Markov-Switching Models with State-Dependent Time-Varying Transition Probabilities

Zacharias Psaradakis*
Department of Economics, Mathematics and Statistics,
Birkbeck, University of London, U.K.

Martin Sola
Department of Economics, Universidad Torcuato di Tella, Argentina

April 2021

Abstract

Markov-switching models with covariate-dependent transition functions that are subject to exogenous discrete stochastic changes are considered. These changes are associated with simultaneous stochastic changes in the covariance structure of the observable variables. Simulation experiments are carried out to assess the quality of large-sample approximations to the distributions of the maximum-likelihood estimator and of related statistics in such a model, and to examine the implications of misspecification due to unaccounted breaks in the transition mechanism. The practical use of the model is illustrated by analyzing the relationship between Argentinian sovereign bond spreads and output growth.

Keywords: Markov-switching models; Maximum likelihood; Monte Carlo experiments; Time-varying transition probabilities.

*Corresponding author: Zacharias Psaradakis, Department of Economics, Mathematics and Statistics, Birkbeck, University of London, Malet Street, London WC1E 7HX, United Kingdom; e-mail: z.psaradakis@bbk.ac.uk.

1 Introduction

Stochastic models that allow the conditional distribution of observable variables of interest to depend on an unobservable temporally inhomogeneous Markov chain (the regime or state sequence) have attracted much attention. Applications involving such models can be found in biology (e.g., Ghavidel, Claesen, and Burzykowski (2015)), engineering (e.g., Ramesh and Wilpon (1992), Hughes and Guttorp (1994)) and, most prominently, economics and finance. Examples from the latter literature include, among many others, the study of: business-cycle fluctuations (Filardo (1994), Filardo and Gordon (1998), Ravn and Sola (1999), Simpson, Osborn, and Sensier (2001), Gadea Rivas and Perez-Quiros (2015)); exchange rates (Diebold, Lee, and Weinbach (1994), Engel and Hakkio (1996)); interest-rate dynamics (Bekaert and Harvey (1995), Gray (1996)); asset allocation (Bekaert and Harvey (1995), Ang and Bekaert (2002)); asset returns (Hall, Psaradakis, and Sola (1997), Schaller and van Norden (1997)); financial and exchange-rate crises (Peria (2002), Alvarez-Plata and Schrooten (2006), Brunetti, Scotti, Mariano, and Tan (2008)).

A question that is often addressed in empirical studies that make use of Markov-switching models with time-varying transition probabilities is whether there are observable variables which help to predict transitions between different regimes (a period of relative calm and a financial crisis, say). In such applications the sample typically includes data from all regimes and one of the aims of the modelling exercise is to identify the different regimes on the basis of sample information. An implicit assumption usually made is that the association between the covariates that determine the evolution of the transition probabilities (referred to hereafter as 'information variables') and the outcome variable (or 'target variable') of the model is not altered by changes in regime, so that the parameters associated with the transition functions of the regime sequence are fixed. However, this may not necessarily be a plausible assumption in some empirical applications. Ravn and Sola (1999), for example, observe that a change in the definition of M2 money stock in the U.S., and therefore in the correlation between M2 and output growth, had a dramatic impact on the separation of regimes implied by a Markov-switching model with time-varying transition probabilities. When a short sample that ended before the change in the definition of M2 is considered, the separation of regimes into booms and recessions is found to be consistent with the National Bureau of Economic Research (NBER) dating of business-cycle peaks and troughs, and M2 growth has a statistically significant effect on the transition probabilities. When the sample is extended, M2 no longer has a significant impact on the transition probabilities and the separation of regimes implied by the model is unrelated to the NBER dating. In view of the fact that in many applications the main reason for specifying transition probabilities as functions of exogenous variables is to use the latter as leading indicators of a given event (e.g., a financial crisis), the problem just described is likely to be the rule rather than the exception. As a result, potential leading indicators may be found to have no significant effect on the probabilities governing transitions between regimes, or may be found to have significant effects because of the strength of their covariation with the target variable after the event (which can hardly justify their use as leading indicators of the event). In addition, it is likely that in-sample separation of regimes will be adversely affected, leading to erroneous conclusions.

This paper considers a class of Markov-switching models designed to capture possible regime-dependence of the parameters of the transition functions of the hidden regime sequence by explicitly allowing for discrete stochastic breaks in these functions. It is reasonable to expect that changes in the (nonlinear) relationship between the target variable being modelled and the information variables that determine the evolution of the covariate-dependent transition matrices are likely to result in changes in the covariance between these variables as well as in the relationship between the information variables and the transition probabilities. Such changes in the covariance structure of the observable variables may, therefore, be used as an identification device for the stochastic changes that the transition mechanism of the unobservable regime sequence may be subject to. In many applications in which an information variable (e.g., money supply) is used as a leading indicator (e.g., of an exchange-rate crisis), although the covariation between the information variable and the target variable of interest (e.g., exchange rate) is likely to have changed over the sample period, the potential resulting changes in the parameters associated with the transition mechanism and the noise covariance matrix are rarely taken into account. The principal idea behind our modelling strategy is to use such breaks to identify changes in the (indirect) relationship between the target variable(s) and the information variable(s). In the formulation of the multivariate model under consideration this is achieved by allowing the noise covariance matrix and the transition functions of the regime sequence to be subject to stochastic breaks driven by an exogenous finite Markov chain.

Apart from the fact that the contemporaneous covariance structure of the target variables and the information variables is of direct interest in our model specification, joint modelling of these variables is essential in order to avoid the difficulties associated with likelihood-based inference procedures that rely on an incomplete model for the conditional distribution of the target variables alone given past information. The consequences for statistical inference of possible endogeneity of the information variables are investigated in Pouzo, Psaradakis, and Sola (2020), under the maintained assumption that the time-varying transition matrices of the hidden Markov regimes are not subject to the type of stochastic breaks considered here. The implications of the failure of this assumption, as well as a possible way of robustifying against them, are the focus of our analysis in this paper.

Section 2 recalls the structure of a typical Markov-switching autoregressive model and motivates our modelling approach. Section 3 introduces the multivariate model of interest and discusses likelihood-based estimation of its parameters. Section 4 contains a simulation study that assesses the properties of estimators and test statistics in the presence of simultaneous discrete changes in the regime transition functions and the noise covariance matrix. It is demonstrated that ignoring breaks of this type has significant adverse effects on ML-based inference procedures, even in what are very large samples by the standards of empirical applications. Furthermore, inference based on a model that allows for unobservable Markov breaks in the transition functions and the noise covariance matrix is almost as accurate as inference based on an 'oracle' model which incorporates full information on the number and location of such breaks. Section 5 presents an illustrative application that analyzes the relationship between Argentinian sovereign bond spreads and output growth.

The analysis reveals that this relationship underwent significant changes as a result of the default on sovereign debt in January 2002 and the subsequent debt restructuring, and that ignoring such a change limits substantially the ability of output growth to predict transitions between regimes. Finally, Section 6 summarizes and concludes.

2 Markov-Switching Models and Motivation

A prototypical Markov-switching autoregressive model for a univariate time series $\{Y_t\}$ is given by

$$Y_t = \mu(S_t) + \boldsymbol{\phi}'\mathbf{y}_{t-1} + \sigma(S_t)\varepsilon_t, \qquad t = 1, 2, \ldots, \tag{1}$$

where $\mathbf{y}_{t-1} := (Y_{t-1}, \ldots, Y_{t-k})'$ for some positive integer k, $\boldsymbol{\phi} := (\phi_1, \ldots, \phi_k)'$ is a vector of unknown coefficients, $\{\varepsilon_t\}$ are independent and identically distributed (i.i.d.) random variables with zero mean and unit variance, and $\{S_t\}$ are latent random variables that take values in the set $\mathbb{S} := \{0, 1\}$ and indicate the unobservable state, or regime, prevailing at each time t. Here and in the sequel,

$$a(U) := a_0 + (a_1 - a_0)U, \qquad a_0, a_1 \in \mathbb{R},$$

for any scalar parameter $a(U)$ whose value depends on the realization of an $\mathbb{S}$-valued random variable U. The regime sequence $\{S_t\}$ in (1) is assumed to be independent of $\{\varepsilon_t\}$ and to form a temporally inhomogeneous Markov chain with transition probabilities

$$\mathsf{P}(S_t = 0 | S_{t-1} = 0, Z_{t-1}) =: q_t = \Lambda\left(\alpha_q + \beta_q Z_{t-1}\right), \tag{2}$$

$$\mathsf{P}(S_t = 1 | S_{t-1} = 1, Z_{t-1}) =: p_t = \Lambda(\alpha_p + \beta_p Z_{t-1}), \tag{3}$$

where Z_{t-1} is an observable exogenous variable upon which the transition probabilities depend and $\Lambda(z) := 1/(1+e^{-z})$, $z \in \mathbb{R}$, is the logistic distribution function. Needless to say, such a model may be generalized to allow for state-dependent autoregressive coefficients $\boldsymbol{\phi}$, more than two regimes, and multiple information variables in (2)–(3); moreover, Λ may be replaced by some other continuous, monotone function whose range is contained in the interval $[0, 1]$.

The specification in (1)–(3) allows for a nonlinear relationship between Y_t and Z_t, in the sense that Z_{t-1} has a direct effect on the probability with which $\{Y_t\}$ switches between the two regimes. The effect of Z_{t-1} on the transition probabilities q_t and p_t need not be symmetric since β_q and β_p are not required to be equal or have the same sign. A Markov-switching autoregressive model with a temporally homogeneous transition mechanism is, of course, a special case of (1)–(3) with $\beta_q = \beta_p = 0$.

An important issue that has not received much attention in the literature concerns the effects on inference of potential changes in the parameters associated with the time-varying transition probabilities. For example, a number of empirical studies have documented that several monetary relationships display instability because of changes in monetary policy and in innovations in the financial sector (e.g., Ravn and Sola (1999)). Such changes may lead to instability of the parameters associated with the Markov transition functions, which are likely to have deleterious effects on the properties of conventional inferential procedures if they are not accounted for.

In the simple setup considered in Psaradakis, Sola, Spagnolo, and Spagnolo (2013), the Markov chain $\{S_t\}$ in (1) is allowed to be governed by the transition probabilities

$$q_t = \Lambda\left(\alpha_q + \beta_{q,t} Z_{t-1}\right), \qquad p_t = \Lambda\left(\alpha_p + \beta_{p,t} Z_{t-1}\right), \tag{4}$$

where, for some fixed integer $t^* > 1$,

$$\beta_{i,t} = \beta_i + (\beta_i^* - \beta_i)\mathbf{1}_{\{t-t^*>0\}}, \qquad i = p, q, \qquad t \geqslant 1, \tag{5}$$

with $\mathbf{1}_{\{C\}}$ denoting an indicator that takes the value 1 when condition C is met and 0 otherwise. Under (4)–(5), the relationship between the transition probabilities and the information variable Z_{t-1} undergoes an one-off change at observation t^*. The transition mechanism (2)–(3) is misspecified when $\beta_q^* \neq \beta_q$ and/or $\beta_p^* \neq \beta_p$, and inference on the parameters of the model and the hidden regimes is demonstrated in Psaradakis, Sola, Spagnolo, and Spagnolo (2013) to be adversely affected as a result.

A more general formulation, which can accommodate an unspecified number of changes in the transition functions of $\{S_t\}$ at unspecified points in the sample, may be obtained by allowing the parameters $(\alpha_q, \alpha_p, \beta_q, \beta_p)$ in (2)–(3) to vary over time as (exogenous) finite Markov chains. Under such a formulation, the relationship between the transition probabilities of $\{S_t\}$ and the variable Z_{t-1} undergoes discrete stochastic changes which are likely to be associated with simultaneous stochastic changes in the (conditional) covariance between Y_t and Z_t. A multivariate model which allows for such behaviour is discussed next.

3 Changes in Transition Functions

For simplicity and clarity of exposition, we will present and discuss in the sequel a bivariate model. The model aims to capture changes in the relationship between Y_t and Z_t by conjecturing that such changes would manifest as changes in the covariance structure between Y_t and Z_t as well as simultaneous changes in the transition mechanism of $\{S_t\}$. This covariance structure and the transition functions of $\{S_t\}$ are then modelled as being subject to stochastic breaks governed by an exogenous finite Markov chain.

3.1 Model

Let $\{\boldsymbol{\xi}_t := (S_t, X_t)'\}$ be an unobservable sequence of random vectors taking values in $\mathbb{S} \times \mathbb{S}$ and $\{\boldsymbol{\eta}_t\}$ be an unobservable sequence of i.i.d. two-dimensional real random vectors having zero mean and identity covariance matrix. We consider the following model for the observable bivariate time series $\{\mathbf{w}_t := (Y_t, Z_t)'\}$:

$$Y_t = \mu(S_t) + \boldsymbol{\phi}'\mathbf{y}_{t-1} + \sigma(S_t)\varepsilon_{yt}, \qquad t = 1, 2, \ldots, \tag{6}$$

$$Z_t = \mu_z + \boldsymbol{\psi}'\mathbf{z}_{t-1} + \sigma_z\varepsilon_{zt}, \qquad t = 1, 2, \ldots, \tag{7}$$

where $\mathbf{y}_{t-1} := (Y_{t-1}, \ldots, Y_{t-k})'$ and $\mathbf{z}_{t-1} := (Z_{t-1}, \ldots, Z_{t-m})'$ for some positive integers k and m, $\boldsymbol{\phi} := (\phi_1, \ldots, \phi_k)'$ and $\boldsymbol{\psi} := (\psi_1, \ldots, \psi_m)'$ are vectors of unknown parameters, and the noise $\boldsymbol{\varepsilon}_t := (\varepsilon_{yt}, \varepsilon_{zt})'$ satisfies

$$\boldsymbol{\varepsilon}_t = \mathbf{R}(X_t)^{1/2}\boldsymbol{\eta}_t, \tag{8}$$

with

$$\mathbf{R}(X_t) := \begin{bmatrix} 1 & \rho(X_t) \\ \rho(X_t) & 1 \end{bmatrix}, \tag{9}$$

and $\max\{|\rho_0|, |\rho_1|\} < 1$. (Here and elsewhere, for any nonnegative definite matrix $\mathbf{A}$, $\mathbf{A}^{1/2}$ denotes a square matrix such that $\mathbf{A}^{1/2}(\mathbf{A}^{1/2})' = \mathbf{A}$). The model is completed by postulating that, conditionally on $\{X_t\}$, $\{S_t\}$ is a temporally inhomogeneous Markov chain with transition probabilities that depend on Z_{t-1} and X_t according to the functional relationships:

$$\mathsf{P}\left(S_t = 0 | S_{t-1} = 0, Z_{t-1}, X_t\right) =: q_t(X_t) = \Lambda\left(\alpha_q(X_t) + \beta_q(X_t)Z_{t-1}\right), \tag{10}$$

$$\mathsf{P}\left(S_t = 1 | S_{t-1} = 1, Z_{t-1}, X_t\right) =: p_t(X_t) = \Lambda\left(\alpha_p(X_t) + \beta_p(X_t)Z_{t-1}\right). \tag{11}$$

In addition, $\{X_t\}$ is a temporally homogeneous Markov chain with transition probabilities

$$q_x := \mathsf{P}(X_t = 0 | X_{t-1} = 0), \qquad p_x := \mathsf{P}(X_t = 1 | X_{t-1} = 1), \tag{12}$$

which is exogenous in the sense that the conditional distribution of X_t given $\{\boldsymbol{\xi}_r : r < t\}$ depends only on X_{t-1}. Finally, $\{\boldsymbol{\xi}_t\}$, $\{\boldsymbol{\eta}_t\}$ and $\ddot{\mathbf{w}}_0 := (\mathbf{y}_0', \mathbf{z}_0')'$ are independent of each other.

Unlike standard formulations of Markov-switching models such as (1)–(3), in which the parameters associated with the transition functions are fixed, under (6)–(12) the transition mechanism that governs the regime process $\{S_t\}$ driving the changes in the intercept and the error variance in the equation for $\{Y_t\}$ is itself dependent upon the state of nature implied by the exogenous process $\{X_t\}$. In other words, the transition matrices of $\{S_t\}$ are not only time-varying, being as they are functions of Z_{t-1}, but also state-dependent, owing to the dependence of the transition probabilities of $\{S_t\}$ at each time t on the state the process $\{X_t\}$ is in. At the same time, the covariance matrix $\mathbf{R}(X_t)$ of $\boldsymbol{\varepsilon}_t$ is subject to Markov changes governed by $\{X_t\}$. Such changes in the conditional covariance between Y_t and Z_t allow us to identify the changes in the indirect nonlinear relationship between Y_t and Z_t that exists because of the dependence of the dynamics of $\{Y_t\}$ on $\{S_t\}$. The model may, of course, be generalized to allow for state-dependent parameters $(\boldsymbol{\phi}, \boldsymbol{\psi}, \mu_z, \sigma_z)$ and more than two states and/or variables.

It is worth emphasizing that joint modelling of (Y_t, Z_t) is essential in order to ensure that likelihood-based inferential procedures have optimal statistical properties. As demonstrated in Pouzo, Psaradakis, and Sola (2020), in the presence of contemporaneous correlation between Y_t and Z_t, ML estimation of the parameters of the equation for Y_t alone yields estimates which are severely biased and inconsistent, and associated hypothesis tests that have unsatisfactory properties, despite the fact that the transition probabilities of the Markov process that drives changes in the parameters of the equation for Y_t depend only on the lagged information variable Z_{t-1}.

Also note that a standard vector autoregressive model for $\{\mathbf{w}_t\}$ with Markov-switching parameters is not particularly useful in cases where Z_t is viewed as a leading indicator for changes in the structure of Y_t. In such cases, it seems more desirable to allow lagged values of Z_t to affect Y_t indirectly through a covariate-dependent Markov transition mechanism. An indirect link of this kind between Y_t and other observable variables that may affect transitions between different states of nature is also absent from observation-driven Markov-switching models in which transition dynamics are driven by the score of the predictive likelihood function, as is the case, for example, in the models considered by Bazzi, Blasques, Koopman, and Lucas (2017) and Bernardi and Catania (2019). This is also true for vector autoregressive models of the type considered in Otranto (2005), in which transition probabilities are dependent not on observable variables but on the unobservable regimes associated with different variables of the system. In the context of volatility modelling, examples of multivariate models in which the conditional correlations between variables are subject to discrete stochastic changes include those of Pelletier (2006) and Bauwens and Otranto (2016). In such models, the regime-dependent conditional correlations are the focus of interest, whereas in our model changes in the conditional correlation between Y_t and Z_t are primarily viewed as an identification device for stochastic changes in the transition mechanism of the unobservable Markov chain that determines the regime-dependent dynamics of the target variable.

3.2 Inference

Given observations $\ddot{\mathbf{w}}_0, \mathbf{w}_1, \ldots, \mathbf{w}_T$, inference in the model defined by (6)–(12) can be carried out by using a recursive algorithm analogous to that discussed in Hamilton (1994, pp. 692–694). This entails iterating on the equations

$$\boldsymbol{\delta}_{t|t} = [\boldsymbol{\iota}'(\boldsymbol{\delta}_{t|t-1} \odot \mathbf{g}_t)]^{-1}(\boldsymbol{\delta}_{t|t-1} \odot \mathbf{g}_t), \qquad t = 1, 2, \ldots, T, \tag{13}$$

and

$$\boldsymbol{\delta}_{t+1|t} = \mathbf{P}_t \boldsymbol{\delta}_{t|t}, \qquad t = 1, 2, \ldots, T, \tag{14}$$

where

$$\boldsymbol{\delta}_{t|l} := \begin{bmatrix} \mathsf{P}(\boldsymbol{\xi}_t' = (0,0)|\mathcal{F}_0^l; \boldsymbol{\theta}) \\ \mathsf{P}(\boldsymbol{\xi}_t' = (0,1)|\mathcal{F}_0^l; \boldsymbol{\theta}) \\ \mathsf{P}(\boldsymbol{\xi}_t' = (1,0)|\mathcal{F}_0^l; \boldsymbol{\theta}) \\ \mathsf{P}(\boldsymbol{\xi}_t' = (1,1)|\mathcal{F}_0^l; \boldsymbol{\theta}) \end{bmatrix}, \qquad \mathbf{g}_t := \begin{bmatrix} g_{00}(\mathbf{w}_t|\boldsymbol{\xi}_t' = (0,0), \mathcal{F}_0^{t-1}; \boldsymbol{\theta}) \\ g_{01}(\mathbf{w}_t|\boldsymbol{\xi}_t' = (0,1), \mathcal{F}_0^{t-1}; \boldsymbol{\theta}) \\ g_{10}(\mathbf{w}_t|\boldsymbol{\xi}_t' = (1,0), \mathcal{F}_0^{t-1}; \boldsymbol{\theta}) \\ g_{11}(\mathbf{w}_t|\boldsymbol{\xi}_t' = (1,1), \mathcal{F}_0^{t-1}; \boldsymbol{\theta}) \end{bmatrix},$$

and

$$\mathbf{P}_t := \begin{bmatrix} q_{t0}q_x & q_{t0}(1-p_x) & (1-p_{t0})q_x & (1-p_{t0})(1-p_x) \\ q_{t1}(1-q_x) & q_{t1}p_x & (1-p_{t1})(1-q_x) & (1-p_{t1})p_x \\ (1-q_{t0})q_x & (1-q_{t0})(1-p_x) & p_{t0}q_x & p_{t0}(1-p_x) \\ (1-q_{t1})(1-q_x) & (1-q_{t1})p_x & p_{t1}(1-q_x) & p_{t1}p_x \end{bmatrix}.$$

Here, $\boldsymbol{\theta}$ denotes the vector of all free parameters of the model, $\mathcal{F}_0^t := \{\ddot{\mathbf{w}}_0, \mathbf{w}_1, \ldots, \mathbf{w}_t\}$ is the information set available at time t, $g_{ij}(\mathbf{w}_t|\boldsymbol{\xi}_t' = (i,j), \mathcal{F}_0^{t-1}; \boldsymbol{\theta})$, $i, j \in \mathbb{S}$, is the conditional density of $\mathbf{w}_t$ given $\boldsymbol{\xi}_t' = (i,j)$ and $\mathcal{F}_0^{t-1}$, $\boldsymbol{\iota}$ is a four-dimensional all-ones column vector, and $\odot$ denotes element-wise multiplication. It is easily checked that the elements of $\mathbf{P}_t$ are the one-step transition probabilities of the bivariate Markov chain $\{\boldsymbol{\xi}_t\}$ (recall that $q_{ti} = \Lambda\left(\alpha_{qi} + \beta_{qi}Z_{t-i}\right)$ and $p_{ti} = \Lambda\left(\alpha_{pi} + \beta_{pi}Z_{t-1}\right)$ for $i \in \mathbb{S}$). Furthermore, under the assumption of Gaussianity of $\boldsymbol{\eta}_t$, $g_{ij}(\mathbf{w}_t|\boldsymbol{\xi}_t' = (i,j), \mathcal{F}_{t-1}; \boldsymbol{\theta})$, $i, j \in \mathbb{S}$, is the bivariate normal density with mean vector

$$\left(\mu_0 + (\mu_1 - \mu_0)i + \boldsymbol{\phi}'\mathbf{y}_{t-1},\ \mu_z + \boldsymbol{\psi}'\mathbf{z}_{t-1}\right)',$$

and covariance matrix

$$\begin{bmatrix} \sigma_0^2 + (\sigma_1^2 - \sigma_0^2)i & \sigma_z\{\sigma_0 + (\sigma_1 - \sigma_0)i\}\{\rho_0 + (\rho_1 - \rho_0)j\} \\ \sigma_z\{\sigma_0 + (\sigma_1 - \sigma_0)i\}\{\rho_0 + (\rho_1 - \rho_0)j\} & \sigma_z^2 \end{bmatrix}.$$

The log-likelihood of $\boldsymbol{\theta}$ associated with the observed data can be computed from the iteration of (13)–(14) as

$$\ell(\boldsymbol{\theta}) := \sum_{t=1}^{T} \ln(\boldsymbol{\iota}'[\boldsymbol{\delta}_{t|t-1} \odot \mathbf{g}_t]).$$

The ML estimator $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is obtained as the maximizer of $\ell(\boldsymbol{\theta})$. Furthermore, inferences about the hidden regimes $\{\boldsymbol{\xi}_t\}$ may be made on the basis of the filtered state probabilities $\boldsymbol{\delta}_{t|t}$, or the smoothed state probabilities $\boldsymbol{\delta}_{t|T}$, evaluated at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$.

Results relating to local asymptotic normality of a general class of Markov-switching models with time-varying transition probabilities and to consistency of ML estimators in such models (in the presence or absence of misspecification) were recently established in Pouzo, Psaradakis, and Sola (2020). These results ensure that, under the assumption of correct model specification and suitable regularity conditions, the ML estimator in a model like (6)–(12) has standard large-sample properties. Specifically, as T diverges to infinity, $\widehat{\boldsymbol{\theta}}$ is consistent for the true parameter vector $\boldsymbol{\theta}_0$ (say) and $\mathbf{I}(\boldsymbol{\theta}_0)^{1/2}\sqrt{T}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ is asymptotically normal with zero mean and identity covariance matrix, $\mathbf{I}(\boldsymbol{\theta}_0)$ being the asymptotic Fisher information matrix (the definition of which can be found in Pouzo, Psaradakis, and Sola (2020, Corollary 1(a))). We note that, under appropriate regularity conditions, such results continue to hold for more general versions of the model (6)–(12) that allow for more than two states and/or variables, as long as the unobservable regime sequences take values in a finite state space and the conditional distribution of the observable variables, given past information and current states, admits a density (with respect to some suitable dominating measure).

4 Monte Carlo Simulations

Simulation experiments are carried out to assess the properties of the ML estimator and of related test statistics both in correctly specified models and in misspecified models which ignore changes in the parameters associated with the Markov transition functions. We begin by describing the experimental design and simulations, and proceed to report and discuss the results of the experiments.

4.1 Experimental Design and Simulation

The data-generating mechanism used in the experiments is the bivariate model defined by (6)–(12), with first-order dynamics ($k = m = 1$), Gaussian noise $\boldsymbol{\eta}_t$, and the following parameter values:

$$\mu_0 = 3, \qquad \mu_1 = 0.5, \qquad \phi = 0.9, \qquad \sigma_0 = 0.5, \qquad \sigma_1 = 1,$$

$$\mu_z = 0.1, \qquad \psi = 0.8, \qquad \sigma_z = 1, \qquad \rho_0 = 0.8, \qquad \rho_1 = -0.8,$$

$$p_x = q_x = 0.95, \qquad \alpha_{p0} = \alpha_{p1} = 1, \qquad \alpha_{q0} = \alpha_{q1} = 2,$$

$$\beta_{q_0} = -1.5, \qquad \beta_{q1} = 3, \qquad \beta_{p0} = 2.5, \qquad \beta_{p1} = -5.$$

Figure 1 shows the regime-specific marginal densities of $\mathbf{w}_t$, that is, bivariate normal densities with mean vector

$$\left(\frac{\mu_0 + (\mu_1 - \mu_0)i}{1-\phi}, \ \frac{\mu_z}{1-\psi} \right)',$$

and covariance matrix

$$\begin{bmatrix} \dfrac{\sigma_0^2 + (\sigma_1^2 - \sigma_0^2)i}{1-\phi^2} & \{\rho_0 + (\rho_1 - \rho_0)j\}\sqrt{\dfrac{\{\sigma_0^2 + (\sigma_1^2 - \sigma_0^2)i\}\sigma_z^2}{(1-\phi^2)(1-\psi^2)}} \\ \{\rho_0 + (\rho_1 - \rho_0)j\}\sqrt{\dfrac{\{\sigma_0^2 + (\sigma_1^2 - \sigma_0^2)i\}\sigma_z^2}{(1-\phi^2)(1-\psi^2)}} & \dfrac{\sigma_z^2}{1-\psi^2} \end{bmatrix},$$

with $i, j \in \mathbb{S}$.

In each of 1000 Monte Carlo replications, $100 + T$ data points for $\mathbf{w}_t$ are generated, with $\mathbf{w}_0' = (0.5, 0.40118)$ and $T \in \{100, 200, 400, 800, 1600, 3200, 6400\}$, but only the last T points of each realization are used in order to attenuate start-up effects. For each realization, we compute ML estimates of the parameters of three models for (Y_t, Z_t):

(i) Model M-1: defined by (6)–(9), with $\rho(X_t) = \rho$ for all t, coupled with the state-independent Markov-switching mechanism associated with (2)–(3);

(ii) Model M-2: defined by (6)–(12), with $\{S_t\}$ and $\{X_t\}$ treated as unobservable;

(iii) Model M-3: defined by (6)–(12), but with the additional assumption that the realization of the Markov process $\{X_t\}$ driving the changes in the transition probabilities $q_t(X_t)$ and $p_t(X_t)$ and in the covariance matrix $\mathbf{R}(X_t)$ is observable.

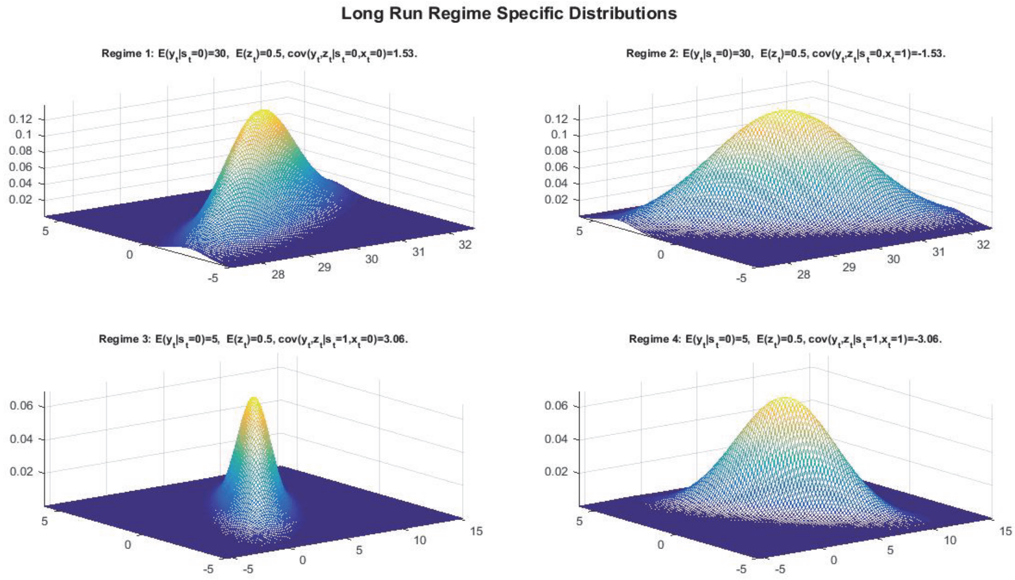

Figure 1: Regime-specific marginal densities of $\{\mathbf{w}_t\}$

Model M-1 is evidently misspecified since it ignores changes in the covariance matrix of $\boldsymbol{\varepsilon}_t$ and in the transition functions of $\{S_t\}$. Model M-3 accounts for all the changes in $q_t(X_t)$, $p_t(X_t)$ and $\rho(X_t)$ correctly, but an observable regime sequence $\{X_t\}$ is not typically available in situations involving real-world data. However, a comparison of parameter estimates obtained from model M-3 and from the empirically relevant model M-2 will reveal whether statistical inference suffers as a result of the additional randomness introduced by the unobservable Markov process $\{X_t\}$.

For all three models, we set $k = m = 1$ and obtain the maximizer of the relevant estimator objective function by means of the Broyden–Fletcher–Goldfarb–Shanno (BFGS) quasi-Newton algorithm using numerically computed derivatives. A grid of seven initial values for each parameter is used to initialize the BFGS iterations; those initial values that result in the highest value of the objective function are then selected.

Figure 2 shows plots of: (i) a typical realization of $\{Y_t\}$ and $\{Z_t\}$; (ii) the time-varying transition probabilities $p_t(X_t)$ and $q_t(X_t)$ and their estimated values – the latter are computed as $\bar{p}_t = \mathsf{P}(X_t = 0|\mathcal{F}_0^t)p_{0t} + \mathsf{P}(X_t = 1|\mathcal{F}_0^t)p_{1t}$ and $\bar{q}_t = \mathsf{P}(X_t = 0|\mathcal{F}_0^t)q_{0t} + \mathsf{P}(X_t =$

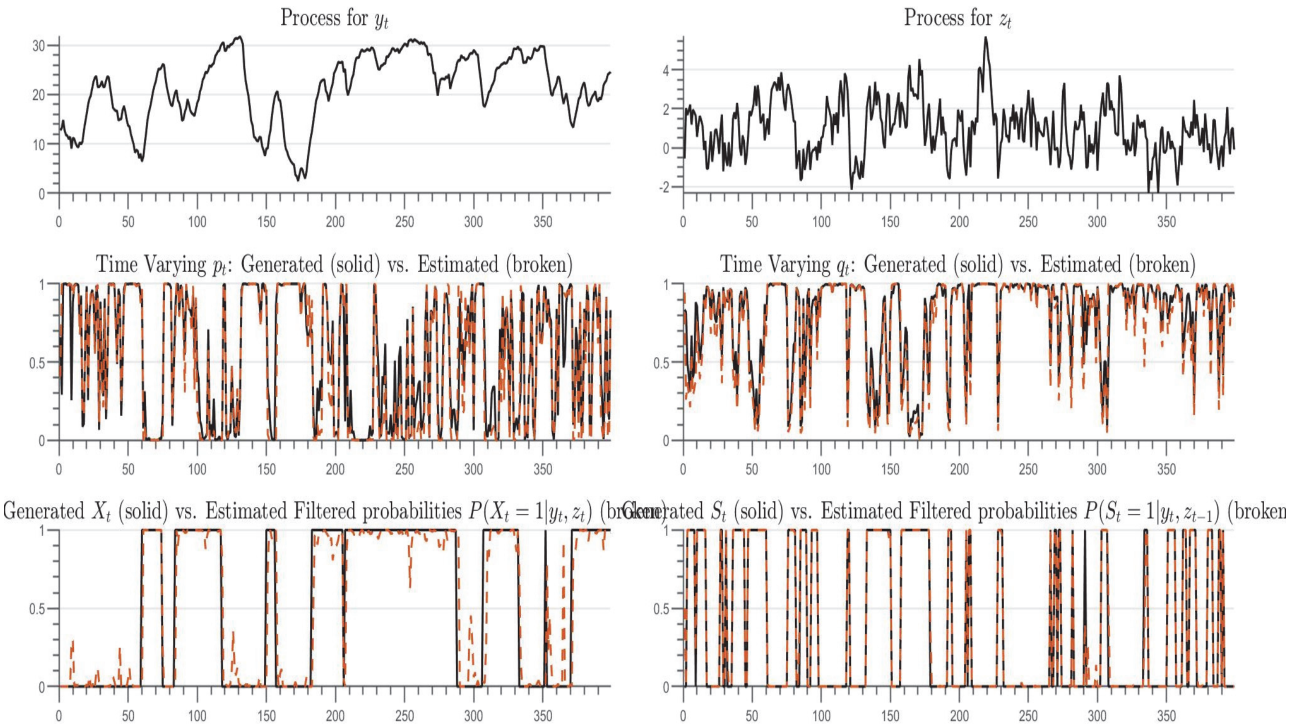

Figure 2: Realization of $\{Y_t\}$ and $\{Z_t\}$; true and estimated transition probabilities p_t and q_t; realization of $\{S_t\}$ and $\{X_t\}$; inferred state probabilities $\mathsf{P}(X_t = 1|\mathcal{F}_0^t)$ and $\mathsf{P}(S_t = 1|\mathcal{F}_0^t)$ (based on Model M-2)

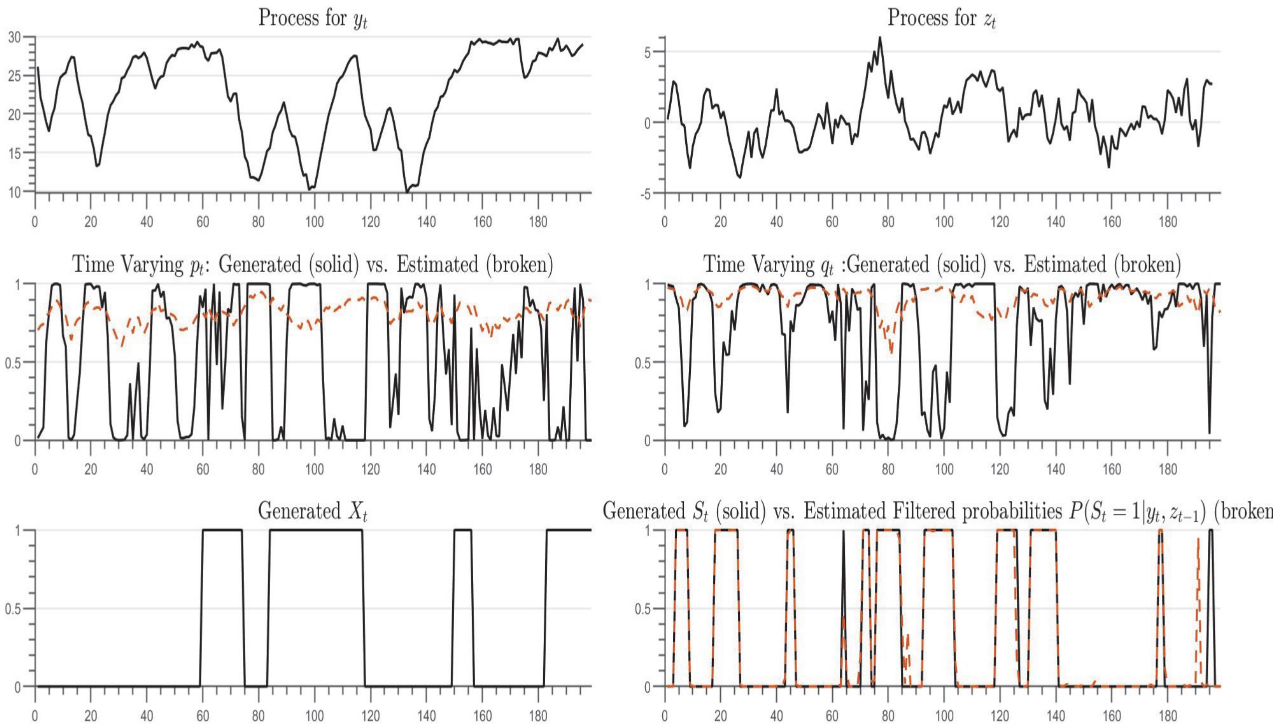

Figure 3: Realization of $\{Y_t\}$ and $\{Z_t\}$; true and estimated transition probabilities p_t and q_t; realization of $\{S_t\}$ and $\{X_t\}$; inferred state probabilities $\mathsf{P}(S_t = 1|\mathcal{F}_0^t)$ (based on Model M-1)

$1|\mathcal{F}_0^t)q_{1t}$; (iii) the associated realizations of the Markov chains $\{S_t\}$ and $\{X_t\}$; (iv) the filtered state probabilities $\mathsf{P}(S_t = 1|\mathcal{F}_0^t)$ and $\mathsf{P}(X_t = 1|\mathcal{F}_0^t)$; all quantities depending on the unknown parameter $\boldsymbol{\theta}$ are computed using the ML estimate $\widehat{\boldsymbol{\theta}}$ from model M-2. For this typical realization, model M-2 is remarkably successful at capturing the changes in regime and the movements in the transition probabilities of $\{S_t\}$. It is interesting to compare results for the same realization of $\{Y_t\}$ and $\{Z_t\}$ when model M-1 is used instead. The relevant plots, shown in Figure 3, reveal that, although model M-1 is equally successful in identifying the changes in regime associated with $\{S_t\}$, it fails dramatically in describing the dynamics of $p_t(X_t)$ and $q_t(X_t)$. This suggests that it is unlikely that Z_{t-1} would be found to be useful in predicting the shifts implied by the regime sequence $\{S_t\}$. As we shall see below, this is not a peculiar feature of the particular realization shown in Figure 3 but is true in general. In light of the fact that models such as M-1 are often used in the empirical literature as a way of identifying leading indicators, their inability to describe correctly the dynamics of the transition probabilities of $\{S_t\}$ presents a serious challenge.

4.2 Simulation Results

In order to conserve space, only a selection of the simulation results are reported here (the full set of results is available upon request). Table 1 records some of the characteristics of the finite-sample distribution of the ML estimator of the parameters of the equation for Y_t in model M-1. Specifically, we report the estimated bias and the ratio of the sampling standard deviation of the estimators to the estimated standard errors (obtained from the empirical Hessian estimator) averaged across Monte Carlo replications for each design point. Note that in Table 1 (as well as in Tables 2 and 3) the figures reported under $\alpha_{p(i)}$, $\beta_{p(i)}$, $\alpha_{q(i)}$ and $\beta_{q(i)}$, $i \in \mathbb{S}$, refer to deviations of the ML estimates of the parameters $(\alpha_p, \beta_p, \alpha_q, \beta_q)$ in (2)–(3) from the true regime-specific values of the corresponding parameters. The most noteworthy finding is the significant bias in the estimation of parameters associated with the transition probabilities, especially β_q and β_p. Such biases are clearly not a small-sample issue and are present even in the largest of the samples under consideration ($T = 6400$). Although the estimated standard errors are downwards

T	μ_0	μ_1	$\alpha_{p(0)}$	$\beta_{p(0)}$	$\beta_{p(1)}$	$\alpha_{q(0)}$	$\beta_{q(0)}$	$\beta_{q(1)}$	σ_0	σ_1	ϕ_1
	Bias										
100	0.080	0.073	0.134	-1.311	-0.689	0.133	-0.689	-1.311	-0.044	-0.044	-0.015
200	0.051	0.048	0.105	-1.293	-0.707	0.090	-0.666	-1.334	-0.018	-0.023	-0.009
400	0.031	0.033	0.069	-1.347	-0.653	0.076	-0.641	-1.359	-0.009	-0.013	-0.006
800	0.020	0.021	0.048	-1.292	-0.708	0.054	-0.685	-1.315	-0.003	-0.007	-0.004
1600	0.012	0.015	0.061	-1.349	-0.651	0.068	-0.651	-1.349	-0.002	-0.002	-0.004
3200	0.012	0.010	0.072	-1.373	-0.627	0.070	-0.634	-1.366	-0.003	-0.002	-0.004
6400	0.008	0.010	0.065	-1.346	-0.654	0.065	-0.658	-1.342	-0.002	-0.002	-0.003
	Ratio of sampling standard deviation to estimated standard error										
100	1.333	1.487	1.516	1.622	1.622	1.419	1.520	1.520	1.273	1.262	1.409
200	1.248	1.184	1.412	1.587	1.587	1.421	1.326	1.326	1.155	1.100	1.165
400	1.073	1.047	1.157	1.137	1.137	1.231	1.154	1.154	1.102	1.036	1.059
800	1.025	1.048	1.075	1.115	1.115	1.113	1.118	1.118	1.051	1.007	1.032
1600	0.980	0.991	1.075	1.121	1.121	1.045	1.043	1.043	1.037	1.027	0.956
3200	1.056	0.993	0.970	0.934	0.934	1.005	0.962	0.962	1.070	1.051	1.000
6400	0.967	1.001	0.965	0.882	0.882	0.943	0.979	0.979	1.035	1.097	0.982

Table 1: Bias and standard deviation of ML estimators (Model M-1)

biased in most cases, the bias is not generally substantial and decreases as the sample size increases.

Table 2 contains information on the sampling distributions of conventional t-type statistics, computed as the ratio of the estimation error to the corresponding estimated standard error, which are typically treated as having an approximate standard normal distribution (this is true under the assumption of a correctly specified likelihood). It is immediately obvious that the mean of these distributions can differ substantially from zero, something which is especially true for t-statistics associated with the parameters of the transition functions. Conventional measures of skewness and kurtosis based on standardized third and fourth empirical central moments (not shown) reveal that the Studentized statistics generally tend to have skewed distributions and, in the case of β_q and β_p, highly leptokurtic (especially for the larger sample sizes).

To examine the implications of these results for hypothesis testing, we report in Table 3 the empirical rejection frequencies (based on standard normal critical regions with nominal

T	α_0	α_1	$\alpha_{p(0)}$	$\beta_{p(0)}$	$\beta_{p(1)}$	$\alpha_{q(0)}$	$\beta_{q(0)}$	$\beta_{q(1)}$	σ_0	σ_1	ϕ_1
	Mean										
100	0.406	-0.368	-0.092	-0.976	0.643	-0.090	0.618	-1.015	-0.551	-0.533	-0.383
200	0.367	-0.356	0.049	-1.516	0.854	0.005	0.874	-1.516	-0.311	-0.373	-0.373
400	0.310	-0.332	0.142	-2.329	1.242	0.161	1.246	-2.382	-0.231	-0.286	-0.339
800	0.282	-0.295	0.211	-3.303	1.923	0.240	1.900	-3.387	-0.134	-0.202	-0.348
1600	0.246	-0.293	0.461	-5.060	2.757	0.537	2.762	-5.064	-0.107	-0.114	-0.470
3200	0.323	-0.289	0.848	-6.918	3.472	0.807	3.630	-6.990	-0.167	-0.141	-0.668
6400	0.327	-0.375	1.114	-9.682	5.095	1.124	5.548	-10.195	-0.132	-0.147	-0.863
	Standard deviation										
100	1.239	1.213	1.137	1.203	1.231	1.096	1.237	1.181	1.174	1.149	1.161
200	1.239	1.213	1.137	1.203	1.231	1.096	1.237	1.181	1.174	1.149	1.161
400	1.070	1.049	1.100	1.389	1.506	1.119	1.641	1.531	1.109	1.051	1.052
800	1.040	1.064	1.064	1.769	1.916	1.071	2.055	1.909	1.064	1.007	1.033
1600	0.978	0.993	1.080	2.736	3.107	1.059	3.274	2.932	1.033	1.029	0.949
3200	1.054	0.999	1.012	3.516	3.824	1.039	4.293	3.878	1.070	1.047	0.995
6400	0.970	0.999	1.019	5.009	5.251	0.989	6.256	6.043	1.036	1.100	0.977

Table 2: Empirical moments of t-statistics (Model M-1)

probability of Type I error 0.05) of: (i) a t-type test of $\mathcal{H}_0 : \vartheta_j = \vartheta_j^*$ versus $\mathcal{H}_1 : \vartheta_j \neq \vartheta_j^*$, where ϑ_j is the j-th element of the parameter vector $\boldsymbol{\vartheta}$ of the model under consideration and ϑ_j^* is its true value; (ii) a t-type test of $\mathcal{H}_0 : \vartheta_j = 0$ versus $\mathcal{H}_1 : \vartheta_j \neq 0$; we refer to the estimated rejection frequencies as 'size' and 'power', respectively. (Note that results should be interpreted with caution in the case of $\mathcal{H}_0 : \sigma_j = 0$, $j \in \mathbb{S}$, because the null value of σ_j is on the boundary of the maintained hypothesis.) It can be seen that the hypothesis that β_q or β_p is equal to either of the two regime-specific true values is rejected either very rarely or very frequently. Furthermore, tests for the statistical significance of these parameters have very low power. As a result, one may be wrongly led to conclude that significant leading indicators have no effect on the transition probabilities.

These findings are not very surprising. As shown in Pouzo, Psaradakis, and Sola (2020), when the true data-generating mechanism is not a special case of the model under consideration, the ML estimator can at best be consistent for the parameter value which provides the best approximating model, in the sense of minimizing the Kullback–Leibler divergence from the data-generating mechanism. Furthermore, the ML estimator and

T	α_0	α_1	$\alpha_{p(0)}$	$\beta_{p(0)}$	$\beta_{p(1)}$	$\alpha_{q(0)}$	$\beta_{q(0)}$	$\beta_{q(1)}$	σ_0	σ_1	ϕ_1
	Size										
100	0.058	0.165	0.096	0.318	0.020	0.102	0.020	0.340	0.169	0.172	0.158
200	0.044	0.128	0.069	0.499	0.012	0.066	0.017	0.490	0.125	0.124	0.136
400	0.027	0.102	0.064	0.708	0.003	0.067	0.010	0.712	0.103	0.094	0.101
800	0.031	0.098	0.054	0.828	0.004	0.055	0.001	0.858	0.083	0.077	0.113
1600	0.035	0.082	0.039	0.965	0.000	0.033	0.000	0.971	0.065	0.065	0.104
3200	0.033	0.093	0.014	0.986	0.000	0.016	0.000	0.992	0.082	0.072	0.180
6400	0.022	0.119	0.004	0.993	0.000	0.002	0.000	0.996	0.070	0.088	0.218
	Power										
100	0.985	0.977	0.678	0.064	0.064	0.677	0.067	0.067	1.000	1.000	1.000
200	0.999	0.997	0.917	0.071	0.071	0.924	0.082	0.082	1.000	1.000	1.000
400	1.000	1.000	0.992	0.117	0.117	0.985	0.107	0.107	1.000	1.000	1.000
800	0.999	0.999	0.999	0.114	0.114	0.999	0.141	0.141	1.000	1.000	1.000
1600	1.000	1.000	1.000	0.228	0.228	0.999	0.228	0.228	1.000	1.000	1.000
3200	1.000	1.000	0.997	0.404	0.404	0.998	0.378	0.378	1.000	1.000	1.000
6400	1.000	1.000	1.000	0.603	0.603	1.000	0.636	0.636	1.000	1.000	1.000

Table 3: Empirical size and power of t-type tests (Model M-1)

associated test statistics that rely on conventionally computed standard errors do not have the usual asymptotic distributions.

Table 4 records some of the characteristics of the finite-sample distribution of the ML estimator of the parameters of the equation for Y_t in the correctly specified model M-2. In sharp contrast to the results obtained under model M-1, the ML estimator exhibits some bias only in small samples and, even for the parameters associated with the transition probabilities, bias is insignificant for $T > 400$. Furthermore, bias in the estimated standard errors is not generally substantial and decreases as the sample size increases.

Encouraging results are also contained in Table 5, in which information on the empirical distributions of conventional t-type statistics is reported. The Studentized statistics tend to have distributions with mean and variance that do not differ substantially from their expected values in most cases. Rather surprisingly, the mean of the Studentized statistics associated with β_{q1} and β_{p1} is significantly different from zero when $T \geqslant 1600$. This, however, does not appear to have an adverse effect on the size and power properties of

T	μ_0	μ_1	α_{p0}	β_{p0}	β_{p1}	α_{q0}	β_{q0}	β_{q1}	σ_0	σ_1	ϕ_1
	Bias										
100	0.017	0.017	0.165	0.130	0.456	0.683	1.101	0.530	-0.019	-0.013	-0.011
200	0.002	-0.001	0.021	0.021	0.113	0.731	0.661	0.118	-0.009	-0.006	-0.007
400	0.003	0.007	0.028	0.013	0.089	0.161	0.210	0.055	-0.003	-0.006	-0.004
800	0.001	-0.001	0.003	0.001	-0.016	0.079	0.062	-0.031	-0.004	-0.003	-0.004
1600	0.001	0.000	0.002	-0.001	-0.056	-0.020	0.022	-0.062	0.000	0.001	-0.001
3200	0.001	0.000	0.004	0.004	-0.066	-0.034	-0.012	-0.063	-0.001	-0.001	-0.002
6400	0.000	0.000	0.002	0.001	-0.067	-0.033	-0.029	-0.066	-0.001	0.000	-0.002
	Ratio of sampling standard deviation to estimated standard error										
100	1.220	1.211	1.327	1.369	1.183	0.881	0.880	1.216	1.141	1.105	1.036
200	1.057	1.088	1.138	1.105	1.538	1.509	1.517	1.257	1.064	1.030	1.026
400	1.032	1.013	1.039	1.019	1.090	1.390	1.092	1.063	0.928	0.951	1.002
800	0.980	1.006	1.016	0.976	0.988	1.059	1.041	0.980	0.948	0.959	0.974
1600	0.965	1.015	0.969	0.946	0.969	1.028	1.023	0.939	0.942	0.868	0.942
3200	0.990	0.989	1.002	0.937	1.010	1.037	0.957	0.951	0.994	1.002	1.006
6400	1.055	1.006	0.939	0.963	0.935	0.893	0.882	0.946	1.015	1.015	1.027

Table 4: Bias and standard deviation of ML estimators (Model M-2)

T	μ_0	μ_1	α_{p0}	β_{p0}	β_{p1}	α_{q0}	β_{q0}	β_{q1}	σ_0	σ_1	ϕ_1
	Mean										
100	0.158	-0.136	0.089	0.061	-0.084	-0.171	-0.116	-0.096	-0.301	-0.237	-0.228
200	0.017	0.009	-0.079	-0.062	0.027	0.037	0.020	-0.020	-0.199	-0.157	-0.192
400	0.066	-0.129	0.065	-0.005	-0.049	-0.106	-0.040	-0.006	-0.099	-0.172	-0.160
800	0.015	0.033	-0.032	-0.041	0.086	-0.024	-0.074	0.115	-0.151	-0.112	-0.189
1600	0.035	-0.008	-0.021	-0.051	0.205	-0.171	-0.046	0.225	-0.022	-0.007	-0.083
3200	0.070	0.003	0.029	0.048	0.314	-0.165	-0.089	0.299	-0.069	-0.087	-0.189
6400	-0.028	0.039	0.018	0.005	0.434	-0.191	-0.160	0.431	-0.083	-0.068	-0.215
	Standard deviation										
100	1.050	1.085	1.017	1.023	1.042	1.083	1.078	0.989	1.100	1.079	1.052
200	1.050	1.085	1.017	1.023	1.042	1.083	1.078	0.989	1.100	1.079	1.052
400	1.031	1.016	0.966	0.989	1.083	1.056	1.212	0.978	0.982	0.989	1.030
800	0.983	1.011	1.007	0.986	1.031	1.130	1.218	1.057	1.025	1.046	1.019
1600	0.969	1.016	0.971	0.943	0.993	1.466	1.131	0.984	1.082	1.023	1.024
3200	0.989	0.993	0.999	0.936	1.000	1.058	0.954	0.941	1.069	1.055	1.043
6400	1.055	1.006	0.939	0.963	0.935	0.893	0.882	0.946	1.015	1.015	1.027

Table 5: Empirical moments of t-statistics (Model M-2)

T	μ_0	μ_1	α_{p0}	β_{p0}	β_{p1}	α_{q0}	β_{q0}	β_{q1}	σ_0	σ_1	ϕ_1
	Size										
100	0.073	0.105	0.039	0.063	0.030	0.111	0.094	0.035	0.114	0.112	0.112
200	0.047	0.062	0.062	0.070	0.039	0.079	0.075	0.041	0.093	0.086	0.082
400	0.052	0.068	0.041	0.062	0.054	0.082	0.061	0.042	0.061	0.076	0.079
800	0.046	0.039	0.057	0.057	0.043	0.054	0.043	0.033	0.067	0.068	0.081
1600	0.042	0.052	0.052	0.049	0.031	0.052	0.060	0.022	0.063	0.054	0.064
3200	0.049	0.045	0.056	0.045	0.037	0.075	0.045	0.019	0.077	0.059	0.091
6400	0.072	0.044	0.042	0.042	0.025	0.061	0.050	0.017	0.127	0.125	0.108
	Power										
100	1.000	0.997	0.757	0.733	0.078	0.039	0.029	0.065	1.000	0.999	1.000
200	1.000	1.000	0.953	0.949	0.168	0.094	0.082	0.160	1.000	0.999	1.000
400	1.000	1.000	1.000	0.999	0.369	0.154	0.197	0.339	1.000	1.000	1.000
800	1.000	1.000	0.999	1.000	0.608	0.445	0.409	0.596	1.000	1.000	1.000
1600	1.000	1.000	1.000	1.000	0.860	0.743	0.750	0.875	1.000	1.000	1.000
3200	1.000	1.000	1.000	1.000	0.983	0.944	0.969	0.986	1.000	1.000	1.000
6400	1.000	1.000	1.000	1.000	1.000	0.989	0.986	0.997	1.000	1.000	1.000
6400	1.000	1.000	1.000	1.000	1.000	0.992	0.994	1.000	1.000	1.000	1.000

Table 6: Empirical size and power of t-tests (Model M-2)

t-type tests, the empirical rejection frequencies of which are reported in Table 6. Unlike the case of model M-1, tests under model M-2 tend to have an empirical Type I error probability which is generally close to the nominal level, especially for $T > 200$. In terms of power, tests involving the parameters associated with the time-varying transition probabilities tend to fare somewhat worse than tests involving other parameters, but rejection frequencies improve with an increasing sample size.

The simulation results reported in Tables 7–9 for model M-3 are generally quite similar to those obtained under model M-2. Perhaps the only noteworthy difference concerns the mean of the Studentized statistics associated with β_{q1} and β_{p1}, which is closer to zero under model M-3. It is worth emphasizing that model M-3 is not empirically relevant since it is based on the assumption that changes in the transition functions of the regime sequence and in the covariance matrix of the noise are observable. However, comparison with the simulation results obtained under model M-2 reveals that not much is lost by treating the aforementioned changes as unobservable stochastic events governed by an

T	μ_0	μ_1	α_{p0}	β_{p0}	β_{p1}	α_{q0}	β_{q0}	β_{q1}	σ_0	σ_1	ϕ_1
	Bias										
100	0.010	0.013	0.092	0.929	0.545	0.088	0.333	0.772	-0.011	-0.009	-0.006
200	0.008	0.011	0.044	0.403	0.253	0.029	0.180	0.465	-0.002	-0.006	-0.003
400	0.001	0.004	0.017	0.123	0.119	0.015	0.082	0.189	-0.001	-0.003	-0.001
800	0.001	0.002	0.000	0.074	0.025	0.001	0.039	0.087	0.001	-0.001	0.000
1600	0.000	0.001	-0.004	0.043	0.011	0.000	0.014	0.038	0.001	-0.001	0.000
3200	0.000	0.000	0.003	0.012	0.016	0.002	0.013	0.002	0.000	-0.002	0.000
6400	0.001	0.000	0.000	0.013	0.006	0.001	0.010	0.005	-0.001	-0.001	0.000
	Ratio of sampling standard deviation to estimated standard error										
100	1.094	1.149	1.342	0.512	1.354	1.183	0.990	0.705	1.083	1.090	1.113
200	1.061	1.017	1.155	0.539	1.441	1.100	1.177	1.440	0.999	1.045	1.059
400	1.016	1.022	1.029	1.169	1.125	1.048	1.054	1.191	1.012	0.993	1.021
800	0.982	1.026	1.016	1.039	1.049	0.998	1.015	1.079	0.988	0.990	1.039
1600	0.991	1.023	1.023	1.011	1.035	1.002	1.022	1.015	0.993	1.004	0.977
3200	1.000	1.024	0.991	1.016	0.995	1.038	1.092	1.039	0.964	0.972	0.969
6400	0.995	1.020	1.016	0.977	1.005	1.034	1.064	0.984	0.985	1.017	1.001

Table 7: Bias and standard deviation of ML estimators (Model M-3)

exogenous Markov process.

To sum up, the results from the Monte Carlo experiments suggest that, in the presence of unaccounted changes in the parameters of the transition functions of the regime sequence and in the covariance matrix of the noise, ML produces severely biased estimates, especially for parameters that are associated with transition probabilities. These biases are present even for what are very large sample sizes by the standards of empirical applications. Hypothesis tests based on such ML estimates also have unsatisfactory properties. By contrast, the ML estimator in a correctly specified model that allows for hidden changes in the transition functions and in the noise covariance matrix has very good finite-sample properties and performs almost as well as an (infeasible) ML estimator which utilizes full information on the number and location of such changes.

5 Empirical Application

To illustrate the practical use of the model discussed in Section 3, we analyze the relationship between Argentinian sovereign bond spreads (over U.S. Treasury rates) and output

T	μ_0	μ_1	α_{p0}	β_{p0}	β_{p1}	α_{q0}	β_{q0}	β_{q1}	σ_0	σ_1	ϕ_1
	Mean										
100	0.095	-0.111	-0.009	-0.048	-0.100	0.037	-0.089	-0.087	-0.215	-0.188	-0.279
200	0.107	-0.154	0.033	-0.039	-0.115	-0.033	-0.061	0.008	-0.084	-0.156	-0.210
400	0.026	-0.083	0.009	-0.031	-0.081	0.004	-0.045	0.031	-0.060	-0.113	-0.141
800	0.044	-0.056	-0.048	0.033	-0.017	-0.042	-0.033	0.025	-0.006	-0.072	-0.035
1600	0.003	-0.050	-0.083	0.044	-0.019	-0.037	-0.025	0.036	0.012	-0.080	-0.025
3200	0.023	0.003	0.019	0.004	-0.061	0.011	-0.047	-0.040	-0.036	-0.113	-0.102
6400	0.055	-0.039	-0.008	0.042	-0.031	-0.002	-0.057	-0.005	-0.109	-0.103	-0.049
	Standard deviation										
100	1.060	1.020	1.029	1.108	1.081	0.963	1.049	1.041	1.021	1.056	1.049
200	1.060	1.020	1.029	1.108	1.081	0.963	1.049	1.041	1.021	1.056	1.049
400	1.022	1.019	0.983	1.026	1.179	1.015	1.120	1.047	1.017	1.001	1.029
800	0.985	1.029	0.993	1.037	1.067	0.987	1.131	1.225	0.989	0.995	1.038
1600	0.991	1.027	1.014	1.016	1.028	1.000	1.010	0.996	0.992	1.006	0.977
3200	1.000	1.026	0.987	0.996	0.986	1.036	1.079	1.033	0.964	0.972	0.969
6400	0.994	1.020	1.018	0.971	1.003	1.033	1.064	0.979	0.985	1.019	1.001

Table 8: Empirical moments of t-statistics (Model M-3)

T	μ_0	μ_1	α_{p0}	β_{p0}	β_{p1}	α_{q0}	β_{q0}	β_{q1}	σ_0	σ_1	ϕ_1
	Size										
100	0.054	0.088	0.068	0.075	0.048	0.053	0.045	0.091	0.094	0.111	0.098
200	0.047	0.075	0.061	0.075	0.069	0.059	0.055	0.084	0.075	0.092	0.081
400	0.042	0.055	0.058	0.061	0.066	0.064	0.050	0.055	0.057	0.062	0.069
800	0.040	0.058	0.056	0.038	0.057	0.054	0.057	0.053	0.050	0.064	0.058
1600	0.041	0.070	0.064	0.044	0.051	0.056	0.057	0.052	0.046	0.066	0.048
3200	0.041	0.052	0.048	0.043	0.055	0.062	0.077	0.061	0.052	0.053	0.056
6400	0.038	0.059	0.056	0.035	0.054	0.062	0.063	0.052	0.054	0.059	0.063
	Power										
100	1.000	1.000	0.760	0.038	0.084	0.779	0.103	0.036	1.000	1.000	1.000
200	1.000	1.000	0.962	0.103	0.225	0.978	0.208	0.106	1.000	1.000	1.000
400	1.000	1.000	1.000	0.224	0.416	1.000	0.419	0.260	1.000	1.000	1.000
800	1.000	1.000	1.000	0.559	0.693	1.000	0.695	0.560	1.000	1.000	1.000
1600	1.000	1.000	1.000	0.879	0.927	1.000	0.938	0.864	1.000	1.000	1.000
3200	1.000	1.000	1.000	0.990	0.999	1.000	0.995	0.991	1.000	1.000	1.000
6400	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 9: Empirical size and power of t-tests (Model M-3)

growth. These two variables are generally expected to be negatively correlated since the higher output growth is, the higher is the capacity of a country to repay debt, leading to lower bond spreads. However, if a structural break occurs as a result of a default on sovereign debt, then correlation between the two variables will typically change because the economy may start to grow after the default, as resources are no longer used to service the debt, while the country's risk continues to increase, resulting in higher bond spreads. Such a change in correlation took place following Argentina's default on sovereign debt.

The data set used in our analysis consists of quarterly observations, from 1995:1 to 2019:1, on the J. P. Morgan Emerging Markets Bond Index (EMBI) of dollar-denominated sovereign bonds issued by Argentina (denoted by Y_t) and the growth rate of real GDP (denoted by Z_t). The latter is computed over an eight-quarter horizon so that the long-run trend and strictly seasonal components of real GDP are removed (cf. Hamilton (2018)).

Changes in the correlation between bond spreads and output growth are evident in Figure 4, which shows the correlation coefficient between the EMBI and GDP growth computed over rolling windows (of 16 observations). It is clear that several changes took place following Argentina's default on sovereign debt and the subsequent debt renegotiations, at a time when the economy was fueled by a commodity-prices bonanza. Such changes in correlation cannot be accounted for by the standard model (1)–(3) and are very likely to lead to changes in the parameters of fixed transition functions such as (2)–(3).

The analysis here focuses on whether output growth has predictive content for regime changes associated with the bond spread in the presence of a potential break in the relationship between the two variables that is associated with the sovereign default. We consider two models: Model 1 is the standard model given by (1)–(3); Model 2 is given by (6)–(12) and allows for stochastic discrete breaks in the noise covariance matrix and in the transition functions. In both cases, we set $k = m = 4$. Model 1 is considered because of its popularity in applied work. It should be borne in mind, however, that, even in the absence of breaks in the transition mechanism, the use of such a specification is problematic if Y_t and Z_t are contemporaneously correlated since ML estimation of the parameters of the equation for Y_t alone yields estimates which are inconsistent and severely biased, as the

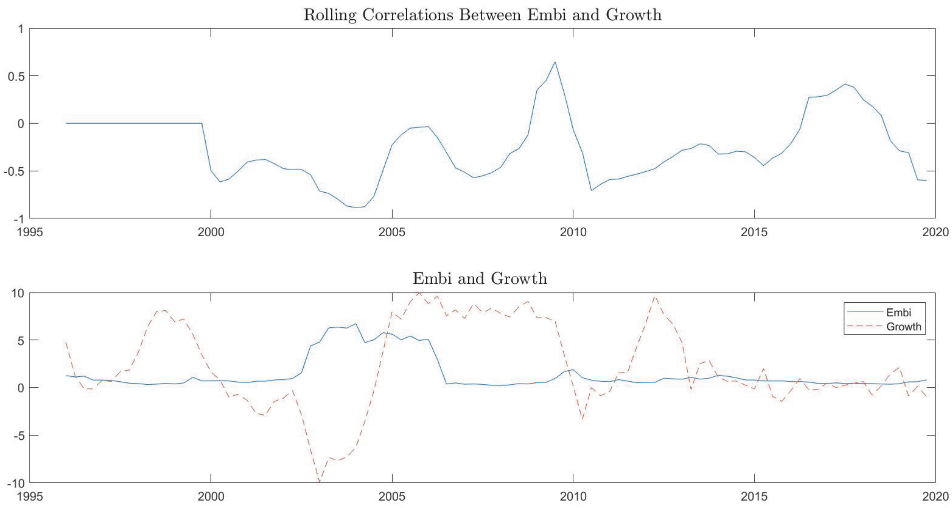

Figure 4: Rolling correlation coefficients (EMBI and real GDP growth)

analysis in Pouzo, Psaradakis, and Sola (2020) demonstrates.

Gaussian ML estimates of the parameters of the two models are reported in Table 10, along with corresponding Huber–White standard errors (in parentheses) to allow for the possibility that it is a pseudo-likelihood that is being maximized. Table 11 contains the values of Ljung–Box statistics (with 4 or 8 autocorrelations) based on the standardized residuals of each equation, along with approximate P-values (in parentheses). There is no significant evidence of residual serial correlation in either of the two models.

The estimated coefficients for the Markov transition functions for Model 1, show that an increase in output growth decreases the probability of remaining in a given regime since both β_q and β_p are negative and would appear to be statistically significant at the 10% level. We note, however, that inference based on these ML estimates is potentially misleading because of the likely inconsistency of the parameter estimator in the presence of changes in the parameters of the Markov transition functions and/or endogeneity of the transition information variable. The results for Model 2 show that the estimated values of β_{q0} and β_{q1} have different signs (although β_{q0} does not differ significantly from zero), so one would expect the evolution of q_t implied by Model 1 and Model 2 to be quite different. The estimates of β_{p0} and β_{p1} are both negative (with β_{p1} being insignificantly different from zero), thus one would not expect the evolution of p_t implied by the two models to be substantially different. Finally, note that the estimated ρ_0 and ρ_1 have opposite signs (ρ_1 being insignificantly different from zero), which suggests that the model is capable of identifying the expected characteristic of pre-default and post-default periods mentioned earlier.

Figure 5 shows plots of the estimated time-varying transition probabilities $\hat{q}_t$ and $\hat{p}_t$ associated with the regimes $\{S_t\}$, as well as the filtered state probabilities $\mathsf{P}(S_t = 1|\mathcal{F}_0^t; \widehat{\boldsymbol{\theta}})$, for Model 1. We also compute the quantity $\hat{\pi}_t := (1 - \hat{q}_t)/(2 - \hat{q}_t - \hat{p}_t)$, which may be thought of as a proxy for the estimated probability that $S_t = 1$, given currently available information on Z_{t-1}, and gives an indication of the contribution of Z_{t-1} in predicting regime changes. It is immediately obvious that the evolution of $\hat{q}_t$ and $\hat{p}_t$ mimics the movements in the output growth rate and does not seem to be particularly informative regarding shifts to the regime that is associated with high EMBI. Furthermore, movements in $\hat{\pi}_t$ are unrelated to movements in $\mathsf{P}(S_t = 1|\mathcal{F}_0^t; \widehat{\boldsymbol{\theta}})$, suggesting that output growth does not have much predictive ability for regime changes.

For Model 2, Figure 6 shows plots of the estimated time-varying transition probabilities $\bar{p}_t = \mathsf{P}(X_t = 0|\mathcal{F}_0^t)p_{0t} + \mathsf{P}(X_t = 1|\mathcal{F}_0^t)p_{1t}$ and $\bar{q}_t = \mathsf{P}(X_t = 0|\mathcal{F}_0^t)q_{0t} + \mathsf{P}(X_t = 1|\mathcal{F}_0^t)q_{1t}$ (evaluated at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$) associated with the regime sequence $\{S_t\}$, the quantity $\bar{\pi}_t := (1 - \bar{q}_t)/(2 - \bar{q}_t - \bar{p}_t)$, and the filtered state probabilities $\mathsf{P}(S_t = 1|\mathcal{F}_0^t; \widehat{\boldsymbol{\theta}})$ and $\mathsf{P}(X_t = 1|\mathcal{F}_0^t; \widehat{\boldsymbol{\theta}})$. The main difference with Model 1 is that both $\bar{q}_t$ and $\bar{\pi}_t$ now seem to be informative about the regime changes associated with EMBI, tracking the movements of $\mathsf{P}(S_t = 1|\mathcal{F}_0^t; \widehat{\boldsymbol{\theta}})$ more closely and preceding its changes on several occasions.

The filtered probabilities $\mathsf{P}(S_t = 1|\mathcal{F}_t; \widehat{\boldsymbol{\theta}})$ associated with high EMBI values are quite similar for both models. For Model 1 (Model 2), the (implied) standard deviation of EMBI in the regime associated with $S_t = 1$, which includes the default on sovereign debt, is approximately 11 times (10 times) larger than it is in the regime associated with $S_t = 0$. For Model 1, the (implied) long-run mean of EMBI is 0.5779 (577.9 points since the series was divided by 1000 prior to estimation) in the regime associated with $S_t = 0$ and 3.583 (3583 points) in the regime associated with $S_t = 1$; the corresponding figures

Model 1		Model 2			
μ_0	0.0686	μ_0	0.0706		
	(0.0179)		(0.0183)		
μ_1	0.4253	μ_1	0.3711		
	(0.1922)		(0.1939)		
α_q	3.4757	α_{q0}	3.4752	ϕ_1	1.0129
	(1.0983)		(1.1852)		(0.0994)
β_q	-0.3172	α_{q1}	3.4752	ϕ_2	-0.0773
	(0.1486)		(1.1852)		(0.0886)
α_p	1.3714	α_{p0}	1.4924	ϕ_3	-0.0584
	(0.4592)		(0.6188)		(0.0474)
β_p	-0.1385	α_{p1}	1.4924	ϕ_4	0.0028
	(0.0807)		(0.6188)		(0.0243)
ϕ_1	1.0284	β_{q0}	0.2644	μ_z	0.2912
	(0.0696)		(0.5485)		(0.2089)
ϕ_2	-0.1002	β_{q1}	-0.5124	ψ_1	1.1104
	(0.0651)		(0.2738)		(0.1370)
ϕ_3	-0.0337	β_{p0}	-6.3976	ψ_2	-0.0673
	(0.0454)		(2.4471)		(0.1420)
ϕ_4	-0.0132	β_{p1}	-0.0748	ψ_3	-0.1040
	(0.0267)		(0.1015)		(0.1486)
σ_0	0.0918	σ_0	0.0960	ψ_4	-0.0616
	(0.0173)		(0.0159)		(0.1253)
σ_1	0.9782	σ_1	0.9644	p_x	0.8099
	(0.1869)		(0.1606)		(0.1014)
		σ_z	1.6611	q_x	0.7796
			(0.1211)		(0.2599)
		ρ_0	-0.5155	ρ_1	0.0796
			(0.1993)		(0.1912)

Table 10: ML estimates of the parameters of Model 1 and Model 2 (with estimated standard errors in parentheses)

	Model 1	Model 2	
	Spread	Spread	Growth
Log-likelihood	1.83035	-172.3706	
Ljung–Box (4)	4.4340	4.4796	0.4733
	(0.3504)	(0.3450)	(0.9760)
Ljung–Box (8)	7.6667	6.8902	1.7571
	(0.4667)	(0.5485)	(0.9721)

Table 11: Residual diagnostics for bond-spread and output-growth equations (with approximate P-values in parentheses)

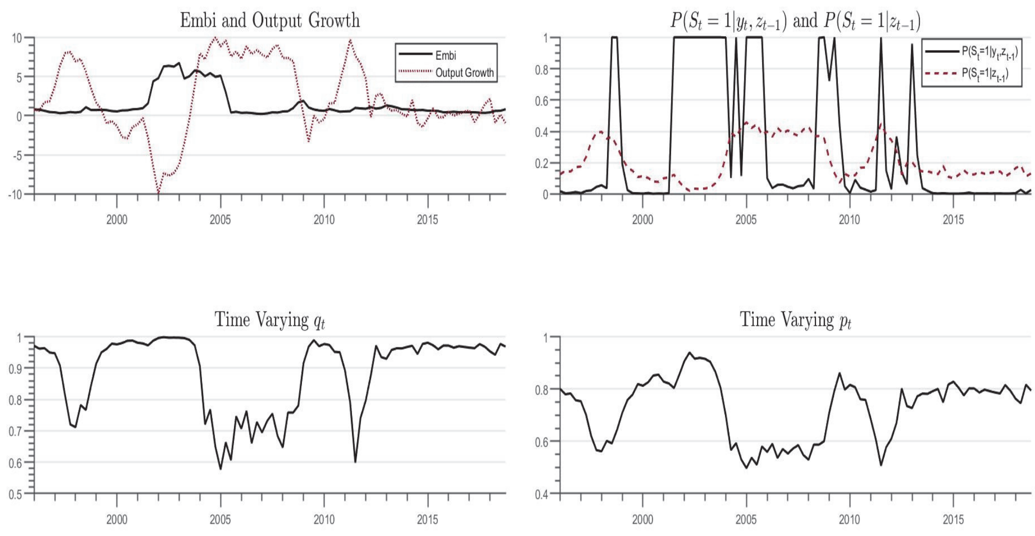

Figure 5: Data; inferred state probabilities $\mathsf{P}(S_t = 1|\mathcal{F}_0^t)$ and $\mathsf{P}(S_t = 1|Z_0, \ldots, Z_{t-1})$; estimated transition probabilities q_t and p_t (based on Model 1)

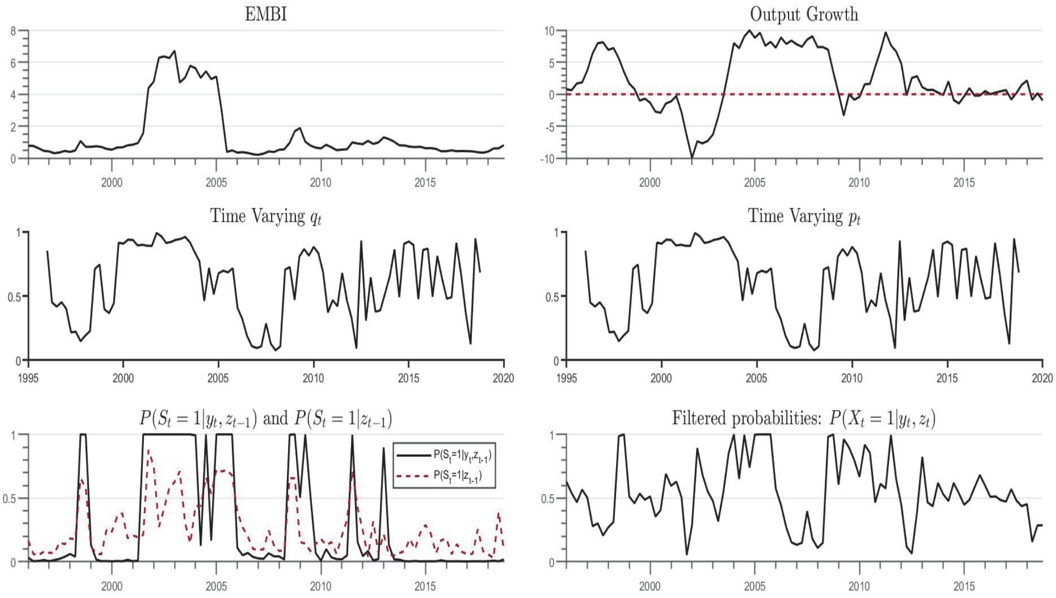

Figure 6: Data; estimated transition probabilities q_t and p_t; inferred state probabilities $\mathsf{P}(S_t = 1|\mathcal{F}_0^t)$, $\mathsf{P}(S_t = 1|Z_0, \ldots, Z_{t-1})$ and $\mathsf{P}(X_t = 1|\mathcal{F}_0^t)$ (based on Model 2)

for Model 2 are 0.6086 (608.6 points) and 3.1991 (3199.1 points), respectively. Finally, the filtered probabilities $\mathsf{P}(X_t = 1|\mathcal{F}_0^t; \widehat{\boldsymbol{\theta}})$ shown in Figure 5 reveal several transitions between a regime associated with weak positive correlation between output growth and EMBI ($X_t = 1$) and a regime that is characterized by negative correlation ($X_t = 0$). This is consistent with the evidence presented in Figure 4.

These results are likely to reflect the fact that the Argentinian government introduced several measures that included renegotiation of the debt structure (starting with the Brady Plan), which were thought to be successful at the time, and allowed a reduction in interest rates during periods in which the rate of growth was positive (albeit decreasing). The situation changed dramatically with the Russian crisis in 1998, which affected the Argentinian economy substantially (and resulted in a change in the correlation between output growth and the EMBI). Even though the government introduced several measures, such as the 'megacanje', or 'megaswap', announced in April 2001, the Argentinian economy entered into recession and it became increasingly more difficult to finance the economy.

This ended in a default on sovereign debt in January 2002. A similar episode occurred in 2008, with the beginning of a recession and a clear increase in the EMBI.

6 Summary

We have discussed a class of Markov-switching models with covariate-dependent transition probability functions in which the parameters associated with the latter are subject to stochastic changes driven by an exogenous Markov process. Such changes are related to changes in the covariance structure between the time series of interest and the variables which drive the evolution of the time-varying transition probabilities. A simulation study has demonstrated the pitfalls of ignoring these changes, pitfalls which include biased parameter estimates and hypothesis tests which exhibit level distortions and low power. The simulations have also shown that the proposed model and related parameter estimator share the same desirable characteristics with a model which incorporates perfect information about the number and location of the breaks associated with the transition functions of the hidden regime sequence. As an illustration of the practical use of the class of models under consideration, we have analyzed the relationship between sovereign bond spreads and output growth in Argentina. The correlation structure between these two variables, and hence the parameters associated with the time-varying transition probabilities of a related regime-switching model, changed as a result of the 2001 economic and financial crisis, a regime shift which the proposed model is well equipped to handle.

We end by noting that there are many other possible applications of the proposed class of models. In economics, these will often involve situations in which the correlation between some variables of interest may have changed over time as a consequence of changes in the credibility of the government. For example, as argued by Giavazzi and Pagano (1990), although a reduction in government expenditure is typically regarded as contractionary in normal times, fiscal contractions may have an expansionary effect if the economy becomes highly indebted.

Acknowledgements

The authors are grateful to Tomas Caravello, Agustin Gutierrez, Ivan Petrella, Demian

Pouzo, Morten Ravn, Ron Smith, an associate editor, and three referees for helpful discussions and/or comments.

References

Alvarez-Plata, P., and M. Schrooten (2006): "The Argentinean currency crisis: a Markov-switching model estimation," *Developing Economies*, 44, 79–91.

Ang, A., and G. Bekaert (2002): "International asset allocation with regime shifts," *Review of Financial Studies*, 15, 1137–1187.

Bauwens, L., and E. Otranto (2016): "Modeling the dependence of conditional correlations on market volatility," *Journal of Business and Economic Statistics*, 34, 254–268.

Bazzi, M., F. Blasques, S. J. Koopman, and A. Lucas (2017): "Time-varying transition probabilities for Markov regime switching models," *Journal of Time Series Analysis*, 38, 458–478.

Bekaert, G., and C. R. Harvey (1995): "Time-varying world market integration," *Journal of Finance*, 50, 403–444.

Bernardi, M., and L. Catania (2019): "Switching generalized autoregressive score copula models with application to systemic risk," *Journal of Applied Econometrics*, 34, 43–65.

Brunetti, C., C. Scotti, R. S. Mariano, and A. H. H. Tan (2008): "Markov switching GARCH models of currency turmoil in southeast Asia," *Emerging Markets Review*, 9, 104–128.

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

Engel, C., and C. S. Hakkio (1996): "The distribution of the exchange rate in the EMS," *International Journal of Finance and Economics*, 1, 55–67.

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

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

Gadea Rivas, M. D., and G. Perez-Quiros (2015): "The failure to predict the Great Recession–A view through the role of credit," *Journal of the European Economic Association*, 13, 534–559.

Ghavidel, F. Z., J. Claesen, and T. Burzykowski (2015): "A nonhomogeneous hidden Markov model for gene mapping based on next-generation sequencing data," *Journal of Computational Biology*, 22, 178–188.

Giavazzi, F., and M. Pagano (1990): "Can severe fiscal contractions be expansionary? Tales of two small European countries," in *NBER Macroeconomics Annual 1990*, ed. by O. J. Blanchard, and S. Fischer, pp. 75–111. MIT Press, Cambridge, Massachusetts.

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

Hall, S. G., Z. Psaradakis, and M. Sola (1997): "Switching error-correction models of house prices in the United Kingdom," *Economic Modelling*, 14, 517–527.

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

——— (2018): "Why you should never use the Hodrick-Prescott filter," *Review of Economics and Statistics*, 100, 831–843.

Hughes, J. P., and P. Guttorp (1994): "A class of stochastic models for relating synoptic atmospheric patterns to regional hydrologic phenomena," *Water Resources Research*, 30, 1535–1546.

Otranto, E. (2005): "The multi-chain Markov switching model," *Journal of Forecasting*, 24, 523–537.

Pelletier, D. (2006): "Regime switching for dynamic correlations," *Journal of Econometrics*, 131, 445–473.

Peria, M. S. M. (2002): "A regime-switching aproach to the study of speculative attacks: a focus on the EMS crisis," *Empirical Economics*, 27, 299–334.

Pouzo, D., Z. Psaradakis, and M. Sola (2020): "Maximum likelihood estimation in Markov regime-switching models with covariate-dependent transition probabilities," Department of Economics, University of California, Berkeley.

Psaradakis, Z., M. Sola, F. Spagnolo, and N. Spagnolo (2013): "Some cautionary results concerning Markov-switching models with time-varying transition probabilities," Department of Economics, Mathematics and Statistics, Birkbeck, University of London.

Ramesh, P., and J. G. Wilpon (1992): "Modeling state durations in hidden Markov models for automatic speech recognition," in *ICASSP-92: 1992 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 1, pp. 381–384. IEEE.

Ravn, M. O., and M. Sola (1999): "Business cycle dynamics: predicting transitions with macrovariables," in *Nonlinear Time Series Analysis of Economic and Financial Data*, ed. by P. Rothman, pp. 231–265. Kluwer Academic Publishers, Dordrecht.

Schaller, H., and S. van Norden (1997): "Regime switching in stock market returns," *Applied Financial Economics*, 7, 177–191.

Simpson, P. W., D. R. Osborn, and M. Sensier (2001): "Modelling business cycle movements in the UK Economy," *Economica*, 68, 243–267.