Studies in Nonlinear Dynamics and Econometrics

Quarterly Journal
April 1998, Volume 3, Number 1
The MIT Press

Studies in Nonlinear Dynamics and Econometrics (ISSN 1081-1826) is a quarterly journal published electronically on the Internet by The MIT Press, Cambridge, Massachusetts, 02142. Subscriptions and address changes should be addressed to MIT Press Journals, Five Cambridge Center, Cambridge, MA 02142; tel.: (617) 253-2889; fax: (617) 577-1545; e-mail: journals-orders@mit.edu. Subscription rates are: Individuals $40.00, Institutions $130.00. Canadians add 7% GST. Prices subject to change without notice.

Avoiding the Pitfalls: Can Regime-Switching Tests Reliably Detect Bubbles?

Simon van Norden

Ecole des Hautes Etudes Commerciales
3000 Chemin de la Côte-Ste.-Catherine
Montreal, Quebec, Canada H3T 2A7
svn@alum.mit.edu

Robert Vigfusson

Department of Economics
Northwestern University
Evanston, Illinois, U.S.A.
r-vigfusson@nwu.edu

Abstract. *Our paper uses simulation methods to examine the size and power of regime-switching tests for bubbles. We find that even with several hundred observations, the tests show sometimes considerable size distortion. This distortion makes the tests conservative; they understate the significance of the evidence of bubbles. Despite this, the tests display considerable power to detect bubbles even when using the conservative asymptotic critical values. We also find that the frequency with which bubbles collapse has an important influence on the tests' power. An application to monthly Canadian and American stock-price data provides mixed evidence of bubbles.*

Keywords. bubbles, regime switching, stock market, crash, multiple equilibria

Acknowledgments. While retaining sole responsibility for any errors or oversights, the authors would like to thank their colleagues at the Bank of Canada, the editor, an anonymous referee, and seminar participants at the 1996 meetings of the SNDE and the CEA for their comments and suggestions.

1 Introduction

Stock market crashes, such as the one that took place in October 1997 in Asian markets, or the equally serious crash in Japan 10 years before, are an important puzzle. They are important because of their sometimes large and persistent macroeconomic effects, but puzzling owing to the difficulty of reconciling such sudden shifts in asset prices with observable market fundamentals. One approach has been to argue that multiple equilibria are possible, and that crashes simply represent a move between alternative equilibrium prices for a given set of fundamentals. Since the early 1980s, most of the discussion has centered on the possibility of rational speculative bubbles. While demonstrating the presence of such bubbles is problematic, the question has prompted much empirical work and debate.

Early work testing for the presence of such bubbles used variance-bound tests, until various econometric problems with this approach were noted (see LeRoy 1989). The misspecification test suggested by West (1987) has fallen out of favor, because misspecified fundamentals can cause it to detect bubbles, and there is little agreement on how to specify the fundamentals. Diba and Grossman (1988) and Hamilton and Whiteman (1985) recommend the use of tests for stationarity and for cointegration to test for the absence of rational speculative bubbles. However, Monte Carlo simulations reported in Evans (1991) show that standard tests for unit roots and cointegration frequently reject the presence of bubbles, even when such bubbles are present by construction.[1] Evans refers to this problem as the "pitfall" of testing for bubbles.

Since Evans's note, new tests for rational speculative bubbles that rely on regime switching have been proposed. Van Norden and Schaller (1993, 1996) and van Norden (1996) use a switching regression to look for a time-varying relationship between returns and deviations from an approximate fundamental price. Hall and Sola (1993) and Funke, Hall, and Sola (1994) test whether asset prices seem to switch between explosive growth and stationary behavior.

Given the problems that have plagued earlier methods of testing for bubbles, the robustness and reliability of these regime-switching methods should be scrutinized. Among the plausible grounds for concern are:

- The presence of rational speculative bubbles implies that the data are nonstationary, but the properties of regime-switching estimators in this instance are unknown. Since nonstationarity exists only under the alternative hypothesis of bubbles, this raises the question of whether the regime-switching tests have the power to detect bubbles when they exist. This is similar to the pitfall that Evans (1991) found with the unit-root and cointegration tests.[2]

- Little is known about the finite-sample properties of regime-switching estimators. In particular, little has been done to determine whether the use of tests whose distribution is known only asymptotically leads to reliable inference. It is conceivable that asymptotically correct tests could experience size distortion in small samples, which would tend to produce evidence of speculative bubbles even when none is present.

Our paper is a first step in addressing these questions. We examine the size and power properties of regime-switching bubble tests in finite samples using simulation methods. Our work relies on publicly available programs that use a combination of EM and gradient-based maximum-likelihood algorithms for fast and fairly robust estimation that makes simulation studies practical.[3] We are also the first to compare and contrast the performance of the two different kinds of regime-switching tests.

While we present results on the power and size of both tests, we should note at the outset some of the limitations of our study. The most important of these is its specificity. Like all simulation studies, the power and size results will be functions of the data and data generating processes (DGPs) we consider. For our study of power, we simulate asset bubbles using the same DGP used by Evans (1991). To study size, we apply a randomization strategy to post-WWII monthly data on stock prices. We choose these data sets and DGPs because they have been studied elsewhere and so our work may give useful benchmarks. We also think that these are reasonable (and, we hope, typical) examples of the kind of data to which these tests are commonly applied. This in no way means that our results are universally applicable.[4]

Our study is focused on the econometric aspects of these tests; therefore, an important issue that we do not investigate is that of "economic" size. In other words, we confine ourselves to studying how well these tests can discriminate between the presence and absence of certain kinds of regime switching. We do not investigate whether such regime switching is necessarily evidence of bubbles. This is because we think the limitations of econometric testing for bubbles are well understood and can be summarized as follows.

It has been generally accepted that tests of the null hypothesis of no bubbles will everywhere and always be joint tests of (1) the hypothesis that there are no bubbles, and (2) that the model of fundamentals is correctly specified. Put another way, we can always find an alternative model of fundamentals that will be

[1]Charemza and Deadman (1995) show that this problem extends to a broader range of processes than those considered by Evans (1991).

[2]Both problems could lead us to conclude that bubbles are absent when they are in fact present. The difference is that with the cointegration and unit-root tests, this result is caused by size distortion, while with the regime-switching tests it is caused by a lack of power. This difference arises because the two kinds of tests reverse the null and alternative hypotheses.

[3]Details on the programs may be found in van Norden and Vigfusson (1996). The GAUSS programs may be downloaded from *http://mitpress.mit.edu/e-journals/SNDE/003/articles/vannorden_datacode.zip*. The van Norden and Vigfusson paper may also be downloaded from this site.

[4]Of course, our methodology is much more generally applicable, since one could apply our simulation experiments to any data set of interest.

observationally equivalent to bubbles. Statistically, this means that any test of the null hypothesis of no bubbles will have severe size distortion for at least some models of fundamentals.[5] In evaluating statistical evidence of bubbles, therefore, we must always consider how reasonable the model of fundamental prices appears. The results of this study give no guidance on this issue.

In the following section, we explain the relationship between speculative bubbles and regime switching, and then review the tests proposed by Hall and Sola and by van Norden. Section 3 examines the size of these tests in the context of their application to a particular data set; one on North American stock market monthly returns in the post-WWII period. After applying the tests to the data as a benchmark, we do randomization experiments to determine appropriate finite-sample critical values, and use them to reconsider the evidence of bubbles. Section 4 considers the power of these tests by simulating the Evans (1991) DGP and examining the ability of each test to find significant evidence of bubbles. Section 5 concludes.

2 Tests for Speculative Bubbles

This section has three goals. We first describe what a bubble is. We next describe the two regime-switching tests used in this paper to detect bubbles. Finally, we discuss the relationship between these two tests.

2.1 Bubbles and Regime Switching

Consider a simple asset-pricing model, which only requires that

$$p_t = f(X_t) + a \cdot E_t(p_{t+1}), \tag{1}$$

where p_t is the logarithm of the asset price, E_t is the operator for expectations conditional on information at time t, $0 < a < 1$, and X_t is a vector of other variables. Solving the equation forward gives the general result

$$p_t = \left(\sum_{j=0}^{T} a^j \cdot E_t(f(X_{t+j})) \right) + a^{T+1} \cdot E_t(p_{T+1}). \tag{2}$$

One solution to Equation 1, which we will denote p_t^*, occurs when

$$\lim_{T \to \infty} a^{T+1} \cdot E_t(p_{T+1}) = 0, \tag{3}$$

so

$$p_t^* = \sum_{j=0}^{\infty} a^j \cdot E_t(f(X_{t+j})). \tag{4}$$

We refer to Equation 4 as the fundamental solution, since it determines the asset price solely as a function of the current and expected behavior of other variables.

However, Equation 4 is not the only solution to Equation 1, since such solutions need not satisfy Equation 3. We define bubble solutions to be any other set of asset prices and expected asset prices that satisfies Equation 1 but where Equation 3 does not hold, so $p_t \neq p_t^*$. We define the size of the bubble B_t as

$$B_t \equiv p_t - p_t^*. \tag{5}$$

[5] See the overview by Flood and Hodrick (1990) for a nontechnical exposition and additional references. Van Norden and Schaller (1996) present Monte Carlo evidence that shows that regime switching in fundamentals can make the van Norden bubble test mistakenly indicate the presence of bubbles.

Note that since p_t^* satisfies Equation 1, it follows[6] from Equations 1 and 5 that

$$B_t = a \cdot E_t(B_{t+1}). \tag{6}$$

Since $a < 1$, the bubble must be expected to grow over time.[7]

Nothing in the above model has any implications for regime switching. (Some of the early literature on rational speculative bubbles even considered purely deterministic bubbles.) Regime switching stems from descriptions of asset-market behavior (for example, those surveyed in Kindleberger 1989) to which the above model of bubbles is often applied. The first example of regime switching in the rational speculative bubble framework is given by Blanchard (1979), who proposes a bubble that moves randomly between two states, C and S. In state C, the bubble will collapse, so[8]

$$E_t(B_{t+1} \mid C) = 0. \tag{7}$$

State S, where the bubble survives and continues to grow, occurs with a fixed probability q. Since

$$E_t(B_{t+1}) = (1 - q) \cdot E_t(B_{t+1} \mid C) + q \cdot E_t(B_{t+1} \mid S), \tag{8}$$

it follows from Equation 6 that

$$E_t(B_{t+1} \mid S) = \frac{B_t \cdot (1 + r)}{q}. \tag{9}$$

This model was subsequently generalized by Evans (1991) and van Norden and Schaller (1993) to consider the case where both the size of collapses and their probability were functions of the size of the bubble.

The distinguishing feature of these regime-switching models is that the behavior of the asset price is now state-dependent, and that the state itself is unobservable. However, these models may differ in the way the probability of observing a given regime varies over time. In Blanchard (1979), this is simply a constant. In the van Norden bubble test, the probability of observing the collapsing regime is assumed to be an increasing function of the size of the bubble. In the Hall and Sola test, this probability is assumed to follow a first-order Markov process, where the probability of remaining in a given regime is constant.[9] To distinguish the latter two kinds of switching models, we will refer to the case where the probability of observing a given state is independent of past states as "simple switching." In the case of a two-state model, the simple-switching model is simply the special case of the Markov-switching model, where

$$Pr(S_t = 0 \mid (S_{t-1} = 0)) = 1 - Pr(S_t = 1 \mid (S_{t-1} = 1)), \tag{10}$$

where $Pr(S_t = k \mid S_{t-1} = k)$ is the probability of remaining in state k, given that the last period's state was k.

As we detail below, the two regime-switching tests for bubbles do not test the null hypothesis of no regime switching against the alternative of regime switching. Instead, they test whether specific relationships differ significantly across the two regimes. To be sure, the hypothesis that there is no regime switching is encompassed by the null hypothesis that there is no significant difference in these relationships across the two regimes. Therefore, if we can reject the null that the relationships are the same across regimes, we can also

---

[6]Blanchard (1979) has a more complete derivation of this and subsequent steps found in this section.

[7]A considerable literature exists on the conditions under which such bubbles are feasible rational expectations solutions. Important contributions to this debate have been made by Obstfeld and Rogoff (1983, 1986), Diba and Grossman (1987), Tirole (1982, 1985), Weil (1990), Buiter and Pesenti (1990), Allen and Gorton (1991), and Gilles and LeRoy (1992). In single-representative-agent models, a truly rational agent cannot expect to sell an over-valued asset (one with a positive bubble) before the bubble bursts. Therefore, bubbles should exist in such models only if they can be expected to grow without limit. Some researchers, such as Froot and Obstfeld (1991), have therefore suggested interpreting empirical tests for bubbles as tests of whether agents are fully rational, or whether they instead exhibit some form of myopia when considering events that are either very far in the future or of very low probability. An alternative interpretation would be to consider evidence of bubbles as suggesting that non-representative-agent models (such as those of De Long et al. 1990, Allen and Gorton 1991, or Bulow and Klemperer 1994) are required.

[8]The notation $E_t(X_j \mid C)$ (or $E_t(X_j \mid S)$) denotes the expectation of X_j conditional on the fact that the state at $t+1$ is C (or S) and on all other information available at time t.

[9]As noted by Evans and Lewis (1995), a two-state first-order Markov process is not compatible with Equation 6. They reconcile this by modifying the usual two-state Markov model to allow for jumps in asset prices when the regime changes.

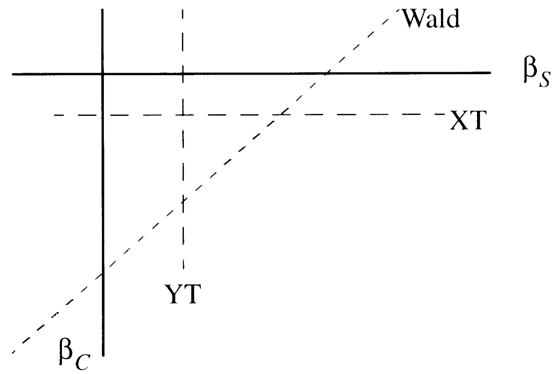

Figure 1
Comparison of Wald and t-test.

reject the null that there is no regime switching. By using a null hypothesis that is more general than simply the absence of regime switching, these tests should be more conservative (less likely to reject the null) than tests for the number of regimes.[10]

2.2 The Hall and Sola Test for Bubbles

Diba and Grossman (1988) suggested using tests for stationarity to rule out the existence of bubbles. This method could be useful in the case of a noncollapsing bubble, but as shown in Evans (1991), these tests tend to reject the presence of bubbles when regime-switching bubbles are present. Hall and Sola (1993) address this problem by extending the standard augmented Dickey-Fuller (ADF) test,

$$\Delta p_t = \alpha + \beta p_{t-1} + \sum_{k=1}^{n} \Psi_k \Delta p_{t-k} + \nu_t$$
$$\text{where } \nu_t \sim N(0, \sigma), \tag{11}$$

to allow the parameters to vary between two regimes, giving

$$\Delta p_t = \alpha_i + \beta_i p_{t-1} + \sum_{k=1}^{n} \Psi_{ki} \Delta p_{t-k} + \nu_{t,i}$$
$$\text{where } \nu_{t,i} \sim N(0, \sigma_i). \tag{12}$$

The slope coefficients β_S and β_C are the basis of the bubble test. Evidence that one regime is nonstationary (i.e., $\beta_S > 0$) while the other is stationary (i.e., $\beta_C < 0$) indicates the presence of a bubble.[11] We test these restrictions using a pair of t-tests. We also do a Wald test to see whether the two coefficients are jointly different from each other. The relationships between these tests can be seen in Figure 1. Requiring two significant t-statistics (each with the correct sign) restricts us to that part of the parameter space which lies to the right of the YT line and below the XT line. Using the Wald test for equality of the coefficients restricts us to the area below and to the right of the dotted line labeled “Wald.”

Funke, Hall, and Sola (1994) use the Markov-switching ADF test to find evidence for bubbles in the Polish economy in the late 1980s and early 1990s. Hall and Sola (1993) performed a brief study of the test’s

[10] It also means that all the parameters in the alternative model are identified under the null. This avoids the need for nonstandard distribution theory to test for the number of regimes (see Hansen 1996). Some have suggested that it would be more intuitive to test for the presence of bubbles by simply testing for the number of regimes. However, the presence of multiple regimes in such data is not disputed, and many fundamental asset-pricing models have been presented to rationalize the presence of regime switching. Examples of these would include Clark’s (1973) model of differing information-arrival rates, or Ceccetti, Lam, and Mark’s (1990) model of regime switching in fundamentals. Evans (1997) discusses regime switching in asset pricing at length.

[11] One property of switching regressions is that such models are identified only up to the particular relabeling of parameters that has the effect of swapping the names of the S and C regimes. This means that one should find either $\beta_S > 0$, $\beta_C < 0$, or $\beta_S < 0$, $\beta_C > 0$. In all that follows, the two regimes were normalized by setting regime S to be the regime with the greater slope coefficient.

properties. However, they only estimated a single realization of each of five different data-generating processes, including Evans bubble process (described below) with the probability of continuing to grow, π, equal to 0.75.[12]

2.3 The Van Norden Bubble Test

Van Norden (1996) and van Norden and Schaller (1993, 1996) modify the Blanchard model to allow for the possibility that the bubble is expected to collapse only partially in state C by replacing Equation 7 with

$$E_t(B_{t+1} \mid C) = u(B_t), \tag{13}$$

where $u(.)$ is a continuous and everywhere differentiable function such that $u(0) = 0$ and $1 \geq u' \geq 0$. Hence, the expected size of collapse will be a function of the relative size of the bubble, B_t, and the bubble is also not expected to grow (and may be expected to shrink) in state C. They also suggest that the probability of the bubble's continued growth falls as the bubble grows, so that

$$q = q(B_t); \qquad \frac{d}{d|B_t|} q(B_t) < 0. \tag{14}$$

Van Norden (1996) and van Norden and Schaller (1996) show that a first-order Taylor-series approximation of this process gives the following two-state switching regression system:[13]

$$\begin{aligned} E_t(\Delta B_{t+1} \mid S) &= \alpha_S + \beta_S B_t, \\ E_t(\Delta B_{t+1} \mid C) &= \alpha_C + \beta_C B_t, \\ Pr(\text{state}_{t+1} = S) &= \Phi(\lambda + \eta B_t), \text{ and} \\ Pr(\text{state}_{t+1} = C) &= 1 - Pr(\text{state}_{t+1} = S), \end{aligned} \tag{15}$$

where the model implies that $\beta_S > 0$, $\beta_C < 0$, and $\eta < 0$, and Φ is the Gaussian cdf function.[14] Again, one property of switching regressions is that such models are identified only up to a particular renaming of parameters that has the effect of swapping the names of the S and C regimes. In this case, this equivalence implies that

$$\text{llf}(\alpha_S, \beta_S, \alpha_C, \beta_C, \lambda, \eta, \sigma_S, \sigma_C) = \text{llf}(\alpha_C, \beta_C, \alpha_S, \beta_s, -\lambda, -\eta, \sigma_C, \sigma_S), \tag{16}$$

where llf() is the log-likelihood function, indicating that these alternative parameterizations cannot be distinguished without additional information. The van Norden bubble model implies that one should find either $[\beta_S > 0,\ \beta_C < 0,\ \eta < 0]$ or $[\beta_S < 0,\ \beta_C > 0,\ \eta > 0]$.

In addition to testing the above restrictions implied by the bubble model, van Norden (1993) and van Norden and Schaller (1993, 1996) test whether the bubble-motivated switching regression model gives significantly more information about the behavior of ΔB_{t+1}, than two simpler models.[15] Significant evidence of bubbles requires that the switching regression model can reject these simpler models. One of these is the normal-mixture model (NM)

$$\begin{aligned} \Delta B_{t+1} &\sim N(\alpha_S, \sigma_S) \qquad \text{when state}_{t+1} = S, \\ \Delta B_{t+1} &\sim N(\alpha_C, \sigma_C) \qquad \text{when state}_{t+1} = C, \\ Pr(\text{state}_{t+1} &= S) = \Phi(\lambda), \end{aligned} \tag{17}$$

[12] Note that Hall and Sola (1993) multiply the bubble term by 20 in constructing their simulated asset prices. This implies, unlike the case mentioned above, that their bubble will have a rate of return 20 times greater than that of the fundamental.

[13] The original model uses the exchange-rate innovation R_{t+1} as the dependent variable. This variable in turn consists of innovations in fundamentals ε'_{t+1} and innovations in the bubble. Hence $R_{t+1} = \varepsilon'_{t+1} + B_{t+1} - E_t(B_{t+1})$. If we assume that in this model $\varepsilon'_{t+1} = 0$ and use Equation 6, then $R_{t+1} = \Delta B_{t+1} - rB_t$. Since r is small, the use of ΔB_t as the dependent variable is a good approximation of the earlier model.

[14] This model differs trivially from that considered in van Norden (1996) and van Norden and Schaller (1993). The former assumed that $\Phi(x)$ was the logistic cdf rather than the Gaussian. Both papers also used slightly different classifying equations (using either $|B_t|$ or B_t^2) to allow for the possibility of negative bubbles.

[15] Van Norden (1996) also considers a third model. Since it nests within the normal-mixture model, rejections of the normal-mixture model imply a rejection of the third model.

which is simply the special case of Equation 15 where $\beta_S = \beta_C = \eta = 0$. A rejection of this null hypothesis implies that there is a significant link between B_t and the behavior of the mixing distributions, because B_t captures shifts either in their means, or in their mixing probabilities, or both.

Equation 16 also nests the linear regression model as the special case where $\beta_S = \beta_C$, $\alpha_S = \alpha_C$, and $\eta = 0$, giving the error contamination model (EC):

$$\begin{aligned} \Delta B_{t+1} &= \alpha + \beta B_t + e_{t+1}, \\ e_{t+1} &\sim N(0, \sigma_S) \text{ with prob } \Phi(\lambda_q), \\ e_{t+1} &\sim N(0, \sigma_C) \text{ with prob } 1 - \Phi(\lambda_q). \end{aligned} \tag{18}$$

Any rejection of this model can be interpreted as evidence of nonlinear predictability in asset prices. Note that if the variances differ across the two regimes, all parameters will be identified under the null.

The test statistics used to test for bubbles in the van Norden framework are simply the likelihood ratio tests of the null of Equation 17 or 18 against the alternative of Equation 15. The former tests the hypothesis that $\beta_S = \beta_C = \eta = 0$, while the latter tests $\beta_S = \beta_C$, $\alpha_S = \alpha_C$, and $\eta = 0$. A third, less restrictive, statistic that is also used is simply the Wald statistic testing the hypothesis that $\beta_S = \beta_C$. These are the three test statistics whose behavior we evaluate below.

Van Norden and Schaller (1993) use this test framework to show evidence of bubbles in monthly returns from the Toronto Stock Exchange (TSE). Van Norden (1996) looks for evidence of bubbles in post-Bretton-Woods floating exchange-rate data, and van Norden and Schaller (1996) examine the behavior of New York Stock Exchange (NYSE) monthly stock returns from 1926 to 1989. The latter paper also asks whether regime switching in fundamentals can account for the evident regime switching in stock returns.

2.4 Comparing Hall and Sola's Test with van Norden's Test

By comparing the last two sections, the reader can see that the Hall and Sola test and the van Norden test show some important similarities and differences in both parts of the regime-switching model: the level equations and the transition equations. Each of the two level equations gives the relationship between the observable dependent and explanatory variables for a particular regime. The transition equations give the probability of changing regimes at a given period of time.

When both tests have the same dependent variable (i.e., $B_t = P_t$ for all t) the level equation of van Norden's test (Equation 15) is a simpler version of the level equation of Hall and Sola's test (Equation 11), where $\Psi_{i,k} = 0$ for all i and k. In applications of these tests, several different kinds of dependent variables have been examined. Funke, Hall, and Sola (1994) used the actual changes in asset prices and the residuals from a regression of fundamentals on the assets. The van Norden test has been applied to excess returns on exchange rates and the rates of returns on stock market indices. Thus the applied researcher can choose from a number of different transformations when using these switching models.

The transition equations, however, are not necessarily the same. If van Norden's coefficient η equals zero, then the van Norden test becomes a constant-probability simple-switching model. Such a model is a special case of a Markov-switching model, implying that the van Norden test would then be nested inside of Hall and Sola's test. However, for a large majority of the bubbles examined below, estimates of η do not equal zero. Hence, the tests are not nested.

Not being nested doesn't mean that the tests are unrelated. For the Hall and Sola test, the probability of being in a given regime is dependent on an unobserved state variable that follows an AR(1) process with the autoregressive coefficient ρ equal to $Pr(S_t = S \mid S_{t-1} = S) + Pr(S_t = C \mid S_{t-1} = C) - 1$ (Hamilton 1989). In the van Norden test, the probability of being in a given regime is dependent on the level of the observed variable B_t. As B_t usually shows positive serial correlations,[16] the dynamics of the two models can be quite similar.

The theory on bubbles is ambiguous about how the probability of collapse should be modeled. The degree of uncertainty on how to model these transition probabilities suggests that either model may be useful. One could test which model would be more appropriate by estimating a Markov-switching model where the transition probabilities were dependent on the size of the bubble. This nonconstant transition probability

[16]This is as we would expect, since theory implies that rational speculative bubbles should exhibit explosive AR(1) dynamics.

Markov-switching model would encompass the other two models, but estimating such a model could prove difficult.

3 Size

As was noted in the introduction, relatively little is known about the finite-sample properties of estimators of switching regressions, particularly when applied to data that is non-Gaussian, as is the case with most financial time series. Given the econometric problems that have been shown to plague other tests for bubbles, it seems prudent to further investigate the behavior of the tests described in the previous section. In this section we will investigate their size properties while in the next section we will consider their power.

Since the Hall and Sola and the van Norden tests for bubbles have the absence of bubbles as their null hypothesis, knowing the size of these tests is important to determine the reliability of purportedly significant evidence of bubbles. This evidence will typically consist of comparing the test statistics mentioned in Section 2 (specifically, the t- and Wald statistics for the Hall and Sola test, and the likelihood-ratio (LR) and Wald statistics for the van Norden test) to their asymptotic distributions (the standard normal for the t-statistic and the χ^2 for the others). To determine whether these tests are correctly sized, we need to find a way to determine their finite-sample distribution under the null hypothesis of no bubbles. As was discussed in Section 1, we know a priori that this distribution will depend on the behavior of the fundamental determinants of asset prices. Therefore, care needs to be taken both in selecting the DGP for the null hypothesis of no bubbles, and in deciding whether this DGP is a reasonable guide for particular applications.

To examine test size, we begin in Section 3.1 by describing our method for simulating the distribution of the test statistics under the null hypothesis of no bubbles. Section 3.2 then discusses the two data sets we use as the basis for our randomization experiments. Section 3.3 calculates actual test statistics from these data sets and compares their interpretations based on asymptotic and finite-sample distribution.

3.1 Methodology

As was mentioned in Section 2, neither the Hall and Sola nor the van Norden bubble tests test for the presence of regime switching per se. Instead, the Hall and Sola test tests for the presence of a particular dynamic behavior of prices that differs across regimes, while the van Norden test tests for a relationship between returns and measures of purported bubble size that differ across regimes.

To simulate prices or returns under the null hypothesis of no bubbles, we therefore need to choose a DGP so that these particular relationships do not differ across any possible regimes that may or may not be present in the data. Having done so, we could then simulate the DGP to get artificial data and calculate the appropriate test statistics. Repeating this procedure for multiple independent draws of our DGP would then allow us to simulate the finite sample distribution of our test statistics for this particular setting, and compare them to their usual asymptotic approximations.

We use a randomization strategy to reorder the data in such as way as to destroy any of the possible relationships for which we are testing while maintaining the exact unconditional distribution of our original data sample. Since the Hall and Sola and van Norden tests use different data series, this strategy was implemented somewhat differently for each test.

3.1.1 Hall and Sola For the Hall and Sola test, we begin with a data series of log asset prices p_t for $t = 0, \dots, T$, and its first difference Δp_t for $t = 1, \dots, T$. We then generate an artificial data series $\tilde{p}_t$ in two steps:

1. shuffle[17] the series Δp_t to generate a new series, $\Delta \tilde{p}_t$, and
2. recursively construct the new series $\tilde{p}_t$ by setting $\tilde{p}_0 = p_0$, and then iteratively using the relationship $\tilde{p}_t = \tilde{p}_{t-1} + \Delta \tilde{p}_t$.

The result is that the artificial price series $\tilde{p}_t$ will be a martingale (since its increments are independent by construction), and its increments will have exactly the same unconditional distribution as Δp_t. The martingale

[17] To be precise, we randomly draw without replacement from the series $\{\Delta p_t\}$ a new series $\{\Delta \tilde{p}_t\}$ with the same number of elements, T.

structure ensures that the true autoregressive coefficients (the β_i terms in Equation 12) must be the same and equal to zero in both regimes.[18]

We performed 2,500 replications of this randomization experiment for each of our two data sets (described in Section 3.2) and tabulated the results. Note, however, that our estimation algorithms converged in only 2,124 and 2,233 of these cases for the TSE Composite (300) Index and Standard & Poor's 500 Index (S&P500) data, respectively. We were unable to calculate standard errors for the remaining cases, and so based our analysis only on those cases where we were able to get convergence. We take some comfort from the fact that parameter estimates on the actual (i.e., nonrandomized) data set converged rapidly, so studying those cases where convergence was achieved should be representative.

3.1.2 Van Norden For the van Norden test, we begin with a data series of log excess returns on an asset x_t for $t = 1, \ldots, T$, and a measure of the purported bubble B_t for $t = 1, \ldots, T$. We then generate an artificial data series $\tilde{B}_t$ by shuffling the original series B_t.[19] The result is that the artificial bubble series $\tilde{B}_t$ will be unrelated to the excess return series x_t by construction, and will have exactly the same unconditional distribution as B_t. This ensures that the true slope coefficients (β_S, β_C, and η) in Equation 15 satisfy the restriction that $\beta_S = \beta_C = \eta = 0$.[20] Note, however, that there is no requirement that $\alpha_S = \alpha_C$, so tests of the error contamination (EC) model (which imposes this restriction) may continue to reject. Therefore, this experiment will only give us a useful assessment of the size of the Wald test of $\beta_S = \beta_C$ and of the LR test of the NM model (which imposes $\beta_S = \beta_C = \eta = 0$).

We again performed 2,500 replications of this randomization experiment for each of our two data sets and tabulated the results. In this case, our estimation algorithms converged normally in 2,479 and 2,459 of these cases for the TSE and the S&P500 data, respectively. We were unable to calculate standard errors for the remaining cases, and so based our analysis only on those cases where we were able to get convergence. As a check, we compared the results obtained for the two LR statistics (which may be computed even in the absence of estimated standard errors) on the full 2,500 replication sample and on the slightly smaller sample where standard errors could be computed. The differences were trivial; 1% and 99% critical values never differed above the second decimal place.

3.2 Data

Stock prices are perhaps the series most commonly investigated for bubbles, and are the series we choose to use in our randomization experiments. The asset prices we analyze are those of the Toronto Stock Exchange Composite (300) Index and Standard and Poor's 500 Index, which we continue to refer to as the TSE and the S&P500. Both are monthly series of the end-of-month closing quotation. We used the longest available span of data from our source,[21] which gave us a series covering the period from 1947M1 to 1997M7 (607 observations) for the S&P500 and from 1956M1 to 1997M7 (499 observations) for the TSE. The post-1956 period for the TSE was previously analyzed in van Norden and Schaller (1993), while van Norden and Schaller (1996) and Schaller and van Norden (1997) used monthly data on the CRSP (Center for the Study of Securities Prices) stock index from 1926 to 1989. We feel our experiments should therefore provide a particularly useful benchmark for these studies.

While the Hall and Sola test requires no additional data, the van Norden test needs data on the excess returns on these indices and a purported measure of the bubble. Constructing excess returns from prices requires that we have additional information on dividends and interest rates.[22] Dividend information came

[18] This randomization also destroys any possible first-order Markovian dependence in the data. While this does not rule out the presence of regime switching in the simulated data, it means that the probability of being in any given state at time t conditional on all information available at $t-1$ must be constant.

[19] Again, *shuffling* means that we randomly draw without replacement from the series $\{B_t\}$ a new series $\{\tilde{B}_t\}$ with the same number of elements, T.

[20] A reasonable alternative to this procedure would to be randomize x instead of B, which would impose the same restrictions on the true DGP. Note, however, that the likelihood function of the simple (i.e., non-Markovian) switching regression model is invariant to a reordering of the observations. This means this alternative randomization must produce results identical to the strategy we described above.

[21] Our data source is the *Bank of Canada Review*, series B4237 for the TSE and B4291 for the S&P500.

[22] Excess returns are given by the formula $x_t = (P_t + D_t)/P_{t-1} - i_t = (P_t/P_{t-1}) \cdot DY_t - i_t$, where x_t is the monthly excess return, P_t is the level of the stock price index, D_t is the level of the stock dividend index, DY_t is the (monthly) dividend yield, and i_t is the monthly yield on alternative investments.

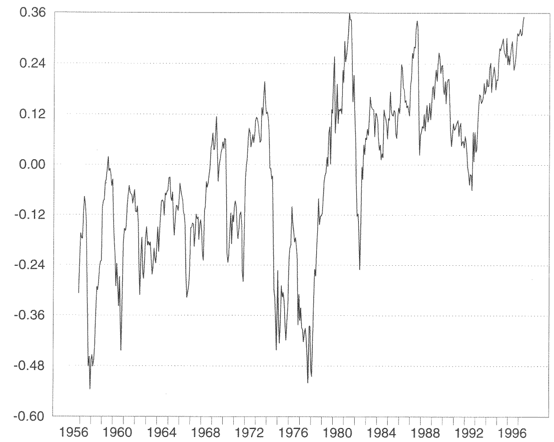

Figure 2
Implied bubble in TSE300.

from the corresponding dividend-yield series for each stock index.[23] The interest-rate series was the yield on three-month corporate paper for the Canadian data, and the rate on three-month commercial paper for the U.S. data.[24] These same series were also used to construct the bubble measure, using a methodology developed by Campbell and Shiller (1987, 1988) and previously used by van Norden and Schaller (1996) and Schaller and van Norden (1997).[25]

Briefly put, the Campbell-Shiller method estimates a VAR which is then used to construct dynamic forecasts of dividends and interest rates. Fundamental prices are calculated as the present value of the forecast dividend series, discounted using the time-varying interest-rate forecasts. The fundamental price at time t uses dynamic forecasts conditioned only on the information available at time t, so unexpected movements in dividends or interest rates can cause potentially large changes in fundamental prices. Prices are included in the VAR system since in a forward-looking market they will embody information about the expected evolution of interest rates and dividends beyond that available to the econometrician. The implied measures of the bubble are shown in Figures 2 and 3 for the TSE and the S&P500, respectively.[26]

3.3 Results

Because our primary focus in this section is the econometric size of regime-switching tests for bubbles, our discussion of the model's estimates for the actual data will focus on the test statistics for the null hypothesis of no bubble, which were previously discussed in Section 2. We omit a full reporting of the estimated model parameters and an evaluation of their fit. However, all models passed a series of score matrix tests for

---

[23] Our data source was again the *Bank of Canada Review*, series B4245 for the TSE and B4226 for the S&P500.

[24] Our data source was again the *Bank of Canada Review*, series b14017 for Canada and b54412 for the U.S. Consistent longer term Canadian interest-rate series were only available for a much shorter time span. However, the relative volatility of the price series implies that at monthly frequencies, the behavior of excess returns is not very sensitive to the choice of interest-rate series. We were also unable to extend our U.S. interest-rate series back beyond 1962, forcing us to use a shorter sample for our U.S. estimates of this model.

[25] See Campbell and Shiller (1988) for a detailed description and justification. Although not used for our randomization experiments, we also estimated the bubble model on the true data using an alternative measure of the purported bubble based on the same cointegrating relationships used in van Norden and Schaller (1993). This measure gave weaker evidence against the null hypothesis of no bubbles.

[26] This method requires some judgment in choosing the lag length for the VAR. Although we selected 10 lags for the TSE data and 8 for the S&P500, we found that the resulting bubble series were sensitive to the number of lags used and that different statistical criteria gave different results. Generally, lags around or less than 12 implied that the stock market was overvalued in the late 1990s. The degree of overvaluation is reduced as lags are added. For example, using 24 lags, we would conclude both markets were severely *under*valued in the late 1990s. We felt that the most plausible results were those that implied that the market was overvalued during this period, as this seemed to match the opinions of many market observers. As argued by Schaller and van Norden (1997), misspecification of the bubble series is presumed to reduce the test's power rather than distort its size (a claim our randomization experiment allows us to examine critically.) Therefore, if we have misspecified the number of lags, we would expect that this would make it less likely that we could reject the null hypothesis of no bubbles.

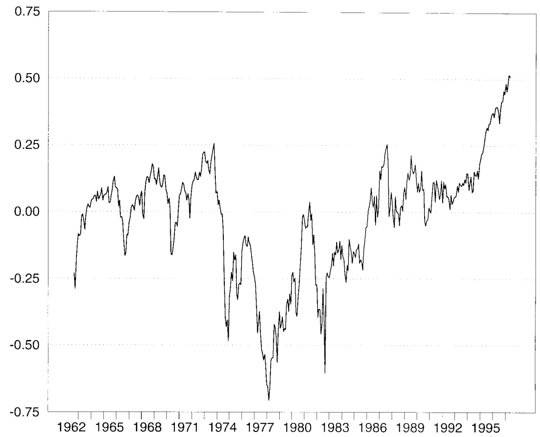

Figure 3
Implied bubble in S&P500.

Table 1
Bubble test results.

Test	Statistic	TSE	S&P500	Asymptotic 5% Critical Value
Hall & Sola	t: $\beta_C \geq 0$	−5.65	−5.17	−1.65
	t: $\beta_S \leq 0$	0.30	0.84	1.65
	Wald: $\beta_S = \beta_C$	31.59	27.20	3.84
Van Norden	LR: $\beta_S = \beta_C$ $\alpha_S = \alpha_C,\ \eta = 0$	6.58	2.43	7.82
	LR: $\beta_S = \beta_C = \eta = 0$	9.59	6.23	7.82
	Wald: $\beta_S = \beta_C$	0.299	0.008	3.84

misspecification, including tests for residual serial correlation, omitted regime persistence, and neglected ARCH effects.

The results of the bubble tests for the TSE and S&P500 data are reported in Table 1. The evidence is mixed. The Hall and Sola test finds strong evidence that the autoregressive coefficients differ across the two regimes. The Wald statistic testing the hypothesis that $\beta_S = \beta_C$ is greater than 25, which may be compared to its asymptotic 5% critical value of 3.84. However, while we find $\beta_C < 0$ as the bubble model would predict (the t-statistic testing the null is greater than 5), estimates of β_S seem to be insignificantly different from zero instead of positive.

Although results for the Hall and Sola tests are similar for the two data sets, results using the van Norden tests give stronger evidence of bubbles with the TSE than with the S&P500 data. However, the van Norden tests are not very supportive of the bubble hypothesis. The LR tests of the EC model ($\beta_S = \beta_C$, $\alpha_S = \alpha_C$, and $\eta = 0$) give statistics below their asymptotic 5% critical values, and Wald tests give no evidence to reject the hypothesis that $\beta_S = \beta_C$. However, LR statistics testing $\beta_S = \beta_C = \eta = 0$ are not far from their standard critical values; that for the TSE is slightly above, and that for the S&P500 is slightly below.

Tables 2 and 3 compare these statistics to their finite-sample critical values as tabulated from our randomization experiments. The use of these exact critical values leads to precisely the same conclusions.

In Table 2, we see the 90%, 95%, and 99% empirical critical values for the three Hall and Sola test statistics and for each of our two data sets. The first point to note is that the empirical critical values shown are (with only one exception) always closer to zero than the asymptotic critical values. This means that the use of asymptotic critical values makes all three tests conservative; rejections of the null hypothesis are more significant than they appear to be, or equivalently, asymptotic p-values will be too high. The size distortion is most pronounced for the Wald statistic, where empirical critical values for the S&P500 are only about 60% of the asymptotic counterparts. The second point to note is that the differences between the finite-sample and the asymptotic critical values is quite small compared to the difference between any of these critical values and

Table 2
Empirical critical values for the Hall & Sola tests.[a]

Test Statistic	Source	90%	95%	99%	Actual
t: $\beta_C \geq 0$[b]	TSE	−1.29	−1.68	−2.51	−5.65
	S&P500	−1.12	−1.37	−1.93	−5.17
	$N(0, 1)$	−1.28	−1.65	−2.33	
t: $\beta_S \leq 0$[c]	TSE	1.22	1.56	2.28	0.30
	S&P500	1.11	1.40	2.09	0.84
	$N(0, 1)$	1.28	1.65	2.33	
Wald: $\beta_S = \beta_C$	TSE	1.84	2.76	5.86	31.59
	S&P500	1.42	2.16	4.28	27.20
	$\chi^2(1)$	2.71	3.84	6.64	

a. Trials for which the covariance matrix of the parameter estimates could not be computed were omitted from this tabulation. This was generally due to a failure to achieve convergence, and accounted for 376 and 267 of the 2,500 trials of the TSE and S&P500 data, respectively.
b. Critical values based on the area in the left tail of the distribution.
c. Critical values based on the area in the right tail of the distribution.

Table 3
Empirical critical values for the van Norden tests.[d]

Test Statistic	Source	90%	95%	99%	Actual
LR: $\beta_S = \beta_C$ $\alpha_S = \alpha_C,\ \eta = 0$	TSE	10.98	12.54	16.45	6.58
	S&P500	8.35	9.91	12.92	2.43
	$\chi^2(3)$	6.25	7.82	11.35	
LR: $\beta_S = \beta_C = \eta = 0$	TSE	4.92	6.51	10.42	9.59
	S&P500	5.10	6.56	10.09	6.23
	$\chi^2(3)$	6.25	7.82	11.35	
Wald: $\beta_S = \beta_C$	TSE	2.21	3.23	6.49	0.299
	S&P500	2.38	3.65	8.00	0.008
	$\chi^2(1)$	2.71	3.84	6.64	

d. Trials that did not have normal convergences and for which the covariance matrix of the parameter estimates could not be computed were omitted from this tabulation. This amounted to 21 and 41 of the 2,500 trials for the TSE and S&P500 data, respectively. Using all 2,500 trials gave identical results to at least the first decimal place.

the test statistics we wish to evaluate. All six test statistics are either more than double the 99% critical value or less than half of the 90% critical value. In this case, therefore, the size distortion is of no practical importance.

In Table 3, we see the 90%, 95%, and 99% empirical critical values for the three van Norden test statistics and for each of our two data sets. As expected, the empirical critical values for the LR test of the EC model's restrictions are higher than conventional critical values.[27] For the other two tests, the empirical critical values are (with again only one exception) always closer to zero than the asymptotic critical values; this means that the use of asymptotic critical values tends to make both tests conservative. However, the empirical critical values appear closer to their theoretical values than was the case with the Hall and Sola test; they are typically over 80% of their theoretical counterparts.

We should again note that the use of finite-sample critical values instead of their asymptotic counterparts never changes the conclusions we draw from the actual tests. Our Wald statistics are always much less than one-quarter of the 90% critical value. The LR statistic for the NM model's restrictions is always significant at the 95% but not the 99% level for the TSE data, and at the 90% but not the 95% level for the S&P500.

To better assess the general extent of the size distortion, we constructed the p-value and p-value discrepancy plots proposed by Davidson and MacKinnon (1997). The former is simply a scatterplot of the theoretical and empirical p-values of a given statistic. If the empirical and theoretical distributions are the same, this will simply be a 45° line. Points above the 45° line mean that the test statistic is liberal at that

[27] As we mentioned in Section 3.1, the restrictions imposed by the EC model (shown in the uppermost block of rows) are not necessarily satisfied by our randomized DGP. Therefore, we are unable to evaluate the size of this test; we simply report the results of our tabulations for completeness.

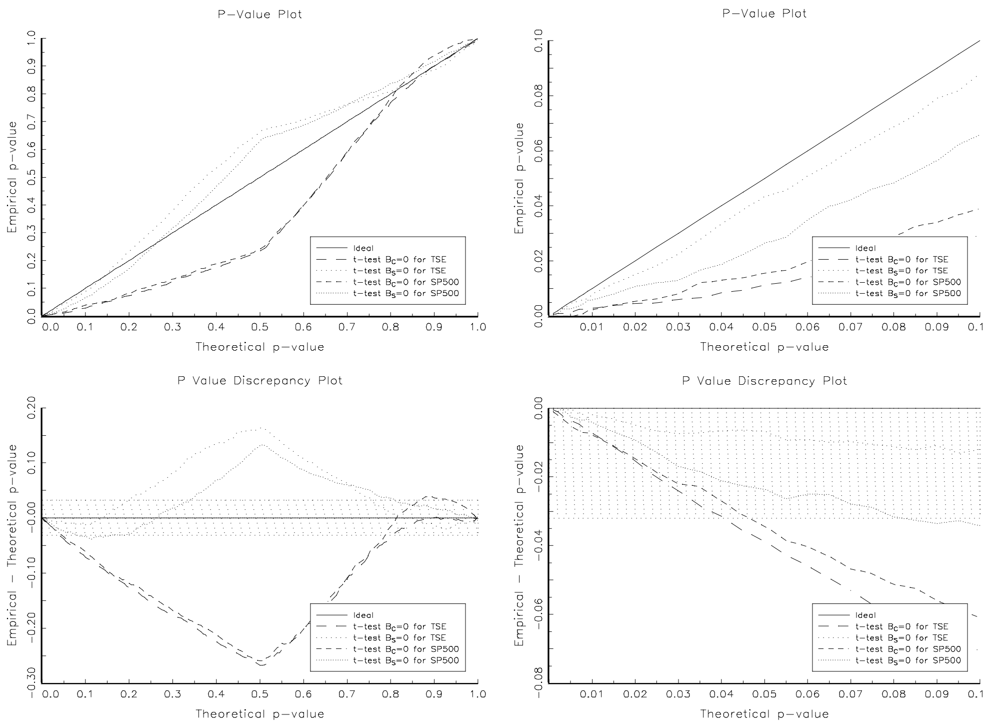

Figure 4
Size distortion in Hall and Sola t-tests.

nominal p-value, while points below the line imply that the statistic is conservative. The p-value discrepancy plots simply graph the difference between the empirical and theoretical p-values as a function of the theoretical p-value. A perfectly sized test will give a horizontal line at 0; a liberal test will lie above this line, and a conservative test will lie below.

We present these graphs in pairs: the upper row shows the p-value plot, while the lower row is the p-value discrepancy plot (Figures 4–7). The left column shows the plot for the full [0, 1] interval, while the right column shows the same results for the [0, 0.1] interval to provide more detail on the region in which we are usually most interested. The discrepancy plots also show a shaded rectangle surrounding the x-axis, which gives the 95% confidence for the Kolmogorov-Smirnov test of the hypothesis that these statistics are drawn from their theoretical distribution. In other words, p-value discrepancies that lie within the shaded region can be attributed to experimental error; larger discrepancies are significant evidence that the estimated statistics are drawn from a different population than that implied by theory.

Results for the t-statistics from the Hall and Sola model are shown in Figure 4. Results are very similar for the S&P500 and the TSE data sets. If we restrict our attention to the 0–0.1 interval, which is the most relevant range for hypothesis testing, we see that the tests are always conservative. Theoretical p-values are sometimes double their true values in this range. This suggests that the Hall and Sola tests have a tendency to understate the significance of the evidence of bubbles when using conventional critical values. The size distortion is less pronounced for the t-tests of β_S, which also tend to be liberal if we use p-values in the 20%–80% range. We have little difficulty rejecting the hypothesis that the distribution of our simulated statistics differs from their predicted asymptotic distribution. The tendency for tests of β_S to show higher empirical p-values than tests for β_C is presumably the result of our identifying restriction, which imposes that $\beta_S > \beta_C$.

Figure 5 shows the results for Wald tests of the hypothesis that $\beta_C = \beta_S$ in both the Hall and Sola and the van Norden models. The results are similar across models and data sets. Size distortion is generally smaller than for the t-tests. The tests tend to be conservative, although they sometimes appeared to be liberal for very low (< 0.05) or very high (> 0.7) p-values. Conservative size distortion was sometimes large enough to be significant, but liberal size distortion was always insignificant.

Figure 6 compares the above results for the Wald test in the van Norden model to the LR tests of the restrictions imposed by the NM model. Again, results are very similar across the two data sets and both

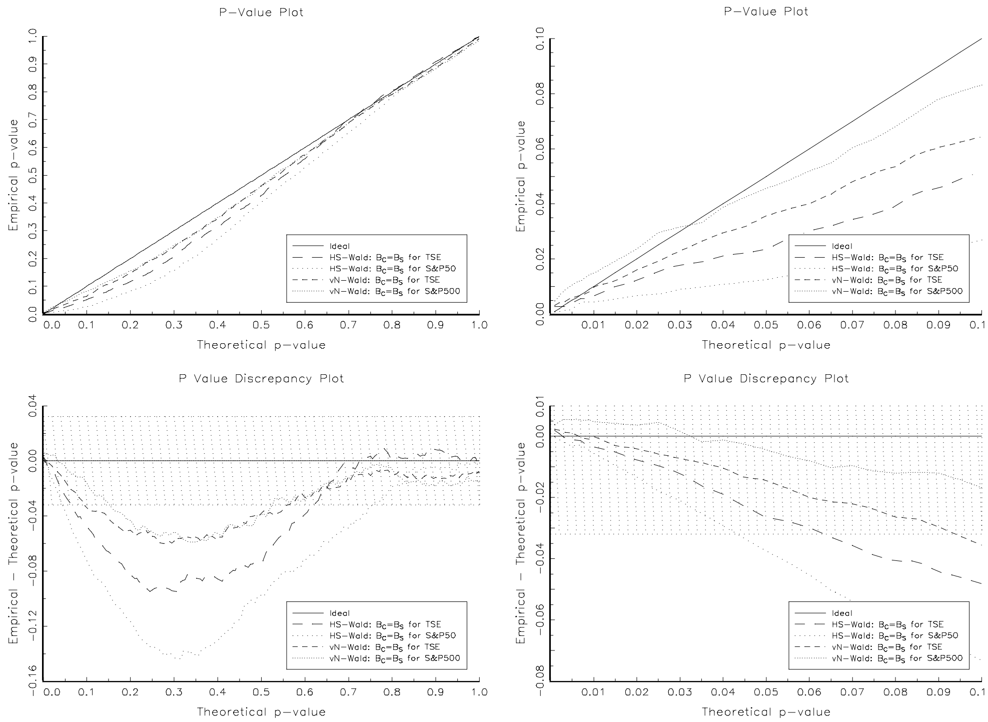

Figure 5
Size distortion in Wald tests.

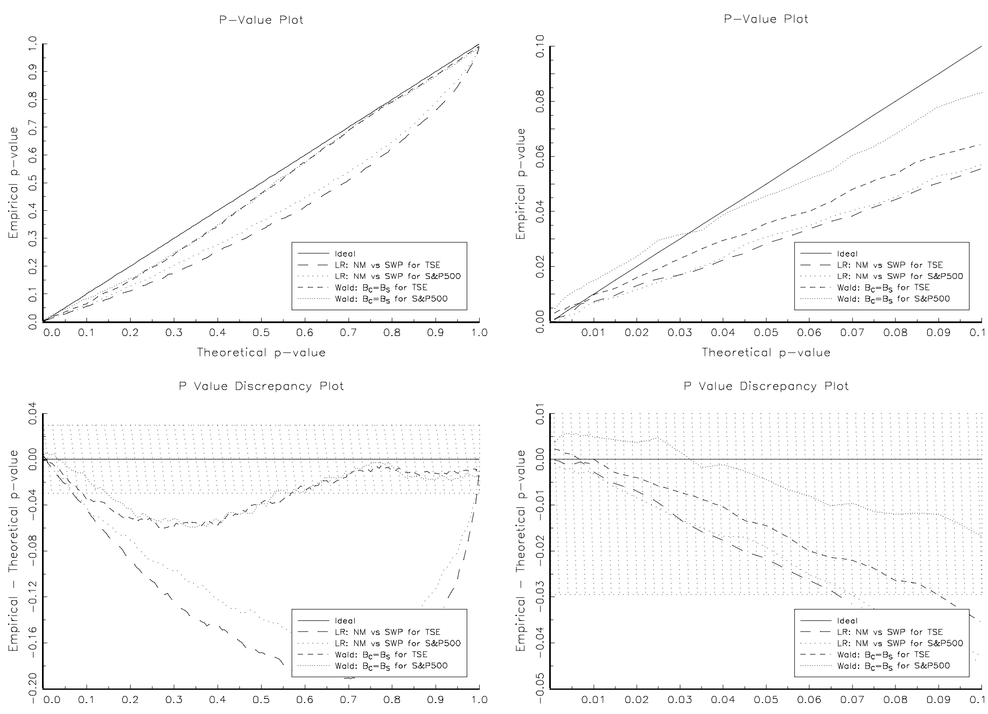

Figure 6
Size distortion in van Norden tests.

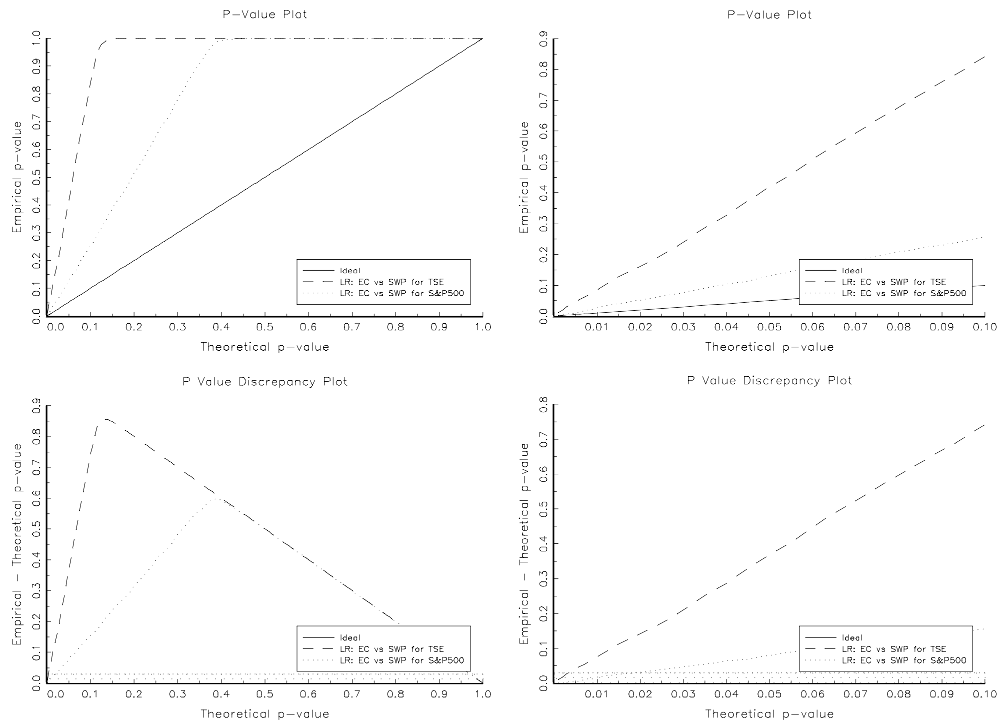

Figure 7
Results for LR test of EC model restrictions.

statistics tend to show conservative size distortion. Although this conservative distortion is often significant, there is never significant evidence that the tests are liberal. The LR tests consistently display more size distortion than the Wald tests, with the difference largest for p-values over 30%.

Finally, Figure 7 shows the results for the test of the restrictions imposed by the EC model. These reflect the considerable distortion noted above in Table 3, which in turn is owing to the fact that the data were not generated under the null hypothesis of the EC model.

To summarize, our randomization experiment suggests that the tests of either model are unlikely to produce false evidence of bubbles (i.e., type-II errors.) While there are sometimes statistically significant size distortions, these have the effect of making the test conservative,[28] and are never large enough to alter the conclusions in our applied example.

4 Power

Having examined the size of the Hall and Sola and the van Norden bubble tests, we now want to use Monte Carlo experiments to evaluate and compare the power of the two testing methodologies. This involves specifying a data-generating process that creates bubbles, generating multiple time series from this process, estimating the regime-switching models, and applying the tests described above. The frequency with which we find significant evidence of bubbles gives us the power of the test. Since the finite-sample critical values we derived in the previous section are DGP-specific, we revert to using asymptotic critical values to evaluate test power. Given our results that this seems to lead to overly conservative inference (i.e., to reduce the chances of type-I error), using these critical values should understate the size-adjusted power of the tests.

4.1 Data-Generating Process

We use various parameterizations of Evans's (1991) bubble model as our data-generating process. This DGP has several attractive features. First, the problems of unit-root and cointegration-based tests on this data set are

[28]The exception is the EC test, where this is not truly "size distortion," since we did not simulate the data under the null hypothesis.

well documented, facilitating a comparison of the regime-switching tests with earlier tests.[29] Second, Charemza and Deadman (1995) study the performance of the earlier tests on other data-generating processes and reach conclusions broadly similar to those of Evans, suggesting that the Evans process might not produce atypical results. Third, as we explain below, the Evans model is not precisely nested within either the Hall and Sola or the van Norden bubble-testing models. We think this introduces an interesting amount of misspecification into the experimental design, and may give a better indication of how the tests are likely to perform when confronted with real data that may not nest perfectly within either model. We also feel that it offers a neutral "middle ground" on which to compare the performance of the two tests.

As we noted in Section 2, the Evans model is a generalization of the Blanchard (1979) model where both the size of collapses and their probability are functions of the size of the bubble; it incorporates partial rather than total collapses, and sets the probability of collapse equal to zero when $B_t \leq \alpha$.

Initially, the bubble grows at an average rate $1 + r$, but the realized rate of growth differs from the expected value by serially uncorrelated mean-zero errors. We will refer to this phase of steady expected growth as regime G. Once the bubble's size reaches a threshold level of α, its behavior changes. It continues to grow at an expected rate of $1 + r$, but there is now a probability $1 - \pi$ of collapse to a level δ (regime C). To compensate, if the bubble does not collapse (regime E), it is expected to grow at a rate greater than $1 + r$.

This model can be written as:

$$\begin{aligned} B_{t+1} &= (1+r)B_t u_{t+1} \qquad \text{for } B_t \leq \alpha, \text{ and} \\ B_{t+1} &= (\delta + \theta_{t+1}\pi^{-1}(1+r)(B_t - \delta(1+r)^{-1}))u_{t+1} \qquad \text{for } B_t > \alpha, \end{aligned} \tag{19}$$

where α and δ are positive parameters with $\delta < (1+r)\alpha$, u_t is an exogenous independently and identically distributed strictly positive random variable with $E_t u_{t+1} = 1$, and θ_t is an exogenous independently and identically distributed Bernoulli process that takes the value 1 with probability π and 0 with probability $1 - \pi$. Evans's bubble satisfies Equation 6.

There are two points to note about this model. First, since u_t is strictly positive, the bubble will never change sign and will never entirely vanish. Second, regime G is distinguished from the mixture of the other two regimes only by the distribution of innovations in the bubble. For a particular distribution of u_t, the innovations in the mixture of regimes C and E will simply appear to be more volatile than in G.

For estimation, we rewrite Equation 19 in first differences as:

$$\begin{aligned} &\text{G} \qquad \Delta B_t = \{(1+r)u_t - 1\}B_{t-1} \qquad \text{for } B_{t-1} \leq \alpha, \\ &\text{E} \qquad \Delta B_t = \left\{\frac{(1+r)}{\pi}u_t - 1\right\}B_{t-1} + \frac{(\pi-1)\delta u_t}{\pi} \qquad \text{for } B_{t-1} > \alpha \text{ and } \theta_t = 1, \text{ and} \\ &\text{C} \qquad \Delta B_t = \delta u_t - B_{t-1} \qquad \text{for } B_{t-1} > \alpha \text{ and } \theta_t = 0. \end{aligned} \tag{20}$$

For our Monte Carlo experiments, we generate 5,000 draws of the above process with 100 observations.[30] We use the same parameter values as Evans, setting $r = 0.05$, $\alpha = 1$, $\delta = 0.5$, $B_1 = \delta$, and $u_t = exp(y_t - \frac{\tau^2}{2})$ where $y_t \sim IIN(0, \tau^2)$, and $\tau = 0.05$. We allow the probability of the bubble continuation, π, to vary over the same interval as Evans: $[0.999, 0.25]$.

To simplify estimation, all data series were standardized to have a mean of zero and a variance of one. (For graphing, they were also centered at (0, 0).) The relationship between ΔB_t and B_t can be seen in Figure 8. At high levels of π, the graph appears to be composed of two branches. The left branch corresponds to state C, where the bubble collapses, and the right with states G and E, where the bubble continues to grow. As π decreases, state G becomes more distinct from state E. State G can be identified as the large mass centered at

[29] Hooker (1996) uses the Evans DGP to examine a bubble test proposed by Durlaf and Hooker (1994) that differs from the regime-switching tests in testing both for specification error and for a bubble term separately and sequentially. Hooker conducts Monte Carlos for both the size and power of the tests. For the Evans DGP, the test performs well for all values of π with the percentage of correct detections ranging from 55%–45%, and decreasing slightly with π. Since the regime-switching tests are better with lower values of π, as shown in the next section, these two kinds of tests may be considered complementary. Our results and Hooker's results, however, are not directly comparable. Here we test the bubble series alone. Hooker tests an I(2) series where fundamentals and bubble are combined. Furthermore, the parameter values used by Hooker differ from those used by Evans and by us.

[30] Our application of the bubble test will be conducted on artificial data for the bubble, where the original authors tested asset prices (i.e., the bubble term plus the fundamental term). Since both must satisfy the same dynamic relationships, this change should be innocuous.

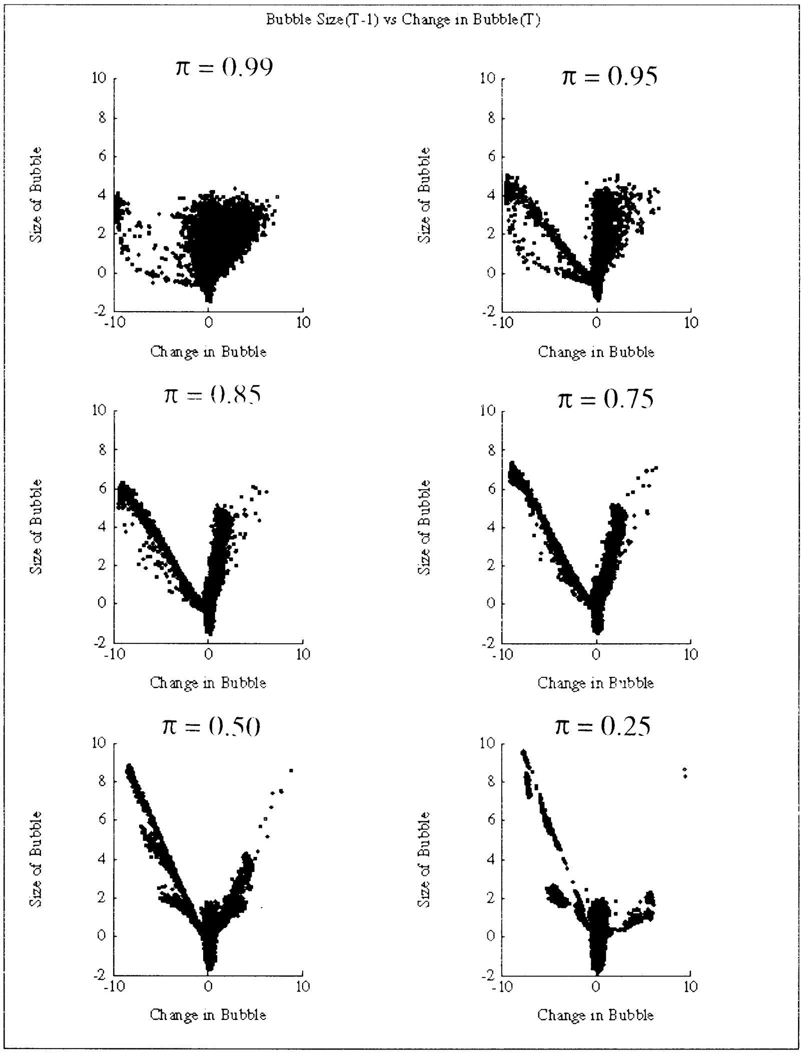

Figure 8
Evans's bubble process.

0 on the horizontal axis. It is most prominent when $\pi = 0.25$. The decrease in π also causes a change in the slopes of the two branches. This is because the growth rate in state E increases as π decreases. This increase in the growth rate results in the decline in the slope of the right branch.

4.2 Monte Carlo Results

The two regime-switching tests frequently detect bubbles that the unit-root tests incorrectly reject. When comparing regime-switching with unit-root tests, one must remember that these two tests have different null hypotheses: one is that bubbles are present, and the other is that they are not. Since unit-root tests take as their null that bubbles are present, and since critical values are chosen so as to minimize the probability of type-I (but not necessarily type-II) error, a reliable unit-root test might frequently lead us to conclude that bubbles are present when they are not (type-II error) but should rarely do the reverse (type-I error.) As seen in Table 4, this is clearly not the case.

The relationship between the value of π and the ability to detect bubbles varies among the different tests. For values of π less than 0.99, the Bhargava (1986) N_1 and N_2 unit-root tests frequently and incorrectly reject

Table 4
Summary table: ability of tests to detect bubbles.

Test		π						
		0.999	**0.99**	**0.95**	**0.85**	**0.75**	**0.50**	**0.25**
Bhargava $N1$	Rejection in favor of explosive alternative	78.5	32.5	0	0	0	0	0
	Rejection in favor of stable alternative	0	1	65.5	94.6	98	100	100
Bhargava $N2$	Rejection in favor of explosive alternative	95	58	15	4.5	2	1	0
	Rejection in favor of stable alternative	0	1	18.5	90	94.5	97	97
van Norden	t-test % rejections	1	5	16	48.5	77	28.5	3
Hall & Sola	t-test % rejections		25	50	64	64	58	35

Table 5
Percentage of draws that failed to converge: Hall and Sola (HS), van Norden (vN, NM, EC).

π	**0.999**	**0.99**	**0.95**	**0.85**	**0.75**	**0.50**	**0.25**
Hall & Sola (HS)	NA[e]	7.60	10.78	4.80	3.40	3.46	10.42
van Norden (vN)	3.32	5.32	3.30	3.82	4.62	4.82	2.92
Normal mixture (NM)	0.92	3.56	1.74	1.22	0.90	0.58	0.58
Error contamination (EC)	3.4	4.42	3.78	2.78	3.38	2.92	2.76
vN & NM	0.14	0.62	0.06	0.18	0.12	0.08	0.02
vN & EC	0.52	0.46	0.36	0.64	0.50	0.18	0.22
NM & EC	0.06	0.22	0.32	0.18	0.18	0.08	0.06
HS ∪ vN	NA	11.98	13.78	8.32	7.880	8.160	13.06
HS & vN	NA	1.00	0.34	0.32	0.16	0.18	0.32

e. The Hall and Sola test has not been done for $\pi = 0.999$.

the null of a bubble in favor of a stationary stable alternative.[31] The van Norden test does best when π equals 0.75, and does a poorer job for other values. The Hall and Sola test detects more bubbles than van Norden according to the t-tests for all values of π except when π equals 0.75.

The following sections give more details on our results. The next section discusses the difficulties experienced in trying to get the maximum-likelihood estimation methods to convergence. Following sections discuss each test individually.

4.2.1 Convergence A standard problem in performing Monte Carlo or other simulation experiments with iterative estimators is that some fraction of the estimates will typically fail to converge. This in turn puts limits on the confidence we should attach to our experimental results. Fortunately, this was not a serious problem in practice. Table 5 shows that for the Hall and Sola test, we achieved convergence for 90% or more of the simulated data sets, regardless of the parameterization of the DGP considered. For the van Norden test, we needed to estimate as many as three switching models on each data sample. Fortunately, convergence rates were generally higher than for the Hall and Sola model, as shown in Table 5. We also occasionally have the problem that the restricted models give values of the higher likelihood function than the unrestricted model (which may reflect false convergence or the presence of multiple local maxima). As shown in Table 6, this problem was also rare, except when $\pi = 0.5$. (We explore the case where $\pi = 0.5$ in greater detail below.) In all subsequent tables, the reported fraction of cases in which bubbles were detected counts as nondetection cases those where some models failed to converge and those where restricted models gave the highest values of the likelihood function.

4.2.2 Hall and Sola Test The individual tests (Table 7) show considerable power to detect bubbles. The Wald test shows that the autoregressive coefficients are significantly different 40%–70% of the time, and t-tests show that the individual coefficients are significantly positive/negative even more frequently (with one exception). Overall, the tests seem to have the greatest power when the probability of the bubble surviving is around

[31] The greatest difference between a percentage that we report and Evans is less than 5%.

Table 6
Percentage of draws with restricted LLF greater than unrestricted: van Norden.

π	0.999	0.99	0.95	0.85	0.75	0.5	0.25
NM > vN[f]	1.122	0.4862	0.04138	0.2080	5.789	88.33	2.514
EC > vN	6.594	3.974	1.179	1.414	8.578	87.87	2.267

f. The NM > vN, etc., are those draws that returned likelihood function values higher for the restricted model than the unrestricted case. The two restricted models are the normal mixture model (NM) and the error contamination model (EC).

Table 7
Hall and Sola: Percentage of draws passing single tests.

π	t: $\beta_C \geq 0$	t: $\beta_S \leq 0$	Wald: $\beta_S = \beta_C$
0.25	65.24	46.13	50.17
0.50	71.00	68.95	65.07
0.75	75.18	76.15	67.43
0.85	71.20	84.58	66.45
0.95	53.75	84.90	53.22
0.99	27.32	89.78	42.29

Table 8
Hall and Sola: Percentage of draws passing multiple tests.

π	Reject $\beta_C \geq 0$ and $\beta_S \leq 0$	Reject $\beta_C \geq 0$, $\beta_S \leq 0$, and β_S=β_C
0.25	34.78	17.26
0.50	57.97	37.79
0.75	63.79	42.51
0.85	64.37	42.82
0.95	50.16	27.03
0.99	24.91	10.10

80%. At the highest probabilities, the estimated coefficient in the collapsing regime performs poorly, perhaps because so few collapses would be observed in a sample of 100 observations. At the lowest probabilities, the estimated coefficient in the surviving regime performs poorly.

The joint test results (Table 8) also show that the bubble test is most successful for mid-range levels of π. If we require that all the coefficients are both statistically different from zero and from each other, we find significant evidence of bubbles as much as 43% of the time. This power again drops off considerably as π approaches 0 or 1.

4.2.3 Van Norden Bubble Test Table 9 shows the results of the likelihood ratio tests for bubbles, which compare the fit of the regime-switching model, vN (Equation 15) to that of the two simpler models, NM (Equation 17) and EC (Equation 18). For both high and low values of π, the nulls are almost always rejected in favor of the switching model.[32]

We next examine the parameters of our estimated model (Table 10). To avoid the identification problem noted in Equation 16, the parameters are normalized by setting regime S as the regime with the greater slope coefficient. The bubble model therefore implies that one should find $[\beta_C < 0,\ \beta_S > 0,\ \eta > 0]$.

When $\pi = 0.999$ or $\pi = 0.99$, we may not observe a bubble collapse in our relatively small sample of 100 observations. This could cause the estimates to miss the behavior of regime C, so that the two regimes in the regime-switching model would be based upon the slight difference between regimes E and G. This is consistent with the Monte Carlo results. We find that the median of $B_C < 0$ and the median of $B_S > 0$ except when $\pi = \{0.999, 0.99\}$.

[32]Remember that those draws where the restricted models had likelihood function values greater than the unrestricted models were included as nonrejections of the null! Including these draws did not have a great effect on the rejection rates except for $\pi = 0.5$. Here, there were very few rejections of the null, reflecting the very high (over 85%) frequency with which the restricted models gave higher values of the likelihood function than the unrestricted model.

Table 9
van Norden: Adjusted LR tests, percentage rejections.

Restriction	NM		EC	
π	5%	1%	5%	1%
.999	98.03	97.45	91.47	90.62
.99	99.24	98.90	94.97	93.81
.95	99.96	99.96	98.45	97.66
.85	99.73	99.73	98.02	97.44
.75	91.00	89.26	87.44	84.92
.5	11.46	11.40	12.01	11.96
.25	97.49	97.49	97.73	97.73

Table 10
van Norden test: Median value of parameters.

	π						
Parameter	.999	.99	.95	.85	.75	.5	.25
α_1	−0.0023	0.126	0.552	0.468	−0.541	−0.338	1.02
β_1	0.363	0.221	−0.667	−0.941	−1.09	−0.937	−1.03
α_2	0.0971	0.101	0.136	0.174	0.205	0.242	0.0897
β_2	0.749	0.628	0.221	0.339	0.470	0.854	0.0438
λ	0.211	−1.47	−2.69	−3.11	−3.23	−2.31	−185.
η	2.54	2.01	1.76	1.45	1.18	1.10	351.
σ_1	1.05	1.24	2.23	2.21	0.0987	0.0567	2.25
σ_2	0.154	0.0947	0.0643	0.0726	0.107	0.212	0.0914

Table 11
van Norden test: Percentage of draws with parameters significantly different from zero[g]

	π						
Parameter	.999	.99	.95	.85	.75	.5	.25
α_1	26.17	**16.19**	7.98	**23.1**	**53.9**	**81.58**	10.28
β_1	**55.54**	47.50	**16.9**	**51.2**	**84.20**	**96.00**	**95.98**
α_2	36.02	56.98	86.3	96.98	99.56	98.89	98.8
β_2	95.8	94.54	96.7	98.33	97.97	93.35	48.83
λ	16.58	**44.1**	**76.9**	**90.52**	**94.3**	**52.35**	**8.02**
η	47.7	60.4	83.6	91.88	90.7	29.24	6.76
σ_1	100	99.95	100	99.9	99.9	99.5	100
σ_2	99.8	99.97	100	100	100	99.9	100

g. Shaded cells are significantly less than zero, and unshaded cells are significantly greater than zero. "Significantly" different means that the t-statistic is greater than 2 or less than −2. Both percentages were calculated for each parameter. The higher percentage is reported.

As shown in Table 11, the t-statistics also provide evidence of bubbles. Independent of the level of π, regime S usually has significantly positive slope and intercept terms, while the slope term in the equation for the probability of being in regime C is usually (correctly) negative for all values of π. However, the actual percentage varies greatly. The change in π greatly affects regime C's slope. At high levels of π the slope is often found to be significantly greater than zero. The lack of actual collapses at high levels of π may be responsible for this failure to detect a collapsing regime. As π decreases, the slope is found to be significantly less than zero. This corresponds well with the presence of a bubble.

5 Conclusions

After discussing and comparing two relatively recent tests for bubbles based on regime switching, we have tried to better understand their reliability and usefulness by investigating their econometric properties. We reached broadly similar conclusions for both tests.

First, both tests seem to be conservative, underestimating the significance of the evidence of bubbles when used with the usual asymptotic critical values. This size distortion could be large, even in samples with over 400 observations. Second, the tests seem to have good power against some (but not all) bubble processes despite using the conservative asymptotic critical values.

This result contrasts with the conventional wisdom among practitioners that *any* misspecification of the model of fundamentals will tend to produce spurious evidence of bubbles. We found the opposite: by randomizing our data, we showed that using completely spurious measures of deviations from fundamentals led us to conclude that no bubbles were present. To be sure, we must carefully qualify this result. These tests are not testing for bubbles directly; they are instead testing for a particular regime-switching structure predicted by simple models of rational speculative bubbles. One must keep in mind that such structure may have other causes.

In our particular application to Canadian and U.S. stock markets, we failed to find consistent evidence of bubbles, although the Hall and Sola test revealed more evidence than the van Norden test. The size distortion was inconsequential in this case; we reached exactly the same conclusions using asymptotic and finite-sample critical values. Whether the failure to find stronger evidence of bubbles reflects a lack of power (as opposed to the absence of bubbles) likely depends on the number of bubble collapses we would have expected to observe over the sample period. If bubble collapses occurred only once or twice in the post-WWII data, these tests would be unlikely to detect them. We would expect they would be much more likely to detect bubbles that collapsed every few years.

References

Allen, Franklin, and Gary Gorton (1991). "Rational finite bubbles." Working Paper 3707, National Bureau for Economic Research, Cambridge, MA.

Bhargava, Alok (1986). "On the theory of testing for unit roots in observed time series." *Review of Economic Studies*, 53:369–384.

Blanchard, Oliver J. (1979). "Speculative bubbles, crashes, and rational expectations." *Economic Letters*, 3(4):387–389.

Buiter, Willem H., and Paolo A. Pesenti (1990). "Rational speculative bubbles in an exchange rate target zone." Discussion Paper 479, Centre for Economic Policy Research, London.

Bulow, Jeremy, and Paul Klemperer (1994). "Rational frenzies and crashes." *Journal of Political Economy*, 102(1):1–23.

Campbell, John Y., and Robert J. Shiller (1987). "Cointegration and tests of present value models." *Journal of Political Economy*, 95:1062–1088.

Campbell, John Y., and Robert J. Shiller (1988). "The dividend-price ratio and expectations of future dividends and discount factors." *Review of Financial Studies*, 1:195–228.

Ceccetti, Stephen G., Pok-Sang Lam, and Nelson C. Mark (1990). "Mean reversion in equilibrium asset prices." *American Economic Review*, 80(3):398–418.

Charemza, Wojciech W., and Derek F. Deadman (1995). "Speculative bubbles with stochastic explosive roots: The failure of unit root testing." *Journal of Empirical Finance*, 2:153–163.

Clark, Peter K. (1973). "A subordinated stochastic process model with finite variance for speculative prices." *Econometrica*, 41:135–355.

Davidson, Russell, and James G. MacKinnon (1997). "Graphical methods for investigating the size and power of test statistics." Queen's University, Discussion Paper 903 revised. (Also forthcoming in *The Manchester School*.)

De Long, J. B., Andrei Shleifer, Lawrence Summers, and R. Waldman (1990). "Noise trader risk in financial markets." *Journal of Political Economy*, 98:703–738.

Diba, Behzad T., and Herschel I. Grossman (1987). "On the inception of rational bubbles." *Quarterly Journal of Economics*, 102:697–700.

Diba, Behzad T., and Herschel I. Grossman (1988). "Explosive rational bubbles in stock prices?" *American Economic Review*, 78:520–530.

Diebold, Francis X., Joon Haeng Lee, and Gretchen C. Weinbach (1994). "Regime switching with time varying transition probabilities." In C. Hargreaves, *Nonstationary Time Series Analysis and Cointegration*. Oxford: Oxford University Press.

Durlaf, Stephen, and Mark A. Hooker (1994). "Misspecification versus bubbles in the Cagan hyperinflation model." In C. Hargreaves, *Nonstationary Time Series Analysis and Cointegration*. Oxford: Oxford University Press.

Evans, George W. (1991). "Pitfalls in testing for explosive bubbles in asset prices." *American Economic Review*, 4:922–930.

Evans, Martin D. D. (1997). "Peso problems: Their theoretical and empirical implications." In Maddala and Rao (eds.), *Handbook of Statistics: Statistical Methods in Finance*.

Evans, Martin D. D., and Karen Lewis (1995). "Do long-term swings in the dollar affect estimates of the risk premia?" *Review of Financial Studies*, 8(3):709–742.

Flood, Robert P., and Robert J. Hodrick (1990). "On testing for speculative bubbles." *Journal of Economic Perspectives*, 4(2):85–102.

Froot, K. A., and M. Obstfeld (1991). "Intrinsic bubbles: The case of stock prices." *The American Economic Review*, 81(5):1189–1214.

Funke, Michael, Stephen Hall, and Martin Sola (1994). "Rational bubbles during Poland's hyperinflation: Implications and empirical evidence." *European Economic Review*, 38:1257–1276.

Gilles, C., and S. F. LeRoy (1992). "Bubbles and charges." *International Economic Review*, 33(2):323–339.

Hall, Stephen, and Martin Sola (1993). "Testing for collapsing bubbles: An endogenous switching ADF test." Discussion Paper 15-93, London Business School.

Hamilton, James D. (1989). "A new approach to the economic analysis of nonstationary time series and the business cycle." *Econometrica*, 57(2):357–384.

Hamilton, James D. (1994). *Time Series Analysis*. Princeton: Princeton University Press.

Hamilton, James, and Charles Whiteman (1985). "The observable implications of self-fulfilling expectations." *Journal of Monetary Economics*, 16:353–374.

Hansen, Bruce (1996). "Inference when a nuisance parameter is not identified under the null hypothesis." *Econometrica*, 64(2), 413–430.

Hooker, Mark A. (1996). "Misspecification versus bubbles in hyperinflation data: Monte Carlo and interwar European evidence." Manuscript.

Kindleberger, C. P. (1989). *Manias, Panics and Crashes: A History of Financial Crises* (revised). New York: Basic Books.

LeRoy, Stephen F. (1989). "Efficient capital markets and martingales." *Journal of Economic Literature*, XXVII(4):583–622.

Obstfeld, M., and K. Rogoff (1983). "Speculative hyperinflations in maximizing models: Can we rule them out?" *Journal of Political Economy*, 91:675–687.

Obstfeld, M., and K. Rogoff (1986). "Ruling out divergent speculative bubbles." *Journal of Monetary Economics*, 17(3):349–362.

Schaller, Huntley, and Simon van Norden. (1997). "Regime switching in stock market returns." *Applied Financial Economics*, 7:177–191.

Shiller, Robert J. (1981). "Do stock prices move too much to be justified by subsequent changes in dividends?" *American Economic Review*, 71:421–436.

Tirole, J. (1982). "On the possibility of speculation under rational expectations." *Econometrica*, 50(5):1163–1181.

Tirole, J. (1985). "Asset bubbles and overlapping generations." *Econometrica*, 53:1071–1100.

van Norden, Simon (1996). "Regime switching as a test for exchange rate bubbles." *Journal of Applied Econometrics*, 11:219–251.

van Norden, S., and H. Schaller (1993). "The predictability of stock market regime: Evidence from the Toronto stock exchange." *The Review of Economics and Statistics*, 75(3):505–510.

van Norden, S., and H. Schaller (1996). "Speculative behaviour, regime-switching, and stock market crashes." Working Paper 96-13, Bank of Canada.

van Norden, Simon, and Robert Vigfusson (1996). "Regime-switching models: A guide to the Bank of Canada Gauss procedures." Working Paper 96-3, Bank of Canada.

Weil, P. (1990). "On the possibility of price decreasing bubbles." *Econometrica*, 58(6):1467–1474.

West, Kenneth D. (1987). "A specification test for speculative bubbles." *Quarterly Journal of Economics*, 102:553–580.

Editorial Policy

The *SNDE* is formed in recognition that advances in statistics and dynamical systems theory may increase our understanding of economic and financial markets. The journal will seek both theoretical and applied papers that characterize and motivate nonlinear phenomena. Researchers will be encouraged to assist replication of empirical results by providing copies of data and programs online. Algorithms and rapid communications will also be published.

ISSN 1081-1826