Red Nacional de Investigadores en Economía

A Note on Quasi-Maximum-Likelihood Estimation in Hidden Markov Models with Covariate-Dependent Transition Probabilities

Demian Pouzo (University of California)

Zacharias Psaradakis (University of London)

Martin Sola (Universidad Torcuato di Tella)

DOCUMENTO DE TRABAJO N° 234

Abril de 2023

Citar como:
Pouzo, Demian, Zacharias Psaradakis y Martin Sola (2023). A Note on Quasi-Maximum-Likelihood Estimation in Hidden Markov Models with Covariate-Dependent Transition Probabilities. *Documento de trabajo RedNIE N°234.*

A Note on Quasi-Maximum-Likelihood Estimation in Hidden Markov Models with Covariate-Dependent Transition Probabilities

Demian Pouzo

Department of Economics, University of California, Berkeley, U.S.A.

Zacharias Psaradakis

Department of Economics, Mathematics and Statistics,

Birkbeck, University of London, U.K.

Martin Sola

Department of Economics, Universidad Torcuato di Tella, Argentina

March 8, 2023

Abstract

We consider hidden Markov models with a discrete-valued regime sequence whose transition probabilities are covariate-dependent. We show that consistent estimation of the parameters of the conditional distribution of the observable variables is possible via quasi-maximum-likelihood based on a (misspecified) mixture model without Markov dependence. 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

Statistical models with parameters that are subject to random changes driven by an unobservable Markov chain, typically referred to as the regime (or state) sequence, have attracted much attention in many different areas of application. An important subclass of such models, so-called hidden Markov models (HMMs), in which observable variables are conditionally independent, given the underlying regime sequence, with conditional distribution which depends on the current regime only, are also widely used in a variety of disciplines. Statistical inference in these models is typically likelihood-based, and the properties of relevant inferential procedures are, naturally, of much interest. In recent work, Pouzo, Psaradakis, and Sola (2022) considered the asymptotic properties of the quasi-maximum-likelihood (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 (discrete-valued) regime sequence, and potential model misspecification. For misspecified models, that is, models that do not contain the data-generating mechanism, the QML estimator has been shown to be consistent for the pseudo-true parameter (set) that minimizes the Kullback–Leibler information measure. In the special case of misspecified HMMs with temporally homogeneous regime sequences, related consistency results were given by Mevel and Finesso (2004) and Douc and Moulines (2012). 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 a general set-up.

In this note, we consider the case where the data-generating mechanism is an HMM with 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. This is a case of practical interest given the widespread use of both HMMs and mixture models. 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 random 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 depend on observable variables.

In the next section, we introduce the setting and notation, and give an overview of some relevant results on the consistency of the QML estimator in potentially misspecified models with Markov regimes. Our main result is established in Section 3, where QML estimation of the parameters of the outcome equation of a misspecified HMM is considered. Section 4 discusses some numerical results from a simulation study. Section 5 summarizes and concludes.

2 Notation and Consistency of QML Estimator

Let $\{(X_t, S_t)\}_{t=0}^{\infty}$ be a discrete-time stochastic process such that X_t is an observable variable with values in $\mathbb{X} \subset \mathbb{R}^h$, $h \geq 1$, and S_t is a latent variable with values in $\mathbb{S} := \{s_1, \dots, s_d\} \subset \mathbb{R}$, $d \geq 2$. As usual, 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. A general setting that encompasses many multiple-regime models specifies that, for each integer $t \geq 1$, the conditional distribution of X_t, given $X_0^{t-1} := (X_0, \dots, X_{t-1})$ and $S_0^t := (S_0, \dots, S_t)$, depends only on X_{t-1} and S_t, and the conditional distribution of S_t, given X_0^{t-1} and S_0^{t-1}, depends only on X_{t-1} and S_{t-1}, so that

$$X_t \mid (X_0^{t-1}, S_0^t) \sim P_*(X_{t-1}, S_t, \cdot),$$
$$S_t \mid (X_0^{t-1}, S_0^{t-1}) \sim Q_*(X_{t-1}, S_{t-1}, \cdot),$$

with $(x, s) \mapsto P_*(x, s, \cdot) \in \mathcal{P}(\mathbb{X})$ and $(x, s) \mapsto Q_*(x, s, \cdot) \in \mathcal{P}(\mathbb{S})$ being the true transition probabilities.[1] The researcher's probability model is given by a family of transition probabilities $(x, s) \mapsto P_\theta(x, s, \cdot) \in \mathcal{P}(\mathbb{X})$ and $(x, s) \mapsto Q_\theta(x, s, \cdot) \in \mathcal{P}(\mathbb{S})$

[1]For an arbitrary Polish space $\mathbb{V}$, $\mathcal{P}(\mathbb{V})$ denotes the family of Borel probability measures on $\mathbb{V}$.

indexed by a parameter θ, taking values in some compact set $\Theta \subset \mathbb{R}^q$, $q \geq 1$, such that, for each $\theta \in \Theta$,

$$X_t \mid (X_0^{t-1}, S_0^t) \sim P_\theta(X_{t-1}, S_t, \cdot),$$
$$S_t \mid (X_0^{t-1}, S_0^{t-1}) \sim Q_\theta(X_{t-1}, S_{t-1}, \cdot).$$

For each $(x, s) \in \mathbb{X} \times \mathbb{S}$ and all $\theta \in \Theta$, $P_*(x, s, \cdot)$ and $P_\theta(x, s, \cdot)$ are assumed to be absolutely continuous with respect to a common σ-finite Borel measure on $\mathbb{X}$.

This setting includes, among others, HMMs, Markov-switching autoregressive models, and certain types of mixture autoregressive models. The two most important features of the setting are that: (i) the hidden regimes $\{S_t\}_{t=0}^\infty$ are a temporally inhomogeneous Markov chain whose transition probabilities may depend on the lagged value of observable variables; (ii) the model $\{(P_\theta, Q_\theta) \colon \theta \in \Theta\}$ is potentially misspecified, in the sense that (P_*, Q_*) may not be a member of the family $\{(P_\theta, Q_\theta) \colon \theta \in \Theta\}$. Mild assumptions about (P_*, Q_*) guarantee the existence of a (unique) probability measure $\nu \in \mathcal{P}(\mathbb{X} \times \mathbb{S})$ such that the Markov chain $\{(X_t, S_t)\}_{t=0}^\infty$ is strictly stationary, ergodic and absolutely regular under the distribution $\bar{P}_*^\nu$ of $\{(X_t, S_t)\}_{t=0}^\infty$ induced by (P_*, Q_*) and $(X_0, S_0) \sim \nu$ (see Lemma 1 of Pouzo, Psaradakis, and Sola (2022)).[2]

Given observations $X_0^T := (X_0, \ldots, X_T)$, $T \geq 1$, and a measure $\kappa \in \mathcal{P}(\mathbb{X} \times \mathbb{S})$, the log-likelihood function for the parameter θ is

$$\theta \mapsto \ell_T^\kappa(\theta) := T^{-1} \sum_{t=1}^T \ln p_t^\kappa(X_t \mid X_0^{t-1}, \theta),$$

where $p_t^\kappa(X_t \mid X_0^{t-1}, \theta)$ denotes the conditional density of X_t, given X_0^{t-1}, under $\theta \in \Theta$ and $(X_0, S_0) \sim \kappa$. The QML estimator $\hat{\theta}_{\kappa,T}$ of θ is defined as an approximate maximizer of $\ell_T^\kappa(\theta)$ over Θ, so that

$$\ell_T^\kappa(\hat{\theta}_{\kappa,T}) \geq \sup_{\theta \in \Theta} \ell_T^\kappa(\theta) - \eta_T,$$

[2]For simplicity, and with a slight abuse of notation, we will also use $\bar{P}_*^\nu$ to denote the probability distribution of the two-sided, strictly stationary extention $\{(X_t, S_t)\}_{t=-\infty}^\infty$ induced by (P_*, Q_*, ν), as well as any marginal or conditional distributions associated with it.

for some $\eta_T \geq 0$ with $\eta_T \to 0$ as $T \to \infty$.

Under suitable conditions, the QML estimator $\hat{\theta}_{\nu,T}$ tends to the pseudo-true parameter (set) that minimizes the Kullback–Leibler information function $\theta \mapsto H^*(\theta)$, defined by

$$H^*(\theta) := E_{\bar{P}_*^\nu}\left(\ln \frac{p_*^\nu(X_0 \mid X_{-\infty}^{-1})}{p^\nu(X_0 \mid X_{-\infty}^{-1}, \theta)}\right),$$

where, for any $\theta \in \Theta$, $p^\nu(X_0 \mid X_{-\infty}^{-1}, \theta)$ denotes the conditional density of X_0, given $X_{-\infty}^{-1} := (X_{-\infty}, \ldots, X_{-1})$, induced by $(P_\theta, Q_\theta, \nu)$, and $p_*^\nu(X_0 \mid X_{-\infty}^{-1})$ is its counterpart induced by (P_*, Q_*, ν). More specifically,

$$\inf_{\theta \in \Theta_*} ||\hat{\theta}_{\nu,T} - \theta|| \to 0 \quad \text{as } T \to \infty, \tag{1}$$

in $\bar{P}_*^\nu$-probability, where $\Theta_* := \arg\min_{\theta \in \Theta} H^*(\theta)$ and $\|\cdot\|$ denotes the Euclidean norm on $\mathbb{R}^q$ (see Theorem 1 of Pouzo, Psaradakis, and Sola (2022)).[3] In the case where the model is correctly specified and point-identified, i.e., there exists a $\theta_0 \in \Theta$ such that $(P_*, Q_*) = (P_\theta, Q_\theta)$ if and only if $\theta = \theta_0$, then $\Theta_* = \{\theta_0\}$ and $\hat{\theta}_{\nu,T}$ is consistent for θ_0.

In the general setting described above, obtaining a complete characterization and interpretation of the pseudo-true parameter set Θ_* is far from straightforward, and must be done on a case-by-case basis. We discuss one such case next.

3 QML in a Misspecified HMM

In this Section, we characterize the pseudo-true parameter set Θ_* in the case of a (misspecified) mixture model (P_θ, Q_θ) when the true data-generating mechanism (P_*, Q_*) is an HMM with covariate-dependent transition probabilities. Specifically, we show that, even though the mixture model misspecifies the dependence structure of the hidden regimes, and thus of the overall system, parameters related to the outcome equation are consistently estimable.

[3] If Θ_* is a singleton, it can also be shown that, under suitable regularity conditions, the quasi-log-likelihood function has a local asymptotic normality property and the QML estimator $\hat{\theta}_{\nu,T}$ admits an asymptotic linear representation (see Pouzo, Psaradakis, and Sola (2022)).

For the sake of simplicity, we consider a bivariate data-generating process with $X_t = (Y_t, Z_t) \in \mathbb{X} = \mathbb{R}^2$, $S_t \in \mathbb{S}$, and $P_*(X_{t-1}, S_t, \cdot)$ specified through the equations

$$Y_t = \mu_1^*(S_t) + \sigma_1^*(S_t)U_{1,t}, \tag{2}$$

$$Z_t = \mu_2^* + \psi^* Z_{t-1} + \sigma_2^* U_{2,t}. \tag{3}$$

Here, $\{(U_{1,t}, U_{2,t})\}_{t\geq 0}$ are i.i.d. centered random variables, independent of $\{S_t\}_{t\geq 0}$, with absolutely continuous common distribution (with respect to Lebesgue measure on $\mathbb{R}^2$) and covariance matrix $[\omega_{ij}]$, with $\omega_{11} = \omega_{22} = 1$ and $\omega_{12} = \rho^*$ for some $|\rho^*| < 1$. The transition probabilities of the regimes are assumed to depend on (Z_{t-1}, S_{t-1}) so that, for each $t \geq 1$, $Q_*(Z_{t-1}, S_{t-1}, \cdot)$ is identified with a Markovian matrix $[q_*^t(s, s')]_{s,s' \in \mathbb{S}}$, where $q_*^t(s, s') := \Pr(S_t = s' \mid S_{t-1} = s, Z_{t-1})$. It is further assumed that $|\psi^*| < 1$ and $\sigma_1^*, \sigma_2^* \in (0, \infty)$, and that the process $\{(Z_t, S_t)\}_{t\geq 0}$ is strictly stationary with invariant distribution $\nu_{ZS} \in \mathcal{P}(\mathbb{R} \times \mathbb{S})$.

Instead of the HMM above, the researcher's postulated model (P_θ, Q_θ) is assumed to be a finite mixture model (without Markov dependence) of the form

$$Y_t = \mu(S_t) + \sigma(S_t)\varepsilon_t, \tag{4}$$

$$S_t \sim Q_{\bar{\vartheta}}, \tag{5}$$

where $Q_{\bar{\vartheta}}(s) = \bar{\vartheta}_s \in (0, 1)$ for each $s \in \mathbb{S}$, $\sigma \in (0, \infty)$, and $\{\varepsilon_t\}_{t\geq 1}$ are i.i.d. random variables independent of $\{S_t\}_{t\geq 1}$. It is further assumed that ε_1 and $U_{1,1}$ are identically distributed, with common density f.

It is worth emphasizing that this relatively simple set-up is of practical interest. HMMs like (2)–(3), in which the transition probabilities of hidden regimes are functions of observable variables, have found a variety of applications in different areas (see, inter alia, Ramesh and Wilpon (1992), Diebold, Lee, and Weinbach (1994), Hughes and Guttorp (1994), Engel and Hakkio (1996), Ghavidel, Claesen, and Burzykowski (2015)). Furthermore, temporally homogeneous versions of these models, in which $q_*^t(s, s')$ are invariant with respect to the index t, are used extensively in economics and finance (see, e.g, Engel and Hamilton (1990), Rydén, Teräsvirta, and Åsbrink (1998), Bollen, Gray, and Whaley (2008)), as well as in biology, computing, engineering and statistics (see the overview by Ephraim and Merhav (2002) and references therein). Mixture models with i.i.d. regimes are also popular in many different fields (see, e.g., McLachlan and Peel (2000)).

The mixture model defined by (4)–(5) evidently misspecifies the dependence structure of the regime sequence. It is not too onerous to verify that, under assumptions about $(U_{1,t}, U_{2,t})$ and $[q_*^t(s, s')]$ that are common in the literature (e.g., $(U_{1,1}, U_{2,1})$ having a Gaussian distribution and $q_*^t(s, s') = \Lambda(\alpha_{s,s'} + \beta_{s,s'} Z_{t-1})$ for some continuous function Λ on $\mathbb{R}$ with range in $(0, 1]$), the conditions of Pouzo, Psaradakis, and Sola (2022) required for the convergence result in (1) to hold for the QML estimator of $\theta = (\mu(s), \sigma(s), \bar{\vartheta}_s)_{s \in \mathbb{S}}$ are satisfied.

A sharper result can, in fact, be established by considering the pseudo-true parameter set under (2)–(5). As the following theorem shows, 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 $\mu = \mu_1^*$, $\sigma = \sigma_1^*$ and $\bar{\vartheta}$ such that $Q_{\bar{\vartheta}} = E_{\nu_{ZS}}[Q_*(Z, S, \cdot)]$ is a pseudo-true parameter, i.e., it minimizes*

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

Proof. Observe that H^* is proportional to

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

where

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

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

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

It is well known that the minimizers of this function are all θ such that

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

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

This result is quite general, in the sense that we consider a misspecified HMM with a temporally inhomogeneous regime sequence and arbitrary observation conditional densities. It implies that dependence of the regimes in such an HMM may be safely ignored as long as the parameters of interest are only those of the conditional density of the observations given the regimes. Treating the regimes (and consequently the observations) as an i.i.d. sequence simplifies likelihood-based inference somewhat compared to the case of correlated Markov regimes.

The result in Theorem 1 is reminiscent of results of Levine (1983) and Bates and White (1985) on the possible consistency of QML estimators in the presence of dynamic misspecification. For a class of regime-switching models in which the regime sequence $\{S_t\}$ is a temporally homogeneous, two-state Markov chain, a similar point was also 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 does not, in fact, hold if the model (and the data-generating process) contain an autoregressive component. This observation is also true in our more general setting with temporally inhomogeneous regime sequences. Specifically, a result analogous to that in Theorem 1 does not hold when the right-hand side of the outcome equation contains lagged values of the dependent variable (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.

4 Numerical Examples

As a numerical illustration of the results discussed in the previous section, 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 simulations, 1000 artificial samples of size $T \in \{200, 800, 1600, 3200\}$ are generated according to the HMM (2)–(3), with the regimes $\{S_t\}$ being a Markov chain on $\mathbb{S} = \{1, 2\}$ such that

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

where $\alpha_1^* = 2$, $\alpha_2^* = 2$, $\beta_1^* = -0.5$ and $\beta_2^* = 0.5$; the distribution of $(U_{1,t}, U_{2,t})$ is Gaussian, $\rho^* \in \{0, 0.8\}$, $\mu_1^*(1) = 1$, $\mu_1^*(2) = -1$, $\sigma_1^*(1) = 1$, $\sigma_1^*(2) = 1$, $\mu_2^* = 0.2$, $\psi^* = 0.8$ and $\sigma_2^* = 0.6$. The parameters of the outcome equation are then estimated for each sample by maximizing the quasi- log-likelihood function associated with the mixture model (4)–(5), with $\Pr(S_t = 1) = \bar{\vartheta}$ and $\varepsilon_t \sim N(0, 1)$.

Monte Carlo estimates of the bias of the QML estimators of $\mu(1)$, $\mu(2)$, $\sigma(1)$ 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). 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

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

T	$\mu(1)$	$\mu(2)$	$\sigma(1)$	$\sigma(2)$	$\mu(1)$	$\mu(2)$	$\sigma(1)$	$\sigma(2)$
	$\rho^* = 0$				$\rho^* = 0.8$			
				Bias				
200	0.191	-0.015	-0.138	-0.020	0.116	0.013	-0.131	-0.025
800	0.032	0.004	-0.032	-0.005	0.043	0.005	-0.032	-0.005
1600	0.009	0.002	-0.012	-0.001	0.008	0.003	-0.011	-0.002
3200	0.010	0.000	-0.009	-0.001	0.017	-0.003	-0.012	0.000
				Standard Deviation / Standard Error				
200	1.162	1.147	1.186	1.155	1.104	1.162	1.131	1.101
800	0.994	0.995	1.011	1.013	0.990	1.017	1.007	1.029
1600	1.017	1.017	1.032	1.032	1.009	1.028	1.011	0.997
3200	0.919	0.924	0.900	0.992	0.939	0.928	0.961	0.934

the asymptotic covariance matrix of the QML estimator in misspecified models (cf. Theorem 5 of Pouzo, Psaradakis, and Sola (2022)).

Ignoring Markov dependence of the regimes, as pointed out in the previous section, is not costless when the outcome equation contains 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

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

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 based on the HMM. For each artificial sample, the parameters of the model

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

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 $\varepsilon_t \sim N(0, 1)$.

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 is not perhaps very

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

surprising as 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, 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.

5 Conclusion

In this note, we have considered QML estimation of the parameters of an HMM with 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 the parameters of the outcome equation can be consistently estimated using the quasi-likelihood function associated with a misspecified HMM with independent regimes.

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, following a QML-based approach analogous to that of Cho and White (2007). As is well known, such testing problems are non-standard and can 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 *Nonstationary Time Series Analysis and Cointegration*, ed. by C. P. Hargreaves, pp. 283–302. Oxford University Press, Oxford.

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.

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.

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.

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*. Wiley, New York.

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

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*, vol. 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.