STUDY CENTER GERZENSEE

Hidden Markov models in time series, with applications in economics

Sylvia Kaufmann

Working Paper 16.06

Hidden Markov models in time series, with applications in economics

Sylvia Kaufmann*†

September 2016

Abstract

Markov models introduce persistence in the mixture distribution. In time series analysis, the mixture components relate to different persistent states characterizing the state-specific time series process. Model specification is discussed in a general form. Emphasis is put on the functional form and the parametrization of time-invariant and time-varying specifications of the state transition distribution. The concept of mean-square stability is introduced to discuss the condition under which Markov switching processes have finite first and second moments in the indefinite future. Not surprisingly, a time series process may be mean-square stable even if it switches between bounded and unbounded state-specific processes. Surprisingly, switching between stable state-specific processes is neither necessary nor sufficient to obtain a mean-square stable time series process. Model estimation proceeds by data augmentation. We derive the basic forward-filtering backward-smoothing/sampling algorithm to infer on the latent state indicator in maximum likelihood and Bayesian estimation procedures. Emphasis is again laid on the state transition distribution. We discuss the specification of state-invariant prior parameter distributions and posterior parameter inference under either a logit or probit functional form of the state transition distribution. With simulated data, we show that the estimation of parameters under a probit functional form is more efficient. However, a probit functional form renders estimation extremely slow if more than two states drive the time series process. Finally, various applications illustrate how to obtain informative switching in Markov switching models with time-invariant and time-varying transition distributions.

JEL classification: C11,C22,C24,C32,C34,E24,E32,E52

Key words: Bayesian inference, EM algorithm, Markov switching, prior information.

*Study Center Gerzensee, Foundation of the Swiss National Bank, Dorfstrasse 2, P.O. Box 21, CH-3115 Gerzensee, sylvia.kaufmann@szgerzensee.ch

†I thank Markus Pape for valuable comments and suggestions. All remaining errors and omissions are mine.

1 Introduction

Hidden Markov models are mixture models with sequential dependence or persistence in the mixture distribution. For a finite, discrete number of G components, persistence in distribution is induced by specifying a latent component indicator which follows a Markov process. The transition probabilities for the Markov process may either be time-invariant or time-varying. In the latter case, Markov models extend mixture of experts model (see chapter II.5 of this volume) by introducing persistence in the mixtures.

Hidden Markov models in time series econometrics became very popular after the publications of Hamilton (1989, 1990). He transferred earlier regression based approaches like Goldfeld and Quandt (1973) into time series analysis by recognizing their usefulness in capturing asymmetric conditional moments or asymmetric dynamic properties of time series. In section 2 we start by setting out the framework and terminology. In time series analysis, components are usually called states or regimes, and the transition between states is termed regime switch or regime change. This wording will be used in this chapter to be consistent with the econometrics literature. We discuss in separate sections the basic modelling choice of specifying the transition distribution of states. Hamilton (1989, 1990) introduced the model with time-invariant or constant transition distribution, and most of the following literature stayed with this specification.

This is not as restrictive as it may seem at first sight, given that more sophisticated models can be built by imposing either restrictions on the state transition probabilities or by combining multiple latent state indicators in a dynamical or hierarchical way. Change-point models (Chib 1996, Pesaran et al. 2007, Bauwens et al. 2015) are nested in Markov switching models by imposing appropriate zero restrictions on the transition distribution. Linking multiple latent state indicators dynamically, we can capture many leading/lagging features in multivariate analysis (Phillips 1991, Paap et al. 2009, Kaufmann 2010). Linking state indicators hierarchically, we obtain hierarchical Markov mixture models e.g. to disentangle long-term from short-term changing dynamics (Geweke and Amisano 2011, Bai and Wang 2011). Nevertheless, constant or exogenous transition distributions do not incorporate an explicit explanation or interpretation of the driving forces underlying the transition distribution.

Including covariates effects into the transition distribution renders it time-varying and yields at least an indication, if not a driving cause, of the regime switches. One of the first proposals is Diebold et al. (1994). Applications followed in business cycle analysis in Filardo (1994) and Filardo and Gordon (1998). Both probit and logit functional forms were used for the transition distribution. Under the assumption of independence between state alternatives, both parameterizations yield essentially the same estimation results. Later on, Koop and Potter (2007) introduced duration dependent time-varying probabilities into a change-point model. An interesting alternative is presented in Billio and Casarin (2011), who use a beta autoregressive process to model time-varying transition probabilities.

Against this background, we outline various extensions that are available within the general framework we present. Given that covariates may have state-dependent effects on the transition distribution, we elaborate on various considerations that may flow into the specific parametrization of time-varying transition probabilities. Section 2 closes with the discussion of an attractive feature of Markov switching models that has so far, to our knowledge, not been exploited in time series analysis. So far, these models have been applied under the assumption that the conditional, i.e. state-dependent, distributions are stationary or, in other words, have finite moments in every period t. This need not be the case, however. Many real phenomena are consistent with a process that alternates between a stationary and a non-stationary state-specific distribution. Think of the recent financial crisis, during which dynamics across economic variables may have engaged transitorily on an unsustainable path. Francq and Zakoïan (2001), and more recently Farmer et al. (2009a), derive conditions under which the unconditional distribution of multivariate time series processes in the indefinite future has finite moments, even if some state- and period-specific conditional distributions may not have finite moments. Moreover, they show that state-specific stationary distributions are not sufficient for a multivariate process to approach finite moments in the indefinite future.

In section 3 we outline the estimation of Markov switching models, where the emphasis is on Bayesian estimation. Maximum likelihood estimation and variants of it are based on the EM algorithm, in which the 'E' step takes explicitly into account the state-dependence in the mixture to infer about the state indicator (Hamilton 1990). Extensions to multivariate models followed in Krolzig et al. (2002) and Clements and Krolzig (2003). The forward-filtering backward-sampling algorithm provides the basis for data augmentation in Bayesian estimation (McCulloch and Tsay 1994; Chib 1996). Markov chain Monte Carlo methods prove very useful to estimate models with multiple latent variables, like factor models with Markov switching factor mean or factor volatility (Kim and Nelson 1998).

Hidden Markov models endorse all issues concerning mixture modelling, as comprehensively exposed in Frühwirth-Schnatter (2006). In the present chapter, we therefore discuss in detail the design of state-invariant prior distributions for time-invariant and time-varying transition probabilities (Kaufmann 2015; Burgette and Hahn 2010). We then set out the posterior random permutation sampler to obtain draws from the unconstrained, multimodal posterior (Frühwirth-Schnatter 2001). To sample the parameters of the logit functional form, we borrow from data augmentation algorithms outlined in Frühwirth-Schnatter and Frühwirth (2010) which render the non-linear, non-Gaussian model in latent utilities linear Gaussian. Parameters are sampled from full conditional distributions rather than by Metropolis-Hastings (Scott 2011; Holmes and Held 2006). The approach of Burgette and Hahn (2010) proves very useful to sample parameters of the probit functional form. Instead of normalizing the error covariance of latent utilities with respect to a specific element (McCulloch et al. 2000; Imai and van Dyk 2005), they propose to restrict the trace of the normalized error covariance of the latent utilities, whereby normalization occurs in each iteration of the sampler with respect to a randomly chosen latent state utility. To conclude section 3, we compare estimation time and sampler efficiency between using the logit and the probit functional form to estimate the data generating process of a univariate series driven by 2 and 3 hidden Markov mixtures. We briefly illustrate that posterior state-identification is obtained by post-processing the posterior draws.

In this chapter, we do not discuss prior design and model choice with respect to the number of regimes. The same considerations as outlined in chapter I.7 of this volume apply to Markov mixture models in time series analysis and the interested reader may refer to it. In brief, model choice with respect to the number of regimes can be addressed by means of marginal likelihood (Chib 1995; Frühwirth-Schnatter 2004). In the maximum likelihood framework, the issue can only be addressed in a proper statistical way by simulating the test statistic, see Hansen (1992). The likelihood ratio statistic violates regularity conditions, because models with different number of states are not nested within each other. For similar reasons, the widely used information criteria are not an alternative, either, or at least should be used with more care than usually done.

In section 4, by discussing informative regime switching we illustrate how explicit economic interpretations of results are obtained from posterior inference. For example, structural restrictions on time-invariant transition probabilities yield explicit interpretations about dynamic relationships across variables. One of the first contributions is Phillips (1991) who analyzed country-specific output series in a multivariate setting. Recently, Sims et al. (2008) proposed a general framework to implement and estimate restricted transition distributions in large multiple equation systems. Including covariates effects into the transition distribution provides an explicit interpretation of the driving factors underlying the latent state indicator. Additionally, prior knowledge may flow into the parametrization of the transition distribution by imposing parameter restrictions (Gaggl and Kaufmann 2014, Bäurle et al. 2016). In the latter case, this induces a restricted, state-identified prior and may call for some restricted estimation procedures. The list of papers used for illustration is by far non-exhaustive and refers mainly to business cycle analysis. Nevertheless, the provided examples are straightforward to apply in other areas like financial econometrics (Hamilton and Susmel 1994; Bauwens and Lubrano 1998). Finally, section 5 concludes the chapter.

The methods discussed in this chapter apply generally to Markov switching models if the dependence on past states is fixed. Models with infinite dependence on past states, like in regime switching generalized ARCH models, are not treated in this chapter. The interested reader may refer to the specific literature (Klaassen 2002; Gray 1996; Bauwens et al. 2014) and to chapter III.5 of this volume for an overview. Forecasting is not treated in this chapter, either. The interested reader may refer to e.g. Elliott and Timmermann (2005), Pesaran et al. (2007) and Chauvet and Piger (2008). Scenario-based forecasting is used in Kaufmann and Kugler (2010).

2 Regime switching – Mixture modelling over time

2.1 Preliminaries and model specification

Hidden Markov models or Markov switching models are mixture models with typical feature of sequential (time) persistence in the mixture distribution. These models are often applied in time series analysis, where a scalar or a vector of observations is denoted by y_t, $t = 1, \ldots, T$, and t indexes the observation period.

In a general model with period-specific observation densities

$$y_t \sim f(y_t|x_{1t}, \theta_t) \tag{1}$$

persistence is introduced by assuming a time-dependent process for θ_t, $\theta_t|x_{2t}, \theta_{t-1}$. In (1), x_{1t} denotes covariates which may also include lagged observations of y_t. In the present chapter, x_{2t} denotes covariates that influence the transition distribution of the parameters. In time series analysis, the mixture components are called states, and in hidden Markov models one typically assumes that the set of parameter states is discrete, $\theta_t \in \{\theta_1 \ldots, \theta_G\}$. The latent component indicator $z_t \in \{1, \ldots, G\}$ is called the state indicator and the binary indicator is defined by $z_{tg} = 1$ iff $z_t = g$. Conditional on z_t, $\theta_t|z_t = \sum_{g=1}^{G} z_{tg}\theta_g = \theta_{z_t}$ and $y_t|z_t \sim f\left(y_t|x_{1t}, \theta_{z_t}\right)$.

State persistence is introduced by formulating a Markov process for z_t:

$$P(z_t = g|z_{t-1} = g', x_{2t}) = \eta_{t,g'g} \tag{2}$$

with $\sum_{g'=1}^{G} \eta_{t,g'g} = 1$. In the most general specification, covariates x_{2t} render the state transition probabilities $\eta_{t,g'g}$ time-specific or time-varying. Hidden Markov models thus extend mixture of experts models by introducing persistence in the mixture distribution.

In this chapter, we will denote the set of states by $\mathcal{G} = \{1, \ldots, G\}$. Specific functional forms of state transition call for specific identification restrictions, which are usually set on a reference state. We will denote this reference state by g_0. Finally, we define $\mathcal{G}_{-g} = \mathcal{G} \backslash g$.

2.2 The functional form of state transition

Different functional forms are available to model Markov state transition η_t, and each of them needs careful specification. In particular, to extract the information of interest from the data, the researcher has to form expectations about parameterizations to shape the functional form in a sensible way, and, if of interest, to design state-invariant prior distributions. Therefore, in this and other sections, we discuss each functional form in turn.

2.2.1 Time-invariant switching

The simplest way to parameterize time-invariant switching

$$P(z_t = g|z_{t-1} = g') = \eta_{g'g} \tag{3}$$

is to define $\eta_{g'g}$ directly as transition probability. In this setup, one has to ensure that $0 < \eta_{g'g} < 1$ while estimating the model. Persistence probabilities have to lie strictly between 0 and 1 to avoid absorbing states. If absorbing states are present in the data, they should follow non-absorbing states, to be able to identify the state-specific parameters of the latter ones. In change-point models with a finite and an infinite number of change-points (Chib 1998; Pesaran et al. 2007; Koop and Potter 2007), $\eta_{g'g} = 0$ for $g < g'$. They represent a sequence of non-recurrent states, where after a switch to state g no recurrence to state $g' < g$ is allowed any more.

The alternative is to work with a logit functional form

$$P(z_t = g|z_{t-1} = g') = \eta_{g'g} = \frac{\exp\left(\gamma_{g'g}\right)}{\sum_{j=1}^{G} \exp\left(\gamma_{jg}\right)}$$

in which, for identification purposes, $\gamma_{g'g_0} = 0$ for some reference state $g_0 \in \mathcal{G}$. No restriction on $\gamma = \{\gamma_{g'g}|g' \in \mathcal{G}, g \in \mathcal{G}_{-g_0}\}$ is needed to ensure that the transition probabilities lie between 0 and 1. In general, working with a functional form has also the advantage that covariates can be included to design time-varying or *informative* regime switching.

2.2.2 Time-varying switching

To design time-varying switching, we introduce covariates in the transition distribution $\eta_{t,g'g} = \eta(x_{2t}, \gamma)$. Depending on the functional form, different restrictions are imposed on γ for identification purposes. In the following, we work with a scalar notation of x_{2t}. With appropriate adjustments, the generalization to a vector of covariates is straightforward.

The logit functional form with covariates writes

$$\eta_{t,g'g} = \frac{\exp\left(x_{2t}\gamma^x_{g'g} + \gamma_{g'g}\right)}{\sum_{j=1}^{G} \exp\left(x_{2t}\gamma^x_{g'g} + \gamma_{g'g}\right)} \tag{4}$$

where, for identification purposes we impose $\gamma_{g'g_0} = 0$ for some reference state $g_0 \in \mathcal{G}$, see e.g. Diebold, Lee, and Weinbach (1994) for an early contribution.

An alternative is the probit functional form

$$\eta_{t,g'g} = \Phi\left(x_{2t}\gamma^x_{g'g} + \gamma_{g'g}\right) \tag{5}$$

where $\Phi(x) = \int_{-\infty}^{x} \phi(u)du$ is the cumulative distribution function with respect to the standard normal density $\phi(\cdot)$. For $G = 2$, the restriction $\gamma_{g'g_0} = -\gamma_{g'g}$ provides identification, see Filardo (1994) and Filardo and Gordon (1998) for early contributions in economics.

Remark 1: In both functional forms, we do not fix the reference state to $g_0 = 1$ as is usually done. In estimation, this generalization allows us to apply the permutation sampler to the reference state as well.

Remark 2: When $G = 2$, the advantage of the probit specification is obviously that the associated latent random utility model is standard normal

$$z^*_{gt} = x_{2t}\gamma^x_{z_{t-1}g} + \gamma_{z_{t-1}g} + \nu_{gt},\ \nu_{gt} \sim N(0, 1) \tag{6}$$

for $g \neq g_0$, which renders parameter estimation straightforward.

The situation is more intricate for multi-state regime switching models. Conditional on the state indicator z, parameter estimation in a multinomial probit model for η_t is not the issue (McCulloch et al. 2000; Imai and van Dyk 2005; Nobile 1998; Burgette and Hahn 2010). Instead, the state indicator is not observed and needs to be inferred from the data. This needs evaluation of $TG(G-1)$ multivariate integrals (see also Nobile 1998)

$$\eta_{t,g'g} = \int_{\mathcal{G}_{g'g}} \phi(\nu_t, \Sigma) d\nu_t, \; g \in \mathcal{G}_{-g_0}$$

where $\phi(\cdot)$ is the density of the $(G-1)$-variate normal with mean 0 and covariance Σ, $\nu_t = (\nu_{1t}, \ldots, \nu_{g_0-1,t}, \nu_{g_0+1,t}, \ldots, \nu_{Gt})'$. The set $\mathcal{G}_{g'g}$ is given by

$$\mathcal{G}_{g'g} = \bigcap_{j \neq g} \left\{ \nu_{gt} - \nu_{jt} > x_{2t}(\gamma^x_{g'j} - \gamma^x_{g'g}) + (\gamma_{g'j} - \gamma_{g'g}) \right\} \cap \left\{ \nu_{gt} > -(x_{2t}\gamma^x_{g'g} + \gamma_{g'g}) \right\}$$

Various procedures have been proposed to evaluate these integrals (Geweke et al. 1994). They all represent approximations to the transition probabilities, however. Moreover, estimation gets very slow, see section 3.4.

2.2.3 Nested alternatives

Some alternatives are nested in both the logit and the probit functional form. If $\gamma^x_{g'g} = 0$, we recover the specification with time-invariant transition probabilities, e.g. $\eta_{g'g} = \exp(\gamma_{g'g}) / \left(\sum_{j=1}^G \exp(\gamma_{g'j})\right)$.

State persistence is maintained, even if we restrict the covariate effect to be state-independent $\gamma^x_{g'g} = \gamma^x_g \; \forall g'$. If we additionally restrict $\gamma_{g'g} = \gamma_g$, we obtain a mixture model with time-varying weights:

$$f(y_t|x_{1t}, x_{2t}, \theta) = \sum_{g=1}^G \eta_{tg}(x_{2t}, \gamma_g) f(y_t|x_{1t}, \theta_g)$$

The relevance of differences across state-dependent parameters or the relevance of covariates can be evaluated using the Savage-Dickey density ratio:

$$\log BF(M_0|M) = \log \frac{\pi(\gamma|y, x_1, x_2)\,|_{\gamma \in \mathcal{R}}}{\pi(\gamma)\,|_{\gamma \in \mathcal{R}}}$$

where M_0 and M indicate, respectively, the restricted and the unrestricted model and $\mathcal{R}$ represents a single or a combination of restrictions on γ mentioned above.

2.3 Generalizations

The framework (1) encompasses linear regressions and dynamic models as well, given that x_{1t} may also include lagged observations of y_t. The Markov process in (2) is of order one. This is not restrictive, as pth-order Markov processes can be reparameterized by defining an encompassing G^p first-order Markov state process with appropriate design of the transition distribution. Likewise, current and (fixed) p lagged state dependence in $f(y_t|\cdot)$ may be reparameterized to current state dependence by defining an encompassing G^{1+p} state variable with appropriately designed transition distribution and enlarging the state-specific parameter to $\theta_t = \{\theta_{z_{t-j}}|j = 0, \ldots, p, z_t \in \mathcal{G}\,\forall t\}$, see also Hamilton (1994, chapter 22).

Model (1) is generic in terms of the state-dependence of y_t and θ_t, as well as in terms of parametrization of the state indicator. Some elements of θ_t might be state-independent, $\theta_t = \{\{\theta_g, \theta_0\}|g = 1, \ldots, G\}$, e.g. Frühwirth-Schnatter and Kaufmann (2006). The model encompasses situations where some elements of the appropriately partitioned vector $y_t = (y'_{1t}, y'_{2t})'$ follow a state-independent distribution, i.e. $f(y_t|x_t, \theta_{z_t}) = f(y_{2t}|y_{1t}, x_t, \theta_{2,z_t})f(y_{1t}|x_t, \theta_1)$.

Multiple states may affect y_t. The simplest situation is the case where independent state processes determine the elements in y_t, $y_t \sim f(y_{2t}|y_{1t}, x_t, \theta_{z_t^2})f(y_{1t}|x_t, \theta_{z_t^1})$, Psaradakis, Ravn, and Sola (2005). If observations are independent and driven by independent processes, the models for y_{1t} and y_{2t} might even be analyzed separately, $y_t \sim f(y_{1t}|x_t, \theta_{1,z_t^1})f(y_{2t}|x_t, \theta_{2,z_t^2})$. Finally, state indicators z_t^j, $j = 1, \ldots, p$, may be linked by a dynamic or hierarchical structure, (Kaufmann 2010; Geweke and Amisano 2011; Bai and Wang 2011). These models are analyzed by defining an encompassing state indicator z_t^*, which captures all of the possible state combinations of the underlying p state indicators, see subsection 4 for some examples and references to applications.

2.4 Some considerations on parametrization

Nothing has been said about scaling of covariates x_{2t}. In fact, in specification (4), the covariate is assumed to be mean-adjusted or normalized, $x_{2t} = \tilde{x}_{2t} - \bar{x}_2$, where $\tilde{x}_{2t}$ and $\bar{x}_2$ are, respectively, the level series and the mean or the normalizing level. We call this the *centered* parametrization, in which the time-invariant part of the transition probabilities, $\gamma_{g'g}$, gets scale-independent. In estimation, this scales the range of sensible values for $\gamma_{g'g}$ and in Bayesian estimation this allows to design a scale-invariant prior.

To illustrate this, assume $G = 2$, state-independent state probability $\eta = \exp(x_{2t}\gamma^x + \gamma) / (1 + \exp(x_{2t}\gamma^x + \gamma))$, with $\gamma^x = 0.5$. The range of γ against η to obtain $\eta = 0.5$ when $\tilde{x}_{2t} = \bar{x}_2$ is not scale-invariant with respect to $\tilde{x}_{2t}$, see the dashed and solid lines in figure 1. Working with the centered version, removes scale-dependence of γ, see the dash-dotted line in figure 1.

[Insert figure 1 around here]

Moreover, it is worthwhile to form expectations about sensible parameter configurations prior to estimation.[1] Assume state-dependent covariate effects, $g_0 = 1$, such that $\gamma_{g'2} = \gamma = (\gamma_1^x, \gamma_2^x, \gamma_1, \gamma_2)$. Figure 2 plots the state persistence against the covariate for two settings of γ. In both settings, state persistence is $\eta_{gg} = 0.88$ if $x_{2t} = 0$. If (γ_1^x, γ_2^x)

[1] The example follows the one in Kaufmann (2015).

deviate with equal sign from zero, in the limit, as $x_{2t} \to \pm\infty$, one of the states becomes absorbing. On the other hand, if the parameters differ in sign both states become absorbing if $x_{2t} \to -\infty$ or the indicator switches back and forth between states if $x_{2t} \to \infty$. The conclusion we draw from this very simple illustration is that if unconstrained covariates have state-dependent effects in the transition distribution, the parameter configuration for γ^x should be such that the probability mass is shifted mainly towards one of the states as x_{2t} varies in size.

[Insert figure 2 around here]

2.5 Mean-square stability: Combining stable and unstable processes

To motivate the discussion, assume a two-state Markov switching univariate autoregressive process of order one:

$$y_t = \rho_{z_t} y_{t-1} + \varepsilon_t,\ \ \varepsilon_t \sim N(0, \sigma^2) \tag{7}$$

with a given transition matrix η. Usually, it is assumed that $|\rho_{z_t}| < 1$, for $z_t = 1, 2$. In this case, the unconditional state-specific distribution $f(y_t|z_t)$ has bounded moments $\forall t$, and $\lim_{t\to\infty} f(y_t)$ as well. The latter condition is weaker than the former and is defined as mean-square stability in the engineering literature (Costa et al. 2004). For multivariate processes, the situation becomes more intricate. Francq and Zakoïan (2001) derive stationarity conditions for Markov switching multivariate autoregressive moving-average models and show that stationary state-specific processes need not be sufficient for mean-square stability.

Mean-square stability requires that the first and second moments of the process y_t converge to a well defined limit as the time horizon extends to infinity (see Definition 1 in Farmer et al. 2009b):

$$\lim_{t\to\infty} E(y_t) = \mu, \quad \lim_{t\to\infty} E(y_t y_t') = \Sigma$$

Mean-square stability is weaker than bounded stability which additionally requires bounded moments $\forall t$. In linear systems, mean-square stability is equivalent to bounded stability (for bounded shocks). However, in Markov switching models the concepts are not the same.

In the above example, a univariate stationary state process $|\rho_1| < 1$ may be combined with a non-stationary one, $|\rho_2| > 1$, if the second state does not recur too often and does not persist for too long. Bounded stability is not given, because the second moment of $f(y_t|z_t = 2)$ is unbounded, while mean-square stability may hold.

One might wonder why we should care about mean-square stability. Usually, Bayesian estimation involves the conditional distribution $f(y_t|z_t, y_{t-1})$, which has finite moments even for unstable processes. However, taking a forecasting perspective, it might be of interest to ensure that the estimated model implies a forecast density with bounded moments in the long run, although unstable periods would produce forecast densities with unbounded moments at some forecast horizon. In the macroeconomic literature, the concept is of interest in solving for the long-run equilibrium in Markov switching dynamic stochastic general equilibrium models, see Farmer et al. (2009a, 2009b, 2011).

Francq and Zakoïan (2001) and Costa et al. (2004) show that mean-square stability is given for a Markov switching process if all roots of the matrix (we drop the subscript t on η)

$$\begin{bmatrix} \eta_{11}\rho_1 \otimes \rho_1 & \dots & \eta_{G1}\rho_G \otimes \rho_G \\ \vdots & & \vdots \\ \eta_{1G}\rho_1 \otimes \rho_1 & \dots & \eta_{GG}\rho_G \otimes \rho_G \end{bmatrix} \tag{8}$$

lie inside the unit circle.

Clearly, mean-square stability depends non-linearly on the state-specific processes ρ_g as well as on the transition distribution η. In particular, state-specific bounded processes (i.e. when the roots of ρ_g, $\forall g \in \mathcal{G}$ lie inside the unit circle) do not always ensure a mean-square stable process for y_t.

The example in Farmer et al. (2009b, p. 1854-55) (another is given in Francq and Zakoïan 2001) illustrates this surprising result and we reproduce it here. Assume a bivariate two state Markov switching autoregressive process y_t with state-dependent autoregressive matrices

$$\rho_1 = \begin{bmatrix} 0 & 2 \\ 0 & 0.5 \end{bmatrix} \qquad \rho_2 = \begin{bmatrix} 0.5 & 0 \\ 2 & 0 \end{bmatrix}$$

Both state-specific processes are stable (and covariance stationary) and yield each unconditional distributions with bounded moments. Assume that the transition between the two processes is characterized by either of the two transition matrices

$$\eta = \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix}, \qquad \tilde{\eta} = \begin{bmatrix} 0.9 & 0.1 \\ 0.4 & 0.6 \end{bmatrix}$$

The intriguing thing is that in combination with transition probability matrix η, the roots of the matrix (8) lie outside the unit circle while with transition matrix $\tilde{\eta}$ the roots lie inside the unit circle. The first process is not mean-square stable, although both state-specific processes are stable. Decreasing the transition frequency between states restores mean-square stability.

There are many situations where one or more regimes may be unstable. In economics for example, periods of hyperinflation are clearly unstable. It is conceivable that the economic variables engaged on an unstable path during the period in which the recent financial crisis unraveled. The concept of mean-square stability allows us to combine unstable and stable processes, if the unstable process, relatively to the stable process, does not persist for too long and/or does not recur too often. Figure 3 plots the boundary values for (ρ_1, ρ_2) conditional on various combinations for (η_{11}, η_{22}) for which the univariate process in (7) is still mean-square stable. In general, the boundary value for ρ_g decreases the more persistent state g is. We observe that many combinations allow for an explosive root in either state-specific process, even if both states are highly persistent, see e.g. the most inner boundary circle conditional on $(\eta_{11}, \eta_{22}) = (0.9, 0.7)$ in the left panel of the figure. For illustration, figure 4 plots a generated series with $(\rho_1, \rho_2) = (1.1, 0.7)$ conditional on $(\eta_{11}, \eta_{22}) = (0.5, 0.7)$.

[Insert figure 3 around here]

[Insert figure 4 around here]

Although this specific feature is appealing for time series and macroeconomic modelling, it has so far been applied very rarely. The work of Davig and Leeper (2007) and of Farmer and co-authors derived conditions for unique equilibrium determination in DSGE models. Foerster et al. (2016) present advances to obtain solutions in Markov switching DSGE models. In time series modelling, most applications estimate Markov switching models with stable state-specific processes. Allowing Markov switching between state-specific stable and unstable, explosive processes represents an interesting avenue for future research.

3 Estimation

As in general for mixture modeling, the latent states are a priori not identified in the mixture and the likelihood will be invariant to permutations of states, $\rho = (\rho_1, \ldots, \rho_G)$, $L\,(y|x_1, z, \theta) = L\,(y|x_1, \rho(z, \theta))$. Thus, state-identification is obtained by imposing ordering restrictions on state-specific parameters θ or η. Obviously, state-identifying restrictions have to be imposed on parameters which indeed differ between states. Knowledge about state-specific parameters can also be imposed to estimate a state-identified model. If there is uncertainty about which parameters differ between states, we may first estimate the unidentified model and apply an appropriate state-identifying restriction ex-post to obtain a state-identified model.

In Bayesian estimation, in the first case we also impose the state-identifying restriction on the prior and apply restricted sampling. In the second case, we design a state-invariant prior to explore the state-invariant posterior $\pi\,(z, \theta, \eta|y, x) = \pi\,(\rho(z, \theta, \eta)|y, x)$. Posterior inference is obtained by random permutation sampling (Frühwirth-Schnatter 2001). State identification is then obtained by post-processing the posterior output (Frühwirth-Schnatter 2011).

We discuss estimation conditional on G. Testing or model evaluation with respect to G is not discussed, given that methods described in chapter ?? of this handbook can be applied with appropriate adjustments.

3.1 The complete data likelihood and the FFBS algorithm

Estimation of hidden Markov models has to take into account the sequential dependence in z_t. The complete data likelihood factorizes into the conditional likelihood:

$$L\,(y|x_1, z, \theta) = \prod_{t=1}^{T} f(y_t|x_{1t}, \theta_{z_t}) \tag{9}$$

where for both frequentist and Bayesian estimation we need an inference on the latent indicator z_t. Given the time dependence in z_t, the 'E' step in the EM algorithm and the data augmentation step in Bayesian inference are slightly more involved than described in chapter I.2 in this volume. Both, however, start out with the factorization of the state distribution conditional on all data

$$\pi\left(z|y,\theta,\eta\right)=\pi\left(z_T|\mathcal{I}_T\right)\prod_{t=1}^{T-1}\pi\left(z_t|\mathcal{I}_T,z_{t+1}\right)\pi\left(z_0\right)$$

where the dependence on (θ,η) is suppressed on the right-hand side for notational convenience, $\mathcal{I}_t$ denotes information up to time t and $\pi\left(z_0\right)$ denotes the starting state probability distribution (Chib 1996). The factorization includes the typical element

$$\begin{aligned}\pi\left(z_t|\mathcal{I}_T,z_{t+1}\right) &\propto \pi\left(z_t|\mathcal{I}_t\right)\pi\left(z_{t+1}|z_t,\mathcal{I}_T\right) \\ &\propto \pi\left(z_t|\mathcal{I}_t\right)\eta_{t+1,z_t z_{t+1}}\end{aligned} \tag{10}$$

The term $\pi\left(z_t|\mathcal{I}_t\right) \propto f(y_t|x_t,z_t,\theta)\pi\left(z_t|\mathcal{I}_{t-1}\right)$ consists of the likelihood $f(y_t|\cdot)$ and $\pi\left(z_t|\mathcal{I}_{t-1}\right)=\eta_t\pi\left(z_{t-1}|\mathcal{I}_{t-1}\right)$ is obtained by extrapolation.

Inference on $\pi\left(z|y,x,\theta,\eta\right)$ is obtained by a forward-filtering backward-smoothing algorithm (FFBS), which is based on the Kalman filter.

1. Run forward in time $t=1,\ldots,T$ to obtain the filter densities $\pi\left(z_t|\mathcal{I}_t\right)$ or the filter probabilities $P\left(z_t=g|\mathcal{I}_t\right)$:

$$P\left(z_t=g|\mathcal{I}_t\right) = \frac{f(y_t|x_t,\theta_g)P\left(z_t=g|\mathcal{I}_{t-1}\right)}{\sum_{g'=1}^{G} f(y_t|x_t,\theta_{g'})P\left(z_t=g'|\mathcal{I}_{t-1}\right)} \tag{11}$$

$$P\left(z_t=g|\mathcal{I}_{t-1}\right) = \sum_{g'=1}^{G}\eta_{t,g'g}P\left(z_{t-1}=g'|\mathcal{I}_{t-1}\right) \tag{12}$$

   At T we obtain $\pi\left(z_T|\mathcal{I}_T\right)$.

2. Run backward in time $t=T-1,\ldots,1$ to obtain the smoothed densities $\pi\left(z_t|\mathcal{I}_T\right)$ or smoothed probabilities $P\left(z_t=g|\mathcal{I}_T\right)$

$$P\left(z_t=g|\mathcal{I}_T\right)=\sum_{g'=1}^{G}P\left(z_t=g|\mathcal{I}_t,z_{t+1}=g'\right)P\left(z_{t+1}=g'|\mathcal{I}_T\right) \tag{13}$$

   where

$$P\left(z_t=g|\mathcal{I}_t,z_{t+1}=g'\right)=\frac{P\left(z_t=g|\mathcal{I}_t\right)\eta_{t+1,gg'}}{\sum_{g=1}^{G}P\left(z_t=g|\mathcal{I}_t\right)\eta_{t+1,gg'}} \tag{14}$$

3.2 Maximum likelihood estimation

Conditional on the smoothed probabilities $P(z_t = g|\mathcal{I}_T)$, $t = 1, \ldots, T$, the complete data likelihood

$$L(y|x_1, z, \theta) = \prod_{t=1}^{T} \sum_{g=1}^{G} P(z_t = g|\mathcal{I}_T) f(y_t|x_{1t}, \theta_g) \tag{15}$$

is maximized with respect to the model parameters (the 'M' step). With respect to θ, we solve

$$\sum_{t=1}^{T} \sum_{g=1}^{G} \frac{\partial \ \log f(y_t|x_{1t}, \theta_g)}{\partial \theta_g'} P(z_t = g|\mathcal{I}_T) = 0$$

For example, in a multiple regression setup with normally distributed error terms, $\theta_g = \{\beta_g, \sigma_g^2\}$, the observation density is

$$f(y_t|x_{1t}, \theta_g) = \frac{1}{\sqrt{2\pi}\sigma_g} \exp\left\{-\frac{1}{2\sigma_g^2}\left(y_t - x_{1t}'\beta_g\right)^2\right\}$$

and the maximum likelihood estimate of β_g corresponds to a weighted least squares estimate

$$\hat{\beta}_g = \left(\sum_{t=1}^{T} \tilde{x}_t \tilde{x}_t'\right)^{-1} \left(\sum_{t=1}^{T} \tilde{x}_t \tilde{y}_t\right)$$

where $\tilde{x}_t = x_t\sqrt{P(z_t = g|\mathcal{I}_T)}$ and $\tilde{y}_t = y_t\sqrt{P(z_t = g|\mathcal{I}_T)}$ are the observations weighted by the square root of the smoothed probabilities. The estimate of the state-dependent variances equals

$$\hat{\sigma}_g^2 = \frac{\sum_{t=1}^{T}\left(\tilde{y}_t - \tilde{x}_t'\hat{\beta}_g\right)^2}{\sum_{t=1}^{T} P(z_t = g|\mathcal{I}_T)}$$

The estimate of time-invariant Markov transition probabilities (3) is given by

$$\hat{\eta}_{g'g} = \sum_{t=2}^{T} \frac{P(z_t = g, z_{t-1} = g'|\mathcal{I}_T)}{P(z_{t-1} = g'|\mathcal{I}_T)}$$

where the numerator equals the terms in (13) for given (g, g') (Hamilton 1994, chapter 22). The first-order conditions with respect to γ in case of time-varying transition probabilities are non-linear, see e.g. Diebold et al. (1994). To estimate multi-state time-varying transition probabilities one may borrow from recent advances in modelling latent class multinomial logit models, Greene and Hensher (2003, 2013), Hess (2014).

Conditional on estimates $(\hat{\theta}, \hat{\eta})$, the maximal value of the likelihood function is

$$L(y|x_1, z, \hat{\theta}) = \prod_{t=1}^{T} \sum_{g=1}^{G} P(z_t = g|\mathcal{I}_T) f(y_t|x_{1t}, \hat{\theta}_g) \tag{16}$$

3.3 Bayesian estimation

To make inference on the joint posterior

$$\pi\left(z, \theta, \eta | y, x\right) \propto L(y|x_1, z, \theta)\pi(z|\eta)\pi(\eta|x_2)\pi(\theta) \tag{17}$$

we use data augmentation based on the FFBS procedure derived in section 3.1. The first term in (17) represents the conditional data likelihood (9). The prior $\pi(z|\eta) = \prod_{t=1}^{T} \pi(z_t|z_{t-1}, \eta_t)\pi(z_0)$ takes into account time dependence in z_t.

To sample from the state-invariant posterior, we have to design state-invariant prior distributions $\pi(\eta|x_2)$ and $\pi(\theta)$. A state-invariant prior $\pi(\theta)$ is often quite straightforward to design, while the design of a state-invariant prior $\pi(\eta|x_2)$ is slightly more intricate depending on the level of complexity in the parametrization (see below).

If prior information is known about state-identifying restrictions, it may be sensible to integrate these restrictions into the prior $\theta \sim \pi\left(\theta\right) 1_{\mathcal{R}}$ or into $\eta \sim \pi\left(\eta\right) 1_{\mathcal{R}}$, where $1_{\mathcal{R}}$ is one if the set of restrictions $\mathcal{R}$ is fulfilled. This, however, destroys state-invariance of the prior distribution and estimation proceeds with restricted sampling.

3.3.1 Prior specification $\pi(\eta|x_2)$

For time-invariant switching (3) a conjugate prior is the Dirichlet distribution $(\eta_{g'1}, \ldots, \eta_{g'G}) \sim D\left(e_{0,g'1}, \ldots, e_{0,g'G}\right)$, $g' = 1, \ldots, G$. To obtain a state-invariant prior

$$\pi\left(\eta\right) \sim \prod_{g'=1}^{G} D\left(e_{0,g'1}, \ldots, e_{0,g'G}\right)$$

we set $e_{0,gg} = \kappa_0 \ \forall g$, and $e_{0,g'g} = \kappa_1 \ \forall g' \neq g$, where $\kappa_0, \kappa_1 > 0$. An informative prior usually puts more weight on the persistence probabilities, i.e. $\kappa_0 > \kappa_1$. If $G = 2$, the prior distribution is beta distribution $B(e_{0,gg}, e_{0,g'g})$, $g = 1, 2$, $g' \neq g$. Sims et al. (2008) derive a general framework to model Markov switching transition probabilities in large multiple-equation systems, including a framework to design a prior Dirichlet distribution which induces restrictions on the transition probability matrix.

For the logit and the probit functional form (see section 2.2.2) the prior for $\gamma = \{\gamma_g | g \in \mathcal{G}_{-g_0}\ \gamma_g = \left(\gamma^x_{1g}, \ldots, \gamma^x_{Gg}, \gamma_{1g}, \ldots, \gamma_{Gg}\right)\}$ is assumed normal

$$\pi\left(\gamma\right) = \prod_{g \in \mathcal{G}_{-g_0}} \pi\left(\gamma_g\right) = \prod_{g \in \mathcal{G}_{-g_0}} N\left(e_{0,g}, E_0\right)$$

where the hyperparameters in $e_{0,g}$ relate respectively to γ_g.

Generally, a state invariant prior for the logistic functional form is designed in the following way (Kaufmann 2015). The hyperparameters relating to state persistence are set to $\{e^x_{0,gg}, e_{0,gg}\} = \{\kappa^x, \kappa\}$. Then, the hyperparameters referring to transition parameters from the reference state to state g are set to $\{e^x_{0,g_0g}, e_{0,g_0g}\} = \{-\kappa^x, -\kappa\}$ and those referring to transition parameters from other states to state g to zero, $\{e^x_{0,jg}, e_{0,jg}\} = \{0, 0\}$, $j \neq$

g, g_0. Thus, the dependence on g in $N(e_{0,g}, E_0)$ is not suppressed because the ordering of hyperparameters in $e_{0,g}$ varies across g. When random permutation sampling (see below) includes the reference state g_0, hyperparameters $e_{0,g}$ have to be permuted accordingly, see the appendix in Kaufmann (2015). Conditional on g_0, i.e. keeping g_0 fixed for estimation, hyperparameters $e_{0,g}$ do not have to be permuted, however.

Under the consideration that relevant parameters in γ_g^x should be shifted away from zero in the same direction (see 2.4), in fact the only sensible parametrization for a state-invariant prior is $\kappa^x = 0$.

The prior specification proposed in Burgette and Hahn (2010) proves especially useful to apply the permutation sampler when we use a probit functional form for η. They define a prior which is state-invariant in the sense that normalization is independent of a reference state. The parameter γ_{g_0} equals the negative of the sum over the parameters governing transitions to the other states, $\gamma_{g_0} = -\sum_{g \neq g_0} \gamma_g$. Additionally, instead of normalizing element (g_0, g_0) of the covariance in latent utilities, the trace $tr(\Sigma^*)$ of the normalized error covariance in latent utilities

$$\begin{aligned} z_{gt}^* &= x_{2t}\gamma_{z_{t-1}g}^x + \gamma_{z_{t-1}g} + \nu_{gt}, \ g \in \mathcal{G}_{-g_0} \\ & \nu_t = (\nu_{1t}, \ldots, \nu_{g_0-1,t}, \nu_{g_0+1,t}, \ldots, \nu_{Gt})' \sim N(0, \Sigma^*) \end{aligned} \tag{18}$$

is normalized to $tr(\Sigma^*) = G - 1$. To obtain a state-invariant prior, we first set all hyperparameters relating to state persistence to $\{e_{0,gg}^x, e_{0,gg}\} = \{\kappa^x, \kappa\}$, $\forall g \neq g_0$, and then all other elements to $\{e_{0,g'g}^x, e_{0,g'g}\} = \{-\kappa^x/(G-1), -\kappa/(G-1)\}$, $g' \neq g$, $\forall g', g \neq g_0$. Again, the only sensible parametrization for a state-invariant prior is $\kappa^x = 0$.

3.3.2 Posterior inference

To obtain a sample from the posterior (17), we repeatedly draw from

1. $\pi(z|y, x, \theta, \eta)$: A draw from the latent indicator z is obtained by applying the FFBS algorithm (see section 3.1), where the 'BS' step becomes a backward-sampling step. Given the filter densities $\pi(z_t = g|\mathcal{I}_t)$, we first sample z_T from $\pi(z_T|\mathcal{I}_T)$. For $t = T-1, \ldots, 1$, we sample from

$$\pi(z_t = g|\mathcal{I}_T) \propto \pi(z_t = g|\mathcal{I}_t)\, \eta_{t+1, g z_{t+1}}$$

2. $\pi(\eta|z)$ or from $\pi(\eta|z)\, 1_{\mathcal{R}}$ in case state-identifying restrictions are imposed on the prior: Details are given in the next sub-section.

3. $\pi(\theta|y, x, z)$ or from $\pi(\theta|y, x, z)\, 1_{\mathcal{R}}$.

If the prior is state-invariant we terminate the sampler by

4(i). randomly permuting the states and state-specific parameters (random permutation sampler): We obtain a sample from the unconstrained multimodal posterior (17)

4(ii). re-ordering the states and state-specific parameters according to a pre-defined state-identifying restriction (restricted sampler): We obtain a sample from the constrained posterior.

Sampling from restricted posterior distributions $\pi(\eta|z) 1_{\mathcal{R}}$ and $\pi(\theta|y,x,z) 1_{\mathcal{R}}$, calls either for restricted or some rejection sampling procedures, see section 4 for some examples.

3.3.3 Posterior sampling of transition parameters

In sampling step 2, under time-invariant switching and a Dirichlet prior, the posterior is also Dirichlet

$$\pi(\eta|z) \sim \prod_{g'=1}^{G} D(e_{g'1}, \ldots, e_{g'G})$$

with $e_{g'g} = e_{0,g'g} + \#\{z_t = g, z_{t-1} = g'\}$, where the prior hyperparameter is updated by the number of times state g is preceded by state g'.

Sampling from $\pi(\gamma|x_{2t}, z)$ under the logit functional form is based on introducing latent state-specific random utilities z^u_{gt} for all but the reference state g_0 (McFadden 1974)

$$\begin{aligned} z^u_{gt} &= X'_t\gamma_g + \nu_{gt}, \ \forall g \in \mathcal{G}_{-g_0} \\ & \nu_{gt} \text{ i.i.d. Type I extreme value} \end{aligned} \tag{19}$$

where $X'_t = (x_{2t}z_{g,t-1}, x_{2t}z_{g,t-1}, \ldots, x_{2t}z_{g,t-1}, z_{g,t-1}, z_{g,t-1}, \ldots, z_{g,t-1})$. Given that maximum utility is obtained for the observed state, we use a partial representation of the model, in which the latent utilities are expressed in difference to the maximum utility of all other states:

$$\begin{aligned} z^*_{gt} &:= z^u_{gt} - z^u_{-g,t} = c + X'_t\gamma_g + \epsilon_{gt}, \ \forall g \in \mathcal{G}_{-g_0} \\ & \epsilon_{gt} \text{ i.i.d. Logistic} \end{aligned} \tag{20}$$

where $z^u_{-g,t} = \max_{j \in \mathcal{G}_{-g}} z^u_{jt}$ and c is a constant.

The latent utilities model (20) is linear in γ, with non-normal errors, however. Frühwirth-Schnatter and Frühwirth (2007) suggest to introduce a second layer of data augmentation to approximate the Logistic distribution of ϵ_{gt} by a mixture of normals with M components. Conditional on $\boldsymbol{z}^*_g = (z^*_{g1}, \ldots, z^*_{gT})$ and the mixture components $\boldsymbol{R}_g = (R_{g1}, \ldots, R_{gT})$, $R_{gt} \in \{1, \ldots, M\}$, we obtain a normal posterior for γ

$$\gamma_g|z, x_2, \boldsymbol{z}^*_g, \boldsymbol{R}_g \sim N\left(e_g\left(\boldsymbol{z}^*_g, \boldsymbol{R}_g\right), E_g(\boldsymbol{R}_g)\right)$$

with moments $e_g(\cdot)$ and $E_g(\cdot)$ explicitly derived in the appendix of Kaufmann (2015).

The first layer of data augmentation (20) renders the model linear in γ and posterior sampling of γ can also be based on a Metropolis-Hastings step, see Scott (2011), Holmes and Held (2006), which is also used in economic applications, see Hamilton and Owyang (2012), Owyang et al. (2015). However, Frühwirth-Schnatter and Frühwirth (2007) show that sampling efficiency is considerably improved by introducing the second layer of data augmentation. In addition, Kaufmann (2015) shows that higher sampling efficiency is achieved when the partial representation (20) rather than specification (19) is used.

When the multinomial probit functional form is used, data augmentation and parameter estimation can generally proceed as in Albert and Chib (1993), McCulloch, Polson, and Rossi (2000), Nobile (1998). The (normalized) model (18) for the latent random utilities is linear Gaussian and the posterior $\pi(\gamma|z, x_2, z^*, \Sigma^*)$ is normal. Burgette and Hahn (2010) propose a sampler that is particularly useful to apply random permutation sampling. Given their detailed description, we do not reproduce the sampler in detail here.

3.4 Sampler efficiency: Logit versus Probit

In this section we simulate series y_t of length $T = 500$ with autoregressive processes subject to $G = \{2, 3\}$ regimes with time varying transition probabilities.

$$y_t = \rho_{z_t} y_{t-1} + \varepsilon_t, \ \varepsilon_t \sim N(0, \sigma^2_{z_t}) \tag{21}$$

with parameter settings

$$(\rho_{z_t}, \sigma^2_{z_t}) = \begin{cases} (0.2, 0.1) & \text{if } z_t = 1 \\ (0.8, 0.01) & \text{if } z_t = 2 \\ (0, 1) & \text{if } z_t = 3 \end{cases}$$

We simulate z_t using the logit functional form for the transition probabilities and use the following parametrization for γ:

	$G = 2$	$G = 3$	
	$g = 2$	$g = 2$	$g = 3$
γ^x_{1g}	4	4	0
γ^x_{2g}	1	3	2
γ^x_{3g}		0	4
γ_{1g}	−2	−1	−1
γ_{2g}	2	1	0
γ_{3g}		0	2

For the covariate x_{2t}, we assume a persistent process, $x_{2t} = 0.8x_{2,t-1} + \epsilon_t$, $\epsilon_t \sim N(0, 1)$.

We estimate the model using either the logit or the probit functional form for the transition probabilities and evaluate estimation time and sampler efficiency. For estimation, we include a constant θ_{0g} and use independent parameter prior specifications, $\theta_{0g} \sim N(0, 0.0225)$, $\rho_g \sim N(0, 0.21)$, $\sigma^2_g \sim IG(5, 0.25)$. For γ, we use $E_0 = I_{2G}$, with I_{2G} the $2G \times 2G$ identity matrix, $(\kappa^x, \kappa) = (0, 2)$ and $(\kappa^x, \kappa) = (0, 0)$ when working with, respectively, the logit and the probit functional form. When we use the probit functional form, the trace-restricted Wishart prior for Σ^* is parameterized according to Burgette and Hahn (2010).

[Insert table 1 around here]

We draw $M = 30,000$ times out of the posterior and retain the last 15,000 draws for posterior inference. In table 1 we report the time to obtain 1,000 draws from the random permutation sampler and the inefficiency factor for the state-identified posterior draws of $\gamma^{(m)}$, $m = 1, \dots, M$. Obviously, there is a trade-off. For $G = 2$, estimation time and efficiency are comparable across both functional forms. By thinning out the posterior sample, we reduce inefficiency roughly by a factor of four. When $G = 3$, while estimation time remains nearly unchanged when working with the logit functional form, there is a tenfold increase in estimation time when working with the probit functional form. Obviously, this is due to the fact that in each iteration, $TG(G-1)$ multivariate integrals have to be evaluated to compute the transition probabilities η_t. On the other hand, the probit sampler of Burgette and Hahn (2010) performs much better in terms of efficiency. This is also reflected in figure 5, in which we depict the autocorrelation function of the state-identified posterior draws for $\gamma_g^{(m)}$, $g = \{2, 3\}$ and $m = 1, \dots, M$. While autocorrelation functions drop at about the same rate when $G = 2$, the autocorrelations drop very quickly to zero when working with the probit functional form when $G = 3$. Given these results, the researcher may use the logit functional form to save on estimation time, and thin out considerably the posterior sample to adjust for the relative inefficiency.

[Insert figure 5 around here]

3.5 Posterior state identification

We illustrate posterior state identification based on the posterior output of the model estimated with the logit functional form (the one obtained using the probit functional form is identical). Figure 6 reproduces scatter plots of unsorted posterior draws on the first line. In panel (a) $\rho_g^{(m)}$ is plotted against $\rho_{g'}^{(m)}$, $g \neq g'$. In panel (b), $\rho_g^{(m)}$ is plotted against $\log \sigma_g^{2(m)}$. Panel (a) reflects the $K!$ modes of the unidentified posterior output of the random permutation sampler. On the other hand, panel (b) reveals that states may be identified by imposing the restriction $(\rho_g, \sigma_g^2)|g = 1, 2, 3, 1_{(\rho_1 < \rho_2 < \rho_3)} 1_{(\sigma_1^2 > \sigma_2^2 > \sigma_3^2)}$. All those draws which can be permuted uniquely to fulfill the state-identifying restriction are retained to do posterior inference. For illustration, panels (c) and (d) in figure 6 reproduce the state-identified marginal posterior distributions of θ_{0g}, ρ_g and $\gamma_{g'g}^x$, respectively.

[Insert figure 6 around here]

4 Informative regime switching in applications

In this section, we discuss various possibilities to introduce information in the transition distribution. The usual critique to Markov switching models with time-invariant transition distributions is that the switches of the state indicator remain unexplained. In economic analysis, the usual approach to give an interpretation to the state indicator is to relate ex-post estimated state-specific periods to some statistical measures of investigated series, like state-specific means or volatilities. Another possibility is to relate the estimated state indicator to some officially released indicator, like the business cycle turning point dates released by the National Bureau of Economic Research (NBER).

To address the critique, one can directly include information into the transition distribution. Being more specific on the design of the transition distribution, one can obtain informative switching for time-invariant specification. Including explicitly covariates which affect the transition probabilities renders the transition distribution informative and time-varying. The last possibility to include information is to impose restrictions on the parameters of the transition distribution.

4.1 Time-invariant switching

Consider the time-invariant specification of the transition probabilities $P(z_t = g|z_{t-1} = g') = \eta_{g'g}$, and collect the probabilities in the matrix

$$\eta = \begin{bmatrix} \eta_{11} & \cdots & \eta_{1G} \\ \vdots & & \\ \eta_{G1} & \cdots & \eta_{GG} \end{bmatrix} \tag{22}$$

Information can be included by explicitly designing or imposing restrictions on the transition matrix. In this sense, the time-invariant transition distribution becomes informative. In the following we discuss various examples. In Sims et al. (2008), the interested reader finds an encompassing framework to impose and analyze restrictions on Markov transition probabilities in large multiple-equation models.

4.1.1 Unconditional switching

Imposing the restriction $\eta_{g'g} = \eta_g$ in (22) renders state switching unconditional. The Markov mixture model simplifies to a simple mixture model

$$f(y|\theta, \eta) = \prod_{t=1}^{T} \sum_{g=1}^{G} \eta_g f(y_t|\theta_g), \;\; P(z_t = g|\eta) = \eta_g$$

4.1.2 Structured Markov switching

The general setup (22) introduces state persistence with expected state persistence of $(1-p_{gg})^{-1}$. It does not ensure minimum time duration, which is defined for some economic features like e.g. a recession. A recession is usually defined as two consecutive quarters (half a year) of negative gross domestic product (GDP) growth. When working with quarterly GDP data, we might include this minimum cycle duration by designing an encompassing state indicator z_t^* with transition matrix:

$$\eta^* = \begin{bmatrix} \eta_{11} & \eta_{12} & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & \eta_{21} & \eta_{22} \end{bmatrix} \quad \begin{array}{l} z_t^* = 1, z_t = 1 \text{ recession} \\ z_t^* = 2, z_t = 1 \text{ trough} \\ z_t^* = 3, z_t = 2 \text{ peak} \\ z_t^* = 4, z_t = 2 \text{ expansion} \end{array}$$

which imposes a minimum cycle duration from peak-to-peak or trough-to-through of five quarters (Artis et al. 2004). Generalizing to longer state durations or to state-specific cycle durations is straightforward.

In multivariate analysis, sub-vectors of data in $y_t = (y'_{1t}, y'_{2t})'$ may be affected by different state indicators. The simplest setup is the case where both indicators follow independent transition distributions, $f(y_t|x_t, z_t, \theta, \eta) = f(y_{2t}|y_{1t}, x_t, z_t^2, \theta, \eta_2) f(y_{1t}|x_t, z_t^1, \theta, \eta_1)$. Taking business cycle analysis again as an example, a common feature in macroeconomic data is that a group of variables is perceived as leading the business cycle, while another group of variables moves contemporaneously with GDP. To include this feature into the model, we may impose that states of z_t^1 should lead states of z_t^2 by designing the encompassing state z_t^* with transition matrix:

$$\eta^* = \begin{bmatrix} \eta^*_{11} & \eta^*_{12} & 0 & 0 \\ 0 & \eta^*_{22} & 0 & \eta^*_{24} \\ \eta^*_{31} & 0 & \eta^*_{33} & 0 \\ 0 & 0 & \eta^*_{43} & \eta^*_{44} \end{bmatrix} \quad \begin{array}{l} z_t^* = 1,\ z_t^1 = 1 \text{ and } z_t^2 = 1 \\ z_t^* = 2,\ z_t^1 = 1 \text{ and } z_t^2 = 2 \\ z_t^* = 3,\ z_t^1 = 2 \text{ and } z_t^2 = 1 \\ z_t^* = 4,\ z_t^1 = 2 \text{ and } z_t^2 = 2 \end{array} \tag{23}$$

where $\eta^*_{g'g}$ are appropriately scaled convolutions of the transition probabilities of the underlying states z_t^1 and z_t^2. The restrictions impose a minimum duration of five periods for a full cycle. They also impose that the leading state indicator z_t^2 can only switch across states when z_t^1 has reached the same state. The approach was used by Phillips (1991) to model international data. Kaufmann (2010) uses the setup to cluster a large panel of macroeconomic data into a group of series leading the business cycle and a group of series contemporaneously moving with GDP growth. The posterior evaluation of (23) yields additional interpretations. For example, the expected lead of z_t^2 into a recovery $(z_t^1 = z_t^2 = 2)$ is $1/(1 - \eta^*_{22})$. Probabilistic forecasts about e.g. the probability of reaching the trough $(z^*_{T+2} = 2)$ within the next half year conditional on being in recession currently $(z_T^* = 1)$ are available from the forecast density $P(z^*_{T+h}|y_T, x_T, z_T^*)$.

Dynamic structure between states can also be designed by using varying time leads for transition. Paap et al. (2009) define asymmetric leads across two-state business cycle phases ($z_t \in \{0, 1\}$ in their setup):

$$z_t^2 = \begin{cases} \prod_{j=\kappa_1}^{\kappa_2} z^1_{t-j} & \text{if } \kappa_1 \geq \kappa_2 \\ 1 - \prod_{j=\kappa_2}^{\kappa_1} \left(1 - z^1_{t-j}\right) & \text{if } \kappa_1 < \kappa_2 \end{cases}$$

where for specific values like $\kappa_1 = 8$ and $\kappa_2 = 5$ one may even obtain state dynamics with overlapping phases of different cycles. In Paap et al. (2009), κ_1 and κ_2 are also part of model estimation.

Yet another example is found in Kim et al. (2005) who link hierarchically two binary state variables z_t^1, z_t^2 to a base 3-state indicator z_t with restricted transition probability

$$\eta = \begin{bmatrix} \eta_{11} & 1-\eta_{11} & 0 \\ 0 & \eta_{22} & 1-\eta_{22} \\ 1-\eta_{33} & 0 & \eta_{33} \end{bmatrix}$$

and define $z_t^1 = 1$ if $z_t = 2$ and $z_t^2 = 1$ if $z_t = 3$. In the measurement equation

$$\rho(L)\left(y_t - \theta_0 - \theta_1 z_t^1 - \theta_2 z_t^2 - \theta_3 \sum_{j=1}^{p} z_{t-j}^1\right) = \varepsilon_t$$

the last term over the sum of a fixed number p of lags is able to capture a bounce-back effect in GDP growth after recessions.

4.2 Time-varying switching

4.2.1 Duration dependence and state-identifying restrictions

Covariates affecting the transition distribution render it informative in the sense that they explain what drives the latent state. Early contributions, mainly in business cycle analysis, are Diebold et al. (1994), Filardo (1994), McCulloch and Tsay (1994) and Filardo and Gordon (1998)). Duration dependence is obtained when the persistence of states depends on the number of periods that the current regime has been prevailing. In change-point modelling, a time-varying extension is presented in Koop and Potter (2007), where regime duration is modelled by a Poisson distribution.

Recently, a multi-country, multi-state extension in business cycle analysis is proposed in Billio et al. (2016), in which the authors include weighted information on lagged business cycle states of all euro area countries and the United States into the transition distribution. The covariate x_{2t} is a weighted average of lagged country-specific state indicators, $x_{2t} = \sum_{i=1}^{n} w_{it} z_{i,g,t-1}$, with w_{it} being the trade-weight of country i.

Gaggl and Kaufmann (2014) work with a panel of 21 groups of US occupation data to analyze the phenomena of jobless recoveries that characterizes the US labor market since the early 1990s. They formulate a dedicated factor model for occupational growth with a latent four-state indicator process in the mean factor growth rate, with transition matrix

$$\eta_t = \begin{bmatrix} \eta_{t,11} & \eta_{t,12} & 0 & \eta_{t,14} \\ \eta_{t,21} & \eta_{t,22} & \eta_{t,23} & 0 \\ \multicolumn{2}{c}{0} & \eta_{t,33} & \eta_{t,34} \\ & & \eta_{t,43} & \eta_{t,44} \end{bmatrix}$$

Interpreting states 1 and 3 as recessions and states 2 and 4 as expansions, the zero restrictions imply a one-time change in business cycle phase-specific growth rates. The change can only occur at a turning point, i.e. when exiting or falling into a recession.

The explicit parametrization for the transition probabilities sets the reference state to $g_0 = 1$ and writes

$$\eta_{t,12} = \frac{\exp\left(\gamma_{12,0} + \gamma_{12,1}x_{2t} + \gamma_{12,2}t\right)}{1 + \sum_{g=\{2,4\}} \exp\left(X_t'\gamma_{1g}\right)}, \; \eta_{t,14} = \frac{\exp\left(\gamma_{14,2}t\right)}{1 + \sum_{g=\{2,4\}} \exp\left(X_t'\gamma_{1g}\right)}$$

with $X_t = (1, \ x_{2t}, \ t)$, where x_{2t} is GDP growth, and similar specifications for $\eta_{t,22}$ and $\eta_{t,23}$. The restrictions $\gamma_{j2,2} \leq 0$ and $\gamma_{j4,2} \leq 0$ identify state 2 and 4 as expansions, given that positive GDP growth $x_{2t} > 0$ increases the probability of switching to state 2 or 4. Time t helps identifying the break point. Therefore $\gamma_{14,2} \geq 0$ and $\gamma_{23,2} \geq 0$.

The results document that since the early 1990s recession, routine jobs have experienced stronger job losses during recessions while non-routine jobs experienced weaker job growth during expansions.

4.2.2 Shape restrictions

For some applications, e.g. when only few observations are expected to be assigned to a state or to be available to estimate a transition to a state, it may be useful to impose explicit restrictions on parameters in γ. This results in formulating a restricted prior $\pi(\gamma)1_{\mathcal{R}}$. For example, Bäurle et al. (2016) analyze changing dynamics in a vector autoregression for Swiss macroeconomic variables when the interest rate approaches the zero lower bound. The model specification allows for two states, $G = 2$ and sets $g_0 = 1$. The state probabilities are assumed state-independent

$$\eta_{2t} = \frac{\exp\left(\gamma^x x_{2t} + \gamma\right)}{1 + \exp\left(\gamma^x x_{2t} + \gamma\right)}$$

where x_{2t} is the lagged level of the interest rate. By restricting $\gamma^x < 0$ the probability of state 2 is increasing as the interest rate is approaching the zero lower bound.

Implicitly, the parameters $\gamma^x \neq 0$ and γ define a threshold value for x_{2t}, $\tilde{\gamma} = -\gamma/\gamma^x$, i.e. the value at which $\eta_{2t} = 0.5$. If we may have an idea about an upper bound $\overline{\gamma}$ (which should not be too high) and a lower bound $\underline{\gamma}$ (which could be e.g. between 0 and 1) for the threshold $\tilde{\gamma}$, then

$$\begin{array}{l} \underline{\gamma} < \tilde{\gamma} \leq \overline{\gamma} \\ \underline{\gamma} < -\gamma/\gamma^x \leq \overline{\gamma} \text{ or } -\gamma^x\underline{\gamma} < \gamma \leq -\gamma^x\overline{\gamma} \end{array}$$

These restrictions may be imposed on the prior distribution for γ^x, γ:

$$\pi\left(\gamma^x, \gamma\right) 1_{\mathcal{R}} = N\left(g_0, G_0\right) 1\left(\gamma^x < 0\right) 1\left(-\gamma^x\underline{\gamma} < \gamma \leq -\gamma^x\overline{\gamma}\right)$$

where the restriction $\gamma^x < 0$ is a state-identifying restriction. These restrictions render γ and γ^x highly correlated, which may be taken into account when specifying the prior moments. Figure 7 plots η_{2t} against values for x_{2t}. To implement a prior threshold value at $\tilde{\gamma} = 0.8$, stronger effects of the covariate imply a higher value for

γ. Depending on the informativeness of the data, we may be more or less informative about γ^x. For example, in Bäurle et al. (2016) only few observations for x_{2t}, the interest rate, are available near the zero lower bound to estimate the transition to state 2. Therefore, the authors use an informative prior on γ^x. Posterior draws are obtained by sampling from conditional constrained posterior distributions $\pi(\gamma^x, \gamma|x_{2t}, z_t) = \pi(\gamma|\gamma^x, x_{2t}, z_t)\pi(\gamma^x|x_{2t}, z_t)1(\gamma^x < 0)\, 1\left(-\gamma^x\underline{\gamma} < \gamma \leq -\gamma^x\overline{\gamma}\right)$, see Bäurle et al. (2016) for details.

[Insert figure 7 around here]

5 Conclusion

In time series analysis, hidden Markov models introduce persistence in the mixture distribution. The persistence is induced by defining a latent state process, which evolves according to a Markov transition distribution. This distribution may either be parameterized in a time-invariant or a time-varying way. We discuss the parametrization of the logit or probit functional form to model time-varying transition distributions. Emphasis is put on Bayesian estimation. We discuss in detail the design of state-invariant prior distributions, in particular those of the parameters of the transition distribution. We describe the random permutation sampler with which we obtain a sample from the unconstrained posterior distribution. The evaluation of estimation time and sampler efficiency between using the logit or the probit functional form reveals a strong trade-off. While estimation time does not increase significantly with the number of latent states when working with the logit functional form, there is a tenfold increase in estimation time when working with the probit functional form when increasing the number of hidden states from 2 to 3. On the other hand, draws from the probit symmetric sampler of Burgette and Hahn (2010), are as efficient as from the thinned out logit posterior. The researcher may therefore opt to work with the logit functional form to save on estimation time for models with more than 2 latent states, and to thin out considerably the posterior draws to reduce sampling inefficiency.

We illustrate how explicit economic interpretation may be obtained from posterior inference by imposing structural or dynamic restrictions on the transition distribution of the state process. Prior knowledge may also be imposed in form of restrictions onto the prior distribution for the parameters of the transition distribution. A restricted, state-identified prior then calls for restricted sampling procedures to draw from the posterior.

An attractive feature of Markov switching models has not been applied so far in time series analysis. Results in the literature derive conditions under which a stationary state-specific process can be combined with a non-stationary, explosive process. As long as the latter one does not prevail for too long nor does recur too often, the process of time series may still converge to finite moments in the indefinite future. This represents an interesting avenue for future research.

References

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

Artis, M. J., M. Marcellino, and T. Proietti (2004). Dating business cycles: A methodological contribution with an application to the euro area. *Oxford Bulletin of Economics and Statistics 66*, 537–565.

Bai, J. and P. Wang (2011). Conditional markov chain and its application in economic time series analysis. *Journal of Applied Econometrics 26*, 715–734.

Bäurle, G., D. Kaufmann, S. Kaufmann, and R. W. Strachan (2016). Changing dynamics at the zero lower bound. Working Paper 16.02, Study Center Gerzensee.

Bauwens, L., B. De Backer, and A. Dufays (2014). A bayesian method of change-point estimation with recurrent regimes: Application to GARCH models. *Journal of Empirical Finance 29*, 207–229.

Bauwens, L., G. Koop, D. Korobilis, and J. V. Rombouts (2015). The contribution of structural break models to forecasting macroeconomic series. *Journal of Applied Econometrics 30*, 596–620.

Bauwens, L. and M. Lubrano (1998). Bayesian inference on GARCH models using the Gibbs sampler. *The Econometrics Journal 1*, C23–C46.

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

Billio, M., R. Casarin, F. Ravazzolo, and H. K. van Dijk (2016). Interactions between eurozone and US booms and busts: A Bayesian panel Markov-switching VAR model. *Journal of Applied Econometrics*, forthcoming.

Burgette, L. F. and P. R. Hahn (2010). Symmetric Bayesian multinomial probit models. Working Paper 10-26, Department of Statistical Science, Duke University.

Chauvet, M. and J. Piger (2008). A comparison of the real-time performance of business cycle dating methods. *Journal of Business & Economic Statistics 26*, 42–49.

Chib, S. (1995). Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association 90*, 1313–1321.

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.

Clements, M. and H.-M. Krolzig (2003). Business cycle asymmetries: Characterization and testing based on markov-switching autoregressions. *Journal of Business & Economic Statistics 21*, 196–211.

Costa, O., M. Fragoso, and R. Marques (2004). *Discrete-Time Markov Jump Linear Systems*. New York: Springer.

Davig, T. and E. M. Leeper (2007). Generalizing the Taylor principle. *The American Economic Review 97*, 607–635.

Diebold, F. X., J.-H. Lee, and G. C. Weinbach (1994). Regime switching with time-varying transition probabilities. In C. P. Hargreaves (Ed.), *Nonstationary Time Series Analysis and Cointegration*, pp. 283–302. Oxford University Press.

Elliott, G. and A. Timmermann (2005). Optimal forecast combination under regime switching. *International Economic Review 46*, 1081–1102.

Farmer, R. E., D. F. Waggoner, and T. Zha (2009a). Indeterminacy in a forward-looking regime switching model. *International Journal of Economic Theory 5*, 69–84.

Farmer, R. E., D. F. Waggoner, and T. Zha (2009b). Understanding Markov-switching rational expectations models. *Journal of Economic Theory 144*, 1849–1867.

Farmer, R. E., D. F. Waggoner, and T. Zha (2011). Minimal state variable solutions to Markov-switching rational expectations models. *Journal of Economic Dynamics and Control 35*, 2150–2166.

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.

Foerster, A., J. Rubio-Ramrez, D. F. Waggoner, and T. Zha (2016). Perturbation methods for Markov-switching dynamic stochastic general equilibrium models. *Quantitative Economics 7*, 637669.

Francq, C. and J.-M. Zakoïan (2001). Stationarity of multivariate Markov switching ARMA models. *Journal of Econometrics 102*, 339 – 364.

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. (2004). Estimating marginal likelihoods for mixture and markov switching models using bridge sampling techniques. *The Econometrics Journal 7*, 143–167.

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

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

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

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

Frühwirth-Schnatter, S. and S. Kaufmann (2006). How do changes in monetary policy affect bank lending? An analysis of Austrian bank data. *Journal of Applied Econometrics 21*, 275–305.

Gaggl, P. and S. Kaufmann (2014). The cyclical component of labor market polarization and jobless recoveries in the U.S. Working Paper 14.03, Study Center Gerzensee.

Geweke, J. and G. Amisano (2011). Hierarchical markov normal mixture models with applications to financial asset returns. *Journal of Applied Econometrics 26*, 1–29.

Geweke, J., M. Keane, and D. Runkle (1994). Alternative computational approaches to inference in the multinomial probit model. *Review of Economics & Statistics 76*, 609–632.

Goldfeld, S. M. and R. E. Quandt (1973). A Markov model for switching regressions. *Journal of Econometrics 1*, 3–16.

Gray, S. (1996). Modelling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics 42*, 27–62.

Greene, W. H. and D. A. Hensher (2003). A latent class model for discrete choice analysis: contrasts with mixed logit. *Transportation Research Part B 37*, 681–698.

Greene, W. H. and D. A. Hensher (2013). Revealing additional dimensions of preference heterogeneity in a latent class mixed multinomial logit model. *Applied Economics 45*, 1897–1902.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica 57*, 357–384.

Hamilton, J. D. (1990). Analysis of time series subject to changes in regime. *Journal of Econometrics 45*, 39–70.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.

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

Hamilton, J. D. and R. Susmel (1994). ARCH and changes in regime. *Journal of Econometrics 45*, 307–333.

Hansen, B. E. (1992). The likelihood ratio test under nonstandard conditions: Testing the Markov switching model of GNP. *Journal of Applied Econometrics 7*, S61–S82.

Hess, S. (2014). Latent class structures: taste heterogeneity and beyond. In S. Hess and A. Daly (Eds.), *Handbook of Choice Modelling*, pp. 311–329. The Lypiatts, 15 Lansdown Road, Cheltenham, Glos GL50 2JA, UK: Edward Elgar Publishing Limited.

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

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

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

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

Kaufmann, S. and P. Kugler (2010). A monetary real-time forecast of euro area inflation. *Journal of Forecasting 29*, 388–405.

Kim, C.-J., J. Morley, and J. Piger (2005). Nonlinearity and the permanent effects of recessions. *Journal of Applied Econometrics 20*, 291–309.

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.

Klaassen, F. (2002). Improving GARCH volatility forecasts with regime-switching GARCH. *Empirical Economics 27*, 363–394.

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

Krolzig, H.-M., M. Marcellino, and G. Mizon (2002). A Markov-switching vector equilibrium correction model of the UK labour market. *Empirical Economics 27*, 233–254.

McCulloch, R. and R. Tsay (1994). Statistical analysis of economic time series via Markov switching models. *Journal of Time Series Analysis 15*, 523–539.

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

McFadden, D. (1974). Conditional logit analysis of qualitative choice behaviour. In P. Zarembka (Ed.), *Frontiers of Econometrics*, pp. 105–142. New York: Adacemic.

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

Owyang, M. T., J. M. Piger, and H. J. Wall (2015). Forecasting national recessions using state level data. *Journal of Money Credit and Banking 47*, 847–866.

Paap, R., R. Segers, and D. van Dijk (2009). Do leading indicators lead peaks more than troughs? *Journal of Business & Economic Statistics 27*, 528–543.

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

Phillips, K. L. (1991). A two-country model of stochastic output with changes in regime. *Journal of International Economics 31*, 121–142.

Psaradakis, Z., M. O. Ravn, and M. Sola (2005). Markov switching causality and the moneyoutput relationship. *Journal of Applied Econometrics 20*, 665–683.

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

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

Tables

Table 1: Time in minutes for 1,000 draws and inefficiency factors for posterior samples retaining all draws and only every 5th draw.

		logit				probit			
$G = 2$									
Time		3.2				3.4			
Inefficiency		$g = 2$				$g = 2$			
		all	5th			all	5th		
$\gamma_{g'g} =$	γ^x_{1g}	80.83	22.31			86.64	18.95		
	γ^x_{2g}	53.65	9.60			44.59	9.64		
	γ_{1g}	39.85	10.00			36.37	6.00		
	γ_{2g}	41.07	8.19			17.61	3.17		
$G = 3$									
Time		3.4				38.6			
Inefficiency		$g = 2$		$g = 3$		$g = 2$		$g = 3$	
		all	5th	all	5th	all	5th	all	5th
$\gamma_{g'g} =$	γ^x_{1g}	19.85	2.60	12.59	2.76	3.66	1.68	2.28	0.87
	γ^x_{2g}	13.19	3.83	19.43	2.37	2.15	0.65	3.44	1.13
	γ^x_{3g}	16.95	2.78	19.53	4.30	3.49	1.03	2.86	1.08
	γ_{1g}	9.12	2.19	16.56	2.81	2.94	0.84	2.65	1.05
	γ_{2g}	14.82	3.85	21.53	4.13	3.01	1.22	3.17	0.94
	γ_{3g}	22.05	5.76	9.86	2.38	2.55	0.60	3.13	0.61

Figures

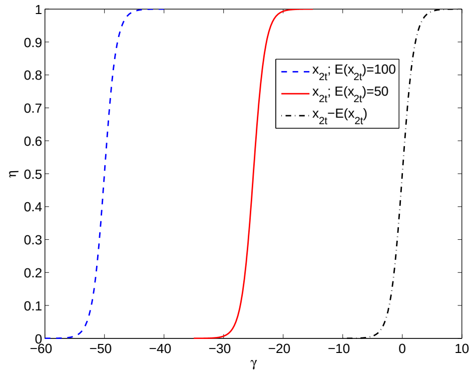

Figure 1: Range of γ against $\eta = \exp\left(x_{2t}\gamma^x + \gamma\right) / \left(1 + \exp(x_{2t}\gamma^x + \gamma)\right)$, with $\gamma^x = 0.5$, if $\eta = 0.5$ when x_{2t} equals its mean of, respectively, 100 and 50 (dashed and solid line). The range of γ is scale-invariant (dash-dotted line) if the covariate $x_{2t} - E(x_{2t})$ is mean-adjusted up-front.

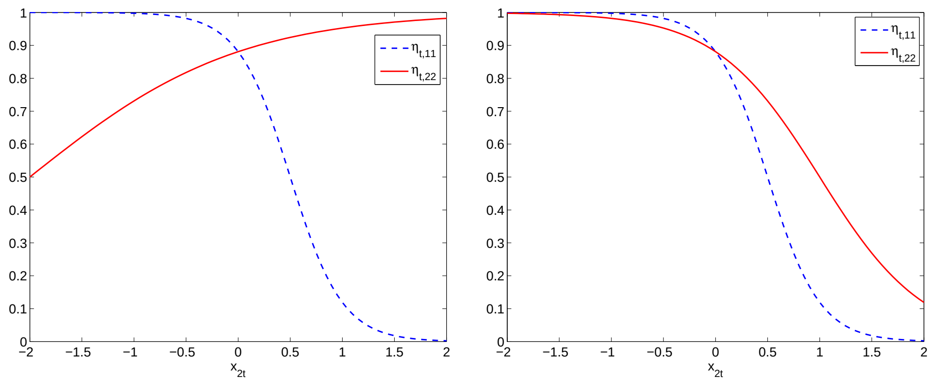

Figure 2: Persistence probabilities $\eta_{t,g'g} \propto \exp\left(x_{2t}\gamma^x_{g'g} + \gamma_{g'g}\right)$. Left panel: $\gamma = (\gamma^x_{12}, \gamma^x_{22}, \gamma_{12}, \gamma_{22}) = (4, 1, -2, 2)$, right panel $\gamma = (4, -2, -2, 2)$.

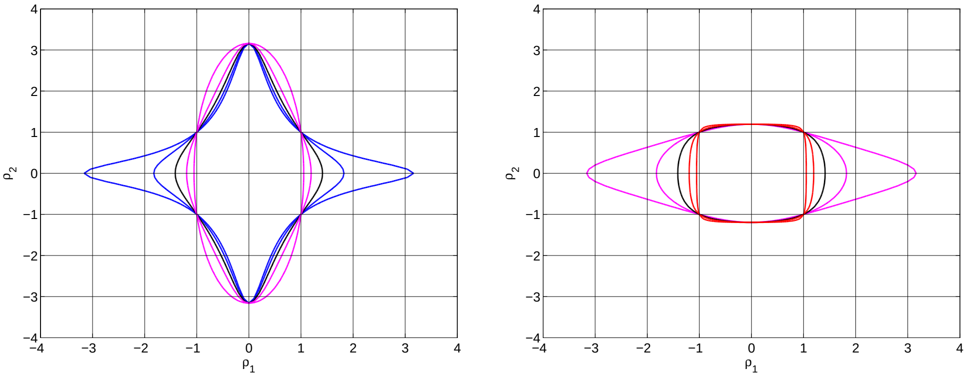

Figure 3: Boundary values for $\rho_1, \rho_2|\eta$ to obtain a mean-square stable Markov switching univariate autoregressive process. Left panel: Conditional on $(\eta_{11}, \eta_{22} = 0.1)$, where $\eta_{11} = \{0.1, 0.3, \ldots, 0.9\}$ from, respectively, the peaked to the oval boundary. Right panel: Conditional on $(\eta_{11}, \eta_{22} = 0.7)$, where $\eta_{11} = \{0.1, 0.3, \ldots, 0.9\}$ from, respectively, the oval-peaked to the square-like boundary.

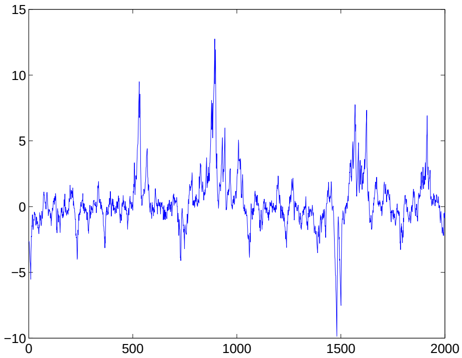

Figure 4: Simulated time series with $\eta_{11} = 0.5$, $\rho_1 = 1.1$, $\eta_{22} = 0.7$, $\rho_2 = 0.7$, $\sigma^2 = 0.1$.

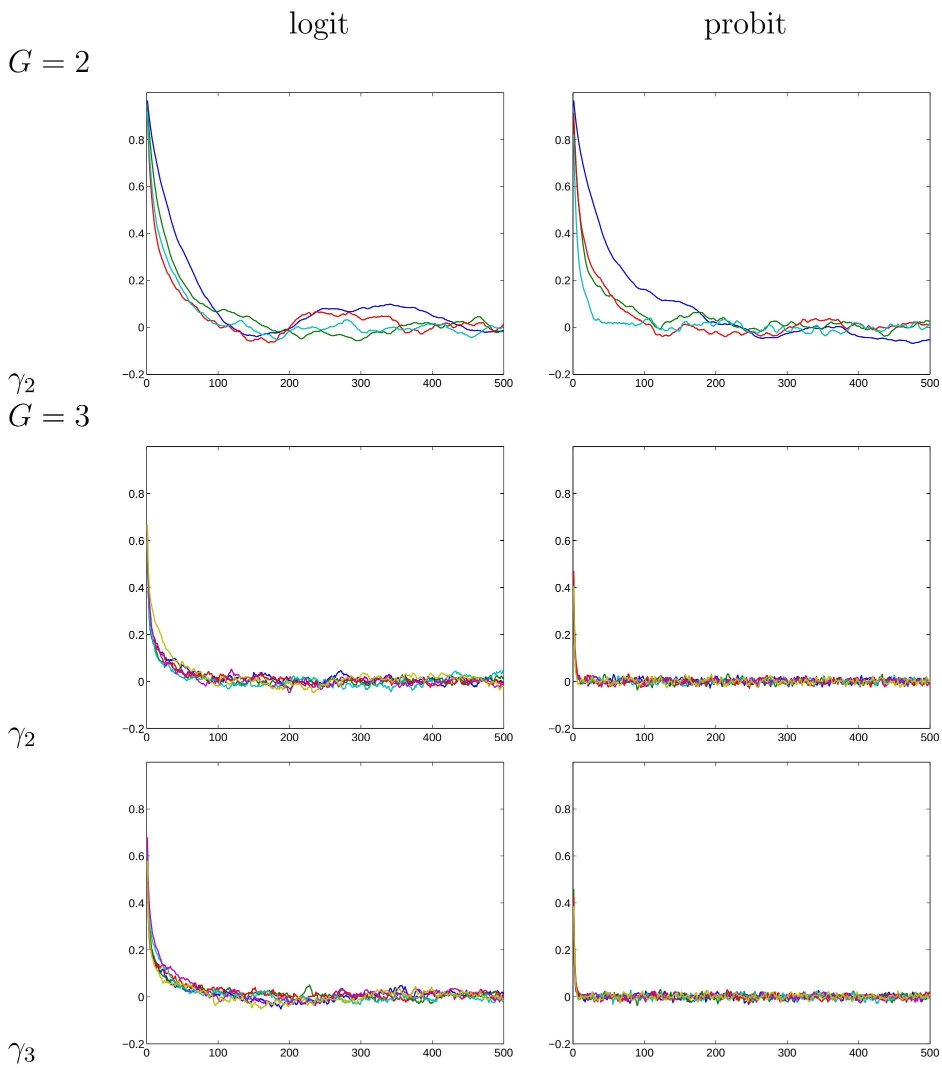

Figure 5: Autocorrelation of state-identified posterior draws for γ_g, $G = \{2, 3\}$.

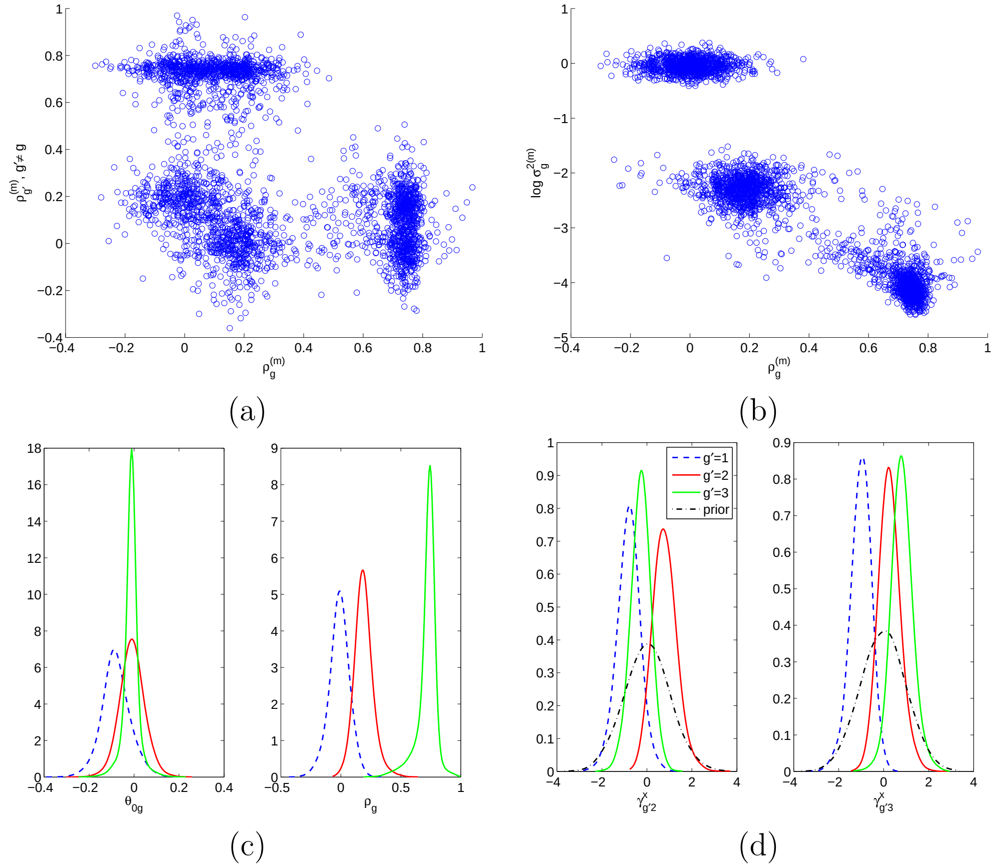

Figure 6: Scatter plots of posterior output, $G = 3$. First line: Unsorted output. Second line: State-identified output, marginal posterior distributions and prior distribution for γ^x.

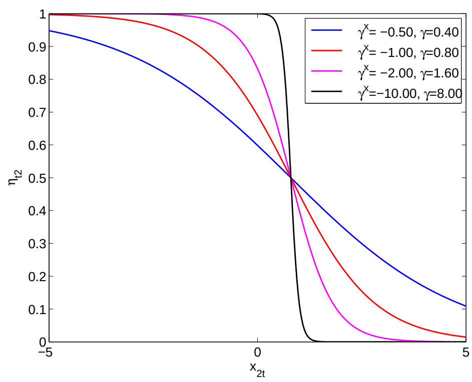

Figure 7: State probabilities $\eta_{t,2}$, implemented to keep the threshold at $\tilde{\gamma} = 0.8$ for increasing covariate effect.