ON THE DETERMINATION OF THE NUMBER OF REGIMES IN MARKOV-SWITCHING AUTOREGRESSIVE MODELS

By Zacharias Psaradakis and Nicola Spagnolo

Birkbeck College, Brunel University

First Version received March 2002

Abstract. Dynamic models with parameters that are allowed to depend on the state of a hidden Markov chain have become a popular tool for modelling time series subject to changes in regime. An important question that arises in applications involving such models is how to determine the number of states required for the model to be an adequate characterization of the observed data. In this paper, we investigate the properties of alternative procedures that can be used to determine the state dimension of a Markov-switching autoregressive model. These include procedures that exploit the ARMA representation which Markov-switching processes admit, as well as procedures that are based on optimization of complexity-penalized likelihood measures. Our Monte Carlo analysis reveals that such procedures estimate the state dimension correctly, provided that the parameter changes are not too small and the hidden Markov chain is fairly persistent. The use of the various methods is also illustrated by means of empirical examples.

Keywords. ARMA representation; complexity-penalized likelihood criteria; Markov-switching models; Monte Carlo experiment; three-pattern identification method.

1. Introduction

Dynamic models with parameters that are subject to Markov regime switching have attracted a great deal of interest in the recent econometric and statistical literature. Such models extend the class of so-called hidden Markov models, in which observations are assumed to be independent conditional on a hidden Markov chain, and applications that involve Markov-switching models abound in areas ranging from economics and finance to engineering and biology.

A key problem which arises in applications is how to determine the number of Markov states (or regimes) required for a switching model to be an adequate characterization of the observed data. Surprisingly, this problem has received little attention in the literature in spite of its obvious theoretical and practical importance. In practice, the state dimension of the hidden Markov chain that drives regime changes is sometimes dictated by the actual application or is determined in an informal manner by visual inspection of plots of the data. Such an approach is evidently subjective and somewhat arbitrary and leaves much to be desired.

0143-9782/03/02 237–252 JOURNAL OF TIME SERIES ANALYSIS Vol. 24, No. 2

A more formal and statistically sound procedure for determining the state dimension can be based on likelihood ratio tests. Such tests are problematic, however, because the usual regularity conditions are not fulfilled under the null hypothesis (some parameters are unidentified and the information matrix is singular) and hence the asymptotic null distribution of the likelihood ratio test statistic is not a χ^2 one. Hansen (1992) proposed a test which does not suffer from this problem, but his test procedure is computationally burdensome and only delivers P-values which are an upper bound for the true P-values. A computationally less demanding test procedure is discussed in Garcia (1998), but this method is theoretically less attractive than Hansen's (1992) since it overlooks the problem of the singular information matrix. An alternative way of circumventing these difficulties would be to use critical values and/or P-values from a bootstrap approximation to the null distribution of the likelihood ratio test statistic, as suggested by McLachlan (1987) for independent mixture models. It should be borne in mind, however, that the asymptotic correctness of such bootstrap tests has not been established yet and is far from obvious.

Instead of relying on hypothesis testing, one can consider methods based on complexity-penalized likelihood criteria. Such methods have enjoyed much popularity in statistics as a means of choosing among competing models and, under appropriate regularity conditions, are known to be capable of selecting with probability 1 the model with lowest Kullback–Leibler divergence from the data-generating mechanism (Nishii, 1988; Sin and White, 1996). Furthermore, as Granger *et al.* (1995) pointed out, these methods are arguably more appropriate for model selection than procedures based on formal hypothesis testing, partly because, unlike testing, they do not favour unfairly the model chosen to be the null hypothesis. The use of complexity-penalized likelihood criteria, such as the popular Akaike information criterion (AIC) and Bayesian information criterion (BIC), as a means of selecting the number of components in independent and Markov-dependent finite mixture models has been studied by Leroux (1992), Leroux and Puterman (1992) and Rydén (1995). To our knowledge, there has so far been no work on the use of such criteria for the estimation of the state dimension of Markov-switching dynamic models.

More recently, Anděl (1993), Poskitt and Chung (1996), Francq and Zakoïan (1998, 2001) and Zhang and Stine (2001) have shown that weakly stationary processes generated by various Markov-switching models admit linear autoregressive moving-average (ARMA) representations.[1] Since the order of these ARMA representations is generally a function of the number of Markov regimes, Poskitt and Chung (1996) and Zhang and Stine (2001) have argued that the state dimension of the hidden Markov chain can be determined by estimating the order of the equivalent ARMA representation of the observable process.

[1]A weakly stationary process is said to admit and ARMA(p, q) representation if it has the same autocovariance structure as a causal and invertible ARMA(p, q) process, i.e. if and only if its autocovariances satisfy a difference equation of minimal order p with minimal rank $q+1$; see, for example, Zhang and Stine (2001, Thm 1).

Zhang and Stine (2001) used simulation experiments to examine the properties of the proposed method as a means of identifying the state order in Markov-switching models with no dynamics and compared them with procedures based on the AIC and BIC.

The present paper contributes to this discussion by investigating the properties of state-dimension determination procedures based on ARMA representations and complexity-penalized likelihood criteria in the case of Markov-switching autoregressive models. Such models have been used extensively in the applied econometric literature since the influential work of Hamilton (1988, 1989) and are empirically more relevant than the simpler Markov-dependent mixture models examined in Rydén (1995) and Zhang and Stine (2001). The paper proceeds by first giving a brief description of the models of interest and the methods that will be used to determine the state dimension of the underlying hidden Markov chain. Then, Section 3 discusses the design of the Monte Carlo experiments that are used to investigate the small-sample performance of various order selection methods and presents the results of the experiments. Section 4 provides a few illustrative examples involving real-world data. Section 5 summarizes and concludes.

2. DETERMINATION OF THE STATE DIMENSION

The focus of this paper is the class of Markov-switching autoregressive models of the form

$$
\begin{aligned}
X_t &= \mu(S_t) + \sum_{\tau=1}^{m} \phi_\tau(S_t)\{X_{t-\tau} - \mu(S_{t-\tau})\} + \sigma(S_t)\varepsilon_t \\
\mu(S_t) &= \sum_{i=1}^{r} \mu^{(i)} \mathbb{I}(S_t = i) \\
\sigma(S_t) &= \sum_{i=1}^{r} \sigma^{(i)} \mathbb{I}(S_t = i) \\
\phi_\tau(S_t) &= \sum_{i=1}^{r} \phi_\tau^{(i)} \mathbb{I}(S_t = i) \qquad \tau = 1, \ldots, m
\end{aligned}
\tag{1}
$$

Here, m is a positive integer, $\mu^{(i)}$, $\sigma^{(i)}$, and $\phi_\tau^{(i)} (i = 1, \ldots, r)$ are real constants, $\{\varepsilon_t\}$ are independent, identically distributed (i.i.d.) random variables with $\mathbb{E}(\varepsilon_t) = \mathbb{E}(\varepsilon_t^2 - 1) = 0$, S_t is a random variable that takes values in the finite set $\Omega = \{1,\ldots,r\}$ and which indicates the unobservable state at date t, and $\mathbb{I}(A)$ is the indicator of the event A. The state variables $\{S_t\}$ are assumed to form a strictly stationary, time-homogeneous, first-order Markov chain on $\Omega = \{1,\ldots,r\}$ with transition probability matrix $\mathbf{P} = (p_{ij})'_{i,j\in\Omega}$, where $p_{ij} = \mathbb{P}(S_{t+1} = j | S_t = i)$. It is also assumed that $\{S_t\}$ is independent of $\{\varepsilon_t\}$ and that $\mathbf{P}$ is ergodic. The model defined by (1) can be thought of as a Markov mixture of r autoregressive models, and we shall refer to it as an r-state m-order Markov-switching autoregressive [MSAR(r, m)] model.

The autoregressive order m is assumed to be fixed and known, so that the problem of interest is to determine the state dimension r from a realization (X_1,..., X_T) of T consecutive observations from (1).[2] Since the selection methods we are interested in either directly exploit the second-order properties of the MSAR(r, m) model or are based on maximum likelihood estimation procedures that rely heavily on stationarity assumptions, we require the process $\{X_t\}$ satisfying (1) to be weakly stationary. This requirement will be met if the spectral radius of the matrix $\Lambda = \mathbf{D}(\mathbf{P} \otimes \mathbf{I}_{m^2})$ is strictly less than 1, where $\mathbf{D} = \mathrm{diag}\{\mathbf{\Phi}^{(1)} \otimes \mathbf{\Phi}^{(1)}, \ldots, \mathbf{\Phi}^{(r)} \otimes \mathbf{\Phi}^{(r)}\}$,

$$\mathbf{\Phi}^{(i)} = \begin{pmatrix} \phi_1^{(i)} & \phi_2^{(i)} & \cdots & \phi_{m-1}^{(i)} & \phi_m^{(i)} \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix} \qquad i = 1, \ldots, r$$

and $\mathbf{I}_k$ denotes the k-dimensional identity matrix (Yao, 2001; Zhang and Stine, 2001).

The results in Zhang and Stine (2001) imply that a weakly stationary process satisfying an MSAR(r, m) model admits an ARMA(p, q) representation with $p \leqslant rm^2$ and $q \leqslant rm^2 - 1$. Therefore, the state dimension r satisfies $r \geqslant \max\{p/m^2, (q+1)/m^2\}$. This result offers a way of setting a lower bound on the number of Markov regimes by determining the orders p and q of the ARMA representation of the MSAR(r, m) process. This is a problem that has attracted considerable attention in the literature and a variety of order identification methods for ARMA processes are available (Choi, 1992). In this paper, we make use of the so-called three-pattern method (TPM) and the associated χ^2 tests proposed by Choi (1993). (A brief description of TPM can be found in the Appendix.)

An alternative approach is to work directly the multiple-regime representation of the MSAR(r, m) process and determine the state dimension r by means of a complexity-penalized likelihood criterion. More specifically, an estimate of r can be obtained as

$$\hat{r} = \arg\max_{1 \leqslant k \leqslant k^*} \{\ln \mathcal{L}(\hat{\boldsymbol{\theta}}_k; X_1, \ldots, X_T) - C_T \dim(\mathbf{\Theta}_k)\} \tag{2}$$

where $\mathcal{L}(\boldsymbol{\theta}_k; X_1, \ldots, X_T)$ and $\mathbf{\Theta}_k$ are the likelihood function and parameter space, respectively, for the MSAR(k, m) model, $\hat{\boldsymbol{\theta}}_k = \arg\max_{\boldsymbol{\theta}_k \in \mathbf{\Theta}_k} \mathcal{L}(\boldsymbol{\theta}_k; X_1, \ldots, X_T)$ is

[2] The problem of how to determine the autogressive order m when the state dimension r is known was studied by Kapetanios (2001).

the maximum likelihood estimator of the model's parameter vector $\boldsymbol{\theta}_k$, and C_T is a real constant.[3] The three most popular choices for C_T in the literature are $C_T = 1, C_T = \frac{1}{2}\ln T$, and $C_T = c \ln\ln T$ $(c > 1)$, yielding the AIC (Akaike, 1974), the BIC (Rissanen, 1978; Schwarz, 1978), and the Hannan–Quinn criterion (HQC) (Hannan and Quinn, 1979), respectively.

Given that an MSAR(r, m) process can be shown to be near-epoch dependent (in L_2-norm) on some mixing base (Kapetanios, 1999; Davidson, 2001), $\hat{r}$ will be weakly or strongly consistent for r under the conditions on the penalty factor $C_T \dim(\boldsymbol{\Theta}_k)$ given in Theorems 3 and 4, respectively, of Kapetanios (2001), assuming the maximum state dimension (k*) allowed for in (2) is not smaller than the true state dimension (r).[4] It is worth noting that AIC does not satisfy these conditions since its penalty factor does not diverge to infinity with T and $\lim_{T\to\infty} C_T/\ln\ln T = 0$.

3. MONTE-CARLO EXPERIMENTS

This section reports on the Monte Carlo experiments that we carry out so as to evaluate the small-sample performance of the model selection methods described in the previous section. After a brief description of the experimental design, a discussion of the results follows.

3.1. *Experimental Design*

In out experiments, we use as the data-generating process (DGP) both an MSAR(2, 1) and an MSAR(3, 1) model with Gaussian i.i.d. errors. The parameter values for the MSAR(2, 1) DGP are

$$\begin{gathered} \mu^{(1)} = 0 \qquad \mu^{(2)} \in \{1, 3\} \\ \sigma^{(1)} = 1 \qquad \sigma^{(2)} \in \{1, \sqrt{1.5}\} \\ \phi_1^{(1)} = 0.3 \qquad \phi_1^{(2)} \in \{0.3, 0.6, 0.9\} \\ (p_{11},\ p_{22}) \in \{(0.6, 0.4), (0.9, 0.9), (0.9, 0.98)\} \end{gathered}$$

In the case of the MSAR(3, 1) DGP, the parameter values for the regimes corresponding to $S_t = 1$ and $S_t = 2$ are the same as in the two-state case, while

$$\mu^{(3)} \in \{2, 6\} \qquad \sigma^{(3)} \in \{1, \sqrt{1.5}\} \qquad \phi_1^{(3)} \in \{0.3, 0.9\}$$

[3] Note that $\dim(\boldsymbol{\Theta}_k) = k(k + m + 1)$ for an MSAR(k, m) model defined as in (1).
[4] It also needs to be assumed that conditions analogous to those in Kapetanios (1999, pp. 189–92) hold so that the asymptotic results of Sin and White (1996) can be exploited.

The transition probability matrix is chosen as

$$\mathbf{P}_1 = \begin{pmatrix} 0.60 & 0.02 & 0.02 \\ 0.20 & 0.90 & 0.08 \\ 0.20 & 0.08 & 0.90 \end{pmatrix} \quad \text{or} \quad \mathbf{P}_2 = \begin{pmatrix} 0.90 & 0.02 & 0.02 \\ 0.02 & 0.90 & 0.08 \\ 0.08 & 0.08 & 0.90 \end{pmatrix}$$

The transition probabilities $(p_{11}, p_{22}) = (0.6, 0.4)$ used for the MSAR(2, 1) DGP imply that the regime-indicator variables $\{S_t\}$ are uncorrelated, so x_t does not depend on the regime that prevailed at date $t-1$. The transition probabilities $(p_{11}, p_{22}) = (0.9, 0.9)$ and $(p_{11}, p_{22}) = (0.9, 0.98)$, allow, on the other hand, the two regimes to be highly persistent, with the regime corresponding to $S_t = 2$ being almost absorbing in the latter case (the stationary distribution of $\{S_t\}$ is (0.5, 0.5) and (0.1667, 0.8333), respectively). When $\{S_t\}$ is a three-state chain, the transition matrix $\mathbf{P}_1$ allows the states corresponding to $S_t = 2$ and $S_t = 3$ to be equally persistent and much more so than the regime corresponding to $S_t = 1$; the stationary distribution of the chain is (0.0476, 0.04762, 0.4762). The transition matrix $\mathbf{P}_2$ implies that the average duration of all three regimes is the same; the stationary distribution of $\{S_t\}$ is (0.1667, 0.3889, 0.4444). All design points satisfy the weak stationarity condition stated in the previous section.

The experiments proceed by generating an artificial time series of length $T+50$ according to (1) with $T \in \{100, 200, 400\}$ and initial values set to zero.[5] The first 50 pseudo-data points are then discarded to minimize the effect of initial conditions and the remaining T points are used to determine the state dimension r by means of the TPM and the complexity-penalized likelihood criteria AIC, BIC and HQC (with $k^* = 3$).[6] In the latter case, the (conditional) likelihood function for the MSAR(2, 1) and MSAR(3, 1) models is computed recursively using the procedure discussed in Hamilton (1994, ch. 22.4). Maximum likelihood estimates are then found by a quasi-Newton optimization algorithm that uses the Broyden–Fletcher–Goldfarb–Shano secant update to the Hessian (with the true DGP parameters used as the staring point for the optimizer). For each design point, 1000 Monte-Carlo replications are carried out.

3.2. *Numerical results*

We now report and discuss the results of our simulation experiments. Since the results we obtain when the innovation variance is kept constant across regimes are similar to those obtained with state-dependent variances, we report only the latter. It should also be borne in mind throughout the subsequent discussion that,

[5]The initial value of the Markov chain $\{S_t\}$ is drawn randomly from its stationary distribution.

[6]Although Hannan and Quinn (1979) suggested setting $C_T = c \ln\ln T$ in (2) with $c > 1$, we take $c = 1$, which is the value of c typically used in practice (and in many econometrics and statistics computer packages).

as explained in Section 2, the estimate of the state dimension obtained by means of the TPM is only a lower bound for the number of Markov regimes.

Starting with the MSAR(2, 1) DGP, Table I and II show the percentage frequencies of the state dimensions selected by the various procedures. The first noteworthy feature of the results is their sensitivity with respect to the magnitude of parameter change. It is immediately apparent that the performance of all selection procedures improves as the difference between the values of parameters in the two regimes increases. This is true for both changes in the autoregressive coefficient $\phi_1(S_t)$ and changes in the mean parameter $\mu(S_t)$. For the largest changes considered in the experiments ($\mu^{(2)} - \mu^{(1)} = 3, \phi_1^{(2)} - \phi_1^{(1)} = 0.6$), all four procedures choose the right state dimension perfectly (or almost perfectly) even when $T = 100$. Their performance when $\mu^{(2)} - \mu^{(1)} = 1$ and $\phi_1^{(2)} - \phi_1^{(1)} = 0.6$ remains very good, provided that $T = 200$ or $T = 400$.

The magnitude of the transition probabilities also affects significantly the properties of the selection procedures. To be precise, our selection methods, especially those based on complexity-penalized likelihood measures, often have difficulty with choosing the correct number of regimes in the case where $p_{11} = 0.6$ and $p_{22} = 0.4$ (i.e. when the regime-indicator variable $\{S_t\}$ are uncorrelated). The problem is pronounced when the sample size and parameter changes are relatively small. This is probably due to the fact that the frequent regime transitions that take place when $p_{11} = 1 - p_{22}$, combined with small changes in parameters, tend to make the MSAR(2, 1) series look very much like heteroscedastic white noise. Fortunately, matters improve considerably as the magnitude of parameter changes and the persistence of Markov regimes increase, especially when the state dimension is selected by the TMP.

Turning to the properties of the individual selection procedures, it is clear that complexity-penalized likelihood methods based on the BIC and HQC often fail to estimate the state dimension correctly. This is particularly evident in the case of the BIC, which tends to underestimate the number of regimes considerably when the changes is parameters are moderate, a problem which is exacerbated by low serial correlation in the hidden Markov chain. The AIC is generally more successful than the BIC and HQC in selecting the correct number of Markov regimes, especially in those cases where the BIC and HQC are prone to underestimation. (For example, even with $T = 400$, the BIC wrongly selects the single-regime model in 806 out of 1000 replications when $\mu^{(2)} - \mu^{(1)} = 3$, $\phi_1^{(2)} - \phi_1^{(1)} = 0$, and $p_{11} = 1 - p_{22} = 0.6$, while the AIC makes a mistake in only 65 replications.)[7] This is an interesting finding since, as mentioned in Section 2, the AIC does not satisfy the conditions required for consistent estimation of the state dimension, and highlights the importance of examining the finite-sample behaviour of model selection procedures in addition to their asymptotic properties. Finally, the TPM performs very well overall. When the

[7]This is contrast to the results in Kapetanios (2001), which show that the HQC and BIC generally outperform AIC as a means of estimating the autoregressive order m of an MSAR(r, m) process when the state dimension r is fixed and known.

TABLE I

MONTE CARLO RESULTS FOR MSAR(2, 1)
DGP $\mu^{(1)} = 0,\ \mu^{(2)} = 1,\ \sigma^{(1)} = 1,\ \sigma^{(2)} = \sqrt{1.5}$

	$T = 100$			$T = 200$			$T = 400$		
r	1	2	3	1	2	3	1	2	3
$\phi_1^{(1)} = \phi_1^{(2)} = 0.3$									
$(p_{11}, p_{22}) = (0.6, 0.4)$									
AIC	71.5	25.4	3.10	75.0	23.6	1.40	75.9	22.6	1.50
BIC	98.3	0.90	0.80	98.9	0.04	0.07	99.3	0.20	0.50
HQC	92.2	6.70	1.10	94.9	4.40	0.07	95.8	3.60	0.60
TPM	34.1	45.9	20.0	5.90	71.1	23.0	0.10	79.8	20.1
$(p_{11}, p_{22}) = (0.9, 0.9)$									
AIC	65.8	32.9	1.30	66.8	32.7	0.50	65.5	33.9	0.60
BIC	98.7	1.20	0.01	99.3	0.70	0.00	99.8	0.10	0.10
HQC	89.1	10.0	0.09	91.7	8.30	0.00	92.8	7.10	0.10
TPM	8.10	84.3	7.60	1.00	85.1	13.9	0.00	87.9	12.1
$(p_{11}, p_{22}) = (0.9, 0.98)$									
AIC	67.4	31.2	1.40	64.9	33.7	1.40	66.4	32.8	0.80
BIC	98.7	1.30	0.00	99.0	0.60	0.40	99.6	0.30	0.10
HQC	90.4	9.60	0.00	92.3	7.30	0.40	92.3	7.60	0.10
TPM	24.1	57.8	18.1	2.60	80.7	16.7	0.00	86.2	13.8
$\phi_1^{(1)} = 0.3, \phi_1^{(2)} = 0.6$									
$(p_{11}, p_{22}) = (0.6, 0.4)$									
AIC	50.9	35.9	13.2	29.5	60.2	10.3	4.80	83.7	11.5
BIC	96.8	2.00	1.20	91.6	7.70	0.70	72.6	27.2	0.20
HQC	82.6	14.3	3.10	65.3	33.5	1.20	25.7	73.4	0.90
TPM	2.30	81.4	16.3	0.00	94.2	5.80	0.00	92.3	7.70
$(p_{11}, p_{22}) = (0.9, 0.9)$									
AIC	23.2	66.8	10.0	5.60	90.6	3.80	0.10	98.6	1.30
BIC	79.8	18.8	1.40	47.6	51.9	0.50	7.20	92.0	0.80
HQC	49.6	47.5	2.90	19.3	79.4	1.30	0.80	98.2	0.10
TPM	0.30	92.0	7.70	0.00	92.3	7.70	0.00	87.6	12.4
$(p_{11}, p_{22}) = (0.9, 0.98)$									
AIC	43.5	45.3	11.2	18.6	77.4	4.00	0.30	96.3	0.70
BIC	84.5	14.4	32.3	56.2	43.2	0.60	15.4	84.4	0.20
HQC	65.1	32.3	2.60	33.9	65.4	0.70	6.90	92.7	0.40
TPM	10.8	74.2	15.0	0.40	83.1	16.5	0.00	68.5	31.5
$\phi_1^{(1)} = 0.3, \phi_1^{(2)} = 0.9$									
$(p_{11}, p_{22}) = (0.6, 0.4)$									
AIC	9.90	79.9	10.2	0.70	88.7	10.6	0.00	85.8	14.2
BIC	58.6	40.9	0.50	20.9	78.9	0.20	0.50	99.4	0.10
HQC	27.3	70.7	0.20	4.50	94.4	1.10	0.00	99.0	0.10
TPM	0.30	90.8	8.90	0.00	94.2	5.80	0.00	94.1	5.90
$(p_{11}, p_{22}) = (0.9, 0.9)$									
AIC	1.80	93.6	4.60	0.00	99.1	0.90	0.00	99.1	0.90
BIC	19.2	80.1	0.70	1.20	98.6	0.20	0.00	99.4	0.60
HQC	7.30	91.1	1.60	0.10	99.7	0.20	0.00	99.3	0.70
TPM	0.00	93.2	6.80	0.00	90.2	9.80	0.00	82.7	17.3
$(p_{11}, p_{22}) = (0.9, 0.98)$									
AIC	27.2	68.2	4.60	740	90.5	2.10	0.60	99.3	0.10
BIC	44.7	54.2	1.10	17.0	82.7	0.30	2.20	97.8	0.00
HQC	35.6	63.0	1.40	11.8	87.8	0.40	1.30	98.7	0.00
TPM	10.8	74.2	15.0	0.10	84.7	15.2	0.00	68.5	31.5

Note: Entries give the percentage of Monte Carlo replications in which an r-state model is selected.

TABLE II

MONTE CARLO RESULTS FOR MSAR(2, 1)
DGP $\mu^{(1)} = 0$, $\mu^{(2)} = 3$, $\sigma^{(1)} = 1$, $\sigma^{(2)} = \sqrt{1.5}$

	$T = 100$			$T = 200$			$T = 400$		
r	1	2	3	1	2	3	1	2	3
$\phi_1^{(1)} = \phi_1^{(2)} = 0.3$									
$(p_{11}, p_{22}) = (0.6, 0.4)$									
AIC	43.0	55.3	1.70	24.7	74.1	1.20	5.70	93.5	0.80
BIC	94.7	5.20	0.10	91.6	8.40	0.00	80.6	19.4	0.00
HQC	73.7	26.2	0.10	62.5	37.5	0.00	33.1	66.9	0.00
TPM	34.3	45.1	20.6	5.40	70.4	24.2	0.00	83.3	16.7
$(p_{11}, p_{22}) = (0.9, 0.9)$									
AIC	8.20	91.1	0.70	0.60	99.3	0.10	0.00	100	0.00
BIC	63.1	36.9	0.00	27.6	72.4	0.00	1.10	98.9	0.00
HQC	26.1	73.9	0.00	4.50	95.5	0.00	0.10	99.9	0.00
TPM	0.00	94.6	5.40	0.00	90.5	9.50	0.00	83.4	16.6
$(p_{11}, p_{22}) = (0.9, 0.98)$									
AIC	22.3	75.6	2.10	4.50	95.0	0.50	0.10	99.9	0.00
BIC	65.4	43.6	0.00	23.4	76.6	0.00	1.70	98.3	0.00
HQC	36.7	63.2	0.10	9.90	90.0	0.10	0.50	99.5	0.00
TPM	5.70	77.9	16.4	0.10	67.3	32.6	0.00	65.2	35.8
$\phi_1^{(1)} = 0.3, \phi_1^{(2)} = 0.6$									
$(p_{11}, p_{22}) = (0.6, 0.4)$									
AIC	14.7	84.7	0.60	2.50	97.4	0.01	0.00	99.6	0.40
BIC	20.9	79.1	0.00	4.90	95.1	0.00	0.00	100	0.00
HQC	18.2	81.8	0.00	3.40	96.6	0.00	0.00	100	0.00
TPM	0.00	91.8	20.6	0.00	95.5	4.50	0.00	95.9	4.10
$(p_{11}, p_{22}) = (0.9, 0.9)$									
AIC	0.00	99.6	0.40	0.00	99.5	0.50	0.00	100	0.00
BIC	0.10	99.9	0.00	0.00	100	0.00	0.00	100	0.00
HQC	0.00	100	0.00	0.00	100	0.00	0.00	100	0.00
TPM	0.00	90.9	9.10	0.00	84.4	15.6	0.00	87.6	12.4
$(p_{11}, p_{22}) = (0.9, 0.98)$									
AIC	0.00	99.9	0.10	0.00	100	0.00	0.00	100	0.00
BIC	0.10	99.9	0.00	0.00	100	0.00	0.00	100	0.00
HQC	0.00	100	0.00	0.00	100	0.00	0.00	100	0.00
TPM	5.50	84.8	9.70	0.00	87.6	12.4	0.00	86.4	13.6
$\phi_1^{(1)} = 0.3, \phi_1^{(2)} = 0.9$									
$(p_{11}, p_{22}) = (0.6, 0.4)$									
AIC	1.60	98.4	0.00	0.00	100	0.00	0.00	100	0.00
BIC	2.00	98.0	0.00	0.00	100	0.00	0.00	100	0.00
HQC	1.90	98.1	0.00	0.00	100	0.00	0.00	100	0.00
TPM	0.00	99.2	0.08	0.00	100	0.00	0.00	100	0.00
$(p_{11}, p_{22}) = (0.9, 0.9)$									
AIC	0.00	100	0.00	0.00	100	0.00	0.00	100	0.00
BIC	0.00	100	0.00	0.00	100	0.00	0.00	100	0.00
HQC	0.00	100	0.00	0.00	100	0.00	0.00	100	0.00
TPM	0.00	100	0.00	0.00	100	0.00	0.00	100	0.00
$(p_{11}, p_{22}) = (0.9, 0.98)$									
AIC	0.00	100	0.00	0.00	100	0.00	0.00	100	0.00
BIC	0.00	100	0.00	0.00	100	0.00	0.00	100	0.00
HQC	0.00	100	0.00	0.00	100	0.00	0.00	100	0.00
TPM	0.00	100	0.00	0.00	100	0.00	0.00	100	0.00

Note: Entries give the percentage of Monte Carlo replications in which an r-state model is selected.

sample size is small, parameter changes are moderate, or the Markov chain is not persistent, the TPM is far superior to the procedures based on the AIC, BIC or HQC. The latter represent an improvement over the TPM only in cases where the changes in parameters are substantial and the number of observations is large. Unlike procedures based on complexity-penalized likelihood measures, underestimation of the state dimension by the TPM is extremely rare.

The results of the experiments for the MSAR(3, 1) DGP are summarized in Table III. The most striking feature of the results is the poor performance of the selection procedures when the autoregressive coefficient is constant. In such cases, the AIC, BIC, HQC and TPM all underestimate the state dimension most of the time (the best result being a 26% success rate for the TPM when $\mathbf{P} = \mathbf{P}_2$ and $T = 400$). The performance of the selection procedures improves considerably when the autoregressive coefficient is state-dependent. The TPM does much better than the complexity-penalized likelihood criteria for the smaller change in the mean parameter, making very few mistakes when $T = 400$. Even for larger changes in the mean parameter, the TPM performs better than the AIC, BIC and HQC, which underestimate the state dimension considerably when $T = 100$.

In summary, two general conclusions can be drawn from the discussion above:

(i) All selection procedures have difficulty in choosing the correct number of Markov regimes for time series of relatively short length and/or DGPs with state-independent autoregressive coefficients.
(ii) Unless the changes in parameters between regimes are quite substantial, procedures based on the BIC, HQC and, to a lesser degree, the AIC have a tendency to underestimate the state dimension; the TPM is the best selection method overall.

4. EMPIRICAL EXAMPLES

We now use some examples involving real-world data sets to illustrate the practical use of the selection procedures discussed before. More specifically, we analyse three time series that recent studies have found to be described well by Markov-switching autoregressive models.

1. The quarterly three-month US Treasury bill rates for the period 1962:1–1987:3.
2. The quarterly percentage changes in real US GNP for the period 1953:2–1984:4.
3. The quarterly real three-month US Treasury bill rates for the period 1961:1–1986:4 (calculated as the difference between the nominal interest rate and the inflation rate)[8].

[8]This is the time series obtained by extracting the end-of-quarter figures from the monthly data set of Mishkin (1990).

TABLE III

Monte Carlo Results for MSAR (3, 1) DGP $\sigma^{(1)} = 1$, $\sigma^{(2)} = \sqrt{1.5}$, $\sigma^{(2)} = 1.5$

	$T = 100$			$T = 200$			$T = 400$		
r	1	2	3	1	2	3	1	2	3
$\mu^{(1)} = 0$, $\mu^{(2)} = 1$, $\mu^{(3)} = 2$, $\phi_1^{(1)} = \phi_1^{(2)} = \phi_1^{(3)} = 0.3$									
$\mathbf{P} = \mathbf{P}_1$									
AIC	65.5	33.4	1.50	59.5	40.0	0.50	52.5	47.2	0.30
BIC	97.9	2.10	0.00	99.0	0.10	0.00	99.1	0.90	0.00
HQC	89.1	10.8	0.10	89.9	10.1	0.00	87.9	12.1	0.00
TPM	86.4	4.80	8.80	88.5	2.80	8.70	84.6	4.70	19.7
$\mathbf{P} = \mathbf{P}_2$									
AIC	52.4	45.4	2.20	32.1	66.8	1.10	7.80	92.1	0.10
BIC	93.2	6.80	0.00	88.8	11.2	0.00	68.9	31.1	0.00
HQC	77.7	22.1	0.20	62.9	37.1	0.00	31.9	68.1	0.00
TPM	86.0	4.60	9.40	80.3	6.40	13.3	59.9	13.7	26.4
$\mu^{(1)} = 0$, $\mu^{(2)} = 1$, $\mu^{(3)} = 2$, $\phi_1^{(1)} = 0.3$, $\phi_1^{(2)} = 0.6$, $\phi_1^{(3)} = 0.9$									
$\mathbf{P} = \mathbf{P}_1$									
AIC	5.00	87.6	7.40	0.50	83.4	16.1	0.40	79.7	19.9
BIC	20.6	77.9	1.50	8.90	83.7	7.40	2.50	80.0	17.5
HQC	10.6	86.2	3.20	2.80	84.2	13.0	0.40	79.9	19.7
TPM	0.00	65.1	34.9	0.00	34.8	65.2	0.00	7.50	92.5
$\mathbf{P} = \mathbf{P}_2$									
AIC	6.50	76.8	16.7	0.80	73.4	25.8	0.00	72.2	27.8
BIC	26.8	70.7	2.50	12.0	78.1	9.90	2.00	77.3	20.7
HQC	15.4	77.8	6.80	4.00	77.2	18.8	0.10	75.6	24.3
TPM	0.00	63.7	36.3	0.00	32.4	67.6	0.00	6.00	94.0
$\mu^{(1)} = 0$, $\mu^{(2)} = 3$, $\mu^{(3)} = 6$, $\phi_1^{(1)} = \phi_1^{(2)} = \phi_1^{(3)} = 0.3$									
$\mathbf{P} = \mathbf{P}_1$									
AIC	56.4	39.0	4.60	47.3	47.8	4.90	27.4	66.6	0.00
BIC	92.7	7.30	0.00	88.9	11.0	0.10	80.0	20.0	0.00
HQC	77.0	22.6	0.40	69.2	30.7	0.10	50.8	48.8	0.40
TPM	92.3	2.00	5.70	90.9	2.50	6.60	85.9	4.40	9.70
$\mathbf{P} = \mathbf{P}_2$									
AIC	41.5	48.7	9.80	27.1	57.9	15.0	8.20	67.1	24.7
BIC	79.7	20.2	0.10	67.2	32.6	0.20	43.5	56.1	0.40
HQC	60.8	36.7	2.50	46.0	50.6	3.40	23.9	70.5	5.60
TPM	92.4	2.50	5.10	90.8	2.80	6.40	85.1	5.10	9.80
$\mu^{(1)} = 0$, $\mu^{(2)} = 3$, $\mu^{(3)} = 6$, $\phi_1^{(1)} = 0.3$, $\phi_1^{(2)} = 0.6$, $\phi_1^{(3)} = 0.9$									
$\mathbf{P} = \mathbf{P}_1$									
AIC	0.00	39.8	60.2	18.5	4.20	77.3	7.00	0.10	92.9
BIC	0.00	47.3	52.7	19.7	5.40	74.9	7.10	0.10	92.8
HQC	0.00	59.7	40.3	18.9	4.80	76.3	7.00	0.10	92.9
TPM	0.00	19.5	80.5	0.00	0.30	97.0	0.00	0.00	100
$\mathbf{P} = \mathbf{P}_2$									
AIC	0.00	57.0	43.0	5.50	6.40	88.1	1.70	0.20	98.1
BIC	0.00	72.7	27.3	7.00	8.20	84.8	1.70	0.20	98.1
HQC	0.00	64.7	35.3	5.70	7.00	87.3	1.70	0.20	98.1
TPM	0.00	15.7	84.3	0.00	1.20	98.8	1.90	0.00	98.1

Note: Entries give the percentage of Monte Carlo replications in which an r-state model is selected.

For the nominal interest rate data, Hamilton (1988) found support for Markov regime switching using an MSAR(2, 4) model with state-independent autoregressive coefficients. For the real GNP series, Hamilton (1989) fitted a simplified version of the MSAR(2, 4) model with no switching in the innovation variance or the autoregressive coefficients. Hansen (1992) and McCulloch and Tsay (1994) opted, on the other hand, for a general MSAR(2, 4) model like the one in (1), while Albert and Chib (1993) argued in favour of a specification with mean switching but no autoregressive dynamics; see also Psaradakis (1998). Finally, Garcia and Perron (1996) fitted MSAR(3, 2) models, with constant and state-dependent autoregressive coefficients, to the real interest rate series.

Since the objective here is to determine the state dimension rather than the autoregressive order, we fix the latter at the value used in the papers mentioned before (i.e. $m = 4$ for the GNP and nominal interest rate, and $m = 2$ for the real interest rate). The estimated number of regimes obtained by complexity-penalized likelihood criteria and the TPM are reported in Table IV. For the nominal interest rate series, all three likelihood-based criteria select a linear single-regime authoregressive model, whereas the TPM selects a model with at least two Markov regimes. Consistent with much of the related literature, a model with two regimes is chosen by AIC and TPM in the case of the real GNP data, while the more conservative BIC and HQC select the single-regime model. Finally, the preferred specification for the real interest rate is a Markov-switching model with three (or more) regimes according to the AIC and TPM, while BIC and HQC choose a single-regime specification.

These results are consistent with out earlier simulation findings, which revealed BIC and HQC to have a tendency to underestimate the true state dimension. To throw more light on the issue, we carried out some more Monte Carlo experiments using the models estimated for the three real-world time series as the DGP (assuming the errors $\{\varepsilon_t\}$ are Gaussian i.i.d.). More specifically, the parameter values were chosen as follows:

- Model 1 (Hamilton, 1988, Table 3)

  $\mu^{(1)} = 1.633,\ \mu^{(2)} = 2.821,\ \sigma^{(1)} = 0.1752,\ \sigma^{(2)} = 0.7326,\ \phi_1^{(1)} = \phi_1^{(2)} = 0.8778,$
  $\phi_2^{(1)} = \phi_2^{(2)} = 0.0675,\ \phi_3^{(1)} = \phi_3^{(2)} = 0.1692,\ \phi_4^{(1)} = \phi_4^{(2)} = -0.2108,$
  $p_{11} = 0.9899,\ p_{22} = 0.9087$

- Model 2 (Hansen, 1992, Table 8)

  $\mu^{(1)} = -0.690,\ \mu^{(2)} = 1.125,\ \sigma^{(1)} = 0.657,\ \sigma^{(2)} = 0.670,\ \phi_1^{(1)} = 0.321,$
  $\phi_1^{(2)} = 0.316,\ \phi_2^{(1)} = 0.510,\ \phi_2^{(2)} = -0.086,\ \phi_3^{(1)} = -0.078,\ \phi_3^{(2)} = -0.072$
  $\phi_4^{(1)} = -0.022,\ \phi_4^{(2)} = -0.012,\ p_{11} = 0.388,\ p_{22} = 0.638$

TABLE IV
ESTIMATED STATE DIMENSION OF EMPRICAL MODELS

Time series	AIC	BIC	HQC	TPM
Nominal interest rate	1	1	1	2
Real output	2	1	1	2
Real interest rate	3	1	1	3

- Model 3 (Garcia and Perron, 1996, Table 1)
  $\mu^{(1)} = -0.375,\ \mu^{(2)} = 1.868,\ \mu^{(3)} = 5.675,\ \sigma^{(1)} = 1.873,\ \sigma^{(2)} = 1.150,$
  $\sigma^{(3)} = 1.8731,\ \phi_1^{(1)} = \phi_1^{(2)} = \phi_1^{(3)} = 0.029,\ \phi_2^{(1)} = \phi_2^{(2)} = \phi_2^{(3)} = -0.043,$
  $p_{11} = 0.967,\ p_{22} = 0.985, p_{12} = 0.000, p_{21} = 0.015,$
  $p_{31} = 0.000,\ p_{32} = 0.048$

- Model 4 (Garcia and Perron, 1996, Table 4)
  $\mu^{(1)} = -0.498,\ \mu^{(2)} = -0.164,\ \mu^{(3)} = 2.582,\ \sigma^{(1)} = 1.817,\ \sigma^{(2)} = 1.163,$
  $\sigma^{(3)} = 1.778,\ \phi_1^{(1)} = 0.012,\ \phi_1^{(2)} = 0.057,\ \phi_1^{(3)} = -0.009,\ \phi_2^{(1)} = -0.245,$
  $\phi_2^{(2)} = 0.860,\ \phi_2^{(3)} = 0.470, p_{11} = 0.966,\ p_{22} = 0.985, p_{12} = 0.000,$
  $p_{21} = 0.015, p_{31} = 0.000,\ p_{32} = 0.050$

Table V presents the results of the experiments for three different sample sizes (based on 1000 Monte-Carlo replications). In the case of Models 1 and 2, it is clearly the AIC that does the best job in selecting the two-regime model, especially for time series of relatively short length. For Models 3 and 4, the TPM is by far the most successful method of choosing the correct number of Markov regimes. As before, the performance the BIC and HQC leaves much to be desired as they generally tend to underestimate the state dimension.

TABLE V
MONTE CARLO RESULTS FOR EMPIRICAL MODELS

	T								
	100			150			300		
r	1	2	3	1	2	3	1	2	3
Model 1									
AIC	13	87	0	4	96	0	3	97	0
BIC	99	1	0	90	10	0	27	73	0
HQC	98	2	0	8	92	0	3	97	0
TPM	45	55	0	27	73	0	8	92	0
Model 2									
AIC	27	73	0	3	97	0	1	99	0
BIC	99	1	0	53	47	0	2	98	0
HQC	62	38	0	10	90	0	1	99	0
TPM	37	55	8	15	78	7	5	93	2
Model 3									
AIC	17	66	17	15	42	43	4	22	74
BIC	34	64	2	35	64	1	21	36	43
HQC	21	68	11	22	49	29	12	28	60
TPM	16	68	16	10	12	78	1	3	96
Model 4									
AIC	20	63	17	18	39	43	3	19	78
BIC	55	44	1	46	47	7	20	37	43
HQC	35	60	5	31	44	25	9	30	61
TPM	2	13	85	1	10	89	0	4	96

Note: Entries give the percentage of Monte Carlo replications in which an r-state model is selected.

5. CONCLUSION

In this paper we have examined the performance of various procedures for selecting the number of regimes in Markov-switching autoregressive models. We have considered selection procedures based on the ARMA representation that a Markov-switching autoregressive process admits as well as procedures based on complexity-penalized likelihood criteria. Our Monte Carlo analysis has revealed that selection procedures based on the TPM and the AIC are generally successful in choosing the correct state dimension, provided that the sample size and parameter changes are not too small. The BIC and HQC have a tendency to underestimate the state dimension. The good overall performance of the TPM method, combined with very low computational costs, makes it the best choice for selecting a lower bound for the number of Markov-regimes in switching models.

In closing, it should be noted that a limiting feature of our analysis has been the assumption that the autoregressive order of the Markov-switching model is known. In practice, one is typically faced with the problem of determining both the autoregressive order of the model and the state dimension of the hidden Markov chain. Extending the procedures examined here to deal with the problem of joint order determination is an important topic for future research.

APPENDIX: STATISTICAL ARMA ORDER IDENTIFICATION

For completeness, this Appendix provides a brief description of Choi's (1993) method of determining the order of an ARMA(p, q) model by means of χ^2 test criteria associated with the so-called three-pattern identification method.

Given an observed time series $\{X_t, 1 \leqslant t \leqslant T\}$, Let $\hat{\rho}_\tau$ be the associated sample autocorrelations defined in the usual way as

$$\hat{\rho}_\tau = \frac{\hat{\gamma}_\tau}{\hat{\gamma}_0}$$

$$\hat{\gamma}_\tau = \frac{1}{T}\sum_{t=1}^{T-|\tau|}(X_t - \bar{X})(X_{t+|\tau|} - \bar{X}) \qquad \tau = 0, \pm 1, \ldots, \pm(T-1)$$

with $\bar{X} = (1/T)\sum_{t=1}^{T} X_t$. Further, for $0 \leqslant k,\ i \leqslant M \ll T$, let $\hat{\boldsymbol{\rho}}(k,i) = (\hat{\rho}_{i+1}, \ldots, \hat{\rho}_{i+k})'$ and

$$\hat{\mathbf{B}}(k,i) = \begin{pmatrix} \hat{\rho}_i & \hat{\rho}_{i-1} & \cdots & \hat{\rho}_{i-k+1} \\ \hat{\rho}_{i+1} & \hat{\rho}_i & \cdots & \hat{\rho}_{i-k+2} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\rho}_{i+k-1} & \hat{\rho}_{i+k-2} & \cdots & \hat{\rho}_i \end{pmatrix}$$

Then, if $\hat{\mathbf{B}}(k,i)$ is nonsingular, define the quantities

$$\hat{\boldsymbol{\varphi}}(k,i) = \hat{\mathbf{B}}(k,i)^{-1}\hat{\boldsymbol{\rho}}(k,i) \qquad \text{and} \qquad \hat{\theta}(k,i) = \hat{\rho}_{i+k+1} - \hat{\boldsymbol{\varphi}}^*(k,i)'\hat{\boldsymbol{\rho}}(k,i)$$

where $\mathbf{z}^* = (z_n, z_{n-1}, \ldots, z_1)'$ denotes the reversed entry vector for $\mathbf{z} = (z_1, \ldots, z_{n-1}, z_n)'$. If $\hat{\mathbf{B}}(k,i)$ is singular, $\hat{\boldsymbol{\varphi}}(k,i)$ and $\hat{\theta}(k,i)$ are instead defined as

$$\hat{\boldsymbol{\varphi}}(k,i) = \hat{\mathbf{B}}(k,i)^{+}\hat{\boldsymbol{\rho}}(k,i) \qquad \text{and} \qquad \hat{\theta}(k,i) = \hat{\rho}_{i+k+1} - \hat{\boldsymbol{\rho}}^*(k,i)'\hat{\mathbf{B}}(k,i)^{+}\hat{\boldsymbol{\rho}}(k,i)$$

where $\mathbf{A}^+$ denotes the Moore–Penrose pseudo-inverse of the matrix $\mathbf{A}$.

The statistical identification of p and q is based on the statistics

$$E_\ell(k,i) = T\mathbf{u}_\ell(k,i)'\hat{\mathbf{W}}_\ell^+\mathbf{u}_\ell(k,i) \qquad \text{and} \qquad J_\ell(k,i) = T\mathbf{v}_\ell(k,i)'\hat{\mathbf{W}}_\ell^+\mathbf{v}_\ell(k,i)$$

where, for some small integer $\ell > 0$,

$$\mathbf{u}_\ell(k,i) = (\hat{\theta}(k,i), \hat{\theta}(k,i+1), \ldots, \hat{\theta}(k,i+\ell-1))'$$
$$\mathbf{v}_\ell(k,i) = (\hat{\theta}(k,i), \hat{\theta}(k+1,i), \ldots, \hat{\theta}(k+\ell-1,i))'$$

and $\hat{\mathbf{W}}_\ell(k,i)$ is an ℓ-dimensional symmetric Toeplitz matrix whose (a, b) element is

$$\frac{1}{\hat{\gamma}_0^2}\sum_{j=-i}^{i}\left(\sum_{h=0}^{k}\sum_{s=0}^{k}\hat{\varphi}_{k,h}^{(i)}\hat{\varphi}_{k,s}^{(i)}\hat{\gamma}_{j+h-s}\right)\left(\sum_{h=0}^{k}\sum_{s=0}^{k}\hat{\varphi}_{k,h}^{(i)}\hat{\varphi}_{k,s}^{(i)}\hat{\gamma}_{j+a-b+h-s}\right)$$

with $\hat{\varphi}_{k,s}^{(i)}$ being the sth element of $\hat{\boldsymbol{\varphi}}(k,i)$ if $s \geqslant 1$ and $\hat{\varphi}_{k,0}^{(i)} = -1$.

If $\{X_t, 1 \leqslant t \leqslant T\}$ is a realization from an ARMA(p, q) process, then $E_\ell(k, i)$ has a χ_ℓ^2 asymptotic distribution if $k = p$ and $i \geqslant q$, while $J_\ell(k,i)$ has χ_ℓ^2 asymptotic distribution if $k \geqslant p$ and $i = q$. Hence, the estimated orders $\hat{p}$ and $\hat{q}$ are obtained as the smallest k and i, respectively, for which $E_\ell(k, i) < \chi_\ell^2(\alpha)$ and $J_\ell(k, i) < \chi_\ell^2(\alpha)$, where $\chi_\ell^2(\alpha)$ is the αth quantile of the χ_ℓ^2 distribution. The selection is then confirmed by checking that

$$E_\ell(\hat{p}, \hat{q}-1) \geqslant \chi_\ell^2(\alpha)$$
$$J_\ell(\hat{p}-1, \hat{q}) \geqslant \chi_\ell^2(\alpha)$$
$$E_\ell(\hat{p}, i) < \chi_\ell^2(\alpha) \qquad \text{for } i \geqslant \hat{q}+1$$

and

$$J_\ell(k, \hat{q}) < \chi_\ell^2(\alpha) \qquad \text{for } k \geqslant \hat{p}+1$$

In out implementation of this selection procedure, we follow Choi (1993) in taking $M = 8$, $\ell = 3$, and $\alpha = 0.95$.

ACKNOWLEDGEMENT

The second author is grateful for the hospitality he received in the Department of Economics at the University of California at San Diego, where part of the research was carried out. Conversations with Clive Granger and James Hamilton on the issues discussed in this paper were very helpful and are gratefully acknowledged.

REFERENCES

Akaike, H. (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control* AC–19, 716–23.

Albert, J. H. and Chib, S. (1993) Bayes inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts. *Journal of Business and Economic Statistics* 11, 1–15.

Anděl, J. (1993) A time series model with suddenly changing parameters. *Journal of Time Series Analysis* 14, 111–23.

Choi, B. S. (1992) *ARMA Model Identification*. New York: Springer-Verlag.
——— (1993) Two chi-square statistics for determining the orders p and q of an ARMA(p, q) process. *IEEE Transactions on Signal Processing* 41, 2165–76.
Davidson, J. (2001) Establishing conditions for the functional central limit theorem in nonlinear and semiparametric time series processes. *Journal of Econometrics* 106, 243–69.
Francq, C. and Zakoïan, J.-M. (1998) Estimating linear representations of nonlinear processes. *Journal of Statistical Planning and Inference* 68, 145–65.
——— and ——— (2001) Stationarity of multivariate Markov-switching ARMA models. *Journal of Econometrics* 102, 339–64.
Garcia, R. (1998) Asymptotic null distribution of the likelihood ratio test in Markov switching models. *International Economic Review* 39, 763–88.
——— and Perron, P. (1996) An analysis of the real interest rate under regime shifts. *Review of Economics and Statistics* 78, 111–25.
Granger, C. W. J., King, M. L. and White, H. (1995) Comments on testing economic theories and the use of model selection criteria. *Journal of Econometrics* 67, 173–87.
Hamilton, J. D. (1988) Rational-expectations econometric analysis of changes in regime: An investigation of the term structure of interest rates. *Journal of Economic Dynamics and Control* 12, 385–423.
——— (1989) A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* 57, 357–84.
——— (1994) *Time Series Analysis*. Princeton: Princeton University Press.
Hannan, E. J. and Quinn, B. G. (1979) The determination of the order of an autoregression. *Journal of the Royal Statistical Society B* 41, 190–5.
Hansen, B. E. (1992) The likelihood ratio test under nonstandard conditions: Testing the Markov switching model of GNB. *Journal of Applied Econometrics* 7, S61–S82 (Erratum: 11, 195–8).
Kapetanios, G. (1999) Essays on the econometric analysis of threshold models, PhD thesis, University of Cambridge.
——— (2001) Model selection in threshold models. *Journal of Time Series Analysis* 22, 733–54.
Leroux, B. G. (1992) Consistent estimation of a mixing distribution. *Annals of Statistics* 20, 1350–60.
——— and Puterman, M. L. (1992) Maximum-penalized-likelihood estimation for independent and Markov-dependent mixture models. *Biometrics* 48, 545–58.
McCulloch, R. E. and Tsay, R. S. (1994) Statistical analysis of economic time series via Markov switching models. *Journal of Time Series Analysis* 15, 523–39.
McLachlan, G. J. (1987) On bootstrapping the likelihood ratio test statistic for the number of components in a normal mixture. *Applied Statistics* 36, 318–24.
Mishkin, F. S. (1990) What does the term structure tell us about future inflation? *Journal of Monetary Economics* 25, 77–95
Nishii, R. (1988) Maximum likelihood principle and model selection when the true model is unspecified. *Journal of Multivariate Analysis* 27, 392–403.
Poskitt, D. S. and Chung, S.-H. (1996) Markov chain models, time series analysis and extreme value theory. *Advances in Applied Probability* 28, 405–25.
Psaradakis, Z. (1998) Bootstrap-based evaluation of Markov-switching time series models. *Econometric Reviews* 17, 275–88.
Rissanen, J. (1978) Modelling by shortest data description. *Automatica* 14, 465–71.
Rydén, T. (1995) Estimating the order of hidden Markov models. *Statistics* 26, 345–54.
Schwarz, G. (1978) Estimating the dimension of a model. *Annals of Statistics* 6, 461–4.
Sin, C.-Y. and White, H. (1996) Information criteria for selecting possibly misspecified parametric models. *Journal of Econometrics* 71, 207–25.
Yao, J. (2001) On square-integrability of an AR process with Markov switching. *Statistics and Probability Letters* 52, 265–70.
Zhang, J. and Stine, R. A. (2001) Autocovariance structure of Markov regime switching models and model selection. *Journal of Time Series Analysis* 22, 107–24.