SERVE BANK OF NEW YORK

Staff Reports

Testing Under Non-standard Conditions in Frequency Domain: With Applications to Markov Regime Switching Models of Exchange Rates and the Federal Funds Rate

Fangxiong Gong
Roberto S. Mariano

Number 23
April 1997

Testing Under Non-standard Conditions in Frequency Domain: With Applications to Markov Regime Switching Models of Exchange Rates and the Federal Funds Rate

Fangxiong Gong

Capital Markets Function
Federal Reserve Bank of New York
New York, NY 10045-0001
Tel: (212)720-6943
E-mail: frank.gong@frbny.sprint.com

Roberto S. Mariano

Department of Economics
University of Pennsylvania
Philadelphia, PA 19104

The views expressed in the paper are those of the authors and are not necessarily reflective of views at the Federal Reserve Bank of New York or the Federal Reserve System. The authors thank Michael Boldin, Valentina Corradi, Francis X. Diebold, Rene Garcia, Jinyong Hahn, James D. Hamilton, and anonymous referees for their helpful comments. Useful suggestions from seminar participants at Federal Reserve Bank of New York and at the University of Pennsylvania are gratefully acknowledged.

Testing Under Non-standard Conditions in Frequency Domain: With Applications to Markov Regime Switching Models of Exchange Rates and the Federal Funds Rate

Abstract

We propose two test statistics in the frequency domain and derive their exact asymptotic null distributions under the condition of unidentified nuisance parameters. The proposed methods are particularly applicable in unobserved components models. Also, it is shown that the tests have considerable power when applied to a class of Markov regime switching models. We show that, after transforming the Markov regime switching model into the frequency domain representation, we only have to face the issue of unidentified nuisance parameters in a nonlinear context. The singularity problem disappears.

Compared to Hansen's (1992,1996) LR-bound test of the same Markov regime switching model, our LM test performs better in terms of finite sample power, except in the special case of the Markov-switching model in which the model becomes a Normal mixture model. Our test needs only a one-dimensional grid search while Hansen's (1992,1996) test requires a three-dimensional grid search.

The LM test is applied to Markov regime switching models of exchange rates and the Federal Funds rate. We used the same exchange rates data in Engel and Hamilton (1990). The null of random walk is not rejected in the exchange rates model. The null is rejected for the Federal Funds rate in subsample periods 1955:1-1979:9 and 1982:10-1995:11.

JEL Codes: C12,C15,C22.

Key Words: Markov Regime Switching, Fourier Transformation, Periodogram, Power Spectrum, Asymptotics, Simulation, Bootstrap.

1 Introduction

In this paper, we propose two statistics for testing under the condition of unidentified nuisance parameters. The proposed methods are particularly applicable in unobserved components models. Also, it is shown that the tests have considerable power when applied to a class of Markov regime switching models.

Markov regime switching models pose two special econometric problems: unidentified nuisance parameters under the null and identical zero scores. Because of these irregularities, formal statistical testing of the null hypothesis of no switching against the alternative of Markov regime switching has not been commonly carried out. These irregularities render the conventional tests (likelihood ratio, Lagrange multiplier, and Wald) inapplicable because they no longer have standard χ^2 distributions. Because of these irregularities, formal statistical testing of the null hypothesis of no switching against the alternative of Markov regime switching has not been commonly carried out.

In the literature there are non-standard tests available for testing Markov regime switching models. Hansen (1992,1996) uses a likelihood ratio bound to test Hamilton's (1989) model of GNP. Garcia (1992) derives an asymptotic distribution for the likelihood ratio statistic by excluding the singularity points from the null hypothesis. In Garcia's (1992) framework, the null is not the "true null" where there is no Markov regime switching, but a point close to it. Thus, the asymptotic distribution is not the exact distribution but an approximation which is found to be close to the empirical distribution.

The difficulty associated with testing Markov switching models is not so much the problem of unidentified parameters under the null. The difficulty is that, the information matrix is generically singular under the null for all values of the transition probabilities.

In this paper, we show that, after transforming the model into a frequency domain representation, we only have to face the issue of unidentified nuisance parameters in a nonlinear context. The intuition is somewhat similar to the suggestion by Lee and Chesher (1986), who propose examining higher-order derivatives at the null in the case of identical zero scores. Here we go to higher moments: the spectrum is the summation of the second moments, and our tests are based on the derivatives of the spectrum function. While the second-moment based methods allow us to bypass the singularity problem, we can not apply the paper's approach to test for a special case of Markov-switching model in which the model becomes a Normal mixture model. Likelihood-based approaches such as Hansen (1992,1996) do not have this problem.

Two test statistics – a Difference Test (DT), which is analogous to the LR test, and a Lagrange multiplier (LM)-type test – are proposed here and their exact asymptotic distributions are derived. The tests can be carried out either in a nonlinear regression framework or in Whittle's maximum likelihood framework. Under a set of regularity assumptions, the two tests are shown to be asymptotically equivalent.

The asymptotic distribution can be obtained by simulating a Gaussian process with a model-specific covariance function. Since the DT test requires estimating parameters both under the null and the alternative, the LM test has an advantage in terms of simulating and bootstrapping finite sample empirical distributions. We find that the simulated and bootstrapped finite sample approximate distribution is very close to the simulated asymptotic distribution.

The LM test is applied to Markov regime switching models of exchange rates and the Federal Funds rate. We used the same exchange rates data in Engel and Hamilton (1990). The null of random walk is not rejected in the exchange rates model for Engel and Hamilton (1990)'s sample period 73:IV-88:I. In an expanded sample 73:IV-96:I, the null cannot be rejected at the 1% level for all the three currencies. At the 5% level of the asymptotic distribution, the null can be rejected for the French Franc and the German Mark, but this rejection cannot hold at the 5% level of the bootstrapped finite sample distribution. For the Federal Funds rate model, the null of random walk is rejected at the 1% level in subsample periods 55:1-79:9 and 82:10-95:11.

Compared to Hansen's (1992,1996) LR bound test, the LM test performs better in terms of finite sample power.[1] However, at the 10 percent and 5 percent nominal sizes, our tests have higher rejection rates than Hansen's (1992,1996) test. At the 20 percent nominal size, the rejection rate of our test is lower than that of Hansen's (1992,1996) test. In terms of computional burden, our test is easier to implement since the test needs only a one dimensional grid search while Hansen's (1992,1996) test requires a three-dimensional grid search.

To motivate and show our work in the frequency domain, we first review the existing literature on testing when nuisance parameters are not identified under the null and the problems associated with testing Markov switching models. In section 3 we discuss how the nonstandard problem can be transformed into a frequency domain testing problem with only unidentified nuisance parameters under the null. Section 4 discusses the Difference Test (DT) and the LM test and their asymptotics in a general setting. The implementation and applications of the tests are discussed in sections 5 and 6.

[1]Except for a special case of the Markov-switching model in which the model becomes a Normal mixture model.

2 A Short Review of the Literature

The problem of unidentified nuisance parameters appears in many applications (see Hansen (1991)). For conventional tests to have a standard χ^2 distribution, the scores are required to have zero expected values and positive variance, and the likelihood function must be locally quadratic, so that the information matrix is locally constant and positive semi-definite. After a Taylor expansion of the score function around the true parameter values (evaluated at the maximum likelihood estimates), the Central Limit Theorem can be applied to conclude that the maximum likelihood estimates have an asymptotic normal distribution. Furthermore, a Taylor expansion of the likelihood function around the unrestricted estimates and evaluation at the restricted estimates lead to the conclusion that the standard likelihood ratio has a χ^2 distribution.

Davies (1977, 1987) is one of the first to investigate the asymptotic theory for testing in the context of unidentified nuisance parameters. His test statistic concentrates on the score function, which is a function of an unidentified parameter; in the limit, the statistic converges to a Gaussian random process.

The weakness of Davies' (1977,1987) test is that he does not derive the exact asymptotic distribution of the supremum test statistic. Using the bound of the statistic in actual testing may have low power depending on how sharp the bound is. The weak asymptotic optimality when sample size $T \to \infty$ and the *size* $\to 0$ does not give us any hint about the finite sample performance.

Following Davies' (1977,1987) work, there has been some recent development in testing under unidentified nuisance parameters. Hansen (1991) investigates the problem in testing nonlinear factors in linear regression models. The test statistic he employs converges asymptotically to a function of a chi-square process. Thus, Monte

Carlo simulation can be used to approximate the distribution of the statistic. Note that the distribution of the statistic depends on the covariance function of the chi-square process, as well as the domain of unidentified parameters and the functional form of the test statistic. Thus, the distribution is data- and model-dependent, which makes general tabulation impossible. Nevertheless Hansen's result is quite general, because the functional form of the test statistic is quite general, subject only to some regularity restrictions.

There has been some empirical work which applies the above stated theorems to test Hamilton's Markov switching model. Boldin (1990) uses Davies's (1987) upper bound test to determine the number of regimes; Garcia and Perron (1991) apply Gallant's (1977) test and a J-test of Davidson and MacKinnon (1981) for non-nested models, and use Davies' (1977, 1987) test to determine the number of regimes.

However, as Hansen (1992) points out, the asymptotic theories and testing procedures developed by Davies (1977, 1987) and Hansen (1991), which account for unidentified nuisance parameters, do not allow for identical zero scores and singular information matrix. Thus, these theorems cannot be applied to test Markov regime switching models. Andrews and Ploberger (1992) also point out that their optimal test procedure could not cover the problem of testing regime switching models with unobserved states, since the singularity of the information matrix under the null violates one of the regularity assumptions (Assumption 1 (f) in their paper).

Recognizing the difficulties associated with the first- and second-order derivatives of the likelihood function under the null, Hansen (1992, 1996) focuses on the likelihood surface and treats the likelihood as a function of unknown parameters. He then applies empirical process theory to bound the asymptotic distribution of the standard likelihood ratio statistic. He finds that the distribution of the likelihood ratio can be bounded by the distribution of the supremum of a Gaussian process. Like

Davies' (1977, 1987) tests, Hansen's (1992,1996) test could have power problems. Non-rejection of the null may be due to the use of a bound as the critical value. Also, Hansen's (1992,1996) test is computationally intensive. For Markov regime switching models, the test requires a three-dimensional grid search.

Garcia (1992) derived an asymptotic distribution for a likelihood ratio statistic by restricting the Markov transition probabilities away from zero and one. However, the singularity problem still exists in his test. His test is valid only if the null is not the "true null" where there is no Markov regime switching, but a point close to it. Thus, the asymptotic distribution is not the exact distribution but an approximation which is found to be close to the empirical distribution.

To see this point, let us look at a simple Markov regime switching model:

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

where $\{y_t\}$ is a stationary process and ϵ_t is i.i.d. $N(0, \sigma^2)$. As in Hamilton (1996), we assume that the state S_t follows a Markov chain that is independent of lagged y.

Let $\theta = \{\alpha_0, \alpha_1, \sigma^2\}$ and $\gamma = \{p, q\}$. Once the transition probabilities p, q are restricted away from zero, the only way to test the null of no regime switching against the alternative of Markov regime switching is to set the null to be $\alpha_1 = 0$. The problem is, even if p and q are restricted away from zero and one, under the null of $\alpha_1 = 0$, the information matrix will still be generically singular. To see this, we note that the scores can be written as (for a proof, see Equation A.5 and Appendix A of Hamilton 1996):

$$
S_\theta^T(\gamma) = \frac{\partial \log L(Y_T \mid \theta, \gamma)}{\partial \theta}
$$

$$= \sum_{t=1}^{T}\sum_{s_t=0}^{1} \frac{\partial \log P(y_t \mid s_t; \theta, \gamma)}{\partial \theta} P(s_t \mid Y_T; \theta, \gamma), \tag{2}$$

where $Y_T = \{y_1, y_2, \ldots, y_T\}$. For model (1), the scores are:

$$\begin{aligned}
S_{\alpha_0}^T &= \frac{\partial \log f(y_T, \ldots, y_1; \theta_0, \alpha)}{\partial \alpha_0} \\
&= \sum_{t=1}^{T}\sum_{s_t=0}^{1} \frac{\epsilon_t}{\sigma^2} p(s_t \mid y_T, \ldots, y_1; \theta_0, \gamma) \\
S_{\alpha_1}^T &= \frac{\partial \log f(y_T, \ldots, y_1; \theta_0, \alpha)}{\partial \alpha_1} \\
&= \sum_{t=1}^{T}\sum_{s_t=0}^{1} \frac{\epsilon_t(s_t)}{\sigma^2} s_t p(s_t \mid y_T, \ldots, y_1; \theta_0, \gamma) \qquad (3) \\
S_{\sigma^2}^T &= \frac{\partial \log f(y_T, \ldots, y_1; \theta_0, \alpha)}{\partial \sigma^2} \\
&= \sum_{t=1}^{T}\sum_{s_t=0}^{1} \frac{\epsilon_t^2 - \sigma^2}{2\sigma^4} p(s_t \mid y_T, \ldots, y_1; \theta_0, \gamma).
\end{aligned}$$

We see that, even if we restrict p and q away from zero or one, under the null of $\alpha_1 = 0$, $p(s_t = 1 \mid y_T, \ldots, y_1; \theta_0, \gamma) = constant$, $S_{\alpha_0}^T$ and $S_{\alpha_1}^T$ are proportional to each other and the information matrix is singular. Garcia's asymptotic distribution will hold only if the null is not the *true null* $\alpha_1 = 0$. However, if the null is a point where $\alpha_1 \neq 0$, all the parameters in the model are identifiable under the null and we would have a standard testing problem; i.e., all the conventional test statistics should work.

So far there is still no exact asymptotic theory regarding the distribution of the likelihood ratio test statistic for Markov regime switching models. In the following sections, we show that, after transforming the model into a frequency domain representation, the singularity issue is avoided and only the problem of unidentified nuisance parameters in a nonlinear context remains to be addressed.

3 The Frequency Domain Problem

3.1 A Class of Markov Regime Switching Models

Let us consider the following type of a Markov regime switching model:

$$\Phi(L)(y_t - \alpha_0 - \alpha_1 s_t) = \epsilon_t$$

or

$$y_t = \alpha_0 + \alpha_1 s_t + \Phi^{-1}(L)\epsilon_t \tag{4}$$

where

$$\begin{aligned} P(s_t = 1 \mid s_{t-1} = 1) &= p \\ P(s_t = 0 \mid s_{t-1} = 0) &= q \end{aligned} \tag{5}$$

ϵ_t is white noise $(0, \sigma^2)$, and $\Phi(L)$ is the polynomial lag operator

$$\Phi(L) = 1 - \phi_1 L - \ldots - \phi_k L^k \tag{6}$$

Note that $E(s_t \mid s_{t-1} = 1) = p$ and $E(s_t \mid s_{t-1} = 0) = 1 - q$, and one can write the AR(1) representation of s_t as follows,

$$\begin{aligned} s_t &= 1 - q + (p + q - 1)s_{t-1} + \mu_t \\ &= \gamma_0 + \gamma s_{t-1} + \mu_t, \end{aligned} \tag{7}$$

where $\gamma = p + q - 1$. Given the transition process of s_t, the distribution of μ_t is the following,

$$\begin{aligned} if \quad s_{t-1} &= 1, \\ \mu_t &= \begin{cases} 1 - p & \text{with probability p} \\ -p & \text{with probability 1-p} \end{cases} \end{aligned}$$

$$\begin{aligned} if \quad s_{t-1} &= 0, \\ \mu_t &= \begin{cases} q & \text{with probability 1-q} \\ q-1 & \text{with probability q} \end{cases} \end{aligned} \tag{8}$$

We see that $E(\mu_t \mid s_{t-1}) = 0$ and $E\mu_t = 0$. It is also easy to show that μ_t is uncorrelated with $s_{t-\tau}$ for all $\tau \geq 1$. Thus μ_t is serially uncorrelated white noise with mean zero. The conditional variance is:

$$\begin{aligned} E(\mu_t^2 \mid s_{t-1} = 1) &= p(1-p) \\ E(\mu_t^2 \mid s_{t-1} = 0) &= q(1-q) \end{aligned}$$

Thus, the variance of μ_t is:

$$var(\mu_t) = E(\mu_t^2) = p(1-p)\pi + q(1-q)(1-\pi), \tag{9}$$

where π is the unconditional probability of state "1"; $\pi = (1-q)/(2-p-q)$. With this, we can write the spectrum of y_t as,

$$\begin{aligned} F_y(w) &= \frac{\alpha_1^2 var(\mu_t)}{(1-\gamma exp(-iw))(1-\gamma exp(iw))} + \mid \Phi^{-1}(exp(-iw)) \mid^2 var(\epsilon_t) \\ &= \frac{\alpha_1^2(p(1-p)\pi + q(1-q)(1-\pi))}{1+\gamma^2 - 2\gamma cos(w)} + \mid \Phi^{-1}(exp(-iw)) \mid^2 \sigma^2 \end{aligned} \tag{10}$$

Note that for the Markov chain s_t, the expected duration in state $s_t = 1$ is $\frac{1}{1-p}$, and the expected duration in state $s_t = 0$ is $\frac{1}{1-q}$; thus the expected period of each cycle is:

$$\lambda = \frac{1}{1-p} + \frac{1}{1-q} = \frac{2-p-q}{(1-p)(1-q)} \tag{11}$$

and we can write,

$$\begin{aligned} F_y(w) &= \frac{\alpha_1^2(p+q)(1-p)(1-q)}{(1+\gamma^2 - 2\gamma cos(w))(2-p-q)} + \mid \Phi^{-1}(exp(-iw)) \mid^2 \sigma^2 \\ &= \frac{\alpha_1^2}{\lambda} \frac{\gamma+1}{1+\gamma^2 - 2\gamma cos(w)} + \mid \Phi^{-1}(exp(-iw)) \mid^2 \sigma^2 \\ &\equiv \delta g(w,\gamma) + f(w,\theta) \end{aligned} \tag{12}$$

3.2 Unobserved Factor Models

The Markov regime switching model outlined in section 3.1 is a special type of unobserved-factor models. A example of linear unobserved factor models is the following:

$$\begin{aligned} y_t &= \alpha_0 + \alpha_1 z_t + \phi(L)\epsilon_t \\ z_t &= \psi(L)v_t, \end{aligned}$$

where z_t is the latent factor and $\phi(L)$, $\psi(L)$ are polynomial lag operator functions. As in general state-space models, the shocks to the factor, v_t, is assumed to be uncorrelated with the "measurement error" ϵ_t, at all leads and lags.

An interesting testing problem would be: whether there is such an unobserved factor z_t that significantly affects the variation of y_t.

In this case, the spectrum of y:

$$\begin{aligned} F_y(w) &= \alpha_1^2|\psi(e^{-iw})|^2\sigma_v^2 + |\phi(e^{-iw})|^2\sigma_\epsilon^2 \\ &\equiv \delta g(w,\gamma) + f(w,\theta), \end{aligned}$$

where $\delta \equiv \alpha^2$, $\gamma = (\psi, \sigma_v)$, $\theta = (\phi, \sigma_\epsilon)$. We see that, under the null $\delta = 0$, γ is not identified. The functional form of the spectrum has the same structure as that of Markov regime switching models. Statistical testing for models of this type can be addressed in the frequency domain as shown in the following sections.

3.3 The Testing Problem in the Frequency Domain

Suppose we have T observations $\{y_t\}, t = 1, 2, \ldots, T$, and the periodogram of y_t is $I_y(w_j)$, where $w_j = \frac{2\pi j}{T}$, $j = 0, 1, \ldots, [T/2]$.

Define $N \equiv [T/2] = T/2$ if T is even, $N \equiv [T/2] = (T-1)/2$ if T is odd. Let

$$\xi_j \equiv I_y(w_j)/F_y(w_j) \tag{13}$$

If we assume $\epsilon_t \sim N(0, \sigma^2)$, under the null, $\xi_j \sim \chi_2^2/2$ for $w_j \neq 0, \pi$; and $\xi_j \sim \chi_1^2$ for $w_j = 0, \pi$ (see Harvey, 1991). Thus ξ_j is independent and identically distributed except for $j = 0$ and $j = T/2$ when T is even.

Even if we do not assume a Gaussian process for y_t, as long as y_t is a stationary process, the distributional results for ξ_j hold asymptotically (Brillinger, 1981; Brockwell and Davis, 1995; Harvey, 1991). Thus our asymptotic theorems developed in the next section always hold for a general stationary process y_t under suitable conditions outlined in Brillinger (1981). Franke and Hardle (1992) use the fact that ξ_j is i.i.d in large samples and the relation (13) to address the sampling uncertainty of spectrum estimates by a bootstrap procedure of resampling from $I_y(w_j)$.

For the Markov regime switching model, the testing problem can be addressed in the following nonlinear regression framework:

$$\begin{aligned} z_j &\equiv \log(I_y(w_j)) = \log(F_y(w_j)) + E\log(\xi_j) + (\log(\xi_j) - E\log(\xi_j)) \\ &= c + log\{\frac{\alpha_1^2}{\lambda}\frac{\gamma+1}{1+\gamma^2-2\gamma cos(w_j)} + \mid \Phi^{-1}(exp(-iw_j)) \mid^2 \sigma^2\} + u_j, \end{aligned} \tag{14}$$

where u_j is a zero-mean i.i.d. process.

Now the test we want to conduct is:

$$\begin{aligned} H_0: \quad \alpha_1 &= 0 \\ H_1: \quad \alpha_1 &\neq 0 \end{aligned} \tag{15}$$

Note that, after the transformation into a frequency domain problem, the test can be easily carried out in a nonlinear regression context. It is still a nonstandard problem in the sense that, under the null, we have unidentified nuisance parameters γ and λ.

It is interesting that, from the above specification, we can see when the test would have low power against the Markov switching alternative in finite sample: If in the DGP the α_1 is very small or the expected period λ of the Markov chain is very long compared to the sample length, then $\frac{\alpha_1^2}{\lambda}$ would be very small and the power would be poor. When the expected duration of each state is large, the whole finite sample may have been collected from one regime and there would be no information about the Markov chain in the sample periodogram.

Hansen (1991) derives the Wald and LM type statistics for testing under unidentified nuisance parameters in a linear regression context with additive nonlinearity. For the preceding testing problem in nonlinear regressions with unidentified nuisance parameters under the null, we provide general asymptotic theory in the next section.

4 The General Framework

4.1 Tests Based on Nonlinear Regressions

The testing problem can be cast in the following general form:

$$\begin{aligned} z_j &= log\{(f(x_j,\theta) + \delta g(x_j,\gamma))\} + u_j \\ &= F(x_j;\theta,\delta,\gamma) + u_j \end{aligned} \qquad (16)$$

$$\begin{aligned} H_0: \quad \delta &= 0 \\ H_1: \quad \delta &\neq 0 \end{aligned} \qquad (17)$$

Here, θ is a vector of identified parameters under the null, and γ is a vector of unidentified nuisance parameters. $\{x_j\}, j = 1, \ldots, N$ is a vector of nonstochastic exogenous variables. u_j is i.i.d. white noise $WN(0, \sigma_u^2)$.

Define

$$Q_N = \frac{1}{N}\sum_{j=1}^{N}\{z_j - F(x_j; \theta, \delta, \gamma)\}^2 \tag{18}$$

Given a set of regularity assumptions A1 in Appendix A, the parameters in the model can be estimated by minimizing Q_N. Statistical testing for $\delta = 0$ can then be carried out. To implement the test, we need to make some regularity assumptions to ensure the consistency and asymptotic normality of the estimates.

Assumption 1 *Let $\psi = (\theta, \delta)$, suppose the true parameter is ψ_0, and denote $f_j(\theta) = f(x_j, \theta)$, $g_j(\gamma) = g(x_j, \gamma)$, and $F_j(\psi, \gamma) = log\{f_j(\theta) + \delta g_j(\gamma)\}$. There exists an open neigborhood N_{ψ_0} of ψ_0 such that:*

1.1 $\frac{\partial f_j}{\partial \theta}$ exists and is continuous on N_{ψ_0}.

1.2 f_j is continuous in $\theta \in \Theta$ uniformly in j. That is, for every $\epsilon > 0$, there is a $v > 0$ such that $| f_j(\theta_0) - f_j(\theta) | < \epsilon$ whenever $(\theta - \theta_0)'(\theta - \theta_0) < v$.

1.3 g_j is continuous in the compact set Γ.

1.4 $\frac{1}{N}\sum_{j=1}^{N} F_j(\psi_1, \gamma)F_j(\psi_2, \gamma)$ converges uniformly in $(\psi_1, \psi_2) \in N_{\psi_0}$ for all $\gamma \times \Gamma$.

1.5 $\lim_{N\to\infty} N^{-1}\sum_{j=1}^{N}(F_j(\psi_0, \gamma) - F_j(\psi, \gamma))^2 \neq 0$ if $\psi \neq \psi_0$.

Given these assumptions, the NLS estimates of the parameters based on minimizing Q_N are consistent (see Amemiya, 1985, Theorem 4.3.1).

The scores evaluated at the null $\psi_0 = (\theta_0, 0)$ are:

$$
\begin{aligned}
S_\theta^N(\gamma) &= \sqrt{N}\frac{\partial Q_N}{\partial \theta} \mid_{\psi_0} = -\frac{2}{\sqrt{N}} \sum_{j=1}^{N} \frac{f_j'(\theta_0)}{f_j(\theta_0)} u_j = \sum_{j=1}^{N} G_j^1 u_j = G^{1\prime} u \\
S_\delta^N(\gamma) &= \sqrt{N}\frac{\partial Q_N}{\partial \delta} \mid_{\psi_0} = -\frac{2}{\sqrt{N}} \sum_{j=1}^{N} \frac{g_j(\gamma)}{f_j(\theta_0)} u_j = \sum_{j=1}^{N} G_j^2(\gamma) u_j = G^{2\prime} u,
\end{aligned}
\tag{19}
$$

where

$$
\begin{aligned}
G_j^1 &\equiv -\frac{2}{\sqrt{N}} \frac{\partial F_j}{\partial \theta} \mid_{\psi_0} = -\frac{2}{\sqrt{N}} \frac{f_j'(\theta_0)}{f_j(\theta_0)} \\
G_j^2(\gamma) &\equiv -\frac{2}{\sqrt{N}} \frac{\partial F_j}{\partial \delta} \mid_{\psi_0} = -\frac{2}{\sqrt{N}} \frac{g_j(\gamma)}{f_j(\theta_0)}
\end{aligned}
\tag{20}
$$

Note that, the score $S_\delta^N(\gamma)$ measures the correlation between the residual u_j and the variable $g_j(\gamma)$. Under the null, u_j is uncorrelated with both G_j^1 and G_j^2. Thus, for any given γ,

$$ES^N(\gamma) = E\{S_\theta^N \;\; S_\delta^N(\gamma)\}' = 0$$

Let:

$$
\begin{aligned}
G_j(\gamma) &\equiv \{G_j^1 \;\; G_j^2(\gamma)\}' \\
G^N(\gamma) &\equiv \{G_1(\gamma), \ldots, G_N(\gamma)\} \\
G^I &\equiv \{G_1^i, \ldots, G_N^i\}' \quad I = 1, 2
\end{aligned}
$$

so that:

$$S^N(\gamma) = \{S_\theta^N, S_\gamma^N(\gamma)\}' = G^N(\gamma) * u$$

Now, define the covariance function of the scores:

$$
\begin{aligned}
V_N(\gamma_1, \gamma_2) &\equiv E(S^N(\gamma_1) S^N(\gamma_2)') = E(G^N(\gamma_1) u u' G^N(\gamma_2)') \\
&= \sigma_u^2 (G^N(\gamma_1) G^N(\gamma_2)') \\
&= \sigma_u^2 \begin{bmatrix} G^1(\gamma_1)' G^1(\gamma_2) & G^1(\gamma_1)' G^2(\gamma_2) \\ G^2(\gamma_1)' G^1(\gamma_2) & G^2(\gamma_1)' G^2(\gamma_2) \end{bmatrix}
\end{aligned}
\tag{21}
$$

$$V_N(\gamma) \equiv V_N(\gamma, \gamma) \tag{22}$$

We assume the following:

Assumption 2 *Let* $\psi \in \Psi$, $\gamma \in \Gamma$,

2.1 $\lim_{N\to\infty} V_N(\psi,\gamma) = V(\psi,\gamma)$ *exists, and is continuous uniformly in* $(\psi,\gamma) \in N_{\psi_0} \times \Gamma$*, where* N_{ψ_0} *is a neighborhood of* ψ_0.

2.2 $V(\gamma) = V(\psi_0,\gamma)$ *is positive definite uniformly over* $\gamma \in \Gamma$.

Let $S^N(\gamma) = (S_\theta^N(\gamma), S_\delta^N(\gamma))' \mid_{\psi_0}$,

2.3(a) $S^N(\gamma) \Rightarrow S(\gamma)$ *on* $\gamma \in \Gamma$*, where* $S(\gamma)$ *is a mean-zero Gaussian process with covariance function* $V(\gamma_1,\gamma_2) = \lim_{N\to\infty} V_N(\gamma_1,\gamma_2)$*, "*$\Rightarrow$*" denotes convergence in distribution.*

Or instead of 2.3(a), we assume

2.3(b) $S^N(\gamma)$ *is stochastically equicontinuous.*

Assumption 2.3(b) ensures 2.3(a) (see Pollard, 1990, Theorem 10.2).

4.1.1 The Difference Test

Given γ, let $\hat{\psi}(\gamma) = argmin Q_N(\psi,\gamma)$, the unrestricted estimates of ψ given γ; let $\tilde{\psi} = argmin Q_N(\psi,\gamma) \mid_{H_0} = argmin Q_N(\psi)$, the null estimates of ψ; and let $(\hat{\psi},\hat{\gamma}) = argmin Q_N(\psi,\gamma)$ be the unrestricted estimates of (ψ,γ).

Define the difference test statistic DT_N (analogous to the likelihood test) as follows:

$$\begin{aligned} DT_N(\gamma) &= N\frac{Q_N(\tilde{\psi}) - Q_N(\hat{\psi}(\gamma),\gamma)}{Q_N(\tilde{\psi})} \\ DT_N &= h(DT_N(\gamma)), \end{aligned} \tag{23}$$

where $h(.)$ takes the form of either a supremum or an average over $\gamma \in \Gamma$ as follows:

$$DT_N = h(DT_N(\gamma)) = \sup_{\gamma \in \Gamma} DT_N(\gamma)$$

or

$$DT_N = h(DT_N(\gamma)) = \int_{\gamma \in \Gamma} H(DT_N(\gamma))\Pi(\gamma)d\gamma \tag{24}$$

$H(.)$ is a integrable function and $\Pi(\gamma)$ is a density measure over Γ.

In order to get the asymptotic distribution of $DT_N(\gamma)$, we need to make the following regularity assumptions about the second moments of the objective function:

Assumption 3 *Let* $M_N(\psi, \gamma) = \frac{\partial^2}{\partial \psi \partial \psi'} Q_N(\psi, \gamma)$, *and* N_{ψ_0} *is a neighborhood of* ψ_0,

3.1 $Q_N(\psi, \gamma)$ *is twice differentiable in* ψ *for all* $\gamma \in \Gamma$.

3.2 $M(\psi, \gamma) = \lim_{N \to \infty} EM_N(\psi, \gamma)$ *is continuous uniformly in* $(\psi, \gamma) \in N_{\psi_0} \times \Gamma$.

3.3 $M_N(\psi, \gamma) \to_p M(\psi, \gamma)$ *for all* $(\psi, \gamma) \in N_{\psi_0} \times \Gamma$, *where "$\to$" means convergence in probability.*

3.4 $M(\gamma) = M(\psi_0, \gamma)$ *is positive definite uniformly over* $\gamma \in \Gamma$.

Recall that $\psi = (\theta, \delta)'$, and let $R = (0, 1)'$, so that the null hypothesis can be written as:

$$H_0: \quad R' * \psi = 0$$

Theorem 1 *Under the set of assumptions 1, 2, 3, we have*

$$\begin{aligned} DT_N(\gamma) &\Rightarrow \frac{1}{2\sigma_u^2} S(\gamma)' M(\gamma)^{-1} R (R' M(\gamma)^{-1} R)^{-1} R' M(\gamma)^{-1} S(\gamma) \\ &= \bar{S}(\gamma)' (2\sigma_u^2 R' M(\gamma)^{-1} R)^{-1} \bar{S}(\gamma) \end{aligned} \tag{25}$$

and

$$\begin{aligned} DT_N &= h\{DT_N(\gamma)\} \\ &\Rightarrow h\{\bar{S}(\gamma)'(2\sigma_u^2 R' M(\gamma)^{-1} R)^{-1} \bar{S}(\gamma)\}, \end{aligned}$$

where " $\Rightarrow$*" means convergence in distribution,* $S(\gamma)$ *is a zero mean Gaussian process with covariance function* $V(\gamma_1, \gamma_2)$ *as defined in equation (21) and assumption 2.1, and*

$$\bar{S}(\gamma) \equiv R' M(\gamma)^{-1} S(\gamma)$$

The proof of Theorem 1 is shown in Appendix B. Note that, if γ is known, then the statistic $DT_N(\gamma)$ converges to a χ^2 asymptotically.

4.1.2 The LM Test

As mentioned earlier, the scores measure the correlation between the residual and the random variable $g(\gamma)$. If the alternative is true and we only estimate the model under the null, then the information of $g(\gamma)$ will be left in the residual, and the scores will able to pick up the correlation. The LM test, like standard LM tests, is just based on this observation.

Define

$$LM_N(\gamma) = \tilde{S}^N(\gamma)' \tilde{V}(\gamma)^{-1} \tilde{S}^N(\gamma) \tag{26}$$

to be the LM statistic estimated under the null given $\gamma \in \Gamma$.

The LM test statistic proposed is $h(LM_T(\gamma))$, where $h(.)$ is a continuous function which takes two forms:

T1: The supremum test:

$$LM = h(LM_N(\gamma)) = \sup_{\gamma \in \Gamma}\{LM_N(\gamma)\} \tag{27}$$

T2: Average Score test:

$$LM = h(LM_N(\gamma)) = \int_{\gamma \in \Gamma} H(LM_N(\gamma))\Pi(\gamma)d\gamma, \tag{28}$$

where $H(.)$ is a integrable function and $\Pi(\gamma)$ is a density measure over Γ.

Theorem 2 *Under the set of assumptions 1, 2, 3, we have that*

$$\begin{aligned} LM_N(\gamma) &\Rightarrow \bar{S}(\gamma)'(R'M(\gamma)^{-1}R)^{-1}R'V(\gamma)^{-1}R(R'M(\gamma)^{-1}R)^{-1}\bar{S}(\gamma) \\ h(LM_N(\gamma)) &\Rightarrow h\{\bar{S}(\gamma)'(R'M(\gamma)^{-1}R)^{-1}R'V(\gamma)^{-1}R(R'M(\gamma)^{-1}R)^{-1}\bar{S}(\gamma)\} \end{aligned}$$

where "$\Rightarrow$" means convergence in distribution.

The proof of Theorem 2 is provided in Appendix C.

Theorem 3 *Under the set of regularity assumptions A2 as outlined in Appendix A, the Difference Test DT in Theorem 1 is asymptotically equivalent to the LM test in Theorem 2.*

The proof of Theorem 3 is provided in Appendix D.

When γ is given, both the $LM_N(\gamma)$ and $DT_N(\gamma)$ have asymptotic χ^2 distribution. We have to estimate the model only under the null to carry out the LM test. As the usual LR test, the DT test requires estimation both under the null and the alternative.

4.2 Tests Based on Asymptotic Efficient Estimates

We can also base our tests on an efficient estimation procedure.

Suppose we have T observations $\{y_t\}, t = 1, 2, \ldots, T$, and the periodogram of y_t is $I_y(w_j)$, where $w_j = \frac{2\pi j}{T}$, $j = 0, 1, \ldots, [T/2]$. Let $F(w_j; \theta, \delta, \gamma)$ denote the spectrum density outlined in section 3.1, and $\delta \equiv \frac{\alpha_1^2}{\lambda}$. Define $N \equiv [T/2]$.

Whittle's (1951) Maximum likelihood function in frequency domain is the following:

$$\begin{aligned} W_N(\theta, \delta; \gamma) &= \log L = -N \log 2\pi - \frac{1}{2} \sum_{j=1}^{N} \log F_y(w_j; \theta, \delta, \gamma) \\ &- \frac{1}{2} \sum_{j=1}^{N} \frac{I_y(w_j)}{F_y(w_j; \theta, \delta, \gamma)}, \end{aligned} \tag{29}$$

where

$$\begin{aligned} F_y(w_j; \theta, \delta, \gamma) &= f(w_j; \theta) + \delta g(w_j; \gamma) \\ g(w_j; \gamma) &\equiv \frac{\gamma + 1}{1 + \gamma^2 - 2\gamma cos(w_j)} \\ f(w_j; \theta) &\equiv \mid \Phi^{-1}(exp(-iw_j)) \mid^2 \sigma^2 \\ \theta &\equiv (\Phi, \sigma^2) \end{aligned}$$

For a general stationary process, estimation based on maximizing W_N is asymptotically equivalent to the ML estimator (see Harvey, 1991). Tests can be based on this efficient estimation procedure.

We want to test $\delta = 0$. Under this null, γ is not identified. Thus we still have a nonstandard testing problem.

Denote $\psi = (\theta, \delta)'$, and the true parameter $\psi_0 = (\theta_0, 0)'$. Given a nuisance parameter γ, the estimation of ψ is a standard procedure. We define the scores evaluated at the null $\psi_0 = (\theta_0, 0)$ as the following:

$$\begin{aligned} S_\theta^N(\gamma) &= \frac{1}{\sqrt{N}}\frac{\partial W_N}{\partial \theta}\mid_{\psi_0} = \frac{1}{2\sqrt{N}}\sum_{j=1}^{N}(\frac{I_j}{f(w_j,\theta_0)} - 1)\frac{\partial \log f(w_j,\theta_0)}{\partial \theta} \\ &= \frac{1}{2\sqrt{N}}\sum_{j=1}^{N}(\frac{I_j}{f(w_j,\theta_0)} - 1)G_j^1 \\ S_\delta^N(\gamma) &= \frac{1}{\sqrt{N}}\frac{\partial W_N}{\partial \delta}\mid_{\psi_0} = \frac{1}{2\sqrt{N}}\sum_{j=1}^{N}(\frac{I_j}{f(w_j,\theta_0)} - 1)\frac{g(w_j,\gamma)}{f(w_j,\theta_0)} \\ &= \frac{1}{2\sqrt{N}}\sum_{j=1}^{N}(\frac{I_j}{f(w_j,\theta_0)} - 1)G_j^2(\gamma) \end{aligned}$$

where

$$\begin{aligned} G_j^1 &\equiv \frac{\partial \log F_j}{\partial \theta}\mid_{\psi_0} = \frac{\partial \log f_j(\theta_0)}{\partial \theta} \\ G_j^2(\gamma) &\equiv \frac{\partial \log F_j}{\partial \delta}\mid_{\psi_0} = \frac{g_j(\gamma)}{f_j(\theta_0)} \end{aligned} \tag{30}$$

Let

$$\xi_j \equiv \frac{I_j}{f(w_j,\theta_0)} - 1$$

It is shown in Harvey (1991) that $\xi_j's$ are i.i.d and $E\xi_j = 0$, $var(\xi_j) = 1$. If we do not assume normality, these results hold asymptotically (see Brillinger, 1981; Harvey, 1991).

Define the score vector

$$S^N(\gamma) = \{S_\theta^N \;\; S_\delta^N(\gamma)\}'$$

The covariance matrix of the scores (the information matrix) is given by:

$$\begin{aligned} V_N(\gamma) &\equiv E(S^N(\gamma)S^N(\gamma)') \\ &= \frac{1}{4N}\begin{bmatrix} \sum_{j=1}^N G_j^1{}'G_j^1 & \sum_{j=1}^N G_j^1{}'G_j^2(\gamma) \\ \sum_{j=1}^N G_j^2(\gamma)'G_j^1 & \sum_{j=1}^N G_j^2(\gamma)'G_j^2(\gamma) \end{bmatrix} \end{aligned}$$

and

$$
\begin{aligned}
V(\gamma) &\equiv \lim_{N\to\infty} V_N(\gamma) \\
&= \frac{1}{8\pi}\begin{bmatrix} \int_{-\pi}^{\pi} G^1(w)'G^1(w)dw & \int_{-\pi}^{\pi} G^1(w)'G^2(w,\gamma)dw \\ \int_{-\pi}^{\pi} G^2(w,\gamma)'G^1(w)dw & \int_{-\pi}^{\pi} G^2(w,\gamma)'G^2(w,\gamma)dw \end{bmatrix}
\end{aligned}
$$

Further, in Harvey (1991), it is shown that the second moment of the objective function W_N is the following:

$$
\begin{aligned}
M_N(\gamma) \equiv M_N(\psi_0,\gamma) &= E\frac{1}{N}\frac{\partial^2}{\partial\psi\partial\psi'}W_N(\psi,\gamma) = \frac{1}{2N}\sum_{j=1}^{N} G_j(\gamma)G_j(\gamma)' = 2V_N(\gamma) \\
M(\gamma) &= \lim_{N\to\infty} M_N(\gamma) = 2V(\gamma)
\end{aligned}
$$

With all these results, we can define the DT test and LM test as follows:

4.2.1 The Difference Test (Likelihood Ratio Test):

$$
\begin{aligned}
DT_N(\gamma) &= 2N(W_N(\hat{\psi}(\gamma),\gamma) - W_N(\tilde{\psi})) \\
DT_N &= h(DT_N(\gamma)) \qquad (31)
\end{aligned}
$$

where $h(.)$ takes the form of either a supremum or an average over $\gamma \in \Gamma$ as follows:

$$
\begin{aligned}
DT_N &= h(DT_N(\gamma)) = \sup_{\gamma\in\Gamma} DT_N(\gamma) \\
\text{or} \quad DT_N &= h(DT_N(\gamma)) = \int_{\gamma\in\Gamma} H(DT_N(\gamma))\Pi(\gamma)d\gamma \qquad (32)
\end{aligned}
$$

where $H(.)$ is a integrable function and $\Pi(\gamma)$ is a density measure over Γ.

4.2.2 The LM Test

Define

$$
LM_N(\gamma) \equiv \tilde{S}^N(\gamma)'\tilde{M}(\gamma)^{-1}\tilde{S}^N(\gamma) = \frac{1}{2}\tilde{S}^N(\gamma)'\tilde{V}(\gamma)^{-1}\tilde{S}^N(\gamma) \qquad (33)
$$

to be the LM statistic estimated under the null given $\gamma \in \Gamma$.

The LM test statistic proposed is $h(LM_N(\gamma))$. $h(.)$ is a continuous function which takes the two forms as in the DT tests.

As in the previous section, define

$$V_N(\gamma_1, \gamma_2) \equiv E(S^N(\gamma_1)S^N(\gamma_2)'),$$

so that $V(\gamma) = \lim_{N\to\infty} V_N(\gamma, \gamma)$. If we maintain the same assumptions about $V_N(\gamma_1, \gamma_2)$, $M_N(\gamma)$ and $S^N(\gamma)$ as outlined in the previous section, we would have the following theorem.

Theorem 4 *Under the set of assumptions 1, 2, 3, the test statistics $DT_N(\gamma)$ and $LM_N(\gamma)$ converge in distribution to*

$$\begin{aligned} T(\gamma) &\equiv S(\gamma)'M(\gamma)^{-1}R(R'M(\gamma)^{-1}R)^{-1}R'M(\gamma)^{-1}S(\gamma) \\ &= \bar{S}(\gamma)'(R'M(\gamma)^{-1}R)^{-1}\bar{S}(\gamma) \end{aligned} \tag{34}$$

DT_N and LM_N converge in distribution to

$$T \equiv h\{\bar{S}(\gamma)'(R'M(\gamma)^{-1}R)^{-1}\bar{S}(\gamma)\}$$

where $S(\gamma)$ is a zero mean Gaussian process with covariance function $V(\gamma_1, \gamma_2)$,

$$\begin{aligned} V(\gamma_1, \gamma_2) &\equiv \lim_{N\to\infty} V_N(\gamma_1, \gamma_2) \\ &= \frac{1}{4N} \lim_{N\to\infty} \begin{bmatrix} \sum_{j=1}^N G_j^1(\gamma_1)'G_j^1(\gamma_2) & \sum_{j=1}^N G_j^1(\gamma_1)'G_j^2(\gamma_2) \\ \sum_{j=1}^N G_j^2(\gamma_1)'G_j^1(\gamma_2) & \sum_{j=1}^N G_j^2(\gamma_1)'G_j^2(\gamma_2) \end{bmatrix} \\ &= \frac{1}{8\pi} \begin{bmatrix} \int_{-\pi}^{\pi} G^1(w,\gamma_1)'G^1(w,\gamma_2)dw & \int_{-\pi}^{\pi} G^1(w,\gamma_1)'G^2(w,\gamma_2)dw \\ \int_{-\pi}^{\pi} G^2(w,\gamma_1)'G^1(w,\gamma_2)dw & \int_{-\pi}^{\pi} G^2(w,\gamma_1)'G^2(w,\gamma_2)dw \end{bmatrix} \end{aligned} \tag{35}$$

and

$$\bar{S}(\gamma) \equiv R'M(\gamma)^{-1}S(\gamma)$$

We can basically follow the same procedure as in the previous section to prove the theorems. In proving the convergence of $DT_N(\gamma)$ and $LM_N(\gamma)$, one needs to use the results regarding the consistency of the estimates from Brockwell and Davis (1995).

To summarize this section, we propose two test statistics, LM_N and DT_N, for testing the null of $\delta = 0$. The LM test, like the usual LM tests, is designed to capture the possible correlation between the residual and the omitted variable $g(., \gamma)$. The DT test, like the usual LR test, reflect the difference between the objective function under the null and the alternative. Under the null, γ is not identified and the statistics have non-standard distributions. We have derived their asymptotic distributions and shown that under some regularity assumptions the two tests are asymptotically equivalent.

5 Implementation of the Tests

In this section we discuss procedures to implement the tests in a simple Markov regime switching model. The model are chosen for illustration purpose. We have shown earlier that, the distribution of the test statistics is model-specific, since the covariance function of the χ^2 process is model specific. This prevents the general tabulation of the distribution. Based on the specific model in hand, one has to find the covariance function of the χ^2 process, either analytically or numerically, and simulate or bootstrap the asymptotic and finite sample distributions. This section serves as an illustration of the whole testing process.

Since the DT test requires estimating parameters both under the null and the alternative, the LM test has an advantage in terms of simulating and bootstrapping finite sample empirical distributions. The trade off can be the power, as in the situation of usual classical tests. For illustrative simplicity, we opt to simulate and bootstrap the LM test in this section.

In the following subsections, we first illustrate the model and derive the covariance function for the χ^2 process (Lemmas 1 and 2). We then show procedures to simulate the asymptotic distribution, simulate and bootstrap the finite sample distribution. After the null distribution is obtained, we proceed to use some actual alternative DGPs to investigate the power of the LM test.

5.1 Switching Between Two Normals with Different Means: The Test

The model (hereafter we call it Model I) is given by:

$$\begin{aligned} y_t &= \alpha_0 + \alpha_1 S_t + \epsilon_t \\ P(S_t &= 1 \mid S_t = 1) = p \\ P(S_t &= 0 \mid S_t = 0) = q \end{aligned} \tag{36}$$

where ϵ_t is $N(0, \sigma^2)$. The test we want to conduct is:

$$\begin{aligned} H_0: \quad \alpha_1 &= 0 \\ H_1: \quad \alpha_1 &\neq 0 \end{aligned}$$

We restrict $\gamma \equiv p + q - 1 \in [-a, a]$, where a is some positive number less than one. Note that in the time-domain maximum likelihood framework, under the null, (p, q) are not identified and the information matrix is singular even if we treat (p, q) to be nuisance parameters and restrict them in (0, 1).

From equation (12), the spectrum of y_t is

$$f_y(w) = \frac{\alpha_1^2}{\lambda} \frac{\gamma+1}{1+\gamma^2-2\gamma cos(w)} + \sigma^2 \tag{37}$$

where λ is the expected duration of each cycle for the Markov chain S_t, and $\gamma = p+q-1$, as defined before.

Following the form of equation (14), we have

$$\begin{aligned} z_j &\equiv log(I_y(w_j)) = log(f_y(w_j)) + log(\xi_j) \\ &= c + log\{\frac{\alpha_1^2}{\lambda}\frac{\gamma+1}{1+\gamma^2-2\gamma cos(w_j)} + \sigma^2\} + u_j \\ &= log\{\frac{\alpha_1^2}{\lambda_1}\frac{\gamma+1}{1+\gamma^2-2\gamma cos(w_j)} + \sigma_1^2\} + u_j \\ &= log\{\delta\frac{\gamma+1}{1+\gamma^2-2\gamma cos(w_j)} + \sigma_1^2\} + u_j \end{aligned} \tag{38}$$

where u_j is a mean zero white noise, $\delta \equiv \frac{\alpha_1^2}{\lambda_1}$.

The original test is equivalent to the following test,

$$\begin{aligned} H_0: \quad \delta &= 0 \\ H_1: \quad \delta &\neq 0 \end{aligned} \tag{39}$$

Following the form of equation (16), we have

$$\begin{aligned} f_j(\theta) &= \sigma_1^2 \\ g_j(\gamma) &= \frac{\gamma+1}{1+\gamma^2-2\gamma cos(w_j)} \end{aligned} \tag{40}$$

The scores are:

$$\begin{aligned} S_{\sigma_1^2}^N(\gamma)\mid_{\psi_0} &= \sqrt{N}\frac{\partial Q_N}{\partial \sigma_1^2}\mid_{\psi_0} = -\frac{2}{\sqrt{N}}\sum_{j=1}^N \frac{u_j}{\sigma_1^2} = \sum_{j=1}^N G_j^1 u_j \\ S_\delta^N(\gamma)\mid_{\psi_0} &= \sqrt{N}\frac{\partial Q_N}{\partial \delta}\mid_{\psi_0} = -\frac{2}{\sqrt{N}}\sum_{j=1}^N \frac{g_j(\gamma)}{\sigma_1^2}u_j = \sum_{j=1}^N G_j^2(\gamma)u_j \end{aligned} \tag{41}$$

We have the following lemma:

Lemma 1 *For Model 1, the covariance function for the Gaussian process $S(\gamma)$ in Theorem 1 is equal to:*

$$V(\gamma_1,\gamma_2) = \frac{4\sigma_u^2}{\sigma_1^4}\begin{bmatrix} 1 & \frac{1}{1-\gamma_2} \\ \frac{1}{1-\gamma_1} & \frac{1+\gamma_1\gamma_2}{(1-\gamma_1\gamma_2)(1-\gamma_1)(1-\gamma_2)} \end{bmatrix} \tag{42}$$

The proof of Lemma 1 is achieved by using contour integral technique in the limit, as shown in Appendix E.

From equation (42) we see that if $\gamma = 0$, the covariance function is singular. So, in addition to restricting γ to be strictly less than 1 and greater than -1, we need to restrict $\gamma \neq 0$.

It is also easy to check that all the assumptions *A2* in Appendix A are satisfied by Model 1. In particular, in *A2.1*, $\lim_{N\to\infty}\frac{1}{N}\sum_{j=1}^{N}\frac{\partial F_j(\gamma)}{\partial\psi}\mid_{\psi_0}\frac{\partial F_j(\gamma)}{\partial\psi'}\mid_{\psi_0}= \frac{1}{4}V(\gamma,\gamma)$ is a finite nonsingular matrix. For *A2.2* and *A2.5*, $\frac{1}{N}\sum_{j=1}^{N}\frac{\partial F_j(\gamma)}{\partial\psi}\frac{\partial F_j(\gamma)}{\partial\psi'}$, $\frac{1}{N}\sum_{j=1}^{N}F_j(\psi_1,\gamma)\frac{\partial^2 F_j(\psi,\gamma)}{\partial\psi\partial\psi'}\mid_{\psi_2}$ will converge to integrals of some integrable, finite and continuous function of $w\in[0,\pi]$ as $N\to\infty$. *A2.3* and *A2.4* can also be easily checked. Thus, we have the following lemma:

Lemma 2 *For Model 1, the following relationship holds for the moment matrix*

$$M(r) = \frac{1}{2\sigma_u^2}V(\gamma,\gamma) = \frac{1}{2\sigma_u^2}V(\gamma)$$

Thus, the LM test is equivalent to the DT test, and both of these tests have asymptotic distribution of a Chi-square process

$$T(\gamma) \equiv \bar{S}(\gamma)'\bar{V}(\gamma)^{-1}\bar{S}(\gamma), \tag{43}$$

where $\bar{S}(\gamma)$ is a Gaussian process with covariance function

$$\bar{V}(\gamma_1,\gamma_2) = R'M(\gamma_1)^{-1}V(\gamma_1,\gamma_2)M(\gamma_2)^{-1}R \tag{44}$$

and

$$\bar{V}(\gamma) \equiv \bar{V}(\gamma,\gamma) = 2\sigma_u^2 R' M(\gamma)^{-1} R$$

The proof of the lemma is shown in Appendix F.

To obtain the asymptotic distribution, one needs to simulate the Gaussian process $\bar{S}(\gamma)$ with covariance function $\bar{V}(\gamma_1,\gamma_2)$. Now,

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

Note that

$$\begin{aligned}
M(\gamma)^{-1}\begin{bmatrix} 0 \\ 1 \end{bmatrix} &= \{\frac{1}{2\sigma_u^2} V(\gamma)\}^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\
&= \frac{\sigma_1^4}{2}\begin{bmatrix} 1 & \frac{1}{1-\gamma} \\ \frac{1}{1-\gamma} & \frac{1+\gamma^2}{(1-\gamma^2)(1-\gamma)^2} \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\
&= \frac{\sigma_1^4}{2}\frac{(1-\gamma^2)(1-\gamma)^2}{2\gamma^2} \begin{bmatrix} \frac{1+\gamma^2}{(1-\gamma^2)(1-\gamma)^2} & -\frac{1}{1-\gamma} \\ -\frac{1}{1-\gamma} & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\
&= \frac{\sigma_1^4}{4}\frac{(1-\gamma)^2(1-\gamma^2)}{\gamma^2} \begin{bmatrix} -\frac{1}{1-\gamma} \\ 1 \end{bmatrix}
\end{aligned}$$

Hence, one gets:

$$\begin{aligned}
\bar{V}(\gamma_1,\gamma_2) &= \frac{\sigma_1^4\sigma_u^2}{2}\frac{(1-\gamma_1)^2}{\gamma_1}\frac{(1-\gamma_2)^2}{\gamma_2}\frac{(1+\gamma_1)(1+\gamma_2)}{1-\gamma_1\gamma_2} \\
&= \frac{\sigma_1^4\sigma_u^2}{2}\Omega(\gamma_1,\gamma_2) \qquad (45)
\end{aligned}$$

We can see that the distribution of the Chi-square process

$$T(\gamma) \equiv \bar{S}(\gamma)' \bar{V}(\gamma)^{-1} \bar{S}(\gamma)$$

does not depend on model parameters σ_1^2 and σ_u^2. The distribution of $T(\gamma)$ is the same as the distribution of the Chi-square process

$$S_\Omega(\gamma)'\Omega(\gamma)^{-1}S_\Omega(\gamma),$$

where $S_\Omega(\gamma)$ is a Gaussian process with covariance function $\Omega(\gamma_1,\gamma_2)$ that depends on γ only.

The simulated asymptotic null distribution of the DT and LM test statistic for Model I is provided in Table 1.[2]

5.2 Simulating and Bootstrapping the Approximating Distribution in Finite Sample

Having found the covariance function of the χ^2 process, we now discuss procedures to simulate the distributions.

As discussed earlier, although for Model I the distribution of the Chi-square process does not depend on other identifiable model parameters ψ except γ, in general the form of the covariance function $V(\gamma_1,\gamma_2)$ is model-specific and depends on the region Γ as well as other model parameters ψ. One has to simulate the null distribution of the tests according to a specific model in hand. Sometimes it may not be possible to find an explicit expressions for $V(\gamma_1,\gamma_2)$ and $M(\gamma)$, and one has to approximate them with $V_N(\gamma_1,\gamma_2)$ and $M_N(\gamma)$. In what follows, we discuss ways to approximate the null distribution in finite sample.

Note that, since our testing problem is in a nonlinear framework with unidentified nuisance parameters under the null, to simulate the null distribution it is always desir-

[2]Garcia (1992) discusses a method for simulating a Chi-square process given a covariance function. The algorithm is presented in Appendix G.

able to estimate the model only under the null to avoid the local minimum/maximum problem under the alternative; otherwise one has to search the whole parameter space to ensure that the loss is minimized globally. A numerical minimization program can easily end up with a local minimum, with a loss that is higher under the alternative than under the null. Thus, in the following, we obtain the finite sample distribution based on the estimates under the null only.

5.2.1 Procedure I: Approximate Distribution by Gaussian Parametric Simulation

(1) First, estimate the model under the null and get the parameter estimates $\tilde{\psi}$ and the residuals $\tilde{u}_j, j = 1, \ldots, N$.

(2) Calculate

$$\tilde{\tilde{V}}_N(\gamma_1, \gamma_2) = R'\tilde{M}_N(\gamma_1)^{-1}\tilde{V}_N(\gamma_1, \gamma_2)\tilde{M}_N(\gamma_2)^{-1}R, \tag{46}$$

where

$$\begin{aligned} \tilde{V}_N(\gamma_1, \gamma_2) &= E(\tilde{S}^N(\gamma_1)\tilde{S}^N(\gamma_2)') \\ &= \tilde{\sigma}_u^2 \begin{bmatrix} \sum_{j=1}^N \tilde{G}_j^1(\gamma_1)\tilde{G}_j^1(\gamma_2) & \sum_{j=1}^N \tilde{G}_j^1(\gamma_1)\tilde{G}_j^2(\gamma_2) \\ \sum_{j=1}^N \tilde{G}_j^2(\gamma_1)\tilde{G}_j^1(\gamma_2) & \sum_{j=1}^N \tilde{G}_j^2(\gamma_1)\tilde{G}_j^2(\gamma_2) \end{bmatrix}, \end{aligned} \tag{47}$$

and $\tilde{M}_N(\gamma)$ is accordingly defined.

(3) Simulate $S^{Nk}(\gamma)$ according to:

$$\tilde{S}^{Nk}(\gamma) = \begin{bmatrix} \sum_{j=1}^N \tilde{G}_j^1(\gamma)\tilde{\sigma}_u\epsilon_j^k \\ \sum_{j=1}^N \tilde{G}_j^2(\gamma)\tilde{\sigma}_u\epsilon_j^k \end{bmatrix} \tag{48}$$

where $\epsilon_j^k, j = 1, \ldots, N$ are from $N(0,1)$. Gaussian simulation is justified here because $\tilde{S}^N(\gamma)$ converges to a Gaussian process.

Calculate

$$\tilde{\tilde{S}}^{Nk}(\gamma) = R'\tilde{M}_N(\gamma)^{-1}\tilde{S}^{Nk}(\gamma) \tag{49}$$

and find the statistic

$$T_N^k = \sup_{\gamma\in\Gamma} \tilde{\tilde{S}}^{Nk}(\gamma)'\tilde{\tilde{V}}_N(\gamma)^{-1}\tilde{\tilde{S}}^{Nk}(\gamma) \tag{50}$$

Repeating step (3) K times, $k = 1, \ldots, K$, one gets an approximate distribution for the test statistic T_N with sample size N ($N \equiv [T/2]$).

The simulated distributions for Model I with $K = 1000$, different sample size are reported in Table 1. We restrict $\gamma \in \Gamma = [0.05, 0.95] + [-0.95, -0.05]$ with increments of 0.01. We see that the simulated distributions are very close to the asymptotic distribution, even with a finite sample of 200 observations. Moreover, the distributions are not sensitive to the increments of γ.

5.2.2 Procedure II: Approximate Distribution by log Chi-square(2) Parametric Simulation

Under the null, we know that u_j is a mean adjusted $\log\chi^2(2)$ for Model I. Thus, we can simply generate ϵ_j^k from the $\log\chi^2(2)$ distribution in step (3) above and find the distribution of the statistic. The simulated distribution for the LM test with sample size of 200 is reported in Table 2. We see that, this distribution is very close to the asymptotic distribution and the distribution simulated by the Gaussian simulation procedure. But the upper-end of the distribution is higher than the asymptotic distribution and Gaussian simulated distribution.

5.2.3 Procedure III: Empirical Finite Sample Distribution by Bootstrap

Procedure I uses the fact that asymptotically, $S(\gamma)$ is a Gaussian process, and thus $S^T(\gamma)$ is simulated by a Gaussian simulation procedure. This approximation may not be good in finite sample. Procedure II is better in finite sample in the sense that, if y_t is normal under the null, u_j will come from a $\log \chi^2(2)$ distribution, even in finite sample. However, if we do not assume y_t to be Gaussian under the null, the finite sample distribution of u_j is unknown.[3] In this case, the empirical finite sample null distribution of the LM test can still be generated by a bootstrap procedure as follows:

(1) Estimate the model under the null and get the residual u_j and estimated $\tilde{\sigma}_1^2$.

(2) Resample $\{u_j^k\}$ from $\{u_j\}$ with replacement, and calculate the LM statistic.

Repeating step (2) K-times for a large K, one gets the empirical null distribution of the LM statistic.

The finite sample (with 200 observations) LM distribution by the bootstrap procedure is reported in Table 2.

Here, the bootstrapped distribution of the LM statistic can be viewed as an empirical distribution as compared to the simulated asymptotic distribution. It is only for comparison. In general, the bootstrapped distribution could be biased if the finite sample distribution of LM_N depends on parameters σ_1^2, since we do not know the true σ_1^2 and we are bootstrapping $LM_N(\tilde{\sigma}_1^2)$. However, for Model I we know that the asymptotic distribution does not depend on model parameter σ_1^2, so the bias problem is not particularly serious here in large sample.

[3] u_j is still i.i.d. distributed asymptotically. See Brillinger (1981).

5.3 Finite Sample Size and Power of the LM test

Having the null distributions in hand, we want to know how powerful the LM test is. We also want to know the actual rejection rates for data generated under the null when the test is implemented by the proposed procedure. In following we generate data under the null and under several alternatives and report the rejection rates.

5.3.1 When data are generated under the null

We generate data of three sample sizes: 200 observations, 500 observations, and 1000 observations from the null. The DGP is a standard normal with mean zero and a standard deviation of one. We calculate the LM statistic and report the actual rejection rates in Table 3. The rejection rates are calculated from 1000 replications.

We see that the actual rejection rates are higher than the nominal size of the simulated asymptotic distribution. But the rejection rates are close to the nominal size, specially when we have a large sample.

5.3.2 When data are generated under the homoskedastic alternative

To see the power of the LM test against the Markov switching alternative specified in Model I (Equation 36), we generate two types of observations according to:

Sample Type I: In equation (36), we set $\alpha_0 = -1$, $\alpha_1 = 2$, $p = 0.9$, $q = 0.75$, with $\sigma_\epsilon = 0.5$. In this sample, the data is less noisy and the two regimes have clear separation (the mean of one regime is more than two standard deviations away from the mean of the other regime).

Sample Type II: In equation (36), we set $\alpha_0 = -1$, $\alpha_1 = 2$, $p = 0.9$, $q = 0.75$, with $\sigma_\epsilon = 1$. In this sample, the data is more noisy than type I sample and the two regimes have less clear separation (the mean of one regime is exactly two standard deviations away from the other).

We generate a sample of Type I with 200 observations, three samples of Type II with 200, 500, and 1000 observations each, estimate the model under the null, and calculate the LM statistic (28). We repeat the process for 1000 replications. The rejection rates for both sample types under the alternative are reported in Table 4.

For Sample Type I with 200 observations, at the significance level of 5%, the power is 100% according to the asymptotic distribution, and the power is greater than 99% according to the finite sample bootstrapped distribution.

For Sample Type II with 200 observations, at the significance level of 5%, the power is 96% using the asymptotic distribution, and the power is 93% if we use the finite sample bootstrapped distribution. The higher noise reduces the power of the LM test somewhat in the finite sample, but the test still has considerable power.

Now if we draw 500 and 1000 observations from Sample Type II, the power of the LM test increases a lot even for the noisy data. In fact, at the 1% significance level of the asymptotic distribution, the rejection rate is 100%; and at the 1% level of the bootstrapped distribution, the rejection rate is 99.5% for 500-obs sample and 100% for the 1000-obs sample.

5.3.3 Comparison between our test and Hansen's (1992,1996) test

The comparison of the finite sample size and power of the LM test and Hansen's (1992,1996) LR bound test is reported in Table 5. We see that the LM test has worse sizes than Hansen's (1992,1996) test at the nominal sizes of 5 and 10 percent, but it has a better size than Hansen's (1992,1996) at the nominal size of 20 percent. The finite sample power performance of the LM test is better than that of the Hansen's (1992,1996) test.[4] Note that our test is much easier to compute, we only have to do an one-dimensional grid search, while Hansen's (1992,1996) test needs grid search in three-dimensions. This is why we are able to report rejection rates for 1000 trials.

5.3.4 When data are generated under the heteroskedastic alternative

A limitation of the test in frequency domain is that, the distribution is derived under homoskedastic assumption. The spectrum is the summation of the information about the unconditional second moments. It can't identify conditional heteroskedasticity. However, if the alternative DGP is Markov switching in both mean and variance, we show that the LM test has power against this alternative and would reject the null. When the null is rejected, further testing of heteroskedasticity can be carried out in a standard fashion, using the regular LR, LM, or Wald tests. This is discussed in detail by Hamilton (1996).

We generate three samples of 200, 500, and 1000 observations each according to the following parameter setting:

Sample Type III: In equation (36), we set $\alpha_0 = -1$, $\alpha_1 = 2$, $p = 0.9$, $q = 0.75$. The volatility in each regime is different, with $\sigma_{0\epsilon} = 1.3$, $\sigma_{1\epsilon} = 1$. In this sample, the data is very noisy and the mean of one regime is within two standard deviations of the other regime.

For these three samples, the rejection rates are shown in Table 6.

[4]One could otherwise say that comparison of the size and power of the LM test and those of Hansen's (1992,1996) test is not exactly accurate–the rejection rates reported in Hansen (1992,1996) is only from 50 trials, so his rejection rates would have large standard errors.

As we see, the LM test has power against the heteroskedastic alternative. The statistic can detect the effect of the switching mean. Once the null is rejected, one can proceed to carry other specification tests in a standard fashion.

5.3.5 When data are generated under the alternative with $p+q$ close to one

As discussed in section 5, our test wound not work when $p+q=1$, or $\gamma=p+q-1=0$. How well will the LM test performance in the neighborhood of $p+q=1$? Here we generate a Sample Type IV: In equation (36), we set $\alpha_0=-1$, $\alpha_1=2$, $p=0.6$, $q=0.395$, and $\sigma_\epsilon=1$. The sample size is 200.

The actual rejection rates from 1000 replications are reported in Table 7. We see that the LM test has no power when $p+q$ is close to one.

6 Applications: Test Results of Two Actual Data Examples

6.1 Testing Whether Exchange Rates Have Two Regimes With Different Means

Engel and Hamilton (1990) propose a Markov regime switching model for exchange rates. The changes in exchange rate $y_t=e_t-e_{t-1}$ are modeled as draws from two different Normals depending on the value of unobserved "state" or "regime" s_t, where s_t is modeled as a Markov chain process as in Model 1, e_t equal to 100 times the log of the exchange rate measured in Dollars per unit of foreign currency. Here we ask the question: Is the two-regime model for exchange rates significant against the null of a single regime model?

The null here is that the log of exchange rates follow a random walk process. The alternative is the Markov switching process as in Model I. We use the same data sets as the one used in Engel and Hamilton (1990).[5] They use three currencies: the German Mark, the French Franc, and the British Pound. The sample period of their data set is 73:IV to 88:I.

The estimated LM statistics for the three currencies are reported in Table 8. We see that, for all the three currencies, the LM statistics are not significant, no matter whether we use the asymptotic distribution in Table 1 or the bootstrapped finite sample distribution in Table 2.

Since the test also has power against the heteroskedastic alternative as specified in Engel and Hamilton (1990), the tests tell us that the estimated Markov regime switching models for the three exchange rates as reported in Engel and Hamilton (1990) are not significant against the null of a single state random walk process, at least for the given sample they used in their paper.

To check the robustness of the results, we conduct the same test to an expanded sample: 73:IV to 96:I for the three currencies. The estimated LM statistics are reported in Table 8. We see that, at the 1% significance level, the null of random walk cannot be rejected, either by the simulated asymptotic distribution or by the bootstrapped distribution, for all the three currencies. Even at the 5% significance level of the bootstrapped distribution, the null cannot be rejected for the three currencies. The null can be rejected at the 5% level of the asymptotic distribution, for French Franc and German Mark, in the sample period 73:IV–96:I.

---

[5] We would like to thank Hamilton for providing the data sets.

6.2 Testing a Two-Regime Markov Switching Model of the Federal Funds Rate

The Federal Reserve changed the method of monetary control twice in October 1979 and October 1982. In the last change, the Federal Reserve altered the way of implementing monetary policy from monitoring money supply to managing interest rates. Specifically, the Fed has carried out monetary policy by either slowly easing or slowly tightening the Federal fund target rates. Rudebusch (1995) finds that a change in the Fed funds target is likely to be followed by another change in the same direction.

The Model in section 5.1 can be a proper statistical model to characterize the two regimes in the Fed's monetary policy, where regime 0 and 1 represent easing or tightening. Let $y_t = r_t - r_{t-1}$ be the monthly changes in the Federal Funds rate, where r_t equal to 100 times the log of Federal Funds rate. The easing regime would correspond to a negative mean state for y_t; and the tightening regime would be a positive mean state. The expected policy durations are characterized by the transition probabilities p and q.

Monthly data of Federal Funds rate are obtained from the Federal Reserve Bank of New York. We have data from January 1955 to November 1995. The sample is divided into three subsamples–Sample 1: January, 1955 to September, 1979; Sample 2: October 1979 to September 1982; Sample 3: October 1982 to November, 1995. The sample divisions are set according to the timing of the two major changes in the Federal Reserve's policy, which are likely to cause structural changes in the model parameters. Huizinga and Mishkin (1986) and Roley (1986) have documented shifts in the stochastic process of interest rates in October 1979 and October 1982. We choose the subsamples because we want to test if there was significant policy regimes (easing/tightening) during each policy period (before October 1979 and after October

1982) and we don't want the rejection of random walk simply because of structural breaks.

The estimated LM statistic for the three subsamples are: $LM = 21.98$ for Sample 1; $LM = 6.74$ for Sample 2; $LM = 28.38$ for Sample 3. Thus, even at the 1% significance level of both asymptotic distribution and the bootstrapped distribution, the null of single policy regime can be rejected for both the periods 55:1-79:9 and 82:10-95:11, indicating different policy regimes in which the Federal Funds rate were either slowly raising or falling. For the subsample 2, the null cannot be rejected at the 1% level, but the sample is too small (only 36 observations) for any meaningful conclusions.

Note that, the first sample has much more observations than the third sample, yet the LM statistic is higher in the last sample period, during which the Fed has carried out monetary policy by slowly adjusting the Federal Funds rate. This again indicates the power of the LM test.

7 Conclusion

This paper proposes a framework for testing under nonstandard conditions in the frequency domain, which can be applied to test the null of single state in a class of Markov regime switching models. It is shown that if we transform the Markov switching model into a frequency domain testing problem, we only have to face the issue of unidentified nuisance parameters in a nonlinear context. Two tests, the DT test and the LM test, are proposed and the exact asymptotic distributions are derived. Under a set of regularity assumptions, the two tests are asymptotically equivalent.

The asymptotic distribution can be obtained by simulating a Gaussian process with a model-specific covariance function. Since the DT test requires estimating parameters both under the null and the alternative, the LM test has an advantage in terms of simulating and bootstrapping finite sample empirical distributions.

The LM test is applied to Markov regime switching models of exchange rates and Federal Funds rate. We used the same exchange rates data in Engel and Hamilton (1990). The null of random walk is not rejected in the exchange rates model for Engel and Hamilton (1990)'s sample period 73:IV-88:I. In an expanded sample 73:IV-96:I, the null cannot be rejected at the 1% level for all the three currencies; at the 5% level of the asymptotic distribution, the null can be rejected for French Franc and German Mark, but this rejection cannot hold at the 5% level of the bootstrapped finite sample distribution. For the Federal Funds rate model, the null of random walk is rejected at the 1% level in subsample periods 55:1-79:9 and 82:10-95:11.

Compared to Hansen's (1992,1996) LR bound test, under the null the LM test has higher rejection rate at the nominal sizes of 5 and 10 percent but has lower rejection rate at the nominal size of 20 percent. The LM test performs better in term of finite sample power, except in a special case of the Markov-switching model in which the model becomes a Normal mixture model, where the LM test is inapplicable. For our test statistics, we have the exact asymptotic distribution, while Hansen (1992,1996) has only the asymptotic bound for his LR statistic. Also his test requires three-dimensional grid search while the tests here only need a one-dimensional grid search.

Finally, it should be noted that, testing the Markov regime switching models is only a special application of the tests, the framework is particularly applicable to a class of models with unobserved state variables.

8 Appendix A

Assumption A1

Regularity assumptions about the objective function Q_N:

For $\psi \equiv (\theta, \delta) \in \Psi$, $\gamma \in \Gamma$, assume

(1) $Q_N(\psi,\gamma) \rightarrow_p Q(\psi,\gamma)$ for all $(\psi,\gamma) \in (\Psi,\Gamma)$, where (Ψ,Γ) are compact set, and $Q(\psi,\Gamma) = \lim_{N\to\infty} EQ_N(\psi,\gamma)$ is continuous uniformly in (ψ,γ) over $\Psi \times \Gamma$.

(2) $Q_N(\psi,\gamma) - Q(\psi,\gamma)$ is stochastically equicontinuous in (ψ,γ) over (Ψ,Γ).

(3) For all $\gamma \in \Gamma$, $Q(\psi,\gamma)$ is uniquely minimized over $\psi \in \Psi$ at ψ_0.

Assumption A2

Regularity assumptions about the second order derivatives of $Q_N(\psi,\gamma)$ given γ:

Let N_{ψ_0} be an open neighborhood of ψ, given $\gamma \in \Gamma$, we assume

(1) $\lim_{N\to\infty} \frac{1}{N}\sum_{j=1}^{N} \frac{\partial F_j(\gamma)}{\partial \psi} \mid_{\psi_0} \frac{\partial F_j(\gamma)}{\partial \psi'} \mid_{\psi_0} \equiv C(\gamma) = \frac{1}{4}V(\gamma)$ exists, and is a finite nonsingular matrix. $V(\gamma)$ is defined in Assumption 2.

(2) $\frac{1}{N}\sum_{j=1}^{N} \frac{\partial F_j(\gamma)}{\partial \psi}\frac{\partial F_j(\gamma)}{\partial \psi'}$ converges to a finite matrix uniformly for all $\psi \in N_{\psi_0}$.

(3) $\frac{\partial^2 F_j(\psi,\gamma)}{\partial\psi\partial\psi'}$ is continuous in $(\psi,\gamma) \in N_{\psi_0} \times \Gamma$ uniformly in j.

(4) $\lim_{N\to\infty} \frac{1}{N^2}\sum_{j=1}^{N}\{\frac{\partial^2 F_j(\psi,\gamma)}{\partial\psi\partial\psi'}\}^2 = 0$ for all $(\psi,\gamma) \in N_{\psi_0} \times \Gamma$.

(5) $\frac{1}{N}\sum_{j=1}^{N} F_j(\psi_1,\gamma)\frac{\partial^2 F_j(\psi,\gamma)}{\partial\psi\partial\psi'} \mid_{\psi_2}$ converges to an finite matrix uniformly for all $(\psi_1,\psi_2) \in N_{\psi_0}$.

9 Appendix B: Proof of Theorem 1

First expand $Q_N(\psi,\gamma)$ around $\hat{\psi}$ given γ,

$$Q_N(\psi,\gamma) = Q_N(\hat{\psi},\gamma) + \frac{\partial Q_N(\gamma)}{\partial\psi} \mid_{\hat{\psi}} (\psi - \hat{\psi}) + \frac{1}{2}(\psi - \hat{\psi})' M_N(\gamma) \mid_{\bar{\psi}} (\psi - \hat{\psi}),$$

where $\bar{\psi}$ is a point between ψ and $\hat{\psi}$.

By definition, $\frac{\partial Q_N(\gamma)}{\partial\psi} \mid_{\hat{\psi}} = 0$. Thus:

$$Q_N(\tilde{\psi}) = Q_N(\tilde{\psi},\gamma) = Q_N(\hat{\psi},\gamma) + \frac{1}{2}(\tilde{\psi} - \hat{\psi})' M_N(\gamma) \mid_{\bar{\psi}} (\tilde{\psi} - \hat{\psi})$$

We then have:

$$N\frac{Q_N(\tilde{\psi}) - Q_N(\hat{\psi},\gamma)}{Q_N(\tilde{\psi})} = \frac{1}{2Q_N(\tilde{\psi})}\sqrt{N}(\tilde{\psi} - \hat{\psi})' M_N(\gamma) \mid_{\bar{\psi}} \sqrt{N}(\tilde{\psi} - \hat{\psi}) \tag{51}$$

Now, expanding the score

$$S^N(\psi,\gamma) = \sqrt{N}\frac{\partial Q_N(\psi,\gamma)}{\partial\psi}$$

around ψ_0, we have

$$S^N(\psi,\gamma) = S^N(\psi_0,\gamma) + \sqrt{N} M_N(\bar{\psi},\gamma)(\psi - \psi_0) \tag{52}$$

and

$$\begin{aligned} 0 &= S^N(\hat{\psi},\gamma) = S^N(\psi_0,\gamma) + \sqrt{N} M_N(\bar{\psi},\gamma)(\hat{\psi} - \psi_0) \\ \sqrt{N}(\hat{\psi} - \psi_0) &= -M_N(\bar{\psi},\gamma)^{-1} S^N(\gamma) \Rightarrow -M(\gamma)^{-1} S(\gamma) \quad as \quad N \rightarrow \infty \end{aligned} \tag{53}$$

As for the limit of $\sqrt{N}(\tilde{\psi} - \psi_0)$, note that $\tilde{\psi}$ is the solution when

$$Q_N(\psi,\gamma) - \lambda R'\psi$$

is minimized.

$\tilde{\psi}$ must satisfy:

$$\begin{aligned} \sqrt{N} S^N(\tilde{\psi},\gamma) - \lambda R' &= 0 \\ R'\tilde{\psi} &= 0 \end{aligned}$$

Expanding the last two equations around ψ_0, we have

$$\begin{aligned} \sqrt{N} S^N(\psi_0,\gamma) + M_N(\bar{\bar{\psi}},\gamma)\sqrt{N}(\tilde{\psi}-\psi_0) - \lambda R' &= 0 \\ R'(\tilde{\psi}-\psi_0) &= 0 \end{aligned}$$

and

$$\begin{bmatrix} -M_N(\bar{\bar{\psi}},\gamma) & \mathrm{R} \\ R' & 0 \end{bmatrix} \begin{bmatrix} \sqrt{N}(\tilde{\psi}-\psi_0) \\ \lambda' \end{bmatrix} = \begin{bmatrix} S^N(\psi_0,\gamma) \\ 0 \end{bmatrix}$$

Thus:

$$\begin{aligned} \sqrt{N}(\tilde{\psi}-\psi_0) &= -M_N(\gamma)^{-1}(I - R(R'M_N(\gamma)^{-1}R)^{-1}R'M_N(\gamma)^{-1})S^N(\gamma) \\ &\Rightarrow -M(\gamma)^{-1}(I - R(R'M(\gamma)^{-1}R)^{-1}R'M(\gamma)^{-1})S(\gamma) \qquad (54) \end{aligned}$$

From equations (56) and (57), we have

$$\begin{aligned} \sqrt{N}(\tilde{\psi}-\hat{\psi}_0) &= M_N(\gamma)^{-1}R(R'M_N(\gamma)^{-1}R)^{-1}R'M_N(\gamma)^{-1}S_N(\gamma) \\ &\Rightarrow M(\gamma)^{-1}R(R'M(\gamma)^{-1}R)^{-1}R'M(\gamma)^{-1}S(\gamma) \qquad (55) \end{aligned}$$

Plugging the above equation into equation (54), we get the result in Theorem 1 as desired. Q.E.D.

10 Appendix C: Proof of Theorem 2

For the LM test, we estimate the score under the null, and denote the score as $\tilde{S}^N(\gamma) = S^N(\tilde{\psi}, \gamma)$. From equations (55) and (57) we have:

$$\begin{aligned}
S^N(\tilde{\psi}, \gamma) &= S^N(\psi_0, \gamma) + \sqrt{N} M_N(\bar{\psi}, \gamma)(\tilde{\psi} - \psi_0) \\
&= S^N(\gamma) + M_N(\gamma) * \{-M_N(\gamma)^{-1}(I + R(R'M_N(\gamma)^{-1}R)^{-1}R'M_N(\gamma)^{-1})S^N(\gamma)\} \\
&= (R(R'M_N(\gamma)^{-1}R)^{-1}R'M_N(\gamma)^{-1})S^N(\gamma) \\
&= R(R'M_N(\gamma)^{-1}R)^{-1}\bar{S}^N(\gamma) \\
&\Rightarrow R(R'M(\gamma)^{-1}R)^{-1}\bar{S}(\gamma),
\end{aligned} \tag{56}$$

where

$$\bar{S}^N(\gamma) \equiv R'M_N(\gamma)^{-1}S^N(\gamma) \Rightarrow R'M(\gamma)^{-1}S(\gamma) \equiv \bar{S}(\gamma) \tag{57}$$

Thus:

$$\begin{aligned}
LM_N(\gamma) &= \tilde{S}^N(\gamma)'\tilde{V}(\gamma)^{-1}\tilde{S}^N(\gamma) \\
&\Rightarrow \bar{S}(\gamma)'(R'M(\gamma)^{-1}R)^{-1}R'V(\gamma)^{-1}R(R'M(\gamma)^{-1}R)^{-1}\bar{S}(\gamma)
\end{aligned} \tag{58}$$

Q.E.D.

11 Appendix D: Proof of Theorem 3

Given the assumptions A2 in Appendix A, one can show that

$$\begin{aligned} M(\gamma) &= plim\{M_N(\gamma)\} = plim\{\frac{\partial^2}{\partial\psi\partial\psi'}Q_N(\psi,\gamma)\} \\ &= \lim_{N\to\infty}\frac{2}{N}\sum_{j=1}^{N}\frac{\partial F_j}{\partial\psi}\mid_{\psi_0}\frac{\partial F_j}{\partial\psi'}\mid_{\psi_0} \\ &= \lim_{N\to\infty}\frac{1}{2}(G(\gamma)G(\gamma)') = \lim_{T\to\infty}\frac{1}{2\sigma_u^2}V_N(\gamma) = \frac{1}{2\sigma_u^2}V(\gamma) \end{aligned} \tag{59}$$

Now replace $V(\gamma)$ by $2\sigma_u^2 M(r)$ in Theorem 2 to get

$$\begin{aligned} LM_N(\gamma) &\Rightarrow \bar{S}(\gamma)'(R'M(\gamma)^{-1}R)^{-1}R'(2\sigma_u^2 M(\gamma))^{-1}R(R'M(\gamma)^{-1}R)^{-1}\bar{S}(\gamma) \\ &= \bar{S}(\gamma)'(2\sigma_u^2 R'M(\gamma)^{-1}R)^{-1}\bar{S}(\gamma) \end{aligned}$$

Thus, $LM = h(LM_N(\gamma))$ has the same asymptotic distribution as $DT = h(DT_N(\gamma))$.

Q.E.D.

12 Appendix E: Proof of Lemma 1

In this case,

$$\begin{aligned}
\lim_{N\to\infty} G^1(\gamma_1)'G^1(\gamma_2) &= \lim_{N\to\infty} \frac{4}{\sigma_1^4 N}\sum_{j=1}^{N} 1 = \frac{4}{\sigma_1^4} \\
\lim_{N\to\infty} G^1(\gamma_1)'G^2(\gamma_2) &= \lim_{N\to\infty} \frac{4}{\sigma_1^4 N}\sum_{j=1}^{N} \frac{1+\gamma_2}{1+\gamma_2^2-2\gamma_2 cos(w_j)} \\
&= \frac{4}{\pi\sigma_1^4}\int_0^{\pi} \frac{1+\gamma_2}{1+\gamma_2^2-2\gamma_2 cos(w)}dw \\
&= \frac{4}{2\pi\sigma_1^4}\int_{-\pi}^{\pi} \frac{1+\gamma_2}{1+\gamma_2^2-2\gamma_2 cos(w)}dw \\
\lim_{N\to\infty} G^2(\gamma_1)'G^1(\gamma_2) &= \frac{4}{2\pi\sigma_1^4}\int_{-\pi}^{\pi} \frac{1+\gamma_1}{1+\gamma_1^2-2\gamma_1 cos(w)}dw \\
\lim_{N\to\infty} G^2(\gamma_1)'G^2(\gamma_2) &= \lim_{N\to\infty} \frac{4}{\sigma_1^4 N}\sum_{j=1}^{N} (\frac{1+\gamma_1}{1+\gamma_1^2-2\gamma_1 cos(w_j)})(\frac{1+\gamma_2}{1+\gamma_2^2-2\gamma_2 cos(w_j)}) \\
&= \frac{4}{2\pi\sigma_1^4}\int_{-\pi}^{\pi} (\frac{1+\gamma_1}{1+\gamma_1^2-2\gamma_1 cos(w)})(\frac{1+\gamma_2}{1+\gamma_2^2-2\gamma_2 cos(w)})dw
\end{aligned}$$

To calculate the integral

$$I_1 = \int_{-\pi}^{\pi} \frac{1+\gamma}{1+\gamma^2-2\gamma cos(w)}dw$$

we set $z = e^{iw}$; thus $dw = -\frac{i}{z}dz$ and $2cos(w) = z + z^{-1}$, and we have the following contour integral along the unit circle in the complex domain,

$$\begin{aligned}
I_1 &= \oint \frac{1+\gamma}{1+\gamma^2-\gamma(z+z^{-1})}\frac{-I}{z}dz \\
&= I\frac{1+\gamma}{\gamma}\oint \frac{dz}{z^2-\frac{1+\gamma^2}{\gamma}z+1} \\
&= I\frac{1+\gamma}{\gamma}\oint \frac{dz}{(z-\gamma^{-1})(z-\gamma)}
\end{aligned}$$

Since $|\gamma| < 1$, by the residual theory, we have

$$\begin{aligned}
I_1 &= I\frac{1+\gamma}{\gamma} * 2\pi I * \lim_{z\to\gamma}\frac{1}{z-\gamma^{-1}} \\
&= 2\pi\frac{1+\gamma}{\gamma}\frac{\gamma}{1-\gamma^2} = \frac{2\pi}{1-\gamma}
\end{aligned}$$

Similarly,

$$\begin{aligned}
I_2 &= \int_{-\pi}^{\pi} (\frac{1+\gamma_1}{1+\gamma_1^2-2\gamma_1 cos(w)})(\frac{1+\gamma_2}{1+\gamma_2^2-2\gamma_2 cos(w_j)})dw \\
&= \oint (\frac{1+\gamma_1}{1+\gamma_1^2-\gamma_1(z+z^{-1})})(\frac{1+\gamma_2}{1+\gamma_2^2-\gamma_2(z+z^{-1})})\frac{-I}{z}dz \\
&= -I\frac{(1+\gamma_1)(1+\gamma_2)}{\gamma_1\gamma_2}\oint \frac{z}{(z^2-\frac{1+\gamma_1^2}{\gamma_1}z+1)(z^2-\frac{1+\gamma_2^2}{\gamma_2}z+1)}dz \\
&= I\frac{(1+\gamma_1)(1+\gamma_2)}{\gamma_1\gamma_2}\oint \frac{z}{(z-\gamma_1^{-1})(z-\gamma_1)(z-\gamma_2^{-1})(1-\gamma_2)}dz \\
&= I\frac{(1+\gamma_1)(1+\gamma_2)}{\gamma_1\gamma_2}\oint f(z)dz \\
where \quad f(z) &\equiv \frac{z}{(z-\gamma_1^{-1})(z-\gamma_1)(z-\gamma_2^{-1})(z-\gamma_2)}
\end{aligned}$$

Again, by the residual theory, we have

$$I_2 = -I\frac{(1+\gamma_1)(1+\gamma_2)}{\gamma_1\gamma_2} * 2\pi I * \sum_{k=1}^{K} \lim_{z\to z_k} \frac{1}{(m-1)!}\frac{\partial^{m-1}}{\partial z^{m-1}}(z-z_k)^m f(z)$$

where $z = z_k$ is the k'th singular point for f(z) inside the unit circle. Since all the singular points are of order one, and $m \geq 1$, we choose $m = 1$, and

$$\begin{aligned}
I_2 &= -I\frac{(1+\gamma_1)(1+\gamma_2)}{\gamma_1\gamma_2} * 2\pi I * \{\lim_{z\to\gamma_1} \frac{z}{(z-\gamma_1^{-1})(z-\gamma_2)(z-\gamma_2^{-1})} \\
&+ \lim_{z\to\gamma_2} \frac{z}{(z-\gamma_1^{-1})(z-\gamma_1)(z-\gamma_2^{-1})}\} \\
&= 2\pi\frac{1+\gamma_1\gamma_2}{(1-\gamma_1\gamma_2)(1-\gamma_1)(1-\gamma_2)}
\end{aligned}$$

Thus, we get the variance matrix in lemma 1.

13 Appendix F: Proof of Lemma 2

Note that in Theorem 1,

$$\bar{S}(\gamma) \equiv R'M(\gamma)^{-1}S(\gamma)$$

and

$$\begin{aligned} DT_N(\gamma) &\Rightarrow \frac{1}{2\sigma_u^2}\bar{S}(\gamma)'(R'M(\gamma)^{-1}R)^{-1}\bar{S}(\gamma) \\ &= \bar{S}(\gamma)'(2\sigma_u^2 R'M(\gamma)^{-1}R)^{-1}\bar{S}(\gamma) \end{aligned}$$

and

$$\begin{aligned} E(\bar{S}(\gamma_1)\bar{S}(\gamma_2)') &= R'M(\gamma_1)^{-1}E(S(\gamma_1)S(\gamma_2)')M(\gamma_2)^{-1}R \\ &= R'M(\gamma_1)^{-1}V(\gamma_1,\gamma_2)M(\gamma_2)^{-1}R \end{aligned}$$

Thus,

$$\begin{aligned} E(\bar{S}(\gamma)\bar{S}(\gamma)') &= R'M(\gamma)^{-1}V(\gamma,\gamma)M(\gamma)^{-1}R \\ &= 2\sigma_u^2 R'M(\gamma)^{-1}R \end{aligned}$$

Q.E.D.

14 Appendix G

The algorithm discussed in Garcia (1992) to simulate a Chi-square process with covariance function $\bar{V}(\gamma_1,\gamma_2)$ is the following:

(1) Select a set of N values in the parameter space Γ, say, $\gamma_1,\gamma_2,\ldots,\gamma_N$.

(2) Calculate the Cholesky decomposition P of the matrix:

$$\Sigma = \begin{bmatrix} \bar{V}(\gamma_1,\gamma_1) & \bar{V}(\gamma_1,\gamma_2) & \dots & \bar{V}(\gamma_1,\gamma_N) \\ \bar{V}(\gamma_2,\gamma_1) & \bar{V}(\gamma_2,\gamma_2) & \dots & \bar{V}(\gamma_2,\gamma_N) \\ \vdots & \vdots & \vdots & \vdots \\ \bar{V}(\gamma_N,\gamma_1) & \bar{V}(\gamma_N,\gamma_2) & \dots & \bar{V}(\gamma_N,\gamma_N) \end{bmatrix}$$

where $PP' = \Sigma$.

(3) Generate the Gaussian process as follows,

$$\begin{aligned} \bar{S}(\gamma_1) &= P_{11}\epsilon(1) \\ \bar{S}(\gamma_2) &= P_{21}\epsilon(1) + P_{22}\epsilon(2) \\ \vdots \quad & \quad \vdots \\ \bar{S}(\gamma_N) &= P_{N1}\epsilon(1) + P_{N2}\epsilon(2) + \ldots + P_{NN}\epsilon(N) \end{aligned}$$

where $\epsilon(I)$ are i.i.d N(0,1).

With the generated $\bar{S}(\gamma)$, find $\sup_{\gamma\in\{\gamma_1,\ldots,\gamma_N\}} T(\gamma)$, and repeat the process many times to generate a distribution of $\sup T(\gamma)$. Thus, we get the simulated asymptotic null distribution of the DT and LM tests.

% of Distribution	Asymptotic Distribution Critical Value	Approximate Dist $SMPL = 200$ Critical Value	Approximate Dist $SMPL = 1000$ Critical Value
99%	9.65	8.12	8.60
95%	6.27	6.14	6.48
90%	5.12	5.07	5.16
80%	3.75	3.84	3.96
70%	3.04	2.99	3.11
60%	2.44	2.44	2.40
50%	2.04	2.07	1.93
40%	1.63	1.68	1.55
30%	1.32	1.33	1.20
20%	1.02	0.97	0.94
10%	0.65	0.64	0.60
5%	0.43	0.42	0.45
1%	0.16	0.20	0.24

Table 1: Asymptotic and Approximate Finite Sample Null Distribution of the Tests For Model 1

Note: For Table 1, the approximate distribution is simulated by the normal parametric simulation as discussed in the paper. $\gamma \in [0.05, 0.95] + [-0.95, -0.05]$ with $grid = 0.01$

% of Distribution	Asymptotic Distribution CriticalValues	Approximate Dist by χ^2 $SMPL = 200$ Critical Values	Bootstrapped Distribution $SMPL = 200$ Critical Values
99%	9.65	10.51	11.44
95%	6.27	6.47	8.40
90%	5.12	5.13	6.83
80%	3.75	3.79	5.31
70%	3.04	3.01	4.28
60%	2.44	2.46	3.53
50%	2.04	2.04	3.03
40%	1.63	1.63	2.47
30%	1.32	1.32	2.02
20%	1.02	0.99	1.53
10%	0.65	0.67	1.04
5%	0.43	0.47	0.76
1%	0.16	0.27	0.41

Table 2: Comparison of Simulated and Bootstrapped Finite Sample Null Distribution For Model 1

Note: For Table 2, $\gamma \in [0.05, 0.95] + [-0.95, -0.05]$ with $grid = 0.01$

Actual Rejection Rates Under the null			
Nominal Size	$SMPL = 200$	$SMPL = 500$	$SMPL = 1000$
1%	4%	3%	1%
5%	12%	11%	8%
10%	17%	15%	13%
20%	26%	25%	22%

Table 3: The Rejection Rates under the Null Using Asymptotic Null Distribution

Note: For Table 3, the LM statistics are calculated by setting $\gamma \in [0.05, 0.95] + [-0.95, -0.05]$ with $grid = 0.01$. Rejection rates are calculated by 1000 replications of each sample

Significance Level	Type I Sample $SMPL = 200$	Type II Sample $SMPL = 200$	Type II Sample $SMPL = 500$	Type II Sample $SMPL = 1000$
1%	98%	88%	100%	100%
5%	100%	96%	100%	100%
10%	100%	97%	100%	100%
20%	100%	99%	100%	100%

Table 4: The Power of The LM Test Against Markov Switching Alternative For Model 1: Rejection Rates Using Asymptotic Null Distribution

Note: For Table 4, the LM statistics are calculated by setting $\gamma \in [0.05, 0.95] + [-0.95, -0.05]$ with $grid = 0.01$. Rejection rates are calculated by 1000 replications of each sample

Nominal size	20%	10%	5%
Size of the LM test	22.8%	13.5%	7.7%
Power of the LM test	95%	91%	88%
Size of Hansen's test	26%	10%	2%
Power of Hansen's test	86%	80%	74%

Table 5: Comparison of the Finite Sample Size and Power of the LM Test and Hansen's (1992) LR Bound Test For Model I

Note: in Table 5, for the LM test, the sample data is generated by using Hamilton's (1989) point estimates of the GNP model, setting the autoregressive parameters to zero. The sample length is 131. The power is calculated by 1000 replications. The size and power of Hansen's test is from Hansen (1995), Table IV. He uses the same DGP.

Significance Level	Type III Sample $SMPL = 200$	Type III Sample $SMPL = 500$	Type III Sample $SMPL = 1000$
1%	85%	99.5%	100%
5%	92%	100%	100%
10%	96%	100%	100%
20%	98%	100%	100%

Table 6: The Power of The LM Test Against Markov Switching Alternative For Model 1: Rejection Rates When the True DGP is Regime Switching both in Mean and Variance

Note: For Table 6, the LM statistics are calculated by setting $\gamma \in [0.05, 0.95] + [-0.95, -0.05]$ with $grid = 0.01$. Rejection rates are calculated by 1000 replications of each sample

Nominal Size	1%	5%	10%	20%
Actual Rejection Rates	6%	12%	18%	28%

Table 7: Actual rejection rates when sample is generated under the alternative with $p + q = 0.995$, or $\gamma = -0.005$, Sample Size is 200

Note: For Table 7, the LM statistics are calculated by setting $\gamma \in [0.05, 0.95] + [-0.95, -0.05]$ with $grid = 0.01$. Rejection rates are calculated by 1000 replications

Currency	LM statistic	$p-value$ by Asymptotic Distribution	$p-value$ by Bootstrapped Distribution
Engel & Hamilton (1990) Samples:73:IV-88:I			
British Pound	5.09	0.11	0.23
French Franc	2.38	0.42	0.61
German Mark	3.18	0.29	0.48
Expanded Samples: 73:IV-96:I			
British Pound	5.36	0.09	0.20
French Franc	7.30	0.03	0.07
German Mark	6.74	0.04	0.11

Table 8: The LM Test for $\alpha_1 = 0$ in Model I

Note: For Table 8, $\gamma \in [0.05, 0.95] + [-0.95, -0.05]$ with $grid = 0.01$

15 References

1. Amemiya, T., 1985, *Advanced Econometrics*, Cambridge MA: Harvard University Press.

2. Andrews, D. W. K., and W. Ploberger, 1992, "Optimal Tests when Nuisance Parameter Is Present Only Under the Alternative", Cowles Foundation Discussion Paper No. 1015.

3. Boldin, M. D., 1990, "Characterizing Business Cycles with a Markov Switching Model", Federal Reserve Bank of New York Research Paper, 1990-37.

4. Brillinger, D. R., 1981, *Time Series Analysis: Data Analysis and Theory*, New York: Holt, Rinehart, and Winston.

5. Brockwell, P. J., and R. A. Davis, 1995, *Time Series: Theory and Methods*, New York: Springer-Verlag.

6. Davies, R. B., 1977, "Hypothesis Testing When a Nuisance Parameter Is Present Only Under the Alternative", *Biometrika*, 64(2): 247-54.

7. Davies, R. B., 1987, "Hypothesis Testing When a Nuisance Parameter Is Present Only under the Alternative", *Biometrika*, 74(1): 33-43.

8. Davidson, R., and J.G. MacKinnon, 1981, "Several Tests for Model Specification in the Presence of Alternative Hypothesis", *Econometrica*, 49: 781-93.

9. Engel, C., and J. D. Hamilton, 1990, "Long Swings in the Dollar: Are They in the Data and Do Markets Know It?", *The American Economic Review*, 80(4): 689-713.

10. Franke, J., and W. Hardle, 1992, "On Bootstrapping Kernel Spectral Estimates", *The Annals of Statistics*, Vol. 20, No. 1, 121-45.

11. Gallant, A.R., 1977, "Testing for Nonlinear Regression Specification: A Non-regular Case", *Journal of the American Statistical Association*, 72: 523-30.

12. Garcia, R., and P. Perron, 1991, "An Analysis of the Real Interest Rate Under Regime Shifts", C.R.D.E. Discussion Paper, No. 2391, Universite de Montreal.

13. Garcia, R., 1992, "Asymptotic Null Distribution of the Likelihood Ratio test in Markov Switching Models", Working Paper, Universite de Montreal.

14. Hamilton, J. D., 1989, "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle", *Econometrica*, 57(2): 357-84.

15. Hamilton, J. D., 1996, "Specification Testing in Markov-Switching Time-Series Models", *Journal of Econometrics* 70, 127-57.

16. Hansen, B. E., 1991, "Inference When a Nuisance Parameter Is not Identified Under the Null Hypothesis", Mimeo, revised June 1993, Rochester University.

17. Hansen, B. E., 1992, "The Likelihood Ratio Test Under Nonstandard Conditions: Testing the Markov Switching Model of GNP", *Journal of Applied Econometrics*, 7: S62-S83.

18. Hansen, B. E., 1996, "Erratum: The Likelihood Ratio Test Under Nonstandard Conditions: Testing the Markov Switching Model of GNP", *Journal of Applied Econometrics*,Vol.11, 195-98.

19. Harvey, A. C., 1991, *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge: Cambridge University Press.

20. Huizinga, J., and F. S. Mishkin, 1986, "Monetary Policy Regime Shifts and the Unusual Behavior of Real Interest Rates", *Carnegie-Rochester Conference Series on Public Policy*, 24, 231-74.

21. Lee, L. F., and A. Chesher, 1986, "Specification Testing When Score Test Statistics are Identically Zero", *Journal of Econometrics*, 31, 121-49.

22. Pollard, D., 1984, *Convergence of Stochastic Processes*, New York: Springer-Verlag.

23. Pollard, D., 1990, *Empirical Processes: Theory and Applications*, Institute of Mathematical Statistics.

24. Roley, V. V., 1986, "The Response of Interest Rates to Money Announcements under Alternative Operating Procedures and Reserve Retirement Systems", NBER Working Paper 1812.

25. Rudebusch, G., 1995, "Federal Reserve Interest Rate Targeting, Rational Expectations, and the Term Structure", *Journal of Monetary Economics* Vol.35, No.2, 245-74.

26. Whittle, P., 1951, "Hypothesis Testing in Time Series Analysis", Thesis, Uppsala University, Almqvist and Wiksell, Uppsala, Hafner, New York.