INTERNATIONAL ECONOMIC REVIEW
Vol. 39, No. 3, August 1998

ASYMPTOTIC NULL DISTRIBUTION OF THE LIKELIHOOD RATIO TEST IN MARKOV SWITCHING MODELS*

BY RENÉ GARCIA[†1]

Université de Montréal, Canada

Markov switching (MS) models raise a problem known as testing hypotheses when a nuisance parameter is not identified under the null hypothesis. I show that the asymptotic distribution theory used for testing in presence of such a problem appears to work also for MS models, even though its validity can be questioned because of identically zero scores under the null estimates. Assuming the validity of this distribution theory, I derive the asymptotic null distribution of the likelihood ratio (LR) test for various MS models. Monte Carlo experiments show that the LR asymptotic distributions approximate the empirical distributions very well.

1. INTRODUCTION

The Markov Switching (MS) Model, introduced by Hamilton (1988, 1989), has been used in numerous economic and financial applications where changes in regime potentially play an important role.[2] In the most general form of this nonlinear model, the mean, the variance and the autoregressive structure of a time series can be made dependent upon a state or regime, the realization of which is governed by a discrete-time, discrete-state Markov stochastic process. While estimation methods for these models are by now well established (Coslett and Lee 1984, Hamilton 1988, 1989, and Boldin 1989), such is not the case for testing procedures, apart from a few exceptions: Hamilton (1996) proposes some specification tests based on the Lagrange multiplier principle to test, for example, various forms of autocorrelation, generalized ARCH effects, and omitted explanatory variables in both the mean and variance; Boldin (1989) uses the Davies' (1987) upper bound test to determine the number of regimes; and Garcia and Perron (1996) use Gallant's

* Manuscript received January 1995; revised February 1997.

† E-mail: *garcia@plgcn.umontreal.ca*

[1] This paper is based on chapters 3 and 4 of my Ph.D. dissertation at Princeton University. I would like to thank Pierre Perron, my thesis adviser, for his insightful advice, and Michael D. Boldin, Eric Ghysels, Fangxiong Gong, Christian Gouriéroux, James D. Hamilton, Bruce E. Hansen, Werner Ploberger, and Michael Rockinger for helpful discussions and comments. I am also grateful to three anonymous referees for their helpful and constructive comments that improved both the presentation and content of an earlier version of the paper. Financial support from the Social Sciences and Humanities Research Council of Canada, the Fonds pour la Formation des chercheurs et l'Aide à la Recherche, and from the PARADI project funded by the Canadian International Development Agency is gratefully acknowledged.

[2] Cecchetti et al., (1990), Engel and Hamilton (1990), Garcia and Perron (1996), Hamilton (1988, 1989), Hassett (1990), Turner et al., (1989).

(1977) test and a J-test for nonnested models (Davidson and MacKinnon 1981), along with the Davies' test, to determine the number of regimes. The use of these nonstandard tests can be explained by the fact that MS models raise a special problem, known in the statistics literature as testing hypotheses in models where a nuisance parameter is not identified under the null hypothesis.

Davies (1977, 1987), Andrews and Ploberger (1994), and Hansen (1991, 1996) study the problem of inference in the presence of nuisance parameters under the null hypothesis. The asymptotic distributions of the usual tests are nonstandard and generally depend upon the covariance function of chi-square processes. This covariance is often model and data-dependent, which precludes the tabulation of critical values. Hansen (1991, 1996) proposes a simulation method to approximate the asymptotic distributions and applies it to the threshold model. In Hansen (1992), the author proposes another approach when, in addition to the problem of unidentified nuisance parameters, the econometrician is faced with identically zero scores under the null hypothesis, which is the case in MS models. The author considers the likelihood function as a function of unknown parameters and uses empirical process theory to bound the asymptotic distribution of a standardized likelihood ratio statistic. For the simplest MS model where the mean of the process switches with the state, the method involves the evaluation of the likelihood across a grid of different values for the mean and transition probability parameters, the parameters of interest in this model. For each set of such values, the constrained likelihood function needs to be optimized with respect to the nuisance parameters of the model. Even for the small dimension of the space of parameters of interest in this simple model, the computational burden can be important if the grid search over the parameter space is extensive. Therefore the use of this method for the more elaborate MS models in practice does not appear very promising. Moreover, Hansen's (1992) method provides a bound for the likelihood ratio statistic and not a critical value, which means that the tests may be conservative.

In this paper, I avoid this computational burden by treating the transition probability parameters as nuisance parameters and by setting the null hypothesis solely in terms of the parameters (mean, variance or autoregressive coefficients) governed by the Markov variable. By reformulating the problem in this way, I proceed as if the distribution theory proposed by Davies (1987), Andrews and Ploberger (1994) and Hansen (1991, 1996) applied to the MS model. According to this theory, the LR test can be written as the supremum of a chi-square functional, asymptotically under the null hypothesis. Under the assumed validity[3] of this framework, I derive analytically the asymptotic null distribution of the likelihood ratio test and the related covariance functions for various two-state MS models. Monte-Carlo experiments show that the derived asymptotic distribution offers a very good approximation to the empirical distribution.

In Section 2, I present a general two-state MS model, explain the problem of nonidentification of some nuisance parameters under the null hypothesis, and set up

[3] According to Hansen (1991) and Andrews and Ploberger (1994), the sufficient assumptions needed for this to hold are not fulfilled in the MS model. Most notably, the matrix of second derivatives evaluated at the null estimates appears to be singular.

the testing problem as the supremum of likelihood ratio statistics. I also make clear how this approach differs from Hansen's methodology based on empirical process theory. In Section 3, I briefly present the asymptotic distribution theory proposed by Davies (1987), Andrews and Ploberger (1994), and Hansen (1991, 1996) for models where nuisance parameters are not identified under the null, and derive the covariance function for the general two-state MS model. Section 4 provides the asymptotic null distributions of the LR statistic for three specific MS models used by various authors to capture changes in regime in economic and financial time series. These models differ by the specification of the noise function in the general two-state MS model. We also compute the power of the LR test for these three models. Section 5 concludes.

2. TESTING IN THE CONTEXT OF MARKOV SWITCHING MODELS

Consider the following two-state MS model:[4]

$$y_t = \alpha_0 + \alpha_1 S_t + z_t \tag{1}$$

$$z_t = \phi_1 z_{t-1} + \cdots + \phi_r z_{t-r} + (\omega_0 + \omega_1 S_t)\epsilon_t$$

$$P(S_t = 1 | S_{t-1} = 1) = p$$

$$P(S_t = 0 | S_{t-1} = 0) = q \qquad (t = 1, \ldots, n)$$

where $\{y_t\}_{-r}^{n}$ is a stationary process and ϵ_t is i.i.d. $\mathbb{N}(0,1)$. Assume one wants to test the null hypothesis of a linear model against the alternative hypothesis of a MS model. The null hypothesis can be expressed as either $\{\alpha_1 = 0,\ \omega_1 = 0\}$ or $\{p = 0\}$ or $\{p = 1\}$. It can readily be seen that if α_1 and ω_1 are equal to zero, the transition probability parameters p and q are unidentified. Moreover, at the constrained maximum likelihood estimates, the score vector (the derivative of the log-likelihood function with respect to $(\alpha_0, \alpha_1, \omega_0, \omega_1, p, q)$ is identically zero. Under these conditions, the likelihood ratio, Lagrange multiplier, and Wald tests do not have a standard asymptotic distribution. This is the point of departure of Hansen's (1992) analysis regarding the likelihood ratio test under nonstandard conditions, since the two problems of unidentified nuisance parameters under the null and identically zero scores are present. In his analysis, the parameter vector is first split into two sub-vectors: a sub-vector of parameters of interest $\alpha = (\alpha_1, \omega_1, p, q)$, and a sub-vector of nuisance parameters $\theta = (\alpha_0, \omega_0, \phi_1, \ldots, \phi_r)$. The sub-vector α is further partitioned into $\beta = (\alpha_1, \omega_1)$ and $\gamma = (p, q)$. The null hypothesis takes the form: H_0: $\beta = 0$, and under H_0 γ is not identified. The next step consists in concentrating the log-likelihood function $L_n(\alpha, \theta)$ with respect to θ given fixed values of α. The concentrated log-likelihood function can then be written as: $\hat{L}_n(\alpha) = L_n(\alpha, \hat{\theta}(\alpha))$.

[4] This specification encompasses the two-state MS models most frequently used in the literature. Two exceptions are noteworthy: the state-dependent autoregressive specification used in Garcia and Perron (1996) and Hansen (1992), or the time-varying transition probability model used in Diebold et al., (1993) and Filardo (1994).

The likelihood ratio process can therefore be defined as: $L\hat{R}_n(\alpha) = \hat{L}_n(\beta, \gamma) - \hat{L}_n(0, \gamma)$ and the LR statistic as: $L\hat{R}_n = \text{Sup}_{\alpha \in A} \hat{LR}_n(\alpha)$. To construct his test, Hansen (1992) standardizes this statistic so that all values of α yield the same variance, and bounds the distribution function of this standardized statistic.

Although the method is appealing since it addresses both above-mentioned violations of the conventional regularity conditions, it has two main shortcomings. First, it runs rapidly into computational limitations. Hansen's (1992) testing procedure requires setting a grid for each element of the parameter vector α, that is for each parameter depending on the Markov variable S_t, plus p and q. In a model where the mean and the variance of the series change with the state, this means a grid over four parameters. To stay computationally tractable, it is necessary to limit the number of grid points. Although the test results do not seem to be too sensitive to the choice of grid for the simple models presented by the author, this limitation will become serious for more elaborate models. Second, Hansen's (1992) method provides a bound for the likelihood ratio statistic and not a critical value, which means that the tests may be conservative (under-rejection of the null hypothesis when it is true).

In this approach, I treat the transition probability parameters p and q as nuisance parameters and the remaining parameters as parameters of interest. Hence, I partition the parameter vector differently. The sub-vector θ is augmented by α_1 and ω_1, that is, $\theta = (\alpha_0, \omega_0, \alpha_1, \omega_1, \phi_1, \ldots, \phi_r)$, while p and q are kept in the γ vector. The null hypothesis still takes the form: H_0: $\alpha_1 = 0$, $\omega_1 = 0$, and under H_0 γ is not identified. I can therefore derive the likelihood ratio statistic for each such set of values of the two transition probability parameters over a certain parameter space, say Γ, where p and q lie. The likelihood ratio of the original problem is then the supremum over Γ of the likelihood ratios obtained for each particular set of values of the p and q parameters. Formally, define LR_n and $LR_n(\gamma)$ as follows:

$$LR_n = 2n\left[Q_n\left(\hat{\theta}, \hat{\gamma}\right) - Q_n(\tilde{\theta})\right] \tag{2}$$

$$LR_n(\gamma) = 2n\left[Q_n\left(\hat{\theta}(\gamma), \gamma\right) - Q_n(\tilde{\theta})\right]$$

where Q_n, the average log-likelihood function of a sample of n observations, is given by:

$$Q_n(\theta, \gamma) = \frac{1}{n} \log p(y_n, \ldots, y_1; \theta, \gamma) \tag{3}$$

The first statistic, LR_n, refers to the difference between the estimated unconstrained $(\hat{\theta}, \hat{\gamma})$ and constrained $(\tilde{\theta})$ models. For the second $LR_n(\gamma)$, the maximizing value of θ under the alternative $(\hat{\theta}(\gamma))$ is obtained for a given value of γ. The statistics LR_n and $LR_n(\gamma)$ are related as follows:

$$LR_n = \sup_{\gamma \in \Gamma} LR_n(\gamma) \tag{4}$$

Therefore, the main difference between Hansen's (1992) approach and my own approach is in the treatment of α_1. Once this parameter is included in the θ vector, we are in the general framework of testing when a nuisance parameter is present only under the alternative (Davies 1987, Andrews and Ploberger 1994, Hansen 1991) and can rely apparently on their distributional theory. Moreover, I will show that one can derive the asymptotic covariance of $LR_n(\gamma)$ analytically. Hence, it becomes much easier computationally to simulate the asymptotic distribution of LR_n over as fine a grid as desired for γ. Of course, to keep the information matrix positive definite, the values of 0 and 1 have to be excluded from the metric space Γ. This problem arises also in structural change models when the timing of the change is an unknown fraction of the sample size (Andrews and Ploberger 1994). There remains one problem. For any value of γ, the score with respect to α_1 evaluated at the null estimates is identically zero. This means that some of the assumptions on which the above-mentioned distributional theory is based are not satisfied. In particular, the assumption of stochastic equicontinuity of the second derivative matrix does not appear to be satisfied, since this matrix is singular at the null estimates but not at the true parameter values. In the rest of the paper, I will proceed as if this distributional theory applied and see if the results obtained are reasonable.

Other results in the literature seem to suggest that the likelihood ratio test is robust to these types of irregularities, but not the LM or Wald tests. Lee and Chescher (1986) have investigated the problem of specification testing when the score vector evaluated at the restricted maximum likelihood estimates is identically zero. They show that even under such irregularities, the usual asymptotic distribution of the LR statistic is still valid. This is not the case of the LM or Wald tests which have to be modified to be asymptotically equivalent to the LR test. In a more general context, Dufour (1995) shows that likelihood ratio statistics behave relatively smoothly in the presence of identification problems, contrary to Wald-type statistics.

3. THE LIKELIHOOD RATIO TEST IN A GENERAL TWO-STATE MARKOV SWITCHING MODEL

In this section, I first state a theorem based on Davies (1977, 1987), Andrews and Ploberger (1994) and Hansen (1991, 1996), according to which the LR test is the supremum of a chi-square functional. In the second subsection, I derive analytically the covariance kernel of this chi-square functional for a general two-state MS model.

3.1. *The Asymptotic Distribution of the Likelihood Ratio Test.*

Several papers have studied the problem of inference in the presence of nuisance parameters that are not identified under the null hypothesis. Davies (1977, 1987) suggests the Sup LR or Sup LM tests, while Andrews and Ploberger (1994) consider a class of tests (average exponential LM, Wald and LR tests) that exhibit optimality properties in terms of weighted average power for particular weight functions (multivariate normal densities). The likelihood ratio test of the form $\sup_{\gamma \in \Gamma} LR_n(\gamma)$, which is not of the average exponential form, is not an optimal test using these weight functions. However, Andrews and Ploberger (1995) show that the sup LR test is a best test

against alternatives that are sufficiently distant from the null hypothesis. Hansen (1991, 1996) studies the asymptotic distribution theory of standard test statistics when nuisance parameters are not identified under the null hypothesis. Under certain regularity conditions, the asymptotic distributions of these standard test statistics are shown to be functionals of chi-square processes. In particular, for the likelihood ratio statistic, one can state the following theorem[5].

THEOREM 1.

$$LR_n \Rightarrow \operatorname{Sup} C \equiv \sup_{\gamma \in \Gamma} C(\gamma) \tag{5}$$

where $C(\gamma)$ *is a chi-square process with covariance matrix* $\bar{K}(.,.)$,[6] *defined as follows*:[7]

$$\bar{K}(\gamma_1, \gamma_2) = \iota_k V(\gamma_1)^{-1} K(\gamma_1, \gamma_2) V(\gamma_2)^{-1} \iota_k' \tag{6}$$

where ι_k *is a vector of dimension* k *(the dimension of the parameter space under the alternative) with ones in the positions of the parameters constrained to be zero under the null and*:

$$K(\gamma_1, \gamma_2) = \lim_{n \to \infty} nE\left[S_n^c(\theta_0, \gamma_1) S_n^c(\theta_0, \gamma_2)\right]$$

$$S_n^c(\theta, \gamma) = \frac{\partial}{\partial \theta} Q_n(\theta, \gamma)$$

$$V(\theta, \gamma) = \lim_{n \to \infty} nE\left[S_n^c(\theta, \gamma) S_n^c(\theta, \gamma)'\right]$$

$$V(\gamma) = V(\theta_0, \gamma). \tag{7}$$

This means that for any set of values $\{\gamma_1, \ldots, \gamma_T\}$ included in Γ, the set of score vectors $\{S_n^c(\theta_0, \gamma_1), \ldots, S_n^c(\theta_0, \gamma_T)\}$ is multivariate normal with mean zero and covariances $K(\gamma_j, \gamma_k) = \lim_{n \to \infty} nE[S_n^c(\theta_0, \gamma_j) S_n^c(\theta_0, \gamma_k)]$. I write these covariances as $K(\gamma_1, \gamma_2)$. One important condition to derive the asymptotic distribution of $LR_n(\gamma)$ is that $V(\gamma)$ is positive definite uniformly over Γ. If $V(\gamma)$ is singular for some values of γ, one must redefine Γ to exclude these values. As mentioned in the previous

[5] For the proof, see Hansen (1991).

[6] A process $Z(\gamma)$ is a chi-square process of order k in $\gamma \in \Gamma$ if it can be represented as $Z(\gamma) = G(\gamma)' K(\gamma, \gamma)^{-1} G(\gamma)$, where $G(\gamma)$ is a mean zero k-vector Gaussian process with covariance function $K(\gamma_1, \gamma_2) = E[G(\gamma_1) G(\gamma_2)']$.

[7] This assumes that $V(\theta, \gamma) = M(\theta, \gamma) = \lim_{n \to \infty} EM_n(\theta, \gamma)$, where $M_n(\theta, \gamma) = (\partial^2 / \partial \theta \partial \theta') Q_n(\theta, \gamma)$. This is frequently the case in the absence of serial correlation and heteroskedasticity.

section, this is the case in this model.[8] The other conditions deal mainly with compactness of the parameter spaces Γ and Θ (where γ and θ respectively lie), continuity of $Q(\theta, \gamma)$ and $V(\theta, \gamma)$, and stochastic equicontinuity in (θ, γ) of $Q_n(\theta, \gamma) - Q(\theta, \gamma)$ and $V_n(\theta, \gamma) - V(\theta, \gamma)$ over the corresponding spaces. We have already mentioned that, because of a singular information matrix at the null estimates, the stochastic equicontinuity of $V_n(\theta, \gamma) - V(\theta, \gamma)$ does not hold. However, in similar problems where the score is identically zero at the null estimates, Lee and Chescher (1986) show that the likelihood ratio statistic still has a chi-square distribution.[9] This means in my case that $LR_n(\gamma)$ would have a chi-square distribution. I will assume it is the case and proceed as if the distributional theory developed in Hansen (1991, 1996) or in Andrews and Ploberger (1994) held. To verify if my assumption is reasonable, I will simulate the derived asymptotic distributions of the likelihood ratio statistic for the various models considered and compare them to the empirical distributions obtained by Monte Carlo methods. To obtain the asymptotic distribution of the LR test, the first step consists in deriving the covariance function of the chi-square process $C(\gamma)$.

3.2. *Derivation of the Covariance Function for a General Two-State MS Model.* The two-state MS model (1) can be rewritten as follows:

$$(8) \qquad y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \cdots - \phi_r y_{t-r} -$$

$$\alpha_0(1 - \phi_1 - \phi_2 - \cdots - \phi_r) - \alpha_1(S_t - \phi_1 S_{t-1} - \phi_2 S_{t-2} - \cdots - \phi_r S_{t-r}) = (\omega_0 + \omega_1 S_t)$$

The likelihood of observation t, conditional upon Ψ_t, the information at time t, is therefore given by:

$$(9) \qquad p(y_t|\Psi_t, \theta, \gamma) = \frac{1}{(2\pi)^{1/2}|\omega_0 + \omega_1 s_t(\gamma)|} \exp\left\{ - \frac{\epsilon_t^2}{2(\omega_0 + \omega_1 s_t(\gamma))^2} \right\}$$

Using the following equalities:

$$(10) \qquad \frac{\partial Q_n}{\partial \theta_i} = S_n(\theta_0, \gamma)_{\theta_i} = \frac{1}{n} \frac{\partial \log p(y_n, \ldots, y_1; \theta_0, \gamma)}{\partial \theta_i}$$

I derive, in the following lemma, the elements of the score vector.

[8] Although the two points 0 and 1 represent part of the null hypothesis, in practice if the econometrician finds these values as estimates for p while estimating the Markov switching alternative after having tried many starting values for the parameters, he will conclude that there is not much evidence for a nonlinearity of this type in the series and accept the null of a linear model or try another nonlinear model. The more interesting issue arises when the estimated value for p is different from 0 or 1, since one has to establish whether or not the parameters governed by the Markov process are significantly different from zero.

[9] But they show that this is not the case for the LM or Wald statistic. Therefore it is conjectured that the asymptotic distribution derived for the LR statistic in MS models would not be adequate for the Sup LM or Sup Wald statistic.

LEMMA 1. The elements of the score vector $S_n^c(\theta_0, \gamma)$, evaluated at the true value θ_0 of the parameters of interest and at a particular given value $\gamma \in \Gamma$ of the nuisance parameters, are given by:

$$(11)\quad S_n^c(\theta_0,\gamma)_{\alpha_0} = \frac{1}{n}\sum_{t=1}^{n}\left(1-\sum_{i=1}^{r}\phi_i\right)\frac{\epsilon_t}{\omega_0^2}$$

$$S_n^c(\theta_0,\gamma)_{\alpha_1} = \frac{1}{n}\sum_{t=1}^{n}\sum_{s_t(\gamma)=0}^{1}\cdots\sum_{s_{t-r}(\gamma)=0}^{1}\left(s_t(\gamma)-\sum_{i=1}^{r}\phi_i s_{t-i}(\gamma)\right)\frac{\epsilon_t}{\omega_0^2}\cdot p_t$$

$$S_n^c(\theta_0,\gamma)_{\phi_i} = \frac{1}{n}\sum_{t=1}^{n} -(\alpha_0 - y_{t-i})\frac{\epsilon_t}{\omega_0^2} \qquad i=1,\ldots,r$$

$$S_n^c(\theta_0,\gamma)_{\omega_0^2} = \frac{1}{n}\sum_{t=1}^{n}\frac{1}{2\omega_0^2}\left[\frac{\epsilon_t^2}{\omega_0^2}-1\right]$$

$$S_n^c(\theta_0,\gamma)_{\omega_1^2} = \frac{1}{n}\sum_{t=1}^{n}\sum_{s_t(\gamma)=0}^{1}\cdots\sum_{s_{t-r}(\gamma)=0}^{1}\frac{s_t(\gamma)^2}{2\omega_0^2}\left[\frac{\epsilon_t^2}{\omega_0^2}-1\right]\cdot p_t,$$

where $p_t = p(S_t(\gamma)=s_t(\gamma),\cdots,S_{t-r}(\gamma)=s_{t-r}(\gamma)|y_n,\cdots,y_{-r+1};\theta_0,\gamma)$. The sums over $s_t(\gamma)$ do not appear in the score elements with respect to α_0, $\phi_i (i=1,\cdots,r)$, and ω_0 since the probabilities sum to 1. Proof of the lemma is provided in Section 1 of the Appendix. The next step involves deriving the covariance kernel $K(\gamma_1,\gamma_2)$ of the mean-zero Gaussian process $S_n^c(\theta_0,\gamma)$, which is done in Lemma 2. I will express $K(\gamma_1,\gamma_2)$ in terms of functions of $\gamma=(p,q)$, namely the unconditional probability of being in state 1 for the Markov variable S_t $(\pi(\gamma))$ and another function A, which depends also on the autoregressive parameters ϕ_i $(i=1,\ldots,r)$.

LEMMA 2. The covariance matrix $K(\gamma_1,\gamma_2)$ of the score vectors, as defined in Section 2, is equal to:

$$(12)\quad K(\gamma_1,\gamma_2) =$$

$$\frac{1}{\omega_0^2}\begin{bmatrix}
\left(1-\sum_{i=1}^{r}\phi_i\right)^2 & \pi_2(\gamma_2)\left(1-\sum_{i=1}^{r}\phi_i\right)^2 & 0 & \cdots & 0 & 0 \\
\pi_1(\gamma_1)\left(1-\sum_{i=1}^{r}\phi_i\right)^2 & \min[\pi_1(\gamma_1),\pi_2(\gamma_2)]A(\gamma_2) & 0 & \cdots & 0 & 0 \\
0\cdots & 0 & \omega_0^2 R & \cdots & 0 & 0 \\
0\cdots & 0 & 0 & \cdots & \frac{1}{2\omega_0^2} & \frac{\pi_2(\gamma_2)}{2\omega_0^2} \\
0\cdots & 0 & 0 & \cdots & \frac{\pi_1(\gamma_1)}{2\omega_0^2} & \frac{\min(\pi_1(\gamma_1),\pi_2(\gamma_2))}{2\omega_0^2}
\end{bmatrix}$$

where $\pi(\gamma) = (1-q)/(2-p-q)$ and $A(\gamma_2)$ is given by:

$$(13)\quad A(\gamma) = 1 + \phi_1^2 + \phi_2^2 + \cdots + \phi_r^2 - 2T_2^1(\gamma)\left[\phi_1 - \sum_{i=2}^{r} \phi_i\phi_{i-1}\right]$$

$$- 2T_2^2(\gamma)\left[\phi_2 - \sum_{i=3}^{r} \phi_i\phi_{i-2}\right] - \cdots - 2T_2^r(\gamma)\phi_r$$

with $T_2^r(\gamma)$ denoting the second row, second column element of the transition probability matrix $T_2(\gamma)$ raised to the power r:

$$T_2^r(\gamma) = \left\{\begin{bmatrix} q & 1-q \\ 1-p & p \end{bmatrix}^r\right\}_{2,2}$$

Finally, R denotes the autocovariance matrix of the vector:

$$\bar{y}_{t-1} = (y_{t-1} - \alpha_0, \cdots, y_{t-r} - \alpha_0)'.$$

Proof of Lemma 2 is provided in Section 2 of the Appendix. The covariance kernel $\bar{K}(\gamma_1, \gamma_2)$ of $C(\gamma)$ is easily derived by noting that $V(\gamma_i) = K(\gamma_i, \gamma_i)$, $i = 1, 2$.

Based on this general covariance function, I derive next the asymptotic null distribution of LR_n for a two-mean model with an uncorrelated and homoskedastic noise, used by Cecchetti et al., (1990) to model the annual growth rate of consumption in an asset-pricing application; a two-mean model with an AR(r) homoskedastic noise, used by Hamilton (1989) to model the quarterly growth rate of output; and finally a two-mean, two-variance model with an uncorrelated noise, used by Turner et al., (1989) to model a series of stock returns. This will allow me to test formally the linear null against the MS alternative in each case. I also compute the power of the Sup LR test for the three models.

4. ASYMPTOTIC NULL DISTRIBUTIONS OF THE SUP LR STATISTIC

Based on Theorem 1 and Lemmas 1 and 2, I derive for each of the three above-mentioned MS models the matrix $\bar{K}(\gamma_1, \gamma_2)$, the covariance functional of the chi-square process $C(\gamma)$, defined in (6).

4.1. *Two-mean Model with an Uncorrelated and Homoskedastic Noise Component.* The two-state MSM with an uncorrelated and homoskedastic noise component is given by:

$$(14)\quad y_t = \alpha_0 + \alpha_1 S_t + \omega_0 \epsilon_t$$

$$P(S_t = 1 | S_{t-1} = 1) = p$$

$$P(S_t = 0 | S_{t-1} = 0) = q$$

THEOREM 2. *The limiting distribution of* LR_n *is* $\operatorname{Sup} C = \operatorname{Sup}_{\gamma=(p,q)\in\Gamma} C(\gamma)$, *where* $C(\gamma)$ *is a chi-square process with covariance*:

$$(15)\quad \bar{K}(\gamma_1,\gamma_2) = \omega_0^2 \frac{\min(\pi_1(\gamma_1),\pi_2(\gamma_2)) - \pi_1(\gamma_1)\pi_2(\gamma_2)}{\pi_1(\gamma_1)\pi_2(\gamma_2)(1-\pi_1(\gamma_1))(1-\pi_2(\gamma_2))}$$

$$\text{where } \pi_i(\gamma_i) = \frac{1-q_i}{2-p_i-q_i}.$$

The proof is a straightforward application of formula (6). In Section 3 of the Appendix, I show that the asymptotic null distribution of the LR test in this particular model is identical to the distribution of the LR test in a one-dimensional threshold model (derived by Chan 1990) and in a structural change model with an unknown change point (Andrews 1993). As shown in Corollary 1 in Andrews (1993), this means that the LR test will diverge to infinity as n goes to infinity, unless Γ is bounded away from 0 and 1. To obtain the supremum of $C(\gamma)$ over Γ, one must therefore fix the bounds of the parameter space Γ, but it remains to be determined how far away from 0 and 1 one must set γ_1 and γ_T to obtain a good approximation to the finite sample distribution. In the context of structural change models with an unknown change point, Andrews (1993) chooses 0.15 and 0.85.

In order to simulate the distribution of $\operatorname{Sup} C$, I follow the general method for simulating chi-square processes described in Section 4 of the Appendix. In this case, the $G(\gamma)$ and ε vectors reduce to scalars. For this particular MS model, γ_1 and γ_2 enter the $g(\gamma)$ expressions only through the function $\pi(\gamma) = (1-q)/(2-p-q)$. Therefore, it is sufficient to set a grid for π, that is, $\{\pi_1, \pi_2, \ldots, \pi_T\}$. In this case, the chi-square process $C(\gamma)$ is equal to $C(\gamma) = g(\gamma)^2\pi(1-\pi)$. To see which bounds will produce the best asymptotic approximation to the finite sample distribution, I simulated the empirical distribution by generating a 1,000 series of 100 observations under the null H_0: $\alpha_1 = 0$ and estimated the likelihood under both the null and the MS alternative in (15). Since the true model is a model with no change in regime, one might expect when estimating the MS model that some sets of optimizing values for the parameters correspond to local maxima of the likelihood function. This problem has been reported in Hamilton (1989) and Garcia and Perron (1996). This means that 1,000 replications will typically produce only a fraction of positive log likelihood ratios, and among these a lot of values close to zero. A way to minimize this problem is to optimize the likelihood function under the alternative by using many sets of starting values for each generated series and take the maximum of the likelihood function over the values so obtained. By proceeding in this fashion, I hope to eliminate or at least reduce the number of local maxima. I applied this method using twelve sets of starting values. The success rate in obtaining a positive likelihood ratio was 100%. The results are shown in the first column of Table 1A. The 99% and 95% critical values are 14.02 and 10.89, respectively. The next two columns of Table 1A show the critical values of the asymptotic distribution obtained with 10,000 replications of $\operatorname{Sup} C$ for $\pi \in [0.01, 0.99]$ and $\pi \in [0.15, 0.85]$, with increments of 0.001. The goal is to assess which trimming produces the best

TABLE 1A
EMPIRICAL AND ASYMPTOTIC CRITICAL VALUES OF THE LIKELIHOOD RATIO
IN A TWO-STATE MARKOV SWITCHING MODEL WITH UNCORRELATED AND HOMOSKEDASTIC ERRORS

% of Dist.	Empirical Distribution Critical Value	Asympt. Dist. Critical Value [0.01–0.99]	Asympt. Dist. Critical Value [0.15–0.85]
99%	14.02	13.64	12.45
95%	10.89	10.18	8.60
90%	8.92	8.68	7.08
80%	6.71	7.04	5.47
70%	5.53	6.05	4.50
50%	4.03	4.72	3.25
10%	1.39	2.55	1.48
5%	0.99	2.15	1.18
1%	0.45	1.56	0.81

TABLE 1B
SIZE AND POWER OF THE SUP LR TEST IN A TWO-STATE MARKOV
SWITCHING MODEL WITH UNCORRELATED AND HOMOSKEDASTIC ERRORS*

Nominal Size	1%	5%
Sup LR	1.4%	6.5%
Power-Sup LR	38.2%	49.5%

*The power of Sup LR test has been computed with respect to the model estimated by Cecchetti et al., (1990).

approximation to the empirical distribution, which is itself obtained without any trimming.[10] It appears that the asymptotic values up to the 65% percentile of the distribution for the [0.01, 0.99] bounds are very close to the corresponding empirical values. The left tail of the distribution is not approximated as well and this could be due to the presence of local maxima. Note also that these critical values for the likelihood ratio statistic are considerably higher than the values of a $X^2(1)$, the distribution of the LR test in the classical theory.

In Table 1B, I report the actual size of the Sup LR test for nominal sizes of 1% and 5% under the $\pi \in [0.01, 0.99]$ range, as well as the power of the Sup LR test for the MS model of consumption growth estimated by Cecchetti et al., (1990). I generated a 1,000 series based on the following estimates ($\alpha_0 = 0.228$, $\alpha_1 = -0.0926$, $\omega_0 = 0.0320$, $p = 0.5279$, $q = 0.9761$), and estimated for each series both the linear and the MS models. A 5% Sup LR test will have in this case a power of about 50%.

4.2. *Two-mean Model with an AR(r) Homoskedastic Noise Component.* With an autoregressive structure of order r and no heteroskedasticity for the noise term,

[10] Trimming the empirical distribution by excluding the points in the distribution where π is outside the [0.01, 0.99] or [0.15, 0.85] intervals does not produce very different distributions from the untrimmed distribution. In particular, the 99% and the 95% values are almost identical. I therefore report only the untrimmed distributions throughout the paper.

we obtain the following specification for the MS model:

$$y_t = \alpha_0 + \alpha_1 S_t + z_t \tag{16}$$

$$z_t = \phi_1 z_{t-1} + \cdots + \phi_r z_{t-r} + \epsilon_t.$$

THEOREM 3. *The limiting distribution of* LR_n *is* $Sup\, C = Sup_{\gamma \in \Gamma} C(\gamma)$, *where* $C(\gamma)$ *is a chi-square process with covariance function*:

$$\bar{K}(\gamma_1, \gamma_2) = \omega_0^2 \frac{\min(\pi_1, \pi_2) A_2 - \pi_1 \pi_2 \left(1 - \sum_{i=1}^{r} \phi_i\right)^2}{\pi_1 \pi_2 \left[A_2 - \left(1 - \sum_{i=1}^{r} \phi_i\right)^2 \pi_2\right]\left[A_1 - \left(1 - \sum_{i=1}^{r} \phi_i\right)^2 \pi_1\right]}, \tag{17}$$

where, to simplify notation, we do not repeat the dependence of π_i *on* γ_i, *and where* $A(\gamma)$ *is as defined in Lemma 2. The expression follows from* (6) *with*:

$$V(\gamma)^{-1} = \omega_0^2 \begin{bmatrix} \dfrac{\pi A(\gamma)}{\Delta(\gamma)} & \dfrac{-\left(1 - \sum_{i=1}^{r} \phi_i\right)^2 \pi}{\Delta(\gamma)} & 0 & 0 \\ \dfrac{-\left(1 - \sum_{i=1}^{r} \phi_i\right)^2 \pi}{\Delta(\gamma)} & \dfrac{\left(1 - \sum_{i=1}^{r} \phi_i\right)^2}{\Delta(\gamma)} & 0 & 0 \\ 0 & 0 & \dfrac{R^{-1}}{\omega_0^2} & 0 \\ 0 & 0 & 0 & 2\omega_0^2 \end{bmatrix}$$

where: $\Delta(\gamma) = \pi(1 - \sum_{i=1}^{r} \phi_i)^2 [A(\gamma) - (1 - \sum_{i=1}^{r} \phi_i)^2 \pi]$ *and* $K(\gamma_1, \gamma_2)$ *is deduced from the expression given in* (12) *by canceling the last row and the last column corresponding to* ω_1 *since it is equal to zero in this particular homoskedastic model.*

In this MS model with an autoregressive noise function, I note that the covariance function does not depend on γ only through the unconditional probability π. Both π and p (in $A(\gamma)$) appear on the right-hand side, and the covariance function depends also on the autoregressive parameters ϕ. Note that when the ϕ's are all zero, the distribution simplifies to the Andrews' (1993) distribution.[11] To assess the performance of the sup LR test in the autoregressive case, I first study the AR(1) case in detail. I determine the bounds of the parameter space over π and p that

[11] This suggests that if the ϕ's are nonzero but small, the Andrews' distribution might be a good approximation.

TABLE 2A
COMPARISON OF EMPIRICAL AND ASYMPTOTIC CRITICAL VALUES OF THE LIKELIHOOD RATIO IN A TWO-STATE MARKOV SWITCHING MODEL WITH FIRST-ORDER AUTOREGRESSIVE AND HOMOSKEDASTIC NOISE FUNCTION FOR DIFFERENT VALUES OF THE AUTOREGRESSIVE PARAMETER

% of Dist.	Asymptotic Distribution Critical Values $\phi_1 = 0.337$	Empirical Distribution Critical Values $\phi_1 = 0.337$	Asymptotic Distribution Critical Values $\phi_1 = -0.5$	Empirical Distribution Critical Values $\phi_1 = -0.5$
99%	12.00	11.82	11.88	13.08
95%	8.68	8.72	8.62	8.82
90%	7.05	7.21	7.06	7.27
80%	5.48	5.52	5.44	5.74
70%	4.49	4.33	4.43	4.73
50%	3.26	2.83	3.25	3.28
10%	1.45	0.66	1.45	0.86
5%	1.18	0.31	1.17	0.57
1%	0.79	0.03	0.81	0.15

TABLE 2B
SIZE OF THE SUP LR TEST IN A TWO-STATE MARKOV SWITCHING MODEL WITH FIRST-ORDER AUTOREGRESSIVE AND HOMOSKEDASTIC NOISE FUNCTION

Nominal Size	1%	5%
	Autoregressive Coefficient = 0.337	
Sup LR	0.97%	5.1%
	Autoregressive Coefficient = −0.5	
Sup LR	1.6%	5.4%

give the asymptotic distribution that provides the best approximation to the empirical distribution. I also simulate the asymptotic distribution for a range of values of the autoregressive parameter to establish whether the distribution changes or remains stable.

4.2.1. *The AR(1) case.* To determine the bounds that give the best approximation to the empirical distribution and to see if the empirical distribution varies as a function of ϕ_1, we simulated the empirical distribution of the likelihood ratio for two AR(1) models with ϕ_1, the autoregressive parameter, equal to 0.337[12] and −0.5. The true model is the AR(1) model and the alternative is the two-state Markov model in (16) with $r = 1$. We generated the distribution using 1,000 replications of the true model and estimating the alternative MS model, starting with six different sets of values of the six parameters for each series to avoid as much as possible the problem of local maxima explained in Section 4.1. The critical values obtained for the empirical distribution of the likelihood ratio are shown in Table 2. They appear to be smaller than the values obtained for the uncorrelated and homoskedastic case. Also it has to be noted that the empirical critical values do not seem to depend on the value of the autoregressive parameter.

[12] This value corresponds to the value of the autoregressive coefficient in an AR(1) model for log GNP, estimated from 1952 II to 1984:IV, the period chosen by Hamilton (1989).

TABLE 3
ASYMPTOTIC CRITICAL VALUES OF THE LIKELIHOOD RATIO IN A TWO-STATE MARKOV SWITCHING MODEL WITH FIRST-ORDER AUTOREGRESSIVE AND HOMOSKEDASTIC NOISE FUNCTION FOR VARIOUS VALUES OF THE AUTOREGRESSIVE PARAMETER

% of Dist.	Critical Values $\phi_1 = 0.3$	Critical Values $\phi_1 = 0.5$	Critical Values $\phi_1 = 0.8$	Critical Values $\phi_1 = 0.95$
99%	11.92	12.07	11.95	12.08
95%	8.74	8.57	8.48	8.48
90%	7.20	7.06	7.00	7.06
80%	5.49	5.44	5.44	5.51
70%	4.55	4.52	4.49	4.52
50%	3.28	3.27	3.26	3.22
10%	1.46	1.46	1.46	1.45
5%	1.21	1.17	1.17	1.18
1%	0.79	0.78	0.81	0.83
% Dist.	Critical Values $\phi_1 = -0.3$	Critical Values $\phi_1 = -0.5$	Critical Values $\phi_1 = -0.8$	Critical Values $\phi_1 = -0.95$
99%	12.26	11.88	12.45	11.79
95%	8.66	8.62	8.68	8.50
90%	7.08	7.06	7.08	7.00
80%	5.43	5.44	5.51	5.50
70%	4.52	4.43	4.55	4.54
50%	3.28	3.25	3.28	3.28
10%	1.47	1.45	1.49	1.47
5%	1.19	1.17	1.19	1.18
1%	0.82	0.81	0.79	0.81

To generate the asymptotic distribution of Sup C, we used the interval $[0.15, 0.85]$ for both parameters p and π, with increments of 0.002 for p and 0.001 for π. In this model, the chi-square process $C(\gamma)$ is equal to: $C(\gamma) = g(\gamma)^2\pi(1 + \phi_1^2 - 2\phi_1 p - (1 - \phi_1)^2\pi)$. The distributions are based on 10,000 replications. The asymptotic critical values are fairly close to the empirical ones, except again for the left tail. Moreover, the distributions do not seem to depend on the value of the autoregressive parameter. This is confirmed in Table 3, where we present the asymptotic critical values corresponding to the various percentage levels for eight different values of ϕ_1: 0.3, 0.5, 0.8, 0.95 and -0.3, -0.5, -0.8, and -0.95. It has to be noted, however, that the distribution is not invariant to the value of the autoregres sive parameter. As Table 2B shows, the actual sizes of the Sup LR test are pretty close to the nominal sizes of 1% and 5%, both for $\phi = 0.337$ and $\phi = -0.5$. To assess the power of the Sup LR test for a model with an autoregressive structure, I use the AR(4) model estimated by Hamilton (1989).

4.2.2. *The AR(4) case: the Hamilton (1989) GNP Model.* To capture the asymmetry in the growth rate of GNP between booms and recessions, Hamilton (1989) chose a MS model identical to the model in (16) with a fourth-order autoregressive noise function. The maximum likelihood estimation results are presented in Table 4 along with the maximum likelihood estimates of the AR(4) model. Note that the likelihood ratio statistic $(2(L_1 - L_0))$ is equal to 4.812. If judged with respect to a chi-square distribution with one degree of freedom, the null of an AR(4) in first differences will be rejected at about the 3% level.

TABLE 4
MAXIMUM LIKELIHOOD ESTIMATES: US REAL GNP 1952:2–1984:4

Parameters	AR(4) Model	Markov Trend Model*
α_0	0.720(0.112)	−0.359(0.265) [0.465]
α_1	—	1.522(0.264) [0.464]
p	—	0.904(0.037) [0.033]
q	—	0.755(0.097) [0.101]
ϕ_1	0.310(0.088)	0.014(0.120) [0.164]
ϕ_2	0.127(0.091)	−0.058(0.137) [0.219]
ϕ_3	−0.121(0.091)	−0.247(0.107) [0.148]
ϕ_4	−0.089(0.087)	−0.213(0.110) [0.136]
σ	0.983(0.061)	0.769(0.067) [0.094]
L	−63.29	−60.88

*The standard errors between parentheses correspond to the values of the numerically computed Hessian. The standard errors between brackets are taken from Hansen (1990a) and correspond to heteroskedastically consistent values.

To assess the estimation results of Hamilton, we need to generate the distribution of Sup C defined at the beginning of the section for $r = 4$. For the autoregressive parameters, I use the estimated values shown in Table 4 for the Markov trend model. The results are shown in the first column of Table 5 for bounds of 0.15 and 0.85 for the π and p parameters. As shown in 4.2.1, these bounds give the best approximation for the empirical distribution of the AR(1) case. Judged by this distribution, one cannot reject at usual levels the null of an AR(4) against the Markov trend model for the first differences of US log GNP, since the p-value is

TABLE 5
DISTRIBUTIONS OF THE LIKELIHOOD RATIO STATISTIC-LINEAR AR(4) AGAINST MARKOV TREND MODEL AR(4)*

	Asymptotic Distribution under the Linear Null	Empirical Distribution under the Markov Trend Null
99%	12.24	33.35
95%	8.59	26.70
90%	7.10	23.99
70%	4.52	17.26
50%	3.27	13.47
10%	1.47	6.19
5%	1.18	4.44
1%	0.80	2.10

*These critical values were computed using for the autoregressive parameters the estimated values with the Markov trend model for US GNP (see Table 1).

around 30%. Therefore, the same conclusion as Hansen (1992) with his simulation-based bound method is reached. However, he obtained a much higher p-value (around 70%), which is consistent with the more conservative nature of his bound test.

Finally, in the right-hand side part of Table 5, I report the empirical distribution of the LR statistic when the data generating process is the MS model of GNP estimated by Hamilton (1989) with the parameter values shown in Table 4. The LR statistic distribution is obtained by estimating both the linear AR(4) model and the Markov trend model with an autoregressive structure of order 4 for a 1,000 series produced by the data generating process. A 5% Sup LR test will have a power close to 80%.

4.3. *A Two-mean Model with an Uncorrelated and Heteroskedastic Noise Component.* The two-state MSM with an uncorrelated and heteroskedastic noise component is given by:

$$y_t = \alpha_0 + \alpha_1 S_t + z_t \tag{18}$$

$$z_t = (\omega_0 + \omega_1 S_t)\epsilon_t$$

$$P(S_t = 1 | S_{t-1} = 1) = p$$

$$P(S_t = 0 | S_{t-1} = 0) = q$$

Because the covariance matrix $K(\gamma_1, \gamma_2)$ of the score vectors given in (12) reduces to a block-diagonal matrix, the limiting distribution of LR_n is $\operatorname{Sup} C = \operatorname{Sup}_{\gamma \in \Gamma} C_1(\gamma) + C_2(\gamma)$, where $C_1(\gamma)$ and $C_2(\gamma)$ are chi-square processes with respective covariances:[13]

$$\bar{K}_1(\gamma_1, \gamma_2) = \omega_0^2 \frac{\min(\pi_1, \pi_2) - \pi_1\pi_2}{\pi_1\pi_2(1-\pi_1)(1-\pi_2)} \tag{19}$$

$$\bar{K}_2(\gamma_1, \gamma_2) = 2\omega_0^4 \frac{\min(\pi_1, \pi_2) - \pi_1\pi_2}{\pi_1\pi_2(1-\pi_1)(1-\pi_2)}$$

where π_i is a function of γ_i, as defined in (15). Therefore, $C(\gamma)$ can be represented as the sum of two chi-square processes with the covariance functions shown above. To simulate the distribution of $\operatorname{Sup} C$, I therefore follow the method described in Section 3.1 to generate two independent Gaussian vectors $g(\gamma)$. Table 6A shows the asymptotic critical values generated with the set of bounds $[0.01 - 0.99]$ for Γ, which gives the best approximation to the empirical distribution, along with the empirical critical values. The empirical distribution was obtained by generating a 1,000 series

[13] This is identical to the Andrews' (1993) distribution with two degrees of freedom. Note that if only a variance switch is estimated, the approximate distribution is the Andrews' distribution with one degree of freedom, just as for the mean switch with no serial correlation.

TABLE 6A
COMPARISON OF EMPIRICAL AND ASYMPTOTIC CRITICAL VALUES OF THE LIKELIHOOD RATIO IN A TWO-STATE MARKOV SWITCHING MODEL WITH UNCORRELATED AND HETEROSKEDASTIC NOISE FUNCTION

% of Dist.	Empirical Distribution Critical Values	Asymptotic Distribution Critical Values [0.01–0.99]
99%	17.38	17.52
95%	14.11	13.68
90%	12.23	11.88
80%	10.02	9.99
70%	8.84	8.86
50%	6.84	7.22
10%	3.25	4.43
5%	2.68	3.87
1%	1.74	3.05

TABLE 6B
SIZE AND POWER OF THE SUP LR TEST IN A TWO-STATE MARKOV SWITCHING MODEL WITH UNCORRELATED AND HETERASKEDASTIC ERRORS*

Nominal Size	1%	5%
Sup LR	0.98%	6.2%
Power-Sup LR	46%	60.9%

*The power of Sup LR test has been computed with respect to the model estimated by Turner et al., (1989).

under the null hypothesis ($\alpha_1 = 0,\ \omega_1 = 0$) and estimating the likelihood ratio between the linear homoskedastic model and the heteroskedastic MS model, again using six sets of starting values for the parameters.[14]

In Table 6B, I report the actual sizes of the Sup LR test for nominal sizes of 1% and 5%. As before, the actual sizes are close to the nominal sizes. I also report the power of the Sup LR test for the MS model of stock returns estimated by Turner et al., (1989). I generated a 1,000 series based on the following parameter estimates ($\alpha_0 = 0.677,\ \alpha_1 = -2.652,\ \omega_0 = 2.693,\ \omega_1 = 2.396,\ p = 0.767,\ q = 0.950$), and estimated for each series both the linear and the MS model. A 5% Sup LR test will have in this case a power of about 60%.

5. CONCLUSION

In this paper, I have shown that the asymptotic distribution theory used for problems where a nuisance parameter is not identified under the null also seems to work for MS models, even though its validity can be questioned because of the additional problem of identically zero scores under the null estimates. Under the

[14] A word of caution about the generation of the empirical distribution is in order. In about 5% of the cases, the optimizing program reaches singularity points, where either ω_0 or $\omega_0 + \omega_1$ are close to zero, giving high values for the likelihood ratio. These values have been excluded from the empirical distribution.

assumed validity of this distribution theory, I have derived the asymptotic distribution of the Sup LR test for a two-state MS model with an autoregressive noise function of any order. I have also provided asymptotic critical values for the likelihood ratio test for three particular models used in the economic literature. According to the Monte Carlo study, these values are close to the critical values of the empirical distribution of the test in each of the three models. The critical values reported for the two-mean and two-mean, two-variance models with an uncorrelated noise function can be used directly to assess the validity of MS models with the same specification for various economic and financial time series. For models with a correlated noise function, I propose a general simulation method that researchers can use to generate the asymptotic distribution of the Sup LR test, given their estimates of the autoregressive parameters. I have shown, however, that this distribution is not sensitive to the values of the autoregressive parameters.

For the AR(4) GNP model estimated by Hamilton (1989), I generated the asymptotic distribution of the Sup LR test and showed that, based on this test, the null of an AR(4) cannot be rejected. In other words, there is no evidence in the period chosen by Hamilton for a MS model in the GNP growth series. I also assessed the power of the test to be around 80% for this particular model. For the other models studied with an uncorrelated noise function, the power of the Sup LR test was in the 50–60% range.

The MS models studied in this paper are the simplest ones, in the sense that the number of states is limited to two, the autoregressive structure does not depend on the state, the transition probabilities are not time-varying, and the model is univariate. Many extensions exist in the literature where there are more than two states (Garcia and Perron 1996, and Sichel 1994), where the autoregressive structure is state-dependent (Hansen 1992), where the transition probabilities are functions of exogenous information (Diebold et al., 1993, and Filardo 1994), of seasonals (Ghysels 1994) or of duration (Durland and McCurdy 1994), and where a bivariate system is estimated (Cecchetti et al., 1993, and Bonomo and Garcia 1996). The methodology presented in this paper can be extended without too much difficulty to models with a higher number of states, a state-dependent autoregressive structure, or a multivariate system. The extension to models with time-varying probabilities is less obvious since the Markov process of the state variable varies from one observation to the other. These developments are left for future work.

APPENDIX

1. *Derivation of the Scores.* Start with the following equality:

$$\frac{\partial \log p(y_n,\ldots,y_1,s_n(\gamma),\ldots,s_1(\gamma);\gamma,\theta)}{\partial\theta_i}$$

$$= \frac{\partial p(y_n,\ldots,y_1,s_n(\gamma),\ldots,s_1(\gamma);\gamma,\theta)}{\partial\theta_i} \frac{1}{p(y_n,\ldots,y_1,s_n(\gamma),\ldots,s_1(\gamma);\gamma,\theta)}$$

Therefore:

$$\frac{\partial p(y_n,\ldots,y_1,s_n(\gamma),\ldots,s_1(\gamma);\gamma,\theta)}{\partial\theta_i}$$

$$= \frac{\partial \log p(y_n,\ldots,y_1,s_n(\gamma),\ldots,s_1(\gamma);\gamma,\theta)}{\partial\theta_i}$$

$$\times p(y_n,\ldots,y_1,s_n(\gamma),\ldots,s_1(\gamma);\gamma,\theta)$$

$$= \sum_{t=1}^{n} \frac{\partial \log p(y_t|\Psi_t,\gamma,\theta)}{\partial\theta_i} p(y_n,\ldots,y_1,\gamma,\theta)\, p(S_n(\gamma)=s_n(\gamma),\ldots,S_1(\gamma)$$

$$= s_1(\gamma)|y_n,\ldots,y_1,\gamma,\theta)$$

where: $\Psi_t = \{y_{t-1},\ldots,y_{t-r},s_{t-1}(\gamma),\ldots,s_{t-r}(\gamma)\}$

Then, summing over $s_t(\gamma),\ldots,s_{t-r}(\gamma)=0,1$ for $t=1,\ldots,n$ and dividing by $p(y_n,\ldots,y_1,\gamma,\theta)$, I obtain:

$$\frac{\partial p(y_n,\ldots,y_1;\gamma,\theta)}{\partial\theta_i}\frac{1}{p(y_n,\ldots,y_1;\gamma,\theta)}$$

$$= \sum_{s_n(\gamma)=0}^{1}\cdots\sum_{s_{n-r}(\gamma)=0}^{1} \frac{\partial \log p(y_n|\Psi_n,\gamma,\theta)}{\partial\theta_i} p(S_n(\gamma)$$

$$= s_n(\gamma),\ldots,S_n(\gamma)=s_{n-r}(\gamma)|y_n,\ldots,y_1,\gamma,\theta)$$

$$+ \sum_{s_1(\gamma)=0}^{1}\cdots\sum_{s_{-r+1}(\gamma)=0}^{1} \frac{\partial \log p(y_1|\Psi_1,\gamma,\theta)}{\partial\theta_i} p(S_1(\gamma)$$

$$= s_1(\gamma),\ldots,S_{-r+1}(\gamma)=s_{-r+1}(\gamma)|y_n,\ldots,y_1,\gamma,\theta)$$

since $p(y_t|\Psi_t,\gamma,\theta)$ depends only on $s_t(\gamma),\ldots,s_{t-r}(\gamma)$. The conditional probabilities $p(S_t(\gamma)=s_t(\gamma),\ldots,S_{t-r}(\gamma)=s_{t-r}(\gamma)|y_n,\ldots,y_1,\gamma,\theta)$ are the so-called smoothed probabilities (see Hamilton 1989).

2. PROOF OF LEMMA 2. I develop below the computation of each element of the covariance kernel $K(\gamma_1,\gamma_2)$ of the score process $S_n^c(\theta_0,\gamma)$. In what follows, the pair (γ_1,γ_2) characterize any two elements of the set of values $\{\gamma_1,\ldots,\gamma_T\}$ included in Γ. In the derivation, the function $\pi=(1-q)/(2-p-q)$ of $\gamma=(p,q)$ (the unconditional probability of the Markov variable) will appear. By convention, I

assume that $\pi_1 < \pi_2$. I start with the (α_0, α_0) element:

$$nE\left[S_n^c(\theta_0, \gamma_1)_{\alpha_0} S_n^c(\theta_0, \gamma_2)_{\alpha_0}\right] = nE \sum_{t=1}^{n} \sum_{s=1}^{n} \frac{1}{n^2} \left(1 - \sum_{i=1}^{r} \phi_i\right)^2 \frac{\epsilon_t \epsilon_s}{\omega_0^4}$$

By the serial independence assumption about the ϵ, we are left with:

$$nE\left[S_n^c(\theta_0, \gamma_1)_{\alpha_0} S_n^c(\theta_0, \gamma_2)_{\alpha_0}\right] = n \sum_{t=1}^{n} \left(1 - \sum_{i=1}^{r} \phi_i\right)^2 \frac{E(\epsilon_t^2)}{n^2 \omega_0^4} = \frac{\left(1 - \sum_{i=1}^{r} \phi_i\right)^2}{\omega_0^2}$$

I now derive the formula for the expectation of the cross-product of the scores with respect to α_0 and α_1:

$$nE\left[S_n^c(\theta_0, \gamma_1)_{\alpha_0} S_n^c(\theta_0, \gamma_2)_{\alpha_1}\right]$$

$$= nE \sum_{t=1}^{n} \sum_{s=1}^{n} \sum_{s_s=0}^{1} \cdots \sum_{s_{s-r}=0}^{1} \frac{\epsilon_t \epsilon_s}{n^2 \omega_0^4} \left(1 - \sum_{i=1}^{r} \phi_i\right) \left(s_s(\gamma_2) - \sum_{i=1}^{r} \phi_i s_{s-i}(\gamma_2)\right) \cdot p_s$$

This can be rewritten as:

$$nE\left[S_n^c(\theta_0, \gamma_1)_{\alpha_0} S_n^c(\theta_0, \gamma_2)_{\alpha_1}\right]$$

$$= nE\left\{ \sum_{t=1}^{n} \sum_{s=1}^{n} \frac{\epsilon_t \epsilon_s}{n^2 \omega_0^4} \left(1 - \sum_{i=1}^{r} \phi_i\right) E\left[s_s(\gamma_2) - \sum_{i=1}^{r} \phi_i s_{s-i}(\gamma_2) | \Psi_n, \theta_0, (\gamma_2)\right]\right\}$$

where I have used the independence assumption between ϵ and $s(\gamma)$. Next, I apply the law of iterated expectations $E\{E[s_s(\gamma_2) - \Sigma_{i=1}^r \phi_i s_{s-i}(\gamma_2) | \Psi_n, \theta_0, \gamma_2]\} = E[s_s(\gamma_2) - \Sigma_{i=1}^r \phi_i s_{s-i}(\gamma_2)]$, and note that $E(s_s(\gamma_2)) = E(s_{s-1}(\gamma_2)) = \cdots = E(s_{j-r}(\gamma_2)) = \pi_2 = (1 - q_2)/(2 - q_2 - p_2)$. Therefore, by the serial independence assumption about ϵ, the final expression is given by:

$$nE\left[S_n^c(\theta_0, \gamma_1)_{\alpha_0} S_n^c(\theta_0, \gamma_2)_{\alpha_1}\right] = \frac{\left(1 - \sum_{i=1}^{r} \phi_i\right)^2}{\omega_0^2} \pi_2$$

Similarly for the (α_1, α_0) element:

$$nE\left[S_n^c(\theta_0, \gamma_1)_{\alpha_1} S_n^c(\theta_0, \gamma_2)_{\alpha_0}\right] = \frac{\left(1 - \sum_{i=1}^{r} \phi_i\right)^2}{\omega_0^2} \pi_1$$

Proceeding in the same way, I obtain the following expression for the (α_1, α_1) element:

$$
\begin{aligned}
&nE\left[S_n^c(\theta_0,\gamma_1)_{\alpha_1} S_n^c(\theta_0,\gamma_2)_{\alpha_1}\right] \\
&\quad = nE\left\{\sum_{t=1}^{n}\sum_{s=1}^{n}\frac{\epsilon_t\epsilon_s}{n^2\omega_0^4}E\left[\left(s_t(\gamma_1)-\sum_{i=1}^{r}\phi_i s_{t-i}(\gamma_1)\right)\right.\right. \\
&\qquad\qquad \left.\left.\times\left(s_s(\gamma_2)-\sum_{i=1}^{r}\phi_i s_{s-i}(\gamma_2)\right)\middle|\Psi_n,\theta_0,\gamma_1,\gamma_2\right]\right\} \\
&\quad = nE\left\{\sum_{t=1}^{n}\sum_{s=1}^{n}\frac{\epsilon_t\epsilon_s}{n^2\omega_0^4}\left[E(s_t(\gamma_1)s_s(\gamma_2))+\phi_1^2E(s_{t-1}(\gamma_1)s_{s-1}(\gamma_2))\right.\right. \\
&\qquad\qquad +\cdots+\phi_r^2E(s_{t-r}(\gamma_1)s_{s-r}(\gamma_2)) \\
&\qquad\qquad -\phi_1E(s_t(\gamma_1)s_{s-1}(\gamma_2))-\phi_1E(s_{t-1}(\gamma_1)s_s(\gamma_2)) \\
&\qquad +\sum_{i=2}^{r}\phi_i\phi_{i-1}E(s_t(\gamma_1)s_{s-1}(\gamma_2))+\sum_{i=2}^{r}\phi_i\phi_{i-1}E(s_{t-1}(\gamma_1)s_s(\gamma_2) \\
&\qquad\qquad -\phi_2E(s_t(\gamma_1)s_{s-2}(\gamma_2))-\phi_2E(s_{t-2}(\gamma_1)s_s(\gamma_2)) \\
&\qquad +\sum_{i=3}^{r}\phi_i\phi_{i-2}E(s_t(\gamma_1)s_{s-2}(\gamma_2))+\sum_{i=3}^{r}\phi_i\phi_{i-2}E(s_{t-2}(\gamma_1)s_s(\gamma_2)) \\
&\qquad\qquad -\cdots- \\
&\qquad\qquad \left.\left.-\phi_rE(s_t(\gamma_1)s_{s-r}(\gamma_2)-\phi_rE(s_{t-r}(\gamma_1)s_s(\gamma_2))\right]\right\}
\end{aligned}
$$

Now, I state the following results for dependent Markov variables:

$$
\begin{aligned}
E\left[s_t(\gamma_1)s_s(\gamma_2)\right] &= \min(\pi_1,\pi_2) \\
E\left[s_t(\gamma_1)s_{s-1}(\gamma_2)\right] &= \min(\pi_1,\pi_2)T_2^1(\gamma_2) \\
E\left[s_t(\gamma_1)s_{s-2}(\gamma_2)\right] &= \min(\pi_1,\pi_2)T_2^2(\gamma_2) \\
\vdots \qquad &= \qquad \vdots \\
E\left[s_t(\gamma_1)s_{s-r}(\gamma_2)\right] &= \min(\pi_1,\pi_2)T_2^r(\gamma_2)
\end{aligned}
$$

where $T_2^i(\gamma_2)$ denotes the second row, second column element of the transition probability matrix T_2 corresponding to γ_2 raised to the power i. Using the

independence assumption of the ϵ, I finally obtain:

$$nE\left[S_n(\theta_0,\gamma_1)_{\alpha_1}S_n(\theta_0,\gamma_2)_{\alpha_1}\right]=\frac{1}{\omega_0^2}\min(\pi_1,\pi_2)A(\gamma_2)$$

where $A(\gamma_2)$ is as defined in Section 2.2 in the text.
Since the expectations of ϵ_t and ϵ_t^3 are 0, the expectations of the cross-products of the scores with respect to α_0 and ω_0^2 on one hand, and α_1 and ω_0^2 on the other, are both zero.

The limits as n tends to infinity of the expectations of the scores with respect to α_0 and ϕ_i, and α_1 and ϕ_i are both zero, as I show below for ϕ_1:

$$nE\left[S_n^c(\theta_0,\gamma_1)_{\alpha_0}S_n^c(\theta_0,\gamma_2)_{\phi_1}\right]$$

$$=\frac{1}{\omega_0^2}\left[-\alpha_0\left(1-\sum_{i=1}^{r}\phi_i\right)+\left(1-\sum_{i=1}^{r}\phi_i\right)\frac{\sum_{t=1}^{n}y_{t-1}}{n}\right]$$

$$=\frac{1}{\omega_0^2}\left[-\alpha_0\left(1-\sum_{i=1}^{r}\phi_i\right)+\left(1-\sum_{i=1}^{r}\phi_i\right)\frac{y_0-y_n}{n}+\left(1-\sum_{i=1}^{r}\phi_i\right)\frac{\sum_{t=1}^{n}y_t}{n}\right]$$

The limit as n tends to infinity of the average of the y is $E(y)$, that is, α_0 under the null hypothesis. Therefore, the whole expression tends to zero. The development is similar for the (α_1,ϕ_1) element. For the (ϕ_i,ϕ_j) element, proceed similarly and arrive at:

$$nE\left[S_n^c(\theta_0,\gamma_1)_{\phi_i}S_n^c(\theta_0,\gamma_2)_{\phi_j}\right]=nE\sum_{t=1}^{n}\frac{\epsilon_t^2}{n^2\omega_0^4}(y_{t-i}-\alpha_0)(y_{t-j}-\alpha_0)$$

$$=\frac{1}{\omega_0^2}\sum_{t=1}^{n}\frac{(y_{t-i}-\alpha_0)(y_{t-j}-\alpha_0)}{n}$$

As n tends to infinity, the sum goes to the corresponding element of the asymptotic autocovariance matrix of the vector: $\bar{y}_{t-1}=(y_{t-1}-\alpha_0\cdots y_{t-r}-\alpha_0)'$. All the expectations of the cross-products of the scores with respect to ω_0^2 and ω_1^2 on the one hand, and to α_0, α_1 and ϕ_i $(i=1,\ldots,r)$ on the other hand, are zero since the expectation of ϵ_t and ϵ_t^3 are 0.

We are therefore left with the cross-products of the scores with respect to the variance parameters ω_0^2 and ω_1^2. I derive first the expectation of the cross-product

of the scores with respect to ω_0^2:

$$nE\left[S_n^c(\theta_0,\gamma_1)_{\omega_0^2}S_n^c(\theta_0,\gamma_2)_{\omega_0^2}\right]=\frac{1}{n}E\sum_{t=1}^{n}\frac{1}{4\omega_0^4}\left[\frac{\epsilon_t^2}{\omega_0^2}-1\right]^2=\frac{1}{2\omega_0^4}$$

where the last equality follows from: $E(\epsilon_t^4)=3\omega_0^4$. The (ω_0^2,ω_1^2) element is given by:

$$nE\left[S_n^c(\theta_0,\gamma_1)_{\omega_0^2}S_n^c(\theta_0,\gamma_2)_{\omega_1^2}\right]$$
$$=\frac{1}{n}E\sum_{t=1}^{n}\sum_{s=1}^{n}\sum_{s_s(\gamma_2)=0}^{1}\cdots\sum_{s_{s-r}(\gamma_2)=0}^{1}\frac{1}{4\omega_0^4}\left[\frac{\epsilon_t^2}{\omega_0^2}-1\right]\left[s_s^2(\gamma_2)\left[\frac{\epsilon_s^2}{\omega_0^2}-1\right]\right]p_s$$

Proceeding as before, I obtain:

$$nE\left[S_n^c(\theta_0,\gamma_1)_{\omega_0^2}S_n^c(\theta_0,\gamma_2)_{\omega_1^2}\right]$$
$$=\frac{1}{n}E\sum_{s=1}^{n}\frac{1}{4\omega_0^4}\left[\frac{\epsilon_s^2}{\omega_0^2}-1\right]^2\cdot E\left(s_s^2(\gamma_2)|\Psi_n;\theta_0,\gamma_2\right)=\frac{\pi_2}{2\omega_0^4}$$

where I have used the i.i.d. assumption about ϵ, the independence between ϵ and $s(\gamma)$ and the law of iterated expectations for $s_s(\gamma_2)$. Similarly,

$$nE\left[S_n^c(\theta_0,\gamma_1)_{\omega_1^2}S_n^c(\theta_0,\gamma_2)_{\omega_0^2}\right]=\frac{\pi_1}{2\omega_0^4}$$

and

$$nE\left[S_n^c(\theta_0,\gamma_1)_{\omega_1^2}S_n^c(\theta_0,\gamma_2)_{\omega_1^2}\right]=\frac{\min(\pi_1,\pi_2)}{2\omega_0^4}.$$

3. *Asymptotic Null Distribution of LR Test: MS Model, One-dimensional Threshold and Structural Change Model with an Unknown Change Point.* In this section, I show the equivalence of the asymptotic null distribution derived for the LR test in the MS model to the one derived for the one-dimensional threshold model (by Chan 1990) and to a structural change model with an unknown change point (Andrews 1993), when the error term z_t is white noise.

When z_t is white noise, the MS model is given by (13) in Section 4.1. The one-dimensional threshold model is given by: $y_t=\alpha_0+I(y_{t-d}\le r)\alpha_1+e_t$, with $I(y_{t-d}\le r)=1$ if $y_{t-d}\le r$, 0 otherwise. The structural change model with one change in mean at an unknown point is as follows:

$$y_t=\alpha_0+\alpha_1D_t+\epsilon_t$$
$$D_t=0 \text{ for } t\le T_B$$
$$D_t=1 \text{ for } t>T_B$$
$$E(D_t)=\lambda$$

where λ is the percentage of the population where $D_t=1$.

For the one-dimensional threshold model, Chan (1990) shows that the asymptotic null distribution of the likelihood ratio is given by: $\sup_{s \in \tilde{S}} B_s^2/(s - s^2)$, where B_s stands for a Brownian bridge, $s(r) = E(I(X_t \leq r))$, with I the indicator function, and $\tilde{S} \subseteq [0,1]$. The parameter s is a function of the threshold r and gives the proportion of points on or below the threshold. The covariance kernel of the Brownian bridge B_s is given by: $s[\min(r,r')] - s(r)s(r')$. This is the numerator of the covariance function we have found for $g(\gamma)$, which was: $\omega_0^2[\min(\pi_1, \pi_2) - \pi_1\pi_2)]/\pi_1\pi_2(1 - \pi_1)(1 - \pi_2)$. I also showed that $C(\gamma) = g(\gamma)^2\pi(1 - \pi)$. Therefore, the asymptotic null distribution of the likelihood ratio statistic is the same for both models. In the structural change problem with an unknown change point, Andrews (1993) also finds that the covariance function is that of a vector Brownian bridge. In the one-dimensional change in mean at an unknown point, the asymptotic distribution of the LR test is also identical to the previous distribution (see Andrews 1993).

4. *Method for Simulating Chi-Square Processes.* According to the definition in footnote 5, a chi-square process is the product of Gaussian vector processes with a certain covariance matrix. I therefore propose a general method to generate Gaussian vector processes with a given covariance matrix. Assume that I select a set of T values in the parameter space Γ to generate the distribution of Sup C, say $\gamma_1, \gamma_2, \cdots, \gamma_T$. Then the first step consists in drawing T vectors of i.i.d. $\mathbb{N}(0,1)$ variates of dimension k, $\varepsilon(1), \varepsilon(2), \cdots, \varepsilon(T)$. As a second step, construct T Gaussian vectors of dimension k, $G(\gamma_1), \cdots, G(\gamma_T)$, as follows:

$$G(\gamma_1) = A(1,1)\varepsilon(1)$$

$$\vdots$$

$$G(\gamma_T) = A(T,1)\varepsilon(1) + A(T,2)\varepsilon(2) + \cdots + A(T,T)\varepsilon(T)$$

The $G(\gamma)$ vectors are Gaussian vectors, and have by construction variance and covariance matrices that are functions of the $A(.,.)$ matrices. Given the covariance function $K(\gamma_i, \gamma_j)$, one can find the corresponding $A(.,.)$ by the following steps:

1. Start with:

$$E[G(\gamma_1)G(\gamma_1)'] = K(\gamma_1, \gamma_1) = A(1,1)A'(1,1)$$

The last equality allows to compute the k^2 elements of the $A(1,1)$ matrix, given the k^2 elements of the $K(\gamma_1, \gamma_1)$ variance matrix.

2a. Determine the k^2 elements of the $A(2,1)$ matrix by:

$$E[G(\gamma_1)G(\gamma_2)'] = K(\gamma_1, \gamma_2) = A(1,1)A(2,1)'$$

given the k^2 elements of $A(1,1)$ computed in step 1.

2b. Determine the k^2 elements of the $A(2,2)$ matrix by:

$$E[G(\gamma_2)G(\gamma_2)'] = K(\gamma_2, \gamma_2)$$

$$= A(2,1)A'(2,1) + A(2,2)A'(2,2)$$

Given the elements of $A(2,1)$ computed at step 2a, one can find the elements of $A(2,2)$, given the $K(\gamma_2, \gamma_2)$ matrix.

3. For any $j = 3,\cdots,T$, determine the elements of the matrix $A(j,1)$ by:

$$E\left[G(\gamma_i)G(\gamma_j)'\right] = K(\gamma_i, \gamma_j) = A(i,1)A'(j,1) + A(i,2)A'(j,2) + \cdots + A(i,i)A'(j,i)$$

The $G(\cdot)s$ constructed in this way are Gaussian with covariance matrices $K(\gamma_i, \gamma_j)$, $i = 1,\cdots,T$, $j = 1,\cdots,T$. This algorithm is equivalent to calculating the Cholesky decomposition of the matrix:

$$\Omega = \begin{bmatrix} K(\gamma_1,\gamma_1) & K(\gamma_1,\gamma_2) & \cdots & K(\gamma_1,\gamma_T) \\ K(\gamma_2,\gamma_1) & K(\gamma_2,\gamma_2) & \cdots & K(\gamma_2,\gamma_T) \\ \vdots & \vdots & \vdots & \vdots \\ K(\gamma_T,\gamma_1) & K(\gamma_T,\gamma_2) & \cdots & K(\gamma_T,\gamma_T) \end{bmatrix}$$

to obtain: $\Omega = PP'$ and generate the vector $P\epsilon$, where ϵ is a $(Tk \times 1)$ vector of i.i.d. $N(0,1)$ variates. With large k, this numerical approach might be the only way to generate the covariance matrix, but for the relatively simple models studied in this paper, I derive analytically the elements of the P matrix (see the working paper version of the article for details).

REFERENCES

ANDREWS, D.W.K., "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica* 61 (1993), 821–856.

——— AND W. PLOBERGER, "Optimal Tests when a Nuisance Parameter is Present Only under the Alternative," *Econometrica* 62 (1994), 1383–1414.

——— AND ———, "Admissibility of the Likelihood Ratio Test when a Nuisance Parameter is Present Only under the Alternative," *The Annals of Statistics* 23 (1995), 1609–1629.

BOLDIN, M.D., "Characterizing Business Cycles with a Markov Switching Model," mimeo, University of Pennsylvania, 1989.

BONOMO, M. AND R. GARCIA, "Consumption and Equilibrium Asset Pricing: An Empirical Assessment," *Journal of Empirical Finance* 3 (1996), 239–265.

CECCHETTI, S.G., P. LAM, AND N.G. MARK, "Mean Reversion in Equilibrium Asset Prices," *American Economic Review* 80 (1990), 398–418.

———, ———, AND ———, "The Equity Premium and the Risk-Free Rate: Matching the Moments," *Journal of Monetary Economics* 31 (1993), 21–45.

CHAN, K.S., "Testing for Threshold Autoregression," *The Annals of Statistics* 18 (1990), 1886–1894.

COSLETT, S. AND L. LEE, "Serial Correlation in Discrete Variable Models," *Journal of Econometrics* 85 (1984), 79–97.

DAVIES, R.B., "Hypothesis Testing when a Nuisance Parameter is Present Only under the Alternative," *Biometrika* 64 (1977), 247–254.

———, "Hypothesis Testing when a Nuisance Parameter is Present Only under the Alternative," *Biometrika* 74 (1987), 33–43.

DAVIDSON, R. AND J.G. MACKINNON, "Several Tests for Model Specification in the Presence of Alternative Hypotheses," *Econometrica* 49 (1981), 781–793.

DIEBOLD, F.X., J.-H. LEE, AND G.C. WEINBACH, "Regime Switching with Time-Varying Transition Probabilities," in *Nonstationary Time Series Analysis and Cointegration*, C. Hargreaves ed., Series *Advanced Texts in Econometrics*, eds., C.W.J. Granger and G. Mizo (Oxford: Oxford University Press, 1993).

DUFOUR, J.-M., "Some Impossibility Theorems in Econometrics with Applications to Instrumental Variables, Dynamic Models and Cointegration," CRDE Working Paper No. 3695, Université de Montréal, 1995.

DURLAND, J.M. AND T.H. MCCURDY, "Modelling Duration Dependence in Cyclical Data Using a Restricted Semi-Markov Process," *Journal of Business and Economic Statistics* 12 (1994), 279–288.

ENGEL, C.M., AND J.D. HAMILTON, "Long Swings in the Dollar: Are They in the Data and Do Markets Know It?" *American Economic Review* 80 (1990), 689–713.

FILARDO, A.J., "Business Cycle Phases and Their Transitional Dynamics," *Journal of Business and Economic Statistics* 12 (1994), 299–308.

GALLANT, A.R., "Testing for a Nonlinear Regression Specification: A Nonregular Case," *Journal of the American Statistical Association* 72 (1977), 523–530.

GARCIA, R. AND P. PERRON, "An Analysis of the Real Interest Rate under Regime Shifts," *Review of Economics and Statistics* 78 (1996), 111–125.

GHYSELS, E., "On the Periodic Structure of the Business Cycle," *Journal of Business and Economic Statistics*, 12 (1994), 289–298.

HAMILTON, J.D., "Rational Expectations Econometric Analysis of Changes in Regime: an Investigation of the Term Structure of Interest Rates," *Journal of Economic Dynamics and Control* 12 (1988), 385–423.

———, "A new Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica* 57 (1989), 357–384.

———, "Specification testing in Markov-Switching Time Series Models," *Journal of Econometrics* 70 (1996), 127–157.

HANSEN, B.E., "Inference when a Nuisance Parameter is not Identified under the Null Hypothesis," Working Paper No. 296, Center for Economic Research, University of Rochester, 1991.

———, "The Likelihood Ratio Test under Non-standard Conditions: Testing the Markov Switching Model of GNP," *Journal of Applied Econometrics* 7 (1992), S61–S82.

———, "Inference when a Nuisance Parameter is not Identified under the Null Hypothesis," *Econometrica* 64 (1996), 413–430.

HASSETT, K., "Markov Switching and the Changing Cyclicality of the Aggregate Real Wage," mimeo, Columbia University, 1990.

LEE, L.-F. AND A. CHESHER, "Specification Testing When Score Test Statistics are Identically Zero," *Journal of Econometrics* 31 (1986), 121–149.

SICHEL, D.E., "Inventories and the Three Phases of the Business Cycles," *Journal of Business and Economic Statistics* 12 (1994), 269–277.

TURNER, C.M., R. STARTZ AND C.T. NELSON, "A Markov Model of Heteroskedasticity, Risk, and Learning in the Stock Market," *Journal of Financial Economics* 25 (1989), 3–22.