WORKING PAPER SERIES

Markov Regime Switching and Unit Root Tests

Charles Nelson, Eric Zivot and
Jeremy M. Piger

Working Paper 2001-013A
http://research.stlouisfed.org/wp/2001/2001-013.pdf

October 2001

FEDERAL RESERVE BANK OF ST. LOUIS
Research Division
411 Locust Street
St. Louis, MO 63102

Markov Regime Switching and Unit Root Tests

Charles R. Nelson
University of Washington, Seattle, WA 98195
cnelson@u.washington.edu

Jeremy Piger
Federal Reserve Board, Washington, DC 20551
Jeremy.M.Piger@frb.gov

Eric Zivot
University of Washington, Seattle, WA 98195
ezivot@u.washington.edu

Abstract

We investigate the power and size performance of unit root tests when the data undergo Markov regime switching. All tests, including those robust to a single break in trend growth rate, have low power against a process with a Markov-switching trend. Under the null hypothesis, we find previously documented size distortions in Dickey-Fuller type tests caused by a single break in trend growth rate or variance do not generalize to most parameterizations of Markov switching in trend or variance. However, Markov switching in variance can lead to overrejection in tests allowing for a single break in the level of trend.

Keywords: Stochastic Trend, Deterministic Trend, Heteroskedasticity, Business Cycle Asymmetry

For the past 20 years the question of whether various economic time series have a unit root or are (trend) stationary has generated much research. Using standard tests many researchers are unable to reject the unit root null hypothesis for macroeconomic and financial time series such as GDP, interest rates, and exchange rates (Nelson and Plosser 1982). Perron (1989) argues that the evidence in favor of unit roots has been overstated, as standard tests have low power against trend stationary alternatives with structural breaks in trend level or growth rate. Perron remedied this problem by modifying the Augmented Dickey-Fuller test with dummy variables to account for a single structural break. Christiano (1992), Banerjee, Lumsdaine and Stock (1992), and Zivot and Andrews (1992) extend this methodology to endogenous estimation of the break date while Lumsdaine and Papell (1997) consider a test robust to two structural breaks. Hereafter we will refer to this class of tests as Perron-type tests. Leybourne, Mills and Newbold (1998) and Hamori and Tokihisa (1997) demonstrate a converse problem, that standard unit root tests reject too often when there is a single structural break in trend or variance under the null hypothesis.

While most of the literature has focused on the effects of a fixed number of structural breaks on unit root tests, there is a growing consensus that the number of regime changes in economic time series might be better modeled as arising from a probabilistic process. To this end, many authors have successfully used Hamilton’s (1989) Markov-switching model to capture regime change in a diverse set of macroeconomic and financial time series. It is thus natural to ask what effects Markov-switching regime change might have on unit root tests, including the Perron-type tests developed to mitigate the effects of a fixed number of structural breaks.

Examples where this issue might be relevant are not hard to find. Evans and Wachtel (1993) suggest an I(1) Markov-switching trend model for prices after standard unit root tests on the price level failed to reject. Garcia and Perron (1996) argue for an I(0) Markov-switching trend and variance model of inflation and real interest rates based on unit root tests performed by Perron (1990) suggesting these series were I(0) if one break in the level of trend is allowed. Finally, many studies that employ a Markov-switching variance or trend growth rate simply assume a unit root in the series of interest without any pretesting, most likely because unit root tests from previous studies suggest the series are I(1). Examples include Hamilton's original paper for GNP, Cecchetti and Mark (1990) for consumption and dividends, and Engel (1994) for the nominal exchange rate.

In this study we investigate the effects of several types of Markov regime switching on unit root tests, focusing on regime change in trend growth rate and variance, the form of structural change most often considered in the macroeconomics and finance literature. The literature surrounding structural breaks and unit root tests provides insight into the size and power effects of a fixed number of breaks in trend growth rate on standard unit root tests. However, it is not clear that these results generalize to the case of endogenous, Markov-switching breaks in trend. Perhaps the closest to addressing this question is Balke and Fomby (1991) who demonstrate that standard unit root tests continue to have low power when a series has endogenous, probabilistic breaks in trend growth rate. However, the process driving their breaks is an independent Bernoulli process, not a Markov-switching process, and they do not consider the performance of Perron-type tests. With regards to regime change in variance, several authors have considered the effects of GARCH type heteroskedasticity on unit root tests, for example Pantula (1988), Kim and Schmidt (1993), Seo (1999) and Hecq (1995), the latter considering the effects on Perron-type tests. However, the effects of Markov-switching in variance has not been considered. The only studies the authors are aware of investigating the effects of Markov regime change in a testing framework are Evans and Lewis (1993) and Hall, Psaradakis and Sola (1997) who conclude that Markov switching in trend growth rate or in the cointegrating vector will weaken the evidence in favor of cointegration in a bivariate system.

The paper is organized as follows: In Section 1 we evaluate the performance of unit root tests when the true data generating process undergoes regime switching in trend growth rate but is otherwise I(0). In line with previous literature, we find that standard unit root tests do a poor job of distinguishing this model from an integrated process. However, we also find that Perron-type tests have low power in this case. The Markov-switching trend model has often been used to model business cycle asymmetry. Thus, we also consider alternative Markov-switching models of business cycle asymmetry, in particular a model by Kim and Nelson (1999) which allows regime switching in the transitory component. Unit root tests have very good power against this generating process, indicating that the true nature of nonlinearities in the business cycle is very important for what effects these nonlinearities have on unit root tests. Finally, we briefly consider a model with Markov-switching autoregressive parameters. Such a model, with one regime an I(1) process and the other stationary, has been used by several authors, for example Ang and Bekaert (1998), to model interest rates. Standard tests have very low power against this process for empirically plausible parameterizations. In Section 2 we evaluate the performance of unit root tests when the true data generating process is I(1) in addition to the Markov switching. The size distortions pointed out in the literature for a single break in trend growth rate or variance do not generalize to most parameterizations of Markov switching. However, similar to the findings of Hecq (1995) for IGARCH errors, Markov switching in variance can cause significant overrejection in Perron-type tests that allow for a single structural break in level. Section 3 concludes.

1. THE POWER OF UNIT ROOT TESTS AGAINST REGIME-SWITCHING ALTERNATIVES

1.1 Regime Switching in the Trend Component

In this section we investigate the power of unit root tests, including Perron-type tests, when the true process is I(0) conditional on a Markov-switching trend growth rate. To begin, consider the following data generating process:

$$\begin{aligned} y_t &= \tau_t + c_t \\ \tau_t &= \mu_t + \tau_{t-1} \\ \mu_t &= \mu_1 S_t + \mu_0 (1 - S_t) \\ \phi(L) c_t &= \varepsilon_t, \varepsilon_t \sim i.i.d.(0, \sigma_\varepsilon^2) \end{aligned} \qquad (1)$$

where S_t is a discrete, unobserved state variable that takes on the value 0 or 1, τ_t is a trend component with a switching growth rate, and $\phi(L)$ is a lag polynomial with either all roots outside the unit circle or one root on the unit circle and the rest outside. In this paper we consider the case where S_t is first order Markov switching. Here, the value of S_t at time *t* depends only on its value at time *t-1*, such that $P\,(S_t = 1 \mid S_{t-1} = 1) = p_{11}$ and $P\,(S_t = 0 \mid S_{t-1} = 0) = p_{00}$.

The model in (1) is a version of the models given in Hamilton (1989) and Lam (1990). In Hamilton (1989), one root of $\phi(L)$ is restricted to unity, that is c_t has a stochastic trend. We will consider Hamilton's version of (1) in Section 2. Lam (1990) generalizes Hamilton's model to allow c_t to (possibly) be a stationary autoregressive process. In this section we consider the performance of unit root tests in this case, where all roots of $\phi(L)$ lie outside the unit circle. Here, innovations do not have permanent effects in the periods between shifts in the growth rate of trend. For some intuition into how unit root tests will perform at distinguishing this model from the I(1) null, consider the alternative representation of the Markov trend function, τ_t:

$$\tau_t = \tau_0 + \mu_0 * t + (\mu_1 - \mu_0)\sum_{j=1}^{t} S_j$$

Then, setting $\tau_0 = 0$:

$$\begin{aligned} y_t &= \mu_0 * t + RT_t + c_t \\ RT_t &= RT_{t-1} + (\mu_1 - \mu_0) * S_t \end{aligned} \tag{2}$$

Here y_t is written as the sum of a deterministic trend, $\mu_0 * t$, a stochastic trend, RT_t, and a stationary component, c_t. The stochastic trend is introduced because the effects of the discrete shocks from the switching trend, $(\mu_1 - \mu_0) * S_t$, are permanently reflected in the level of RT_t. This stochastic trend is different from an integrated process in the traditional sense in that it does not necessarily change each period. It is similar to the integrated case in that first differencing y_t eliminates the stochastic trend, leaving only a Markov switching mean.

To assess the power of unit root tests against the process given in (2) we perform Monte Carlo simulations for both standard and Perron-type unit root tests. We parameterize the experiments based on the observation that the tests should do a poor job of identifying the alternative given by (2) when the proportion of the variance of changes in y_t given by the stochastic trend, RT_t, is smaller rather than larger. The variance of innovations to RT_t is given by $(\mu_1 - \mu_0)^2 (p - p^2)$ where $p = E(S_t = 1) = \dfrac{1 - p_{00}}{2 - p_{00} - p_{11}}$. Lam (1990) finds that 37% of the variance of growth rates in real U.S. GNP is due to RT_t. We thus chose parameter values that will yield this 37% proportion when $p_{11} = 0.5$ and $p_{00} = 0.95$, the transition probability estimates found by Lam. These parameter values are $\mu_0 = 1$, $\mu_1 = -1.5$, $\sigma_\varepsilon^2 \sim N(0, 0.4)$, and $\phi(L) = 1$. For each unit root test 1000 Monte Carlo simulations were performed with two sample sizes, T = 200 and T = 500, and the initial values of S_t and y_t set equal to zero. To set p_{11} and p_{00} we appeal to an existing literature (Hamilton 1989, Lam 1990, Diebold and Rudebusch 1996, Engel 1994) which finds for various monthly and quarterly series that one state is highly persistent, generally having a transition probability above 0.9, while the other is somewhat less persistent, although still usually having a transition probability of 0.5 or greater. We thus consider the following values of p_{00}: 0.9, 0.95, 0.98 and of p_{11}: 0.5, 0.6, 0.7, 0.8, 0.9, 0.95, 0.98.

Augmented Dickey-Fuller Test.

We first consider the power of the Augmented Dickey Fuller, hereafter ADF, test (Dickey and Fuller 1979, Said and Dickey 1984) against the alternative hypothesis given in (2). We consider the ADF test based on the t-statistic associated with the null hypothesis $\rho = 1$ from the test regression:

$$y_t = c + \rho y_{t-1} + \beta t + \sum_{j=1}^{k} \phi_j \Delta y_{t-j} + \eta_t \tag{3}$$

with the lag length, k, chosen by the backward lag-length selection procedure given in Campbell and Perron (1991) with a maximum lag length, $\bar{k}$, set equal to the lower integer bound of $T^{1/3}$ as suggested by Said and Dickey (1984).

As would be expected from the existing literature, the ability of the ADF tests to distinguish the regime-switching trend stationary alternative given in (2) is quite poor. Table 1 shows the rejection probabilities for the 5% nominal size ADF test. For the T = 200 case the test never rejects above 35%, only rejects above 20% for 6 of the 21 combinations of the transition probabilities considered, and often rejects in the 5-10% range. The test tends to perform better when one transition probability dominates the other, for example, for the values of the transition

Table 1. Empirical Power of a 5% Augmented Dickey Fuller Test, True Process has Markov Switching in Trend Growth Rate

T							
200				500			
	Power				*Power*		
p_{11}	$p_{00} = 0.9$	$p_{00} = 0.95$	$p_{00} = 0.98$	p_{11}	$p_{00} = 0.9$	$p_{00} = 0.95$	$p_{00} = 0.98$
0.5	0.13	0.31	0.33	0.5	0.05	0.06	0.08
0.6	0.24	0.31	0.16	0.6	0.05	0.05	0.09
0.7	0.15	0.08	0.25	0.7	0.08	0.10	0.07
0.8	0.06	0.05	0.12	0.8	0.05	0.06	0.04
0.9	0.06	0.04	0.20	0.9	0.05	0.04	0.06
0.95	0.06	0.05	0.10	0.95	0.05	0.06	0.06
0.98	0.14	0.11	0.08	0.98	0.09	0.09	0.03

probabilities estimated by Lam (1990) for U.S. real GDP, $p_{00} = .95$ and $p_{11} = 0.5$, the test rejects with a 31% frequency. This is because the variance of innovations to the stochastic trend, RT_t, is smaller the larger the difference between the transition probabilities, that is $(\mu_1 - \mu_0)^2 (p - p^2)$ is a decreasing function of $|p_{00} - p_{11}|$. Intuitively, as one state becomes increasingly dominant, the process more closely resembles one with constant trend growth rate. For the larger sample size the ADF test has even lower power, rejecting at 10% or less frequency in all cases. This is not surprising as the larger sample size gives the ADF test more opportunity to detect the stochastic trend, RT_t.

Perron-Type Tests.

Since the influential work of Perron (1989) a large number of unit root tests that allow for structural breaks in trend growth rate or level under the alternative have been developed. The objective of this research program is to develop tests with higher power against broken-trend stationary alternatives. These tests are robust to a fixed number of structural breaks, usually one. However, there has been some argument in the literature that when there are multiple structural breaks in trend growth rate it may be sufficient to simply account for the largest of these breaks, see for example Garcia and Perron (1996), pg. 113. We are thus interested in whether such tests provide increased power against an alternative with a Markov-switching trend growth rate. Here we consider two such tests that assume a single break in the growth rate of the trend function occurring at an unknown date, one given in Perron (1994, 1997), hereafter the Perron test, and the other given in Zivot and Andrews (1992), hereafter the ZA test. The Perron test assumes a single break in trend growth rate under both the null and alternative hypothesis and specifies the break as an additive outlier, meaning the full effects of the break are immediately reflected. The test is based on the regressions in equations (3a) and (3b) of Perron (1997). The ZA test assumes a single break in trend growth rate under only the alternative hypothesis and specifies the break as an innovational outlier, meaning the full effects of the change are felt over time. The test is based on the regression in equation 2' in Zivot and Andrews (1992). For both tests the date of the structural break was estimated as the date that provides the most evidence against the null hypothesis, see Zivot and Andrews (1992) for details.

Tables 2-3 contain the rejection frequencies for 5% nominal size Perron and ZA tests. Interestingly, the Perron test performs worse than the ADF test for many of the cases considered. For example, when $T = 200$ the ADF test rejects more frequently for 17 of the 21 combinations of transition probabilities. For the transition probabilities estimated by Lam (1990) for real GDP, $p_{00} = .95$ and $p_{11} = 0.5$, the Perron test rejects 16% of the time vs. 31% for the ADF test. The ZA test performs somewhat better, rejecting more frequently than the ADF test for 18 of the 21 combinations of transition probabilities considered when $T = 200$. However, the difference is not decisive: in 10 of these 18 cases the ZA test is within 15% of the ADF test. In addition, the ZA test only rejects more than 40% of the time on four occasions and for over half the cases rejects at a less than 25% frequency. For the Lam (1990) transition probability estimates for real

GDP the ZA test rejects at a 19% frequency vs. 31% for the ADF test. When $T = 500$ the tests have even lower power, usually rejecting at close to their nominal size.

Table 2. Empirical Power of a 5% Perron (1994, 1997) Test, True Process has Markov Switching in Trend Growth Rate

T							
200				500			
	Power				*Power*		
p_{11}	$p_{00} = 0.9$	$p_{00} = 0.95$	$p_{00} = 0.98$	p_{11}	$p_{00} = 0.9$	$p_{00} = 0.95$	$p_{00} = 0.98$
0.5	0.14	0.16	0.22	0.5	0.05	0.06	0.07
0.6	0.13	0.15	0.24	0.6	0.03	0.05	0.06
0.7	0.10	0.11	0.19	0.7	0.04	0.10	0.08
0.8	0.05	0.05	0.12	0.8	0.03	0.05	0.05
0.9	0.04	0.03	0.09	0.9	0.05	0.06	0.03
0.95	0.05	0.03	0.09	0.95	0.04	0.04	0.05
0.98	0.11	0.09	0.17	0.98	0.06	0.06	0.02

Table 3. Empirical Power of a 5% Zivot-Andrews (1992) Test, True Process has Markov Switching in Trend Growth Rate

T							
200				500			
	Power				*Power*		
p_{11}	$p_{00} = 0.9$	$p_{00} = 0.95$	$p_{00} = 0.98$	p_{11}	$p_{00} = 0.9$	$p_{00} = 0.95$	$p_{00} = 0.98$
0.5	0.24	0.19	0.73	0.5	0.07	0.10	0.13
0.6	0.14	0.43	0.43	0.6	0.05	0.07	0.10
0.7	0.40	0.25	0.35	0.7	0.01	0.06	0.10
0.8	0.10	0.36	0.35	0.8	0.10	0.10	0.14
0.9	0.16	0.09	0.20	0.9	0.05	0.07	0.10
0.95	0.12	0.09	0.18	0.95	0.04	0.07	0.10
0.98	0.24	0.28	0.33	0.98	0.12	0.14	0.10

1.2 Regime Switching in the Transitory Component

Models with two-state Markov switching in trend growth rate, such as that discussed in the previous section, have been used extensively to model business cycle asymmetry. One reason for its popularity is the ability of a regime switching trend growth rate to capture the empirical observation that recessions are steeper and shorter than expansions. However, one implication of the two-state Markov-switching trend model is that recessions have permanent effects on the level of output, that is the economy never recovers output lost during a recession. Many authors have provided evidence that this implication is not consistent with the data, instead, following steep, short recessions the economy seems to undergo a high-growth recovery phase to gain back what was lost, see for example Friedman (1969, 1993), Wynne and Balke (1992, 1996), and Sichel (1994). In other words, the business cycle is better characterized with three phases rather than two. Recently, Kim and Nelson (1999) used Markov regime switching in the transitory component of real GDP to capture this pattern of business cycle asymmetry. Here we consider a trend stationary version of their model:

$$\begin{aligned} y_t &= \tau_t + c_t \\ \tau_t &= \mu + \tau_{t-1} \\ \phi(L)c_t &= \gamma * S_t + \varepsilon_t, \varepsilon_t \sim i.i.d.(0, \sigma_\varepsilon^2) \end{aligned} \tag{4}$$

where $\phi(L)$ has all roots outside the unit circle. Here, unlike the model in (1), the average growth rate of the deterministic trend, μ, is constant. Instead, regime switching occurs in the transitory component, c_t. If $\gamma < 0$, when S_t = 1 the level of the series is driven down into a steep recession. However, the recession is not permanent as past shocks from γ disappear through the autoregressive dynamics in the transitory component, causing a high growth recovery phase once S_t returns to zero. In the words of Friedman(1969, 1993), the economy is "plucked" downward during recession, bouncing back to trend following the recession.

The results of Kim and Nelson (1999) suggest that a model specifying recessions as "plucking" episodes provides as good as or better description of U.S. real GDP than a model with regime shifts in the trend component. However, given that the regime switching in (4) works through the transitory component we would expect unit root tests to have much better power against this alternative than the model in Section 1.1. To investigate this we perform a Monte Carlo experiment with the ADF test. We parameterize the simulation based on the percentage of the variance of c_t coming from the "plucks" γ. Kim and Nelson (1999) find this percentage to be approximately 80% for real GDP for estimated transition probabilities of $p_{11} = .95$ and $p_{00} = .70$. When $\phi(L) = 1$ this percentage is given by:

$$\frac{\gamma^2(p-p^2)}{\gamma^2(p-p^2)+\sigma_\varepsilon^2} \tag{5}$$

To meet the 80% metric when $p_{11} = .95$ and $p_{00} = .70$ we parameterize the simulation with $\gamma = -1.0$ and $\varepsilon_t \sim N(0,.04)$. We set $\mu = 0.8$, the average growth rate of real GDP over the Kim and Nelson sample. Again, we perform 1000 Monte Carlo trials for the same range of transition probabilities as in Section 1.1.

Table 4 contains the rejection frequencies for the 5% ADF test. As expected, the ADF test performs very well, rejecting at close to 100% for the most empirically relevant values of the transition probabilities. For example, for the estimated transition probabilities found by Kim and Nelson for real GDP, $p_{11} = .95$ and $p_{00} = .70$, the ADF test rejects at a 99% frequency when $T = 200$ and a 100% frequency when $T = 500$. The power remains above 50% in all but one of the 21 combinations considered for $T = 200$ and in all cases for $T = 500$.

The differing performance of unit root tests for the model in (1) vs. the model in (4) is important in answering the question of whether real GDP has a unit root. If we believe that business cycle nonlinearities are shifts in trend as in Lam (1990) these shifts will have significant deleterious effects on the power of unit root tests, including Perron-type tests. If however, these nonlinearities are better characterized as Friedman's "plucks" the power of unit root tests will be unaffected. Instead, the only remaining sort of structural change relevant to unit root tests will be long run breaks, such as the much discussed productivity slowdown. In this case Perron-type tests will still have an advantage over standard tests such as the ADF test. This points us to the importance of determining the true nature of business cycle nonlinearities for deciding what classes of unit root tests should be used in studies of real GDP.

Table 4. Empirical Power of a 5% Augmented Dickey Fuller Test, True Process has Markov Switching in the Transitory Component

		T					
	200				500		
	Power				*Power*		
p_{11}	$p_{00} = 0.9$	$p_{00} = 0.95$	$p_{00} = 0.98$	p_{11}	$p_{00} = 0.9$	$p_{00} = 0.95$	$p_{00} = 0.98$
0.5	1.00	1.00	1.00	0.5	1.00	1.00	1.00
0.6	1.00	1.00	0.99	0.6	1.00	1.00	1.00
0.7	1.00	0.99	0.99	0.7	1.00	1.00	1.00
0.8	0.99	0.98	0.97	0.8	1.00	1.00	1.00
0.9	0.95	0.87	0.79	0.9	1.00	1.00	0.99
0.95	0.88	0.70	0.53	0.95	1.00	0.99	0.94
0.98	0.81	0.52	0.30	0.98	1.00	0.94	0.62

1.3 Regime-Switching Autoregressive Coefficients

To this point we have investigated Markov switching taking the form of discrete disturbances to the trend or transitory component of a time series. Another popular formulation is Markov switching in the autoregressive parameters of a time series, an example of which is:

$$\begin{aligned} y_t &= \mu_t + \rho_t y_{t-1} + \varepsilon_t \\ \mu_t &= \mu_1 S_t + \mu_0 (1 - S_t) \\ \rho_t &= \rho_1 S_t + \rho_0 (1 - S_t) \\ \varepsilon_t &\sim i.i.d.(0, \sigma_{\varepsilon t}^2) \\ \sigma_{\varepsilon t}^2 &= \sigma_{\varepsilon 1}^2 S_t + \sigma_{\varepsilon 0}^2 (1 - S_t) \end{aligned} \tag{6}$$

In (6), y_t follows an AR(1) process in which the autoregressive parameter, the constant term, and the variance of the error term all switch between two regimes. A popular version of (6) in the empirical literature specifies y_t to be I(1) in one regime and I(0) in the other, for example $\rho_0 = 1$

and $|\rho_1| < 1$. Ang and Bekaert (1998) demonstrate that as long as the I(0) regime has positive probabilities of occurring and persisting, in this case $(1 - p_{00}) \neq 0$ and $p_{11} \neq 0$, y_t is covariance stationary. This occasionally integrated model has been usefully employed to model interest rates. For example, Ang and Bekaert (1998) point out that the U.S. Federal Reserve tends to move short term interest rates in a very persistent fashion during low inflation periods. However, during high inflation times Federal Reserve interest rate changes become less persistent and have higher variance.

For our purposes, we are interested in the ability of unit root tests to distinguish the occasionally integrated model from the I(1) null hypothesis. To investigate this issue we performed Monte Carlo simulations with the ADF test when the generating process is (6). We parameterize the Monte Carlo experiments to mimic the pattern of Federal Reserve interest rate movements discussed above. Thus, when $S_t = 0$ (low inflation times) y_t is a random walk with no drift, that is $\rho_0 = 1$, $\mu_0 = 0$, and $\varepsilon_t \mid S_t = 0 \sim N(0, .25)$. When $S_t = 1$ (high inflation times) y_t is a stationary AR(1) with positive mean and $\varepsilon_t \mid S_t = 1 \sim N(0, 2.0)$. One would expect that unit root tests would perform worse for more persistent values of the autoregressive parameter when $S_t = 1$. Thus, we consider three pairs of μ_1, ρ_1: (1.0, 0.8); (0.5, 0.9); (0.25, 0.95). In these pairs μ_1 is altered to maintain a constant mean of 5 for y_t in the stationary state.

Tables 5-7 present the Monte Carlo simulations for the three pairings of μ_1, ρ_1 and the sample sizes $T = 200$ and $T = 500$. As would be expected, the tests perform better as ρ_1 decreases, as p_{11} increases relative to p_{00} (the less time that is spent in the I(1) state), and the larger the sample size (the more data available for the test to detect the I(0) state). In general however the tests perform very poorly for empirically plausible parameterizations. Of the 63 power statistics reported for the $T = 200$ cases the test has power greater than 50% on only 3 occasions (all for the smallest value of ρ_1), and greater than 20% on only 17 occasions (10 of these for the smallest value of ρ_1). As the sample size increases the performance of the test is fairly good for the lowest value of ρ_1 considered but is still poor for larger values of ρ_1. For example, Ang and Bekaert (1998) show that the regime switches in U.S. interest rates roughly correspond to business cycle frequencies. Depending on the frequency of the data this corresponds to values of p_{00} between 0.9 and 0.95 and values of p_{11} between 0.5 and 0.9. For $\rho_1 = .9$ and $T = 500$ the ADF test has power greater than 40% over this range of transition probabilities on only one occasion.

Table 5. Empirical Power of a 5% Augmented Dickey Fuller Test, True Process has Markov Switching in the Autoregressive Parameters and $\mu_1 = 1.0, \rho_1 = 0.8$

T							
200				500			
	Power				*Power*		
p_{11}	$p_{00} = 0.9$	$p_{00} = 0.95$	$p_{00} = 0.98$	p_{11}	$p_{00} = 0.9$	$p_{00} = 0.95$	$p_{00} = 0.98$
0.5	0.12	0.09	0.08	0.5	0.43	0.14	0.08
0.6	0.17	0.09	0.09	0.6	0.50	0.19	0.09
0.7	0.20	0.12	0.09	0.7	0.63	0.26	0.10
0.8	0.28	0.17	0.12	0.8	0.74	0.35	0.13
0.9	0.48	0.30	0.19	0.9	0.93	0.64	0.23
0.95	0.64	0.47	0.27	0.95	0.99	0.83	0.41
0.98	0.81	0.66	0.47	0.98	1.00	0.96	0.73

Table 6. Empirical Power of a 5% Augmented Dickey Fuller Test, True Process has Markov Switching in the Autoregressive Parameters and $\mu_1 = 0.5, \rho_1 = 0.9$

T							
200				500			
	Power				*Power*		
p_{11}	p_{00} = 0.9	p_{00} = 0.95	p_{00} = 0.98	p_{11}	p_{00} = 0.9	p_{00} = 0.95	p_{00} = 0.98
0.5	0.08	0.07	0.07	0.5	0.15	0.09	0.05
0.6	0.10	0.10	0.07	0.6	0.19	0.09	0.07
0.7	0.11	0.10	0.10	0.7	0.27	0.13	0.09
0.8	0.15	0.12	0.09	0.8	0.39	0.17	0.10
0.9	0.22	0.17	0.13	0.9	0.64	0.35	0.16
0.95	0.32	0.27	0.21	0.95	0.86	0.61	0.32
0.98	0.42	0.37	0.29	0.98	0.95	0.85	0.59

Table 7. Empirical Power of a 5% Augmented Dickey Fuller Test, True Process has Markov Switching in the Autoregressive Parameters and $\mu_1 = 0.25, \rho_1 = 0.95$

T							
200				500			
	Power				*Power*		
p_{11}	p_{00} = 0.9	p_{00} = 0.95	p_{00} = 0.98	p_{11}	p_{00} = 0.9	p_{00} = 0.95	p_{00} = 0.98
0.5	0.07	0.06	0.08	0.5	0.10	0.08	0.06
0.6	0.09	0.08	0.09	0.6	0.09	0.07	0.05
0.7	0.09	0.08	0.09	0.7	0.12	0.08	0.06
0.8	0.09	0.11	0.11	0.8	0.15	0.11	0.09
0.9	0.11	0.13	0.10	0.9	0.29	0.15	0.13
0.95	0.17	0.14	0.15	0.95	0.44	0.27	0.19
0.98	0.17	0.17	0.17	0.98	0.59	0.45	0.34

2. REGIME-SWITCHING I(1) PROCESSES AND THE SIZE OF UNIT ROOT TESTS

2.1 Regime Switching in the Trend Component and Variance

In Section 1.1 we were interested in the ability of unit root tests to distinguish a process that was I(0) with a Markov-switching trend growth rate from an I(1) process. Here we will investigate what deleterious size effects a Markov-switching trend growth rate and variance in an otherwise I(1) process might have on unit root tests. Consider the following model motivated by Hamilton (1989):

$$
\begin{aligned}
y_t &= \tau_t + c_t \\
\tau_t &= \mu_t + \tau_{t-1} \\
\mu_t &= \mu_0(1-S_t) + \mu_1 S_t \\
\phi(L)c_t &= \varepsilon_t, \varepsilon_t \sim i.i.d.(0, \sigma_{\varepsilon t}^2) \\
\sigma_{\varepsilon t}^2 &= \sigma_{\varepsilon 1}^2 S_t + \sigma_{\varepsilon 0}^2 (1-S_t)
\end{aligned}
\tag{7}
$$

Again, S_t is first order Markov switching and τ_t is a deterministic trend component with a switching growth rate. As in Hamilton (1989) we specify $\phi(L)$ to have one root on the unit circle and all other roots outside the unit circle, so that shocks to y_t in between the Markov-switching trend breaks have permanent effects on the level of the series. We also allow the variance of the error term to undergo regime switching.

To simplify matters we set $\phi(L) = (1-L)$. The model in (7) can then be written with a constant growth rate and serially correlated, conditionally heteroskedastic errors:

$$
\begin{aligned}
\Delta y_t &= \mu + e_t \\
e_t &= (\mu_1 - \mu)S_t + (\mu_0 - \mu)(1-S_t) + \varepsilon_t
\end{aligned}
\tag{8}
$$

To make $E(e_t) = 0$ choose $\mu = (\mu_1 - \mu_0)p + \mu_0$. Substituting in the chosen expression for μ we arrive at the autocovariance function:

$$
\begin{aligned}
Cov(e_t, e_{t-k}) &= (\mu_1 - \mu_0)^2 E(S_t - p)(S_{t-k} - p) \\
&= (\mu_1 - \mu_0)^2 Cov(S_t, S_{t-k})
\end{aligned}
\tag{9}
$$

Also, conditional on S_t, e_t has a time varying variance due to the heteroskedasticiy of ε_t:

$$
Var(e_t \mid S_t = j) = \sigma_{\varepsilon j}^2, \; j = 0,1 \tag{10}
$$

A result from the theory of Markov processes tells us that $P(S_t = 1 \mid S_{t-k} = 1)$ and $P(S_t = 0 \mid S_{t-k} = 0)$ converge to the unconditional probabilities p and (1- p) at a geometric rate. Then, noting that $Cov(S_t S_{t-k}) = (p * P(S_t = 1 \mid S_{t-k} = 1) - p^2)$ we have $Cov(e_t, e_{t-k}) =$

$(\mu_1-\mu_0)^2(p*P(S_t=1|S_{t-k}=1)-p^2)\to 0$ geometrically. Thus, the model in (7) can be written with constant trend growth rate and errors exhibiting serial correlation that dies off geometrically. It should be noted that this result is entirely due to the modeling of breaks in the trend function as endogenous, probabilistic events. It does not hold true in models assuming a fixed number of structural breaks in trend growth rate such as the cases considered by Perron (1989) and Zivot and Andrews (1992) among others.

Several previous studies, for example Schwert (1989), investigate the properties of unit root tests under various forms of ARMA innovations. Therefore, we will find an ARMA process a useful alternative representation of e_t. Consider the following stationary AR(1) representation of S_t given by Hamilton (1989):

$$\begin{aligned} S_t &= (1-p_{00})+\theta S_{t-1}+\omega_t \\ \theta &= -1+p_{00}+p_{11} \end{aligned} \tag{11}$$

where, conditional on $S_{t-1}=1$, $\omega_t=(1-p_{11})$ with probability p_{11} and $\omega_t=-p_{11}$ with probability $1-p_{11}$ and conditional on $S_{t-1}=0$, $\omega_t=-(1-p_{00})$ with probability p_{00} and $\omega_t=p_{00}$ with probability $1-p_{00}$. Hamilton (1989) shows that the error term, ω_t, has $E(\omega_t)=0$, $E(\omega_t^2)=\sigma_\omega^2=p_{11}(1-p_{11})p+p_{00}(1-p_{00})(1-p)$, and is uncorrelated in that $E(\omega_t\,|\,\omega_{t-j})=0$ for all four possible values of ω_{t-j} and $j=1,2,.....$ Using (11) note that:

$$\begin{aligned} e_t-\theta e_{t-1} &= d+b\omega_t+\varepsilon_t-\theta\varepsilon_{t-1} \\ d &= (\mu_1-\mu)(1-p_{00})+(\mu_0-\mu)(1-p_{11}) \\ b &= (\mu_1-\mu_0) \end{aligned} \tag{12}$$

The term on the left hand side of (12) is an AR(1) while the term on the right hand side has the autocovariance function of an MA(1) in that it is zero after the first lag. Thus, e_t follows an ARMA(1,1) process.

To determine the effects of the regime switching in trend growth rate and variance on unit root tests we perform Monte Carlo experiments for the three tests discussed in Section 1: the ADF test, Perron test, and ZA test. We consider two cases, one in which there is only regime switching in trend growth rate and one in which there is only switching in variance. To parameterize the trend switching case we set the parameters to yield a specified amount of serial correlation as measured by the first order autocorrelation of e_t:

$$corr\,(e_t, e_{t-1}) = \frac{(\mu_1 - \mu_0)^2 (p * p_{11} - p^2)}{(\mu_1 - \mu_0)^2 (p - p^2) + p\sigma_{v1}^2 + (1-p)\sigma_{v0}^2} \tag{13}$$

where the denominator is the unconditional variance of e_t. We set $\mu_0 = 1$, $\mu_1 = 5$, $\varepsilon_t \sim N(0, \sigma_{\varepsilon t}^2)$, and $\sigma_{\varepsilon 0} = \sigma_{\varepsilon 1} = 1$ to yield a value of (13) equal to 0.50 for $p_{11} = 0.9$ and $p_{00} = 0.7$, the transition probability estimates for U.S. real GNP found by Hamilton (1989). This level of autocorrelation is similar to that found in the existing literature. For example, the value of (13) for U.S. real GNP reported by Hamilton is 0.38 while Engel's (1994) parameter estimates for the Japanese / French exchange rate suggest a value of (13) equal to 0.50.

For the variance switching case we set $\varepsilon_t \sim N(0, \sigma_{\varepsilon t}^2)$, $\dfrac{\sigma_{\varepsilon 1}}{\sigma_{\varepsilon 0}} = \dfrac{3}{1}$, and $\mu_0 = \mu_1 = 1.0$. This level of heteroskedasticity is quite reasonable for asset prices, for example Turner, Startz, and Nelson (1989) report $\dfrac{\sigma_{\varepsilon 1}}{\sigma_{\varepsilon 0}} = 2.6$ for stock returns while Engel (1994) reports much higher ratios for several U.S. exchange rates. However, this level of heteroskedasticity is overstated for series such as real GDP. Thus, our results for the switching variance case have more relevance for financial time series than for macroeconomic quantities. Again, we consider the same range of transition probabilities and sample sizes as in Section 1. Each Monte Carlo experiment is comprised of 1000 trials with initial values of S_t and y_t set equal to zero.

We begin by considering the effects of the Markov-switching trend growth rate in (7). Because this regime switching simply introduces serial correlation into an otherwise I(1) process we can appeal to the large literature evaluating the effects of serial correlation on unit root tests. Schwert (1989) demonstrates the ADF test performs well in the presence of ARMA errors such as those in (12). However, Leybourne, Mills, and Newbold (1998) show that the ADF test tends to overreject the null hypothesis when there is a single break in trend growth rate that occurs early in the sample. Thus, we expect the ADF test to overreject for parameterizations of (7) that yield few breaks, with one occurring early in the sample. The question of interest is for how broad a range of the Markov-switching parameterizations this result holds. Table 8 presents the rejection frequencies for the 5% ADF test. Note that only for $T = 200$ and $p_{11} = 0.98$ are the size distortions pointed out by Leybourne, Mills, and Newbold present. For most parameterizations the ADF test has size close to its nominal size and in general is slightly oversized. This is likely due to the Campbell-Perron lag selection procedure which has been documented by Hall (1994) to cause slight overrejection.

Table 8. Empirical Size of a 5% Augmented Dickey Fuller Test, True Process has Markov Switching in Trend Growth Rate

T							
200				500			
	Size				*Size*		
p_{11}	$p_{00} = 0.9$	$p_{00} = 0.95$	$p_{00} = 0.98$	p_{11}	$p_{00} = 0.9$	$p_{00} = 0.95$	$p_{00} = 0.98$
0.5	0.05	0.07	0.05	0.5	0.07	0.06	0.05
0.6	0.07	0.07	0.07	0.6	0.05	0.05	0.05
0.7	0.05	0.07	0.06	0.7	0.05	0.05	0.05
0.8	0.06	0.06	0.07	0.8	0.04	0.06	0.05
0.9	0.06	0.03	0.06	0.9	0.05	0.04	0.05
0.95	0.06	0.05	0.06	0.95	0.05	0.05	0.04
0.98	0.15	0.10	0.06	0.98	0.07	0.07	0.03

Next we consider the Perron and ZA test that allow for a single break in trend growth rate under the alternative. Table 9 contains the rejection frequencies for the 5% Perron test. The Perron test performs similarly to the ADF test for most parameterizations, which is not surprising given that it captures serial correlation in the same way as the ADF test. Notably, the Perron test performs better than the ADF test when $p_{11} = 0.98$. This is most likely because the Perron test is robust to a single break in trend growth rate under the null hypothesis as well as the alternative, making the Leybourne, Mills and Newbold (1998) critique not as relevant. Table 10 demonstrates that the ZA test can be significantly oversized when there are only a small number of breaks, that is for large values of p_{00} or p_{11}. This is because the distribution of the ZA test is derived assuming a null with no structural change, the presence of a small number of structural breaks under the null hypothesis will violate this null hypothesis and lead to overrejections. This issue is not as serious for the larger sample size, $T = 500$. Both the Perron and ZA test perform similarly to the ADF tests for this larger sample size.

Table 9. Empirical Size of a 5% Perron (1994, 1997) Test, True Process has Markov Switching in Trend Growth Rate

T							
200				500			
	Size				*Size*		
p_{11}	$p_{00} = 0.9$	$p_{00} = 0.95$	$p_{00} = 0.98$	p_{11}	$p_{00} = 0.9$	$p_{00} = 0.95$	$p_{00} = 0.98$
0.5	0.07	0.05	0.06	0.5	0.05	0.03	0.08
0.6	0.06	0.04	0.06	0.6	0.04	0.04	0.04
0.7	0.05	0.05	0.06	0.7	0.05	0.06	0.04
0.8	0.05	0.06	0.05	0.8	0.04	0.06	0.03
0.9	0.05	0.04	0.06	0.9	0.05	0.05	0.03
0.95	0.05	0.03	0.05	0.95	0.03	0.05	0.02
0.98	0.09	0.05	0.04	0.98	0.05	0.05	0.02

Table 10. Empirical Size of a 5% Zivot-Andrews (1992) Test, True Process has Markov Switching in Trend Growth Rate

T							
200				500			
	Size				*Size*		
p_{11}	$p_{00} = 0.9$	$p_{00} = 0.95$	$p_{00} = 0.98$	p_{11}	$p_{00} = 0.9$	$p_{00} = 0.95$	$p_{00} = 0.98$
0.5	0.09	0.10	0.09	0.5	0.07	0.04	0.05
0.6	0.08	0.09	0.10	0.6	0.07	0.06	0.07
0.7	0.09	0.10	0.13	0.7	0.06	0.09	0.07
0.8	0.09	0.09	0.13	0.8	0.05	0.08	0.09
0.9	0.08	0.09	0.15	0.9	0.11	0.08	0.09
0.95	0.11	0.10	0.19	0.95	0.05	0.07	0.08
0.98	0.24	0.23	0.29	0.98	0.10	0.10	0.08

We now move to the simulations investigating Markov switching in variance. Many authors have investigated the effects of various forms of heteroskedasticity on unit root tests, including Pantula (1988), Kim and Schmidt (1993), and Seo (1999). Provided that the heteroskedasticity meets certain conditions, given explicitly by Hamori and Tokihisa (1997), heteroskedasticity does not create size distortions for standard unit root tests. In Piger (2000) it is shown that Markov-switching heteroskedasticity meets these conditions, suggesting that standard unit root tests should perform well. However, we are still interested in investigating two scenarios. First, Hamori and Tokihisa (1997) have shown that a single break in variance causes Dickey-Fuller type tests to be oversized. Thus, we might expect that certain parameterizations of Markov switching in variance that yield a small number of breaks will cause size distortions in the ADF test. Table 11 demonstrates that this is not the case. The ADF test is reasonably sized for even large values of p_{00} and p_{11}, suggesting the result of Hamori and Tokihisa fades quickly when more than one break is allowed.

Table 11. Empirical Size of a 5% Augmented Dickey Fuller Test, True Process has Markov Switching in Variance

T							
200				500			
	Size				*Size*		
p_{11}	$p_{00} = 0.9$	$p_{00} = 0.95$	$p_{00} = 0.98$	p_{11}	$p_{00} = 0.9$	$p_{00} = 0.95$	$p_{00} = 0.98$
0.5	0.07	0.07	0.05	0.5	0.06	0.05	0.05
0.6	0.08	0.08	0.06	0.6	0.07	0.06	0.04
0.7	0.06	0.07	0.07	0.7	0.05	0.05	0.06
0.8	0.07	0.07	0.08	0.8	0.06	0.06	0.07
0.9	0.07	0.09	0.08	0.9	0.05	0.06	0.05
0.95	0.06	0.08	0.08	0.95	0.06	0.04	0.07
0.98	0.05	0.06	0.08	0.98	0.05	0.06	0.06

Secondly, Hecq (1995) points out for the case of IGARCH errors that periods of high and low variance in an integrated process can lead to the illusion of breaks in the level of trend. Tests that are robust to a structural break in level under the alternative can spuriously detect such breaks and overreject as a result. We thus might expect versions of the Perron and ZA test that allow for a break in the level of trend to be oversized in the presence of Markov-switching heteroskedasticity. To investigate this issue we consider the performance of the Perron test allowing for a single break in the level of trend under both the null and the alternative, based on equations (14) and (17) in Perron (1994), and the ZA test allowing for a single break in the level of trend under the alternative, given by equation 1' in Zivot and Andrews (1992). As Tables 12-13 make clear, the size distortions can be significant for certain parameterizations of p_{00} and p_{11}. For example, for $T = 200$ the 5% nominal size Perron test rejects at a greater than 10% frequency in all but one of the 21 combinations of transition probabilities considered and greater than 15% for 8 of the 21 combinations. The ZA test rejects at a greater than 10% frequency in all but one case and greater than 15% in more than half of the cases when $T = 200$. Both tests perform somewhat better when $T = 500$ but are still oversized.

Table 12. Empirical Size of a 5% Perron (1994, 1997) Test, True Process has Markov Switching in Variance

		T					
		200				500	
		Size				*Size*	
p_{11}	p_{00} = 0.9	p_{00} = 0.95	p_{00} = 0.98	p_{11}	p_{00} = 0.9	p_{00} = 0.95	p_{00} = 0.98
0.5	0.12	0.13	0.12	0.5	0.11	0.06	0.11
0.6	0.14	0.13	0.14	0.6	0.10	0.09	0.08
0.7	0.12	0.17	0.13	0.7	0.11	0.15	0.11
0.8	0.13	0.17	0.19	0.8	0.11	0.16	0.14
0.9	0.12	0.17	0.21	0.9	0.15	0.10	0.16
0.95	0.10	0.16	0.20	0.95	0.06	0.14	0.18
0.98	0.09	0.12	0.16	0.98	0.06	0.06	0.17

Table 13. Empirical Size of a 5% Zivot-Andrews (1992) Test, True Process has Markov Switching in Variance

		T					
		200				500	
		Size				*Size*	
p_{11}	p_{00} = 0.9	p_{00} = 0.95	p_{00} = 0.98	p_{11}	p_{00} = 0.9	p_{00} = 0.95	p_{00} = 0.98
0.5	0.15	0.14	0.14	0.5	0.07	0.11	0.09
0.6	0.15	0.17	0.15	0.6	0.11	0.10	0.11
0.7	0.16	0.19	0.16	0.7	0.08	0.16	0.14
0.8	0.15	0.22	0.21	0.8	0.10	0.20	0.15
0.9	0.14	0.21	0.24	0.9	0.13	0.16	0.20
0.95	0.10	0.17	0.24	0.95	0.10	0.13	0.16
0.98	0.08	0.12	0.19	0.98	0.04	0.11	0.16

2.2 Regime Switching in the Transitory Component

In Section 1.2 we discussed how different Markov-switching models of business cycle asymmetry can have very different implications for the effects of asymmetry on the power of unit root tests. Here we examine the difference this modeling choice has for the size of unit root tests. Consider the following I(1) version of the model presented in Section 1.2:

$$\begin{aligned} y_t &= \tau_t + z_t + c_t \\ \tau_t &= \mu + \tau_{t-1} \\ z_t &= z_{t-1} + v_t, v_t \sim i.i.d.(0, \sigma_v^2) \\ \phi(L)c_t &= \gamma * S_t + \varepsilon_t, \varepsilon_t \sim i.i.d.(0, \sigma_\varepsilon^2) \end{aligned} \tag{14}$$

where $\phi(L)$ has all roots outside the unit circle. Here y_t is the sum of a deterministic trend with constant drift, a random walk component, and a stationary autoregressive component that, assuming $\gamma < 0$, is "plucked" downward whenever $S_t = 1$. To see the effects the process in (14) might have on the size of unit root tests rewrite (14) in first differences assuming $\phi(L) = 1$:

$$\begin{aligned} \Delta y_t &= \mu + e_t^* \\ e_t^* &= v_t + \Delta\varepsilon_t + \gamma * \Delta S_t \end{aligned} \tag{15}$$

The process can thus be written in first differences with constant drift and an error term that is augmented by a Markov-switching component. The Markov switching introduces additional serial correlation into the process, namely the first difference of S_t. One interesting note is the similarity of this case to the additive outlier literature discussed by Franses and Haldrup (1994) among others. The parameter γ would correspond to an additive outlier in the case where S_t was serially uncorrelated as opposed to being a Markov-switching process. As Madalla and Yin (1997) and Vogelsang (1999) point out, the first difference of S_t in (15) would then introduce an MA(1) component into the first difference of y_t. For smaller values of γ the additional serial correlation introduced in both the Markov switching and additive outlier cases is captured by tests such as the ADF test and does not cause overrejections. However, as the size of γ increases, the contribution of the transitory component to the variance of Δy_t increases relative to the contribution of the stochastic trend component. This can eventually lead to spurious rejections from unit root tests if the variance of the transitory component begins to dominate. The question is whether parameterizations of (14) corresponding to U.S. business cycles generate such spurious rejections.

Table 14. Empirical Size of a 5% Augmented Dickey Fuller Test, True Process has Markov Switching in the Transitory Component

T							
200				500			
	Size				*Size*		
p_{11}	p_{00} = 0.9	p_{00} = 0.95	p_{00} = 0.98	p_{11}	p_{00} = 0.9	p_{00} = 0.95	p_{00} = 0.98
0.5	0.36	0.28	0.16	0.5	0.22	0.19	0.12
0.6	0.40	0.26	0.19	0.6	0.29	0.24	0.15
0.7	0.42	0.35	0.23	0.7	0.37	0.29	0.19
0.8	0.49	0.34	0.24	0.8	0.47	0.37	0.24
0.9	0.46	0.34	0.24	0.9	0.51	0.46	0.28
0.95	0.38	0.27	0.15	0.95	0.48	0.44	0.28
0.98	0.24	0.18	0.09	0.98	0.34	0.27	0.19

To investigate this issue we performed Monte Carlo experiments to investigate the performance of the ADF tests when the generating process is (14). We parameterize the simulation using parameter estimates from Kim and Nelson (1999) for U.S. real GDP. That is, we set $\mu = .8$, $v_t \sim N(0, 0.4)$, $\gamma = -1.1$, $\varepsilon_t \sim N(0, 0.04)$, and the lag order of $\phi(L)$ set equal to 2 with $\phi_1 = 1.26$ and $\phi_2 = -0.46$. Table 14 demonstrates that this level of "plucking" is indeed large enough to cause spurious rejections in the ADF test. These rejections are fairly severe, the 5% ADF test rejects at a more than 10% frequency for all but one of the combinations of the transition probabilities considered in Table 14. For $T = 200$ the rejections climb above 30% for 9 of the 21 cases while for $T = 500$ rejections are larger than 30% on 8 occasions. Again, this points out that whether nonlinearities in the U.S. business cycle take the form of shifts in trend or "plucks" in the transitory component can have large implications for the performance of unit root tests applied to U.S. output series.

3. CONCLUSION

We have investigated the performance of unit root tests when the true process undergoes various types of Markov-switching regime change. We consider both processes that are I(0) and I(1) in the periods between the regime switching. Our main findings are:

1) In line with previous literature, the Augmented Dickey-Fuller test does a poor job of distinguishing an I(0) process with Markov-switching breaks in trend growth rate from an I(1) process. Interestingly however, tests designed to be robust to a single structural break in trend growth rate under the alternative also have very low power in this case.
2) When the true process is I(1) and undergoes Markov switching in both trend growth rate and variance ADF tests have approximately the correct size for almost all combinations of transition probabilities. This demonstrates that studies documenting size distortions from a single break in trend growth and variance do not generalize to multiple, probabilistic breaks. Also, tests robust to a single break in level overreject the null hypothesis when there is Markov switching in variance.
3) When modeling business cycle asymmetry, an alternative to Markov switching in trend growth rate as in Lam (1990) is to allow for Markov-switching "plucks" in the transitory component of GDP as in Kim and Nelson (1999). The ADF test has good power when these "plucks" occur under the alternative hypothesis. However, the ADF test can be oversized when the regime switching occurs under the null, mainly because the "plucks" increase the contribution of the transitory component to the series. This demonstrates that the true nature of business cycle asymmetry has serious implications for the performance of unit root tests on output series.

Acknowledgements

Nelson and Piger acknowledge support from the Van Voorhis endowment at the University of Washington. Nelson and Zivot acknowledge support from the National Science Foundation under grant SBR-9711301 and Nelson from grant SES9818789. Piger acknowledges support from the Grover and Creta Ensley Fellowship at the University of Washington. The authors are grateful to Chang-Jin Kim, James Morley, Chris Murray, Dick Startz, participants at the 1999 NBER/NSF time series conference, and two anonymous referees for helpful comments. Responsibility for errors are entirely the authors'. This paper is based on chapter 1 of Piger's Ph.D. dissertation at the University of Washington. The views in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or of any other person associated with the Federal Reserve System.

References

Ang, A., and Bekaert, G. (1998), "Regime Switches in Interest Rates," Working Paper, NBER.

Balke, N.S., and Fomby, T.B. (1991), "Shifting Trends, Segmented Trends and Infrequent Permanent Shocks," *Journal of Monetary Economics*, 28, 61-85.

Banerjee, A., Lumsdaine, R.L., and Stock, J.H. (1992), "Recursive and Sequential Tests for a Unit Root: Theory and International Evidence," *Journal of Business and Economic Statistics*, 10, 271-287.

Campbell, J.Y., and Perron, P. (1991), "Pitfalls and Opportunities: What Macroeconomists Should Know about Unit Roots," *NBER Macroeconomics Annual*, 141-201.

Cecchetti, S.G., and Mark, N.C. (1990), "Evaluating Empirical Tests of Asset Pricing Models – Alternative Interpretations," *American Economic Review*, 80, 48-51.

Christiano, L.J. (1992), "Searching for a Break in GNP," *Journal of Business and Economic Statistics*, 10, 237-250.

Dickey, D.A., and Fuller, W.A. (1979), "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association*, 74, 427-31.

Diebold, F.X., and Rudebusch, G.D. (1996), "Measuring Business Cycles: A Modern Perspective," *The Review of Economics and Statistics*, 78, 67-77.

Engel, C. (1994), "Can the Markov Switching Model Forecast Exchange Rates?," *Journal of International Economics*, 36, 151-165.

Evans, M.D.D., and Lewis, K.K. (1993), "Trend in Excess Returns in Currency and Bond Markets," *European Economic Review*, 37, 1005-1019.

Evans, M.D.D., and Wachtel, P. (1993), "Were Price Changes During the Great-Depression Anticipated? Evidence from Nominal Interest Rates," *Journal of Monetary Economics*, 32, 3-34.

Franses, P.H., and Haldrup, N. (1994), "The Effects of Additive Outliers on Tests for Unit Roots and Cointegration," *Journal of Business and Economic Statistics*, 12, 471-478.

Friedman, M. (1969), *The Optimum Quantity of Money and Other Essays*, Chicago: Aldine.

------ (1993), "The "Plucking Model" of Business Fluctuations Revisited," *Economic Inquiry*, 31, 171-177.

Garcia, R., and Perron, P. (1996), "An Analysis of the Real Interest Rate under Regime Shifts," *Review of Economics and Statistics*, 78, 111-125.

Hall, A. (1994), “Testing for a Unit Root in Time Series with Pretest Data-Based Model Selection,” *Journal of Business & Economic Statistics,* 12, 461-470.

Hall, S., Psaradakis, Z. and Sola, M. (1997), “Cointegration and Changes in Regime: The Japanese Consumption Function,” *Journal of Applied Econometrics*, 12, 151-168.

Hamilton, J.D. (1989), “A New Approach to the Economic Analysis of Non-stationary Time Series and the Business Cycle,” *Econometrica*, 57, 357-384.

Hamori, S., and Tokihisa, A. (1997), “Testing for a Unit Root in the Presence of a Variance Shift,” *Economics Letters*, 57, 245-253.

Hecq, A. (1995), “Unit Root Tests with Level Shift in the Presence of GARCH,” *Economics Letters*, 49, 125-130.

Kim, C.-J., and Nelson, C.R. (1999), “Friedman’s Plucking Model of Business Fluctuations: Tests and Estimates of Permanent and Transitory Components,” *Journal of Money, Credit and Banking*, 31, 317-34.

Kim, K., and Schmidt, P. (1993), “Unit Root Tests with Conditional Heteroskedasticity,” *Journal of Econometrics*, 59, 287-300.

Lam, P.-S. (1990), “The Hamilton Model with a General Autoregressive Component,” *Journal of Monetary Economics*, 26, 409-432.

Leybourne, S.J., Mills, T.C., and Newbold, P. (1998), “Spurious Rejections by Dickey-Fuller Tests in the Presence of a Break Under the Null,” *Journal of Econometrics,* 87, 191-203.

Lumsdaine, R.L., and Papell, D.H. (1997), “Multiple Trend Breaks and the Unit Root Hypothesis,” *The Review of Economics and Statistics*, 79, 212-218.

Maddala, G.S. and Y. Yin (1997), “Outliers, Unit Roots and Robust Estimation of Nonstationary Time Series,” in *Handbook of Statistics* (Vol. 15), eds. G.S. Maddala and C.R. Rao, Amsterdam: Elsevier, pp. 237-266.

Nelson, C.R., and Plosser, C.I. (1982), “Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications,” *Journal of Monetary Economics,* 10, 139-162.

Pantula, S.G. (1988), “Estimation of Autoregressive Models with ARCH Errors,” *Sankhya B,* 50, 119-138.

Perron, P. (1989), “The Great Crash, the Oil Price Shock and the Unit Root Hypothesis,” *Econometrica*, 57, 1361-1401.

------ (1990), “Testing for a Unit Root in a Time-Series with a Changing Mean,” *Journal of Business and Economic Statistics,* 8, 153-162.

------ (1994), “Trend, Unit Root and Structural Change in Macroeconomic Time Series,” In *Cointegration for the Applied Economist*, ed. B.B. Rao, New York: St. Martin’s Press, 113-146.

------ (1997), “Further Evidence on Breaking Trend Functions in Macroeconomic Variables,” *Journal of Econometrics*, 80, 355-385.

Piger, J. (2000), “Essays on Business Cycle Asymmetry,” Ph.D. Dissertation, University of Washington, Dept. of Economics.

Said, S.E., and Dickey, D.A. (1984), “Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order,” *Biometrika,* 71, 599-607.

Schwert, W.G. (1989), “Tests for Unit Roots: A Monte Carlo Investigation,” *Journal of Business and Economic Statistics,* 7, 147-159.

Seo, B. (1999), “Distribution Theory for Unit Root Tests with Conditional Heteroskedasticity,”. *Journal of Econometrics*, 91, 113-144.

Sichel, D. E. (1994), “Inventories and the Three Phases of the Business Cycle,” *Journal of Business and Economic Statistics*, 12, 269-277.

Turner, C.M., Startz, R., and Nelson, C.R. (1989), “A Markov Model of Heteroskedasticity, Risk, and Learning in the Stock Market,” *Journal of Financial Economics*, 25, 3-22.

Vogelsang, T. (1999), “Two Simple Procedures for Testing for a Unit Root When There are Additive Outliers” *Journal of Time Series Analysis*, 20, 237-252.

Wynne, M.A., and Balke, N.S. (1992), “Are Deep Recessions Followed by Strong Recoveries?,” *Economics Letters*, 39, 183-189.

Wynne, M.A., and Balke, N.S. (1996), “Are Deep Recessions Followed by Strong Recoveries? Results for the G-7 Countries,” *Applied Economics*, 28, 889-897.

Zivot, E., and Andrews, D.W.K. (1992), “Further Evidence on the Great Crash, the Oil Price Shock, and the Unit Root Hypothesis,” *Journal of Business and Economic Statistics,* 10, 251-270.