OPTIMAL ESTIMATION OF COINTEGRATED SYSTEMS WITH IRRELEVANT INSTRUMENTS

By

Peter C.B. Phillips

COWLES FOUNDATION PAPER NO. 1400

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

2013

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

Journal of Econometrics 178 (2014) 210–224

Contents lists available at ScienceDirect

Journal of Econometrics

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

Optimal estimation of cointegrated systems with irrelevant instruments[☆]

Peter C.B. Phillips [*]

Yale University, United States
University of Auckland, New Zealand
University of Southampton, United Kingdom
Singapore Management University, Singapore

ARTICLE INFO

Article history:
Available online 16 September 2013

JEL classification:
C22

Keywords:
Asymptotic efficiency
Cointegrated system
Coverage probability
Instrumental variables
Irrelevant instrument
Karhunen–Loève representation
Optimal estimation
Orthonormal basis
Sieve estimation of stochastic processes
Trend basis
Trend likelihood

ABSTRACT

It has been known since Phillips and Hansen (1990) that cointegrated systems can be consistently estimated using stochastic trend instruments that are independent of the system variables. A similar phenomenon occurs with deterministically trending instruments. The present work shows that such "irrelevant" deterministic trend instruments may be systematically used to produce asymptotically efficient estimates of a cointegrated system. The approach is convenient in practice, involves only linear instrumental variables estimation, and is a straightforward one step procedure with no loss of degrees of freedom in estimation. Simulations reveal that the procedure works well in practice both in terms of point and interval estimation, having little finite sample bias and less finite sample dispersion than other popular cointegrating regression procedures such as reduced rank VAR regression, fully modified least squares, and dynamic OLS. The procedure is a form of maximum likelihood estimation where the likelihood is constructed for data projected onto the trending instruments. This "trend likelihood" is related to the notion of the local Whittle likelihood but avoids frequency domain issues.

1. Introduction

Clive Granger's gift for useful conceptualization is nowhere more evident than in his work on cointegration which massively impacted econometric practice and strengthened time series linkages with economic theory. The econometric methods associated with cointegration are now part of the central edifice of econometrics and have become one of its major exports to statistics and to empirical practice in the social and business sciences. By the early 1990s the methodology of rank determination, cointegrating system estimation and inference had all been worked out and incorporated into regression software facilitating widespread adoption in applications.

Efficient estimation of the cointegration space requires that estimation addresses the effects of both joint dependence and serial dependence. This is done parametrically in the reduced rank regression VAR approach (Johansen, 1988, 1995), and semiparametrically by fully modified least squares in Phillips and Hansen (1990) and by frequency domain techniques in Phillips (1991a). These methods require full system estimation and, in semiparametric cases, two-step estimation that utilizes consistent estimates of the equation errors. Two-sided dynamic least squares (Phillips and Loretan, 1991; Saikkonen, 1991; Stock and Watson, 1993) and narrow-band frequency domain methods (Phillips, 1991a; Phillips and Loretan, 1991) also produce efficient estimates, using single equation one-step regressions that are augmented with differences as well as levels.

The contribution of the present paper is to introduce an entirely different approach to efficient estimation. The linear IV regression approach developed here provides direct one-step efficient estimation of cointegrating coefficients as well as consistent estimates of the long-run regression coefficients that embody the effects of joint dependence. Furthermore, since the instrument variables are chosen to be deterministic functions of time, there is no need for further corrections for serial dependence. In consequence, the approach provides an extremely simple mechanism for optimally estimating long-run coefficients in cointegrated systems while making weak assumptions about the generating mechanism so that the procedure has wide applicability.

☆ My thanks to Werner Ploberger, three referees and the editor for helpful comments. Some of the ideas herein were first mentioned by the author in a lecture at the York Econometrics Meeting, June 2003. The complete paper was later presented at the Faro Conference, September 2005 (Phillips, 2006). Partial support is acknowledged from a Kelly Fellowship and the NSF under Grant Nos. SES 04-142254 and SES 09-56687.

* Correspondence to: Yale University, United States.
E-mail address: peter.phillips@yale.edu.

0304-4076/$ – see front matter
http://dx.doi.org/10.1016/j.jeconom.2013.08.022

The fact that efficient estimation using what may be regarded as irrelevant instruments is possible may appear somewhat magical, especially in view of existing results on IV estimation in stationary systems where relevance of the instruments is critical to asymptotic efficiency and can even jeopardize consistency when the instruments are weak[1] (Phillips, 1989; Staiger and Stock, 1997). Furthermore, the results here make clear that what is often regarded as potentially dangerous spurious correlation among trending variables can itself be used in a systematic way to produce rather startling positive results. In this respect, the results of the present paper extend some earlier findings by the author (1998, 2002, 2005a) on the usefulness of apparently spurious trend regressions.

The essential idea can be explained as follows. We start by constructing a basis for a suitably defined space of trending variables using as basis functions what might initially be regarded as irrelevant deterministic trends that have no direct bearing on the generation of the stochastically trending system variables. In conducting IV estimation with these basis functions, we project all the system variables on the trend space and, in doing so, isolate the long-run behavior of the system variables and their differences, thereby enabling estimation of all the long-run parameters, including the cointegrating coefficients and the long-run conditional mean of the equilibrium error. The estimates are efficient because the set of basis functions is complete in the limit, so that all possible forms of trend behavior are accounted for, and because the procedure automatically adjusts for the endogeneity of the system regressors by consistently estimating the long-run conditional mean of the equilibrium error.

The idea amounts to sieve estimation of endogenous stochastic processes using deterministic basis functions. The approach turns out to be economical as well as general because only linear instrumental variable methods are needed and the trend instruments are straightforward deterministic functions of time. The approach is also agnostic about the form of the trend behavior in the system variables, provided it can in the limit be captured by the basis functions. In effect, this approach simply uses a basis for the trend space to focus attention on long-run behavior in a linear cointegrating regression.

An interesting by-product of the asymptotic analysis is that regression of a stationary time series on apparently irrelevant trending instruments provides a new way of consistently estimating long-run covariance matrices and long-run regression coefficients. The approach can be used in quite general HAC estimation contexts, an application of the idea that is systematically explored elsewhere (Phillips, 2005b).

The procedure developed here may be regarded as a form of maximum likelihood estimation where the likelihood is constructed to focus on trend or long-run features in the data. Such a "trend likelihood" is closely related to the notion of the local Whittle likelihood (Künsch, 1987) where only those frequencies in a narrow band around the origin are used in the construction of the Whittle likelihood. Accordingly, the IV cointegration estimator given here is most closely related to the narrow band technique suggested in the author's earlier work (1991a), although there is no need for frequency domain calculations or techniques.

Trend likelihood methods will be useful in contexts other than those studied here. One application that is particularly relevant to recent econometric research is long memory parameter estimation. In this context, the approach delivers a general purpose long memory estimator that is applicable in both stationary and nonstationary cases in a manner that is analogous to the frequency domain approach studied recently by Shimotsu and Phillips (2005). This particular application is discussed briefly at the end of the paper. Other potential applications are to cointegrated regression models with nearly integrated and fractionally integrated regressors. Dealing efficiently with endogeneity issues in such models is more complex, however, and is not pursued in the present work. The reader is referred to Magdalinos and Phillips (2009) for another IV approach to estimating cointegrated systems with roots in the vicinity of unity.

The paper is organized as follows. Section 2 lays out the model and preliminaries. The main results are given in Section 3. Selection of the number of instruments is considered in Section 4. Section 5 provides some simulation findings for cointegrated systems. The concept of a trend likelihood is introduced in Section 6 and applications to long memory estimation are discussed. Section 7 concludes. Proofs and other technical material, including some lemmas of independent interest, are given in the Appendix.

2. Model and preliminaries

We consider the following cointegrated system

$$y_t = Ax_t + u_{0t} \tag{1}$$

$$\Delta x_t = u_{xt} \tag{2}$$

relating the observable time series y_t $(m_y \times 1)$ and x_t $(m_x \times 1)$ with initial conditions at $t = 0$ and $x_0 = O_p(1)$. The composite error $u_t = \left(u'_{0t}, u'_{xt}\right)'$ is a weakly dependent time series satisfying

$$u_t = C(L)\,\varepsilon_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j}, \qquad \sum_{j=0}^{\infty} j^a \left|\left|c_j\right|\right| < \infty, \quad a > 3, \tag{L}$$

where $\varepsilon_t = iid\,(0, \Sigma)$ with $\Sigma > 0$ and $E(\|\varepsilon_t\|^{\upsilon}) < \infty$, for some $\upsilon > 2$ and matrix norm $\|\|$. The long-run moving average coefficient matrix $C(1)$ is assumed to be nonsingular, so that x_t is a full rank integrated process. The time series u_t is stationary with variance matrix $\Sigma_u = \sum_{j=0}^{\infty} c_j \Sigma c_j'$, autocovariance function $\Gamma_u(h) = E\left(u_t u'_{t+h}\right) = \sum_{j=0}^{\infty} c_j \Sigma c'_{j+h}$, finite υ'th absolute moment $E\,\|u_t\|^{\upsilon} \le \left(\sum_{j=0}^{\infty} \left|c_j\right|\right)^{\upsilon} E\,||\varepsilon_t||^{\upsilon} < \infty$, spectrum $f_u(\lambda) = (1/2\pi)\,C\left(e^{i\lambda}\right)\Sigma C\left(e^{-i\lambda}\right)'$, and long-run variance matrix $\Omega = 2\pi f_u(0) = C(1)\,\Sigma\, C(1)'$, which is partitioned conformably with u_t as

$$\Omega = \begin{bmatrix} \Omega_{00} & \Omega_{0x} \\ \Omega_{x0} & \Omega_{xx}. \end{bmatrix}$$

We define the conditional long-run covariance matrix $\Omega_{00.x} = \Omega_{00} - \Omega_{0x}\Omega_{xx}^{-1}\Omega_{x0}$.

The summability condition **L** implies that

$$\sum_{h=-\infty}^{\infty} h^3\,||\Gamma_u(h)|| < \infty, \tag{3}$$

so that $f_u(\lambda)$ has continuous second derivative $f_u^{(2)}(\lambda) = -\frac{\sigma^2}{2\pi} \sum_{h=-\infty}^{\infty} h^2 \Gamma_u(h)\, e^{-i\lambda h}$. While this framework assumes stationary u_t, allowance for some heterogeneity in ε_t and u_t is possible and can be made in the usual way with minor modifications to **L** as in Phillips and Solo (1992) without affecting the results given below in an essential way.

Under **L**, partial sums $S_t = \sum_{i=1}^{t} u_i$ satisfy the functional law (e.g., Phillips and Solo (1992))

$$B_n(\cdot) := \frac{S_{\lfloor n\cdot\rfloor}}{\sqrt{n}} = \frac{\sum_{i=1}^{\lfloor n\cdot\rfloor} u_i}{\sqrt{n}} \Rightarrow B(\cdot), \tag{4}$$

where $\lfloor a\rfloor$ signifies the integer part of a, $\Rightarrow$ is weak convergence, and $B(\cdot)$ is vector Brownian motion with variance matrix Ω. We

[1] However, recent results of Chao and Swanson (2005) and Han and Phillips (2006) show that it is possible to compensate for the effects of weak (and even irrelevant instruments in some cases) through the use of large numbers of instruments and moment conditions.

partition B conformably with u_t by setting $B = \left(B_0', B_x'\right)'$ and define the Brownian motion $B_{0.x} = B_0 - \Omega_{0x}\Omega_{xx}^{-1}B_x$, a Brownian motion with variance matrix $\Omega_{00.x}$ that is independent of B_x.

The limit process $B(r)$ has an almost sure unique representation in terms of deterministic functions over the interval $r \in [0, 1]$. It is particularly convenient in the mathematical derivations that follow to use the orthonormal functions corresponding to the covariance kernel of B and this leads to the following vector Karhunen–Loève (KL) representation (see Phillips (1998, 2005b))

$$B(r) = \sqrt{2}\sum_{k=1}^{\infty}\frac{\sin[(k-1/2)\pi r]}{(k-1/2)\pi}\xi_k = \sum_{k=1}^{\infty}\lambda_k^{\frac{1}{2}}\varphi_k(r)\xi_k, \tag{5}$$

where the components ξ_k are *iid* $N(0, \Omega)$, $\lambda_k = 1/((k-1/2)\pi)^2$, and $\varphi_k(r) = \sqrt{2}\sin[(k-1/2)\pi r]$. This series representation of $B(r)$ is convergent almost surely and uniformly in $r \in [0, 1]$. We may write (5) as a system of equations with partitioned regressors as follows

$$B(r) = \Xi_K\Lambda_K^{\frac{1}{2}}\tilde{\varphi}_K(r) + \Xi_\perp\Lambda_\perp^{\frac{1}{2}}\tilde{\varphi}_\perp(r), \tag{6}$$

where $\Lambda_K = diag(\lambda_1, \ldots, \lambda_K)$, $\Lambda_\perp = diag(\lambda_{K+1}, \lambda_{K+2}, \ldots)$,

$$\Xi_K = [\xi_1, \ldots, \xi_K], \qquad \Xi_\perp = [\xi_{K+1}, \xi_{K+2}, \ldots],$$
$$\tilde{\varphi}_K(r) = [\varphi_1(r), \ldots, \varphi_K(r)]', \quad \text{and}$$
$$\tilde{\varphi}_\perp(r) = [\varphi_{K+1}(r), \varphi_{K+2}(r), \ldots]'.$$

We further partition these matrices conformably with u_t as

$$\Xi_K = \begin{bmatrix}\Xi_{0K}\\ \Xi_{xK}\end{bmatrix} = \begin{bmatrix}[\xi_{01}, \ldots, \xi_{0K}]\\ [\xi_{x1}, \ldots, \xi_{xK}]\end{bmatrix},$$
$$\Xi_\perp = \begin{bmatrix}\Xi_{0\perp}\\ \Xi_{x\perp}\end{bmatrix} = \begin{bmatrix}[\xi_{0K+1}, \xi_{0K+2}, \ldots]\\ [\xi_{xK+1}, \xi_{xK+2}, \ldots]\end{bmatrix}.$$

Note that the coefficient of the deterministic function $\varphi_k(r)$ in (5) is of order $O_p(\frac{1}{k})$, so that weighted functions in the KL representation become less important as k gets large.

Using the Phillips and Solo (1992) approach and extending the probability space, it is possible to develop a convenient weak approximation to the partial sum process $B_n(\cdot)$ in terms of a Brownian motion B with variance matrix Ω

$$\sup_{t\in[0,1]}\|B_n(t) - B(t)\| = o_p\left(\frac{1}{n^{\frac{1}{2}-\frac{1}{\nu}}}\right) \quad \text{as } n \to \infty, \tag{7}$$

as detailed in Lemma A in the Appendix, which is a multivariate extension of Phillips (2006, Lemma 3.1) and Akonom (1993, Theorem 3). In what follows, we will assume that the probability space has been expanded as necessary in order for (7) to apply. The moment condition $\nu > 2$ in $\mathbf{L}$ ensures that $o_p\left(1/n^{\frac{1}{2}-\frac{1}{\nu}}\right) = o_p(1)$ in (7). The larger the moment exponent υ the smaller is the error magnitude in (7). This weak approximation helps to simplify the limit theory.

3. Estimation with many irrelevant instruments

Define the augmented regression equation

$$y_t = Ax_t + \Omega_{0x}\Omega_{xx}^{-1}\Delta x_t + u_{0.xt}, \tag{8}$$

where $u_{0.xt} = u_{0t} - \Omega_{0x}\Omega_{xx}^{-1}u_{xt}$, and write the equation in observation format as

$$Y' = AX' + \Omega_{0x}\Omega_{xx}^{-1}\Delta X' + U_{0.x}',$$

where $Y' = [y_1, \ldots, y_n]$ with similar definitions for $\Delta X'$, and $U_{0.x}'$.

Let $\{\varphi_k\}_{k=1}^{\infty}$ be an orthonormal basis of the space $L_2[0, 1]$ of square integrable deterministic functions on the interval $[0, 1]$. All functions $f \in L_2[0, 1]$ can then be written in terms of the functions $\{\varphi_k\}_{k=1}^{\infty}$ as $f(x) =_{L_2} \sum_{k=1}^{\infty} c_k\varphi_k(x)$, where $=_{L_2}$ signifies equality in the L_2 sense. Our approach to estimation of (8) is to use as instrumental variables for both x_t and Δx_t a (potentially infinite) sequence of deterministic functions of the form $\{\varphi_k\left(\frac{t}{n}\right) : k = 1, \ldots, K\}$. Thus, we allow K to pass to infinity with n, so that in the limit an infinite number of instruments are being employed. Since these instruments are all deterministic functions and are uncorrelated with x_t and Δx_t they might be regarded as irrelevant to the regression. Indeed, such deterministic functions of time would, in conventional econometric parlance, be regarded as being spurious for both x_t and Δx_t.

In what follows, it will be convenient for the development to use the orthonormal sequence

$$\varphi_k(r) = \sqrt{2}\sin[(k-1/2)\pi r], \tag{9}$$

used in the KL representation (5). In practice, there is little difference in regression results when other sequences of orthonormal instruments are used, and some illustrative simulation results to this effect will be given later. Let $\varphi_{Kt} = \left(\varphi_1\left(\frac{t}{n}\right), \ldots, \varphi_K\left(\frac{t}{n}\right)\right)'$, $\Phi_K' = [\varphi_{K1}, \ldots, \varphi_{Kn}]$ and $P_K = \Phi_K\left(\Phi_K'\Phi_K\right)^{-1}\Phi_K'$ be the orthogonal projector to the space spanned by the columns of Φ_K. Assume the order condition $K \geq 2m_x$ holds and apply instrumental variables linear regression to (8) using the matrix of instruments Φ_K. As indicated, the instruments are being used here for both the levels x_t and the differences Δx_t in (8). In the regression we can treat $C = \Omega_{0x}\Omega_{xx}^{-1}$ simply as an unknown coefficient matrix.

The IV estimator of the cointegrating matrix A and regression coefficient C satisfy

$$(A_{IV}, C_{IV}) = \arg\min_{A,C}\left(Y' - AX' + C\Delta X'\right)P_K \times \left(Y - XA' + \Delta XC'\right). \tag{10}$$

Accordingly,

$$A_{IV} = \arg\min_{A}\left(Y' - AX'\right)R_K(Y - AX),$$

where $R_K = P_K - P_K\Delta X\left(\Delta X'P_K\Delta X\right)^{-1}\Delta X'P_K$, leading to the explicit partitioned regression formula

$$A_{IV} = \left(Y'R_KX\right)\left(X'R_KX\right)^{-1}, \tag{11}$$

and the corresponding residual moment matrix

$$\Omega_{0.xx}^{IV} = K^{-1}\hat{U}_{0.x}'P_K\hat{U}_{0.x}' = K^{-1}\left(Y' - A_{IV}X' + C_{IV}\Delta X'\right)P_K \times \left(Y - XA_{IV}' + \Delta XC_{IV}'\right) \tag{12}$$

from this regression, where $\hat{U}_{0.x}' = Y' - A_{IV}X' + C_{IV}\Delta X'$ is the matrix of regression residuals. In (12), the matrix is weighted by the dimension (K) of the instrument space rather than the number of observations (n).

The estimator A_{IV} has the advantage that it can be calculated by straightforward linear regression and does not involve any preliminary steps or regression. There is also no need to take complex data transformations, as in the narrow-band frequency domain approach of Phillips (1991a), which was earlier recognized to be a one-step approach to efficient cointegrating regression. On the other hand, the latter estimator may itself be interpreted in terms of an IV regression. In particular, we may replace the projector P_K in (10) above with $P_{cK} = \Phi_{cK}\left(\Phi_{cK}^*\Phi_{cK}\right)^{-1}\Phi_{cK}^*$, where $*$ denotes complex conjugate transpose, $\Phi_{cK}' = [\varphi_{cK1}, \ldots, \varphi_{cKn}]$, and $\varphi_{cKt} = \varphi_{cK}\left(\frac{t}{n}\right)$, where the latter has complex sinusoidal components $\varphi_{ck}\left(\frac{t}{n}\right) = (2\pi n)^{-1/2}e^{2\pi ik\frac{t}{n}}$. Then, $X'\Phi_{cK} = ((2\pi n)^{-1/2}\sum_{t=1}^{n}X_t e^{2\pi ik\frac{t}{n}})$ is a vector of K discrete Fourier transforms (dfts) of X_t, and it is immediately apparent that IV regression in (10) with P_K replaced by P_{cK} is equivalent to a narrow-band frequency domain regression involving the K harmonic frequencies $\{\lambda_k = \frac{2\pi k}{n} : k = 1, \ldots, n\}$. What (10) and the results below show, is that it is not necessary to take dfts and do regression in the frequency domain. What is important in the regression is that the instruments serve as a basis for the trend space and, for efficient estimation, that when $K \to \infty$ the basis be complete. This may just as well be achieved with real polynomials as with complex sinusoid polynomials. So the conceptual framework goes beyond frequency domain regression.

The idea behind the IV estimate in (11) is as follows. The deterministic trend variables φ_{Kt} serve as instruments for the levels of the integrated regressors x_t. As remarked in the introduction, even when using a fixed number of instruments and without employing the additional regressors Δx_t in the regression equation (8), such an IV regression is well-known to produce a consistent estimate of the cointegrating matrix A because of the spurious regression phenomena (Phillips, 1986; Phillips and Hansen, 1990). However, as we demonstrate below, some particularly interesting effects emerge as K increases when the regression equation is augmented as in (8).

First, in view of the KL representation (5), it is known from Phillips (1998, 2002) that deterministic instruments like φ_{Kt} become more effective in modeling integrated regressors as $K \to \infty$. Indeed, in the limit these instruments are capable of capturing the full KL representation of the limiting Brownian motion that corresponds to the level regressors x_t in (8). Thus, for large K, these regressors are strongly relevant for x_t, while at the same time clearly satisfying the orthogonality condition. Second, and perhaps more interesting and unexpected, is that in the augmented regression equation (8), it turns out that, as K increases, the instruments also become more effective in estimating the precise form of the coefficient matrix $C = \Omega_{0x}\Omega_{xx}^{-1}$, which is the long-run regression coefficient of u_{0t} on Δx_t.

Thus, two different effects work simultaneously in the IV regression leading to (11)—one capturing the movements of the nonstationary regressor x_t, while retaining orthogonality with the equation errors, the other capturing the long-run regression effects associated with the stationary regressor Δx_t and adjusting the conditional mean for the endogeneity of the regressor. In fact, as the main result below shows, as $K \to \infty$ and $n \to \infty$ the IV regression estimate is asymptotically efficient in the sense of Phillips (1991a,b) and the IV regression estimate of C is consistent. Thus, in the same one-step regression and with the same instrument set, we achieve an asymptotically efficient estimate of the cointegrating matrix A, a consistent estimate of the long-run regression coefficient $\Omega_{0x}\Omega_{xx}^{-1}$, and (as shown below) a consistent estimate of the long-run conditional error variance matrix $\Omega_{00.x}$. So, all the long-run parameters are consistently estimated in this one step regression.

The limit theory for A_{IV} is given in the following result, confirming that the estimate is efficient and asymptotically equivalent to full maximum likelihood under Gaussian errors in finite dimensional cases and achieves semiparametric efficiency bounds when u_t is a Gaussian linear process of the general form **L** (c.f., Phillips (1991b) and Jeganathan (1995)). Inference can be conducted in the usual fashion for mixed normal limit theory using appropriate error variance matrix estimates combined with the usual inverse of the moment matrix in the partitioned regression, $\left(X'R_KX\right)^{-1}$. In the present case, the long-run variance matrix of $u_{0.xt}$ is consistently estimated by the standardized residual moment matrix $\Omega_{0.xx}^{IV}$, as shown below.

Theorem. *Under L and the rate condition*

$$\frac{1}{K} + \frac{K}{n^{\left(1-\frac{2}{v}\right)\wedge\left(\frac{5}{6}-\frac{1}{3v}\right)}} + \frac{K^5}{n^4} \to 0, \tag{R}$$

as $n \to \infty$, the following hold:

(a) $n\left(A_{IV} - A\right) \Rightarrow \left(\int_0^1 dB_{0.x}B_x'\right)\left(\int_0^1 B_xB_x'\right)^{-1} \equiv MN\left(0, \Omega_{00.x} \otimes \left(\int_0^1 B_xB_x'\right)^{-1}\right)$.

(b) $n^{-2}X'R_KX \Rightarrow \int_0^1 B_xB_x'$.

(c) $\Omega_{0.xx}^{IV} \to_p \Omega_{00.x}$.

Remarks. (a) Condition **R** requires that $K \to \infty$ but at a rate that is slower than $n^{4/5}$ and the smaller of $n^{1-\frac{2}{v}}$ and $n^{5/6-1/3v}$. The latter restriction is likely to be stronger than is necessary. However, the restriction is convenient for the proof of the theorem and it arises because the proof makes direct use of the approximation (7) in determining error magnitudes. For large v, of course, the condition is hardly restrictive and amounts to $K = o\left(n^{4/5-\delta}\right)$ for small $\delta > 0$.

(b) An interesting by-product of the proof of the theorem is that we have the convergence $n^{-1}U_{0.x}'P_KX \Rightarrow \int_0^1 dB_{0.x}(r)\, B_x(r)'\, dr$. In fact, the following weak convergence to a stochastic integral is established in (57)

$$\left(\frac{U_{0.x}'\Phi_K}{\sqrt{n}}\right)\left(\frac{\Phi_K'X}{n^{3/2}}\right) \Rightarrow \int_0^1 dB_{0.x}(r)\, B_x(r)',$$

as $n, K \to \infty$. An important aspect of this result is that the limit processes $B_{0.x}$ and B_x are independent and have zero quadratic covariation. Of course, this orthogonality is central to the successful removal of endogeneity in the IV regression and leads to the mixed normal limit distribution of the IV estimator A_{IV}. On the other hand, convergence of the corresponding matrix quadratic form $n^{-1}U_x'P_KX$ to the stochastic integral $\int_0^1 dB_x(r)\, B_x(r)'$ does not occur, so that

$$\left(\frac{U_x'\Phi_K}{\sqrt{n}}\right)\left(\frac{\Phi_K'X}{n^{3/2}}\right) \not\Rightarrow \int_0^1 dB_x(r)\, B_x(r)'.$$

Indeed, as shown in Phillips (2002), in the scalar case (i.e., when x_t is scalar and B_x is scalar Brownian motion) we have

$$\left(\frac{u_x'\Phi_K}{\sqrt{n}}\right)\left(\frac{\Phi_K'x}{n^{3/2}}\right) \Rightarrow \frac{1}{2}B_x(1)^2 \neq \int_0^1 B_x dB_x, \tag{13}$$

so that the quadratic variation component of the integral is omitted in the limit. In fact, the weak convergence (13) is to the Stratonovich integral (e.g., Protter (1990)) rather than the Ito integral. Thus, when they are applied to unit root models or vector autoregressions with some unit roots, IV regressions of the type considered here do not lead to estimates that have the usual unit root limit distributions.

4. Instrument number selection

Phillips (2005a,b) gave formulae for the optimal choice of K in the context of long-run variance estimation in terms of minimizing the asymptotic mean square error of estimation. The optimal rate in that case is $K = O\left(n^{4/5}\right)$. We may extend that result to the multivariate case, as in Lemma C of the Appendix, to accommodate estimation of the long-run variance matrix Ω. This approach may be employed in the present regression context with a focus on finding the optimal choice of K for estimating the long-run regression coefficient $C = \Omega_{0x}\Omega_{xx}^{-1}$, which appears in the augmented regression model (8). Again, the optimal rate is $K = O\left(n^{4/5}\right)$, as is shown in (62) in the Appendix.

While this approach has some justification in the present context because C is a regression coefficient in (8), it by no means implies that the asymptotic mean squared error (AMSE) of estimation of the cointegrating matrix A is optimized by this choice. To analyze

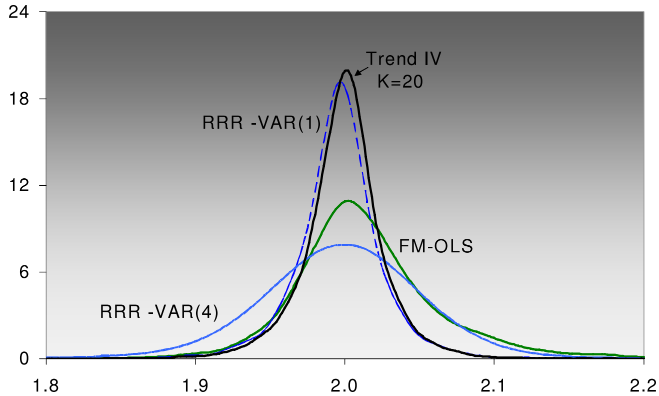

Fig. 1. Distributions of cointegrating coefficient estimators with $\rho = 0.75$, $b = 2.0$, $n = 50$, and $K = 20$. Moving average errors with $\theta_1 = \theta_2 = 0.4$.

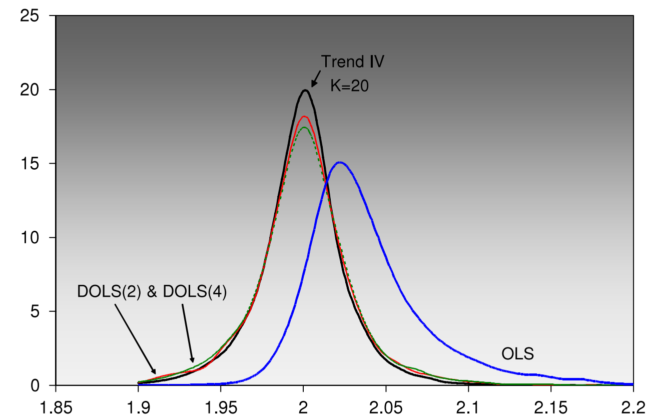

Fig. 2. Distributions of cointegrating coefficient estimators with $\rho = 0.75$, $b = 2.0$, $n = 50$, and $K = 20$. Moving average errors with $\theta_1 = \theta_2 = 0.4$.

the AMSE of estimation of A, it is necessary to develop an asymptotic expansion of the estimate A_{IV}. The situation is analogous to that considered by Linton (1995) and Xiao and Phillips (1998, 1999) in semiparametric regression problems where a smoothing parameter needs to be selected for the nonparametric estimation component. While the first order limit distribution, as in part (a) of the theorem above, is invariant to the precise choice of smoothing parameter that is employed (provided the smoothing parameter obeys some general rate restriction such as condition **R**), the second order expansion is affected and higher order AMSE comparisons might be conducted to develop an optimal criterion.

In the cointegrating regression context studied here, higher order expansions are complicated by the mixed normal limit theory of A_{IV} and the use of functional limit theory in the first order asymptotics. The same complications arise with respect to other semiparametric estimates of A. These issues are yet to be fully explored in the literature, although Xiao and Phillips (2002) provide some higher order analysis for the expected value of Wald tests in a related setting. We shall leave the development of a higher order asymptotic expansion for A_{IV} to future research. Intuition indicates that the primary need in the estimation of the cointegration matrix A is for bias control and preliminary calculations undertaken by the author indicate that the optimal expansion rate for K in terms of the AMSE of A_{IV} will be slower than the $O\left(n^{4/5}\right)$ rate for long run variance and regression coefficient estimation, discussed above.

5. Simulations

To illustrate, we briefly report some cointegrating regression simulations with Trend IV methods and compare its performance

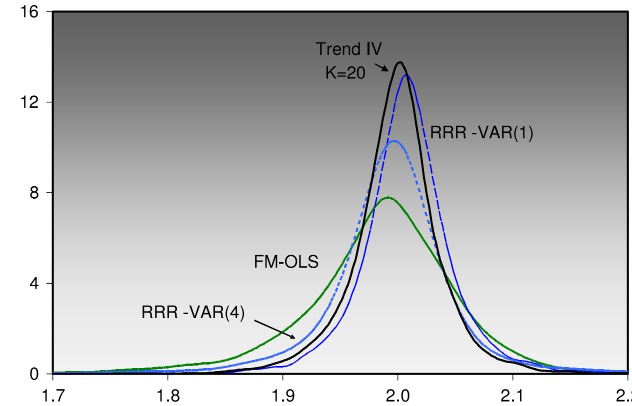

Fig. 3. Distributions of cointegrating coefficient estimators with $\rho = -0.75$, $b = 2.0$, $n = 50$, and $K = 20$. Moving average errors with $\theta_1 = \theta_2 = 0.4$.

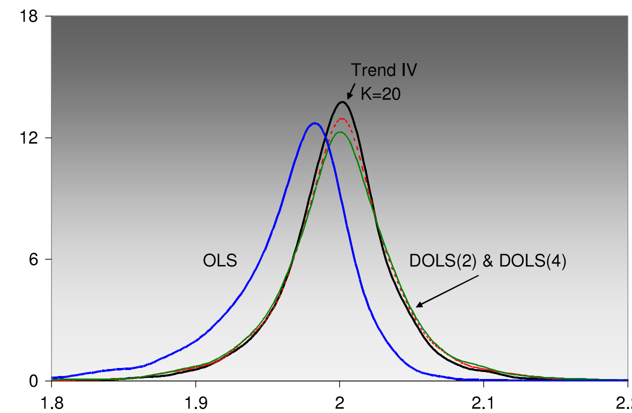

Fig. 4. Distributions of cointegrating coefficient estimators with $\rho = -0.75$, $b = 2.0$, $n = 50$, and $K = 20$. Moving average errors with $\theta_1 = \theta_2 = 0.4$.

with the most popular existing techniques, notably reduced rank regression (RRR) in a VAR system, fully modified least squares (FM-OLS), and dynamic least squares (DOLS) in regressions augmented with leads and lagged differences. A two variable system is used in these simulations. It will be useful to extend these to higher order systems to assess the impact of more variables and a higher dimensional cointegrating space on IV regression.

Figs. 1–6 provide some typical findings from the cointegrated model with moving average and autoregressive errors

$$\begin{array}{l} X_{1t} = bX_{2t} + u_{1t} \\ X_{2t} = X_{2t-1} + u_{2t}, \end{array} \qquad u_t = \begin{cases} \varepsilon_t + \Theta\varepsilon_{t-1} & \text{MA}(1) \\ \Theta u_{t-1} + \varepsilon_t & \text{VAR}(1) \end{cases}, \tag{14}$$

$$\varepsilon_t = \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \sim iidN\left(0, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right), \qquad \Theta = \begin{bmatrix} \theta_1 & 0 \\ 0 & \theta_2 \end{bmatrix}, \tag{15}$$

for $b = 2.0$, $n = 50$, $K = 20$ and the cases $\rho \in \{0.75, -0.75\}$, each with 10,000 replications. The figures show kernel density estimates of the probability densities of each of the cointegration estimators. We use DOLS(p) to signify DOLS with p leads and lags, and RRR(p) to signify RRR with p lags in the corresponding VAR. Tables I and II provide a summary of the findings for a wider selection of the parameter values (θ_1, θ_2). Similar results were obtained for $n = 100$ but with smaller differences between procedures and they are not reported here.

It is apparent from both the figures and the tables that the trending IV estimator works extremely well against this competing group of cointegrating regression procedures. From the summary statistics in the tables, the root mean squared error (RMSE) of the Trend IV estimates is, with just two exceptions, uniformly smaller than the RMSE of all the other estimates. The exceptions

Table I
Finite sample performance of cointegration estimators with AR errors, $b = 2$, $\rho = 0.75$, $T = 50$, $N = 10{,}000$ replications.

AR coeffs (θ_1, θ_2)	Estimator	Bias	SD	RMSE	AR coeffs (θ_1, θ_2)	Bias	SD	RMSE
	OLS	0.023	0.040	0.046		0.256	0.222	0.339
	FMOLS	0.010	0.081	0.082		0.200	0.210	0.290
	RRR_1	−0.015	0.057	0.059		0.011	0.038	0.040
(0.8, 0.8)	RRR_4	−0.009	1.261	1.261	(−0.8, −0.8)	0.001	0.086	0.086
	$DOLS_2$	0.000	0.044	0.044		0.001	0.042	0.042
	$DOLS_4$	0.000	0.048	0.048		0.000	0.046	0.046
	Trend IV	0.000	0.037	0.037		0.001	0.035	0.035
	OLS	0.034	0.041	0.053		0.098	0.090	0.133
	FMOLS	0.012	0.055	0.056		0.045	0.075	0.088
	RRR_1	−0.006	0.035	0.036		0.004	0.035	0.035
(0.4, 0.4)	RRR_4	0.004	1.062	1.062	(−0.4, −0.4)	0.000	0.209	0.209
	$DOLS_2$	0.000	0.038	0.038		−0.001	0.038	0.038
	$DOLS_4$	0.000	0.038	0.038		−0.001	0.039	0.039
	Trend IV	0.000	0.0034	0.034		0.000	0.034	0.034
	OLS	0.055	0.054	0.077		0.335	0.220	0.401
	FMOLS	0.020	0.058	0.061		0.204	0.329	0.387
	RRR_1	−0.001	0.035	0.035		2.003	148.97	148.98
(0, 0)	RRR_4	0.001	0.203	0.203	(−0.8, 0.8)	0.019	5.586	5.586
	$DOLS_2$	0.000	0.036	0.036		0.378	0.312	0.490
	$DOLS_4$	−0.001	0.038	0.038		0.383	0.322	0.500
	Trend IV	0.000	0.034	0.034		0.330	0.301	0.447

Table II
Finite sample performance of cointegration estimators with MA errors, $b = 2$, $\rho = 0.75$, $T = 50$, $N = 10{,}000$ replications.

MA coeffs (θ_1, θ_2)	Estimator	Bias	SD	RMSE	MA coeffs (θ_1, θ_2)	Bias	SD	RMSE
	OLS	0.036	0.041	0.054		0.366	0.208	0.421
	FMOLS	0.014	0.059	0.060		0.393	0.196	0.439
	RRR_1	−0.007	0.036	0.036		0.004	0.057	0.057
(0.8, 0.8)	RRR_4	0.000	0.149	0.149	(−0.8, −0.8)	0.000	0.053	0.053
	$DOLS_2$	0.000	0.038	0.038		0.001	0.067	0.067
	$DOLS_4$	0.000	0.038	0.038		0.001	0.070	0.070
	Trend IV	0.000	0.033	0.033		0.001	0.047	0.047
	OLS	0.039	0.043	0.058		0.117	0.102	0.155
	FMOLS	0.016	0.056	0.059		0.065	0.089	0.110
	RRR_1	−0.005	0.035	0.036		0.004	0.036	0.036
(0.4, 0.4)	RRR_4	0.001	0.142	0.142	(−0.4, −0.4)	0.000	0.069	0.069
	$DOLS_2$	0.000	0.038	0.038		0.000	0.039	0.039
	$DOLS_4$	0.000	0.038	0.038		−0.001	0.040	0.040
	Trend IV	0.000	0.033	0.033		0.000	0.034	0.034
	OLS	0.170	0.132	0.215		0.444	0.221	0.496
	FMOLS	0.087	0.145	0.169		0.483	0.276	0.556
	RRR_1	−0.025	0.435	0.440		−15.028	1667	1667
(−0.4, 0.4)	RRR_4	−0.002	1.030	1.030	(−0.8, 0.8)	0.048	3.200	3.200
	$DOLS_2$	0.103	0.106	0.148		0.536	0.339	0.635
	$DOLS_4$	0.103	0.110	0.151		0.541	0.349	0.644
	Trend IV	0.065	0.090	0.111		0.430	0.321	0.537

occur when $\theta_1 = -0.8, \theta_2 = 0.8$ for both MA and AR errors, in which case the OLS estimator has smaller RMSE than all the other estimates, but the Trend IV estimator has the next best RMSE and has smaller bias in both these cases.

In the case of both MA and AR errors, the IV estimator shows very little finite sample bias in general and has smaller dispersion than all the other procedures, except for the case $\theta_1 = -0.8, \theta_2 = 0.8$ just mentioned. Similar results for the trending IV estimator were obtained for different values of K in the range $20 \leq K \leq 40$, so there seems to be reasonable robustness to the dimension of the instrument space, although when the serial dependence coefficients θ_1 and θ_2 have very different magnitudes there appears to be more sensitivity as K increases – see Figs. 7 and 8 – and in such cases the bias of OLS is much greater and typically the other procedures perform poorly.

Dynamic OLS and reduced rank regression (RRR) appear to be the next best procedures. Dynamic OLS has more variance than trending IV and RRR shows evidence of finite sample bias,

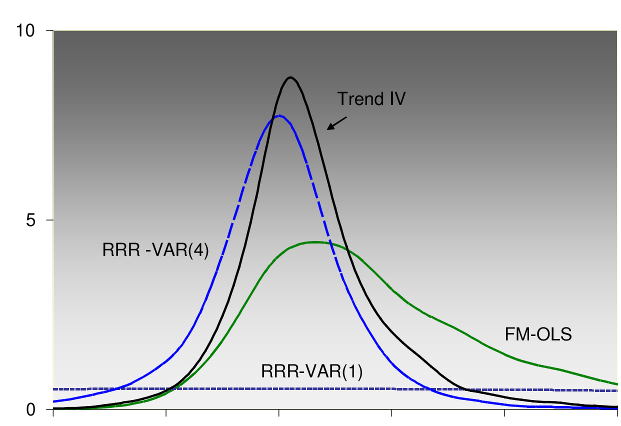

Fig. 5. Distributions of cointegrating coefficient estimators with $\rho = 0.75, b = 2.0$, $n = 50$, and $K = 20$. Autoregressive errors with $\theta_1 = 0, \theta_2 = -0.6$.

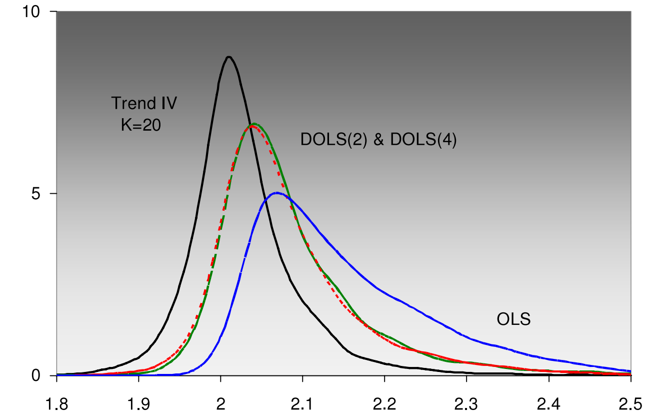

Fig. 6. Distributions of cointegrating coefficient estimators with $\rho = 0.75, b = 2.0$, $n = 50$, and $K = 20$. Autoregressive errors with $\theta_1 = 0, \theta_2 = -0.6$.

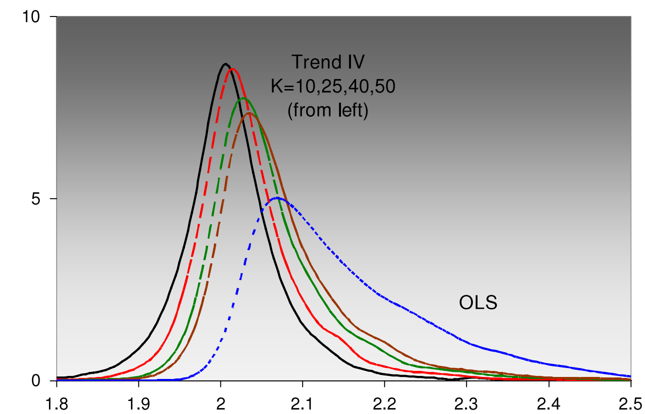

Fig. 7. Distributions of the Trend IV cointegrating coefficient estimator for various K and AR errors. All other parameters are as in Fig. 6.

especially when ρ is negative. Increasing the order of the VAR reduces the bias but also increases the dispersion of the RRR estimator. FM-OLS shows the most dispersion of these procedures, but is generally well centered. OLS is clearly biased and, interestingly, seems to have more dispersion than trending IV in almost all cases.

For VAR errors, Trend IV regression works very well and sometimes outperforms the other methods by a substantial margin. We observe that DOLS can perform quite poorly under VAR errors and can have substantial finite sample bias, as indicated in Fig. 6. This seems to be explained by the need for a large number of leads and lags to control for feedback and serial correlation, especially when the serial dependence coefficients are of different magnitudes and sign. Similarly, RRR needs four lags in order to perform adequately in such cases and, as is apparent here from the flat nature of the density in Fig. 5 and several cases in the tables, RRR is very susceptible to extreme outliers in some cases, particularly when the AR or MA coefficients are of different magnitude.

Observe that when u_t is *iid* $N(0, \Sigma)$, we have $\Omega = \Sigma$ and the equation error $u_{0.xt} = u_{0t} - \Sigma_{0x}\Sigma_{xx}^{-1}u_{xt}$ is independent of u_{xt} and is normally distributed. In this case it follows from the calculation in the Appendix that the error in the trending IV estimator has a leading term whose finite sample distribution is symmetric about the origin and mixed normal, analogous to the limit distribution. This helps to explain the good finite sample performance of A_{IV}.

Figs. 7–9 show the effects of varying K on the distribution of the Trend IV estimator. Not surprisingly, as K increases (for given n) the bias in the estimator increases and the distribution tends to the distribution of the least squares regression estimate in the augmented regression model (8), i.e., regression of y_t on x_t and Δx_t. Since $n = 50$, the curves corresponding to $K = 50$ in Figs. 7–8 correspond to OLS on (8). The curves labeled OLS in the figures correspond to OLS regression of y_t on x_t (i.e., model (1)). Thus, augmenting the regression equation itself helps to reduce the least squares regression bias. This figure shows that trend IV regression has virtually no bias when $K = 10$ in both these cases but nonnegligible bias for large values of K. These simulations therefore seem to support the conjecture made earlier that the optimal expansion rate for K in cointegrating regression is less than the optimal rate for HAC estimation. Fig. 9 shows that very similar finite sample results hold for the Trend IV estimator when it is constructed from time polynomial instruments (here, we use Legendre polynomials) rather than the sinusoidal polynomials (9).

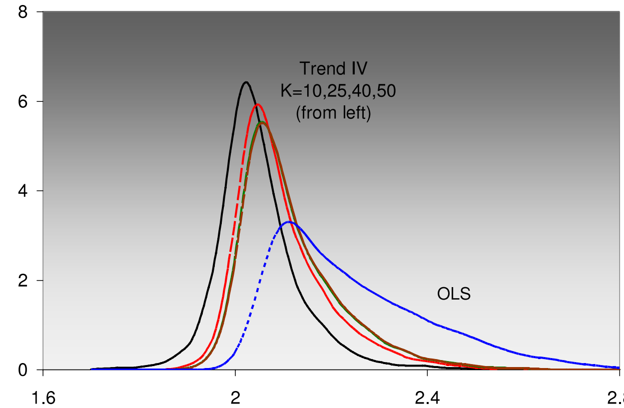

Fig. 8. Distributions of the Trend IV cointegrating coefficient estimator for various K and MA errors. All other parameters are as in Fig. 6.

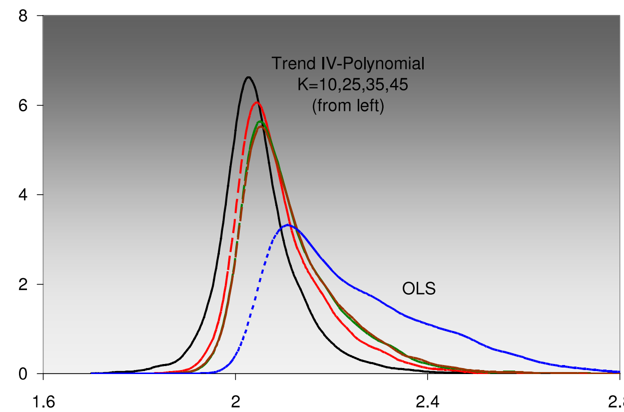

Fig. 9. Distributions of the Trend IV cointegrating coefficient estimator for various K and MA errors using Legendre polynomial instruments. All other parameters are as in Fig. 8.

Tables III and IV report simulation findings on the finite sample performance of interval estimates based on the fitted standard errors for each estimator, a 95% nominal confidence level, and an assumed normal distribution for the randomly standardized and centered statistic. To reduce space, the tables only show results for RRR (4 lags), DOLS (4 leads and lags) and Trend IV (with $K = 20$ instruments), which from the earlier results appear to give the best estimation results in finite samples. With few exceptions RRR produces confidence intervals with the longest length, often by a very large margin, while DOLS and Trend IV produce confidence intervals that are of similar length. In terms of coverage probability, Trend IV has empirical coverage closer to the nominal 95% than DOLS in every case, and is often much closer to the nominal. The coverage probability of Trend IV is also better than that of the RRR intervals. Overall, these findings suggest that Trend IV interval estimates are superior to those of RRR and DOLS, the former producing intervals that are longer and the latter producing intervals with less accurate coverage probability.

Table III
Finite sample performance of interval estimators with AR errors, $b = 2$, $\rho = 0.75$, $T = 50$, $N = 10{,}000$ replications.

AR coeffs (θ_1, θ_2)	Estimator	Coverage prob	Length	AR coeffs (θ_1, θ_2)	Coverage prob	Length
(0.8, 0.8)	RRR_4	0.644	0.324	(−0.8, −0.8)	0.858	0.109
	$DOLS_4$	0.471	0.055		0.999	0.311
	Trend IV	0.633	0.062		0.930	0.117
(0.4, 0.4)	RRR_4	0.792	0.116	(−0.4, −0.4)	0.846	0.108
	$DOLS_4$	0.783	0.087		0.984	0.179
	Trend IV	0.860	0.092		0.927	0.113
(−0.4, 0.4)	RRR_4	0.818	0.256	(−0.8, 0.8)	0.706	1.220
	$DOLS_4$	0.626	0.221		0.246	0.421
	Trend IV	0.798	0.231		0.402	0.547

Table IV
Finite sample performance of interval estimators with MA errors, $b = 2$, $\rho = 0.75$, $T = 50$, $N = 10{,}000$ replications.

AR coeffs (θ_1, θ_2)	Estimator	Coverage prob	Length	AR coeffs (θ_1, θ_2)	Coverage prob	Length
(0.8, 0.8)	RRR_4	0.805	0.132	(−0.8, −0.8)	0.971	0.147
	$DOLS_4$	0.802	0.091		0.998	0.428
	Trend IV	0.899	0.102		0.988	0.218
(0.4, 0.4)	RRR_4	0.815	0.134	(−0.4, −0.4)	0.869	0.110
	$DOLS_4$	0.841	0.098		0.994	0.201
	Trend IV	0.902	0.103		0.955	0.129
(−0.4, 0.4)	RRR_4	0.841	0.254	(−0.8, 0.8)	0.800	1.050
	$DOLS_4$	0.638	0.248		0.171	0.558
	Trend IV	0.830	0.260		0.445	0.722

6. Trend likelihood

Define the trend (φ) transform of a multiple time series a_t as $\xi_k^a = \sum_{t=1}^n \frac{a_t}{\sqrt{n}}\varphi_k(\frac{t}{n})$ and the corresponding matrix transform as $\xi_K^a = \sum_{t=1}^n \frac{u_t}{\sqrt{n}}\tilde{\varphi}_K(\frac{t}{n})'$. This transform simply projects the observations onto the space of the instruments Φ_K.

It follows just as in the proof of (54) in the Theorem that we have the representation

$$\xi_K^u = \frac{U'\Phi_K}{\sqrt{n}} = \int_0^1 dB(r)\,\tilde{\varphi}_K(r)' + O\left(\frac{K^2}{n^2}\right) + o_p\left(n^{-\frac{1}{2}+\frac{1}{\nu}}\right), \quad (16)$$

where the error order holds uniformly over the columns for $k = 1, \ldots, K$. Since the first component of (16) is Gaussian with covariance matrix $\Omega \otimes I_K$, the (negative) log likelihood function of ξ_K^u is approximately (up to scaling and an error that can be neglected in view of (16))

$$L(\Omega) = K \log|\Omega| + \operatorname{tr}\left\{\Omega^{-1}\xi_K^u \xi_K^{u\prime}\right\}, \quad (17)$$

which we may regard as a *trend likelihood* because the φ transform ξ_K^u focuses attention on the long-run components of u_t. If u_t were observed, minimization of (17) would lead directly to the long-run covariance matrix estimate $\hat{\Omega} = K^{-1}\xi_K^u\xi_K^{u\prime} = K^{-1}\sum_{t=1}^n \xi_k^u \xi_k^{u\prime}$, which is the HAC estimator developed in Phillips (2005a,b). So, $\hat{\Omega}$ may be considered a trend MLE in the sense that it optimizes the trend likelihood (17).

Since

$$\xi_K^u \sim \int_0^1 dB(r)\,\tilde{\varphi}_K(r)' = \begin{bmatrix} I & \Omega_{0x}\Omega_{xx}^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} \int_0^1 dB_{0.x}(r)\,\tilde{\varphi}_K(r)' \\ \int_0^1 dB_x(r)\,\tilde{\varphi}_K(r)' \end{bmatrix}$$

$$:= \begin{bmatrix} I & \Omega_{0x}\Omega_{xx}^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} \xi_K^{u_{0.x}} \\ \xi_K^{u_x} \end{bmatrix},$$

and $B_{0.x}$ is independent of B_x, the likelihood (17) transforms to the sum

$$\underbrace{K\log|\Omega_{0.xx}| + \operatorname{tr}\left\{\Omega_{0.xx}^{-1}\xi_K^{u_{0.x}}\xi_K^{u_{0.x}\prime}\right\}}_{L(\Omega_{0.xx})} + \underbrace{K\log|\Omega_{xx}| + \operatorname{tr}\left\{\Omega_{xx}^{-1}\xi_K^{u_x}\xi_K^{u_x\prime}\right\}}_{L(\Omega_{xx})}. \quad (18)$$

To make this likelihood data dependent, we use the fact that in φ transform form the model is

$$\xi_K^y = A\xi_K^x + \Omega_{0x}\Omega_{xx}^{-1}\xi_K^{\Delta x} + \xi_K^{u_{0.x}}, \quad \xi_K^{\Delta x} = \xi_K^{u_{xt}}. \quad (19)$$

The Jacobian of the transformation in (19) is unity and $L(\Omega_{xx})$ does not depend on the cointegrating matrix A or the long-run regression coefficient matrix $C = \Omega_{0x}\Omega_{xx}^{-1}$. Hence, the trend MLE estimator satisfies

$$\left(\hat{A}, \hat{C}, \hat{\Omega}_{0.xx}\right) = \arg\min_{A, C, \Omega_{0.xx}} L(A, C, \Omega_{0.xx}),$$

where

$$L(A, C, \Omega_{0.xx}) = K\log|\Omega_{0.xx}| + \operatorname{tr}\left\{\Omega_{0.xx}^{-1}\left(\xi_K^y - A\xi_K^x - C\xi_K^{\Delta x}\right) \times \left(\xi_K^y - A\xi_K^x - C\xi_K^{\Delta x}\right)'\right\}.$$

Concentrating out $\Omega_{0.xx}$ leads directly to the IV estimator given in (10). Thus, the estimates A_{IV}, C_{IV} and $\Omega_{0.xx}^{IV}$ may all be regarded as trend maximum likelihood estimates.

The trend likelihood (17) is Gaussian because it makes use of the asymptotic normality of the transformed variables ξ_K^a. So one advantage of projecting on the trend instrument space is that the data become approximately normal, just as discrete Fourier transforms of stationary time series are approximately normal. In this regard, the trend likelihood is analogous to the local Whittle likelihood for frequencies in the vicinity of the origin. This means that common applications of narrow-band frequency domain techniques, may also be approached using trend likelihood methods that do not involve complex arithmetic.

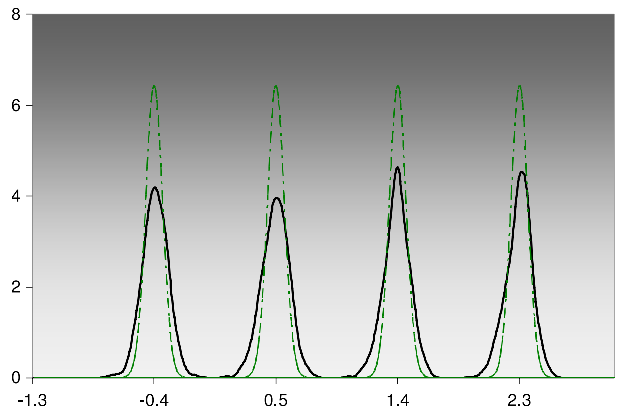

Fig. 10. Trend MLE estimates of d for $n = 200$, $K = 65$, and $d_0 = -0.4, 0.5, 1.4, 2.3$. The solid curves show kernel estimates of the densities of $\hat{d}$ for each of these four cases and the broken curves represent the limit distribution $N\left(d_0, \frac{1}{4K}\right)$.

One example that is important in recent econometric research is the semiparametric estimation of long memory. Phillips (1999) and Shimotsu and Phillips (2005) show how to construct an exact form of the local Whittle (LW) likelihood for a long memory process X_t generated by the model $(1-L)^{d_0} X_t = u_t \mathbf{1}\{t \geq 1\}$ allowing for the memory parameter to take any value d_0 on the real line and where u_t is a short memory process with spectrum $f_u(\lambda)$. The exact LW likelihood has the form

$$\frac{1}{m}\sum_{j=1}^{m}\left[\log\left(G\lambda_j^{-2d}\right)+\frac{1}{G}I_{\Delta^d x}\left(\lambda_j\right)\right], \tag{20}$$

where $G = f_u(0) = \frac{1}{2\pi}\omega^2$, $\lambda_j = \frac{2\pi j}{n}$, m defines the upper limit of the frequency band, and $I_{\Delta^d x}\left(\lambda_j\right)$ is the periodogram of $\Delta^d X_t = (1-L)^d X_t$. Shimotsu and Phillips (2005) show that under broad regularity conditions the exact LW (ELW) estimator $\hat{d}$ that minimizes (20) is consistent and has the limit distribution $\sqrt{m}\left(\hat{d}-d_0\right) \Rightarrow N\left(0, \frac{1}{4}\right)$. The ELW estimator is a good general purpose estimator of the long memory parameter, covers stationary and nonstationary cases and is well-suited to confidence interval construction.

Analogous to (20) and using the operator algebra from Phillips (1999), we may construct a trend likelihood for the φ transform $\xi_K^{\Delta^d X}$ using the trend instruments (9). The trend likelihood turns out to have the following form

$$\frac{1}{K}\sum_{j=1}^{K}\left[\log\omega^2-2d\log\left(\frac{\pi\left(k-\frac{1}{2}\right)}{n}\right)\right]+\frac{1}{\omega^2}\xi_K^{\Delta^d X\prime}\xi_K^{\Delta^d X}, \tag{21}$$

which may be minimized with respect to ω^2 (the long-run variance of u_t) and d to get the corresponding Trend IV estimates $\hat{\omega}_{IV}^2$ and $\hat{d}_{IV}$. Simulations reveal that these estimates have performance characteristics close to those of the ELW estimates and we conjecture that $\hat{d}_{IV}$ has the same limit distribution as the ELW estimator for all values of d_0. Again, the form of (21) avoids the use of complex arithmetic. Fig. 10 illustrates by showing the densities of $\hat{d}_{IV}$ calculated from 10,000 replications when $n = 200$, $K = 65$ and the true memory parameter has the four values $d_0 = -0.4, 0.5, 1.4, 2.3$. Also shown in the figure are the corresponding normal densities $N\left(d_0, \frac{1}{4K}\right)$ for each of these cases.

7. Conclusion and extensions

The results of this paper highlight some of the advantages of working with an agnostic set of basis functions in capturing the effects of trend. The use of trend basis functions as instruments in regression focuses attention on the long-run behavior of the system variables in both levels and differences. For the nonstationary variables in levels, the regression provides optimal estimates of the cointegrating coefficients. For the stationary variables that appear as differences of the system variables, the regression produces the long-run covariance and regression coefficients that capture and adjust for the effects of simultaneity in the system.

Thus, using instrumental variables from an agnostic set of trend basis functions can be viewed as a simple regression device for detecting long-run effects in an econometric model. In this respect, the device operates in the same way as narrow-band frequency domain techniques that concentrate solely on low frequencies. But it has the advantages of completely avoiding the complications of the frequency domain and having a very simple interpretation that should be appealing to applied researchers. For practical purposes, the approach is very easy to implement, provides asymptotically valid standard errors and tests from the usual regression output, and requires only basic econometric software packages to implement. Finite sample performance seems superior to existing procedures in terms of estimation (where there is generally less bias) and inference (where the coverage probability of confidence intervals is more accurate and length is often shorter).

While it is not mentioned earlier, it should be clear that the approach applies without modification when the cointegrating regression involves an intercept or when there is deterministic trend cointegration. In both cases, the trend basis instruments continue to provide asymptotically efficient estimates of the cointegrating coefficients and no other instruments are required. The methods may also be extended to systems with different normalizations than the triangular system used in (1)–(2) and as such may be used for testing the dimension of the cointegrating space.

The instruments considered in this paper are deterministic functions. We might also consider the use of a collection of integrated series as instruments, following the original analysis in Phillips and Hansen (1990). Large numbers of such instruments are also capable of modeling trend regressors with an R^2 that approaches unity, as shown in Phillips (1998), but are obviously harder to justify in practical work. In consequence, it seems possible that the results given here may be extended to include such regressors. However, it is also necessary that such instruments be capable of modeling the long-run regression coefficients of stationary series and that remains to be proved.

The use of agnostic deterministic instruments opens up the interesting question of instrument selection in practical work. While the KL representation (5) gives a natural sequential ordering for the deterministic instruments in terms of the eigenvalue magnitude, we may also use data determined techniques to select instruments. In this regard, modern shrinkage methods such as those suggested in Liao (2013) for moment condition selection seem promising for use in the present context where there is instrument selection and potential rank reduction—see Liao and Phillips (2012). We hope to explore some of these extensions of the present methods in later work.

Appendix. Lemmas and proofs

Lemma A (*Phillips (2007, Lemma 3.1)*). *If u_t satisfies* **L**, *the probability space which supports u_t can be expanded in such a way that there exists a process distributionally equivalent to $B_n(\cdot) = n^{-1/2}\sum_{i=1}^{\lfloor n\cdot\rfloor} u_i$ and a Brownian motion $B(\cdot)$ with variance matrix Ω on the new space for which*

$$\sup_{t\in[0,1]}\|B_n(t)-B(t)\| = o_p\left(\frac{1}{n^{\frac{1}{2}-\frac{1}{\nu}}}\right) \quad \text{as } n\to\infty. \tag{22}$$

Proof. The result follows as in Phillips (2007, Lemma 3.1). An "in probability" approximation is all that is needed here. But, as discussed in that reference, a strong approximation of the same form is also possible, albeit under stronger moment conditions.[2] ■

The following two results are based on results proved in Phillips (2005b).

Lemma B (*Phillips (2005b, Lemma A)*). *Under* **R**, $n^{-1}\sum_{t=1}^{n}\varphi_{Kt}\varphi_{Kt}' = I_K + O\left(\frac{1}{n}\right)$, *and* $\left(n^{-1}\sum_{t=1}^{n}\varphi_{Kt}\varphi_{Kt}'\right)^{-1} = I_K + O\left(\frac{1}{n}\right)$, *as* $n, K \to \infty$.

Remark. The proof of Lemma A in Phillips (2005b) establishes the explicit form

$$n^{-1}\sum_{t=1}^{n}\varphi_{Kt}\varphi_{Kt}' = I_K + \frac{1}{n}G, \tag{23}$$

where $G = (g_{mk})$ has elements

$$g_{mk} = \begin{cases} 2 & k = m \\ 1 & k \neq m \text{ and } m-k \text{ odd} \\ -1 & k \neq m \text{ and } m-k \text{ even.} \end{cases} \tag{24}$$

Thus, G has uniformly bounded elements as $n, K \to \infty$.

Lemma C. *Let* $\hat{\Omega}_K = K^{-1}U'P_KU$. *Then, under* **L** *and when* $\frac{1}{K} + \frac{K}{n} \to 0$ *we have:*

(a) $\lim_{n\to\infty}\left(\frac{n}{K}\right)^2 E\left(\hat{\Omega}_K - \Omega\right) = -\frac{\pi^2}{6}\sum_{h=-\infty}^{\infty} h^2\Gamma_u(h) := D$;

(b) *If* $K = o\left(n^{4/5}\right)$, *then* $\sqrt{K}\left(vec\left(\hat{\Omega}_K\right) - vec(\Omega)\right) \Rightarrow N(0, 2P_D(\Omega\otimes\Omega))$ *where* $P_D = D\left(D'D\right)^{-1}D'$ *projects onto the range of the duplicator matrix* D *for which* $D\omega = vec(\Omega)$ *where* ω *is the vector of nonredundant elements of* Ω;

(c) *If* $K^5/n^4 \to 1$, *then* $\lim_{n\to\infty}\left(\frac{n}{K}\right)^4 E\left(vec\left(\hat{\Omega}_K\right) - vec(\Omega)\right)\left(vec\left(\hat{\Omega}_K\right) - vec(\Omega)\right)' = vec(D)vec(D)' + 2P_D(\Omega\otimes\Omega.)$

(d) $K^{-1}U'P_KU \to_p \Omega$.

Proof. These results are simply matrix generalizations of the results in Phillips (2005b). ■

Lemma D.

$$Z_{n,k} := n^{-1}\sum_{t=1}^{n}B\left(\frac{t}{n}\right)\varphi_k\left(\frac{t}{n}\right) - \int_0^1 B(r)\varphi_k(r)dr = \frac{1}{n}\left(\sum_{m=1}^{\infty}\lambda_m^{\frac{1}{2}}\xi_m g_{mk}\right),$$

and $\max_{1\le k\le K}\left\|Z_{n,k}\right\| = O_p\left(\frac{(\log K)^{1/2}}{n}\right)$.

Proof. $Z_{n,k}$ is Gaussian with zero mean. Using the KL expansion (5) and that fact that φ_k is an eigenfunction with eigenvalue λ_k of the covariance kernel of $B(r)$, we have $\int_0^1 B(r)\varphi_k(r)dr = \lambda_k^{1/2}\xi_k$ and then, following a suggestion of a referee,

$$\begin{aligned} Z_{n,k} &= n^{-1}\sum_{t=1}^{n}B\left(\frac{t}{n}\right)\varphi_k\left(\frac{t}{n}\right) - \lambda_k^{1/2}\xi_k \\ &= n^{-1}\sum_{t=1}^{n}\left(\sum_{m=1}^{\infty}\lambda_m^{\frac{1}{2}}\varphi_m\left(\frac{t}{n}\right)\xi_m\right)\varphi_k\left(\frac{t}{n}\right) - \lambda_k^{1/2}\xi_k \\ &= \sum_{m=1}^{\infty}\lambda_m^{\frac{1}{2}}\xi_m\left\{n^{-1}\sum_{t=1}^{n}\varphi_m\left(\frac{t}{n}\right)\varphi_k\left(\frac{t}{n}\right) - \delta_{mk}\right\} \\ &= \frac{1}{n}\left(\sum_{m=1}^{\infty}\lambda_m^{\frac{1}{2}}\xi_m g_{mk}\right) = O_p\left(n^{-1}\right), \end{aligned}$$

since $U_k = \sum_{m=1}^{\infty}\lambda_m^{\frac{1}{2}}\xi_m g_{mk}$ is $N(0, \lambda_k\Omega)$ with $\lambda_k = \sum_{m=1}^{\infty}\lambda_m g_{mk}^2 \le 4\sum_{m=1}^{\infty}\lambda_m < \infty$ because $|g_{mk}| \le 2$ from (24), and $\sum_{m=1}^{\infty}\lambda_m^{\frac{1}{2}}\xi_m g_{mk}$ converges almost surely by virtue of the Martingale convergence theorem.

Next, in view of (24) we have the equivalence

$$Z_{n,k} =_d Z_{n,k}' = \frac{1}{n}\left(\sum_{m=1}^{\infty}\lambda_m^{\frac{1}{2}}\xi_m\right) + \frac{1}{n}\lambda_k^{\frac{1}{2}}\xi_k$$

and

$$\begin{aligned} \max_{1\le k\le K}\left\|Z_{n,k}'\right\| &\le \frac{1}{n}\left\|\sum_{m=1}^{\infty}\lambda_m^{\frac{1}{2}}\xi_m\right\| + \frac{1}{n}\max_{1\le k\le K}\lambda_k^{\frac{1}{2}}\left\|\xi_k\right\| \\ &= O_p\left(\frac{(\log K)^{1/2}}{n}\right), \end{aligned}$$

since the maximum extreme value of K independent normal random variables is $O_{a.s}\left((\log K)^{1/2}\right)$—see Galambos (1978). Hence, $Z_{n,k}$ is of order $O_p\left((\log K)^{1/2}/n\right)$ uniformly in $k \le K$.

A.1. Proof of the Theorem

Write

$$n\left(A_{IV} - A\right) = \left(n^{-1}U_0'R_KX\right)\left(n^{-2}X'R_KX\right)^{-1}. \tag{25}$$

We begin by considering the various terms in the denominator of this matrix quotient, which we may expand as follows

$$n^{-2}X'R_KX = n^{-2}X'P_KX - K^{-1}\left(n^{-1}X'P_K\Delta X\right)\left(K^{-1}\Delta X'P_K\Delta X\right)^{-1} \times \left(n^{-1}\Delta X'P_KX\right). \tag{26}$$

Starting with the first term in (26) we have

$$\frac{1}{n^2}X'P_KX = \left(\frac{1}{n}\frac{X'\Phi_K}{\sqrt{n}}\right)\left(\frac{\Phi_K'\Phi_K}{n}\right)^{-1}\left(\frac{1}{n}\frac{\Phi_K'X}{\sqrt{n}}\right). \tag{27}$$

From the approximation (22) we can write

$$\frac{x_t}{\sqrt{n}} = B_x\left(\frac{t}{n}\right) + \zeta_{xt} = B_x\left(\frac{t}{n}\right) + o_p\left(\frac{1}{n^{\frac{1}{2}-\frac{1}{v}}}\right),$$

uniformly over $t = 1, \ldots, n$, and by Lemma D

$$n^{-1}\sum_{t=1}^{n}B_x\left(\frac{t}{n}\right)\varphi_k\left(\frac{t}{n}\right) = \int_0^1 B_x(r)\varphi_k(r)dr + O_p\left(\frac{(\log K)^{1/2}}{n}\right),$$

[2] In particular, recent results on multivariate strong approximation (e.g., Zaitsev (1998)) for partial sums of iid vectors can be used in combination with the Phillips and Solo (1992) device to prove a strong approximation for partial sums of a multivariate linear process. These results ensure a uniform error of $O_{a.s.}\left(\frac{\log n}{\sqrt{n}}\right)$ when the variates have exponential moments.

uniformly in $k \leq K$. Then, as in the proof of Lemma 2.2 of Phillips (2002), the first factor of (27) is

$$\begin{aligned}
\frac{1}{n}\frac{X'\Phi_K}{\sqrt{n}} &= \frac{1}{n}\sum_{t=1}^{n}\frac{x_t}{\sqrt{n}}\varphi_{Kt}' \\
&= \left(n^{-1}\sum_{t=1}^{n}\left[B_x\left(\frac{t}{n}\right)+\zeta_{xt}\right]\varphi_{Kt}'\right) \\
&= \int_0^1 B_x(r)\,\tilde{\varphi}_K(r)'\,dr + n^{-1}\sum_{t=1}^{n}\zeta_{xt}\varphi_{Kt}' \\
&\quad + O_p\left(\frac{(\log K)^{1/2}}{n}\right) \\
&= \Xi_{xK}\Lambda_K^{\frac{1}{2}} + n^{-1}\sum_{t=1}^{n}\zeta_{xt}\varphi_{Kt}' + O_p\left(\frac{(\log K)^{1/2}}{n}\right) \\
&:= \Xi_{xK}\Lambda_K^{\frac{1}{2}} + \eta_{Kn}', \\
&= \Xi_{xK}\Lambda_K^{\frac{1}{2}} + o_p\left(\frac{1}{n^{\frac{1}{2}-\frac{1}{v}}}\right)
\end{aligned} \tag{28}$$

since $\zeta_{xt} = o_p\left(\frac{1}{n^{\frac{1}{2}-\frac{1}{v}}}\right)$ uniformly over t. It follows that

$$\begin{aligned}
\left(\frac{X'\Phi_K}{n^{3/2}}\right)\left(\frac{\Phi_K'X}{n^{3/2}}\right) &= \left\{\Xi_{xK}\Lambda_K^{\frac{1}{2}} + \eta_{Kn}'\right\}\left\{\Lambda_K^{\frac{1}{2}}\Xi_{xK}' + \eta_{Kn}\right\} \\
&= \Xi_{xK}\Lambda_K\Xi_{xK}' + \eta_{Kn}'\Lambda_K^{1/2}\Xi_{xK}' \\
&\quad + \Xi_{xK}\Lambda_K^{\frac{1}{2}}\eta_{Kn} + \eta_{Kn}'\eta_{Kn} \\
&= \Xi_{xK}\Lambda_K\Xi_{xK}' + o_p\left(\frac{K^{1/2}}{n^{\frac{1}{2}-\frac{1}{v}}} + \frac{K}{n^{1-\frac{2}{v}}}\right).
\end{aligned} \tag{29}$$

The error magnitude in (29) holds because, taking the i'th row, η_{iKn}', of η_{Kn}', we have

$$\eta_{iKn}'\eta_{iKn} = o_p\left(\frac{K}{n^{1-\frac{2}{v}}}\right), \tag{30}$$

and, denoting the j'th column of Ξ_{xK}' by Ξ_{xjK}',

$$\begin{aligned}
\left|\eta_{iKn}'\Lambda_K^{1/2}\Xi_{xjK}'\right| &\leq \left(\eta_{iKn}'\eta_{iKn}\right)^{1/2}\left(\Xi_{xjK}\Lambda_K\Xi_{xjK}'\right)^{1/2} \\
&= o_p\left(\frac{K^{1/2}}{n^{\frac{1}{2}-\frac{1}{v}}}\right),
\end{aligned} \tag{31}$$

which combine to give (29) and hence

$$\left(\frac{X'\Phi_K}{n^{3/2}}\right)\left(\frac{\Phi_K'X}{n^{3/2}}\right) = \Xi_{xK}\Lambda_K\Xi_{xK}' + o_p\left(\frac{K^{1/2}}{n^{\frac{1}{2}-\frac{1}{v}}}\right).$$

Next, observe that

$$\begin{aligned}
\Xi_{xK}\Lambda_K\Xi_{xK}' &= \sum_{k=1}^{K}\lambda_k\xi_{xk}\xi_{xk}' = \sum_{k=1}^{\infty}\lambda_k\xi_{xk}\xi_{xk}' - \sum_{k=K+1}^{\infty}\lambda_k\xi_{xk}\xi_{xk}' \\
&= \int_0^1 B_xB_x' - \sum_{k=K+1}^{\infty}\lambda_k\xi_{xk}\xi_{xk}',
\end{aligned} \tag{32}$$

since, by orthonormality of the $\varphi_k(r)$, we have the following alternate representation

$$\begin{aligned}
\int_0^1 B_xB_x' &= \int_0^1\left(\sum_{k=1}^{\infty}\lambda_k^{\frac{1}{2}}\varphi_k(r)\xi_{xk}\right)\left(\sum_{k=1}^{\infty}\lambda_k^{\frac{1}{2}}\varphi_k(r)\xi_{xk}'\right)dr \\
&= \sum_{k=1}^{\infty}\lambda_k\xi_{xk}\xi_{xk}'.
\end{aligned}$$

Thus,

$$\begin{aligned}
\left(\frac{X'\Phi_K}{n^{3/2}}\right)\left(\frac{\Phi_K'X}{n^{3/2}}\right) &= \int_0^1 B_xB_x' - \sum_{k=K+1}^{\infty}\lambda_k\xi_{xk}\xi_{xk}' + o_p\left(\frac{K^{1/2}}{n^{\frac{1}{2}-\frac{1}{v}}}\right) \\
&= \int_0^1 B_xB_x' + O_p\left(\frac{1}{K}\right) + o_p\left(\frac{K^{1/2}}{n^{\frac{1}{2}-\frac{1}{v}}}\right),
\end{aligned} \tag{33}$$

since $E\left\|\sum_{k=K+1}^{\infty}\lambda_k\xi_{xk}\xi_{xk}'\right\| \leq \text{const.}\sum_{k=K+1}^{\infty}\frac{1}{k^2}E\xi_{xk}'\xi_{xk} = O\left(\frac{1}{K}\right)$.

Next observe that

$$\left(n^{-1}\sum_{t=1}^{n}\varphi_{Kt}\varphi_{Kt}'\right)^{-1} = \left(I_K + \frac{1}{n}G\right)^{-1} = I_K - \frac{1}{n}G + O\left(n^{-2}\right)$$

where the elements of the $K \times K$ matrix G are uniformly $O(1)$ as shown in (24). Then

$$\begin{aligned}
&\left(\frac{1}{n}\frac{X'\Phi_K}{\sqrt{n}}\right)\left(n^{-1}\sum_{t=1}^{n}\varphi_{Kt}\varphi_{Kt}'\right)^{-1}\left(\frac{1}{n}\frac{\Phi_K'X}{\sqrt{n}}\right) \\
&= \left(\frac{1}{n}\frac{X'\Phi_K}{\sqrt{n}}\right)\left(\frac{1}{n}\frac{\Phi_K'X}{\sqrt{n}}\right) \\
&\quad + \frac{1}{n}\left(\frac{1}{n}\frac{X'\Phi_K}{\sqrt{n}}\right)\left(G + O\left(\frac{1}{n}\right)\right)\left(\frac{1}{n}\frac{\Phi_K'X}{\sqrt{n}}\right) \\
&= \int_0^1 B_xB_x' + o_p\left(\frac{K^{1/2}}{n^{\frac{1}{2}-\frac{1}{v}}}\right) + O_p\left(\frac{1}{K}\right) \\
&\quad + O_p\left(\frac{1}{n}\left(\Xi_{xK}\Lambda_K^{\frac{1}{2}} + \eta_{Kn}'\right)\right. \\
&\quad \left.\times\left(G + O\left(\frac{1}{n}\right)\right)\left(\Lambda_K^{\frac{1}{2}}\Xi_{xK}' + \eta_{Kn}\right)\right).
\end{aligned}$$

As in (30) and (31) we have

$$\left|\eta_{iKn}'G\eta_{jKn}\right| \leq \left(\eta_{iKn}'GG'\eta_{iKn}\right)^{1/2}\left(\eta_{jKn}'\eta_{jKn}\right)^{1/2} = o_p\left(\frac{K}{n^{1-\frac{2}{v}}}\right),$$

and

$$\left|\eta_{iKn}'G\Lambda_K^{\frac{1}{2}}\Xi_{xjK}'\right| \leq \left(\eta_{iKn}'GG'\eta_{iKn}\right)^{1/2}\left(\Xi_{xjK}\Lambda_K\Xi_{xjK}'\right)^{1/2} = o_p\left(\frac{K^{1/2}}{n^{\frac{1}{2}-\frac{1}{v}}}\right)$$

so that

$$\begin{aligned}
&O_p\left(\frac{1}{n}\left(\Xi_{xK}\Lambda_K^{\frac{1}{2}} + \eta_{Kn}'\right)G\left(\Lambda_K^{\frac{1}{2}}\Xi_{xK}' + \eta_{Kn}\right)\right) \\
&= o_p\left(\frac{K^{1/2}}{n^{\frac{3}{2}-\frac{1}{v}}} + \frac{K}{n^{2-\frac{2}{v}}}\right).
\end{aligned}$$

We deduce that

$$\frac{1}{n^2}X'P_KX = \int_0^1 B_xB_x' + o_p\left(\frac{K^{1/2}}{n^{\frac{1}{2}-\frac{1}{v}}}\right) + O_p\left(\frac{1}{K}\right) \Rightarrow \int_0^1 B_xB_x', \tag{34}$$

as $n \to \infty$.

Now consider the second term in (26), viz.

$$K^{-1}\left(n^{-1}X'P_K\Delta X\right)\left(K^{-1}\Delta X'P_K\Delta X\right)^{-1}\left(n^{-1}\Delta X'P_KX\right). \tag{35}$$

First, from Lemma B,

$$K^{-1}\Delta X'P_K\Delta X \to_p \Omega_{xx} > 0. \tag{36}$$

Next,

$$\begin{aligned}
n^{-1}X'P_K\Delta X &= \left(\frac{1}{n}\frac{X'\Phi_K}{\sqrt{n}}\right)\left(\frac{\Phi_K'\Phi_K}{n}\right)^{-1}\left(\frac{\Phi_K'\Delta X}{\sqrt{n}}\right) \\
&= \left(\frac{1}{n}\frac{X'\Phi_K}{\sqrt{n}}\right)\left(I_K - \frac{1}{n}G + O\left(n^{-2}\right)\right)\left(\frac{\Phi_K'\Delta X}{\sqrt{n}}\right) \\
&= \left(\frac{1}{n}\frac{X'\Phi_K}{\sqrt{n}}\right)\left(\frac{\Phi_K'\Delta X}{\sqrt{n}}\right) - \frac{1}{n}\left(\frac{1}{n}\frac{X'\Phi_K}{\sqrt{n}}\right) \\
&\quad \times\left(G + O\left(\frac{1}{n}\right)\right)\left(\frac{\Phi_K'\Delta X}{\sqrt{n}}\right).
\end{aligned} \tag{37}$$

The limit form of $n^{-3/2}X'\Phi_K$ is given above in (28), so we concentrate on the second factor, $n^{-1/2}\Phi_K'\Delta X = n^{-1/2}\Phi_K'U_x$. Using partial summation and setting $S_{xt} = \sum_{s=1}^{t} u_{xs}$, we get

$$\frac{U_x'\Phi_K}{\sqrt{n}} = \sum_{t=1}^{n} \frac{u_{xt}}{\sqrt{n}}\varphi_{Kt}' = \frac{1}{\sqrt{n}}S_{xn}\tilde{\varphi}_K(1)' - \sum_{t=1}^{n}\frac{S_{xt-1}}{\sqrt{n}}\Delta\varphi_{Kt}'. \tag{38}$$

Note that

$$\begin{aligned}
\Delta\varphi_k\left(\frac{t}{n}\right) &= \sqrt{2}\left[\sin\left\{\left(k-\frac{1}{2}\right)\pi\frac{t}{n}\right\} - \sin\left\{\left(k-\frac{1}{2}\right)\pi\frac{t-1}{n}\right\}\right] \\
&= \sqrt{2}2\sin\left\{\frac{1}{2}\left(k-\frac{1}{2}\right)\pi\frac{1}{n}\right\}\cos\left\{\left(k-\frac{1}{2}\right)\pi\frac{t-\frac{1}{2}}{n}\right\} \\
&= \sqrt{2}2\frac{\sin\left\{\frac{1}{2}\left(k-\frac{1}{2}\right)\pi\frac{1}{n}\right\}}{\frac{1}{2}\left(k-\frac{1}{2}\right)\pi\frac{1}{n}}\cos\left\{\left(k-\frac{1}{2}\right)\pi\frac{t-\frac{1}{2}}{n}\right\} \times\frac{1}{2}\left(k-\frac{1}{2}\right)\pi\frac{1}{n} \\
&= \sqrt{2}\left[1+O\left(\frac{K^2}{n^2}\right)\right]\cos\left\{\left(k-\frac{1}{2}\right)\pi\frac{t-\frac{1}{2}}{n}\right\} \times\left(k-\frac{1}{2}\right)\pi\frac{1}{n} \\
&= \varphi_k^{(1)}\left(\frac{t-\frac{1}{2}}{n}\right)\frac{1}{n}\left\{1+O\left(\frac{K^2}{n^2}\right)\right\},
\end{aligned} \tag{39}$$

uniformly in $k = 1, \ldots, K$. The approximation (7) implies that

$$\sup_{r\in[0,1]}\left\|n^{-1/2}\sum_{t=1}^{\lfloor nr\rfloor}u_{xt} - B_x(r)\right\| = o_p\left(n^{-\frac{1}{2}+\frac{1}{v}}\right),$$

as $n\to\infty$, and so, using (39) and Lemma D, we have

$$\begin{aligned}
&\frac{1}{\sqrt{n}}S_{xn}\tilde{\varphi}_K(1)' - \sum_{t=1}^{n}\frac{S_{xt-1}}{\sqrt{n}}\Delta\varphi_{Kt}' \\
&= \left(B_x(1)+o_p\left(\frac{1}{n^{\frac{1}{2}-\frac{1}{v}}}\right)\right)\tilde{\varphi}_K(1)' - \sum_{t=1}^{n}\frac{S_{xt-1}}{\sqrt{n}}\Delta\varphi_{Kt}' \\
&= \left(B_x(1)+o_p\left(\frac{1}{n^{\frac{1}{2}-\frac{1}{v}}}\right)\right)\tilde{\varphi}_K(1)' - \int_0^1 B_x(r)\,\tilde{\varphi}_K^{(1)}(r)'dr\left\{1+O\left(\frac{K^2}{n^2}\right)\right\} + O_p\left(\frac{1}{\sqrt{n}}\right) \\
&= \int_0^1 dB_x(r)\,\tilde{\varphi}_K(r)'dr\left\{1+O\left(\frac{K^2}{n^2}\right)\right\} + o_p\left(\frac{1}{n^{\frac{1}{2}-\frac{1}{v}}}\right).
\end{aligned}$$

Thus, we may write

$$\frac{U_x'\Phi_K}{\sqrt{n}} = \int_0^1 dB_x(r)\,\tilde{\varphi}_K(r)'dr + \vartheta_{xKn}', \tag{40}$$

where the elements of ϑ_{xKn} are uniformly $O\left(\frac{K^2}{n^2}\right) + o_p\left(n^{-\frac{1}{2}+\frac{1}{v}}\right)$ over $k = 1, \ldots, K$.

Combining (28) and (40) we have

$$\begin{aligned}
&\left(\frac{1}{n}\frac{X'\Phi_K}{\sqrt{n}}\right)\left(\frac{\Phi_K'\Delta X}{\sqrt{n}}\right) \\
&= \left(\Xi_{xK}\Lambda_K^{\frac{1}{2}} + \eta_{Kn}'\right)\left(\int_0^1\tilde{\varphi}_K(r)dB_x(r)'\,dr + \vartheta_{xKn}\right) \\
&= \int_0^1 \Xi_{xK}\Lambda_K^{\frac{1}{2}}\tilde{\varphi}_K(r)dB_x(r)' + \Xi_{xK}\Lambda_K^{\frac{1}{2}}\vartheta_{xKn} + \eta_{Kn}'\vartheta_{xKn} \\
&\quad + \eta_{Kn}'\int_0^1\tilde{\varphi}_K(r)dB_x(r)'.
\end{aligned}$$

In view of (32) and the order of the elements of ϑ_{xKn} and η_{Kn}, and denoting the j'th row of ϑ_{xKn}' by ϑ_{x_jKn}', we have

$$\begin{aligned}
\left|\Xi_{x_jK}\Lambda_K^{1/2}\vartheta_{x_jKn}\right| &\le \left(\vartheta_{x_jKn}'\vartheta_{x_jKn}\right)^{1/2}\left(\Xi_{x_jK}\Lambda_K\Xi_{x_jK}'\right)^{1/2} \\
&= O_p\left(\frac{K^{5/2}}{n^2}\right) + o_p\left(\frac{K^{1/2}}{n^{\frac{1}{2}-\frac{1}{v}}}\right),
\end{aligned} \tag{41}$$

$$\begin{aligned}
\left|\eta_{iKn}'\vartheta_{x_jKn}\right| &\le \left(\eta_{iKn}'\eta_{iKn}\right)^{1/2}\left(\vartheta_{x_jKn}'\vartheta_{x_jKn}\right)^{1/2} \\
&= o_p\left(\frac{K^{1/2}}{n^{\frac{1}{2}-\frac{1}{v}}}\right)\left(O_p\left(\frac{K^{5/2}}{n^2}\right) + o_p\left(\frac{K^{1/2}}{n^{\frac{1}{2}-\frac{1}{v}}}\right)\right) \\
&= o_p\left(\frac{K^3}{n^{\frac{5}{2}-\frac{1}{v}}} + \frac{K}{n^{1-\frac{2}{v}}}\right),
\end{aligned} \tag{42}$$

and, at most

$$\eta_{Kn}'\int_0^1\tilde{\varphi}_K(r)dB_x(r)' = o_p\left(\frac{K}{n^{\frac{1}{2}-\frac{1}{v}}}\right).$$

Thus,

$$\begin{aligned}
\left(\frac{1}{n}\frac{X'\Phi_K}{\sqrt{n}}\right)\left(\frac{\Phi_K'\Delta X}{\sqrt{n}}\right) &= \int_0^1 \Xi_{xK}\Lambda_K^{\frac{1}{2}}\tilde{\varphi}_K(r)dB_x(r)' + o_p\left(\frac{K}{n^{\frac{1}{2}-\frac{1}{v}}}\right) \\
&= O_p(1) + o_p\left(\frac{K}{n^{\frac{1}{2}-\frac{1}{v}}}\right),
\end{aligned} \tag{43}$$

since

$$\int_0^1 \Xi_{xK}\Lambda_K^{\frac{1}{2}}\tilde{\varphi}_K(r)dB_x(r)' = \int_0^1\sum_{k=1}^{K}\left\{\xi_{xk}\lambda_k^{1/2}\varphi_k(r)\right\}dB_x(r)' = O_p(1),$$

as $K\to\infty$ because $\sum_{k=1}^{\infty}\xi_{xk}\lambda_k^{1/2}\varphi_k(r) = B_x(r)$ is almost surely convergent. Similarly,

$$\frac{1}{n}\left(\frac{1}{n}\frac{X'\Phi_K}{\sqrt{n}}\right)G\left(\frac{\Phi_K'\Delta X}{\sqrt{n}}\right) = O_p\left(\frac{1}{n}\right) + o_p\left(\frac{K}{n^{\frac{3}{2}-\frac{1}{v}}}\right),$$

and, thus, at most

$$n^{-1}X'P_K\Delta X = O_p(1) + o_p\left(\frac{K}{n^{\frac{1}{2}-\frac{1}{v}}}\right), \tag{44}$$

from (37) and (43). Combining (44) and (36) in (35) we obtain

$$K^{-1}\left(n^{-1}X'P_K\Delta X\right)\left(K^{-1}\Delta X'P_K\Delta X\right)^{-1}\left(n^{-1}\Delta X'P_KX\right) = O_p\left(\frac{1}{K}\right) + o_p\left(\frac{K}{n^{1-\frac{2}{v}}}\right). \tag{45}$$

It follows from (34), (45) and **R** that

$$\frac{1}{n^2}X'R_KX \Rightarrow \int_0^1 B_xB_x'. \tag{46}$$

The next step in the proof is to consider the numerator in the matrix quotient (25), viz.,

$$n^{-1}U_0'R_KX = n^{-1}U_0'P_KX - \left(K^{-1}U_0'P_K\Delta X\right) \times \left(K^{-1}\Delta X'P_K\Delta X\right)^{-1}\left(n^{-1}\Delta X'P_KX\right).$$

From Lemma B, we have $K^{-1}U_0'P_K\Delta X \to_p \Omega_{0x}$ which, combined with (36), gives

$$\begin{aligned} n^{-1}U_0'R_KX &= n^{-1}U_0'P_KX - \left(\Omega_{0x}+o_p(1)\right)\left(\Omega_{xx}^{-1}+o_p(1)\right) \\ &\quad \times \left(n^{-1}\Delta X'P_KX\right) \qquad (47) \\ &= n^{-1}U_0'P_KX - \Omega_{0x}\Omega_{xx}^{-1}\left(n^{-1}\Delta X'P_KX\right)+o_p(1) \\ &= n^{-1}\left(U_0'-\Omega_{0x}\Omega_{xx}^{-1}\Delta X'\right)P_KX+o_p(1) \\ &= n^{-1}U_{0.x}'P_KX+o_p(1). \qquad (48) \end{aligned}$$

Next

$$\begin{aligned} \frac{1}{n}U_{0.x}'P_KX &= \left(\frac{U_{0.x}'\Phi_K}{\sqrt{n}}\right)\left(\frac{\Phi_K'\Phi_K}{n}\right)^{-1}\left(\frac{\Phi_K'X}{n^{3/2}}\right) \\ &= \left(\frac{U_{0.x}'\Phi_K}{\sqrt{n}}\right)\left(I_K+\frac{1}{n}G\right)^{-1}\left(\frac{\Phi_K'X}{n^{3/2}}\right) \\ &= \left(\frac{U_{0.x}'\Phi_K}{\sqrt{n}}\right)\left(\frac{\Phi_K'X}{n^{3/2}}\right) - \frac{1}{n}\left(\frac{U_{0.x}'\Phi_K}{\sqrt{n}}\right) \\ &\quad \times \left(G+O\left(\frac{1}{n}\right)\right)\left(\frac{\Phi_K'X}{n^{3/2}}\right) \\ &= \left(\frac{U_{0.x}'\Phi_K}{\sqrt{n}}\right)\left(\frac{\Phi_K'X}{n^{3/2}}\right)+O_p\left(\frac{K}{n}\right), \qquad (49) \end{aligned}$$

since, as shown below in (54), the elements of $\frac{U_{0.x}'\Phi_K}{\sqrt{n}}$ are $O_p(1)$ as are those of $\frac{\Phi_K'X}{n^{3/2}}$ from (28). Indeed, from (28), we have

$$\frac{\Phi_K'X}{n^{3/2}} = \frac{1}{n}\sum_{t=1}^n \varphi_{Kt}\frac{x_t'}{\sqrt{n}} = \Lambda_K^{\frac{1}{2}}\Xi_{xK}'+\eta_{Kn}, \qquad (50)$$

where the elements of η_{Kn} are uniformly $o_p\left(n^{-1/2+1/v}\right)$. Using partial summation, we have as in (38)

$$\begin{aligned} \frac{U_{0.x}'\Phi_K}{\sqrt{n}} &= \sum_{t=1}^n \frac{u_{0.xt}}{\sqrt{n}}\varphi_{Kt}' \\ &= \left(\frac{1}{\sqrt{n}}\sum_{t=1}^n u_{0.xt}\right)\tilde{\varphi}_K(1)' - \sum_{t=1}^n \frac{S_{0.xt}}{\sqrt{n}}\Delta\varphi_{Kt}', \qquad (51) \end{aligned}$$

and by virtue of the approximation (7), we have

$$\sup_{r\in[0,1]}\left\| n^{-1/2}\sum_{t=1}^{\lfloor nr\rfloor} u_{0.xt} - B_{0.x}(r)\right\| = o_p\left(n^{-\frac{1}{2}+\frac{1}{v}}\right), \qquad (52)$$

as $n \to \infty$. Thus, combining (51), (52) and (39), and using Lemma D, we obtain

$$\begin{aligned} &\left(\frac{1}{\sqrt{n}}\sum_{t=1}^n u_{0.xt}\right)\tilde{\varphi}_K(1)' - \sum_{t=1}^n \frac{S_{0.xt}}{\sqrt{n}}\Delta\varphi_{Kt}' \\ &= \left(B_{0.x}(1)+o_p\left(\frac{1}{n^{\frac{1}{2}-\frac{1}{v}}}\right)\right)\tilde{\varphi}_K(1)' - \sum_{t=1}^n \frac{S_{0.xt}}{\sqrt{n}}\Delta\varphi_{Kt}' \\ &= \left(B_{0.x}(1)+o_p\left(\frac{1}{n^{\frac{1}{2}-\frac{1}{v}}}\right)\right)\tilde{\varphi}_K(1)' - \int_0^1 B_{0.x}(r)\tilde{\varphi}_K^{(1)}(r)'dr \\ &\quad \times \left\{1+O\left(\frac{K^2}{n^2}\right)+o_p\left(\frac{1}{n^{\frac{1}{2}-\frac{1}{v}}}\right)\right\}+O_p\left(\frac{1}{\sqrt{n}}\right) \\ &= \int_0^1 dB_{0.x}(r)\tilde{\varphi}_K(r)'\left\{1+O\left(\frac{K^2}{n^2}\right)\right\}+o_p\left(\frac{1}{n^{\frac{1}{2}-\frac{1}{v}}}\right). \qquad (53) \end{aligned}$$

Thus, we may write

$$\frac{U_{0.x}'\Phi_K}{\sqrt{n}} = \int_0^1 dB_{0.x}(r)\tilde{\varphi}_K(r)'+\vartheta_{0.xKn}', \qquad (54)$$

where the elements of $\vartheta_{0.xKn}$ are uniformly $O\left(\frac{K^2}{n^2}\right)+o_p\left(n^{-\frac{1}{2}+\frac{1}{v}}\right)$ over $k=1,\ldots,K$. Then

$$\begin{aligned} &\left(\frac{U_{0.x}'\Phi_K}{\sqrt{n}}\right)\left(\frac{\Phi_K'X}{n^{3/2}}\right) \\ &= \left(\int_0^1 dB_{0.x}(r)\tilde{\varphi}_K(r)'+\vartheta_{0.xKn}'\right)\left(\Lambda_K^{\frac{1}{2}}\Xi_{xK}'+\eta_{Kn}\right) \\ &= \left(\int_0^1 dB_{0.x}(r)\tilde{\varphi}_K(r)'\right)\Lambda_K^{\frac{1}{2}}\Xi_{xK}'+\vartheta_{0.xKn}'\left(\Lambda_K^{\frac{1}{2}}\Xi_{xK}'+\eta_{Kn}\right) \\ &\quad +\int_0^1 dB_{0.x}(r)\tilde{\varphi}_K(r)'\eta_{Kn} \\ &= \left(\int_0^1 dB_{0.x}(r)\tilde{\varphi}_K(r)'\right)\Lambda_K^{\frac{1}{2}}\Xi_{xK}' \\ &\quad +O_p\left(\frac{K^{5/2}}{n^2}\right)+o_p\left(\frac{K^3}{n^{\frac{5}{2}-\frac{1}{v}}}\right)+o_p\left(\frac{K^{1/2}}{n^{\frac{1}{2}-\frac{1}{v}}}\right). \qquad (55) \end{aligned}$$

The error orders in (55) are justified as follows: first,

$$\begin{aligned} \vartheta_{0.xKn}'\left(\Lambda_K^{\frac{1}{2}}\Xi_{xK}'+\eta_{Kn}\right) &= O_p\left(\frac{K^{5/2}}{n^2}\right)+o_p\left(\frac{K^{1/2}}{n^{\frac{1}{2}-\frac{1}{v}}}\right) \\ &\quad +o_p\left(\frac{K^3}{n^{\frac{5}{2}-\frac{1}{v}}}+\frac{K}{n^{1-\frac{2}{v}}}\right), \end{aligned}$$

which is obtained as in (41) and (42); and, second, since $\int_0^1 dB_{0.x}(r)\varphi_K(r)'\eta_{Kn}$ has zero mean and conditional variance matrix $\Omega_{00.x}\otimes \eta_{Kn}'\eta_{Kn} = o_p\left(\frac{K}{n^{1-\frac{2}{v}}}\right)$ in view of (30), it follows that

$$\int_0^1 dB_{0.x}(r)\tilde{\varphi}_K(r)'\eta_{Kn} = o_p\left(\frac{K^{1/2}}{n^{\frac{1}{2}-\frac{1}{v}}}\right).$$

Thus, under the rate condition **R**, we have

$$\left(\frac{U_{0.x}'\Phi_K}{\sqrt{n}}\right)\left(\frac{\Phi_K'X}{n^{3/2}}\right) = \left(\int_0^1 dB_{0.x}(r)\tilde{\varphi}_K(r)'\right)\Lambda_K^{\frac{1}{2}}\Xi_{xK}'+o_p(1).$$

Next, observe that conditional on Ξ_{xK}, we have

$$\begin{aligned} &\left(\int_0^1 dB_{0.x}(r)\tilde{\varphi}_K(r)'\right)\Lambda_K^{\frac{1}{2}}\Xi_{xK}'\Big|_{\Xi_{xK}} \\ &\equiv N\left(0,\Omega_{00.x}\otimes\left(\Xi_{xK}\Lambda_K^{\frac{1}{2}}\Lambda_K^{\frac{1}{2}}\Xi_{xK}'\right)\right) \\ &\Rightarrow N\left(0,\Omega_{00.x}\otimes\int_0^1 B_xB_x'\right), \qquad (56) \end{aligned}$$

since $\Xi_{xK}\Lambda_K\Xi_{xK}' = \int_0^1 B_xB_x' - \sum_{k=K+1}^\infty \lambda_k\xi_{xk}\xi_{xk}' \to_p \int_0^1 B_xB_x'$ as $K\to\infty$ in view of (32) and (33). As $K\to\infty$, we therefore have the weak convergence

$$\left(\int_0^1 dB_{0.x}(r)\tilde{\varphi}_K(r)'\right)\Lambda_K^{\frac{1}{2}}\Xi_{xK}' \Rightarrow \int_0^1 dB_{0.x}(r)B_x(r)'.$$

Thus, as $n\to\infty$

$$\left(\frac{U_{0.x}'\Phi_K}{\sqrt{n}}\right)\left(\frac{\Phi_K'X}{n^{3/2}}\right) \Rightarrow \int_0^1 dB_{0.x}(r)B_x(r)'dr. \qquad (57)$$

Combining (48), (49), and (57) we have

$$n^{-1}U_0'R_KX \Rightarrow \int_0^1 dB_{0.x}(r)B_x(r)'dr. \qquad (58)$$

The stated limit result (a) now follows from (25), (46) and (58). Part (b) is shown in (46). To prove (c), it is sufficient to observe that as in (48)

$$\begin{aligned} &K^{-1}\hat{U}_{0.x}'P_K\hat{U}_{0.x}' \\ &= K^{-1}\left(Y' - A_{IV}X' + C_{IV}\Delta X'\right)P_K\left(Y - XA_{IV}' + \Delta XC_{IV}'\right) \\ &= K^{-1}U_0'R_KU_0 = K^{-1}U_{0.x}'P_KU_{0.x} + o_p(1) \to_p \Omega_{00.x}, \end{aligned}$$

since $K^{-1}U'P_KU \to_p \Omega$ from Lemma C. ■

A.2. An optimal AMSE expansion rate for K

To simplify the presentation, we consider the scalar case, with corresponding adjustments to notation so that (8) becomes $y_t = ax_t + \frac{\omega_{0x}}{\omega_{xx}}\Delta x_t + u_{0.xt}$. Our ultimate object is to expand the estimation error

$$n(a_{IV} - a) = \left(n^{-1}u_0'R_Kx\right)\left(n^{-2}x'R_Kx\right)^{-1} \tag{59}$$

in an asymptotic series. However, here we will be content to examine certain of its leading components. First, consider the numerator. Using (47) and Lemma B, we have

$$\begin{aligned} n^{-1}u_0'R_Kx = {} & n^{-1}u_0'P_Kx - \left(K^{-1}u_0'P_K\Delta x\right)\left(K^{-1}\Delta x'P_K\Delta x\right)^{-1} \\ & \times \left(n^{-1}\Delta x'P_Kx\right). \end{aligned} \tag{60}$$

Since $K^{-1}u_0'P_K\Delta x = K^{-1}u_0'P_Ku_x$ and $K^{-1}\Delta x'P_K\Delta x = K^{-1}u_x{}'P_Ku_x$ are elements of $\hat{\Omega}_K = K^{-1}U'P_KU$, we have the following expansion from Lemma C

$$\hat{\Omega}_K - \Omega = \frac{K^2}{n^2}D + \frac{1}{\sqrt{K}}E_K,$$

where $E_K \Rightarrow N(0, 2P_D(\Omega \otimes \Omega))$, from which we deduce that, in an obvious subscript notation,

$$K^{-1}u_0'P_K\Delta x = K^{-1}u_0'P_Ku_x = \omega_{0x} + \frac{K^2}{n^2}D_{0x} + \frac{1}{\sqrt{K}}E_{K,0x},$$

$$\begin{aligned} K^{-1}\Delta x'P_K\Delta x = K^{-1}u_x{}'P_Ku_x &= \omega_{xx} + \frac{K^2}{n^2}D_{xx} + \frac{1}{\sqrt{K}}E_{K,xx} \\ &= \omega_{xx}\left\{1 + \frac{K^2}{n^2}\frac{D_{xx}}{\omega_{xx}} + \frac{1}{\sqrt{K}}\frac{E_{K,xx}}{\omega_{xx}}\right\}. \end{aligned}$$

Define the long-run regression coefficient $\rho_{0.x} = \omega_{0x}/\omega_{xx}$, which appears as a coefficient in the augmented regression model (8). Observe that expression (60) involves the following implied estimate of $\rho_{0.x}$

$$\begin{aligned} \hat{\rho}_{0.x} &= \left(K^{-1}u_0'P_K\Delta x\right)\left(K^{-1}\Delta x'P_K\Delta x\right)^{-1} \\ &= \rho_{0.x} + \frac{1}{\omega_{xx}}\left(\frac{K^2}{n^2}D_{0x} + \frac{1}{\sqrt{K}}E_{K,0x}\right) \\ &\quad - \frac{\omega_{0x}}{\omega_{xx}}\left(\frac{K^2}{n^2}\frac{D_{xx}}{\omega_{xx}} + \frac{1}{\sqrt{K}}\frac{E_{K,xx}}{\omega_{xx}}\right) + O_p\left(\frac{K^2}{n^2} + \frac{1}{\sqrt{K}}\right)^2 \\ &= \rho_{0.x} + \frac{K^2}{n^2}\frac{1}{\omega_{xx}}\left(D_{0x} - \frac{\omega_{0x}}{\omega_{xx}}D_{xx}\right) \\ &\quad + \frac{1}{\sqrt{K}}\frac{1}{\omega_{xx}}\left(E_{K,0x} - \frac{\omega_{0x}}{\omega_{xx}}E_{K,xx}\right) + O_p\left(\frac{K^2}{n^2} + \frac{1}{\sqrt{K}}\right)^2, \end{aligned} \tag{61}$$

from which we may derive an AMSE optimal formula for the choice of K in estimating $\rho_{0.x}$. In particular, setting $a_\omega' = \left(1, -\frac{\omega_{0x}}{\omega_{xx}}\right)$, $e_2' = (0, 1)$, using row vectorization, and writing

$$E_{K,0x} - \frac{\omega_{0x}}{\omega_{xx}}E_{K,xx} = a_\omega'E_Ke_2 = \left(a_\omega' \otimes e_2'\right)\operatorname{vec}(E_K),$$

$$D_{0x} - \frac{\omega_{0x}}{\omega_{xx}}D_{xx} = a_\omega'De_2 := B,$$

$$V = 2\left(a_\omega' \otimes e_2'\right)P_D(\Omega \otimes \Omega)(a_\omega \otimes e_2),$$

we have from the leading term of (61)

$$\begin{aligned} E\left\{\hat{\rho}_{0.x} - \rho_{0.x}\right\}^2 &\sim E\left\{\frac{K^2}{n^2}\frac{a_\omega'De_2}{\omega_{xx}} + \frac{1}{\sqrt{K}}\frac{a_\omega'E_Ke_2}{\omega_{xx}}\right\}^2 \\ &= \left\{\frac{K^4}{n^4}B^2 + \frac{1}{K}V\right\}. \end{aligned}$$

Minimizing this expression with respect to K gives the following AMSE optimal rule

$$K = n^{4/5}\left[\frac{V}{4B^2}\right]^{1/5}, \tag{62}$$

which is analogous to the usual AMSE optimal rule in HAC estimation with quadratic kernels. ■

References

Akonom, J., 1993. Comportement asymptotique du temps d'occupation du processus des sommes partielles. Annals of the Institute of Henri Poincaré 29, 57–81.
Chao, J.C., Swanson, N.R., 2005. Consistent estimation with a large number of weak instruments. Econometrica 73, 1673–1692.
Galambos, J., 1978. The Asymptotic Theory of Extreme Order Statistics. Wiley, New York.
Han, C., Phillips, P.C.B., 2006. GMM with many moment conditions. Econometrica 74, 147–192.
Jeganathan, P., 1995. Some aspects of asymptotic theory with applications to time series models. Econometric Theory 11, 818–887.
Johansen, S., 1988. Statistical analysis of cointegration vectors. Journal of Economic Dynamics and Control 12, 231–254.
Johansen, S., 1995. Likelihood-based Inference in Cointegrated Vector Autoregressive Models. Oxford University Press, Oxford.
Künsch, H.R., 1987. Statistical aspects of self-similar processes. In: Prohorov, Y., Sazanov, V.V. (Eds.), Proceedings of the First World Congress of the Bernoulli Society, Vol. 1. VNU Science Press, Utrecht, pp. 67–74.
Liao, Z., 2013. Adaptive GMM shrinkage estimation with consistent moment selection. Econometric Theory 29, 1–48.
Liao, Z., Phillips, P.C.B., 2012. Automated Estimation of Vector Error Correction Models, Cowles Foundation Discussion Paper, #1873, Yale University.
Linton, O.B., 1995. Second order approximation in the partially linear regression model. Econometrica 63 (5), 1079–1112.
Magdalinos, A., Phillips, P.C.B., 2009. Econometric Inference in the Vicinity of Unity, unpublished working paper, Yale University.
Phillips, P.C.B., 1986. Understanding spurious regressions in econometrics. Journal of Econometrics 33, 311–340.
Phillips, P.C.B., 1989. Partially identified econometric models. Econometric Theory 5, 181–240.
Phillips, P.C.B., 1991a. Spectral regression for cointegrated time series. In: Barnett, W., Powell, J., Tauchen, G. (Eds.), Nonparametric and Semiparametric Methods in Economics and Statistics. Cambridge University Press, New York, pp. 413–435.
Phillips, P.C.B., 1991b. Optimal inference in cointegrated systems. Econometrica 59, 283–306.
Phillips, P.C.B., 1998. New tools for understanding spurious regressions. Econometrica 66, 1299–1326.
Phillips, P.C.B., 1999. Discrete Fourier transforms of fractional processes. Cowles Foundation Discussion Paper #1243, Yale University.
Phillips, P.C.B, 2002. New unit root asymptotics in the presence of deterministic trends. Journal of Econometrics 111, 323–353.
Phillips, P.C.B., 2005a. Challenges of trending time series econometrics. Mathematics and Computers in Simulation 68, 401–416.
Phillips, P.C.B., 2005b. HAC estimation by automated regression. Econometric Theory 21, 181–240.
Phillips, P.C.B., 2006. Optimal Estimation of Cointegrated Systems with Irrelevant Instruments, Cowles Foundation Working Paper #1547, Yale University.
Phillips, P.C.B., 2007. Unit root log periodogram regression. Journal of Econometrics 138, 104–124.
Phillips, P.C.B., Hansen, B.E., 1990. Statistical inference in instrumental variables regression with I(1) processes. Review of Economic Studies 57, 99–125.
Phillips, P.C.B., Loretan, M., 1991. Estimating long-run economic equilibria. Review of Economic Studies 59, 407–436.
Phillips, P.C.B., Solo, V., 1992. Asymptotics for linear processes. Annals of Statistics 20, 971–1001.
Protter, P., 1990. Stochastic Integration and Differential Equations: A New Approach. Springer-Verlag, New York.
Saikkonen, P., 1991. Asymptotically efficient estimation of cointegration regressions. Econometric Theory 7, 1–21.
Shimotsu, K., Phillips, P.C.B., 2005. Exact local Whittle estimation of fractional integration. Annals of Statistics 33, 1890–1933.

Staiger, D., Stock, J.H., 1997. Instrumental variables regression with weak instruments. Econometrica 65, 557–586.
Stock, J.H., Watson, M.W., 1993. A simple estimator of cointegrating vectors in higher order integrated systems. Econometrica 61, 783–820.
Xiao, Zhijie, Phillips, Peter C.B., 1998. Higher order approximations for frequency domain time series regression. Journal of Econometrics 86, 297–336.
Xiao, Zhijie, Phillips, Peter C.B., 1999. Higher order approximations for Wald statistics in time series regressions with integrated processes. Journal of Econometrics 108, 157–198.
Zaitsev, A.Y., 1998. Multidimensional version of the results of Komlós, Major and Tusnády for vectors with finite exponential moments. ESAIM: Probability and Statistics 2, 41–108.