



Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Song, Y;Woźniak, T

Title:

Markov Switching

Date:

2021

Citation:

Song, Y. & Woźniak, T. (2021). Markov Switching. Oxford Research Encyclopedia of Economics and Finance, pp.1-34. <https://doi.org/10.1093/acrefore/9780190625979.013.174>.

Persistent Link:

<https://hdl.handle.net/11343/268243>

Markov Switching

Yong Song, Tomasz Woźniak

Department of Economics, University of Melbourne

Abstract

Markov switching models are a family of models that introduces time variation in the parameters in the form of their state, or regime-specific values. This time variation is governed by a latent discrete-valued stochastic process with limited memory. More specifically, the current value of the state indicator is determined by the value of the state indicator from the previous period only implying the Markov property. A transition matrix characterizes the properties of the Markov process by determining with what probability each of the states can be visited next period conditionally on the state in the current period. This setup decides on the two main advantages of the Markov switching models: the estimation of the probability of state occurrences in each of the sample periods by using filtering and smoothing methods and the estimation of the state-specific parameters. These two features open the possibility for interpretations of the parameters associated with specific regimes combined with the corresponding regime probabilities.

The most commonly applied models from this family are those that presume a finite number of regimes and the exogeneity of the Markov process, which is defined as its independence of the model unpredictable innovations. In many such applications, the desired properties of the Markov switching model have been obtained either by imposing appropriate restrictions on transition probabilities or by introducing the time dependence of these probabilities determined by explanatory variables or functions of the state indicator. One of the extensions of this basic specification includes infinite hidden Markov models that provide great flexibility and excellent forecasting performance by allowing the number of states to go to infinity. Another extension, the endogenous Markov switching model, explicitly relates the state indicator to the model's innovations, making it more interpretable and offering promising avenues for development.

Keywords: Bull and Bear Markets, Business Cycle Analysis, Change-point Model, Endogenous Markov Switching, Exogenous Markov Switching, Finite Mixture Model, Infinite Hidden Markov Model, Markov Process, Non-homogenous Markov Switching, Transition Probabilities

^{*}This is the authors' accepted copy of a paper
Song, Woźniak (2021) Markov Switching, *Oxford Research Encyclopedia of Economics and Finance*

A Primer on Markov Switching

The Markov switching (MS) methodology was introduced in the seminal works of [Goldfeld and Quandt \(1973\)](#) and [Hamilton \(1989\)](#). It is directly applicable to time series analysis for its dynamic nature. This section presents the benchmark model and the corresponding notation for the data and model parameters. For a more comprehensive textbook exposition see [Hamilton \(1994\)](#), [Krolzig \(1997\)](#), [Kim and Nelson \(1999\)](#) and [Frühwirth-Schnatter \(2006\)](#).

The dependent variable at time t is denoted by y_t , for $t = 1, \dots, T$, where T is the number of periods in the sample. y_t can be either a scalar, vector, or matrix. The independent variable being a scalar, vector, or a matrix is denoted by x_t . x_t does not need to have the same dimension as y_t , and it might include the lagged values of y_t .

Markov Process

A latent state at time t is unobservable to an econometrician and denoted by s_t . It takes the value $k \in \{1, 2, \dots, K\}$, where K is a positive integer representing the total number of states. The latent variable s_t indicates in which state the current system is at time t . Hence, it can be called *state indicator* or *regime indicator*. The terms *state* and *regime* are often used interchangeably.

The dynamics of the state indicator are governed by a Markov process. The probability distribution of s_t given the whole path $\{s_{t-1}, s_{t-2}, \dots, s_1\}$ only depends on the most recent state s_{t-1} . Define a *transition probability* as

$$\Pr(s_t = j \mid s_{t-1} = i, s_{t-2}, \dots, s_1) = \Pr(s_t = j \mid s_{t-1} = i) = p_{ij}, \quad (1)$$

where $i, j = 1, \dots, K$. Given that in period $t - 1$ the process was in state i the probability that the state will switch to state j in period t is equal to p_{ij} . A *transition matrix* organizes these transition probabilities in a $K \times K$ matrix and is defined as

$$P = [p_{ij}]_{K \times K}, \quad (2)$$

where p_{ij} is the element on the i th row and j th column such that the elements in each of the rows of matrix P sum to one. The diagonal elements of the transition matrix determine the expected state duration that is equal to $(1 - p_{ii})^{-1}$ for state i .

A vector of unconditional state probabilities, denoted by $\pi \equiv \Pr[s_t]$, is defined by

$$P' \pi = \pi. \quad (3)$$

This equation indicates time invariance of distribution π , that is, an iteration over one period performed by the premultiplication by the transition probability P' does not change vector π . The solution to equation (3) for π given $\iota_K' \pi = 1$, where ι_K is a K -vector of ones, is given in [Hamilton \(1994, Chapter 22\)](#) and expresses π as a function of P . Define a $(K+1) \times K$ matrix $\mathbf{P} = \begin{bmatrix} I_K - P' \\ \iota_K' \end{bmatrix}$, where I_K denotes an identity matrix of order K . Hamilton's solution

for π is given by the $(K + 1)$ th column of $(\mathbf{P}'\mathbf{P})^{-1}\mathbf{P}'$.

The distribution of the initial state at $t = 0$, denoted by s_0 , is represented by the following K -vector

$$\pi_0 \equiv [\Pr(s_0 = k)]_{K \times 1}. \quad (4)$$

For an ergodic Markov process, the initial distribution can be simply set to the stationary distribution π . If the Markov process is non stationary, there typically exists a theory-guided choice for π_0 . For example, a change-point model requires that at the initial period the process is in the first regime, $s_0 = 1$.

Measurement Equation

A measurement equation lays out the probability law of the observations and is given by

$$y_t \sim F(x_t, s_t), \quad (5)$$

where F represents the distribution of y_t conditional on the observation x_t and the latent state s_t . F should be specified in applications and depends on the structure of the data y_t . Therefore, it can be given by a discrete, continuous, or mixture distribution. For example, consider a time series of financial asset returns and assume that each state admits an autoregressive process of order one with Gaussian innovations. Then F can be expressed as

$$F(x_t, s_t) \equiv N(\mu_{s_t} + \beta_{s_t}y_{t-1}, \sigma_{s_t}^2),$$

where $\mu_{s_t} + \beta_{s_t}y_{t-1}$ is the mean and $\sigma_{s_t}^2$ the variance of this conditional distribution given s_t and y_{t-1} . The same assumption also implies a regression form of equation (5) given by

$$y_t = \mu_{s_t} + \beta_{s_t}y_{t-1} + \sigma_{s_t}\epsilon_t, \quad \epsilon_t \sim N(0, 1). \quad (6)$$

Equations (1) and (5) comprise the foundation of the MS framework. Combining them with the initial condition (4) yields the likelihood function $p(Y|\Theta)$ that is available through the filter technique from [Hamilton \(1989\)](#) called the *Hamilton filter*, where Y denotes a collection of all y_t for $t = 1, \dots, T$; and Θ is the collection of all time-invariant parameters. For instance, in the example from equation (6), $\Theta \equiv \{\mu_k, \beta_k, \sigma_k^2\}_{k=1}^K$.

Filtering and Smoothing the States

The Hamilton filter gives the conditional distribution of the state s_t given the data up to time t , denoted by $p(s_t|Y_{1:t})$. It is used to compute the one-period-ahead probability

density of y_t from the following formula:

$$p(y_{t+1} | Y_{1:t}) = \sum_{s_{t+1}=1}^K p(y_{t+1}, s_{t+1} | Y_{1:t}) \quad (7a)$$

$$= \sum_{s_{t+1}=1}^K p(y_{t+1} | s_{t+1}, Y_{1:t}) p(s_{t+1} | Y_{1:t}) \quad (7b)$$

$$= \sum_{s_{t+1}=1}^K \left[p(y_{t+1} | s_{t+1}, Y_{1:t}) \sum_{s_t=1}^K p(s_{t+1}, s_t | Y_{1:t}) \right] \quad (7c)$$

$$= \sum_{s_{t+1}=1}^K \left[p(y_{t+1} | s_{t+1}, Y_{1:t}) \sum_{s_t=1}^K p(s_{t+1} | s_t, Y_{1:t}) p(s_t | Y_{1:t}) \right] \quad (7d)$$

$$= \sum_{s_{t+1}=1}^K \left[p(y_{t+1} | s_{t+1}, Y_{1:t}) \sum_{s_t=1}^K p(s_{t+1} | s_t) p(s_t | Y_{1:t}) \right] \quad (7e)$$

$$= \sum_{s_{t+1}=1}^K \sum_{s_t=1}^K p(y_{t+1} | s_{t+1}, Y_{1:t}) p(s_{t+1} | s_t) p(s_t | Y_{1:t}), \quad (7f)$$

where $p(y_{t+1}|s_{t+1}, Y_{1:t})$ is obtained from the measurement equation (5) and $p(s_{t+1}|s_t)$ is the transition probability in equation (1). Equation (7b) includes the forecasted state probability that is conveniently decomposed into the transition and filtered probabilities according to

$$p(s_{t+1} | Y_{1:t}) = \sum_{s_t=1}^K p(s_{t+1} | s_t) p(s_t | Y_{1:t}). \quad (8)$$

The filtered probability for s_{t+1} is given by

$$p(s_{t+1} | Y_{1:t+1}) = \frac{p(s_{t+1}, y_{t+1} | Y_{1:t+1})}{p(y_{t+1} | Y_{1:t+1})} \quad (9a)$$

$$= \frac{p(y_{t+1} | s_{t+1}, Y_{1:t}) p(s_{t+1} | Y_{1:t})}{p(y_{t+1} | Y_{1:t})} \quad (9b)$$

$$= \frac{p(y_{t+1} | s_{t+1}, Y_{1:t}) \sum_{s_t=1}^K p(s_{t+1} | s_t) p(s_t | Y_{1:t})}{p(y_{t+1} | Y_{1:t})}. \quad (9c)$$

The filtered probability of s_{t+1} is easy to compute as long as the filtered probability of s_t is known. The conditional density $p(y_{t+1}|Y_{1:t})$ is derived in equation (7) and can be easily computed in an iterative procedure. The filtering algorithm proceeds iteratively for $t = 1, \dots, T$ by the application of equation (9c) that describes the progression of the filtered probabilities from $p(s_t|Y_{1:t})$ to $p(s_{t+1}|Y_{1:t+1})$. Note that the filtering step includes a

prediction step consisting of computing the forecasted probability $p(s_{t+1}|Y_{1:t})$.

The computation of the smoothed probabilities denoted by $p(s_t|Y_{1:T})$ for all t is performed by an iterative backward smoothing procedure proposed by [Chib \(1996\)](#). It is initiated at $t = T$ where the last filtered density is equal to the smoothed probability at time T , $p(s_T|Y_{1:T})$, and then proceeds for $t = T - 1, \dots, 1$ by applying:

$$p(s_t | Y_{1:T}) = \sum_{s_{t+1}=1}^K \frac{p(s_{t+1} | s_t)p(s_t | Y_{1:t})p(s_{t+1} | Y_{1:T})}{\sum_{s_{t+1}=1}^K p(s_{t+1} | s_t)p(s_t | Y_{1:t})}. \quad (10)$$

A comprehensive exposition on the forward filtering and the backward smoothing procedures can be found in [Frühwirth-Schnatter \(2006, Chapter 13\)](#).

Finally, the likelihood function is constructed as the product of conditional probability densities:

$$p(Y | \Theta) = \prod_{t=1}^T p(y_t | Y_{1:t-1}, \Theta), \quad (11)$$

where $Y_{1:0}$ is treated as known and the state indicators, s_t for $t = 1, \dots, T$, are integrated out. An analytical expression for the likelihood function is based on equations (11) and (7b) and is written out in [Frühwirth-Schnatter \(2006\)](#).

The presented derivations are valid for homogenous Markov switching models, that is, the models for which $p(s_{t+1}|s_t, Y_{1:t}) = P$. The filtering and smoothing algorithms complicate for dynamic models with path dependence in which the distribution of the current state depends on the path of all of the past latent state indicators. Some early solutions were proposed by [Billio and Monfort \(1998\)](#) and [Billio, Monfort, and Robert \(1999\)](#) for autoregressive moving average (ARMA) models and by [Haas, Mittnik, and Paoletta \(2004\)](#) for generalized autoregressive conditional heteroskedasticity (GARCH) models. The application of Sequential Monte Carlo methods was proposed by [Bauwens, Dufays, and Rombouts \(2014\)](#).

The interpretation of the MS models is based on the combined analysis of the parameter estimates and the estimate of the state probabilities. The latter should be chosen among the forecasted, filtered, or smoothed probabilities, denoted by $p(s_t|Y_{1:t-1})$, $p(s_t|Y_{1:t})$ and $p(s_t|Y_{1:T})$, respectively, depending on a particular application and the objective of investigation.

Parameter Estimation

The maximum likelihood estimation is based on an iterative expectation-maximization algorithm (see [Hamilton, 1990](#)). In each of its iterations, a filtering-smoothing algorithm is used to propose the current estimate of S , a collection of all s_t for $t = 1, \dots, T$, and a maximization step is applied to compute an estimate of Θ and P . The maximum likelihood estimation is straightforward if regularity conditions are satisfied. However, for larger models, it can become cumbersome due to unbounded likelihood function.

The first results for the asymptotic normality of the maximum likelihood estimator of the MS model parameters were proposed by [Lindgren \(1978\)](#) and [Lehmann \(1983\)](#), whereas the finite sample properties of this estimator were analyzed in [Psaradakis and](#)

Sola (1998). The latter study using Monte Carlo simulations for a simple univariate MS model documents strong irregularities in the shape of the small sample distribution of the parameters, including larger estimation standard errors and skewness, in samples of up to 400 observations, especially for variance parameters and transition probabilities. Such irregularities were much less pronounced for the conditional mean parameters and usually disappeared for samples exceeding 800 observations.

Bayesian estimation relies on a data augmentation technique that requires the specification of a complete-data likelihood function $p(Y, S|\Theta) = p(Y|S, \Theta)p(S|\Theta)$, where the objects on the righthand side of the expression are easily obtainable without integration. The complete-data likelihood function is subsequently used to specify the full conditional posterior distributions $p(\Theta|Y, S)$ and $p(S|Y, \Theta)$. Sampling from the joint posterior distribution of the parameters and states is performed via Markov Chain Monte Carlo (MCMC) methods. Since the former conditional distribution takes S as given, it can be sampled from using standard techniques. The sampling from the latter relies on an iterative procedure, the Forward Filtering and Backward Sampling (FFBS) method proposed by Chib (1996), as the latent state indicators are not conditionally *a posteriori* independent. In order to draw samples from $p(S|Y, \Theta)$, apply the forward filtering algorithm and save the filtered probabilities $p(s_t|Y_{1:t})$ for all t . Then, at the m th iteration of the MCMC algorithm sample $s_T^{(m)}$ from $p(s_T|Y_{1:T})$, and sample $s_t^{(m)}$ backwards for each $t = T - 1, \dots, 1$ from

$$p\left(s_t | s_{t+1}^{(m)}, Y_{1:t}\right) = \frac{p\left(s_{t+1}^{(m)} | s_t\right) p\left(s_t | Y_{1:t}\right)}{\sum_{s_{t+1}=1}^K p\left(s_{t+1}^{(m)} | s_t\right) p\left(s_t | Y_{1:t}\right)}. \quad (12)$$

Frühwirth-Schnatter (2006) surveys various versions of this algorithm.

Inference on the Number of States

The frequentist solution to a problem of selection of the number of states of the Markov process K relies on information criteria such as the Akaike Information Criterion (see Psaradakis and Spagnolo, 2003). Testing of a hypothesis that $K = 1$ against an alternative hypothesis that $K = 2$ is highly cumbersome because the MS model is not identified under the null hypothesis, and the solutions require sophisticated inferential methods some of which are provided by Carrasco, Hu, and Ploberger (2014) and Meitz and Saikkonen (2017).

Bayesian model selection based on marginal data densities provides the solution to the number of states determination problem, including a model with $K = 1$ (see Frühwirth-Schnatter, 2006). However, the main challenge for a correct Bayesian inference in MS models is the problem of label switching which is defined as the invariance of the likelihood function to various labelings of the states. Consider an example in equation (6) with two states characterized by state-specific parameter vectors of state A and state B , denoted respectively by Θ_A and Θ_B . The label switching problem states that irrespective of whether the vector of parameters (Θ_1, Θ_2) is set to (Θ_A, Θ_B) or (Θ_B, Θ_A) the value of the likelihood function evaluated at these parameters stays invariant. Frühwirth-Schnatter

(2001) analyzes a multimodal global shape of the likelihood function and the posterior distribution (see Frühwirth-Schnatter, 2004, for the application of this approach to the computation of marginal data densities). Implementing ordering restrictions on the state-dependent parameters of the model that would provide a unique classification of the states is a solution that is only applicable if such restrictions do not numerically bind the posterior distribution. Alternatively, Geweke (2007) proposes basing the inference on label-switching invariant characteristics such as predictive densities.

Forecasting

MS models offer an improved forecasting accuracy due to assigning a significant weight to the most recent regime, if the states are persistent, and by incorporating the possibility of switching to other states over the forecast period. Consider forecasted state probabilities for an h -period-ahead forecast that are given by

$$\pi_{t+h|t} = p(s_{t+h} | Y_{1:t}) = \sum_{s_{t+h-1}=1}^K \cdots \sum_{s_t=1}^K p(s_{t+h} | s_{t+h-1}) \cdots p(s_{t+1} | s_t) p(s_t | Y_{1:t}) \quad (13a)$$

$$= P^h \begin{bmatrix} p(s_t = 1 | Y_{1:t}) \\ \vdots \\ p(s_t = M | Y_{1:t}) \end{bmatrix} \quad (13b)$$

where the last equality holds for the homogenous MS models. These forecasted probabilities play the role of the weight assigned to the regime-dependent predictive densities. Therefore, the joint predictive density for all the forecasted variables up to forecast horizon h conditional on the data, $Y_{1:t}$, and parameters, Θ , is as follows

$$p(y_{t+h}, \dots, y_{t+1} | Y_{1:t}, \Theta) = \prod_{i=1}^h \left(\sum_{s_{t+i}=1}^K p(y_{t+i} | Y_{1:t+i-1}, s_{t+i}, \Theta) p(s_{t+i} | Y_{1:t}, \Theta) \right). \quad (14)$$

Given specific assumptions of the considered models it is often possible to derive analytical formulae for the conditional expected value point forecast and the associated forecast error variance. For instance, the moments of the one-period-ahead forecast of the model in equation (6) are given by:

$$E[y_{t+h} | y_{t+h-1}, \Theta] = \left(\sum_{s_{t+h}=1}^K \pi_{t+h|t, s_{t+h}} \mu_{s_{t+h}} \right) + \left(\sum_{s_{t+h}=1}^K \pi_{t+h|t, s_{t+h}} \beta_{s_{t+h}} \right) y_{t+h-1} \quad (15a)$$

$$Var[y_{t+h} | y_{t+h-1}, \Theta] = \sum_{s_{t+h}=1}^K \pi_{t+h|t, s_{t+h}} \sigma_{s_{t+h}}^2 \quad (15b)$$

where $\pi_{t+h|t, s_{t+h}}$ denotes the s_{t+h} th element of vector $\pi_{t+h|t}$. Note that the predictive density is not normal but a mixture of normal distributions that might exhibit such features

as leptokurtosis, multimodality, and asymmetry. For more complicated models, the employment of numerical integration methods might be necessary. This and several other aspects of forecasting with MS models are presented in [Krolzig \(1997\)](#) and [Teräsvirta \(2006\)](#).

A concept Granger causality relates the causal link between variables to their predictive power and was proposed by [Granger \(1969\)](#) and [Sims \(1980\)](#). An analysis for a specific state of an MS vector autoregressive model was proposed by [Psaradakis, Ravn, and Sola \(2005\)](#). The framework that investigates Granger causality in the mean unconditionally on the states and treats the MS vector autoregressive model as a stochastic process was proposed by [Droumaguet, Warne, and Woźniak \(2017\)](#), who also introduced parametric conditions under which a variable does not affect the forecast of the hidden Markov process. [Warne \(2000\)](#) proposes a related analysis of Granger causality in the variance.

Selected Applications in Economics and Finance

The first application of the MS models in economics was proposed by [Hamilton \(1989\)](#) and consisted of the analysis of business cycles. Consider data on the gross domestic product growth rates to which the following autoregressive model is fitted:

$$(y_t - \mu_{s_t}) = \beta_{s_t}(y_{t-1} - \mu_{s_{t-1}}) + \epsilon_t, \quad (16)$$

where the error term is conditionally normally distributed. The business cycle interpretation of the model relied on the combined analysis of the signs of the regime-specific intercept terms and the historical narrative about the periods with high values of the smoothed state probabilities for each of the regimes. Accordingly, a negative value of the intercept term coincided with the periods of economic recessions, whereas its positive value was associated with economic expansions. Similar reasoning was applied to financial markets characterized by bull and bear markets occurring one after another by [Hamilton and Lin \(1996\)](#). An important feature of the model in equation (16) is that it implies autocorrelations that are functions of the autoregressive parameter, as is the case in linear autoregressive models with $K = 1$, but also of the elements of the transition matrix (see [Krolzig, 1997](#)).

Another example includes a multivariate model of the effects of monetary policy on the real economy in the United States with conditional heteroskedasticity modeled with the MS process proposed by [Sims and Zha \(2006\)](#). In this model, the volatility states occurred to have high values of smoothed probabilities in the periods corresponding to the terms of subsequent chairs of the Federal Reserve. For instance, the state with the highest value of the volatility had high probabilities of occurrence in the period of Paul Volcker's chairmanship. In contrast, the lowest volatility state spanned Alan Greenspan's term. [Lanne, Lütkepohl, and Maciejowska \(2010\)](#) and [Lütkepohl and Woźniak \(2020\)](#) obtained a similar interpretation of MS heteroskedastic states and used it to identify a monetary policy shock in a structural dynamic model. In another example of multivariate analysis, an explicit form of dependence between two or more Markov processes describing country or regional business cycles was proposed by [Owyang, Piger, and Wall \(2005\)](#), [Hamilton](#)

and Owyang (2012), and Leiva-Leon (2017).

An example of introducing MS in a structural model of an economy was introduced in a new class of structural MS rational expectations models. These models use the MS rule to determine the time variation of the structural shock variances of the Dynamic Stochastic General Equilibrium model and to model the time-varying inflation target of a central bank. In this framework, the considered agents are rational and, therefore they know the MS rule and, consequently, take it into account in their decision-making problem. Farmer, Waggoner, and Zha (2009) provides the theory behind such a formulation of the model while Liu, Waggoner, and Zha (2011) provided a model for macroeconomic fluctuations. Farmer, Waggoner, and Zha (2011) propose a method of solving these models. A novel approach to temporal model selection problem in this context was proposed by Waggoner and Zha (2012). They set as two states of an MS model two macroeconomic models, a Dynamic Stochastic General Equilibrium model and the corresponding vector autoregressive specification, allowing the state allocation to decide which of these two models fits the data better in each of the sample periods.

The time variation of the parameters according to the MS rule was applied with success to dynamic models of conditional variances of financial asset returns. This extension occurred essential from the point of view of making statements about the persistence of the volatility process that changes over time. Hamilton and Susmel (1994) and Kaufmann and Frühwirth-Schnatter (2002) introduce MS in the parameters of the Autoregressive Conditional Heteroskedasticity (ARCH) model whereas Bauwens et al. (2014) and Augustyniak, Bauwens, and Dufays (2019) applied such a process to the parameters of the GARCH models and propose solutions to the path dependence problem. Finally, So, Lam, and Li (1998) and Carvalho and Lopes (2007) implemented the MS mechanism to the parameters of the state equation for the logarithm of conditional variance of financial returns in a Stochastic Volatility model.

The MS model was also used to capture time varying contagion and systematic risk on financial markets as in Gallo and Otranto (2008) and Casarin, Sartore, and Tronzano (2018).

Exogenous Markov Switching

The MS model is defined through the likelihood function in which the predictive densities of the data $p(y_{t+1}|s_{t+1}, Y_{1:t})$ are weighted by the forecasted state probabilities $p(s_{t+1}|Y_{1:t})$ as in equation (7b). The independence of forecasted state probability $p(s_{t+1}|Y_{1:t})$ from the contemporaneous error term of the measurement equation (5) defines a popular family of exogenous MS models with a finite number of states. The properties of the latent Markov process in these models are driven by the form of the transition matrix P . This section revises various forms of this matrix and analyzes the implied properties of the Markov process. Models with an infinite number of states and with endogenous Markov process are discussed.

A Family of Markov Switching Models

A general stationary and aperiodic MS process in which each of the states can be revisited at any time t presumes that all of the K^2 elements of the transition matrix are estimated. In such a model, there is no absorbing state, all of the elements of the ergodic probabilities vector π are greater than zero, and the probabilities of the initial state are most often set to π . This is the most frequently applied MS model in economics and finance.

A class of finite mixture models is nested within the MS models by setting each of the rows of the transition matrix to π' , where all the elements of π are strictly positive. In this model, the forecasted state probabilities are time invariant and equal to $p(s_{t+1}|Y_{1:t}) = \pi$. However, the classification of observations is facilitated through the smoothed probabilities that change over time. Finite mixture models provide a convenient way of modeling nonstandard distributions that are often required for the error terms in economic and finance applications. It can be shown that any distribution of a random variable defined on a real scale can be approximated by a mixture of normal distributions, while distribution of a random variable defined on a positive real scale can be approximated by a mixture of gamma distributions (see [Norets, 2010](#)).

Change-point models can be used to introduce monotonic regime changes as in a model proposed by [Chib \(1998\)](#). The process is initiated in the first state $s_0 = 1$. With probability p_{11} it remains unchanged, and with probability p_{21} it switches to the other regime. The first state is never to be revisited and, thus, $p_{12} = 0$. In general, given that at some period t the Markov process is in state $s_t = k < K$, it remains in this state with probability p_{kk} and is only allowed to switch to the next regime with probability $p_{k+1,k}$. Finally, when the process reaches the K th state, it stays there forever. An example of a transition matrix for such a process with the number of states $K = 3$ is given by

$$\begin{bmatrix} p_{11} & p_{12} & 0 \\ 0 & p_{22} & p_{23} \\ 0 & 0 & 1 \end{bmatrix}.$$

Therefore, these models are capable of estimating the time at which the regime changes occur.

An interesting extension of this model was proposed by [Pesaran, Pettenuzzo, and Timmermann \(2006\)](#) which allows new regimes to occur in the forecasted sample. Finally, the change-point models introduce nonstationarity in the Markov-process and, thus, their ergodic probabilities are all equal to zero except for the last element of π that is equal to one. Therefore, the last state is the absorbing state that gains 100 percent of the probability mass asymptotically with $T \rightarrow \infty$. Finally, [Frühwirth-Schnatter \(2006\)](#) provided a detailed discussion of the nuances of the estimation of stationary and nonstationary Markov processes.

In simple and popular deterministic change-point models, s_t is assumed to be known and provided by the econometrician. In many applications, this model is used to estimate state-dependent parameters in samples predetermined by the investigator or using a test for multiple breaking points for stationary data (see e.g. [Bai and Perron, 1998](#)). Moreover,

it is straightforward to set s_t to obtain monotonic regime changes. Nevertheless, in the deterministic change-point models the transition matrix is redundant given that s_t is known. The regime-change dates are not estimated and, unless the econometrician knows the data generating process and sets s_t accordingly, the model fit deteriorates heavily compared to other models considered in this section.

A more elaborate form of the transition matrix may lead to the desired application-specific properties of the Markov process. Consider a model used by [Sims \(2001\)](#), who introduced symmetric jumping among adjacent regimes. In this example, the desired transition matrix for the case of $K = 4$ states is

$$\begin{bmatrix} p_{11} & (1 - p_{11}) & 0 & 0 \\ (1 - p_{22})/2 & p_{22} & (1 - p_{22})/2 & 0 \\ 0 & (1 - p_{33})/2 & p_{33} & (1 - p_{33})/2 \\ 0 & 0 & (1 - p_{44}) & p_{44} \end{bmatrix}.$$

[Sims, Waggoner, and Zha \(2008\)](#) consider a generalization of this model inspired by the developments in [Cogley and Sargent \(2005\)](#). This model uses a scarce parameterization of the transition matrix that is capable of capturing occasional discontinuous shifts in the values of the regime-dependent parameters when K is small, as well as frequent, incremental changes in these parameters for larger K . A general way of imposing restrictions on the transition matrix was proposed by [Sims et al. \(2008\)](#) and [Woźniak and Droumaguet \(2015\)](#).

Finally, the Markov property of the latent process might be extended by introducing the dependence of the current state, s_t^* , on its several recent realizations. For instance, the original model by [Hamilton \(1989\)](#) assumed the dependence of the model parameters on the current and previous regime, s_t^* and s_{t-1}^* respectively. This dependence can be modeled by a new four-state Markov process, s_t , through the following state representation:

$$\begin{aligned} s_t = 1 & \quad \text{if } s_t^* = 1 \text{ and } s_{t-1}^* = 1, \\ s_t = 2 & \quad \text{if } s_t^* = 2 \text{ and } s_{t-1}^* = 1, \\ s_t = 3 & \quad \text{if } s_t^* = 1 \text{ and } s_{t-1}^* = 2, \\ s_t = 4 & \quad \text{if } s_t^* = 2 \text{ and } s_{t-1}^* = 2, \end{aligned}$$

and an appropriate form of the transition matrix:

$$\begin{bmatrix} p_{11} & p_{12} & 0 & 0 \\ 0 & 0 & p_{21} & p_{22} \\ p_{11} & p_{12} & 0 & 0 \\ 0 & 0 & p_{21} & p_{22} \end{bmatrix}$$

Independent Markov Processes

In this class of MS models, various groups of parameters of the model depend on separate and independent Markov processes (see [Phillips, 1991](#); [Ravn and Sola, 1995](#), for some early

applications). Examples include models in which the parameters of the conditional mean process depend on a different Markov process than the conditional variances (see [Sims et al., 2008](#)), structural models in which the money demand equation depends on different Markov process than other parameters of the models (see [Sims and Zha, 2006](#)), and panel MS model that include a separate Markov process for parameters of equations for each individual (see [Kaufmann, 2010](#); [Billio, Casarin, Ravazzolo, and Van Dijk, 2016](#)). Consider L such independent processes s_{lt} each parameterized by a $K^l \times K^l$ transition matrix P^l , for $l = 1, \dots, L$. This model can be represented by a composite Markov process $s_t = (s_{1t}, \dots, s_{Lt})$ with $\prod_{l=1}^L K^l$ states and the corresponding transition matrix given by

$$P = P^1 \otimes \dots \otimes P^L, \quad (17)$$

where \otimes denotes the Kronecker product. The gain from the tensor product representation of the transition matrix in equation (17), introducing nonlinear restrictions, is an economic parameterization facilitating the estimation.

Nonhomogenous Markov Switching

Finally, this survey of parameterizations of transition matrices is concluded by the presentation of a nonhomogenous MS model in which the transition probabilities change over time (see [Filardo, 1994](#); [Diebold, Lee, and Weinbach, 1994](#)). The introduction of this time variation is often combined with the dependence on some variables v_t that might contain x_t (e.g. [Filardo, 1994](#)) as well as the state indicators s_t (see [Otranto, 2005](#)), or both (see [Billio et al., 2016](#)), as well as a measure of a duration of the state (see [Sichel, 1991](#); [Durland and McCurdy, 1994](#)). The restriction imposed on the rows of the transition matrix leads to the parameterization of the transition probabilities through the multinomial logistic regression as proposed in [Meligkotsidou and Dellaportas \(2011\)](#) in the following form:

$$p_{ij} \equiv P(s_t = j | s_{t-1} = i) = \frac{\exp(v_t \gamma_{ij})}{\sum_{j=1}^K \exp(v_t \gamma_{ij})},$$

where γ_{ij} are parameter vectors to be estimated the dimensions of which correspond to vector v_t . Note that for the identification of transition probabilities for each i there is a j so that γ_{ij} is a vector of zeros (see [Kaufmann, 2015](#); [Koki, Meligkotsidou, and Vrontos, 2020](#), for the detailed model specification and estimation procedures). The selection of the variables in v_t determines the time-dependence in transition probabilities and is subject to empirical verification. Finally, [Billio and Casarin \(2010\)](#) extended this specification by considering a two-state MS model ($K = 2$) and setting the transition probabilities p_{ij} to follow beta distribution with its parameters depending on v_t .

Infinite Hidden Markov Model

The infinite hidden Markov model (IHMM) was developed by [Beal, Ghahramani, and Rasmussen \(2002\)](#) and [Teh, Jordan, Beal, and Blei \(2006\)](#). It builds on the Dirichlet process

mixture (DPM) model of [Escobar and West \(1995\)](#) and extends the finite number of states of the MS model to the case in which this number goes to infinity, $K \rightarrow \infty$. Such an extension introduces a fundamental advancement of econometric modeling by transforming the parametric MS framework into a nonparametric structure.

A direct consequence is that the transition matrix P implied by equation (2) has an infinite dimension and can be presented as

$$P \equiv [p_{ij}] = \begin{bmatrix} p_{11} & p_{12} & p_{13} & \dots \\ p_{21} & p_{22} & p_{23} & \dots \\ p_{31} & p_{32} & p_{33} & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}, \quad (18)$$

where $i, j = 1, 2, 3, \dots$ and $p_{ij} \geq 0$. From the definition, each row of P must sum up to 1, $\sum_{j=1}^{\infty} p_{ij} = 1$. The time-invariant parameters that describe the k th state are defined as θ_k , and there is an infinite number of them. Similar to the finite-state MS models, the parameter space comprises the state indicator $S \equiv \{s_t\}_{t=1}^T$, the time-invariant parameters $\Theta \equiv \{\theta_k\}_{k=1}^{\infty}$ and the transition matrix $P = [p_{ij}]_{\infty \times \infty}$.

Estimation

Due to the parameter saturation problem, the IHMM cannot be estimated by classical methods without regularization. On the contrary, the Bayesian approach is coherent and more appropriate for inference and, thus, most research on the IHMM uses mostly this framework.

Three alternative approaches can be used to draw inference from the IHMM. The first is to integrate out the transition probability P based on the Chinese restaurant representation of the Dirichlet process as in [Fox, Sudderth, Jordan, and Willsky \(2011\)](#). This method works directly on the states, but its derivation is complicated. The second is to apply the beam sampler to truncate stochastically the number of states to a finite one during the MCMC as in [Van Gael, Saatchi, Teh, and Ghahramani \(2008\)](#). This method provides an exact inference similar to the first method. It also utilizes the conditional independence by keeping the transition matrix to a finite dimension, which allows partial parallel computations and is usually much faster than the first approach.

The last method applies the degree- K weak limit approximation from [Ishwaran and Zarepour \(2002\)](#) as in [Bauwens, Carpentier, and Dufays \(2017\)](#). It uses a truncated Dirichlet process so that the IHMM resembles an appropriate finite-state MS model. In practice, [Song \(2014\)](#) found that standard MS models with a large number of states performed similarly to the IHMM, where the number of inactive states should be nonzero and where an inactive state is a state with no data assigned to it. This approach renders the IHMM easier to execute, although two caveats exist. First, the prior distribution on the transition matrix must be chosen so that the concentration parameter from the truncation approximation is consistent with the concentration parameter from the IHMM (see [Ishwaran and Zarepour, 2002](#)). Otherwise, the approximation is not valid. Second,

the number of states in the MCMC must be monitored to avoid poor approximation. A simple rule would be that the number of active states—that is those with a nonzero number of observations classified into it—must always be less than the total number states K in the approximation (see the implementation by [Bauwens et al., 2017](#)).

Flexibility of the IHMM

The IHMM originated from the machine learning literature where it was used to reveal dynamic clustering in applications, including dialogue summary and motion capture. Subsequently, it has been applied to various fields in economics and finance. The seminal developments include [Song \(2014\)](#), [Jochmann \(2015\)](#) and [Dufays \(2015\)](#). The motivation for using the IHMM is that it demonstrates well the trade-off between heuristic economic interpretations and competitive forecast accuracy. This feature constitutes an advantage over many of the machine learning methods that hardly allow for structural interpretations.

Similar to the conventional MS models, the IHMM maintains its first-order Markov chain property. However, due to the differences in the prior distribution setup, and the problem of doubling the states, a phenomenon consisting of the possibility of producing an additional state that mimics an already existing one, the IHMM should not be considered a device for detecting the number of states of the mixture or MS models as was suggested in some early approaches (e.g. [Otranto and Gallo, 2002](#)). [Miller and Harrison \(2013\)](#) stated the argument formally.

An attractive feature of the IHMM is that it jointly captures regime-switching and structural break. Consequently, it grants more flexibility and accommodates data dynamics upon the arrival of new observations. The regime-switching module pools data with similar behavior to borrow statistical strength from each other. In addition, the structural break dynamics can generate a new state when the arriving new observations exhibit a new law of motion. A well-known example is the unprecedented global financial crisis in 2007 and 2008. It began with the subprime mortgage crisis in the United States and became the most severe financial crisis since 1930. Any standard MS model for bull and bear markets such as [Maheu and McCurdy \(2000\)](#), [Lunde and Timmermann \(2004\)](#) and [Maheu, McCurdy, and Song \(2012\)](#) is incapable of capturing the new phenomenon because they are limited by the data history and do not allow structural changes. The IHMM is an appropriate vehicle to achieve such a goal because the capability of generating a new unprecedented state is a feature of the latent process.

Another advantage of the IHMM lies in its superior forecasting performance, which is an empirical observation with plausible intuition. Benchmark models such as autoregressions or linear regression have a rigid assumption of the functional forms. Instead, the IHMM is more flexible and hence robust to model misspecification. Its flexibility comes from its ability to sort data into clusters endogenously, so that behaviorally different data will not affect the state inference. Two vital components to achieve improved forecasting performance are the hierarchical prior structure ([Song, 2014](#)) and certain regime persistence ([Fox et al., 2011](#)) in economic and finance applications. The hierarchical structure exploits more data information by letting the regimes inform one another. At

the same time, regime persistence reflects salient stylized facts that economic time series are prone to local dynamic stability.

The IHMM provides a convenient tool for a control approach in grand modeling frameworks. Consider a modeling framework that utilizes independent variables x_t and error term ϵ_t . Any incorrect distributional assumption about ϵ_t could potentially adversely affect the inference. If this distribution is not the focus of the application, simple estimators are applicable. However, if the distributional assumptions become essential as in the risk analysis, the semiparametric approach has the advantage of imposing a nonparametric distribution on ϵ_t . This assumption releases ϵ_t from any potential misspecification. The examples of such applications are provided by [Jensen and Maheu \(2010\)](#), who modeled ϵ_t using the DPM, as well as [Hou \(2017\)](#) and [Dufays, Zhuo, Rombouts, and Song \(2019\)](#) who applied the IHMM.

Extensions of the IHMM

The IHMM provides a basis for the burgeoning academic literature on its extensions. An IHMM with DPM emission was proposed in [Fox et al. \(2011\)](#) and further generalized to a block IHMM with more extensive in-state dynamics by [Stepleton, Ghahramani, Gordon, and Lee \(2009\)](#). In this approach, each state is a distinct MS model, which makes it suitable to the bull and bear market modeling as in [Maheu et al. \(2012\)](#). To capture long memory, [Gael, Teh, and Ghahramani \(2009\)](#) proposed the factorial IHMM.

Examples of applications include [Nakano, Le Roux, Kameoka, Nakamura, Ono, and Sagayama \(2011\)](#) and [Heller, Teh, and Gorur \(2009\)](#). The factorial IHMM does not impose any restrictions for identification and, thus, structural interpretations are hardly possible, which limits the scope of applications in economics and finance. Another modeling strategy consists of applying the IHMM structure to existing interpretable parametric models can be found in [Liu and Maheu \(2018\)](#) and [Jin, Maheu, and Yang \(2019\)](#). Finally, there is a substantial body of work that extends the IHMM in various directions as in [Shi and Song \(2014\)](#), [Bauwens et al. \(2017\)](#), [Maheu and Yang \(2016\)](#), [Yang \(2019\)](#), [Hou \(2017\)](#) and [Luo, Klein, Ji, and Hou \(2019\)](#).

Endogenous Markov Switching

The endogenous MS model questions the assumption of exogeneity of the Markov process and makes an explicit link between the measurement equation error term and this latent process. In consequence, some form of endogeneity of the Markov process is introduced. [Kim, Piger, and Startz \(2008\)](#) argue that in many applications, endogeneity of the Markov process corresponds to the data properties and theoretical considerations to a larger extent than the exogenous one. While the seminal proposal by [Kim et al. \(2008\)](#) implemented endogeneity in the MS model, this section focuses on a specification proposed by [Chang, Choi, and Park \(2017\)](#). In this model, a discrete-valued latent process, s_t , is driven by a real-valued process, w_t , that is related to the measurement equation error term and then conveniently discretized.

Introducing Endogeneity

Define the dynamics of the real-valued latent factor by an autoregressive process

$$w_t = \alpha w_{t-1} + v_t, \quad (19)$$

where α is the persistence coefficient such that $|\alpha| \leq 1$ and v_t is a standard normal error term. The initial value of the process, w_0 , is recommended to be normally distributed with the zero mean and variance equal to $1/(1 - \alpha^2)$ for $|\alpha| < 1$, or equal to zero, $w_0 = 0$, for $\alpha = 1$. The process in equation (19) is discretized by defining a threshold parameter, τ , and the discrete-valued state indicator for a two-state model, $K = 2$, as

$$s_t = \begin{cases} 1 & \text{for } w_t < \tau, \\ 2 & \text{for } w_t \geq \tau. \end{cases} \quad (20)$$

Therefore, as long as the process w_t is subject to interpretation, its primary role is to define the Markov process s_t .

Chang et al. (2017) define the measurement equation in a general form that made an explicit link between the potential dependence of the conditional mean and standard deviation on the independent variables, x_t , and the latent processes

$$y_t = m(x_t, w_t) + \sigma(x_t, w_t)\epsilon_t = m(x_t, s_t) + \sigma(x_t, s_t)\epsilon_t, \quad (21)$$

where ϵ_t , conditionally on x_t and w_t (or s_t) is a serially uncorrelated standard normal error term. Endogeneity of the Markov process is formally introduced in this model by an appropriate specification of the joint distribution of the error terms from the state and measurement equations (19) and (21) respectively that is given by

$$\begin{bmatrix} \epsilon_t \\ v_{t+1} \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right), \quad (22)$$

where ρ is the correlation coefficient. It can be shown that the exogenous MS model can be obtained by setting $\rho = 0$. In other words, the restricted endogenous and the exogenous MS models are observationally equivalent—that is, they both lead to the same value of the likelihood function given the values of Θ . However, for the values of ρ that are different from zero the relationship between ϵ_t and w_{t+1} and, consequently, between ϵ_t and s_{t+1} is established. Expressions for the implied transition probabilities that are changing over time are given in the original paper.

A different form of endogeneity was proposed by Kim et al. (2008) who specified the joint distribution of error terms, such as the one in equation (22), for vector (ϵ_t, v_t) . This model presumes a contemporaneous effect of ϵ_t on w_t and s_t which Chang and Kwak (2017) point out to be misspecified. Kim et al. (2008) consider an application to modeling a volatility feedback effect in financial time series.

Interpretations Considering Correlation Coefficient

To illustrate the interpretation of nonzero ρ consider a model in equation (6) with s_t specified by the endogenous Markov process and with $|\beta_k| < 1$ for $k = 1, 2$. A possible application in finance includes the modeling of the leverage effect defined as the negative correlation between the current innovation and future conditional variance of the return on a financial asset. Let $\rho < 0$ and $\sigma_1 < \sigma_2$. In this case, a negative realization of ϵ_t increases the probability of the second state in period $t + 1$ and leads to an increase in the conditional volatility.

A simple application in time series analysis includes the modeling of the mean reversion that works differently for ρ of different signs. Let $\mu_1 < \mu_2$ and $\rho < 0$. Then, a positive realization of ϵ_t decreases the probability of the second state in the period $t + 1$. Therefore, the mean reversion is also obtained at the level of the future conditional expected value that now decreases. In the opposite case of $\rho > 0$, a positive realization of ϵ_t increases the probability of the second state in period $t + 1$ and, consequently, increases the conditional expected value of y_t . Therefore, the forecasts of y revert to a mean that is higher, which has a destabilizing effect. Examples of more elaborate applications in economics include the endogenous switching of the parameters modeling the effects of monetary and fiscal policies proposed by [Chang, Maih, and Tan \(2018\)](#) and [Chang and Kwak \(2017\)](#), respectively.

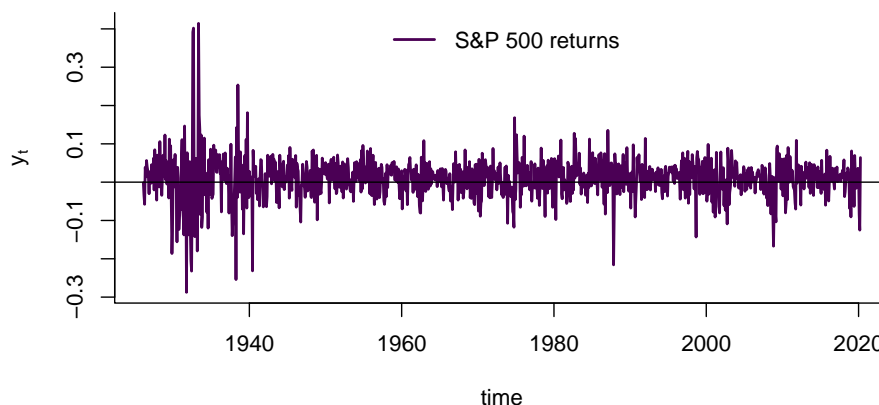
Future Developments

Two suggested paths for future developments in the MS and IHMM frameworks include efficient algorithms with the potential for parallel computations and interpretability incurred through sparsity understood as an automated way of reducing the complexity of a general model. Improvements on both of these fronts are required to make the analysis of big data sets possible by combining sufficient flexibility of the model with scarce parameterization and feasible computations.

Existing approaches to MS models rely on the FFBS technique. This filtering and smoothing method consists of an iterative procedure that implies serial computations. With an increasing number of observations and states required to capture the features of data, the FFBS algorithm becomes computationally unfeasible for practical implementations. Vectorization, tensor algebra, and efficient numerical algorithms provide some of the solutions. Still, they are far from being as computationally fast as available algorithms for real-valued state-space models such as the precision sampler by [Chan and Jeliaskov \(2009\)](#).

Moreover, an increasing interest in heterogeneous MS models in which individual parameters follow independent Markov processes calls for new methods of allowing sparsity in the modeling. In many such studies, the question of whether time variation is required for a particular parameter, and if so then how many MS regimes are required to model, remain unanswered due to the lack of computationally fast algorithms. It is worth mentioning that there are solutions for real-valued state-space models, for

Figure 1: Monthly logarithmic rates of return for the S&P 500 Index for the period January 1926 – March 2020



example, [Frühwirth-Schnatter and Wagner \(2010\)](#), [Bitto and Frühwirth-Schnatter \(2019\)](#), and [Cadonna, Frühwirth-Schnatter, and Knaus \(2020\)](#).

Similar considerations apply to the IHMM, although it should be emphasized that the IHMM provides certain solutions to the challenges singled out in this article for the MS models. Here, as the number of observations increases, the number of states, K , to be modeled in a particular iteration of the estimation algorithm following the beam sampler step may increase as well. This fact increases the computation time geometrically and calls for a more time-efficient estimation method. Variational inference offers faster algorithms at the cost of approximating the posterior distribution at an unknown precision (see [Wainwright and Jordan, 2008](#), for further reference). Variational Bayes correctly captures the central tendency of the approximated distribution. However, it underestimates the posterior variances (see [Wang and Blei, 2019](#), and references therein). Applications of Variational Bayes estimation to Dirichlet process and IHMM can be found in [Blei and Jordan \(2006\)](#), [Kurihara, Welling, and Teh \(2007\)](#), [Teh, Kurihara, and Welling \(2008\)](#), and [Wang, Paisley, and Blei \(2011\)](#). Alternative approaches may seek to improve the computation through parallelization such as [Fearnhead \(2004\)](#), [Rodriguez \(2011\)](#), [Williamson, Dubey, and Xing \(2013\)](#) and [Tripuraneni, Gu, Ge, and Ghahramani \(2015\)](#).

Notable developments granting sparsity in the DPM have been proposed by [Frühwirth-Schnatter and Malsiner-Walli \(2019\)](#). In this approach, the sparsity is accompanied by the choice of concentration parameters and the computations are significantly simplified as the sparse structure heavily penalizes the number of states.

Three Models for Financial Asset Returns

The working and interpretation of the MS models are illustrated by applying three models to the analysis of 1131 monthly logarithmic rates of return of the S&P 500 index that are plotted in [Figure 1](#) for a sample spanning the period starting in January 1926 and finishing

Table 1: Estimation Results for the MS and Endogenous MS Models

θ	$E[\theta Y]$	$sd[\theta Y]$	quantiles		$E[\theta Y]$	$sd[\theta Y]$	quantiles	
			5%	95%			5%	95%
μ_1	-0.012	0.011	-0.029	0.005	-0.014	0.01	-0.032	0.002
μ_2	0.012	0.001	0.009	0.014	0.011	0.00004	0.009	0.013
σ_1^2	0.115	0.008	0.102	0.129	0.014	0.002	0.011	0.018
σ_2^2	0.038	0.001	0.037	0.04	0.0014	0.00008	0.001	0.001
p_{11}	0.909	0.036	0.841	0.959	0.914*	0.049*		
p_{22}	0.987	0.005	0.977	0.994	0.953*	0.039*		
π_1	0.134	0.051	0.068	0.223	0.364*	0.157*		
π_2	0.866	0.051	0.777	0.932	0.636*	0.157*		
ρ	0				-0.371	0.104	-0.534	-0.199
τ					1.978	0.692	0.87	3.021

Note: The reported values include posterior means and standard deviations, as well as the 5th and 95th percentiles of the marginal posterior distribution of the parameters.

* Denotes values computed as characteristics of the Markov chain that are not parameters of the endogenous Markov switching model.

in March 2020. All three models are given by the same predictive density

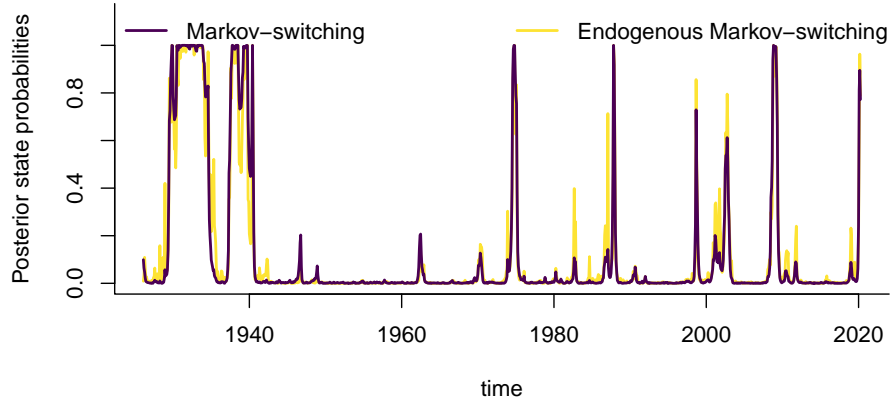
$$y_t|s_t \sim \mathcal{N}(\mu_{s_t}, \sigma_{s_t}^2) \quad (23)$$

The specification of the Markov process s_t differs in the considered models. A two-state Exogenous Markov-switching model, referred to via MS, has its transition matrix defined by equations (2) and (1). The Markov process in a two-state Endogenous MS model is given by a real-valued latent process from equations (20) and (19). Finally, in the IHMM model the number of states is unrestricted and the transition matrix is given by equation (18).

Table 1 reports the estimation results for the MS and Endogenous MS models. In both of the models State 1 is characterized by a negative average return that is equal to -0.012 and -0.014 respectively, and a higher value of the conditional variance than in State 2. The conditional variances in State 1 are equal to 0.115 for the MS and 0.014 for the Endogenous MS models, while in State 2 they take the respective values of 0.038 and 0.0014. The conditional means in State 1 cannot be considered statistically less than zero. However, they should be considered statistically different than the conditional means in State 2 based on the observation that the 95% posterior quantiles of the means in State 1 are smaller than the 5% posterior quantiles of the means in State 2. The opposite ordering is observed for the conditional variances that are larger in State 1 than in State 2.

The interpretation of the parameters, when juxtaposed with the posterior estimates

Figure 2: Posterior probabilities of State 1 for the MS and Endogenous MS models



of the state probabilities across the sample for these two models, results in a clear interpretation of the states. Figure 2 reports these state probabilities for State 1. The corresponding probabilities for State 2 can be computed easily by subtracting the probabilities from the figure from value 1. For both of the models the probabilities of the occurrence of State 1 at each of the sample periods are alike. They have high values for the persistent periods from June 1929 to July 1934, May 1937 – November 1939, from April to June 1940, July 1974 – January 1975, August 2008 – April 2009, and finally, from January to March 2020, as well as for shorter occurrences in late 1987, August 1998, and mid 2002. Each of these periods can be associated with one of the historical economic recessions or downturns in the financial markets. As the average returns in State 1 are negative and the volatility is relatively higher, this state is given the interpretation of the *bear market regime*. On the contrary, since the occurrences of State 2 coincide with calmer periods on financial markets, and since the average return for this regime is positive and the volatility is relatively low, State 2 is assigned a label of the *bull market regime*.

Table 1 reports the estimates of the transition matrices elements and the implied ergodic state probabilities. These values are estimated as parameters for the case of the MS model, whereas they were computed as summary characteristics of the Markov process of the Endogenous MS model in which the transition matrix is a part of the parametric model. The bull market is more persistent than the bear market as the probabilities of remaining in a state for another period are higher for State 2 than for State 1. The implied expected state durations of the bull market is 77 months for the MS model and 22 months for the Endogenous MS model, and the expected duration of the bear market are around 11 months for both of the models. Finally, the MS model estimates that the bull market dominates for over 86% of the sample period based on the ergodic probability estimate for State 2, whereas the same estimate for the Endogenous MS model is equal to around 64%.

Figure 3 presents the posterior means for the real-valued time-varying latent process w_t and the posterior mean of the threshold parameter τ driving the Markov process of

Figure 3: Posterior means of the latent process w_t and the threshold parameter τ for the Endogenous MS models

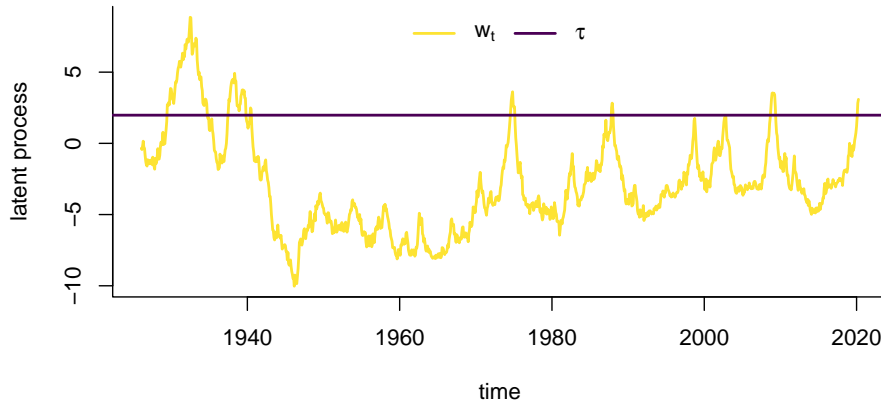
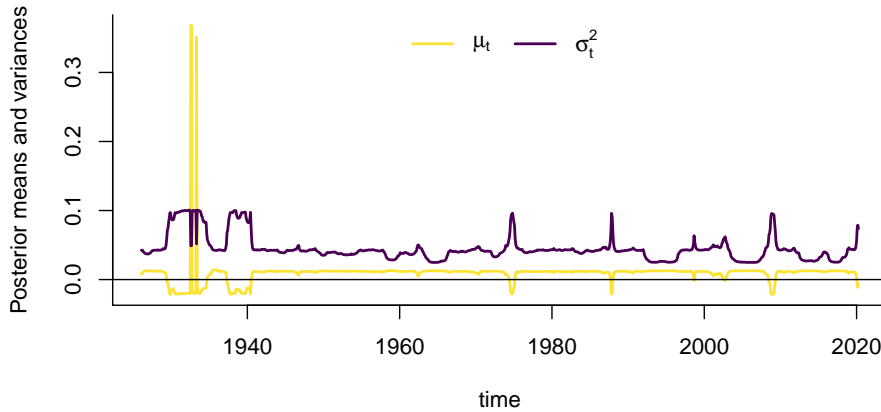


Figure 4: Posterior estimates of time-varying means and variances of the returns for the IHMM

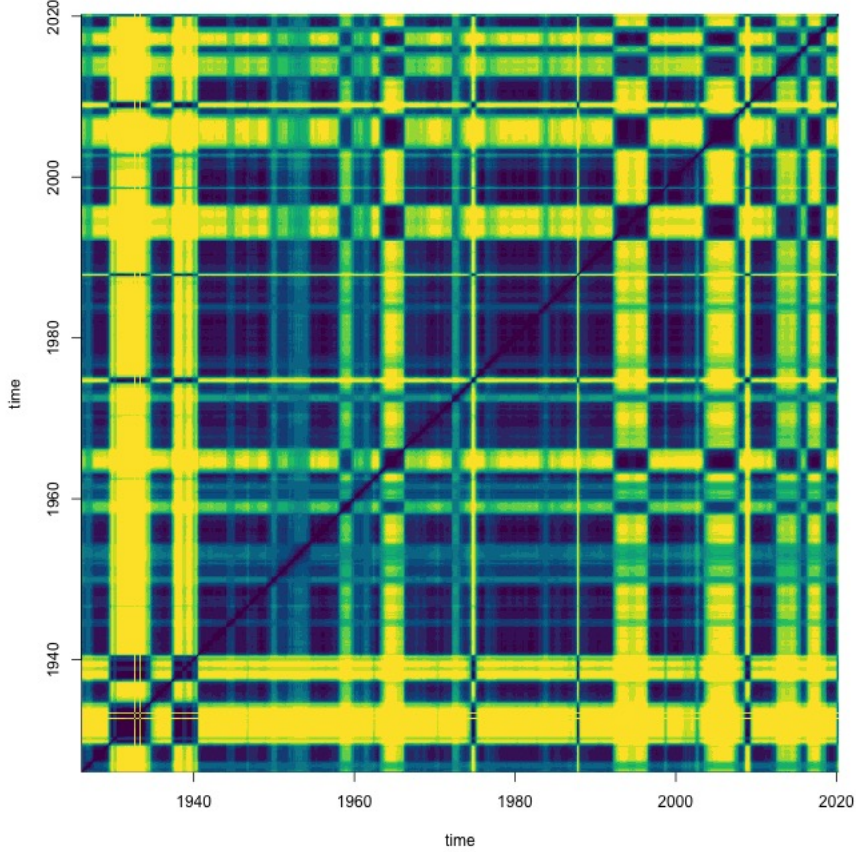


the Endogenous MS model. In this figure, the periods in which the value of the latent process is greater than τ coincide with the periods of high probability of occurrence of State 1 reported in Figure 2.

The presentation of the results for a nonparametric IHMM requires extra efforts as the number of the states is estimated and the states themselves cannot be assigned clear interpretations. However, some additional computations reveal that the characterization of the data in terms of bull and bear markets is also admissible for this model.

Figure 4 presents the posterior means of the time-varying mean and variance parameters. They are computed by an appropriate weighting of the corresponding values of the state-dependent parameters μ_{s_t} and $\sigma_{s_t}^2$ by the estimated probability of the state at period t . In this figure, the periods in which the values of the conditional mean are less than zero and those of the conditional variances are relatively high coincide with the periods of high probability of occurrence of State 1 in Figure 2. Accordingly, the periods of positive

Figure 5: Heat map of IHMM states



Note: The heat map reports probabilities of the event that $s_{t_0} = s_{t_1}$ for all pairs (t_0, t_1) such that $t_0, t_1 = 1, \dots, T$. Blue color indicates high while yellow indicates low values of these probabilities.

conditional mean and relatively low conditional variances coincide with the periods of low probability of occurrence of State 1 of the MS and Endogenous MS models. Despite these similarities, the IHMM offers a richer picture and captures some extraordinary periods such as the two positive outlying observations in the early 1930's.

Individual states of the IHMM are not identified, and therefore are not interpretable, due to the label switching problem. However, the states can be characterized by label invariant characteristics such as the probability of two states being the same in different periods. Each pair of such probabilities, denoted by $p(s_{t_0} = s_{t_1} | Y)$, is estimated by the fraction of posterior draws for which $s_{t_0} = s_{t_1}$ for $t_0, t_1 = 1, \dots, T$. These values form a $T \times T$ matrix that is presented in Figure 5 as a heat map. The diagonal values are equal to 1 by construction as the states must coincide for a given period.

The very top row of the heat map provides the probabilities that the historical states

Table 2: Posterior Probabilities of the Number of States of the IHMM

k	3	4	5	6	7
$Pr[K = k Y]$	0.045	0.711	0.224	0.019	0.001

Note: The probabilities for other values of k are numerically equal to zero in our simulation based on 1000 draws from the posterior distribution.

are the same as the state in March 2020 which is characterized by a negative conditional return and relatively high value of the conditional variance associated with the arrival of early news regarding the spread of coronavirus in the United States. The periods that have high probability of being in the same state are a half-year period following the global financial crisis, a month of the Asian financial crisis of the late 1990s, short periods in mid-1980s and mid-1970s, and a prolonged period throughout the 1930s.

Finally, Table 2 reports the estimated probabilities of the number of states of the IHMM. The posterior probability mass is concentrated for the values of K from 3 to 7, with nearly 94% of the posterior probability assigned jointly to values 4 and 5.

In conclusion, all the three models considered in our empirical illustration, despite exhibiting alternative specifications of the Markov process, capture similar properties of the low-frequency financial log-returns and offer similar classifications of the sample into persistent bull and bear markets.

Acknowledgments

The authors thank the editor of the series, Professor Anindya Banerjee, two anonymous referees for their comments and suggestions that improved the quality of this paper. They also thank the editorial team of the Oxford Research Encyclopedia of Economics and Finance. They are grateful to Roberto Casarin, Bill Griffiths, Chenghan Hou, Zhuo Li, Vance Martin, and Qiao Yang for their useful discussions.

References

- Augustyniak, M., L. Bauwens, and A. Dufays (2019). A New Approach to Volatility Modeling: The Factorial Hidden Markov Volatility Model. *Journal of Business and Economic Statistics* 37(4), 696–709.
- Bai, J. and P. Perron (1998). Estimating and Testing Linear Models with Multiple Structural Changes. *Econometrica* 66(1), 47–78.
- Bauwens, L., J.-F. Carpentier, and A. Dufays (2017). Autoregressive moving average infinite hidden markov-switching models. *Journal of Business & Economic Statistics* 35(2), 162–182.
- Bauwens, L., A. Dufays, and J. V. Rombouts (2014). Marginal likelihood for Markov-switching and change-point GARCH models. *Journal of Econometrics* 178, 508–522.
- Beal, M. J., Z. Ghahramani, and C. E. Rasmussen (2002). The infinite hidden markov model. In *Advances in neural information processing systems*, pp. 577–584.
- Billio, M. and R. Casarin (2010). Identifying Business Cycle Turning Points with Sequential Monte Carlo Methods: An Online and Real-Time Application to the Euro Area. *Journal of Forecasting* 29, 145–167.

- Billio, M., R. Casarin, F. Ravazzolo, and H. K. Van Dijk (2016). Interconnections Between Eurozone and US Booms and Busts Using a Bayesian Panel Markov-Switching VAR Model. *Journal of Applied Econometrics* 31(7), 1352–1370.
- Billio, M. and A. Monfort (1998). Switching state-space models likelihood function, filtering and smoothing. *Journal of Statistical Planning and Inference* 68(1), 65–103.
- Billio, M., A. Monfort, and C. P. Robert (1999). Bayesian estimation of switching ARMA models. *Journal of Econometrics* 93(2), 229–255.
- Bitto, A. and S. Frühwirth-Schnatter (2019). Achieving shrinkage in a time-varying parameter model framework. *Journal of Econometrics* 210(1), 75–97.
- Blei, D. M. and M. I. Jordan (2006). Variational inference for dirichlet process mixtures. *Bayesian analysis* 1(1), 121–143.
- Cadonna, A., S. Frühwirth-Schnatter, and P. Knaus (2020). Triple the gamma—a unifying shrinkage prior for variance and variable selection in sparse state space and TVP models. *Econometrics* 8, 1–36.
- Carrasco, M., L. Hu, and W. Ploberger (2014). Optimal Test for Markov Switching Parameters. *Econometrica* 82(2), 765–784.
- Carvalho, C. M. and H. F. Lopes (2007). Simulation-based sequential analysis of Markov switching stochastic volatility models. *Computational Statistics and Data Analysis* 51(9), 4526–4542.
- Casarin, R., D. Sartore, and M. Tronzano (2018). A Bayesian Markov-Switching Correlation Model for Contagion Analysis on Exchange Rate Markets. *Journal of Business and Economic Statistics* 36(1), 101–114.
- Chan, J. and I. Jeliazkov (2009). Efficient Simulation and Integrated Likelihood Estimation in State Space Models. *International Journal of Mathematical Modelling and Numerical Optimisation* 1, 101–120.
- Chang, Y., Y. Choi, and J. Y. Park (2017). A new approach to model regime switching. *Journal of Econometrics* 196(1), 127–143.
- Chang, Y. and B. Kwak (2017). U.S. monetary-fiscal regime changes in the presence of endogenous feedback in policy rules. *Halle Institute for Economic Research Discussion Papers*.
- Chang, Y., J. Maih, and F. Tan (2018). State Space Models with Endogenous Regime Switching. *Norges Bank Working Papers*.
- Chib, S. (1996). Calculating posterior distributions and modal estimates in markov mixture models. *Journal of Econometrics* 75(1), 79–97.
- Chib, S. (1998). Estimation and comparison of multiple change-point models. *Journal of Econometrics* 86(2), 221–241.
- Cogley, T. and T. J. Sargent (2005). Drifts and volatilities: Monetary policies and outcomes in the post WWII US. *Review of Economic Dynamics* 8, 262–302.
- Diebold, F. X., J.-H. Lee, and G. C. Weinbach (1994). Regime switching with time-varying transition probabilities. In C. Hargreaves (Ed.), *Nonstationary Time Series Analysis and Cointegration*, pp. 283–302. Oxford: Oxford University Press.
- Droumaguet, M., A. Warne, and T. Woźniak (2017). Granger Causality and Regime Inference in Markov Switching VAR Models with Bayesian Methods. *Journal of Applied Econometrics* 32, 802–818.
- Dufays, A. (2015). Infinite-state markov-switching for dynamic volatility. *Jnl of Financial Econometrics* 14(2), 418–460.
- Dufays, A., L. Zhuo, J. Rombouts, and Y. Song (2019). Sparse change-point var models. *University of Melbourne Working Papers*.
- Durland, J. M. and T. H. McCurdy (1994). Duration-dependent transitions in a Markov model of U.S. GNP growth. *Journal of Business and Economic Statistics* 12(3), 279–288.
- Escobar, M. D. and M. West (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association* 90(430), 577–588.
- Farmer, R. E., D. F. Waggoner, and T. Zha (2009). Understanding Markov-switching rational expectations models. *Journal of Economic Theory* 144(5), 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(12), 2150–2166.
- Fearnhead, P. (2004). Particle filters for mixture models with an unknown number of components. *Statistics*

- and Computing 14(1), 11–21.
- Filardo, A. J. (1994). Business-Cycle Phases and Their Transitional Dynamics. *Journal of Business & Economic Statistics* 12(3), 299–308.
- Fox, E. B., E. B. Sudderth, M. I. Jordan, and A. S. Willsky (2011). A sticky hdp-hmm with application to speaker diarization. *The Annals of Applied Statistics* 5(2A), 1020–1056.
- Frühwirth-Schnatter, S. (2001). Markov Chain Monte Carlo Estimation of Classical and Dynamic Switching and Mixture Models. *Journal of the American Statistical Association* 96(453), 194–209.
- Frühwirth-Schnatter, S. (2004). Estimating marginal likelihoods for mixture and Markov switching models using bridge sampling techniques. *Econometrics Journal* 7(1), 143–167.
- Frühwirth-Schnatter, S. (2006). *Finite mixture and Markov switching models*. Springer Science & Business Media.
- Frühwirth-Schnatter, S. and G. Malsiner-Walli (2019). From here to infinity: sparse finite versus dirichlet process mixtures in model-based clustering. *Advances in data analysis and classification* 13(1), 33–64.
- Frühwirth-Schnatter, S. and H. Wagner (2010). Stochastic model specification search for Gaussian and partial non-Gaussian state space models. *Journal of Econometrics* 154(1), 85–100.
- Gael, J. V., Y. W. Teh, and Z. Ghahramani (2009). The infinite factorial hidden markov model. In *Advances in Neural Information Processing Systems*, pp. 1697–1704.
- Gallo, G. M. and E. Otranto (2008). Volatility spillovers , interdependence and comovements : A Markov Switching approach. *Computational Statistics & Data Analysis* 52, 3011 – 3026.
- Geweke, J. (2007). Interpretation and inference in mixture models: Simple MCMC works. *Computational Statistics & Data Analysis* 51(7), 3529–3550.
- Goldfeld, S. M. and R. E. Quandt (1973). A Markov Model for Switching Regressions. *Journal of Econometrics* 1(1), 3–16.
- Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica* 37(3), 424–438.
- Haas, M., S. Mittnik, and M. S. Paolella (2004). A new approach to markov-switching garch models. *Journal of Financial Econometrics* 2(4), 493–530.
- Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica: Journal of the Econometric Society*, 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, Princeton, NJ.
- Hamilton, J. D. and G. Lin (1996). Stock Market Volatility and the Business Cycle. *Journal of Applied Econometrics* 11(5), 573–593.
- Hamilton, J. D. and M. T. Owyang (2012). The Propagation of Regional Recessions. *Review of Economics and Statistics* 94(4), 935–47.
- Hamilton, J. D. and R. Susmel (1994). Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics* 64, 307–333.
- Heller, K., Y. W. Teh, and D. Gorur (2009). Infinite hierarchical hidden markov models. In *Artificial Intelligence and Statistics*, pp. 224–231.
- Hou, C. (2017). Infinite hidden markov switching vars with application to macroeconomic forecast. *International Journal of Forecasting* 33(4), 1025–1043.
- Ishwaran, H. and M. Zarepour (2002). Exact and approximate sum representations for the Dirichlet process. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique* 30(2), 269–283.
- Jensen, M. J. and J. M. Maheu (2010). Bayesian semiparametric stochastic volatility modeling. *Journal of Econometrics* 157(2), 306–316.
- Jin, X., J. M. Maheu, and Q. Yang (2019). Bayesian parametric and semiparametric factor models for large realized covariance matrices. *Journal of Applied Econometrics* 34(5), 641–660.
- Jochmann, M. (2015). Modeling us inflation dynamics: A bayesian nonparametric approach. *Econometric Reviews* 34(5), 537–558.
- 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(2), 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(1), 82–94.
- Kaufmann, S. and S. Frühwirth-Schnatter (2002). Bayesian analysis of switching ARCH models. *Journal of Time Series Analysis* 23(4), 425–458.
- Kim, C.-J. and C. R. Nelson (1999). State-space models with regime switching: classical and gibbs-sampling approaches with applications. *MIT Press Books* 1.
- Kim, C. J., J. Piger, and R. Startz (2008). Estimation of Markov regime-switching regression models with endogenous switching. *Journal of Econometrics* 143(2), 263–273.
- Koki, C., L. Meligkotsidou, and I. D. Vrontos (2020). Forecasting under model uncertainty : Non-homogeneous hidden Markov models with Pölya-Gamma data augmentation. *Journal of Forecasting*, 1–19.
- Krolzig, H.-M. (1997). *Markov-Switching Vector Autoregressions. Modelling, Statistical Inference, and Application to Business Cycle Analysis*. Springer.
- Kurihara, K., M. Welling, and Y. W. Teh (2007). Collapsed variational dirichlet process mixture models. In *IJCAI*, Volume 7, pp. 2796–2801.
- Lanne, M., H. Lütkepohl, and K. Maciejowska (2010). Structural vector autoregressions with Markov switching. *Journal of Economic Dynamics and Control* 34(2), 121–131.
- Lehmann, E. L. (1983). *Theory of Point Estimation*. Wiley, New York.
- Leiva-Leon, D. (2017). Measuring Business Cycles Intra-Synchronization in US: A Regime-switching Interdependence Framework. *Oxford Bulletin of Economics and Statistics* 79(4), 513–545.
- Lindgren, G. (1978). Markov Regime Models for Mixed Distributions and Switching Regressions. *Scandinavian Journal of Statistics* 5(2), 81–91.
- Liu, J. and J. M. Maheu (2018). Improving markov switching models using realized variance. *Journal of Applied Econometrics* 33(3), 297–318.
- Liu, Z., D. F. Waggoner, and T. Zha (2011). Sources of macroeconomic fluctuations: A regime-switching DSGE approach. 2(2), 251–301.
- Lunde, A. and A. Timmermann (2004). Duration dependence in stock prices: An analysis of bull and bear markets. *Journal of Business & Economic Statistics* 22(3), 253–273.
- Luo, J., T. Klein, Q. Ji, and C. Hou (2019). Forecasting realized volatility of agricultural commodity futures with infinite hidden markov har models. *International Journal of Forecasting*, forthcoming.
- Lütkepohl, H. and T. Woźniak (2020). Bayesian Inference for Structural Vector Autoregressions Identified by Markov-Switching Heteroskedasticity. *Journal of Economic Dynamics and Control* 113.
- Maheu, J. M. and T. H. McCurdy (2000). Identifying bull and bear markets in stock returns. *Journal of Business & Economic Statistics* 18(1), 100–112.
- Maheu, J. M., T. H. McCurdy, and Y. Song (2012). Components of bull and bear markets: bull corrections and bear rallies. *Journal of Business & Economic Statistics* 30(3), 391–403.
- Maheu, J. M. and Q. Yang (2016). An infinite hidden markov model for short-term interest rates. *Journal of Empirical Finance* 38, 202–220.
- Meitz, M. and P. Saikkonen (2017). Testing for observation-dependent regime switching in mixture autoregressive models. *HECER Discussion Paper* (420), 1–56.
- Meligkotsidou, L. and P. Dellaportas (2011). Forecasting with non-homogenous hidden Markov models. *Statistics and Computing* 21, 439–449.
- Miller, J. W. and M. T. Harrison (2013). A simple example of dirichlet process mixture inconsistency for the number of components. In *Advances in neural information processing systems*, pp. 199–206.
- Nakano, M., J. Le Roux, H. Kameoka, T. Nakamura, N. Ono, and S. Sagayama (2011). Bayesian nonparametric spectrogram modeling based on infinite factorial infinite hidden markov model. In *2011 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 325–328. IEEE.
- Norets, A. (2010). Approximation of conditional densities by smooth mixtures of regressions. *Annals of*

- Statistics* 38(3), 1733–1766.
- Otranto, E. (2005). The multi-chain Markov switching model. *Journal of Forecasting* 24, 523–537.
- Otranto, E. and G. M. Gallo (2002). A Nonparametric Bayesian Approach To Detect the Number of Regimes in Markov Switching Models. *Econometric Reviews* 21(4), 477–496.
- Owyang, M. T., J. Piger, and H. J. Wall (2005). Business Cycle Phases in U.S. States. *Review of Economics and Statistics* 87(4), 604–616.
- Pesaran, M. H., D. Pettenuzzo, and A. Timmermann (2006). Forecasting time series subject to multiple structural breaks. *Review of Economic Studies* 73, 1057–1084.
- 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 money-output relationship. *Journal of Applied Econometrics* 20(5), 665–683.
- Psaradakis, Z. and M. Sola (1998). Finite-sample properties of the maximum likelihood estimator in autoregressive models with Markov switching. *Journal of Econometrics* 86(2), 369–386.
- Psaradakis, Z. and N. Spagnolo (2003). On the determination of the number of regimes in Markov-switching autoregressive models. *Journal of Time Series Analysis* 24(2), 237–252.
- Ravn, M. O. and M. Sola (1995). Stylized facts and regime changes: Are prices procyclical? *Journal of Monetary Economics* 36(3), 497–526.
- Rodriguez, A. (2011). On-line learning for the infinite hidden markov model. *Communications in Statistics-Simulation and Computation* 40(6), 879–893.
- Shi, S. and Y. Song (2014). Identifying speculative bubbles using an infinite hidden markov model. *Journal of Financial Econometrics* 14(1), 159–184.
- Sichel, D. E. (1991). Business cycle duration dependence: a parametric approach. *Review of Economics and Statistics* 73, 254–256.
- Sims, C. A. (1980). Macroeconomics and Reality. *Econometrica* 48(1), 1–48.
- Sims, C. A. (2001). Stability and instability in us monetary policy behavior. *Princeton University Working Papers*.
- Sims, C. A., D. F. Waggoner, and T. Zha (2008). Methods for inference in large multiple-equation Markov-switching models. *Journal of Econometrics* 146(2), 255–274.
- Sims, C. A. and T. Zha (2006). Were There Regime Switches in U.S. Monetary Policy? *American Economic Review* 96(1), 54–81.
- So, M. E. P., K. Lam, and W. K. Li (1998). A stochastic volatility model with markov switching. *Journal of Business and Economic Statistics* 16(2), 244–253.
- Song, Y. (2014). Modelling regime switching and structural breaks with an infinite hidden markov model. *Journal of Applied Econometrics* 29(5), 825–842.
- Stepleton, T., Z. Ghahramani, G. Gordon, and T.-S. Lee (2009). The block diagonal infinite hidden markov model. In *Artificial Intelligence and Statistics*, pp. 552–559.
- Teh, Y. W., M. I. Jordan, M. J. Beal, and D. M. Blei (2006). Hierarchical dirichlet processes. *Journal of the American Statistical Association* 101, 1566–1581.
- Teh, Y. W., K. Kurihara, and M. Welling (2008). Collapsed variational inference for hdp. In *Advances in neural information processing systems*, pp. 1481–1488.
- Teräsvirta, T. (2006). Forecasting economic variables with nonlinear models. In G. Elliott, C. W. Granger, and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Volume 1, Chapter 8, pp. 413–457. Elsevier B.V.
- Tripuraneni, N., S. S. Gu, H. Ge, and Z. Ghahramani (2015). Particle gibbs for infinite hidden markov models. In *Advances in Neural Information Processing Systems*, pp. 2395–2403.
- Van Gael, J., Y. Saatchi, Y. W. Teh, and Z. Ghahramani (2008). Beam sampling for the infinite hidden markov model. In *Proceedings of the 25th international conference on Machine learning*, pp. 1088–1095. ACM.
- Waggoner, D. F. and T. Zha (2012). Confronting model misspecification in macroeconomics. *Journal of Econometrics* 171(2), 167–184.
- Wainwright, M. J. and M. I. Jordan (2008). Graphical models, exponential families, and variational inference.

Foundations and Trends® in Machine Learning 1(1–2), 1–305.

- Wang, C., J. Paisley, and D. Blei (2011). Online variational inference for the hierarchical dirichlet process. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pp. 752–760.
- Wang, Y. and D. M. Blei (2019). Frequentist Consistency of Variational Bayes. *Journal of the American Statistical Association* 114(527), 1147–1161.
- Warne, A. (2000). Causality and Regime Inference in a Markov Switching VAR.
- Williamson, S., A. Dubey, and E. Xing (2013). Parallel markov chain monte carlo for nonparametric mixture models. In *International Conference on Machine Learning*, pp. 98–106.
- Woźniak, T. and M. Droumaguet (2015). Assessing Monetary Policy Models : Bayesian Inference for Heteroskedastic Structural VARs. *University of Melbourne Working Paper Series*, 1–28.
- Yang, Q. (2019). Stock returns and real growth: A bayesian nonparametric approach. *Journal of Empirical Finance* 53, 53–69.