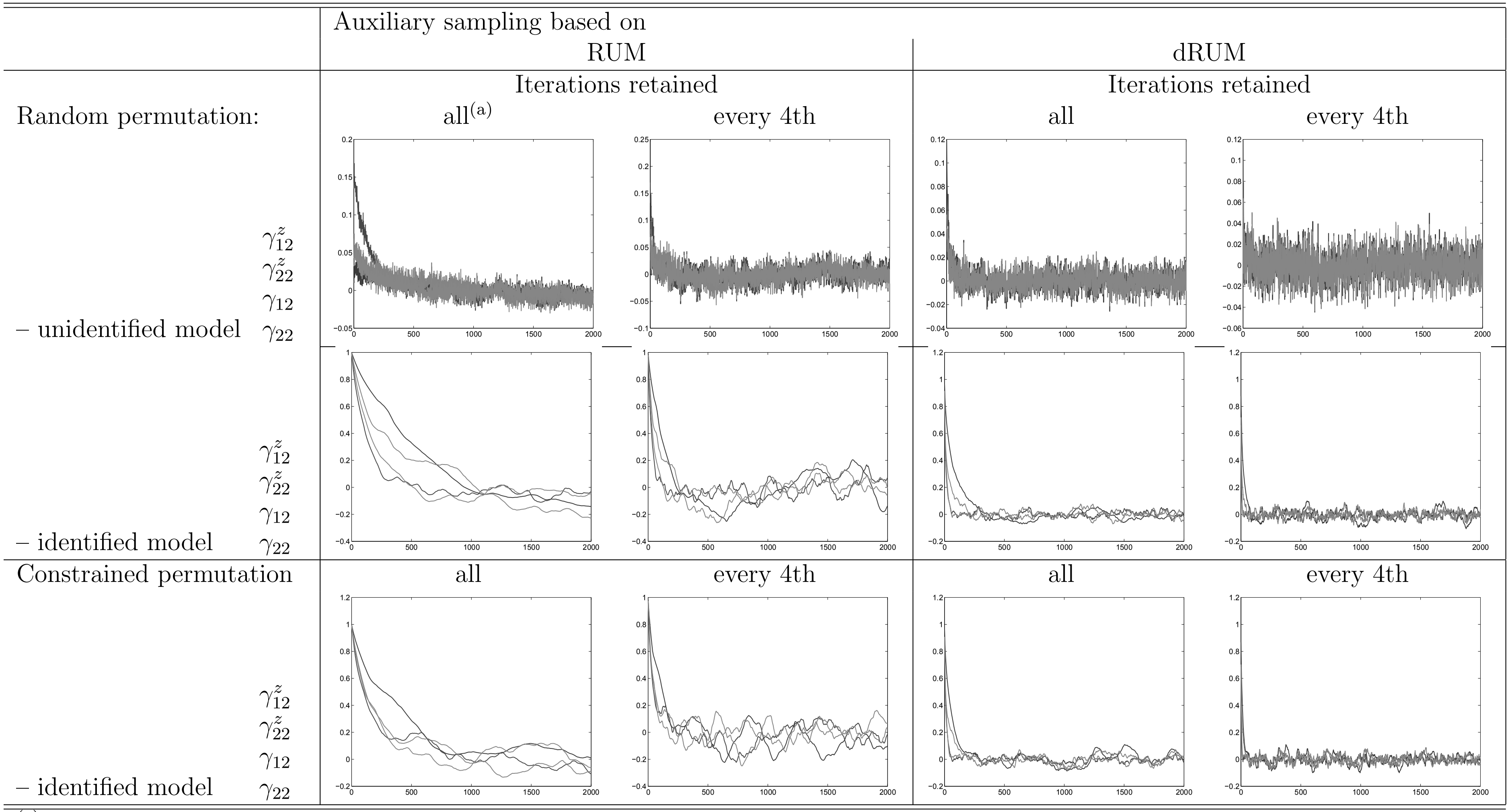


Figure 8: Simulated data. Autocorrelation function of sampled values γ.



[a] The last 20,000 of a total of 50,000 iterations.

Figure 9: M3 growth, HP-trend and low-frequency component (> 6 years).

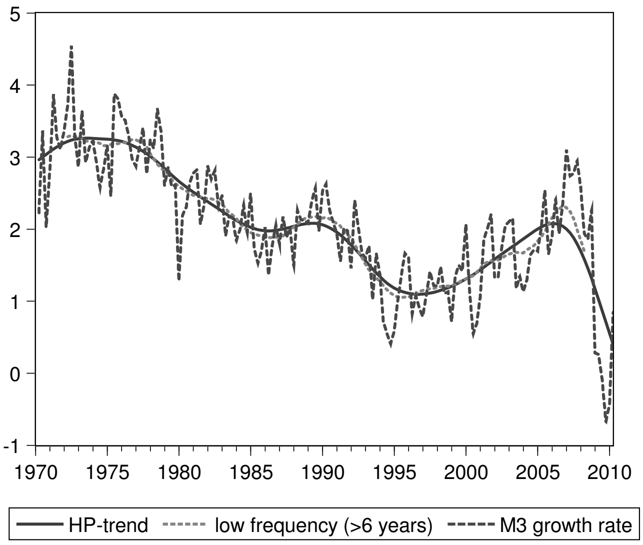


Figure 10: Scatter plot of sampled regression parameter against constant transition effect.

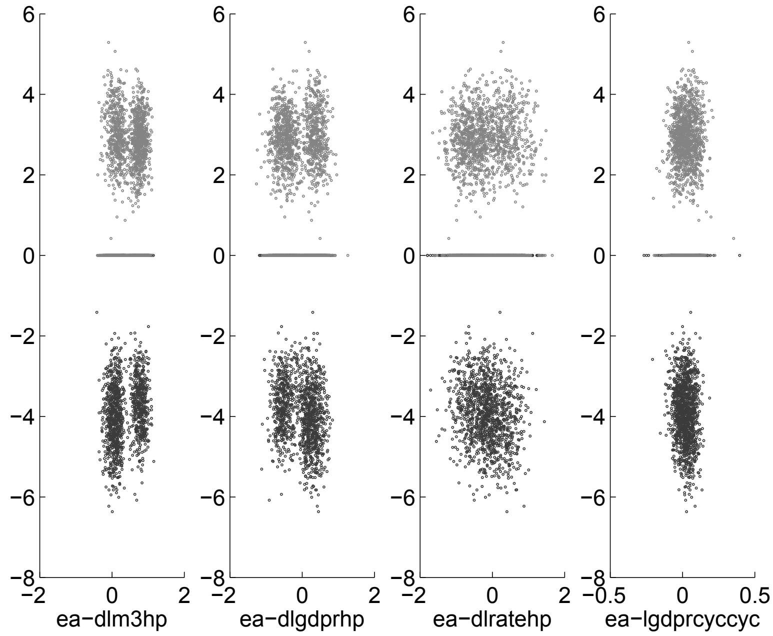


Figure 11: Marginal posterior distribution of state-identified regression coefficients (solid line, regime 2).

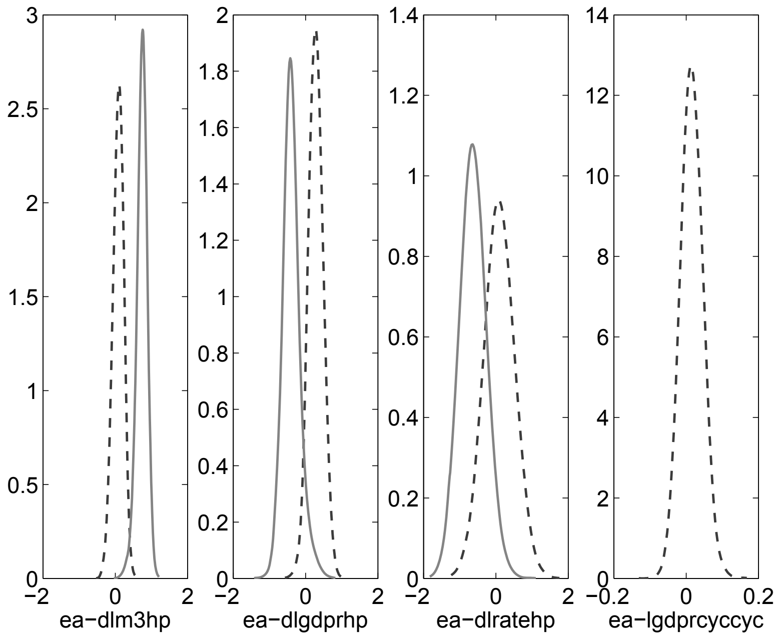


Figure 12: Marginal posterior distribution of error variance and state-identified covariate effects on the transition probability (solid line, regime 2).

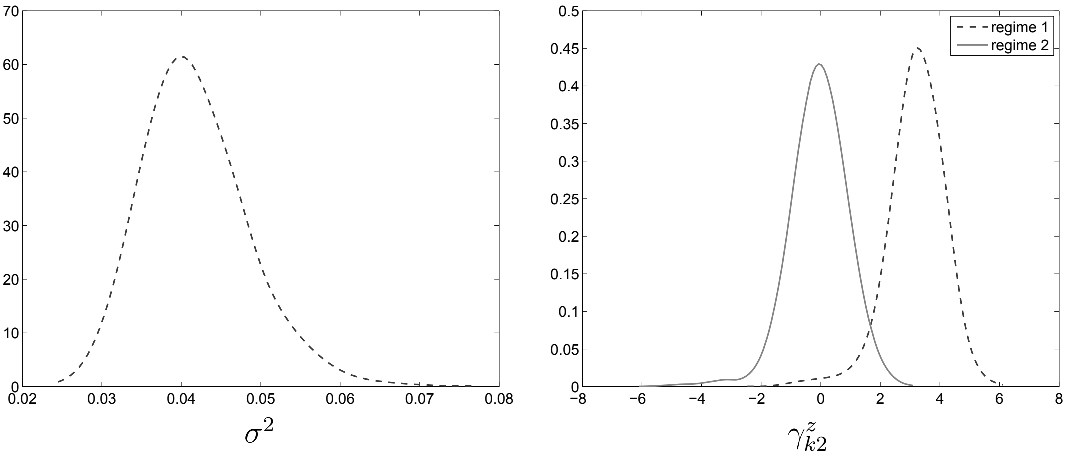


Figure 13: Posterior state probabilities along with HCIP inflation, mean-adjusted loans growth and trend M3 growth. The horizontal line corresponds to a threshold level of 2.0 % quarterly credit growth rate, composed from an average of 1.7% growth rate and an inferred 0.3% according to Definition 1 (see section ??)

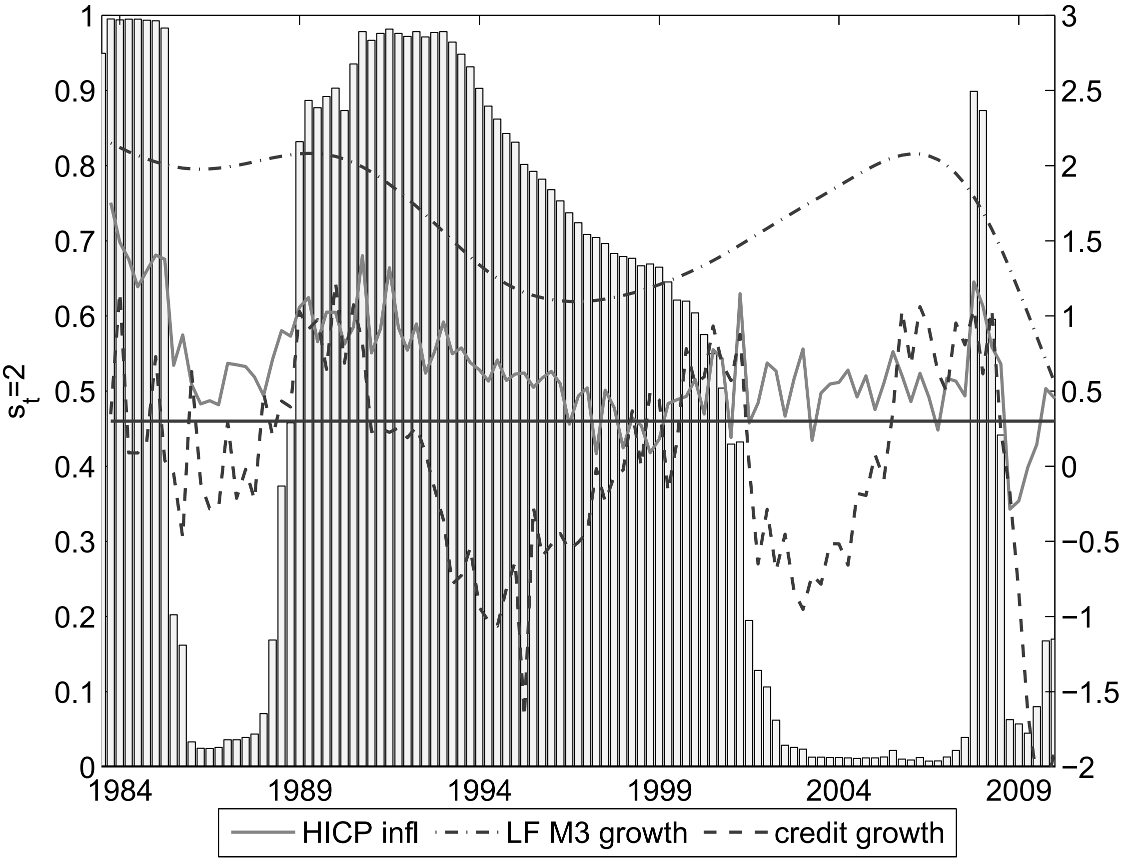
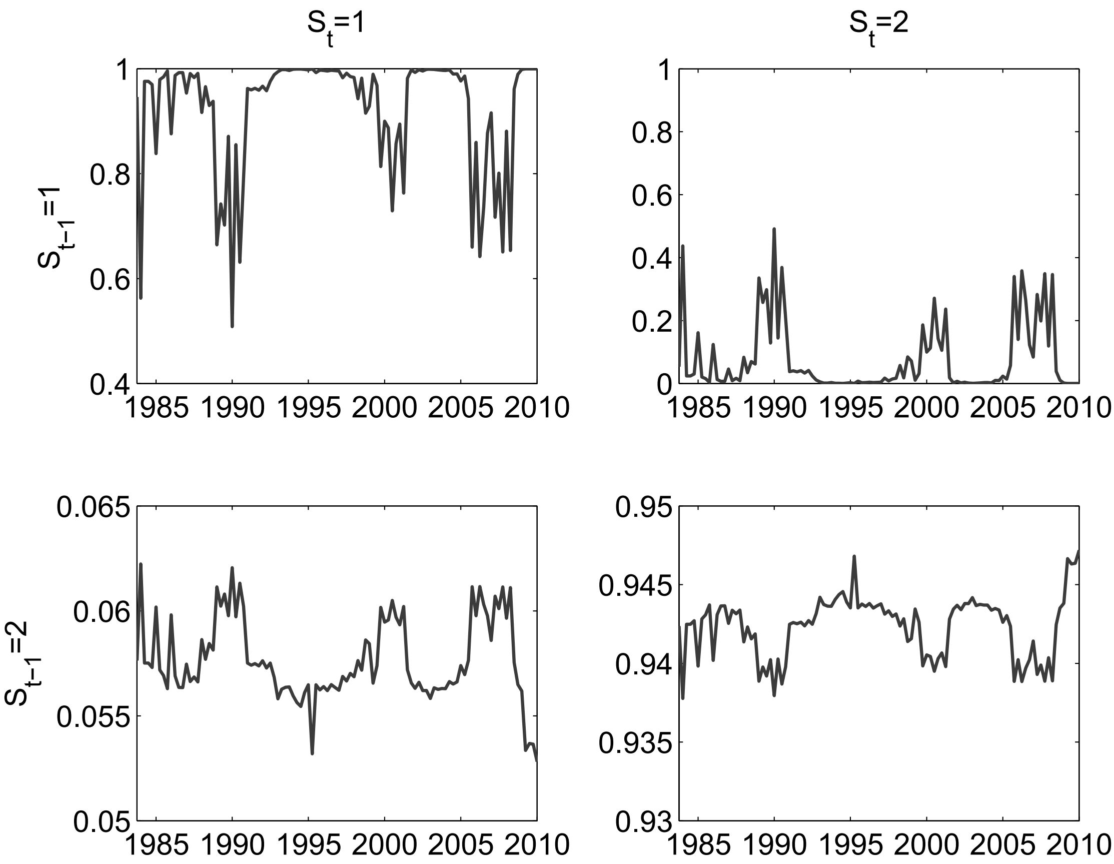


Figure 14: Median posterior transition probabilities.



Swiss National Bank Working Papers published since 2004:

2004-1 Samuel Reynard: Financial Market Participation and the Apparent Instability of Money Demand

2004-2 Urs W. Birchler and Diana Hancock: What Does the Yield on Subordinated Bank Debt Measure?

2005-1 Hasan Bakhshi, Hashmat Khan and Barbara Rudolf: The Phillips curve under state-dependent pricing

2005-2 Andreas M. Fischer: On the Inadequacy of Newswire Reports for Empirical Research on Foreign Exchange Interventions

2006-1 Andreas M. Fischer: Measuring Income Elasticity for Swiss Money Demand: What do the Cantons say about Financial Innovation?

2006-2 Charlotte Christiansen and Angelo Ranaldo: Realized Bond-Stock Correlation: Macroeconomic Announcement Effects

2006-3 Martin Brown and Christian Zehnder: Credit Reporting, Relationship Banking, and Loan Repayment

2006-4 Hansjörg Lehmann and Michael Manz: The Exposure of Swiss Banks to Macroeconomic Shocks – an Empirical Investigation

2006-5 Katrin Assenmacher-Wesche and Stefan Gerlach: Money Growth, Output Gaps and Inflation at Low and High Frequency: Spectral Estimates for Switzerland

2006-6 Marlene Amstad and Andreas M. Fischer: Time-Varying Pass-Through from Import Prices to Consumer Prices: Evidence from an Event Study with Real-Time Data

2006-7 Samuel Reynard: Money and the Great Disinflation

2006-8 Urs W. Birchler and Matteo Facchinetti: Can bank supervisors rely on market data? A critical assessment from a Swiss perspective

2006-9 Petra Gerlach-Kristen: A Two-Pillar Phillips Curve for Switzerland

2006-10 Kevin J. Fox and Mathias Zurlinden: On Understanding Sources of Growth and Output Gaps for Switzerland

2006-11 Angelo Ranaldo: Intraday Market Dynamics Around Public Information Arrivals

2007-1 Andreas M. Fischer, Gulzina Isakova and Ulan Termechikov: Do FX traders in Bishkek have similar perceptions to their London colleagues? Survey evidence of market practitioners' views

2007-2 Ibrahim Chowdhury and Andreas Schabert: Federal Reserve Policy viewed through a Money Supply Lens

2007-3 Angelo Ranaldo: Segmentation and Time-of-Day Patterns in Foreign Exchange Markets

2007-4 Jürg M. Blum: Why 'Basel II' May Need a Leverage Ratio Restriction

2007-5 Samuel Reynard: Maintaining Low Inflation: Money, Interest Rates, and Policy Stance

2007-6 Rina Rosenblatt-Wisch: Loss Aversion in Aggregate Macroeconomic Time Series

2007-7 Martin Brown, Maria Rueda Maurer, Tamara Pak and Nurlanbek Tynaev: Banking Sector Reform and Interest Rates in Transition Economies: Bank-Level Evidence from Kyrgyzstan

2007-8 Hans-Jürg Büttler: An Orthogonal Polynomial Approach to Estimate the Term Structure of Interest Rates

2007-9 Raphael Auer: The Colonial Origins Of Comparative Development: Comment. A Solution to the Settler Mortality Debate

2007-10 Franziska Bignasca and Enzo Rossi: Applying the Hirose-Kamada filter to Swiss data: Output gap and exchange rate pass-through estimates

2007-11 Angelo Ranaldo and Enzo Rossi: The reaction of asset markets to Swiss National Bank communication

2007-12 Lukas Burkhard and Andreas M. Fischer: Communicating Policy Options at the Zero Bound

2007-13 Katrin Assenmacher-Wesche, Stefan Gerlach, and Toshitaka Sekine: Monetary Factors and Inflation in Japan

2007-14 Jean-Marc Natal and Nicolas Stoffels: Globalization, markups and the natural rate of interest

2007-15 Martin Brown, Tullio Jappelli and Marco Pagano: Information Sharing and Credit: Firm-Level Evidence from Transition Countries

2007-16 Andreas M. Fischer, Matthias Lutz and Manuel Wälti: Who Prices Locally? Survey Evidence of Swiss Exporters

2007-17 Angelo Ranaldo and Paul Söderlind: Safe Haven Currencies

2008-1 Martin Brown and Christian Zehnder: The Emergence of Information Sharing in Credit Markets

2008-2 Yvan Lengwiler and Carlos Lenz: Intelligible Factors for the Yield Curve

2008-3 Katrin Assenmacher-Wesche and M. Hashem Pesaran: Forecasting the Swiss Economy Using VECX* Models: An Exercise in Forecast Combination Across Models and Observation Windows

2008-4 Maria Clara Rueda Maurer: Foreign bank entry, institutional development and credit access: firm-level evidence from 22 transition countries

2008-5 Marlene Amstad and Andreas M. Fischer: Are Weekly Inflation Forecasts Informative?

2008-6 Raphael Auer and Thomas Chaney: Cost Pass Through in a Competitive Model of Pricing-to-Market

2008-7 Martin Brown, Armin Falk and Ernst Fehr: Competition and Relational Contracts: The Role of Unemployment as a Disciplinary Device

2008-8 Raphael Auer: The Colonial and Geographic Origins of Comparative Development

2008-9 Andreas M. Fischer and Angelo Ranaldo: Does FOMC News Increase Global FX Trading?

2008-10 Charlotte Christiansen and Angelo Ranaldo: Extreme Coexceedances in New EU Member States' Stock Markets

2008-11 Barbara Rudolf and Mathias Zurlinden: Measuring capital stocks and capital services in Switzerland

2008-12 Philip Sauré: How to Use Industrial Policy to Sustain Trade Agreements

2008-13 Thomas Bolli and Mathias Zurlinden: Measuring growth of labour quality and the quality-adjusted unemployment rate in Switzerland

2008-14 Samuel Reynard: What Drives the Swiss Franc?

2008-15 Daniel Kaufmann: Price-Setting Behaviour in Switzerland – Evidence from CPI Micro Data

2008-16 Katrin Assenmacher-Wesche and Stefan Gerlach: Financial Structure and the Impact of Monetary Policy on Asset Prices

2008-17 Ernst Fehr, Martin Brown and Christian Zehnder: On Reputation: A Microfoundation of Contract Enforcement and Price Rigidity

2008-18 Raphael Auer and Andreas M. Fischer: The Effect of Low-Wage Import Competition on U.S. Inflationary Pressure

2008-19 Christian Beer, Steven Ongena and Marcel Peter: Borrowing in Foreign Currency: Austrian Households as Carry Traders

2009-1 Thomas Bolli and Mathias Zurlinden: Measurement of labor quality growth caused by unobservable characteristics

2009-2 Martin Brown, Steven Ongena and Pinar Yeşin: Foreign Currency Borrowing by Small Firms

2009-3 Matteo Bonato, Massimiliano Caporin and Angelo Ranaldo: Forecasting realized (co)variances with a block structure Wishart autoregressive model

2009-4 Paul Söderlind: Inflation Risk Premia and Survey Evidence on Macroeconomic Uncertainty

2009-5 Christian Hott: Explaining House Price Fluctuations

2009-6 Sarah M. Lein and Eva Köberl: Capacity Utilisation, Constraints and Price Adjustments under the Microscope

2009-7 Philipp Haene and Andy Sturm: Optimal Central Counterparty Risk Management

2009-8 Christian Hott: Banks and Real Estate Prices

2009-9 Terhi Jokipii and Alistair Milne: Bank Capital Buffer and Risk Adjustment Decisions

2009-10 Philip Sauré: Bounded Love of Variety and Patterns of Trade

2009-11 Nicole Allenspach: Banking and Transparency: Is More Information Always Better?

2009-12 Philip Sauré and Hosny Zoabi: Effects of Trade on Female Labor Force Participation

2009-13 Barbara Rudolf and Mathias Zurlinden: Productivity and economic growth in Switzerland 1991-2005

2009-14 Sébastien Kraenzlin and Martin Schlegel: Bidding Behavior in the SNB's Repo Auctions

2009-15 Martin Schlegel and Sébastien Kraenzlin: Demand for Reserves and the Central Bank's Management of Interest Rates

2009-16 Carlos Lenz and Marcel Savioz: Monetary determinants of the Swiss franc

2010-1 Charlotte Christiansen, Angelo Ranaldo and Paul Söderlind: The Time-Varying Systematic Risk of Carry Trade Strategies

2010-2 Daniel Kaufmann: The Timing of Price Changes and the Role of Heterogeneity

2010-3 Loriano Mancini, Angelo Ranaldo and Jan Wrampelmeyer: Liquidity in the Foreign Exchange Market: Measurement, Commonality, and Risk Premiums

2010-4 Samuel Reynard and Andreas Schabert: Modeling Monetary Policy

2010-5 Pierre Monnin and Terhi Jokipii: The Impact of Banking Sector Stability on the Real Economy

2010-6 Sébastien Kraenzlin and Thomas Nellen: Daytime is money

2010-7 Philip Sauré: Overreporting Oil Reserves

2010-8 Elizabeth Steiner: Estimating a stock-flow model for the Swiss housing market

2010-9 Martin Brown, Steven Ongena, Alexander Popov, and Pinar Yeşin: Who Needs Credit and Who Gets Credit in Eastern Europe?

2010-10 Jean-Pierre Danthine and André Kurmann: The Business Cycle Implications of Reciprocity in Labor Relations

2010-11 Thomas Nitschka: Momentum in stock market returns: Implications for risk premia on foreign currencies

2010-12 Petra Gerlach-Kristen and Barbara Rudolf: Macroeconomic and interest rate volatility under alternative monetary operating procedures

2010-13 Raphael Auer: Consumer Heterogeneity and the Impact of Trade Liberalization: How Representative is the Representative Agent Framework?

2010-14 Tommaso Mancini Griffoli and Angelo Ranaldo: Limits to arbitrage during the crisis: funding liquidity constraints and covered interest parity

2010-15 Jean-Marc Natal: Monetary Policy Response to Oil Price Shocks

2010-16 Kathrin Degen and Andreas M. Fischer: Immigration and Swiss House Prices

2010-17 Andreas M. Fischer: Immigration and large banknotes

2010-18 Raphael Auer: Are Imports from Rich Nations Deskilling Emerging Economies? Human Capital and the Dynamic Effects of Trade

2010-19 Jean-Pierre Danthine and John B. Donaldson: Executive Compensation: A General Equilibrium Perspective

2011-1 Thorsten Beck and Martin Brown: Which Households Use Banks? Evidence from the Transition Economies

2011-2 Martin Brown, Karolin Kirschenmann and Steven Ongena: Foreign Currency Loans – Demand or Supply Driven?

2011-3 Victoria Galsband and Thomas Nitschka: Foreign currency returns and systematic risks

2011-4 Francis Breedon and Angelo Ranaldo: Intraday patterns in FX returns and order flow

2011-5 Basil Guggenheim, Sébastien Kraenzlin and Silvio Schumacher: Exploring an uncharted market: Evidence on the unsecured Swiss franc money market

2011-6 Pamela Hall: Is there any evidence of a Greenspan put?

2011-7 Daniel Kaufmann and Sarah Lein: Sectoral Inflation Dynamics, Idiosyncratic Shocks and Monetary Policy

2011-8 Iva Cecchin: Mortgage Rate Pass-Through in Switzerland

2011-9 Raphael A. Auer, Kathrin Degen and Andreas M. Fischer: Low-Wage Import Competition, Inflationary Pressure, and Industry Dynamics in Europe

2011-10 Raphael A. Auer and Philip Sauré: Spatial Competition in Quality, Demand-Induced Innovation, and Schumpeterian Growth

2011-11 Massimiliano Caporin , Angelo Ranaldo and Paolo Santucci de Magistris: On the Predictability of Stock Prices: a Case for High and Low Prices

2011-12 Jürg Mägerle and Thomas Nellen: Interoperability between central counterparties

2011-13 Sylvia Kaufmann: K-state switching models with endogenous transition distributions