CIRANO
Centre interuniversitaire de recherche
en analyse des organisations

Série Scientifique
Scientific Series

Nº 95s-7

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

René Garcia

Montréal
Février 1995

CIRANO

Le CIRANO est une corporation privée à but non lucratif constituée en vertu de la Loi des compagnies du Québec. Le financement de son infrastructure et de ses activités de recherche provient des cotisations de ses organisations-membres, d'une subvention d'infrastructure du ministère de l'Industrie, du Commerce, de la Science et de la Technologie, de même que des subventions et mandats obtenus par ses équipes de recherche. La *Série Scientifique* est la réalisation d'une des missions que s'est données le CIRANO, soit de développer l'analyse scientifique des organisations et des comportements stratégiques.

CIRANO is a private non-profit organization incorporated under the Québec Companies Act. Its infrastructure and research activities are funded through fees paid by member organizations, an infrastructure grant from the Ministère de l'Industrie, du Commerce, de la Science et de la Technologie, and grants and research mandates obtained by its research teams. The Scientific Series *fulfils one of the missions of CIRANO: to develop the scientific analysis of organizations and strategic behaviour.*

Les organisations-partenaires / The Partner Organizations

- Ministère de l'Industrie, du Commerce, de la Science et de la Technologie.
- École des Hautes Études Commerciales.
- École Polytechnique.
- Université de Montréal.
- Université Laval.
- McGill University.
- Université du Québec à Montréal.
- Bell Québec.
- La Caisse de dépôt et de placement du Québec.
- Hydro-Québec.
- Banque Laurentienne du Canada.
- Fédération des caisses populaires de Montréal et de l'Ouest-du-Québec.
- Téléglobe Canada.

ISSN 1198-8177

Asymptotic Null Distribution of the Likelihood Ratio Test in Markov Switching Models*

René Garcia†

Abstract / Résumé

The Markov Switching Model, introduced by Hamilton (1988, 1989), has been used in various economic and financial applications where changes in regime play potentially an important role. While estimation methods for these models are by now well established, such is not the case for the corresponding testing procedures. The Markov switching 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. In these circumstances, the asymptotic distributions of the usual tests (likelihood ratio, Lagrange multiplier, Wald tests) are non-standard.

In this paper, we show that, if we treat the transition probabilities as nuisance parameters in a Markov switching model and set the null hypothesis in terms uniquely of the parameters governed by the Markov variable, the distributional theory proposed by Hansen (1991a) is applicable to Markov switching models under certain assumptions. Based on this framework, we derive analytically, in the context of two-state Markov switching models, the asymptotic null distribution of the likelihood ratio test (which is shown to be also valid for the Lagrange multiplier and Wald tests under certain conditions) and the related covariance functions. Monte Carlo simulations show that the asymptotic distributions offer a very good approximation to the corresponding empirical distributions.

Les modèles à changements de régime markoviens soulèvent un problème particulier connu dans la littérature statistique sous la rubrique des tests d'hypothèse dans les modèles où un paramètre de nuisance n'est pas identifié sous l'hypothèse nulle. Dans ces cas, les distributions asymptotiques des tests usuels (ratio de vraisemblance, multiplicateur de Lagrange, Wald) ne sont pas standard. Dans le présent article, nous montrons que, si nous traitons les probabilités de transition comme des paramètres de nuisance dans un modèle à changements de régime markoviens et fixons l'hypothèse nulle uniquement en fonction des paramètres régis par la variable de Markov, la théorie distributionnelle proposée par Hansen (1991) est applicable aux modèles à changements de régime markoviens sous certaines hypothèses. Dans ce cadre, nous dérivons analytiquement la distribution asymptotique du ratio de vraisemblance sous l'hypothèse nulle ainsi que les fonctions de covariance correspondantes pour divers modèles à changements de régime markoviens : un modèle à 2 moyennes avec erreurs non corrélées et homoscédastiques; un modèle à 2 moyennes avec des erreurs suivant un processus AR(r) ; et finalement un modèle à 2 moyennes et 2 variances avec des erreurs non corrélées. Dans les trois cas, des expériences de Monte Carlo montrent que les distributions asymptotiques dérivées offrent une très bonne approximation de la distribution empirique. La dérivation de la distribution asymptotique de la statistique du ratio de vraisemblance pour ces trois modèles simples markoviens à 2 états sera utile pour évaluer la signification statistique des résultats qui sont apparus dans la littérature et plus généralement pour offrir un ensemble de valeurs critiques aux chercheurs futurs dans ce domaine. Les valeurs critiques de la distribution asymptotique du test du ratio de vraisemblance dans les modèles à changements de régime markoviens sont considérablement plus élevées que les valeurs critiques impliquées par la distribution chi-carré standard.

---

* 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, Fangxiong Gong, Christian Gouriéroux, James D. Hamilton , Bruce E. Hansen, Werner Ploberger, and Michael Rockinger for helpful discussions and comments. Any errors are solely my responsability. Financial support from the Government of Québec (Fonds FCAR) and from the PARADI project funded by the Canadian International Development Agency is gratefully acknowledged.

† CIRANO, Département de sciences économiques, and CRDE, Université de Montréal, C.P. 6128, succursale A, Montréal (Québec) H3C 3J7.

The Markov Switching Model, introduced by Hamilton (1988, 1989), has been used in numerous economic and financial applications where changes in regime play potentially an important role[1]. In the most general form of this non-linear 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, 1991), Boldin (1989)), such is not the case for the testing procedures. There are however a few exceptions: Hamilton (1991) 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 for both the mean and variance; Boldin (1989) uses the Davies' (1987) upper bound test to determine the number of regimes; Garcia and Perron (1995) use Gallant's (1977) test and a J-test for non-nested models (Davidson and Mac-Kinnon (1981)), along with the Davies' test, to determine also the number of regimes. The use of these non-standard tests can be explained by the fact that the Markov switching 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. In these circumstances, the asymptotic distributions of the usual tests (likelihood ratio, Lagrange multiplier, Wald tests) are non-standard.

Hansen (1991a) provided a series of examples of economic models where this problem of unidentified nuisance parameters arises, and studied the corresponding asymptotic distribution theory. In general, the distributions of the tests are shown to depend upon the covariance function of chi-square processes. Since this covariance is model and data dependent, Hansen (1991a) proposes a simulation method to approximate the asymptotic null distribution and applies it to the threshold model. In Hansen (1993), the author proposes another method based on empirical process theory for the case where, in addition to the problem of unidentified nuisance parameters, the econometrician is faced with identically null scores. The use of this method is motivated by the existence of this double problem in Markov switching models. In this paper, we show that, if we treat the transition probabilities as nuisance parameters and set the null hypothesis in terms uniquely of the parameters (mean, variance or autoregressive coefficients) governed by the Markov variable, the distributional theory proposed by Hansen (1991a) is applicable to Markov switching models since the problem regarding the nullity of the scores can be side-stepped once some assumptions are made about the conditional state probabilities. Within Hansen's (1991a) framework, we derive analytically the asymptotic null distribution of the likelihood ratio test and the related covariance functions for various two-state Markov

[1] Cecchetti, Lam, and Mark (1990), Engel and Hamilton (1990), Garcia and Perron (1995), Hamilton (1988,1989), Hassett (1990), Turner, Startz, and Nelson (1989).

switching models: a two-mean model with an uncorrelated and homoskedastic noise component; a two-mean model with an AR(r) homoskedastic noise component; and finally a two-mean, two-variance model with an uncorrelated noise component. In all three cases, Monte-Carlo experiments show that the derived asymptotic distributions offer a very good approximation to the empirical distribution. The derivation of the asymptotic distribution of the likelihood ratio statistic for these three simple two-state Markov models will prove useful to assess the statistical significance of the results that appeared in the literature (i.e. Cechetti, Lam, and Mark (1990), Hamilton (1989), Turner, Startz and Nelson (1989)) and to generally offer a set of critical values to future researchers. This method offers a useful alternative to Hansen's (1993) methodology, the application of which is limited by computational requirements.

In Section 1, we present a general two-state Markov switching model, explain the problem of non-identification of some nuisance parameters under the null hypothesis, and set up the testing problem as the supremum of likelihood ratio statistics. In Section 2, we briefly present Hansen's (1991a) asymptotic distribution theory for the trinity of tests (likelihood ratio, Lagrange multiplier, and Wald) in models where nuisance parameters are not identified under the null. In Section 3, the covariance function for the general two-state Markov switching model is derived. Section 4 provides the asymptotic null distributions of the LR statistic for three specific Markov switching 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 Markov switching model. We also compute the power of the LR test for these three models. Section 5 concludes.

1. Testing in the Context of Markov Switching Models

The two-state Markov switching model is defined as follows[2]:

$$\begin{aligned} &y_t=\alpha_0+\alpha_1 S_t+z_t \\ &z_t=\phi_1 z_{t-1}+\dots+\phi_r z_{t-r}+(\omega_0+\omega_1 S_t)\epsilon_t \\ &P(S_t=1|S_{t-1}=1)=p \\ &P(S_t=0|S_{t-1}=0)=q \qquad (t=1,\dots,n) \end{aligned} \tag{1}$$

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 Markov

[2] This specification encompasses the specifications most frequently used in the literature for two-state Markov switching models. Two exceptions are noteworthy: the state-dependent autoregressive specification used in Garcia and Perron (1995) or the time-varying transition probability model used in Diebold, Lee, and Weinbach (1993) and Filardo (1992).

switching model. The null hypothesis can be expressed as either $\{\alpha_1=0,\omega_1=0\}$ or $\{p=0\}$ or $\{p=1\}$. To see the problem of unidentified nuisance parameters under the null, note that if α_1 and ω_1 are equal to zero, the transition probability parameters p and q are unidentified since any value between 0 and 1 will leave the likelihood function unchanged. As for the problem of identically zero scores, note that under $\{p=1\}$, the scores with respect to p , q , α_1 and ω_1 will be identically zero under the null and the asymptotic information matrix will be singular.[3] 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 (1993) analysis regarding the likelihood ratio test under non-standard conditions, since the two problems of unidentified nuisance parameters under the null and identically zero scores are present. He uses empirical process theory to derive a bound for the asymptotic distribution of a standardized likelihood ratio statistic. Although the method is appealing since it addresses both above-mentioned violations of the conventional regularity conditions, it seems to run rapidly into computational limitations. Hansen's testing procedure requires to set a grid 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 and to stay computationally tractable, it is necessary to limit the number of grid points. Also, Hansen's method provides a bound for the likelihood ratio statistic and not a critical value.

The problem comes from the fact that at $\{p=0\}$ or $\{p=1\}$, the scores with respect to α_1 and ω_1 are zero. Although these two points 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 non-linearity of this type in the series and accept the null of a linear model or try another non-linear 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. A way to approach the problem is to treat the transition probability parameters p and q truly as nuisance parameters, since if we fix them at predetermined values other than 0 and 1, there are no scores with respect to these probability parameters. Moreover, it is shown that the information matrix for the remaining parameters becomes non-singular at $\{\alpha_1=0,\omega_1=0\}$ once some assumptions are made about the conditional state probabilities. We can therefore derive the likelihood ratio statistic for each such set of values for the two transition probability parameters over a certain parameter space, say Γ where p and q lie. The likelihood

---

[3] This point is clear when looking at the element of the score vector corresponding to α_1 derived in Lemma 1.

ratio of the original problem is therefore 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:

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

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

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

ith $\gamma=(p,q)$ and $\theta=(\alpha_0,\alpha_1,\phi_1,\cdots,\phi_r,\omega_0^2,\omega_1^2)$. 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 (see Hansen (1991), theorem (3)):

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

where Γ is a metric space from which the values 0 and 1 have to be excluded to keep the information matrix positive definite as mentioned in Section 1.

In the context of hypothesis tests when a nuisance parameter is present only under the alternative, Andrews and Ploberger (1993) show that the sup LR test is a best test, in a certain sense, against alternatives that are sufficiently distant from the null hypothesis. In Andrews and Ploberger (1994), they 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 LR test is not admissible in this class of tests. Although the Markov switching model is not included in the examples covered in the paper[4], we compare in terms of size the sup LR test to the exponential LR test for various two-state Markov switching models.

---

[4] Andrews and Ploberger (1994) mention specifically that their test is not applicable to the Markov switching model because the information matrix is not uniformly positive definite over the Γ space. This is due to the identically null score problem mentioned in the introduction. We will see in Section 3 what assumptions are necessary to apply the Sup LR test as well as the average Exp LR test to Markov Switching models.

2. The Asymptotic Distribution of the Likelihood Ratio Test

In this section, we state a restricted version of a theorem appearing in Hansen (1991a).

Theorem 1: Under the set of assumptions stated in Appendix A and in the absence of serial correlation and heteroskedasticity in the noise function:[5]

$$LR_n \rightarrow SupC \equiv \sup_{\gamma \in \Gamma} C(\gamma) \tag{5}$$

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

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

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:

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

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

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

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

Under the assumptions of the theorem, Hansen (1991a) shows that, as in the classical theory, the LR, LM and Wald statistics all have the same asymptotic distribution.

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 section, this is the case in our model. This problem arises also in structural change models when the timing of the change is an unknown fraction of the sample size. In this case,

---

[5] By taking the absence of serial correlation and heteroskedasticity as given, we assume that the specification tests related to the Markov switching specification have been run. The goal is to focus on the mispecification test of the linear null against the Markov switching alternative.

[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 $K(\gamma_1,\gamma_2) = E[G(\gamma_1) G(\gamma_2)']$.

the fraction has to be bounded away from 0 and 1. 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[7] in (θ,γ) of $Q_n(\theta,\gamma)-Q(\theta,\gamma)$ and $V_n(\theta,\gamma)-V(\theta,\gamma)$ over the corresponding spaces. Verifying whether all the regularity conditions hold in the present context appears difficult. We will therefore assume that these conditions hold and, short of a proof, 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.

3. Derivation of the Covariance Function for the General Two-State Markov Switching Model

The two-state Markov switching model (1) can be rewritten as follows:

$$y_t-\phi_1 y_{t-1}-\phi_2 y_{t-2}-\cdots-\phi_r y_{t-r}-\alpha_0(1-\phi_1-\phi_2-\cdots-\phi_r)-\alpha_1(S_t-\phi_1 S_{t-1}-\phi_2 S_{t-2}-\cdots-\phi_r S_{t-r})=(\omega_0+\omega_1 S_t)\epsilon_t \tag{8}$$

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

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

The derivatives with respect to each parameter of the logarithm of the probability of observation t, conditional upon Ψ_t, the information at time t, and evaluated at θ_0 are therefore given by:

---

[7] In Hansen (1991a)(footnote 1, page 10), stochastic equicontinuity is defined as follows: $\{G_n(\lambda)\}$ is stochastically equicontinuous on Λ if for all $\epsilon>0$ and $\eta>0$ there exists some $\delta>0$ such that $\lim_n P\left[\sup_{\lambda\in\Lambda}\ \sup_{\rho(\lambda,\lambda')<\delta} |G_n(\lambda')-G_n(\lambda)|>\epsilon\right]<\eta$, where $\rho(.,.)$ denotes the distance metric defined on Λ.

$$\frac{\partial \log\ p(y_t|\Psi_t,\theta,\gamma)}{\partial \alpha_0} = (1-\sum_{i=1}^{r}\phi_i)\frac{\epsilon_t}{(\omega_0+\omega_1 s_t(\gamma))^2}$$

$$\frac{\partial \log\ p(y_t|\Psi_t,\theta,\gamma)}{\partial \alpha_1} = (s_t(\gamma)-\sum_{i=1}^{r}\phi_i s_{t-i}(\gamma))\frac{\epsilon_t}{(\omega_0+\omega_1 s_t(\gamma))^2}$$

$$\frac{\partial \log\ p(y_t|\Psi_t,\theta,\gamma)}{\partial \phi_i} = -(\alpha_0+\alpha_1 s_{t-i}(\gamma)-y_{t-i})\frac{\epsilon_t}{(\omega_0+\omega_1 s_t(\gamma))^2}$$

$$i=1,\ldots,r$$

$$\frac{\partial \log\ p(y_t|\Psi_t,\theta,\gamma)}{\partial \omega_0^2} = -\frac{1}{2}\frac{[1+(\omega_0^2)^{-1/2}\omega_1 s_t(\gamma)]}{(\omega_0+\omega_1 s_t(\gamma))^2}+\frac{\epsilon_t^2[1+(\omega_0^2)^{-1/2}\omega_1 s_t(\gamma)]}{2(\omega_0+\omega_1 s_t(\gamma))^4}$$

$$\frac{\partial \log\ p(y_t|\Psi_t,\theta,\gamma)}{\partial \omega_1^2} = -\frac{1}{2}\frac{[s_t(\gamma)^2+(\omega_1^2)^{-1/2}\omega_0 s_t(\gamma)]}{(\omega_0+\omega_1 s_t(\gamma))^2}+\frac{\epsilon_t^2[s_t(\gamma)^2+(\omega_1^2)^{-1/2}\omega_0 s_t(\gamma)]}{2(\omega_0+\omega_1 s_t(\gamma))^4} \quad (10)$$

Using the following equalities:

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

we derive, in the following lemma, the corresponding scores.

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:

$$S_n^c(\theta_0,\gamma)_{\alpha_0} = \frac{1}{n}\sum_{t=1}^{n}\sum_{s_t(\gamma)=0}^{1}\cdots\sum_{s_{t-r}(\gamma)=0}^{1}(1-\sum_{i=1}^{r}\phi_i)\frac{\epsilon_t}{(\omega_0+\omega_1 s_t(\gamma))^2}\cdot p_t$$

$$S_n^c(\theta_0,\gamma)_{\alpha_1} = \frac{1}{n}\sum_{t=1}^{n}\sum_{s_t(\gamma)=0}^{1}\cdots\sum_{s_{t-r}(\gamma)=0}^{1}(s_t(\gamma)-\sum_{i=1}^{r}\phi_i s_{t-i}(\gamma))\frac{\epsilon_t}{(\omega_0+w_1 s_t(\gamma))^2}\cdot p_t$$

$$S_n^c(\theta_0,\gamma)_{\phi_i} = \frac{1}{n}\sum_{t=1}^{n}\sum_{s_t(\gamma)=0}^{1}\cdots\sum_{s_{t-r}(\gamma)=0}^{1}-(\alpha_0-y_{t-i})\frac{\epsilon_t}{(\omega_0+\omega_1 s_t(\gamma))^2}\cdot p_t$$

$$i=1,\ldots,r$$

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

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

where $p_t = p(S_t(\gamma)=s_t(\gamma),\cdots,S_{t-r}(\gamma)=s_{t-r}(\gamma)|y_n,\cdots,y_{-r+1};\theta_0,\gamma)$.

A proof is provided in Appendix B. In the element of the score vector with respect to α_1, note that $p_t = p(S_t(\gamma)=s_t(\gamma),\cdots,S_{t-r}(\gamma)=s_{t-r}(\gamma)|y_n,\cdots,y_{-r+1};\theta_0,\gamma)$ is equal to $\pi=(1-q)/(2-p-q)$, the unconditional probability of being in state 1, for each t since under H_0 ($\alpha_1=0$, $\omega_1=0$) these probabilities cannot be filtered out. Therefore, the score becomes identically null. The way we side-step this problem of identically null scores is to assume that if the unconditional probability is π, then $\pi \times n$ of the points will be affected to one regime, and the remainder to the other regime[8]. In other words, a proportion π of p_t will have value 1, while for the complement $(1-\pi)$ of the points, p_t will be zero. Intuitively, this reassignment of probabilities reflects the way the filtering algorithm will most often assign the probabilities in a finite sample drawn under the null assumption. Asymptotically, as n tends to infinity, the score will still be zero at the null θ_0. To see whether the assumption is valid or not, we will compare the asymptotic distribution of the LR ratio test derived under this assumption to the empirical distribution for various models. Since the score is no longer identically zero, we can state the following lemma.

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

$$K(\gamma_1,\gamma_2)=\frac{1}{\omega_0^2}\begin{bmatrix} (1-\sum_{i=1}^{r}\phi_i)^2 & \pi_2(1-\sum_{i=1}^{r}\phi_i)^2 & 0 & \cdots & 0 & 0 \\ \pi_1(1-\sum_{i=1}^{r}\phi_i)^2 & \min(\pi_1,\pi_2)A^* & 0 & \cdots & 0 & 0 \\ 0 \ \cdots & 0 & \omega_0^2 R & & 0 & 0 \\ 0 \ \cdots & 0 & 0 & \cdots & \frac{1}{2\omega_0^2} & \frac{\pi_2}{2\omega_0^2} \\ 0 \ \cdots & 0 & 0 & \cdots & \frac{\pi_1}{2\omega_0^2} & \frac{\min(\pi_1,\pi_2)}{2\omega_0^2} \end{bmatrix} \qquad (12)$$

where A^* is given by:

[8] Another way to side-step the problem of identically null scores will be to set the null arbitrarily close to zero, since in this case the p_t could be filtered out.

$$
\begin{aligned}
A^{*} = \; & 1+\phi_1^2+\phi_2^2+\cdots\phi_r^2 \\
& -2T^{*1}[\phi_1-\sum_{i=2}^{r}\phi_i\phi_{i-1}] \\
& -2T^{*2}[2,2][\phi_2-\sum_{i=3}^{r}\phi_i\phi_{i-2}] \\
& - \quad \cdots \quad - \\
& -2T^{*r}[2,2]\phi_r
\end{aligned}
$$

with $T^{*i}[2,2]$ denoting the second row, second column element of the matrix T raised to the power i, where T^{*} is the transition probability matrix of the Markov variable S_t:

$$
T^{*}=\begin{bmatrix} q^{*} & 1\text{-}q^{*} \\ 1\text{-}p^{*} & p^{*} \end{bmatrix}
$$

with the star denoting the transition probabilities corresponding to $\max(\pi_1,\pi_2)$. The $\min(\pi_1,\pi_2)$ element in the $K(\pi_1,\pi_2)$ covariance matrix is obtained under the assumption that the points of the sample which are classified in state 1 under π_1 will also be classified in state 1 under π_2, with π_2 greater than π_1. Finally, R denotes the autocovariance matrix of $\{y_t\}$, where $\{y_t\}$ is an AR(r). A proof of Lemma 2 is provided in Appendix C.

In the next section, we will derive, based on this general covariance function, the asymptotic null distribution of LR_n for three specific models: a two-mean model with an uncorrelated and homoskedastic noise component, used by Cecchetti, Lam, and Mark (1990) to model the annual growth rate of consumption; a two-mean model with an AR(r) homoskedastic noise component, used by Hamilton (1989) to model the quarterly growth rate of output; and finally a two-mean, two-variance model with an uncorrelated noise component, used by Turner, Startz and Nelson (1989) to model a series of stock returns. The derived critical values will allow us to test formally the linear null againt the Markov switching alternative in all three cases. We also compute the power of the Sup LR test for these three models.

4. Asymptotic null distributions of the Sup LR statistic

Based on theorem 1 and Lemmas 1 and 2, we derive for each of the three above-mentioned Markov switching models $\bar{K}(\gamma_1,\gamma_2)$, the covariance function of the chi-square process $C(\gamma)$.

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:

$$\begin{aligned} & y_t = \alpha_0 + \alpha_1 S_t + \epsilon_t \\ & P(S_t = 1 \mid S_{t-1} = 1) = p \\ & P(S_t = 0 \mid S_{t-1} = 0) = q \end{aligned} \tag{13}$$

The limiting distribution of LR_n[9] is $SupC = \underset{\gamma \in \Gamma}{Sup}\, C(\gamma)$, where $C(\gamma)$ is a chi-square process with covariance:

$$\bar{K}(\gamma_1,\gamma_2) = \omega_0^2 \frac{\min(\pi_1,\pi_2) - \pi_1\pi_2}{\pi_1\pi_2(1-\pi_1)(1-\pi_2)} \quad \text{where} \quad \pi_i = \frac{1-q_i}{2-p_i-q_i} \tag{14}$$

Proof: By theorem 1, the covariance of $C(\gamma)$ is given by:

$$\bar{K}(\gamma_1,\gamma_2) = [0 \;\; 1 \;\; 0] V(\gamma_1)^{-1} K(\gamma_1,\gamma_2) V(\gamma_2)^{-1} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

$$= \omega_0^2 \begin{bmatrix} \frac{-1}{1-\pi_1} & \frac{1}{\pi_1(1-\pi_1)} & 0 \end{bmatrix} \frac{1}{\omega_0^2} \begin{bmatrix} 1 & \pi_2 & 0 \\ \pi_1 & \min(\pi_1,\pi_2) & 0 \\ 0 & 0 & \frac{1}{2\omega_0^2} \end{bmatrix} \omega_0^2 \begin{bmatrix} \frac{-1}{1-\pi_2} \\ \frac{1}{\pi_2(1-\pi_2)} \\ 0 \end{bmatrix}$$

$$= \omega_0^2 \frac{\min(\pi_1,\pi_2) - \pi_1\pi_2}{\pi_1\pi_2(1-\pi_1)(1-\pi_2)}$$

In order to simulate the distribution of $SupC$, we follow the general method for simulating chi-square processes described in Appendix D. In this case, the $G(\gamma)$ and ε vectors reduce to scalars and we generate at each draw a Tx1 vector of $g(\gamma)$ as follows:

---

[9] In earlier versions of the paper, it was shown that the asymptotic null distribution of the LR test in this particular model was identical to the ditribution of the LR test in a one-dimensional threshold model (derived by Chan (1990)) and to a structural change model with an unknown change point (Andrews (1993)).

$$g(\gamma_1)=\left[\frac{1}{\pi_1(1-\pi_1)}\right]^{1/2}\epsilon(1)$$
$$g(\gamma_2)=\left[\frac{\pi_1}{1-\pi_1}\right]^{1/2}\frac{1}{\pi_2}\epsilon(1)+\frac{(\pi_2-\pi_1)^{1/2}}{\pi_2[(1-\pi_1)(1-\pi_2)]^{1/2}}\epsilon(2) \qquad (15)$$
$$\vdots \qquad \vdots$$
$$g(\gamma_T)=\left[\frac{\pi_1}{1-\pi_1}\right]^{1/2}\frac{1}{\pi_T}\epsilon(1)+\frac{(\pi_2-\pi_1)^{1/2}}{\pi_T[(1-\pi_1)(1-\pi_2)]^{1/2}}\epsilon(2)+\dots+\frac{(\pi_T-\pi_{T-1})^{1/2}}{\pi_T[(1-\pi_{T-1})(1-\pi_T)]^{1/2}}\epsilon(T)$$

The chi-square process $C(\gamma)$ is therefore equal to:

$$C(\gamma)=g(\gamma)^2\pi(1-\pi) \qquad (16)$$

To obtain the supremum of $C(\gamma)$ over Γ, we must fix the bounds of the parameter space Γ or, in our case, of the function π of the parameters p and q. Since the $K(.,.)$ are singular for π equal to 0 or 1, we must fix the bounds away from 0 and 1 for the theory to be valid.

However, it remains to be determined how far away from 0 and 1 we 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. To see which bounds will produce the best asymptotic approximation to the finite sample distribution, we 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 Markov switching alternative in (13).

Since the true model is a model with no change in regime, one might expect when estimating the Markov switching 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 (1995). This means that a 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, we hope to eliminate or at least reduce the number of local maxima. We 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 1. The 99% and 95% critical values are 14.02 et 10.89 respectively. The next two columns of Table 1 show the critical values of the asymptotic distribution obtained with 10,000 replications of $SupC$ for $\pi\in[0.01,0.99]$ and $\pi\in[0.15,0.85]$, with increments of 0.001. 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 $\chi^2(1)$, the distribution of the LR test in the classical theory.

To compare the exponential LR test to the Sup LR test, we report in the upper part of Table 2 the actual size of both tests for nominal sizes of 1% and 5% under the $\pi \in [0.01, 0.99]$ range. A nominal 5% test with the exponential LR test will have an actual size of around 30%, compared with 6.5% for the Sup LR.

Finally, in the lower part of Table 2, we report the power of the Sup LR test for the Markov switching model of consumption growth estimated by Cecchetti, Lam, and Mark (1990). We generated a 1,000 series based on the following estimates (α_0=0.228, α_1=-0.0926, ω_0=0.0320, p=0.5279, q=0.9761), and estimated for each series both the linear and the Markov switching 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, we obtain the following specification for the Markov switching model:

$$\begin{aligned} y_t &= \alpha_0 + \alpha_1 S_t + z_t \\ z_t &= \phi_1 z_{t-1} + \dots + \phi_r z_{t-r} + \epsilon_t \end{aligned} \tag{17}$$

The limiting distribution of LR_n is $SupC = \sup_{\gamma \in \Gamma} C(\gamma)$, where $C(\gamma)$ is a chi-square process with covariance:

$$\bar{K}(\gamma_1,\gamma_2) = \omega_0^2 \frac{\min(\pi_1,\pi_2)A^* - \pi_1\pi_2(1-\sum_{i=1}^{r}\phi_i)^2}{\pi_1\pi_2[A_2-(1-\sum_{i=1}^{r}\phi_i)^2\pi_2][A_1-(1-\sum_{i=1}^{r}\phi_i)^2\pi_1]} \tag{18}$$

Proof: The expression follows from:

$$\bar{K}(\gamma_1,\gamma_2) = \iota_{\alpha_1} V(\gamma_1)^{-1} K(\gamma_1,\gamma_2) V(\gamma_2)^{-1} \iota'_{\alpha_1}$$

and:

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

where: $\Delta=\pi(1-\sum_{i=1}^{r}\phi_i)^2[A-(1-\sum_{i=1}^{r}\phi_i)^2\pi]$ and $K(\gamma_1,\gamma_2)$ is deduced from the expression given in (12) for an homoskedastic process.

We note that the parameter space Γ is now two-dimensional since both p and π are present in the covariance function, and also that the covariance function depends on the autoregressive parameters. To assess the performance of the sup LR test in the autoregressive case, we first study the AR(1) case in detail. We determine the bounds of the parameter space over p and π that give the asymptotic distribution which provides the best approximation to the empirical distribution. We 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[10] and -0.5. The true model is the AR(1) model and the alternative is the two-state Markov model in (17) with r=1. We generated the distribution using 1,000 replications of the true model and estimating the alternative Markov switching model, starting with six different sets of values for 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 3. They appear to be smaller than the values obtained for the uncorrelated and homoskedastic case. Also it has to

[10] 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).

be noted that the empirical critical values do not seem to depend on the value of the autoregressive parameter.

To generate the asymptotic distribution of ***SupC***, 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.

To simulate the distribution of ***SupC***, we generate at each draw a random Tx1 vector of $g(\gamma)$, as follows:

$$g(\gamma_1)=\left[\frac{1}{\pi_1[p_1,\pi_1]}\right]^{1/2}\epsilon(1)$$

$$g(\gamma_2)=\left[\frac{\pi_1}{[p_1,\pi_1]}\right]^{1/2}\frac{1}{\pi_2}\epsilon(1)+\frac{(\pi_2[p_1,\pi_1]-\pi_1[p_2,\pi_2])^{1/2}}{\pi_2[p_1,\pi_1]^{1/2}[p_2,\pi_2]^{1/2}}\epsilon(2)$$

$$\vdots$$

$$g(\gamma_T)=\left[\frac{\pi_1}{[p_1,\pi_1]}\right]^{1/2}\frac{1}{\pi_T}\epsilon(1)+\frac{(\pi_2[p_1,\pi_1]-\pi_1[p_2,\pi_2])^{1/2}}{\pi_T[p_1,\pi_1]^{1/2}[p_2,\pi_2]^{1/2}}\epsilon(2)+\dots+\frac{(\pi_T[p_{T-1},\pi_{T-1}]-\pi_{T-1}[p_T,\pi_T])^{1/2}}{\pi_T[p_{T-1},\pi_{T-1}]^{1/2}[p_T,\pi_T]^{1/2}}\epsilon(T)$$

(19)

where $[p,\pi]$ is defined as follows:

$$[p,\pi]=1+\phi_1^2-2\phi_1 p-(1-\phi_1)^2\pi. \qquad (20)$$

The chi-square process $C(\gamma)$ is therefore equal to:

$$C(\gamma)=g(\gamma)^2\pi(1+\phi_1^2-2\phi_1 p-(1-\phi_1)^2\pi) \qquad (21)$$

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 distribution do not seem to depend on the value of the autoregressive parameter. This is confirmed in Table 4, 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 autoregressive parameter.

To compare the exponential LR test to the Sup LR test, we report in Table 5 the actual sizes of both tests for nominal sizes of 1% and 5%, for ϕ=0.337 and ϕ=-0.5. For both values of ϕ_1, a nominal 5% test with the exponential LR test will have an actual size of around 25%, compared with a value close to 5% for the Sup LR test.

To assess the power of the Sup LR test for a model with an autoregressive structure, we will 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 Markov switching model identical to the model in (17) with a fourth-order autoregressive noise function. The maximum likelihood estimation results are presented in Table 6 along with the maximum likelihood estimates of the AR(4) model. We can first note that the likelihood ratio statistic (2L) 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.

To assess the estimation results of Hamilton, we need to generate the distribution of ***SupC*** defined at the beginning of the section for r=4. For the autoregressive parameters, we use the estimated values shown in Table 6 for the Markov trend model. The results are shown in Table 7 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, we cannot reject at usual levels the null of an AR(4) against the Markov trend model for the first differences of US log GNP. We reach therefore the same conclusion as Hansen (1993) with his simulation-based bound method.

Finally, in the right hand side part of Table 7, we report the empirical distribution of the LR statistic when the data generating process is the Markov switching model of GNP estimated by Hamilton (1989) with the parameter values shown in Table 6. 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 Two-mean model with an uncorrelated and heteroskedastic noise component

The two-state MSM with an uncorrelated and heteroskedastic noise component is given by:

$$\begin{aligned} &y_t=\alpha_0+\alpha_1 S_t+z_t \\ &z_t=(\omega_0+\omega_1 S_t)\epsilon_t \\ &P(S_t=1|S_{t-1}=1)=p \\ &P(S_t=0|S_{t-1}=0)=q \end{aligned} \tag{22}$$

In this case, the covariance matrix $K(\gamma_1,\gamma_2)$ of the score vectors given in (12) reduces to:

$$V=\frac{1}{\omega_0^2}\begin{bmatrix} 1 & \pi_2 & 0 & 0 \\ \pi_1 & \min(\pi_1,\pi_2) & 0 & 0 \\ 0 & 0 & \frac{1}{2\omega_0^2} & \frac{\pi_2}{2\omega_0^2} \\ 0 & 0 & \frac{\pi_1}{2\omega_0^2} & \frac{\min(\pi_1,\pi_2)}{2\omega_0^2} \end{bmatrix} \tag{23}$$

The limiting distribution of LR_n is therefore $SupC = \underset{\gamma \in \Gamma}{Sup}\, C_1(\gamma)+C_2(\gamma)$, where $C_1(\gamma)$ and $C_2(\gamma)$ are chi-square processes with respective covariances:

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

$$\bar{K}_2(\gamma_1,\gamma_2)=2\omega_0^4 \frac{\min(\pi_1,\pi_2)-\pi_1\pi_2}{\pi_1\pi_2(1-\pi_1)(1-\pi_2)} \tag{24}$$

The proof follows exactly the steps described before for the homoskedastic case. Finally, we arrive at the following covariance matrix for $C(\gamma)$.

$$\bar{K}(\gamma_1,\gamma_2)=\begin{bmatrix} \frac{\min(\pi_1,\pi_2)-\pi_1\pi_2}{\pi_1\pi_2[1-\pi_1][1-\pi_2]} & 0 \\ 0 & 2\omega_0^2 \left[\frac{\min(\pi_1,\pi_2)-\pi_1\pi_2}{\pi_1\pi_2[1-\pi_1][1-\pi_2]}\right] \end{bmatrix} \tag{25}$$

Therefore, $C(\gamma)$ can be represented as the sum of two chi-square processes with the covariances shown above. To simulate the distribution of $SupC$, we therefore follow the method described in section 4.1 to generate two independent Gaussian vectors $g_1(\gamma)$ and $g_2(\gamma)$. Table 8 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 1,000 series under the null hypothesis ($\alpha_1=0,\ \omega_1=0$) and estimating the likelihood ratio between the linear homoskedastic model and the heteroskedastic Markov switching model, using again six sets of starting values for the parameters.[11]

The comparison of the exponential LR test and the Sup LR test is reported in Table 9, which shows the actual sizes of both tests for nominal sizes of 1% and 5%. Like in the homoskedastic case, the size of the exponential LR will be distorted, but the distorsion will be stronger: a nominal 5% test will have an actual size of around 60%, compared with a value close to 5% for the Sup LR.

Finally, in the lower part of Table 9, we report the power of the Sup LR test for the Markov switching model of stock returns estimated by Turner, Startz, and Nelson (1989). We generated a 1,000 series based on the following parameter

---

[11] 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.

estimates (α_0=0.677, α_1=-2.652, ω_0=2.693, ω_1=2.396, p=0.767, q=0.950), and estimated for each series both the linear and the Markov switching model. A 5% Sup LR test will have in this case a power of about 60%.

5. Conclusion

This paper has shown that the critical values of the asymptotic null distribution of the likelihood ratio test in Markov switching models (also valid, under certain conditions, for the Lagrange multiplier and Wald tests) are considerably higher than the critical values implied by the standard $\chi^2(1)$ distribution. The paper also shows, for a series of two-state Markov switching models that the asymptotic distribution is very close to the small sample distribution.

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 Markov switching models with the same specification for various economic and financial time series. For models with a correlated noise function, we 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. We have shown however that this distribution is insensitive to the values of the autoregressive parameters.

For the AR(4) GNP model estimated by Hamilton (1989), we generated the asymptotic distribution of the Sup LR test and shown 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 Markov switching model in the GNP growth series. We 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.

REFERENCES

Andrews, D.W.K. (1993), “Tests for parameter instability and structural change with unknown change point”, *Econometrica,* 61, 821-856.

Andrews, D.W.K., and Ploberger W. (1993), “Admissibility of the likelihood ratio test when a nuisance parameter is present only under the alternative”, unpublished manuscript, Cowles Foundation, Yale University.

Andrews, D.W.K. and Ploberger W. (1994), “Optimal tests when a nuisance parameter is present only under the alternative”, *Econometrica*,

Boldin, M.D. (1989), “Characterizing business cycles with a Markov switching model”, mimeo, University of Pennsylvania.

Cechetti, S.G., Lam P., and Mark, N.G. (1990), “Mean reversion in equilibrium asset prices”, *American Economic Review*, 80, 398-418.

Chan, K.S. (1990), “Testing for threshold autoregression”, *The Annals of Statistics*, 18, 1886-1894.

Coslett, S. and L. Lee (1984), “Serial correlation in discrete variable models”, *Journal of Econometrics*, 85, 79-97.

Davies, R.B. (1987), “Hypothesis testing when a nuisance parameter is present only under the alternative”, *Biometrika*, 74, 33-43.

Davidson, R. and MacKinnon, J.G. (1981), “Several tests for model specification in the presence of alternative hypotheses”, *Econometrica*, 49, 781-793.

Engel, C.M., and Hamilton, J.D. (1990), “Long swings in the dollar:are they in the data and do markets know it? *American Economic review*, 80, 689-713.

Gallant, A.R. (1977), “Testing for a nonlinear regression specification: A nonregular case”, *Journal of the American Statistical Association*, 72, 523-530.

Garcia, R. and Perron, P. (1995), “An analysis of the real interest rate under regime shifts”, forthcoming in the *Review of Economics and Statistics*.

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.

Hamilton, J.D. (1989), “A new approach to the economic analysis of nonstationary time series and the business cycle”, *Econometrica*, 57, 357-384.

Hamilton, J.D. (1991), “Specification testing in Markov-switching time series models”, mimeo, University of Virginia.

Hansen, B.E. (1991a), "Inference when a nuisance parameter is not identified under the null hypothesis", mimeo, Rochester University.

Hansen, B.E. (1993), "The likelihood ratio test under non-standard conditions: testing the Markov switching model of GNP", *Journal of Applied Econometrics*,

Hassett, K. (1990), "Markov switching and the changing cyclicality of the aggregate real wage", mimeo, Columbia University.

Turner, C.M., Startz, R. and Nelson, C.T. (1989), "A Markov model of heteroskedasticity, risk, and learning in the stock market," National Bureau of Economic Research, Inc.

TABLE 1
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
85%	7.47	7.72	6.13
80%	6.71	7.04	5.47
75%	6.08	6.49	4.93
70%	5.53	6.05	4.50
65%	5.04	5.67	4.15
60%	4.73	5.33	3.83
55%	4.35	5.00	3.51
50%	4.03	4.72	3.25
45%	3.70	4.45	3.00
40%	3.45	4.18	2.77
35%	3.12	3.91	2.55
30%	2.81	3.66	2.35
25%	2.52	3.41	2.14
20%	2.23	3.14	1.93
15%	1.78	2.88	1.72
10%	1.39	2.55	1.48
5%	0.99	2.15	1.18
1%	0.45	1.56	0.81

TABLE 2
Size and Power of the Sup LR test
in a Two-state Markov Switching Model with
<u>Uncorrelated and Homoskedastic Errors</u>

Nominal Size	1%	5%
Sup LR	1.4%	6.5%
Exp LR	10.2%	28.6%
Power - Sup LR	38.2%	49.5%

Note: The power of Sup LR test has been computed with respect to the model estimated by Cecchetti, Lam, and Mark (1990).

TABLE 3
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

% Dist.	Asymptotic Distribution Critical Values ϕ_1=0.337	Empirical Distribution Critical Values ϕ_1=0.337	Asymptotic Distribution Critical Values ϕ_1=-0.5	Empirical Distribution Critical Values ϕ_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
85%	6.15	6.12	6.11	6.39
80%	5.48	5.52	5.44	5.74
75%	4.92	4.98	4.91	5.27
70%	4.49	4.33	4.43	4.73
65%	4.13	3.92	4.10	4.34
60%	3.81	3.54	3.80	3.93
55%	3.50	3.11	3.51	3.61
50%	3.26	2.83	3.25	3.28
45%	3.00	2.49	3.00	3.00
40%	2.78	2.29	2.76	2.73
35%	2.55	2.01	2.54	2.43
30%	2.33	1.74	2.33	2.09
25%	2.12	1.45	2.11	1.82
20%	1.89	1.20	1.92	1.53
15%	1.70	0.96	1.70	1.19
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 4
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

% Dist.	Critical Values ϕ_1=0.3	Critical Values ϕ_1=0.5	Critical Values ϕ_1=0.8	Critical Values ϕ_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
85%	6.20	6.10	6.12	6.17
80%	5.49	5.44	5.44	5.51
75%	4.97	4.94	4.92	4.97
70%	4.55	4.52	4.49	4.52
65%	4.17	4.16	4.10	4.12
60%	3.83	3.83	3.80	3.79
55%	3.54	3.54	3.51	3.49
50%	3.28	3.27	3.26	3.22
45%	3.04	3.01	3.02	2.99
40%	2.80	2.77	2.78	2.75
35%	2.56	2.55	2.56	2.53
30%	2.35	2.32	2.34	2.33
25%	2.15	2.11	2.12	2.12
20%	1.93	1.91	1.92	1.90
15%	1.70	1.70	1.70	1.68
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

TABLE 4 (Cont'd)
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

% Dist.	Critical Values ϕ_1=-0.3	Critical Values ϕ_1=-0.5	Critical Values ϕ_1=-0.8	Critical Values ϕ_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
85%	6.11	6.11	6.14	6.15
80%	5.43	5.44	5.51	5.50
75%	4.94	4.91	5.00	4.96
70%	4.52	4.43	4.55	4.54
65%	4.15	4.10	4.18	4.17
60%	3.82	3.80	3.85	3.84
55%	3.55	3.51	3.55	3.55
50%	3.28	3.25	3.28	3.28
45%	3.03	3.00	3.04	3.03
40%	2.80	2.76	2.81	2.79
35%	2.56	2.54	2.59	2.57
30%	2.35	2.33	2.37	2.35
25%	2.14	2.11	2.15	2.13
20%	1.93	1.92	1.94	1.91
15%	1.72	1.70	1.72	1.70
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

TABLE 5 **Size of the Sup LR and Exp LR tests** **in a Two-state Markov Switching Model with** **<u>First-Order Autoregressive and Homoskedastic Noise Function</u>**		
Nominal Size	1%	5%
Autoregressive Coefficient = 0.337		
Sup LR	0.97%	5.1%
Exp LR	6.1%	22.9%
Autoregressive Coefficient = -0.5		
Sup LR	1.6%	5.4%
Exp LR	7.1%	26%

TABLE 6

Maximum Likelihood Estimates - US Real GNP 1952:2-1984:4

Parameters	AR(4) Model	Markov Trend Model[1]
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

Note 1: 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.

TABLE 7
Distributions of the Likelihood
Ratio Statistic - Linear AR(4) against Markov Trend Model AR(4)[1]

	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
85%	6.16	21.38
80%	5.52	19.63
75%	4.99	18.30
70%	4.52	17.26
65%	4.17	16.42
60%	3.84	15.47
55%	3.55	14.38
50%	3.27	13.47
45%	3.03	12.78
40%	2.79	11.95
35%	2.56	11.08
30%	2.33	10.21
25%	2.12	9.18
20%	1.93	8.24
15%	1.72	7.52
10%	1.47	6.19
5%	1.18	4.44
1%	0.80	2.10

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

TABLE 8 **Comparison of Empirical and Asymptotic Critical Values** **of the Likelihood Ratio in a Two-state Markov Switching Model** **with <u>Uncorrelated and Heteroskedastic Noise Function</u>**		
% 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
85%	10.93	10.78
80%	10.02	9.99
75%	9.42	9.41
70%	8.84	8.86
65%	8.22	8.39
60%	7.80	7.97
55%	7.37	7.59
50%	6.84	7.22
45%	6.34	6.87
40%	5.98	6.55
35%	5.58	6.23
30%	5.22	5.90
25%	4.82	5.58
20%	4.37	5.25
15%	3.84	4.87
10%	3.25	4.43
5%	2.68	3.87
1%	1.74	3.05

TABLE 9
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%
Exp LR	25.7%	41.7%
Power - Sup LR	46%	60.9%

Note: The power of Sup LR test has been computed with respect to the model estimated by Turner, Startz, and Nelson (1989).

APPENDIX A

We reproduce below the sets of assumptions 1, 2, and 3 in Hansen (1991a).

Assumption 1

i) Θ and Γ are compact.

ii) $Q(\theta,\gamma)=\lim_{n\to\infty} EQ_n(\theta,\gamma)$ is continuous in (θ,γ) uniformly over $\Theta \times \Gamma$.

iii) $Q_n(\theta,\gamma) \to_p Q(\theta,\gamma)$ for all $(\theta,\gamma) \in \Theta\times\Gamma$.

iv) $Q_n(\theta,\gamma)-Q(\theta,\gamma)$ is stochastically equicontinuous in (θ,γ) over $\Theta\times\Gamma$.

v) For all $\gamma \in\Gamma,\ Q(\theta,\gamma)$ is uniquely maximized over $\theta\in\Theta$ at θ_0 .

Assumption 2

For $\theta\in\Theta_0=\{\theta\in\Theta:h(\theta)=0\},\ Q_n(\theta,\gamma)$ does not depend upon γ.

Assumption 3

i) $M(\theta,\gamma)=\lim_{n\to\infty} EM_n(\theta,\gamma)$ and $V(\theta,\gamma)=\lim_{n\to\infty} n\ E\ S_n^c(\theta,\gamma)S_n^c(\theta,\gamma)'$ are continuous in (θ,γ) uniformly over $\Theta_c\times\Gamma$, where θ_c is some neighborhood of θ_0 .

ii) $[M_n(\theta,\gamma),V_n(\theta,\gamma)] \to_p [M(\theta,\gamma),V(\theta,\gamma)]$ for all $(\theta,\gamma)\in\theta_c\times\Gamma$.

iii) $M_n(\theta,\gamma)-M(\theta,\gamma)$ and $V_n(\theta,\gamma)-V(\theta,\gamma)$ are stochastically equicontinuous in (θ,γ) over $\theta_c\times\Gamma$.

iv) $M(\gamma)=M(\theta_0,,\gamma)$ and $V(\gamma)=V(\theta_0,\gamma)$ are positive definite uniformly over $\gamma\in\Gamma$.

v) $\sqrt{n}S_n^c(\theta_0,\gamma)\Rightarrow S^c(\gamma)$ on $\gamma\in\Gamma$, where $S^c(.)$ is a mean zero Gaussian process with the

covariance function:

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

where $\Rightarrow$ denotes weak convergence of probability measures with respect to the uniform metric.

APPENDIX B

Derivation of the scores

Start with the following equality:

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

Therefore:

$$\begin{aligned}\frac{\partial p(y_n,\ldots,y_1,s_n(\gamma),\ldots,s_1(\gamma);\gamma,\theta)}{\partial\theta_l}&=\frac{\partial \log p(y_n,\ldots,y_1,s_n(\gamma),\ldots,s_1(\gamma);\gamma,\theta)}{\partial\theta_l}p(y_n,\ldots,y_1,s_n(\gamma),\ldots,s_1(\gamma);\gamma,\theta)\\&=\sum_{t=1}^{n}\frac{\partial \log p(y_t|\Psi_t,\gamma,\theta)}{\partial\theta_l}p(y_n,\ldots,y_1,\gamma,\theta)p(S_n(\gamma)=s_n(\gamma),\ldots,S_1(\gamma)=s_1(\gamma)|y_n,\ldots,y_1,\gamma,\theta)\end{aligned}$$

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 $p(y_n,\ldots,y_1,\gamma,\theta)$, we obtain:

$$\begin{aligned}\frac{\partial p(y_n,\ldots,y_1;\gamma,\theta)}{\partial\theta_l}\frac{1}{p(y_n,\ldots,y_1;\gamma,\theta)}&=\sum_{s_n(\gamma)=0}^{1}\cdots\sum_{s_{n-r}(\gamma)=0}^{1}\frac{\partial \log p(y_n|\Psi_n,\gamma,\theta)}{\partial\theta_l}p(S_n(\gamma)=s_n(\gamma),\ldots,S_n(\gamma)=s_{n-r}(\gamma)|y_n,\ldots,y_1,\gamma,\theta)\\&+\\&\vdots\\&+\sum_{s_1(\gamma)=0}^{1}\cdots\sum_{s_{-r+1}(\gamma)=0}^{1}\frac{\partial \log p(y_1|\Psi_1,\gamma,\theta)}{\partial\theta_l}p(S_1(\gamma)=s_1(\gamma),\ldots,S_{-r+1}(\gamma)=s_{-r+1}(\gamma)|y_n,\ldots,y_1,\gamma,\theta)\end{aligned}$$

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)).

APPENDIX C

Proof of Lemma 2

We will develop below the computation of each element of the covariance matrix of the scores $K(\gamma_1,\gamma_2)$, starting with the (α_0,α_0) element:

$$n\ E\left[S_n^c(\theta_0,\gamma_1)_{\alpha_0} S_n^c(\theta_0,\gamma_2)_{\alpha_0}\right] = n\ E \sum_{t=1}^{n} \sum_{s=1}^{n} \sum_{s_t(\gamma_1)=0}^{1} \cdots \sum_{s_{t-r}(\gamma_1)=0}^{1} \sum_{s_s(\gamma_2)=0}^{1} \cdots \sum_{s_{s-r}(\gamma_2)=0}^{1} \frac{1}{n^2}(1-\sum_{l=1}^{r}\phi_l)^2 \frac{\epsilon_t \epsilon_s}{\omega_0^4}\ p_t\ p_s$$

where:

$$p_t = p(S_t(\gamma_1)=s_t(\gamma_1),\ldots,S_{t-r}(\gamma_1)=s_{t-r}(\gamma_1)|y_n,\ldots,y_{-r+1};\theta_0,\gamma_1)$$

$$p_s = p(S_s(\gamma_2)=s_s(\gamma_2),\ldots,S_{s-r}(\gamma_2)=s_{s-r}(\gamma_2)|y_n,\ldots,y_{-r+1};s_s(\gamma_1);\theta_0,\gamma_1)$$

The conditioning of p_s on $s_s(\gamma_1)$ reflects the fact that the filters based on γ_1 and γ_2 are not independent since they are inferred from the same series $\{y_t\}$.

First note that the sums of the products of the probabilities is equal to 1. Also, by the serial independence assumption about the ϵ, we are left with:

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

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

$$nE[S_n^c(\theta_0,\gamma_1)_{\alpha_0} S_n^c(\theta_0,\gamma_2)_{\alpha_1}] = nE\sum_{t=1}^{n} \sum_{s=1}^{n} \sum_{s_t(\gamma_1)=0}^{1} \cdots \sum_{s_{t-r}(\gamma_1)=0}^{1} \sum_{s_s(\gamma_2)=0}^{1} \cdots \sum_{s_{s-r}(\gamma_2)=0}^{1} \frac{1}{n^2}(1-\sum_{l=1}^{r}\phi_l)(s_s(\gamma_2)-\sum_{l=1}^{r}\phi_l s_{s-l}(\gamma_2))\frac{\epsilon_t \epsilon_s}{\omega_0^4} p_t\ p_s$$

Since the conditional probabilities p_t sum to one, we are left with:

$$nE[S_n^c(\theta_0,\gamma_1)_{\alpha_0} S_n^c(\theta_0,\gamma_2)_{\alpha_1}] = nE\sum_{t=1}^{n} \sum_{s=1}^{n} \sum_{s_s=0}^{1} \cdots \sum_{s_{s-r}=0}^{1} \frac{\epsilon_t \epsilon_s}{n^2\omega_0^4}(1-\sum_{l=1}^{r}\phi_l)(s_s(\gamma_2)-\sum_{l=1}^{r}\phi_l s_{s-l}(\gamma_2)).p_s$$

This can be rewritten as:

$$nE[S_n^c(\theta_0,\gamma_1)_{\alpha_0} S_n^c(\theta_0,\gamma_2)_{\alpha_1}] = nE\ \{\sum_{t=1}^{n} \sum_{s=1}^{n} \frac{\epsilon_t \epsilon_s}{n^2\omega_0^4}(1-\sum_{l=1}^{r}\phi_l)E[s_s(\gamma_2)-\sum_{l=1}^{r}\phi_l s_{s-l}(\gamma_2)]|\Psi_n,\theta_0,\gamma_2)]\}$$

where we have used the independence assumption between ϵ and $s(\gamma)$. Next, we apply the law of iterated expectations $E\{E[s_s(\gamma_2)-\sum_{l=1}^{r}\phi_l s_{s-l}(\gamma_2)|\Psi_n,\theta_0,\gamma_2]\} = E[s_s(\gamma_2)-\sum_{l=1}^{r}\phi_l s_{s-l}(\gamma_2)]$ and note that

$E(s_s(\gamma_2)) = E(s_{s-1}(\gamma_2)) = \cdots = E(s_{j-r}(\gamma_2)) = \pi_2$. Therefore, by the serial independence assumption about ϵ, the final expression is given by:

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

Similarly for the (α_1,α_0) element:

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

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

$$nE[S_n^c(\theta_0,\gamma_1)_{\alpha_1} S_n^c(\theta_0,\gamma_2)_{\alpha_1}] = nE\ \{\sum_{t=1}^{n}\sum_{s=1}^{n}\frac{\epsilon_t\epsilon_s}{n^2\omega_0^4}E[(s_t(\gamma_1)-\sum_{l=1}^{r}\phi_l s_{t-l}(\gamma_1))(s_s(\gamma_2)-\sum_{l=1}^{r}\phi_l s_{s-l}(\gamma_2))\,|\,\Psi_n,\theta_0,\gamma_1,\gamma_2]\}$$

$$nE[S_n^c(\theta_0,\gamma_1)_{\alpha_1} S_n^c(\theta_0,\gamma_2)_{\alpha_1}] = nE\{\sum_{t=1}^{n}\sum_{s=1}^{n}\frac{\epsilon_t\epsilon_s}{n^2\omega_0^4}E[(s_t(\gamma_1)-\sum_{l=1}^{r}\phi_l s_{t-l}(\gamma_1))(s_s(\gamma_2)-\sum_{l=1}^{r}\phi_l s_{s-l}(\gamma_2))|\Psi_n,\theta_0,\gamma_1,\gamma_2]\}$$

$$= nE\{\sum_{t=1}^{n}\sum_{s=1}^{n}\frac{\epsilon_t\epsilon_s}{n^2\omega_0^4}[E(s_t(\gamma_1)s_s(\gamma_2)) + \phi_1^2 E(s_{t-1}(\gamma_1)s_{s-1}(\gamma_2)) + \cdots + \phi_r^2 E(s_{t-r}(\gamma_1)s_{s-r}(\gamma_2))$$

$$-\phi_1 E(s_t(\gamma_1)s_{s-1}(\gamma_2)) - \phi_1 E(s_{t-1}(\gamma_1)s_s(\gamma_2)) + \sum_{l=2}^{r}\phi_l\phi_{l-1}E(s_t(\gamma_1)s_{s-1}(\gamma_2)) + \sum_{l=2}^{r}\phi_l\phi_{l-1}E(s_{t-1}(\gamma_1)s_s(\gamma_2))$$

$$-\phi_2 E(s_t(\gamma_1)s_{s-2}(\gamma_2)) - \phi_2 E(s_{t-2}(\gamma_1)s_s(\gamma_2)) + \sum_{l=3}^{r}\phi_l\phi_{l-2}E(s_t(\gamma_1)s_{s-2}(\gamma_2)) + \sum_{l=3}^{r}\phi_l\phi_{l-2}E(s_{t-2}(\gamma_1)s_s(\gamma_2))$$

$$-\cdots-$$

$$-\phi_r E(s_t(\gamma_1)s_{s-r}(\gamma_2) - \phi_r E(s_{t-r}(\gamma_1)s_s(\gamma_2))]\}$$

Now, we state the following results for Markov variables:

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

where T^* corresponds to the transition probability matrix for the Markov variable having the highest probability limit ($\max(\pi_1,\pi_2)$); $T^{i*}[2,2]$ denotes the second row, second column element of the matrix T^* raised to the power i. Using the independence assumption of the ϵ, we finally obtain:

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

with A^* given by:

$$
\begin{aligned}
A^{*} = \; & 1+\phi_1^2+\phi_2^2+\cdots\phi_r^2 \\
& -2T^{*1}[2,2][\phi_1-\sum_{l=2}^{r}\phi_l\phi_{l-1}] \\
& -2T^{*2}[2,2][\phi_2-\sum_{l=3}^{r}\phi_l\phi_{l-2}] \\
& - \quad \cdots \quad - \\
& -2T^{*r}[2,2]\phi_r
\end{aligned}
$$

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 we will show below for ϕ_1:

$$
n\;E[S_n^c(\theta_0,\gamma_1)_{\alpha_0}S_n^c(\theta_0,\gamma_2)_{\phi_1}]=nE\sum_{t=1}^{n}\sum_{s=1}^{n}\sum_{s_t(\gamma_1)=0}^{1}\cdots\sum_{s_{t-r}(\gamma_1)=0}^{1}\sum_{s_s(\gamma_2)=0}^{1}\cdots\sum_{s_{s-r}(\gamma_2)=0}^{1}-(1-\sum_{l=1}^{r}\phi_l)\frac{\epsilon_t\epsilon_s}{n^2\omega_0^4}(\alpha_0-y_{s-1})\;p_t\;p_s
$$

Since all the conditional probabilities sum to one and the ϵ are serially independent, we finally obtain:

$$
\begin{aligned}
n\;E[S_n^c(\theta_0,\gamma_1)_{\alpha_0}S_n^c(\theta_0,\gamma_2)_{\phi_1}] &= \frac{1}{\omega_0^2}\left[-\alpha_0(1-\sum_{l=1}^{r}\phi_l)+(1-\sum_{l=1}^{r}\phi_l)\frac{\sum_{t=1}^{n}y_{t-1}}{n}\right] \\
&= \frac{1}{\omega_0^2}\left[-\alpha_0(1-\sum_{l=1}^{r}\phi_l)+(1-\sum_{l=1}^{r}\phi_l)\frac{y_0-y_n}{n}+(1-\sum_{l=1}^{r}\phi_l)\frac{\sum_{t=1}^{n}y_t}{n}\right]
\end{aligned}
$$

The limit as n tends to infinity of the average of the y is E(y), i.e. 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, we proceed similarly and arrive at:

$$
\begin{aligned}
n\;E[S_n^c(\theta_0,\gamma_1)_{\phi_i}S_n^c(\theta_0,\gamma_2)_{\phi_j}] &= n\;E\sum_{t=1}^{n}\frac{\epsilon_i^2}{n^2\omega_0^4}(y_{t-i}-\alpha_0)(y_{t-j}-\alpha_0) \\
&= \frac{1}{\omega_0^2}\sum_{t=1}^{n}\frac{(y_{t-i}-\alpha_0)(y_{t-j}-\alpha_0)}{n}
\end{aligned}
$$

As n tends to infinity, the sum goes to the corresponding element of the asymptotic autocovariance matrix of the $\{y_t\}$.

All the expectations of the cross-products of the scores with respect to ω_0^2 and ω_1^2 on one hand, and to α_0, α_1 and ϕ_i (i=1,...,r) on the other, 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. We derive first the expectation of the cross-product of the scores with respect to ω_0^2:

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

Note that the probabilities sum to 1 and that, by the serial independence assumption on the ϵ, the expectation of the cross-products of the square-bracketed terms are zero when t is different from s. We are therefore left with:

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

where the last equality follows from: $E(\epsilon_t^4)=3\omega_0^4$.

The $(\omega_0^2,\ \omega_1^2)$ element is given by:

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

Proceeding as before, we obtain:

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

$$=\frac{\pi_2}{2\omega_0^4}$$

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

APPENDIX D

Method for Simulating Chi-Square Processes

According to the definition in footnote 5, a chi-square process is the product of Gaussian vector processes which have a certain covariance matrix. We therefore propose below a general method to generate Gaussian vector processes with a given covariance matrix. Assume that we select a set of T values in the parameter space Γ to generate the distribution of $SupC$, say $\gamma_1, \gamma_2, \cdots, \gamma_T$.

Then the first step consists in drawing T vectors of i.i.d. $N(0,1)$ variates of dimension k, $\varepsilon(1), \varepsilon(2), \cdots, \varepsilon(T)$, i.e.:

$$E[\varepsilon_i(j)]=0 \qquad \forall\ i,j \qquad i=1,\cdots,k \quad j=1,\cdots,T$$
$$E[\varepsilon_i^2(j)]=1$$
$$E[\varepsilon_i(j)\varepsilon_l(m)]=0 \qquad i\neq l \qquad i,l=1,\cdots,k \quad j,m=1,\cdots,T$$

As a second step, construct T Gaussian vectors of dimension k, $G(\gamma_1), \cdots, G(\gamma_T)$, as follows:

$$G(\gamma_1)=A(1,1)\varepsilon(1)$$
$$G(\gamma_2)=A(2,1)\varepsilon(1)+A(2,2)\varepsilon(2)$$
$$G(\gamma_T)=A(T,1)\varepsilon(1)+A(T,2)\varepsilon(2)+\cdots+A(T,T)\varepsilon(T)$$

The $G(\gamma)$ vectors are Gaussian vectors and have by construction variance and covariance matrices which 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:

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

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.

2. a. 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. The last equality results from the orthogonality of $\varepsilon(1)$ and $\varepsilon(2)$ and by $E[\varepsilon(1)\varepsilon(1)']=I$, where I is the identity matrix of dimension k.

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

$$E[G(\gamma_2)G(\gamma_2)'] = K(\gamma_2,\gamma_2) = E\{[A(2,1)\varepsilon(1) + A(2,2)\varepsilon(2)].[A(2,1)\varepsilon(1) + A(2,2)\varepsilon(2)]'\}$$

$$= E[A(2,1)\varepsilon(1)\varepsilon(1)'A'(2,1) + A(2,2)\varepsilon(2)\varepsilon(2)'A'(2,2)]$$

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

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:

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

The $G(.)s$ so constructed 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 following 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 x 1) vector of i.i.d. N(0,1) variates. When k is large, this numerical approach might be the only way to generate the covariance matrix, but for the relatively simple models studied in this paper, we will derive analytically the elements of the P matrix.