Limit Theory for VARs with Mixed Roots Near Unity*

Peter C. B. Phillips

Yale University, University of Auckland,

Singapore Management University & University of Southampton

Ji Hyung Lee

Yale University

August 20, 2012

Abstract

Limit theory is developed for nonstationary vector autoregression (VAR) with mixed roots in the vicinity of unity involving persistent and explosive components. Statistical tests for common roots are examined and model selection approaches for discriminating roots are explored. The results are useful in empirical testing for multiple manifestations of nonstationarity – in particular for distinguishing mildly explosive roots from roots that are local to unity and for testing commonality in persistence.

Keywords: Common roots, Local to unity, Mildly explosive, Mixed roots, Model selection, Persistence, Tests of common roots.

JEL classification: C22

1 Introduction

Aman Ullah's contributions cover a wide spectrum of econometrics with sustained scientific work over the last four decades in finite sample theory, nonparametric estimation, spatial econometrics, panel data modeling, financial econometrics, time series and applied

*This paper is based on the first part of a Yale take home examination in 2010/2011. The authors thank the Editor and two referees for helpful and constructive comments on the original version. Phillips acknowledges support from the NSF under Grant No. SES-0956687.

econometrics. His advanced textbook on *Nonparametric Econometrics* (1999, with Adrian Pagan) has been particularly influential, helping to educate a generation of econometricians in nonparametric methods and providing an accessible reference for applied researchers. His monograph on *Finite Sample Econometrics* (2004) encapsulates many of his own contributions to this subject and touches some of the wider reaches of this difficult and vitally important field.

One field of econometrics that his work has less frequently touched is nonstationary time series and unit root limit theory. Since the mid 1980s models with autoregressive roots in the vicinity of unity have attracted much attention. These models are particularly useful in empirical work with nonstationary series when it may be too restrictive to insist on the presence of roots precisely at unity or where mildly integrated or mildly explosive behavior may be more relevant than unit roots. When multiple time series are considered, it may be useful to allow simultaneously for various types of behavior in the individual series: some roots that are local to unity (of the form $1 + \frac{c}{n}$ for fixed c and given sample size n) and others that are mildly integrated (of the form $1 + \frac{b}{k_n}$ for fixed $b < 0$ and a sequence $k_n \to \infty$ slower than n) or mildly explosive (of the form $1 + \frac{b}{k_n}$ for fixed $b > 0$ fixed and a similar sequence $k_n \to \infty$ slower than n). Roots of the latter form lie in a wide vicinity of unity of radius $O\left(k_n^{-1}\right)$ and thereby accommodate some interesting alternatives where the behavior of the process, including its limit behavior, differ from that of the random wandering character associated with unit root processes. The mathematical form of these roots involves the localizing coefficient b and a rate sequence k_n which it is often convenient to write in the exponent form $k_n = n^{\alpha}$ for some parameter $\alpha \in (0, 1)$.

Limit theory for regressors with roots local to unity developed early in the literature of this field (Phillips, 1987b; Chan and Wei, 1987). More recent work has considered mildly integrated and mildly explosive cases (Phillips and Magdalinos, 2007a, 2007b; [PM7a&b]). The latter theory has proved particularly relevant in studying data during periods of financial exuberance (Phillips, Wu and Yu, 2011; Phillips and Yu, 2011). In such cases, exuberance can be modeled in terms of a mildly explosive process with an autoregressive root $\rho_n = 1 + \frac{b}{k_n}$ for which $b > 0$ and $\frac{1}{k_n} + \frac{k_n}{n} \to 0$. It is especially interesting in practical work to study transitions between normal market behavior, which can be represented in terms of a unit root or near unit root model, and exuberant behavior. The emergence of market exuberance or an asset price bubble may then be modeled as a structural break in which the (long run) autoregressive coefficient of the model ρ_n shifts from being near to unity to mildly explosive. Dating such a transition amounts to date stamping the emergence of exuberance. A similar transition from an exuberant to a mildly integrated or mean reverting process captures the collapse of an asset price bubble and correspondingly enables the date stamping of bubble termination. Phillips, Wu and Yu (2011) and Phillips and Yu (2011) showed how to perform tests of these hypotheses and construct date stamping algorithms that were empirically implemented to characterize the 1990s Nasdaq bubble and the events leading up to and following the recent global financial crisis. The work in those papers dealt with the special case where the normal period model was a strict unit root process.

The methods of the present paper allow for these methods to be extended to the wider class of local to unity processes (for normal periods) and enable tests to be developed to distinguish such roots from mildly explosive and mildly integrated roots, thereby widening the range of potential empirical applications. In particular, the present paper considers time series models with mixed and common roots in the vicinity of unity. To simplify exposition, we work with a bivariate model and analyze a case of primary interest where there is one local to unit root and one mildly explosive root. Models of this type may be anticipated when there are manifestations of nonstationarity in the data but somewhat different individual characteristics in the two series. Or it may be that the behavior is common across the series – for instance in several asset prices – arising from a single source of persistence or exuberance. We may be particularly interested empirically in testing commonality in persistence or long run behavior across series, which occurs when the autoregressive roots have the same value. The methods of the current paper enable empirical researchers to conduct such tests.

The remainder of the paper is organized as follows. Section 2 considers mixed VARs whose variates have mixed degrees of persistence that allow for a local to unit root and a mildly explosive root. A limit theory for least squares regression and associated Wald tests for commonality in the autoregressive coefficients is developed. Since the null hypothesis is composite and involves the unknown (local to unity) localizing coefficient, standard Wald tests have limit distributions that are parameter dependent and do not have uniform size. Modified Wald statistics for testing commonality in long run behavior are developed and shown to produce consistent tests. In particular, this modification ensures that the tests are completely consistent in the sense that size goes to zero and power to unity asymptotically. Section 3 considers a model selection approach and shows that the BIC criterion can also distinguish persistent and mildly explosive behavior. Section 4 concludes. A technical Appendix includes subsidiary lemmas and proofs of the main results.

2 Mixed Variate VARs

For simplicity of exposition, we consider the bivariate VAR(1) model

$$X_t = R_n X_{t-1} + u_t, \quad t = 1, ..., n, \tag{2.1}$$

$$R_n = \begin{bmatrix} \rho_n & 0 \\ 0 & \theta_n \end{bmatrix}, \ \rho_n = 1 + \frac{c}{n}, \ \theta_n = 1 + \frac{b}{k_n}, \ b > 0, \tag{2.2}$$

which we write in component form as

$$\begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \begin{bmatrix} \rho_n & 0 \\ 0 & \theta_n \end{bmatrix} \begin{bmatrix} X_{1t-1} \\ X_{2t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}, \quad t = 1, ..., n \tag{2.3}$$

with initialization $X_0 = o_p\left(k_n^{1/2}\right)$, and martingale difference innovations u_t satisfying Assumption 1 below. Our results may be extended to systems with weakly dependent errors u_t under conditions like those in the linear process framework of Magdalinos and Phillips (2009), but all the key ideas follow as in the simpler VAR(1) model studied here so we do not provide details. The coefficient $\rho_n = 1 + \frac{c}{n}$ is local to unity, $\theta_n = 1 + \frac{b}{k_n}$ is a mildly explosive coefficient with $b > 0$ and the sequence k_n satisfies $\frac{1}{k_n} + \frac{k_n}{n} \to 0$ as $n \to \infty$. The power rate $k_n = n^\alpha$ for $\alpha \in (0, 1)$ satisfies this latter condition as well as conditions we use later in the paper to develop consistent test procedures that involve slowly varying functions $L_n \to \infty$. In particular $\frac{n^\alpha L_n}{n} \to 0$ for all such functions L_n.

Although $\frac{\theta_n}{\rho_n} \to 1$ as $n \to \infty$ (so both coefficients are in the vicinity of unity), $k_n\left(\frac{\theta_n}{\rho_n} - 1\right) \to b > 0$ and so θ_n is 'further' from unity than ρ_n for all finite c as $n \to \infty$. In order to distinguish the mildly explosive behavior induced by θ_n from the persistence induced by ρ_n, statistical tests need to differentiate θ_n from ρ_n for all finite c as $n \to \infty$. As we will show, consistent tests can be constructed to discriminate between such localizing coefficients. The fact that consistent tests of hypotheses involving localizing coefficients is possible is relevant to practical work where there is substantial interest in identifying exuberance in asset price data. It is also of theoretical interest because it is well known that the localizing coefficient c cannot be consistently estimated in local to unity specifications. By contrast, the localizing coefficient b in mildly integrated and mildly explosive specifications is consistently estimable, and it is this feature of the model that makes possible consistent testing of differences in localizing behavior.

The diagonal form of R_n in (2.1) conforms with standard practice in the stochastically nonstationary literature. The presence of non-zero off diagonal elements in R_n induces higher order stochastic trends or explosive mechanisms in the time series, at least unless those coefficients are local to zero or negligible. Hence, non-zero off diagonal elements in R_n result in amplified feedback across series in nonstationary autoregressions. It is therefore conventional practice to retain a diagonal form R_n in developing a limit theory, as we do here. Of course, if the context and characteristics of the series suggest the presence of such feedbacks then they may be included and their effects on the limit theory can be analyzed.

Assumption 1. *The errors $\{u_t\}$ in (2.1) form a martingale difference sequence with respect to the natural filtration $\mathcal{F}_t = \sigma(u_t, u_{t-1}, ...)$ satisfying*

$$E_{\mathcal{F}_{t-1}}\left(u_t u_t'\right) = \Sigma \quad \text{and} \quad E_{\mathcal{F}_{t-1}}\|u_t\| \geq \delta \quad \text{a.s. for all } t \tag{2.4}$$

for some $\delta > 0$ and positive definite matrix $\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$, $\sup_t E\|u_t\|^4 < \infty$, and

$$\max_{1 \leq t \leq n} E\left(\|u_t\|^2 \mathbf{1}\{\|u_t\| > \lambda_n\}\right) \to 0 \quad \text{as } n \to \infty \tag{2.5}$$

for any sequence $(\lambda_n)_{n \in \mathbb{N}}$ such that $\lambda_n \to \infty$, and where

$$\|M\| = \max_i \left\{\lambda_i^{1/2} : \lambda_i \text{ is an eigenvalue of } M'M\right\}$$

is the spectral norm of the matrix M.

As expected from the differences in the coefficients ρ_n and θ_n in (2.3), the time series components X_{1t} and X_{2t} have different orders of magnitude as $n \to \infty$. These differences translate into different rates of convergence of the sample moments of X_t and the least squares regression components. To accommodate these differences we employ the (asymptotically equivalent) normalizing matrices

$$D_n := \begin{bmatrix} n & 0 \\ 0 & k_n \theta_n^n \end{bmatrix} \text{ and } F_n := \begin{bmatrix} n & 0 \\ 0 & \frac{\theta_n^n}{(\theta_n^2 - 1)} \end{bmatrix}.$$

The unrestricted least squares regression estimate of R_n in (2.1) is written in standard notation as $\hat{R}_n = X'X_{-1}\left(X_{-1}'X_{-1}\right)^{-1}$. This estimate is consistent and has a limit distribution that is obtained from a combination of functional limit theory that applies to the persistent components and central limit theory that applies to the mildly explosive components, as detailed in the following result.

Theorem 2.1 *As* $n \to \infty$,

$$\left(\hat{R}_n - R_n\right) F_n \Rightarrow \left[\begin{array}{cc} \frac{\int_0^1 J_{1c}(r)dB(r)}{\int_0^1 J_{1c}(r)^2 dr} & \frac{Y(b)}{X_2(b)} \end{array} \right] := \Phi, \tag{2.6}$$

where $J_{1c}(r) = \int_0^r e^{c(r-s)} dB_1(s)$, *which is an Ornstein-Uhlenbeck (O-U) process,* $B(r) = (B_1(r), B_2(r))'$ *is bivariate Brownian motion with variance matrix* Σ, $X(b) = (X_1(b), X_2(b))' \equiv N(0, \frac{1}{2b}\Sigma)$, $Y(b) =^d X(b)$, *and* $X(b)$ *and* $Y(b)$ *are independent. The two column components* $\frac{\int_0^1 J_{1c}(r)dB(r)}{\int_0^1 J_{1c}(r)^2 dr}$ *and* $\frac{Y(b)}{X_2(b)}$ *of the limiting matric variate* Φ *are independent.*

Remarks

1. The two columns of $\hat{R}_n - R_n$ converge at different rates, the first at the usual $O(n)$ rate for near integrated regressions and the second at the mildly explosive rate $\frac{\theta_n^n}{(\theta_n^2 - 1)} = O(k_n \theta_n^n) = O\left(k_n e^{bn/k_n}\right)$. In particular, writing $\Phi = (\Phi_{ij})$, we have

$$n(\widehat{r}_{11} - r_{11}) \;\Rightarrow\; \Phi_{11} = \frac{\int_0^1 J_{1c}(r)dB_1(r)}{\int_0^1 J_{1c}(r)^2 dr}, \tag{2.7}$$

$$n(\widehat{r}_{21} - r_{21}) \;\Rightarrow\; \Phi_{21} = \frac{\int_0^1 J_{1c}(r)dB_2(r)}{\int_0^1 J_{1c}(r)^2 dr}, \tag{2.8}$$

$$\frac{\theta_n^n}{(\theta_n^2 - 1)} (\widehat{r}_{22} - r_{22}) \Rightarrow \Phi_{22} = \frac{Y_2(b)}{X_2(b)}, \quad \frac{\theta_n^n}{(\theta_n^2 - 1)} (\widehat{r}_{12} - r_{12}) \Rightarrow \Phi_{12} = \frac{Y_1(b)}{X_2(b)}. \tag{2.9}$$

2. The process $J_{1c}(r) = \int_0^r e^{c(r-s)} dB_1(s)$ that appears in the limit variate Φ_{11} involves component $B_1(r)$ of $B(r)$, so that the limit variate $\int_0^1 J_{1c}(r)dB_1(r) / \int_0^1 J_{1c}(r)^2 dr$ has a standard local unit root distribution that is independent of σ_{11} but is dependent on c.

3. The limit variate $\frac{Y(b)}{X_2(b)} = \frac{(2b)^{1/2} Y(b)}{(2b)^{1/2} X_2(b)} =: \frac{Y}{X_2}$ is independent of b and we can therefore write $\frac{Y(b)}{X_2(b)} =: \frac{Y}{X_2}$, where $Y \equiv N(0, \Sigma)$, $X = (X_1, X_2)' \equiv N(0, \Sigma)$, and X and Y are independent.

As indicated earlier, we may be interested in testing commonality of persistence characteristics in the component series X_{1t} and X_{2t}. In the present case, setting $R_n = (r_{ij})$ and under a maintained hypothesis that R_n is diagonal with roots local to unity, commonality amounts to testing the hypothesis $H_0 : r_{11} = r_{22} = 1 + \frac{c}{n}$ for some finite $c \in (-\infty, \infty)$. The null can be written as $H_0 : a_1{}'vec\,(R_n) = 0$ where $a_1' = [1, 0, 0, -1]$ without explicitly specifying a common persistence parameter $r_n = 1 + c/n$. H_0 may also be subsumed in a block test of $R_n = r_n I$ for some $r_n = 1 + \frac{c}{n}$, which we can write in the form H_0^A : $A'vec\,(R_n) = 0$ where we use row vectorization in the vec operator and

$$A' = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} =: \begin{bmatrix} a_1' \\ a_2' \\ a_3' \end{bmatrix}.$$

The standard Wald test of H_0 uses the statistic

$$W_n = \left(a_1' vec\left(\hat{R}_n\right)\right)^2 / a_1' \left\{\hat{\Sigma} \otimes \left(X_{-1}' X_{-1}\right)^{-1}\right\} a_1,$$

and the corresponding block test of H_0^A has the form

$$\begin{aligned} W_n^A &= \left(A'vec\left(\hat{R}_n\right)\right)' \left(A' \left\{\hat{\Sigma} \otimes \left(X_{-1}' X_{-1}\right)^{-1}\right\} A\right)^{-1} \left(A'vec\left(\hat{R}_n\right)\right) \\ &= \left(A'vec\left(n\hat{R}_n\right)\right)' \left(A' \left\{\hat{\Sigma} \otimes n^2 \left(X_{-1}' X_{-1}\right)^{-1}\right\} A\right)^{-1} \left(A'vec\left(n\hat{R}_n\right)\right), \end{aligned}$$

where $\hat{\Sigma} = n^{-1} \sum_{t=1}^{n} \hat{u}_t \hat{u}_t'$ is a consistent estimator of Σ based on the least squares residuals $\hat{u}_t = X_t - \hat{R}_n X_{t-1}$.

Under (2.3) the coefficients $r_{11} = \rho_n$ and $r_{22} = \theta_n$, so that $r_{11} - r_{22} = \frac{c}{n} - \frac{b}{k_n} \sim -\frac{b}{k_n} = o(1)$, which is local to zero. Hence the model (2.2) actually corresponds to a local alternative to the null H_0.

Theorem 2.2 *Under the null hypothesis $H_0 : R_n = r_n I$ with $r_n = 1 + \frac{c}{n}$, as $n \to \infty$*

$$W_n \Rightarrow \frac{\left(a_1' \xi\right)^2}{a_1' \left\{\Sigma \otimes \left(\int_0^1 J_c(r) J_c(r)'\right)^{-1}\right\} a_1}, \tag{2.10}$$

and

$$W_n^A \Rightarrow \xi' A \left(A' \left\{ \Sigma \otimes \left(\int_0^1 J_c(r) J_c(r)' dr \right)^{-1} \right\} A \right)^{-1} A'\xi, \tag{2.11}$$

where $J_c(r) = \int_0^r e^{c(r-s)} dB(s)$, $\xi = vec\,(\Xi)$ *and* $\Xi = \int_0^1 dBJ_c{}' \left(\int_0^1 J_c J_c{}' \right)^{-1}$. *Under the alternative* $H_1 : R_n = \text{diag}\,(\rho_n, \theta_n)$

$$W_n, W_n^A \sim \frac{\left(-\frac{n}{k_n} b\right)^2}{\sigma_{11} \left(\int_0^1 J_{1c}(r)^2 dr \right)^{-1}} \{1 + o_p(1)\} = O_p\left(\frac{n}{k_n}\right)^2. \tag{2.12}$$

Remarks

4. The null limit distributions (2.10) and (2.11) are parameter dependent. The dependence involves the localizing coefficient c and the variance matrix Σ. When $c = 0$,

$$\Xi = \int_0^1 dBB' \left(\int_0^1 BB' \right)^{-1} = \Sigma^{1/2} \int_0^1 dVV' \left(\int_0^1 VV' \right)^{-1} \Sigma^{-1/2} =: \Sigma^{1/2} \Xi_V \Sigma^{-1/2}$$

   where $V \equiv BM\,(I_2)$ is standard vector Brownian motion. The limit distribution of the Wald statistic is then

$$W_n \Rightarrow \frac{\left(a_1' \left(\Sigma^{1/2} \otimes \Sigma^{-1/2}\right) \xi_V\right)^2}{a_1' \left\{ \Sigma \otimes \Sigma^{-1/2} \left(\int_0^1 VV' \right)^{-1} \Sigma^{-1/2} \right\} a_1} = \frac{(b'\xi_V)^2}{b' \left\{ I \otimes \left(\int_0^1 VV' \right)^{-1} \right\} b}, \tag{2.13}$$

   where $\xi_V = vec\,(\Xi_V)$ and

$$b = \frac{\left(\Sigma^{1/2} \otimes \Sigma^{-1/2}\right) a_1}{\left(a_1' \left(\Sigma \otimes \Sigma^{-1}\right) a_1\right)^{1/2}}$$

   lies on the unit sphere $b'b = 1$. Thus, even in the case of a common unit root, the null limit distribution of the test depends on Σ, although this matrix is consistently estimable by the residual moment matrix $\hat{\Sigma}$. In the general case, the limit distributions (2.10) and (2.11) both have nuisance parameters (c, Σ).

5. The parameter c is not consistently estimable and it is therefore not possible to construct a standard test of the composite H_0. However, modified tests are available to distinguish H_0 from alternatives that involve a mildly explosive component. For instance, for some (possibly slowly varying) sequence $L_n \to \infty$, the statistic $W_{L_n} =$

$W_n/L_n \to_p 0$ under H_0 for all finite c. Then, under the alternative hypothesis H_1, $W_{L_n} = O_p\left(\frac{n^2}{k_n^2 L_n}\right)$ which diverges for all sequences $L_n \to \infty$ such that $\frac{k_n^2 L_n}{n^2} \to 0$. In particular, if $k_n = O(n^\alpha)$ for some $\alpha \in (0,1)$ and L_n is slowly varying at infinity, then $W_{L_n} = O_p\left(\frac{n^{2(1-\alpha)}}{L_n}\right) \to \infty$ as $n \to \infty$ and tests based on the statistic W_{L_n} with any fixed critical value[1] are consistent and have zero size asymptotically. Similar remarks apply to the block test based on $W_{L_n}^A = W_n^A/L_n$.

6. In view of (2.12), $W_n, W_n^A = O_p\left(\frac{n^2}{k_n^2}\right)$ and the Wald statistics diverge, as do the scaled statistics W_{L_n} and $W_{L_n}^A$. So there is discriminatory power under the local alternative $H_1 : r_{11} = \rho_n = 1 + \frac{c}{n}$, $r_{22} = \theta_n = 1 + \frac{b}{k_n}$.

3 Model Selection

Another approach to testing for common roots in (2.1) is to apply model selection methods. This involves estimating (2.1) in the restricted case under the null of a common root and under the alternative of unrestricted roots.

Estimating (2.1) under the restriction $R_n = r_n I$ gives the pooled least squares estimator $\hat{r}_n = \left(\sum_{t=1}^n X_t' X_{t-1}\right)\left(\sum_{t=1}^n X_{t-1}' X_{t-1}\right)^{-1}$ of the common root r_n. We have the following limit theory for $\hat{r}_n$ under the null hypothesis and alternative.

Lemma 3.1 *(i) Under the null $R_n = r_n I$ with $r_n = 1 + \frac{c}{n}$, $\hat{r}_n$ has the limit distribution*

$$n\left(\hat{r}_n - r_n\right) \Rightarrow \left(\int_0^1 J_c(r)' dB\right) / \left(\int_0^1 J_c(r)' J_c(r) dr\right), \tag{3.1}$$

and the residual moment matrix $\tilde{\Sigma} = n^{-1} \sum_{t=1}^n \tilde{u}_t \tilde{u}_t' \to_p \Sigma$, where $\tilde{u}_t = X_t - \hat{r}_n X_{t-1}$, has the form

$$\tilde{\Sigma} = \frac{1}{n} \sum_{t=1}^n u_t u_t' + O\left(n^{-1}\right). \tag{3.2}$$

(ii) Under the alternative hypothesis where $R_n = \operatorname{diag}(\rho_n, \theta_n)$, $\hat{r}_n$ has the limit distribution

$$k_n \theta_n^n \left(\hat{r}_n - \theta_n\right) \Rightarrow 2b \frac{Y_2(b)}{X_2(b)}, \tag{3.3}$$

[1] For example, asymptotic critical values might be computed for the limit distribution (2.13) with $\Sigma = I$ and

$$b = \frac{a_1}{\left(a_1' a_1\right)^{1/2}}$$

where $Y_2(b) =^d X_2(b) \equiv N\left(0, \frac{\sigma_{22}}{2b}\right)$ *and* $Y_2(b)$ *and* $X_2(b)$ *are independent. The residual moment matrix* $\tilde{\Sigma}$ *of the restricted regression has the following asymptotic behavior under the alternative hypothesis:*

$$\tilde{\Sigma} = \frac{1}{n}\sum_{t=1}^{n} u_t u_t' + \frac{b^2 n}{k_n^2}\left[\begin{array}{cc} \frac{1}{n^2}\sum_{t=1}^{n} X_{1t-1}^2 & 0 \\ 0 & 0 \end{array}\right]\{1 + o_p(1)\}. \tag{3.4}$$

Since $\breve{\Sigma} = n^{-1}\sum_{t=1}^{n} u_t u_t' \to_p \Sigma$, it follows from (3.2) that $\tilde{\Sigma}$ is consistent for Σ under the null. However, from (3.4) and the fact that $n^{-2}\sum_{t=1}^{n} X_{1t-1}^2 \Rightarrow \int_0^1 J_{1c}^2$, it is apparent that $\tilde{\Sigma}$ is consistent for Σ when $n = o\left(k_n^2\right)$ but is inconsistent when $\frac{k_n^2}{n} = O(1)$ and, in particular, when $k_n = o\left(n^{1/2}\right)$. These results enable us to determine conditions for the consistency of model selection criteria such as the Schwarz criterion (BIC).

For the model (2.1), the restricted regression and unrestricted regression BIC criteria are:

$$\mathrm{BIC}_r = \log\left|\tilde{\Sigma}\right| + \frac{\log n}{n}, \quad \mathrm{BIC}_u = \log\left|\hat{\Sigma}\right| + 4\frac{\log n}{n}.$$

When the null holds and $R_n = r_n I$ it is evident that

$$\mathrm{BIC}_r = \log\left|\tilde{\Sigma}\right| + \frac{\log n}{n} = \log\left|\breve{\Sigma}\right| + \frac{\log n}{n} + O_p\left(\frac{1}{n}\right), \tag{3.5}$$

whereas for the unrestricted regression

$$\mathrm{BIC}_u = \log\left|\hat{\Sigma}\right| + 4\frac{\log n}{n} = \log\left|\breve{\Sigma}\right| + 4\frac{\log n}{n} + O_p\left(\frac{1}{n}\right) \tag{3.6}$$

since $\hat{\Sigma} = \breve{\Sigma} + O_p\left(n^{-1}\right)$ analogous to the proof of (3.2). In view of (3.5) and (3.6), $\mathrm{BIC}_r < \mathrm{BIC}_u$ up to a term of $O_p\left(\frac{1}{n}\right)$. The restricted model will therefore be correctly chosen with probability approaching unity under the null.

When the alternative holds, (3.6) continues to apply for the unrestricted regression. But under the alternative for the restricted regression we have from (3.4)

$$\begin{aligned} \log\left|\tilde{\Sigma}\right| &= \log\left|\breve{\Sigma} + \frac{b^2 n}{k_n^2}\left[\begin{array}{cc} n^{-2}\sum_{t=1}^{n} X_{1t-1}^2 & 0 \\ 0 & 0 \end{array}\right]\{1 + o_p(1)\}\right| \\ &= \log\left|\breve{\Sigma}\right| + \log\left|I + \frac{b^2 n}{k_n^2}\breve{\Sigma}^{-1}\left[\begin{array}{cc} n^{-2}\sum_{t=1}^{n} X_{1t-1}^2 & 0 \\ 0 & 0 \end{array}\right]\{1 + o_p(1)\}\right| \end{aligned}$$

$$\begin{aligned}
&= \log\left|\breve{\Sigma}\right| + \frac{b^2 n}{k_n^2}\text{tr}\left\{\breve{\Sigma}^{-1}\left[\begin{array}{cc} n^{-2}\sum_{t=1}^{n}X_{1t-1}^2 & 0 \\ 0 & 0\end{array}\right]\right\}\{1+o_p(1)\} \\
&= \log|\Sigma| + \frac{b^2 n}{k_n^2}\text{tr}\left\{\Sigma^{-1}\left[\begin{array}{cc} n^{-2}\sum_{t=1}^{n}X_{1t-1}^2 & 0 \\ 0 & 0\end{array}\right]\right\}\{1+o_p(1)\} \\
&= \log|\Sigma| + \frac{b^2 n}{k_n^2}\frac{n^{-2}\sum_{t=1}^{n}X_{1t-1}^2}{\sigma_{11.2}}\{1+o_p(1)\},
\end{aligned}$$

where $\sigma_{11.2} = \sigma_{11} - \sigma_{12}/\sigma_{22}$. Then

$$\begin{aligned}
\text{BIC}_r &= \log\left|\tilde{\Sigma}\right| + \frac{\log n}{n} \\
&= \log\left|\breve{\Sigma}\right| + \frac{b^2 n}{k_n^2}\frac{n^{-2}\sum_{t=1}^{n}X_{1t-1}^2}{\sigma_{11.2}}\{1+o_p(1)\} + \frac{\log n}{n}.
\end{aligned}$$

It follows that $\text{BIC}_r > \text{BIC}_u$ under the alternative as $n \to \infty$ whenever

$$\frac{b^2 n}{k_n^2}\frac{n^{-2}\sum_{t=1}^{n}X_{1t-1}^2}{\sigma_{11.2}} > 3\frac{\log n}{n},$$

which inequality holds with probability approaching unity provided $\frac{n^2}{k_n^2 \log n} \to \infty$ as $n \to \infty$ because $n^{-2}\sum_{t=1}^{n}X_{1t-1}^2 \Rightarrow \int_0^1 J_{1c}^2 > 0$ with probability one. Hence, under the alternative, the unrestricted model will be chosen with probability approaching unity as $n \to \infty$ provided k_n goes to infinity slower than $n/(\log n)^{1/2}$, that is provided $\frac{k_n(\log n)^{1/2}}{n} \to 0$.

It follows that model selection by BIC is consistent and as $n \to \infty$ the criterion will successfully distinguish roots in the vicinity of unity provided one of the roots $\theta_n = 1 + \frac{b}{k_n}$ is mildly explosive and sufficiently different from local to unity in the sense that $k_n \to \infty$ slower than $O\left(\frac{n}{L_n}\right)$ where L_n is a slowly varying function that diverges at least as fast as $(\log n)^{1/2}$, i.e., $\liminf_{n\to\infty} \frac{L_n}{(\log n)^{1/2}} > 0$. In this respect, the discriminatory capability of model selection is analogous to that of classical Wald testing.

4 Conclusion

Model selection by BIC is well known to be blind to local alternatives in general (see Ploberger and Phillips, 2003; and Leeb and Poetscher, 2005). For instance, in the current set up, BIC cannot consistently distinguish between a model with a unit root ($\rho_n = 1$) and models with roots local to unity ($\rho_n = 1 + \frac{c}{n}$), just as localizing coefficients such as the parameter c are not consistently estimable. On the other hand, as shown here, BIC and classical tests can successfully distinguish roots in the immediate locality of unity like ρ_n from roots that are in the wider vicinity of unity like θ_n, which opens the door to distinguishing mildly explosive behavior in data. We expect these model selection results to be generalizable to models with weakly dependent innovations, analogous to the findings in Phillips (2008) on unit root discrimination and Cheng and Phillips (2009) for cointegrating rank determination.

Tests of this type will be useful in empirical work where it is of interest to differentiate between the behavioral time series character of financial data such as asset prices and the fundamentals that are believed to determine prices, like dividends and earnings. In such cases, the primary maintained hypothesis is that the series have roots that are local to unity (without being specific about the localizing coefficient) and the alternative is that one or other of the series may be mildly explosive at least over subperiods of data (see Phillips, Wu and Yu, 2011; Phillips and Yu, 2011). On the other hand, if the primary maintained hypothesis is that both series may be mildly explosive and the null hypothesis is commonality in the roots, then problems of bias and inconsistency may arise in testing and model selection. Recent work by Nielsen (2009) and Phillips and Magdalinos (2011) provide a limit theory for least squares regression in the case of purely explosive common roots and show that least squares regression is inconsistent. That work may be extended to the case of common mildly explosive roots and will be explored in later work.

5 Appendix

5.1 Preliminary Lemmas

We start with some lemmas that assist in the asymptotic development. These results rely on existing limit theory so we only sketch the main details here for convenience. We repeatedly use the fact that $k_n(\theta_n^2 - 1) = 2b + O(\frac{1}{k_n})$ and $\theta_n^{-n} = \exp(-b\frac{n}{k_n})\{1 + o(1)\} = o(1)$. The first result is from PM7a. See also Phillips and Magdalinos (2008) and Magdalinos and Phillips (2009) for related results on systems with explosive and mildly explosive processes.

Lemma 5.1 *(PM7a) Define*

$$
\begin{aligned}
X_n(b) &= \begin{bmatrix} X_{1n}(b) \\ X_{2n}(b) \end{bmatrix} := \frac{1}{\sqrt{k_n}} \sum_{j=1}^{n} \theta_n^{-j} u_j, \\
Y_n(b) &= \begin{bmatrix} Y_{1n}(b) \\ Y_{2n}(b) \end{bmatrix} := \frac{1}{\sqrt{k_n}} \sum_{j=1}^{n} \theta_n^{-(n-j)-1} u_j.
\end{aligned}
$$

Then, as $n \to \infty$, $X_n(b) \Rightarrow X(b) = (X_1(b), X_2(b))' \equiv N(0, \frac{1}{2b}\Sigma)$, *and* $Y_n(b) \Rightarrow Y(b) = (Y_1(b), Y_2(b))'$, *where* $Y(b) =^d X(b)$, *and* $X(b)$ *and* $Y(b)$ *are independent.*

Lemma 5.2 *Define* $S_n(r) := \frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor nr \rfloor} u_j$ *and*

$$
\begin{aligned}
X_{1n}^c(r) &= \frac{X_{1\lfloor nr \rfloor}}{\sqrt{n}} = \frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor nr \rfloor} \rho_n^j u_{1\lfloor nr \rfloor - j}, \\
X_{2n}(b) &= \frac{X_{2n}}{\sqrt{k_n}\theta_n^n} = \frac{1}{\sqrt{k_n}} \sum_{j=1}^{n} \frac{u_{2j}}{\theta_n^j}.
\end{aligned}
$$

Then, as $n \to \infty$,

(i) $S_n(r) = \begin{bmatrix} \frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor nr \rfloor} u_{1j} \\ \frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor nr \rfloor} u_{2j} \end{bmatrix} \Rightarrow \begin{bmatrix} B_1(r) \\ B_2(r) \end{bmatrix} = B(r) \equiv BM(\Sigma)$;

(ii) $X_{1n}^c(r) \Rightarrow J_{1c}(r) = \int_0^r e^{c(r-s)} dB_1(s)$ *and* $n^{-1} \sum_{j=1}^{n} X_{1t-1} u_t \Rightarrow \int_0^1 J_{1c}(r) dB(r)$;

(iii) $X_{2n}(b) \Rightarrow X_2(b)$, *where* $X_2(b) \equiv N\left(0, \frac{\sigma_{22}}{2b}\right)$;

(iv) $J_{1c}(r)$ *and* $X_2(b)$ *are independent.*

(v) For all $s, r > 0$ *the following joint convergence applies:*

$$
\left[\frac{X_{1\lfloor nr \rfloor}}{\sqrt{n}}, \frac{X_{2\lfloor ns \rfloor}}{\sqrt{k_n}\theta_n^{\lfloor ns \rfloor}} \right] \Rightarrow \left[J_{1c}(r), X_2(b) \right], \quad \text{as } n \to \infty.
$$

Proof. Result (i) is standard, (ii) is from Phillips (1987b), and (iii) is from lemma 5.1. To prove (iv), it suffices to show that $B_1(r)$ and $X_2(b)$ are independent, since $J_{1c}(r)$ is a functional of $\{B_1(s)\}_{s\leq r}$. Note that the covariance

$$\begin{aligned} E\left(S_{1n}(1)X_{2n}(b)\right) &= E\left[\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{n}u_{1j}\right)\left(\frac{1}{\sqrt{k_n}}\sum_{k=1}^{n}\frac{u_{2k}}{\theta_n^k}\right)\right] \\ &= \frac{\sigma_{12}}{\sqrt{nk_n}}\sum_{k=1}^{n}\frac{1}{\theta_n^k} = \frac{\sigma_{12}}{\sqrt{nk_n}}\frac{1}{\theta_n}\left(\frac{1-\theta_n^{-n}}{1-\theta_n^{-1}}\right) \\ &= \frac{\sigma_{12}}{\sqrt{nk_n}}\frac{1}{\theta_n-1}\left\{1+o(1)\right\} = \frac{\sigma_{12}}{b}\sqrt{\frac{k_n}{n}}\left\{1+o(1)\right\} = o(1), \end{aligned}$$

as $n\to\infty$. Independence of the limit processes $J_{1c}(r)$ and $X_2(b)$ follows. To prove (v), first observe that for any (integer sequence) $L_n\to\infty$ such that $\frac{L_n}{k_n}\to\infty$, we have $\frac{X_{2L_n}}{\sqrt{k_n}\theta_n^{L_n}}\Rightarrow X_2(b)$. Note that $X_{2n}(b)=\frac{X_{2L_n}}{\sqrt{k_n}\theta_n^{L_n}}+\frac{1}{\sqrt{k_n}}\sum_{j=L_n+1}^{n}\frac{u_{2j}}{\theta_n^j}$ and

$$\begin{aligned} E\left|\frac{1}{\sqrt{k_n}}\sum_{j=L_n+1}^{n}\frac{u_{2j}}{\theta_n^j}\right|^2 &= \frac{1}{k_n}\sum_{j=L_n+1}^{n}\frac{\sigma_{22}}{\theta_n^{2j}} = \frac{\sigma_{22}}{k_n}\frac{1}{\theta_n^{2L_n+2}}\left(\frac{1-\theta_n^{-2n+2L_n}}{1-\theta_n^{-2}}\right) \\ &= \frac{\sigma_{22}}{k_n\left(\theta_n^2-1\right)}\left(\theta_n^{-2L_n}-\theta_n^{-2n}\right) = o(1), \end{aligned}$$

since $\theta_n^{-2L_n}=\left(1+\frac{b}{k_n}\right)^{-2L_n}=\left\{\left(1+\frac{b}{k_n}\right)^{k_n}\right\}^{-2\frac{L_n}{k_n}}=\exp(-2b\frac{L_n}{k_n})+o(1)=o(1)$. Hence, $\frac{X_{2L_n}}{\sqrt{k_n}\theta_n^{L_n}}\Rightarrow X_2(b)$ by lemma 5.1. Now let $L_n=\lfloor ns\rfloor$ for any $s>0$ and then $\frac{X_{2\lfloor ns\rfloor}}{\sqrt{k_n}\theta_n^{\lfloor ns\rfloor}}\Rightarrow X_2(b)$. Joint convergence and (v) follow from marginal convergence and asymptotic independence of the components. ■

Lemma 5.3 *As* $n\to\infty$,

(i) $\frac{1}{k_n^2\theta_n^{2n}}\sum_{t=1}^{n}X_{2t-1}^2\Rightarrow\frac{(X_2(b))^2}{2b}$,

(ii) $\frac{1}{n^2}\sum_{t=1}^{n}X_{1t-1}^2\Rightarrow\int_0^1 J_{1c}(r)^2dr$,

(iii) $\frac{1}{nk_n\theta_n^n}\sum_{t=1}^{n}X_{1t-1}X_{2t-1}=o_p(1)$.

Proof. (i) follows from PM7a and (ii) is standard (Phillips, 1987a&b). For (iii), it is convenient to take a probability space where $\left[\frac{X_{1\lfloor nr\rfloor}}{\sqrt{n}},\frac{X_{2\lfloor ns\rfloor}}{\sqrt{k_n}\theta_n^{\lfloor ns\rfloor}}\right]\to_p\left[J_{1c}(r),X_2(b)\right]$. Then, for any sequence $L_n \to \infty$ such that $\frac{L_n}{n} \to 0$, we have

$$\begin{aligned}
\frac{1}{nk_n\theta_n^n}\sum_{t=1}^{n} X_{1t-1}X_{2t-1} &= \frac{1}{\sqrt{nk_n}\theta_n^n}\left\{\sum_{t=1}^{L_n} + \sum_{t=L_n+1}^{n}\right\}\left(\frac{X_{1t-1}}{\sqrt{n}}\right)\left(\frac{X_{2t-1}}{\sqrt{k_n}\theta_n^{t-1}}\right)\theta_n^{t-1} \\
&= \frac{X_2(b)}{\sqrt{nk_n}\theta_n^n}\sum_{t=L_n+1}^{n}\left(J_{1c}\left(\frac{t}{n}\right)\right)\theta_n^{t-1}\left\{1+o_p(1)\right\} \\
&\quad + \frac{\theta_n^{L_n}}{\sqrt{nk_n}\theta_n^n}\sum_{t=1}^{L_n}\left(\frac{X_{1t-1}}{\sqrt{n}}\right)\left(\frac{X_{2t-1}}{\sqrt{k_n}\theta_n^{t-1}}\right)\frac{\theta_n^{t-1}}{\theta_n^{L_n}}
\end{aligned}$$

$$\begin{aligned}
&= \frac{X_2(b)}{\sqrt{nk_n}\theta_n^n}\sum_{t=L_n+1}^{n}\left(J_{1c}\left(\frac{t}{n}\right)\right)\theta_n^{t-1}\left\{1+o_p(1)\right\} + O_p\left(\frac{L_n\theta_n^{L_n}}{\sqrt{nk_n}\theta_n^n}\right) \\
&= \frac{X_2(b)}{\sqrt{nk_n}\theta_n^n}\sum_{t=1}^{n}\left(J_{1c}\left(\frac{t}{n}\right)\right)\theta_n^{t-1} + o_p(1).
\end{aligned}$$

Now $\sum_{t=1}^{n}\left(J_{1c}\left(\frac{t}{n}\right)\right)\theta_n^{t-1}$ has zero mean and variance

$$\begin{aligned}
E\left(\sum_{t=1}^{n}\left(J_{1c}\left(\frac{t}{n}\right)\right)\theta_n^{t-1}\right)^2 &= \sum_{t=1}^{n}\sum_{s=1}^{n}E\left(J_{1c}\left(\frac{t}{n}\right)J_{1c}\left(\frac{s}{n}\right)\right)\theta_n^{t+s-2} \\
&\leq M\left(\frac{\theta_n^n-1}{\theta_n-1}\right)^2 \leq M'\frac{k_n^2\theta_n^{2n}}{b^2},
\end{aligned}$$

for some finite constants M and M'. It follows that

$$\mathrm{Var}\left(\frac{1}{\sqrt{nk_n}\theta_n^n}\sum_{t=L_n}^{n}\left(J_{1c}\left(\frac{t}{n}\right)\right)\theta_n^{t-1}\right) = O\left(\frac{k_n^2\theta_n^{2n}}{nk_n\theta_n^{2n}}\right) = O\left(\frac{k_n}{n}\right) = o(1),$$

leading to $\frac{1}{\sqrt{nk_n}\theta_n^n}\sum_{t=L_n}^{n}\left(J_{1c}\left(\frac{t}{n}\right)\right)\theta_n^{t-1} = o_p(1)$, which implies that $\frac{1}{nk_n\theta_n^n}\sum_{t=1}^{n}X_{1t-1}X_{2t-1} = o_p(1)$ and this also holds in the original probability space, giving the required result. ■

Lemma 5.4 *As $n \to \infty$,*

(i) $D_n^{-1}X_{-1}'X_{-1}D_n^{-1} \Rightarrow \begin{bmatrix} \int_0^1 J_{1c}(r)^2dr & 0 \\ 0 & \frac{(X_2(b))^2}{2b} \end{bmatrix}$,

(ii) $u'X_{-1}D_n^{-1} \Rightarrow \begin{bmatrix} \int_0^1 J_{1c}(r)dB(r) & X_2(b)Y(b) \end{bmatrix}$.

Proof. Using lemma 5.3

$$
\begin{aligned}
D_n^{-1} X_{-1}' X_{-1} D_n^{-1} &= D_n^{-1} \left(\sum_{t=1}^{n} X_{t-1} X_{t-1}' \right) D_n^{-1} \\
&= \begin{bmatrix} \frac{1}{n^2} \sum_{t=1}^{n} X_{1t-1}^2 & \frac{1}{n k_n \theta_n^n} \sum_{t=1}^{n} X_{1t-1} X_{2t-1} \\ \frac{1}{n k_n \theta_n^n} \sum_{t=1}^{n} X_{2t-1} X_{1t-1} & \frac{1}{k_n^2 \theta_n^{2n}} \sum_{t=1}^{n} X_{2t-1}^2 \end{bmatrix} \\
&\Rightarrow \begin{bmatrix} \int_0^1 J_{1c}(r)^2 dr & 0 \\ 0 & \frac{(X_2(b))^2}{2b} \end{bmatrix},
\end{aligned}
$$

giving (i). Result (ii) follows directly from lemmas 5.2 and 5.3 as

$$
\begin{aligned}
u' X_{-1} D_n^{-1} &= \begin{bmatrix} \frac{1}{n} \sum_{t=1}^{n} X_{1t-1} u_t & \frac{1}{k_n \theta_n^n} \sum_{t=1}^{n} X_{2t-1} u_t \end{bmatrix} \\
&= \begin{bmatrix} \sum_{t=1}^{n} \left(\frac{X_{1t-1}}{\sqrt{n}} \right) \left(\frac{u_t}{\sqrt{n}} \right) & \frac{1}{\sqrt{k_n} \theta_n^n} \sum_{t=1}^{n} \left(\frac{X_{2t-1}}{\sqrt{k_n} \theta_n^{t-1}} \right) u_t \theta_n^{t-1} \end{bmatrix} \\
&\Rightarrow \begin{bmatrix} \int_0^1 J_{1c}(r) dB(r) & X_2(b) Y(b) \end{bmatrix}.
\end{aligned}
$$

Joint convergence follows from the independence between $B(r)$ and $(X_2(b), Y(b))$. ■

5.2 Proofs of the Main Results

Proof of Theorem 2.1. Using Lemma 5.4, continuous mapping and joint convergence, we have

$$
\left(\hat{R}_n - R_n \right) D_n = \left(u' X_{-1} D_n^{-1} \right) \left(D_n^{-1} X_{-1}' X_{-1} D_n^{-1} \right)^{-1} \Rightarrow \begin{bmatrix} \frac{\int_0^1 J_{1c}(r) dB(r)}{\int_0^1 J_{1c}(r)^2 dr} & \frac{Y(b)}{X_2(b)/2b} \end{bmatrix}.
$$

Since $(\theta_n^2 - 1) = \frac{2b}{k_n}(1 + o(1))$ the equivalent result

$$
\left(\hat{R}_n - R_n \right) F_n \Rightarrow \begin{bmatrix} \frac{\int_0^1 J_{1c}(r) dB(r)}{\int_0^1 J_{1c}(r)^2 dr} & \frac{Y(b)}{X_2(b)} \end{bmatrix},
$$

holds as stated. ■

Proof of Theorem 2.2. We first prove (2.10) and (2.12) for the statistic W_n. Under the null we have by standard theory

$$
n \left(\hat{R}_n - R_n \right) \Rightarrow \int_0^1 dB J_c' \left(\int_0^1 J_c J_c' \right)^{-1} =: \Xi, \quad n^2 \left(X_{-1}' X_{-1} \right)^{-1} \Rightarrow \int_0^1 J_c J_c' \tag{5.1}
$$

$\hat{\Sigma} = n^{-1}\sum_{t=1}^{n}\hat{u}_t\hat{u}_t' \to_p \Sigma$, and (2.10) follows directly for W_n and (2.11) for W_n^A. Under the alternative from theorem 2.1 with correct centering we have

$$a_1' vec\left\{\left(\hat{R}_n - R_n\right)F_n\right\} = \; n\left(\widehat{r}_{11} - r_{11}\right) - \tfrac{\theta_n^n}{(\theta_n^2-1)}\left(\widehat{r}_{22} - r_{22}\right) \Rightarrow a_1' vec\Phi,$$

whereas under (2.2) with $b > 0$, the null centred linear combination behaves as

$$\begin{aligned} a_1' vec\left(n\hat{R}_n\right) &= n(\widehat{r}_{11} - \widehat{r}_{22}) = n(\widehat{r}_{11} - r_{11}) - n\left(\widehat{r}_{22} - r_{22}\right) + n(r_{11} - r_{22}) \\ &= n(\widehat{r}_{11} - r_{11}) - \frac{\theta_n^n}{(\theta_n^2 - 1)}\left(\widehat{r}_{22} - r_{22}\right)\frac{n(\theta_n^2 - 1)}{\theta_n^n} + \left(c - \frac{nb}{k_n}\right) \\ &= n(\widehat{r}_{11} - r_{11}) + \left(c - \frac{nb}{k_n}\right) + o_p(1) \\ &= n(\widehat{r}_{11} - r_{11}) + O_p(\frac{n}{k_n}) \to -\infty, \text{ as } n \to \infty, \end{aligned}$$

in view of (2.7) - (2.9) and since $\frac{n(\theta_n^2-1)}{\theta_n^n} = \frac{\frac{n}{k_n}}{\theta_n^n}k_n(\theta_n^2 - 1) = O(\frac{\frac{n}{k_n}}{\exp(b\frac{n}{k_n})}) = o(1)$. Next, setting $d_n = \left(\sum_{t=1}^n X_{1t-1}^2 \sum_{t=1}^n X_{2t-1}^2\right) - \left(\sum_{t=1}^n X_{1t-1}X_{2t-1}\right)^2$ and using Lemma 4.3 we find that

$$\begin{aligned} d_n &= \sum_{t=1}^n X_{1t-1}^2 \sum_{t=1}^n X_{2t-1}^2 \left\{1 - \frac{\left(\frac{1}{nk_n\theta_n^n}\sum_{t=1}^n X_{1t-1}X_{2t-1}\right)^2}{\frac{1}{n^2}\sum_{t=1}^n X_{1t-1}^2 \frac{1}{k_n^2\theta_n^{2n}}\sum_{t=1}^n X_{2t-1}^2}\right\} \\ &= \sum_{t=1}^n X_{1t-1}^2 \sum_{t=1}^n X_{2t-1}^2 \left\{1 - o_p(1)\right\}, \end{aligned} \tag{5.2}$$

and

$$\begin{aligned} \frac{d_n}{n^2 k_n^2 \theta_n^{2n}} &= \frac{1}{n^2}\sum_{t=1}^n X_{1t-1}^2 \frac{1}{k_n^2\theta_n^{2n}}\sum_{t=1}^n X_{2t-1}^2 \left\{1 - o_p(1)\right\} \\ &\Rightarrow \left(\int_0^1 J_{1c}(r)^2 dr\right)\left(\frac{X(b)^2}{2b}\right). \end{aligned}$$

It follows that

$$
\begin{aligned}
n^2\left(X_{-1}'X_{-1}\right)^{-1} &= \frac{n^2}{d_n}\left[\begin{array}{cc} \sum_{t=1}^n X_{2t-1}^2 & -\sum_{t=1}^n X_{1t-1}X_{2t-1} \\ -\sum_{t=1}^n X_{1t-1}X_{2t-1} & \sum_{t=1}^n X_{1t-1}^2 \end{array}\right] \\
&= \left[\begin{array}{cc} \frac{n^2}{\sum_{t=1}^n X_{1t-1}^2} & -\frac{n^2\sum_{t=1}^n X_{1t-1}X_{2t-1}}{\sum_{t=1}^n X_{1t-1}^2\sum_{t=1}^n X_{2t-1}^2} \\ -\frac{n^2\sum_{t=1}^n X_{1t-1}X_{2t-1}}{\sum_{t=1}^n X_{1t-1}^2\sum_{t=1}^n X_{2t-1}^2} & \frac{n^2}{\sum_{t=1}^n X_{2t-1}^2} \end{array}\right]\{1+o_p(1)\} \qquad (5.3) \\
&= \left[\begin{array}{cc} \left(\frac{\sum_{t=1}^n X_{1t-1}^2}{n_2}\right)^{-1}+o_p(1) & o_p(1) \\ o_p(1) & o_p(1) \end{array}\right] \Rightarrow \left[\begin{array}{cc} \left(\int_0^1 J_{1c}(r)^2dr\right)^{-1} & 0 \\ 0 & 0 \end{array}\right].
\end{aligned}
$$

Since $\hat{\Sigma} \rightarrow_p \Sigma$, we have

$$
\begin{aligned}
& n^2 a_1'\left\{\hat{\Sigma}\otimes\left(X_{-1}'X_{-1}\right)^{-1}\right\}a_1 \\
=\ & a_1'\left\{(\Sigma+o_p(1))\otimes\left[\begin{array}{cc} \left(\frac{\sum_{t=1}^n X_{1t-1}^2}{n_2}\right)^{-1}+o_p(1) & o_p(1) \\ o_p(1) & o_p(1) \end{array}\right]\right\}a_1 \\
\Rightarrow\ & a_1'\left\{\Sigma\otimes\left[\begin{array}{cc} \left(\int_0^1 J_{1c}(r)^2dr\right)^{-1} & 0 \\ 0 & 0 \end{array}\right]\right\}a_1 \\
=\ & \sigma_{11}\left(\int_0^1 J_{1c}(r)^2dr\right)^{-1}.
\end{aligned}
$$

It follows that

$$
\begin{aligned}
W_n &= \left(a_1' vec\left(\hat{R}_n\right)\right)^2 / a_1'\left\{\hat{\Sigma}\otimes\left(X_{-1}'X_{-1}\right)^{-1}\right\}a_1 \\
&= \frac{\left\{n(\widehat{r}_{11}-r_{11})+O_p(\frac{n}{k_n})\right\}^2}{\sigma_{11}\left(\int_0^1 J_{1c}(r)^2dr\right)^{-1}+o_p(1)} = O_p\left(\frac{n^2}{k_n^2}\right),
\end{aligned}
$$

giving the stated result.

The proof of (2.12) for the statistic W_n^A under the alternative follows the same lines but involves more complex calculations to cope with different orders of magnitude in the components. First consider the behavior of the centred elements under the alternative. By

(2.7) - (2.9) we have

$$A'vec\left\{\left(\hat{R}_n - R_n\right)F_n\right\} = \begin{bmatrix} n\left(\widehat{r}_{11} - r_{11}\right) - \frac{\theta_n^n}{(\theta_n^2-1)}\left(\widehat{r}_{22} - r_{22}\right) \\ \frac{\theta_n^n}{(\theta_n^2-1)}\left(\widehat{r}_{12} - r_{12}\right) \\ n\left(\widehat{r}_{21} - r_{21}\right) \end{bmatrix}' \Rightarrow A'vec\Phi.$$

On the other hand under (2.2) with $b > 0$, the null-centred linear combinations behave as follows. First,

$$\begin{aligned} a_1'vec\left(n\hat{R}_n\right) &= n(\widehat{r}_{11} - \widehat{r}_{22}) = n(\widehat{r}_{11} - r_{11}) - n\left(\widehat{r}_{22} - r_{22}\right) + n(r_{11} - r_{22}) \\ &= n(\widehat{r}_{11} - r_{11}) - \frac{\theta_n^n}{(\theta_n^2 - 1)}\left(\widehat{r}_{22} - r_{22}\right)\frac{n(\theta_n^2 - 1)}{\theta_n^n} + \left(c - \frac{nb}{k_n}\right) \\ &= n(\widehat{r}_{11} - r_{11}) + O_p(\frac{n}{k_n}) \to -\infty, \text{ as } n \to \infty, \end{aligned}$$

as for W_n. Second

$$a_2'vec\left(n\hat{R}_n\right) = n\widehat{r}_{12} = \frac{\theta_n^n}{(\theta_n^2 - 1)}\widehat{r}_{12}\frac{n(\theta_n^2 - 1)}{\theta_n^n} = O_p\left(\frac{\frac{n}{k_n}}{\exp(b\frac{n}{k_n})}\right) = o_p\left(1\right),$$

and third

$$a_3{}'vec\left(n\hat{R}_n\right) = n\widehat{r}_{21} \Rightarrow a_3'vec\Phi, \text{ as } n \to \infty.$$

Also, as in (5.3)

$$\left(X_{-1}'X_{-1}\right)^{-1} = \begin{bmatrix} \frac{1}{\sum_{t=1}^n X_{1t-1}^2} & -\frac{\sum_{t=1}^n X_{1t-1}X_{2t-1}}{\sum_{t=1}^n X_{1t-1}^2\sum_{t=1}^n X_{2t-1}^2} \\ -\frac{\sum_{t=1}^n X_{1t-1}X_{2t-1}}{\sum_{t=1}^n X_{1t-1}^2\sum_{t=1}^n X_{2t-1}^2} & \frac{1}{\sum_{t=1}^n X_{2t-1}^2} \end{bmatrix}\left\{1 + o_p(1)\right\}.$$

We now evaluate each of the components of the matrix

$$\begin{aligned} & A'\left\{\hat{\Sigma} \otimes \left(X_{-1}'X_{-1}\right)^{-1}\right\}A \\ = & \begin{bmatrix} a_1' \\ a_2' \\ a_3' \end{bmatrix}\begin{bmatrix} \hat{\sigma}_{11}\left(X_{-1}'X_{-1}\right)^{-1} & \hat{\sigma}_{12}\left(X_{-1}'X_{-1}\right)^{-1} \\ \hat{\sigma}_{12}\left(X_{-1}'X_{-1}\right)^{-1} & \hat{\sigma}_{22}\left(X_{-1}'X_{-1}\right)^{-1} \end{bmatrix}\left[a_1, a_2, a_3\right]. \end{aligned}$$

Using lemma 4.3 we find

$$
\begin{aligned}
& a_1' \begin{bmatrix} \hat{\sigma}_{11}\left(X_{-1}'X_{-1}\right)^{-1} & \hat{\sigma}_{12}\left(X_{-1}'X_{-1}\right)^{-1} \\ \hat{\sigma}_{12}\left(X_{-1}'X_{-1}\right)^{-1} & \hat{\sigma}_{22}\left(X_{-1}'X_{-1}\right)^{-1} \end{bmatrix} a_1 \\
= & \left(\hat{\sigma}_{11}\frac{1}{\sum_{t=1}^n X_{1t-1}^2} + 2\hat{\sigma}_{12}\frac{\sum_{t=1}^n X_{1t-1}X_{2t-1}}{\sum_{t=1}^n X_{1t-1}^2 \sum_{t=1}^n X_{2t-1}^2} + \hat{\sigma}_{22}\frac{1}{\sum_{t=1}^n X_{2t-1}^2}\right)\{1+o_p(1)\} \\
= & \hat{\sigma}_{11}\frac{1}{\sum_{t=1}^n X_{1t-1}^2}\{1+o_p(1)\},
\end{aligned}
$$

$$
\begin{aligned}
& a_1' \begin{bmatrix} \hat{\sigma}_{11}\left(X_{-1}'X_{-1}\right)^{-1} & \hat{\sigma}_{12}\left(X_{-1}'X_{-1}\right)^{-1} \\ \hat{\sigma}_{12}\left(X_{-1}'X_{-1}\right)^{-1} & \hat{\sigma}_{22}\left(X_{-1}'X_{-1}\right)^{-1} \end{bmatrix} a_2 \\
= & -\left(\hat{\sigma}_{11}\frac{\sum_{t=1}^n X_{1t-1}X_{2t-1}}{\sum_{t=1}^n X_{1t-1}^2 \sum_{t=1}^n X_{2t-1}^2} + \hat{\sigma}_{12}\frac{1}{\sum_{t=1}^n X_{2t-1}^2}\right)\{1+o_p(1)\} \\
= & -\hat{\sigma}_{11}\frac{\sum_{t=1}^n X_{1t-1}X_{2t-1}}{\sum_{t=1}^n X_{1t-1}^2 \sum_{t=1}^n X_{2t-1}^2}\{1+o_p(1)\},
\end{aligned}
$$

$$
a_2' \begin{bmatrix} \hat{\sigma}_{11}\left(X_{-1}'X_{-1}\right)^{-1} & \hat{\sigma}_{12}\left(X_{-1}'X_{-1}\right)^{-1} \\ \hat{\sigma}_{12}\left(X_{-1}'X_{-1}\right)^{-1} & \hat{\sigma}_{22}\left(X_{-1}'X_{-1}\right)^{-1} \end{bmatrix} a_2 = \hat{\sigma}_{11}\frac{1}{\sum_{t=1}^n X_{2t-1}^2}\{1+o_p(1)\},
$$

$$
a_3' \begin{bmatrix} \hat{\sigma}_{11}\left(X_{-1}'X_{-1}\right)^{-1} & \hat{\sigma}_{12}\left(X_{-1}'X_{-1}\right)^{-1} \\ \hat{\sigma}_{12}\left(X_{-1}'X_{-1}\right)^{-1} & \hat{\sigma}_{22}\left(X_{-1}'X_{-1}\right)^{-1} \end{bmatrix} a_3 = \hat{\sigma}_{22}\frac{1}{\sum_{t=1}^n X_{1t-1}^2}\{1+o_p(1)\},
$$

$$
\begin{aligned}
& a_1' \begin{bmatrix} \hat{\sigma}_{11}\left(X_{-1}'X_{-1}\right)^{-1} & \hat{\sigma}_{12}\left(X_{-1}'X_{-1}\right)^{-1} \\ \hat{\sigma}_{12}\left(X_{-1}'X_{-1}\right)^{-1} & \hat{\sigma}_{22}\left(X_{-1}'X_{-1}\right)^{-1} \end{bmatrix} a_3 \\
= & \left(\hat{\sigma}_{12}\frac{1}{\sum_{t=1}^n X_{2t-1}^2} - \hat{\sigma}_{22}\frac{\sum_{t=1}^n X_{1t-1}X_{2t-1}}{\sum_{t=1}^n X_{1t-1}^2 \sum_{t=1}^n X_{2t-1}^2}\right)\{1+o_p(1)\} \\
= & -\hat{\sigma}_{22}\frac{\sum_{t=1}^n X_{1t-1}X_{2t-1}}{\sum_{t=1}^n X_{1t-1}^2 \sum_{t=1}^n X_{2t-1}^2}\{1+o_p(1)\},
\end{aligned}
$$

$$
a_2' \begin{bmatrix} \hat{\sigma}_{11}\left(X_{-1}'X_{-1}\right)^{-1} & \hat{\sigma}_{12}\left(X_{-1}'X_{-1}\right)^{-1} \\ \hat{\sigma}_{12}\left(X_{-1}'X_{-1}\right)^{-1} & \hat{\sigma}_{22}\left(X_{-1}'X_{-1}\right)^{-1} \end{bmatrix} a_3 = -\hat{\sigma}_{12}\frac{\sum_{t=1}^n X_{1t-1}X_{2t-1}}{\sum_{t=1}^n X_{1t-1}^2 \sum_{t=1}^n X_{2t-1}^2}\{1+o_p(1)\},
$$

and

$$a_3' \begin{bmatrix} \hat{\sigma}_{11}\left(X_{-1}'X_{-1}\right)^{-1} & \hat{\sigma}_{12}\left(X_{-1}'X_{-1}\right)^{-1} \\ \hat{\sigma}_{12}\left(X_{-1}'X_{-1}\right)^{-1} & \hat{\sigma}_{22}\left(X_{-1}'X_{-1}\right)^{-1} \end{bmatrix} a_3 = \hat{\sigma}_{22}\frac{1}{\sum_{t=1}^{n} X_{2t-1}^2}\left\{1+o_p\left(1\right)\right\}.$$

Hence

$$\begin{aligned} & A'\left\{\hat{\Sigma}\otimes\left(X_{-1}'X_{-1}\right)^{-1}\right\}A \\ = & \begin{bmatrix} \hat{\sigma}_{11}\frac{1}{\sum_{t=1}^{n}X_{1t-1}^2} & -\hat{\sigma}_{11}\frac{\sum_{t=1}^{n}X_{1t-1}X_{2t-1}}{\sum_{t=1}^{n}X_{1t-1}^2\sum_{t=1}^{n}X_{2t-1}^2} & -\hat{\sigma}_{22}\frac{\sum_{t=1}^{n}X_{1t-1}X_{2t-1}}{\sum_{t=1}^{n}X_{1t-1}^2\sum_{t=1}^{n}X_{2t-1}^2} \\ & \hat{\sigma}_{22}\frac{1}{\sum_{t=1}^{n}X_{1t-1}^2} & -\hat{\sigma}_{12}\frac{\sum_{t=1}^{n}X_{1t-1}X_{2t-1}}{\sum_{t=1}^{n}X_{1t-1}^2\sum_{t=1}^{n}X_{2t-1}^2} \\ & & \hat{\sigma}_{22}\frac{1}{\sum_{t=1}^{n}X_{2t-1}^2} \end{bmatrix}\left\{1+o_p\left(1\right)\right\}. \end{aligned}$$

Set $K_n = diag(n, n, k_n\theta_n^n)$ and observe that

$$\begin{aligned} & K_nA'\left\{\hat{\Sigma}\otimes\left(X_{-1}'X_{-1}\right)^{-1}\right\}AK_n \\ = & \begin{bmatrix} \hat{\sigma}_{11}\frac{n^2}{\sum_{t=1}^{n}X_{1t-1}^2} & o_p\left(1\right) & o_p\left(1\right) \\ & \hat{\sigma}_{22}\frac{n^2}{\sum_{t=1}^{n}X_{1t-1}^2} & o_p\left(1\right) \\ & & \hat{\sigma}_{22}\frac{k_n^2\theta_n^{2n}}{\sum_{t=1}^{n}X_{2t-1}^2} \end{bmatrix}\left\{1+o_p\left(1\right)\right\} \end{aligned}$$

since

$$\begin{aligned} \frac{n^2\sum_{t=1}^{n}X_{1t-1}X_{2t-1}}{\sum_{t=1}^{n}X_{1t-1}^2\sum_{t=1}^{n}X_{2t-1}^2} &= o_p\left(1\right), \\ nk_n\theta_n^n\frac{\sum_{t=1}^{n}X_{1t-1}X_{2t-1}}{\sum_{t=1}^{n}X_{1t-1}^2\sum_{t=1}^{n}X_{2t-1}^2} &= \frac{\frac{1}{nk_n\theta_n^n}\sum_{t=1}^{n}X_{1t-1}X_{2t-1}}{\frac{1}{n^2}\sum_{t=1}^{n}X_{1t-1}^2\frac{1}{k_n^2\theta_n^{2n}}\sum_{t=1}^{n}X_{2t-1}^2} = o_p\left(1\right), \end{aligned}$$

by lemma 4.3(iii). We deduce that

$$
\begin{aligned}
W_n^A &= \left(A'vec\left(\hat{R}_n\right)\right)'\left(A'\left\{\hat{\Sigma}\otimes\left(X_{-1}'X_{-1}\right)^{-1}\right\}A\right)^{-1}\left(A'vec\left(\hat{R}_n\right)\right) \\
&= \left(A'vec\left(\hat{R}_n\right)\right)'K_n\left(K_nA'\left\{\hat{\Sigma}\otimes\left(X_{-1}'X_{-1}\right)^{-1}\right\}AK_n\right)^{-1}K_n\left(A'vec\left(\hat{R}_n\right)\right) \\
&= \left(A'vec\left(\hat{R}_n\right)\right)'K_n\begin{bmatrix}\sigma_{11}\frac{n^2}{\sum_{t=1}^n X_{1t-1}^2} & 0 & 0 \\ 0 & \sigma_{22}\frac{n^2}{\sum_{t=1}^n X_{1t-1}^2} & 0 \\ 0 & & \sigma_{22}\frac{k_n^2\theta_n^{2n}}{\sum_{t=1}^n X_{2t-1}^2}\end{bmatrix}^{-1} \\
&\quad\times K_n\left(A'vec\left(\hat{R}_n\right)\right)\left\{1+o_p(1)\right\}. \qquad (5.4)
\end{aligned}
$$

Next

$$
\begin{aligned}
A'vec\left(\hat{R}_n\right)K_n &= A'vec\left(\hat{R}_n-R_n\right)K_n+A'vec\left(R_n\right)K_n \\
&= \begin{bmatrix} n\left(\widehat{r}_{11}-r_{11}\right)-n\left(\widehat{r}_{22}-r_{22}\right) \\ n\left(\widehat{r}_{12}-r_{12}\right) \\ k_n\theta_n^n\left(\widehat{r}_{21}-r_{21}\right)\end{bmatrix}'+\begin{bmatrix} n\left(r_{11}-r_{22}\right) \\ 0 \\ 0\end{bmatrix}' \\
&= \begin{bmatrix} n\left(\widehat{r}_{11}-r_{11}\right)+O_p\left(\frac{n}{k_n\theta_n^n}\right) \\ O_p\left(\frac{n}{k_n\theta_n^n}\right) \\ k_n\theta_n^n\left(\widehat{r}_{21}-r_{21}\right)\end{bmatrix}'+\begin{bmatrix} c-\frac{n}{k_n}b \\ 0 \\ 0\end{bmatrix}', \qquad (5.5)
\end{aligned}
$$

from Theorem 2.1 and (2.7) - (2.9). It now follows from (5.4) and (5.5) that

$$
\begin{aligned}
W_n^A &= \left(-\frac{n}{k_n}b+O_p\left(1\right),o_p\left(1\right),O_p\left(1\right)\right)\begin{bmatrix}\sigma_{11}\frac{n^2}{\sum_{t=1}^n X_{1t-1}^2} & 0 & 0 \\ 0 & \sigma_{22}\frac{n^2}{\sum_{t=1}^n X_{1t-1}^2} & 0 \\ 0 & & \sigma_{22}\frac{k_n^2\theta_n^{2n}}{\sum_{t=1}^n X_{2t-1}^2}\end{bmatrix}^{-1} \\
&\quad\times\begin{bmatrix}-\frac{n}{k_n}b+O_p\left(1\right) \\ o_p\left(1\right) \\ O_p\left(1\right)\end{bmatrix} \\
&= \left(-\frac{n}{k_n}b\right)^2\left(\int_0^1 J_{1c}(r)^2dr\right)/\sigma_{11}\left\{1+o_p\left(1\right)\right\},
\end{aligned}
$$

giving the stated result. ■

Proof of Lemma 3.1. Part (i) follows by standard methods in view of Lemmas 5.2 - 5.5. Also $\tilde{u}_t = X_t - \hat{r}_n X_{t-1} = u_t - (\hat{r}_n - r_n) X_{t-1}$, and so we have

$$\begin{aligned}\tilde{\Sigma} &= n^{-1}\sum_{t=1}^{n} u_t u_t' + (\hat{r}_n - r_n) n^{-1} \sum_{t=1}^{n} \left(X_{t-1}u_t' + u_t X_{t-1}'\right) + (\hat{r}_n - r_n)^2 n^{-1} \sum_{t=1}^{n} X_{t-1}X_{t-1}' \\ &= n^{-1}\sum_{t=1}^{n} u_t u_t' + O_p\left(\frac{1}{n}\right), \qquad (5.6)\end{aligned}$$

as stated. For part (ii) to obtain the limit distribution under the alternative, write $\hat{r}_n$ as

$$\begin{aligned}\hat{r}_n &= \left(\sum_{t=1}^{n} X_{1t}X_{1t-1} + \sum_{t=1}^{n} X_{2t}X_{2t-1}\right)\left(\sum_{t=1}^{n} X_{1t-1}^2 + \sum_{t=1}^{n} X_{2t-1}^2\right)^{-1} \\ &= \frac{\rho_n \sum_{t=1}^n X_{1t-1}^2 + \theta_n \sum_{t=1}^n X_{2t-1}^2}{\sum_{t=1}^n X_{1t-1}^2 + \sum_{t=1}^n X_{2t-1}^2} + \frac{\sum_{t=1}^n u_{1t}X_{1t-1} + \sum_{t=1}^n u_{2t}X_{2t-1}}{\sum_{t=1}^n X_{1t-1}^2 + \sum_{t=1}^n X_{2t-1}^2} \\ &= \frac{\theta_n + \rho_n \sum_{t=1}^n X_{1t-1}^2 / \sum_{t=1}^n X_{2t-1}^2}{1 + \sum_{t=1}^n X_{1t-1}^2 / \sum_{t=1}^n X_{2t-1}^2} \\ &\quad + \frac{\sum_{t=1}^n u_{2t}X_{2t-1} / \sum_{t=1}^n X_{2t-1}^2 + \sum_{t=1}^n u_{1t}X_{1t-1} / \sum_{t=1}^n X_{2t-1}^2}{1 + \sum_{t=1}^n X_{1t-1}^2 / \sum_{t=1}^n X_{2t-1}^2} \\ &= \theta_n \left\{1 + \frac{\rho_n}{\theta_n}\frac{\sum_{t=1}^n X_{1t-1}^2}{\sum_{t=1}^n X_{2t-1}^2}\right\}\left\{1 + \frac{\sum_{t=1}^n X_{1t-1}^2}{\sum_{t=1}^n X_{2t-1}^2}\right\}^{-1} \\ &\quad + \left\{\frac{\sum_{t=1}^n u_{2t}X_{2t-1} + \sum_{t=1}^n u_{1t}X_{1t-1}}{\sum_{t=1}^n X_{2t-1}^2}\right\}\left\{1 + \frac{\sum_{t=1}^n X_{1t-1}^2}{\sum_{t=1}^n X_{2t-1}^2}\right\}^{-1}.\end{aligned}$$

Then, using Lemma 5.3

$$\begin{aligned}\hat{r}_n - \theta_n &= (\rho_n - \theta_n) \frac{\sum_{t=1}^n X_{1t-1}^2}{\sum_{t=1}^n X_{2t-1}^2}\{1 + o_p(1)\} + \frac{\sum_{t=1}^n u_{2t}X_{2t-1} + \sum_{t=1}^n u_{1t}X_{1t-1}}{\sum_{t=1}^n X_{2t-1}^2}\{1 + o_p(1)\} \\ &= \frac{1}{k_n\theta_n^n} \frac{\frac{1}{k_n\theta_n^n}\sum_{t=1}^n u_{2t}X_{2t-1} + \frac{n}{k_n\theta_n^n}\frac{1}{n}\sum_{t=1}^n u_{1t}X_{1t-1}}{\frac{1}{k_n^2\theta_n^{2n}}\sum_{t=1}^n X_{2t-1}^2}\{1 + o_p(1)\} \\ &\quad + \frac{n^2}{k_n^2\theta_n^{2n}}\left(\frac{c}{n} - \frac{b}{k_n}\right)\frac{\frac{1}{n^2}\sum_{t=1}^n X_{1t-1}^2}{\frac{1}{k_n^2\theta_n^2}\sum_{t=1}^n X_{2t-1}^2}\{1 + o_p(1)\} \\ &= \frac{1}{k_n\theta_n^n}\frac{\frac{1}{k_n\theta_n^n}\sum_{t=1}^n u_{2t}X_{2t-1}}{\frac{1}{k_n^2\theta_n^{2n}}\sum_{t=1}^n X_{2t-1}^2}\{1 + o_p(1)\},\end{aligned}$$

and in view of Lemma 5.1

$$k_n\theta_n^n(\hat{r}_n-\theta_n)\Rightarrow\frac{X_2(b)\,Y_2(b)}{X_2(b)^2/2b}=2b\frac{Y_2(b)}{X_2(b)},$$

giving the stated result (3.3). To prove (3.4), first note that

$$\hat{r}_n-\rho_n=(\hat{r}_n-\theta_n)+(\theta_n-\rho_n)=\left(\frac{b}{k_n}-\frac{c}{n}\right)+O_p\left(\frac{1}{k_n\theta_n^n}\right).$$

The restricted regression residuals are

$$\begin{aligned}\tilde{u}_t &= X_t-\hat{r}_nX_{t-1}=u_t-(\hat{r}_nI-R_n)X_{t-1}=u_t-\begin{bmatrix}\hat{r}_n-\rho_n & 0\\ 0 & \hat{r}_n-\theta_n\end{bmatrix}X_{t-1}\\ &= u_t-(\hat{r}_n-\theta_n)X_{t-1}+\begin{bmatrix}(\rho_n-\theta_n)X_{1t-1}\\ 0\end{bmatrix}\\ &= u_t+\frac{b}{k_n}\begin{bmatrix}X_{1t-1}\\ 0\end{bmatrix}\{1+o_p(1)\}.\end{aligned}$$

Let $\breve{\Sigma}=n^{-1}\sum_{t=1}^n u_tu_t'$ and then $\breve{\Sigma}\to_p\Sigma$ and

$$\begin{aligned}\tilde{\Sigma} &= \breve{\Sigma}+\frac{b}{k_nn}\sum_{t=1}^n\left\{\begin{bmatrix}X_{1t-1}\\ 0\end{bmatrix}u_t'+u_t\begin{bmatrix}X_{1t-1} & 0\end{bmatrix}\right\}\\ &\quad+\frac{b^2}{k_n^2n}\sum_{t=1}^n\begin{bmatrix}X_{1t-1}\\ 0\end{bmatrix}\begin{bmatrix}X_{1t-1} & 0\end{bmatrix}\\ &= \breve{\Sigma}+\frac{b^2n}{k_n^2}\begin{bmatrix}n^{-2}\sum_{t=1}^n X_{1t-1}^2 & 0\\ 0 & 0\end{bmatrix}\{1+o_p(1)\},\end{aligned}$$

since $n^{-1}\sum_{t=1}^n X_{1t-1}u_t=O_p(1)$ by Lemma 5.2(ii) and $\frac{k_n}{n}\to 0$. ■

6 References

Chan, N. H. and C. Z. Wei (1987). "Asymptotic Inference for Nearly Nonstationary AR(1) Processes, *Annals of Statistics* 15, 1050–1063.

Cheng, X and P. C. B. Phillips (2009). "Semiparametric cointegrating rank selection," *The Econometrics Journal*, 12, pp. S83-S104.

Hall, P. and C.C. Heyde (1980). *Martingale Limit Theory and its Application.* Academic Press.

Lai, T.L. and C.Z. Wei (1982). "Least Squares Estimates in Stochastic Regression Models with Applications to Identification and Control of Dynamic Systems". *Annals of Statistics*, 10, 154-166.

Leeb H. and B. Pötscher, "Model selection and inference: facts and fiction," *Econometric Theory*, 21, pp. 21-59.

Magdalinos, T. and P. C. B Phillips (2009). "Limit Theory for Cointegrated Systems with Moderately Integrated and Moderately Explosive Regressors". *Econometric Theory*, 25, 482-526.

Nielsen, B. (2009). "Singular vector autoregressions with deterministic terms: Strong consistency and lag order determination." University of Oxford working paper.

Pagan, A. and A. Ullah (1999). *Nonparametric Econonometrics.* Cambridge University Press.

Ploberger, W. and P. C. B. Phillips (2003). "Empirical limits for time series econometric models" *Econometrica,* 71, 627-673.

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, 535–547.

Phillips, P. C. B. (2008). "Unit root model selection," *Journal of the Japan Statistical Society,* 38, 65-74.

Phillips, P. C. B., T. Magdalinos (2007a), "Limit Theory for Moderate Deviations from a Unit Root," *Journal of Econometrics* 136, 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. and T. Magdalinos (2008). "Limit theory for explosively cointegrated systems", *Econometric Theory*, 24, 865-887.

Phillips, P. C. B., T. Magdalinos (2011), "Inconsistent VAR Regression with Common Explosive Roots," Working Paper, Yale University.

Phillips P. C. B., Y. Wu and J. Yu (2011). "Explosive behavior in the 1990s Nasdaq: When did exuberance escalate asset values?", *International Economic Review,* 52, pp. 201-226.

Phillips P. C. B. and J. Yu (2011). "Dating the Timeline of Financial Bubbles during the Subprime Crisis," *Quantitative Economics,* 2, pp. 455-491.

Ullah, A. (2004). *Finite Sample Econometrics.* Oxford: Oxford University Press.