

**LAG LENGTH SELECTION FOR UNIT ROOT TESTS  
IN THE PRESENCE OF NONSTATIONARY VOLATILITY**

by

**Giuseppe Cavaliere, Peter C. B. Phillips,  
Stephan Smeekes, and A. M. Robert Taylor**

**COWLES FOUNDATION PAPER NO. 1460**



**COWLES FOUNDATION FOR RESEARCH IN ECONOMICS  
YALE UNIVERSITY  
Box 208281  
New Haven, Connecticut 06520-8281**

**2015**

**<http://cowles.econ.yale.edu/>**

# Lag Length Selection for Unit Root Tests in the Presence of Nonstationary Volatility

Giuseppe Cavaliere<sup>1</sup>, Peter C. B. Phillips<sup>2</sup>, Stephan Smeekes<sup>3</sup>, and  
A. M. Robert Taylor<sup>4</sup>

<sup>1</sup>*Department of Statistical Sciences, University of Bologna, Bologna, Italy*

<sup>2</sup>*Cowles Foundation for Research in Economics, Yale University, New Haven, Connecticut, USA; Department of Economics, University of Auckland, New Zealand; Department of Economics, University of Southampton, U.K.; and Singapore Management University, Singapore*

<sup>3</sup>*Department of Quantitative Economics, Maastricht University, Maastricht, The Netherlands*

<sup>4</sup>*Essex Business School, University of Essex, Colchester, U.K.*

A number of recent papers have focused on the problem of testing for a unit root in the case where the driving shocks may be unconditionally heteroskedastic. These papers have, however, taken the lag length in the unit root test regression to be a deterministic function of the sample size, rather than data-determined, the latter being standard empirical practice. We investigate the finite sample impact of unconditional heteroskedasticity on conventional data-dependent lag selection methods in augmented Dickey–Fuller type regressions and propose new lag selection criteria which allow for unconditional heteroskedasticity. Standard lag selection methods are shown to have a tendency to over-fit the lag order under heteroskedasticity, resulting in significant power losses in the (wild bootstrap implementation of the) augmented Dickey–Fuller tests under the alternative. The proposed new lag selection criteria are shown to avoid this problem yet deliver unit root tests with almost identical finite sample properties as the corresponding tests based on conventional lag selection when the shocks are homoskedastic.

**Keywords** Information criteria; Lag selection; Nonstationary volatility; Unit root test; Wild bootstrap.

**JEL Classification** C22; C15.

## 1. INTRODUCTION

Applied researchers have recently focused attention on the question of whether or not the variability in the shocks driving macroeconomic time series has changed over time; see, e.g., the literature review in Buseti and Taylor (2003). The empirical evidence has suggested that time-varying behaviour (specifically, a general decline) in unconditional volatility in the shocks driving macroeconomic time series over the past two decades or so is a relatively common phenomena, consonant with the so-called great moderation; see, inter alia, Kim and Nelson (1999), McConnell and Perez Quiros (2000), Van Dijk et al. (2002), Sensier and Van Dijk (2004), and references therein.<sup>1</sup> Sensier and Van Dijk (2004), e.g., report that over 80% of the real and price variables in the Stock and Watson (1999) data-set reject the null of constant unconditional innovation variance. Empirical evidence also suggests that data are often characterized by smooth volatility changes rather than by abrupt changes (see, inter alia, Van Dijk et al., 2002).

Such, nonstationary volatility, effects can significantly impact on the size of standard unit root tests, even asymptotically, as has been shown by Cavaliere and Taylor (2007, 2008), among others. A solution to this problem is analyzed by Cavaliere and Taylor (2008, 2009b), who employ the wild bootstrap to capture the nonstationary volatility within the re-sampled data. They show that the wild bootstrap correctly reproduces the first-order limiting null distribution under nonstationary volatility, thereby allowing for the construction of asymptotically valid bootstrap tests.

The analysis in Cavaliere and Taylor (2008, 2009b) is based on the use of a lag length in the augmented Dickey–Fuller (ADF) test regression which is a deterministic function of the sample size. In practice, however, applied researchers usually choose the lag length by data-dependent methods. Often this is done using standard information criteria or by sequential  $t$ -testing (using conventional critical values) on the significance on the highest lag. However, both of these approaches are misspecified in the presence of nonstationary volatility: standard information criteria are based on the assumption of constant volatility, while the limit distributions used in sequential  $t$ -testing are affected by the presence of nonstationary volatility. As such, if nonstationary volatility is present in the data, the lag length selected by the applied researcher may not be appropriate. While not necessarily invalidating the asymptotic properties of the unit root test, this may nonetheless have a significant impact on finite sample performance.

In this paper, we analyze the finite sample effects of nonstationary volatility on the selection of the lag order in (bootstrap) unit root testing. Using Monte Carlo simulation methods we will show that, under certain time-varying volatility specifications, standard information criteria select too many lags and that this has a significant negative effect on

---

<sup>1</sup>The recent financial turmoil and disruption of economic activity associated with the onset of the 2008 credit crisis has undoubtedly reversed this decline and produced a corresponding rise in unconditional volatility. Such changes reinforce the need to allow for the possibility of non-constancy in unconditional volatility.

the power of the resulting unit root test. As a consequence, we also propose a modification of the standard information criteria, based on the approach of Beare (2008) which rescales the data by an estimate of the underlying volatility process. Again using Monte Carlo methods, we show that these new criteria are considerably more robust, in terms of the lag length they select, than the standard criteria in the presence of nonstationary volatility and perform very similarly to the standard criteria when volatility is constant. We show that this results in unit root tests which display significantly more power than those based on the standard lag selection criteria under nonstationary volatility yet do not lose power relative to these tests when volatility is constant. Moreover, the sizes of the tests based on the standard and new criteria are shown to be broadly the same under both constant and nonconstant volatility environments.

The structure of the paper is as follows. In Section 2 we introduce our reference data generating process (DGP) and detail the class of heteroskedastic volatility processes under which we will work. The (wild bootstrap) unit root tests, and associated lag selection criteria with the new heteroskedasticity-robust modification thereof, are discussed in Section 3. The finite-sample properties of the standard and new lag selection criteria, along with the size and power properties of the associated (wild bootstrap) unit root tests, are explored through Monte Carlo simulation in Section 4. Section 5 concludes the article.

## 2. THE HETEROSKEDASTIC MODEL

Consider the case where we have  $T + 1$  observations generated according to the DGP,

$$y_t = x_t + \beta' z_t, \quad t = 0, 1, \dots, T, \tag{1a}$$

$$x_t = \rho x_{t-1} + u_t, \tag{1b}$$

$$u_t = \varepsilon_t + \sum_{j=1}^{\infty} \psi_j \varepsilon_{t-j} =: \psi(L)\varepsilon_t, \tag{1c}$$

$$\varepsilon_t = \sigma_t e_t \tag{1d}$$

with  $E(x_0^2) < \infty$ . Our focus in this paper is on tests for whether or not  $y_t$  contains a unit root; that is, on testing  $H_0 : \rho = 1$  against  $H_1 : |\rho| < 1$  in (1).

In (1a),  $z_t$  is a vector of deterministic components. As in Ng and Perron (2001) we focus on the  $\kappa$ th-order trend function,  $z_t := (1, t, \dots, t^\kappa)'$ , with special focus on the leading cases of a constant ( $\kappa = 0$ ) and linear trend ( $\kappa = 1$ ). We also make the following assumptions on the shocks  $u_t$ , where  $\mathcal{D} := D[0, 1]$  denotes the space of right continuous with left limit (càdlàg) processes:

**Assumption 1.** (i)  $\psi(z) \neq 0$  for all  $|z| \leq 1$ , and  $\sum_{j=1}^{\infty} j|\psi_j| < \infty$ . (ii)  $e_t$  is i.i.d. with  $Ee_t = 0$ ,  $Ee_t^2 = 1$  and  $E|e_t|^{4+a} < \infty$  for some  $a > 0$ . (iii) The volatility term  $\sigma_t$  satisfies  $\sigma_{\lfloor Tr \rfloor} = \omega(r)$  for all  $r \in [0, 1]$ , where  $\omega(\cdot) \in \mathcal{D}$  is nonstochastic, twice-differentiable and strictly positive.

**Remark 1.** Assumption 1 corresponds to the set of conditions imposed on the shocks in Cavaliere and Taylor (2008) and Smeekes and Taylor (2012), strengthened by the addition of condition (iii). This additional condition ensures that the new heteroskedasticity-robust information criteria, which we propose in section 3 below, are based on a consistent estimate of the volatility process; see Beare (2008) who shows that (iii) suffices for this purpose when using a nonparametric kernel estimator. As the conditions in Assumption 1 are stronger than those in Cavaliere and Taylor (2008) and Smeekes and Taylor (2012), the large sample validity of the bootstrap unit root tests discussed in the next section is guaranteed. The reader is directed to Cavaliere and Taylor (2008) and Beare (2008) for further discussion of the conditions imposed by Assumption 1. Notice that Assumption 1 contains unconditional homoskedasticity as a special case, but does not allow for models of conditional heteroskedasticity.

### 3. UNIT ROOT TESTING AND INFORMATION CRITERIA

#### 3.1. Bootstrap Unit Root Tests

In this paper we focus attention on wild bootstrap implementations of the ADF tests because of their enduring popularity with practitioners. However, the analysis provided in this paper is also valid for any unit root test that requires an autoregressive lag order to be selected. The ADF  $t$ -statistic is the usual regression  $t$ -statistic of significance on  $\gamma$ , denoted  $t_\gamma^d$  in what follows, in the ADF regression

$$\Delta y_t^d = \gamma y_{t-1}^d + \sum_{j=1}^p \phi_{p,j} \Delta y_{t-j}^d + \varepsilon_{p,t}^d, \quad t = p+1, \dots, T. \quad (2)$$

where  $y_t^d := y_t - \hat{\beta}' z_t$  is the de-trended analogue of  $y_t$ , where the parameter estimate  $\hat{\beta}$  can be obtained either by the ordinary least squares (OLS) or the quasi-difference (QD) regression of  $y_t$  on  $z_t$ ; see, e.g., Elliott et al. (1996). In the context of (2),  $p$  is the lag truncation order. We defer a discussion of the criteria that will be used to estimate  $p$  until Sections 3.2 and 3.3.

Under nonstationary volatility, the  $t_\gamma^d$  statistic is not asymptotically pivotal and the associated ADF test can display very large size distortions; see Cavaliere and Taylor (2008, 2009b). One solution to this problem, studied by Cavaliere and Taylor (2008, 2009b) and Smeekes and Taylor (2012) among others, is to apply the wild bootstrap principle. Cavaliere and Taylor (2008, 2009b) demonstrate the asymptotic validity of this approach, for the case of a deterministic lag length satisfying Assumption 2 below, and give simulation results which show that the method works well in finite samples. Hence, our focus in what follows will be on wild bootstrap implementations of the ADF test where data-dependent methods are used to select the lag length in (2). We now outline the wild bootstrap algorithm we will use.

**Algorithm 1.**

1. Calculate  $y_t^d := y_t - \hat{\beta}'z_t$ , where  $\hat{\beta}$  is obtained either by the OLS or QD regression of  $y_t$  on  $z_t$ ,  $t = 0, \dots, T$ .
2. Estimate by OLS the ADF regression in (2) using a lag order,  $q$ , to obtain the ADF residuals

$$\check{\varepsilon}_{q,t}^d := \Delta y_t^d - \check{\gamma}y_{t-1}^d - \sum_{j=1}^q \check{\phi}_{q,j} \Delta y_{t-j}^d, \quad t = 1, \dots, T, \tag{3}$$

by defining  $y_{-1}^d, \dots, y_{-q}^d := 0$ .<sup>2</sup>

3. Construct (wild) bootstrap errors  $\varepsilon_t^*$  according to the device  $\varepsilon_t^* := \xi_t \check{\varepsilon}_{q,t}^d$ , where  $\xi_t$  satisfies  $E(\xi_t) = 0$  and  $E(\xi_t^2) = 1$ .<sup>3</sup>
4. Build  $u_t^*$  recursively as  $u_t^* = \sum_{j=1}^q \check{\phi}_{q,j} u_{t-j}^* + \varepsilon_t^*$ , using the estimated parameters  $\check{\phi}_{q,j}$  from Step 2 (initialized at  $u_0^*, \dots, u_{1-q}^* = 0$ ), and build  $y_t^*$  as  $y_t^* = y_{t-1}^* + u_t^*$ ,  $t = 1, \dots, T$ , initialized at  $y_0^* = 0$ .
5. Using the bootstrap sample  $y_t^*$ , apply the same method of detrending as applied to the original sample in step 1 to obtain the detrended bootstrap series  $y_t^{d*} := y_t^* - \hat{\beta}^{*'}z_t$ , where  $\hat{\beta}^*$  is defined analogously as in step 1, but with the bootstrap data. Calculate the bootstrap augmented ADF statistic, denoted  $t_\gamma^{d*}$ , from the bootstrap analogue of the ADF regression, with lag truncation  $p^*$ ,

$$\Delta y_t^{d*} = \gamma^* y_{t-1}^{d*} + \sum_{j=1}^{p^*} \phi_{p^*,j} \Delta y_{t-j}^{d*} + \varepsilon_{p^*,t}^{d*}, \quad t = p^* + 1, \dots, T. \tag{4}$$

6. Repeat Steps 3 to 5  $N$  times, obtaining bootstrap test statistics,  $t_{\gamma,b}^{d*}$  say, for  $b = 1, \dots, N$ , and calculate the bootstrap critical value

$$cv^{d*}(\pi) := \max\{x : N^{-1} \sum_{b=1}^N I(t_{\gamma,b}^{d*} < x) \leq \pi\}$$

or, equivalently, as the  $\pi$ -quantile of the ordered  $\{t_{\gamma,b}^{d*}\}_{b=1}^N$  statistics. Reject the null of a unit root if  $t_\gamma^d$  is smaller than  $cv^{d*}(\pi)$ , where  $\pi$  is the nominal level of the test.

**Remark 2.** In this paper we do not consider the question of whether OLS or QD detrending should be preferred in unit root testing. This depends critically on the

---

<sup>2</sup>This initialization ensures that we obtain  $T$  residuals and, hence,  $T$  bootstrap errors in Step 4. An asymptotically equivalent alternative is to omit this initialisation thereby yielding only  $T - q$  residuals, but to then initialize the recursion in Step 4 with the first  $q$  detrended sample values. We found virtually no differences between the two schemes for the sample sizes considered.

<sup>3</sup>In this paper, we take  $\xi_t$  to be standard normal. Other choices are also possible, although Cavaliere and Taylor (2008, Remark 6) mention that this has almost no impact on finite sample behavior.

initial condition, as is now well known in the unit root literature, see, e.g., Müller and Elliott (2003). Harvey et al. (2009) propose a union of OLS and QD detrended tests if there is uncertainty about the initial condition. This approach is extended to allow for nonstationary volatility using the wild bootstrap by Smeekes and Taylor (2012). However, using such a union-based approach still requires one to select lag lengths for use in ADF regressions. As such the problem and remedies considered in the current paper directly extend to the union tests, which is why we do not treat them explicitly in this paper. A different question is whether OLS or QD detrending should be used in the lag length selection procedure itself; as elaborated on below, Perron and Qu (2007) find that OLS is superior in this context.

### 3.2. Standard Lag Selection Criteria

While the wild bootstrap procedure outlined in Algorithm 1 takes account of any possible nonstationary volatility in the shocks without the need to parametrically model the volatility process, the presence of the lagged dependent variables in (2) is required to parametrically account for any serial correlation in the shocks. Consequently, in order to implement the ADF test, the selection of an appropriate lag length in (2), and indeed in (3) and (4), is required.

It is unrealistic to assume that the true value of  $p$ ,  $p_0$  say, in (2) is known to the practitioner, since the nature of the serial correlation in  $u_t$  cannot be reasonably assumed known. Indeed,  $p_0$  may be infinite, as is the case, for example, if  $u_t$  is a finite-order moving average (MA) process. In such cases, it is well known, see for example Chang and Park (2002), that if the lag truncation order in (2) satisfies the following deterministic rate condition,

**Assumption 2.** *Let  $p \rightarrow \infty$  and  $p = o(T^{1/3})$  as  $T \rightarrow \infty$ .*

Then, provided  $\varepsilon_t$  in (1c) is either homoskedastic or conditionally heteroskedastic (but unconditionally homoskedastic), the resulting ADF statistic,  $t_\gamma^d$ , will have the usual Dickey–Fuller limiting null distribution free of serial correlation nuisance parameters; as tabulated for the case of OLS detrending in Fuller (1996, p. 642) and for QD detrending in Elliott et al. (1996, p. 825). As noted in Section 3.1, Cavaliere and Taylor (2008, 2009b) demonstrate a corresponding result for the case where  $\varepsilon_t$  is unconditionally heteroskedastic; here the limiting null distribution of  $t_\gamma^d$  remains free of serial correlation nuisance parameters but does now depend on the form of the underlying volatility process.

As pointed out by Cavaliere and Taylor (2009a, Section 3.3), the sieve, or re-colouring, device in Step 4 of Algorithm 1 is motivated purely by finite sample concerns, and  $q$  does

not therefore have to increase to infinity with the sample size.<sup>4</sup> Also, although  $p^*$  is not required to diverge with  $T$ , we do require that  $q \leq p^*$  for large  $T$ .<sup>5</sup> Specifically, we make the following assumptions on  $q$  and  $p^*$ .

**Assumption 3.** (i) Let  $p^* = o(T^{1/3})$ ; (ii) there is a  $T^*$  such that  $q \leq p^*$  for all  $T > T^*$ .

For a given sample size, Assumptions 2 and 3 give no practical guidance on how to select the lag length in (2), (3), and (4). A popular choice, which permits a trade-off between the size distortions that result from including too few lags and the power losses that obtain when too many lags are included, is to base it on an information criterion (see also Remark 3). This estimates the lag length as

$$\hat{p} := \arg \min_{p_{\min} \leq k \leq p_{\max}} IC(k), \quad IC(k) := \ln \hat{\sigma}_k^2 + k \frac{C_T}{T}, \tag{5}$$

where  $\hat{\sigma}_k := (T - p_{\max})^{-1} \sum_{t=p_{\max}+1}^T (\hat{\epsilon}_{k,t}^d)^2$  with  $\hat{\epsilon}_{k,t}^d$  the OLS residuals from the  $k$ th order ADF regression for  $y_t^d$  in (2); that is,  $\hat{\epsilon}_{k,t}^d := \Delta y_t^d - \hat{\gamma} y_{t-1}^d - \sum_{j=1}^k \hat{\phi}_{k,j} \Delta y_{t-j}^d$ , and where  $p_{\min} \leq p_{\max}$  are selected such that  $p_{\min}, p_{\max} \rightarrow \infty$  as  $T \rightarrow \infty$  with  $p_{\max}$  satisfying the rate condition in Assumption 2. In (5),  $C_T$  is a penalty function that differs according to the specific information criterion to be used; for the Akaike information criterion (AIC)  $C_T := 2$ , for the Bayesian information criterion (BIC)  $C_T := \ln T$ . Tsay (1984) shows that for finite  $p$ , the properties of AIC and BIC in the stationary case remain the same in the presence of unit roots; i.e., BIC is consistent while AIC is not (it overestimates with a positive probability). Pötscher (1989) extends these results to allow for nonconstant volatility in the errors, and finds that BIC is still consistent for the setting considered in this paper.

Ng and Perron (2001) propose a class of modified information criteria (MIC), motivated specifically for selecting the lag length in the ADF regression, (2), of the form

$$MIC(k) := \ln \hat{\sigma}_k^2 + k \frac{C_T + \tau_T(k)}{T},$$

where  $\tau_T(k) := (\hat{\sigma}_k^2)^{-1} \hat{\gamma}^2 \sum_{t=p_{\max}+1}^T (y_{t-1}^d)^2$ . The associated lag length estimate is then defined as in (5) but replacing  $IC(k)$  by  $MIC(k)$  in the definition of  $\hat{p}$ . The penalty function  $C_T$  is selected as for the original criteria; e.g.,  $C_T := 2$  and  $C_T := \ln T$  yield the modified AIC (MAIC) and criterion and the modified BIC (MBIC) criterion respectively. Although

<sup>4</sup>This differs from the approach taken by Smeekes and Taylor (2012, Assumption 5) for reasons explained in their Remark 15.

<sup>5</sup>Cavaliere and Taylor (2009a) assume that  $q \leq p^*$  for all  $T$  but this is not necessary for the validity of the bootstrap. By allowing  $p^*$  to be smaller than  $q$  one can replicate the effect of underfitting the lag length in the bootstrap, which may improve finite sample performance (cf. Richard, 2009).



asymptotically the properties of the original criteria will be maintained, Ng and Perron (2001) show that these modified criteria yield large improvements over the standard criteria for the purpose of unit root testing, in particular if a negative moving average parameter is present in the short-run dynamics. Perron and Qu (2007) propose a further modification of these criteria, by suggesting that they should always be applied to OLS rather than QD detrended data even if the unit root test itself is based on QD detrended data. This will improve the power properties of the test, in particular for alternatives further from the null.

Note that provided  $p_{\min}$  and  $p_{\max}$  satisfy the conditions stated above, the limiting null distributions of  $t_{\gamma}^d$  and  $t_{\gamma}^{d*}$  will not be affected by the short-run dynamics irrespective of the asymptotic properties of the selected information criterion. Our investigation is therefore purely related to the performance of the (wild bootstrap) ADF test in finite samples, since in finite samples the lag selection criteria are misspecified if the volatility process is time-varying and cannot necessarily be relied upon to yield an appropriate estimate of the required lag lengths. This is confirmed by the simulation results we present in Section 4.

**Remark 3.** We focus here on lag length selection through information criteria rather than through sequential  $t$ -testing, as this approach has proven to be more popular and also more successful; unreported simulations show that sequential  $t$ -testing tends to select too many lags on average. Sequential  $t$ -testing can be adapted to the setting of nonstationary volatility by either using heteroskedasticity-robust standard errors, or by again applying the wild bootstrap principle.

### 3.3. Heteroskedasticity-Robust Lag Selection Criteria

In this subsection, we propose a method for lag length selection based on information criteria that is designed to be robust to heteroskedasticity. Rather than modifying the information criteria themselves, we modify the series that is the input to the information criteria. We adapt the idea proposed in Beare (2008) to lag length selection; that is, we estimate the volatility nonparametrically and then re-scale the series with the estimated volatility.

To estimate the volatility nonparametrically we use the local constant, or Nadaraya–Watson, estimator used by Beare (2008).<sup>6</sup> The volatility estimator at time  $t$  is defined as

$$\hat{\sigma}_{m,t} := \sqrt{\hat{\omega}_m^2(t/T)}, \quad \hat{\omega}_k^2(r) := \frac{\sum_{t=1}^T K\left(\frac{t/T-r}{h}\right) (\check{\mathcal{E}}_{m,t}^d)^2}{\sum_{t=1}^T K\left(\frac{t/T-r}{h}\right)}, \quad (6)$$

---

<sup>6</sup>We also considered the re-weighted local constant estimator proposed by Xu and Phillips (2011). As discussed by Xu and Phillips (2011), this estimator shares all the advantages of the local linear estimator. However, unlike the local linear estimator (but like the local constant estimator), it cannot be negative. The simulation results with this estimator were virtually identical to the results reported here with the local constant estimator and, hence, are omitted in the interests of space.

where  $\{\check{\epsilon}_{m,t}^d\}$  are defined in (3) with a lag truncation of  $m$ ,  $K(\cdot)$  is a kernel function and  $h$  is a bandwidth parameter. As in Beare (2008), the following assumption is needed on the kernel  $K(\cdot)$  and the bandwidth  $h$  in order to ensure that (6) consistently estimates the volatility.

**Assumption 4.** (i)  $K(\cdot)$  is continuously differentiable and satisfies  $\int K(x)dx > 0$ ,  $\int |xK(x)|dx < \infty$ , and  $\int |xK'(x)|dx < \infty$ . Moreover, the Fourier transform of  $K(\cdot)$ , denoted  $\psi(\cdot)$ , satisfies  $\int |x\psi(x)|dx < \infty$ . (ii)  $h \rightarrow 0$  and  $Th^4 \rightarrow \infty$  as  $T \rightarrow \infty$ .

The volatility estimates from (6) are then used to re-scale the series of interest as follows:

$$\tilde{y}_t := \sum_{s=1}^t \frac{\Delta y_s^d}{\hat{\sigma}_{m,s}}, \quad \tilde{y}_0 := 0. \tag{7}$$

The idea behind (7) is that  $\tilde{y}_t$  will be rendered (approximately) homoskedastic. The re-scaled series  $\tilde{y}_t$  is then used as input to the information criteria. The corresponding re-scaled (modified) information criteria, denoted RS(M)IC in what follows, are then calculated as

$$RSIC(k) := \ln \tilde{\sigma}_k^2 + k \frac{C_T}{T}, \quad RSMIC(k) := \ln \tilde{\sigma}_k^2 + k \frac{C_T + \tilde{\tau}_T(k)}{T},$$

where  $\tilde{\tau}_T(k) := (\tilde{\sigma}_k^2)^{-1} \tilde{\gamma}^2 \sum_{t=p_{\max}+1}^T (\tilde{y}_{t-1}^d)^2$ ,  $\tilde{\sigma}_k = (T - p_{\max})^{-1} \sum_{t=p_{\max}+1}^T (\tilde{\epsilon}_{k,t}^d)^2$ , and where  $\tilde{\epsilon}_{k,t}^d$  is the OLS residual from a  $k$ -th order ADF regression on  $\tilde{y}_t^d$ , which is either the OLS or QD detrended analogue of  $\tilde{y}_t$ .

In practice one must also select a value for the lag truncation  $m$  used in the construction of the volatility estimator in (6). The choice  $m = 0$  corresponds to Beare (2008), while taking  $m = p_{\max}$  would also seem to be a sensible choice in the lag selection framework. In this paper we will follow Beare (2008) and set  $m = 0$ , but unreported simulations showed that setting  $m = p_{\max}$  gave virtually identical results.<sup>7</sup>

#### 4. MONTE CARLO SIMULATIONS

In this section we will use Monte Carlo simulation methods to investigate the finite sample performance of the standard information criteria and their new heteroskedasticity-robust analogues developed in the previous section. Comparison is made both of the lag order selected by these criteria and of the size and power properties of the associated wild bootstrap ADF tests for a variety of homoskedastic and heteroskedastic ARMA models.

---

<sup>7</sup>Similarly, it is possible to use the residuals which are obtained when imposing the unit root null hypothesis. Unreported simulation results indicated that the results do not change in this case either.

#### 4.1. The Monte Carlo Design

In the simulation study, we use the following DGP:

$$y_t = x_t + \beta' z_t, \quad t = 0, 1, \dots, T, \quad (8a)$$

$$x_t = \rho_T x_{t-1} + u_t, \quad x_0 = 0 \quad (8b)$$

$$u_t = \phi_1 u_{t-1} + \phi_2 u_{t-2} + \phi_3 u_{t-3} + \varepsilon_t + \theta \varepsilon_{t-1} \quad (8c)$$

$$\varepsilon_t = \sigma_t e_t, \quad e_t \sim i.i.d. N(0, 1), \quad (8d)$$

for the local-to-unity setting where  $\rho_T = 1 - c/T$ , such that  $c = 0$  corresponds to the unit root null hypothesis and  $c > 0$  to local alternatives. Without loss of generality, we set  $\beta = 0$ .

We report results for the combinations of the AR and MA parameters  $\phi_1, \phi_2, \phi_3$  and  $\theta$  in (8c) given in Table 1. Table 1 also reports for each model the true value,  $p_0$ , of the associated lag augmentation in (2). These ARMA parameters allow for a range of different dynamic models, ranging from near I(2) data (models 8 and, to a lesser extent, 5 and 10, with  $\rho_T = 1$ ) to near over-differenced data (model 11 with  $\rho_T = 1$ ). The range of models is very similar to that considered by Ng and Perron (2005) and allows both finite AR (of orders 1, 2, and 3) and MA(1) models.

We report results for the following two volatility models:

1. Volatility Model 1 — smooth transition:  $\sigma_t^2 = \sigma_0^2 + (\sigma_1^2 - \sigma_0^2)S_t$ , where  $S_t = (1 + \exp(-\gamma(t - \lfloor \tau T \rfloor)/T))^{-1}$  with  $\sigma_0 = 1$ . We consider parameters  $\delta = 1/3, 3$  and  $\tau = 0.2, 0.8$  with  $\gamma = 25$ .

TABLE 1  
ARMA Models Considered

<i>Model</i>	$p_0$	$\phi_1$	$\phi_2$	$\phi_3$	$\theta$
1	0	0.00	0.00	0.00	0.00
2	1	-0.80	0.00	0.00	0.00
3	1	-0.50	0.00	0.00	0.00
4	1	0.50	0.00	0.00	0.00
5	1	0.80	0.00	0.00	0.00
6	2	0.40	0.20	0.00	0.00
7	2	1.10	-0.35	0.00	0.00
8	2	1.30	-0.35	0.00	0.00
9	3	0.30	0.20	0.10	0.00
10	3	0.10	0.20	0.30	0.00
11	$\infty$	0.00	0.00	0.00	-0.80
12	$\infty$	0.00	0.00	0.00	-0.50
13	$\infty$	0.00	0.00	0.00	0.50
14	$\infty$	0.00	0.00	0.00	0.80

2. Volatility Model 2 — stochastic volatility:  $\sigma_t^2 = \omega^2(t/T)$ , where  $\omega^2(s) = \sigma_0^2 \exp(vJ_{\tilde{c}}(s))$  and  $J_{\tilde{c}}$  is an Ornstein–Uhlenbeck process. Again we set  $\sigma_0 = 1$ , and we consider parameters  $\tilde{c} = 0, 10$  and  $v = 4, 9$ .

In the first model, a smooth (logistic) transition occurs in the variance from  $\sigma_0^2$  to  $\sigma_1^2$  centred at time  $\lfloor T\tau \rfloor$ ,  $\lfloor \cdot \rfloor$  denoting the integer part of its argument, with  $\gamma$  the speed of transition. The results from this model were very similar to those from a model with a single abrupt change in volatility (obtained by setting  $\gamma = \infty$ ). In turn these were also qualitatively similar to the results obtained from a model with two abrupt breaks in volatility when the dominant break of the two coincided with that in the single break model. Consequently, we do not report these results here; they can be obtained from the accompanying working paper, Cavaliere et al. (2012), which also contains results for a mid-sample break,  $\tau = 0.5$ .

Notice that the stochastic volatility model is not formally allowed under the assumptions needed on the kernel estimation. We chose to consider this model of volatility as it is a popular model in the literature. Moreover, good performance by the new lag selection criteria for models such as this which falls outside the class of models they are intended for can be argued to reinforce their potential. Also observe that the homoskedastic case is contained in the first models when  $\delta = 1$ , and in the second model when  $v = 0$ .

In this analysis we present results only for the MAIC criterion of Ng and Perron (2001) and the heteroskedasticity-robust analogue thereof, RSMAIC, from section 3.3. We do so because MAIC is the most popular and successful criterion used in unit root testing. However, a summary of the corresponding results for other popular lag selection methods is given at the end of this section. In the context of the MAIC and RSMAIC criteria the minimum lag length,  $p_{\min}$  was set to zero throughout, while the maximum lag length was set to  $p_{\max} = \lfloor A(T/100)^{1/4} \rfloor$ , with the choice of the constant  $A$  specified in the subsections which follow. We report results for the sample sizes  $T = 150$  and  $T = 250$ .<sup>8</sup> Throughout this section, we will only report results for the specification where a constant is included in  $z_t$  in (8a); that is demeaned data. Results for the constant and trend case are very similar, and are available on request. As recommended by Perron and Qu (2007), we apply the information criteria to OLS demeaned data. As mentioned before, the volatility estimator used in the RSMAIC is  $\hat{\sigma}_{0,t}$ , with the kernel  $K(\cdot)$  taken as the Gaussian kernel and the bandwidth set equal to  $h = 0.1$ , a value that produced good results in Beare (2008).<sup>9</sup>

<sup>8</sup>For smaller sample sizes the differences between the regular and re-scaled IC are less obviously seen, at least in part because the maximum lag length parameters will be smaller; see, for example, the additional results for  $T = 50$  reported in Cavaliere et al. (2012).

<sup>9</sup>Different specifications again gave very similar results.

#### 4.2. Selected Lag Lengths

We first focus on the lag lengths selected by the standard MAIC and the new heteroskedasticity-robust RSMAIC criteria. As part of our analysis we vary the maximum lag length,  $p_{\max}$ , by considering results for both  $A = 6$  and  $A = 12$ . In large samples and for the (low-order) autoregressive models the lag selection should not be significantly affected by changing the upper bound. If, however, a criterion is seriously affected by the choice of  $p_{\max}$  then this provides clear evidence that the criterion is not selecting the lag length appropriately for the sample sizes considered. We report lag selection results for the model under the alternative hypothesis with  $c = 7$ ; results under the null ( $c = 0$ ) are very similar and can be found in Cavaliere et al. (2012). All results are based on 5,000 simulations.

Table 2 reports the average (taken across the Monte Carlo replications) selected lag lengths obtained under homoskedasticity. It can be seen from these results that the MAIC and RSMAIC criteria perform very similarly to one another here for all of the AR and MA models considered. These results suggest that the rescaling approach used in calculating the RSMAIC criterion does not fundamentally change its properties from those of the MAIC criterion under homoskedasticity, which is a necessary condition to

TABLE 2  
Average Lag Lengths Selected by MAIC and RSMAIC. Homoskedastic Errors

Model	$p_0$	A = 6				A = 12			
		T = 150		T = 250		T = 150		T = 250	
		MAIC	RSMAIC	MAIC	RSMAIC	MAIC	RSMAIC	MAIC	RSMAIC
1	0	0.75	0.71	0.76	0.71	0.96	0.91	1.01	0.93
2	1	1.85	1.92	1.86	1.89	2.33	2.43	2.18	2.27
3	1	1.73	1.72	1.77	1.75	2.06	2.05	2.07	2.05
4	1	1.64	1.61	1.71	1.68	1.92	1.83	1.97	1.89
5	1	1.63	1.61	1.68	1.69	1.94	1.90	1.94	1.95
6	2	2.03	1.97	2.43	2.39	2.37	2.25	2.72	2.65
7	2	2.59	2.59	2.66	2.65	2.97	2.94	2.98	2.96
8	2	2.54	2.62	2.63	2.75	2.88	3.02	2.93	3.16
9	3	2.13	2.08	2.66	2.63	2.40	2.31	3.02	2.95
10	3	3.15	3.07	3.54	3.52	3.60	3.42	3.93	3.84
11	$\infty$	5.06	5.05	6.34	6.34	8.35	8.34	9.97	10.01
12	$\infty$	3.45	3.43	3.86	3.86	4.12	4.09	4.40	4.40
13	$\infty$	2.67	2.62	3.05	3.02	3.03	2.96	3.42	3.33
14	$\infty$	4.69	4.66	5.47	5.46	5.74	5.66	7.04	6.98

The DGP is (8a)–(8d) with  $\beta = 0$  and  $c = 7$ , for each of the 14 ARMA models considered in Table 1 and where  $\sigma_t$  is constant for  $T = 150$  and  $T = 250$ . The MAIC and RSMAIC lag selection criteria are as outlined in sections 3.2 and 3.3, respectively, with  $p_{\max} = \lfloor A(T/100)^{1/4} \rfloor$ . As in Perron and Qu (2007), the information criteria are applied to OLS demeaned data, for the case of  $z_t = 1$ . Results are based on 5,000 replications.

TABLE 3  
Average Lag Lengths Selected by MAIC and RSMAIC. Smooth Transition Volatility Model

Model	$p_0$	A = 6				A = 12			
		T = 150		T = 250		T = 150		T = 250	
		MAIC	RSMAIC	MAIC	RSMAIC	MAIC	RSMAIC	MAIC	RSMAIC
$\delta = 1/3, \tau = 0.8$									
1	0	1.95	0.76	2.43	0.81	3.81	1.02	5.02	1.05
2	1	2.67	1.98	3.07	1.96	4.71	2.55	5.63	2.49
3	1	2.65	1.80	3.09	1.85	4.54	2.20	5.69	2.24
4	1	2.59	1.68	3.09	1.73	4.54	1.99	5.71	2.06
5	1	2.63	1.71	3.06	1.79	4.57	2.04	5.77	2.11
6	2	2.70	1.97	3.36	2.45	4.48	2.25	6.06	2.80
7	2	3.28	2.63	3.73	2.70	5.18	3.02	6.20	3.11
8	2	3.27	2.67	3.72	2.83	5.30	3.27	6.16	3.38
9	3	2.69	2.06	3.57	2.69	4.65	2.39	6.10	3.00
10	3	3.40	3.03	4.25	3.58	5.29	3.31	6.85	3.97
11	$\infty$	4.77	5.07	6.09	6.35	8.26	8.35	10.48	10.07
12	$\infty$	3.75	3.48	4.36	3.85	5.86	4.18	7.13	4.54
13	$\infty$	3.22	2.64	3.83	3.02	5.22	3.09	6.56	3.46
14	$\infty$	4.66	4.64	5.49	5.46	7.13	5.76	8.74	6.98
$\delta = 3, \tau = 0.2$									
1	0	1.96	0.73	2.50	0.77	3.69	0.91	5.06	0.93
2	1	2.67	1.95	3.10	1.91	4.51	2.41	5.85	2.37
3	1	2.64	1.75	3.10	1.78	4.44	2.05	5.65	2.07
4	1	2.63	1.63	3.13	1.74	4.46	1.94	5.65	1.96
5	1	2.57	1.65	3.06	1.75	4.37	1.96	5.55	2.05
6	2	2.77	2.09	3.48	2.49	4.56	2.36	6.14	2.80
7	2	3.27	2.59	3.69	2.63	5.15	2.99	6.32	3.03
8	2	3.27	2.67	3.74	2.77	4.99	3.17	6.26	3.30
9	3	2.85	2.24	3.64	2.78	4.70	2.56	6.22	3.08
10	3	3.57	3.18	4.27	3.51	5.54	3.55	6.89	3.90
11	$\infty$	4.84	5.18	6.10	6.32	8.55	8.48	10.61	9.89
12	$\infty$	3.78	3.43	4.41	3.77	5.80	4.04	7.08	4.34
13	$\infty$	3.29	2.66	3.91	3.06	5.20	3.03	6.57	3.46
14	$\infty$	4.78	4.79	5.54	5.59	7.22	5.87	9.05	7.18

As for Table 2 except that  $\sigma_t$  follows the smooth transition model in Volatility Model 1.

apply it successfully. It can also be seen that for the AR models considered, other things being equal, changing the maximum lag length (through the choice of the constant  $A$ ) has only a minor impact on the average lag length selected for both criteria, as expected.

Table 3 presents the corresponding results for the case of a smooth transition break in volatility. We do not report the results here for either early positive ( $\delta = 1/3, \tau = 0.2$ ) and late negative break ( $\delta = 3, \tau = 0.8$ ) as for these models MAIC is virtually unaffected by the heteroskedasticity present, with MAIC and RSMAIC performing very similarly.<sup>10</sup>

<sup>10</sup>Full results can be found in Cavaliere et al. (2012).

On the other hand, the effect on MAIC of either a late positive or early negative break as reported here, is substantial. For these volatility models MAIC selects considerably higher lag lengths than it does under homoskedasticity. This effect can be seen for all of the ARMA models considered. Moreover, in these cases changing the maximum lag length now has a major impact on the performance of MAIC, which is again a clear indication that the standard MAIC criterion selects too many lags, approaching the upper bound as a result. In contrast, the RSMAIC criterion appears to select roughly the same number of lags in the smooth transition break models as it does under homoskedasticity; there only appears to be a minimal increase in some of the cases considered. Also, RSMAIC is far less affected by varying  $p_{\max}$  than MAIC is, which again confirms the robustness of the lag length selected by RSMAIC in this case.

Table 4 presents the corresponding average selected lags under stochastic volatility. Compared with the results in Table 2, it can again be seen that the standard MAIC criterion selects a higher lag length on average than it does under homoskedasticity when  $\tilde{c} = 0$ . Results for  $\tilde{c} = 10$  are qualitatively similar, though somewhat less pronounced, and are again available in Cavaliere et al. (2012). We now also see an increase in the average lag length selected by RSMAIC, although it still selects a considerably lower average lag length than MAIC. Hence, even though RSMAIC is affected to some degree by stochastic volatility, it remains considerably more reliable than MAIC in this setting.

To summarize, our simulation results have shown that lag length selection by MAIC is affected by the presence of nonstationary volatility in the errors. As such it cannot be reliably used to select an appropriate lag length for a unit root test in this setting. The simulation results also show that RSMAIC appears to be significantly more robust to the presence of nonstationary volatility, while its performance under homoskedasticity is almost identical to MAIC. In the context of unit root testing, it is arguably the performance of the unit root test for which lag orders are selected, rather than the actual selected lag order, which is of primary importance. If the lag selection has no effect on the size or power properties of the resulting unit root test, then there is no problem in using a potentially misspecified method such as MAIC. Therefore we will now investigate the impact of nonstationary volatility on the finite sample size and power properties of the wild bootstrap ADF unit root test, when the lag length in the ADF regression has been selected by either MAIC or RSMAIC.

#### 4.3. Rejection Frequencies of Bootstrap Unit Root Tests

In this subsection we investigate the performance of the wild bootstrap ADF unit root test from Algorithm 1, using QD detrending, and where the lag truncation order in the original ADF regression (2), the sieve regression (3), and the bootstrap ADF regression (4), were selected by either MAIC or RSMAIC, using the same tuning parameters as outlined in section 4.1, with results reported for  $A = 12$ . All results in this subsection are based on 5,000 simulations and 199 bootstrap replications.

TABLE 4  
Average Lag Lengths Selected by MAIC and RSMAIC. Stochastic Volatility Model

Model	$p_0$	A = 6				A = 12			
		T = 150		T = 250		T = 150		T = 250	
		MAIC	RSMAIC	MAIC	RSMAIC	MAIC	RSMAIC	MAIC	RSMAIC
$\bar{c} = 0, \nu = 4$									
1	0	2.15	1.06	2.85	1.29	4.12	1.42	5.56	1.75
2	1	2.81	2.21	3.37	2.40	4.80	2.97	6.22	3.30
3	1	2.77	1.98	3.33	2.16	4.74	2.48	6.12	2.79
4	1	2.79	1.88	3.36	2.09	4.67	2.25	6.17	2.55
5	1	2.77	1.95	3.36	2.21	4.68	2.43	6.02	2.79
6	2	2.86	2.20	3.64	2.73	4.84	2.63	6.45	3.32
7	2	3.38	2.79	3.92	3.00	5.35	3.31	6.66	3.63
8	2	3.38	2.89	3.99	3.20	5.32	3.70	6.73	4.23
9	3	2.90	2.33	3.88	3.01	4.81	2.70	6.63	3.60
10	3	3.53	3.15	4.38	3.72	5.51	3.61	7.17	4.33
11	$\infty$	4.70	5.04	6.02	6.27	8.30	8.27	10.42	9.95
12	$\infty$	3.79	3.47	4.50	3.96	5.84	4.24	7.29	4.85
13	$\infty$	3.30	2.79	4.10	3.30	5.27	3.29	6.85	3.93
14	$\infty$	4.75	4.71	5.54	5.55	7.14	5.89	9.02	7.34
$\bar{c} = 0, \nu = 9$									
1	0	3.19	2.23	4.34	3.05	6.13	3.48	8.64	5.12
2	1	3.47	3.08	4.46	3.83	6.39	4.78	8.82	6.42
3	1	3.54	2.88	4.50	3.57	6.47	4.26	8.81	5.77
4	1	3.50	2.71	4.50	3.48	6.41	3.88	8.83	5.59
5	1	3.61	2.86	4.59	3.62	6.54	4.19	9.09	5.77
6	2	3.54	2.91	4.63	3.80	6.56	4.11	9.01	5.83
7	2	3.89	3.39	4.77	4.08	6.81	4.68	9.07	6.21
8	2	4.02	3.54	4.90	4.31	7.10	5.29	9.38	7.06
9	3	3.56	3.02	4.67	3.96	6.61	4.26	8.99	5.97
10	3	3.91	3.56	4.99	4.43	6.87	4.78	9.31	6.53
11	$\infty$	4.37	4.93	5.70	6.08	8.11	8.21	10.62	10.21
12	$\infty$	3.98	3.69	4.90	4.43	6.93	5.26	9.07	6.71
13	$\infty$	3.79	3.33	4.81	4.10	6.74	4.55	9.11	6.21
14	$\infty$	4.75	4.78	5.61	5.66	7.97	6.64	10.32	8.56

As for Table 2 except that  $\sigma_t$  follows the stochastic volatility model in Volatility Model 2.

We first report, in Tables 5 and 6, the size properties of the wild bootstrap ADF tests based on MAIC and RSMAIC lag selection for the same set of ARMA and volatility models as were reported in the previous subsection.<sup>11</sup> Sizes for MAIC and RSMAIC seem to be comparable across the different models; both give sizes close to the nominal level of 5% except for model 11 (which has a large negative MA parameter), where there is some oversize (of roughly the same degree) seen for both methods. Overall, it does not

<sup>11</sup>Results for the remaining volatility models are available in Cavaliere et al. (2012).



TABLE 5  
Empirical Rejection Frequencies of the Wild Bootstrap ADF Test with  
QD Demeaned Data. Homoskedastic Errors

<i>Model</i>	<i>T</i> = 150		<i>T</i> = 250	
	<i>MAIC</i>	<i>RSMAIC</i>	<i>MAIC</i>	<i>RSMAIC</i>
1	0.046	0.045	0.053	0.051
2	0.047	0.041	0.047	0.045
3	0.044	0.044	0.050	0.049
4	0.049	0.057	0.053	0.052
5	0.051	0.051	0.051	0.050
6	0.043	0.042	0.049	0.047
7	0.056	0.051	0.049	0.048
8	0.051	0.052	0.047	0.046
9	0.041	0.035	0.040	0.039
10	0.054	0.052	0.053	0.053
11	0.101	0.098	0.089	0.090
12	0.041	0.059	0.056	0.054
13	0.049	0.046	0.048	0.046
14	0.041	0.040	0.039	0.040

See notes for Table 2. Bootstrap ADF tests constructed as detailed in Algorithm 1 for QD demeaned data ( $\varepsilon_t = 1$ ) using either MAIC or RMAIC lag selection. Results are for the nominal 5% significance level and based on 5,000 simulations.

appear that the choice between using MAIC or RSMAIC when choosing the lag length has a significant impact on the size of the resulting unit root test, regardless of whether the errors are homoskedastic or heteroskedastic.

We next present finite sample local power curves for the bootstrap ADF tests. In order to keep the number of graphs manageable, we need to make a selection of the ARMA models considered. To this end we report results for the i.i.d. model (model 1), the AR(1) model with  $\phi_1 = 0.5$  (model 4) and the MA(1) model with  $\theta = -0.5$  (model 12). We consider the same type of volatility models as before.

In Fig. 1 we first present the finite sample local power curves of the wild bootstrap ADF tests based on MAIC and RSMAIC lag selection for the homoskedastic model. In the homoskedastic case the power of the tests using MAIC and RSMAIC are almost identical to one another, which is again as expected given the results from Section 4.2. This shows that the power losses incurred by using the RSMAIC criterion to select the lag length when in fact the MAIC criterion is correctly specified are negligible even for  $T = 150$ .

Figures 2 and 3 give the corresponding local power curves for the smooth transition variance break model with a late positive break and an early negative break,

TABLE 6  
Empirical Rejection Frequencies of the Wild Bootstrap ADF Test with QD Demeaned Data,  
Heteroskedastic Errors

Model	Smooth transition				Stochastic volatility			
	T = 150		T = 250		T = 150		T = 250	
	MAIC	RSMAIC	MAIC	RSMAIC	MAIC	RSMAIC	MAIC	RSMAIC
	$\delta = 1/3, \tau = 0.8$				$\tilde{c} = 0, \nu = 4$			
1	0.046	0.046	0.045	0.047	0.050	0.049	0.051	0.052
2	0.044	0.047	0.051	0.051	0.046	0.047	0.046	0.047
3	0.049	0.052	0.049	0.050	0.056	0.054	0.051	0.049
4	0.049	0.055	0.045	0.050	0.048	0.055	0.052	0.055
5	0.044	0.053	0.047	0.050	0.054	0.059	0.057	0.059
6	0.047	0.048	0.049	0.050	0.048	0.050	0.051	0.057
7	0.049	0.054	0.046	0.049	0.050	0.050	0.047	0.048
8	0.044	0.051	0.049	0.050	0.048	0.052	0.047	0.050
9	0.046	0.044	0.044	0.044	0.046	0.043	0.050	0.046
10	0.043	0.050	0.046	0.052	0.048	0.048	0.055	0.053
11	0.111	0.113	0.093	0.103	0.117	0.111	0.083	0.084
12	0.058	0.064	0.054	0.064	0.065	0.067	0.056	0.060
13	0.047	0.047	0.046	0.048	0.050	0.050	0.049	0.048
14	0.043	0.043	0.048	0.053	0.049	0.048	0.050	0.051
	$\delta = 3, \tau = 0.2$				$\tilde{c} = 0, \nu = 9$			
1	0.051	0.050	0.046	0.052	0.057	0.057	0.051	0.050
2	0.043	0.047	0.046	0.057	0.057	0.055	0.050	0.051
3	0.047	0.051	0.041	0.049	0.051	0.052	0.052	0.054
4	0.054	0.054	0.053	0.058	0.050	0.053	0.053	0.059
5	0.051	0.058	0.053	0.054	0.052	0.050	0.042	0.050
6	0.050	0.055	0.050	0.057	0.048	0.045	0.050	0.054
7	0.063	0.058	0.052	0.052	0.052	0.048	0.055	0.061
8	0.046	0.049	0.045	0.049	0.044	0.049	0.046	0.047
9	0.049	0.047	0.051	0.049	0.038	0.046	0.044	0.046
10	0.046	0.056	0.046	0.050	0.043	0.052	0.049	0.056
11	0.103	0.107	0.057	0.068	0.136	0.106	0.088	0.084
12	0.061	0.066	0.049	0.056	0.065	0.059	0.049	0.056
13	0.049	0.048	0.047	0.049	0.049	0.054	0.046	0.045
14	0.060	0.060	0.054	0.059	0.050	0.052	0.053	0.055

See notes for Tables 3 and 4. Wild bootstrap ADF tests constructed as detailed in Algorithm 1 for QD demeaned data ( $z_t = 1$ ) using either MAIC or RMAIC lag selection. Results are for the nominal 5% significance level and based on 5000 simulations.

respectively.<sup>12</sup> For these models, the bootstrap ADF test based on the use of RSMAIC is clearly more powerful than the corresponding test based on MAIC. This is a direct

<sup>12</sup>Notice that the local power curves for these models are quite different from the corresponding local power curves seen in Fig. 1 under homoskedasticity, even for  $T = 250$ . This is not an effect of the lag order selection method but rather a consequence of the result that if nonstationary volatility is present, then the limiting distributions of the ADF statistic,  $t_{\tau}^d$ , under both the null hypothesis and local alternatives, and hence the asymptotic local power function of the associated bootstrap test, are functions of the underlying volatility process (cf. Cavaliere and Taylor, 2008, p. 8).

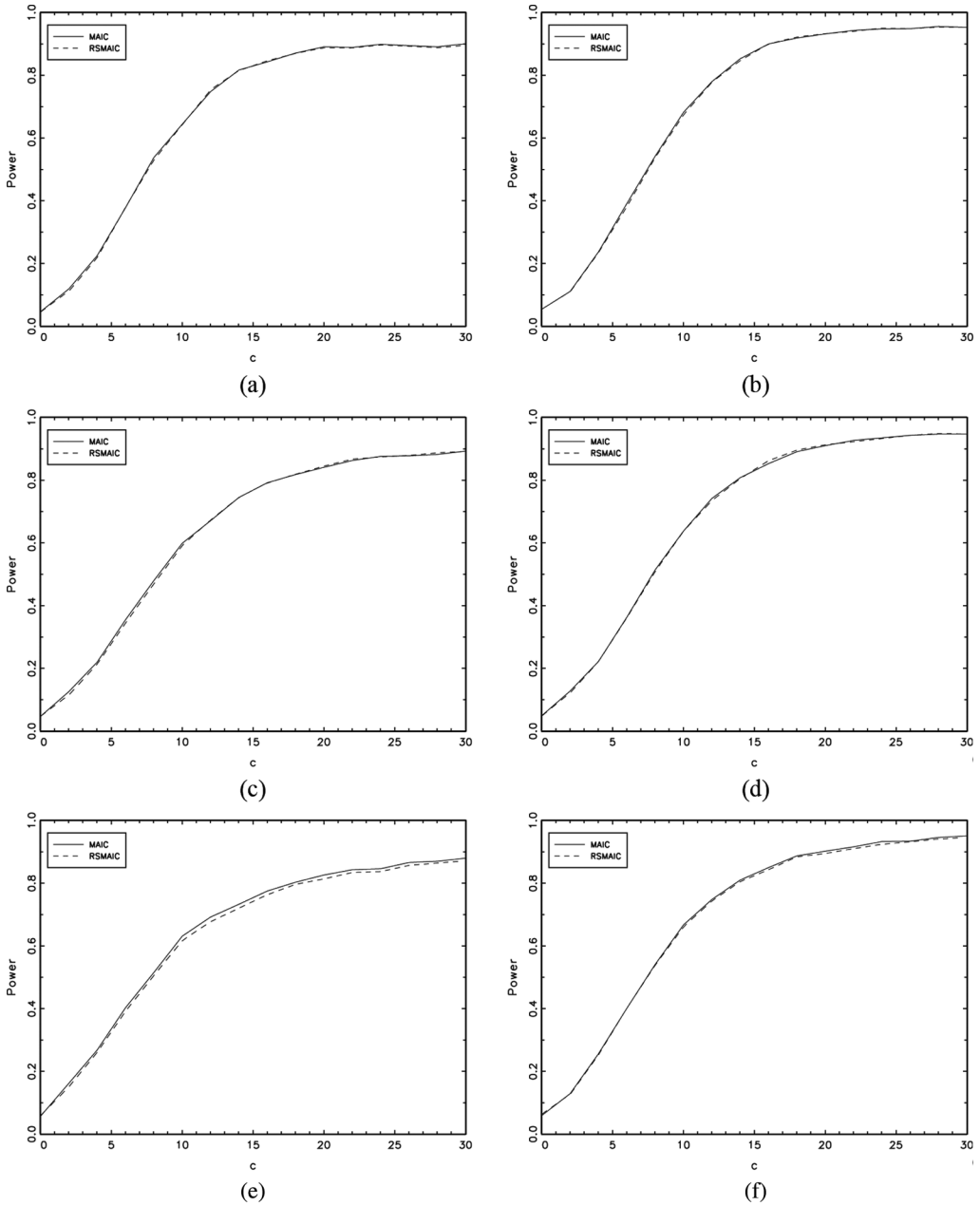


FIGURE 1 Power of the wild bootstrap ADF test with QD demeaned data. Homoskedastic errors: (a)  $T = 150$ , ARMA model 1, (b)  $T = 250$ , ARMA model 1, (c)  $T = 150$ , ARMA model 4, (d)  $T = 250$ , ARMA model 4, (e)  $T = 150$ , ARMA model 12, and (f)  $T = 250$ , ARMA model 12.

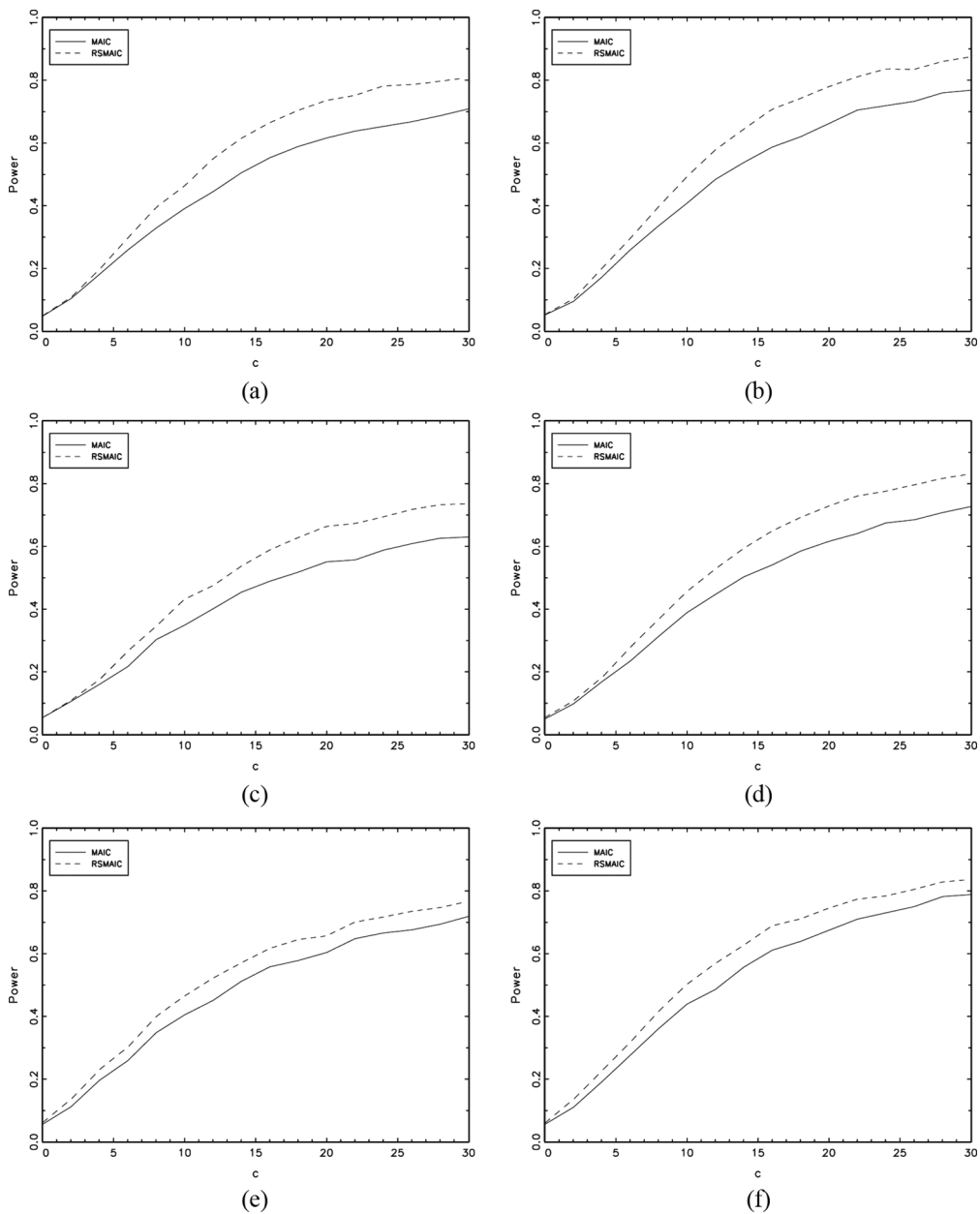


FIGURE 2 Power of the wild bootstrap ADF test with QD demeaned data. Smooth transition volatility model with  $\delta = 1/3$ ,  $\tau = 0.8$ : (a)  $T = 150$ , ARMA model 1, (b)  $T = 250$ , ARMA model 1, (c)  $T = 150$ , ARMA model 4, (d)  $T = 250$ , ARMA model 4, (e)  $T = 150$ , ARMA model 12, and (f)  $T = 250$ , ARMA model 12.

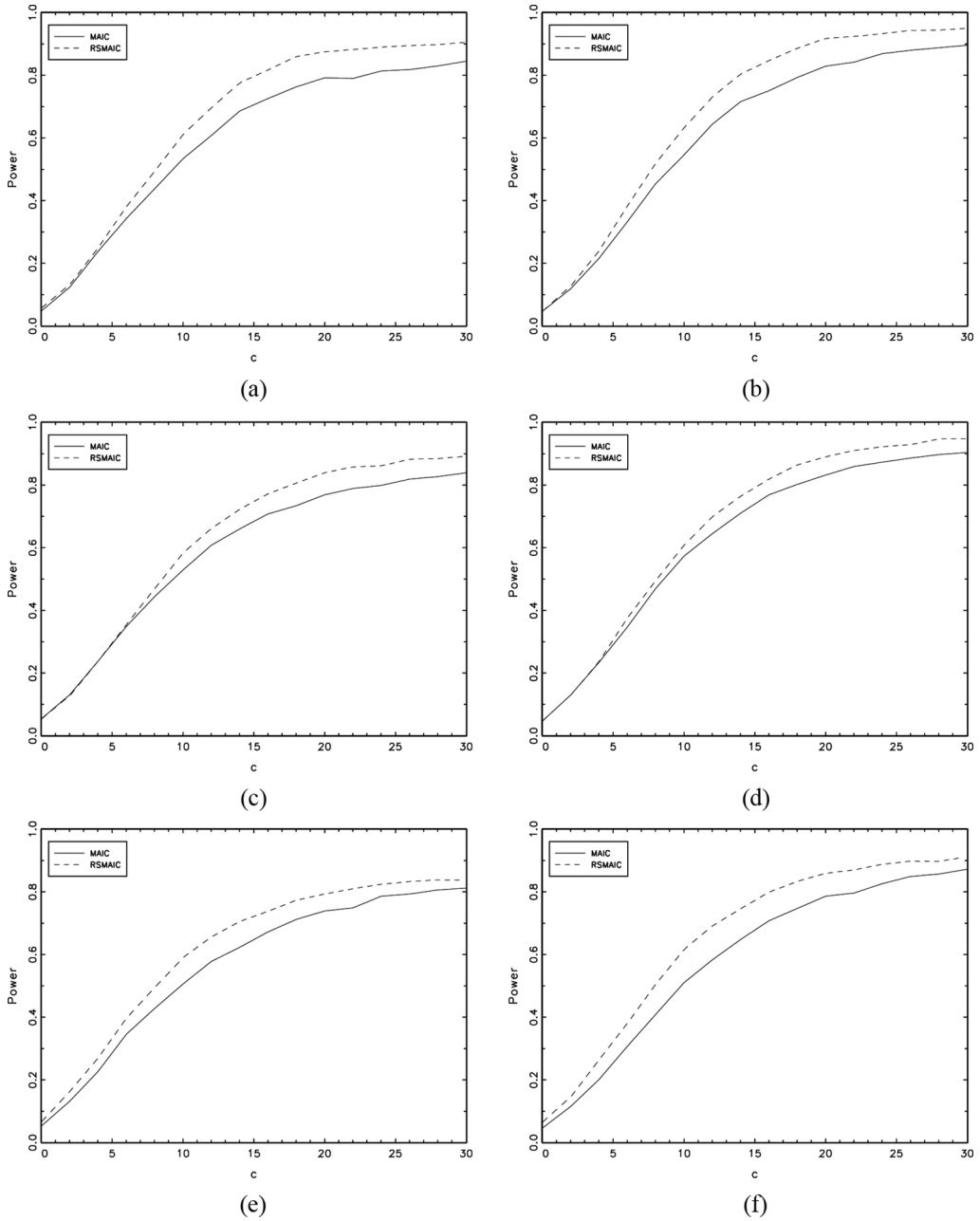


FIGURE 3 Power of the wild bootstrap ADF test with QD demeaned data. Smooth transition volatility model with  $\delta = 3$ ,  $\tau = 0.2$ : (a)  $T = 150$ , ARMA model 1, (b)  $T = 250$ , ARMA model 1, (c)  $T = 150$ , ARMA model 4, (d)  $T = 250$ , ARMA model 4, (e)  $T = 150$ , ARMA model 12, and (f)  $T = 250$ , ARMA model 12.

consequence of the results reported in Section 4.2 which showed that the MAIC criterion significantly over-fits the lag order relative to the RSMAIC criterion for these designs. It is clear that in these cases there are considerable finite sample power gains available by using RSMAIC. Moreover, the power differences between using MAIC and RSMAIC lag selection even increase slightly between  $T = 150$  and  $T = 250$ , which appears to be related to the associated increase in the maximum lag length,  $p_{\max}$ , between the two sample sizes.

Figures 4 and 5 graph the finite sample local power curves for the stochastic volatility models with  $\tilde{c} = 0$  and  $\nu = 4, 9$ . While the bootstrap ADF test based on RSMAIC is still more powerful than the corresponding test based on MAIC, the difference between the two is now rather smaller than was seen for the smooth transition break in volatility models. This is to be expected from the results on the average lag length selected by these two criteria in section 4.2, which showed that RSMAIC has a tendency to over-fit the lag length in this case, although not to the same extent as is seen with MAIC. While the gains of using RSMAIC may be smaller for the stochastic volatility case, it is nonetheless important to note that there is never a loss in power when using RSMAIC rather than MAIC to select the lag length.

We can summarize the results in this subsection by observing that lag order selection based on MAIC has a negative impact on the finite sample power of the resulting wild bootstrap ADF unit root test if nonstationary volatility is present, with the extent of this effect depending on the specific volatility model. Based on our results, we recommend the use of the RSMAIC lag selection criterion for selecting the lag length in the context of ADF unit root testing, given its greater degree of robustness to nonstationary volatility than the standard MAIC lag selection criterion, and the resulting higher finite sample power which is achievable when using RSMAIC over MAIC. These power gains are most strongly seen for the smooth break in volatility models. Moreover, under homoskedasticity we found almost no differences in power between the unit root tests which use RSMAIC and MAIC to select the lag order. Under all of the volatility and ARMA models considered the finite sample size properties of the unit root tests based on MAIC and RSMAIC were virtually identical. As such we believe it provides a reliable practical alternative to MAIC.

We conclude this section by noting that the conclusions drawn above concerning wild bootstrap ADF tests based on the MAIC lag selection method and its re-scaled analogue, RSMAIC, all carry through qualitatively to the corresponding ADF tests based on other information criteria such as AIC and BIC (where the re-scaling in computing their heteroskedasticity-robust analogues is done identically). We also considered sequential  $t$ -tests for specifying the lag truncation order, as in Ng and Perron (1995), comparing their standard approach with modifications thereof based on either the use of White (1980) heteroskedasticity-robust standard errors or the wild bootstrap. Simulations indicated that sequential  $t$ -testing is affected by nonstationary volatility in much the same way as the information criteria reported here. Using White standard errors helps to alleviate the problems, but does not erase them. Wild bootstrap ADF tests using lag selection

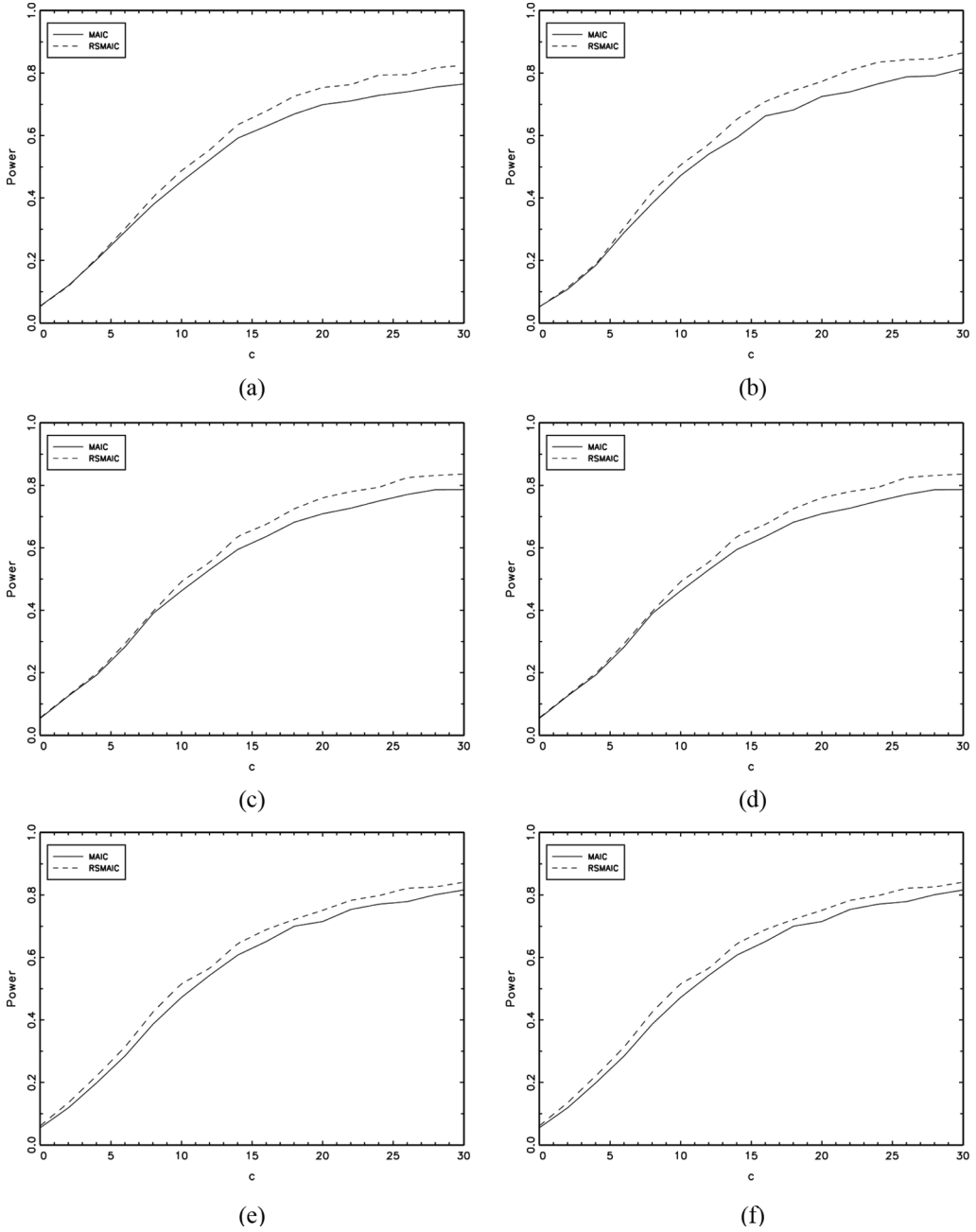


FIGURE 4 Power of the wild bootstrap ADF test with QD demeaned data. Stochastic volatility model with  $\tilde{c} = 0$ ,  $\nu = 4$ : (a)  $T = 150$ , ARMA model 1, (b)  $T = 250$ , ARMA model 1, (c)  $T = 150$ , ARMA model 4, (d)  $T = 250$ , ARMA model 4, (e)  $T = 150$ , ARMA model 12, and (f)  $T = 250$ , ARMA model 12.

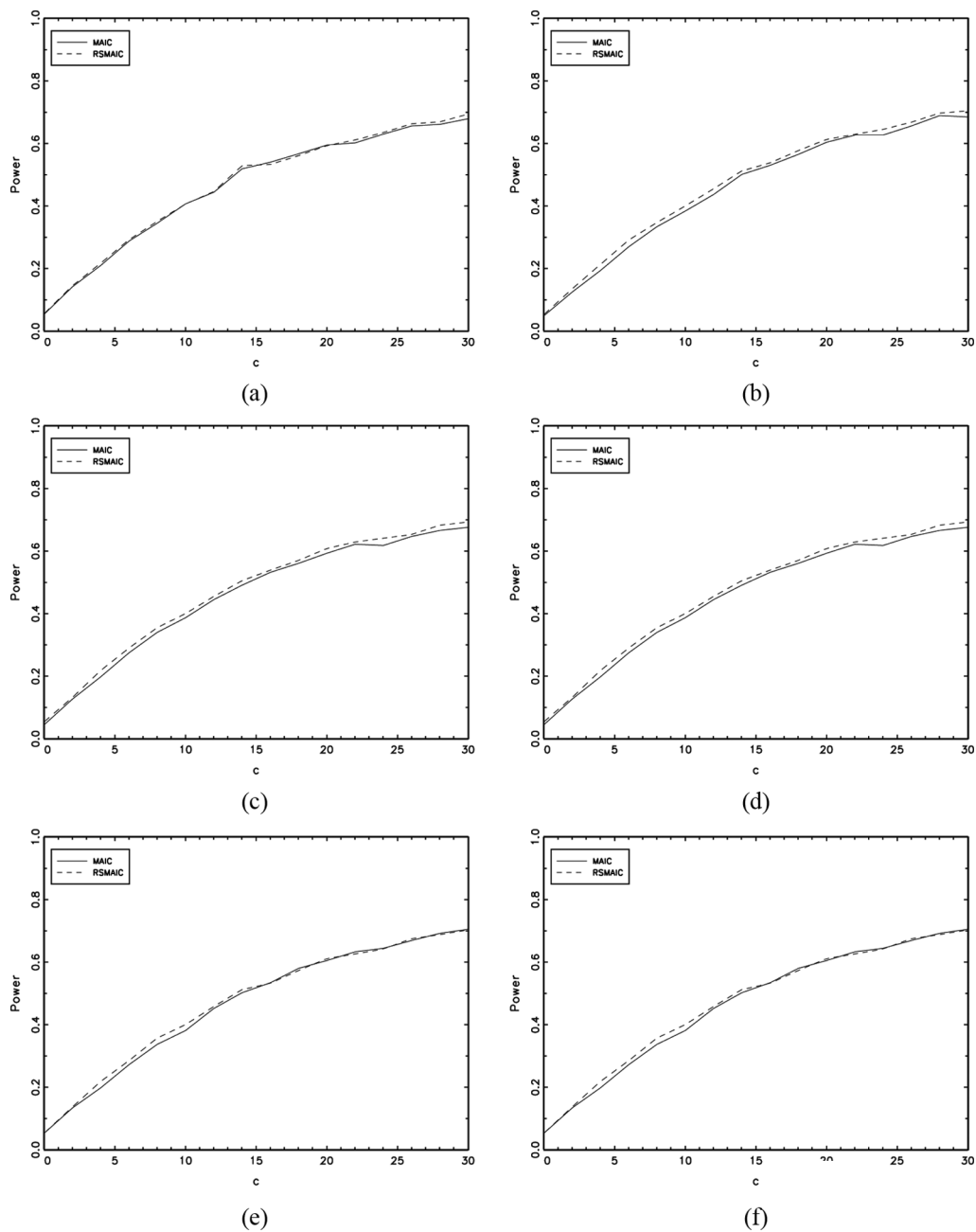


FIGURE 5 Power of the wild bootstrap ADF test with QD demeaned data. Stochastic volatility model with  $\tilde{c} = 0$ ,  $v = 9$ : (a)  $T = 150$ , ARMA model 1, (b)  $T = 250$ , ARMA model 1, (c)  $T = 150$ , ARMA model 4, (d)  $T = 250$ , ARMA model 4, (e)  $T = 150$ , ARMA model 12, and (f)  $T = 250$ , ARMA model 12.



based on wild bootstrap sequential  $t$ -tests, like the tests based on the RSMAIC method, achieve higher power than the tests based on the standard sequential  $t$ -tests but have the considerable drawback that they take a very long time to compute. Moreover, we found them to be generally inferior than the tests based on RSMAIC, and so we do not report these results in detail. They are, however, available on request.

## 5. CONCLUSION

We have investigated the effect of nonstationary volatility on lag length selection in the context of unit root testing, proposing a modification of the popular information criteria used for lag length selection, designed to be robust against nonstationary volatility. The modification consisted of rescaling the data by a nonparametric estimate of the volatility process before computing the information criterion of interest.

Focusing on the popular MAIC criterion, we found that nonstationary volatility can have a significant impact on lag length selection in finite samples. Simulations for several volatility models showed that the lag order was often overfitted, with the selected lag length being highly dependent on the maximum lag length allowed in certain cases. Our proposed re-scaled MAIC, labeled RSMAIC, criterion did not demonstrate this feature and was shown to be robust to nonstationary volatility, most notably a break in volatility. Moreover, the RSMAIC criterion was shown to perform almost identically to the MAIC criterion in terms of the lag order selected under homoskedasticity.

We then investigated the relative behaviour of the wild bootstrap ADF unit root tests obtained for these two different lag selection criteria. It was found that using MAIC in the presence of nonstationary volatility leads to a loss of finite sample power in the associated unit root test, caused by the tendency of MAIC to fit significantly more lags than RSMAIC. This despite the fact that size properties of the unit root tests based on MAIC and RSMAIC lag selection were shown to be broadly comparable. Moreover, under homoskedasticity no significant losses in power were observed for the unit root tests based on RSMAIC relative to those based on MAIC.

## ACKNOWLEDGMENT

We thank three anonymous referees for their helpful and constructive comments on an earlier draft of this paper.

## FUNDING

Cavaliere and Taylor thank the Danish Council for Independent Research, Sapere Aude | DFF Advanced Grant (Grant nr: 12-124980) for financial support.

## REFERENCES

- Beare, B. K. (2008). Unit root testing with unstable volatility. Nuffield College Economics Working Paper No. 2008-06. Oxford University.
- Busetti, F., Taylor, A. M. R. (2003). Variance shifts, structural breaks, and stationarity tests. *Journal of Business and Economic Statistics* 21:510–531.
- Cavaliere, G., Phillips, P. C. B., Smeekes, S., Taylor, A. M. R. (2012). Lag length selection for unit root tests in the presence of nonstationary volatility. Working paper. Available at <http://www.personeel.unimaas.nl/s.smeekes/CPSTWP.pdf>. Last accessed 11 November 2013.
- Cavaliere, G., Taylor, A. M. R. (2007). Testing for unit roots in time series models with nonstationary volatility. *Journal of Econometrics* 140:919–947.
- Cavaliere, G., Taylor, A. M. R. (2008). Bootstrap unit root tests for time series with nonstationary volatility. *Econometric Theory* 24:43–71.
- Cavaliere, G., Taylor, A. M. R. (2009a). Bootstrap  $M$  unit root tests. *Econometric Reviews* 28:393–421.
- Cavaliere, G., Taylor, A. M. R. (2009b). Heteroskedastic time series with a unit root. *Econometric Theory* 25:1228–1276.
- Chang, Y., Park, J. Y. (2002). On the asymptotics of ADF tests for unit roots. *Econometric Reviews* 21:431–447.
- Elliott, G., Rothenberg, T. J., Stock, J. H. (1996). Efficient tests for an autoregressive unit root. *Econometrica* 64:813–836.
- Fuller, W. A. (1996). *Introduction to Statistical Time Series* 2nd (ed.). New York: Wiley.
- Harvey, D. I., Leybourne, S. J., Taylor, A. M. R. (2009). Unit root testing in practice: dealing with uncertainty over the trend and initial condition. *Econometric Theory* 25:587–636.
- Kim, C.-J., Nelson, C. R. (1999). Has the US economy become more stable? A Bayesian approach based on a Markov-switching model of the business cycle. *Review of Economics and Statistics* 81:608–616.
- McConnell, M. M., Perez Quiros, G. (2000). Output fluctuations in the United States: what has changed since the early 1980s?. *American Economic Review* 90:1464–1476.
- Müller, U. K., Elliott, G. (2003). Tests for unit roots and the initial condition. *Econometrica* 71:1269–1286.
- Ng, S., Perron, P. (1995). Unit root tests in ARMA models with data dependent methods for selection of the truncation lag. *Journal of the American Statistical Association* 90:268–281.
- Ng, S., Perron, P. (2001). Lag length selection and the construction of unit root tests with good size and power. *Econometrica* 69:1519–1554.
- Ng, S., Perron, P. (2005). A note on the selection of time series models. *Oxford Bulletin of Economics and Statistics* 67:115–134.
- Perron, P., Qu, Z. (2007). A simple modification to improve the finite sample properties of Ng and Perron's unit root tests. *Economics Letters* 94:12–19.
- Pötscher, B. M. (1989). Model selection under nonstationarity: autoregressive models and stochastic linear regression models. *Annals of Statistics* 17:1257–1274.
- Richard, P. (2009). Modified fast double sieve bootstraps for ADF tests. *Computational Statistics & Data Analysis* 53:4490–4499.
- Sensier, M., Van Dijk, D. (2004). Testing for volatility changes in U.S. macroeconomic time series. *Review of Economics and Statistics* 86:833–839.
- Smeekes, S., Taylor, A. M. R. (2012). Bootstrap union tests for unit roots in the presence of nonstationary volatility. *Econometric Theory* 28:422–456.
- Stock, J. H., Watson, M. W. (1999). A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series. In: Engle, R. F., White, H. (eds.) *Cointegration, Causality and Forecasting: A Festschrift in Honour of Clive W.J. Granger*. Oxford: Oxford University Press, pp. 1–44.
- Tsay, R. S. (1984). Order selection in nonstationary autoregressive models. *Annals of Statistics* 12:1425–1433.
- Van Dijk, D., Osborn, D. R., Sensier, M. (2002). Changes in variability of the business cycle in the G7 countries. Econometric Institute Report EI 2002-28. Erasmus University Rotterdam.
- White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48:817–838.
- Xu, K.-L., Phillips, P. C. B. (2011). Tilted nonparametric estimation of volatility functions with empirical applications. *Journal of Business and Economic Statistics* 29:518–528.