Típo de documento: Working Paper N° 4

On the Robustness of Mixture Models in the Presence of Hidden Markov Regimes with Covariate-Dependent Transition Probabilities

Autoría Ditelliana: Sola, Martín (*Universidad Torcuato Di Tella. Departamento de Economía*)

Otras autorías: Pouzo, Demian (*Department of Economics, University of California, Berkeley, U.S.A.*); Psaradakis, Zacharias (*Birkbeck Business School, Birkbeck, University of London, U.K.*)

Fecha de publicación: 06/2024

On the Robustness of Mixture Models in the Presence of Hidden Markov Regimes with Covariate-Dependent Transition Probabilities

Demian Pouzo
Department of Economics, University of California, Berkeley, U.S.A.
Email: dpouzo@berkeley.edu

Zacharias Psaradakis
Birkbeck Business School, Birkbeck, University of London, U.K.
Email: z.psaradakis@bbk.ac.uk

Martin Sola
Department of Economics, Universidad Torcuato di Tella, Argentina
Email: msola@utdt.edu

Abstract

We consider general hidden Markov models that may include exogenous covariates and whose discrete-state-space regime sequence has transition probabilities that are functions of observable variables. We show that the parameters of the observation conditional distribution are consistently estimated by quasi-maximum-likelihood even if the Markov dependence of the hidden regime sequence is not taken into account. Some related numerical results are also discussed.

Key words and phrases: Consistency; covariate-dependent transition probabilities; hidden Markov model; mixture model; quasi-maximum-likelihood; misspecified model.

1 Introduction

Consistency and asymptotic normality of least-squares estimators in regression models in the presence of potential model misspecification (e.g., misspecification of the response function or misspecification of the dynamic structure of the errors) are well-established facts (see, e.g., Domowitz and White (1982)). Such fundamental results, together with the related classical work of Huber (1967), underpin a large body of literature exploring the feasibility of drawing valid and meaningful inferences from parametric models that need not necessarily contain the true data-generating process (DGP). Numerous results of this kind have been established for a wide variety of models and estimators, both in static and dynamic settings, ranging from inference procedures based on estimating equations and moment conditions (e.g., Bates and White (1985)) to quasi-maximum-likelihood (QML) procedures for conditional mean and/or conditional variance models (e.g., White (1982, 1994), Levine (1983), Gourieroux et al. (1984), Newey and Steigerwald (1997)).

This paper adds to the literature by presenting another example of robustness with respect to misspecification. Specifically, we consider the case of Hidden Markov Models (HMMs), where observable variables exhibit conditional independence given an underlying unobservable regime sequence (and, possibly, an exogenous covariate sequence), focusing on situations where the dependence structure of the regime sequence is misspecified. In our set-up, the DGP is taken to be a generalized HMM that may include covariates and has a finite number of Markov regimes, but the postulated probability model is a finite mixture model, that is, an HMM with independent, identically distributed (i.i.d.) regimes. By considering the pseudo-true parameter set for the QML estimator in the (misspecified) mixture model, it is shown that the parameters of the conditional distribution of the observable response variables are consistently estimable even if the dependence of the unobservable regime sequence is not taken into account. An important distinguishing feature of our analysis is that it allows the true regime sequence to be a temporally inhomogeneous Markov chain whose transition probabilities are functions of observable variables.

This case holds practical significance given the widespread use of both HMMs and mixture models. HMMs with temporally inhomogeneous regime sequences have found applications in diverse areas such as biology (e.g., Ghavidel et al. (2015)), economics (e.g., Diebold et al. (1994), Engel and Hakkio (1996)), earth sciences (e.g., Hughes et al. (1999)) and engineering (e.g., Ramesh and Wilpon (1992)). Temporally homogenous variants of HMMs and of Markov-switching regression models are also used extensively in economics and finance (e.g., Engel and Hamilton (1990), Rydén et al. (1998), Jeanne and Masson (2000), Bollen et al. (2008)), as well as in biology, computing, engineering and statistics (see Ephraim and Merhav (2002) and references therein). Statistical inference in such models is typically likelihood-based and the properties of QML procedures are, naturally, of much interest. Nevertheless, HMMs are inherently intricate and computationally demanding due to the need to account for the underlying correlated regime sequence and for the dependence of the conditional distribution on the current hidden regime. By demonstrating that it is feasible to use a mixture model — a simpler and computationally less demanding framework — while still estimating consistently the parameters of the conditional distribution of the observations, this paper offers a more accessible avenue for practitioners to follow without sacrificing the accuracy of parameter estimates.

In related recent work, Pouzo et al. (2022) considered the asymptotic properties of the QML estimator in a rich class of models with Markov regimes under general conditions which allow for autoregressive dynamics in the observation sequence, covariate-dependence in the transition probabilities of the hidden regime sequence, and potential model misspecification. The QML estimator was shown to be consistent for the pseudo-true parameter (set) that minimizes the Kullback–Leibler information measure. Unsurprisingly, identifying the possible limit of the QML estimator when the true probability structure of the data does not necessarily lie within the parametric family of distributions specified by the model is not always a feasible task within such a general set-up. This paper provides an answer in the simpler case of switching-regression models, HMMs and related mixture models.

Consistency results for misspecified HMMs (without covariates in the outcome equation) can also be found in Mevel and Finesso (2004) and Douc and Moulines (2012). Unlike our analysis, however, which allows the regime transition probabilities to be time-dependent and driven by observable variables, these papers restrict attention to the case of time-invariant transition mechanisms.

In the next section, we introduce the DGP and statistical model of interest, and consider QML estimation of the parameters of the outcome equation of a misspecified HMM. Section 3 discusses numerical results from a simulation study. Section 4 summarizes and concludes.

2 Framework, Results and Discussion

2.1 DGP and Model

Consider a discrete-time stochastic process $\{(X_t, S_t)\}_{t\geq 0}$ such that $X_t = (Y_t, Z_t, W_t)'$ is an observable variable with values in $\mathbb{X} \subset \mathbb{R}^3$ and S_t is a latent variable with values in $\mathbb{S} := \{s_1, \dots, s_d\} \subset \mathbb{R}$, $d \geq 2$. The variable S_t is viewed as the hidden regime (or state) associated with index t, which is "observable" only indirectly through its effect on X_t. The following assumptions are made about the DGP:

(i) For each $t \geq 1$, the conditional distribution of S_t given $X_0^{t-1} := (X_0, \dots, X_{t-1})$ and $S_0^{t-1} := (S_0, \dots, S_{t-1})$, denoted by $Q_*(Z_{t-1}, S_{t-1}, \cdot)$, depends only on (Z_{t-1}, S_{t-1}) and is such that $Q_*(z, s, s') > 0$ for all $(z, s, s') \in \mathcal{Z} \times \mathbb{S}^2$, where $\mathcal{Z} \subset \mathbb{R}$ is the state space of $\{Z_t\}_{t\geq 0}$.

(ii) For each $t \geq 1$, the conditional distribution of (Y_t, Z_t) given (X_0^{t-1}, S_0^t, W_t), denoted by $P_*(W_t, S_t, Z_{t-1}, \cdot)$, depends only on (W_t, S_t, Z_{t-1}) and is specified via the equations

$$Y_t = \mu_1^*(S_t) + \gamma^*(S_t)W_t + \sigma_1^*(S_t)U_{1,t}, \tag{1}$$

$$Z_t = \mu_2^* + \psi^* Z_{t-1} + \sigma_2^* U_{2,t}, \tag{2}$$

where μ_1^*, γ^* and σ_1^* are known real functions on $\mathbb{S}$, $\sigma_1^*(s), \sigma_2^*(s) \in (0, \infty)$ for all $s \in \mathbb{S}$, and $\psi \in (-1, 1)$. The noise variables $\{(U_{1,t}, U_{2,t})\}_{t\geq 0}$ are i.i.d., independent of

$\{(S_t, W_t)\}_{t\geq 0}$, and have mean zero, covariance matrix $(\omega_{ij})_{i,j=1}^2$, with $\omega_{11} = \omega_{22} = 1$ and $\omega_{12} = \rho^* \in (-1, 1)$, and distribution which is absolutely continuous with respect to some σ-finite Borel measure. In addition, $\{(Z_t, S_t)\}_{t\geq 0}$ is strictly stationary with invariant distribution ν_{ZS}.

(iii) For each $t \geq 1$, the conditional distribution of W_t given (X_0^{t-1}, S_0^t) depends only on W_{t-1}; moreover, $\{W_t\}_{t\geq 0}$ is strictly stationary.

Instead of the Markov-switching structure of the DGP, the researcher's postulated parametric model is a family of finite mixture models (without Markov dependence). Specifically, the model is specified by assuming that the regime variables $\{S_t\}_{t\geq 1}$ are i.i.d. with common distribution

$$Q_{\bar{\vartheta}}(s) = \bar{\vartheta}_s \in (0, 1), \quad s \in \mathbb{S}. \tag{3}$$

In addition, the observable variables $\{Y_t\}_{t\geq 1}$ are assumed to satisfy the equations

$$Y_t = \mu(S_t) + \gamma(S_t)W_t + \sigma(S_t)\varepsilon_t, \quad t \geq 1, \tag{4}$$

where μ, γ and σ are known real functions on $\mathbb{S}$ (such that $\sigma(s) > 0$ for all $s \in \mathbb{S}$) and $\{\varepsilon_t\}_{t\geq 1}$ are i.i.d. random variables, independent of $\{(S_t, W_t)\}_{t\geq 1}$, with ε_1 having the same distribution as $U_{1,1}$. The mixture model defined by (3) and (4) is parameterized by $\theta := (\mu(s), \gamma(s), \sigma(s), \bar{\vartheta}_s)_{s\in\mathbb{S}}$, which is assumed to take values in a compact set $\Theta \subset \mathbb{R}^q$, $q > 1$. We denote by $P_\theta(W_t, S_t, \cdot)$ the conditional distribution of Y_t given (W_t, S_t) that is implied by (4) under each $\theta \in \Theta$; the corresponding conditional density is denoted by $p_\theta(W_t, S_t, \cdot)$.

The DGP has a (generalized) HMM structure in which $\{Y_t\}_{t\geq 0}$ are independent, conditionally on the regime sequence $\{S_t\}_{t\geq 0}$ and an exogenous covariate sequence $\{W_t\}_{t\geq 0}$ (having the Markov property), so that the conditional distribution of Y_t given the regime and covariate sequences depends only on (S_t, W_t). The standard HMM formulation is a special case in which W_t is absent from the outcome equation (1). The inclusion of the exogenous covariate W_t in (1) and (4) allows the study of the causal effect of W on Y under different regimes; these causal effects are captured by γ^* and are estimable via the mixture specification (3)–(4). It is worth noting that, although we focus on scalar responses and covariates for the sake of simplicity, all our results can be extended straightforwardly to cases where $X_t \in \mathbb{X} \subset \mathbb{R}^h$ with $h > 3$. For example, W_t may be a vector of covariates, which may include lagged values of W_t in cases where dynamic causal effects are of interest.

The two most important features of our set-up are that: (a) the true hidden regimes $\{S_t\}_{t\geq 0}$ are a temporally inhomogeneous Markov chain whose transition probabilities depend on the lagged value of the observable variable Z_t; (b) the statistical model is misspecified, in the sense that (P_*, Q_*) is not a member of the family $\{(P_\theta, Q_{\bar{\vartheta}})\colon \theta \in \Theta\}$; this is because the dependence structure of the regime sequence is misspecified. As already discussed in Section 1, this relatively simple set-up is of much practical interest since HMMs with temporally inhomogeneous regime sequences have found many applications. Mixture models with i.i.d. regimes are also popular in many different fields (see McLachlan and Peel (2000), Frühwirth-Schnatter (2006)).

2.2 QML Estimation

Given observations $(X_1, \ldots, X_T)$, $T \geq 1$, the quasi-log-likelihood function for the parameter θ is

$$\theta \mapsto \ell_T(\theta) := T^{-1} \sum_{t=1}^{T} \ln \left(\sum_{s\in\mathbb{S}} \bar{\vartheta}_s p_\theta(W_t, s, Y_t) \right). \tag{5}$$

The QML estimator $\hat{\theta}_T$ of θ is defined as an approximate maximizer of $\ell_T(\theta)$ over Θ, so that

$$\ell_T(\hat{\theta}_T) \geq \sup_{\theta\in\Theta} \ell_T(\theta) - \eta_T,$$

for some sequence $\{\eta_T\}_{T\geq 1} \subset [0, \infty)$ converging to zero.

It is not too onerous to verify that, under assumptions about $(U_{1,1}, U_{2,1})$ and Q_* that are common in the literature (e.g., $(U_{1,1}, U_{2,1})$ being Gaussian and $Q_*(z, s, s') = F(\alpha_{s,s'} + \beta_{s,s'} z)$ for some continuous distribution function F on $\mathbb{R}$ whose support is all of $\mathbb{R}$), the conditions of Pouzo et al. (2022) required for convergence of the QML estimator of θ to a well-defined limit are satisfied. Specifically, let

$$\theta \mapsto H^*(\theta) := E_{\bar{P}_*}\left(\ln \frac{p_*(Y_1|W_1)}{p(Y_1|W_1,\theta)}\right),$$

be the Kullback–Leibler information function, where $p(Y_1|W_1,\theta)$ denotes the conditional density of Y_1 given W_1 induced by $(P_\theta, Q_{\bar{\vartheta}})$ for each $\theta \in \Theta$, $p_*(Y_1|W_1)$ denotes the conditional density of Y_1 given W_1 induced by (P_*, Q_*), and the expectation $E_{\bar{P}_*}(\cdot)$ is with respect to the distribution $\bar{P}_*$ of $\{(X_t, S_t)\}_{t\geq 0}$ induced by (P_*, Q_*). Then, we have

$$\inf_{\theta\in\Theta_*} ||\hat{\theta}_T - \theta|| \to 0 \quad \text{as } T \to \infty, \tag{6}$$

in $\bar{P}_*$-probability, where

$$\Theta_* := \arg\min_{\theta\in\Theta} H^*(\theta), \tag{7}$$

is the pseudo-true parameter (set) and $\|\cdot\|$ denotes the Euclidean norm on $\mathbb{R}^q$ (cf. Theorem 1 of Pouzo et al. (2022)).

A sharper result can be established by considering the pseudo-true parameter Θ_* under the specified DGP. Together with (6)–(7), the following theorem shows, that despite the erroneous treatment of hidden regimes as independent, QML based on the (misspecified) mixture model provides consistent estimators of the true parameters of the outcome equation.

Theorem 1. *The choice $\mu = \mu_1^*$, $\sigma = \sigma_1^*$, $\gamma = \gamma^*$, and $(\bar{\vartheta}_s)_{s\in\mathbb{S}}$ such that $Q_{\bar{\vartheta}} = E_{\nu_{ZS}}[Q_*(Z,S,\cdot)]$ is a pseudo-true parameter, i.e., it minimizes the function*

$$\theta \mapsto E_{\bar{P}_*}\left[\ln \sum_{s\in\mathbb{S}} \frac{Q_{\bar{\vartheta}}(s)}{\sigma(s)} f\left(\frac{Y_1 - \mu(s) - \gamma(s)W_1}{\sigma(s)}\right)\right],$$

where f is the common density of $U_{1,1}$ and ε_1.

Proof. Observe that the Kullback–Leibler information function H^* is proportional to

$$\theta \mapsto -\int_{\mathbb{R}^2} \ln\left(\frac{\sum_{s\in\mathbb{S}} Q_{\bar{\vartheta}}(s)\sigma(s)^{-1} f\left((y-\mu(s)-\gamma(s)w)/\sigma(s)\right)}{f_*(y,w)}\right) f_*(y,w)\, dydw, \tag{8}$$

where

$$(y, w) \mapsto f_*(y, w) = \sum_{s \in \mathbb{S}} \Pr_*(S_1 = s) \sigma_1^*(s)^{-1} f\left((y - \mu_1^*(s) - \gamma^*(s)w)/\sigma_1^*(s)\right),$$

and $\Pr_*$ stands for the true probability over the hidden regimes, given by

$$s \mapsto \Pr_*(S_1 = s) = \int_{\mathbb{R}\times\mathbb{S}} \sum_{s' \in \mathbb{S}} Q_*(z, s', s) \nu_{ZS}(dz, ds').$$

The minimizers of the function in (8) are all θ such that

$$\sum_{s \in \mathbb{S}} Q_{\bar{\vartheta}}(s) \sigma(s)^{-1} f\left((\cdot - \mu(s) - \gamma(s)\cdot)/\sigma(s)\right) = f_*(\cdot).$$

It is straightforward to verify that the equality above holds for $\mu = \mu_1^*$, $\sigma = \sigma_1^*$, $\gamma = \gamma^*$, and $\bar{\vartheta}$ such that $Q_{\bar{\vartheta}}(s) = \Pr_*(S_1 = s)$. □

It is worth remarking that when the minimizer θ_* identified in Theorem 1 is unique (and an interior point of Θ), asymptotic normality of $\sqrt{T}(\hat{\theta}_T - \theta_*)$ may be deduced from the results of Pouzo et al. (2022) under suitable differentiability and moment conditions. These conditions are satisfied, for example, in the case where P_* is Gaussian and Q_* is such that $Q_*(z, s, s') = F(\alpha_{s,s'} + \beta_{s,s'} z)$ for some continuous distribution function F on $\mathbb{R}$ whose support is all of $\mathbb{R}$.

2.3 Discussion

The consistency results in (6)–(7) and in Theorem 1 are quite general, in the sense that they cover misspecified generalized HMMs with temporally inhomogeneous regime sequences and arbitrary observation conditional densities. They imply that dependence of the regimes in such HMMs may be safely ignored as long as the parameters of interest are those of the conditional density of the observations given the regimes and the covariates.

Treating the regimes as an independent sequence simplifies likelihood-based inference compared to the case of correlated Markov regimes. In the latter case, an added difficulty, as demonstrated by Pouzo et al. (2022), is that consistent QML estimation of the true parameter values in a model with Markov regimes having covariate-dependent transition functions typically requires joint analysis of equations such as (1) and (2) with $\rho^* \neq 0$, even if the parameters of interest are only those associated with (1). Furthermore, as pointed out by Hamilton (2016), rich parameterizations of the transition mechanism of the regime sequence may not necessarily be desirable when working with relatively short time series because of legitimate concerns relating to potential overfitting and inaccurate statistical inference. In such cases, parsimonious specifications which provide good approximations to key features of the data — and, in our setting, consistent estimates of the parameters of interest — can be attractive and useful.

We note that, for a class of regime-switching models in which the regime sequence $\{S_t\}$ is a temporally homogeneous, two-state Markov chain, an observation analogous to that implied by Theorem 1 was made by Cho and White (2007). They argued that the parameters of a model for the conditional distribution of the observable variable X_t, given (X_0^{t-1}, S_0^t), can be consistently estimated by QML based on a misspecified version of the model with i.i.d. regimes – and exploited this result to construct a quasi-likelihood-ratio test of the null hypothesis of a single regime against the alternative hypothesis of two regimes. However, as Carter and Steigerwald (2012) demonstrated, consistency of the QML estimator for the true parameters in such a setting does not, in fact, hold if the model and the DGP contain an autoregressive component. This observation remains true in our more general set-up with temporally inhomogeneous hidden regime sequences. Specifically, a result analogous to that in Theorem 1 does not hold when lagged values of Y_t are present as covariates in the outcome equations (1) and (3) (e.g., as in Markov-switching autoregressive models). In this case, misspecification of the dependence structure of the regimes will affect estimation of all the parameters, not just those associated with the transition functions of the regime sequence.

3 Numerical Examples

As a numerical illustration of the results discussed in Section 2, we report here findings from a small Monte Carlo simulation study in which the effect on QML estimators of ignoring Markov dependence of hidden regimes is assessed.

In the experiments, artificial data are generated according to the generalized HMM defined by (1)–(2), with the regimes $\{S_t\}$ forming a Markov chain on $\mathbb{S} = \{1, 2\}$ such that

$$\Pr(S_t = s | S_{t-1} = s, Z_{t-1} = z) = [1 + \exp(-\alpha_s^* - \beta_s^* z)]^{-1}, \quad s \in \{1, 2\}, \quad z \in \mathbb{R},$$

and the covariates $\{W_t\}$ satisfying the autoregressive model

$$W_t = \mu_3^* + \delta^* W_{t-1} + \sigma_3^* U_{3,t}. \tag{9}$$

The noise variables $\{(U_{1,t}, U_{2,t}, U_{3,t})\}$ are i.i.d, Gaussian, independent of $\{S_t\}$, with mean zero and covariance matrix

$$\begin{bmatrix} 1 & \rho^* & \omega^* \\ \rho^* & 1 & 0 \\ \omega^* & 0 & 1 \end{bmatrix}.$$

The parameter values are $\alpha_1^* = \alpha_2^* = 2$, $\beta_1^* = -\beta_2^* = 0.5$, $\mu_1^*(1) = -\mu_1^*(2) = 1$, $\gamma^*(1) = 0.5$, $\gamma^*(2) = 1$, $\sigma_1^*(1) = \sigma_1^*(2) = 1$, $\mu_2^* = \mu_3^* = 0.2$, $\psi^* = \delta^* = 0.8$, $\sigma_2^* = \sigma_3^* = 1$, $\rho^* \in \{0, 0.65\}$, and $\omega^* \in \{0, 0.65\}$.

For each of 1000 samples of size $T \in \{200, 800, 1600, 3200\}$ from this DGP, estimates of the parameters of the outcome equation are obtained by maximizing the quasi-log-likelihood function (5) associated with the mixture model (3)–(4), with $\Pr(S_t = 1) = \bar{\vartheta}$ and $\varepsilon_t \sim \mathcal{N}(0, 1)$. Monte Carlo estimates of the bias of the QML estimators of $\mu(1)$, $\mu(2)$, $\gamma(1)$, $\gamma(2)$, $\sigma(1)$ and $\sigma(2)$ are reported in Table 1. We also report the ratio of the sampling standard deviation of the estimators to estimated standard errors (averaged across replications for each design point), with the latter computed from the observed information matrix, that is, the negative Hessian of the quasi-log-likelihood function.

The results for $\omega^* = 0$ shown in the top panel of Table 1 reveal that, although the estimators of $\mu(1)$ and $\mu(2)$ are somewhat biased in the smallest of the sample sizes considered, finite-sample bias becomes insignificant in the rest of the cases (regardless of the value of the correlation parameter ρ^*), as is to be expected in light of the result in Theorem 1. Furthermore, unless the sample size is small, estimated standard errors are very accurate as approximations to the standard deviation of the QML estimators. This finding is perhaps somewhat surprising since the inverse of the observed information matrix is not necessarily a consistent estimator for the asymptotic covariance matrix of the QML estimator in a misspecified model (cf. Theorem 5 of Pouzo et al. (2022)).

The bottom panel of Table 1 contains results for a DGP with $\omega^* = 0.65$. A non-zero value for the correlation parameter ω^* violates the exogeneity assumption about W_t that is maintained throughout Section 2 (and it is not obvious what the limit point of the QML estimator based on (5) might be in this case). The simulation results show that estimators of the parameters of the outcome equation are significantly biased, even for the largest sample size considered in the simulations. Biases in this case are clearly a consequence of the mixture model being misspecified beyond the assumption of i.i.d. regimes, the additional source of misspecification being the incorrect assumption of uncorrelatedness of the covariate W_t and the noise variable $U_{1,t}$. The results relating to the accuracy of the estimated standard errors are not substantially different from those obtained with $\omega^* = 0$.

As pointed out in Section 2.3, another situation in which ignoring Markov dependence of the regimes is costly involves outcome equations that contain autoregressive dynamics. To demonstrate numerically the difficulties in such a case, 1000 artificial samples of various sizes are generated according to the Markov-switching autoregression

$$Y_t = \mu_1^*(S_t) + \phi^* Y_{t-1} + \sigma_1^*(S_t) U_{1,t}, \tag{10}$$

with $\phi^* = 0.9$. The remaining parameter values and the generating mechanisms of $\{Z_t\}$, $\{S_t\}$ and $\{(U_{1,t}, U_{2,t})\}$ are the same as in earlier simulation experiments. For

Table 1: Bias and Standard Deviation of QML Estimators (HMM)

T	$\mu(1)$	$\mu(2)$	$\gamma(1)$	$\gamma(2)$	$\sigma(1)$	$\sigma(2)$	$\mu(1)$	$\mu(2)$	$\gamma(1)$	$\gamma(2)$	$\sigma(1)$	$\sigma(2)$
	$\rho^* = 0$, $\omega^* = 0$						$\rho^* = 0.65$, $\omega^* = 0$					
	Bias											
200	0.063	-0.021	-0.028	0.019	-0.113	-0.045	0.030	-0.062	-0.011	0.026	–0.084	-0.055
800	0.025	-0.001	-0.006	0.001	-0.027	-0.009	0.021	-0.009	-0.002	0.001	-0.026	-0.011
1600	0.010	0.001	-0.001	0.001	-0.014	-0.004	0.012	-0.003	0.000	0.000	-0.014	-0.007
3200	0.001	0.000	0.000	0.000	-0.005	-0.002	0.002	-0.004	-0.001	0.002	-0.005	-0.003
	Standard Deviation / Standard Error											
200	1.527	1.360	1.492	1.285	1.562	1.401	1.451	1.342	1.488	1.282	1.450	1.273
800	1.155	1.082	1.136	1.084	1.132	1.036	1.156	1.097	1.109	1.053	1.121	1.065
1600	1.095	1.102	1.055	1.031	1.048	1.067	1.083	1.089	1.066	1.006	1.022	1.025
3200	1.070	1.060	1.059	1.017	1.075	1.033	1.075	1.062	1.035	0.997	1.043	1.010
	$\rho^* = 0$, $\omega^* = 0.65$						$\rho^* = 0.65$, $\omega^* = 0.65$					
	Bias											
200	-0.193	-0.265	0.229	0.257	-0.164	-0.119	-0.201	-0.273	0.229	0.259	-0.156	-0.120
800	-0.243	-0.243	0.239	0.240	-0.089	-0.089	-0.225	-0.237	0.240	0.238	-0.096	-0.087
1600	-0.235	-0.242	0.235	0.238	-0.086	-0.086	-0.228	-0.238	0.235	0.236	-0.088	-0.084
3200	-0.233	-0.236	0.235	0.234	-0.083	-0.082	-0.231	-0.238	0.235	0.235	-0.083	-0.082
	Standard Deviation / Standard Error											
200	1.402	1.310	1.324	1.220	1.540	1.210	1.246	1.313	1.377	1.229	1.422	1.182
800	1.112	1.118	1.014	1.076	1.109	1.018	1.075	1.205	1.069	1.100	1.115	1.092
1600	1.025	1.162	1.034	1.059	1.041	1.069	1.034	1.144	1.027	1.086	-1.039	1.089
3200	1.069	1.197	0.994	1.058	1.014	1.060	1.072	1.179	0.981	1.056	1.082	1.065

Table 2: Bias and Standard Deviation of QML Estimators (Markov-Switching Autoregressive Model)

T	$\mu(1)$	$\mu(2)$	$\sigma(1)$	$\sigma(2)$	ϕ	$\mu(1)$	$\mu(2)$	$\sigma(1)$	$\sigma(2)$	ϕ
	$\rho^* = 0$					$\rho^* = 0.8$				
	Bias									
200	0.424	-0.288	-0.209	0.008	0.060	0.244	-0.142	-0.136	-0.004	0.052
800	0.263	-0.333	-0.062	0.036	0.068	0.140	-0.240	-0.017	0.010	0.059
1600	0.186	-0.356	-0.006	0.046	0.069	0.128	-0.271	0.008	0.027	0.060
3200	0.135	-0.347	0.018	0.048	0.069	0.125	-0.285	0.015	0.033	0.061
	Standard Deviation / Standard Error									
200	1.320	1.469	1.445	1.461	1.025	1.422	1.527	1.429	1.441	0.975
800	1.318	1.359	1.345	1.332	0.855	1.082	1.085	1.161	1.145	0.817
1600	1.213	1.243	1.239	1.320	0.881	1.179	1.153	1.170	1.150	0.838
3200	1.165	1.220	1.093	1.224	1.003	1.126	1.119	1.089	1.167	0.831

each artificial sample, the parameters of the regime-switching autoregressive model

$$Y_t = \mu(S_t) + \phi Y_{t-1} + \sigma(S_t)\varepsilon_t, \tag{11}$$

are estimated by maximizing the quasi-log-likelihood function associated with it under the assumption that the regime variables $\{S_t\}$ are i.i.d., with $\Pr(S_t = 1) = \bar{\vartheta}$, and the noise variables $\{\varepsilon_t\}$ are i.i.d., independent of $\{S_t\}$, with $\varepsilon_t \sim \mathcal{N}(0, 1)$.

The Monte Carlo results reported in Table 2 reveal substantial finite-sample bias in the case of the QML estimators of the intercepts $\mu(1)$ and $\mu(2)$. The QML estimators of $\sigma(1)$, $\sigma(2)$ and ϕ generally exhibit little bias, which may be partly due to the fact that the simulation design is such that the values of ϕ^* and σ_1^* are the same regardless of the realized regime. Unlike the HMM case considered before, estimated standard errors obtained from the observed information matrix tend to be inaccurate as approximations to the finite-sample standard deviation of the QML estimators in the autoregressive model, even for those parameters that are estimated with little bias. We note that qualitatively similar results are obtained when, in addition to Y_{t-1}, an exogenous covariate W_t, generated as in (9), is included in the right-hand side of both (10) and (11).

4 Conclusion

In this paper, we have considered QML estimation of the parameters of a generalized HMM with exogenous covariates and a finite hidden state space. A distinguishing feature of our approach is that it allows the regime sequence to be a temporally inhomogeneous Markov chain with covariate-dependent transition probabilities. It has been shown that a mixture model with independent regimes is robust in the presence of correlated Markov regimes, in the sense that the parameters of the outcome equation can be consistently estimated by maximizing the quasi-likelihood function associated with the misspecified mixture model.

One possible application of our main result is to exploit it to construct tests for the number of regimes in HMMs with covariate-dependent transition probabilities, adopting a QML-based approach analogous to that of Cho and White (2007). As is well known, such testing problems are non-standard and typically involve unidentifiable nuisance parameters, parameters that lie on the boundary of the parameter space, singularity of the Fisher information matrix, and non-quadratic approximations to the log-likelihood function.

References

Bates, C. and H. White (1985). A unified theory of consistent estimation for parametric models. *Econometric Theory 1*, 151–178.

Bollen, N. P. B., S. F. Gray, and R. E. Whaley (2008). Regime switching in foreign exchange rates: Evidence from currency option prices. *Journal of Econometrics 94*, 239–276.

Carter, A. V. and D. G. Steigerwald (2012). Testing for regime switching: A comment. *Econometrica 80*, 1809–1812.

Cho, J. S. and H. White (2007). Testing for regime switching. *Econometrica 75*, 1671–1720.

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

Domowitz, I. and H. White (1982). Misspecified models with dependent observations. *Journal of Econometrics 20*, 35–58.

Douc, R. and E. Moulines (2012). Asymptotic properties of the maximum likelihood estimation in misspecified hidden Markov models. *Annals of Statistics 40*, 2697–2732.

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.

Engel, C. and J. D. Hamilton (1990). Long swings in the Dollar: Are they in the data and do markets know it? *American Economic Review 80*, 689–713.

Ephraim, Y. and N. Merhav (2002). Hidden Markov processes. *IEEE Transactions on Information Theory 48*, 1518–1569.

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

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.

Gourieroux, C., A. Monfort, and A. Trognon (1984). Pseudo maximum likelihood methods: Theory. *Econometrica 52*, 681–700.

Hamilton, J. D. (2016). Macroeconomic regimes and regime shifts. In J. B. Taylor and H. Uhlig (Eds.), *Handbook of Macroeconomics, Vol. 2A*, pp. 163–201. Amsterdam: North-Holland.

Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In L. M. Le Cam and J. Neyman (Eds.), *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Volume 1, pp. 221–233. University of California Press, Berkeley, CA.

Hughes, J. P., P. Guttorp, and S. P. Charles (1999). A non-homogeneous hidden Markov model for precipitation occurrence. *Applied Statistics 48*, 15–30.

Jeanne, O. and P. Masson (2000). Currency crises, sunspots and Markov-switching models. *Journal of International Economics 50*, 327–350.

Levine, D. (1983). A remark on serial correlation in maximum likelihood. *Journal of Econometrics 23*, 337–342.

McLachlan, G. J. and D. Peel (2000). *Finite Mixture Models.* New York: Wiley.

Mevel, L. and L. Finesso (2004). Asymptotical statistics of misspecified hidden Markov models. *IEEE Transactions on Automatic Control 49*, 1123–1132.

Newey, W. K. and D. G. Steigerwald (1997). Asymptotic bias for quasi-maximum-likelihood estimators in conditional heteroskedasticity models. *Econometrica 65*, 587–599.

Pouzo, D., Z. Psaradakis, and M. Sola (2022). Maximum likelihood estimation in Markov regime-switching models with covariate-dependent transition probabilities. *Econometrica 90*, 1681–1710.

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*, Volume 1, pp. 381–384. IEEE.

Rydén, T., T. Teräsvirta, and S. Åsbrink (1998). Stylized facts of daily returns series and the hidden Markov model. *Journal of Applied Econometrics 13*, 217–244.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica 50*, 1–25.

White, H. (1994). *Estimation, Inference and Specification Analysis.* Cambridge, U.K.: Cambridge University Press.