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

Autorías: Pouzo, Demian (*University of California, Berkeley*); Psaradakis, Zacharias (*University of London*); Sola, Martín (*Universidad Torcuato Di Tella. Departamento de Economía)*

Fecha de publicación: 16/06/2025

Publicado originalmente en: Econometric Theory (e-ISSN: 1469-4360)

¿Cómo citar este trabajo?

Pouzo, D., Psaradakis, Z., & Sola, M. (2025). ON THE ROBUSTNESS OF MIXTURE MODELS IN THE PRESENCE OF HIDDEN MARKOV REGIMES WITH COVARIATE-DEPENDENT TRANSITION PROBABILITIES. Econometric Theory, 1–15. https://doi.org/10.1017/s0266466625100017

Econometric Theory, **0**, 2025, 1–15.
doi:10.1017/S0266466625100017

ON THE ROBUSTNESS OF MIXTURE MODELS IN THE PRESENCE OF HIDDEN MARKOV REGIMES WITH COVARIATE-DEPENDENT TRANSITION PROBABILITIES

Demian Pouzo
University of California

Zacharias Psaradakis
Birkbeck, University of London

Martin Sola
Universidad Torcuato di Tella

This article studies the robustness of quasi-maximum-likelihood estimation in hidden Markov models when the regime-switching structure is misspecified. Specifically, we examine the case where the data-generating process features a hidden Markov regime sequence with covariate-dependent transition probabilities, but estimation proceeds under a simplified mixture model that assumes regimes are independent and identically distributed. We show that the parameters governing the conditional distribution of the observables can still be consistently estimated under this misspecification, provided certain regularity conditions hold. Our results highlight a practical benefit of using computationally simpler mixture models in settings where regime dependence is complex or difficult to model directly.

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

The authors wish to thank Patrik Guggenberger, Peter Phillips, and two referees for their helpful comments and suggestions. For the purposes of open access, the corresponding author has applied a CC-BY public copyright licence to any accepted manuscript version arising from this submission. Address correspondence to Zacharias Psaradakis, Birkbeck Business School, Birkbeck, University of London, London, UK, e-mail: z.psaradakis@bbk.ac.uk

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, conditional variance, and conditional quantile models (e.g., Gourieroux, Monfort, and Trognon, 1984; Komunjer, 2005; Levine, 1983; Newey and Steigerwald, 1997; White, 1982, 1994).

This article 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, exogenous covariate sequences), 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. A condition on the tail behavior of the characteristic function of the (standardized) conditional distribution of the observable responses is also provided under which the pseudo-true parameter for the QML estimator is a singleton set. An important distinguishing feature of our analysis is that the true regime sequence is allowed 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, Claesen, and Burzykowski, 2015), economics (e.g., Diebold, Lee, and Weinbach, 1994; Engel and Hakkio, 1996), earth sciences (e.g., Hughes, Guttorp, and Charles, 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., Bollen, Gray, and Whaley, 2008; Engel and Hamilton, 1990; Jeanne and Masson, 2000; Rydén, Teräsvirta, and Åsbrink, 1998), 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 article offers a more accessible avenue for practitioners to follow without sacrificing the accuracy of parameter estimates.

In related recent work, Pouzo, Psaradakis, and Sola (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 in such a general set-up. This article provides an answer in the simpler case of switching-regression models, HMMs and related mixture models. Consistency results for misspecified pure HMMs (with no covariates in the outcome equation) can also be found in Mevel and Finesso (2004) and Douc and Moulines (2012). Unlike our analysis, 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 generalized 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 $\mathbb{S} := \{1, 2, \ldots, d\} \subset \mathbb{N}$ for some $d \geq 2$. 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:

1. For each $t \geq 1$, the conditional distribution of S_t given $X_0^{t-1} := (X_0, \ldots, X_{t-1})$ and $S_0^{t-1} := (S_0, \ldots, S_{t-1})$, denoted by $Q_*(\cdot|Z_{t-1}, S_{t-1})$, depends only on (Z_{t-1}, S_{t-1}) and is such that $Q_*(s|z, s') > 0$ for all $(s, s', z) \in \mathbb{S}^2 \times \mathcal{Z}$, where $\mathcal{Z} \subset \mathbb{R}$ is the state space of Z_t.
2. For each $t \geq 1$, the conditional distribution of Z_t given (X_0^{t-1}, S_0^t) depends only on Z_{t-1}; furthermore, $\{(Z_t, S_t)\}_{t\geq 0}$ is strictly stationary with invariant distribution ν_{ZS}.
3. For each $t \geq 1$, the conditional distribution of Y_t given (X_0^{t-1}, S_0^t, W_t) depends only on (W_t, S_t) and is specified via the equation

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

where μ_1^*, γ^*, and $\sigma_1^* > 0$ are known real functions on $\mathbb{S}$; the noise variables $\{U_{1,t}\}_{t\geq 0}$ are i.i.d., independent of $\{S_t\}_{t\geq 0}$, with mean zero, variance one, and density f.

https://doi.org/10.1017/S0266466625100017 Published online by Cambridge University Press

4. 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}, and W_t is independent of (Z_{t-1}, S_{t-1}); furthermore, $\{W_t\}_{t\geq 0}$ is strictly exogenous in (1) and strictly stationary with invariant distribution ν_W.

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 distribution

$$Q_{\bar{\vartheta}}(s) = \bar{\vartheta}_s \in (0,1), \quad s \in \mathbb{S}. \tag{2}$$

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

$$Y_t = \mu(S_t) + \gamma(S_t) W_t + \sigma(S_t)\varepsilon_t, \quad t \geq 1, \tag{3}$$

where μ, γ, and $\sigma > 0$ are known real functions on $\mathbb{S}$ and $\{\varepsilon_t\}_{t\geq 1}$ are i.i.d. random variables, independent of $\{(S_t, W_t)\}_{t\geq 1}$, such that ε_1 has the same density f as $U_{1,1}$. The mixture model defined by (2) and (3) is parameterized by $\theta := (\pi(s), \bar{\vartheta}_s)_{s\in\mathbb{S}}$, with $\pi(s) := (\mu(s), \gamma(s), \sigma(s))$, which is assumed to take values in a compact set $\Theta \subset \mathbb{R}^q$, $q > 1$. We denote by $P_\pi(\cdot|W_t, S_t)$ the conditional distribution of Y_t given (W_t, S_t) that is implied by (3); the corresponding conditional density is denoted by $p_\pi(\cdot|W_t, S_t)$.

Key aspects of our set-up: First, 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 inclusion of the exogenous covariate W_t in (1) and (3) allows the study of the causal effect of W on Y under different regimes; this causal effect is captured by γ^* and is estimable via the mixture specification (2) and (3). Exogeneity of W (Assumption 4 above) is essential for the results discussed in Section 2.2 to hold and, hence, for consistent estimation of the causal effect γ^* under the (erroneous) assumption of independent regimes.[1] Second, 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. The sequence $\{Z_t\}_{t\geq 0}$ has the Markov property and is not required to be exogenous, in the sense that Z_t may be contemporaneously correlated with $U_{1,t}$. Third, the statistical model is misspecified, in the sense that the DGP is not a member of the family $\{(P_\pi, Q_{\bar{\vartheta}}) : (\pi, \bar{\vartheta}) \in \Theta\}$; this is because the dynamic structure of the regimes 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 widely used in many different fields (see McLachlan and Peel, 2000 and Frühwirth-Schnatter, 2006), including economics and econometrics (see Compiani and Kitamura, 2016).

[1] The standard HMM formulation is a special case in which W_t is absent from the outcome equation (1).

https://doi.org/10.1017/S0266466625100017 Published online by Cambridge University Press

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. Similarly, Z_t may be a vector of information variables that affect the dynamic profile of the regime transition probabilities, whose generating mechanism has a finite-order autoregressive structure.

2.2. QML Estimation

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

$$\theta \mapsto \ell_T(\theta) := T^{-1} \sum_{t=1}^{T} \ln\left(\sum_{s \in \mathbb{S}} \bar{\vartheta}_s p_\pi(Y_t | W_t, s)\right). \tag{4}$$

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 $\{\eta_T\}_{T \geq 1} \subset \mathbb{R}_+$ converging to zero.

It is not too onerous to verify that, under assumptions that are common in the literature (e.g., Gaussianity of $U_{1,1}$ and $Q_*(s|z,s') = G(\alpha_{s,s'} + \beta_{s,s'} z)$ for some continuous distribution function G 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

$$\theta \mapsto H^*(\theta) := \mathbb{E}_{\bar{P}_*}\left[\ln\left(\frac{p_*(Y_1|W_1)}{p_\theta(Y_1|W_1)}\right)\right]$$

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

$$\inf_{\theta \in \Theta_*} ||\hat{\theta}_T - \theta|| \to 0 \quad \text{as } T \to \infty, \tag{5}$$

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

$$\Theta_* := \arg\min_{\theta \in \Theta} H^*(\theta) \tag{6}$$

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

A sharper result can be established by considering the pseudo-true parameter Θ_* under the specified DGP. Together with (5) and (6), the following theorem

https://doi.org/10.1017/S0266466625100017 Published online by Cambridge University Press

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* $\bar{\vartheta}_s^* = \mathbb{E}_{\nu_{ZS}}[Q_*(s|Z,S)]$ *for all* $s \in \mathbb{S}$ *is a pseudo-true parameter, that is, it maximizes the function*

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

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

$$\theta \mapsto -\int_{\mathbb{R}^2} \ln\left(\frac{\sum_{s\in\mathbb{S}} \bar{\vartheta}_s \sigma(s)^{-1} f((y-\mu(s)-\gamma(s)w)/\sigma(s))}{p_*(y|w)}\right) p_*(y|w) dy \nu_W(dw), \tag{7}$$

where

$$(y,w) \mapsto p_*(y|w) = \sum_{s\in\mathbb{S}} \Pr_*(S_1 = s \mid W_1 = w) \sigma_1^*(s)^{-1} f\left((y - \mu_1^*(s) - \gamma^*(s)w)/\sigma_1^*(s)\right).$$

Under our assumptions about (W_t, Z_{t-1}), $\Pr_*(S_1 = \cdot | W_1 = w) = \Pr_*(S_1 = \cdot)$, where $\Pr_*$ stands for the true probability over the hidden regimes, given by

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

The minimizers of the function in (7) are all θ such that

$$\sum_{s\in\mathbb{S}} \bar{\vartheta}_s \sigma(s)^{-1} f((\cdot - \mu(s) - \gamma(s)\cdot)/\sigma(s)) = p_*(\cdot|\cdot).$$

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

Theorem 1 establishes that the true parameters $\pi^*(s) := (\mu_1^*(s), \sigma_1^*(s), \gamma^*(s))$, $s \in \mathbb{S}$, associated with the observation equation (1), together with $(\bar{\vartheta}_s^*)_{s\in\mathbb{S}}$, minimize the Kullback–Leibler information $H^*(\theta)$. The corollary that follows shows that this minimizer is unique, as long as $\theta_* := (\pi^*(s), \bar{\vartheta}_s^*)_{s\in\mathbb{S}}$ is identified in Θ. In the present context, θ_* is said to be identified in Θ if, for any $\theta \in \Theta$ such that $p_\theta(\cdot|\cdot) = p_{\theta_*}(\cdot|\cdot)$ $\bar{P}_*$-almost surely, $\theta = \theta_*$ up to permutations.[2]

[2] Formally, $(\pi(\mathfrak{p}[s]), \bar{\vartheta}_{\mathfrak{p}[s]}) = (\pi^*(s), \bar{\vartheta}_s^*)$ for all $s \in \mathbb{S}$ and any permutation $\mathfrak{p} : \mathbb{S} \to \mathbb{S}$. The qualifier "up to permutations" reflects the fact that the problem remains unchanged if the indices of the regimes are permuted.

https://doi.org/10.1017/S0266466625100017 Published online by Cambridge University Press

COROLLARY 1. *If θ_* is identified in Θ, then it is the unique maximizer (up to permutations) of*

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

Proof. By Theorem 1, it suffices to show that

$$\int_{\mathbb{R}^2}\ln(p_*(y|w))\,p_*(y|w)dy\nu_W(dw) > \int_{\mathbb{R}^2}\ln(p_\theta(y|w))\,p_*(y|w)dy\nu_W(dw),$$

for any $\theta \in \Theta$ which is not equal to a permutation of θ_*. Since θ_* is identified, for any such θ, $p_\theta(\cdot|\cdot) \neq p_{\theta_*}(\cdot|\cdot)$ with positive probability under $\bar{P}_*$. Therefore, the strict inequality above follows from the (strict) Jensen inequality. □

To provide some (non-technical) intuition behind Theorem 1 and its corollary, recall that when the misspecified mixture model is fitted to data using maximum likelihood techniques, the objective function which is maximized is a quasi-likelihood. The resulting QML estimator is an estimator for the parameters which make the model's implied conditional distribution of the response as close as possible—measured in Kullback–Leibler divergence—to the true conditional distribution, even when the model is misspecified. Despite the fact that the researcher ignores dependence of the regimes, the conditional distribution of the response variable given the covariates remains a mixture of distributions under both the true and misspecified models. This structural similarity allows the mixture model to match the key features of the conditional distribution of interest correctly and, thus, the QML estimator consistently recovers the parameters of the outcome equation. This result relies heavily on two key conditions: stationarity of the regimes (Assumption 2) and exogeneity of the covariates (Assumption 4). The former stationarity condition is crucial because it ensures stability of the marginal distribution of the regimes, which the mixture model tries to fit. Without stationarity, the limiting object of the QML estimation procedure could vary over time, and consistency would break down. The latter exogeneity condition is fundamental because it guarantees that the covariates do not "carry" information about future states or future shocks into the noise of the outcome equation. If exogeneity failed, the conditional distribution of the response would not be properly captured by the simple mixture model and the QML estimator would be biased and inconsistent for the parameters of interest.

We conclude by remarking that when the minimizer θ_* of the Kullback–Leibler information function is identified in Θ (and belongs to the interior 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 f is Gaussian and $Q_*(s|z,s') = G(\alpha_{s,s'} + \beta_{s,s'}z)$ for some continuous distribution function G on $\mathbb{R}$ whose support is all of $\mathbb{R}$. In the next subsection, we discuss a sufficient condition for the high-level identifiability requirement of Corollary 1 and show that this condition holds in some commonly used models.

2.3. Discussion on the Identifiability Condition

Identifiability of mixture models has been studied extensively in the literature following the original contribution of Teicher (1963), who established identifiability of finite mixtures of distributions such as the one-dimensional Gaussian and gamma. Yakowitz and Spragins (1968) gave a necessary and sufficient condition for identifiability, which holds, for example, in the case of finite mixtures of multivariate Gaussian distributions. This condition was exploited by Holzmann, Munk, and Stratman (2004) and Holzmann, Munk, and Gneiting (2006) to provide sufficient low-level identifiability conditions based on the tail behavior of the characteristic function of the component distributions.[3]

Using the approach of Holzmann et al. (2004) and Holzmann et al. (2006), we now present a low-level condition, based only on features of f, which is sufficient for the identifiability of θ_* required in Corollary 1.

LEMMA 1. *Let φ be the characteristic function associated with f. If, for any $a_1 > a_2$,*

$$\lim_{\tau\to\infty}\frac{\varphi(a_1\tau)}{\varphi(a_2\tau)}=0,$$

then θ_* is identified in Θ.

Proof. By the classical result of Yakowitz and Spragins (1968), to establish unique identifiability, it suffices to show that $\left\{\frac{1}{\sigma(s)}f\left(\frac{y-m(s,w)}{\sigma(s)}\right)\right\}_{s\in\mathbb{S}}$ are linearly independent, where $m(s,w) := \mu(s)+\gamma(s)w$. To this end, observe that

$$\int_{\mathbb{R}} e^{\mathrm{i}\tau y}\frac{1}{\sigma(s)}f\left(\frac{y-m(s,w)}{\sigma(s)}\right)dy = e^{\mathrm{i}\tau m(s,w)}\int_{\mathbb{R}} e^{\mathrm{i}\tau u\sigma(s)}f(u)\,du = e^{\mathrm{i}\tau m(s,w)}\varphi(\tau\sigma(s)).$$

Hence, for any $\lambda_1,\ldots,\lambda_d$ in $\mathbb{R}$, $\sum_{s\in\mathbb{S}}\lambda_s\frac{1}{\sigma(s)}f\left(\frac{y-m(s,w)}{\sigma(s)}\right)=0$ implies

$$\sum_{s\in\mathbb{S}}\lambda_s e^{\mathrm{i}\tau m(s,w)}\varphi(\tau\sigma(s))=0, \tag{8}$$

for any $\tau\in\mathbb{R}$ and any $w\in\mathbb{R}$. Without loss of generality, let $\sigma(1)\le\sigma(2)\le\cdots\le\sigma(d)$, with $m\in\mathbb{S}$ being such that $\sigma(1)=\cdots=\sigma(m)<\sigma(m+1)$. Then, (8) is equivalent to

$$\lambda_1+\sum_{s=2}^{m}\lambda_s e^{\mathrm{i}\tau[m(s,w)-m(1,w)]}+\sum_{s=m+1}^{d}\lambda_s e^{\mathrm{i}\tau[m(s,w)-m(1,w)]}\frac{\varphi(\tau\sigma(s))}{\varphi(\tau\sigma(1))}=0. \tag{9}$$

[3] A review of related results for parametric and nonparameteric models that incorporate mixture distributions can be found in Compiani and Kitamura (2016).

https://doi.org/10.1017/S0266466625100017 Published online by Cambridge University Press

Since $\sigma(s) > \sigma(1)$ for any $s \in \{m+1,\ldots,d\}$ and $e^{\mathrm{i}\tau[m(s,w)-m(1,w)]}$ is uniformly bounded in τ, it follows by our assumption that $\sum_{s=m+1}^{d} \lambda_s e^{\mathrm{i}\tau[m(s,w)-m(1,w)]} \frac{\varphi(\tau\sigma(s))}{\varphi(\tau\sigma(1))} \to 0$ as $\tau \to \infty$. This result readily implies that $n^{-1}\sum_{l=1}^{n}\sum_{s=m+1}^{d} \lambda_s e^{\mathrm{i}lu_0[m(s,w)-m(1,w)]} \frac{\varphi(lu_0\sigma(s))}{\varphi(lu_0\sigma(1))} \to 0$ as $n \to \infty$, for any $u_0 > 0$. Regarding the term $\sum_{s=2}^{m} \lambda_s e^{\mathrm{i}\tau[m(s,w)-m(1,w)]}$, observe that $m(s,w) \neq m(1,w)$ for any $s \in \{1,\ldots,m\}$—otherwise, since $\sigma(s) = \sigma(1)$, the regimes associated with $S_t = s$ and $S_t = 1$ would be identical rather than distinct. Thus, by choosing $\tau = lu_0$, where $l \in \mathbb{N}$ and $u_0 \in \mathbb{R}_+$ is such that $u_0[m(s,w)-m(1,w)] \in (-\pi,\pi)$ for all $s \in \{1,\ldots,m\}$, it follows by Holzmann et al. (2004, Lem. 2.1) that $n^{-1}\sum_{l=1}^{n}\sum_{s=2}^{m} \lambda_s e^{\mathrm{i}lu_0[m(s,w)-m(1,w)]} \to 0$ as $n \to \infty$. By these two results and (9),

$$\lambda_1 = -\lim_{n\to\infty} n^{-1}\sum_{l=1}^{n}\left(\sum_{s=2}^{m}\lambda_s e^{\mathrm{i}lu_0[m(s,w)-m(1,w)]} + \sum_{s=m+1}^{d}\lambda_s e^{\mathrm{i}lu_0[m(s,w)-m(1,w)]}\frac{\varphi(lu_0\sigma(s))}{\varphi(lu_0\sigma(1))}\right)$$
$$= 0.$$

By iterating on this procedure, it follows that $\lambda_1 = \lambda_2 = \cdots = \lambda_d = 0$, thereby establishing the desired result. □

As an example, consider what is, arguably, the most widely used class of mixture models, namely those in which f is Gaussian. In this case, $\varphi(\tau) = e^{-\tau^2/2}$, $\tau \in \mathbb{R}$, and $\varphi(a_1\tau)/\varphi(a_2\tau) = e^{-(a_1^2-a_2^2)\tau^2/2}$, $a_1 > a_2$, so the condition of Lemma 1 is satisfied. Thus, in the Gaussian case, the QML estimator of the parameters of the mixture model (2) and (3) converges, in $\bar{P}_*$-probability, to θ_*. This result remains valid for non-Gaussian distributions, including distributions with heavy tails (and finite variance). For instance, the result holds if f is the density of a (rescaled) Student-t distribution with degrees of freedom $\upsilon > 2$ (see Holzmann et al. 2006, Ex. 1).

2.4. Discussion on the Main Theorem

The consistency results in (5) and (6) 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. It is important to note, however, that care should be taken in estimating the asymptotic covariance matrix of the QML estimator since the inverse of the observed information matrix is not necessarily a consistent estimator in a misspecified model. Consistent estimation in this case typically requires the use of an empirical sandwich estimator that does not rely on the information matrix equality (cf. Pouzo et al. 2022, Thm. 5).

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 the generating mechanism of $\{Z_t\}$, 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 over-fitting 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.

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, Carter and Steigerwald (2012) demonstrated that 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 (2) (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), with regimes $\{S_t\}$ which form 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 with $\{Z_t\}$ and $\{W_t\}$ satisfying the autoregressive equations

$$Z_t = \mu_2^* + \psi^* Z_{t-1} + \sigma_2^* U_{2,t},$$
$$W_t = \mu_3^* + \delta^* W_{t-1} + \sigma_3^* U_{3,t}.$$

https://doi.org/10.1017/S0266466625100017 Published online by Cambridge University Press

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

For each of 1,000 samples of size $T \in \{200, 800, 1,600, 3,200\}$ from this DGP, estimates of the parameters of the outcome equation are obtained by maximizing the quasi-log-likelihood function (4) associated with the mixture model (2) and (3), 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). The latter are computed using a sandwich estimator based on the Hessian and the gradient of the quasi-log-likelihood function (cf. Pouzo et al., 2022, Thm. 5), with weights obtained from the Parzen kernel and a data-dependent bandwidth selected by the plug-in method of Andrews (1991).

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.

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 (4) 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 2.4, 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, 1,000 artificial samples of various sizes are generated according to the

https://doi.org/10.1017/S0266466625100017 Published online by Cambridge University Press

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	0.093	−0.022	−0.033	0.018	−0.125	−0.045	0.090	−0.032	−0.029	0.013	−0.111	−0.041
800	0.017	−0.006	−0.003	0.002	−0.028	−0.013	0.021	−0.006	−0.003	0.004	−0.023	−0.010
1,600	0.009	−0.001	−0.001	0.000	−0.014	−0.006	−0.001	−0.008	0.002	0.003	−0.008	−0.008
3,200	−0.001	−0.004	−0.001	0.001	−0.005	−0.003	0.010	0.000	−0.003	0.000	−0.006	−0.002
	Standard deviation / standard error											
200	1.361	1.129	1.338	1.142	1.452	1.179	1.365	1.234	1.520	1.168	1.484	1.157
800	1.049	0.954	1.031	1.022	1.040	0.984	1.035	0.979	1.036	1.010	1.033	0.990
1,600	1.054	1.031	1.022	1.008	1.024	1.006	0.998	0.966	1.005	0.972	1.016	0.975
3,200	1.031	0.973	0.974	0.997	1.040	0.992	1.045	1.041	1.015	0.948	0.976	1.014
	$\rho^* = 0, \omega^* = 0.65$						$\rho^* = 0.65, \omega^* = 0.65$					
	Bias											
200	−0.190	−0.262	0.228	0.254	−0.171	−0.120	−0.205	−0.280	0.225	0.261	−0.162	−0.116
800	−0.226	−0.241	0.238	0.238	−0.098	−0.090	−0.233	−0.243	0.235	0.240	−0.096	−0.090
1,600	−0.231	−0.238	0.236	0.235	−0.089	−0.084	−0.227	−0.237	0.233	0.235	−0.089	−0.084
3,200	−0.235	−0.237	0.236	0.235	−0.083	−0.082	−0.230	−0.238	0.234	0.236	−0.085	−0.082
	Standard deviation / standard error											
200	1.183	1.151	1.328	1.129	1.328	1.075	1.182	1.189	1.322	1.111	1.317	1.183
800	0.812	0.888	1.008	0.819	0.826	0.861	0.979	1.012	1.035	0.991	1.038	1.025
1,600	0.997	1.042	0.998	0.988	1.000	0.975	0.989	1.071	0.982	0.965	1.011	1.003
3,200	1.021	1.080	1.002	0.995	0.971	1.022	1.084	1.162	1.024	1.030	1.018	1.032

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	−0.500	0.222	−0.093	−0.141	−0.012	−0.765	0.468	−0.114	−0.135	−0.049
800	−0.400	0.039	0.042	−0.067	0.002	−0.626	0.288	0.027	−0.060	−0.034
1,600	−0.440	0.023	0.086	−0.049	0.004	−0.753	0.276	0.096	−0.049	−0.031
3,200	−0.462	0.013	0.115	−0.036	0.005	−0.699	0.249	0.103	−0.035	−0.028
	Standard deviation / standard error									
200	1.546	1.214	1.764	1.563	1.140	0.438	1.261	0.542	1.008	1.023
800	1.282	0.907	1.335	1.177	0.997	1.228	0.928	1.347	1.121	0.954
1,600	1.396	0.916	1.335	0.708	0.988	1.542	1.011	1.504	1.066	1.008
3,200	1.424	0.897	1.398	0.913	0.968	1.078	0.803	1.102	0.775	0.976

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 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 are not always accurate as approximations to the finite-sample standard deviation of the QML estimators in the autoregressive model, even for a parameter such as $\sigma(1)$, which is 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 the previous experiments, is included in the right-hand sides of (10) and (11).

4. CONCLUSION

In this article, 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 estimated consistently 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 information matrix, and non-quadratic approximations to the log-likelihood function.

REFERENCES

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59, 817–858.

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

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

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

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

Compiani, G., & Kitamura, Y. (2016). Using mixtures in econometric models: A brief review and some new results. *Econometrics Journal*, 19, C95–C127.

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

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

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

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

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

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

Frühwirth-Schnatter, S. (2006). *Finite mixture and Markov switching models*. Springer.

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

Gourieroux, C., Monfort, A., & Trognon, A. (1984). Pseudo maximum likelihood methods: Theory. *Econometrica*, 52, 681–700.
Hamilton, J. D. (2016). Macroeconomic regimes and regime shifts. In J. B. Taylor, & H. Uhlig (Eds.), *Handbook of macroeconomics* (Vol. 2A, pp. 163–201). North-Holland.
Holzmann, H., Munk, A., & Gneiting, T. (2006). Identifiability of finite mixtures of elliptical distributions. *Scandinavian Journal of Statistics*, 33, 753–763.
Holzmann, H., Munk, A., & Stratman, B. (2004). Identifiability of finite mixtures - with applications to circular distributions. *Sankhyā*, 66, 440–449.
Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In L. M. Le Cam, & J. Neyman (Eds.), *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* (Vol. 1, pp. 221–233). University of California Press.
Hughes, J. P., Guttorp, P., & Charles, S. P. (1999). A non-homogeneous hidden Markov model for precipitation occurrence. *Applied Statistics*, 48, 15–30.
Jeanne, O., & Masson, P. (2000). Currency crises, sunspots and Markov-switching models. *Journal of International Economics*, 50, 327–350.
Komunjer, I. (2005). Quasi-maximum likelihood estimation for conditional quantiles. *Journal of Econometrics*, 128(1), 137–164.
Levine, D. (1983). A remark on serial correlation in maximum likelihood. *Journal of Econometrics*, 23, 337–342.
McLachlan, G. J., & Peel, D. (2000). *Finite mixture models*. Wiley.
Mevel, L., & Finesso, L. (2004). Asymptotical statistics of misspecified hidden Markov models. *IEEE Transactions on Automatic Control*, 49, 1123–1132.
Newey, W. K., & Steigerwald, D. G. (1997). Asymptotic bias for quasi-maximum-likelihood estimators in conditional heteroskedasticity models. *Econometrica*, 65, 587–599.
Pouzo, D., Psaradakis, Z., & Sola, M. (2022). Maximum likelihood estimation in Markov regime-switching models with covariate-dependent transition probabilities. *Econometrica*, 90, 1681–1710.
Ramesh, P., & Wilpon, J. G. (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.
Rydén, T., Teräsvirta, T., & Åsbrink, S. (1998). Stylized facts of daily returns series and the hidden Markov model. *Journal of Applied Econometrics*, 13, 217–244.
Teicher, H. (1963). Identifiability of finite mixtures. *Annals of Mathematical Statistics*, 36, 1265–1269.
White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, 50, 1–25.
White, H. (1994). *Estimation, inference and specification analysis*. Cambridge University Press.
Yakowitz, S. J., & Spragins, J. D. (1968). On the identifiability of finite mixtures. *Annals of Mathematical Statistics*, 39, 209–214.