

# The Impact of the Initial Condition on Covariate Augmented Unit Root Tests

Chrystalleni Aristidou, David I. Harvey\* and Stephen J. Leybourne  
School of Economics, University of Nottingham

January 2016

## Abstract

We examine the behaviour of OLS-demeaned/detrended and GLS-demeaned/detrended unit root tests that employ stationary covariates, as proposed by Hansen (1995) and Elliott and Jansson (2003), respectively, in situations where the magnitude of the initial condition of the time series under consideration may be non-negligible. We show that the asymptotic power of such tests is very sensitive to the initial condition; OLS- and GLS-based tests achieve relatively high power for large and small magnitudes of the initial condition, respectively. Combining information from both types of test via a simple union of rejections strategy is shown to effectively capture the higher power available across all initial condition magnitudes.

**Keywords:** Unit root tests; stationary covariates; initial condition uncertainty; asymptotic power.

**JEL Classification:** C22.

## 1 Introduction

Conventional testing for a unit root in a time series is typically carried out using the OLS-demeaning/detrending procedure of Dickey and Fuller (1979), or the GLS-demeaning/detrending procedure of Elliott, Rothenberg and Stock (1996). When the series under consideration covaries with an available stationary variable, Hansen (1995)

---

\*Corresponding author: School of Economics, University of Nottingham, University Park, Nottingham NG7 2RD, U.K. Email: [dave.harvey@nottingham.ac.uk](mailto:dave.harvey@nottingham.ac.uk)

showed that it is possible to substantially increase the power of the OLS-based unit root tests by augmenting the underlying OLS regression model with that stationary covariate. Elliott and Jansson (2003) and Westerlund (2013) show that incorporating covariates in a GLS-demeaning/detrending setting also improves the power of GLS-based unit root tests.

As shown in Müller and Elliott (2003), the powers of conventional OLS-based and GLS-based unit root tests are sensitive to the magnitude of the unobserved initial condition of a time series. For a small initial condition, GLS-based tests can have substantially more power than their OLS-based counterparts, while the reverse is true for a large initial condition. Typically, the power of OLS-based tests is an increasing function of this magnitude, whereas GLS-based tests demonstrate the opposite behaviour. In any practical testing situation, the magnitude of the initial condition is not known (nor can it be consistently estimated) and it is therefore unclear whether it is best to apply an OLS- or GLS-based unit root test in order to extract the most information about the presence, or otherwise, of a unit root. Harvey *et al.* (2009) examine the behaviour of a simple union of rejections strategy whereby (in its simplest guise) the unit root null hypothesis is rejected whenever either of the individual OLS- or GLS-based unit root tests rejects. This procedure is shown to perform well in practice since it captures the superior power of the GLS-based test for a small initial condition and the superior power of the OLS-based test for a large initial condition.

In this paper we show that the patterns of sensitivity of the power of OLS- and GLS-based covariate augmented unit root tests to the magnitude of the initial condition are actually very similar to that of their non-covariate augmented counterparts. This implies that the same considerations are relevant as in the non-covariate augmented case, when deciding which of the OLS- or GLS-based covariate augmented unit root tests to apply. Our proposed solution is once again to employ a union of rejections strategy, which we demonstrate is very effective in the covariate augmented context.

The plan of the paper is as follows. The next section sets out the model and describes the Hansen (1995) and Elliott and Jansson (2003) covariate augmented unit root tests. Here we also consider a simpler variant of the Elliott and Jansson (2003) GLS-based test following Westerlund (2013) which proves useful in the context of the union of rejections strategy. Section 3 derives the local asymptotic power functions of the tests in the presence of possibly non-negligible initial conditions and examines their asymptotic local powers. Section 4 introduces the union of rejections strategies and examines their large sample power properties. Finite sample power comparisons are shown in section 5, which also includes discussion of issues regarding the practical implementation of the recommended procedure. Section 6 concludes the paper. In

what follows,  $\mathbb{I}(\cdot)$  denotes the indicator function,  $L$  denotes the lag operator,  $\xrightarrow{p}$  denotes convergence in probability, and  $\Rightarrow$  denotes weak convergence.

## 2 The model and covariate augmented unit root tests

For purposes of transparency we will conduct our analysis within the context of a fairly simple model that admits a single covariate and abstracts from serial correlation in the innovations; we subsequently discuss elaborations to the case of more general serial correlation in section 5 below. We consider the following model for the series  $y_t$  and the covariate  $x_t$ ,  $t = 1, \dots, T$ :

$$\begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} \mu_y + \beta_y t \\ \mu_x + \beta_x t \end{bmatrix} + \begin{bmatrix} u_{y,t} \\ u_{x,t} \end{bmatrix} \quad (1)$$

where

$$\begin{bmatrix} u_{y,t} - \rho u_{y,t-1} \\ u_{x,t} \end{bmatrix} = \begin{bmatrix} v_t \\ e_t \end{bmatrix}. \quad (2)$$

Within this generic data generating process (DGP) specification we identify three alternative specifications for the deterministic components of  $y_t$  and  $x_t$ , with varying restrictions concerning the trend component of  $y_t$  and  $x_t$ :

$$\text{Model A : } \beta_y = \beta_x = 0$$

$$\text{Model B : } \beta_y \neq 0, \beta_x = 0$$

$$\text{Model C : } \beta_y \neq 0, \beta_x \neq 0$$

In Model A, no trends are assumed present in either  $y_t$  or  $x_t$ ; in Model B, a trend is permitted in  $y_t$  alone, while both  $y_t$  and  $x_t$  admit a trend in Model C. We make the following assumption regarding the innovations  $v_t$  and  $e_t$ :

**Assumption 1.** *The stochastic process  $\varepsilon_t = [v_t \ e_t]'$  is a martingale difference sequence with variance  $E(\varepsilon_t \varepsilon_t') = \Omega$  where*

$$\Omega = \begin{bmatrix} \sigma_v^2 & \sigma_{ev} \\ \sigma_{ev} & \sigma_e^2 \end{bmatrix}$$

and  $\sup_t E(\|e_t\|^4) < \infty$ . Let the squared correlation between the innovation  $v_t$  and the covariate  $u_{x,t} = e_t$  be denoted by

$$R^2 = \frac{\sigma_{ev}^2}{\sigma_v^2 \sigma_e^2}.$$

Within (2), for the autoregressive process  $u_{y,t}$  we set  $\rho = 1 + c/T$  for  $c \leq 0$ , with  $c = 0$  and  $c < 0$  corresponding to unit root and local-to-unit root autoregressive processes, respectively. Here  $u_{x,t} = e_t$  is the stationary covariate which is correlated with the innovation term of  $u_{y,t}$  when  $\sigma_{ev} \neq 0$  (i.e. when  $R^2 > 0$ ).

In this paper we wish to allow for the possibility that the initial condition of the autoregressive process  $u_{y,t}$ , i.e.  $u_{y,1}$ , is asymptotically non-negligible, so that its limiting effect on covariate augmented unit root tests can be ascertained. Specifically, the following assumption is made regarding the behaviour of  $u_{y,1}$ :

**Assumption 2.** *For  $c < 0$ , the initial condition is generated according to  $u_{y,1} = \alpha \sqrt{\sigma_v^2 / (1 - \rho^2)}$ , where  $\alpha$  is a fixed parameter. For  $c = 0$ , we may set  $u_{y,1} = 0$  without loss of generality, due to the exact similarity of the covariate augmented unit root tests to the initial condition in this case.*

In Assumption 2, the parameter  $\alpha$  controls the magnitude of the fixed initial condition  $u_{y,1}$  (i.e. the deviation of the initial observation from the underlying mean/trend in the data) relative to the standard deviation of a stationary  $AR(1)$  process with parameter  $\rho$  and innovation variance  $\sigma_v^2$ . This form for the initial condition is closely related to that given in Müller and Elliott (2003) and Harvey and Leybourne (2005). Notice also that, when  $c < 0$ , the initial value is not asymptotically negligible because  $T^{-1/2}u_{y,1} \rightarrow \alpha\sigma_v/\sqrt{-2c}$  as  $T \rightarrow \infty$ .

Our focus in this paper is on testing the unit root null  $H_0 : \rho = 1$  against the stationary alternative  $H_1 : \rho < 1$ , in the case where a stationary covariate is available. In the context of the model (1)-(2) and Assumption 1 we now outline statistics that derive from the Hansen (1995) and Elliott and Jansson (2003) approaches to covariate augmented unit root testing, which are respectively based on OLS and GLS detrending of the  $y_t$  data.

## 2.1 OLS-based statistics

The Hansen (1995) approach tests for a unit root in  $y_t$  using a Dickey-Fuller-type regression, augmented by the stationary covariate as an additional regressor, and implicitly employs OLS demeaning/detrending of the  $y_t$  and  $x_t$  series (note that Hansen does not consider Model C, but extension to this case is trivial). Based on our components representation of the DGP in (1), we express this type of statistic as follows:

$$t_{\hat{\phi}} = \frac{\hat{\phi}}{s.e.(\hat{\phi})}$$

where  $\hat{\phi}$  and  $s.e.(\hat{\phi})$  are the OLS estimate and associated standard error of  $\phi$  obtained from the regression

$$\Delta\hat{u}_{y,t} = \phi\hat{u}_{y,t-1} + \delta\hat{u}_{x,t} + \eta_t \quad (3)$$

with  $\hat{u}_{y,t}$  and  $\hat{u}_{x,t}$  denoting residuals from the OLS demeaned/detrended  $y_t$  and  $x_t$  series

$$\hat{u}_{y,t} = \begin{cases} y_t - \hat{\mu}_y & \text{for Model A} \\ y_t - \hat{\mu}_y - \hat{\beta}_y t & \text{for Models B, C} \end{cases}$$

$$\hat{u}_{x,t} = \begin{cases} x_t - \hat{\mu}_x & \text{for Models A, B} \\ x_t - \hat{\mu}_x - \hat{\beta}_x t & \text{for Model C} \end{cases}$$

where in the demeaned cases,  $\hat{\mu}_y$  and  $\hat{\mu}_x$  denote the estimated intercepts in the regressions of  $y_t$  and  $x_t$ , respectively, on a constant, while in the detrended cases,  $\hat{\mu}_y, \hat{\beta}_y$  and  $\hat{\mu}_x, \hat{\beta}_x$  denote the intercept, trend coefficient estimates in the regressions of  $y_t$  and  $x_t$ , respectively, on a constant and linear trend.

## 2.2 GLS-based statistics

Elliott and Jansson (2003) propose an approach to covariate augmented unit root testing based on a likelihood ratio principle combined with GLS demeaning/detrending for  $y_t$  but retaining OLS demeaning/detrending for the stationary covariate  $x_t$ . Specifically, for our basic model, their statistic is given by

$$\hat{\Lambda} = T \left\{ tr \left( \left[ \sum_{t=1}^T \hat{u}_t(1)\hat{u}_t(1)' \right]^{-1} \left[ \sum_{t=1}^T \hat{u}_t(\bar{\rho})\hat{u}_t(\bar{\rho})' \right] \right) - 1 - \bar{\rho} \right\}$$

where, for  $r = \bar{\rho} = 1 + \bar{c}/T$  (for some chosen  $\bar{c} < 0$ ) and  $r = 1$ ,

$$\hat{u}_t(r) = z_t(r) - d_t(r)'\hat{\beta}(r)$$

with

$$z_t(r) = \begin{bmatrix} (1 - r\mathbb{I}(t > 1)L)y_t \\ x_t \end{bmatrix}$$

$$d_t(r)' = \begin{cases} \begin{bmatrix} 1 - r\mathbb{I}(t > 1) & 0 \\ 0 & 1 \end{bmatrix} & \text{for Model A} \\ \begin{bmatrix} 1 - r\mathbb{I}(t > 1) & 0 & (1 - r\mathbb{I}(t > 1)L)t \\ 0 & 1 & 0 \end{bmatrix} & \text{for Model B} \\ \begin{bmatrix} 1 - r\mathbb{I}(t > 1) & 0 & (1 - r\mathbb{I}(t > 1)L)t & 0 \\ 0 & 1 & 0 & t \end{bmatrix} & \text{for Model C} \end{cases}$$

and

$$\hat{\beta}(r) = \left[ \sum_{t=1}^T d_t(r) \hat{\Omega}^{-1} d_t(r)' \right]^{-1} \left[ \sum_{t=1}^T d_t(r) \hat{\Omega}^{-1} z_t(r) \right]$$

where  $\hat{\Omega}$  is a consistent estimator of  $\Omega$ .

An alternative approach to covariate augmented unit root testing that also makes use of GLS demeaning/detrending for  $y_t$  is to adapt the Hansen (1995) Dickey-Fuller-based statistic, where the deterministic coefficients in (1) are estimated using GLS rather than OLS, an approach suggested by Westerlund (2013). Specifically, we consider the following GLS-based variant of Hansen's statistic:

$$t_{\tilde{\phi}} = \frac{\tilde{\phi}}{s.e.(\tilde{\phi})}$$

where  $\tilde{\phi}$  and  $s.e.(\tilde{\phi})$  are obtained from the fitted OLS regression

$$\Delta \tilde{u}_{y,t} = \tilde{\phi} \tilde{u}_{y,t-1} + \tilde{\delta} \hat{u}_{x,t} + \tilde{\eta}_t \quad (4)$$

with  $\hat{u}_{x,t}$  denoting residuals from the OLS demeaned/detrended  $x_t$  series as before, but now  $\tilde{u}_{y,t}$  denoting the GLS demeaned/detrended  $y_t$  series, obtained from an OLS regression of  $(1 - \bar{\rho}\mathbb{I}(t > 1)L)y_t$  on  $1 - \bar{\rho}\mathbb{I}(t > 1)$  for Model A, and  $(1 - \bar{\rho}\mathbb{I}(t > 1)L)y_t$  on  $[1 - \bar{\rho}\mathbb{I}(t > 1), (1 - \bar{\rho}\mathbb{I}(t > 1)L)t]'$  for Models B and C.

Both the  $\hat{\Lambda}$  and  $t_{\tilde{\phi}}$  GLS-based statistics rely on specifying a value of  $\bar{c}$ . Elliott and Jansson (2003) and Westerlund (2013) suggest using the Elliott *et al.* (1996) values of  $\bar{c} = -7$  for Model A and  $\bar{c} = -13.5$  for Models B and C. These choices are motivated by the value of  $c = \bar{c}$  for which the nominal 0.05-level asymptotic Gaussian local power envelope is at 0.50 in the non-covariate augmented case, which corresponds to the case of  $R^2 = 0$  in the context of the covariate augmented tests. As Elliott and Jansson (2003) and Westerlund (2013) note, it is also possible to select  $\bar{c}$  according to the value of  $R^2$ , so that the asymptotic Gaussian local power envelope is at 0.50 for any given  $R^2$ , but these authors do not recommend such an approach, arguing that unit root test power is increasing in  $R^2$  (for a given  $c$ ), and so base their choice of a single  $\bar{c}$  parameter on the lowest power scenario ( $R^2 = 0$ ). In what follows, we follow such previous work and set  $\bar{c} = -7$  for Model A and  $\bar{c} = -13.5$  for Models B and C.

### 3 Asymptotic results

In this section we derive the local asymptotic distributions for  $t_{\hat{\phi}}$ ,  $t_{\tilde{\phi}}$  and  $\hat{\Lambda}$  under Assumptions 1 and 2, when  $\rho = 1 + c/T$ ,  $c \leq 0$ . We make use of the following weak

convergence result

$$\begin{aligned} T^{-1/2} \sum_{t=1}^{\lfloor rT \rfloor} \begin{bmatrix} v_t \\ e_t \end{bmatrix} &\Rightarrow \begin{bmatrix} \sigma_v & 0 \\ \sigma_e R & \sqrt{\sigma_e^2(1-R^2)} \end{bmatrix} \begin{bmatrix} W_1(r) \\ W_2(r) \end{bmatrix} \\ &= \begin{bmatrix} \sigma_v W_1(r) \\ \sigma_e \{RW_1(r) + \sqrt{1-R^2}W_2(r)\} \end{bmatrix} \end{aligned}$$

where  $W_1(r)$  and  $W_2(r)$  are independent Brownian motions. The initial condition manifests itself via the result (see, for example, Müller and Elliott, 2003)

$$T^{-1/2}(u_{y, \lfloor rT \rfloor} - u_{y,1}) \Rightarrow \sigma_v K_c(r) \quad (5)$$

where

$$K_c(r) = \begin{cases} W_1(r) & c = 0 \\ \alpha(e^{rc} - 1)/\sqrt{-2c} + W_{1c}(r) & c < 0 \end{cases} \quad (6)$$

and  $W_{1c}(r)$  is the Ornstein-Uhlenbeck process

$$W_{1c}(r) = c \int_0^r e^{c(r-s)} W_1(s) ds + W_1(r).$$

The following theorem now provides the limit distributions of the three covariate augmented unit root statistics, the proof of which can be found in the companion working paper version Aristidou *et al.* (2016) [ALT].<sup>1</sup>

**Theorem 1** *For the DGP given by (1)-(2), under Assumptions 1 and 2, with  $\rho = 1 + c/T$ ,  $c \leq 0$ ,*

*(i) For Model  $i$  ( $i = A, B, C$ ),*

$$t_{\hat{\phi}} \Rightarrow \frac{c}{\sqrt{1-R^2}} \sqrt{\int_0^1 L_c^i(r)^2 dr} + \sqrt{1-R^2} \frac{\int_0^1 L_c^i(r) dW_1(r)}{\sqrt{\int_0^1 L_c^i(r)^2 dr}} - R \frac{\int_0^1 L_c^i(r) dW_2(r)}{\sqrt{\int_0^1 L_c^i(r)^2 dr}}$$

where

$$\begin{aligned} L_c^A(r) &= K_c(r) - \int_0^1 K_c(s) ds \\ L_c^B(r) &= L_c^C(r) = K_c(r) - \left\{ 4 \int_0^1 K_c(s) ds - 6 \int_0^1 s K_c(s) ds \right\} - \left\{ 12 \int_0^1 s K_c(s) ds - 6 \int_0^1 K_c(s) ds \right\} r. \end{aligned}$$

---

<sup>1</sup>Note that the key result (6) which shows how the initial condition enters the limit distributions is unchanged if  $\alpha$  is not a fixed parameter but is instead a random variable. Of course, the limit  $K_c(r)$  will depend on the distribution assumed for  $\alpha$  in such a case.

(ii) For Model  $i$  ( $i = A, B, C$ ),

$$\begin{aligned}
t_{\bar{\phi}} \Rightarrow & \frac{c}{\sqrt{1-R^2}} \sqrt{\int_0^1 M_{c,\bar{c}}^i(r)^2 dr} + \sqrt{1-R^2} \frac{\int_0^1 M_{c,\bar{c}}^i(r) dW_1(r)}{\sqrt{\int_0^1 M_{c,\bar{c}}^i(r)^2 dr}} - R \frac{\int_0^1 M_{c,\bar{c}}^i(r) dW_2(r)}{\sqrt{\int_0^1 M_{c,\bar{c}}^i(r)^2 dr}} \\
& + \frac{c\alpha}{\sqrt{-2c(1-R^2)}} \frac{\int_0^1 M_{c,\bar{c}}^i(r) dr}{\sqrt{\int_0^1 M_{c,\bar{c}}^i(r)^2 dr}} + \frac{1}{\sqrt{1-R^2}} \frac{N_{c,\bar{c}}^i \left\{ c \int_0^1 r M_{c,\bar{c}}^i(r) dr - \int_0^1 M_{c,\bar{c}}^i(r) dr \right\}}{\sqrt{\int_0^1 M_{c,\bar{c}}^i(r)^2 dr}} \\
& + \frac{R}{\sqrt{1-R^2}} \frac{P^i \int_0^1 M_{c,\bar{c}}^i(r) dr + Q^i \int_0^1 r M_{c,\bar{c}}^i(r) dr}{\sqrt{\int_0^1 M_{c,\bar{c}}^i(r)^2 dr}}
\end{aligned}$$

where

$$M_{c,\bar{c}}^A(r) = K_c(r)$$

$$M_{c,\bar{c}}^B(r) = M_{c,\bar{c}}^C(r) = K_c(r) - \left\{ \bar{c}^* K_c(1) + 3(1-\bar{c}^*) \int_0^1 s K_c(s) ds \right\} r$$

$$N_{c,\bar{c}}^A = 0$$

$$N_{c,\bar{c}}^B = N_{c,\bar{c}}^C = \bar{c}^* K_c(1) + 3(1-\bar{c}^*) \int_0^1 r K_c(r) dr$$

$$P^A = P^B = RW_1(1) + \sqrt{1-R^2} W_2(1)$$

$$P^C = 4 \left\{ RW_1(1) + \sqrt{1-R^2} W_2(1) \right\} - 6 \left\{ R \int_0^1 r dW_1(r) + \sqrt{(1-R^2)} \int_0^1 r dW_2(r) \right\}$$

$$Q^A = Q^B = 0$$

$$Q^C = 12 \left\{ R \int_0^1 r dW_1(r) + \sqrt{(1-R^2)} \int_0^1 r dW_2(r) \right\} - 6 \left\{ RW_1(1) + \sqrt{1-R^2} W_2(1) \right\}$$

with  $\bar{c}^* = (1-\bar{c} + \bar{c}^2/3)^{-1}(1-\bar{c})$ .

(iii) For Model  $i$  ( $i = A, B, C$ ),

$$\hat{\Lambda} \Rightarrow G_{c,\bar{c}}^i + H_{c,\bar{c}}^i + \frac{R^2}{1-R^2} (\bar{c}^2 - 2c\bar{c}) \int_0^1 S_c^i(r)^2 dr + \frac{R}{\sqrt{1-R^2}} 2\bar{c} \int_0^1 S_c^i(r) dW_2(r)$$

where

$$G_{c,\bar{c}}^A = \bar{c}^2 \int_0^1 K_c(r)^2 dr - \bar{c} K_c(1)$$

$$G_{c,\bar{c}}^B = G_{c,\bar{c}}^C = \bar{c}^2 \int_0^1 K_c(r)^2 dr + (1-\bar{c}) K_c(1)^2 - k_{\bar{c}}^{-1} \left\{ (1-\bar{c}) K_c(1) + \bar{c}^2 \int_0^1 r K_c(r) dr \right\}^2$$

$$H_{c,\bar{c}}^A = H_{c,\bar{c}}^C = 0$$

$$\begin{aligned}
H_{c,\bar{c}}^B = & k_{\bar{c}}^{-1} \left\{ (1-\bar{c}) K_c(1) + \bar{c}^2 \int_0^1 r K_c(r) dr \right\}^2 - \left\{ k_{\bar{c}} + \frac{\bar{c}^2 R^2}{12(1-R^2)} \right\}^{-1} \times \\
& \left[ (1-\bar{c}) K_c(1) + \bar{c}^2 \int_0^1 r K_c(r) dr + \frac{R^2}{1-R^2} \left\{ \frac{\bar{c}(c-\bar{c})}{2} \int_0^1 K_c(r) dr - \bar{c}(c-\bar{c}) \int_0^1 r K_c(r) dr \right\} \right. \\
& \left. + \frac{R}{\sqrt{1-R^2}} \left\{ \bar{c} \int_0^1 r dW_2(r) - \frac{\bar{c}}{2} \int_0^1 dW_2(r) \right\} \right]^2
\end{aligned}$$

$$\begin{aligned}
S_c^A(r) &= S_c^B(r) = K_c(r) - \int_0^1 K_c(s) ds \\
S_c^C(r) &= K_c(r) - (4 - 6r) \int_0^1 K_c(s) ds - (12r - 6) \int_0^1 s K_c(s) ds
\end{aligned}$$

with  $k_{\bar{c}} = 1 - \bar{c} + \bar{c}^2/3$ .

We now consider numerical results for the asymptotic properties of the tests presented so far in this paper. In Table 1 we report asymptotic null (left-tail) critical values for  $R^2 = \{0, 0.1, \dots, 0.9\}$  for all tests at the nominal 0.10-, 0.05- and 0.01-levels, which were obtained by direct simulation of the limit representations of Theorem 1 with  $c = 0$  (note from (6) that the limits are not dependent on  $\alpha$  when  $c = 0$ ). For all asymptotic results in this paper, we conducted Monte Carlo simulations using Gauss 9.0 with 50,000 replications, approximating the Brownian motion processes  $W_1(r)$  and  $W_2(r)$  using independent  $NIID(0, 1)$  random variates for each, and approximating the corresponding integrals by normalized sums of 2000 steps.

To gain some insight into the relative power performance of the three tests  $t_{\hat{\phi}}$ ,  $t_{\tilde{\phi}}$  and  $\hat{\Lambda}$ , we first abstract from the effect of the initial condition by making the usual assumption that it is asymptotically negligible. The limit distributions of the statistics are then as given in Theorem 1 on setting  $\alpha = 0$ . Figures 1 and 2 show the local asymptotic powers of the tests conducted at the nominal 0.05-level as functions of  $-c = \{0, 0.5, \dots, 20.0\}$  (with  $c = 0$  corresponding to asymptotic size, i.e. 0.05) for Models A and B, respectively, for  $R^2 = \{0.2, 0.4, 0.6, 0.8\}$ .

Consider first the results for Model A in Figure 1. We observe that for smaller values of  $R^2$ , the familiar result of the GLS-based tests delivering a substantial power advantage relative to the OLS-based test  $t_{\hat{\phi}}$  is borne out. These power advantages, however, diminish as  $R^2$  increases, so that by  $R^2 = 0.8$ , there is considerably less difference between the power profiles of the three tests. For the two GLS-based tests, there is very little to choose between them for small to moderate  $R^2$ , while for larger  $R^2$ , the  $\hat{\Lambda}$  test has slightly lower power than  $t_{\tilde{\phi}}$  for small  $-c$  but modestly higher power for some larger  $-c$ . In Figure 2 (Model B), as expected we see a reduction in power of all tests relative to Model A, brought about by the allowance of a trend in  $y_t$ . However, it is still the case that for small  $R^2$ , the GLS-based tests outperform  $t_{\hat{\phi}}$ , and have similar levels of power to each other. Interestingly, as  $R^2$  increases,  $t_{\tilde{\phi}}$  becomes generally more powerful than  $\hat{\Lambda}$ , and by  $R^2 = 0.8$ , the relative power levels of  $\hat{\Lambda}$  have reduced to values generally below those of the OLS-based test  $t_{\hat{\phi}}$ . In contrast,  $t_{\tilde{\phi}}$  retains a power advantage over  $t_{\hat{\phi}}$  for all  $R^2$  considered. Overall, from the results of Figures 1-2 we conclude that, if we abstract from potentially non-negligible initial conditions, on balance the  $t_{\tilde{\phi}}$  test offers arguably the most appealing power profile of the tests considered.

We now examine the effects of an asymptotically non-negligible initial condition on local asymptotic power. Figures 3 and 4 report local asymptotic powers of the five tests conducted at the nominal 0.05-level, as functions of  $\alpha = \{0, 0.1, \dots, 4.0\}$  ( $\alpha = 0$  corresponding to an asymptotically negligible initial condition); note that replacing  $\alpha$  with  $-\alpha$  would give the same results. Figure 3 presents results for Model A, where a representative local alternative setting of  $c = -5$  is used, while Figure 4 gives results for Model B using  $c = -10$ .

Considering first Model A in Figure 3, the stand out feature of these power curves is that while the power of the OLS-based  $t_{\hat{\phi}}$  test is increasing in the magnitude of the initial condition  $\alpha$ , the powers of both GLS-based tests decrease to zero as  $\alpha$  increases. Hence, while the GLS tests are more powerful for  $\alpha = 0$  (cf. Figure 1) and for small values of  $\alpha$ , they are much less powerful than  $t_{\hat{\phi}}$  for larger initial conditions. This pattern of results closely mirrors what is found when analyzing the effects of initial conditions on standard, non-covariate augmented OLS and GLS demeaned/detrended unit root tests, and highlights the fact that GLS-based unit root tests do not deliver reliable unit root test inference in the presence of large initial conditions. Between the two GLS-based tests, there is little difference between the power profiles for  $R^2 = 0.2$  and  $R^2 = 0.4$ , while as  $R^2$  increases to 0.6 and then 0.8, it is clear that  $\hat{\Lambda}$  emerges as the more powerful procedure. For Model B (Figure 4), we observe the same broad patterns of results vis-à-vis the power of  $t_{\hat{\phi}}$  compared to the GLS-based tests. Once again,  $t_{\hat{\phi}}$  has power that increases in  $\alpha$ , while the GLS-based tests have higher power when  $\alpha = 0$ , but then a decreasing power profile as  $\alpha$  rises. Between the GLS-based tests, in contrast to Model A, here we see that  $t_{\hat{\phi}}$  offers generally the best levels of power across  $\alpha$ , particularly for larger  $\alpha$  and  $R^2$  values.<sup>2</sup>

## 4 A union of rejections strategy

The results of the previous section demonstrate that when the initial condition is small, we would want to apply one of the two GLS-based tests ( $t_{\hat{\phi}}$  or  $\hat{\Lambda}$ ); on the other hand, when the initial condition is larger, applying such a test would result in a (potentially substantial) loss of power relative to applying the OLS-based test  $t_{\hat{\phi}}$ . In practice, given

---

<sup>2</sup>In ALT, results are also provided for Model C. Similar comments apply as for Model B in Figure 2, although for  $R^2 = 0.8$ , the power of  $t_{\hat{\phi}}$  also drops slightly below that of  $t_{\hat{\phi}}$ , so that here  $t_{\hat{\phi}}$  generally outperforms both GLS-based tests. The power profiles for Model C across  $\alpha$  again highlight that the GLS-based tests are typically more powerful than  $t_{\hat{\phi}}$  for zero and small  $\alpha$ , with the reverse ranking for larger  $\alpha$ . As with Model A, the powers of the two GLS-based tests are very similar for  $R^2 = \{0.2, 0.4\}$ , while  $\hat{\Lambda}$  performs better for the larger  $R^2$ .

uncertainty regarding the magnitude of the initial condition, we wish to have available a procedure that capitalizes on both the relatively high power of the GLS approach when  $\alpha$  is small, and the relatively high power of  $t_{\hat{\phi}}$  otherwise. A similar issue arises in the case of non-covariate augmented unit root testing, and the approach proposed by Harvey *et al.* (2009) is to take a *union of rejections* of the OLS- and GLS-based tests, whereby the null hypothesis is rejected if either of the individual tests rejects. In the present context, this implies taking a union of rejections between  $t_{\hat{\phi}}$  and either one of the GLS-based tests.

We now set out the union of rejections approach based on  $t_{\hat{\phi}}$  (here denoted by  $t_{OLS}$ ) and a GLS-based test ( $t_{\hat{\phi}}$  or  $\hat{\Lambda}$ ) denoted by  $t_{GLS}$ . Denoting the asymptotic  $\xi$ -level critical values of these tests by  $cv_{\xi}^{OLS}$  and  $cv_{\xi}^{GLS}$ , respectively, we can define the simple union of rejections strategy by the decision rule

$$\text{Reject } H_0 \text{ if } \{t_{OLS} < cv_{\xi}^{OLS} \text{ or } t_{GLS} < cv_{\xi}^{GLS}\}.$$

An alternative way of representing this decision rule is to express it in terms a single test statistic,  $t_{UR}$ , as follows:

$$\text{Reject } H_0 \text{ if } \left\{ t_{UR} = \min \left( t_{OLS}, \frac{cv_{\xi}^{OLS}}{cv_{\xi}^{GLS}} t_{GLS} \right) < cv_{\xi}^{OLS} \right\}.$$

If we use  $\mathcal{L}^{OLS}$  and  $\mathcal{L}^{GLS}$  to denote the generic joint limit distributions of  $t_{OLS}$  and  $t_{GLS}$ , respectively (i.e. the right-hand-side expressions given in Theorem 1), an application of the continuous mapping theorem establishes that

$$t_{UR} \Rightarrow \min \left( \mathcal{L}^{OLS}, \frac{cv_{\xi}^{OLS}}{cv_{\xi}^{GLS}} \mathcal{L}^{GLS} \right).$$

The Bonferroni bound for the asymptotic size of this procedure under the null is  $2\xi$  (since it simply involves rejecting the null when either of the individual tests reject). Harvey *et al.* (2009) suggest restoring the union of rejections asymptotic size to the nominal level  $\xi$  by applying a common positive scaling constant,  $\psi_{\xi} > 1$ , to the (negative) critical values  $cv_{\xi}^{OLS}$  and  $cv_{\xi}^{GLS}$  (so that  $t_{OLS}$  is compared with  $\psi_{\xi}cv_{\xi}^{OLS}$  and  $t_{GLS}$  with  $\psi_{\xi}cv_{\xi}^{GLS}$ ), such that in the limit, rejection of the null occurs with probability  $\xi$ . While this approach extends naturally to the covariate augmented unit root testing problem when using  $t_{\hat{\phi}}$  for  $t_{GLS}$ , since here both  $cv_{\xi}^{OLS}$  and  $cv_{\xi}^{GLS}$  are negative, we cannot apply such a simple adjustment when using  $\hat{\Lambda}$ , since the latter test has *positive* critical values, and scaling by  $\psi_{\xi} > 1$  would induce *greater* asymptotic size. An adjustment that is applicable in all cases is to first apply a common additive adjustment

to both  $t_{GLS}$  and  $cv_{\xi}^{GLS}$ , say  $t_{GLS} - \lambda_{\xi}$  and  $cv_{\xi}^{GLS} - \lambda_{\xi}$ , such that the GLS-based test decision rule is unchanged, but that the adjusted critical value  $cv_{\xi}^{GLS} - \lambda_{\xi}$  equals  $cv_{\xi}^{OLS}$ , i.e.  $\lambda_{\xi} = cv_{\xi}^{GLS} - cv_{\xi}^{OLS}$ . Once the critical values are lined up in this way, the critical values are both negative, and a Harvey *et al.* (2009)-type multiplicative scaling can be applied to control the asymptotic size. More formally we propose the following union of rejections decision rule:

$$\text{Reject } H_0 \text{ if } \{t_{OLS} < \psi_{\xi} cv_{\xi}^{OLS} \text{ or } t_{GLS} - \lambda_{\xi} < \psi_{\xi} cv_{\xi}^{OLS}\}$$

or, equivalently,

$$\text{Reject } H_0 \text{ if } \{t_{UR}^* = \min(t_{OLS}, t_{GLS} - \lambda_{\xi}) < \psi_{\xi} cv_{\xi}^{OLS}\}.$$

In the limit we obtain

$$t_{UR}^* \Rightarrow \min(\mathcal{L}^{OLS}, \mathcal{L}^{GLS} - \lambda_{\xi})$$

and we compute  $\psi_{\xi}$  by simulation of the limit distribution of  $t_{UR}^*$ , calculating the  $\xi$ -level null critical value for this distribution, say  $cv_{\xi}^{UR}$ , and then evaluate  $\psi_{\xi}$  as  $\psi_{\xi} = cv_{\xi}^{UR} / cv_{\xi}^{OLS}$ . In what follows, we consider two union of rejections procedures, one based on a union of  $t_{\hat{\phi}}$  and  $t_{\tilde{\phi}}$ , the other based on a union of  $t_{\hat{\phi}}$  and  $\hat{\Lambda}$ ; hereafter we denote these unions by  $UR(t_{\hat{\phi}}, t_{\tilde{\phi}})$  and  $UR(t_{\hat{\phi}}, \hat{\Lambda})$  respectively. Values for  $\psi_{\xi}$  for  $R^2 = \{0, 0.1, \dots, 0.9\}$  at the nominal 0.10-, 0.05- and 0.01-levels are given in Table 2 for each of these union of rejections strategies.

The union of rejection strategies  $UR(t_{\hat{\phi}}, t_{\tilde{\phi}})$  and  $UR(t_{\hat{\phi}}, \hat{\Lambda})$  are by construction asymptotically correctly sized. We now consider the asymptotic local power properties of  $UR(t_{\hat{\phi}}, t_{\tilde{\phi}})$  and  $UR(t_{\hat{\phi}}, \hat{\Lambda})$  in relation to the powers of the individual tests, the results for which are also displayed in Figures 3-4. Consider first Model A in Figure 3, and to aid comparison of the union of rejections procedures, consider an informal (and infeasible) power ‘‘envelope’’ formed from the limit power of  $\hat{\Lambda}$  for values of  $\alpha$  up to the point where  $\hat{\Lambda}$  and  $t_{\hat{\phi}}$  have the same power, and from the limit power of  $t_{\hat{\phi}}$  for  $\alpha$  beyond this point. With reference to this envelope, both  $UR(t_{\hat{\phi}}, t_{\tilde{\phi}})$  and  $UR(t_{\hat{\phi}}, \hat{\Lambda})$  track its broad shape, offering decent power levels across the range of  $\alpha$  values considered. Both  $UR(t_{\hat{\phi}}, t_{\tilde{\phi}})$  and  $UR(t_{\hat{\phi}}, \hat{\Lambda})$  capture much of the power advantage that  $t_{\tilde{\phi}}$  and  $\hat{\Lambda}$  hold over  $t_{\hat{\phi}}$  for small  $\alpha$ , while also achieving the substantial power gain that  $t_{\hat{\phi}}$  holds over either of the GLS-based tests for larger  $\alpha$ , with power profiles that are increasing in  $\alpha$  as opposed to approaching zero. Of the two union of rejections procedures, a trade-off clearly exists between the higher power for small  $\alpha$  that  $UR(t_{\hat{\phi}}, \hat{\Lambda})$  achieves, and the higher power for larger  $\alpha$  that  $UR(t_{\hat{\phi}}, t_{\tilde{\phi}})$  displays. However, in the small- $\alpha$  region where  $UR(t_{\hat{\phi}}, \hat{\Lambda})$  outperforms  $UR(t_{\hat{\phi}}, t_{\tilde{\phi}})$ , the power differences are relatively modest,

while in the larger- $\alpha$  range where  $UR(t_{\hat{\phi}}, t_{\hat{\tau}})$  outperforms  $UR(t_{\hat{\phi}}, \hat{\Lambda})$ , the gains can be quite substantial. For this reason, we consider that  $UR(t_{\hat{\phi}}, t_{\hat{\tau}})$  offers the more preferable power profile of the two procedures across the range of  $\alpha$  considered. Indeed, it could be argued that overall  $UR(t_{\hat{\phi}}, \hat{\Lambda})$  has an inferior power profile to that of  $t_{\hat{\phi}}$  alone, due to the relatively low power levels for modest to large  $\alpha$ . On the other hand,  $UR(t_{\hat{\phi}}, t_{\hat{\tau}})$  has a power profile closer to that of  $t_{\hat{\phi}}$  than does  $UR(t_{\hat{\phi}}, \hat{\Lambda})$  for this region of  $\alpha$ , while still achieving much of the power gains of  $t_{\hat{\tau}}$  over  $t_{\hat{\phi}}$  when  $\alpha$  is closer to zero. This union of rejections therefore offers decent power gains over  $t_{\hat{\phi}}$  for the arguably more typical case of small  $\alpha$ , while simultaneously providing insurance against the low power for large  $\alpha$  that is associated with  $t_{\hat{\phi}}$ . For Model B in Figure 4, broadly similar comments can be applied, with  $UR(t_{\hat{\phi}}, t_{\hat{\tau}})$  outperforming  $UR(t_{\hat{\phi}}, \hat{\Lambda})$  overall, and  $UR(t_{\hat{\phi}}, t_{\hat{\tau}})$  having a power profile that more closely tracks the shape of the informal envelope comprised of the best performing tests for each region of  $\alpha$ . If anything, compared to Model A, the power differences between  $UR(t_{\hat{\phi}}, t_{\hat{\tau}})$  and the informal envelope are less marked, adding weight to our recommendation for  $UR(t_{\hat{\phi}}, t_{\hat{\tau}})$ .<sup>3,4</sup>

## 5 Finite sample comparisons

In this section we consider the finite sample behaviour of the individual tests of section 2 and the proposed union of rejections procedures  $UR(t_{\hat{\phi}}, t_{\hat{\tau}})$  and  $UR(t_{\hat{\phi}}, \hat{\Lambda})$  under Assumptions 1 and 2. In order to implement the tests in such a setting, we first require a consistent estimator of  $R^2$ , given that all the tests have critical values that depend on this unknown quantity (the union of rejections procedures also require  $R^2$ -dependent scaling values  $\psi_{\xi}$ ). Under our assumptions, the estimator

$$\hat{R}^2 = \frac{\hat{\sigma}_{ev}^2}{\hat{\sigma}_v^2 \hat{\sigma}_e^2}$$

where

$$\hat{\sigma}_v^2 = T^{-1} \sum_{t=2}^T \hat{v}_t^2, \quad \hat{\sigma}_e^2 = T^{-1} \sum_{t=1}^T \hat{e}_t^2, \quad \hat{\sigma}_{ev} = T^{-1} \sum_{t=2}^T \hat{e}_t \hat{v}_t$$

---

<sup>3</sup>Results for Model C are again provided in ALT, and similar comments apply as for Model B in Figure 4.

<sup>4</sup>We also considered a variant of the  $UR(t_{\hat{\phi}}, t_{\hat{\tau}})$  procedure where the asymptotic size is controlled using only a multiplicative scaling to  $cv_{\xi}^{OLS}$  and  $cv_{\xi}^{GLS}$  as in Harvey *et al.* (2009). We found that such a variant led to near identical local asymptotic power profiles to those displayed in Figure 4 (Model B). In the case of Figure 3 (Model A),  $UR(t_{\hat{\phi}}, t_{\hat{\tau}})$  yielded slight local asymptotic power advantages over this variant for larger values of  $\alpha$ , while surrendering very little power to  $UR(t_{\hat{\phi}}, t_{\hat{\tau}})$  for small  $\alpha$ , hence we do not consider this variant further.

with  $\hat{e}_t = \hat{u}_{x,t}$  and  $\hat{v}_t$  being the residual from a regression of  $\hat{u}_{y,t}$  on  $\hat{u}_{y,t-1}$  can be shown to provide a consistent estimator of  $R^2$ . Additionally, as highlighted in section 2.2, the  $\hat{\Lambda}$  statistic requires a consistent estimator of  $\Omega$ . Given that  $\Omega$  is also comprised of  $\sigma_v^2$ ,  $\sigma_e^2$  and  $\sigma_{ev}$ , a natural estimator is to use

$$\hat{\Omega} = \begin{bmatrix} \hat{\sigma}_v^2 & \hat{\sigma}_{ev} \\ \hat{\sigma}_{ev} & \hat{\sigma}_e^2 \end{bmatrix}$$

which can be shown to be consistent for  $\Omega$ . Note that both  $\hat{R}^2$  and  $\hat{\Omega}$  remain consistent when  $\alpha$  in Assumption 2 is not equal to zero; in contrast, the corresponding estimators outlined in Elliott and Jansson (2003, p.81) are only consistent under the local-to-unit root alternative when the initial condition is asymptotically negligible (i.e.  $\alpha = 0$ ), due to their reliance on a first difference-based (cf. GLS-based) demeaning/detrending of  $y_t$ .

Our Monte Carlo simulations are based on generating (1)-(2) for  $T = 150$  using 50,000 replications, with  $\varepsilon_t = [v_t \ e_t]' \sim IIN(0, \Omega)$ ,  $\sigma_v^2 = \sigma_e^2 = 1$ ,  $\sigma_{ev}^2 = R^2 = \{0.2, 0.4, 0.6, 0.8\}$ , and with  $\mu_y = \beta_y = \mu_x = \beta_x = 0$ . We first simulated the empirical size of the  $t_{\hat{\phi}}$ ,  $t_{\tilde{\phi}}$ ,  $\hat{\Lambda}$  tests and the  $UR(t_{\hat{\phi}}, t_{\tilde{\phi}})$ ,  $UR(t_{\hat{\phi}}, \hat{\Lambda})$  procedures at the nominal 0.05-level, setting  $\rho = 1$  in (2). Asymptotic critical values and  $\psi_\xi$  values were used, linearly interpolating between the values in Tables 1 and 2 on the basis of  $\hat{R}^2$ . The results for Models A and B are reported in Table 3, and we observe only modest finite sample size distortions across the different tests and values of  $R^2$ . For larger  $R^2$  in the case of Model B,  $\hat{\Lambda}$  and  $UR(t_{\hat{\phi}}, \hat{\Lambda})$  are a little under-sized, while  $t_{\tilde{\phi}}$  and  $UR(t_{\tilde{\phi}}, t_{\tilde{\phi}})$  are a little over-sized for all cases. However, all sizes for  $t_{\tilde{\phi}}$  and  $UR(t_{\tilde{\phi}}, t_{\tilde{\phi}})$  are below 0.07 and 0.06 respectively, hence finite sample size distortion does not appear to be a major concern for these procedures.<sup>5</sup>

Of most interest are the relative finite sample powers of the procedures, and Figure 5 presents results for Model A, for settings that correspond to the local asymptotic power results in Figure 3. Here, we set  $\rho = 1 + c/T$  with  $c = -5$ , and report the estimated powers of nominal 0.05-level tests across  $\alpha = \{0, 0.1, \dots, 4.0\}$ . We find that the relative finite sample powers bear a very close resemblance to the corresponding local asymptotic results, with the powers of  $t_{\hat{\phi}}$  increasing in  $\alpha$ , the powers of  $t_{\tilde{\phi}}$  and  $\hat{\Lambda}$  initially higher than that for  $t_{\hat{\phi}}$  for small  $\alpha$ , but then falling towards zero as  $\alpha$  increases, and the  $UR(t_{\hat{\phi}}, t_{\tilde{\phi}})$  and  $UR(t_{\hat{\phi}}, \hat{\Lambda})$  procedures capturing a proportion of the higher  $t_{\tilde{\phi}}$  and  $\hat{\Lambda}$  power for small  $\alpha$ , and a proportion of the higher power of  $t_{\hat{\phi}}$  for larger  $\alpha$ . Compared to the asymptotic results,  $t_{\tilde{\phi}}$  and  $UR(t_{\hat{\phi}}, t_{\tilde{\phi}})$  appear to have higher

---

<sup>5</sup>Results for Model C are very similar to those for Model B, as shown in ALT.

relative power for  $T = 150$ , which arises as a result of the small over-size seen for these procedures, but otherwise the finite sample and large sample results are very similar. What is clear is that the union of rejections procedures offer robust power profiles across the full range of initial conditions, avoiding the low power that can arise from use of the GLS-based tests alone while retaining a good proportion of the additional power offered by the GLS-based tests over the OLS-based variant. Of the two,  $UR(t_{\hat{\phi}}, t_{\tilde{\phi}})$  emerges as the test with arguably the most attractive power properties overall, and on the basis of both the asymptotic and finite sample results, it is this procedure that we recommend for practical applications.<sup>6</sup>

In practice, when implementing  $UR(t_{\hat{\phi}}, t_{\tilde{\phi}})$  we would want to allow for additional serial correlation in  $u_{y,t}$  and  $u_{x,t}$ . In order to admit more general serial correlation into our DGP, we consider the following simple autoregressive-based extension to (2):

$$\begin{bmatrix} a(L)(u_{y,t} - \rho u_{y,t-1}) \\ b(L)u_{x,t} \end{bmatrix} = \begin{bmatrix} v_t \\ e_t \end{bmatrix}$$

with

$$\begin{aligned} a(L) &= 1 - a_1L - \dots - a_pL^p, \\ b(L) &= 1 - b_1L - \dots - b_qL^q \end{aligned}$$

where the roots of  $a(L)$  and  $b(L)$  all lie outside the unit circle, and where  $\varepsilon_t = [v_t \ e_t]'$  continues to satisfy Assumption 1. We also modify Assumption 2 so that, when  $c < 0$ ,  $u_{y,1} = \alpha \sqrt{\omega^2 / (1 - \rho^2)}$ , where  $\omega^2$  denotes the long run variance of  $a(L)^{-1}v_t$ , i.e.  $\omega^2 = \sigma_v^2 / a(1)^2$ .

In this setting, consider  $t_{\hat{\phi}}$  and  $t_{\tilde{\phi}}$  statistics, computed as in section 2 but on replacing (3) and (4) with the fitted regressions

$$\Delta \hat{u}_{y,t} = \hat{\phi} \hat{u}_{y,t-1} + \sum_{j=1}^p \hat{\pi}_j \Delta \hat{u}_{y,t-j} + \sum_{j=0}^q \hat{\delta}_j \hat{u}_{x,t-j} + \hat{\eta}_t, \quad (7)$$

$$\Delta \tilde{u}_{y,t} = \tilde{\phi} \tilde{u}_{y,t-1} + \sum_{j=1}^p \tilde{\pi}_j \Delta \tilde{u}_{y,t-j} + \sum_{j=0}^q \tilde{\delta}_j \tilde{u}_{x,t-j} + \tilde{\eta}_t. \quad (8)$$

It can then be shown that the large sample results for  $t_{\hat{\phi}}$ ,  $t_{\tilde{\phi}}$  and  $UR(t_{\hat{\phi}}, t_{\tilde{\phi}})$  from sections 3 and 4 continue to hold.<sup>7</sup> Moreover,  $R^2$  is now consistently estimated using the form

<sup>6</sup>Results for Models B and C are reported in ALT, and similar comments apply, confirming our preference for  $UR(t_{\hat{\phi}}, t_{\tilde{\phi}})$ .

<sup>7</sup>Note that extension of the  $\hat{\Lambda}$  statistic to the case of additional serial correlation is more involved, requiring construction of estimates of long-run versions of  $\Omega$  and  $R^2$ .

of  $\hat{R}^2$  given in the previous section, but with  $\hat{e}_t$  and  $\hat{v}_t$  replaced with residuals from  $q$ 'th and  $p + 1$ 'th order autoregressions fitted to  $\hat{u}_{x,t}$  and  $\hat{u}_{y,t}$ , respectively. In practice, since  $p$  and  $q$  are unknown, they can be determined endogenously using typical lag order selection rules such as downward testing or application of an information criterion. Finally, note that the fitted regressions (7) and (8) also cover the case where  $[u_{y,t}, u_{x,t}]'$  follows a standard vector autoregression (VAR) process of order  $\min(p + 1, q)$ . In this case, (7) and (8) then model the single equation for  $u_{y,t}$  from the VAR, cf. Hansen (1995), and the contemporaneous regressor  $\hat{u}_{x,t}$  is not required in (7) and (8).

## 6 Conclusion

In this paper we have considered the power of covariate augmented unit root tests, based on OLS demeaning/detrending and GLS demeaning/detrending, in the presence of asymptotically non-negligible initial conditions. We have shown that while the GLS-based approaches display superior finite sample and local asymptotic power for zero and small initial conditions, the power of such procedures falls towards zero as the initial condition increases in magnitude. Since we cannot be sure that such large initial conditions will not arise, this limits the reliability of such GLS-based tests in practice. On the other hand, while the OLS-based variants lose power for small initial conditions relative to their GLS-based counterparts, this ranking is reversed for larger initial conditions as the power of the OLS-based tests increases with the initial condition magnitude. We have then proposed a union of rejections based procedure, which detects evidence in favour of the alternative hypothesis taken from both OLS- and GLS-based demeaned/detrended variants, and find that such a procedure works very well, retaining attractive power levels across zero, small and large initial condition magnitudes. Our findings mirror those found in the standard non-covariate augmented unit root testing environment, and our recommended procedure adds to the suite of available unit root testing procedures a covariate augmented approach that offers reliable power levels across the range of possible (unknown) initial conditions.

## References

- Aristidou, C., Harvey, D.I. and Leybourne, S.J. (2016). The impact of the initial condition on covariate augmented unit root tests. Granger Centre Discussion Paper No. 16/01, School of Economics, University of Nottingham.
- Dickey, D.A. and Fuller, W.A. (1979). Distribution of estimators for autoregressive

- time series with a unit root. *Journal of the American Statistical Association* 74, 427–431.
- Elliott, G. and Jansson, M. (2003). Testing for unit roots with stationary covariates. *Journal of Econometrics* 115, 75–89.
- Elliott, G., Rothenberg, T.J. and Stock, J.H. (1996). Efficient tests for an autoregressive unit root. *Econometrica* 64, 813–836.
- Hansen, B.E. (1995). Rethinking the univariate approach to unit root testing. *Econometric Theory* 11, 1148–1171.
- Harvey, D.I. and Leybourne, S.J. (2005). On testing for unit roots and the initial observation. *Econometrics Journal* 8, 97–111.
- Harvey, D.I., Leybourne, S.J. and Taylor, A.M.R. (2009). Unit root testing in practice: dealing with uncertainty over the trend and initial condition (with commentaries and rejoinder). *Econometric Theory* 25, 587–667.
- Müller, U.K. and Elliott, G. (2003). Tests for unit roots and the initial condition. *Econometrica* 71, 1269–1286.
- Westerlund, J. (2013). A computationally convenient unit root test with covariates, conditional heteroskedasticity and efficient detrending. *Journal of Time Series Analysis* 34, 477–495.

Table 1. Asymptotic  $\xi$ -level critical values of covariate augmented unit root tests

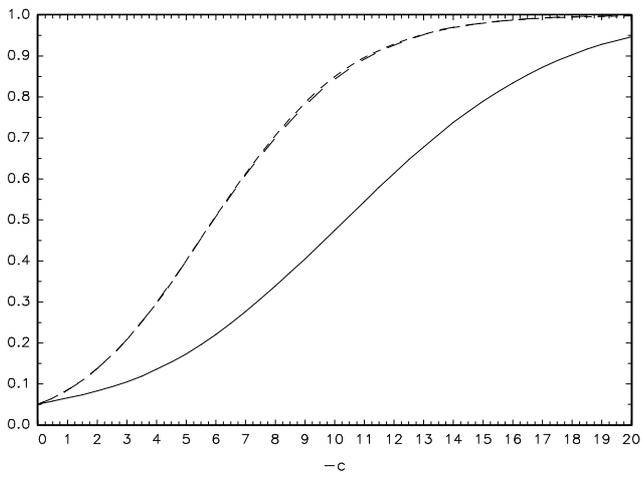
$R^2$	$t_{\hat{\phi}}$			$t_{\bar{\phi}}$			$\hat{\Lambda}$		
	$\xi = 0.10$	$\xi = 0.05$	$\xi = 0.01$	$\xi = 0.10$	$\xi = 0.05$	$\xi = 0.01$	$\xi = 0.10$	$\xi = 0.05$	$\xi = 0.01$
<i>Model A</i>									
0.0	-2.57	-2.86	-3.40	-1.61	-1.94	-2.60	4.60	3.30	1.92
0.1	-2.52	-2.82	-3.39	-1.57	-1.91	-2.57	4.80	3.36	1.67
0.2	-2.46	-2.77	-3.37	-1.52	-1.88	-2.53	5.08	3.44	1.42
0.3	-2.40	-2.72	-3.33	-1.47	-1.82	-2.51	5.45	3.60	1.22
0.4	-2.33	-2.65	-3.28	-1.41	-1.77	-2.46	5.95	3.85	1.06
0.5	-2.25	-2.58	-3.21	-1.34	-1.71	-2.41	6.64	4.28	0.98
0.6	-2.16	-2.50	-3.15	-1.27	-1.64	-2.35	7.72	4.99	1.08
0.7	-2.05	-2.40	-3.06	-1.18	-1.57	-2.29	9.61	6.25	1.52
0.8	-1.92	-2.27	-2.95	-1.07	-1.48	-2.21	13.37	8.99	2.90
0.9	-1.74	-2.10	-2.78	-0.95	-1.39	-2.16	24.82	17.64	7.97
<i>Model B</i>									
0.0	-3.13	-3.42	-3.98	-2.56	-2.85	-3.43	6.90	5.66	3.92
0.1	-3.05	-3.35	-3.90	-2.52	-2.81	-3.37	7.22	5.70	3.55
0.2	-2.98	-3.28	-3.83	-2.46	-2.77	-3.32	7.71	5.90	3.30
0.3	-2.89	-3.20	-3.76	-2.41	-2.71	-3.28	8.43	6.23	3.14
0.4	-2.79	-3.10	-3.69	-2.34	-2.65	-3.23	9.46	6.88	3.15
0.5	-2.68	-3.00	-3.59	-2.27	-2.58	-3.16	11.01	7.96	3.48
0.6	-2.54	-2.88	-3.49	-2.19	-2.50	-3.10	13.47	9.75	4.30
0.7	-2.39	-2.73	-3.36	-2.10	-2.43	-3.03	17.68	12.99	6.20
0.8	-2.20	-2.55	-3.19	-2.02	-2.34	-2.95	26.34	19.96	10.79
0.9	-1.94	-2.30	-2.97	-1.97	-2.31	-2.91	52.20	41.10	25.74
<i>Model C</i>									
0.0	-3.13	-3.42	-3.98	-2.56	-2.85	-3.43	6.90	5.66	3.92
0.1	-3.05	-3.35	-3.90	-2.50	-2.79	-3.35	7.24	5.71	3.64
0.2	-2.98	-3.28	-3.83	-2.43	-2.74	-3.30	7.71	5.90	3.38
0.3	-2.89	-3.20	-3.76	-2.35	-2.67	-3.24	8.36	6.27	3.29
0.4	-2.79	-3.10	-3.69	-2.27	-2.59	-3.18	9.30	6.86	3.34
0.5	-2.68	-3.00	-3.59	-2.17	-2.49	-3.10	10.70	7.82	3.65
0.6	-2.54	-2.88	-3.49	-2.06	-2.39	-3.01	12.90	9.47	4.39
0.7	-2.39	-2.73	-3.36	-1.93	-2.27	-2.91	16.75	12.52	6.15
0.8	-2.20	-2.55	-3.19	-1.78	-2.14	-2.81	24.79	18.97	10.61
0.9	-1.94	-2.30	-2.97	-1.60	-2.00	-2.70	49.76	39.30	25.13

Table 2. Asymptotic  $\psi_\xi$  values for  $\xi$ -level union of rejections procedures

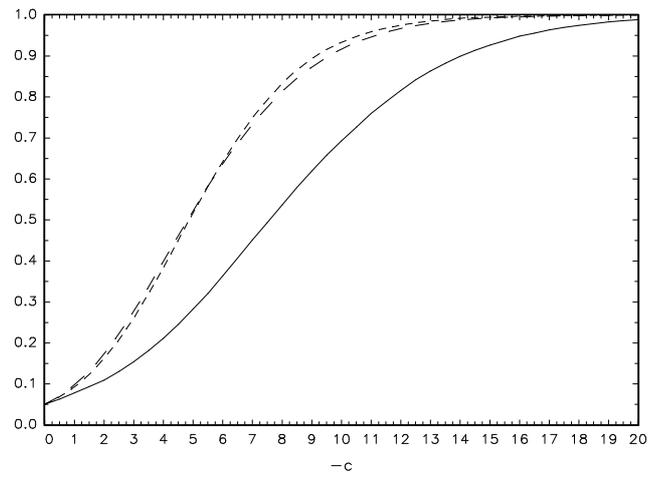
$R^2$	$UR(t_{\hat{\phi}}, t_{\hat{\tau}})$			$UR(t_{\hat{\phi}}, \hat{\Lambda})$		
	$\xi = 0.10$	$\xi = 0.05$	$\xi = 0.01$	$\xi = 0.10$	$\xi = 0.05$	$\xi = 0.01$
<i>Model A</i>						
0.0	1.099	1.081	1.062	1.200	1.132	1.071
0.1	1.105	1.083	1.057	1.212	1.143	1.080
0.2	1.108	1.086	1.056	1.226	1.157	1.090
0.3	1.111	1.091	1.057	1.240	1.168	1.099
0.4	1.113	1.093	1.064	1.261	1.184	1.106
0.5	1.119	1.097	1.068	1.276	1.205	1.117
0.6	1.126	1.102	1.070	1.297	1.223	1.135
0.7	1.135	1.106	1.073	1.336	1.236	1.149
0.8	1.152	1.115	1.076	1.389	1.271	1.172
0.9	1.179	1.136	1.082	1.502	1.352	1.207
<i>Model B</i>						
0.0	1.063	1.053	1.039	1.134	1.107	1.058
0.1	1.063	1.055	1.042	1.148	1.115	1.064
0.2	1.065	1.055	1.041	1.162	1.128	1.082
0.3	1.065	1.055	1.039	1.179	1.130	1.091
0.4	1.066	1.057	1.040	1.197	1.149	1.098
0.5	1.066	1.058	1.047	1.220	1.160	1.107
0.6	1.068	1.061	1.045	1.248	1.178	1.111
0.7	1.070	1.060	1.045	1.285	1.204	1.133
0.8	1.075	1.060	1.044	1.364	1.261	1.152
0.9	1.088	1.067	1.047	1.499	1.362	1.205
<i>Model C</i>						
0.0	1.063	1.053	1.039	1.134	1.107	1.058
0.1	1.065	1.055	1.042	1.147	1.113	1.067
0.2	1.067	1.056	1.040	1.157	1.119	1.074
0.3	1.069	1.056	1.042	1.169	1.130	1.085
0.4	1.072	1.061	1.041	1.183	1.135	1.095
0.5	1.075	1.063	1.048	1.197	1.144	1.102
0.6	1.080	1.065	1.047	1.222	1.155	1.107
0.7	1.085	1.071	1.048	1.251	1.186	1.121
0.8	1.095	1.075	1.049	1.318	1.235	1.147
0.9	1.126	1.093	1.061	1.491	1.353	1.202

Table 3. Finite sample size of nominal 0.05-level covariate augmented unit root tests:  $T = 150$

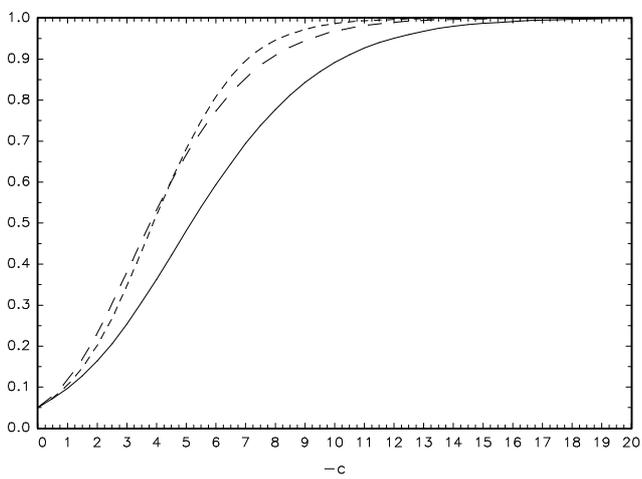
$R^2$	$t_{\hat{\phi}}$	$t_{\tilde{\phi}}$	$\hat{\Lambda}$	$UR(t_{\hat{\phi}}, t_{\tilde{\phi}})$	$UR(t_{\hat{\phi}}, \hat{\Lambda})$
<i>Model A</i>					
0.2	0.053	0.064	0.053	0.059	0.053
0.4	0.053	0.065	0.051	0.057	0.050
0.6	0.050	0.064	0.048	0.056	0.048
0.8	0.048	0.061	0.041	0.054	0.042
<i>Model B</i>					
0.2	0.054	0.064	0.049	0.059	0.050
0.4	0.053	0.063	0.047	0.057	0.049
0.6	0.052	0.060	0.043	0.055	0.044
0.8	0.048	0.053	0.029	0.049	0.032



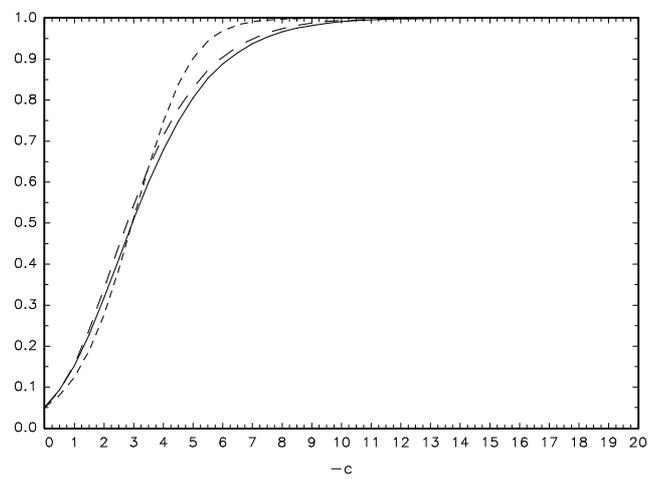
(a)  $R^2 = 0.2$



(b)  $R^2 = 0.4$

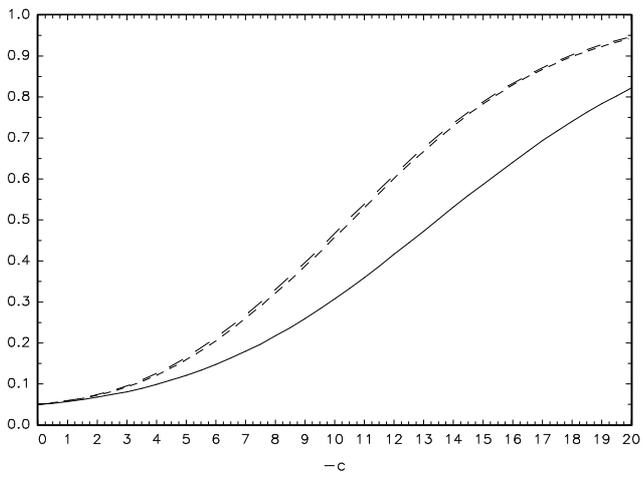


(c)  $R^2 = 0.6$

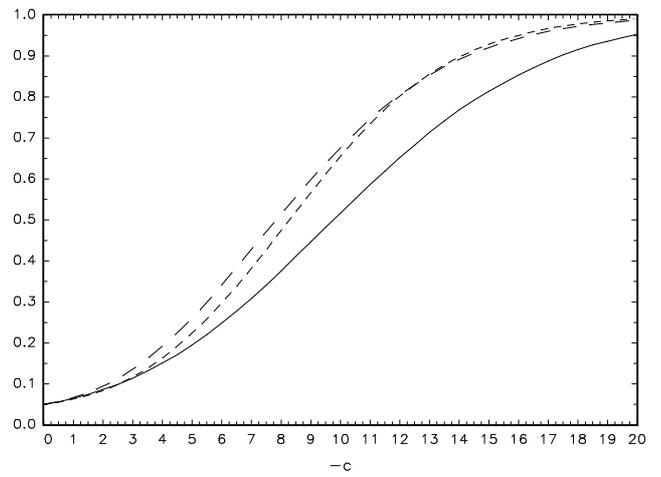


(d)  $R^2 = 0.8$

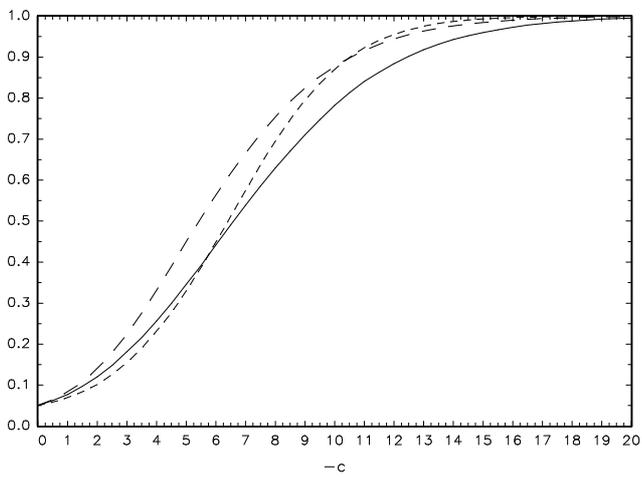
Figure 1. Local asymptotic power of nominal 0.05-level tests: Model A,  $\alpha = 0$ ;  
 $t_{\hat{\phi}}$ : —,  $t_{\tilde{\phi}}$ : --,  $\hat{\Lambda}$ : - - -



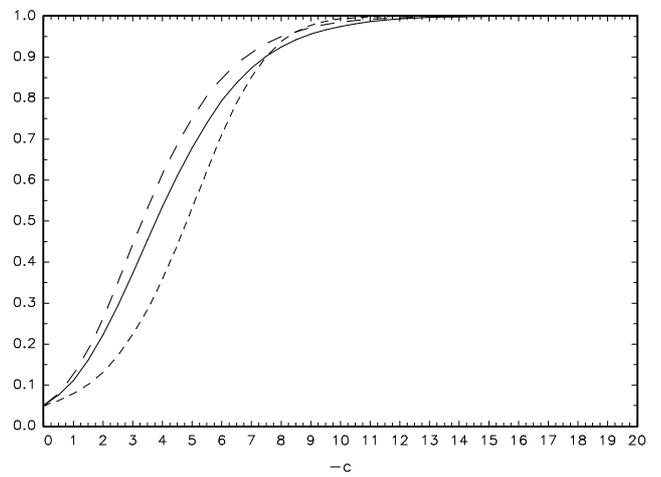
(a)  $R^2 = 0.2$



(b)  $R^2 = 0.4$

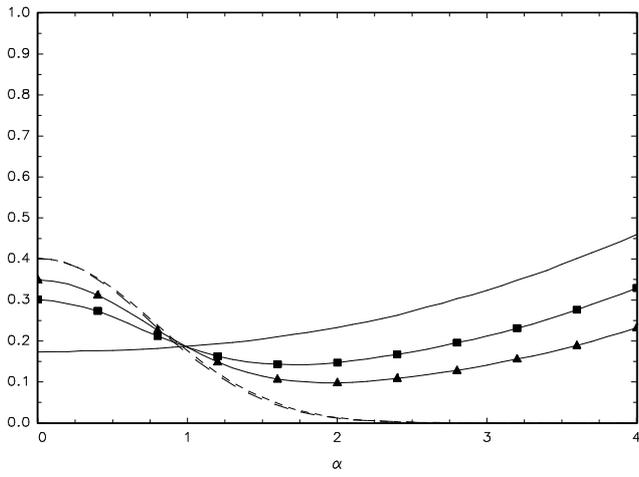


(c)  $R^2 = 0.6$

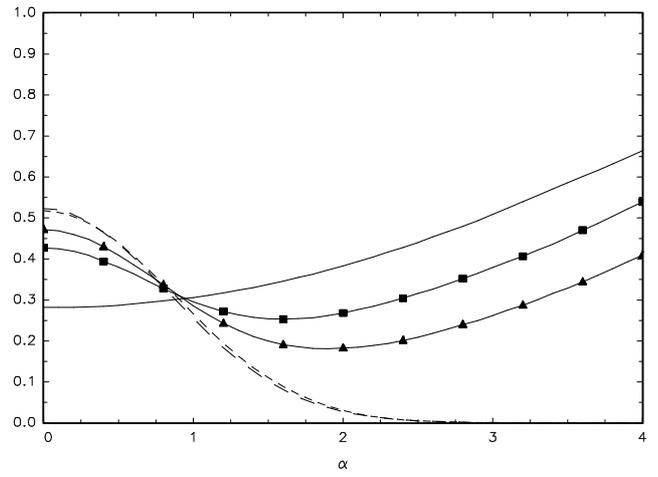


(d)  $R^2 = 0.8$

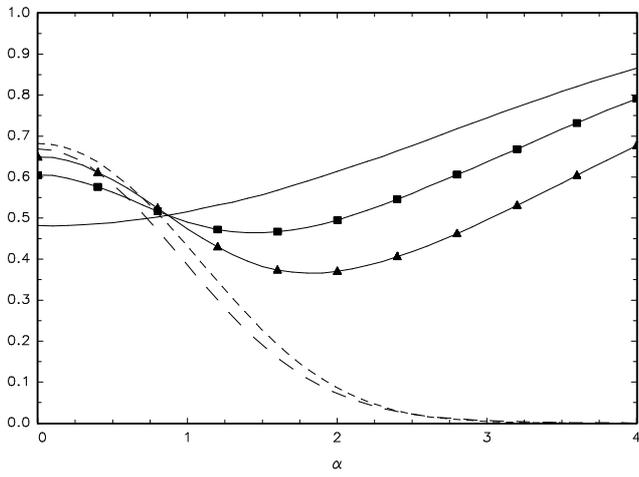
Figure 2. Local asymptotic power of nominal 0.05-level tests: Model B,  $\alpha = 0$ ;  
 $t_{\hat{\phi}}$ : —,  $t_{\tilde{\phi}}$ : --,  $\hat{\Lambda}$ : - - -



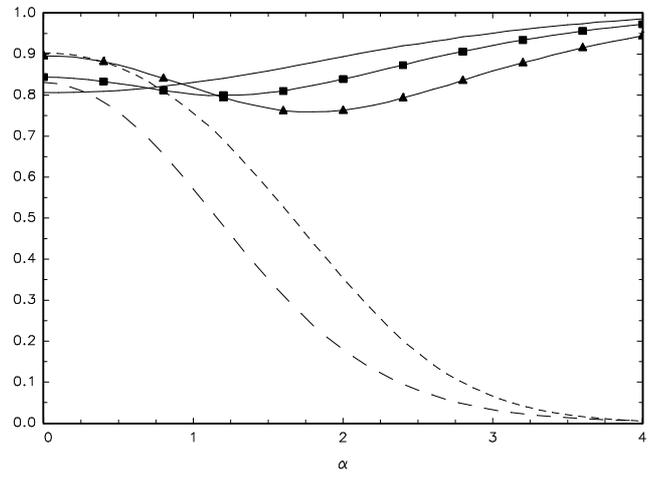
(a)  $R^2 = 0.2$



(b)  $R^2 = 0.4$

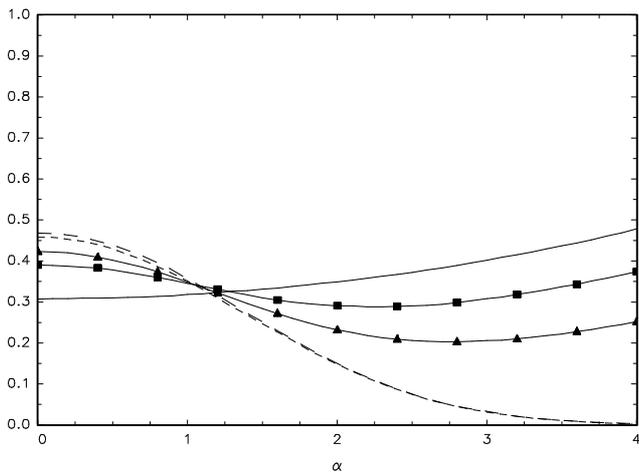


(c)  $R^2 = 0.6$

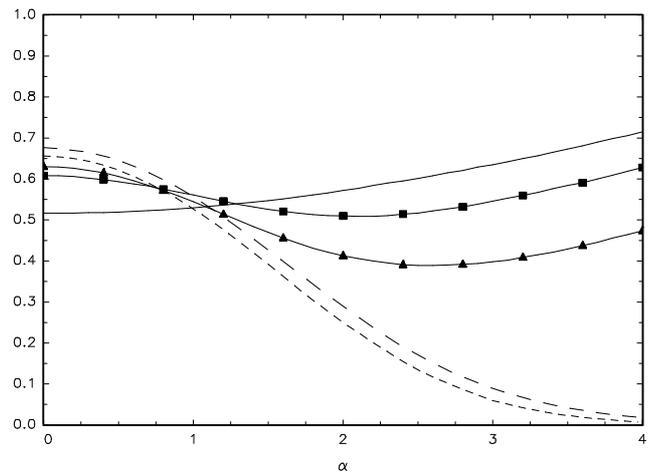


(d)  $R^2 = 0.8$

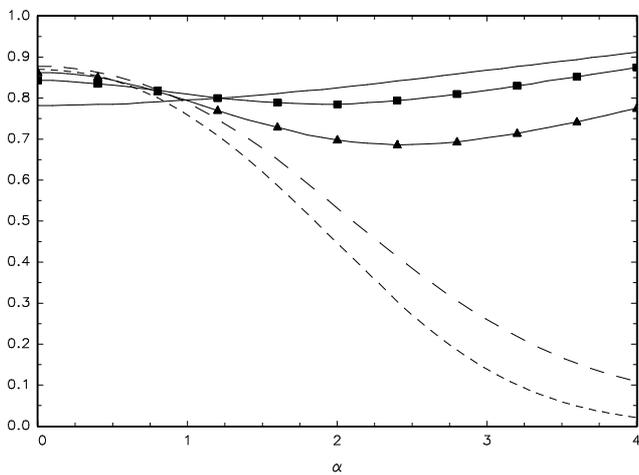
Figure 3. Local asymptotic power of nominal 0.05-level tests: Model A,  $c = -5$ ;  
 $t_{\hat{\phi}}$ : —,  $t_{\tilde{\phi}}$ : --,  $\hat{\Lambda}$ : - - -,  $UR(t_{\hat{\phi}}, t_{\tilde{\phi}})$ : ■,  $UR(t_{\hat{\phi}}, \hat{\Lambda})$ : ▲



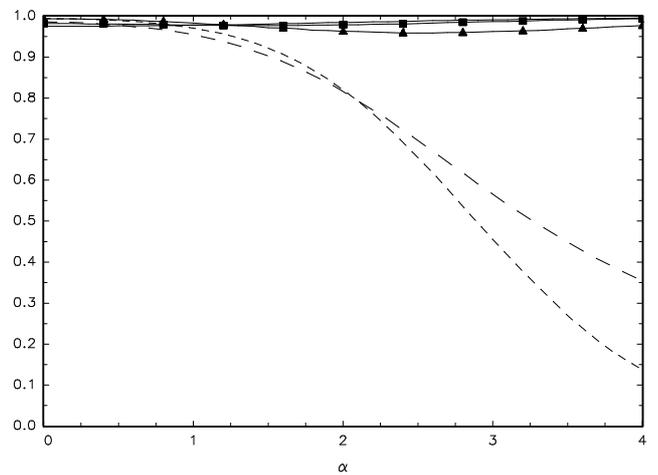
(a)  $R^2 = 0.2$



(b)  $R^2 = 0.4$

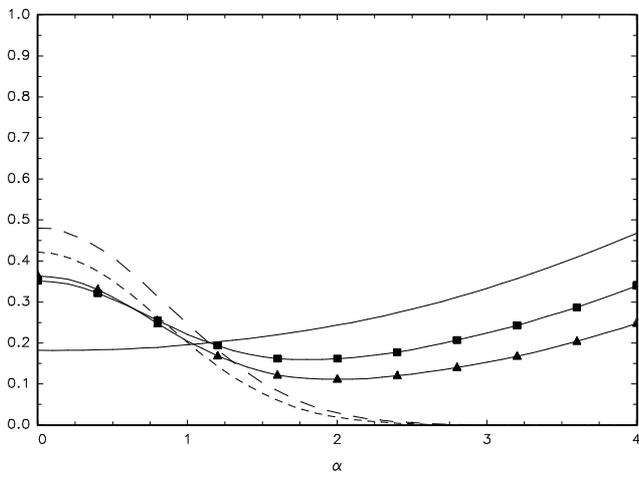


(c)  $R^2 = 0.6$

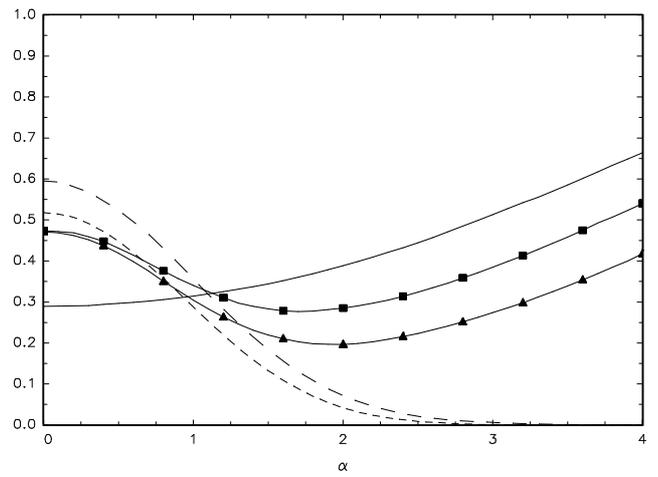


(d)  $R^2 = 0.8$

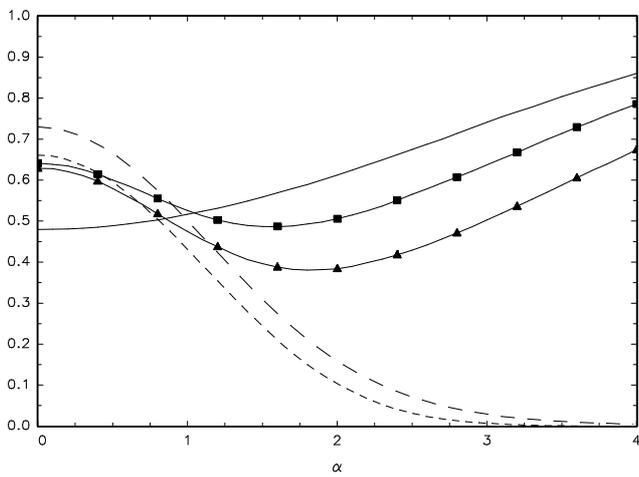
Figure 4. Local asymptotic power of nominal 0.05-level tests: Model B,  $c = -10$ ;  
 $t_{\hat{\phi}}$ : —,  $t_{\tilde{\phi}}$ : --,  $\hat{\Lambda}$ : - - -,  $UR(t_{\hat{\phi}}, t_{\tilde{\phi}})$ : ■,  $UR(t_{\hat{\phi}}, \hat{\Lambda})$ : ▲



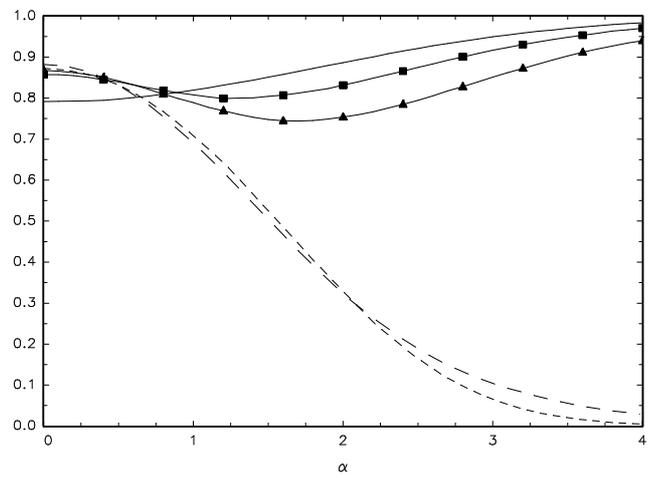
(a)  $R^2 = 0.2$



(b)  $R^2 = 0.4$



(c)  $R^2 = 0.6$



(d)  $R^2 = 0.8$

Figure 5. Finite sample power of nominal 0.05-level tests: Model A,  $T = 150$ ,  $c = -5$ ;  
 $t_{\hat{\phi}}$ : —,  $t_{\tilde{\phi}}$ : --,  $\hat{\Lambda}$ : - - -,  $UR(t_{\hat{\phi}}, t_{\tilde{\phi}})$ : ■,  $UR(t_{\hat{\phi}}, \hat{\Lambda})$ : ▲