12-1-2014

Restricted Likelihood Ratio Tests in Predictive Regression

Peter C.B. Phillips

Ye Chen

RESTRICTED LIKELIHOOD RATIO TESTS IN PREDICTIVE REGRESSION

By

Peter C. B. Phillips and Ye Chen

December 2014

COWLES FOUNDATION DISCUSSION PAPER NO. 1968

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

http://cowles.econ.yale.edu/

Restricted Likelihood Ratio Tests in Predictive Regression*

Peter C. B. Phillips[†] and Ye Chen[††]
[†]Yale University, University of Auckland,
Singapore Management University & University of Southampton
[††]Singapore Management University

September 19, 2014

Abstract

Chen and Deo (2009a) proposed procedures based on restricted maximum likelihood (REML) for estimation and inference in the context of predictive regression. Their method achieves bias reduction in both estimation and inference which assists in overcoming size distortion in predictive hypothesis testing. This paper provides extensions of the REML approach to more general cases which allow for drift in the predictive regressor and multiple regressors. It is shown that without modification the REML approach is seriously oversized and can have unit rejection probability in the limit under the null when the drift in the regressor is dominant. A limit theory for the modified REML test is given under a localized drift specification that accommodates predictors with varying degrees of persistence. The extension is useful in empirical work where predictors typically involve stochastic trends with drift and where there are multiple regressors. Simulations show that with these modifications, the good performance of the restricted likelihood ratio test (RLRT) is preserved and that RLRT outperforms other predictive tests in terms of size and power even when there is no drift in the regressor.

Keywords: Localized drift, Predictive regression, Restricted likelihood ratio test, Size distortion.

JEL classification: C12, C13, C58

*Peter C. B. Phillips, Cowles Foundation, Yale University, University of Auckland, University of Southampton & Singapore Management University, New Haven, CT,USA Email: peter.phillips@yale.edu. Ye Chen, School of Economics, Singapore Management University, 90 Stamford Road, Singapore, 178903; Email: chen.ye.2009@phdecons.smu.edu.sg. Phillips acknowledges support from the NSF under Grant No. SES 12-58258.

It seemed a plausible assumption that if we could demonstrate the existence in individuals or organizations of the ability to foretell the elusive fluctuations, either of particular stocks, or of stocks in general, this might lead to the identification of economic theories or statistical practices whose soundness had been established by successful prediction. (Alfred Cowles 3rd,1933)

There is no way to predict the price of stocks and bonds over the next few days or weeks. But it is quite possible to foresee the broad course of these prices over longer periods, such as the next three to five years. (Royal Swedish Academy of Sciences, Press Release on The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel, 14 October, 2013)

1 Introduction

Stock return predictability is a significant area of empirical research in both economics and finance with a vast literature stretching back at least to the early work of Cowles (1933), who took pioneering steps in the statistical modeling of stock prices. Since Fama's 1970 seminal paper, many works have studied the martingale hypothesis of stock returns to assess whether future returns can be predicted by past returns. Systematic accumulated evidence over the intervening years has shown that at least in the short run stock returns are difficult to predict. But since 1980 there has been considerable research on the question of whether financial asset returns may be predictable over the long term.

Among the many empirical methods now available, a return forecast regression framework has been particularly popular. An important contribution to this literature was made by Campbell and Shiller (1988) in developing a log linear return approximation to stock returns based on the following relationship:

$$p_t - d_t \approx -r_{t+1} + \Delta d_{t+1} + k + \rho \left(p_{t+1} - d_{t+1} \right), \tag{1.1}$$

where p_t is the log stock price at the end of period t, d_t is the log dividend paid during the period t, r_{t+1} is the log stock return, and the two parameters (k, ρ) are determined by sample averages relating to the dividend-price ratio. The linear approximation has wide validity and has been shown to be valid even in the presence of bubble effects under certain conditions (Phillips and Lee, 2013). Engsted, Pedersen and Tanggaard (2012) provide simulation results to support the use of the linear approximation when the log dividend-price ratio is subject to bubbles.

The approximate linear relationship (1.1) between stock returns and dividends has encouraged extensive use of linear econometric methods in forecasting stock returns. For example, based on (1.1), Cochrane (2008) considered the following linear VAR representation

$$\begin{aligned} r_{t+1} &= a_r + b_r (d_t - p_t) + \epsilon_{t+1}^r, \\ \Delta d_{t+1} &= a_d + b_d (d_t - p_t) + \epsilon_{t+1}^d, \\ d_{t+1} - p_{t+1} &= a_{dp} + \phi (d_t - p_t) + \epsilon_{t+1}^{dp}. \end{aligned} \tag{1.2}$$

Notably in (1.2), each equation has an intercept, which is motivated in part by the existence of the constant term in (1.1). Of course, models such as (1.2) without an intercept (or the implied drift in the case where d_t follows a martingale) are nested as a special case.

The model (1.1) makes the second equation of (1.2) redundant, leading to the following predictive regression framework (1.3) for testing stock return (y_t) predictability by means of a predictor (x_{t-1}) that typically includes variables such as dividends, book-to-market ratios and earnings-price ratios

$$\begin{aligned} y_t &= \pi + \beta x_{t-1} + u_{0t}, & (1.3) \\ x_t &= \mu + \rho_n x_{t-1} + u_{xt}, \end{aligned}$$

where the disturbances (u_{0t}, u_{xt}) are *iid* with zero means, covariance matrix $\begin{pmatrix} \sigma_{00}^2 & \sigma_{0x} \\ \sigma_{0x} & \sigma_{xx}^2 \end{pmatrix}$ and contemporaneous correlation $\lambda = \frac{\sigma_{0x}}{\sigma_{00}\sigma_{xx}}$. Let $u_{0t} = \phi u_{xt} + u_{0.xt}$ with $\phi = \mathbb{E}(u_{0t}u_{xt}) / \mathbb{E}(u_{xt}^2) = \lambda \frac{\sigma_{00}}{\sigma_{xx}}$ and suppose the initialization satisfies $x_0 \sim_d (0, \sigma_{xx}^2)$. In empirical work the correlation λ is usually negative and close to -1, capturing the negative relationship between stock returns and predictors such as dividend-price ratios. For our purposes in the present paper, we use this predictive regression model as the true generating mechanism with sample data $\{y_t, x_t\}_{t=1}^n$. Gaussianity is frequently assumed in setting up the likelihood function but is not required for the limit theory.

Two econometric issues with predictive regression have been a recent focus of attention, one being the bias of the OLS estimator of the regression coefficient β and the other being oversizing in the conventional t-test of the null of no predictability ($\beta = 0$). Both these problems arise from the presence of a (typically) highly persistent predictor x_t and (typically) strong error correlation between u_{xt} and u_{0t}. The estimation bias of $\widehat{\beta}$ is $-\phi \mathbb{E}(\widehat{\rho}_n - \rho_n)$ (Stambaugh, 1999) and this bias transmits forecast bias in predicting stock returns. For a predictor with a root that is the local to unity, the limit distribution of the t-test statistic is right skewed (Phillips, 2014). Hence, the null hypothesis of no predictability is often rejected when one uses critical values from a symmetric t-distribution even when the null hypothesis is correct. This explains oversizing in the standard t-test. The likelihood ratio test (LRT) also suffers from oversizing, because the LRT statistic is approximately the squared t-test under some regularity conditions (Chen and Deo, 2009a).

Chen and Deo (2009a) examined these econometric problems and proposed restricted maximum likelihood (REML) and restricted likelihood ratio test (RLRT) procedures to resolve them. The approach has the following desirable properties. First, REML produces a bias-reduced estimate without loss of efficiency compared to MLE. REML is able to produce a less biased estimator of β because the restricted likelihood (RL) is the likelihood of first differenced y_t and x_t. Through differencing the adverse impact (on estimation bias) of the intercept in the regression is reduced. This type of reduction occurs in the stationary AR(1) case $x_t = \mu + \rho_n x_{t-1} + u_{xt}$, where the bias of the OLS estimator of ρ_n is $\frac{-(1+3\rho_n)}{n}$ when an intercept is fitted but only $\frac{-2\rho_n}{n}$ when there is no intercept in the regression. Second, RLRT leads to improved inference compared to a standard t-test, because RL possesses smaller curvature (Chen and Deo, 2009a), thereby reducing the value of the statistic and reducing size distortion in predictive regression. Also, the RLRT statistic is asymptotically χ^2 for stationary predictors and the approximation error is small, as shown in Chen and Deo (2009a). Chen, Deo and Yi (2013) give the limit distribution for nearly nonstationary predictors by using a weighted least squares approximation to the restricted likelihood (WLSRL, hereafter), following Chen and Deo (2010). The corresponding test statistic is called QRLRT. A sup bound critical value for QRLRT is suggested for implementation, since the QRLRT statistic is nonpivotal.

While the restricted likelihood procedure has many advantages, its empirical size turns out to be close to unity, as we show here, when the true DGP of x_t has drift in the predictive regression (1.3). This size distortion occurs for both the RLRT of Chen and Deo (2009a) and the QRLRT of Chen et al. (2013). The reason for the distortion is that the associated RLRT has a sample size and drift dependent distribution in the presence of time series drift. This dependence applies even when the drift is small or local to zero in the sense of Phillips, Shi and Yu (2014). The associated critical value of the RLRT is then an increasing function of the drift, which produces size distortion in the use of standard χ^2 critical values. Since many financial time series and potential predictors manifest drift, use of the restricted likelihood approach in predictive regression seems likely to involve drift-induced distortions in testing.

The goal of this paper is to extend the Chen and Deo (2009a) method to allow for the presence of drift in the predictive regression (1.3) and develop a new approach to implementing the RLRT procedure. The limit distribution of the new RLRT is provided and test critical values are shown to be no longer sample size or drift dependent. The limit distribution of the new RLRT is not pivotal for all forms of localized drift in the generating mechanism but simulation results demonstrate robustness in the procedure and a sup bound critical value is recommended for implementation. In these simulations, we find that when the true DGP has drift our RLRT procedure has stable size and good power compared with the methoesd of Campbell and Yogo (2006) and Jansson and Moreira (2006) and improvements also hold when there is no drift in the time series.

Our approach to the RLRT is developed here primarily for the case of a scalar predictor and is shown to be easily applied when there are multiple predictors, as a generalization of Chen et al. (2013). Another advantage of our intercept-robust method is that it conveniently accommodates a range of assumptions about initial conditions in the derivation of limit distribution of the RLRT.

The paper is organized as follows. Section 2 extends Chen and Deo (2009a) to include time series drift and demonstrates the impact of drift on the limit theory. We then develop a drift-corrected RL function and provide the limit theory under the null of no predictability using this procedure. Section 3 presents the limit distribution of the new RLRT for the case of a multivariate predictor and shows that the asymptotics are equivalent to those of Chen et al. (2013). Section 4 reports Monte Carlo simulations for estimation and inference, and examines robustness under non-Gaussian errors. Section 5 concludes. Proofs of the main results are collected in the Appendix. Detailed derivations and additional technical material are provided in a supplementary document (Phillips and Chen, 2014) that is available online.

Throughout the paper, we adopt the following notations. The symbol "$\Rightarrow$" indicates weak convergence as $n \to \infty$, "$\to_p$" and "$\to^p$" stand for convergence in probability, "$\to_L$" denotes convergence in L_2 norm, "$\sim$" represents asymptotically equivalence, "$:=$" is definitional equality, CD signifies either estimation using REML or inference using RLRT as proposed by Chen and Deo (2009a), and PC signifies the corresponding counterparts of CD proposed in the present paper.

2 Restricted Likelihood Methods in Predictive Regression

This section explores the use of the RLRT to a framework that allows for drift in the true DGP. For the reasons already discussed, extension to this case is likely to be relevant in much empirical work. Following Phillips et al. (2014), it is convenient to use a localized drift mechanism in formulating the stochastic process x_t so that

$$x_t = \mu + \rho_n x_{t-1} + u_{xt}, \text{ with } \mu = \widetilde{\mu} n^{-\gamma} \text{ and } \gamma \in [0, \infty) \tag{2.1}$$

When $\gamma \in [0, \frac{1}{2}]$, the intercept has a non-negligible effect on the asymptotic behavior of x_t when ρ_n is local to unity. When $\gamma > \frac{1}{2}$, the intercept has negligible effects, so that the drift component of x_t (which is of order $O(t/n^\gamma)$) is dominated asymptotically by the stochastic trend, which is of order $O_p\left(t^{1/2}\right)$.

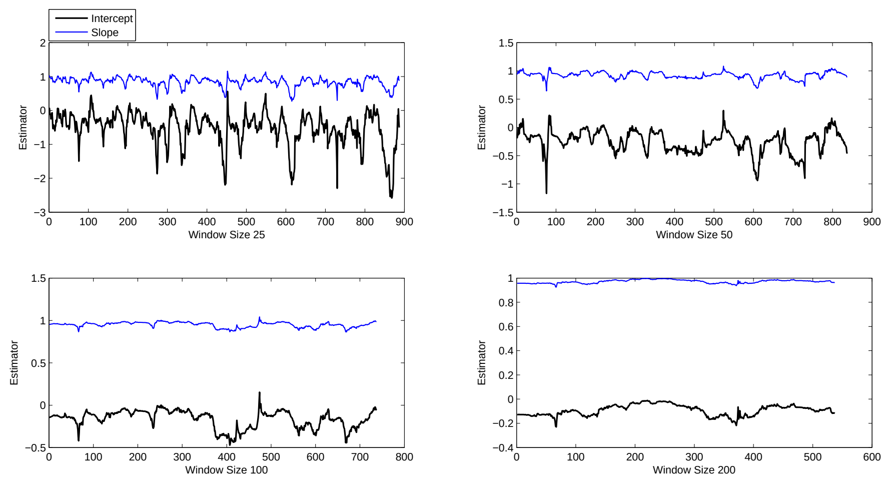

Figure 1: Estimates of an AR(1) model with fitted intercept using monthly D/P ratio of NYSE/AMEX data from Campbell and Yogo (2006)

Empirical studies demonstrate that the drift is sample size and frequency dependent for many financial time series (e.g. see Phillips et al., 2014). To illustrate, we fit the monthly dividend-price ratio of the NYSE/AMEX value-weighted index data (1871-2002) from Campbell and Yogo (2006) to an AR(1) model with intercept, under different rolling window sizes (25, 50, 100 and 250). Figure 1 shows that for these data the slope coefficient estimates fluctuate around unity, and the intercept estimates lie in the interval $(-2.58, 0.56)$. As the window sample size increases from $T = 25$ to $T = 200$, the magnitude of the drift tends on average to become smaller. Similar results hold for estimates from quarterly data.

The localized drift specification $\mu = \widetilde{\mu} n^{-\gamma}$ allows for such phenomena. In particular, smaller window sizes can be associated with larger values of the drift and indicate the possibility that values of $\gamma \in [0, \frac{1}{2}]$ may be relevant in some applications. Small sample sizes are particularly relevant in case of structural breaks, which may be empirically important over subperiods where financial time series exhibit exuberant, crisis, or post-crisis behavior, leaving less data to study predictability of returns.

Chen and Deo (2009a) proposed the use of an RLRT statistic obtained under the fitted model (2.2) given below, when the true DGP follows the predictive regression (1.3) with neither π nor μ intercept present. However, when the true DGP follows (1.3) with drift, simulations show that the rejection probability of the RLRT test is close to unity under the null hypothesis $\beta = 0$. The rejection rate of the QRLRT test under this null is also close to unity. To explain this phenomena, we show how the limit behavior of the RLRT test statistic is sample size dependent for γ under three categories of localized drift ($\gamma \in [0, \frac{1}{2})$, $\gamma = \frac{1}{2}$, $\gamma > \frac{1}{2}$), making a unit rejection rate inevitable for this test as $n \to \infty$. A new procedure to address oversizing is then provided combined with the limit distribution for the associated RLRT.

2.1 Oversizing in the RLRT

Chen et al. (2013) establish the limit distribution of RLRT under the following fitted regression estimation model (CD):

$$\begin{aligned} y_t &= \pi + \beta x_{t-1} + u_{0t}, \\ x_t &= \mu + \widetilde{x}_t, \\ \widetilde{x}_t &= \rho_n \widetilde{x}_{t-1} + u_{xt}. \end{aligned} \tag{2.2}$$

with $x_0 = 0$, while the true DGP follows (1.3) with zero intercepts π and μ. The limit behavior of the RLRT under these conditions is of order $O_p(1)$. However, the following theorem shows that, when the true DGP follows (1.3) with drift, the limit behavior of the RLRT using the CD procedure is actually sample size dependent, and the corresponding 95% critical value tends to infinity as the value of the drift and the sample size increase.

Let $L(\Theta)$ with $\Theta = \left(\beta, \rho_n, \phi, \sigma_{xx}^2, \sigma_{00.x}^2\right)$ be the log RL function for the predictive regression model with corresponding RLRT statistic defined as

$$R_n = -2L\left(\widehat{\Theta}_0\right) + 2L\left(\widehat{\Theta}\right)$$

where the REML estimator is $\widehat{\Theta}_0$ under the null and $\widehat{\Theta}$ under the alternative. Following Hayakawa (1977), under the null of $\beta = 0$, the asymptotic expansion of the RLRT statistic has the form

$$R_n = -h^{11} s_1^2 - 2h^{12} s_1 s_2 - \left(h^{22} - h_{22}^{-1}\right) s_2^2 + O_p\left(n^{-1/2}\right). \tag{2.3}$$

where s_i denotes the score function with respect to the i'th parameter in (β, ρ_n), and (h_{ij})

and $\left(h^{ij}\right)$ are elements of the Hessian matrix and inverse Hessian matrix, respectively.

Theorem 2.1 *In predictive regression (1.3) with $x_t = \mu + \rho_n x_{t-1} + u_{x,t}$, $\mu = \widetilde{\mu} n^{-\gamma}$ under $H_0 : \beta = 0$, the asymptotic behavior of the RLRT using the CD procedure is as follows.*

(1) If $\rho_n = 1$,

$$R_n \begin{cases} \Longrightarrow \left\{ \sqrt{1-\lambda^2 g_\lambda} Z + \lambda \left(g_\lambda\right)^{1/2} \tau \right\}^2, & \text{if } \gamma > \frac{1}{2} \\ = O_p(1), & \text{if } \gamma = \frac{1}{2} \\ = O_p\left(n^{1-2\gamma}\right). & \text{if } \gamma < \frac{1}{2} \end{cases}$$

(2) If $\rho_n = 1 + c/n$,

$$R_n \begin{cases} \Longrightarrow \left\{ \sqrt{1-\lambda^2 g_{c,\lambda,\widetilde{\mu}}^{1/2}} p_{c,\lambda,\widetilde{\mu}} + \lambda \left(g_{c,\lambda,\widetilde{\mu}}\right)^{1/2} \tau_{c,\lambda,\widetilde{\mu}} \right\}^2, & \text{if } \gamma > \frac{1}{2} \\ = O_p(1), & \text{if } \gamma = \frac{1}{2} \\ = O_p\left(n^{1-2\gamma}\right). & \text{if } \gamma < \frac{1}{2} \end{cases}$$

(3) If $\rho_n = 1 + \frac{c}{k_n}$ where $c < 0, k_n = n^\alpha$ with $\alpha \in (0,1)$,

$$R_n \begin{cases} \Longrightarrow \chi_1^2, & \text{if } \gamma \geqslant \frac{1}{2} \text{ and } 0 < \frac{\alpha}{2} < \gamma < \frac{1}{2} \\ = O_p\left(n^{1-2\gamma}\right). & \text{if } 0 < \gamma \leqslant \frac{\alpha}{2} < \frac{1}{2} \end{cases}$$

where $\lambda^2 = \frac{\phi^2 \sigma_{xx}^2}{\sigma_{00}^2}$ and Z is standard $N(0,1)$ independent of the (random) quantities g_λ, $g_{c,\lambda}$, $g_{c,\lambda,\widetilde{\mu}}$, $p_{c,\lambda,\widetilde{\mu}}$, τ, τ_c, and $\tau_{c,\lambda,\widetilde{\mu}}$, which are given in the Appendix in the proof of Theorem 2.1.

Remark 1 *Specifically, using the CD procedure leads to an expansion for the RLRT statistic R_n of the exlicit form*

$$R_n = \left\{ \sqrt{1-\lambda^2 g^n} p^n + \lambda \left(g^n\right)^{1/2} \tau^n \right\}^2 + O_p\left(n^{-1/2}\right), \tag{2.4}$$

with

$$\begin{aligned} p^n &= \frac{\left(\sum_{t=1}^n x_{t-1} u_{0.xt} - \frac{1}{n} \sum_{t=1}^n x_{t-1} \sum_{t=1}^n u_{0.xt}\right)}{\left\{\sum_{t=1}^n x_{t-1}^2 - \frac{1}{n}\left(\sum_{t=1}^n x_{t-1}\right)^2\right\}^{1/2}} \frac{1}{\sigma_{00.x}}, \\ g^n &= \frac{1 - \frac{1}{n}\left(\sum_{t=1}^n x_{t-1}\right)^2 / \sum_{t=1}^n x_{t-1}^2}{1 - \lambda^2 \frac{1}{n}\left(\sum_{t=1}^n x_{t-1}\right)^2 / \sum_{t=1}^n x_{t-1}^2}, \end{aligned}$$

$$\tau^n = \frac{\mu \sum_{t=1}^n x_{t-1} + \sum_{t=1}^n x_{t-1} u_{xt}}{\left(\sum_{t=1}^n x_{t-1}^2\right)^{1/2}} \frac{1}{\sigma_{xx}}.$$

The asymptotic behavior of the quantities p^n, g^n, and τ^n follow by using results in the lemmas given in Appendix A. The asymptotic behavior of the RLRT R_n then follows directly.

Remark 2 *For a predictor with a unit root ($\rho_n = 1$) and an asymptotically negligible drift satisfying $\gamma > \frac{1}{2}$, the associated limit distribution of the RLRT statistic is the same as when the true DGP has no drift. The inclusion of such a drift in the generating mechanism introduces additional terms, whose asymptotic order is smaller than $O_p\left(n^{-1/2}\right)$ and these smaller order terms may be retained in simulating critical values in order to improve inferential precision.*

Remark 3 *If $\gamma = \frac{1}{2}$, then $\mu = \frac{\tilde{\mu}}{\sqrt{n}}$ and $n^{-1/2} x_{t=\lfloor nr \rfloor} = \frac{\tilde{\mu}\lfloor nr \rfloor}{n} + \frac{\tilde{x}_{\lfloor nr \rfloor}}{\sqrt{n}} \sim \tilde{\mu} r + B_x(r)$, where B_x is Brownian motion with variance σ_{xx}^2. The limit behavior of τ_n in this case is easily seen to be*

$$\frac{\frac{\widetilde{\mu}^2}{2} + \widetilde{\mu} \int_0^1 B_x(r)\, dr + \int_0^1 B_x(r)\, dB_x(r)}{\sigma_{xx} \left\{ \frac{\widetilde{\mu}^2}{3} + 2\widetilde{\mu} \int_0^1 r B_x(r)\, dr + \int_0^1 B_x^2(r)\, dr \right\}^{1/2}}. \tag{2.5}$$

Importantly, note that when μ is fixed and not local to zero (2.5) then

$$\tau_n \sim \frac{\frac{\mu^2 n}{2} + \mu n^{1/2} B_x(1) + \int_0^1 B_x(r)\, dB_x(r)}{\sigma_{xx} \left\{ \frac{\mu^2 n}{3} + 2\mu n^{1/2} \int_0^1 r B_x(r)\, dr + \int_0^1 B_x^2(r)\, dr \right\}^{1/2}} = O_p\left(n^{1/2}\right),$$

since the numerator of τ_n has a higher asymptotic order than the denominator. Hence, for the fixed intercept case, the quantity τ_n in (2.4) gives rise to sample size dependent critical values of RLRT and the RLRT statistic diverges as $n \to \infty$ at rate $O_p(n)$. Similarly, if $\gamma < \frac{1}{2}$, the order of magnitude of the RLRT statistic is $O_p\left(n^{1-2\gamma}\right)$ since the asymptotic order of τ_n in this case is $O_p\left(n^{1/2-\gamma}\right)$. Precise results, including asymptotic representations of (p^n, g^n), are given in the Appendix. From this analysis over the three regions of γ, we find that the quantity τ_n in R_n drives the critical values of the RLRT statistic to infinity as $n \to \infty$ whenever the intercept dominates the asymptotics. Similar results can be found for predictors with roots local to unity or with roots moderate deviation from unity. These results are all given in the Appendix.

Remark 4 *The limit theory is provided under the assumption that $x_0 = 0$ to simplify derivations. Simulated critical values under $x_0 = 0$ are almost the same as those for $x_0 \sim_d N\left(0, \sigma_{xx}^2\right) = O_p(1)$. The limit theory for the RLRT statistic can be extended to deal with various other initial value conditions, such as distant past originations of the type considered* *in Phillips and Magdalinos (2009b). To perform these extensions some modifications of the RL function are required and for brevity these are not considered in the present paper.*

Remark 5 *The errors in (1.3) are assumed to be iid for the asymptotic development and are taken to be Gaussian for setting up the likelihood. The asymptotics continue to hold under weaker conditions such as those in Phillips and Magdalinos (2007a), where the errors are martingale differences satisfying the moment conditions* $\mathbb{E}\left(u_{xt}\right)^{2+\delta_1} < \infty$ *for some* $\delta_1 > 0$*, and* $\mathbb{E}\left(u_{0.xt}\right)^{2+\delta_2} < \infty$ *for some* $\delta_2 > 0$*. These moment conditions are used when the predictors are moderately integrated and martingale central limit theory is used for the asymptotic development. In that case, we have quasi restricted maximum likelihood estimation and likelihood ratio tests.*

2.2 Modifying the RLRT Procedure

We propose a new procedure to resolve the oversizing caused by the presence of drift. This procedure starts with the exact formulation of the predictive regression model (2.8), from which an associated restricted likelihood function is obtained and used to implement the corresponding RLRT. We first introduce the procedure leading to the restricted maximum likelihood (REML) estimator in the framework of the general linear model.

Patterson and Thompson (1971) introduced REML estimation into a mixed linear effect model of the form

$$\begin{aligned} Y_{n\times 1} &= X_{n\times p} b_{p\times 1} + \epsilon_{n\times 1}, \\ \epsilon &\sim N(0, H). \end{aligned} \tag{2.6}$$

REML estimation was proposed to estimate the parameters in the variance component while removing the impact of estimating the regression coefficients. Harville (1974) presented a convenient representation of the likelihood function with respect to parameters in the variance component as follows:

$$\begin{aligned} &(2\pi)^{-\frac{1}{2}(n-p)} \left\{\det\left(X'X\right)\right\}^{\frac{1}{2}} \left\{\det\left(H\right)\right\}^{-\frac{1}{2}} \left\{\det\left(X'H^{-1}X\right)\right\}^{-\frac{1}{2}} \\ &\exp\left\{-\frac{1}{2}\left(y - X\widehat{b}\right)' H^{-1}\left(y - X\widehat{b}\right)\right\}. \end{aligned} \tag{2.7}$$

with $\widehat{b} = (X'H^{-1}X)^{-1}X'H^{-1}y$. Verbyla (1990) provided an interpretation of REML in terms of marginal likelihood and Smyth and Verbyla (1996) further extended REML to generalized linear model with varying dispersion.

There is now a growing literature on the use and advantages of REML estimation in time series models which we briefly discuss. In particular, REML is known to produce substantial bias reductions in time series models without loss of efficiency, to improve forecast performance, and to provide a good base for inference. For example, Cooper and Thompson (1977) applied REML estimation in estimating autoregressive moving average models. Tunnicliffe Wilson (1989) used REML estimation in linear regression with ARMA process errors. Cheang and Reinsel (2000) gave an approximate representation of the REML estimator and a bias formula based on the OLS estimator of the linear regression model with stationary noise series that follows an AR(p) process without drift. In this case, the bias involved in REML estimation is shown to be much smaller than that of MLE for the estimate of the AR parameters. For vector autoregressions, Chen and Deo (2010) derived a weighted least squares estimator (WLSRL) as an approximation to the REML estimator and derived the bias formula and limit distribution of this WLSRL estimator. They show, for the VAR(p) model with drift, that the bias from REML estimation in this model equals that of OLS in the same model without drift up to $o\left(n^{-1}\right)$. The REML estimator also improves forecasts. Deo (2012) compares forecast performance in terms of absolute forecast error using both WLSRL and OLS. The simulation results reveal that for all forecast horizons the variance of the absolute forecast errors of OLS is much higher than that of REML. With regard to inference, Chen and Deo (2009b) prove that the RLRT distribution is well approximated as χ^2 in both the AR(1) process with unit root and near unit root coefficients. Chen and Deo (2011) show that RLRT produce confidence intervals with good coverage probabilities in an AR(p) process. In addition to this literature on estimation and forecasting, the restricted maximum likelihood is also used for model selection, where the approach leads to the residual information criterion (RIC) – see Shi and Tsai (2002), and Leng, Shi and Tsai (2008).

When the true DGP of x_t in (1.3) has a drift, we develop the predictive regression estimator by applying REML exactly to the following model (2.8):

$$\begin{aligned} y_t &= \pi + \beta x_{t-1} + u_{0t} \\ x_t &= \rho_n^t x_0 + \mu \sum_{i=0}^{t-1} \rho_n^i + \widetilde{x}_t, \\ \widetilde{x}_t &= \rho_n \widetilde{x}_{t-1} + u_{xt}. \end{aligned} \tag{2.8}$$

with $x_0 \sim_{iid} N\left(0, \sigma_{xx}^2\right)$ and $\widetilde{x}_0 = 0$. In comparison to the model of CD (2.2), observe that there is a key difference in the specification of the stochastic process x_t between the two models. The representation of x_t in (2.2) delivers an equivalent representation of x_t to that of

(1.3) only when t goes to infinity and $|\rho_n| < 1$. By contrast, in setting up REML using (2.8) our approach builds in an exact representation of the process x_t in (1.3) for all the possible value of ρ_n and t allowing for both intercepts (π, μ) in the system.

Following Chen and Deo (2009a), the RL function for the predictive regression (1.3) is factorized as the conditional RL function for Y_t with respect to X_t and the RL function for X_t. These RL functions relate to the models used in estimation. Hence, the representation of x_t is important in building the associated RL functions. The following Lemma presents the RL function for predictive regression (1.3) using our procedure.

Lemma 2.2 *The restricted maximum log-likelihood function for predictive regression (1.3) under the estimation model (2.8) is given by*

$$L(\Theta, Y, X) = -\frac{n-1}{2}\log\sigma_{00.x}^2 - \frac{1}{2\sigma_{00.x}^2}S(\beta, \rho_n, \phi) - \frac{n}{2}\log\sigma_{xx}^2 + \frac{1}{2}P(\rho_n) - \frac{1}{2\sigma_{xx}^2}Q(\rho_n),$$

where $Y = (y_1, y_2, \ldots y_n)$, $X = (x_0, x_1, \ldots x_n)$. The quantities $S(\beta, \rho_n, \phi), P(\rho_n)$ and $Q(\rho_n)$ are given in the Appendix in the proof of Lemma 2.2.

Remark 6 *Lemma 2.2 corresponds to Theorem 2 in Chen and Deo (2009a). The difference in the representation of x_t corresponds to the difference between CD's estimation model (2.2) and our model (2.8). In particular, the vector $\mathbf{1} = [1, \ldots 1]'$ in the RL function in Theorem 2 in Chen and Deo (2009a) is replaced by*

$$Z' = \begin{pmatrix} 1 & \rho_n^1 & \rho_n^2 & \cdots & \rho_n^n \\ 0 & 1 & 1+\rho_n^1 & \cdots & 1+\rho_n^1+\ldots+\rho_n^{n-1} \end{pmatrix},$$

when formulating the associated restricted likelihood for x_t. Simply put, this leads to a different regressor X in (2.7).

Remark 7 *The profile likelihood with respect to ρ_n is:*

$$(n-1)\log Q(\rho_n) - P(\rho_n). \tag{2.9}$$

Remark 8 *For AR(1) model with drift, i.e., $x_t = \mu + \rho_n x_{t-1} + u_{x,t}$, the associated RL function is*

$$L(\Theta, X) = -\frac{n}{2}\log\sigma_{xx}^2 + \frac{1}{2}P(\rho_n) - \frac{1}{2\sigma_{xx}^2}Q(\rho_n). \tag{2.10}$$

Chen and Deo (2009a) discuss the role of the intercept in REML and demean the series x_t

first as

$$x_t - \frac{\mu}{1-\rho_n} = \rho_n \left(x_{t-1} - \frac{\mu}{1-\rho_n} \right) + u_{xt},$$

giving a representation as a stationary AR(1) zero mean process $\widetilde{x}_{t-1}$ *plus* $\frac{\mu}{1-\rho_n}$. *This procedure requires the existence of the mean for the stochastic process* x_t *and thereby limits the use of REML to stationary processes. In contrast, our representation of* x_t *can deal with stationary processes with fixed mean and processes such as the unit root process, local to unity processes, and mildly integrated processes (Phillips and Magdalinos, 2007a&b). .*

Remark 9 *Our approach also works with vector autoregressions with intercept. The estimating procedure is similar to that for the univariate case. For specification of the variance matrix required in REML, readers may refer to the Appendix in Chen and Deo (2010).*

2.3 Limit Distribution of the RLRT

Using the RL function of Lemma (2.2), we obtain the associated limit theory for the RLRT test under the null hypothesis as follows.

Theorem 2.3 *In predictive regression model (1.3) with* $x_t = \widetilde{\mu} n^{-\gamma} + \rho_n x_{t-1} + u_{xt}$ *and under* $H_0 : \beta = 0$, *the limit behavior of the RLRT statistic is as follows:*

(1) If $\rho_n = 1$,

$$R_n \Longrightarrow \begin{cases} \left\{ \int_0^1 W_x^m(r)\, dW_0(r) \right\}^2 / \left\{ \int_0^1 [W_x^m(r)]^2\, dr \right\}, & \text{if } \gamma > \frac{1}{2} \\ \frac{\left\{ \int_0^1 W_x^m(r) dW_0(r) + \frac{\widetilde{\mu}}{\sigma_{xx}} \left\{ \frac{1}{2} W_0(1) - \int_0^1 W_0(r) dr \right\} \right\}^2}{\int_0^1 \{W_x^m(r)\}^2 dr + \frac{\widetilde{\mu}^2}{12\sigma_{xx}^2} + \frac{\widetilde{\mu}}{\sigma_{xx}} \left\{ 2\int_0^1 r W_x(r) dr - \int_0^1 W_x(r) dr \right\}}, & \text{if } \gamma = \frac{1}{2} \\ 3\left\{ W_0^m(1) - \int_0^1 W_0(r) dr \right\}^2 = \chi_1^2. & \text{if } \gamma < \frac{1}{2} \end{cases}.$$

(2) If $\rho_n = 1 + c/n$,

$$R_n \Longrightarrow \begin{cases} \left\{ \int_0^1 J_c^m(r)\, dW_0(r) \right\}^2 / \left\{ \int_0^1 [J_c^m(r)]^2\, dr \right\}, & \text{if } \gamma > \frac{1}{2} \\ \frac{\left\{ \frac{\widetilde{\mu}}{\sigma_{xx}} \int_0^1 F_c^m(r) dW_0(r) + \int_0^1 J_c^m(r) dW_0(r) \right\}^2}{\int_0^1 \left\{ \frac{\widetilde{\mu}}{\sigma_{xx}} F_c^m(r) dr + J_c^m(r) \right\}^2 dr}, & \text{if } \gamma = \frac{1}{2} \\ \left\{ \int_0^1 F_c^m(r)\, dW_0(r) \right\}^2 / \left\{ \int_0^1 \{F_c^m(r)\}^2\, dr \right\} = \chi_1^2. & \text{if } \gamma < \frac{1}{2} \end{cases}.$$

(3) If $\rho_n = 1 + \frac{c}{k_n}$ *with* $c < 0, k_n = n^\alpha$ *and* $\alpha \in (0,1)$,

$$R_n \Longrightarrow \chi_1^2.$$

(4) If $\rho_n \in (-1,1)$,

$$R_n \Longrightarrow \chi_1^2,$$

where $F_c^m(r)$ is defined in the proof of Theorem 2.3.

Remark 10 *Using our procedure, the limit distribution of RLRT is not sample size dependent, in contrast to Theorem 2.1 which shows the sample size dependence of the limit theory of RLRT under CD's estimation model. In particular, as discussed in Remark 2, the value of τ_n increases as either the value of the drift or the sample size increases. The corresponding finite sample quantity in our limit theory is q^n (given in (A.3) the Appendix in the proof of Theorem 2.3) which is $O_p(1)$ as $n \to \infty$ because its numerator and denominator are balanced in asymptotic order. It follows that the new RLRT procedure provides appropriate first differencing of the data to correct for bias and delivers a stable limit theory for RLRT over values of ρ_n and drift parameters.*

Remark 11 *In particular, uniform RLRT χ_1^2 inference holds when localized drift with $\gamma < \frac{1}{2}$ is present in predictive regression. Moreover, the limit theory for R_n with a local-to-unit root predictor is the same as $c \to 0$ to that for a unit root predictor and for all values of γ in the three categories considered. Further, as $c \to -\infty$, $R_n \Longrightarrow \chi_1^2$, corresponding to the case of both stationary and mildly integrated predictors.*

Remark 12 *The test statistic R_n is pivotal asymptotic χ_1^2 for all $\gamma < \frac{1}{2}$.*

Remark 13 *For $\gamma \geqslant \frac{1}{2}$, the test statistic R_n is nonpivotal asymptotically and its limit distribution is a function of c, λ, and $\widetilde{\mu}$. We examine the sensitivity of test critical values to different parameter configurations in both the CD and new procedures. Figure 2 demonstrates the sensitivity of 5% right tailed critical value to $\widetilde{\mu}$. We set $\sigma_{xx}^2 = \sigma_{00}^2 = 1$, $\phi = -0.95$ and allow $\widetilde{\mu} \in [-10, 10]$. The columns correspond to values of $\gamma \in \{0.6, 0.5, 0.3\}$. The rows correspond to unit root, local unit root and mildly integrated predictors. The graphs in Figure 2 reveal that the RLRT test critical values are highly sensitive to the drift value $\widetilde{\mu}$ in all cases for the CD procedure. The graphs show broad robustness of the critical values in the new procedure to drift, with some small sensitivity arising only in cases where $\mu \sim 0$.*

Remark 14 *Figure 3 shows the sensitivity of 5% right tailed critical values to ϕ for RLRT with columns corresponding to values of $\gamma \in \{0.6, 0.5, 0.3\}$. The rows show critical values corresponding to unit root (UR), local unit root (LUR), and mildly integrated (MI) processes. We set $\sigma_{xx}^2 = \sigma_{00}^2 = 1$, $\widetilde{\mu} = 8$ and allow $\phi \in [-0.9, -0.99]$. The findings reveal that the CD* *implementation of RLRT is quite sensitive to changes in ϕ whereas the 5% right tailed critical values are stable using the new procedure.*

Remark 15 *Figure 4 shows the sensitivity of 5% right tailed critical values to the signal to noise ratio $\frac{\sigma_{xx}^2}{\sigma_{00}^2}$ for RLRT using the CD and new (PC) procedures. We let $\sigma_{00}^2 = 1$, $\widetilde{\mu} = 8$, $\phi = -0.95$ and allow $\sigma_{xx}^2 \in [0.1, 2]$. The column panels correspond to different values of $\gamma \in \{0.6, 0.5, 0.3\}$. The row panels give critical values for predictors following UR, LUR, and MI processes. The results confirm that the CD implementation of the RLRT has critical values that are quite sensitive to changes in the signal to noise ratio in contrast to PC.*

Remark 16 *Table 1 reports* 5% *right tailed critical values of RLRT using the CD and PC procedures, all obtained by simulations with* 10,000 *replications and sample size* $n = 5000$. *The results in Table 1 show the sensitivity of the CD critical values to the parameter configuration, especially to the drift parameter $\tilde{\mu}$ for which the CD critical value is apparently an increasing function. In constrast, the PC implementation leads to critical values that are very stable with respect $\tilde{\mu}$, although there is some mild sensitivity to $\tilde{\mu}$ for the UR case when $\mu \sim 0$. These differences corroborate the asymptotic results in Theorems 2.1 and 2.3. The PC critical values appear to decrease as $\widetilde{\mu}$ increases in the UR case for $\gamma \geqslant 0.5$. Figure 5 presents both the density and the associated 5% right tailed critical values (given by the vertical lines in the graphs) of RLRT using the CD and PC implementations for $\widetilde{\mu} = 8$. The first and second panels give the RLRT densities for the CD and PC procedures for selected values of $\gamma \in \{0.6, 0.5, 0.3\}$. The graphs show the sample size and intercept dependence of the RLRT densities using the CD procedure against the stability of these densities to the drift parameter using the PC procedure.*

2.4 Implementation

To implement RLRT using our procedure we suggest a plug-in method to compute the associated critical value. In order to avoid estimating the localized coefficient (slope) c, following Chen et al. (2013), we suggest using a sup bound critical value with respect to ρ_n, which we define as

$$\overline{CV}_{\rho_n,\mu,\lambda} = \sup_{\rho_n \leqslant 1} \left\{ CV_{\rho_n,\mu,\lambda} : P\left(R_n > CV_{\rho_n,\mu,\lambda}\right) = \delta \right\}.$$

Table 1: Critical values of Alternative RLRT Statistics

			$\gamma = 0.6$			$\gamma = 0.5$			$\gamma = 0.3$		
	ρ_n	$\widetilde{\mu}$	μ	CD	PC	μ	CD	PC	μ	CD	PC
UR											
	1	1	0.006	4.344	7.902	0.014	4.847	7.611	0.078	31.553	3.926
	1	2	0.012	4.706	7.756	0.028	8.112	6.199	0.155	92.009	3.926
	1	4	0.024	6.833	6.636	0.057	19.389	4.592	0.311	307.724	3.926
	1	8	0.048	15.683	4.867	0.113	54.144	4.059	0.621	1118.362	3.926
	1	12	0.072	27.516	4.309	0.170	105.477	3.940	0.932	2434.610	3.926
	1	20	0.121	60.037	4.041	0.283	258.558	3.917	1.554	6583.936	3.926
LUR											
c = -200	0.96	1	0.006	3.981	3.860	0.014	3.967	3.857	0.078	6.057	3.925
	0.96	2	0.012	3.950	3.860	0.028	4.532	3.846	0.155	10.613	3.925
	0.96	4	0.024	4.285	3.850	0.057	9.868	3.839	0.311	23.752	3.925
	0.96	8	0.048	7.638	3.833	0.113	40.956	3.837	0.621	65.699	3.925
	0.96	12	0.072	15.688	3.844	0.170	98.232	3.846	0.932	128.535	3.925
	0.96	20	0.121	47.234	3.842	0.283	255.139	3.860	1.554	316.876	3.925
MIUR											
c = -200	0.906	1	0.006	3.809	3.892	0.014	3.865	3.892	0.078	12.576	3.892
$\alpha = 0.9$	0.906	2	0.012	3.847	3.892	0.028	4.136	3.892	0.155	61.388	3.892
	0.906	4	0.024	4.028	3.892	0.057	7.217	3.892	0.311	299.213	3.892
	0.906	8	0.048	5.873	3.892	0.113	28.641	3.892	0.621	942.352	3.892
	0.906	12	0.072	11.008	3.892	0.170	76.092	3.892	0.932	1426.162	3.892
	0.906	20	0.121	33.405	3.892	0.283	246.253	3.892	1.554	1899.606	3.892

In the predictive regression model (1.3) with $x_t = \widetilde{\mu} n^{-\gamma} + x_{t-1} + u_{xt}$, under the null $H_0 : \beta = 0$, the sup bound RLRT statistic then satisfies

$$\lim_{n\to\infty} \sup_{\rho_n \leqslant 1} \left\{ P\left(R_n > \overline{CV}_{\rho_n,\mu,\lambda} \right) \right\} \leqslant \delta.$$

The critical value $\overline{CV}_{\rho_n,\mu,\lambda}$ depends on the intercept parameter $\widetilde{\mu}$ and localizing exponent γ in the representation $\mu = \widetilde{\mu} n^{-\gamma}$. We propose setting $\gamma = \frac{1}{2}$ (for which the drift effect has the same order as the stochastic trend in the UR and LUR cases).[1] We then estimate the coefficient $\widetilde{\mu}$ from the fitted intercept. The associated sup bound critical value that we use for implementation is then defined as

$$\overline{CV}^{\gamma=\frac{1}{2}}_{\rho_n,\lambda} = \sup_{\rho_n \leqslant 1} \left\{ CV_{\rho_n,\lambda} : P\left(R_n > CV^{\gamma=\frac{1}{2}}_{\rho_n,\lambda} \right) = \delta \right\}. \tag{2.11}$$

[1] In fact, the parameter γ can be consistently estimated when $0 \leqslant \gamma \leqslant \frac{1}{2}$ and the estimator of γ converges in probability to $\frac{1}{2}$ for $\gamma > \frac{1}{2}$ for predictors of different persistence, as is shown in the Appendix following the argument given in Phillips et al. (2014). Simulations (not reported here) show that there are no effective gains in inference from such estimation of γ in comparison to the use of the sup bound critical value with setting $\gamma = \frac{1}{2}$ that is implemented here.

Simulations show that $\overline{CV}_{\rho_n,\lambda}^{\gamma=\frac{1}{2}}$ is an increasing function of ρ_n for given λ. This monotonicity is demonstrated in Figure 6.

The estimate of $\widetilde{\mu}$ carries the conditional normalization effect of the exponent setting $\gamma = \frac{1}{2}$ in the sup bound critical value (2.11). In the Monte Carlo study, we report the result using sup bound critical value $\overline{CV}_{\rho_n,\lambda}^{\gamma=\frac{1}{2}}$. We also report the result in the Appendix using exact critical values simulated under the true parameter configurations. As the results show, inference using critical values from the sup bound critical value are very close to those based on exact critical values computed under the true parameter configurations.

2.5 Scale-Invariant RLRT

The estimator $\widehat{\mu}$ is affected by scaling the data. However, inference results on predictability are scale-invariant. This is because the test statistic itself is scale-invariant. The scale-invariance property can be verified from the asymptotic expansion of the RLRT statistic in the new implementation. Specifically, if the data are multiplied by a common factor m, then the resulting the REML estimates $\widehat{\mu}$, $\widehat{\sigma}_{xx}^2$ and $\widehat{\sigma}_{00.x}^2$ are scaled by m, m^2 and m^2 respectively. Hence, the RLRT statistic and asymptotic critical values remain invariant and inferences are unaffected by the data scaling. In addition, the REML estimates $\widehat{\beta}$, $\widehat{\rho}_n$ and $\widehat{\phi}$ are all scale-invariant.

The following example illustrates the scale-invariance property of the RLRT statistic. Data are generated under the DGP (1.3) under the null for sample size $n = 50$ with a local to unit root predictor and with $\gamma = 0.6$, $\widetilde{\mu} = 10$, and $c = -5$ as in Table 2. We scale the data (X, Y) by 1/100. The results shown in Table 2 reveal the scale-invariance of the REML estimates $\widehat{\beta}$, $\widehat{\rho}_n$ and $\widehat{\phi}$, and the scale effects on the estimates $\widehat{\mu}$, $\widehat{\sigma}_{xx}^2$ and $\widehat{\sigma}_{00.x}^2$ of $10^{-2}, 10^{-4}$, and 10^{-4}, respectively. Table 3 shows that inferences are unaffected. The PC RLRT statistic is 0.122, which is much smaller than the 5% right tailed critical value (CV-PC$_{\gamma=\frac{1}{2}}$) of 3.899, suggesting the null cannot be rejected, which is consonant with the null $\beta = 0$ under which the data is generated. The CD RLRT statistic is 6.408, which exceeds the 5% right tailed critical value (CV-CD) of 4.177, thereby rejecting the null. This example corroborates the finding of Theorem 2.1, which suggests a tendency of the CD procedure towards rejection under the null when the true DGP has drift.

3 Multivariate Predictors

This section details the limit distribution of RLRT for multivariate predictors allowing for UR, LUR and MI predictor processors. This theory completes the asymptotic findings of

Table 2: Estimation results for alternative (CD and PC) REML estimators under $H_0 : \beta = 0$, when the predictor x_t has intercept and the data are scaled.

Parameters	CD	PC	CD-Scaled	PC-Scaled
μ	0.000	0.524	0.000	0.005
ρ	0.989	0.920	0.989	0.920
β	-0.087	-0.023	-0.087	-0.023
ϕ	-0.939	-0.939	-0.939	-0.939
$\sigma^2_{00.x}$	0.093	0.093	0.000	0.000
σ^2_{xx}	1.028	0.962	0.000	0.000

Table 3: Inference results under CD and PC implementations of RLRT for the null $H_0 : \beta = 0$, when the predictor x_t has an intercept, using sup bound critical values.

Data	RLRT-CD	RLRT-PC	CV-CD	CV-PC$_{\gamma=\frac{1}{2}}$	$\widehat{\widehat{\mu}}$
Unscaled	6.408	0.122	4.177	3.899	0.037
Scaled	6.408	0.122	4.177	3.899	0.037

Chen et al. (2013) for a wide class of predictors. To simplify presentation, we concentrate on the case of a bivariate predictor, which covers many cases of practical interest.

For the case of a univariate predictor in predictive regression without drift, as in predictive regression (1.3), Chen et al. (2013) provide the limit distribution of the QRLRT. The QRLRT is based on the weighted restricted likelihood function (WLSRL) function and has the advantage that it delivers the estimators of β and ρ_n in closed form. Chen and Deo (2010) showed that this estimator is asymptotically equivalent to the exact REML estimator.

As in Theorem 3.2 for a univariate predictor, the limit distribution of RLRT is derived from a direct Taylor expansion of R_n. Following Theorem(3.2, we start with the asymptotic expansion of R_n, whose representation depends on the specification of the estimation model. We first present asymptotics for the CD version of RLRT with bivariate predictors. The CD estimation model is as follows:

$$\begin{aligned} y_t &= \pi + \beta' \mathbf{x}_t + u_{0t}, \\ \mathbf{x}_t &= \mu + \widetilde{\mathbf{x}}_t, \\ \widetilde{\mathbf{x}}_t &= \boldsymbol{\rho}' \widetilde{\mathbf{x}}_{t-1} + \mathbf{u}_{x,t}, \\ u_{0t} &= \boldsymbol{\phi}' \mathbf{u}_{x,t} + u_{0.xt}, \end{aligned} \tag{3.1}$$

The true DGP in Chen and Deo (2009a) has the simpler form

$$\begin{aligned} y_t &= \beta' \mathbf{x}_t + u_{0t}, \\ \mathbf{x}_t &= \boldsymbol{\rho}' \mathbf{x}_{t-1} + \mathbf{u}_{x,t}, \\ u_{0t} &= \boldsymbol{\phi}' \mathbf{u}_{x,t} + u_{0.xt}. \end{aligned} \tag{3.2}$$

Based on Hayakawa (1977), under the null of $\beta_1 = \beta_2 = 0$, we have the following asymptotic expansion of the RLRT

$$R_n = S' Z S + O_p\left(n^{-1/2}\right)$$

where S is a 9×1 vector with s_i denoting the score function with respect to the i'th parameter in

$$\Theta = \left(\beta_1, \beta_2, \rho_{1n}, \rho_{2n}, \phi_1, \phi_2, \sigma_{x1x2}, \sigma^2_{xx1}, \sigma^2_{xx2}\right),$$

H indicates the Hessian matrix with elements (h_{ij}), and $H_0^{-1} = \begin{pmatrix} 0 & 0 \\ 0 & H_{11}^{-1} \end{pmatrix}$ is a 9×9 matrix with H_{11} containing the elements of second derivative with respect to the unrestricted parameters. It is noted that for each parameter in $\Theta^2 = \left(\phi_1, \phi_2, \sigma_{x1x2}, \sigma^2_{xx1}, \sigma^2_{xx2}\right)$, the associated normalized second partial derivatives such as $\frac{\partial RL}{\partial \theta_i^2 \partial \theta_j}$ are of the order $o_p(1)$, where $\theta_i^2 \neq \theta_j$, θ_i^2 indicates the i'th parameter in Θ^2, and θ_j is the j'th parameter in Θ . Hence, we retain only the score functions and Hessian elements with respect to the four parameters $(\beta_1, \beta_2, \rho_{1n}, \rho_{2n})$ involved in the asymptotic expansion of R_n.

Theorem 3.1 *In the context of the predictive regression (3.2), under $H_0 : \beta_1 = \beta_2 = 0$, the asymptotic distribution of the RLRT using the CD procedure is given by*

$$\left(A_S^k\right)' A_Z^k A_S^k \tag{3.3}$$

where $k = 1, 2 \ldots 6$ stands for the following cases: (1) $\rho_{1n} = 1$ and $\rho_{2n} = 1$; (2) $\rho_{1n} = 1 + \frac{c_1}{n}$ and $\rho_{2n} = 1 + \frac{c_2}{n}$; (3) $\rho_{1n} = 1 + \frac{c_1}{k_n}$ and $\rho_{2n} = 1 + \frac{c_2}{k_n}$ $(c_1, c_2 < 0, k_n = n^\alpha$ with $\alpha \in (0,1))$; (4) $\rho_{1n} = 1$ and $\rho_{2n} = 1 + \frac{c_2}{n}$; (5) $\rho_{1n} = 1$ and $\rho_{2n} = 1 + \frac{c_2}{k_n}$; and (6) $\rho_{1n} = 1 + \frac{c_1}{n}$ and $\rho_{2n} = 1 + \frac{c_2}{n}$. In (3.3) the notation $A^k_{(\cdot)}$ stands for the asymptotic distribution of the quantity in parentheses $(\cdot)$ under normalization, as given in the proof of Theorem 3.1 in the Appendix.

For completeness, we present the corresponding result for the univariate predictor case where the notation is the same as that in Theorem 2.1.

Theorem 3.2 *In the predictive regression (1.3) when the univariate predictor x_t has no drift, under $H_0 : \beta = 0$, the asymptotic distribution of the RLRT using the CD procedure is as follows:*

(1) If $\rho_n \in (-1, 1)$, $R_n \Longrightarrow \chi_1^2$;

(2) If $\rho_n = 1, R_n \Longrightarrow \left\{ \sqrt{1 - \lambda^2 g_\lambda} Z + \lambda (g_\lambda)^{1/2} \tau \right\}^2$;

(3) If $\rho_n = 1 + c/n, R_n \Longrightarrow \left\{ \sqrt{1 - \lambda^2 g_{c,\lambda}} Z + \lambda (g_{c,\lambda})^{1/2} \tau_c \right\}^2$;

(4) If $\rho_n = 1 + \frac{c}{k_n}$ with $c < 0, k_n = n^\alpha$ with $\alpha \in (0, 1)$, $R_n \Longrightarrow \chi_1^2$.

Remark 17 *This result matches that of Chen et al. (2013). In Theorem 2 of Chen et al. (2013), the QRLRT is based on the loss function $Q_n(\beta, \rho_n)$, which shares a similar estimating equation (or first order condition) with the RL function. The resulting WLSRL estimator is the exact REML estimator. Unlike the REML estimator, the WLSRL estimator has a closed form expression but retains its good finite sample properties. In deriving the limit distribution of QRLRT, which equals the difference between $Q_n(\beta, \rho_n)$ evaluated at the null and the alternative, Chen et.al (2013)expand the difference by a Taylor series at the WLSRL estimate, and obtain the limit distribution of QRLRT using its expanded representation. On the contrary, our REML estimator is based on the RL function. The RLRT statistic is equal to the difference between the log RL function evaluated at the null and that at the alternative. In deriving the limit distribution of the RLRT statistic, a closed form of the REML estimator is not required.*

We present the corresponding asymptotics of the PC version of the RLRT based on the following estimation model

$$\begin{aligned} y_t &= \boldsymbol{\pi} + \boldsymbol{\beta}' \mathbf{x}_t + u_{0t}, \\ \mathbf{x}_t &= \boldsymbol{\rho}_n^t \mathbf{x}_0 + \boldsymbol{\mu} \sum_{i=0}^{t-1} \boldsymbol{\rho}_n^i + \widetilde{\mathbf{x}}_t, \\ \widetilde{\mathbf{x}}_t &= \boldsymbol{\rho}' \widetilde{\mathbf{x}}_{t-1} + \mathbf{u}_{x,t}, \\ u_{0t} &= \boldsymbol{\phi}' \mathbf{u}_{x,t} + \mathbf{u}_{0.xt}, \end{aligned} \tag{3.4}$$

where in the bivariate predictor case $\boldsymbol{\beta} = [\beta_1, \beta_2]'$, $\boldsymbol{\rho} = \begin{pmatrix} \rho_{1n} & 0 \\ 0 & \rho_{1n} \end{pmatrix}$, $\Sigma = \begin{pmatrix} \sigma_{x_1x_1}^2 & \sigma_{x_1x_2} \\ \sigma_{x_1x_2} & \sigma_{x_2x_2}^2 \end{pmatrix}$, $\boldsymbol{\phi} = [\phi_1, \phi_2]'$ and $\boldsymbol{u}_{x,t} = [u_{x1,t}\ u_{x2,t}]$. It is further assumed that $[u_{0.xt}, \mathbf{u}_{x,t}] \sim_{iid} N\left(\mathbf{0}, diag\left(\sigma_{00.x}^2, \Sigma\right)\right)$ and $x_0 \sim N(\mathbf{0}, \Sigma)$. The diagonal structure of $\boldsymbol{\rho}$ can be achieved by transformation and is supported in some cases by empirical evidence, as in Amihud and Hurvich (2004). The true

DGP in Chen and Deo (2009a) has the following form

$$\begin{aligned} y_t &= \boldsymbol{\pi} + \boldsymbol{\beta}'\mathbf{x}_t + u_{0t}, \\ \mathbf{x}_t &= \boldsymbol{\mu} + \boldsymbol{\rho}'\mathbf{x}_{t-1} + \mathbf{u}_{x,t}, \\ u_{0t} &= \boldsymbol{\phi}'\mathbf{u}_{x,t} + u_{0.xt}. \end{aligned} \tag{3.5}$$

Lemma 3.3 *The restricted log-likelihood function for predictive regression (3.5) under the estimation model (3.4) is given by*

$$L(\Theta, Y, X) = -\frac{n-1}{2}\log\sigma_{00.x}^2 - \frac{1}{2\sigma_{00.x}^2}S(\boldsymbol{\beta}, \boldsymbol{\rho}, \boldsymbol{\phi}) + \frac{1}{2}P(\boldsymbol{\rho}, \Sigma) - \frac{1}{2\sigma_{xx}^2}Q(\boldsymbol{\rho}, \Sigma),$$

where $Y = (y_1, y_2, \ldots y_n)$*,* $X = (\mathbf{x}_0, \mathbf{x}_1, \ldots \mathbf{x}_n)'$*. The quantities* $S(\boldsymbol{\beta}, \boldsymbol{\rho}, \boldsymbol{\phi}), P(\boldsymbol{\rho}, \Sigma)$ *and* $Q(\boldsymbol{\rho}, \Sigma)$ *are given in the Appendix in the proof of Lemma 3.3.*

Remark 18 *Lemma 3.3 gives the restricted likelihood function using CD's procedure. To ease implementation of REML and RLRT we suggest, following Chen and Deo (2009a), we reparameterize the parameters set as* $\left(\Sigma, \sigma_{00.x}^2, \boldsymbol{\rho}, \boldsymbol{\phi}, \boldsymbol{\gamma}\right)$ *with* $\boldsymbol{\gamma} = \boldsymbol{\beta} - \boldsymbol{\rho}\boldsymbol{\phi}$*. The reparameterization allows estimating* $\boldsymbol{\rho}$ *and* Σ *first by concentrating other parameters out of the likelihood and then obtaining the remaining REML estimators by minimizing* $S(\boldsymbol{\phi}, \boldsymbol{\gamma})$*.*

Theorem 3.4 *For the predictive regression (3.5) under* $H_0 : \beta_1 = \beta_2 = 0$*, the asymptotic null distribution of the RLRT using the PC procedure is represented as*

$$\left(B_S^k\right)' B_Z^k B_S^k, \tag{3.6}$$

where $k = 1, 2 \ldots 6$ *stands for the following cases: (1)* $\rho_{1n} = 1$ *and* $\rho_{2n} = 1$*; (2)* $\rho_{1n} = 1 + \frac{c_1}{n}$ *and* $\rho_{2n} = 1 + \frac{c_2}{n}$*; (3)* $\rho_{1n} = 1 + \frac{c_1}{k_n}$ *and* $\rho_{2n} = 1 + \frac{c_2}{k_n}$ $(c_1, c_2 < 0, k_n = n^\alpha$ *with* $\alpha \in (0,1))$*; (4)* $\rho_{1n} = 1$ *and* $\rho_{2n} = 1 + \frac{c_2}{n}$*; (5)* $\rho_{1n} = 1$ *and* $\rho_{2n} = 1 + \frac{c_2}{k_n}$*; and (6)* $\rho_{1n} = 1 + \frac{c_1}{n}$ *and* $\rho_{2n} = 1 + \frac{c_2}{n}$*. In (3.6) the notation* $B_{(\cdot)}^k$ *stands for the asymptotic distribution of the quantity in parentheses* $(\cdot)$ *under normalization, as given in the proof of Theorem (3.4) in the Appendix.*

Remark 19 *The predictors are assumed to fall within three categories covering unit root, local unit root, and mildly integrated processes. The representation of the limit distribution of the RLRT is therefore given under six scenarios for the bivariate predictor case, which covers all possible combinations of the predictors with varying degrees of persistence.*

Remark 20 *The limit distribution in Theorem 3.4 is derived under the joint null* $\beta_1 = \beta_2 = 0$ *but may be extended to the marginal null hypotheses* $\beta_1 = 0$ *and* $\beta_2 = 0$*.*

Remark 21 *If both predictors x_{1t} and x_{2t} are mildly integrated processes, the associated limit distribution is derived by assuming $k_n = n^{\alpha}$ for both x_{1t} and x_{2t}. In this case, differences in the deviation from unity are reflected through the localizing coefficients c_1 and c_2. This assumption provides a substantial simplification in the proof but may be extended using recent results on multivariate mildly integrated processes with differing rates by Magdalinos and Phillips (2009).*

4 Simulations

This section examines the finite sample estimation and inferential performance of the new procedures. Section 4.1 compares estimation performance among REML-CD, REML-PC and MLE procedures when the true DGP follows an AR(1) model with drift. Section 4.2 extends this comparative study to an AR(2) model with drift and the predictive regression model with drift. Section 4.3 reports inferential performance on size and power between RLRT-CD and RLRT-PC procedures in predictive regression DGPs with and without intercept. These simulations are complemented by a further comparative study between the predictive regression methods of RLRT-PC, Campbell and Yogo (2006), Jansson and Moreira (2006), and Phillips and Magdalinos (2009a). The number of replications is 5000 throughout.

4.1 Autoregression with Drift

This section considers both AR(1) and AR(2) models with parameter configurations that range from stationary to near unit root and unit root processes. The drift coefficient in both AR(1) and AR(2) models is set to 0.5, which fits within the localized drift form of Phillips et al. (2013) for different values of γ and T. The error variance is normalized to unity. For the AR(1) model, we consider values of ρ_n from 0.1 to 1.03 with sample sizes from 25 to 200. These parameter values capture properties of stationary autoregressions as well as models with persistence characteristics that arise for ρ_n in the vicinity of unity. Strictly speaking, the CD representation of x_t in (2.2) requires $\rho_n < 1$, thereby excluding unit root and local unit root processes on the explosive side of unity. However, Chen and Deo (2009a) show that REML can achieve bias reductions of around 50% even when the AR parameter is close to unity. Hence, for comparative purposes, we include their approach in this study for cases where $\rho_n \geqslant 1$. Autoregressive coefficients greater than unity are associated with explosive behavior, which is demonstrated empirically by some financial time series during periods of market exuberance.

Table 4 reports finite sample estimation performance of REML-CD, REML-PC and MLE procedures in terms of bias, variance, and root mean square error (RMSE hereafter). The

Table 4: Finite sample performance comparison between MLE and alternative REML estimators for the AR(1) model with intercept.

Sample Size		25			50			100			200		
ρ_n	Stats	MLE	PC	CD	MLE	PC	CD	MLE	PC	CD	MLE	PC	CD
0.1	Bias	-0.058	-0.005	-0.013	-0.028	-0.003	-0.005	-0.014	-0.002	-0.003	-0.028	-0.003	-0.005
	Var	0.040	0.050	0.044	0.020	0.022	0.021	0.010	0.011	0.010	0.020	0.022	0.021
	RMSE	0.209	0.224	0.210	0.145	0.148	0.145	0.102	0.103	0.102	0.145	0.148	0.145
0.5	Bias	-0.101	0.011	-0.023	-0.051	-0.009	-0.018	-0.026	-0.005	-0.010	-0.013	-0.003	-0.005
	Var	0.034	0.075	0.044	0.016	0.018	0.018	0.008	0.008	0.008	0.004	0.004	0.004
	RMSE	0.210	0.275	0.211	0.138	0.135	0.134	0.093	0.092	0.091	0.064	0.064	0.063
0.7	Bias	-0.108	0.013	0.096	-0.058	-0.006	0.044	-0.030	-0.006	-0.001	-0.015	-0.003	-0.004
	Var	0.025	0.055	0.049	0.012	0.016	0.027	0.006	0.006	0.007	0.003	0.003	0.003
	RMSE	0.191	0.235	0.242	0.123	0.125	0.169	0.081	0.077	0.086	0.054	0.053	0.054
0.9	Bias	-0.056	-0.010	0.122	-0.035	-0.007	0.105	-0.022	-0.004	0.099	-0.013	-0.002	0.097
	Var	0.008	0.009	0.001	0.003	0.003	0.000	0.002	0.002	0.000	0.001	0.001	0.000
	RMSE	0.106	0.097	0.126	0.066	0.058	0.106	0.045	0.041	0.099	0.032	0.030	0.097
0.95	Bias	-0.030	-0.004	0.090	-0.016	-0.003	0.065	-0.010	-0.002	0.055	-0.007	-0.001	0.051
	Var	0.003	0.004	0.000	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	RMSE	0.066	0.061	0.091	0.035	0.031	0.066	0.023	0.021	0.055	0.017	0.016	0.051
0.99	Bias	-0.013	0.000	0.066	-0.004	0.000	0.036	-0.002	0.000	0.022	-0.001	0.000	0.015
	Var	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	RMSE	0.039	0.038	0.067	0.014	0.013	0.037	0.006	0.006	0.022	0.003	0.003	0.015
1	Bias	-0.011	0.000	0.060	-0.003	0.000	0.030	-0.001	0.000	0.015	-0.003	0.000	0.030
	Var	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	RMSE	0.034	0.033	0.062	0.011	0.011	0.030	0.004	0.004	0.015	0.011	0.011	0.030
1.01	Bias	-0.008	0.001	0.055	-0.002	0.000	0.025	0.000	0.000	0.010	0.000	0.000	0.003
	Var	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	RMSE	0.029	0.029	0.056	0.008	0.008	0.025	0.002	0.002	0.010	0.000	0.000	0.003
1.03	Bias	-0.005	0.001	0.045	0.000	0.000	0.015	0.000	0.000	0.003	0.000	0.000	0.000
	Var	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	RMSE	0.022	0.023	0.046	0.005	0.005	0.016	0.001	0.001	0.003	0.000	0.001	0.001

main findings are summarized as follows. (1) For all cases, REML-PC produces the smallest bias for each ρ_n and for each sample size among three estimators. (2) When ρ_n is smaller than $\frac{1}{2}$, both REML-CD and REML-PC produce estimator with similar variance, which is close to but slightly bigger than that yielded by MLE. When taking into account the reduced bias, all three methods have similar RMSE. (3) When ρ_n is bigger than $\frac{1}{2}$, we can see the advantage of REML-PC. It is able to give rise to not only the greatest bias reduction, but also the smallest RMSE. In addition, there is no trade-off between the bias and variance by REML-PC for many cases.

Table 5 reports the result for AR(2) process with drift, in which the sum of the autoregressive coefficients (the long run AR coefficient) measures the persistence of the process. The

Table 5: Finite sample performance comparison between MLE and alternative REML estimators for AR(2) model with intercept.

Sample Size			25			50			100			200		
Sum	(ρ_{1n}, ρ_{2n})	Stats	MLE	PC	CD	MLE	PC	CD	MLE	PC	CD	MLE	PC	CD
0.2	(0.7,-0.5)	Bias	0.002	0.008	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		Var	0.037	0.106	0.036	0.016	0.018	0.016	0.008	0.008	0.008	0.004	0.004	0.004
		RMSE	0.195	0.331	0.191	0.129	0.135	0.127	0.088	0.090	0.088	0.061	0.062	0.061
0.6	(1.19,-0.59)	Bias	-0.032	-0.017	0.007	-0.015	-0.006	-0.003	-0.007	-0.002	-0.002	-0.003	-0.001	-0.001
		Var	0.033	0.112	0.031	0.014	0.017	0.014	0.007	0.007	0.007	0.003	0.003	0.003
		RMSE	0.189	0.352	0.181	0.123	0.131	0.120	0.084	0.084	0.083	0.058	0.058	0.058
0.8	(0.2,0.6)	Bias	-0.058	-0.010	0.109	-0.036	-0.007	0.099	-0.021	-0.004	0.094	-0.012	-0.002	0.083
		Var	0.032	0.063	0.032	0.014	0.016	0.013	0.007	0.007	0.006	0.003	0.003	0.004
		RMSE	0.190	0.251	0.226	0.126	0.128	0.156	0.085	0.085	0.123	0.058	0.058	0.104
0.91	(1.05,-0.14)	Bias	-0.025	-0.004	0.054	-0.016	-0.003	0.047	-0.010	-0.002	0.044	-0.006	-0.001	0.044
		Var	0.042	0.075	0.036	0.020	0.022	0.017	0.010	0.010	0.009	0.005	0.005	0.004
		RMSE	0.220	0.273	0.223	0.145	0.148	0.155	0.101	0.101	0.111	0.070	0.070	0.082
0.96	(1.1,-0.14)	Bias	-0.012	-0.001	0.036	-0.007	-0.001	0.026	-0.004	-0.001	0.022	0.000	0.000	0.003
		Var	0.042	0.073	0.037	0.020	0.022	0.018	0.010	0.010	0.009	0.005	0.005	0.005
		RMSE	0.224	0.270	0.244	0.147	0.149	0.183	0.101	0.102	0.128	0.071	0.071	0.085
0.99	(1.31,-0.32)	Bias	-0.004	0.000	0.020	-0.002	0.000	0.012	-0.001	0.000	0.008	-0.001	0.000	0.006
		Var	0.041	0.094	0.032	0.019	0.022	0.014	0.009	0.010	0.007	0.005	0.005	0.003
		RMSE	0.237	0.307	0.249	0.151	0.148	0.215	0.102	0.099	0.183	0.070	0.069	0.141
1	(1.47,-0.47)	Bias	-0.001	-0.001	0.012	0.000	0.000	0.005	0.000	0.000	0.002	0.000	0.001	0.001
		Var	0.039	0.120	0.026	0.018	0.021	0.010	0.008	0.009	0.004	0.004	0.004	0.002
		RMSE	0.246	0.348	0.242	0.152	0.145	0.225	0.100	0.095	0.219	0.067	0.065	0.218
1.01	(1.14,-0.13)	Bias	-0.002	0.000	0.019	0.000	0.000	0.007	0.000	0.000	0.001	0.000	0.000	-0.001
		Var	0.042	0.071	0.040	0.020	0.023	0.019	0.010	0.011	0.009	0.005	0.005	0.004
		RMSE	0.229	0.266	0.278	0.152	0.153	0.259	0.104	0.104	0.269	0.072	0.072	0.310
1.03	(1.12,-0.19)	Bias	0.000	0.001	0.014	0.001	0.000	0.001	0.000	0.000	-0.004	0.000	0.000	-0.006
		Var	0.043	0.068	0.043	0.021	0.024	0.020	0.010	0.011	0.009	0.005	0.005	0.005
		RMSE	0.229	0.261	0.295	0.153	0.154	0.296	0.105	0.106	0.343	0.073	0.074	0.408

long run AR coefficients considered range from 0.2 to 1.03. Some parameter configurations are adopted from published empirical studies in order to make the DGP more realistic. To be specific, we set (ρ_{1n}, ρ_{2n}) to be $(1.10, -0.14)$, $(1.05, -0.14)$, $(0.2, 0.6)$ with corresponding long run AR parameters 0.96, 0.91 and 0.8, respectively. The first two pairs are taken from Amihud, Hurvich and Wang (2010), which assumes the predictor variable in predictive regression follow an AR(2) model with intercept, and the associated coefficients are $(1.10, -0.14)$ and $(1.05, -0.14)$. The pair $(0.2, 0.6)$ follows the setting in Zhu (2012). We report the mean squared bias, mean variance and mean RMSE for ρ_{1n} and ρ_{2n}.

A summary of the results from Table 5 is as follows. (1) REML-PC produces the smallest bias in most cases. (2) When the sum of the slope coefficients is smaller than 0.6, REML-CD has a slight advantage in producing an estimator with a smaller RMSE than that of REML-PC. This is similar to the AR(1) case. (3) When $\rho_{1n} + \rho_{2n}$ exceeds 0.6, REML-PC produces the greatest bias reduction, compared to REML-CD and MLE. In addition, REML-PC yields the smallest RMSE for sample sizes of 50, 100 and 200. (4) When sample size is 25, REML-CD has a better finite sample performance in terms of RMSE.

Overall, Tables 4 and 5 suggest that REML-PC provides improvements over the estimation procedure of Chen and Deo (2009a). Specifically, the results show that REML-PC reduces bias, and substantially so in the vicinity of unity, without increasing RMSE for the autoregressive model with drift. Bias reduction is achieved in the PC implementation because the representation of x_t in (2.8) is equivalent to that in the predictive regression (1.3) for any t and ρ_n, whereas the CD representation (2.2) uses an approximate form for x_t in (1.3). Moreover, REML-PC applies to unit root, local unit root, and even mildly explosive processes. This implementation of REML reduces the impact of the intercept on the estimation of the slope parameters in autoregression and provides a foundation for inference using RLRT in predictive regression, which we now discuss.

4.2 Predictive Regression – Estimation

This section and the following section report simulations with predictive regressions. The results cover: (1) finite sample comparisons between MLE, REML-CD, and REML-PC concerning the estimation of β; (2) size and power comparisons using MLE, RLRT-CD, and RLRT-PC when the true DGP has drift or not; (3) and similar size and power comparisons with the Campbell and Yogo (2006), Phillips and Magdalinos (2009a), and Jansson and Moreira (2006) procedures.

We set ρ_n in the interval $[0.8, 1]$, which covers many empirical applications. The error variances are set to unity ($\sigma_{00}^2 = \sigma_{xx}^2 = 1$) and ϕ is set to -0.95 unless otherwise specified, concordant with the high negative correlation between stock returns and many commonly used predictors. Standardizing the variances helps focus attention on the impact of other parameters on the RLRT. We parameterized the predictive slope coefficient in local form as $\beta = b\sqrt{1-\phi^2}/n$, providing size analysis for $b = 0$ and local power analysis with $b \neq 0$. Since financial data are often skewed and have heavy tails, we also considered (in unreported simulations) cases with t distributed errors (with 5 degrees of freedom) to measure the impact of thick tail behavior in stock returns and found results similar to those reported here, so the new procedures appear to be robust to this type of distributional error specification.

Table 6 reports finite sample performance for estimates of β obtained by MLE, REML-CD and REML-PC under the null hypothesis $\beta = 0$. The findings can be summarized as follows. (1) In terms of bias, REML-PC yields the smallest bias of all methods. REML-PC often reduces bias in the other procedures by 80% or more, and in some cases, almost removes bias completely. (2) REML-PC typically achieves bias reduction while maintaining or lowering RMSE. When the sample size exceeds 50, the results Table 6 shows that REML-PC actually produces a slightly smaller RMSE than that by MLE for several cases. It appears from table 6 that REML is able to remove the bias by a large extent without loss of efficiency.

4.3 Predictive Regression – Inference

We first present size comparisons in alternative implementations of RLRTs involving predictors with and without intercepts. The simulations findings show that when the true DGP has an intercept, RLRT-CD suffers size distortion whereas RLRT-PC maintains stable size and has good power. When the predictor follows an autoregression without intercept both RLRT-PC and RLRT-CD perform well in terms of size and power.

Table 7 shows size comparisons between MLE, RLRT-CD and RLRT-PC, when the predictor follows an AR(1) process with intercept. Right sided testing is performed under a 5% nominal significance level and the critical value for RLRT-CD is simulated using the sup bound critical value from Chen et al. (2013). Similar results apply when using the critical value from Theorem 2.1 for RLRT-CD. The findings reveal that RLRT-PC has a uniform advantage over RLRT-CD in terms of size and that the actual size of the CD test is close to unity for local to unity and mildly integrated predictors. For a unit root predictor, the CD test has size closer to the nominal size but can produce negative RLRT values. Table 8 reports complementary results for power, showing that RLRT-PC provides a uniform advantage in terms of power compared to RLRT-CD.

Figure 7 provides power curves of the RLRT using the two implementations in the case of a mildly integrated predictor x_t with intercept. The power curve for RLRT-PC is monotonic and that of RLRT-CD is U shaped, declining to zero before climbing to unity as b increases. The U shaped behavior of the power curve explains the bias in the RLRT-CD test and why the power reported in Table 8 is smaller than the size reported in Table 7 for a mildly integrated predictor.

For AR(1) predictors without an intercept, Table 9 reports finite sample comparisons of rejection rates for RLRT-CD and RLRT-PC tests. This specialization conforms exactly to the CD implementation of RLRT and so reveals any potential disadvantages in the RLRT-PC procedure's allowance for intercept effects. We find the following results. (1) Both RLRT-CD and RLRT-PC generally show good test size overall but the CD test is oversized in the unit root case and the PC test tends to be conservative. (2) In terms of power, RLRT-PC maintains good power except in the case of a mildly integrated predictor when $b = 25$, but its local power is dominated by RLRT-CD for local unit root and mildly integrated predictors.

Figure 8 shows power curves for the mildly unit root predictor when $\gamma = \frac{1}{2}$, $\beta = \frac{b\sqrt{1-\phi^2}}{n}$ with a (shifted) logarithmic scale for $b \in [0, 50]$ and with other parameters following the configurations in Table 8. The power loss in allowing for an intercept in the RLRT-PC procedure is particularly evident in the mildly integrated case.

4.4 Additional Comparisons

We next compare the RLRT procedure with the methods of Campbell and Yogo (2006; CY), Phillips and Magdalinos (2009a) and Jansson and Moreira (2006). These methods are particularly designed to allow for local to unity predictors. CY (2006) used Bonferroni methods to produce feasible tests for a single predictor in such cases, refining the procedure proposed by Cavanagh, Elliott and Stock (1995). Recent work (Phillips, 2014) has shown that in no intercept cases the CY tests are undersized for near unit root predictors and oversized for mildly integrated predictors with rejection probabilities close to unity. We focus here on the effects of intercept corrections on the performance of these tests in comparision with RLRT procedures.

Table 10 provides size comparisons between RLRT-PC and the conventional regression t-test, Bonferroni t-test, Bonferroni Q-test, and Sup Q-test, as recommended in CY (2006). The DGP used here includes intercepts for both y_t and x_t. The results show that for a nominal 5% size, the conventional t-test is oversized and the Bonferroni t-test is undersized for a unit root predictor with a rejection rate less than 1%. The corresponding rejection rate of the Bonferroni Q-test is close to 0%. Hence drift impacts the Q-test statistic performance even though the test is said to be invariant to the presence of an intercept in CY (2006). The Sup Q-test has relatively good size performance for a unit root predictor but is otherwise undersized. In comparison to all these tests, the RLRT-PC test does well, with an overwhelming advantage in size control in the predictive regression context, allowing for different sample sizes and predictors with various degrees of persistence. Power is reported in Table 11 and the findings show that RLRT-PC has uniformly better power than CY (2006) for all parameter configurations and all sample sizes. In addition to improved size, RLRT methods have a computational advantage because they do not rely on confidence belts that have to be prepared to implement the test. Moreover, unlike the Bonferroni-based procedures, RLRT extends readily to the multiple predictor case.

In the case of multiple predictors we compare our procedure with the IVX method of Phillips and Magdalinos (2009a) which has the dual advantages of convenient treatment of multiple predictors and a common chi-square limit theory. The IVX framework is based on the multivariate cointegrated system

$$\begin{aligned} y_t &= Ax_t + u_{ot} \\ x_t &= R_n x_{t-1} + u_{xt}, \quad t = 1, .., n \end{aligned} \tag{4.1}$$

where A is an $m \times k$ coefficient matrix and $R_n = I_K + \frac{C}{n^\alpha}$ is a autoregressive matrix with roots $|\lambda_i (R_n)| \leqslant 1$. The matrix $C = diag\,(c_1, \dots c_k)$ is a diagonal matrix with $c_k \leqslant 0$ for all $i = 1, \dots k$. IVX constructs an instrument z_t directly from the predictor x_t using the autoregression $z_t = R_{nz} z_{t-1} + u_{zt}$ for some known $R_{nz} = I_{K_z} + \frac{C_z}{n^\omega}, \omega \in (0,1)$, and $C_z = diag\,(c_{z,1}, \dots c_{z,K_z})$ with inputs $u_{zt} = \Delta x_t$. The IVX Wald test for predictability is pivotal and follows a standard χ^2 distribution. Kostakis, Magdalinos and Stamatogiannis (2014) employ IVX as a tool for predictive regression and modify the IVX method to accommodate the presence of the intercept in y_t in the true DGP. The following comparative study follows their procedure in implementation.

Table 12 compares both RLRT-CD and RLRT-PC with IVX in the context of bivariate predictors and with a 10% right tailed critical value. The parameter configuration follows Table 4 in Kostakis et al. (2014). In particular, we choose $\varpi = 0.95$ and $C_z = I_k$. When intercepts are present in the true DGP, we set the intercept for y_t, x_{1t} and x_{2t} to be $0.5, 5 \times n^{-0.5}$ and $5 \times n^{-0.5}$ $(n = 100)$ respectively. For power comparisons, we set $\beta_i = \frac{25\sqrt{1-\phi_i^2}}{n}$ $(i = 1, 2)$. In this case, $\phi_1 = -0.8474$ and $\phi_2 = 0.012$. Table 12 shows that RLRT-CD has less size distortion and better power by a small margin compared to IVX when there is no intercept in the true DGP. When there are intercepts in the true DGP, RLRT-PC is oversized by around 1%, whereas IVX is undersized by around 1% so both methods perform well in terms of size control. RLRT-PC demonstrates some advantage in terms of power for stationary predictors over IVX. Thus, for bivariate predictors the results slightly favor RLRT-PC for unit root and stationary predictors and IVX is favored for the remaining cases.

In the appendix of the supplement to this paper, we report size and power comparisons between RLRT-PC and IVX for the case of a univariate predictor in Table S4, S5 and S6 for configurations that corresponds to Table 7, 8 and 9. In the univariate case, when there is an intercept in the true DGP, IVX is undersized with size close to zero yet still has very good power except when T is as small as 50, in which case IVX has less power than RLRT-PC. When there is no drift term in the true DGP, IVX is slightly oversized and has stable size across all generating mechanisms, while RLRT-PC is generally undersized; and IVX dominates RLRT-PC in terms of power particularly for mildly integrated predictors but is slightly dominated by RLRT in the unit root case. Importantly, in this case the power of IVX increases with the sample size, whereast both CD and PC versions of RLRT show decline in power as the sample size increases in the mildly integrated case. In sum, for a univariate predictor and small sample sizes RLRT appears to have a power advantage only for highly persistent predictors. In general, IVX is competitive in terms of both size and power, and is generally far superior in power to both RLRT methods when the predictor is mildly integrated except for small sample sizes.

Some further size and power comparisons were conducted with the Jansson and Moreira (2006) procedure. Chen and Deo (2009a) point out that the likelihood used in Jansson and Moreira (2006; JM) is already restricted because it relates to the likelihood of a maximal invariant. However, RLRT exploits the restricted likelihood in developing the inferential procedure, whereas JM (2006) considers the maximal invariant test statistic without resorting to the associated likelihood. One similarity of the two approaches is that they both reduce or remove curvature in the test problem. The relatively smaller curvature in the likelihood ratio test delivered by the restricted likelihood ensures an improved approximation of the asymptotic to the finite sample distribution, in comparison to the standard likelihood. On the other hand, removing curvature by conditioning on some specific ancillary statistics enables JM (2006) to produce a uniformly most powerful conditionally unbiased test. In particular, JM first derive a maximal invariant statistic based on transforming observations of y_t and x_t to $(y_t - y_{t-1},\, x_t)$. Chen and Deo (2009a) point out that this transformation under the assumption of no intercept in x_t enables the use of the exact likelihood rather than the restricted likelihood. JM (2006) find a sufficient statistic for the distribution of the maximal invariant which is used to construct a test with a uniformly most powerful conditional optimality property.

Table 13 is the counterpart of Table 7 in Chen and Deo (2009a) and we additionally examine the impact of $Corr\,(u_{xt}, u_{0t})$ on size and power. The same parameter configuration is used in Table 2 in Jansson and Moreira (2006). Chen and Deo (2009a) show how, with this particular parameter configuration, extremely high values of $Corr\,(u_{xt}, u_{0t})$ can degrade size performance for RLRT-CD. However, when the predictor has an intercept, it is apparent from Table 13 that the JM (2006) method has size and power both close to zero, indicating the sensitivity of this method to the presence of drift. RLRT-PC, on the other hand, is seen to be robust to $Corr\,(u_{xt}, u_{0t})$.

When the predictor has no drift, we refer to the size performance of RLRT-PC given already in Table 9. Table 6 in Chen and Deo (2009a) shows that the RLRT-CD is oversized when the predictor is nonstationary and $Corr\,(u_{xt}, u_{0x.t}) \simeq -1$. However, RLRT-CD is no longer oversized at the unit root when $Corr\,(u_{xt}, u_{0x.t}) \simeq -0.5$. Chen and Deo suggest that it is the correlation in the innovations that is responsible for the size distortion. Our simulation results indicate that it is the form of the AR(1) model specified for the predictor (rather than the error correlation) that impacts inferential performance by producing size distortion as the RLRT-PC test shows virtually no size distortion for all sample sizes and all ρ_n. Finally, as is now well known (e.g. Kasparis, Andreou, Phillips, 2014) the JM procedure involves substantial computation and encounters numerical difficulties in some parameter configurations. RLRT demonstrates better size and power properties and involves far less computational time.

5 Conclusion

Building on the work of Chen and Deo (2009a), this paper shows the advantages of the use of restricted likelihood techniques in predictive regression models. The REML estimator and RLRT test both exhibit good finite sample properties. The REML estimator reduces the bias of the MLE by around 50% and the RLRT test for predictability corrects size distortion in the standard t-test and outperforms several commonly used predictive regression tests in terms of size and power when drift is present in persistent predictors. The main contribution of the paper lies in the extension of earlier research on RLRT testing by including drift in the specification and by allowing for multiple predictors in the generating mechanism, thereby providing a wider field of potential applications.

The modifications involved in the new procedure directly remove the impact of the drift in implementing the RLRT by using the exact form of the predictor model in building the restricted likelihood. This procedure is shown to easily accommodate multiple predictors and autoregressive predictors with different initializations. Simulations show that the RLRT procedure has superior finite sample performance in terms of size and power compared to both Campbell and Yogo (2006) and Jansson and Moreira (2006) methods, even for a true DGP without drift. Our simulations show that the IVX method of Phillips and Magdalinos (2009a) and Kostakis, Magdalinos and Stamatogiannis (2014) also performs well and is robust to the presence of intercepts and multiple predictors, giving particularly good performance in relation to RLRT methods when the predictors are mildly integrated.

References

Amihud, Y., & Hurvich, C. M. (2004). Predictive regressions: A reduced-bias estimation method. *Journal of Financial and Quantitative Analysis*, 39(04), 813-841.

Amihud, Y., Hurvich, C. M., & Wang, Y. (2010). Predictive regression with order-p autoregressive predictors. *Journal of Empirical Finance*, 17(3), 513-525.

Campbell, J. Y., & Shiller, R. J. (1988). The dividend-price ratio and expectations of future dividends and discount factors. *Review of financial studies,* 1(3), 195-228.

Campbell, J. & M. Yogo, (2006), Efficient tests of stock return predictability. Journal of Financial Economics 81, 27–60.

Cavanagh, C. L., Elliott, G., & James, H. Stock, 1995, Inference in models with nearly integrated regressors. *Econometric theor*y, 11(5), 1131-1147.

Cheang, W. K. & Reinsel, G., (2000), Bias reduction of autoregressive estimates in time series regression model through restricted maximum likelihood, *Journal of American Statistical Association* 95, 1173–84.

Chen, W. & Deo, R., (2009a), Bias reduction and likelihood based almost-exactly sized hypothesis testing in predictive regressions using the restricted likelihood. *Econometric Theory* 25, 1143–79.

Chen, W. & Deo, R., (2009b), The restricted likelihood ratio test at the boundary in autoregressive series. *Journal of Time Series Analysis* 30, 618–30.

Chen, W. & Deo, R., (2010), Weighted least squares approximate restricted likelihood estimation for vector autoregressive processes. *Biometrika* 97,231–7.

Chen, W. & Deo, R., (2011), The restricted likelihood ratio test for autoregressive processes. *Journal of Time Series Analysis* 34, 325–339.

Chen, W. & Deo, R., & Yi, (2013). Uniform Inference in Predictive Regression Models. *Journal of Business & Economic Statistics*, 31(4), 525-533.

Cochrane, J. H. (2008). The dog that did not bark: A defense of return predictability. *Review of Financial Studies*, 21(4), 1533-1575.

Cooper, D. M., & Thompson, R., (1977), A Note on the Estimation of the Parameters of the Autoregressive Moving Average Process, *Biometrika*, 64, 625-628.

Cowles 3rd, A. (1933). Can stock market forecasters forecast?. Econometrica: Journal of the Econometric Society, 309-324.

R. S. Deo, 2012, Improved forecasting of autoregressive series by weighted least squares approximate REML estimation, *International Journal of Forecasting*, vol. 28, no.1, pp.39–43.

Engsted, T., Pedersen, T. Q., & Tanggaard, C. (2012). The Log-Linear Return Approximation, Bubbles, and Predictability. *Journal of Financial and Quantitative Analysis*, 47(03), 643-665.

Hayakawa, T. (1977). The likelihood ratio criterion and the asymptotic expansion of its distribution. *Annals of the Institute of Statistical Mathematics*, 29(1), 359-378.

Harville, D. A. (1974). Bayesian inference for variance components using only error contrasts. Biometrika, 61(2), 383-385.

Jansson, M. & M. Moreira (2006) Optimal inference in regression models with nearly integrated regressors. *Econometrica* 74, 681–714.

Kostakis, A., Magdalinos, A., & Stamatogiannis, M. (2014). Robust econometric inference for stock return predictability. Unpublished Manuscript, University of Nottingham, UK.

Leng & Shi, P. & Tsai C. L., (2008), Clarification: Regression model selection–a residual likelihood approach. *Journal of the Royal Statistical Society,* B70, Part5, p.1067.

Magdalinos, T. & P. C. B Phillips (2009). "Limit Theory for Cointegrated Systems with Moderately Integrated and Moderately Explosive Regressors", *Econometric Theory*, 482-526.

Patterson, H.D., & Thompson, Robin, (1971), Recovery of Inter- Block Information when Block Sizes Are Unequal, *Biometrika*, 58, 545-54.

Phillips, P. C. B. (1987a). "Time Series Regression with a Unit Root," *Econometrica,* 55, 277–302.

Phillips, P. C. B. (1987b). Towards a unified asymptotic theory for autoregression. *Biometrika*, 74(3), 535-547.

Phillips, P. C. B. (2014). On Confidence Intervals for Autoregressive Roots and Predictive Regression, *Econometrica,* 82, 1177-1195.

Phillips, P. C. B. and Y. Chen (2014). Restricted Likelihood Ratio Tests in Predictive Regression: Online Supplement, Working Paper, Yale University and Singapore Management University.

Phillips, P. C. B., & Lee (2013), Asset Pricing with Financial Bubble Risk, Working Paper, Yale University.

Phillips, P. C. B., & Magdalinos, T. (2007a). Limit theory for moderate deviations from a unit root. *Journal of Econometrics*, 136(1), 115-130.

Phillips, P. C. B., T. Magdalinos (2007b), "Limit Theory for Moderate Deviations from a Unit Root Under Weak Dependence," in G. D. A. Phillips and E. Tzavalis (Eds.) *The Refinement of Econometric Estimation and Test Procedures: Finite Sample and Asymptotic Analysis.* Cambridge: Cambridge University Press, pp.123-162.

Phillips, P. C. B., & Magdalinos, T. (2009a). Econometric inference in the vicinity of unity. CoFie Working Paper No. 7, Singapore Management University.

Phillips, P. C. B., & Magdalinos, T. (2009b). Unit root and cointegrating limit theory when initialization is in the infinite past. Econometric Theory, 25(6), 1682.

Phillips, P.C.B., S. Shi and J. Yu (2014). "Specification Sensitivity in Right-Tailed Unit Root Testing for Explosive Behaviour", *Oxford Bulletin of Economics and Statistics*, 76, 315-333.

Shi, P. & Tsai C. L., (2002), Regression model selection–a residual likelihood approach. *Journal of the Royal Statistical Society,* B64, 237–52.

Smyth GK, A P Verbyla, (1996), A conditional approach to residual maximum likelihood estimation in generalized linear models, *Journal of the Royal Statistical Society* B 1996, 58:565-572.

Stambaugh, R. F. (1999). Predictive regressions. *Journal of Financial Economics,* 54(3), 375-421.

Tunnicliffe Wilson, G., (1989), On the Use of Marginal Likelihood in Time Series Model Estimation, *Journal of the Royal Statistical Society*, Ser. B, 51, 15-27.

Verbyla, A. P., (1990), A conditional derivation of residual maximum likelihood. *Australian Journal of Statistics*, 32, 227-230.

Zhu, M. (2012). Jackknife for Bias Reduction in Predictive Regressions. *Journal of Financial Econometrics,* 11, 193-220.

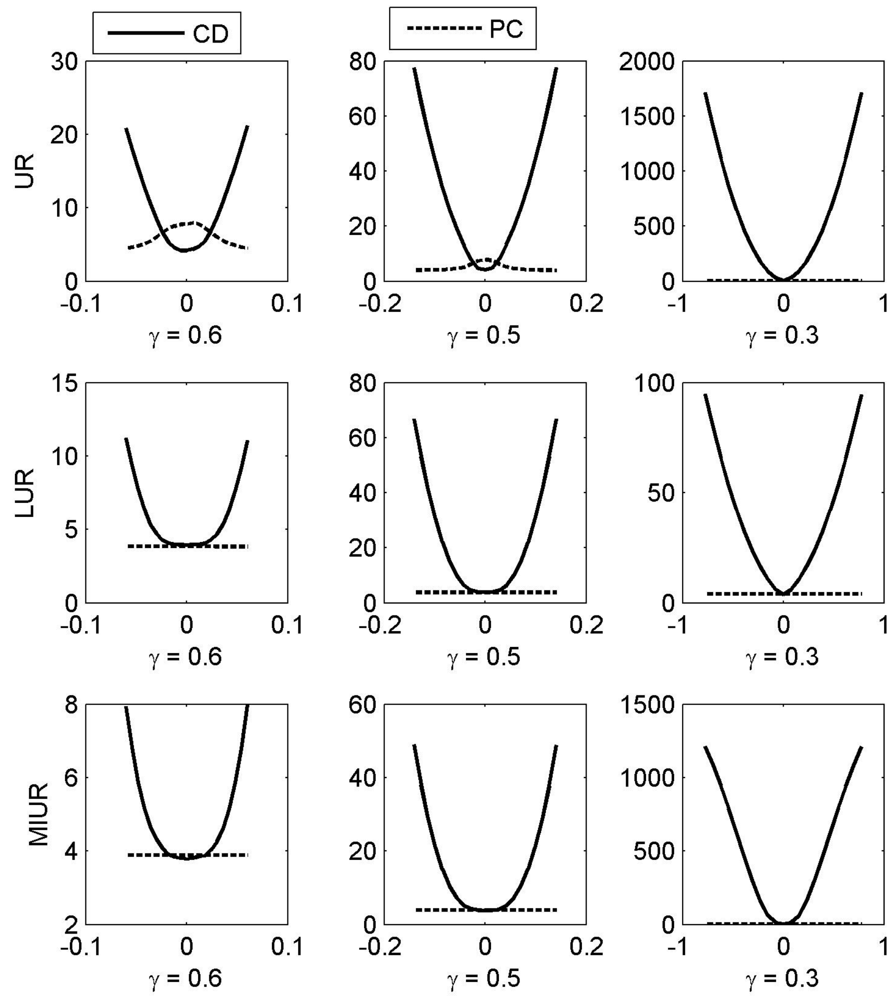

Figure 2: The sensitivity of 5% right tailed critical values to $\widetilde{\mu}$ for RLRT using alternative procedures. The horizontal axis in each case shows $\widetilde{\mu} \in [-10, 10]$ for intercept $\mu = \widetilde{\mu} n^{-\gamma}$. The vertical axis indicates simulated 5% right tailed critical values for $n = 5000$. (UR $\equiv$ Unit root with $\rho_n = 1$, LUR $\equiv$ Local unit root with $\rho_n = 1 + \frac{c}{n}$, MIUR $\equiv$ Moderately integrated root with $\rho_n = 1 + \frac{c}{n^\alpha}$.)

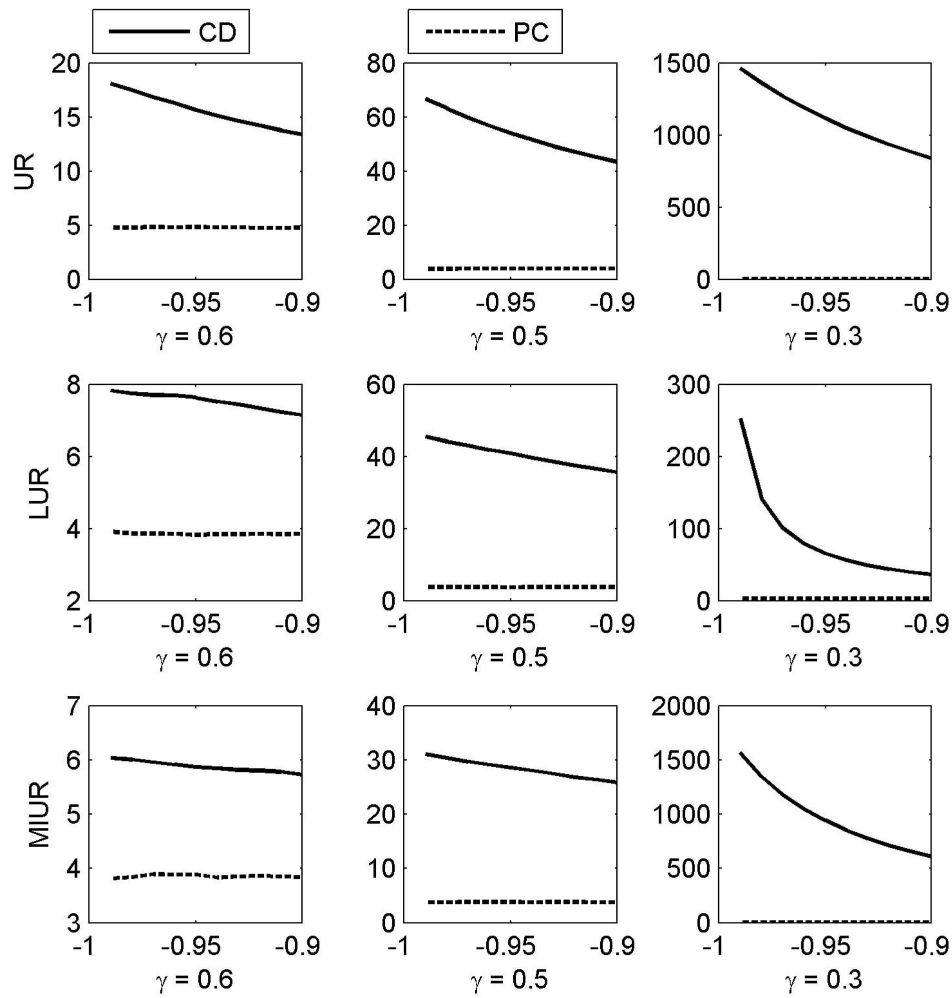

Figure 3: The sensitivity of 5% right tailed critical values to ϕ for RLRT under the CD and new (PC) implementation. The horizontal axis in each case measures values of $\phi \in [-0.99, -0.9]$ with intercept $\mu = 8 \times n^{-\gamma}$ for various γ. The vertical axis shows simulated 5% right tailed critical values for $n = 5000$.

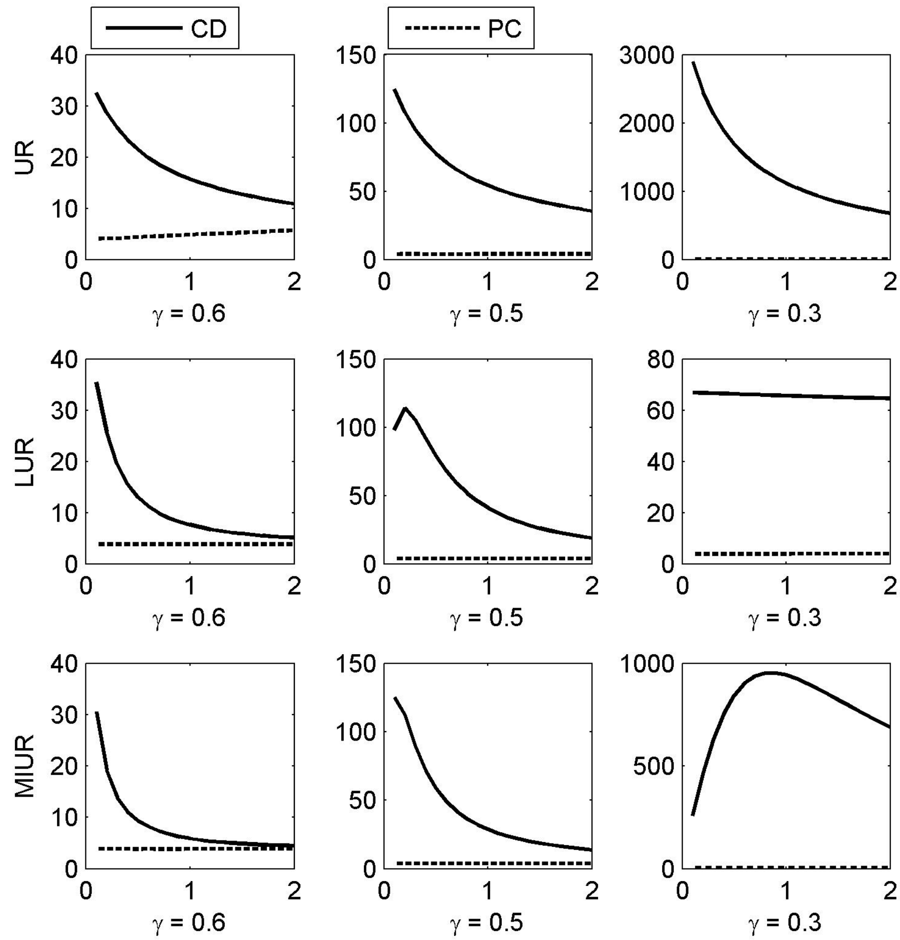

Figure 4: The sensitivity of 5% right tailed critical value to signal to noise ratio $\frac{\sigma_{xx}^2}{\sigma_{00}^2}$ for RLRT using alternative procedures. Horizontal axis in each case indicates $\sigma_{xx}^2 \in [0.1, 2]$ with $\sigma_{00}^2 = 1$ for intercept $\mu = 8 \times n^{-\gamma}$. Vertical axis indicates 5% right tailed critical value of simulated asymptotics of $n = 5000$.

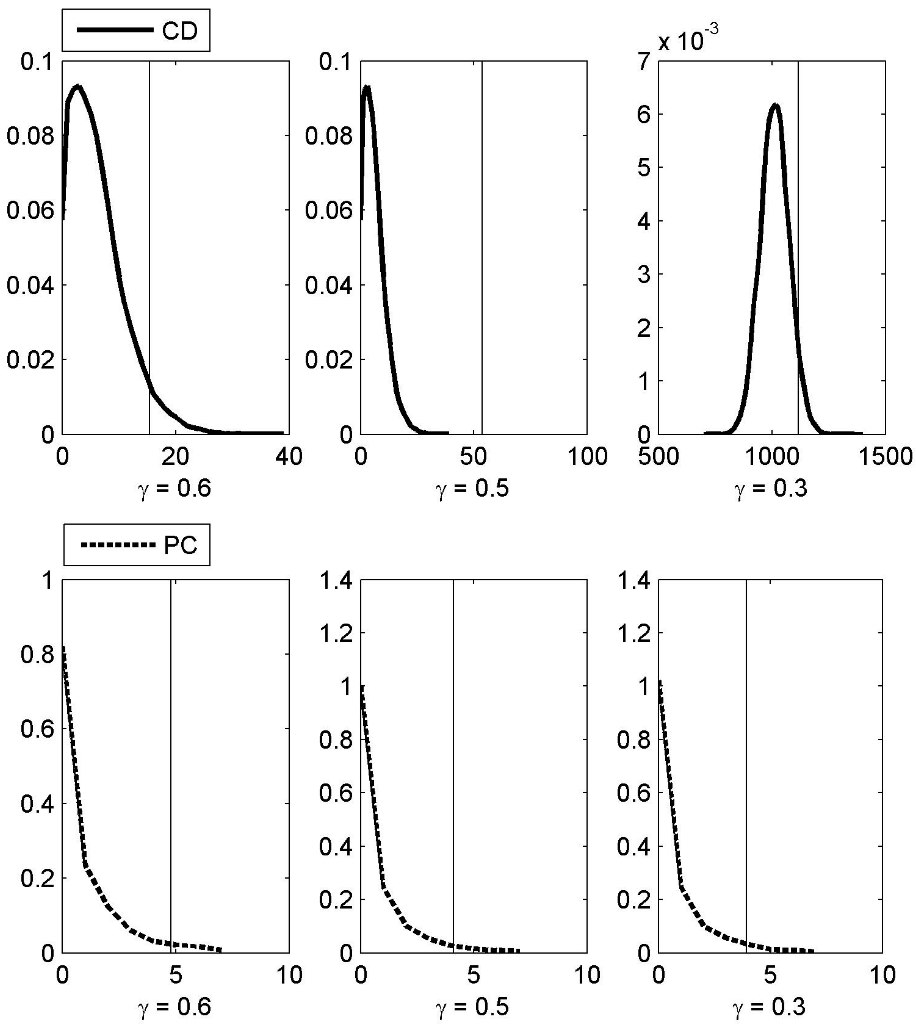

Figure 5: The sensitivity of the RLRT density to μ using alternative procedures for the case of a unit root predictor x_t.

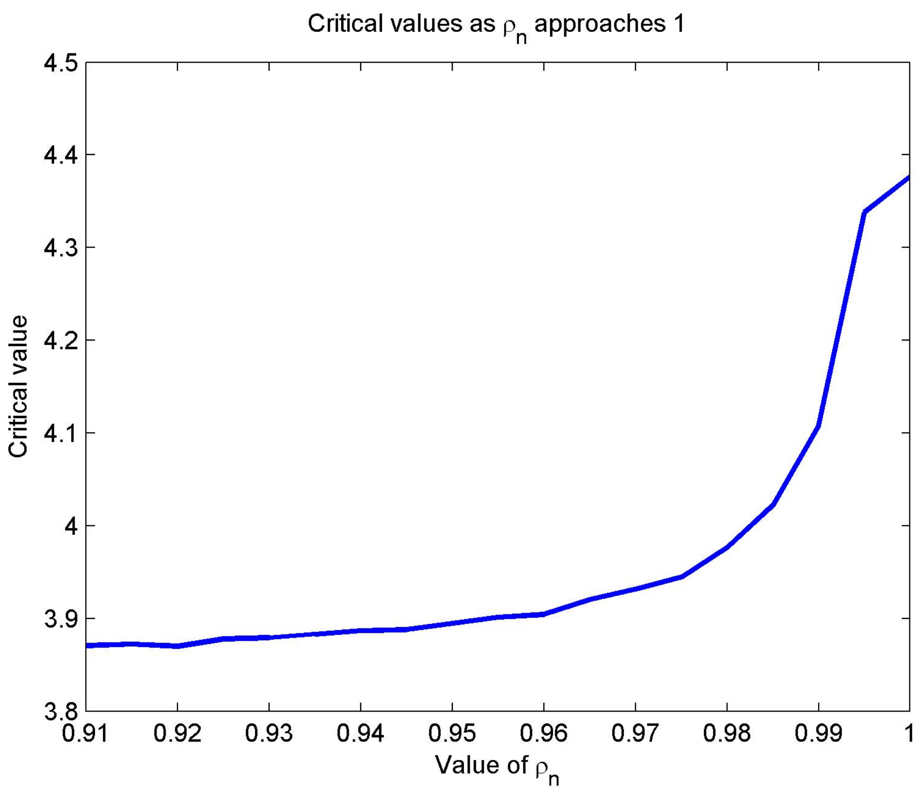

Figure 6: 5% right tailed sup bound critical values ($\overline{CV}_{\rho_n,\lambda}^{\gamma=\frac{1}{2}}$) of RLRT for ρ_n in the vicinity of unity.

Table 6: Finite sample performance comparison between MLE and alternative REML estimators for β under $H_0 : \beta = 0$, when the predictor x_t has intercept.

	ρ_n	n	Statis	$\gamma = 0.6$ $\widetilde{\mu}$=10			$\gamma = 0.5$ $\widetilde{\mu}$=5			$\gamma = 0.3$ $\widetilde{\mu}$=3		
				MLE	CD	PC	MLE	CD	PC	MLE	CD	PC
UR												
c = 0	1	50	Bias	0.003	-0.028	0.000	0.005	-0.028	0.000	0.003	-0.028	0.000
			SD	0.011	0.005	0.011	0.015	0.007	0.015	0.011	0.006	0.011
			RMSE	0.011	0.029	0.011	0.016	0.029	0.015	0.011	0.029	0.011
	1	100	Bias	0.002	-0.014	0.000	0.003	-0.014	0.000	0.001	-0.014	0.000
			SD	0.006	0.003	0.006	0.008	0.004	0.008	0.005	0.002	0.005
			RMSE	0.006	0.014	0.006	0.008	0.014	0.008	0.005	0.014	0.005
	1	200	Bias	0.001	-0.007	0.000	0.001	-0.007	0.000	0.000	-0.007	0.000
			SD	0.003	0.001	0.003	0.004	0.002	0.004	0.002	0.001	0.002
			RMSE	0.003	0.007	0.003	0.004	0.007	0.004	0.002	0.007	0.002
	1	400	Bias	0.001	-0.003	0.000	0.001	-0.003	0.000	0.000	-0.004	0.000
			SD	0.002	0.001	0.002	0.002	0.001	0.002	0.001	0.000	0.001
			RMSE	0.002	0.004	0.002	0.002	0.004	0.002	0.001	0.004	0.001
LUR												
c = -5	0.9	50	Bias	0.033	-0.094	0.007	0.046	-0.088	0.008	0.035	-0.094	0.007
			SD	0.056	0.017	0.058	0.069	0.033	0.072	0.057	0.018	0.059
			RMSE	0.065	0.096	0.058	0.083	0.094	0.073	0.067	0.096	0.059
	0.95	100	Bias	0.019	-0.047	0.003	0.024	-0.045	0.003	0.015	-0.047	0.003
			SD	0.030	0.009	0.031	0.035	0.012	0.036	0.026	0.008	0.026
			RMSE	0.035	0.048	0.031	0.042	0.047	0.037	0.030	0.048	0.026
	0.975	200	Bias	0.011	-0.023	0.002	0.013	-0.023	0.002	0.006	-0.024	0.001
			SD	0.016	0.005	0.017	0.018	0.005	0.019	0.012	0.003	0.012
			RMSE	0.019	0.024	0.017	0.022	0.023	0.019	0.013	0.024	0.012
	0.9875	400	Bias	0.006	-0.012	0.001	0.006	-0.011	0.001	0.003	-0.012	0.001
			SD	0.009	0.002	0.009	0.009	0.003	0.010	0.005	0.002	0.005
			RMSE	0.011	0.012	0.009	0.011	0.012	0.010	0.006	0.012	0.005
MIUR												
c = -5	0.8521	50	Bias	0.044	-0.130	0.007	0.054	-0.103	0.005	0.045	-0.129	0.007
$\alpha = 0.9$			SD	0.074	0.044	0.079	0.084	0.080	0.093	0.075	0.047	0.080
			RMSE	0.086	0.137	0.079	0.100	0.130	0.093	0.088	0.137	0.081
	0.9208	100	Bias	0.025	-0.070	0.003	0.030	-0.063	0.002	0.022	-0.073	0.003
			SD	0.041	0.016	0.044	0.045	0.026	0.050	0.038	0.013	0.039
			RMSE	0.049	0.072	0.044	0.054	0.068	0.050	0.044	0.074	0.040
	0.9575	200	Bias	0.015	-0.037	0.002	0.016	-0.034	0.002	0.011	-0.040	0.002
			SD	0.023	0.008	0.025	0.024	0.009	0.027	0.019	0.006	0.020
			RMSE	0.028	0.038	0.025	0.029	0.036	0.027	0.022	0.040	0.020
	0.9772	400	Bias	0.008	-0.019	0.001	0.009	-0.018	0.001	0.005	-0.021	0.001
			SD	0.013	0.004	0.013	0.013	0.004	0.014	0.009	0.003	0.010
			RMSE	0.015	0.020	0.014	0.015	0.019	0.014	0.011	0.022	0.010

Table 7: Size comparisons of RLRT using alternative procedures to test $H_0 : \beta = 0$, when the predictor x_t has an intercept.

	ρ_n	n	$\gamma = 0.6$ $\widetilde{\mu}$=10 μ	CD	PC	$\gamma = 0.5$ $\widetilde{\mu} = 5$ μ	CD	PC	$\gamma = 0.3$ $\widetilde{\mu} = 3$ μ	CD	PC
UR											
c = 0	1	50	0.956	0.049	0.053	0.707	0.045	0.051	0.928	0.051	0.053
	1	100	0.631	0.049	0.057	0.500	0.043	0.061	0.754	0.048	0.058
	1	200	0.416	0.031	0.048	0.354	0.031	0.049	0.612	0.032	0.049
	1	400	0.275	0.031	0.048	0.250	0.031	0.049	0.497	0.032	0.049
LUR											
c = -5	0.900	50	0.956	0.975	0.052	0.707	0.781	0.051	0.928	0.967	0.053
	0.950	100	0.631	0.960	0.055	0.500	0.847	0.058	0.754	0.991	0.049
	0.975	200	0.416	0.958	0.052	0.354	0.878	0.054	0.612	1.000	0.054
	0.988	400	0.275	0.958	0.052	0.250	0.878	0.054	0.497	1.000	0.054
MIUR											
c = -5	0.852	50	0.956	0.895	0.051	0.707	0.606	0.046	0.928	0.868	0.052
$\alpha = 0.9$	0.921	100	0.631	0.925	0.054	0.500	0.739	0.052	0.754	0.986	0.054
	0.958	200	0.416	0.943	0.047	0.354	0.827	0.041	0.612	1.000	0.055
	0.977	400	0.275	0.943	0.047	0.250	0.827	0.041	0.497	1.000	0.055

Table 8: Power comparisons of RLRT using alternative procedures for true $\beta = \frac{25\sqrt{1-\phi^2}}{n}$, when the predictor x_t has an intercept.

	ρ_n	n	$\gamma = 0.6$ $\widetilde{\mu}$=10 μ	CD	PC	$\gamma = 0.5$ $\widetilde{\mu} = 5$ μ	CD	PC	$\gamma = 0.3$ $\widetilde{\mu} = 3$ μ	CD	PC
UR											
c = 0	1	50	0.956	1.000	1.000	0.707	1.000	1.000	0.928	1.000	1.000
	1	100	0.631	1.000	1.000	0.500	1.000	1.000	0.754	1.000	1.000
	1	200	0.416	1.000	1.000	0.354	1.000	1.000	0.612	1.000	1.000
	1	400	0.275	1.000	1.000	0.250	1.000	1.000	0.497	1.000	1.000
LUR											
c = -5	0.900	50	0.956	0.922	0.993	0.707	0.835	0.944	0.928	0.918	0.991
	0.950	100	0.631	0.895	0.979	0.500	0.821	0.944	0.754	0.941	0.991
	0.975	200	0.416	0.904	0.977	0.354	0.854	0.947	0.612	0.985	1.000
	0.988	400	0.275	0.904	0.977	0.250	0.854	0.947	0.497	0.985	1.000
MIUR											
c = -5	0.852	50	0.956	0.113	0.510	0.707	0.125	0.352	0.928	0.111	0.491
$\alpha = 0.9$	0.921	100	0.631	0.037	0.355	0.500	0.045	0.253	0.754	0.036	0.476
	0.958	200	0.416	0.009	0.248	0.354	0.011	0.206	0.612	0.012	0.466
	0.977	400	0.275	0.009	0.248	0.250	0.011	0.206	0.497	0.012	0.466

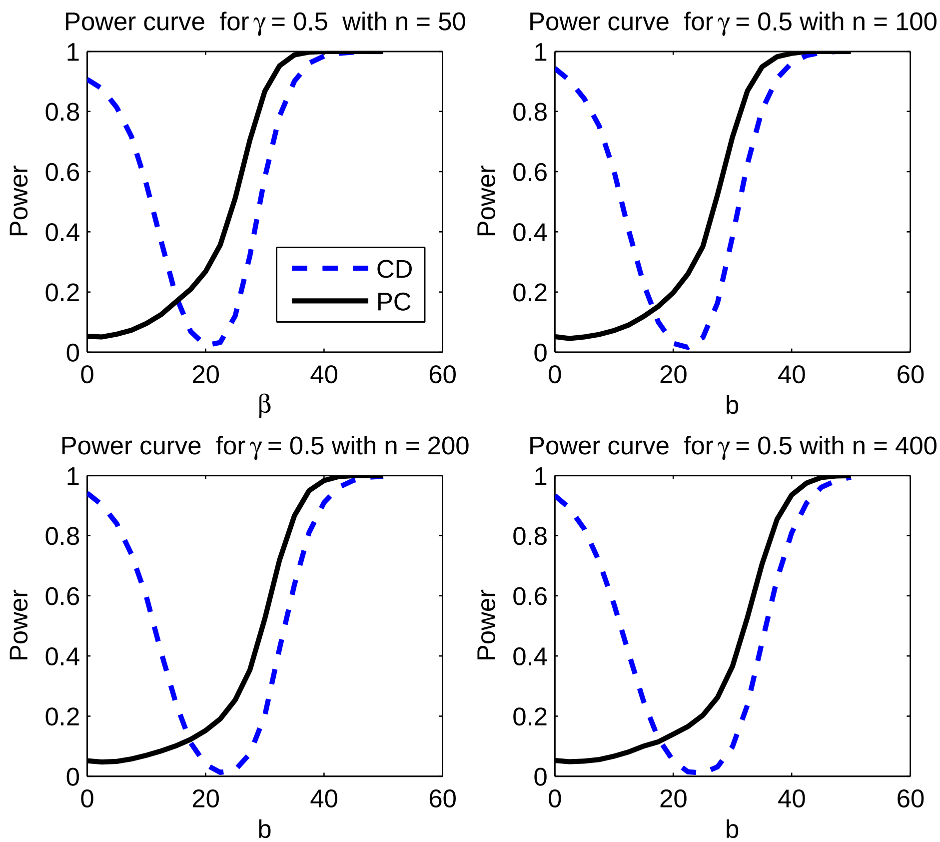

Figure 7: Power curves of RLRT using alternative procedures in the case of a mildly integrated predictor x_t with intercept.

Table 9: Size and power comparisons of RLRT tests of $H_0 : \beta = 0$, with true $\beta = \frac{b\sqrt{1-\phi^2}}{n}$, when the predictor x_t has no intercept.

	ρ_n	n	b = 0		b = 25		b = 50	
			CD	PC	CD	PC	CD	PC
UR								
c = 0	1	50	0.068	0.050	1.000	1.000	1.000	1.000
	1	100	0.064	0.042	1.000	1.000	1.000	1.000
	1	200	0.061	0.042	1.000	0.999	1.000	0.999
	1	400	0.067	0.047	0.999	0.999	0.999	0.999
LUR								
c = -5	0.900	50	0.034	0.029	0.869	0.622	1.000	1.000
	0.950	100	0.043	0.038	0.877	0.588	1.000	1.000
	0.975	200	0.033	0.024	0.896	0.561	1.000	1.000
	0.988	400	0.041	0.023	0.882	0.560	1.000	1.000
MIUR								
c = -5	0.852	50	0.037	0.036	0.480	0.152	1.000	0.998
$\alpha = 0.9$	0.921	100	0.047	0.041	0.435	0.094	0.999	0.999
	0.958	200	0.034	0.027	0.426	0.079	1.000	0.993
	0.977	400	0.043	0.027	0.421	0.078	1.000	0.992

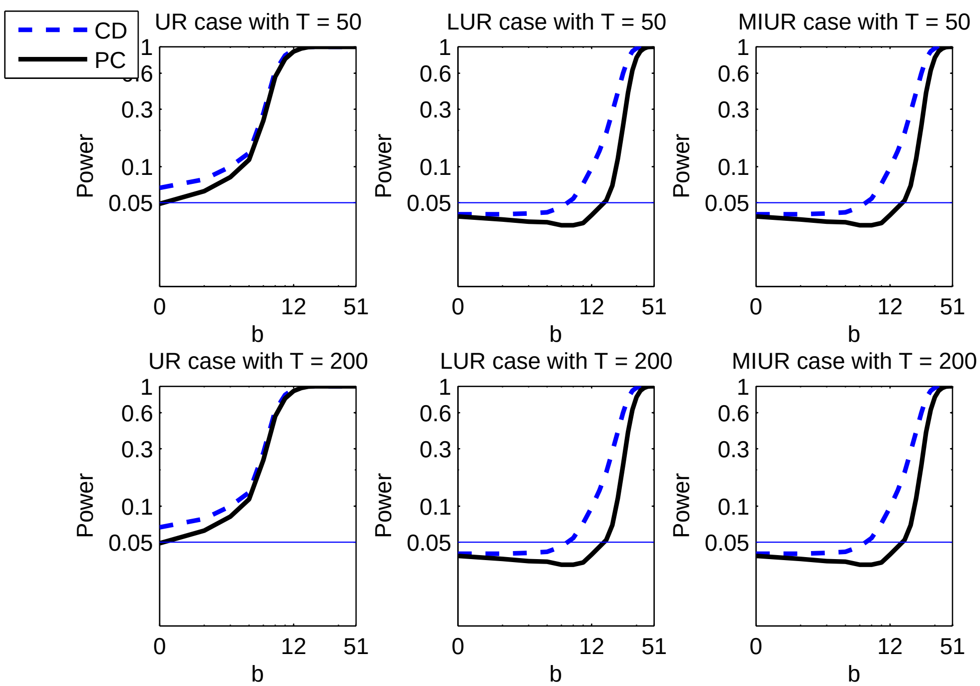

Figure 8: Power curves of RLRT using alternative procedures for a mildly integrated predictor x_t without intercept.

Table 10: Finite sample rejection rate (size) comparison between conventional t-test, Bonferroni t-test, Bonferroni Q-test, Sup Q-test and RLRT-PC for testing $H_0 : \beta = 0$, when the predictor x_t has intercept, using normal critical value for the t ratio and confidence belt based critical values for the Q tests.

		n	c=0	μ	t	Bonf.t	Bonf.Q	Sup Q	PC
$\gamma = 0.6$	$\widetilde{\mu} = 10$								
	c = 0	50	1	0.956	0.088	0.008	0.000	0.059	0.062
		100	1	0.631	0.084	0.008	0.000	0.052	0.057
		200	1	0.416	0.082	0.005	0.000	0.054	0.057
		400	1	0.275	0.090	0.007	0.000	0.054	0.059
	c = -5								
		50	0.900	0.956	0.142	0.055	0.000	0.000	0.052
		100	0.950	0.631	0.139	0.046	0.000	0.000	0.050
		200	0.975	0.416	0.142	0.045	0.000	0.000	0.050
		400	0.988	0.275	0.144	0.050	0.000	0.000	0.055
	c = -5								
	$\alpha = 0.9$	50	0.980	0.956	0.138	0.053	0.001	0.000	0.053
		100	0.990	0.631	0.135	0.048	0.000	0.000	0.052
		200	0.995	0.416	0.133	0.051	0.000	0.000	0.051
		400	0.998	0.275	0.139	0.054	0.000	0.000	0.052
$\gamma = 0.5$	$\widetilde{\mu} = 5$								
	c = 0	50	1	0.707	0.079	0.007	0.000	0.058	0.063
		100	1	0.500	0.070	0.007	0.000	0.053	0.057
		200	1	0.354	0.067	0.004	0.000	0.055	0.058
		400	1	0.250	0.073	0.005	0.000	0.051	0.059
	c = -5								
		50	0.900	0.707	0.123	0.055	0.000	0.000	0.050
		100	0.950	0.500	0.114	0.046	0.000	0.000	0.049
		200	0.975	0.354	0.105	0.046	0.000	0.000	0.050
		400	0.988	0.250	0.112	0.049	0.000	0.000	0.054
	c = -5								
	$\alpha = 0.9$	50	0.980	0.707	0.123	0.058	0.000	0.000	0.051
		100	0.990	0.500	0.117	0.049	0.000	0.000	0.052
		200	0.995	0.354	0.116	0.054	0.000	0.000	0.055
		400	0.998	0.250	0.115	0.054	0.000	0.000	0.052
$\gamma = 0.3$	$\widetilde{\mu} = 3$								
	c = 0	50	1	0.928	0.065	0.005	0.000	0.058	0.062
		100	1	0.754	0.057	0.006	0.000	0.053	0.058
		200	1	0.612	0.055	0.003	0.000	0.055	0.058
		400	1	0.497	0.061	0.003	0.000	0.051	0.059
	c = -5								
		50	0.900	0.928	0.083	0.055	0.000	0.000	0.051
		100	0.950	0.754	0.076	0.049	0.000	0.000	0.050
		200	0.975	0.612	0.065	0.047	0.000	0.000	0.052
		400	0.988	0.497	0.067	0.045	0.000	0.000	0.050
	c = -5								
	$\alpha = 0.9$	50	0.980	0.928	0.091	0.057	0.000	0.000	0.052
		100	0.990	0.754	0.080	0.047	0.000	0.000	0.051
		200	0.995	0.612	0.076	0.050	0.000	0.000	0.052
		400	0.998	0.497	0.073	0.050	0.000	0.000	0.055

Table 11: Power comparison between RLRT-PC and Bonf Q, true $\beta = \frac{25\sqrt{1-\phi^2}}{n}$, when the predictor x_t has an intercept, using confidence belt based critical values for the Q test.

			$\gamma = 0.6$ $\widetilde{\mu}$=10			$\gamma = 0.5$ $\widetilde{\mu}$=5			$\gamma = 0.3$ $\widetilde{\mu}$=3		
	ρ_n	n	μ	Bonf Q	PC	μ	Bonf Q	PC	μ	Bonf Q	PC
UR											
c = 0	1	50	0.956	0.999	1.000	0.707	0.988	1.000	0.928	0.999	1.000
	1	100	0.631	0.999	1.000	0.500	0.991	1.000	0.754	1.000	1.000
	1	200	0.416	0.998	1.000	0.354	0.990	1.000	0.612	1.000	1.000
	1	400	0.275	0.996	1.000	0.250	0.990	1.000	0.497	1.000	1.000
LUR											
c = -5	0.900	50	0.956	0.252	0.993	0.707	0.290	0.944	0.928	0.254	0.991
	0.950	100	0.631	0.160	0.979	0.500	0.183	0.944	0.754	0.154	0.991
	0.975	200	0.416	0.131	0.977	0.354	0.141	0.947	0.612	0.122	1.000
	0.988	400	0.275	0.113	0.977	0.250	0.120	0.947	0.497	0.106	1.000
MIUR											
c = -5	0.852	50	0.956	0.171	0.510	0.707	0.233	0.352	0.928	0.179	0.491
$\alpha = 0.9$	0.921	100	0.631	0.080	0.355	0.500	0.116	0.253	0.754	0.060	0.476
	0.958	200	0.416	0.048	0.248	0.354	0.066	0.206	0.612	0.025	0.466
	0.977	400	0.275	0.034	0.248	0.250	0.043	0.206	0.497	0.014	0.466

Table 12: Size and power comparison between IVX, RLRT-CD and RLRT-PC for testing $H_0 : \beta_1 = \beta_2 = 0$, for a predictor x_t with and without intercept and for a nominal 10% size.

	No intercept				Intercept			
	Size		Power		Size		Power	
C	IVX	CD	IVX	CD	IVX	PC	IVX	PC
0	0.185	0.144	1.000	1.000	0.150	0.150	1.000	1.000
-5	0.143	0.114	1.000	1.000	0.083	0.113	1.000	1.000
-10	0.129	0.113	1.000	1.000	0.080	0.115	1.000	1.000
-15	0.118	0.114	1.000	1.000	0.083	0.118	0.997	1.000
-20	0.110	0.113	0.998	0.999	0.085	0.117	0.989	1.000
-50	0.094	0.112	0.560	0.656	0.093	0.117	0.534	0.669

Table 13: Size and power comparison between Jansson and Moreira(2006) and RLRT-PC for testing $H_0 : \beta = 0$.True $\beta = \frac{b\sqrt{1-\phi^2}}{n}$, $n = 400$, and the predictor x_t has an intercept.

			b=0		b=5		b=10		b=15	
ϕ	c	ρ_n	PC	JM	PC	JM	PC	JM	PC	JM
-0.5	0	1.000	0.053	0.012	0.999	0.013	1.000	0.013	1.000	0.013
	-5	0.988	0.058	0.000	0.270	0.000	0.810	0.000	0.978	0.000
	-10	0.975	0.052	0.000	0.113	0.000	0.371	0.000	0.729	0.000
0.5	0	1.000	0.061	0.000	0.999	0.000	1.000	0.013	1.000	0.013
	-5	0.988	0.048	0.000	0.357	0.000	0.750	0.000	0.932	0.000
	-10	0.975	0.038	0.000	0.198	0.000	0.468	0.000	0.737	0.000

A Appendix A: Univariate predictor

This Appendix provides some basic limit theory that assists in deriving the limit distributions of the RLRT test statistics with a scalar predictor. We first present the limit behavior of standardized forms of the sample moments $\sum_{t=1}^{n} x_{t-1}$, $\sum_{t=1}^{n} x_{t-1}^2$, $\sum_{t=1}^{n} x_{t-1}u_{0.xt}$ and $\sum_{t=1}^{n} x_{t-1}u_{xt}$ for a univariate predictor x_t under a localized drift specification. These results are, in the main, direct applications of results in Phillips (1987a&b) and Phillips and Magdalinos (2007a&b) so we simply list them here. The results from Lemma A.1 to Lemma A.4 are derived assuming the random process x_t is generated by $x_t = \mu + \rho_n x_{t-1} + u_{xt}$ with $\mu = \widetilde{\mu} n^{-\gamma}$. Full details of the derivations are provided in the supplementary document to this paper (Phillips and Chen, 2014).

In the following section, we denote by $W_0(r)$, $W_x(r)$, and $W_{0.x}(r)$ the standard Brownian motions, corresponding to the functional limits standardized partial sums of the errors u_{0t}, u_{xt}, and $u_{0.xt}$, each with variance normalized to unity. At the same time, we have Brownian motions $B_0(r)$, $B_x(r)$, and $B_{0.x}(r)$ associated with the unstandardized errors. We define the linear diffusion $K_c(r) := \sigma_{xx} J_c(r) = \int_0^r e^{c(r-s)} dW_x(r)$, and the demeaned process $W_x^m(r) := W_x(r) - \int_0^1 W_x(r)\,dr$.

A.1 Preliminary Lemmas

Lemma A.1 *If $\rho_n = 1$, we obtain,*
(1)

$$\frac{1}{n^{3/2}}\sum_{t=1}^{n} x_{t-1} \implies \begin{cases} \int_0^1 B_x(r)dr & \text{if } \gamma > \frac{1}{2}, \\ \frac{\widetilde{\mu}}{2} + \int_0^1 B_x(r)dr & \text{if } \gamma = \frac{1}{2}, \end{cases}$$
$$\frac{1}{n^{2-\gamma}}\sum_{t=1}^{n} x_{t-1} \to^p \frac{\widetilde{\mu}}{2} \text{ if } \gamma < \frac{1}{2}.$$

(2)

$$\frac{1}{n^2}\sum_{t=1}^{n} x_{t-1}^2 \implies \begin{cases} \int_0^1 B_x^2(r)dr & \text{if } \gamma > \frac{1}{2}, \\ \frac{\widetilde{\mu}^2}{3} + 2\widetilde{\mu}\int_0^1 rB_x(r)dr + \int_0^1 B_x^2(r)dr & \text{if } \gamma = \frac{1}{2}, \end{cases}$$
$$\frac{1}{n^{3-2\gamma}}\sum_{t=1}^{n} x_{t-1}^2 \to^p \frac{\widetilde{\mu}^2}{3} \text{ if } \gamma < \frac{1}{2}.$$

(3)

$$\frac{1}{n}\sum_{t=1}^{n} x_{t-1}u_{0.xt} \implies \begin{cases} \int_0^1 B_x(r)dB_{0.x}(r) & \text{if } \gamma > \frac{1}{2}, \\ \widetilde{\mu}\left\{B_{0.x}(1) - \int_0^1 B_{0.x}(r)dr\right\} + \int_0^1 B_x(r)dB_{0.x}(r) & \text{if } \gamma = \frac{1}{2}, \end{cases}$$

$$\frac{1}{n^{3/2-\gamma}}\sum_{t=1}^{n} x_{t-1}u_{0.xt} \implies \widetilde{\mu}\left\{B_{0.x}(r) - \int_0^1 B_{0.x}(r)dr\right\} \text{ if } \gamma < \frac{1}{2}.$$

(4) Replacing $B_{0.x}(r)$ with $B_x(r)$ in (3), the corresponding asymptotic results for $\sum_{t=1}^{n} x_{t-1}u_{xt}$ follow.

Lemma A.2 *If $\rho_n = 1 + \frac{c}{n}$, then we have:*

(1)

$$\frac{1}{n^{3/2}}\sum_{t=1}^{n} x_{t-1} \implies \begin{cases} \int_0^1 K_c(r)dr & \text{if } \gamma > \frac{1}{2}, \\ \widetilde{\mu}\int_0^1 F_c(r)\,dr + \int_0^1 K_c(r)dr & \text{if } \gamma = \frac{1}{2}, \end{cases}$$

$$\frac{1}{n^{2-\gamma}}\sum_{t=1}^{n} x_{t-1} \to^{p} \widetilde{\mu}\int_0^1 F_c(r)\,dr \text{ if } \gamma < \frac{1}{2}.$$

(2)

$$\frac{1}{n^{2}}\sum_{t=1}^{n} x_{t-1}^2 \implies \begin{cases} \int_0^1 K_c(r)^2dr & \text{if } \gamma > \frac{1}{2}, \\ \widetilde{\mu}^2\int_0^1 F_c^2(r)\,dr + 2\widetilde{\mu}\int_0^1 F_c(r)\,K_c(r)dr + \int_0^1 K_c(r)^2dr & \text{if } \gamma = \frac{1}{2}, \end{cases}$$

$$\frac{1}{n^{3-2\gamma}}\sum_{t=1}^{n} x_{t-1}^2 \to^{p} \widetilde{\mu}^2\int_0^1 F_c^2(r)\,dr \text{ if } \gamma < \frac{1}{2}.$$

(3)

$$\frac{1}{n}\sum_{t=1}^{n} x_{t-1}u_{0.xt} \implies \begin{cases} \int_0^1 K_c(r)dB_{0.x}(r) & \text{if } \gamma > \frac{1}{2}, \\ \widetilde{\mu}\int_0^1 F_c(r)\,dB_{0.x}(r) + \int_0^1 K_c(r)dB_{0.x}(r) & \text{if } \gamma = \frac{1}{2}, \end{cases}$$

$$\frac{1}{n^{3/2-\gamma}}\sum_{t=1}^{n} x_{t-1}u_{0.xt} \implies \widetilde{\mu}\int_0^1 F_c(r)\,dB_{0.x}(r) \text{ if } \gamma < \frac{1}{2}.$$

(4) The corresponding asymptotic results for the sample covariance $\sum_{t=1}^{n} x_{t-1}u_{xt}$ follow by replacing $B_{0.x}(r)$ with $B_x(r)$ in (3)

Lemma A.3 *If $\rho_n = 1 + \frac{c}{k_n}$ with $c < 0$, and $k_n = n^{\alpha}$ with $\alpha \in (0,1)$, then for some $\delta = \delta(\alpha,\gamma)$ with $0 < \delta(\alpha,\gamma) \leqslant \frac{1}{2}$:*

(1) For $\gamma \geqslant \frac{1}{2}$ *or* $0 < \frac{\alpha}{2} < \gamma < \frac{1}{2}$,
$(i)\frac{1}{\sqrt{n}}\widetilde{x}_n \Rightarrow 0$; $(ii)\frac{1}{n}x_n^2 = O_p\left(n^{-\delta}\right)$; $(iii)\frac{1}{n^{\alpha/2+1}}\sum_{t=1}^{n}\widetilde{x}_{t-1} = O_p\left(n^{-\delta}\right)$.
(2) For $0 < \gamma \leqslant \frac{\alpha}{2} < \frac{1}{2}$,
$(i)\frac{1}{n^{\alpha/2-\gamma+1/2}}\widetilde{x}_n \Rightarrow 0$; $(ii)\frac{1}{n^{\alpha-2\gamma+1}}x_n^2 = O_p\left(n^{-\delta}\right)$; $(iii)\frac{1}{n^{\alpha-\gamma+1}}\sum_{t=1}^{n}\widetilde{x}_{t-1} = O_p\left(n^{-\delta}\right)$.

Lemma A.4 *If* $\rho_n = 1 + \frac{c}{k_n}$ *with* $c < 0$, $k_n = n^\alpha$ *with* $\alpha \in (0,1)$, *then for some* $\delta = \delta(\alpha,\gamma)$ *with* $0 < \delta(\alpha,\gamma) \leqslant \frac{1}{2}$*:*

(1) If $\gamma \geqslant \frac{1}{2}$ *or* $0 < \frac{\alpha}{2} < \gamma < \frac{1}{2}$,

$$
\begin{aligned}
(i)\frac{1}{n^{\alpha/2+1}}\sum_{t=1}^{n}x_{t-1} &= O_p\left(n^{-\delta}\right),\\
(ii)\frac{1}{n^{\alpha+1}}\sum_{t=1}^{n}x_{t-1}^2 &= \frac{\sigma_{xx}^2}{-2c} + O_p\left(n^{-\delta}\right),\\
(iii)\frac{1}{n^{\alpha/2+1/2}}\sum_{t=1}^{n}x_{t-1}u_{0.xt} &\Longrightarrow N\left(0,\frac{\sigma_{xx}^2\sigma_{00.x}^2}{-2c}\right),\\
(iv)\frac{1}{n^{\alpha/2+1/2}}\sum_{t=1}^{n}x_{t-1}u_{xt} &\Longrightarrow N\left(0,\frac{\sigma_{xx}^4}{-2c}\right).
\end{aligned}
$$

(2) If $0 < \gamma \leqslant \frac{\alpha}{2} < \frac{1}{2}$,

$$
\begin{aligned}
(i)\frac{1}{n^{\alpha-\gamma+1}}\sum_{t=1}^{n}x_{t-1} &= \frac{\widetilde{\mu}}{-c} + O_p\left(n^{-\delta}\right),\\
(ii)\frac{1}{n^{2\alpha-2\gamma+1}}\sum_{t=1}^{n}x_{t-1}^2 &= \frac{\widetilde{\mu}^2}{c^2} + O_p\left(n^{-\delta}\right),\\
(iii)\frac{1}{n^{\alpha-\gamma+1/2}}\sum_{t=1}^{n}x_{t-1}u_{0.xt} &\Longrightarrow \frac{\widetilde{\mu}}{-c}B_{0.x}(1),\\
(iv)\frac{1}{n^{\alpha-\gamma+1/2}}\sum_{t=1}^{n}x_{t-1}u_{xt} &\Longrightarrow \frac{\widetilde{\mu}}{-c}B_x(1).
\end{aligned}
$$

A.2 Proofs of the Main Results

A.2.1 Proof of Theorem 2.1

Proof. Under CD's estimation model (2.2), the unnormalized score function and second order partial derivatives are as follows:

$$
s_1 := \frac{1}{\sigma_{00.x}^2}\left(\sum_{t=1}^{n}x_{t-1}u_{0.xt} - \frac{1}{n}\sum_{t=1}^{n}x_{t-1}\sum_{t=1}^{n}u_{0.xt}\right),
$$

$$
\begin{aligned}
s_2 := & \quad -\phi s_1 + \frac{1}{\sigma_{xx}^2} \sum_{t=1}^{n} \left(x_t - \rho_n x_{t-1} \right) x_{t-1}, \\
h^{11} := & \quad -\phi^2 \sigma_{xx}^2 \left(\sum_{t=1}^{n} x_{t-1}^2 \right)^{-1} - \sigma_{00.x}^2 \left\{ \sum_{t=1}^{n} x_{t-1}^2 - \frac{1}{n} \left(\sum_{t=1}^{n} x_{t-1} \right)^2 \right\}^{-1}, \\
h^{22} := & \quad -\sigma_{xx}^2 \left(\sum_{t=1}^{n} x_{t-1}^2 \right)^{-1}, \\
h_{22}^{-1} := & \quad \left\{ \left(-\frac{\phi^2}{\sigma_{00.x}^2} - \frac{1}{\sigma_{xx}^2} \right) \sum_{t=1}^{n} x_{t-1}^2 + \frac{\phi^2}{n \sigma_{00.x}^2} \left(\sum_{t=1}^{n} x_{t-1} \right)^2 \right\}^{-1}, \\
h^{12} := & \quad \phi h^{22} = h^{21}.
\end{aligned}
$$

Let $\lambda^2 = \frac{\phi^2 \sigma_{xx}^2}{\left(\phi^2 \sigma_{xx}^2 + \sigma_{00.x}^2 \right)} = \frac{\phi^2 \sigma_{xx}^2}{\sigma_{00}^2}$, and given the expressions for s_i and h_{ij}, we have

$$
\begin{aligned}
R_n &= -h^{11} s_1^2 - 2h^{12} s_1 s_2 - \left(h^{22} - h_{22}^{-1} \right) s_2^2 + O_p \left(n^{-1/2} \right) \\
&= \left\{ \sqrt{1 - \lambda^2 g^n} p^n + \lambda \left(g^n \right)^{1/2} \tau^n \right\}^2 + O_p \left(n^{-1/2} \right),
\end{aligned}
$$

with

$$
\begin{aligned}
p^n &= \frac{\left(\sum_{t=1}^n x_{t-1} u_{0.xt} - \frac{1}{n} \sum_{t=1}^n x_{t-1} \sum_{t=1}^n u_{0.xt} \right)}{\left\{ \sum_{t=1}^n x_{t-1}^2 - \frac{1}{n} \left(\sum_{t=1}^n x_{t-1} \right)^2 \right\}^{1/2}} \frac{1}{\sigma_{00.x}}, \\
g^n &= \frac{1 - \frac{1}{n} \left(\sum_{t=1}^n x_{t-1} \right)^2 / \sum_{t=1}^n x_{t-1}^2}{1 - \lambda^2 \frac{1}{n} \left(\sum_{t=1}^n x_{t-1} \right)^2 / \sum_{t=1}^n x_{t-1}^2}, \\
\tau^n &= \frac{\sum_{t=1}^n x_{t-1} \left(x_t - \rho_n x_{t-1} \right)}{\left(\sum_{t=1}^n x_{t-1}^2 \right)^{1/2}} \frac{1}{\sigma_{xx}} = \frac{\mu \sum_{t=1}^n x_{t-1} + \sum_{t=1}^n x_{t-1} u_{xt}}{\left(\sum_{t=1}^n x_{t-1}^2 \right)^{1/2}} \frac{1}{\sigma_{xx}}.
\end{aligned}
$$

After normalization for each component in p^n, g^n and τ^n, we obtain the limit distribution of R_n for different parameter scenarios using the preliminary lemmas above. Specifically, we consider the following cases:

(1) If $\rho_n = 1$ we have the following.

(i) When $\gamma > \frac{1}{2}$,

$$
p^n \implies \frac{\int_0^1 W_x^m (r) \, dW_{0.x} (r)}{\left\{ \int_0^1 \left[W_x^m (r) \right]^2 dr \right\}^{1/2}} = Z,
$$

$$g^n \implies g_\lambda = \frac{1 - \frac{\left(\int_0^1 W_x(r)dr\right)^2}{\int_0^1 W_x^2(r)dr}}{1 - \lambda^2 \frac{\left(\int_0^1 W_x(r)dr\right)^2}{\int_0^1 W_x^2(r)dr}},$$

$$\tau^n \implies \tau = \frac{\int_0^1 W_x(r)\, dW_x(r)}{\sqrt{\int_0^1 W_x^2(r)\, dr}}.$$

(ii) When $\gamma = \frac{1}{2}$,

$$p^n \implies p_{\lambda,\widetilde{\mu}} = \frac{\frac{\widetilde{\mu}}{2\sigma_{xx}} W_{0.x}(1) - \frac{\widetilde{\mu}}{\sigma_{xx}} \int_0^1 W_{0.x}(r)\, dr + \int_0^1 W_x^m(r)\, dW_{0.x}(r)}{\left\{\frac{\widetilde{\mu}^2}{12\sigma_{xx}^2} + \frac{\widetilde{\mu}}{\sigma_{xx}}\left\{2\int_0^1 rW_x(r)\, dr - \int_0^1 W_x(r)\, dr\right\} + \int_0^1 \{W_x^m(r)\}^2 dr\right\}^{1/2}},$$

$$g^n \implies g_{\lambda,\widetilde{\mu}} = \frac{1 - \frac{\frac{\widetilde{\mu}^2}{4\sigma_{xx}^2} + \frac{\widetilde{\mu}}{\sigma_{xx}} \int_0^1 W_x(r)dr + \left(\int_0^1 W_x(r)dr\right)^2}{\frac{\widetilde{\mu}^2}{3\sigma_{xx}^2} + \frac{2\widetilde{\mu}}{\sigma_{xx}} \int_0^1 rW_x(r)dr + \int_0^1 W_x^2(r)dr}}{1 - \lambda^2 \frac{\frac{\widetilde{\mu}^2}{4\sigma_{xx}^2} + \frac{\widetilde{\mu}}{\sigma_{xx}} \int_0^1 W_x(r)dr + \left(\int_0^1 W_x(r)dr\right)^2}{\frac{\widetilde{\mu}^2}{3\sigma_{xx}^2} + \frac{2\widetilde{\mu}}{\sigma_{xx}} \int_0^1 rW_x(r)dr + \int_0^1 W_x^2(r)dr}},$$

$$\tau^n \implies \tau_{\lambda,\widetilde{\mu}} = \frac{\frac{\widetilde{\mu}^2}{2\sigma_{xx}^2} + \frac{\widetilde{\mu}}{\sigma_{xx}} W_x(1) + \int_0^1 W_x(r)\, dW_x(r)}{\left\{\frac{\widetilde{\mu}^2}{3\sigma_{xx}^2} + \frac{2\widetilde{\mu}}{\sigma_{xx}} \int_0^1 rW_x(r)\, dr + \int_0^1 W_x^2(r)\, dr\right\}^{1/2}}.$$

(iii) When $\gamma < \frac{1}{2}$,

$$p^n \implies p = \frac{\frac{\widetilde{\mu}}{2} W_{0.x}(1) - \widetilde{\mu} \int_0^1 W_{0.x}(r)\, dr}{\left(\frac{\widetilde{\mu}^2}{12}\right)^{1/2}} = \sqrt{3}\left(W_{0.x}(1) - 2\int_0^1 W_{0.x}(r)\, dr\right),$$

$$g^n \implies g_\lambda = \frac{1 - \frac{\frac{\widetilde{\mu}^2}{4}}{\frac{\widetilde{\mu}^2}{3}}}{1 - \lambda^2 \frac{\frac{\widetilde{\mu}^2}{4}}{\frac{\widetilde{\mu}^2}{3}}} = \frac{1}{4 - 3\lambda^2},$$

$$\begin{aligned}\tau^n \implies \tau_{\lambda,\widetilde{\mu},\gamma} &= \frac{\frac{\widetilde{\mu}^2 n^{1/2-\gamma}}{2\sigma_{xx}} + \widetilde{\mu}\left(W_x(1) - \int_0^1 W_x(r)\, dr\right)}{\left(\frac{\widetilde{\mu}^2}{3}\right)^{1/2}} \\ &= \sqrt{3}\left(\frac{\widetilde{\mu}}{2\sigma_{xx}} n^{1/2-\gamma} + W_x(1) - \int_0^1 W_x(r)\, dr\right).\end{aligned}$$

Hence,

$$\begin{aligned} R_n &= \left\{ \begin{array}{c} \sqrt{3\left(\frac{4-4\lambda^2}{4-3\lambda^2}\right)} \left(W_{0.x}(1) - 2\int_0^1 W_{0.x}(r)\,dr\right) + \\ \sqrt{3}\lambda\left(\frac{1}{4-3\lambda^2}\right)^{1/2}\left(\frac{\widetilde{\mu}}{2\sigma_{xx}} n^{1/2-\gamma} + W_x(1) - \int_0^1 W_x(r)\,dr\right) \end{array} \right\}^2 \\ &= O_p\left(n^{1-2\gamma}\right). \end{aligned}$$

(2) For $\rho_n = 1 + \frac{c}{n}$, the following results hold.

(i) When $\gamma > \frac{1}{2}$,

$$\begin{aligned} p^n &\Longrightarrow \frac{\int_0^1 J_c^m(r)\,dW_{0.x}(r)}{\left\{\int_0^1 \left[J_c^m(r)\right]^2 dr\right\}^{1/2}} = Z, \\ g^n &\Longrightarrow g_\lambda = \frac{1 - \frac{\left(\int_0^1 J_c(r)dr\right)^2}{\int_0^1 J_c(r)^2 dr}}{1 - \lambda^2 \frac{\left(\int_0^1 J_c(r)dr\right)^2}{\int_0^1 J_c(r)^2 dr}}, \\ \tau^n &\Longrightarrow \tau = \frac{\int_0^1 J_c(r)\,dW_x(r)}{\sqrt{\int_0^1 J_c(r)^2\,dr}}. \end{aligned}$$

(ii) When $\gamma = \frac{1}{2}$,

$$\begin{aligned} p^n &\Longrightarrow p_{c,\lambda,\widetilde{\mu}} = \frac{\frac{\widetilde{\mu}}{\sigma_{xx}}\int_0^1 F_c^m(r)\,dW_{0.x}(r) + \int_0^1 J_c^m(r)\,dW_{0.x}(r)}{\left\{\int_0^1 \left\{\frac{\widetilde{\mu}}{\sigma_{xx}} F_c^m(r) + J_c^m(r)\right\}^2 dr\right\}^{1/2}}, \\ g^n &\Longrightarrow g_{c,\lambda,\widetilde{\mu}} = \frac{1 - \frac{\frac{\widetilde{\mu}^2}{\sigma_{xx}^2}\left\{\int_0^1 F_c(r)dr\right\}^2 + \frac{2\widetilde{\mu}}{\sigma_{xx}}\int_0^1 F_c(r)dr\int_0^1 J_c(r)dr + \left(\int_0^1 J_c(r)dr\right)^2}{\frac{\widetilde{\mu}^2}{\sigma_{xx}^2}\int_0^1 F_c^2(r)dr + \frac{2\widetilde{\mu}}{\sigma_{xx}}\int_0^1 F_c(r)J_c(r)dr + \int_0^1 J_c(r)^2 dr}}{1 - \lambda^2\frac{\frac{\widetilde{\mu}^2}{\sigma_{xx}^2}\left\{\int_0^1 F_c(r)dr\right\}^2 + \frac{2\widetilde{\mu}}{\sigma_{xx}}\int_0^1 F_c(r)dr\int_0^1 J_c(r)dr + \left(\int_0^1 J_c(r)dr\right)^2}{\frac{\widetilde{\mu}^2}{\sigma_{xx}^2}\int_0^1 F_c^2(r)dr + \frac{2\widetilde{\mu}}{\sigma_{xx}}\int_0^1 F_c(r)J_c(r)dr + \int_0^1 J_c(r)^2 dr}}, \\ \tau^n &\Longrightarrow \tau_{c,\lambda,\widetilde{\mu}} = \frac{\frac{\widetilde{\mu}^2}{\sigma_{xx}^2}\int_0^1 F_c(r)\,dr + \frac{\widetilde{\mu}}{\sigma_{xx}^2}\int_0^1 J_c(r)\,dr + \frac{\widetilde{\mu}}{\sigma_{xx}}\int_0^1 F_c(r)\,dW_x(r) + \int_0^1 J_c(r)\,dW_x(r)}{\left\{\frac{\widetilde{\mu}^2}{\sigma_{xx}^2}\int_0^1 F_c^2(r)\,dr + \frac{2\widetilde{\mu}}{\sigma_{xx}}\int_0^1 F_c(r)\,J_c(r)dr + \int_0^1 J_c(r)^2 dr\right\}^{1/2}}. \end{aligned}$$

(iii) When $\gamma < \frac{1}{2}$

$$p^n \Longrightarrow p_{c,\widetilde{\mu}} = \frac{\int_0^1 F_c^m(r)\,dW_{0.x}(r)}{\left\{\int_0^1 \left\{F_c^m(r)\right\}^2 dr\right\}^{1/2}},$$

$$
\begin{aligned}
g^n &\implies g_{c,\lambda} = \frac{1-\frac{\left\{\int_0^1 F_c(r)dr\right\}^2}{\int_0^1 F_c^2(r)dr}}{1-\lambda^2\frac{\left\{\int_0^1 F_c(r)dr\right\}^2}{\int_0^1 F_c^2(r)dr}}, \\
\tau^n &\implies \tau_{c,\widetilde{\mu},\lambda,\gamma} = \frac{\frac{\widetilde{\mu}n^{1/2-\gamma}}{\sigma_{xx}}\int_0^1 F_c\left(r\right)dr + \int_0^1 F_c\left(r\right)dW_x\left(r\right)}{\left\{\int_0^1 F_c^2\left(r\right)dr\right\}^{1/2}}.
\end{aligned}
$$

(3) If $\rho_n = 1 + \frac{c}{k_n}$ with $c < 0, k_n = n^{\alpha}$ with $\alpha \in (0,1)$, we have the following results.

(i) When $\gamma \geqslant \frac{1}{2}$ and $0 < \frac{\alpha}{2} < \gamma < \frac{1}{2}$,

$$
\begin{aligned}
p^n &\implies \frac{N\left(0, \frac{\sigma_{xx}^2\sigma_{00.x}^2}{-2c}\right)}{\sqrt{\frac{\sigma_{xx}^2}{-2c}}}\frac{1}{\sigma_{00.x}} = N\left(0,1\right), \\
g^n &\implies 1 + O_p\left(n^{-\delta}\right), \\
\tau^n &\implies \frac{N\left(0, \frac{\sigma_{xx}^4}{-2c}\right)}{\sqrt{\frac{\sigma_{xx}^2}{-2c}}}\frac{1}{\sigma_{xx}} = N\left(0,1\right).
\end{aligned}
$$

Hence,

$$
\begin{aligned}
R_n &= \left\{N\left(0,1\right)\left(1-\lambda^2\right)^{1/2} + \lambda N\left(0,1\right)\right\}^2 + O_p\left(n^{-\delta}\right) \\
&= \chi_1^2 + O_p\left(n^{-\delta}\right).
\end{aligned}
$$

(ii) When $0 < \gamma \leqslant \frac{\alpha}{2} < \frac{1}{2}$,

$$
\begin{aligned}
p^n &= O_p\left(n^{-\delta}\right), \\
g^n &= \frac{1}{2-\lambda^2} + O_p\left(n^{-\delta}\right), \\
\tau^n &\sim \frac{\widetilde{\mu}n^{1/2-\gamma}\frac{\widetilde{\mu}}{\sigma_{xx}c} + \frac{\widetilde{\mu}}{c}W_x\left(1\right)}{\left(\frac{2\widetilde{\mu}^2}{c^2}\right)^{1/2}} + O_p\left(n^{-\delta}\right) = \frac{\left\{\frac{\widetilde{\mu}n^{1/2-\gamma}}{\sigma_{xx}} + W_x\left(1\right)\right\}}{\sqrt{2}} + O_p\left(n^{-\delta}\right).
\end{aligned}
$$

Hence

$$
R_n \sim \left\{\lambda\frac{\left\{\frac{\widetilde{\mu}n^{1/2-\gamma}}{\sigma_{xx}} + W_x\left(1\right)\right\}}{\sqrt{2}}\left(\frac{1}{2-\lambda^2}\right)^{1/2}\right\}^2
$$

$$= \frac{\lambda^2}{2\left(2-\lambda^2\right)}\left(\left\{\frac{\widetilde{\mu}n^{1/2-\gamma}}{\sigma_{xx}}+W_x\left(1\right)\right\}\right)^2.$$

■

A.2.2 Proof of Lemma 2.2

Proof. For the DGP $x_t = \mu + \rho_n x_{t-1} + u_{x,t}$ with constant x_0, by backward iteration, we have

$$x_t = \mu \sum_{i=0}^{t-1} \rho_n^i + \rho_n^t x_0 + \sum_{i=0}^{t-1} \rho_n^i u_{x,t-i}.$$

Hence, the associated restricted likelihood function is

$$L\left(\Theta, Y, X\right) = -\frac{n-1}{2}\log - \frac{1}{2\sigma_{00.x}^2} S\left(\beta, \rho_n, \phi\right) - \frac{n}{2}\log\sigma_{xx}^2 + \frac{1}{2}P\left(\rho_n\right) - \frac{\sigma_{xx}^2}{2}Q\left(\rho_n\right),$$

where

$$\begin{aligned}
S\left(\phi, \beta, \rho_n\right) &= \left[\underline{Y} - \phi\underline{X_t} - \left(\beta - \phi\rho_n\right)\underline{X_{t-1}}\right]'\left[\underline{Y} - \phi\underline{X_t} - \left(\beta - \phi\rho_n\right)\underline{X_{t-1}}\right], \\
P\left(\rho_n\right) &= \log|Z'Z|, \\
Q\left(\rho_n\right) &= \left(X - Z\widehat{\tau}\right)'\Sigma^{-1}\left(X - Z\widehat{\tau}\right),
\end{aligned}$$

with $\widehat{\tau} = \left(Z'\Sigma^{-1}Z\right)^{-1}\left(Z'\Sigma^{-1}X\right)$ such that $\Sigma = var\left(X\right) = \sigma_{xx}^2\left(B'B\right)^{-1}$ where

$$B = \begin{bmatrix} 1 & 0 & 0 & \ldots & 0 & 0 \\ -\rho_n & 1 & 0 & \ldots & 0 & 0 \\ 0 & -\rho_n & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & & & & \vdots \\ 0 & 0 & 0 & \ldots & 1 & 0 \\ 0 & 0 & 0 & \ldots & -\rho_n & 1 \end{bmatrix},$$

and Z is the associated regressor matrix

$$Z' = \begin{pmatrix} 1 & \rho_n^1 & \rho_n^2 & \cdots & \rho_n^n \\ 0 & 1 & 1+\rho_n^1 & \cdots & 1+\rho_n^1+\ldots+\rho_n^{n-1} \end{pmatrix}.$$

■

A.2.3 Proof of Theorem 2.3

Proof. From the REML under the estimation model (2.8), the unormalized score functions and its second derivatives are:

$$
\begin{aligned}
s_1 :=&\ \frac{1}{\sigma_{00.x}^2}\left(\sum_{t=1}^n x_{t-1}u_{0.xt} - \frac{1}{n}\sum_{t=1}^n x_{t-1}\sum_{t=1}^n u_{0.xt}\right),\\
s_2 :=&\ -\phi s_1 + \frac{1}{\sigma_{xx}^2}\left(\sum_{t=1}^n x_{t-1}u_{xt} - \frac{1}{n}\sum_{t=1}^n x_{t-1}\sum_{t=1}^n u_{xt}\right),\\
h^{11} :=&\ -\left(\phi^2\sigma_{xx}^2 + \sigma_{00.x}^2\right)\left(\sum_{t=1}^n x_{t-1}^2 - \frac{1}{n}\left(\sum_{t=1}^n x_{t-1}\right)^2\right)^{-1},\\
h^{22} :=&\ -\sigma_{xx}^2\left(\sum_{t=1}^n x_{t-1}^2 - \frac{1}{n}\left(\sum_{t=1}^n x_{t-1}\right)^2\right)^{-1},\\
h_{22}^{-1} :=&\ -\frac{\left(\phi^2\sigma_{xx}^2 + \sigma_{00.x}^2\right)}{\sigma_{00.x}^2\sigma_{xx}^2}\left(\sum_{t=1}^n x_{t-1}^2 - \frac{1}{n}\left(\sum_{t=1}^n x_{t-1}\right)^2\right)^{-1},\\
h^{12} :=&\ \phi h^{22} = h^{21}.
\end{aligned}
$$

After some calculation, we find that

$$R_n = \left\{\sigma_{00.x}p^n + \phi\sigma_{xx}q^n\right\}^2\left(\phi^2\sigma_{xx}^2 + \sigma_{00.x}^2\right)^{-1} + O_p\left(n^{-1/2}\right), \tag{A.1}$$

with

$$p^n = \frac{\left(\sum_{t=1}^n x_{t-1}u_{0.xt} - \frac{1}{n}\sum_{t=1}^n x_{t-1}\sum_{t=1}^n u_{0.xt}\right)}{\left\{\sum_{t=1}^n x_{t-1}^2 - \frac{1}{n}\left(\sum_{t=1}^n x_{t-1}\right)^2\right\}^{1/2}}\frac{1}{\sigma_{00.x}}, \tag{A.2}$$

$$q^n = \frac{\left(\sum_{t=1}^n x_{t-1}u_{xt} - \frac{1}{n}\sum_{t=1}^n x_{t-1}\sum_{t=1}^n u_{xt}\right)}{\left\{\sum_{t=1}^n x_{t-1}^2 - \frac{1}{n}\left(\sum_{t=1}^n x_{t-1}\right)^2\right\}^{1/2}}\frac{1}{\sigma_{xx}}. \tag{A.3}$$

We consider the following cases of ρ_n in turn. (1) $\rho_n = 1$.
(i) When $\gamma > \frac{1}{2}$,

$$q^n \Longrightarrow q = \frac{\int_0^1 W_x^m(r)\,dW_x(r)}{\left\{\int_0^1 \left[W_x^m(r)\right]^2 dr\right\}^{1/2}}.$$

Hence

$$R_n \Longrightarrow \frac{\int_0^1 W_x^m(r)\, d\left\{\sigma_{00.x} W_{0.x}(r) + \phi \sigma_{xx} W_x(r)\right\}}{\left\{\int_0^1 \left[W_x^m(r)\right]^2 dr\right\}^{1/2}} = \frac{\int_0^1 W_x^m(r)\, dW_0(r)}{\left\{\int_0^1 \left[W_x^m(r)\right]^2 dr\right\}^{1/2}}.$$

(ii) When $\gamma = \frac{1}{2}$,

$$q^n \Longrightarrow q_{\widetilde{\mu}} = \frac{\frac{\widetilde{\mu}}{2\sigma_{xx}} W_x(1) - \frac{\widetilde{\mu}}{\sigma_{xx}} \int_0^1 W_x(r)\, dr + \int_0^1 W_x^m(r)\, dW_x(r)}{\left\{\frac{\widetilde{\mu}^2}{12\sigma_{xx}^2} + \frac{\widetilde{\mu}}{\sigma_{xx}}\left\{2\int_0^1 r W_x(r)\, dr - \int_0^1 W_x(r)\, dr\right\} + \int_0^1 \left\{W_x^m(r)\right\}^2 dr\right\}^{1/2}}.$$

Hence,

$$R_n \Longrightarrow \frac{\frac{\widetilde{\mu}}{2\sigma_{xx}} W_0(1) - \frac{\widetilde{\mu}}{\sigma_{xx}} \int_0^1 W_0(r)\, dr + \int_0^1 W_x^m(r)\, dW_0(r)}{\left\{\frac{\widetilde{\mu}^2}{12\sigma_{xx}^2} + \frac{\widetilde{\mu}}{\sigma_{xx}}\left\{2\int_0^1 r W_x(r)\, dr - \int_0^1 W_x(r)\, dr\right\} + \int_0^1 \left\{W_x^m(r)\right\}^2 dr\right\}^{1/2}}.$$

(iii) When $\gamma < \frac{1}{2}$,

$$q^n \Longrightarrow \frac{\frac{\widetilde{\mu}}{2} W_x(1) - \widetilde{\mu} \int_0^1 W_x(r)\, dr}{\left(\frac{\widetilde{\mu}^2}{12}\right)^{1/2}} = \sqrt{3}\left(W_x(1) - 2\int_0^1 W_x(r)\, dr\right).$$

Hence,

$$\begin{aligned} R_n &\Longrightarrow 3\left\{B_{0.x}(1) + \phi B_x(1) - 2\int_0^1 \left(B_{0.x}(r) + \phi B_x(r)\right) dr\right\}^2 \left(\phi^2 \sigma_{xx}^2 + \sigma_{00.x}^2\right)^{-1} \\ &= 3\left\{W_0(1) - 2\int_0^1 W_0(r) dr\right\}^2. \end{aligned}$$

(2) $\rho_n = 1 + \frac{c}{n}$.

(i) When $\gamma > \frac{1}{2}$, $q^n \Longrightarrow \int_0^1 J_c^m(r)\, dW_x(r) / \left\{\int_0^1 \left\{J_c^m(r)\right\}^2 dr\right\}^{1/2}$.

Hence, $R_n \Longrightarrow \int_0^1 J_c^m(r)\, dW_0(r) / \left\{\int_0^1 \left\{J_c^m(r)\right\}^2 dr\right\}^{1/2}$.

(ii) When $\gamma = \frac{1}{2}$,

$$q^n \Longrightarrow q_{c,\lambda,\widetilde{\mu}} = \frac{\frac{\widetilde{\mu}}{\sigma_{xx}} \int_0^1 F_c^m(r)\, dW_x(r) + \int_0^1 J_c^m(r)\, dW_x(r)}{\left\{\int_0^1 \left\{\frac{\widetilde{\mu}}{\sigma_{xx}} F_c^m(r) + J_c^m(r)\right\}^2 dr\right\}^{1/2}},$$

and then

$$R_n \Longrightarrow \frac{\left\{\frac{\widetilde{\mu}}{\sigma_{xx}}\int_0^1 F_c^m(r)\,dW_0(r)+\int_0^1 J_c^m(r)\,dW_0(r)\right\}^2}{\int_0^1\left\{\frac{\widetilde{\mu}}{\sigma_{xx}}F_c^m(r)+J_c^m(r)\right\}^2 dr}$$

(iii) When $\gamma<\frac{1}{2}$, $q^n \Longrightarrow \int_0^1 F_c^m(r)\,dW_x(r)\,/\left\{\int_0^1\{F_c^m(r)\}^2 dr\right\}^{1/2}$.

Hence, $R_n \Longrightarrow \int_0^1 F_c^m(r)\,dW_0(r)\,/\left\{\int_0^1\{F_c^m(r)\}^2 dr\right\}^{1/2}$

(3) $\rho_n = 1+\frac{c}{k_n}$ with $c<0, k_n=n^\alpha$ with $\alpha\in(0,1)$.

(i) When $\gamma\geqslant\frac{1}{2}$ and $0<\frac{\alpha}{2}<\gamma<\frac{1}{2}$,

$$q^n \Longrightarrow \frac{N\left(0,\frac{\sigma_{xx}^4}{-2c}\right)}{\sqrt{\frac{\sigma_{xx}^2}{-2c}}} = N\left(0,\sigma_{xx}^2\right),$$

and then

$$R_n \Longrightarrow \left\{N\left(0,\sigma_{00.x}^2\right)+\phi N\left(0,\sigma_{xx}^2\right)\right\}^2\left(\phi^2\sigma_{xx}^2+\sigma_{00.x}^2\right)^{-1}=\chi_1^2.$$

(ii) When $0<\gamma\leqslant\frac{\alpha}{2}<\frac{1}{2}$, both the numerator and denominator go to zero and higher order terms need to be considered. In this case, we have

$$p^n \sim \frac{n^{-\frac{\alpha}{2}+\gamma}N\left(0,\frac{\sigma_{xx}^2\sigma_{00.x}^2}{-2c}\right)+n^{\gamma-\frac{1}{2}}\frac{B_x(1)B_{0.x}(1)}{c}}{\left(n^{-\alpha+2\gamma}\frac{\sigma_{xx}^2}{-2c}-n^{2\gamma-1}\frac{\{B_x(1)\}^2}{c^2}\right)^{1/2}} \sim \frac{N\left(0,\frac{\sigma_{xx}^2\sigma_{00.x}^2}{-2c}\right)+o_p(1)}{\left(\frac{\sigma_{xx}^2}{-2c}+o_p(1)\right)^{1/2}} \Longrightarrow N\left(0,\sigma_{00.x}^2\right),$$

$$q^n \sim \frac{n^{-\frac{\alpha}{2}+\gamma}N\left(0,\frac{\sigma_{xx}^4}{-2c}\right)+n^{\gamma-\frac{1}{2}}\frac{\{B_x(1)\}^2}{c}}{\left(n^{-\alpha+2\gamma}\frac{\sigma_{xx}^2}{-2c}-n^{2\gamma-1}\frac{\{B_x(1)\}^2}{c^2}\right)^{1/2}} \sim \frac{N\left(0,\frac{\sigma_{xx}^4}{-2c}\right)+o_p(1)}{\left(\frac{\sigma_{xx}^2}{-2c}+o_p(1)\right)^{1/2}} \Longrightarrow N\left(0,\sigma_{xx}^2\right),$$

and then

$$R_n \Longrightarrow \left\{N\left(0,\sigma_{00.x}^2\right)+\phi N\left(0,\sigma_{xx}^2\right)\right\}^2\left(\phi^2\sigma_{xx}^2+\sigma_{00.x}^2\right)^{-1}=\chi_1^2.$$

■

B Appendix B: Bivariate predictors

This Appendix provides results for deriving the limit distribution of RLRT in the case of bivariate predictors. The Lemmas are given for a bivariate predictor (x_{1t}, x_{2t}) generated as in the predictive regression (3.5). We use the notation for the limit Brownian motion and diffusion processes given in Appendix A.

B.1 Preliminary Lemmas

Lemma B.1 *(1)(i) If* $\rho_{1n}=1$ *and* $\rho_{2n}=1+\frac{c_2}{k_n}$*, define*

$$\widetilde{x}_{1,n}\left(r\right):=\frac{\widetilde{x}_{1,\lfloor nr\rfloor}}{\sqrt{n}}=\frac{1}{\sqrt{n}}\sum_{j=1}^{\lfloor nr\rfloor}\rho_{1n}^{j}u_{x1,\lfloor nr\rfloor-j},$$

$$\widetilde{x}_{2,n}\left(c_2\right):=\frac{\widetilde{x}_{2,\lfloor ns\rfloor}}{\sqrt{k_n}}=\frac{1}{\sqrt{k_n}}\sum_{j=1}^{\lfloor ns\rfloor}\rho_{2n}^{j}u_{x2,\lfloor ns\rfloor-j},$$

and we have joint convergence $\left(\widetilde{x}_{1,n}\left(r\right),\widetilde{x}_{2,\lfloor ns\rfloor}\left(c_2\right)\right)\Rightarrow\left(B_{x1}\left(r\right),N_{x2}\right)$*, with* $N_{x2}\sim N\left(0,\frac{\sigma_{xx2}^2}{-2c_2}\right)$ *for all* $r,s>0$*.*

(ii) If $\rho_{1n}=1+\frac{c}{n}$ *and* $\rho_{2n}=1+\frac{c_2}{k_n}$*, let*

$$\widetilde{x}_{1,n}^{c}\left(r\right):=\frac{\widetilde{x}_{1,\lfloor nr\rfloor}}{\sqrt{n}}=\frac{1}{\sqrt{n}}\sum_{j=1}^{\lfloor nr\rfloor}\rho_{1n}^{j}u_{x1,\lfloor nr\rfloor-j},$$

and we have the joint convergence $\left(\widetilde{x}_{1,n}^{c}\left(r\right),\widetilde{x}_{2,\lfloor ns\rfloor}\left(c_2\right)\right)\Rightarrow\left(J_{c_1}\left(r\right),N_{x2}\right)$*, with* $N_{x2}\sim N\left(0,\frac{\sigma_{xx2}^2}{-2c_2}\right)$ *for all* $r,s>0$*.*

(iii) If $\rho_{1n}=1+\frac{c_1}{k_n}$ *and* $\rho_{2n}=1+\frac{c_2}{k_n}$*, let*

$$\widetilde{x}_{1,n}\left(c_1\right):=\frac{\widetilde{x}_{1,\lfloor nr\rfloor}}{\sqrt{k_n}}=\frac{1}{\sqrt{k_n}}\sum_{j=1}^{\lfloor nr\rfloor}\rho_{1n}^{j}u_{x1,\lfloor nr\rfloor-j},$$

$$\widetilde{x}_{2,n}\left(c_2\right):=\frac{\widetilde{x}_{2,\lfloor ns\rfloor}}{\sqrt{k_n}}=\frac{1}{\sqrt{k_n}}\sum_{j=1}^{\lfloor ns\rfloor}\rho_{2n}^{j}u_{x2,\lfloor ns\rfloor-j},$$

and we have the joint convergence $\left(\widetilde{x}_{1,\lfloor nr\rfloor}\left(c_1\right),\widetilde{x}_{2,\lfloor ns\rfloor}\left(c_2\right)\right)\Rightarrow\left(N_{x1},N_{x2}\right)$*, with* $N_{x1}\sim N\left(0,\frac{\sigma_{xx2}^2}{-2c_c}\right)$ *and* $N_{x2}\sim N\left(0,\frac{\sigma_{xx2}^2}{-2c_2}\right)$*. for all* $r,s>0$*.*

(2)The limit behavior of standardized versions of $\sum_{t=1}^{n}\widetilde{x}_{1,t-1}\widetilde{x}_{2,t-1}$ *is as follows:*

(i)$\frac{1}{n^2}\sum_{t=1}^{n}\widetilde{x}_{1,t-1}\widetilde{x}_{2,t-1}\Rightarrow\int_0^1 B_{x_1}(r)B_{x_2}(r)dr$*, if* $\rho_{1n}=1$ *and* $\rho_{2n}=1$*,*

(ii)$\frac{1}{n^2}\sum_{t=1}^{n}\widetilde{x}_{1,t-1}\widetilde{x}_{2,t-1}\Rightarrow\int_0^1 J_{c_1}(r)J_{c_2}(r)dr$*, if* $\rho_{1n}=1+\frac{c_1}{n}$ *and* $\rho_{2n}=1+\frac{c_2}{n}$*,*

(iii)$\frac{1}{nk_n}\sum_{t=1}^{n}\widetilde{x}_{1,t-1}\widetilde{x}_{2,t-1}\Rightarrow\int_0^1 N_{x_1}N_{x_2}dr$*, if* $\rho_{1n}=1+\frac{c_1}{k_n}$ *and* $\rho_{2n}=1+\frac{c_2}{k_n}(c_1,c_2<0,k_n=n^{\alpha}$ *with* $\alpha\in(0,1))$*,*

(iv)$\frac{1}{n^2}\sum_{t=1}^{n}\widetilde{x}_{1,t-1}\widetilde{x}_{2,t-1}\Rightarrow\int_0^1 B_{x_1}(r)J_{c_2}(r)dr$*, if* $\rho_{1n}=1$ *and* $\rho_{2n}=1+\frac{c_2}{n}$*,*

(v)$\frac{1}{n\sqrt{nk_n}}\sum_{t=1}^{n}\widetilde{x}_{1,t-1}\widetilde{x}_{2,t-1}\to^{p}0$*, if* $\rho_{1n}=1$ *and* $\rho_{2n}=1+\frac{c_2}{k_n}$*,*

(vi)$\frac{1}{n\sqrt{nk_n}}\sum_{t=1}^{n}\widetilde{x}_{1,t-1}\widetilde{x}_{2,t-1}\to^{p}0$ *if* $\rho_{1n}=1+\frac{c_1}{n}$ *and* $\rho_{2n}=1+\frac{c_2}{k_n}$*.*

Lemma B.2 *The limit behavior of suitably standardized versions of* $\sum_{t=1}^{n} x_{1,t-1}x_{2,t-1}$ *is as follows:*

(i) If $\rho_{1n} = 1$ *and* $\rho_{2n} = 1$,

$$\frac{1}{n^2}\sum_{t=1}^{n} x_{1,t-1}x_{2,t-1} \implies \begin{cases} \int_0^1 B_{x_1}(r)B_{x_2}(r)dr & \text{if } \gamma > \frac{1}{2}, \\ \frac{\widetilde{\mu}_1\widetilde{\mu}_2}{3} + \widetilde{\mu}_2 \int_0^1 rB_{x_1}(r)dr + \\ \widetilde{\mu}_1 \int_0^1 rB_{x_2}(r)dr + \int_0^1 B_{x_1}(r)B_{x_2}(r)dr & \text{if } \gamma = \frac{1}{2}, \end{cases}$$

$$\frac{1}{n^{3-2\gamma}}\sum_{t=1}^{n} x_{1,t-1}x_{2,t-1} \to^p \frac{\widetilde{\mu}_1\widetilde{\mu}_2}{3} \text{ if } \gamma < \frac{1}{2}.$$

(ii) If $\rho_{1n} = 1 + \frac{c_1}{n}$ *and* $\rho_{2n} = 1 + \frac{c_2}{n}$,

$$\frac{1}{n^2}\sum_{t=1}^{n} x_{1,t-1}x_{2,t-1} \implies \begin{cases} \int_0^1 J_{c_1}(r)J_{c_2}(r)dr & \text{if } \gamma > \frac{1}{2}, \\ \widetilde{\mu}_1\widetilde{\mu}_2 \int_0^1 F_{c_1}(r)\,F_{c_2}(r)\,dr + \widetilde{\mu}_2 \int_0^1 J_{c_1}(r)F_{c_2}(r)\,dr \\ +\widetilde{\mu}_1 \int_0^1 J_{c_2}(r)F_{c_1}(r)\,dr + \int_0^1 J_{c_1}(r)J_{c_2}(r)dr & \text{if } \gamma = \frac{1}{2}, \end{cases}$$

$$\frac{1}{n^{3-2\gamma}}\sum_{t=1}^{n} x_{1,t-1}x_{2,t-1} \to^p \widetilde{\mu}_1\widetilde{\mu}_2 \int_0^1 F_{c_1}(r)\,F_{c_2}(r)\,dr \text{ if } \gamma < \frac{1}{2}.$$

(iii) If $\rho_{1n} = 1 + \frac{c_1}{k_n}$ *and* $\rho_{2n} = 1 + \frac{c_2}{k_n}(c_1, c_2 < 0, k_n = n^\alpha$ *with* $\alpha \in (0,1))$,

$$\frac{1}{nk_n}\sum_{t=1}^{n} x_{1,t-1}x_{2,t-1} \Rightarrow \int_0^1 N_{x_1}N_{x_2}dr \text{ if } \gamma \geqslant \frac{1}{2}, 0 < \frac{\alpha}{2} < \gamma < \frac{1}{2}$$

$$\frac{1}{n^{1+2\alpha-2\gamma}}\sum_{t=1}^{n} x_{1,t-1}x_{2,t-1} \Rightarrow \frac{\widetilde{\mu}_1\widetilde{\mu}_2}{c_1c_2} \text{ if } 0 < \gamma < \frac{\alpha}{2} < \frac{1}{2}$$

(iv) If $\rho_{1n} = 1$ *and* $\rho_{2n} = 1 + \frac{c_2}{n}$,

$$\frac{1}{n^2}\sum_{t=1}^{n} x_{1,t-1}x_{2,t-1} \implies \begin{cases} \int_0^1 B_{x_1}(r)J_{c_2}(r)dr & \text{if } \gamma > \frac{1}{2}, \\ \widetilde{\mu}_1\widetilde{\mu}_2 \int_0^1 rF_{c_2}(r)\,dr + \widetilde{\mu}_2 \int_0^1 B_{x_1}(r)F_{c_2}(r)\,dr \\ +\widetilde{\mu}_1 \int_0^1 J_{c_2}(r)rdr + \int_0^1 B_{x_1}(r)J_{c_2}(r)dr & \text{if } \gamma = \frac{1}{2}, \end{cases}$$

$$\frac{1}{n^{3-2\gamma}}\sum_{t=1}^{n} x_{1,t-1}x_{2,t-1} \to^p \widetilde{\mu}_1\widetilde{\mu}_2 \int_0^1 rF_{c_2}(r)\,dr \text{ if } \gamma < \frac{1}{2}.$$

(v)If $\rho_{1n} = 1$ and $\rho_{2n} = 1 + \frac{c_2}{k_n}$,

$$\begin{aligned}
\frac{1}{n\sqrt{nk_n}} \sum_{t=1}^{n} x_{1,t-1} x_{2,t-1} &\Rightarrow \int_0^1 B_{x_1}(r) N_{x_2} dr \text{ if } \gamma \geqslant \frac{1}{2} \\
\frac{1}{n^{2+\alpha-2\gamma}} \sum_{t=1}^{n} x_{1,t-1} x_{2,t-1} &\Rightarrow \frac{\widetilde{\mu}_1 \widetilde{\mu}_2}{-2c_2} \text{ if } 0 < \gamma < \frac{\alpha}{2} < \frac{1}{2}
\end{aligned}$$

(vi) If $\rho_{1n} = 1 + \frac{c_1}{n}$ and $\rho_{2n} = 1 + \frac{c_2}{k_n}$.

$$\begin{aligned}
\frac{1}{n\sqrt{nk_n}} \sum_{t=1}^{n} x_{1,t-1} x_{2,t-1} &\Rightarrow \int_0^1 J_{c_1}(r) N_{x_2} dr \text{ if } \gamma \geqslant \frac{1}{2} \\
\frac{1}{n^{2+\alpha-2\gamma}} \sum_{t=1}^{n} x_{1,t-1} x_{2,t-1} &\Rightarrow \widetilde{\mu}_1 \widetilde{\mu}_2 \int_0^1 F_{c_1}(r) dr \text{ if } 0 < \gamma < \frac{\alpha}{2} < \frac{1}{2}
\end{aligned}$$

Lemma B.3 *The following joint convergence results hold: (i) If $\rho_{1n} = 1 + \frac{c_1}{k_n}$ and $\rho_{2n} = 1 + \frac{c_2}{k_n} (c_1, c_2 < 0, k_n = n^{\alpha}$ with $\alpha \in (0,1))$, then*

$$\left(\begin{array}{c} \frac{1}{\sqrt{nk_n}} \sum_{t=1}^{n} \widetilde{x}_{1,t-1} u_{0.xt}, \frac{1}{\sqrt{nk_n}} \sum_{t=1}^{n} \widetilde{x}_{2,t-1} u_{0.xt}, \frac{1}{\sqrt{nk_n}} \sum_{t=1}^{n} \widetilde{x}_{1,t-1} u_{x1,t}, \\ \frac{1}{\sqrt{nk_n}} \sum_{t=1}^{n} \widetilde{x}_{2,t-1} u_{x2,t}, \frac{1}{\sqrt{nk_n}} \sum_{t=1}^{n} \widetilde{x}_{1,t-1} u_{x2,t}, \frac{1}{\sqrt{nk_n}} \sum_{t=1}^{n} \widetilde{x}_{2,t-1} u_{x1,t} \end{array} \right)$$

converges weakly to a multivariate normal with zero mean and corresponding variance matrix M_1.

(ii) If $\rho_{1n} = 1 \left(or\ 1 + \frac{c_1}{n} \right)$ and $\rho_{2n} = 1 + \frac{c_2}{k_n} (c_1, c_2 < 0, k_n = n^{\alpha}$ with $\alpha \in (0,1))$, then

$$\left(\frac{1}{\sqrt{nk_n}} \sum_{t=1}^{n} \widetilde{x}_{2,t-1} u_{0.xt}, \frac{1}{\sqrt{nk_n}} \sum_{t=1}^{n} \widetilde{x}_{2,t-1} u_{x2,t}, \frac{1}{\sqrt{nk_n}} \sum_{t=1}^{n} \widetilde{x}_{2,t-1} u_{x1,t}, \frac{\widetilde{x}_{2,n}}{\sqrt{k_n}} \right)$$

converges weakly to a multivariate normal with zero mean and corresponding variance matrix M_2. Explicit forms of M_1 and M_2 are given in the Supplement.

B.2 Proof of the Main Results

B.2.1 Proof of Theorem 3.1

Proof. Without standardization, we have the following expressions for (s_i) and (h_{ij}) :

$$s_1 := \frac{1}{\sigma_{00.x}^2} \left(\sum_{t=1}^{n} x_{1,t-1} u_{0x.t} - \frac{1}{n} \sum_{t=1}^{n} x_{1,t-1} \sum_{t=1}^{n} u_{0.xt} \right),$$

$$
\begin{aligned}
s_2 :=& \quad \frac{1}{\sigma^2_{00.x}}\left(\sum_{t=1}^{n} x_{2,t-1}u_{0x.t} - \frac{1}{n}\sum_{t=1}^{n} x_{2,t-1}\sum_{t=1}^{n} u_{0.xt}\right), \\
s_3 :=& \quad -\phi_1 s_1 + e\sigma^2_{xx2}\sum_{t=1}^{n} x_{1,t-1}u_{x1,t} - e\sigma_{x1x2}\sum_{t=1}^{n} x_{1,t-1}u_{x2,t}, \\
s_4 :=& \quad -\phi_2 s_2 + e\sigma^2_{xx1}\sum_{t=1}^{n} x_{2,t-1}u_{x2,t} - e\sigma_{x1x2}\sum_{t=1}^{n} x_{2,t-1}u_{x1,t}, \\
h_{11} :=& \quad \frac{-1}{\sigma^2_{00.x}}\left\{\sum_{t=1}^{n} x^2_{1,t-1} - \frac{1}{n}\left(\sum_{t=1}^{n} x_{1,t-1}\right)^2\right\}, \\
h_{12} :=& \quad \frac{-1}{\sigma^2_{00.x}}\left\{\sum_{t=1}^{n} x_{1,t-1}x_{2,t-1} - n\left(\frac{1}{n}\sum_{t=1}^{n} x_{1,t-1}\right)\left(\frac{1}{n}\sum_{t=1}^{n} x_{2,t-1}\right)\right\}, \\
h_{22} :=& \quad \frac{-1}{\sigma^2_{00.x}}\left\{\sum_{t=1}^{n} x^2_{2,t-1} - \frac{1}{n}\left(\sum_{t=1}^{n} x_{2,t-1}\right)^2\right\}, \\
h_{23} :=& \quad -\phi_1 h_{12}, \\
h_{24} \quad :=& -\phi_2 h_{22}, \\
h_{33} :=& \quad \phi_1^2 h_{11} - e\sigma^2_{xx2}\sum_{t=1}^{n} x^2_{1,t-1}, \\
h_{34} :=& \quad \phi_1\phi_2 h_{12} + e\sigma_{x1x2}\sum_{t=1}^{n} x_{1,t-1}x_{2,t-1}, \\
h_{44} :=& \quad \phi_2^2 h_{22} - e\sigma^2_{xx1}\sum_{t=1}^{n} x^2_{2,t-1}, \\
h_{13} :=& \quad -\phi_1 h_{11}, \\
h_{14} :=& \quad -\phi_2 h_{12},
\end{aligned}
$$

with $e = \frac{1}{\sigma^2_{xx1}\sigma^2_{xx2} - \sigma^2_{x1x2}}$. Given the limit results reported in the preliminary lemmas, the limit forms of standardized versions of the (s_i) and (h_{ij}) follow by continuous mapping and joint convergence, as does the corresponding limit distribution of R_n. In particular, for each $k = 1, 2, \ldots 6$, we define $A^k_S = (A_{s_1}, A_{s_2}, A_{s_3}, A_{s_4})'$ and

$$
A^k_Z = \begin{pmatrix} A_{h_{11}} & \cdot & \cdot & \cdot \\ A_{h_{21}} & A_{h_{22}} & \cdot & \cdot \\ A_{h_{31}} & A_{h_{32}} & A_{h_{33}} & \cdot \\ A_{h_{41}} & A_{h_{42}} & A_{h_{43}} & A_{h_{44}} \end{pmatrix}^{-1} - \begin{pmatrix} 0_{2\times 2} & 0_{2\times 2} \\ 0_{2\times 2} & \begin{pmatrix} A_{h_{33}} & \cdot \\ A_{h_{43}} & A_{h_{44}} \end{pmatrix}^{-1} \end{pmatrix}.
$$

We discuss the associated asymptotic results for $A^k_{(\cdot)}$ under the following six scenarios:

(1) If $\rho_{1n} = 1$ and $\rho_{2n} = 1$, notice, the superscript 1, 2 and 3 stand for linear,quadratic, and stochastic integral respectively:

$$\begin{aligned}
n^{-3/2}\sum_{t=1}^{n} x_{j,t-1} &\implies \mathbb{A}_j^1 = \int_0^1 B_{x_j} dr,\ (j=1,2),\\
n^{-2}\sum_{t=1}^{n} x_{j,t-1}^2 &\implies \mathbb{A}_j^2 = \int_0^1 B_{x_j}(r)^2 dr\ (j=1,2)\\
\frac{1}{n}\sum_{t=1}^{n} x_{j,t-1}u_{at} &\implies \mathbb{A}_{j,a}^3 = \int_0^1 B_{x_j}(r)dB_a(r),\ (j=1,2;\ a=x_1,x_2,0.x)
\end{aligned}$$

(2) If $\rho_{1n} = 1 + \frac{c_1}{n}$ and $\rho_{2n} = 1 + \frac{c_2}{n}$,the results follows (1) with replacing $B_{x_j}(r)$ with $K_{c_j}(r)$.

(3) If $\rho_{1n} = 1 + \frac{c_1}{k_n}$ and $\rho_{2n} = 1 + \frac{c_2}{k_n}$,

$$\begin{aligned}
\frac{1}{n\sqrt{k_n}}\sum_{t=1}^{n} x_{j,t-1} &\rightarrow^p \mathbb{A}_j^1 = 0,\ (j=1,2),\\
\frac{1}{nk_n}\sum_{t=1}^{n} x_{j,t-1}^2 &\rightarrow^p \mathbb{A}_j^2 = \int N_{x_j}N_{x_j} dr\ (j=1,2),\\
\frac{1}{\sqrt{nk_n}}\sum_{t=1}^{n} x_{j,t-1}u_{at} &\implies \mathbb{A}_{j,a}^3 = N\left(0, \frac{\sigma^2_{xxj}\sigma^2_{aa}}{-2c_j}\right)\ (j=1,2;\\
a &= x_1, x_2, 0.x \text{ and } aa = xx1, xx2, 00.x)
\end{aligned}$$

where the affix signifier 'a' on the left side in $\mathbb{A}$ and u_{at} corresponds to the element on the right side associated with the affix signifier 'aa'.

(4) If $\rho_{1n} = 1$ and $\rho_{2n} = 1 + \frac{c_2}{n}$, the results follows (1) with replacing $B_{x_2}(r)$ with $K_{c_2}(r)$.

(5) If $\rho_{1n} = 1$ and $\rho_{2n} = 1 + \frac{c_2}{k_n}$,

$$\begin{aligned}
\frac{1}{n}\sum_{t=1}^{n} x_{1,t-1} &\implies \mathbb{A}_1^1 = \int_0^1 B_{x1}(r)\,dr\\
\frac{1}{n\sqrt{k_n}}\sum_{t=1}^{n} x_{2,t-1} &\rightarrow^p \mathbb{A}_2^1 = 0\\
\frac{1}{n}\sum_{t=1}^{n} x_{1,t-1}^2 &\implies \mathbb{A}_1^2 = \int_0^1 B_{x1}(r)^2\,dr
\end{aligned}$$

$$
\begin{aligned}
\frac{1}{nk_n}\sum_{t=1}^{n} x_{2,t-1}^2 &\to^p \mathbb{A}_2^2 = \int (N_{x_2})^2\, dr \\
\frac{1}{\sqrt{nk_n}}\sum_{t=1}^{n} x_{1,t-1}u_{0.xt} &\Longrightarrow A_{1,0.x}^3 = \int_0^1 B_{x1}(r)\, dB_{0.x}(r) \\
\frac{1}{\sqrt{nk_n}}\sum_{t=1}^{n} x_{2,t-1}u_{0.xt} &\Longrightarrow A_{2,0.x}^3 = N\left(0, \frac{\sigma_{xx2}^2\sigma_{00.x}^2}{-2c_2}\right) \\
\frac{1}{n}\sum_{t=1}^{n} x_{1,t-1}u_{x1,t} &\Longrightarrow A_{1,x_1}^3 = \int_0^1 B_{x1}(r)\, dB_{x1}(r) \\
\frac{1}{\sqrt{nk_n}}\sum_{t=1}^{n} x_{2,t-1}u_{x2,t} &\Longrightarrow A_{2,x_2}^3 = N\left(0, \frac{\sigma_{xx2}^4}{-2c_2}\right) \\
\frac{1}{n}\sum_{t=1}^{n} x_{1,t-1}u_{x2,t} &\Longrightarrow A_{1,x_2}^3 = \int_0^1 B_{x1}(r)\, dB_{x2}(r) \\
\frac{1}{\sqrt{nk_n}}\sum_{t=1}^{n} x_{2,t-1}u_{x1,t} &\Longrightarrow A_{2,x_1}^3 = N\left(0, \frac{\sigma_{xx1}^2\sigma_{xx2}^2}{-2c_2}\right) \\
\frac{1}{n\sqrt{nk_n}}\sum_{t=1}^{n} x_{1,t-1}x_{2,t-1} &\Longrightarrow A_{1,2}^2 = \int_0^1 B_{x1}(r)\, N_{x2}dr
\end{aligned}
$$

(6) If $\rho_{1n} = 1 + \frac{c_1}{n}$ and $\rho_{2n} = 1 + \frac{c_2}{k_n}$, the results follows (1) with replacing $B_{x_1}(r)$ with $K_{c_1}(r)$.
■

B.2.2 Proof of Theorem 3.2

Proof. (i) If $\rho_n \in (-1, 1)$,

$$
\begin{aligned}
R_n &\Rightarrow \left\{ \frac{1}{\sigma_{00.x}} \frac{N\left(0, \frac{\sigma_{xx}^2\sigma_{00.x}^2}{1-\rho_n^2}\right)}{\left(\frac{\sigma_{xx}^2}{1-\rho_n^2}\right)^{1/2}} \sqrt{1-\lambda^2} + \frac{\left(N\left(0, \frac{\sigma_{xx}^4}{1-\rho_n^2}\right)\right)}{\left(\frac{\sigma_{xx}^2}{1-\rho_n^2}\right)^{1/2}} \frac{\lambda}{\sigma_{xx}} \right\}^2 \\
&= \left\{ \frac{\sqrt{1-\lambda^2}}{\sigma_{00.x}} N\left(0, \sigma_{00.x}^2\right) + \frac{\lambda N\left(0, \sigma_{xx}^2\right)}{\sigma_{xx}} \right\}^2 = \{N(0,1)\}^2 = \chi_1^2.
\end{aligned}
$$

(ii) If $\rho_n = 1 + \frac{c}{n}$,

$$
R_n \Rightarrow \left\{ \lambda\tau_c g_{c,\lambda}^{1/2} + \sqrt{1-\lambda^2 g_{c,\lambda}} Z \right\}^2.
$$

(iii) If $\rho_n = 1 + \frac{c}{k_n}$,

$$
\begin{aligned}
R_n &\Rightarrow \left\{ \frac{1}{\sigma_{00.x}} \frac{N\left(0, \frac{\sigma_{xx}^2 \sigma_{00.x}^2}{-2c}\right)}{\left(\frac{\sigma_{xx}^2}{-2c}\right)^{1/2}} \sqrt{1-\lambda^2} + \frac{\left(N\left(0, \frac{\sigma_{xx}^4}{-2c}\right)\right)}{\left(\frac{\sigma_{xx}^2}{-2c}\right)^{1/2}} \frac{\lambda}{\sigma_{xx}} \right\}^2 \\
&\Rightarrow \{N(0,1)\}^2 = \chi_1^2.
\end{aligned}
$$

■

B.2.3 Proof of Lemma 3.3

Proof. Follow the proof of Lemma 2.2, we have

$$
L(\Theta, Y, X) = -\frac{n-1}{2} \log \sigma_{00.x}^2 - \frac{1}{2\sigma_{00.x}^2} S(\boldsymbol{\beta}, \boldsymbol{\rho}, \boldsymbol{\phi}) + \frac{1}{2} P(\boldsymbol{\rho}, \Sigma) - \frac{1}{2} Q(\boldsymbol{\rho}, \Sigma),
$$

where

$$
\begin{aligned}
S(\boldsymbol{\beta}, \boldsymbol{\rho}, \boldsymbol{\phi}) &= \sum_{t=1}^{n} \left[\underline{y_t} - \boldsymbol{\phi}' \underline{\mathbf{x}_t} - (\boldsymbol{\beta}' - \boldsymbol{\phi}' \boldsymbol{\rho}) \underline{\mathbf{x}_{t-1}} \right]^2, \\
P(\boldsymbol{\rho}, \Sigma) &= \log |Z'Z| + (n+3) \log |\Sigma|, \\
Q(\boldsymbol{\rho}, \Sigma) &= \sum_{t=1}^{n} (\mathbf{x}_t - \rho \mathbf{x}_{t-1})' \Sigma^{-1} (\mathbf{x}_t - \boldsymbol{\rho} \boldsymbol{x}_{t-1}) - \frac{1}{n} \left\{ \sum_{t=1}^{n} (\mathbf{x}_t - \boldsymbol{\rho} \boldsymbol{x}_{t-1}) \right\}' \Sigma^{-1} \left\{ \sum_{t=1}^{n} (\mathbf{x}_t - \boldsymbol{\rho} \boldsymbol{x}_{t-1}) \right\},
\end{aligned}
$$

and

$$
Z' = \begin{pmatrix} \mathbf{0}_{2\times 2} & \mathbf{I}_{2\times 2} & \mathbf{I}_{2\times 2} + \boldsymbol{\rho} & \ldots & \mathbf{I}_{2\times 2} + \boldsymbol{\rho} + \ldots \boldsymbol{\rho}^{n-2} & \mathbf{I}_{2\times 2} + \boldsymbol{\rho} + \ldots \boldsymbol{\rho}^{n-1} \\ \mathbf{I}_{2\times 2} & \boldsymbol{\rho} & \boldsymbol{\rho}^2 & \ldots & \boldsymbol{\rho}^{n-1} & \boldsymbol{\rho}^n \end{pmatrix}
$$

■

B.2.4 Proof of Theorem 3.4

Proof. Without standardization, we have the following expressions for (s_i) and (h_{ij}) :

$$
\begin{aligned}
s_1 &:= \frac{1}{\sigma_{00.x}^2} \left(\sum_{t=1}^{n} x_{1,t-1} u_{0x.t} - \frac{1}{n} \sum_{t=1}^{n} x_{1,t-1} \sum_{t=1}^{n} u_{0.xt} \right), \\
s_2 &:= \frac{1}{\sigma_{00.x}^2} \left(\sum_{t=1}^{n} x_{2,t-1} u_{0x.t} - \frac{1}{n} \sum_{t=1}^{n} x_{2,t-1} \sum_{t=1}^{n} u_{0.xt} \right),
\end{aligned}
$$

$$
\begin{aligned}
s_3 := & \quad -\phi_1 s_1 + e\sigma_{xx2}^2 \left\{ \sum_{t=1}^{n} x_{1,t-1} u_{x1.t} - \frac{1}{n} \left(\sum_{t=1}^{n} x_{1,t-1} \right) \sum_{t=1}^{n} u_{x1.t} \right\} \\
& \quad + e\sigma_{x1x2} \left\{ \sum_{t=1}^{n} x_{1,t-1} u_{x2.t} - \frac{1}{n} \left(\sum_{t=1}^{n} x_{1,t-1} \right) \sum_{t=1}^{n} u_{x2.t} \right\}, \\
s_4 := & \quad -\phi_2 s_2 + e\sigma_{xx2}^2 \left\{ \sum_{t=1}^{n} x_{2,t-1} u_{x2.t} - \frac{1}{n} \left(\sum_{t=1}^{n} x_{2,t-1} \right) \sum_{t=1}^{n} u_{x2.t} \right\} \\
& \quad + e\sigma_{x1x2} \left\{ \sum_{t=1}^{n} x_{2,t-1} u_{x1.t} - \frac{1}{n} \left(\sum_{t=1}^{n} x_{2,t-1} \right) \sum_{t=1}^{n} u_{x2.t} + (\mu_1 - \mu_2) \sum_{t=1}^{n} x_{2,t-1} \right\}, \\
h_{11} := & \quad \frac{-1}{\sigma_{00.x}^2} \left\{ \sum_{t=1}^{n} x_{1,t-1}^2 - \frac{1}{n} \left(\sum_{t=1}^{n} x_{1,t-1} \right)^2 \right\}, \\
h_{12} := & \quad \frac{-1}{\sigma_{00.x}^2} \left\{ \sum_{t=1}^{n} x_{1,t-1} x_{2,t-1} - n \left(\frac{1}{n} \sum_{t=1}^{n} x_{1,t-1} \right) \left(\frac{1}{n} \sum_{t=1}^{n} x_{2,t-1} \right) \right\}, \\
h_{22} := & \quad \frac{-1}{\sigma_{00.x}^2} \left\{ \sum_{t=1}^{n} x_{2,t-1}^2 - \frac{1}{n} \left(\sum_{t=1}^{n} x_{2,t-1} \right)^2 \right\}, \\
h_{23} := & \quad -\phi_1 h_{12}, \;\; h_{24} := -\phi_2 h_{22}, \\
h_{33} := & \quad \phi_1^2 h_{11} - e\sigma_{xx2}^2 \left\{ \sum_{t=1}^{n} x_{1,t-1}^2 - \frac{1}{n} \left(\sum_{t=1}^{n} x_{1,t-1} \right)^2 \right\}, \\
h_{34} := & \quad \phi_1 \phi_2 h_{12} + e\sigma_{x1x2} \left\{ \sum_{t=1}^{n} x_{1,t-1} x_{2,t-1} - \frac{1}{n} \left(\sum_{t=1}^{n} x_{1,t-1} \right) \left(\sum_{t=1}^{n} x_{2,t-1} \right) \right\}, \\
h_{44} := & \quad \phi_2^2 h_{22} - e\sigma_{xx1}^2 \left\{ \sum_{t=1}^{n} x_{2,t-1}^2 - \frac{1}{n} \left(\sum_{t=1}^{n} x_{2,t-1} \right)^2 \right\}, \\
h_{13} := & \quad -\phi_1 h_{11}, \\
h_{14} := & \quad -\phi_2 h_{12},
\end{aligned}
$$

with $e = \frac{1}{\sigma_{xx1}^2 \sigma_{xx2}^2 - \sigma_{x1x2}^2}$. Given some limit results reported in the preliminary lemmas, the limit forms of standardized versions of the (s_i) and (h_{ij}) follow by continuous mapping and joint convergence, as does the corresponding limit distribution of R_n. ■