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.

© 2013 Elsevier B.V. All rights reserved.

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 times 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.
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 (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 sieving 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

\[ y_t = A x_t + u_t \]

\[ \Delta y_t = \epsilon_t \]

relating the observable time series \( y_t \) and \( x_t \) with initial conditions at \( t = 0 \) and \( x_0 = 0 \). The composite error \( u_t = (u_{x_t}, u_{\epsilon_t})' \) is a weakly dependent time series satisfying

\[ u_t = C(L) \epsilon_t = \sum_{j=0}^{\infty} c_j \epsilon_{t-j}, \quad \sum_{j=0}^{\infty} |c_j| < \infty, \quad a > 3, \quad (L) \]

where \( \epsilon_t \) is iid \( (0, \Sigma) \) with \( \Sigma > 0 \) and \( E(\|\epsilon_t\|) < \infty \), for some \( v > 2 \) and matrix norm \( \| \cdot \| \). The long-run moving average coefficient matrix \( C(1) \) is assumed to be nonsingular, so that \( u_t \) 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 \( \Gamma_u(h) = E(\epsilon_t u_{t+h}) = \sum_{j=0}^{\infty} c_j c_j' \), finite \( v \)th absolute moment \( E(\|u_t\|^v) \leq \sum_{j=0}^{\infty} |c_j|^v E(\|\epsilon_t\|^v) < \infty \), spectrum \( f_u(\lambda) = (1/2\pi) C(\epsilon^i \lambda) C(\epsilon^{-i} \lambda) \) and long-run variance matrix \( \Omega = 2\pi f_u(0) = C(1) \Sigma_C(1) \), which is partitioned conformably with \( u_t \) as

\[ \Omega = \begin{pmatrix} \Omega_{00} & \Omega_{0x} \\ \Omega_{x0} & \Omega_{xx} \end{pmatrix} \]

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

\[ \lim_{h \to -\infty} h^3 \| \Gamma_u(h) \| < \infty, \quad (3) \]

so that \( f_u(\lambda) \) has continuous second derivative \( f''_u(\lambda) = -f'' u(\lambda) e^{-i\lambda h} \). While this framework assumes stationary \( u_t \) allows for some heterogeneity in \( \epsilon_t \) 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))

\[ B_u(\cdot) := \frac{S_{\lfloor \kappa \rfloor}}{\sqrt{n}} \sum_{i=1}^{\lfloor \kappa \rfloor} u_i \Rightarrow B(\cdot), \quad (4) \]

where \( \lfloor a \rfloor \) signifies the integer part of \( a \), \( \Rightarrow \) is weak convergence, and \( B(\cdot) \) is vector Brownian motion with variance matrix \( \Omega \).
partition \( B \) conformably with \( u_t \) by setting \( B = (B_0, B'_0) \) and define the Brownian motion \( B_{0x} = B_0 - \Omega_0 \Omega_0^{-1} B_x \), a Brownian motion with variance matrix \( \Omega_{0x} \), 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))

\[
B(r) = \sqrt{2} \sum_{k=1}^{\infty} \begin{pmatrix} \sin \left( \frac{(k-1/2) \pi r}{r} \right) \\ \sin \left( \frac{(k-1/2) \pi r}{r} \right) \end{pmatrix},
\]

where the components \( \varepsilon_k \) are iid \( N(0, \Omega) \), \( \lambda_k = 1/((k-1/2)^2) \), and \( \varphi_\nu(r) = \sqrt{2} \sin \left( \frac{(k-1/2) \pi r}{r} \right) \). 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

\[
B(r) = \mathbb{S}_k A_k \hat{\varphi}_k (r) + \mathbb{S}_L A_L \hat{\varphi}_L (r),
\]

where \( \mathbb{S}_k = \text{diag}(\lambda_1, \ldots, \lambda_k) \), \( A_L = \text{diag}(\lambda_{k+1}, \lambda_{k+2}, \ldots) \),

\[
\hat{\varphi}_k (r) = \begin{bmatrix} \varphi_1(r), \ldots, \varphi_k(r) \end{bmatrix}^\top, \quad \text{and} \quad \hat{\varphi}_L (r) = \begin{bmatrix} \varphi_{k+1}(r), \varphi_{k+2}(r), \ldots \end{bmatrix}^\top.
\]

Further we partition these matrices conformably with \( u_t \) as

\[
\mathbb{S}_k = \begin{bmatrix} \mathbb{S}_{0k} & \mathbb{S}_{1k} \\ \mathbb{S}_{0k} & \mathbb{S}_{1k} \end{bmatrix}, \quad \mathbb{S}_L = \begin{bmatrix} \mathbb{S}_{0k+1} & \mathbb{S}_{0k+2} & \ldots \\ \mathbb{S}_{0k+1} & \mathbb{S}_{0k+2} & \ldots \end{bmatrix}.
\]

Note that the coefficient of the deterministic function \( \varphi_k(r) \) in (5) is of order \( \Omega_q(\frac{1}{r^2}) \), 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 sum process \( B_n(\cdot) \) in terms of a Brownian motion \( B \) with variance matrix \( \Omega \).

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

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 \( L \) ensures that \( o_p \left( \frac{1}{n^{1/2}} \right) = o_p (1) \) in (7). The larger the moment exponent \( \nu \) 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

\[
y_t = \Delta X_t + \Omega_{0x} \Omega_{ax}^{-1} \Delta X_t + u_{0x},
\]

where \( u_{0x} = u_{0x} - \Omega_{0x} \Omega_{ax}^{-1} u_{ax} \), and write the equation in observation format as

\[
y' = Ax' + \Omega_{0x} \Omega_{ax}^{-1} \Delta X_t + U_{0x},
\]

where \( y = [y_1, \ldots, y_n] \) with similar definitions for \( \Delta X_t \), and \( U_{0x} \).

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 \( f(x) = \sum_{k=1}^{\infty} c_k \varphi_k(x) \), where \( c_k \) signifies equality in the \( L_2 \) sense. Our approach to estimation of (8) is to use as instrumental variables for both \( x_t \) and \( \Delta x_t \) a (potentially infinite) sequence of deterministic functions of the form \( \{\varphi_k(\frac{t}{n}) : k = 1, \ldots, K\} \). 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 \( \Delta x_t \) 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 \( \Delta x_t \).

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

\[
\varphi_k(r) = \sqrt{2} \sin \left( \frac{(k-1/2) \pi r}{r} \right),
\]

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 \( \varphi_0 = (\varphi_1(\frac{1}{n}), \ldots, \varphi_1(\frac{1}{n})) \), \( \varphi'_k = \{\varphi_1, \ldots, \varphi_k\} \), and \( P_{k+1} = \Phi_k \Phi_k'(\varphi_1(\frac{1}{n}), \ldots, \varphi_k(\frac{1}{n})) \) be the orthogonal projector to the space spanned by the columns of \( \varphi_k \). Assume the order condition \( K \geq 2m_h \) holds and apply instrumental variables linear regression to (8) using the matrix of instruments \( \varphi_k \). As indicated, the instruments are being used here for both the levels \( x_t \) and the differences \( \Delta x_t \). In the regression we can treat \( C = \Omega_{0x} \Omega_{ax}^{-1} \) simply as an unknown coefficient matrix.

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

\[
(A_W, C_W) = \arg \min_{A, C} \left( Y' - AX' + C \Delta X' \right) P_K \times \left( Y' - AX' + \Delta X' \right),
\]

Accordingly,

\[
A_W = \arg \min_{A} \left( Y' - AX' \right) R_K \left( Y' - AX' \right),
\]

where \( R_K = P_K \Delta X' P_K' \Delta X \), leading to the explicit partitioned regression formula

\[
A_W = \left( Y'R_K X' \right) \left( X' R_K X \right)^{-1}
\]

and the corresponding residual moment matrix

\[
\Omega_{0x} = K^{-1} \hat{U}'_0 X' P_K \hat{U}'_0 X = K^{-1} \left( Y' - A W X' + C_W \Delta X' \right) P_K \times \left( Y' - AX' + \Delta X' \right)
\]

from this regression, where \( \hat{U}'_0 X = Y' - A W X' + C_W \Delta X' \) 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 \( A_W \) 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 \( \Phi_k = \Phi_k(\varphi_1(\frac{1}{n}), \ldots, \varphi_k(\frac{1}{n})) \), where \( * \) denotes complex conjugate transpose, \( \Phi_k = \{\varphi_1, \ldots, \varphi_k\} \), and \( \hat{\varphi}_{kt} = \varphi_k(\frac{1}{n}) \), where the latter has complex sinusoidal components \( \varphi_k(\frac{1}{n}) = (2\pi n)^{-1/2} e^{2\pi i n k} \). Then, \( X' \Phi_k \) 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 $P_{K}$ 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 $K \rightarrow \infty$ the basis be complete. This may just as well be achieved with real polynomials as with complex sinusoidal 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 $x_t$ 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 $\Delta x_t$, 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 (Phills, 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 (1998a, 2002) that deterministic instruments like $x_t$ become more effective in modeling integrated regressors as $K \rightarrow \infty$. 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 $C = \Omega_{xx}^{-1}$, which is the long-run regression coefficient of $x_t$ on $\Delta x_t$. 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 $\Delta x_t$ and adjusting the conditional mean for the endogeneity of the regressor. In fact, as the main result below shows, as $K \rightarrow \infty$ and $n \rightarrow \infty$ 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 $\Omega_{xx}^{-1}$, and (as shown below) a consistent estimate of the long-run conditional error variance matrix $\Omega_{0x}$. 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 $\mathbf{X}$ is a Gaussian linear process of the general form $\mathbf{L}$ (c.f., Phillips (1991b) and Jegathan (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, $(\mathbf{X}' \mathbf{R}_x \mathbf{X})^{-1}$. In the present case, the long-run variance matrix of $I_{0x}$ is consistently estimated by the standardized residual moment matrix $\Phi_{nx}^{0x}$ as shown below.

**Theorem.** Under $L$ and the rate condition

$$\frac{1}{K} \left( 1 + \frac{K}{n (1 - \frac{1}{n}) (\frac{3}{4} \frac{1}{n})} + \frac{K^2}{n^{2}} \right) \rightarrow 0,$$

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

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

(b) $n^{-2} x' R_x X \Rightarrow \int_0^1 B_x B_x'.

(c) $\Omega_{IV}^{nx} \rightarrow \varphi \Omega_{0x}.$

**Remarks.** (a) Condition $R$ requires that $K \rightarrow \infty$ but at a rate that is slower than $n^{3/5}$ and the smaller of $n^{1 - \frac{1}{n}}$ and $n^{3/6 - 1/3}$. 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 $K = o(n^{3/5-\delta})$ for small $\delta > 0$.

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

$$\left( \frac{U_{nx}^{0x} \Phi_x}{\sqrt{n}} \right) \left( \frac{c_x (r)}{n^{3/2}} \right) \Rightarrow \int_0^1 dB_x (r) B_x (r)'$$

as $n, K \rightarrow \infty$. An important aspect of this result is that the limit processes $B_{0x}$ 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 $n^{-1} U_{nx}^{0x} P_x X$ to the stochastic integral $\int_0^1 dB_x (r) B_x (r)'$ does not occur, so that

$$\left( \frac{U_{nx}^{0x} \Phi_x}{\sqrt{n}} \right) \left( \frac{c_x (r)}{n^{3/2}} \right) \Rightarrow \int_0^1 dB_x (r) B_x (r)' \quad (13)$$

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

$$\left( \frac{U_{nx}^{0x} \Phi_x}{\sqrt{n}} \right) \left( \frac{c_x (r)}{n^{3/2}} \right) \Rightarrow \frac{1}{2} B_x (1)^2 = \int_0^1 B_x dB_x,$$

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., Proctor (1990)) rather than the Itô 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 $K = O \left( n^{3/5} \right)$. 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 $\Omega$. 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 $C = \Omega_{xx}^{-1} \Omega_{0x}$, which appears in the augmented regression model (8). Again, the optimal rate is $K = O \left( n^{3/5} \right)$, 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
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 $O(p^{4/5})$ 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 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

$$
X_{1t} = bX_{2t} + u_{1t}, \quad X_{2t} = X_{2t-1} + u_{2t}, \quad u_{t} = \left\{ \begin{array}{ll}
\varepsilon_{t} + \theta_{1}\varepsilon_{t-1} & \text{MA}(1) \\
\Theta u_{t-1} + \varepsilon_{t} & \text{VAR}(1)
\end{array} \right.,
$$

\begin{equation}
\varepsilon_{t} = \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \sim iidN \left(0, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right), \quad \Theta = \begin{bmatrix} \theta_{1} & 0 \\ 0 & \theta_{2} \end{bmatrix},
\end{equation}

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 ($\theta_{1}, \theta_{2}$). 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
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 $\theta_1$ and $\theta_2$ 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,
especially when $\rho$ 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 signs. 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 $u_{0,\text{xt}} = u_{0} - \Sigma_{\text{xt}}^{\text{xt}}\Sigma^{-1}u_{\text{xt}}$ is independent of $u_{\text{xt}}$ 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 $\Delta x_t$. 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).

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.
which is the HAC estimator developed in Phillips (2005a,b). So, the trend likelihood (17) is Gaussian because it makes use of the asymptotic normality of the transformed variables $\xi_K^v$. 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.

### Table III

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

<table>
<thead>
<tr>
<th>AR coeffs</th>
<th>Estimator</th>
<th>Coverage prob</th>
<th>Length</th>
<th>AR coeffs</th>
<th>Coverage prob</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(\theta_1, \theta_2)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$(0.8, 0.8)$</td>
<td>RRR$_t$</td>
<td>0.644</td>
<td>0.324</td>
<td>0.858</td>
<td>0.109</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DOLS$_d$</td>
<td>0.475</td>
<td>0.055</td>
<td>$(0.8, -0.8)$</td>
<td>0.999</td>
<td>0.311</td>
</tr>
<tr>
<td></td>
<td>Trend IV</td>
<td>0.633</td>
<td>0.062</td>
<td>0.930</td>
<td>0.117</td>
<td></td>
</tr>
<tr>
<td>$(0.4, 0.4)$</td>
<td>RRR$_t$</td>
<td>0.792</td>
<td>0.116</td>
<td>0.846</td>
<td>0.108</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DOLS$_d$</td>
<td>0.783</td>
<td>0.087</td>
<td>$(0.4, -0.4)$</td>
<td>0.984</td>
<td>0.179</td>
</tr>
<tr>
<td></td>
<td>Trend IV</td>
<td>0.860</td>
<td>0.092</td>
<td>0.927</td>
<td>0.113</td>
<td></td>
</tr>
<tr>
<td>$(-0.4, 0.4)$</td>
<td>Trend IV</td>
<td>0.798</td>
<td>0.231</td>
<td>0.847</td>
<td>0.120</td>
<td></td>
</tr>
</tbody>
</table>

### Table IV

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

<table>
<thead>
<tr>
<th>AR coeffs</th>
<th>Estimator</th>
<th>Coverage prob</th>
<th>Length</th>
<th>AR coeffs</th>
<th>Coverage prob</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(\theta_1, \theta_2)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$(0.8, 0.8)$</td>
<td>RRR$_t$</td>
<td>0.805</td>
<td>0.132</td>
<td>0.971</td>
<td>0.147</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DOLS$_d$</td>
<td>0.802</td>
<td>0.091</td>
<td>$(0.8, -0.8)$</td>
<td>0.998</td>
<td>0.428</td>
</tr>
<tr>
<td></td>
<td>Trend IV</td>
<td>0.899</td>
<td>0.102</td>
<td>0.988</td>
<td>0.218</td>
<td></td>
</tr>
<tr>
<td>$(0.4, 0.4)$</td>
<td>RRR$_t$</td>
<td>0.825</td>
<td>0.134</td>
<td>0.869</td>
<td>0.110</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DOLS$_d$</td>
<td>0.841</td>
<td>0.098</td>
<td>$(0.4, -0.4)$</td>
<td>0.994</td>
<td>0.201</td>
</tr>
<tr>
<td></td>
<td>Trend IV</td>
<td>0.902</td>
<td>0.103</td>
<td>0.955</td>
<td>0.129</td>
<td></td>
</tr>
<tr>
<td>$(-0.4, 0.4)$</td>
<td>Trend IV</td>
<td>0.841</td>
<td>0.254</td>
<td>0.800</td>
<td>1.050</td>
<td></td>
</tr>
</tbody>
</table>

### 6. Trend likelihood

Define the trend ($\psi$) transform of a multiple time series $\xi_t$ as

$$
\xi_K^v = \sum_{k=1}^n q_k \phi_k(\frac{r}{n})
$$

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

$$
K \log |\Omega_{0,xx}| + \text{tr} \left\{ \Omega_{0,xx}^{-1} (\xi_{0,xx}^v - \xi_{0,xx}^v \xi_{0,xx}^v \xi_{0,xx}^v) \right\}
$$

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

$$
\xi_K^v = A \xi_K^v + \Omega_{0,xx} \Omega_{xx}^{-1} \xi_{xx}^v + \xi_{0,xx}^v, 
\xi_{xx}^v = \xi_{xx}^v
$$

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 $C = \Omega_{xx}^{-1}$. Hence, the trend MLE estimator satisfies

$$
(\hat A, \hat C, \hat \Omega_{xx}) = \arg \min_{A,C,\Omega_{xx}} L(A, C, \Omega_{xx}),
$$

where

$$
L(A, C, \Omega_{xx}) = K \log |\Omega_{xx}| + \text{tr} \left\{ \Omega_{xx}^{-1} (\xi_{xx}^v - CET \xi_{xx}^v) \right\}
$$

Concentrating out $\Omega_{0,xx}$ leads directly to the IV estimator given in (10). Thus, the estimates $\hat A_{IV}, \hat C_{IV}$ and $\hat \Omega_{0,xx}$ 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 $\xi_K^v$. 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.
Alsoshowninthefigurearethecorrespondingnormaldensities
memoryparameterhasthefourvalues

d
from10,000replicationswhen

d
usesof
characteristicstoclosethoseoftheELWestimatesandweconjecturethat

closelyestimatedareconsistentandhasthelimitdistribution

regularityconditionstheexactLW(ELW)estimator

whichmaybeminimizedwithrespectto

ofthefrequencyband,and

where

usingthetrendinstruments(9).Thetrendlikelihoodturns

toaninterceptwhenthereisdeterministictrend

approachapplieswithoutmodificationwhenthecointegratingre-

gressioninvolvesaninterceptorwhenthereisdeterministictrend

approachapplieswithoutmodificationwhenthecointegratingre-

intervalsismoreaccurateandlengthisoftenshorter).

regressionfocusesattentiononthelong-runbehaviorofthesystem

variablesinbothlevelsanddifferences.Forthenonstationary

variablesinlevels,theregressionprovidesoptimalestimatesofthe

cointegratingcoefficients.Forthestationaryvariablesthatappear

differencesofthesystemvariables,theregressionproducesthe

long-runcovarianceandregressioncoefficientsthatcaptureand

adjustfortheeffectsofsimultaneityinthesystem.

Thus,usinginstrumentalvariablesfromanagnosticsetoftrend

basisfunctionscanbewviewedasasimpleregressiondevicefor

detectinglong-runeffectsinaneconometricmodel.Inthisrespect,

thedeviceworksinthe samewayasnarrow-bandfrequency

domaintechniquesthecentrateonlyonlowfrequencies.But

ithastheadvantagesofcompletelyavoidingthecomplexitiesof

thefrequencydomainandhavingaverysimpleinterpretationthat

shouldbeappealingtoappliedresearchers.Forpracticalpurposes,

theapproachisveryeasytoimplement,providesasymptotically

validstandarderrorsandtestsfromtheusualregressionoutput,

andrequiresonlybasiceconomicsoftwarepackages
toimplement.Finitesampleperformanceassesuperioretoexisting

proceduresintermsofestimation(wherethereisgenerallyless

bias)andedference(wherethecoverageprobabilityofconfidence

intervalsismoreaccurateandlengthisoftenshorter).

Whileitismentionedearlier,italsoshallberecalledthatthe

approachapplieswithoutmodificationwhenthecointegrating

regressionservesaninterceptorwhenthereisdeterministictrend

regression.Inbothcases,thetrendbasisinstrumentscontinu-

etogetherasymptoticallyefficientestimatesofthecointegrating

coefficientsandnootherinstrumentsarerequired.Themethods

mayalsobeextendedtosystemsdifferentnormalizations

thantheusualsystemusedin(1)–(2)andasuchmaybeeused

for testingthedimensionofthecointegratingspace.

Theinstrumentsconsideredinthispaperaredeterministic

functions.Wemightalsoconsidertheuseofacollectionofintegrated

seriesasinstruments,followingtheoriginalanalysisin

PhillsandHansend(1990).Largenumbersofsuchinstruments

arealsocapableofmodelingtrendregressorswithanR

approachesunity,asshowninPhills(1998),butareobviously

hardertojustifyinpracticalwork.Inconsequence,itispossible

thattheresultsgivenheremaybeextendedtoincludesuch

regressors.However,itisalsonecessarythattheso-calledinstruments

be capableofmodelingthelong-runregressioncoefficientsofstan-

aryseriesandthatremainstobeproved.

Theuseofagnosticdeterministicinstrumentsopensthe

interestingquestionofinstrumentselectioninpracticalwork.While

theKlrepresentation(5)givesanaturalsequentialorderingforthe

deterministicinstrumentsintermsoftheeigenvaluemagnitude,

we may also use data determined techniques to select instruments.

Inthisregard,modernshrinkagemethodssuchasthosesuggested

inLiao(2013)formomentconditionselectionseempromisingfor

useinthecontextwherethereisinstrumentselectionand

potentialrankreduction—seeLiaoandPhills(2012).Wehopetoe

exploretheseextensionsofthepresentmethodsinalter-

work.

Appendix. Lemmas and proofs

Lemma A (Phills (2007, Lemma 3.1)). If \( u_n \) satisfies \( L \), the

probability space which supports \( u_n \) can be expanded in such a way

that there exists a process distributionally equivalent to \( B_n (\cdot) = n^{-1/2} \sum_{i=1}^{[nt]} u_i \) and a Brownian motion \( B (\cdot) \) with variance matrix \( \Omega \) on the

new space for which

\[
\sup_{t \in [0, 1]} \| B_n (t) - B (t) \| = o_p \left( \frac{1}{n^{1/2 + \epsilon}} \right) \quad \text{as } n \to \infty.
\]
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, \( n^{-1} \sum_{i=1}^{n} \varphi_{kt} \varphi_{kt}' = I_k + O \left( \frac{1}{n} \right) \), and \( (n^{-1} \sum_{i=1}^{n} \varphi_{ki} \varphi_{ki}' )^{-1} = I_k + O \left( \frac{1}{n} \right) \), as \( n, K \to \infty \).

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

\[
Z_{n,k} = n^{-1} \sum_{i=1}^{n} B \left( \frac{1}{n} \right) \varphi_k \left( \frac{1}{n} \right) - \lambda_k^{1/2} \xi_k
\]

then, following a suggestion of a referee,

\[
Z_{n,k} = n^{-1} \sum_{i=1}^{n} B \left( \frac{1}{n} \right) \varphi_k \left( \frac{1}{n} \right) - \lambda_k^{1/2} \xi_k
\]

\[
= n^{-1} \sum_{i=1}^{n} \left[ \frac{1}{n} \right] \sum_{m=1}^{\infty} \frac{1}{\lambda_m} \varphi_m \left( \frac{1}{n} \right) \varphi_k \left( \frac{1}{n} \right) - \lambda_k^{1/2} \xi_k
\]

\[
= \sum_{m=1}^{\infty} \frac{1}{\lambda_m} \xi_m \left[ \frac{1}{n} \right] \sum_{i=1}^{n} \varphi_m \left( \frac{1}{n} \right) \varphi_k \left( \frac{1}{n} \right) - \delta_{mk}
\]

\[
= \frac{1}{n} \sum_{m=1}^{\infty} \left[ \frac{1}{\lambda_m} \xi_m \xi_m \right] = O_p (n^{-1}),
\]

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

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

\[
Z_{n,k} = \frac{1}{n} \sum_{m=1}^{\infty} \frac{1}{\lambda_m} \xi_m \xi_m + \frac{1}{n} \frac{1}{\lambda_k} \xi_k
\]

and

\[
\max_{1 \leq k \leq K} \| Z_{n,k} \| \leq \frac{1}{n} \sum_{m=1}^{\infty} \frac{1}{\lambda_m} \xi_m \xi_m + \frac{1}{n} \max_{1 \leq k \leq K} \| \xi_k \|
\]

\[
= O_p \left( \frac{(\log K)^{1/2}}{n} \right).
\]

since the maximum extreme value of \( K \) independent normal random variables is \( \tilde{O}_p \left( (\log K)^{1/2} \right) \) see Galambos (1978). Hence, \( Z_{n,k} \) is of order \( O_p \left( (\log K)^{1/2} / n \right) \) uniformly in \( k \leq K \).

A.1. Proof of the Theorem

Write

\[
n (A_u - A) = (n^{-1} U_p X) (n^{-2} X' P X) X^{-1}.
\]

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

\[
n^{-2} X' P X = n^{-2} X' P X - K^{-1} (n^{-1} X' P X) K^{-1} (\Delta X' P X) (\Delta X' P X) X^{-1}
\]

\[
\times (n^{-1} X' P X).
\]

Starting with the first term in (26) we have

\[
\frac{1}{n} \frac{1}{n} X' P X = \left( \begin{array}{cc} 1 & \Phi_K' \Phi_K \\ n & n \end{array} \right)^{-1} \left( \begin{array}{c} \Phi_K \\ n \end{array} \right) \left( \begin{array}{c} \Phi_K \\ n \end{array} \right)^{-1} \left( \begin{array}{c} \Phi_K \\ n \end{array} \right) \left( \begin{array}{c} \Phi_K \\ n \end{array} \right)^{-1} \left( \begin{array}{c} \Phi_K \\ n \end{array} \right).
\]

From the approximation (22) we can write

\[
\frac{1}{n} \frac{1}{n} X' P X = B_x \left( \frac{1}{n} \right) + O_p \left( \frac{1}{n} \right)
\]

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

\[
n^{-1} \sum_{i=1}^{n} B_x \left( \frac{1}{n} \right) \varphi ( \frac{1}{n} ) = \int_0^1 B_x ( t ) \varphi ( t ) dt
\]

\[
+ O_p \left( \frac{(\log K)^{1/2}}{n} \right).
\]
uniformly in \( k \leq K \). Then, as in the proof of Lemma 2.2 of Phillips (2002), the first factor of (27) is

\[
\frac{1}{n} \sum_{i=1}^{n} \frac{x_i}{n^{1/2}} = \frac{1}{n} \sum_{i=1}^{n} \frac{x_i}{n^{1/2}}. 
\]

Thus,

\[
\frac{1}{n} \sum_{i=1}^{n} \frac{x_i}{n^{1/2}} = \frac{1}{n} \sum_{i=1}^{n} \frac{x_i}{n^{1/2}}. 
\]

Next observe that

\[
\frac{1}{n} \sum_{i=1}^{n} \frac{x_i}{n^{1/2}} = \frac{1}{n} \sum_{i=1}^{n} \frac{x_i}{n^{1/2}}. 
\]

The error magnitude in (29) holds because, taking the \( t \)-th row, \( \eta_{t_k} \), of \( G_n \), we have

\[
\eta_{t_k} G_n \eta_{t_k} \leq \left( \eta_{t_k} G_n \eta_{t_k} \right)^{1/2} = \eta_{t_k} (\eta_{t_k} \eta_{t_k})^{1/2} = o_p \left( \frac{K}{n^{1/2}} \right). 
\]

which combine to give (29) and hence

\[
\frac{1}{n} \sum_{i=1}^{n} \frac{x_i}{n^{1/2}} = \frac{1}{n} \sum_{i=1}^{n} \frac{x_i}{n^{1/2}}. 
\]

Finally, observe that

\[
\frac{1}{n} \sum_{i=1}^{n} \frac{x_i}{n^{1/2}} = \frac{1}{n} \sum_{i=1}^{n} \frac{x_i}{n^{1/2}}. 
\]

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

\[
\int_{0}^{1} B_x B_x' = \int_{0}^{1} \left( \sum_{k=1}^{K} \lambda_k \xi_k \phi_k(r) \xi_k' \right) \left( \sum_{k=1}^{K} \lambda_k \phi_k(r) \xi_k' \right) dr 
\]

\[
= \sum_{k=1}^{K} \lambda_k \xi_k \phi_k(r) \xi_k'. 
\]
The limit form of $n^{-1/2}X\phi_K$ is given above in (28), so we concentrate on the second factor, $n^{-1/2}\phi'_K\Delta X = n^{-1/2}\phi'_K U$. Using partial summation and setting $S_{\delta t} = \sum_{t=1}^{\delta t} U_{\delta t}$, we get

$$U'_t\phi_K \sqrt{n} = \sum_{t=1}^{\delta t} \frac{U_{\delta t}}{\sqrt{n}} \psi'_{kt} = \frac{1}{\sqrt{n}} S_{\delta t} \bar{\psi}_k(1)' - \sum_{t=1}^{\delta t-1} \frac{S_{\delta t-1}}{\sqrt{n}} \Delta \psi'_{kt}.$$  \hfill (38)

Note that

$$\Delta \psi_k \left( \frac{t}{n} \right) = \sqrt{2} \sin \left( \left( k - \frac{1}{2} \right) \frac{\pi}{n} \right) - \sin \left( \left( k - \frac{1}{2} \right) \frac{\pi}{n} - \frac{1}{n} \right),$$

$$= \sqrt{2} \sin \left( \frac{1}{2} \left( k - \frac{1}{2} \right) \frac{\pi}{n} \right) \frac{1}{n} \cos \left( \left( k - \frac{1}{2} \right) \frac{\pi}{n} - \frac{1}{n} \right),$$

$$= \frac{1}{2} \left( k - \frac{1}{2} \right) \frac{1}{n} \cos \left( \left( k - \frac{1}{2} \right) \frac{\pi}{n} - \frac{1}{n} \right),$$

$$= \phi_k^{(1)} \left( \frac{t - \frac{1}{2}}{n} \right) \frac{1}{n} + O \left( \frac{K^2}{n^2} \right).$$  \hfill (39)

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

$$\sup_{r \in [0,1]} \left| n^{-1/2} \sum_{t=1}^{n r} U_{\delta t} - B_s(r) \right| = o_p \left( n^{-\frac{1}{4} + \frac{1}{8}} \right).$$

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

$$\frac{1}{\sqrt{n}} S_{\delta t} \bar{\psi}_k(1)' = - \sum_{t=1}^{\delta t-1} \frac{S_{\delta t-1}}{\sqrt{n}} \Delta \psi'_{kt}$$

$$- \left( B_s(1) + o_p \left( \frac{1}{n^{1/2-\tau}} \right) \right) \bar{\psi}_k(1)' - \sum_{t=1}^{\delta t-1} \frac{S_{\delta t-1}}{\sqrt{n}} \Delta \psi'_{kt}$$

$$- \int_0^{1} B_s(1) \bar{\psi}_k(r)' dr \left( 1 + O \left( \frac{K^2}{n^2} \right) \right) + o_p \left( \frac{1}{\sqrt{n}} \right),$$

$$= \int_0^{1} dB_s(r) \bar{\psi}_k(r)' dr \left( 1 + O \left( \frac{K^2}{n^2} \right) \right) + o_p \left( \frac{1}{n^{1/2-\tau}} \right).$$

Thus, we may write

$$U'_t \phi_K \frac{1}{\sqrt{n}} = \int_0^{1} dB_s(r) \bar{\psi}_k(r)' dr + \theta'_{s,Kn},$$  \hfill (40)

where the elements of $\theta'_{s,Kn}$ are uniformly $O \left( \frac{K^2}{n^2} \right) + o_p \left( n^{-\frac{1}{4} + \frac{1}{8}} \right)$ over $k = 1, \ldots, K$.

Combining (28) and (40) we have

$$\left( \frac{1}{n} X' \phi_K \right) \left( \frac{\phi'_K \Delta X}{\sqrt{n}} \right) = \left( \sum_{s=1}^1 A_{sK}^2 + \eta'_s \right) \left( \int_0^{1} \bar{\psi}_k(r) dB_s(r) \right)' dr + \theta'_{s,Kn},$$

$$= \int_0^{1} \sum_{s=1}^1 A_{sK}^2 \bar{\psi}_k(r) dB_s(r) \right)' + \sum_{s=1}^1 A_{sK}^2 \eta'_s \theta_{s,Kn} + \eta'_s \theta_{s,Kn}$$

$$+ \eta'_s \int_0^{1} \bar{\psi}_k(r) dB_s(r) \right)' .$$

In view of (32) and the order of the elements of $\theta'_{s,Kn}$ and $\eta'_{s,Kn}$, and denoting the $j$th row of $\theta'_{s,Kn}$ by $\theta'_{s,Kn}^j$, we have

$$\left| \sum_{s=1}^1 A_{sK}^2 \bar{\psi}_k(r) dB_s(r) \right)' \leq \left( \eta'_s + \eta_1 \right) \bar{\psi}_k(r) dB_s(r)' + \sum_{s=1}^1 A_{sK}^2 \eta'_s \theta_{s,Kn} + \eta'_s \theta_{s,Kn}$$

$$= \int_0^{1} \sum_{s=1}^1 A_{sK}^2 \bar{\psi}_k(r) dB_s(r) \right)' + \sum_{s=1}^1 A_{sK}^2 \eta'_s \theta_{s,Kn} + \eta'_s \theta_{s,Kn}$$

and, at most

$$\int_0^{1} \bar{\psi}_k(r) dB_s(r)' = o_p \left( \frac{1}{n^{1/2-\tau}} \right).$$

Thus,

$$\left( \frac{1}{n} X' \phi_K \right) \left( \frac{\phi'_K \Delta X}{\sqrt{n}} \right) = \int_0^{1} \sum_{s=1}^1 A_{sK}^2 \bar{\psi}_k(r) dB_s(r) \right)' + \sum_{s=1}^1 A_{sK}^2 \eta'_s \theta_{s,Kn} + \eta'_s \theta_{s,Kn}$$

$$= \int_0^{1} \sum_{s=1}^1 \xi_{sK} \phi_k^{(1)}(r) dB_s(r)' \right) + \sum_{s=1}^1 A_{sK}^2 \eta'_s \theta_{s,Kn} + \eta'_s \theta_{s,Kn}$$

$$= \int_0^{1} A_{sK}^2 \phi_k^{(1)}(r) dB_s(r)' \right) + \sum_{s=1}^1 A_{sK}^2 \eta'_s \theta_{s,Kn} + \eta'_s \theta_{s,Kn}$$

since

$$\int_0^{1} \sum_{s=1}^1 \xi_{sK} \phi_k^{(1)}(r) dB_s(r)' \right) = \int_0^1 \sum_{s=1}^1 \xi_{sK} \phi_k^{(1)}(r) dB_s(r)' \right)$$

$$= \int_0^1 \sum_{s=1}^1 \xi_{sK} \phi_k^{(1)}(r) dB_s(r)' \right)$$

as $K \to \infty$ because $\sum_{s=1}^1 \xi_{sK} \phi_k^{(1)}(r) = B_s(r)$ is almost surely convergent. Similarly,

$$\int_0^{1} \frac{1}{n} \left( \frac{1}{n} \right)^{1/2} \left( \frac{K^2}{n^2} \right) \Delta \psi'_{kt} dB_s(r)' \right) = \int_0^1 \left( \frac{1}{n} \right)^{1/2} \left( \frac{K^2}{n^2} \right) \Delta \psi'_{kt} dB_s(r)' \right)$$

$$= \int_0^1 \left( \frac{1}{n} \right)^{1/2} \left( \frac{K^2}{n^2} \right) \Delta \psi'_{kt} dB_s(r)' \right)$$

and, thus, at most

$$n^{-1} X' X = O_p \left( \frac{1}{n^{1/2-\tau}} \right).$$  \hfill (44)

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

$$K^{-1} \left( n^{-1/2} X' X \right) \left( K^{-1} \Delta X' X \right)^{-1} \left( n^{-1/2} X' X \right)$$

$$= O_p \left( \frac{1}{K} \right) + o_p \left( \frac{1}{n^{1/2-\tau}} \right).$$  \hfill (45)

It follows from (34), (45) and $\mathbf{R}$ that

$$\frac{1}{n} X' R X \to \int_0^1 B_s(r)' \right).$$  \hfill (46)
The next step in the proof is to consider the numerator in the matrix quotient (25), viz.,
\[ n^{-1}U'_0R_KX = n^{-1}U'_0P_KX - (K^{-1}U'_0P_K \Delta X) \times (K^{-1} \Delta X'P_KX)^{-1}(n^{-1} \Delta X'P_KX). \]
From Lemma B, we have \( K^{-1}U'_0P_K \Delta X \to_p \Omega_{0K} \) which, combined with (36), gives
\[ n^{-1}U'_0R_KX = n^{-1}U'_0P_KX - (\Omega_{0K} + o_p(1)) \left( (n^{-1}O_{0K} + o_p(1)) \times (n^{-1}O_{0K} + o_p(1)) \right) \]
\[ = n^{-1}U'_0P_KX - (\Omega_{0K}O_{0K}^{-1} + o_p(1))P_KX + o_p(1) \]
\[ = n^{-1}U'_0P_KX + o_p(1). \]
Next we consider
\[ \frac{1}{n}U'_0P_KX = \left( \frac{U'_0P_KX}{\sqrt{n}} \right) \times \left( \frac{\Phi'_KX}{\sqrt{n/2}} \right)^{-1} \]
\[ = \frac{1}{\sqrt{n}} \left( I_K + \frac{1}{n} \right) \frac{\Phi'_KX}{\sqrt{n/2}} \]
\[ = \frac{1}{\sqrt{n}} \left( \frac{U'_0P_KX}{n/2} \right) \times \left( \frac{\Phi'_KX}{\sqrt{n/2}} \right) \times \left( \frac{G + o_p(1)}{n} \right) \frac{\Phi'_KX}{\sqrt{n/2}} + o_p(\frac{K}{n}). \]
(49)

since, as shown below in (54), the elements of \( \frac{U'_0P_KX}{\sqrt{n}} \) are \( O_p(1) \) as are those of \( \frac{\Phi'_KX}{\sqrt{n/2}} \) from (28). Indeed, from (28), we have
\[ \frac{\Phi'_KX}{\sqrt{n/2}} = \frac{1}{n} \sum_{i=1}^{n} \eta_i x_i \frac{1}{\sqrt{n}} = \frac{1}{\sqrt{n}} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\Lambda_{ii}}{x_i}. \]
(50)

where the elements of \( \eta_{0K} \) are uniformly \( O_p(n^{-1/2+1/2}). \) Using partial summation, we have as in (38)
\[ U'_0P_KX \frac{\sqrt{n}}{\sqrt{n}} \sum_{i=1}^{n} \frac{U_{0i}x_i}{\sqrt{n}} \frac{\phi_{0i}x_i}{\sqrt{n}} \]
\[ = \left( \frac{1}{\sqrt{n}} \right) \left( \sum_{i=1}^{n} \frac{U_{0i}x_i}{\sqrt{n}} \right) \phi_{0i}(1) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} S_{0i}x_i \Delta \phi_{0i}. \]
(51)

and by virtue of the approximation (7), we have
\[ \sup_{1/2 \leq |t| \leq 0} \left\| n^{-1/2} \sum_{i=1}^{n} \frac{U_{0i}x_i}{\sqrt{n}} - B_0(x) \right\| \leq o_p \left( n^{-1/2+1/2} \right). \]
(52)
as \( n \to \infty. \) Thus, combining (51), (52) and (39), and using Lemma D, we obtain
\[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{U_{0i}x_i}{\sqrt{n}} \phi_{0i}(1) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} S_{0i}x_i \Delta \phi_{0i} \]
\[ = \left( B_0(x) + o_p \left( \frac{1}{n^{1/2-1/2}} \right) \right) \phi_{0i}(1) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} S_{0i}x_i \Delta \phi_{0i} \]
\[ = \left( B_0(x) + o_p \left( \frac{1}{n^{1/2-1/2}} \right) \right) \phi_{0i}(1) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} B_0(x) \phi_{0i}(1) \phi_{0i}(1) \frac{dx}{dr} \]
\[ \times \left\{ 1 + O \left( \frac{K^2}{n^2} \right) + o_p \left( \frac{1}{n^{2/2-1/2}} \right) \right\} + o_p \left( \frac{1}{n^{1/2-1/2}} \right). \]
(53)

Thus, we may write
\[ \frac{U'_0P_KX}{\sqrt{n}} = \int_0^1 dB_0(x) \phi_{0i}(1) \phi_{0i}(1) dr. \]
(54)

where the elements of \( \theta_{0i,0j} \) are uniformly \( O \left( \frac{K^2}{n^2} \right) + o_p \left( n^{-1/2+1/2} \right) \) over \( k = 1, \ldots, K. \) Then
\[ \left( \frac{U'_0P_KX}{\sqrt{n}} \right) \left( \frac{\Phi'_KX}{\sqrt{n/2}} \right) \]
\[ = \left( \int_0^1 dB_0(x) \left( \phi_{0i}(1) \phi_{0i}(1) \right) \right) \left( \frac{K^2}{n^2} \right) + o_p \left( n^{-1/2+1/2} \right). \]
(55)
The error orders in (55) are justified as follows: first,
\[ \frac{1}{\sqrt{2n}} \left( \frac{K^2}{n^2} \right) + o_p \left( \frac{K^2}{n^2} \right) \]
\[ + o_p \left( n^{-1/2+1/2} \right). \]
which is obtained as in (41) and (42); and, second, since \( \int_0^1 dB_0(x) \phi_{0i}(1) \phi_{0i}(1) \]
\[ \Rightarrow \int_0^1 dB_0(x) \phi_{0i}(1) \phi_{0i}(1) \]
\[ = o_p \left( \frac{K^2}{n^2} \right) + o_p \left( n^{-1/2+1/2} \right). \]

Thus, under the rate condition \( R \), we have
\[ \left( \frac{U'_0P_KX}{\sqrt{n}} \right) \left( \frac{\Phi'_KX}{\sqrt{n/2}} \right) \]
\[ = \left( \int_0^1 dB_0(x) \phi_{0i}(1) \phi_{0i}(1) \right) \frac{A_{iK}^2}{\mathcal{E}^{iK} + \eta_{0K} \mathcal{E}^{iK} + \eta_{0K}} \]
\[ + o_p \left( \frac{K^2}{n^2} \right) + o_p \left( \frac{K^2}{n^2} \right) \]
\[ + o_p \left( n^{-1/2+1/2} \right). \]

Thus, the rate condition \( R \) is satisfied.
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)
\[
K^{-1} \hat{u}_x P_x \hat{u}_x = K^{-1} (Y - A x + \Delta X P_x - Y - \chi A \hat{u}_x + \Delta X \hat{u}_x) = K^{-1} \hat{u}_x P_x \hat{u}_x + o_p(1) \to_p \Omega_{00, x},
\]
since \(K^{-1}U' P_x U \to_p \Omega \) 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 \(y_t = \alpha x_t + \varepsilon_t\). Our ultimate object is to expand the estimation error
\[
n(\hat{\alpha}_t - \alpha) = (n^{-1/2} u'_t R_x' x) (n^{-1} x' R_x x)^{-1}
\]
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
\[
n^{-1/2} u'_t R_x' x = n^{-1/2} u'_t P_x x - (K^{-1} u'_t P_x \Delta x) (K^{-1} \Delta x P_x \Delta x)^{-1} \times (n^{-1} \Delta x P_x x).
\]
Since \(K^{-1} u'_t P_x \Delta x = K^{-1} u'_t P_x u_x\) and \(K^{-1} \Delta x P_x \Delta x = K^{-1} u'_t P_x u_x\) are elements of \(\hat{\Omega}_K = K^{-1} U' P_x U\), we have the following expansion from Lemma C
\[
\hat{\Omega}_K = \Omega + K^{-1} D + \frac{1}{\sqrt{K}} E_K,
\]
where \(E_K \Rightarrow \mathcal{N}(0, 2P_0 (\Omega \otimes \Omega))\), from which we deduce that in an obvious subcript notation,
\[
K^{-1} u'_t P_x \Delta x = K^{-1} u'_t P_x u_x = \omega_{0x} + \frac{K}{n^2} D_{0x} + \frac{1}{\sqrt{K}} E_{K, 0x},
\]
\[
K^{-1} \Delta x P_x \Delta x = K^{-1} u'_t P_x u_x = \omega_{xx} + \frac{K}{n^2} D_{xx} + \frac{1}{\sqrt{K}} E_{K, xx},
\]
\[
= \omega_{xx} \left\{ 1 + \frac{K}{n^2} \frac{D_{xx}}{\omega_{xx}} + \frac{1}{\sqrt{K}} \frac{E_{K, xx}}{\omega_{xx}} \right\}.
\]
Define the long-run regression coefficient \(\rho_{0x} = \omega_{0x}/\omega_{xx}\), which appears as a coefficient in the augmented regression model (8). Observe that expression (60) involves the following implied estimate of \(\rho_{0x}\)
\[
\rho_{0x} = \left( K^{-1} u'_t P_x \Delta x \right) \left( K^{-1} \Delta x P_x \Delta x \right)^{-1} = \rho_{0x} + \frac{1}{\omega_{xx}} \left( \frac{K^2}{n^2} D_{xx} + \frac{1}{\sqrt{K}} E_{K, xx} \right) \frac{1}{\omega_{xx}} + \frac{1}{\sqrt{K}} E_{K, xx} + O_p \left( \frac{K^2}{n^2} + \frac{1}{\sqrt{K}} \right)^2,
\]
from which we may derive an AMSE optimal formula for the choice of \(K\) in estimating \(\rho_{0x}\). In particular, setting \(\alpha'_x = 1 - \frac{\omega_{0x}}{\omega_{xx}}\),
\[
\alpha'_x = (0, 1),
\]
using row vectorization, and writing
\[
E_{K, 0x} = \omega_{0x} \frac{D_{0x}}{\omega_{xx}} \frac{E_{K, 0x}}{E_{K, xx}} = \alpha'_x \frac{E_{K, 0x}}{E_{K, xx}} = (\alpha'_x \otimes e_2) \text{ vec}(E_K),
\]
\[
D_{0x} = \omega_{0x} \frac{D_{0x}}{\omega_{xx}} \frac{E_{K, 0x}}{E_{K, xx}} = \alpha'_x \frac{D_{0x}}{\omega_{xx}} \alpha'_x \frac{D_{0x}}{\omega_{xx}} + \frac{1}{\sqrt{K}} \frac{E_{K, 0x}}{E_{K, xx}} + O_p \left( \frac{K^2}{n^2} + \frac{1}{\sqrt{K}} \right)^2,
\]
\[
V = 2 (\alpha'_x \otimes e_2) P_0 (\Omega \otimes \Omega) (\alpha'_x \otimes e_2),
\]
we have from the leading term of (61)
\[
E \left[ \rho_{0x} - \rho_{0x} \right]^2 \sim E \left\{ \frac{K^2}{n^2} \frac{D_{0x}}{\omega_{xx}} + \frac{1}{\sqrt{K}} \frac{E_{K, 0x}}{E_{K, xx}} \right\}^2 = \left\{ \frac{K^4}{n^4} \frac{D_{0x}}{\omega_{xx}}^2 + \frac{1}{\sqrt{K}} \frac{E_{K, 0x}}{E_{K, xx}} \right\}^2.
\]
Minimizing this expression with respect to \(K\) gives the following AMSE optimal rule
\[
K = n^{4/5} \left( \frac{V}{4B^2} \right)^{1/5},
\]
which is analogous to the usual AMSE optimal rule in HAC estimation with quadratic kernels.

References


