Title
Testing for a trend with persistent errors

Journal
Journal of Econometrics, 219(2)

ISSN
0304-4076

Author
Elliott, Graham

Publication Date
2020-12-01

DOI
10.1016/j.jeconom.2020.03.006

Peer reviewed

Journal of Econometrics xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Testing for a trend with persistent errors☆

Graham Elliott

Department of Economics, University of California, San Diego, LA JOLLA, 92093, United States of America

ARTICLE INFO

Article history:
Available online xxxx

JEL classification:
C12
C21
C22

Keywords:
Composite hypothesis
Trend

ABSTRACT

We develop new tests for the coefficient on a time trend in a regression of a variable on a constant and time trend where there is potentially strong serial correlation. This serial correlation can also include a unit root. We obtain tests under two different assumptions on the initial value for the stochastic component of the variable being examined, either this being zero asymptotically and also allowing the initial condition to be drawn from its unconditional distribution. We find that statistics perform better under the second of these assumptions, which is the more natural assumption to make.

1. Introduction

We examine the problem of conducting hypothesis tests on the coefficient on a time trend in a regression of a variable that is serially correlated on a constant and time trend. Consider the model

$$x_t = a + bt + w_t, \quad (1)$$
$$w_t = rw_{t-1} + u_t,$$
$$u_1 = \xi$$

where $\{x_t\}_{t=1}^T$ are observed and $\{a, b, r, \xi\}$ and the variance of u_t are unobserved (for now consider u_t to be serially uncorrelated, the statistics presented in the Appendix A correct for additional serial correlation using estimators of the spectral density of u_t at frequency zero resulting in asymptotically similar tests to tests based on u_t serially uncorrelated under standard assumptions of the literature).

We are interested in testing the hypothesis

$$H_0 : b = 0 \quad (2)$$
$$H_a : b > 0.$$

The most likely application of these tests is for a pre-test for including a time trend in a regression. For many statistical procedures being able to remove a time trend from consideration helps greatly not only in the precision of estimators but also in the interpretation of results (because spurious correlation between time trends can be ruled out). However there are some instances where theory implies the lack of a time trend in data. In an application of their methods Bunzel and Vogelsang (2005) examine the Prebisch–Singer hypothesis. In hydrology there is a large literature related to testing the potential long run rise or fall of the volume of water in a river despite serially correlated data (see Yue et al. (2002) for an example).

☆ I thank the referees and editor along with participants in seminars at Monash University, Norges Bank and the IAAE 2017 conference in Sapporo and the 1st Financial Econometrics and New Finance Conference at Zheijiang University for comments. All errors are mine.

E-mail address: grelliott@ucsd.edu.

https://doi.org/10.1016/j.jeconom.2020.03.006
0304-4076/

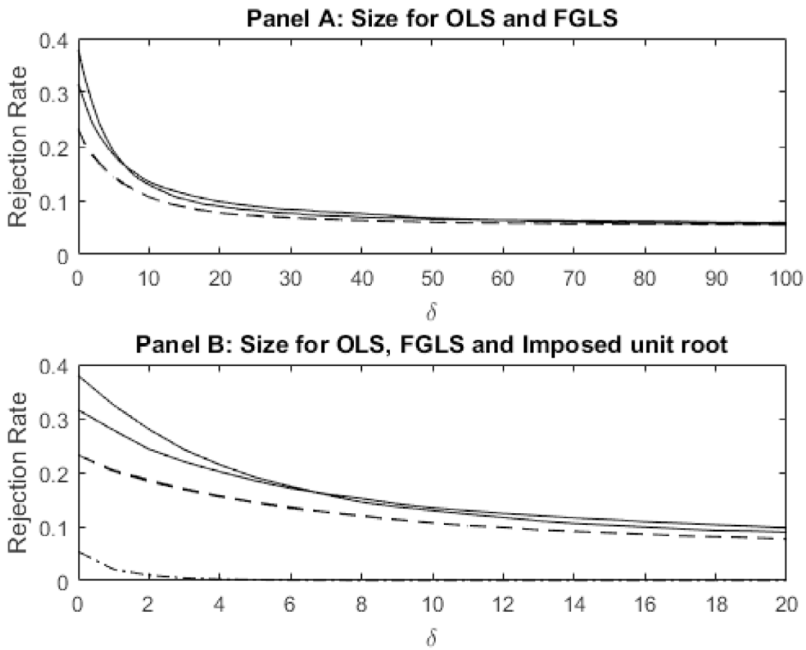

Fig. 1. Solid lines are size for FGLS (lower line is for $\xi = 0$) and the dotted line is for OLS. The dash–dot line in the second panel is GLS with a unit root imposed.

The primary complication in deriving tests that control size and have good properties is due to the nuisance parameter r that describes the largest root in a model of the serial correlation of the regressor. If this parameter were known then x_t can be quasi differenced and (at least asymptotically) a UMP test exists for the one sided test and a UMPU test exists for the two sided hypothesis. Alternatively if this parameter is bounded in absolute value with a bound that is (given the sample size) far enough from one then standard tests discussed below are available and can be employed. These too have asymptotic optimality properties. We are primarily concerned with testing situations where such a bound does not exist, and we would like to allow for r to be close to or equal to one. When the parameter approaches one or could be one, a number of papers have examined the regression coefficient under study (Canjels and Watson (1997) for estimation, and Bunzel and Vogelsang (2005), Harvey et al. (2007, 2010) and Perron and Yabu (2009) have each provided tests).

The following figure (Fig. 1) shows the problem of size control for models where r becomes close to or equal to one and also provides a guide to some of the previous solutions to providing tests that can potentially control size uniformly over values for r. For the model in (1) with $u_t \sim N(0, 1)$ we examine the size of tests of (2) undertaken in a few different ways. The first consideration is to simply construct the usual one sided test on the OLS estimate for b from a regression of x_t on a constant, time trend and lag of x_t using the data from $t = 2, \ldots, T$. The second consideration is to use a t test based on the Feasible GLS estimator for b where rather than knowing r we replace it with an estimator based on a regression of the residuals from a regression of x_t on a constant and time trend on the residuals lagged one period (again dropping the first observation because of the lag). The first panel of Fig. 1 shows the size from a Monte Carlo for various values for r (the x axis reports this as $\delta = T(1-r)$, anticipating the asymptotic problem that is examined below) where $T = 500$ and results reported are averages over 10 000 replications. Two variations on ξ are considered, either setting this value to zero or instead allowing ξ to be zero for $r = 1$ and drawn from a normal distribution with mean zero and variance $1/(1-r^2)$ when $r < 1$. Whilst the numbers are not identical the effect of the different initial conditions on OLS is not noticeable in the figure. It is clear that when δ is closer to zero (r closer to one) that standard tests fail to control size. As δ becomes larger, standard tests do indeed control size. Hence the difficult part of the testing problem to be solved is in the region of $r = 1$, since any test that has good properties in that region could be mixed or 'switched' with the standard one when r is small enough.

The second panel of Fig. 1 reports the same values as in the first panel focussing on values for r closer to one, where $0 \leq \delta \leq 20$. Added to the figure is the size of tests based on the MLE for b when we assume $r = 1$. The test imposing $r = 1$ is undersized for $r < 1$, quickly converging to zero. It is here that one strategy that could be employed becomes clear — constructing a test that mixes the GLS test at $r = 1$, which controls size at the boundary point and is otherwise undersized, with the test that works for r sufficiently less than one could, depending on how the tests are mixed control size. A test could control size by putting enough weight on the tests at $r = 1$ when $r < 1$ but in the region where the regular tests are oversized could control size in this region as it is mixed with a severely undersized test, and then for values for r further from one place more weight on the usual test. This is precisely the insight of Harvey et al. (2007), who suggest a test statistic of the form $z_\lambda^i = z_0(1-\lambda(x)) + \lambda(x)R_i(x)z_1$ where z_0 is the test when r is far less than one and z_1 is the test based on imposing $r = 1$. The mixing statistic $\lambda(x)$ is a random variable depending on the data that has the properties that (pointwise) is such that $\lambda(x) \to^p 1$ if $r = 1$ and $\lambda(x) \to^p 0$ for $r < 1$. The pointwise properties are not sufficient in themselves to ensure uniform size control, however they show in Monte Carlo results that size does appear to be controlled for values of r near one for many sample sizes. The role of $R_i(x)$ is because when they set $R_i(x) = 1$ (as for the z_λ statistic) the test is severely undersized asymptotically for δ near zero, so alternative choices were made to lessen this issue.

Perron and Yabu (2009) criticize this formulation of the mixture on a number of grounds. From a practical point of view the strongest criticism is that the mixing coefficient $\lambda(x)$, although converging to one asymptotically when $r = 1$, has in finite samples a distribution that places positive weight on smaller values for r which can impact the performance of the test. Indeed, to construct $\lambda(x)$ Harvey et al. (2007) make a number of ad hoc choices motivated by the need to control size and choose power curves, resulting in three variations of the test (i.e. choices of $R_i(x)$). The effect of this is that in any sample the test involves both z_0 and z_1. Instead, Perron and Yabu (2009) alter the strategy in two ways. First, rather than have $\lambda(x)$ take values between zero or one, they essentially suggest a statistic that chooses either the values of zero and one, essentially choosing the statistic z_1 whenever r is close enough to one (based on the data) and choosing a version of the standard test otherwise. This has the virtue of essentially always choosing z_1 when r is close to one. The second innovation is to improve the statistic chosen when $r < 1$. This essentially involves using a different estimator for r in the feasible GLS regression (Harvey et al. (2007) use OLS)[1] in order to have better properties for r large but not equal to one.

Each of the above mentioned approaches are designed for the testing problem under the assumption that the nuisance parameter ξ is either zero or asymptotically negligible ($T^{-1/2}\xi \to 0$ either directly or in probability). Since all tests considered either in this paper or above are invariant to translations of the form $\{x_t\}_{t=1}^T \mapsto \{x_t + \tilde{a}\}_{t=1}^T$, $\tilde{a} \in \mathbb{R}$ the choice of ξ when $r = 1$ is immaterial (see Müller and Elliott (2003)), so this is not an issue when $r = 1$. However when $r < 1$, it perhaps makes more sense to assume that ξ is drawn from the stationary distribution of w_t, which we refer to below as ξ drawn from its unconditional distribution. We can speculate that this will matter in the construction of critical values since tests here have limit distributions as functions of Brownian motions driven by partial sums of the shocks to w_t, and the form of these functions are different for different assumptions on the initial value. Under a very different assumption on the initial value than drawing this value from the stationary distribution, namely that ξ is fixed in repeated samples but dependent on r so that it is larger the closer r is to one, Harvey et al. (2010) show that the test statistics considered above are oversized. This occurs when $T^{-1/2}\xi$ converges to a constant and the authors construct a trimmed version of the above statistics to minimize this effect. Note that this assumption is not the same as considering ξ to be a random variable as in the analysis below.

Regardless of the assumption on ξ, for problems where $r \in [0, 1]$, no tests will be optimal uniformly in r. From a hypothesis testing problem we have a nuisance parameter r that renders the null and alternative hypotheses non simple. As is seen in Fig. 1, use of a (pointwise) consistent estimator such as OLS for r will not result in size control. Hence previous approaches, although ad hoc, make sense since constructing tests that control size uniformly for these problems is difficult. For any classical test, choosing a test to use amongst tests that control size is essentially the choice over power functions of the tests. Here, because of the nuisance parameter, we have a surface of power functions from which to choose (or alternatively, asymptotically a power function for each choice of δ). Indeed, examining the power functions presented in either Harvey et al. (2007) (which were compared with the method of Bunzel and Vogelsang (2005)) or Perron and Yabu (2009) (which compared with both the previous approaches) there is no uniformly dominant method. Each paper argued that despite this their approach was preferred over their predecessor based on having better power properties over the majority of parameterizations considered.

We take the same approach here, constructing new tests that pay close attention to the properties of the test when r is large or potentially one, and switching to standard tests for values of r further from one. At issue statistically here is that there is a nuisance parameter, the local size of the largest root of w_t, that complicates the testing. By utilizing the methods[2] of Elliott et al. (2015), denoted below as EMW, we are able to construct tests that (asymptotically) control size directly for a large region for δ including zero rather than rely on a test that has very small size (z_1) in this region. Because we can control size for a large region of r near one we switch to standard methods for values of r much further from 1 than previous tests. Our approach, by virtue of the explicit dependence on how power is directed towards the alternative hypothesis, results not in a single test but a family of tests dependent on these choices. In addition to providing results when ξ is asymptotically irrelevant, the situation assumed in most previous work, we also provide tests that are constructed for the case where ξ is drawn from the unconditional distribution for w_t when $r < 1$. This is a more natural assumption to make in practice and we will show that tests derived under this assumption have better properties.

We develop these tests in the next section, and in Section 3 examine this test along with those discussed above. The actual construction of the tests allowing for serial correlation is presented in a self contained section in the Appendix A.

[1] This alternative estimator is sufficiently complicated that rather than repeating the method here, we refer readers to the original paper.

[2] Other methods are possible, for example King (1996).

2. Constructing the test

The form of the family of tests we propose is a test that has the same switching approach as discussed above, namely

$$TR = \varphi_0(x)(1 - \chi(x)) + \chi(x)\varphi_1(x)$$

where φ_0 is a test function (so takes a value zero for failure to reject and one for rejection) for the hypothesis in (2) when r is far less than one (this is the 'simple' test in the jargon of EMW), $\chi(x)$ is a switching rule that, as in Perron and Yabu (2009), takes values of zero or one depending on whether the data indicates we are in the simple region or not, and finally $\varphi_1(x)$ is a new test function derived in this paper for testing (2) when r is in the neighborhood of one. So for any sample $\{x_t\}_{t=1}^T$ the test function TR as a function of the sample takes on a value of one if the hypothesis is rejected and zero otherwise. A primary difference here as opposed to previous tests is that our test switches to the standard test for alternatives much further from one than earlier tests. We are able to do this and still control size as a result of the test when r is near one being constructed to control size not just when $r = 1$ but also for local alternatives as well. Hence rather than relying on the test assuming a unit root in the neighborhood of a unit root being conservative to keep the mixture test size controlled, we rely on the construction of the test function $\varphi_1(x)$ having good size properties over a wide range of models.

2.1. The test when r is near one

We turn to the construction of $\varphi_1(x)$, which provides a test of the hypothesis when r is near one. The test statistics we propose have the form

$$WLR = \frac{\frac{1}{Jm}\sum_{i=1}^{m}\sum_{j=1}^{J} L(x, \delta_i, \beta_{ij})}{\sum_{i=1}^{m} p_i L(x, \delta_i)} \tag{3}$$

where $\delta_i = T(1 - r_i)$, β_{ij} are chosen points under the alternative hypothesis local to zero, p_i are a set of weights (that do not sum to one generally for reasons discussed below), $L(x, \delta_i, \beta_{ij})$ is an approximation to the likelihood local both to $r = 1$ and $b = 0$ and $L(x, \delta_i) = L(x, \delta_i, 0)$, the approximation to the likelihood when $b = 0$ (i.e. under the null hypothesis). The statistic rejects when $\log(WLR)$ is greater than zero, i.e. when the weighted average likelihood under the alternative is large relative to the weighted average likelihood under the alternative. The critical value here is subsumed into the weights under the null, which explains why these weights do not sum to one. To see this, define $\tilde{p}_i$ as a set of weights that sum to one, and notice that we could write the test function as rejecting for x when

$$\frac{\frac{1}{Jm}\sum_{i=1}^{m}\sum_{i=j}^{J} L(x, \delta_i, \beta_i)}{\sum_{i=1}^{m} \tilde{p}_i L(x, \delta_i)} > cv$$

but this is equivalent to rejecting for x when

$$\frac{\frac{1}{Jm}\sum_{i=1}^{m}\sum_{i=j}^{J} L(x, \delta_i, \beta_i)}{\sum_{i=1}^{m} (cv * \tilde{p}_i) L(x, \delta_i)} > 1$$

so we directly construct the test with $p_i = cv * \tilde{p}_i$ and so the sum of the weights is the critical value in the standard set up of a likelihood ratio testing problem.

The form of the suggested statistic is a likelihood ratio test for (2) where we deal with the nuisance parameter δ describing the serial correlation by taking weighted averages over this nuisance parameter. Under the alternative hypothesis (the numerator of the statistic) the statistic then weights different models, with the choice of weights directing power to the alternatives. This is a standard approach to constructing tests that have optimality properties against the chosen weighted alternative, see Andrews and Ploberger (1994). Under the null hypothesis the weights $\tilde{p}_i$ are an estimate of the least favorable distribution for the testing problem given the choice of the alternative. Tests constructed in this way have, by virtue of their form as a Neyman Pearson statistic, optimality properties. See EMW for an extensive discussion with relation to the approach used here.

A competing and similar approach to the one conducted here is available in Sriananthakumar and King (2006). In this paper rather than choose point masses for the models to sum over they suggest using distributions over the parameters. They find for some applications (they did not consider the application of this paper) that mixing three distributions can provide a test that controls size. In previous work (for other applications) we have found that choosing a reasonably large number of point models provides better approximations for the problems we have examined, however we did not examine their method for this problem. Our method then differs in the choice of mixture distributions, but also in terms of the algorithm used to determine the weights. The test we construct also has very straightforward asymptotic theory results and is simple to adjust for additional serial correlation in u_t.

To construct the test then we need to define the form of the approximate likelihood $L(x, \delta_i, \beta_{ij})$ and then estimate the critical value adjusted weights p_i using the methods in EMW. We first consider $L(x, \delta_i, \beta_{ij})$. For the testing problem addressed here, the form of the likelihood even asymptotically depends on the assumption on the initial condition, in much the same way that it does for testing in the unit root literature. We will consider two assumptions, the first allowing ξ to be fixed so that it is asymptotically irrelevant and can be set to zero in the approximation, and second drawing the initial value w_1 from the unconditional distribution for w_t when $r < 1$. Note that because under both assumptions we remove the nuisance parameter a from the likelihood through invariance to translations of the form $\{x_t\}_{t=1}^T \mapsto \{x_t + \tilde{a}\}_{t=1}^T$, $\tilde{a} \in \mathbb{R}$, then this also removes the initial value at $r = 1$.

2.1.1. The likelihood when $\xi = 0$

Consider the model in (1) with $\xi = 0$ and additionally assume that u_t are serially uncorrelated and have $N(0, \sigma^2)$ marginal distributions. In this case the log of the likelihood for the statistic invariant to the translation discussed in the previous paragraph is approximately (ignoring the determinant component, which converges to one for r local to one[3]) proportional to

$$\sigma^{-2}(x - x_1\iota - bz)'(A(r)'A(r) - A(r)'A(r)\iota(\iota'A(r)'A(r)\iota)^{-1}\iota'A(r)'A(r))(x - x_1\iota - bz)$$

where $x = (x_1, x_2, \ldots, x_T)'$, $z = (1, 2, \ldots, T)'$, ι is a $Tx1$ vector of ones and $A(r)$ is a TxT differencing matrix with ones on the diagonal, $-r$ on the diagonal immediately below the main diagonal and zeros everywhere else (so $A(r)y$ for any vector y is $(y_1, y_2 - ry_1, \ldots, y_T - ry_{T-1})'$). See Müller and Elliott (2003) for a derivation of this likelihood. Let $b_{ij} = \beta_{ij}/\sqrt{T\sigma}$ and $\delta_i = T(1 - r_i)$ then we can rewrite this for choices (δ_i, β_{ij}) as

$$\begin{aligned} & h(x) + \delta_i^2\sigma^{-2}T^{-2}\tilde{x}_{-1}'\tilde{x}_{-1} + 2\delta_i\sigma^{-2}T^{-1}\tilde{x}_{-1}'(\tilde{x} - \tilde{x}_{-1}) \\ & - \frac{\left(\delta_i\sigma^{-1/2}T^{-1/2}\tilde{x}_T + \delta_i^2\sigma^{-1}T^{-3/2}\sum_{t=2}^T \tilde{x}_{t-1}\right)^2}{T(1 + (T-1)(1-r)^2)} \\ & + \beta_{ij}^2\left[T^{-1} + \delta_i^2 s_2 + r^2\left(\frac{T-1}{T}\right) + 2r_i\delta_i s_1 - \frac{T^{-1}\left(1 + r_i\delta_i(\frac{T-1}{T}) + \delta_i^2 s_1\right)^2}{(1 + (T-1)(1-r_i)^2)}\right] \\ & - 2\beta_{ij}\left[\begin{array}{c} (r_i + \delta_i)\sigma^{-1}T^{-1/2}\tilde{x}_T + \delta_i^2\sigma^{-1}T^{-5/2}z'\tilde{x}_{-1} - \delta_i^2\sigma^{-1}T^{-5/2}\tilde{x}_{-1}'\iota \\ -\frac{T^{-1}\left(1 + r_i\delta_i(\frac{T-1}{T}) + \delta_i^2 s_1\right)}{1 + (T-1)(1-r_i)^2}\left(\delta_i\sigma^{-1/2}T^{-1/2}\tilde{x}_T + \delta_i^2\sigma^{-1}T^{-3/2}\sum_{t=2}^T \tilde{x}_{t-1}\right) \end{array}\right]. \end{aligned} \tag{4}$$

where $\tilde{x} = x - x_1\iota$ and $\tilde{x}_{-1} = (0, \tilde{x}_1, \ldots, \tilde{x}_{T-1})'$, $s_1 = T^{-2}\sum_{t=2}^T t$ and $s_2 = T^{-3}\sum_{t=2}^T t^2$. Since $h(x)$ is not a function of the nuisance parameters (apart from the scale parameter which is consistently estimable) we can ignore this as it cancels in the numerator and denominator for the test statistic WLR. The remaining terms are either $O_p(1)$ or disappear asymptotically. After subtracting $h(x)$ then the exponent of the remainder of the terms in (4) multiplied by minus one half is defined as $L_0(x, \delta, \beta)$ where the zero subscript denotes computation under the assumption that $\xi = 0$. Then using standard FCLT and CMT results in

$$\begin{aligned} -2\ln L_0(x, \delta_i, \beta_{ij}) \Rightarrow\ & \delta_i^2 \int W_{\delta,\beta}(s)^2 ds + 2\delta_i \int W_{\delta,\beta}(s)dW_{\delta,\beta}(s) + \beta_{ij}^2\left(1 + \frac{\delta_i^2}{3} + \delta_i\right) \\ & - 2\beta_{ij}\left((1+\delta_i)W_\delta(1) + \delta_i^2\int sW(s)ds\right) \\ =\ & H(\delta_i, \beta_{ij}) \end{aligned}$$

where $W_{\delta,\beta}(s) = \int_0^s e^{-\delta(s-\lambda)}dW(\lambda) + \beta s$ and $W(\lambda)$ is a standard Brownian Motion. The notation suppresses dependence on the true values for (δ, β) that arise through $W_{\delta,\beta}(s)$.

The limit distribution of the test statistics are then equal to

$$WLR \Rightarrow \frac{\frac{1}{Jm}\sum_{i=1}^m \sum_{i=j}^J \exp(-0.5H(\delta_i, \beta_{ij}))}{\sum_{i=1}^m w_i \exp(-0.5H(\delta_i, 0))}$$

for various choices over δ_i and b_{ij}. Each of these choice variables indicate the choice of the alternative in the weighted average under the alternative.

Our actual statistic, i.e. the construction of $L(x, \delta_i, \beta_i)$, is based on the term (4) after removing $h(x)$ but has the same limiting distribution as stated immediately above. Exact calculation details are given in the Appendix A. For the null likelihood we set $\beta = 0$ so the second two terms do not appear. For the tests below we set $\delta_i = (0, .0.1^2, 0.2^2, \ldots, 2^2, 2.5^2, \ldots, 11^2)$ so $m = 39$.

[3] The determinant term is equal to $\left|A(r)'A(r)\right|^{-1/2}\left|\iota'A(r)'A(r)\iota\right|^{-1/2}$ which equals $(1 + (T-1)(1-r)^2)^{-1/2} \to 1$ for $T(1-r)$ being $O(1)$.

2.1.2. The likelihood when ξ comes from the unconditional distribution

Consider now the same model from the previous subsection, however now $\xi \sim N(0, \sigma^2/(1-r^2))$ and independent of $\{u_t\}_{t=2}^T$ for $r < 1$. Note that the statistic invariant to translations of the form $\{x_t\}_{t=1}^T \mapsto \{x_t + \tilde{a}\}_{t=1}^T$, $\tilde{a} \in \mathbb{R}$ is invariant to the value of ξ at $r = 1$. In this case different values for $r < 1$ impact the likelihood in an additional way. For this case the component inside the exponent of the likelihood is proportional to

$$\sigma^{-2}(x - x_1\iota - bz)'(A(r)'V^{-1}A(r) - A(r)'V^{-1}A(r)\iota(\iota'A(r)'V^{-1}A(r)\iota)^{-1}\iota'A(r)'V^{-1}A(r))(x - x_1\iota - bz) \tag{5}$$

where V^{-1} is the identity matrix except that it has $(1-r^2)$ in the (1,1) element if $r < 1$, and is the identity matrix otherwise.

The determinant term is now $\left|A(r)'V^{-1}A(r)\right|^{-1/2}\left|\iota'A(r)'V^{-1}A(r)\iota\right|^{-1/2}$ which after some algebra yields $\left(1 + \frac{(1-r)(T-1)}{1+r}\right)^{-1/2}$. Note that this term is continuous in r at $r = 1$ from below.

Expanding as above for choices (δ_i, β_{ij}) we obtain that (5) can be rewritten for $r_i < 1$ as[4]

$$\begin{aligned}
&\tilde{h}(x) + \delta_i^2\sigma^{-2}T^{-2}\tilde{x}_{-1}'\tilde{x}_{-1} + 2\delta_i\sigma^{-2}T^{-1}\tilde{x}_{-1}'(\tilde{x} - \tilde{x}_{-1}) \\
&- \frac{\left(\delta_i\sigma^{-1/2}T^{-1/2}\tilde{x}_T + \delta_i^2\sigma^{-1}T^{-3/2}\sum_{t=2}^T \tilde{x}_{t-1}\right)^2}{T(1 - r_i^2 + (T-1)(1-r)^2)} \\
&+ \beta_{ij}^2\left[(1-r_i)T^{-1} + \delta_i^2 s_2 + r^2\left(\frac{T-1}{T}\right) + 2r_i\delta_i s_1 - \frac{\left(\delta_i T^{-1} + r_i\delta_i + \delta_i^2 s_1\right)^2}{T(1-r_i^2+(T-1)(1-r_i)^2)}\right] \\
&- 2\beta_{ij}\left[\begin{array}{l}(r_i+\delta_i)\sigma^{-1}T^{-1/2}\tilde{x}_T + \delta_i^2\sigma^{-1}T^{-5/2}z'\tilde{x}_{-1} - \delta_i^2\sigma^{-1}T^{-5/2}\tilde{x}_{-1}'\iota \\ -\frac{\left(\delta_i T^{-1} + r_i\delta_i + \delta_i^2 s_1\right)}{T(1-r_i^2+(T-1)(1-r_i)^2)}\left(\delta_i\sigma^{-1/2}T^{-1/2}\tilde{x}_T + \delta_i^2\sigma^{-1}T^{-3/2}\sum_{t=2}^T\tilde{x}_{t-1}\right)\end{array}\right].
\end{aligned} \tag{6}$$

Again $\tilde{h}(x)$ does not depend on nuisance parameters and so it will cancel from the numerator and denominator in *WLR*. Ignoring this term, multiplying by minus one half and taking limits using standard FCLT and CMT results yields a limit distribution of

$$\begin{aligned}
-2\ln L_1(x, \delta_i, \beta_{ij}) \Rightarrow\ & \delta_i^2\int M_{\delta,\beta}(s)^2 ds + 2\delta_i\int M_{\delta,\beta}(s)dM_{\delta,\beta}(s) - \frac{\left(\delta_i M_{\delta,\beta}(1) + \delta_i^2\int M_{\delta,\beta}(s)ds\right)^2}{2\delta_i + \delta_i^2} \\
& + \beta_{ij}^2\left(1 + \frac{\delta_i^2}{12} + \frac{\delta_i}{2}\right) && (7) \\
& -2\beta_{ij}\left((1 - \frac{\delta_i}{2})M_{\delta,\beta}(1) + \delta_i^2\int sM_{\delta,\beta}(s)ds - \frac{\delta_i^2}{2}\int M_{\delta,\beta}(s)ds\right) \\
& - \ln\left(1 + \frac{\delta_i}{2}\right) && (8) \\
=\ & H_1(\delta_i, \beta_{ij})
\end{aligned}$$

where $M_{\delta,\beta}(s)$ is defined as

$$M_{\delta,\beta}(s) = \begin{cases} W(s) + \beta s \text{ for } \delta = 0 \\ (e^{-\delta s} - 1)(2\delta)^{-1/2} + \int_0^s e^{-\delta(s-\lambda)}dW(\lambda) + \beta s \text{ else.} \end{cases} \tag{9}$$

As in the previous case this establishes the limit distributions for the test statistics. The choice of δ_i over which to construct the statistics we again set $\delta_i = (0, .0.1^2, 0.2^2, \ldots, 2^2, 2.5^2, \ldots, 11^2)$.

2.1.3. Weights for the likelihood under H_a

The choice of the weights for the likelihood under the alternative in each case involves choosing β_{ij}. This choice impacts the statistics in a number of ways. In terms of the approximate optimality properties of the statistics the tests have such properties against the alternatives chosen. More practically we can consider the choice as one that generates useful power properties of the tests. Finally, for the methods of EMW to work well the choices can result in better or poorer estimation of the least favorable weights.

In practice some searching over choices was undertaken to product tests with good asymptotic properties. Since this (and the finite sample performance of the tests) is what we are ultimately interested in then it seems reasonable to make a number of choices that have explicit trade-offs. We experimented with different choices, and report one for each initial condition assumption. The actual choices are available in the description of the test statistic construction in the

4 The expression for $r_i = 1$ is as for the zero initial condition case.

Appendix A. For β_{ij}, we choose $b_{ij} = c_j/\sqrt{1-\delta_i^2/d}$ for some value d which is in the spirit of point optimal test choices (see King (1988) and King and Sriananthkumar (2017)). These contributions by King and later coauthors suggest choosing point alternatives (here we mix two point alternatives) to 'pin' power curves somewhere in the middle of the power curve in order to construct a statistic that has good power properties at least at the chosen point and hopefully elsewhere as well. Our particular choice here achieves a number of objectives. First, β_{ij} becomes smaller as δ_i increases because the power of tests for a trend increase quickly as the data becomes more stationary. This is the primary reason for focussing on power when δ is small, because this is the region where power is low for the tests (this is evident in the large sample figures below, where as δ becomes larger the scale of the alternatives for which power is below one becomes smaller). The rate of decline is increased by choosing a larger value for d. Second, the entire function is shifted up through choices of c_j. For the best results it was found that choosing two values for c_j improved the shape of the asymptotic power functions.

Because we can make a number of choices for both the values for δ_i and also the values under the alternative hypothesis (β_{ij}) we essentially derive a family of tests here indexed by these choices. For both of the tests below we set $\delta_i = (0, .0.1^2, 0.2^2, \ldots, 2^2, 2.5^2, \ldots, 11^2)$ so $m = 39$. For the zero initial value assumption we choose $J = 2$ setting $\beta_{i1} = 1/\sqrt{1-\delta_i^2/30}$ and $\beta_{i2} = 3/\sqrt{1-\delta_i^2/30}$. In the case of the initial value being chosen from its unconditional distribution we again set $J = 2$ with $\beta_{i1} = 1/\sqrt{1-\delta_i^2/50}$ and $\beta_{i2} = 3/\sqrt{1-\delta_i^2/50}$. These values were chosen based on the asymptotic power curves they generate after some experimentation.

2.1.4. Switching

For the tests we need to define $\chi(x)$, which will equal one when we employ the *WLR* statistic and zero when the standard test controls size. We do so effectively through a pre-test for $r = 1$ but one which only rejects for values for δ very far from zero (so a test with almost zero size for a wide range of δ). The idea is to only switch when the standard test is in a range for δ that controls size well. From Fig. 1 we see that this is in the range for δ beyond 80 or so. Consider the regression

$$\Delta x_t = a + bt + p_0 x_{t-1} + \sum_{l=1}^{L} p_l \Delta x_{t-1} + v_t \qquad (10)$$

where if $L = 0$ no additional lagged changes are included and L is chosen by the BIC criterion (one could also choose the MAIC criterion of Perron and Ng (1996)). If $L = 0$ is imposed (because we know there is no serial correlation in u_t) then we construct the statistic $\chi(x) = 1(-T\hat{p}_0 < 100)$. The choice of 100 as a cutoff is made because at this point both the *WLR* test and the test above have essentially maximal power for testing the hypothesis under consideration (indeed, as shown below maximal power is almost attained around $\delta = 30$ for the tests we construct) and so this choice results in effectively only switching for situations where it is extremely obvious from the data that we are not near a unit root in w_t. For situations where there is some potential serial correlation in u_t the statistic needs to be adjusted to account for this. So for $L > 0$ we construct $\chi(x) = 1(-T\hat{p}_0\sqrt{\hat{\omega}_u^2/\hat{\sigma}_v^2} < 100)$ where $\hat{\sigma}_v^2 = (T-L-3)^{-1}\sum_{L+1}^{T}\hat{v}_t^2$ where $\hat{v}_t$ are the OLS residuals from (10) and $\hat{\omega}_u^2$ is an estimate of the spectral density of $\hat{u}_t$ where $\hat{u}_t$ are the OLS residuals of (10) with L set to zero. This approach to switching is employed for both assumptions on the initial condition.

2.1.5. The test when r is far below one

There are two issues in choosing the test function $\varphi_0(x)$ for use when r is far below one. The first is size control, the second is the properties under more general serial correlation. Whilst use of the feasible GLS estimator may allow some power advantages over using least squares when r is far from one, our (unpresented) Monte Carlo evaluation found that using the t test from the regression (10) with L chosen using the BIC criterion had the advantages that it accounts fairly well for serial correlation of various types and it simplifies construction of the test statistic since we require estimation of the same regression as for constructing the switching statistic.

2.1.6. The test statistics TR_0 and TR_1

In each of the assumptions for the initial condition, we effectively construct a family of tests indexed by choices over δ_i and the alternative models, as well as the choices in switching. As we have motivated above, the choices for δ_i and the point at which we switch are less ad hoc than they appear — they are chosen to provide tests that do not switch (and so are based on the *WLR* statistics) over the range of models for which the size of standard tests does not match nominal size. The choices for the local alternatives are more ad hoc but have clear theoretical and practical implications. From the theoretical perspective the tests have optimality properties against the choice of the alternative via the Neyman Pearson lemma. From a practical perspective the alternatives are chosen to have local power functions that provide a reasonable trade-off across values for δ.

Once the choices for the alternative distribution are made, tests are defined. Here we have defined two tests, one for each assumption on the initial condition. These tests are referred to as TR_0 and TR_1 for the test with a zero initial condition and the test assuming the initial value arises from its unconditional distribution respectively. They differ from each other in three ways. First, they are based on different likelihoods as noted in Sections 2.1.1 and 2.1.2. Second, they differ in the choices for β_{ij}, as detailed in Section 2.1.3. Finally, the weights constructed from the EMW method differ between the tests. A full step by step construction of the tests including a detailing of the weights is given in Appendix A.

ECONOM: 4836
ARTICLE IN PRESS

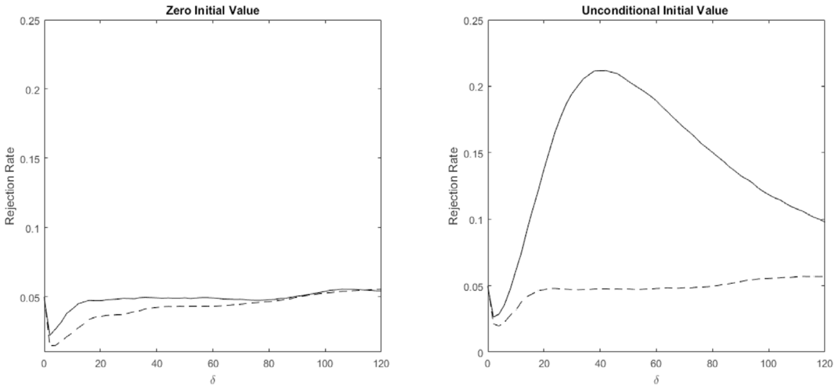

Fig. 2. Solid lines indicate size for TR_0 and dashed lines are size for TR_1. The first panel is for a dgp with $\xi = 0$ and the second panel for $\xi \sim N(0, (1 - r^2)^{-1})$.

3. Evaluation of the methods

We first evaluate approximations to asymptotic size and power, then turn to a small sample Monte Carlo evaluation of the methods.

The size properties of the two proposed statistics are examined in Fig. 2. Approximations to the asymptotic distribution are constructed by setting the sample size to 1000 and use a known variance (set to one), estimates are over 10 000 Monte Carlo simulations. The first panel of Fig. 2 shows the size of TR_0 and TR_1 for various values of the nuisance parameter δ setting the initial value $\xi = 0$. The solid line shows results for TR_0 and the dashed line for TR_1. Firstly, in the first panel we see that the tests optimized for either initial value assumption control size for this model. This is not surprising since the assumptions underlying the construction of TR_1 include a large mass at zero. Secondly, we note that size is below nominal size for δ near zero, but quickly moves close to nominal size. Compared to the other statistics suggested in the literature, the tests here have size near nominal size for a much greater range of δ. This is in part because the test statistic is essentially completely based on the *WLR* component of the tests for δ under 80 or so, and by the EMW approach these methods control size by construction.

The second panel reports results for $\xi \sim N(0, (1 - r^2)^{-1})$ for each of the statistics. Clearly TR_0 no longer controls size, this is typical of tests for this hypothesis constructed under the assumption of a zero null when the initial value is drawn from its unconditional distribution. However TR_1 does control size, with again some undersized behavior for δ close to zero. For larger values of δ in both panels there is some slight oversized behavior — this is because the tests switch around $\delta = 100$ (around rather than at because δ has to be estimated) and this reflects properties not of the tests based on EMW but the standard least squares regression test which is now mixed into the results. Other methods that switch to either OLS or the feasible MLE have similar properties (see Fig. 1). Overall, that TR_1 is able to control size for both assumptions on the initial value is suggestive of this assumption being a better one in the construction of the tests, so long as its use does not come at the cost of a lot of power.

We now turn to large sample power properties of the statistics suggested in this paper and earlier work. We compare them to the z_λ^{m1} statistic of Harvey et al. (2009) and the *MU* statistic in Perron and Yabu (2009). A more detailed examination of other statistics from these two papers is available in a supplemental appendix. Fig. 3 reports results from a Monte Carlo exercise with the sample size set to 1000 and 20 000 Monte Carlo draws. Each panel shows power as a function of $\beta = b/\sqrt{T}$ where the data is drawn from (1) with $\xi = 0$. Different panels accord with different choices of the nuisance parameter δ, which takes the values 0, 5, 10, 15, 20, and 30.

At $\delta = 0$, we note first that the only statistic that is (essentially) equivalent to the infeasible MLE power bound is the (Perron and Yabu, 2009) MU statistic, which comes about by virtue of that statistic forcing itself to be equal to the MLE with a unit root and for roots near one (Harvey et al., 2007) also have a statistic (they denote as z_λ) that achieves this bound however the one reported here has much better power than this statistic for other alternatives). This comes at a cost in power as δ gets larger, for example for $\delta = 10$ through 20 the power of the MU statistic (the dotted line with the flat part) is far below all other statistics for most of the local alternatives. The z_λ^{m1} statistic (the dash–dot line) has lower power at $\delta = 0$ and, because of poor asymptotic size properties for $\delta > 0$ low power for near alternatives. Despite this, at $\delta = 15$ this statistic has a power function that is not dominated by other methods (but does not dominate other methods either), however at other values for δ (larger and smaller) other statistics have better local power.

Similar to the comparison between the MU and z_λ^{m1} statistics, the TR_0 test has power that does not dominate each other across all values for δ. Different choices for the weighted averaging (not reported here) allow construction of tests that, like z_λ^{m1}, have poor size properties for δ greater than but near zero but steep power curves. However the power function for TR_0 compares very favorably. At $\delta = 0$ there is only a very small drop-off in power relative to *MU*, but for more

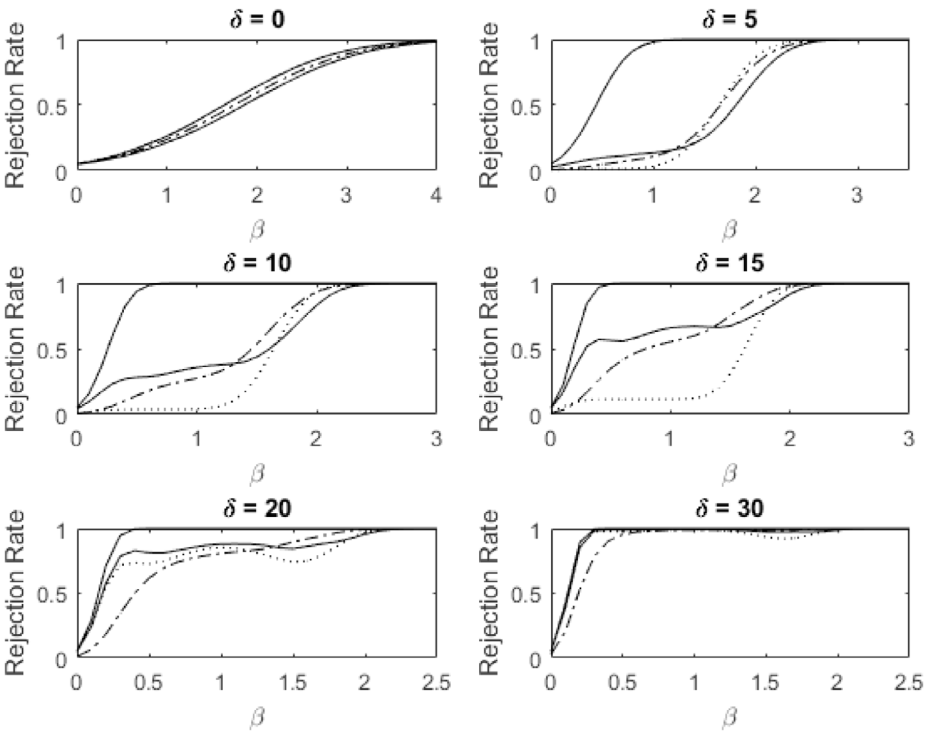

Fig. 3. The upper solid lines show power for the infeasible MLE, dash–dot lines indicate power for the z_λ^{m1} statistics, dotted lines show results for *MU*, and the lower solid line shows power for TR_0.

distant alternatives, especially for moderate δ of 15 or 20, there is a vast gap between the methods. Similarly, comparing to z_λ^{m1}, especially for moderate β and intermediate values for δ, there is a sizeable vertical distance between the power functions. Overall TR_0 strongly dominates MU for smaller values of β and converges faster to the infeasible MLE power as δ becomes larger. At $\delta = 30$, where the statistic is basically WLR_0 since switching is rare for this region of δ, the power of TR_0 is approximately equal to the power of the infeasible MLE test. The main reason for these power gains is that size is better controlled for moderate δ.

Whilst the argument above suggests that TR_0 is a good choice for the zero initial condition problem, the second assumption on the initial condition makes more sense intuitively because models are stationary when w_t has a root greater than one. When for $|r| < 1$ our model assumes $\xi \sim N(0, \sigma^2/(1-r^2))$, as shown in Section 2 the limit distributions under both the null and alternative hypothesis differ from those when we assume $\xi = 0$. And, as we noted above, the optimal statistics generated in this paper will have different weights for the same chosen alternative distributions. So now we turn to large sample results for TR_1 under this assumption.

As in the previous case, we evaluate power for various values for δ. As before, the infeasible MLE (under the new assumption on the initial condition) is provided as a benchmark even though for most δ the power of this statistic cannot be matched as it 'knows' r and the feasible version does not control size, as indicated in Fig. 1. Fig. 4 shows the results for this statistic (solid lines) and TR_1 (dashed lines). TR_1 does not suffer as much from being undersized for small but non zero δ, compared to TR_0 in the case where $\xi = 0$. As in the $\xi = 0$ case the cost of controlling size is lower power than the infeasible bound for small δ, although again in comparison to the $\xi = 0$ case the impact here seems to be smaller. The statistic TR_1 has power close to the infeasible bound at $\delta = 0$ and at $\delta = 30$. It is useful to note that at $\delta = 30$ there is virtually no switching to the feasible MLE test, so this power is all from using WLR_1 as calculated above.

Finally for the large sample results, we turn to comparing the impacts of the tests for each initial condition assumption on power. In the size results reported in Fig. 2 earlier, we saw that using TR_0 when ξ is drawn from its unconditional distribution resulted in size distortions, whereas TR_1 controlled size under both assumptions on ξ. Each panel in Fig. 5 shows power against alternatives β nonzero for the same choices of δ as in the previous power figures. When $\xi = 0$, the solid line shows the power of TR_0 and the dots show power of TR_1. The results show that there is some loss in not using the test statistic designed for $\xi = 0$ here, especially at smaller β, although power for TR_1 in these cases for moderate δ exceed that of TR_0. This implies that it would be difficult to pick between the two tests. This contrasts strongly with the impact on power (and size) when ξ is drawn from its unconditional distribution. The solid line with dots shows the power of TR_0 and the dashed line shows the power of TR_1. Here, as noted in Fig. 2, TR_0 does not control size for moderate δ, it is oversized and the power curves are shifted up due to this effect. However power advantages from being oversized

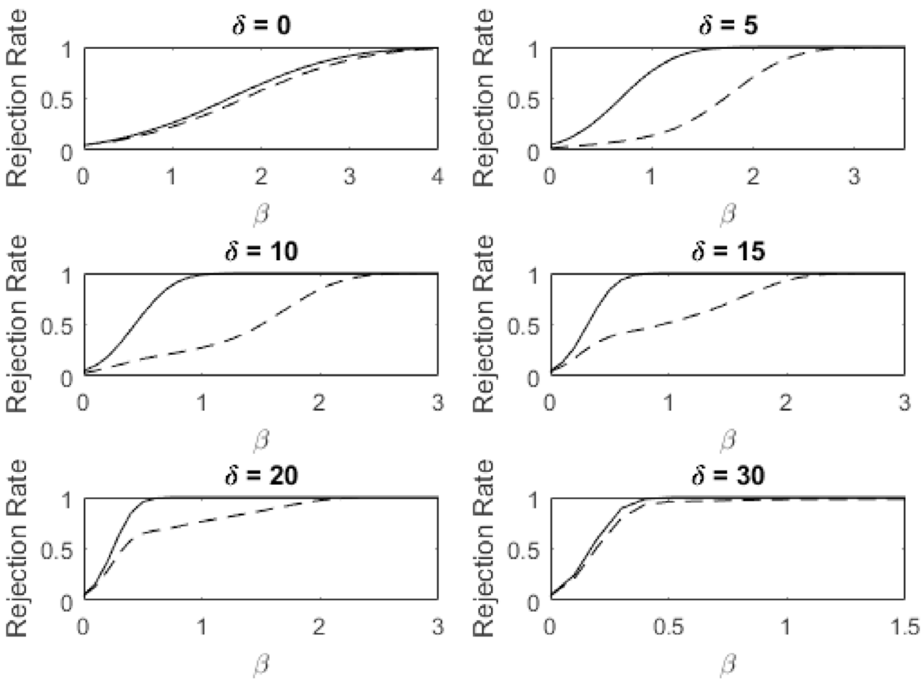

Fig. 4. The solid line indicates power from the infeasible MLE, the dashed line is power for TR_1.

are not really apparent except at very small values for β. For $\delta = 10$ and above the (non sized adjusted) power for TR_0 is only just above the power for TR_1 even though TR_1 controls size, for larger values of β TR_1 outperforms even when controlling size. Choosing TR_1 then is a strong choice given that we cannot tell in practice what the correct assumption on ξ is, it controls size for both assumptions and use of TR_1 does not involve significant size losses.

We now turn to small sample evaluation of the methods. We extend the model in (1) to allow for serial correlation in u_t, we consider data generating processes of the form

$$(1 - \theta_1 L)u_t = (1 + \theta_2 L)\varepsilon_t$$

where ε_t are serially independent $N(0, 1)$ random variables and we consider various values for θ_1 and θ_2. For the test to control size asymptotically we require that the test be adjusted to allow for serial correlation — this is accomplished through using an estimator of the spectral density at frequency zero of u_t in place of the variance as well as (for one of the terms) adopting the correction in Phillips (1987) for one of the terms. The details are in the Appendix A where the construction of the statistics are presented in their full generality. We report results in columns of the table for both the TR_0 and TR_1 statistics for both initial condition assumptions. We approximate the unconditional case by allowing for 500 'burn in' observations prior to the sample that is employed for the tests. In each case we allow up to four lags in estimating robust statistics. The size results reported are averages across the outcomes from 10,000 Monte Carlo simulations. We report values for δ equal to $\{0, 5, 10, 15, 20, 30\}$ with the first three values in Table 1 and the second three in Table 2. Note that because these values for δ are far below the switching level of $\delta = 100$, the results shown essentially have very little switching and so generally reflect the use of the WLR component of the statistic in (3).

Many of the standard results from time series methods when corrections for serial correlation are apparent in the results. In particular there is some oversizedness when there is a large MA component, despite our use of the Ng and Perron (2001) MAIC lag length selector. This is a well known issue in scaling partial sums, see Ng and Perron (2001) for details. Also, when one allows for lags when there is no serial correlation there is a small effect where size is larger than if no lags were allowed for. Finally, differences between the empirical size and nominal size are smaller as the sample is increased (for tests that asymptotically control size, which excludes here TR_0 when the initial value is drawn from its unconditional distribution).

Following from the asymptotic results, empirical size of the tests is smaller than nominal size for values of δ near but not equal to zero, with this effect diminishing as δ gets larger. The same phenomenon appeared[5] in previous tests for this hypothesis. The tests are somewhat undersized relative to the asymptotic results, however this effect is smaller for larger sample sizes. The oversizedness of TR_0 when the initial value is drawn from the unconditional distribution is also

[5] This is true of previously suggested tests, see Table 2 of Harvey et al. (2007) for example.

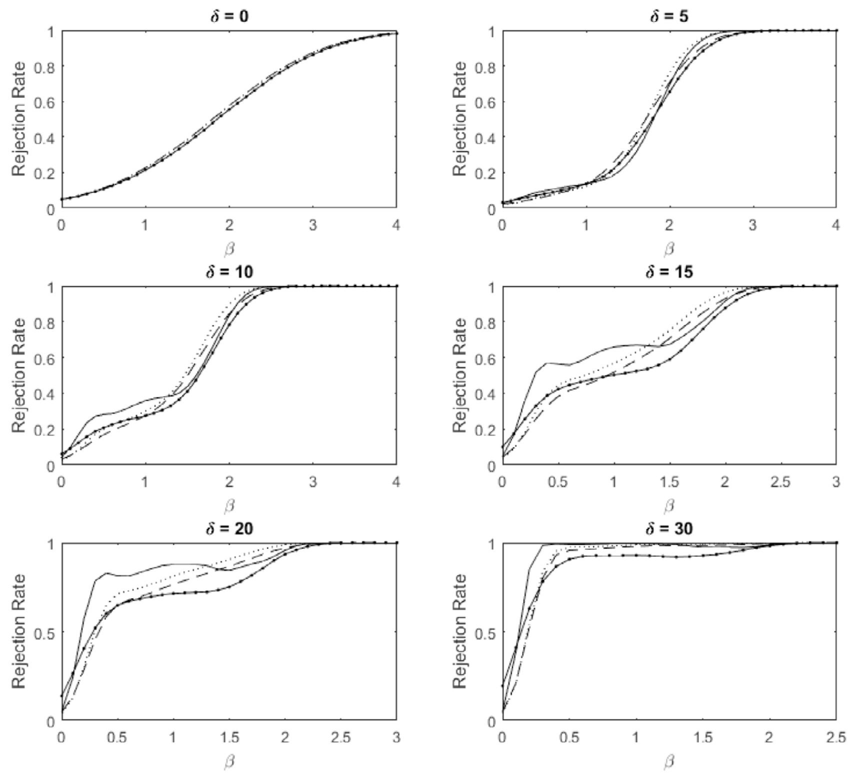

Fig. 5. For $\xi = 0$, the solid line indicates power for TR_1 and dots indicate power for TR_0. For ξ drawn from the unconditional distribution, the dotted solid line indicates power for TR_1 and dashes indicate power for TR_0.

clearly apparent in the simulation results, especially for the larger values for δ. That the TR_1 test controls size for both assumptions on the initial value, expected from the asymptotic results, is also apparent in the small sample results. This provides a strong motivation for using this statistic unless the researcher is sure that the initial value of zero makes sense when $r| < 1$.

4. Conclusion

We considered here the problem of testing for the value of a coefficient on a time trend in a linear model of a variable with a mean and time trend. Testing for this coefficient being zero then tests for the presence of the time trend. To test other null values for b simply subtract $b_0 t$ from the variable to be tested and one can use the tests described in this paper. Multiplying the variable to be tested by minus one yields a test that the coefficient is less than zero instead of greater than zero as in the paper. This test is non trivial because the nuisance parameter describing the largest root in a linear model of the serial correlation is a nuisance parameter under both the null and alternative hypotheses and hence any test must control size over all values of this nuisance parameter. With sufficiently mean reverting serial correlation, this is not much of an issue however when this root is near the unit circle then the value of this coefficient impacts tests greatly. Standard methods do not control size in these cases.

For this problem, power is much higher the more mean reverting the data, and so it makes sense in developing new tests to concentrate on the power properties of the tests for models where the serial correlation has a unit root or near unit root. This is the focus of this paper.

In previous work, authors have constructed tests that appear to control size but at unknown costs to power. These methods exploit the fact that when the residuals of the regression are very serially correlated standard tests are oversized, in this region imposing a unit root results in tests that are heavily undersized. By mixing the two tests that approximately control size can be computed. In this paper we use the methods of Elliott et al. (2015) to construct tests paying attention to the size for these difficult values of the nuisance parameter, where serial correlation is large. Although our tests share the switching behavior of previous tests, our tests only switch for models where roots are far from the unit circle. In the case where the initial condition of the stochastic component of the variable being examined is assumed asymptotically irrelevant – which is the assumption of almost all of the previous literature – we are able to construct a family of tests for the hypothesis. The test we detail (TR_0) has attractive asymptotic power properties relative to previously constructed tests.

Table 1
Presented are small sample sizes for various models. The second row and first column define the model for the pseudo data generation, the third column indicates the statistic for which results are reported. ξ UC refers to drawing the initial value from its unconditional distribution. Monte Carlo standard errors are 0.002.

	θ_1	θ_2	$T = 100$				$T = 250$			
			$\xi = 0$		ξ UC		$\xi = 0$		ξ UC	
			TR_0	TR_1	TR_0	TR_1	TR_0	TR_1	TR_0	TR_1
$\delta = 0$	0	0	0.047	0.048	0.048	0.047	0.047	0.048	0.048	0.047
	0.3	0	0.055	0.055	0.049	0.047	0.055	0.055	0.049	0.047
	0.5	0	0.063	0.065	0.050	0.049	0.063	0.065	0.050	0.049
	−0.3	0	0.051	0.054	0.053	0.055	0.054	0.054	0.053	0.055
	−0.5	0	0.079	0.076	0.072	0.070	0.079	0.076	0.072	0.070
	0	0.3	0.054	0.053	0.051	0.050	0.054	0.053	0.051	0.050
	0	0.5	0.060	0.058	0.059	0.058	0.060	0.058	0.059	0.058
	0	−0.3	0.062	0.059	0.066	0.063	0.062	0.059	0.066	0.063
	0	−0.5	0.146	0.132	0.159	0.146	0.146	0.132	0.159	0.146
$\delta = 5$	0	0	0.015	0.009	0.021	0.012	0.017	0.012	0.025	0.016
	0.3	0	0.011	0.007	0.018	0.010	0.015	0.009	0.021	0.014
	0.5	0	0.012	0.008	0.017	0.010	0.015	0.011	0.022	0.014
	−0.3	0	0.030	0.023	0.031	0.022	0.033	0.025	0.037	0.027
	−0.5	0	0.054	0.046	0.053	0.044	0.059	0.049	0.061	0.050
	0	0.3	0.018	0.012	0.026	0.016	0.022	0.015	0.032	0.022
	0	0.5	0.027	0.021	0.036	0.026	0.036	0.026	0.051	0.035
	0	−0.3	0.044	0.032	0.052	0.039	0.052	0.037	0.064	0.047
	0	−0.5	0.102	0.081	0.130	0.119	0.125	0.095	0.161	0.134
$\delta = 10$	0	0	0.020	0.013	0.029	0.019	0.029	0.018	0.044	0.027
	0.3	0	0.016	0.011	0.024	0.016	0.025	0.015	0.040	0.023
	0.5	0	0.017	0.012	0.024	0.015	0.027	0.016	0.040	0.021
	−0.3	0	0.038	0.026	0.046	0.033	0.048	0.032	0.062	0.039
	−0.5	0	0.062	0.046	0.071	0.057	0.074	0.053	0.087	0.063
	0	0.3	0.024	0.017	0.034	0.024	0.037	0.024	0.057	0.034
	0	0.5	0.035	0.027	0.045	0.037	0.055	0.036	0.076	0.050
	0	−0.3	0.051	0.034	0.066	0.051	0.066	0.041	0.094	0.061
	0	−0.5	0.093	0.064	0.140	0.118	0.110	0.073	0.174	0.130

However rather than constructing tests under the assumption of an asymptotically irrelevant initial condition for the data even when serial correlation is such that we expect the data to mean revert (around the potential trend), we recommend the tests of this paper (TR_1) derived under the alternative assumption that the initial value is drawn from its unconditional distribution in practice. These tests, unlike those derived for the zero initial condition case, control size under both assumptions on the initial condition. Power considerations also suggest use of TR_1, very little power is given up if the assumption on the initial condition is wrong. Assuming the initial condition arises from its unconditional distribution also makes more intuitive sense for this problem since for less serially correlated models the assumption results in the data being covariance stationary.

Appendix A

We detail construction of the tests

A.1. Construction of the tests

For the two versions of the TR statistics, we employ the following steps;

(1) Choose the lag length for constructing robust tests by using the BIC or MAIC with a maximum of $p_{\max}$ lags of $x(t)$ in a regression of $x(t)$ that includes a constant and time trend. Call the chosen lag length p.

(2) From regressions in (10) both including and excluding p lags of Δx_t construct an estimate $\hat{\delta}$ as

$$\hat{\delta} = -T\hat{p}_0\sqrt{\frac{\hat{\omega}^2}{\hat{\sigma}_v^2}}$$

where $\hat{p}_0$ is an estimate from the regression including the lags, $\hat{\omega}^2$ is an estimate of the spectral density at frequency zero using the residuals when the lagged changes of x_t are excluded and $\hat{\sigma}_v^2$ is an estimate of the variance of the residuals from the regression with the lagged changes of x_t included.

(3) If $\hat{\delta} > 100$, the test for $b = 0$ is the usual t-test of the hypothesis in (2) in the regression in step 2 where the lagged changes of x_t have been included in the regression.

Table 2
Presented are small sample sizes for various models. The second row and first column define the model for the pseudo data generation, the third column indicates the statistic for which results are reported. ξ UC refers to drawing the initial value from its unconditional distribution. Monte Carlo standard errors are 0.002.

	θ_1	θ_2	$T = 100$				$T = 250$			
			$\xi = 0$		ξ UC		$\xi = 0$		ξ UC	
			TR_0	TR_1	TR_0	TR_1	TR_0	TR_1	TR_0	TR_1
$\delta = 15$	0	0	0.029	0.017	0.039	0.026	0.042	0.022	0.070	0.036
	0.3	0	0.023	0.015	0.029	0.023	0.038	0.020	0.062	0.031
	0.5	0	0.021	0.015	0.028	0.023	0.036	0.022	0.060	0.031
	−0.3	0	0.047	0.030	0.057	0.039	0.061	0.034	0.086	0.046
	−0.5	0	0.067	0.045	0.087	0.063	0.080	0.049	0.113	0.069
	0	0.3	0.031	0.021	0.040	0.033	0.050	0.027	0.081	0.041
	0	0.5	0.044	0.032	0.050	0.045	0.068	0.039	0.102	0.055
	0	−0.3	0.058	0.035	0.078	0.058	0.072	0.040	0.116	0.065
	0	−0.5	0.079	0.048	0.132	0.109	0.086	0.047	0.161	0.114
$\delta = 20$	0	0	0.035	0.020	0.044	0.030	0.052	0.023	0.091	0.039
	0.3	0	0.028	0.016	0.035	0.027	0.046	0.022	0.085	0.034
	0.5	0	0.026	0.017	0.031	0.026	0.045	0.023	0.079	0.034
	−0.3	0	0.049	0.028	0.066	0.042	0.063	0.031	0.103	0.047
	−0.5	0	0.065	0.039	0.095	0.065	0.073	0.041	0.124	0.069
	0	0.3	0.037	0.022	0.044	0.036	0.057	0.027	0.100	0.045
	0	0.5	0.047	0.031	0.054	0.048	0.071	0.037	0.118	0.056
	0	−0.3	0.055	0.031	0.084	0.056	0.069	0.033	0.125	0.063
	0	−0.5	0.060	0.031	0.119	0.098	0.059	0.030	0.138	0.101
$\delta = 30$	0	0	0.037	0.015	0.052	0.032	0.052	0.017	0.107	0.035
	0.3	0	0.032	0.017	0.040	0.030	0.051	0.020	0.102	0.033
	0.5	0	0.032	0.019	0.037	0.030	0.052	0.023	0.100	0.035
	−0.3	0	0.044	0.018	0.070	0.043	0.051	0.021	0.113	0.044
	−0.5	0	0.048	0.026	0.089	0.059	0.049	0.026	0.112	0.059
	0	0.3	0.039	0.019	0.048	0.037	0.055	0.022	0.112	0.041
	0	0.5	0.045	0.025	0.055	0.047	0.063	0.029	0.118	0.052
	0	−0.3	0.042	0.016	0.078	0.054	0.046	0.019	0.111	0.055
	0	−0.5	0.033	0.017	0.099	0.085	0.029	0.017	0.100	0.083

(4) If $\hat{\delta} \leq 100$, we construct the WLR test. For this test we construct $\tilde{x}_t = x_t - x_1$. We construct from the data the following statistics:

$$y_1 = T^{-2}\hat{\omega}^{-2}\sum_{t=2}^{T}\tilde{x}_{t-1}^2$$

$$y_2 = T^{-2}\hat{\omega}^{-2}\sum_{t=2}^{T}\tilde{x}_{t-1}\Delta\tilde{x}_t + \frac{1}{2}\left(\frac{\hat{\sigma}_{\Delta w}^2}{\hat{\omega}^2} - 1\right)$$

$$y_3 = T^{-1/2}\hat{\omega}^{-1}\tilde{x}_T$$

$$y_4 = T^{-5/2}\hat{\omega}^{-1}\sum_{t=2}^{T}t\tilde{x}_{t-1}$$

$$y_5 = T^{-3/2}\hat{\omega}^{-1}\sum_{t=2}^{T}\tilde{x}_{t-1}$$

where $\hat{\sigma}_{\Delta w}^2$ is an estimator of the variance of the first difference of residuals from a regression of x_t on a constant and time trend and $\hat{\omega}^2$ is as in step 2.

For the case where $\xi = 0$ is assumed we have

$$\begin{aligned}\hat{H}_0(\delta_i, \beta_{ij}) &= \delta_i^2 y_1 + 2\delta_i y_2 - \frac{T^{-1}\left(\delta_i y_3 + \delta_i^2 y_5\right)^2}{(1+(T-1)(1-r_i)^2)} \\ &+ \beta_{ij}^2\left[T^{-1} + \delta_i^2 s_2 + r^2\left(\frac{T-1}{T}\right) + 2r_i\delta_i s_1 - \frac{T^{-1}\left(1 + r_i\delta_i(\frac{T-1}{T}) + \delta_i^2 s_1\right)^2}{(1+(T-1)(1-r_i)^2)}\right] \\ &- 2\beta_{ij}\left[(r_i+\delta_i)y_3 + \delta_i^2 y_4 - \delta_i^2 T^{-1} y_5 - \frac{\left(1 + r_i\delta_i(\frac{T-1}{T}) + \delta_i^2 s_1\right)}{T(1+(T-1)(1-r_i)^2)}\left(\delta_i y_3 + \delta_i^2 y_5\right)\right] \\ &+ \log\left(1 + (1-r_i)^2(T-1)\right).\end{aligned} \quad (11)$$

Table 3
Weights for WLR test construction. p_0 refers to weights for TR_0 and p_1 weights for TR_1.

Delta	p_0	p_1
0.000	1.3553	1.1186
0.010	0.6316	0.7397
0.040	0.0687	0.2181
0.090	0.0021	0.0304
0.160	0.0000	0.0022
0.250	0.0000	0.0001
0.360	0.0000	0.0000
0.490	0.0000	0.0000
0.640	0.0000	0.0000
0.810	0.0000	0.0000
1.000	0.0000	0.0000
1.210	0.0000	0.0000
1.440	0.0000	0.0000
1.690	0.0000	0.0000
1.960	0.0000	0.0000
2.250	0.0000	0.0000
2.560	0.0000	0.0000
2.890	0.0000	0.0000
3.240	0.0000	0.0000
3.610	0.0000	0.0000
4.000	0.0000	0.0000
6.250	0.0000	0.0000
9.000	0.0000	0.0000
12.250	0.0004	0.0000
16.000	0.0244	0.0013
20.250	0.0587	0.0808
25.000	0.0329	0.0893
30.250	0.0243	0.0375
36.000	0.0335	0.0363
42.250	0.0473	0.0515
49.000	0.0485	0.0513
56.250	0.0425	0.0469
64.000	0.0411	0.0489
72.250	0.0501	0.0485
81.000	0.0558	0.0442
90.250	0.0523	0.0448
100.000	0.0559	0.0552
110.250	0.0529	0.0532
121.000	0.0254	0.0156

where $s_1 = T^{-2}\sum_{t=2}^{T} t$ and $s_2 = T^{-3}\sum_{t=2}^{T} t^2$. For the case where the initial observation is drawn from its unconditional distribution we have

$$\begin{aligned}
\hat{H}_1(\delta_i,\beta_{ij}) &= \delta_i^2 y_1 + 2\delta_i y_2 - \frac{\left(\delta_i y_3 + \delta_i^2 y_5\right)^2}{T(1-r_i^2 1(\delta_i>0)+(T-1)(1-r_i)^2)} \\
&+ \beta_{ij}^2\left[(1-r_i 1(\delta_i>0))T^{-1} + \delta_i^2 s_2 + r^2\left(\frac{T-1}{T}\right) + 2r_i\delta_i s_1 - \frac{\left(1-r_i^2 1(\delta_i>0)+r_i\delta_i(\frac{T-1}{T})+\delta_i^2 s_1\right)^2}{T(1-r_i^2 1(\delta_i>0)+(T-1)(1-r_i)^2)}\right] \\
&- 2\beta_{ij}\left[(r_i+\delta_i)y_3 + \delta_i^2 y_4 - \delta_i^2 T^{-1} y_5 - \frac{\left(1-r_i^2 1(\delta_i>0)+r_i\delta_i(\frac{T-1}{T})+\delta_i^2 s_1\right)}{T(1-r_i^2 1(\delta_i>0)+(T-1)(1-r_i)^2)}\left(\delta_i y_3+\delta_i^2 y_5\right)\right] \\
&+ \log\left(1+\frac{(1-r_i)(T-1)}{1+r_i}\right).
\end{aligned} \qquad (12)$$

These forms allow vectorization of the programming of the test statistic which increases speed of computation (which is close to instantaneous on a standard personal computer) and allows a simpler generalization to serial correlation corrections when there is serial correlation in u_t.

The *WLR* statistic

$$WLR_k = \frac{\frac{1}{2m}\sum_{i=1}^{m}\sum_{i=j}^{2}\exp(-0.5\hat{H}_k(x,\delta_i,\beta_{ij}))}{\sum_{i=1}^{m} p_i \exp(-0.5\exp(\hat{H}_k(x,\delta_i,0)))}$$

for either $k = 0$ or $k = 1$ is then constructed using the weights p_{0i} from Table 3 for WLR_0 using the zero initial version ($\hat{H}_0(x,\delta_i,\beta_{ij})$) or weights p_{1i} for the WLR_1 statistic using the unconditional distribution version of $\hat{H}_1(x,\delta_i,\beta_{ij})$.

$\delta_i = \{0, 0.25^2, 0.5^2, \ldots, 9^2\}$ so $m = 39$, $J = 2$ with $\beta_{ij} = c_j/\sqrt{1-\delta_i^2/30}$ with values for $c_j = \{1, 3\}$ for WLR_0 and $\beta_{ij} = c_j/\sqrt{1-\delta_i^2/50}$ with values for $c_j = \{1, 3\}$ for WLR_1. The test rejects if $WLR > 1$.

(5) The test TR is then the test in step (3) if $\hat{\delta} > 100$ and the test in step (4) if $\hat{\delta} \leq 100$. The test TR_0 uses WLR_0 in step (4), the test TR_1 uses WLR_1 in step (4).

Appendix B. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jeconom.2020.03.006.

References

Andrews, D., Ploberger, W., 1994. Optimal tests when a nuisance parameter is present only under the alternative. Econometrica 62, 1383–1414.
Bunzel, H., Vogelsang, T., 2005. Powerful trend function tests that are robust to strong serial correlation with an application to the Prebisch-Singer hypothesis. J. Bus. Econom. Statist. 23, 381–394.
Canjels, E., Watson, M., 1997. Estimating deterministic trends in the presence of serially correlated errors. Rev. Econ. Stat. 79, 184–200.
Elliott, G., Mueller, U., Watson, M., 2015. Nearly optimal tests when a nuisance parameter is present under the null hypothesis. Econometrica 83, 771–811.
Harvey, D., Leybourne, S., Taylor, R., 2007. A simple, robust and powerful test of the trend hypothesis. J. Econometrics.
Harvey, D., Leybourne, S., Taylor, R., 2010. The impact of the initial condition on robust tests for a linear trend. J. Time Series Anal..
King, M., 1988. Towards a theory of point optimal testing. Econometric Rev. 6, 169–218.
King, M., 1996. Hypothesis testing in the presence of nuisance parameters. J. Statist. Plann. Inference 50, 103–120.
King, M., Srinanthkumar, S., 2017. Point Optimal Testing: A Survey of the Post 1987 Literature. working paper.
Müller, U., Elliott, G., 2003. Tests for unit roots and the initial condition. Econometrica 71, 1269–1286.
Perron, P., Ng, S., 1996. Useful modifications to some unit root tests with dependent errors and their local asymptotic properties. Rev. Econom. Stud. 63, 435–463.
Perron, P., Yabu, T., 2009. Estimating deterministic trends with an integrated or stationary noise component. J. Econometrics 151, 56–69.
Phillips, P., 1987. Time series regression with a unit root. Econometrica 55, 277–301.
Srinanthakumar, S., King, M., 2006. A new approximate point optimal test of a composite null hypothesis. J. Econometrics 130, 101–122.
Yue, S., Pilon, P., Phinney, B., Cavadias, G., 2002. The influence of autocorrelation on the ability to detect trend in hydrological series. Hydrol. Process. 16, 1807–1829.