ECONOMETRICA

JOURNAL OF THE ECONOMETRIC SOCIETY

An International Society for the Advancement of Economic Theory in its Relation to Statistics and Mathematics

http://www.econometricsociety.org/

Econometrica, Vol. 82, No. 2 (March, 2014), 765–784

OPTIMAL TEST FOR MARKOV SWITCHING PARAMETERS

MARINE CARRASCO
University of Montreal, Montréal, Quebec H3C3J7, Canada

LIANG HU
Wayne State University, Detroit, MI 48202, U.S.A.

WERNER PLOBERGER
Washington University in St. Louis, St. Louis, MO 63130, U.S.A.

Econometrica, Vol. 82, No. 2 (March, 2014), 765–784

NOTES AND COMMENTS

OPTIMAL TEST FOR MARKOV SWITCHING PARAMETERS

BY MARINE CARRASCO, LIANG HU, AND WERNER PLOBERGER[1]

This paper proposes a class of optimal tests for the constancy of parameters in random coefficients models. Our testing procedure covers the class of Hamilton's models, where the parameters vary according to an unobservable Markov chain, but also applies to nonlinear models where the random coefficients need not be Markov. We show that the contiguous alternatives converge to the null hypothesis at a rate that is slower than the standard rate. Therefore, standard approaches do not apply. We use Bartlett-type identities for the construction of the test statistics. This has several desirable properties. First, it only requires estimating the model under the null hypothesis where the parameters are constant. Second, the proposed test is asymptotically optimal in the sense that it maximizes a weighted power function. We derive the asymptotic distribution of our test under the null and local alternatives. Asymptotically valid bootstrap critical values are also proposed.

KEYWORDS: Information matrix test, optimal test, Markov switching model, Neyman–Pearson lemma, random coefficients model.

1. INTRODUCTION

IN THIS PAPER, WE FOCUS ON testing the constancy of parameters in dynamic models. The parameters are constant under the null hypothesis, whereas they are random and weakly dependent under the alternative. The model of interest is very general and includes as a special case the state space models and the Markov switching model initially introduced by Baum and Petrie (1966) and further studied by Hamilton (1989), where the regime changes in the parameters are driven by an unobservable two-state Markov chain. The model under the null need not be linear and could be a GARCH model, for instance. Moreover, the random coefficients need not be Markov.

Two distinct features make testing the stability of coefficients particularly challenging. The first is that the hyperparameters that enter in the dynamics of the random coefficients are not identified under the null hypothesis. As a result, the usual tests, like the likelihood ratio test, do not have a chi-squared distribution. The second feature is that the information matrix is singular under the null hypothesis. This is due to the fact that the underlying regimes are not observable. The first feature, known as the problem of nuisance parameters that are not identified under the null hypothesis, also arises when testing for structural change or threshold effects. It has been investigated in many papers, for example, Davies (1977, 1987), Andrews (1993), Andrews and

[1]An earlier version of this paper has circulated under the title "Optimal Test for Markov Switching." The authors are grateful to an editor and three referees for helpful comments. Carrasco gratefully acknowledges partial financial support from the National Science Foundation under Grant SES-0211418.

 DOI: 10.3982/ECTA8609

Ploberger (1994), Hansen (1996), among others. However, the second feature of our testing problem implies that the "right" (i.e., contiguous) local alternatives are of order $T^{-1/4}$ (where T is the sample size), while they are of the order $T^{-1/2}$ in the case of structural change and threshold models. The asymptotic local optimality discussed below shows that there do not exist tests with nontrivial power against local alternatives that converge faster than $T^{-1/4}$. Therefore, it is necessary to consider this rate of convergence when discussing power. Consequently, the results of Andrews and Ploberger (1994) do not apply here and we need to expand the likelihood to the fourth order to derive the properties of our test.

Our contribution is twofold. First, we propose a new test for parameter stability. This test is based on functionals of the first two derivatives of the likelihood evaluated under the null and the autocorrelations of the process describing the random parameters. It can be viewed as a time-series extension of White's (1982) information matrix test and shares some of its advantages. In particular, it requires the estimation of the model under the null hypothesis only. This feature is particularly desirable when bootstrapping critical values. Another advantage of our test is that it does not require a full specification of the dynamics of the random coefficients. We only need to know their covariance structure. It means that our test will have power against a wide variety of alternatives. The second contribution of our paper is to show that the proposed test is asymptotically locally optimal in the sense that there exists no test that is more powerful for a specific alternative that we are going to characterize. The proof consists in showing that, for fixed values of the nuisance parameters, our test is asymptotically locally equivalent to the likelihood ratio test. Then, the nuisance parameters are integrated out with respect to some prior distribution. We appeal to the Neyman–Pearson lemma to prove optimality.

There are few papers proposing tests for Markov switching. Garcia (1998) studied the asymptotic distribution of a sup-type likelihood ratio test. Hansen (1992) treated the likelihood as an empirical process indexed by all the parameters (those identified and those unidentified under the null). His test relied on taking the supremum of LR over the nuisance parameters. Both papers require estimating the model under the alternatives, which may be cumbersome. None investigates local powers. Gong and Mariano (1997) reparameterized their linear model in the frequency domain and constructed a test based on the differences in the spectrum between null and alternative. A Bayesian model selection procedure for Markov switching was proposed by Kim and Nelson (2001).

The connection between the information matrix test of White (1982) and the score test for independent mixture has been outlined by Chesher (1984). Lee and Chesher (1986) showed that, when the information matrix is singular, the likelihood ratio test may only have power against local alternatives of order $T^{-1/4}$ but remains optimal provided the true parameter vector under H_0 is an interior point of the parameter space. Recently, Cho and White (2007) proposed a likelihood ratio test for an independent and identically distributed (i.i.d.) mixture in dynamic models and showed that this test has power against Markov switching alternatives even though it ignores the temporal dependence of the Markov chain.

Published applications of our test include Hamilton (2005), Warne and Vredin (2006), Kahn and Rich (2007), Hu and Shin (2008), Morley and Piger (2012), and Dufrénot, Mignon, and Péguin-Feissolle (2011).

The outline of the paper is as follows. Section 2 describes the test statistic. Section 3 derives the null distribution of the test and a bootstrap method to compute its p-values. Section 4 establishes the optimality. Section 5 investigates the power of the test. In Section 6, we describe simulation results and investigate the nonlinearity in the GNP growth. Using U.S. data over the past 60 years, we found that while our test cannot reject the null hypothesis of a linear model against a Markov switching mean, it strongly rejects the same null against varying mean and variance. Finally, Section 7 concludes. Further examples and the proofs are collected in the Supplemental Material (Carrasco, Hu, and Ploberger (2014); CHP hereafter). Moreover, Appendix B of CHP defines the tensor notations used to derive the fourth order expansion of the likelihood. These notations are interesting in their own right, as they could be used in other econometric problems involving higher-order expansions.

2. ASSUMPTIONS AND TEST STATISTIC

The observations are denoted by $y_1, y_2, \dots, y_T$ (univariate or multivariate). Let $f_t(\cdot)$ denote the conditional density (with respect to some dominating measure) of y_t given $y_{t-1}, \dots, y_1$. $f_t(\cdot)$ is known up to a p-dimensional vector of parameters, say θ_t. We are interested in testing whether these parameters are constant over time. Namely, we test

$$H_0 : \theta_t = \theta_0, \quad \text{for some unspecified } \theta_0 \tag{2.1}$$

against

$$H_1 : \theta_t = \theta_0 + \eta_t, \tag{2.2}$$

where the switching variable η_t is not observable. We will construct asymptotically optimal tests for this testing problem under some assumptions on the structure of the η_t. So our basic probability spaces are the sets of all $(y_1, y_2, \dots, y_T, \eta_1, \eta_2, \dots, \eta_T)$.

ASSUMPTION 1: *The latent variable η_t is stationary and its distribution may depend on some unknown parameters β. They are nuisance parameters that are not identified under H_0. We assume that β belongs to a compact set B. Moreover, η_t is strongly exogenous in the sense that the joint likelihood of $(y_1, y_2, \dots, y_T, \eta_1, \eta_2, \dots, \eta_T)$ factorizes as $\prod_{t=1}^{T} f(y_t|\theta_t, y_{t-1}, y_{t-2}, \dots, y_1)q(\eta_t| \eta_{t-1}, \eta_{t-2}, \dots, \eta_1; \beta)$ and the values of θ_t belong to some compact subset of $\mathbb{R}^p$, Θ, containing θ_0.*

So we assume that even under the null there exists a distribution of the η_t. Nevertheless, under the null (2.1), this distribution does not play any role with regard to the distribution of the data $(y_T, y_{T-1}, y_{T-2}, \dots, y_1)$. Under H_0, the y_t and the η_t are mutually independent. Hence we can—when we fix θ_0—consider our null to be simple.

ASSUMPTION 2: *y_t is stationary under H_0 and the following conditions on the conditional log-density of y_t given $y_{t-1}, \dots, y_1$ (under H_0), l_t, are satisfied: $l_t = l_t(\theta)$, as a function of the parameter θ, is at least five times differentiable, and for $k = 1, \dots, 5$,*

$$E_{\theta_0} \sup_{\theta \in \mathcal{N}} \left(\left\| l_t^{(k)}(\theta) \right\|^{20} \right) < \infty,$$

where $l_t^{(k)}$ denotes the kth derivative of the log-likelihood with respect to the parameter θ and $\mathcal{N}$ is a neighborhood around θ_0. E_{θ_0} is the expectation with respect to the probability measure corresponding to the parameter θ_0 and $\| \cdot \|$ denotes the Frobenius norm. Moreover, θ_0 is an interior point of Θ and the information matrix $I(\theta_0) = E_{\theta_0}(l_t^{(1)}(\theta_0) l_t^{(1)'}(\theta_0))$ is nonsingular.

REMARK 1: Assumption 2 is maintained for mathematical convenience. It guarantees, in particular, the uniform convergence of the information matrix and the consistency and asymptotic normality of the maximum likelihood estimator (MLE) of θ under H_0. Suitable generalizations of the Kolmogorov–Cencov theorem (cf. Kunita (1990, Theorem 1.4.7, p. 38)) might yield criteria which only depend on expressions like $\sup_{\theta \in \mathcal{N}} E_{\theta_0}(\| l_t^{(k)}(\theta) \|^M)$. However, we think that the present version is general enough to cover most of the standard cases. In Section A.2 of CHP, we show that Assumption 2 is satisfied for ARCH and GARCH models with normal errors.

REMARK 2: As in Andrews and Ploberger (1994, Section 4.1), the vector of observable variables y_t may include exogenous variables. We do not impose restrictions on the moments of y_t. For instance, y_t could be a stationary IGARCH process. However, we rule out the case where y_t is a random walk. To deal with unit root, we would have to alter the test statistic by proper rescaling, and its asymptotic distribution would be different; see Hu (2011).

For technical reasons, we need to maintain some restrictions on the process η_t.

ASSUMPTION 3: *η_t is a function κ of a latent Markov process ϑ_t. We assume that ϑ_t is stationary and β-mixing with geometric decay. It implies in particular that there exist $0 < \lambda < 1$ and a measurable nonnegative function g such that*

$$\sup_{|h| \le 1} \left| E\left[h(\vartheta_{t+m}) | \vartheta_t\right] - E\left[h(\vartheta_t)\right] \right| \le \lambda^m g(\vartheta_t), \tag{2.3}$$

and $Eg(\vartheta_t) < \infty$. *Furthermore, we assume that* $E\eta_t = 0$ *and* $\max_t \|\eta_t\| \leq M < \infty$. *Moreover, the constant* λ *and the bound* M *are independent of* β *defined in Assumption 1 and* $E\|g\|$ *can be bounded by a constant independent of* β. *Finally,* $E(\eta_t \eta_{t-k})$, *for any integer* k, *is assumed to be continuous in* β.

The assumption $E\eta_t = 0$ is not restrictive, as the model can always be reparameterized to ensure this condition. ϑ_t β-mixing is satisfied by, for example, an irreducible and aperiodic Markov chain with finite state space. The property that ϑ_t is β-mixing with geometric decay will imply that η_t is geometric ergodic. $\max_t \|\eta_t\| \leq M < \infty$ will also be satisfied by any finite state space Markov chain; however, it will not be satisfied by an AR(1) process with normal error. This condition could be relaxed to allow for distributions of η_t with thin tails, but this extension is beyond the scope of the present paper. Although some form of mixing is necessary for the η_t, one should be able to relax condition (2.3).

Note that Assumption 3 does not require η_t itself to be Markov. Our assumption allows, for example, η_t to be a MA-process of the form

$$\eta_t = e_t + a_1 e_{t-1} + \cdots + a_p e_{t-p},$$

where the e_t are i.i.d. In this case, $\vartheta_t = (\eta_t, e_t, \ldots, e_{t-p})'$ is Markov. This example illustrates the potential autocorrelation functions of η_t. In fact, the set of autocorrelations satisfying our assumption approximates an arbitrary autocorrelation in any of the usual topologies.

The test statistic TS_T, for a given β, is of the form

$$\mathrm{TS}_T(\beta) = \mathrm{TS}_T(\beta, \hat{\theta}) = \Gamma_T - \frac{1}{2T}\hat{\varepsilon}(\beta)'\hat{\varepsilon}(\beta), \tag{2.4}$$

where $\Gamma_T = \frac{1}{\sqrt{T}}\sum_t \mu_{2,t}(\beta, \hat{\theta})$ with

$$\mu_{2,t}(\beta, \theta) = \frac{1}{2}\Big(\mathrm{tr}\big(\big(l_t^{(2)} + l_t^{(1)} l_t^{(1)\prime}\big)E^{\beta}\big(\eta_t \eta_t'\big)\big) + 2\sum_{s<t}\mathrm{tr}\big(l_t^{(1)} l_s^{(1)\prime} E^{\beta}\big(\eta_t \eta_s'\big)\big)\Big), \tag{2.5}$$

where tr denotes the trace and $\hat{\varepsilon}(\beta)$ is the residual from the OLS regression of $\mu_{2,t}(\beta, \hat{\theta})$ on $l_t^{(1)}(\hat{\theta})$, and $\hat{\theta}$ is the constrained maximum likelihood estimator of θ under H_0 (i.e., the ML estimator under the assumption of constant parameters). As β is unknown and cannot be estimated consistently under H_0, we use sup-type tests as in Davies (1987):

$$\mathrm{supTS} = \sup_{\beta \in \bar{B}} \mathrm{TS}_T(\beta),$$

or exponential-type tests as in Andrews and Ploberger (1994):

$$\text{(2.6)} \qquad \exp\text{TS} = \int_{\bar{B}} \exp\bigl(\text{TS}_T(\beta)\bigr)\, dJ(\beta),$$

where J is some prior distribution for β with support on $\bar{B}$, a compact subset of B. We will establish admissibility for a class of expTS statistics.

REMARKS: 1. In some applications, it may be of interest to test for the variability of one subset of parameters. To accommodate this case, it suffices to set the elements of η_t corresponding to the constant coefficients equal to zero. Then, Γ_T involves only the elements of the score and Hessian corresponding to the varying coefficients. However, in the computation of $\hat{\varepsilon}(\beta)$, one needs to project $\mu_{2,t}(\beta, \hat{\theta})$ on the whole vector $l_t^{(1)}(\hat{\theta})$ (including the score with respect to the constant coefficients).

2. The test statistic $\text{TS}_T(\beta)$ depends only on the score and derivative of the score under the null and on the estimator of θ under H_0. Therefore, it does not require estimating the model under the alternative. This is a great advantage when applying bootstrap to compute the critical values or p-values of the test.

3. The test relies on the second Bartlett identity (Bartlett (1953a, 1953b)) and is related to the information matrix test introduced by White (1982). Later, Chesher (1984) showed that the score test of the hypothesis that random coefficients have zero variance is equivalent to the information matrix test. Davidson and MacKinnon (1991) derived information-matrix-type tests for testing random parameters. Both papers assumed that the random parameters are independent, whereas we assume that the parameters are serially correlated and we fully exploit this correlation. As shown in Equation (2.5), our test statistics involve the second derivatives of the log-likelihood and the outer products of the scores as in the information matrix test, plus an extra term. This term actually arises from the serial dependence of the time-varying coefficients. Specifically, it captures the correlations between the conditional scores at different time periods.

4. The form of our test depends on the latent process η_t solely through its second-order properties.

5. We assume throughout the paper that the model under the null is correctly specified. The issue of misspecification is not addressed here.

3. NULL DISTRIBUTION AND IMPLEMENTATION

First of all, we establish the distribution of the test statistics under the null. For this purpose, define

$$\text{(3.1)} \qquad d(\beta) = d(\beta, \theta_0) = \bigl(I(\theta_0)\bigr)^{-1} \operatorname{cov}\bigl(\mu_{2,t}(\beta, \theta_0), l_t^{(1)}(\theta_0)\bigr),$$

where $I(\theta_0)$ denotes the information matrix of θ under H_0. Then, we have the following result.

THEOREM 3.1: *Under* H_0,

(i) Γ_T *converges to a Gaussian process*, $N(\beta)$, *with mean* 0 *and covariance* $k(\beta_1, \beta_2) = E_{\theta_0}((\mu_{2,t}(\beta_1, \theta_0) - d(\beta_1)' l_t^{(1)}(\theta_0))(\mu_{2,t}(\beta_2, \theta_0) - d(\beta_2)' l_t^{(1)}(\theta_0)))$.

(ii) $\frac{1}{T} \sum \hat{\varepsilon}_t^2(\beta)$ *converges to* $E_{\theta_0}((\mu_{2,t}(\beta, \theta_0) - d(\beta)' l_t^{(1)}(\theta_0))^2)$ *in probability uniformly in* β.

(iii) $\mathrm{TS}_T(\beta)$ *converges weakly to* $N(\beta) - \frac{1}{2}k(\beta, \beta)$ *and*

$$\mathrm{supTS} \xrightarrow{d} \sup_{\beta \in \bar{B}} \left\{ N(\beta) - \frac{1}{2}k(\beta, \beta) \right\},$$

$$\mathrm{expTS} \xrightarrow{d} \int_{\bar{B}} \exp\left(N(\beta) - \frac{1}{2}k(\beta, \beta) \right) dJ(\beta).$$

The asymptotic distribution of the test statistic might not be nuisance parameter free. We think that the advantage of being optimal outweighs this disadvantage. Modern computing technology makes it possible to get critical values by parametric bootstrapping (see Davidson and MacKinnon (2004) or Hall (1992)). In particular, we propose the following strategy to approximate the α critical values c_α and the p-value of expTS (the same strategy can be used for supTS):

1. Using the observed data, estimate θ_0 by $\hat{\theta}$, the maximum likelihood estimator under H_0. Compute the test statistic expTS.
2. Given $\hat{\theta}$, generate S independent samples $\{y_1^s, \dots, y_T^s\}_{s=1,\dots,S}$ under H_0 with $\theta = \hat{\theta}$.
3. Using each simulated sample $\{y_1^s, \dots, y_T^s\}$, estimate θ by $\hat{\theta}^s$ the MLE under H_0 and compute expTS with $\hat{\theta}^s$. The resulting statistics are denoted expTS^s, $s = 1, 2, \dots, S$.
4. Define $c_{\alpha,S}$ as the empirical critical values using expTS^s, $s = 1, 2, \dots, S$, that is, $c_{\alpha,S}$ is the $(1 - \alpha)$ quantile of expTS^s, $s = 1, 2, \dots, S$. Moreover, the bootstrap p-value is given by $\frac{1}{S}\sum_{s=1}^{S} I(\mathrm{expTS}^s > \mathrm{expTS})$.

Letting $S \to \infty$ will force the empirical critical values to converge to the critical values for the distribution of the test statistic under $P_{\hat{\theta}}$. We will prove that when T goes to infinity, these critical values converge to the true ones (the ones corresponding to the data generated by a P_{θ_0}, $\theta_0 \in H_0$). It should be noted that the consistency of the above bootstrap is also preserved under contiguous local alternatives. Contiguity guarantees that convergence under the null also holds true under the probability measure describing the local alternative.

THEOREM 3.2: *The bootstrap critical values* $c_{\alpha,S}$ *converge to* c_α *in probability as* T *and* S *go to infinity under* H_0 *and under local alternatives* $\theta_T = \theta_0 + h/\sqrt{T}$.

We now derive the expression of the test statistic in the following special case, which is general enough to cover most of the Markov switching models applied in economics.

EXAMPLE 3.3: Assume that η_t can be written as chS_t, where S_t is a scalar Markov chain with $V(S_t) = 1$, h is a vector specifying the direction of the alternative (for identification, h is normalized so that $\|h\| = 1$), and c is a scalar specifying the amplitude of the change. Moreover, assume that

$$\text{(3.2)} \qquad \operatorname{corr}(S_t, S_s) = \rho^{|t-s|}$$

for some $-1 < \rho < 1$. In such case, $\beta = (c^2, h', \rho)'$.

Example 3.3 implies that all the random coefficients jump at the same time. However, it does not impose that all coefficients should be random under the alternative. To deal with the case where a subset of coefficients remain constant under the alternative, it suffices to set the corresponding elements of h equal to zero. Assumption 3 imposes some restrictions on the Markov chain S_t. If S_t has a finite state space, it will be geometric ergodic provided its transition probability matrix satisfies some restrictions described, for example, in Cox and Miller (1965, p. 124). More precisely, if S_t is a two-state Markov chain, which takes the values a and b, and has transition probabilities $p = P(S_t = a|S_{t-1} = a)$ and $q = P(S_t = b|S_{t-1} = b)$, S_t is geometric ergodic if $0 < p < 1$ and $0 < q < 1$. Under this condition, S_t satisfies (3.2) with $\rho = p + q - 1$. S_t can also have a continuous state space as long as it is bounded. Consider an autoregressive model

$$S_t = \rho S_{t-1} + e_t,$$

where e_t is i.i.d. $U[-1, 1]$ and $-1 < \rho < 1$. Then S_t has bounded support $(-1/(1 - |\rho|), 1/(1 - |\rho|))$ and has mean zero. Moreover, it is easy to check that S_t is geometric ergodic using Theorem 3 on page 93 of Doukhan (1994); hence (3.2) is satisfied. For this choice of S_t, y_t follows a random coefficient model under the alternative.

In Example 3.3, $\mu_{2,t}(\beta, \theta)$ can be rewritten as

$$\text{(3.3)} \qquad \mu_{2,t}(\beta, \theta) = \frac{1}{2}c^2 h'\Bigg[\left(\frac{\partial^2 l_t}{\partial\theta\,\partial\theta'} + \left(\frac{\partial l_t}{\partial\theta}\right)\left(\frac{\partial l_t}{\partial\theta}\right)'\right) + 2\sum_{s<t}\rho^{(t-s)}\left(\frac{\partial l_t}{\partial\theta}\right)\left(\frac{\partial l_s}{\partial\theta}\right)'\Bigg]h,$$

and $\bar{B} = \{c^2, h, \rho : c^2 > 0, \|h\| = 1, \underline{\rho} < \rho < \bar{\rho}\}$ and $-1 < \underline{\rho} < \bar{\rho} < 1$.

The maximum of $\mathrm{TS}_T(\beta)$ with respect to c^2 can be computed analytically. Denote $\mu^*_{2,t}(\beta, \theta) = \mu_{2,t}(\beta, \theta)/c^2$, $\Gamma^*_T = \sum \mu^*_{2,t}(\beta, \theta)/\sqrt{T}$, and $\hat{\varepsilon}^*$ the residual of the regression of $\mu^*_{2,t}(\beta, \hat{\theta})$ on $l^{(1)}_t(\hat{\theta})$ so that Γ^*_T and $\hat{\varepsilon}^*$ do not depend on c^2. Then, we have

$$\text{(3.4)} \qquad \text{supTS} = \sup_{\{h,\rho:\|h\|=1,\underline{\rho}<\rho<\bar{\rho}\}} \frac{1}{2}\left(\max\left(0, \frac{\Gamma^*_T}{\sqrt{\hat{\varepsilon}^{*\prime}\hat{\varepsilon}^*}}\right)\right)^2.$$

Note that because $\sum_{t=1}^T l^{(1)}_t(\hat{\theta}) = 0$, the term Γ^*_T is equal to $\mathbf{1}'\hat{\varepsilon}^*$, where $\mathbf{1}$ is a $n \times 1$ vector of 1. Interestingly, the ratio $\Gamma^*_T/\sqrt{\hat{\varepsilon}^{*\prime}\hat{\varepsilon}^*}$ corresponds to the t-statistic for testing $H_0: \lambda = 0$ in the artificial regression

$$\mathbf{1} = \lambda\mu^*_{2,t}(\beta, \hat{\theta}) + \delta l^{(1)}_t(\hat{\theta}) + \text{residual}.$$

For an account on the role of artificial regressions in specification tests, see Davidson and MacKinnon (2004, Chapter 15). The evaluation of the expTS statistic is more involved since we need to pick some prior distributions for all nuisance parameters in $\beta = (c^2, h', \rho)'$. The most commonly used priors are uniform distributions. However, since c^2 is not bounded from above, a uniform prior is not appropriate. We define $G = c^2$ and take an exponential prior for G, which has the p.d.f. of $\tau e^{-\tau G}$. The reason for picking an exponential distribution is that we obtain an analytic form when taking the integral of $\text{TS}_T(\beta)$ with respect to the distribution of G as follows:

$$\text{(3.5)} \qquad \int_0^\infty \exp\bigl(\text{TS}_T(\beta)\bigr)\tau e^{-\tau G}\, dG$$
$$= \begin{cases} \dfrac{\sqrt{2\pi}\tau}{\sqrt{\hat{\varepsilon}^{*\prime}\hat{\varepsilon}^*}} \exp\left[\dfrac{(\Gamma^*_T - \tau)^2}{2\hat{\varepsilon}^{*\prime}\hat{\varepsilon}^*}\right]\Phi\left(\dfrac{\Gamma^*_T - \tau}{\sqrt{\hat{\varepsilon}^{*\prime}\hat{\varepsilon}^*}}\right), & \text{if } \hat{\varepsilon}^{*\prime}\hat{\varepsilon}^* \neq 0, \\ 1, & \text{otherwise,} \end{cases}$$
$$\equiv \Psi(h, \rho),$$

where $\Phi(\cdot)$ is the standard normal c.d.f. Then, taking uniform priors on h and ρ, we obtain

$$\text{(3.6)} \qquad \text{expTS} = \int_{\{\underline{\rho}\leq\rho\leq\bar{\rho},\|h\|<1\}} \Psi(h, \rho)\, d\rho\, dh.$$

Below, we give a class of models to which our estimation procedure applies. Other examples are discussed in Appendix A of CHP.

EXAMPLE 3.4—Markov-Switching Model: Hamilton (1989) proposed to model the change in the log of the real gross national product as a Markov-switching model of the form

$$y_t = \mu + \mu_1 S_t + u_t,$$
$$u_t = \phi_1 u_{t-1} + \cdots + \phi_r u_{t-r} + e_t,$$

where $e_t \sim$ i.i.d. $\mathcal{N}(0, \sigma^2)$ and S_t is a two-state Markov chain that takes the values 0 and 1 with transition probabilities p and q. This model has been used extensively to model the business cycle.

4. LOCAL ALTERNATIVES AND ASYMPTOTIC LOCAL OPTIMALITY

To establish the asymptotic local optimality of expTS, we use a weighted average power criterion similar to that used in Wald (1943) and Andrews and Ploberger (1994). Let J_T be some probability measure on the parameters β describing the dynamics of η_t. We want to show that expTS maximizes

$$\lim_{T\to\infty} \sup \int P\big([\varphi_T \text{ rejects}]|\beta, \theta\big)\, dJ_T(\beta, \theta), \tag{4.1}$$

over all tests φ_T of asymptotic level α. For this criterion to be useful, J_T should concentrate on parameters describing contiguous local alternatives. Otherwise, the power function asymptotically either equals α (if the alternative is too close to the null) or equals 1 (if the alternative is too far from the null). Then one can apply a suitable adaptation of the Neyman–Pearson lemma and conclude that the tests based on the likelihood ratio are optimal (in the sense of maximizing (4.1)).

The null hypothesis for a given θ is denoted as $H_0(\theta): \theta_t = \theta$ and the sequence of local alternatives is given by $H_{1T}(\theta): \theta_t = \theta + \frac{\eta_t}{\sqrt[4]{T}}$. The main difference with structural change and threshold testing is that here the local alternatives are of order $T^{-1/4}$. This is due to the fact that the regimes η_t are unknown and one needs to estimate them at each period. So in some sense, there is a curse of dimensionality where the number of parameters (including the probabilities of each regime) increases with the number of observations. It is also linked to the singularity of the information matrix under the null hypothesis. A local alternative of order $T^{-1/4}$ is also found in Lee and Chesher (1986), Gong and Mariano (1997), and Cho and White (2007).

Let Q_T^β denote the joint distribution of $(\eta_1, \dots, \eta_T)$, indexed by the unknown parameter β. Let $P_{\theta,\beta}$ be the probability measure on $y_1, y_2, \dots, y_T$ corresponding to $H_{1T}(\theta)$, and P_θ be the probability measure on $y_1, y_2, \dots, y_T$ corresponding to $H_0(\theta)$. The ratio of the densities under $H_0(\theta)$ and $H_{1T}(\theta)$ is given by

$$\ell_T^\beta(\theta) \equiv \frac{dP_{\theta,\beta}}{dP_\theta} = \int \prod_{t=1}^T f_t\big(\theta + \eta_t/T^{1/4}\big)\, dQ_T^\beta \Big/ \prod_{t=1}^T f_t(\theta).$$

Theorem 4.1 below shows that, for θ_0 known, a test based on

$$\exp\left(\frac{1}{\sqrt{T}} \sum_{t=1}^T \mu_{2,t}(\beta, \theta_0) - \frac{1}{2} E_{\theta_0}\big(\mu_{2,t}(\beta, \theta_0)^2\big)\right) \tag{4.2}$$

is asymptotically equivalent to the likelihood ratio test for testing $H_0(\theta_0)$ versus $H_{1T}(\theta_0)$ and therefore is admissible for these hypotheses.

THEOREM 4.1: *Under Assumptions 1–3, we have, under $H_0(\theta)$,*

$$\ell_T^\beta(\theta) \Big/ \exp\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mu_{2,t}(\beta,\theta) - \frac{1}{2}E_\theta\big(\mu_{2,t}(\beta,\theta)^2\big)\right) \xrightarrow{P} 1, \tag{4.3}$$

where the convergence in probability is uniform over β and $\theta \in \mathcal{N}$.

The proof of Theorem 4.1 is the main contribution of the paper and is given in Appendix C of CHP.

REMARK 3: Equation (4.3) demonstrates that, asymptotically, the density of the probability measure describing the local alternative depends only on $\mu_{2,t}(\beta,\theta)$. The only characteristic of η_t influencing these terms is its covariance. Therefore, we may conclude that different η_t which have the same covariance will have the same $\mu_{2,t}(\beta,\theta)$, hence generating the same probability measure. Moreover, the likelihood ratio between these two probability measures converges to 1. So any discrimination between them is impossible. For local alternatives (i.e., if the changes in the parameters are $O(1/\sqrt[4]{T})$), it is impossible to discriminate between, for example, an autoregressive process and a Markov switching process, when these processes have the same covariance. The Monte Carlo analysis in Section 6 investigates the two types of alternatives.

In practice, θ_0 is unknown and is therefore replaced by the MLE $\hat{\theta}$. The problem is that the test based on (4.2) is not robust to parameter uncertainty; indeed, the distribution of (4.2) with θ_0 replaced by $\hat{\theta}$ differs from the distribution of (4.2) with θ_0. As a consequence, we modify our test statistic by replacing $\mu_{2,t}(\beta,\hat{\theta})$ by $\hat{\varepsilon}_t(\beta,\hat{\theta})$, its projection on the space orthogonal to $l_t^{(1)}(\hat{\theta})$. The resulting test statistic is robust to parameter uncertainty. To see this, we apply the mean value theorem (β is omitted for convenience):

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\hat{\varepsilon}_t(\hat{\theta}) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\hat{\varepsilon}_t(\theta_0) + \frac{1}{T}\sum_{t=1}^{T}\frac{\partial\hat{\varepsilon}_t(\overline{\theta})}{\partial\theta}(\hat{\theta}-\theta_0),$$

where $\overline{\theta}$ is an intermediate value. Hence, under standard conditions, we have

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\partial\hat{\varepsilon}_t(\overline{\theta})}{\partial\theta} \to E_{\theta_0}\left(\frac{\partial\hat{\varepsilon}_t(\theta_0)}{\partial\theta}\right) = -\operatorname{cov}_{\theta_0}\big(\hat{\varepsilon}_t(\theta_0), l_t^{(1)}(\theta_0)\big) = 0.$$

The new test statistic is robust to parameter uncertainty but is no longer equivalent to the likelihood ratio of $H_0(\theta_0)$ against $H_{1T}(\theta_0)$. We show that it is equivalent to the likelihood ratio of $H_0(\theta_0)$ against

$$H_{1T}(\theta_T): \theta_t = \theta_0 + \frac{\eta_t}{\sqrt[4]{T}} - \frac{d}{\sqrt{T}},$$

where d is defined in (3.1). Hence, it is optimal for this alternative.

Let $P_{\theta_T,\beta}$ be the probability measure on $y_1, y_2, \ldots, y_T$ corresponding to $H_{1T}(\theta_T)$. The following result is useful in analyzing the asymptotic properties of our test statistics under $H_{1T}(\theta_T)$.

COROLLARY 4.2: *For any β, the sequence $P_{\theta_T,\beta}$ is contiguous with respect to P_{θ_0}.*

This result follows from the CLT satisfied by $\mu_{2,t}(\beta, \theta_0)$ and Le Cam's first lemma.

Let $\theta_T = \theta - d/\sqrt{T}$. Define the process

$$\begin{aligned} Z_T(\beta, \theta) = {} & \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \mu_{2,t}(\beta, \theta) - \frac{1}{2} E_\theta\big(\mu_{2,t}(\beta, \theta)^2\big) \\ & - \frac{1}{\sqrt{T}} \sum_{t=1}^{T} d' l_t^{(1)}(\theta) + \frac{1}{2} E_\theta\big(\big(d' l_t^{(1)}(\theta)\big)^2\big). \end{aligned}$$

The asymptotic local optimality of tests based on $\exp(Z_T(\beta, \hat{\theta}))$ is proved in two steps. In the first step, we show that $\exp(Z_T(\beta, \theta_T))$ is essentially the likelihood $\frac{dP_{\theta-d/\sqrt{T},\beta}}{dP_\theta}$, so tests based on mixtures of $\exp(Z_T(\beta, \theta_T))$ optimize a criterion like (4.1). Second, we show that $Z_T(\beta, \theta)$ is almost constant for variations in θ of the order of $O(1/\sqrt{T})$, so $Z_T(\beta, \theta) \approx Z_T(\beta, \hat{\theta})$, where $\hat{\theta}$ is the ML estimator. So, using the Bayesian mixtures of $\exp(Z_T(\beta, \hat{\theta}))$ as test statistics will yield the same results as using $\exp(Z_T(\beta, \theta_T))$, which we know to be asymptotically locally optimal. Then the admissibility of our test expTS will follow from the asymptotic equivalence between expTS and the test based on $Z_T(\beta, \hat{\theta})$. Before establishing the asymptotic local optimality of expTS, we need to define a notation. Let $P_{\hat{\theta}}$ denote the probability measure corresponding to the value of the maximum likelihood estimator. (We can understand our parametric family as a mapping, which assigns to every θ a measure P_θ. Then, the measure $P_{\hat{\theta}}$ results from an evaluation of this mapping at $\hat{\theta}$. It is a random measure.)

THEOREM 4.3: *Assume that J is a measure with mass 1 concentrated on a compact subset of B. Let d be as in (3.1) and define*

$$\mathrm{ST}(\theta) = \int \exp\big(Z_T\big(\beta, \theta - d(\beta, \theta)/\sqrt{T}\big)\big)\, dJ(\beta).$$

Then

$$\text{expTS} - \text{ST}(\theta_0) \to 0 \tag{4.4}$$

in probability under P_{θ_0}, *where* expTS *is defined in* (2.4), (2.5), *and* (2.6).

Let $K(\hat{\theta})$ *be real numbers so that* $P_{\hat{\theta}}([\text{expTS} < K(\hat{\theta})]) \le 1 - \alpha$, $P_{\hat{\theta}}([\text{expTS} > K(\hat{\theta})]) \le \alpha$ *and assume* $K(\hat{\theta}) \to K$. *Then the tests* φ_T, *which reject if* expTS $> K(\hat{\theta})$, *and accept if* expTS $< K(\hat{\theta})$, *are, for all* θ_0, *asymptotically equivalent under* P_{θ_0} *to tests rejecting if* $\text{ST}(\theta_0) > K$, *and accepting if* $\text{ST}(\theta_0) < K$. *Moreover, we have* $P_{\theta_0}([\text{ST}(\theta_0) < K]) \le 1 - \alpha$, *and* $P_{\theta_0}([\text{ST}(\theta_0) > K]) \le \alpha$. *Hence,* φ_T *is admissible in the sense of Definition* D.4 *given in CHP*.

Theorem 4.3 shows two important results: (a) the asymptotic equivalence between expTS and the likelihood ratio test, from which the admissibility of expTS can be inferred, and (b) the validity of the asymptotic critical values where $\hat{\theta}$ replaces the unknown θ_0. The admissibility of the sup test could be proved using an approach similar to that of Andrews and Ploberger (1995).

The restriction to prior measures with compact support might be a bit restrictive. In most cases, we should be able to approximate prior measures with noncompact support by ones with compact support.

5. DISTRIBUTION UNDER LOCAL ALTERNATIVES AND POWER

In this section, we first derive the asymptotic distribution of expTS under local alternatives and then discuss its power.

THEOREM 5.1: *Assume Assumptions* 1 *to* 3 *hold. Under* $H_{1T}(\theta_T)$, $\text{TS}_T(\beta, \hat{\theta})$ *converges to a Gaussian process with mean* $k(\beta, \beta_0) - \frac{1}{2}k(\beta, \beta)$ *and variance* $k(\beta_1, \beta_2)$, *where* β_0 *is the true value of the parameter* β *under the alternative*.

This result follows from Le Cam's third lemma and from the fact that the joint distribution of the $\text{TS}_T(\beta, \hat{\theta})$ and the logarithms of the densities of the local alternatives converges to a joint normal, and these two Gaussian random variables are correlated.

Theorems 4.1 and 4.3 imply that the likelihood ratio test for a given β (denoted $\text{LR}(\beta)$) is asymptotically equivalent to 2 times $\text{TS}_T(\beta, \hat{\theta})$. Hence, Theorem 5.1 provides the asymptotic distribution of $\text{LR}(\beta)$ divided by 2 under $H_{1T}(\theta_T)$.

From Theorem 5.1, we can conclude that our test has nontrivial power against local alternatives if $E_{\theta_0}((\mu_{2,t}(\beta, \theta_0) - d(\beta)' l_t^{(1)}(\theta_0))^2) > 0$. It is, however, also possible that

$$E_{\theta_0}\left(\left(\mu_{2,t}(\beta, \theta_0) - d(\beta)' l_t^{(1)}(\theta_0)\right)^2\right) = 0. \tag{5.1}$$

Some insight about this phenomenon can be gained by noticing that

$$\begin{aligned}&E_{\theta_0}\left[\left(\mu_{2,t}-d'l_t^{(1)}\right)^2\right]\\&\quad=E_{\theta_0}\left(\mu_{2,t}^2\right)-E_{\theta_0}\left(l_t^{(1)}\mu_{2,t}\right)'\left(E_{\theta_0}\left(l_t^{(1)}l_t^{(1)\prime}\right)\right)^{-1}E_{\theta_0}\left(l_t^{(1)}\mu_{2,t}\right)\end{aligned}$$

using (3.1). Hence (5.1) is satisfied if and only if $\mu_{2,t}$ belongs to the linear span of the components of $l_t^{(1)}$. This is unlikely to happen except in very special cases. Let us construct such an example. Assume for a moment that $\rho=0$ and all the other prerequisites of Assumptions 1–3 and Example 3.3 are fulfilled. Then $\mu_{2,t}$ is a linear functional of the second-order derivatives of the log-likelihood, namely, $h'(\frac{\partial^2 l_t}{\partial\theta\partial\theta'}+(\frac{\partial l_t}{\partial\theta})(\frac{\partial l_t}{\partial\theta})')h$. Then (5.1) means that the second-order derivatives can be written as a linear combination of the scores. A typical example where this happens is testing for independent mixture of two normals with different unknown means and same unknown variance. Then $h=(1,0)'$ and (5.1) is fulfilled (for all β because here β is simply c^2).

If (5.1) is fulfilled for all β, then it is impossible to construct a test with nontrivial power against these specific local alternatives. Indeed, the $\mathrm{TS}_T(\beta,\hat{\theta})$ are consistent approximations of the log-density between the measure under the null (corresponding to θ_0) and the measure under the alternative. If this density converges to 1, then any reasonable distance, for example, total variation, converges to zero. So in this kind of situation, null and alternative are not distinct probability measures, which makes it impossible to construct consistent tests. Any test will have trivial local power for an alternative of order $T^{-1/4}$. In Section D.4. of CHP, we discuss in more detail the case of Hamilton's model (Example 3.4) and we characterize the values of ρ for which our test (and any other test) will lack power.

The good news is that this phenomenon is the exception rather than the rule, as illustrated by the following theorem.

THEOREM 5.2: *Suppose Assumptions 1 to 3 hold. Consider Example 3.3. Assume furthermore that, for all t, s, $h'(\frac{\partial l_t}{\partial\theta})(\frac{\partial l_s}{\partial\theta})'h$ cannot be represented as a linear combination of components of $(\frac{\partial l_t}{\partial\theta})$. Then, for each h, there exist at most finitely many ρ so that (5.1) is fulfilled.*

Since, in expTS, we integrate over a range of values for h and ρ, expTS will have local power under the assumptions of Theorem 5.2.

6. MONTE CARLO STUDY AND APPLICATION

In this section, we investigate the power of supTS and expTS in two models: first, the simplest model, where the limiting distribution of our test statistics is pivotal, and second, the AR model used in Hamilton (1989), where critical values are computed via the bootstrap method outlined in Section 3.

For the expTS test, we use an exponential prior as suggested in (3.5) with $\tau = \sqrt{\hat{\varepsilon}^{*\prime}\hat{\varepsilon}^*}$. This yields the following simplification:

$$\Psi(h,\rho) = \begin{cases} \sqrt{2\pi}\exp\left[\frac{1}{2}\left(\frac{\Gamma_T^*}{\sqrt{\hat{\varepsilon}^{*\prime}\hat{\varepsilon}^*}} - 1\right)^2\right]\Phi\left(\frac{\Gamma_T^*}{\sqrt{\hat{\varepsilon}^{*\prime}\hat{\varepsilon}^*}} - 1\right), & \text{if } \hat{\varepsilon}^{*\prime}\hat{\varepsilon}^* \neq 0, \\ 1, & \text{otherwise.} \end{cases}$$

The tests supTS and expTS use a grid search for ρ over the interval $[-0.7, 0.7]$ with increment 0.01.

6.1. *Switching Mean Without Regressors*

We consider the simple model in Garcia (1998) with switching intercept and no regressors:

$$y_t = \mu + \mu_1 S_t + \varepsilon_t, \tag{6.1}$$

where $P(S_t = 1|S_{t-1} = 1) = p$ and $P(S_t = -1|S_{t-1} = -1) = q$, and $\varepsilon_t \sim$ i.i.d. $\mathcal{N}(0, \sigma^2)$. The null hypothesis is $H_0 : \mu_1 = 0$. In Appendix A of CHP, we show that the asymptotic distributions of supTS and expTS are pivotal, and we provide critical values for these tests and compare them to empirical critical values. We now investigate the size performance of our tests and three competing tests, namely, the tests proposed by Cho and White (2007), Hansen (1992), and Garcia (1998) for different sample sizes. We use 3,000 iterations. For supTS and expTS, we use the asymptotic critical values from Table A-I of CHP with $\rho \in [-0.7, 0.7]$. For the three competing tests, we report the size calculated from their proposed asymptotic critical values.[2] From Table I, we see that both supTS and expTS have very good sizes compared to competing tests, even when T is fairly small, for example, $T = 50$.

We now investigate the local power of the five tests. The data generating process is $y_t = \alpha/\sqrt[4]{T} S_t + u_t$, where $u_t \sim$ i.i.d. $\mathcal{N}(0, 1)$. The sample size is chosen to be 200 and the number of replications is 3,000. Two specifications for S_t are investigated: (a) S_t is a two-state Markov chain such that S_t takes the value 1 or -1 with unknown transition probabilities $P(S_t = 1|S_{t-1} = 1) = p$ and $P(S_t = -1|S_{t-1} = -1) = q$, (b) S_t is AR(1) with i.i.d. Uniform error. We select the AR process so that it has the same mean, variance, and covariance as the two-state Markov chain for comparison. Specifically, $p = q = 0.75$ corresponds to $S_t = 0.5S_{t-1} + \epsilon_t$, where $\epsilon_t \sim U[-3/2, 3/2]$; $p = q = 0.25$ corre-

[2]For Cho and White's test, the adopted critical values correspond to $\theta_1, \theta_2 \in [-2, 2]$. For Garcia's test, the critical values correspond to $\pi \in [0.15, 0.85]$. Hansen's asymptotic critical values use a Newey–West estimator for the variance with bandwidth 10, a grid search for α_1 over $[0.1, 2]$ with increment 0.1, and a grid search for p and q over $[0.15, 0.85]$ with increment 0.05.

TABLE I

EMPIRICAL SIZE; NUMBER OF REPLICATIONS: 3,000

T	Nominal Level (%)	1%	5%	10%
50	supTS	1.5	5.5	10.5
	expTS	1.4	5.2	9.2
	Cho & White	0.8	3.9	7.2
	Garcia	1.2	6.2	12.9
	Hansen	14.6	26.4	33.8
100	supTS	1.3	5.4	10.3
	expTS	1.0	5.4	9.8
	Cho & White	0.7	4.3	8.5
	Garcia	1.2	7.1	13.7
	Hansen	10.1	19.2	25.1
200	supTS	1.2	5.3	10.5
	expTS	1.2	5.1	10.3
	Cho & White	1.0	4.1	8.4
	Garcia	1.3	6.6	13.6
	Hansen	6.7	13.2	19.0
500	supTS	0.8	4.6	9.5
	expTS	1.0	4.7	9.2
	Cho & White	0.6	4.2	8.8
	Garcia	1.1	6.8	12.9
	Hansen	3.1	7.7	11.8

sponds to $S_t = -0.5S_{t-1} + \epsilon_t$, where $\epsilon_t \sim U[-3/2, 3/2]$; and $p = 0.9$, $q = 0.5$ corresponds to $S_t = 0.4 + 0.4S_{t-1} + \epsilon_t$ where $\epsilon_t \sim U[-\sqrt{7/5}, \sqrt{7/5}]$.[3]

The size-corrected powers[4] at 5% level for all five tests are reported in Table II.

As expected, the powers increase when the alternative is farther away from H_0. Given that our test statistics depend on the dynamic of η_t only through its autocovariance, their powers should be the same whether η_t is a two-state Markov chain or an AR(1). We observe in the simulations that indeed the powers are very similar. On the other hand, the powers of competing tests sometimes decline sharply when η_t is an AR(1). Both supTS and expTS have comparable power performance with Garcia's test and dominate Cho and White's and Hansen's tests throughout, except for the case of a Markov switching alternative with $(p, q) = (0.9, 0.5)$.

[3]When $S_t = 1$ or -1 with $(p, q) = (0.9, 0.5)$, the ergodic distribution is $(5/6, 1/6)$. Therefore, $E(S_t) = 2/3$ and $\mathrm{Var}(S_t) = 5/9$.

[4]For supTS, empirical critical values are 4.094, 2.508, and 1.856 at 1%, 5%, and 10% levels, respectively. For expTS, they are 4.512, 1.834, and 1.362. For Cho and White, they are 8.477, 5.244, and 4.011. For Garcia, they are 14.319, 10.383, and 8.739. For Hansen, they are 4.741, 3.611, and 3.086.

TABLE II

SIZE-CORRECTED POWERS; $T = 200$, NUMBER OF REPLICATIONS: 3,000

α	(p, q)	(0.75, 0.75)	AR	(0.25, 0.25)	AR	(0.9, 0.5)	AR
1	supTS	7.0	7.6	8.9	10.1	5.3	5.4
	expTS	7.3	8.2	9.5	10.5	5.6	5.7
	Cho & White	5.1	5.9	5.2	5.6	5.4	5.4
	Garcia	6.7	6.4	6.8	7.4	5.3	5.1
	Hansen	5.3	5.3	5.5	6.5	5.3	5.2
2	supTS	32.8	31.9	39.1	39.4	11.1	11.9
	expTS	34.7	34.0	41.6	41.8	11.3	12.4
	Cho & White	7.5	6.3	6.2	6.3	8.1	5.6
	Garcia	27.3	25.1	31.7	25.9	11.2	8.7
	Hansen	13.3	10.7	14.6	13.8	6.9	6.7
3	supTS	76.5	76.7	81.7	80.6	26.3	27.5
	expTS	78.1	78.3	83.8	81.8	28.6	29.2
	Cho & White	19.4	8.6	17.8	7.9	27.4	6.9
	Garcia	79.0	64.4	83.8	66.1	40.8	22.4
	Hansen	46.5	37.9	51.3	44.6	19.9	10.4
4	supTS	95.4	94.5	97.4	96.2	47.4	54.5
	expTS	96.2	94.8	97.7	96.5	49.5	56.1
	Cho & White	64.9	11.9	63.6	12.2	72.3	8.4
	Garcia	99.3	90.1	99.6	90.4	85.6	45.0
	Hansen	87.6	76.2	89.8	79.6	54.9	22.5

6.2. *Testing Fluctuations in U.S. GNP*

Many papers have documented changes in the dynamic of U.S. GNP growth over time. These changes have been modeled as a Markov switching mean by Hamilton (1989). However, the variance of the GNP growth has experienced a sharp decline in the 1980s (the so-called great moderation, as documented by Kim and Nelson (1999) and McConnell and Perez-Quiros (2000)) and a renewed increase with the recent financial crisis, suggesting that the error variance should also be allowed to fluctuate in the model. To the best of our knowledge, this hypothesis has never been formally tested. Our approach provides tractable tools to test it. First, we test $H_0 : \mu_1 = 0$ in Hamilton's model described in Example 3.4 with $r = 4$ using the expression of $\mu_{2,t}$ given in Section A2 of CHP. Second, we test $H_0 : \mu_1 = \sigma_1 = 0$ in the following model:

$$\begin{aligned} y_t &= \mu + \mu_1 S_t + u_t, \\ u_t &= \phi_1 u_{t-1} + \cdots + \phi_4 u_{t-4} + (\sigma + \sigma_1 S_t) e_t, \end{aligned} \tag{6.2}$$

where $e_t \sim$ i.i.d. $\mathcal{N}(0, 1)$ and S_t is a $\{0, 1\}$ Markov chain. The likelihood under H_0 is the same as in the traditional Hamilton's model. However, the construction of the test is different, namely, η_t takes a different form. Following

TABLE III

TEST RESULTS ON REAL GNP

		1952Q2 to 1984Q4		1952Q2 to 2010Q4	
		Test Statistic	p-value	Test Statistic	p-value
Switch in mean	supTS	0.08	0.84	1.11	0.25
	expTS	0.66	0.72	1.00	0.23
Switch in mean & variance	supTS	1.93	0.34	14.34	0.00
	expTS	1.12	0.17	230.61	0.00

Example 3.3, we decompose the nuisance parameter vector as $\beta = (c^2, h', \rho)'$ and apply (3.3), (3.4), and (3.5). The two elements in h that correspond to switching mean and variance are generated uniformly over the unit sphere 100 times; the other elements are all set to be 0. Critical values are bootstrapped using 3,000 iterations. We use Hamilton's original data set and an extended data set including observations from 1952Q2 to 2010Q4.

The test statistics and bootstrapped p-values are collected in Table III.

For Hamilton's data, our test cannot reject the null of constant mean, which is consistent with Hansen (1992) and Garcia (1998). The same result is obtained for the extended series. However, when testing for switching in both the mean and variance, our tests reject the null of parameter constancy, especially for the extended series. This is in line with Stock and Watson (2003), who argued that, over the past 30 years, the most striking change in the business cycle is the dramatic decline in the output volatility.

It is useful to investigate the power of the supTS and expTS tests for the extended series. We simulate the data according to Model (6.2) using parameter values calibrated on real data (see Table A-III of CHP). The power is excellent. Using 3,000 iterations, the powers at 5% of supTS and of expTS are 86.2 and 89.8, respectively.

7. CONCLUSION

This paper presents the first optimal test against Markov switching alternatives. This test applies to a wide range of models that are popular in macroeconomics and finance. It is simple to implement, as it requires only the estimation of the parameters under the null hypothesis of constant parameters, and bootstrap can be used to compute its critical values.

REFERENCES

ANDREWS, D. W. K. (1993): "Tests for Parameter Instability and Structural Change Point," *Econometrica*, 61, 821–856. [765]

ANDREWS, D. W. K., AND W. PLOBERGER (1994): "Optimal Tests When a Nuisance Parameter Is Present Only Under the Alternative," *Econometrica*, 62, 1383–1414. [765,766,768,770,774]

——— (1995): "Admissibility of the Likelihood Ratio Test When a Nuisance Parameter is Present Only Under the Alternative," *The Annals of Statistics*, 23, 1609–1629. [777]

BARTLETT, M. S. (1953a): "Approximate Confidence Intervals," *Biometrika*, 40, 12–19. [770]

——— (1953b): "Approximate Confidence Intervals, II. More Than One Parameter," *Biometrika*, 40, 306–317. [770]

BAUM, L., AND T. PETRIE (1966): "Statistical Inference for Probabilistic Functions of Finite State Markov Chains," *Annals of Mathematical Statistics*, 37, 1554–1563. [765]

CARRASCO, M., L. HU, AND W. PLOBERGER (2014): "Supplement to 'Optimal Test for Markov Switching Parameters'," *Econometrica Supplemental Material*, 82, http://www.econometricsociety.org/ecta/supmat/8609_proofs.pdf; http://www.econometricsociety.org/ecta/supmat/8609_data_and_programs.zip. [767]

CHESHER, A. (1984): "Testing for Neglected Heterogeneity," *Econometrica*, 52, 865–872. [766, 770]

CHO, J. S., AND H. WHITE (2007): "Testing for Regime Switching," *Econometrica*, 75, 1671–1720. [766,774,779]

COX, D. R., AND H. D. MILLER (1965): *The Theory of Stochastic Processes*. London: Methuen. [772]

DAVIDSON, R., AND J. MACKINNON (1991): "Une nouvelle forme du test de la matrice d'information," *Annales d'Economie et de Statistique*, 20/21, 171–192. [770]

——— (2004): *Econometric Theory and Methods*. New York: Oxford University Press. [771,773]

DAVIES, R. B. (1977): "Hypothesis Testing When a Nuisance Parameter Is Present Only Under the Alternatives," *Biometrika*, 64, 247–254. [765]

——— (1987): "Hypothesis Testing When a Nuisance Parameter Is Present Only Under the Alternatives," *Biometrika*, 74, 33–43. [765,769]

DOUKHAN, P. (1994): *Mixing: Properties and Example*. New York: Springer-Verlag. [772]

DUFRÉNOT, G., V. MIGNON, AND A. PÉGUIN-FEISSOLLE (2011): "The Effects of the Subprime Crisis on the Latin American Financial Markets: An Empirical Assessment," *Economic Modelling*, 28, 2342–2357. [767]

GARCIA, R. (1998): "Asymptotic Null Distribution of the Likelihood Ratio Test in Markov Switching Models," *International Economic Review*, 39, 763–788. [766,779,782]

GONG, F., AND R. S. MARIANO (1997): "Testing Under Non-Standard Conditions in Frequency Domain: With Applications to Markov Regime Switching Models of Exchange Rates and the Federal Funds Rate," Staff Report 23, Federal Reserve Bank of New York. [766,774]

HALL, P. (1992): *The Bootstrap and Edgeworth Expansion*. New York: Springer-Verlag. [771]

HAMILTON, J. D. (1989): "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57, 357–384. [765,773,778,781]

——— (2005): "What's Real About the Business Cycle?" *Federal Reserve Bank of St. Louis Review*, 87, 435–452. [767]

HANSEN, B. (1992): "The Likelihood Ratio Test Under Non-Standard Conditions: Testing the Markov Switching Model of GNP," *Journal of Applied Econometrics*, 7, 61–82. [766,779,782]

——— (1996): "Inference When a Nuisance Parameter Is Not Identified Under the Null Hypothesis," *Econometrica*, 64, 413–430. [766]

HU, L. (2011): "Optimal Test for Stochastic Unit Root With Markov Switching," Report, Wayne State University. [768]

HU, L., AND Y. SHIN (2008): "Optimal Testing for Markov Switching GARCH Models," *Studies in Nonlinear Dynamics and Econometrics*, 12 (3), Article 3. [767]

KAHN, J. A., AND R. W. RICH (2007): "Tracking the New Economy: Using Growth Theory to Detect Changes in Trend Productivity," *Journal of Monetary Economics*, 54, 1670–1701. [767]

KIM, C. J., AND C. NELSON (1999): "Has the U.S. Economy Become More Stable? A Bayesian Approach Based on a Markov-Switching Model of the Business Cycle," *The Review of Economics and Statistics*, 81, 608–616. [781]

——— (2001): "A Bayesian Approach to Testing for Markov Switching in Univariate and Dynamic Factor Models," *International Economic Review*, 42, 989–1013. [766]

KUNITA, H. (1990): *Stochastic Flows and Stochastic Differential Equations*. Cambridge: Cambridge University Press. [768]

LEE, L. F., AND A. CHESHER (1986): "Specification Testing When Score Test Statistics Are Identically Zero," *Journal of Econometrics*, 31, 121–149. [766,774]

MCCONNELL, M., AND G. PEREZ-QUIROS (2000): "Output Fluctuations in the United States: What Has Changed Since the Early 1980's?" *American Economic Review*, 90, 1464–1476. [781]

MORLEY, J., AND J. PIGER (2012): "The Asymmetric Business Cycle," *The Review of Economics and Statistics*, 94, 208–221. [767]

STOCK, J., AND M. WATSON (2003): "Has the Business Cycle Changed and Why?" in *NBER Macroeconomics Annual*. Chicago: University of Chicago, 159–218. [782]

WALD, A. (1943): "Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations Is Large," *Transactions of the American Mathematical Society*, 54, 426–482. [774]

WARNE, A., AND A. VREDIN (2006): "Unemployment and Inflation Regimes," *Studies in Nonlinear Dynamics and Econometrics*, 10 (2), 1–52. [767]

WHITE, H. (1982): "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50, 1–25. [766,770]

Département de Sciences Economiques, University of Montreal, CP 6128, succ Centre Ville, Montréal, Quebec H3C3J7, Canada; marine.carrasco@umontreal.ca,

Dept. of Economics, Wayne State University, 656 W. Kirby St., Detroit, MI 48202, U.S.A.; lianghu@wayne.edu,

and

Dept. of Economics, Washington University in St. Louis, Campus Box 1208, St. Louis, MO 63130, U.S.A.; werner@ploberger.com.

Manuscript received June, 2009; final revision received February, 2013.