Journal of Econometrics 178 (2014) 741–760

Contents lists available at ScienceDirect

Journal of Econometrics

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

Integrated modified OLS estimation and fixed-*b* inference for cointegrating regressions

Timothy J. Vogelsang[a,*], Martin Wagner[b,c,d]

[a] Department of Economics, Michigan State University, United States
[b] Faculty of Statistics, Technical University Dortmund, Germany
[c] Institute for Advanced Studies, Vienna, Austria
[d] Frisch Centre for Economic Research, Oslo, Norway

ARTICLE INFO

Article history:
Received 19 April 2011
Received in revised form
23 October 2013
Accepted 29 October 2013
Available online 12 November 2013

JEL classification:
C12
C13
C32

Keywords:
Bandwidth
Cointegration
Fixed-*b* asymptotics
FM-OLS
IM-OLS
Kernel

ABSTRACT

This paper is concerned with parameter estimation and inference in a cointegrating regression, where as usual endogenous regressors as well as serially correlated errors are considered. We propose a simple, new estimation method based on an augmented partial sum (integration) transformation of the regression model. The new estimator is labeled integrated modified ordinary least squares (IM-OLS). IM-OLS is similar in spirit to the fully modified OLS approach of Phillips and Hansen (1990) and also bears similarities to the dynamic OLS approach of Phillips and Loretan (1991), Saikkonen (1991) and Stock and Watson (1993), with the key difference that IM-OLS does not require estimation of long run variance matrices and avoids the need to choose tuning parameters (kernels, bandwidths, lags). Inference does require that a long run variance be scaled out, and we propose traditional and fixed-*b* methods for obtaining critical values for test statistics. The properties of IM-OLS are analyzed using asymptotic theory and finite sample simulations. IM-OLS performs well relative to other approaches in the literature.

1. Introduction

Cointegration methods are widely used in empirical macroeconomics and empirical finance. It is well known that in a cointegrating regression the ordinary least squares (OLS) estimator of the parameters is super-consistent, i.e. converges at rate equal to the sample size T. When the regressors are endogenous, the limiting distribution of the OLS estimator is contaminated by so-called second order bias terms, see e.g. Phillips and Hansen (1990). The presence of these bias terms renders inference difficult. Consequently, several modifications to OLS that lead to zero mean Gaussian mixture limiting distributions have been proposed, which in turn make standard asymptotic inference feasible. These methods include the fully modified OLS (FM-OLS) approach of Phillips and Hansen (1990) and the dynamic OLS (DOLS) approach of Phillips and Loretan (1991); Saikkonen (1991) and Stock and Watson (1993).

The FM-OLS approach uses a two-part transformation to remove the asymptotic bias terms and requires the estimation of long run variance matrices (as discussed in detail in Section 2). The DOLS approach augments the cointegrating regression by leads and lags of the first differences of the regressors to correct for the (second-order) endogeneity bias. Both of these methods require tuning parameter choices. For FM-OLS a kernel function and a bandwidth have to be chosen for long run variance estimation. For DOLS the number of leads and lags has to be chosen and if the DOLS estimates are to be used for inference, a long run variance estimator, with an ensuing choice of kernel and bandwidth, is also required.

Standard asymptotic theory does not capture the impact of kernel and bandwidth choices on the sampling distributions of estimators and test statistics based upon them. In order to shed light on the impact of kernel and bandwidth choice on the FM-OLS estimator, the first result of the paper derives the so-called fixed-*b* limit of the FM-OLS estimator. Fixed-*b* asymptotic theory has been put forward by Kiefer and Vogelsang (2005) in the context of stationary regressions to capture the impact of kernel and bandwidth choices on the sampling distributions of HAC-type test statistics. The benefit of this approach is that critical values that reflect kernel and bandwidth choices are provided. The fixed-*b* limiting distribution of the FM-OLS estimator features highly complicated dependence upon nuisance parameters and does not lend itself towards

* Corresponding author. Tel.: +1 517 353 4582.
E-mail address: tjv@msu.edu (T.J. Vogelsang).

0304-4076/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jeconom.2013.10.015

the development of fixed-b inference. We also show that the limit for $b \to 0$ of the fixed-b limiting distribution of the FM-OLS estimator is the usual asymptotic distribution derived for the FM-OLS estimator in Phillips and Hansen (1990). In deriving the fixed-b limit of the FM-OLS estimator we derive the fixed-b limit of the half long run variance matrix, which may be of interest in itself because such a result is not available in the literature up to now.

After this detailed consideration of the FM-OLS estimator, the paper proceeds to propose a simple, tuning parameter free new estimator of the parameters of a cointegrating regression. This estimator leads to a zero mean Gaussian mixture limiting distribution and implementation does not require the choice of any tuning parameters. The estimator is based on OLS estimation of a partial sum transformation of the cointegrating regression which is augmented by the original regressors, hence the name integrated modified OLS (IM-OLS) estimator. Inference based on this estimator still requires the estimation of a long run variance parameter. In this respect we offer two solutions. First, standard asymptotic inference based on a consistent estimator of the long run variance and second, fixed-b inference. We show that the conditional asymptotic variance of FM-OLS is smaller or equal to the conditional asymptotic variance of the IM-OLS estimator when resorting to standard asymptotic theory. However, when fixed-b asymptotic theory is invoked in order to capture the effects of kernel and bandwidth choices it turns out that the fixed-b limiting distribution of the FM-OLS estimator involves a bias term, whereas IM-OLS is asymptotically bias free. Furthermore, the fixed-b conditional asymptotic variance of the FM-OLS estimator is much more complex than the traditional one for which the above result concerning relative efficiency holds.

There are two other papers in the literature that develop ‘partial’ fixed-b theory for inference in cointegrating regressions. Bunzel (2006) analyzes tests based on the DOLS estimator and derives a fixed-b limit for a long run variance estimator constructed using the DOLS residuals. This fixed-b limit captures the choice of kernel and bandwidth but ignores the impact of lead and lag length choices required to implement DOLS. Jin et al. (2006) develop a partial fixed-b theory for tests based on FM-OLS. For the long run variance estimator needed to carry out the FM-OLS transformation, they appeal to a consistency result which ignores the impact of the kernel and bandwidth choice on the FM-OLS estimator. Conditional on this traditional consistency result, they derive a fixed-b limit for a second long run variance estimator that can be used to construct tests.

Developing useful fixed-b results for tests based on IM-OLS leads to some new hurdles compared to both the partial fixed-b tests developed for the FM-OLS and DOLS estimators tests and to stationary regressions. Specifically, the OLS residuals of the IM-OLS regression need to be further adjusted, as discussed in detail in Sections 3 and 5, in order to obtain pivotal fixed-b test statistics. A similar complication also arises in Vogelsang and Wagner (2013a), who consider fixed-b inference for Phillips and Perron (1988) type unit root tests where the original OLS residuals also cannot be used for fixed-b inference. Thus, unit root and cointegration analysis necessitates different thinking about fixed-b inference compared to stationary regression settings.

The theoretical analysis of the paper is complemented by a simulation study to assess the performance of the estimators and tests. The performance is benchmarked against results obtained with OLS, FM-OLS and DOLS. It turns out that the new estimator performs relatively well, in terms of having smaller bias and only moderately larger RMSE than the FM-OLS estimator, in line with the theoretical findings of the paper. The larger RMSE appears to be the price to be paid for partial summing the cointegrating regression, which leads to a regression with I(2) regressors and I(1) errors. In comparison DOLS has smaller bias but much larger RMSE. The simulations of size and power of the tests show that the developed fixed-b limit theory well describes the test statistics' distributions. In particular fixed-b test statistics based on the IM-OLS estimator lead to the smallest size distortions at the expense of only minor losses in (size-corrected) power. This finding is quite similar to the findings of Kiefer and Vogelsang (2005) for testing in stationary regressions and thus extends one of the major contributions of fixed-b theory to the cointegration literature.

The paper is organized as follows: In Section 2 we present a standard linear cointegrating regression and start by reviewing the OLS and FM-OLS estimators and then give the fixed-b limiting distribution of the FM-OLS estimator. Section 3 presents the new IM-OLS estimator whose finite sample performance is studied by means of simulations in Section 4. In Section 5 inference using the IM-OLS parameter estimates is discussed, both with standard and fixed-b asymptotic theory. The finite sample performance of the resultant test statistics is assessed, again with simulations, in Section 6. Section 7 briefly summarizes and concludes. All proofs are relegated to the Appendix. Supplementary material that can be downloaded from the authors' homepages provides tables with fixed-b critical values for the IM-OLS based tests for up to four integrated regressors and the usual specifications of the deterministic component (intercept, intercept and linear trend) for a variety of kernel functions. Also MATLAB code implementing the discussed methods is available upon request.

2. FM-OLS estimation and inference in cointegrating regressions

Consider the following data generating process

$$y_t = \mu + x_t'\beta + u_t \tag{1}$$

$$x_t = x_{t-1} + v_t, \tag{2}$$

where y_t is a scalar time series and x_t is a $k \times 1$ vector of time series with a sample of observations $t = 1, 2, \ldots, T$ available. For notational brevity here we only include the intercept μ as deterministic component (this restriction is removed later when we discuss the IM-OLS estimator in the following section). Stacking the error processes defines $\eta_t = [u_t, v_t']'$. It is assumed that η_t is a vector of $I(0)$ processes, in which case x_t is a non-cointegrating vector of $I(1)$ processes and there exists a cointegrating relationship among $[y_t, x_t']'$ with cointegrating vector $[1, -\beta']'$.

To review existing theory and to obtain the key theoretical results in the paper, assumptions about η_t are required. It is sufficient to assume that η_t satisfies a functional central limit theorem (FCLT) of the form

$$T^{-1/2}\sum_{t=1}^{[rT]} \eta_t \Rightarrow B(r) = \Omega^{1/2}W(r), \quad r \in [0, 1], \tag{3}$$

where $[rT]$ denotes the integer part of rT and $W(r)$ is a $(k+1)$-dimensional vector of independent standard Brownian motions and

$$\Omega = \sum_{j=-\infty}^{\infty} \mathbb{E}(\eta_t \eta_{t-j}') = \begin{bmatrix} \Omega_{uu} & \Omega_{uv} \\ \Omega_{vu} & \Omega_{vv} \end{bmatrix} > 0,$$

where clearly $\Omega_{vu} = \Omega_{uv}'$. The assumption $\Omega_{vv} > 0$ rules out cointegration in x_t. Partition $B(r)$ as

$$B(r) = \begin{bmatrix} B_u(r) \\ B_v(r) \end{bmatrix}$$

and likewise partition $W(r)$ as $W(r) = [w_{u\cdot v}(r), W_v'(r)]'$, where $w_{u\cdot v}(r)$ and $W_v(r)$ are a scalar and a k-dimensional standard Brownian motion respectively. It will be convenient to use $\Omega^{1/2}$ of the Cholesky form

$$\Omega^{1/2} = \begin{bmatrix} \sigma_{u\cdot v} & \lambda_{uv} \\ \mathbf{0} & \Omega_{vv}^{1/2} \end{bmatrix},$$

where $\sigma^2_{u\cdot v} = \Omega_{uu} - \Omega_{uv}\Omega_{vv}^{-1}\Omega_{vu}$ and $\lambda_{uv} = \Omega_{uv}(\Omega_{vv}^{-1/2})'$. Using this Cholesky decomposition we can write

$$B(r) = \begin{bmatrix} B_u(r) \\ B_v(r) \end{bmatrix} = \begin{bmatrix} \sigma_{u\cdot v} w_{u\cdot v}(r) + \lambda_{uv} W_v(r) \\ \Omega_{vv}^{1/2} W_v(r) \end{bmatrix}.$$

Next define the one-sided long run covariance matrix $\Lambda = \sum_{j=1}^{\infty} \mathbb{E}(\eta_{t-j}\eta_t')$, which is partitioned according to the partitioning of Ω as

$$\Lambda = \begin{bmatrix} \Lambda_{uu} & \Lambda_{uv} \\ \Lambda_{vu} & \Lambda_{vv} \end{bmatrix}.$$

Note that $\Omega = \Sigma + \Lambda + \Lambda'$, with $\Sigma = \mathbb{E}(\eta_t\eta_t')$, which is partitioned as

$$\Sigma = \begin{bmatrix} \Sigma_{uu} & \Sigma_{uv} \\ \Sigma_{vu} & \Sigma_{vv} \end{bmatrix}.$$

To discuss the OLS and FM-OLS estimators for (1) define $\widetilde{x}_t = [1, x_t']'$ and $\theta = [\mu, \beta']'$. Stacking all observations together gives the matrix representation $y = \widetilde{X}\theta + u$ with

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_T \end{bmatrix}, \qquad \widetilde{X} = \begin{bmatrix} \widetilde{x}_1' \\ \vdots \\ \widetilde{x}_T' \end{bmatrix}, \qquad u = \begin{bmatrix} u_1 \\ \vdots \\ u_T \end{bmatrix}.$$

Using this notation, the OLS estimator is given by

$$\widehat{\theta} = \left(\widetilde{X}'\widetilde{X}\right)^{-1}\widetilde{X}'y.$$

To state asymptotic results the following scaling matrix is needed:

$$A = \begin{bmatrix} T^{-1/2} & \mathbf{0} \\ \mathbf{0} & T^{-1}I_k \end{bmatrix}.$$

For the OLS estimator it is well known from Phillips and Durlauf (1986) and Stock (1987) that

$$\begin{pmatrix} T^{1/2}(\widehat{\mu} - \mu) \\ T(\widehat{\beta} - \beta) \end{pmatrix} = A^{-1}\left(\widehat{\theta} - \theta\right) = \left(A\widetilde{X}'\widetilde{X}A\right)^{-1}\left(A\widetilde{X}'u\right)$$

$$\Rightarrow \left(\int B_v^*(r)B_v^*(r)'dr\right)^{-1}\left(\int B_v^*(r)dB_u(r) + \Delta_{vu}^*\right) = \Theta,$$

where

$$B_v^*(r) = \begin{pmatrix} 1 \\ B_v(r) \end{pmatrix}, \qquad \Delta_{vu}^* = \begin{pmatrix} 0 \\ \Delta_{vu} \end{pmatrix}, \quad \text{and}$$

$$\Delta_{vu} = \Sigma_{vu} + \Lambda_{vu}.$$

Unless stated otherwise, the range of integration is [0, 1] throughout the paper.

When u_t is uncorrelated with v_t and hence uncorrelated with x_t, it follows that (i) $\lambda_{uv} = \mathbf{0}$, $\Delta_{vu} = \mathbf{0}$, and (ii) $B_u(r)$ is independent of $B_v(r)$. Because of the independence between $B_u(r)$ and $B_v(r)$ in this case, one can condition on $B_v(r)$ to show that the limiting distribution of $T(\widehat{\beta} - \beta)$ is a zero mean Gaussian mixture. Therefore, one can also show that t and *Wald* statistics for testing hypotheses about β have the usual $N(0, 1)$ and chi-square limits assuming serial correlation in u_t is handled using consistent robust standard errors.

When the regressors are endogenous, the limiting distribution of $T(\widehat{\beta} - \beta)$ is obviously more complicated because of the correlation between $B_u(r)$ and $B_v(r)$ and the presence of the nuisance parameters in the vector Δ_{vu}. One can therefore no longer condition on $B_v(r)$ to obtain an asymptotic normal result and Δ_{vu} introduces an asymptotic bias. Inference is difficult in this situation because nuisance parameters cannot be removed by simple scaling methods.

The FM-OLS estimator of Phillips and Hansen (1990) is designed to asymptotically remove Δ_{vu} and to deal with the correlation between $B_u(r)$ and $B_v(r)$. To understand how the FM-OLS estimator works, consider the stochastic process $B_{u\cdot v}(r) = B_u(r) - B_v(r)'\Omega_{vv}^{-1}\Omega_{vu} = \sigma_{u\cdot v}w_{u\cdot v}(r)$ which, by construction, is independent of $B_v(r) = \Omega_{vv}^{1/2}W_v(r)$. Using $B_{u\cdot v}(r)$, one can write

$$\int B_v^*(r)dB_u(r) + \Delta_{vu}^* = \int B_v^*(r)dB_{u\cdot v}(r) + \int B_v^*(r)dB_v(r)'\Omega_{vv}^{-1}\Omega_{vu} + \Delta_{vu}^*. \quad (4)$$

Because $B_v(r)$ and $B_{u\cdot v}(r)$ are independent, conditioning on $B_v(r)$ can be used to show that $\int B_v^*(r)dB_{u\cdot v}(r)$ is a zero mean Gaussian mixture.

The FM-OLS estimator rests upon two transformations. One transformation removes the term $\int B_v^*(r)dB_v(r)'\Omega_{vv}^{-1}\Omega_{vu}$ in (4), whereas the other removes the Δ_{vu}^* term in (4). Because these terms depend on Ω and Δ, the two transformations require estimates of Ω and Δ_{vu}. Let $\widehat{\Omega}$ denote a nonparametric kernel estimator of Ω of the form

$$\widehat{\Omega} = T^{-1}\sum_{i=1}^{T}\sum_{j=1}^{T} k\left(\frac{|i-j|}{M}\right)\widehat{\eta}_i\widehat{\eta}_j', \quad (5)$$

where $\widehat{\eta}_t = [\widehat{u}_t, \Delta x_t']'$ and $\widehat{u}_t$ are the OLS residuals from (1). The function $k(\cdot)$ is the kernel weighting function and M is the bandwidth. Partition $\widehat{\Omega}$ the same way as Ω and define

$$y_t^+ = y_t - \Delta x_t'\widehat{\Omega}_{vv}^{-1}\widehat{\Omega}_{vu}$$

and

$$u_t^+ = u_t - \Delta x_t'\widehat{\Omega}_{vv}^{-1}\widehat{\Omega}_{vu}.$$

Under conditions such that $\widehat{\Omega}$ is a consistent estimator of Ω (see e.g. Jansson, 2002), it follows that

$$A\widetilde{X}'u^+ \Rightarrow \int B_v^*(r)dB_{u\cdot v}(r) + \Delta_{vu}^{+*},$$

where

$$\Delta_{vu}^{+*} = \begin{pmatrix} 0 \\ \Delta_{vu}^+ \end{pmatrix}, \qquad \Delta_{vu}^+ = \Delta_{vu} - \Delta_{vv}\Omega_{vv}^{-1}\Omega_{vu},$$

$u^+ = [u_2^+, \ldots, u_T^+]'$ and where we use for notational simplicity $\widetilde{X}$ to denote the regressors stacked for $t = 2, \ldots, T$ in FM-OLS estimation. Thus, using y_t^+ in place of y_t to estimate θ removes the $\int B_v^*(r)dB_v(r)'\Omega_{vv}^{-1}\Omega_{vu}$ term, but the modified vector Δ_{vu}^{+*} remains.

The term Δ_{vu}^+ is easy to remove as follows: Define the half long run variance $\Delta = \Sigma + \Lambda$ and define a nonparametric kernel estimator for this quantity as

$$\widehat{\Delta} = T^{-1}\sum_{i=1}^{T}\sum_{j=i}^{T} k\left(\frac{|i-j|}{M}\right)\widehat{\eta}_i\widehat{\eta}_j'. \quad (6)$$

Partition Δ and $\widehat{\Delta}$ in the same way as Ω and define $\widehat{\Delta}_{vu}^+$ as

$$\widehat{\Delta}_{vu}^+ = \widehat{\Delta}_{vu} - \widehat{\Delta}_{vv}\widehat{\Omega}_{vv}^{-1}\widehat{\Omega}_{vu}.$$

The FM-OLS estimator is defined as

$$\widehat{\theta}^+ = (\widetilde{X}'\widetilde{X})^{-1}(\widetilde{X}'y^+ - \mathcal{M}^*),$$

where

$$\mathcal{M}^* = T\widehat{\Delta}_{vu}^{+*} \quad \text{and} \quad y^+ = \begin{pmatrix} y_2^+ \\ \vdots \\ y_T^+ \end{pmatrix}, \qquad \widehat{\Delta}_{vu}^{+*} = \begin{pmatrix} 0 \\ \widehat{\Delta}_{vu}^+ \end{pmatrix}.$$

It is shown in Phillips and Hansen (1990) that

$$\begin{aligned} A^{-1}\left(\widehat{\theta}^{+}-\theta\right) &= \left(A\widetilde{X}'\widetilde{X}A\right)^{-1}\left(A\widetilde{X}'u^{+}-A\mathcal{M}^{*}\right) \\ &\Rightarrow \left(\int B_v^*(r)B_v^*(r)'dr\right)^{-1}\int B_v^*(r)dB_{u\cdot v}(r) \\ &= \sigma_{u\cdot v}\left(\int B_v^*(r)B_v^*(r)'dr\right)^{-1} \\ &\quad \times \int B_v^*(r)dw_{u\cdot v}(r), \end{aligned} \tag{7}$$

provided that $\widehat{\Omega}$ and $\widehat{\Delta}_{vu}^{+}$ are consistent. The second part of the transformation uses $\mathcal{M}^*$ to remove Δ_{vu}^{+*}, and the result for $T(\widehat{\theta}^{+}-\theta)$ is such that conditional on $B_v(r)$, a zero mean normal limit is obtained. The limit given by (7) is a mean zero mixture of normals. Conditional on $B_v(r)$ the asymptotic variance is well known to be

$$V_{FM} = \sigma_{u\cdot v}^2\left(\int B_v^*(r)B_v^*(r)'dr\right)^{-1}. \tag{8}$$

Asymptotically pivotal t and *Wald* statistics with $N(0, 1)$ and chi-square limiting distributions can be constructed by taking into account $\sigma_{u\cdot v}^2$, the long run variance of $B_{u\cdot v}(r)$. The traditional estimator of $\sigma_{u\cdot v}^2$ is

$$\widehat{\sigma}_{u\cdot v}^2 = \widehat{\Omega}_{uu} - \widehat{\Omega}_{uv}\widehat{\Omega}_{vv}^{-1}\widehat{\Omega}_{vu}, \tag{9}$$

whereas an alternative estimator is given by

$$\widehat{\sigma}_{u^+}^2 = T^{-1}\sum_{i=1}^{T}\sum_{j=1}^{T} k\left(\frac{|i-j|}{M}\right)\widehat{u}_i^{+}\widehat{u}_j^{+\prime}, \tag{10}$$

where $\widehat{u}_t^{+} = y_t^{+} - \widehat{\mu}^{+} - x_t'\widehat{\beta}^{+}$. The kernel and bandwidth used to construct $\widehat{\sigma}_{u^+}^2$ are not necessarily the same as those used to construct $\widehat{\Omega}$ and $\widehat{\Delta}$.

In practice FM-OLS estimation requires the choice of bandwidth and kernel. While bandwidth and kernel play no role asymptotically when appealing to consistency results for $\widehat{\Omega}$ and $\widehat{\Delta}$, in finite samples they affect the sampling distributions of $\widehat{\theta}^{+}$ and thus of t and *Wald* statistics based on the FM-OLS estimator $\widehat{\theta}^{+}$. To obtain an approximation for $\widehat{\theta}^{+}$ that reflects the choice of bandwidth and kernel, the natural asymptotic theory to use is the fixed-b theory developed by Kiefer and Vogelsang (2005) and further analyzed by Sun et al. (2008). Fixed-b theory has primarily been developed for models with stationary regressors, which means that some additional work is required to obtain analogous results for cointegrating regressions. As we shall see below, a major difference is that the first component of $\widehat{\eta}_t$, i.e. $\widehat{u}_t$, is the residual from a cointegrating regression, which leads to dependence of the corresponding limit partial sum process (defined as $P_{\widehat{\eta}}(r)$ below) on the number of integrated regressors and the deterministic components.

It is important to note that there are two papers in the literature that develop fixed-b theory for cointegration regressions. Jin et al. (2006) carry out a fixed-b analysis of $\widehat{\sigma}_{u^+}^2$ appealing to consistency of $\widehat{\Omega}$ and $\widehat{\Delta}$ which amounts to using the limit given by (7). They obtain pivotal fixed-b limits for t and *Wald* statistics that differ from the stationary regression case. Similarly, Bunzel (2006) develops a fixed-b theory for t and *Wald* statistics based on DOLS and finds that the fixed-b limits of t and *Wald* statistics are the same as in Jin et al. (2006).

The fixed-b theory developed by Jin et al. (2006) is only a partial fixed-b theory as it ignores the impact of $\widehat{\Omega}$ and $\widehat{\Delta}$ on the test statistics which in turn ignores the impact of the choice of bandwidth and kernel needed to implement $\widehat{\Omega}$ and $\widehat{\Delta}$.[1] Developing fixed-b asymptotic theory for $\widehat{\Omega}$ and $\widehat{\Delta}$ is an important contribution of the current paper.

A brief overview of the fixed-b theory may be useful to some readers. Fixed-b theory obtains limits of nonparametric kernel estimators of long run variance matrices by treating the bandwidth as a fixed proportion of the sample size. Specifically, it is assumed that $M = bT$, where $b \in (0, 1]$ remains fixed as $T \to \infty$. Under this assumption it is possible to obtain a limiting expression for a long run variance estimator that is a random variable depending on the kernel $k(\cdot)$ and b. This is in contrast to a consistency result where the limit is a constant, i.e. the true long run variance. It might be tempting to conclude that using fixed-b theory is equivalent to proposing a long run variance estimator that is inconsistent. This is not the case. The long run variance estimators are given by (5) and (6). Given a sample and a particular choice of M, the estimators given by (5) and (6) can be imbedded in sequences that converge to the population long run variances (consistency) or imbedded in sequences that converge to random limits that are functions of $k(\cdot)$ and b (fixed-b). It becomes a question as to which limit provides a more useful approximation. If one wants to capture the impact of bandwidth and kernel choice on the sampling behavior of (5) and (6), fixed-b theory is informative while a consistency result is not.

Obtaining a fixed-b result for $\widehat{\Omega}$ relies upon algebra in Hashimzade and Vogelsang (2008), extended to a multivariate framework and taking into account the above mentioned differences (in relation to $\widehat{u}_t$ in a cointegration framework). The approach pursued in Hashimzade and Vogelsang (2008) is to rewrite $\widehat{\Omega}$ in terms of partial sums of $\widehat{\eta}_t$. Once the limit behavior of appropriately scaled partial sums of $\widehat{\eta}_t$ is established, the fixed-b limit for $\widehat{\Omega}$ follows from the continuous mapping theorem. Obtaining a fixed-b result for $\widehat{\Delta}$ requires additional calculations beyond what is available in the existing fixed-b literature and may be of independent interest itself.

In order to formulate the fixed-b results for $\widehat{\Omega}$, $\widehat{\Delta}$, and $\widehat{\theta}^{+}$ we need to define some additional quantities. Define $P_{\widehat{\eta}}(r)$ and its instantaneous change $dP_{\widehat{\eta}}(r)$ as

$$P_{\widehat{\eta}}(r) = \begin{bmatrix}\widehat{B}_u(r)\\ B_v(r)\end{bmatrix}, \qquad dP_{\widehat{\eta}}(r) = \begin{bmatrix}d\widehat{B}_u(r)\\ dB_v(r)\end{bmatrix},$$

where $\widehat{B}_u(r) = B_u(r) - \int_0^r B_v^*(s)'ds\Theta$ and $d\widehat{B}_u(r) = dB_u(r) - B_v^*(r)'dr\Theta$. As is shown in the Appendix, $P_{\widehat{\eta}}(r)$ is the limit process of the scaled partial sum process of $\widehat{\eta}_t$.

The fixed-b limits of $\widehat{\Omega}$ and $\widehat{\Delta}$ are expressed in terms of functionals whose forms depend on the smoothness of the kernel. We distinguish two cases for the kernel (a third case, not examined here, can be found in Hashimzade and Vogelsang (2008)). In the first case the kernel function $k(\cdot)$, with $k(0) = 1$, is assumed to be twice continuously differentiable with first and second derivatives given by $k'(\cdot)$ and $k''(\cdot)$. Furthermore $k'_{+}(0)$ denotes the derivative evaluated at zero from the right. An example of kernels of this type is given by the Quadratic Spectral kernel.

Let $P_1(r)$ and $P_2(r)$ denote two stochastic processes and define the stochastic processes $Q_b(P_1(r), P_2(r))$ and $Q_b^{\Delta}(P_1(r), P_2(r))$ as

$$\begin{aligned} Q_b(P_1, P_2) &= -\frac{1}{b^2}\int_0^1\int_0^1 k''\left(\frac{|r-s|}{b}\right)P_1(s)P_2(r)'dsdr \\ &\quad + \frac{1}{b}\int_0^1 k'\left(\frac{|1-s|}{b}\right)\left(P_1(1)P_2(s)' + P_1(s)P_2(1)'\right)ds \\ &\quad + P_1(1)P_2(1)', \end{aligned}$$

[1] Similarly, the Bunzel (2006) fixed-b results do not capture the choices of lag and lead lengths needed to implement DOLS.

$$\begin{aligned} Q_b^{\Delta}(P_1, P_2) \;=\; & -\frac{1}{b^2}\int_0^1\int_r^1 k''\left(\frac{|r-s|}{b}\right) P_1(s)P_2(r)'drds \\ & + \frac{1}{b}\int_0^1 k'\left(\frac{|1-s|}{b}\right) P_1(s)P_2(1)'ds \\ & + \frac{1}{b}k'_+(0)\int_0^1 P_1(s)P_2(s)'ds + P_1(1)P_2(1)' \\ & - \int_0^1 dP_1(s)P_2(s)'. \end{aligned}$$

The second case considered refers to the Bartlett kernel (i.e. $k(x) = 1 - |x|$ for $|x| \le 1$ and 0 otherwise), in which case the stochastic processes $Q_b(P_1, P_2)$ and $Q_b^{\Delta}(P_1, P_2)$ become

$$\begin{aligned} Q_b(P_1, P_2) \;=\; & \frac{2}{b}\int_0^1 P_1(s)P_2(s)'ds - \frac{1}{b}\int_0^{1-b}\big(P_1(s)P_2(s+b)' \\ & + P_1(s+b)P_2(s)'\big)\, ds - \frac{1}{b}\int_{1-b}^1 \big(P_1(1)P_2(s)' \\ & + P_1(s)P_2(1)'\big)\, ds + P_1(1)P_2(1)', \\ Q_b^{\Delta}(P_1, P_2) \;=\; & \frac{1}{b}\int_0^1 P_1(s)P_2(s)'ds - \frac{1}{b}\int_0^{1-b} P_1(s)P_2(s+b)'ds \\ & - \frac{1}{b}\int_{1-b}^1 P_1(s)P_2(1)'ds + P_1(1)P_2(1)' \\ & - \int_0^1 dP_1(s)P_2(s)'. \end{aligned}$$

With all required quantities defined we can now state the fixed-b limit results for $\widehat{\Omega}$ and $\widehat{\Delta}$ which in turn lead to the fixed-b limit of the FM-OLS estimator. In the formulation of the theorem we will not distinguish the two discussed cases with respect to the kernel function, but just use the brief notation Q_b and Q_b^{Δ}. In addition, we now use $\widehat{\theta}_b^+$ to denote $\widehat{\theta}^+$ to indicate the dependence of $\widehat{\theta}^+$ on the bandwidth. The dependence on the kernel remains implicit in the notation.

Theorem 1. *Assume that the data are generated by (1) and (2) and that the FCLT (3) holds. Let $M = bT$, where $b \in (0, 1]$ is held fixed as $T \to \infty$. Then as $T \to \infty$*

$$\widehat{\Omega} \Rightarrow Q_b(P_{\widehat{\eta}}, P_{\widehat{\eta}}), \qquad \widehat{\Delta} \Rightarrow Q_b^{\Delta}(P_{\widehat{\eta}}, P_{\widehat{\eta}}) - \Lambda' \tag{11}$$

and in particular

$$\begin{aligned} &\widehat{\Omega}_{vv} \Rightarrow Q_b(B_v, B_v), \qquad \widehat{\Omega}_{vu} \Rightarrow Q_b(B_v, \widehat{B}_u), \\ &\widehat{\Delta}_{vv} \Rightarrow Q_b^{\Delta}(B_v, B_v) - \Lambda'_{vv}, \qquad \widehat{\Delta}_{vu} \Rightarrow Q_b^{\Delta}(B_v, \widehat{B}_u) - \Lambda'_{uv}. \end{aligned}$$

The fixed-b limit of the FM-OLS estimator $\widehat{\theta}_b^+$ is given by

$$\begin{aligned} A^{-1}\left(\widehat{\theta}_b^+ - \theta\right) \;=\; & \left(A\widetilde{X}'\widetilde{X}A\right)^{-1}\left(A\widetilde{X}'u^+ - A\mathcal{M}^*\right) \\ \Rightarrow\; & \left(\int B_v^*(r)B_v^*(r)'dr\right)^{-1} \\ & \times\left(\int B_v^*(r)dB_{uv}^b(r) + \mathcal{B}_1 - \mathcal{B}_2\right), \end{aligned} \tag{12}$$

with $B_{uv}^b(r) = B_u(r) - B_v(r)'Q_b(B_v, B_v)^{-1}Q_b(B_v, \widehat{B}_u)$ and

$$\mathcal{B}_1 = \begin{pmatrix} 0 \\ \Delta_{vu} - \left(Q_b^{\Delta}(B_v, \widehat{B}_u) - \Lambda'_{uv}\right) \end{pmatrix},$$

$$\mathcal{B}_2 = \begin{pmatrix} 0 \\ \left(\Delta_{vv} - \left(Q_b^{\Delta}(B_v, B_v) - \Lambda'_{vv}\right)\right) Q_b(B_v, B_v)^{-1}Q_b(B_v, \widehat{B}_u) \end{pmatrix}.$$

Theorem 1 shows that under the fixed-b asymptotic approximation, the limit distribution of the FM-OLS estimator depends in a complicated fashion upon nuisance parameters. These nuisance parameters are, by construction, related to the two transformations upon which the FM-OLS estimator relies. The result clearly shows that the zero mean mixed normal approximation for FM-OLS will not be satisfactory if the sampling distributions of $\widehat{\Omega}$ and $\widehat{\Delta}$ are not close to Ω and Δ. Consider the orthogonalization step of FM-OLS. The term $\int B_v^*(r)dB_{uv}^b(r)$ is close to a zero mean Gaussian mixture only if in $B_{uv}^b(r) = B_u(r) - B_v(r)'Q_b(B_v, B_v)^{-1}Q_b(B_v, \widehat{B}_u)$ the Q_b terms are close to the population quantities Ω_{vv}^{-1} and Ω_{vu}, with this proximity depending upon kernel and bandwidth choice. Similar observations hold for the second transformation, i.e. the removal of Δ_{vu}^{+*}. The term $\mathcal{B}_1 - \mathcal{B}_2$ is close to zero when $Q_b^{\Delta}(B_v, \widehat{B}_u) - \Lambda'_{uv}$ and $Q_b^{\Delta}(B_v, B_v) - \Lambda'_{vv}$ are close to Δ_{vu} and Δ_{vv}. If these approximations are not accurate, an additive bias is present.

The following corollary provides an alternative expression for the fixed-b limit of $\widehat{\theta}_b^+$ that can be useful in making additional comparisons between Theorem 1 and the traditional limit distribution given in (7).

Corollary 1. *The fixed-b limit of $\widehat{\theta}_b^+$ can be written as:*

$$\begin{aligned} \lim_{T\to\infty} A^{-1}\left(\widehat{\theta}_b^+ - \theta\right) = & \left(\int_0^1 B_v^*(r)B_v^*(r)'dr\right)^{-1} \\ & \times\left(\int_0^1 B_v^*(r)dB_{u\cdot v}(r) + \mathcal{F}(\widetilde{B}_{u\cdot v}) - \mathcal{F}\left(F(B_v^*)\right)\right), \end{aligned}$$

with

$$\begin{aligned} \widetilde{B}_{u\cdot v}(r) \;=\; & B_{u\cdot v}(r) - \int_0^r B_v^*(s)ds\left(\int_0^1 B_v^*(s)B_v^*(s)'ds\right)^{-1} \\ & \times \int_0^1 B_v^*(s)dB_{u\cdot v}(s) \\ F\left(B_v^*(r)\right) \;=\; & \int_0^r B_v^*(s)'ds\left(\int_0^1 B_v^*(s)B_v^*(s)'ds\right)^{-1} \\ & \times\left(\int_0^1 B_v^*(s)dB_v(s)'\Omega_{vv}^{-1}\Omega_{vu} + \Delta_{vu}^*\right) \end{aligned}$$

and the functional $\mathcal{F}(P)$ defined for some stochastic process $P(r)$ as

$$\begin{aligned} \mathcal{F}(P) \;=\; & -\int_0^1 B_v^*(r)dB_v(r)'Q_b(B_v, B_v)^{-1}Q_b(B_v, P) \\ & - \begin{pmatrix} 0 \\ Q_b^{\Delta}(B_v, P) \end{pmatrix} - \begin{pmatrix} 0 \\ \Omega_{vv}Q_b(B_v, B_v)^{-1}Q_b(B_v, P) \end{pmatrix} \\ & + \begin{pmatrix} 0 \\ Q_b^{\Delta}(B_v, B_v)Q_b(B_v, B_v)^{-1}Q_b(B_v, P) \end{pmatrix}. \end{aligned}$$

The corollary provides several insights. First, the term $\mathcal{F}(F(B_v^*))$, being a function of B_v and the nuisance parameters Ω_{vv}, Ω_{vu} and Δ_{vu}^*, represents a conditional asymptotic bias in FM-OLS. Second, one can show through a tedious calculation that $\mathcal{F}(\widetilde{B}_{u\cdot v})$ is linear in $dB_{u\cdot v}(s)$ with weights that are functions of $B_v(r)$ and Ω_{vv}. Therefore, the fixed-b limit of FM-OLS is mixture normal with a non-zero mean and a conditional variance matrix that is much more complex than V_{FM}. Third, we can show that the fixed-b limit of FM-OLS simplifies to (7) as $b \to 0$ by showing that $\text{plim}_{b\to 0}\mathcal{F}(\widetilde{B}_{u\cdot v}) = 0$ and $\text{plim}_{b\to 0}\mathcal{F}\left(F(B_v^*)\right) = 0$. The following proposition gives the formal result, and a sketch of the proof is given in the Appendix.

Proposition 1. *As $b \to 0$, the fixed-b limiting distribution of $\widehat{\theta}_b^+$ converges in probability to the traditional limit distribution (7).*

These results show that the performance of FM-OLS relies critically on the consistency approximation of the long run variance estimators being accurate and that moving around the bandwidth and kernel impacts the sampling behavior of the FM-OLS estimator. It is reasonable to expect the accuracy of the traditional approximation given by (7) to often be inadequate in practice given the well known bias and sampling variability problems of nonparametric kernel long run variance estimators.

3. The integrated modified OLS estimator

In this section we present a new estimator for which a simple transformation is used to obtain an asymptotically unbiased estimator with a zero mean Gaussian mixture limiting distribution. Like FM-OLS, the transformation has two steps but neither step requires estimates of Ω or Δ_{vu}^{+} and so the choice of bandwidth and kernel is completely avoided. We consider a slightly more general version of (1) given by

$$y_t = f_t'\delta + x_t'\beta + u_t, \tag{13}$$

where x_t continues to follow (2) and where for the deterministic components f_t we merely assume that there is a $p \times p$ matrix τ_F and a vector of functions, $f(s)$, such that

$$T^{-1}\tau_F^{-1}\sum_{t=1}^{[rT]} f_t \to \int_0^r f(s)ds \quad \text{with} \int_0^1 f(s)f(s)'ds > 0. \tag{14}$$

If e.g. $f_t = (1, t, t^2, \ldots, t^{p-1})'$, then τ_F is a diagonal matrix with diagonal elements $1, T, T^2, \ldots, T^{p-1}$ and $f(s) = (1, s, s^2, \ldots, s^{p-1})'$.

Computing the partial sum of both sides of (13) gives

$$S_t^y = S_t^{f\prime}\delta + S_t^{x\prime}\beta + S_t^u, \tag{15}$$

where $S_t^y = \sum_{j=1}^t y_j$ and S_t^f, S_t^x and S_t^u are defined analogously. In vector notation, using similar notation as in the discussion of the OLS estimator, we have

$$S^y = S^{\tilde{x}}\theta + S^u, \tag{16}$$

with $S^{\tilde{x}}$ stacking S_t^f and S_t^x. With this notation the OLS estimator in the partial sum regression is given by

$$\widetilde{\theta} = \left(S^{\tilde{x}\prime}S^{\tilde{x}}\right)^{-1}\left(S^{\tilde{x}\prime}S^y\right) \tag{17}$$

which leads to

$$\widetilde{\theta} - \theta = \left(S^{\tilde{x}\prime}S^{\tilde{x}}\right)^{-1}\left(S^{\tilde{x}\prime}S^u\right). \tag{18}$$

The benefit of partial summing is that sub-matrices of the form

$$\sum_{t=1}^T x_t u_t \tag{19}$$

that appear in $\widehat{\theta}$ and $\widehat{\theta}^+$ are replaced by sub-matrices of the form

$$\sum_{t=1}^T S_t^x S_t^u \tag{20}$$

in $\widetilde{\theta}$. Appropriately scaled sums of the form of (19) have been well studied in the econometrics literature, see Phillips (1988), Hansen (1992), De Jong and Davidson (2000a,b) and the references therein, and are the source of the additive nuisance parameters, Δ_{vu}, that show up in the limit of the OLS estimator. In contrast, scaled sums of the form of (20) do not have such additive terms in their limits. Partial summing before estimating the model thus performs the same role for IM-OLS that $\mathcal{M}^*$ plays for FM-OLS.

This still leaves the problem that correlation between u_t and v_t (x_t) rules out the possibility of conditioning on $B_v(r)$ to obtain a conditional asymptotic normality result. The solution to this problem is simple and only requires that x_t be added as a regressor to the partial sum regression (15):

$$S_t^y = S_t^{f\prime}\delta + S_t^{x\prime}\beta + x_t'\gamma + S_t^u. \tag{21}$$

Adding x_t as regressors is related to DOLS. The simple endogeneity correction by just including the original regressors x_t in the partial summed regression works because both x_t and S_t^u are I(1) processes, which implies that all correlation is soaked up in the long run correlation matrix $\Omega_{vv}^{-1}\Omega_{vu}$ and it is not necessary to include any leads or lags as would be the case e.g. in DOLS estimation of the original regression (with I(1) regressors but I(0) errors). Therefore, the 'centering' parameter for $\widetilde{\gamma}$ in case of endogeneity is $\Omega_{vv}^{-1}\Omega_{vu}$ and not the population value of $\gamma = 0$.

As shall be discussed in greater detail in Section 5, fixed-b inference regarding δ, β and γ using OLS estimators from (21) is complicated by correlation between $\widetilde{\delta}, \widetilde{\beta}, \widetilde{\gamma}$, the OLS estimators from (21), and the OLS residuals of (21) which we denote by

$$\widetilde{S}_t^u = S_t^y - S_t^{f\prime}\widetilde{\delta} - S_t^{x\prime}\widetilde{\beta} - x_t'\widetilde{\gamma}. \tag{22}$$

This correlation depends on unknown nuisance parameters and the correlation remains even asymptotically. In order for HAC based tests to be asymptotically pivotal under fixed-b asymptotics, we need residuals that are asymptotically independent of $\widetilde{\delta}, \widetilde{\beta}, \widetilde{\gamma}$. This can be achieved by adjusting the residuals from (22) as follows. Define the vector of regressors z_t as

$$z_t = t\sum_{j=1}^T \xi_j - \sum_{j=1}^{t-1}\sum_{s=1}^j \xi_s, \quad \xi_t = [S_t^{f\prime}, S_t^{x\prime}, x_t']', \tag{23}$$

and let $z_t^\perp$ denote the vector of residuals from individually regressing each element of z_t on the regressors S_t^f, S_t^x, x_t. The adjusted residuals, denoted by $\widetilde{S}_t^{u*}$, are obtained as the OLS residuals from the regression of $\widetilde{S}_t^u$ on $z_t^\perp$. In other words

$$\widetilde{S}_t^{u*} = \widetilde{S}_t^u - z_t^{\perp\prime}\widehat{\pi}, \tag{24}$$

where $\widehat{\pi} = \left(\sum_{t=1}^T z_t^\perp z_t^{\perp\prime}\right)^{-1}\sum_{t=1}^T z_t^\perp \widetilde{S}_t^u$. We show that (upon appropriate scaling) $\widetilde{S}_t^{u*}$ is asymptotically independent of $\widetilde{\delta}, \widetilde{\beta}, \widetilde{\gamma}$. This asymptotic independence is sufficient to obtain asymptotically pivotal test statistics under fixed-b asymptotics. The adjustment to the residuals used in (24) is similar in spirit to, although mechanically different from, the adjustment used in constructing $\widehat{u}_t^+$ whereby y_t is replaced by $\widehat{y}_t^+$ when computing residuals for (1) using the FM-OLS estimators.

We now focus on the asymptotic behavior of the OLS estimators of δ, β and γ from (21), which we label the IM-OLS estimators of δ, β and γ. Redefine $S^{\tilde{x}}$ so that it stacks S_t^f, S_t^x, x_t and redefine $\widetilde{\theta}$ so that it stacks $\widetilde{\delta}, \widetilde{\beta}, \widetilde{\gamma}$. With this economical use of notation, the matrix form of (21) is still given by (16) and the OLS estimator is still formally given by (17) and (18). Define the scaling matrix

$$A_{IM} = \begin{bmatrix} T^{-1/2}\tau_F^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & T^{-1}I_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & I_k \end{bmatrix}.$$

The following theorem gives the asymptotic distribution of $\widetilde{\delta}, \widetilde{\beta}, \widetilde{\gamma}$.

Theorem 2. *Assume that the data are generated by (1) and (2), that the FCLT (3) holds and that the deterministic components satisfy (14). Define θ by stacking the vectors δ, β and $\Omega_{vv}^{-1}\Omega_{vu}$. Then as $T \to \infty$*

$$\begin{aligned}
\begin{pmatrix} T^{1/2}\tau_F(\widetilde{\delta} - \delta) \\ T(\widetilde{\beta} - \beta) \\ (\widetilde{\gamma} - \Omega_{vv}^{-1}\Omega_{vu}) \end{pmatrix} &= A_{IM}^{-1}\left(\widetilde{\theta} - \theta\right) \\
&= \left(T^{-2}A_{IM}S^{\tilde{x}\prime}S^{\tilde{x}}A_{IM}\right)^{-1}\left(T^{-2}A_{IM}S^{\tilde{x}\prime}S^u\right) - \begin{pmatrix} 0 \\ 0 \\ \Omega_{vv}^{-1}\Omega_{vu} \end{pmatrix} \\
&\Rightarrow \sigma_{u\cdot v}\left(\Pi\int g(s)g(s)'ds\Pi'\right)^{-1}\Pi\int g(s)w_{u\cdot v}(s)ds \\
&= \sigma_{u\cdot v}(\Pi')^{-1}\left(\int g(s)g(s)'ds\right)^{-1}\int [G(1) - G(s)]dw_{u\cdot v}(s) \\
&= \Psi,
\end{aligned} \tag{25}$$

where

$$\Pi = \begin{bmatrix} I_p & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \Omega_{vv}^{1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \Omega_{vv}^{1/2} \end{bmatrix}, \qquad g(r) = \begin{bmatrix} \int_0^r f(s)ds \\ \int_0^r W_v(s)ds \\ W_v(r) \end{bmatrix},$$

$$G(r) = \int_0^r g(s)ds.$$

Conditional on $W_v(r)$, it holds that $\Psi \sim N(\mathbf{0}, V_{IM})$, where V_{IM} is given by

$$V_{IM} = \sigma_{u\cdot v}^2 (\Pi')^{-1} \left(\int g(s)g(s)'ds\right)^{-1} \times \left(\int [G(1) - G(s)][G(1) - G(s)]'ds\right) \times \left(\int g(s)g(s)'ds\right)^{-1} \Pi^{-1}. \tag{26}$$

This conditional asymptotic variance differs from the conditional asymptotic variance of the FM-OLS estimator of δ and β. Denoting with $m(s) = [f(s)', W_v(s)']'$ and with $\Pi_{FM} = \mathrm{diag}(I_p, \Omega_{vv}^{1/2})$ the latter is given by, as already introduced in (8) for the special case $f(s) = 1$ in Section 2:

$$V_{FM} = \sigma_{u\cdot v}^2 \left(\Pi_{FM}'\right)^{-1} \left(\int m(s)m(s)'ds\right)^{-1} (\Pi_{FM})^{-1}. \tag{27}$$

An interesting question arises as to how the conditional variance covariance matrices in (26) and (27) compare. We are unaware of a theory of mixture of normals that would allow for an easy comparison of the two unconditional variance covariance matrices. However, we can make a comparison *conditional* on $B_v(s)$, i.e. conditional on x_t.

Proposition 2. *Conditional on $B_v(s)$, it holds that $V_{FM} \leq V_{IM}$ for the δ and β components.*

The proof is given in the Appendix and is essentially a continuous time version of the Gauss–Markov Theorem applied in the limiting environment. While Proposition 1 suggests that IM-OLS is asymptotically less efficient than FM-OLS, one needs to keep in mind that the traditional FM-OLS variance V_{FM} ignores the impact of the long run variance estimators on the sampling behavior of FM-OLS. A more useful comparison from the perspective of practice is between V_{IM} and the conditional variance implicit in the fixed-b analysis of FM-OLS in Corollary 1. This is true because these asymptotic variances explicitly capture the impact of the transformations used to remove the impact of endogeneity on the estimation of the regression parameters. Given the complicated dependence on nuisance parameters, the kernel and bandwidth, and other features of the model in the limit given by Corollary 1, such a comparison would be very tedious, and there is no reason to think that a clean variance ranking would be obtained. However, we do obtain a clean ranking of asymptotic bias. IM-OLS is asymptotically unbiased whereas FM-OLS is asymptotically biased because of the $\mathcal{F}\left(F(B_v^*)\right)$ term in Corollary 1.

Thus, fixed-b theory predicts that FM-OLS will exhibit more bias in finite samples than IM-OLS whereas rankings of variance and mean square error must be made on a case by case basis. We make some finite sample comparisons between FM-OLS and IM-OLS using a simulation study described in the next section and the patterns we observe are generally consistent with the predictions made by the fixed-b theory.

4. Finite sample bias and root mean squared error

In this section we compare the performance of the OLS, FM-OLS, DOLS and IM-OLS estimators as measured by bias and root mean squared error (RMSE) with a small simulation study. The data generating process is given by

$$y_t = \mu + x_{1t}\beta_1 + x_{2t}\beta_2 + u_t,$$
$$x_{it} = x_{i,t-1} + v_{it}, \qquad x_{i0} = 0, \quad i = 1, 2$$

where

$$u_t = \rho_1 u_{t-1} + \varepsilon_t + \rho_2(e_{1t} + e_{2t}), \qquad u_0 = 0,$$
$$v_{it} = e_{it} + 0.5e_{i,t-1}, \quad i = 1, 2,$$

where ε_t, e_{1t} and e_{2t} are i.i.d. standard normal random variables independent of each other. The parameter values chosen are $\mu = 3$, $\beta_1 = \beta_2 = 1$, where we note that the value of μ has no effect on the results because the estimators of β_1 and β_2 are exactly invariant to the value of μ. The values for ρ_1 and ρ_2 are chosen from the set $\{0.0, 0.3, 0.6, 0.9\}$. The parameter ρ_1 controls serial correlation in the regression error, whereas the parameter ρ_2 controls whether the regressors are endogenous or not. The kernels chosen for FM-OLS are the Bartlett and the Quadratic Spectral kernels and the bandwidths are reported for the grid $M = bT$ with $b \in \{0.06, 0.10, 0.30, 0.50, 0.70, 0.90, 1.00\}$. We also use the data dependent bandwidth chosen according to Andrews (1991). The DOLS estimator is implemented using the information criterion based lead and lag length choice as developed in Kejriwal and Perron (2008), where we use the more flexible version discussed in Choi and Kurozumi (2012) in which the numbers of leads and lags included are not restricted to be equal. The considered sample sizes are $T = 100, 200$ and the number of replications is 5000.

In Table 1 we display for brevity only the results for $T = 100$ for the Bartlett kernel because the results for the Quadratic Spectral kernel and for $T = 200$ are qualitatively very similar. Panel A reports bias and Panel B reports RMSE.

When there is no endogeneity ($\rho_2 = 0$), none of the estimators shows much bias for any value of ρ_1. When the bandwidth is relatively small, FM-OLS and OLS have similar RMSEs, as would be expected since they have the same asymptotic variance when $\rho_2 = 0$. But, as the bandwidth increases, the RMSE of FM-OLS tends to first increase and then decreases, indicating a hump-shape in the RMSE. OLS and FM-OLS have smaller RMSE than IM-OLS and this holds regardless of bandwidth for FM-OLS. This is not surprising because IM-OLS uses a regression with an I(1) error, whereas OLS and FM-OLS are based on a regression with an I(0) error, compare again also Proposition 2 and the discussion thereafter. Nevertheless, DOLS has the largest RMSE.

When $\rho_2 \neq 0$, in which case there is endogeneity, some interesting and different patterns emerge. As ρ_2 increases, the bias of OLS increases. FM-OLS is less biased than OLS, but FM-OLS does show an increase in bias as ρ_2 increases. This pattern of increasing bias is especially pronounced when ρ_1 is far away from zero. The bias of FM-OLS also depends on the bandwidth and is seen to initially fall as the bandwidth increases and then tends to increase as the bandwidth becomes large. The bias of FM-OLS can exceed the bias of OLS when very large bandwidths are used. In contrast the biases of IM-OLS and DOLS are much less sensitive to ρ_2 and are always smaller than the biases of OLS or FM-OLS. The bias of DOLS is similar to the bias of IM-OLS when ρ_1 is small whereas for larger values of ρ_1, the bias of DOLS tends to be smaller than that of IM-OLS. When $\rho_1 = 0.9$, the biases of IM-OLS and DOLS are much smaller than the biases of FM-OLS or OLS. The overall picture depicted by Panel A is that DOLS has smaller bias than IM-OLS which in turn has lower bias than both OLS and FM-OLS. The magnitude of the bias of both DOLS and IM-OLS is less sensitive to the values of ρ_1 and ρ_2 than for OLS and FM-OLS.

Looking at Panel B we see that the RMSEs of DOLS and IM-OLS tend to be larger than the RMSEs of OLS and FM-OLS, although

Table 1
Finite sample bias and RMSE of the various estimators of β_1, $T = 100$.

ρ_1	ρ_2	OLS	IM-OLS	DOLS	FM-OLS, Bartlett kernel							
					$M = 6$	10	30	50	70	90	100	AND
Panel A: Bias												
0.0	0.0	.0002	.0007	.0002	.0005	.0004	.0003	.0003	.0002	.0002	.0002	.0004
	0.3	.0050	−.0001	−.0003	.0018	.0029	.0047	.0053	.0055	.0056	.0056	.0015
	0.6	.0098	−.0008	−.0002	.0031	.0054	.0091	.0104	.0108	.0110	.0110	.0025
	0.9	.0146	−.0015	−.0001	.0043	.0078	.0135	.0154	.0160	.0164	.0165	.0035
0.3	0.0	.0002	.0009	−.0014	.0007	.0006	.0004	.0004	.0003	.0003	.0003	.0006
	0.3	.0107	.0012	−.0010	.0046	.0063	.0101	.0114	.0118	.0120	.0121	.0042
	0.6	.0213	.0014	−.0010	.0085	.0120	.0198	.0224	.0233	.0238	.0239	.0079
	0.9	.0318	.0016	−.0004	.0124	.0177	.0295	.0335	.0348	.0355	.0357	.0115
0.6	0.0	.0004	.0015	−.0059	.0010	.0010	.0006	.0006	.0005	.0004	.0004	.0010
	0.3	.0239	.0063	−.0046	.0130	.0149	.0220	.0249	.0258	.0263	.0265	.0129
	0.6	.0473	.0111	−.0036	.0250	.0287	.0435	.0492	.0512	.0522	.0526	.0248
	0.9	.0708	.0160	−.0031	.0370	.0426	.0650	.0736	.0766	.0781	.0786	.0366
0.9	0.0	−.0001	.0022	−.0032	.0006	.0009	.0002	.0000	−.0006	−.0006	−.0005	.0006
	0.3	.0801	.0560	.0371	.0678	.0664	.0723	.0791	.0817	.0836	.0843	.0682
	0.6	.1603	.1098	.0769	.1349	.1319	.1443	.1581	.1640	.1678	.1691	.1359
	0.9	.2405	.1637	.1189	.2021	.1973	.2163	.2371	.2464	.2519	.2539	.2035
Panel B: RMSE												
0.0	0.0	.0265	.0375	.1301	.0287	.0290	.0299	.0304	.0306	.0302	.0301	.0286
	0.3	.0286	.0376	.1350	.0292	.0299	.0314	.0320	.0324	.0320	.0319	.0289
	0.6	.0345	.0378	.1371	.0308	.0327	.0357	.0368	.0375	.0371	.0369	.0303
	0.9	.0426	.0379	.1388	.0334	.0369	.0420	.0437	.0447	.0442	.0439	.0325
0.3	0.0	.0365	.0532	.2022	.0403	.0407	.0414	.0419	.0422	.0416	.0414	.0401
	0.3	.0408	.0532	.2040	.0414	.0426	.0446	.0455	.0462	.0456	.0454	.0410
	0.6	.0520	.0533	.2076	.0447	.0480	.0536	.0556	.0566	.0561	.0558	.0439
	0.9	.0668	.0534	.2097	.0498	.0559	.0662	.0694	.0708	.0703	.0700	.0483
0.6	0.0	.0589	.0903	.3535	.0671	.0678	.0673	.0678	.0682	.0671	.0667	.0666
	0.3	.0688	.0906	.3552	.0704	.0724	.0750	.0766	.0775	.0766	.0762	.0697
	0.6	.0930	.0916	.3579	.0799	.0851	.0957	.0996	.1012	.1004	.1001	.0787
	0.9	.1233	.0934	.3595	.0937	.1029	.1230	.1294	.1318	.1311	.1307	.0919
0.9	0.0	.1547	.2661	.7758	.1822	.1889	.1847	.1835	.1816	.1774	.1758	.1800
	0.3	.1864	.2780	.7823	.2039	.2102	.2100	.2123	.2117	.2077	.2063	.2019
	0.6	.2607	.3121	.7983	.2595	.2656	.2757	.2843	.2855	.2820	.2806	.2579
	0.9	.3515	.3622	.8228	.3324	.3387	.3604	.3754	.3782	.3749	.3736	.3311

when ρ_1 and ρ_2 are large, IM-OLS can have slightly smaller RMSE than FM-OLS when a large bandwidth is used. In all cases, DOLS has the highest RMSE. For a given value of ρ_1, the RMSE of OLS noticeably increases as ρ_2 increases. When ρ_1 is small, the RMSE of FM-OLS is not very sensitive to ρ_2 unless the bandwidth is large. The RMSE of IM-OLS does not vary with ρ_2 when ρ_1 is small. When ρ_1 is large, the RMSE of FM-OLS increases with ρ_2. The RMSE of IM-OLS shows a similar pattern, but the RMSE of IM-OLS is less sensitive to the value of ρ_2. DOLS has a much larger RMSE than all other estimators when $\rho_1 = 0.9$. Focusing on the bandwidth we see that the RMSE of FM-OLS is sensitive to the bandwidth as was the case with bias. As the bandwidth increases, the RMSE of FM-OLS tends to increase.

The simulations show that IM-OLS is more effective in reducing bias than FM-OLS and both bias and RMSE of IM-OLS are less sensitive to the nuisance parameters ρ_1 and ρ_2 than are the bias and RMSE of FM-OLS. DOLS has less bias than IM-OLS but a higher RMSE. The superior bias properties of IM-OLS and DOLS come at the cost of higher RMSE, unless ρ_1 and ρ_2 are both large in which case IM-OLS has RMSE similar to OLS and FM-OLS. With respect to the FM-OLS estimator, the simulations reflect the predictions of Theorem 1 and the following Corollary 1 and Proposition 1, showing that the performance of the FM-OLS estimator is sensitive to the bandwidth choice (due to its impact on the approximation accuracy of the long run variance estimators).

5. Inference using IM-OLS

This section is devoted to a discussion of hypothesis testing using the IM-OLS estimator. The basis for doing so is the zero mean Gaussian mixture limiting distribution of the IM-OLS estimator given in Theorem 1 and the expression for the conditional asymptotic variance matrix given by (26). In particular we consider *Wald* tests for testing multiple linear hypotheses of the form

$$H_0 : R\theta = r,$$

where $R \in \mathbb{R}^{q\times(p+2k)}$ with full rank q and $r \in \mathbb{R}^q$. Because the vector $\widetilde{\theta}$ has elements that converge at different rates, obtaining formal results for the *Wald* statistics requires a condition on R that is unnecessary when all estimated coefficients converge at the same rate. As is well known in the literature, for a given constraint (a given row of R), the estimator with the slowest rate of convergence dominates the asymptotic distribution of the linear combination implied by the constraint. See, for example, the discussion in Section 4 of Sims et al. (1990). When there are two or more restrictions being tested, it is not necessarily the case that the slowest converging estimator dominates a given restriction. Should another restriction involve that slowest converging estimator, it is usually possible that the restrictions can be rotated so that (i) the slowest rate estimator only appears in one restriction and (ii) the *Wald* statistic has the exact same value. Because of this possibility, we do not state conditions on R related to the rates of convergence of the estimators involved in the constraints. Rather, we state a sufficient condition for R under which the *Wald* statistics have limiting chi-square distributions. We assume that there exists a nonsingular $q \times q$ scaling matrix A_R such that

$$\lim_{T\to\infty} A_R^{-1} R A_{IM} = R^*, \tag{28}$$

where R^* has rank q. Note that A_R typically has elements that are positive powers of T and that it need not be diagonal.

The expression (26) immediately suggests estimators, $\check{V}_{IM}$, for V_{IM} of the form

$$\begin{aligned}\check{V}_{IM} &= \check{\sigma}^2_{u\cdot v} A^{-1}_{IM} \left(S^{\widetilde{x}\prime} S^{\widetilde{x}}\right)^{-1} \left(C'C\right) \left(S^{\widetilde{x}\prime} S^{\widetilde{x}}\right)^{-1} A^{-1}_{IM} \\ &= \check{\sigma}^2_{u\cdot v} \left(T^{-2} A_{IM} S^{\widetilde{x}\prime} S^{\widetilde{x}} A_{IM}\right)^{-1} (T^{-4} A_{IM} C' C A_{IM}) \\ &\quad \times \left(T^{-2} A_{IM} S^{\widetilde{x}\prime} S^{\widetilde{x}} A_{IM}\right)^{-1},\end{aligned}$$

where $\check{\sigma}^2_{u\cdot v}$ is an estimator of $\sigma^2_{u\cdot v}$ and $C = [c_1, \ldots, c_T]'$ with $c_t = S^{S^{\widetilde{x}}}_T - S^{S^{\widetilde{x}}}_{t-1}$ and $S^{S^{\widetilde{x}}}_t = \sum_{j=1}^t S^{\widetilde{x}}_j$.

There are several obvious candidates for $\check{\sigma}^2_{u\cdot v}$. The first is to use $\widehat{\sigma}^2_{u\cdot v}$ as given in (6), whose consistency properties have been studied e.g. in Phillips (1995), see also Jansson (2002). The second obvious idea is to use, $\Delta\widetilde{S}^u_t$, the first differences of the OLS residuals of the IM-OLS regression (21), to directly estimate $\sigma^2_{u\cdot v}$:

$$\widetilde{\sigma}^2_{u\cdot v} = T^{-1} \sum_{i=2}^{T} \sum_{j=2}^{T} k\left(\frac{|i-j|}{M}\right) \Delta\widetilde{S}^u_i \Delta\widetilde{S}^u_j.$$

It turns out (see Theorem 3) that $\widetilde{\sigma}^2_{u\cdot v}$ is not consistent under traditional assumptions on the bandwidth and kernel as discussed e.g. in Jansson (2002). However, under traditional bandwidth assumptions, the limit of $\widetilde{\sigma}^2_{u\cdot v}$ is shown in Theorem 3 to be larger than $\sigma^2_{u\cdot v}$, which implies that test statistics using $\widetilde{\sigma}^2_{u\cdot v}$ are asymptotically conservative when standard normal or chi-square critical values are used.

Given an estimator of $\sigma^2_{u\cdot v}$, we can define a *Wald* statistic as

$$\check{W} = (R\widetilde{\theta} - r)'[R A_{IM} \check{V}_{IM} A_{IM} R']^{-1} (R\widetilde{\theta} - r),$$

where $\check{V}_{IM}$ is either $\widehat{V}_{IM}$ using $\widehat{\sigma}^2_{u\cdot v}$, which defines $\widehat{W}$, or $\widetilde{V}_{IM}$ using $\widetilde{\sigma}^2_{u\cdot v}$, which defines $\widetilde{W}$. The asymptotic null distribution of these test statistics is given in Theorem 3.

Clearly, appealing to a consistency result for $\widehat{\sigma}^2_{u\cdot v}$ justifies standard inference procedures. As discussed earlier, referring to consistency properties of long run variance estimators, however, ignores the impact of bandwidth and kernel choices. In order to capture the effects of these choices fixed-b asymptotic theory needs to be developed. Given the form of the test statistics, and in particular the form of $\widehat{V}_{IM}$ and $\widetilde{V}_{IM}$, what is required is that the estimator of $\sigma^2_{u\cdot v}$ has a fixed-b limit that is proportional to $\sigma^2_{u\cdot v}$ (in order for the long run variance to be scaled out in the test statistics), independent of $\widetilde{\theta}$, and does not depend upon additional nuisance parameters. In the case where a long run variance estimator has such properties, the resulting *Wald* statistics have pivotal asymptotic distributions that only depend upon the kernel and bandwidth (as well as the number of integrated regressors and the deterministic components) and can thus be tabulated.

It follows from Theorem 1 that the fixed-b limit of $\widehat{\sigma}^2_{u\cdot v}$ does not fulfill the stated requirements, because it is not proportional to $\sigma^2_{u\cdot v}$ and it also depends upon nuisance parameters in a rather complicated fashion (see again the result for the fixed-b limit of $\widehat{\Omega}$ in Theorem 1). As follows from the results of Lemma 2, the fixed-b limit of $\widetilde{\sigma}^2_{u\cdot v}$ is proportional to $\sigma^2_{u\cdot v}$ and does not otherwise depend on nuisance parameters. However, the limit is correlated with the limit of $\widetilde{\theta}$, with this correlation itself being a complicated function of unknown nuisance parameters. Thus, under fixed-b asymptotics, *Wald* statistics using $\widetilde{\theta}$ and $\widehat{\sigma}^2_{u\cdot v}$ or $\widetilde{\sigma}^2_{u\cdot v}$ do not have asymptotically pivotal distributions. This presents a new hurdle in cointegrating regressions for fixed-b inference that does not arise in stationary regression settings, because the usual OLS residuals cannot directly be used to form the basis for fixed-b inference.

One solution is to use the adjusted residuals, $\widetilde{S}^{u*}_t$, already defined in (24) in Section 3 to construct an estimator of $\sigma^2_{u\cdot v}$ defined as

$$\widetilde{\sigma}^{2*}_{u\cdot v} = T^{-1} \sum_{i=2}^{T} \sum_{j=2}^{T} k\left(\frac{|i-j|}{M}\right) \Delta\widetilde{S}^{u*}_i \Delta\widetilde{S}^{u*}_j.$$

As we show, this estimator has the required properties to deliver a pivotal fixed-b limit for *Wald* statistics. This leads to a third estimator of V_{IM} given by

$$\widetilde{V}^*_{IM} = \widetilde{\sigma}^{2*}_{u\cdot v} A^{-1}_{IM} \left(S^{\widetilde{x}\prime} S^{\widetilde{x}}\right)^{-1} (C'C) \left(S^{\widetilde{x}\prime} S^{\widetilde{x}}\right)^{-1} A^{-1}_{IM}.$$

When using $\widetilde{\sigma}^{2*}_{u\cdot v}$ and where thus $\check{V}_{IM}$ is given by $\widetilde{V}^*_{IM}$, we denote the *Wald* statistic by $\widetilde{W}^*$.

The following lemma characterizes the asymptotic behavior of the partial sum processes of $\Delta\widetilde{S}^u_t$ and $\Delta\widetilde{S}^{u*}_t$ which is needed to subsequently obtain the fixed-b limits of the *Wald* statistics.

Lemma 2. *Let $\widetilde{S}^u_t$ and $\widetilde{S}^{u*}_t$ denote residuals as given in (22) and (24). The asymptotic behavior of the corresponding partial sum processes is given by*

$$\begin{aligned}&T^{-1/2} \sum_{t=2}^{[rT]} \Delta\widetilde{S}^u_t \Rightarrow \\ &\sigma_{u\cdot v} \left[\int_0^r dw_{u\cdot v}(s) - g(r)' \left(\int_0^1 g(s) g(s)' ds \right)^{-1} \right. \\ &\left. \times \int_0^1 (G(1) - G(s))\, dw_{u\cdot v}(s) \right] = \sigma_{u\cdot v} \widetilde{P}(r), \end{aligned} \tag{29}$$

$$\begin{aligned}&T^{-1/2} \sum_{t=2}^{[rT]} \Delta\widetilde{S}^{u*}_t \Rightarrow \\ &\sigma_{u\cdot v} \left[\int_0^r dw_{u\cdot v}(s) - h(r)' \left(\int_0^1 h(s) h(s)' ds \right)^{-1} \right. \\ &\left. \times \int_0^1 (H(1) - H(s))\, dw_{u\cdot v}(s) \right] = \sigma_{u\cdot v} \widetilde{P}^*(r), \end{aligned} \tag{30}$$

where

$$h(r)' = \left[g(r)', \int_0^r (G(1) - G(s))'\, ds \right], \qquad H(r) = \int_0^r h(s) ds.$$

Furthermore, conditional upon $W_v(r)$ it holds that Ψ, the scaled and centered limit of $\widetilde{\theta}$, is correlated with $\widetilde{P}(r)$ but uncorrelated with $\widetilde{P}^*(r)$.

The correlation between $\widetilde{P}(r)$ and Ψ implies that the fixed-b limit of $\widetilde{\sigma}^2_{u\cdot v}$ is correlated with Ψ and this correlation depends on nuisance parameters through Π. The important result of Lemma 2 is, however, that the random process $\widetilde{P}^*(r)$, defined in (30), is conditional upon $W_v(r)$ uncorrelated with Ψ. Given that both quantities are conditionally Gaussian implies that they are independent of each other. Therefore the fixed-b limit of $\widetilde{\sigma}^{2*}_{u\cdot v}$ is independent of Ψ and pivotal fixed-b inference can be performed.

The asymptotic behavior of the three *Wald* statistics under the null hypothesis is given by Theorem 3. Standard asymptotic results based on traditional bandwidth and kernel assumptions (as detailed in Jansson, 2002) are given for $\widehat{W}$ and $\widetilde{W}$ whereas a fixed-b result is given for $\widetilde{W}^*$.[2]

Theorem 3. *Assume that the data are generated by (1) and (2), that the FCLT (3) holds, that the deterministic components satisfy (14) and that R satisfies (28). Suppose that the bandwidth, M, and kernel, $k(\cdot)$, satisfy conditions such that $\widehat{\sigma}^2_{u\cdot v}$ is consistent. Then as $T \to \infty$*

$$\widehat{W} \Rightarrow \chi^2_q,$$

[2] Note for completeness that we have also worked out the fixed-b limits of $\widehat{W}$ and $\widetilde{W}$. Since these limits are not pivotal, they cannot be used to generate critical values. For this reason we do not report the expressions for these fixed-b limits here.

where χ_q^2 is a chi-square random variable with q degrees of freedom. When $q = 1$,

$$\widehat{t} \Rightarrow Z,$$

where $\widehat{t}$ is the t-statistic version of $\widehat{W}$ and Z is distributed standard normal.

Consider the same assumptions concerning the bandwidth and kernel as before, then as $T \to \infty$

$$\widetilde{\sigma}_{u\cdot v}^2 \Rightarrow \sigma_{u\cdot v}^2(1 + d_\gamma' d_\gamma),$$

with d_γ denoting the last k components of $\left(\int g(s)g(s)'ds\right)^{-1}\int[G(1) - G(s)]dw_{u\cdot v}$. Consequently, it follows that

$$\widetilde{W} \Rightarrow \frac{\chi_q^2}{1 + d_\gamma' d_\gamma},$$

where χ_q^2 is a chi-square random variable with q degrees of freedom that is correlated with d_γ. When $q = 1$,

$$\widetilde{t} \Rightarrow \frac{Z}{\sqrt{1 + d_\gamma' d_\gamma}},$$

where $\widetilde{t}$ is the t-statistic version of $\widetilde{W}$; Z is distributed standard normal and is correlated with d_γ.

If $M = bT$, where $b \in (0, 1]$ is held fixed as $T \to \infty$, then as $T \to \infty$

$$\widetilde{W}^* \Rightarrow \frac{\chi_q^2}{Q_b(\widetilde{P}^*, \widetilde{P}^*)},$$

where χ_q^2 is a chi-square random variable with q degrees freedom independent of $Q_b(\widetilde{P}^*, \widetilde{P}^*)$. When $q = 1$,

$$\widetilde{t}^* = \frac{R\widetilde{\theta} - r}{\sqrt{RA_{IM}\widetilde{V}_{IM}^* A_{IM} R'}} \Rightarrow \frac{Z}{\sqrt{Q_b(\widetilde{P}^*, \widetilde{P}^*)}},$$

where Z is distributed standard normal independent of $Q_b(\widetilde{P}^*, \widetilde{P}^*)$.

When appealing to consistency of $\widehat{\sigma}_{u\cdot v}^2$, inference using $\widehat{W}$ is standard. In contrast $\widetilde{\sigma}_{u\cdot v}^2$ is inconsistent under traditional bandwidth assumptions. But because $d_\gamma' d_\gamma > 0$, $\widetilde{\sigma}_{u\cdot v}^2$ is upwardly biased and the critical values of $\widetilde{W}$ are smaller than those of the χ_q^2 distribution. Therefore, using χ_q^2 critical values for $\widetilde{W}$ leads to a conservative test under traditional bandwidth assumptions. The fixed-*b* limiting distribution of $\widetilde{W}^*$ is similar to what is obtained for *Wald* statistics in stationary regression settings except that the form of $Q_b(\widetilde{P}^*, \widetilde{P}^*)$ is more complicated in the cointegration case. In addition to dependence upon f_t, a feature that occurs also in stationary regressions, the process $Q_b(\widetilde{P}^*, \widetilde{P}^*)$ depends also upon $W_v(r)$, i.e. upon the number of integrated regressors included in the cointegrating regression.[3] Thus, critical values need to be simulated taking into account the specifications of the deterministic components, the number of integrated regressors, the kernel function and the bandwidth choice. In Table 4 we tabulate critical values for the t-statistic for the parameter associated with x_t in a model with an intercept and one x_t regressor only. We provide critical values for the Bartlett and QS kernels and a grid of bandwidths indexed by b. In a supplementary Appendix we provide critical values for a selection of kernels and bandwidths for models with up to 4 integrated regressors and deterministic components consisting of intercept only and intercept plus linear trend.

[3] A similar finding was made by Bunzel (2006) for fixed-*b* inference using DOLS and by Jin et al. (2006) for partial fixed-*b* inference using FM-OLS.

6. Finite sample performance of test statistics

In this section we provide some finite sample results concerning the tests' performance using the simulation design from Section 4. Throughout this section we only report results for cases where $\rho_1 = \rho_2$. We report results for t-statistics for testing the null hypothesis $H_0 : \beta_1 = 1$ and *Wald* statistics for testing the joint null hypothesis $H_0 : \beta_1 = 1, \beta_2 = 1$. The OLS statistics were implemented without taking into account serial correlation in the regression error and serve as a benchmark. The FM-OLS statistics were implemented using $\widehat{\sigma}_{u^+}^2$. The IM-OLS statistics were implemented in three ways: The first uses $\widehat{\sigma}_{u\cdot v}^2$ and is labeled IM(O), the second uses $\widetilde{\sigma}_{u\cdot v}^2$ and is labeled IM(D) and the third uses $\widetilde{\sigma}_{u\cdot v}^{2*}$ and is labeled IM(Fb). We report results for both the Bartlett and the Quadratic Spectral (QS) kernels. With respect to bandwidth choice the FM and IM statistics are implemented in two ways. The first way uses the data dependent bandwidth rule of Andrews (1991). The second way uses a specific bandwidth, M, over the grid $M = 1, 2, \ldots, T$. This grid is indexed by the bandwidth to sample size ratio, $b = M/T$. As in Section 4, again DOLS is included with the leads and lags chosen as described and the bandwidth for the long run variance estimation is chosen according to Andrews (1991). Rejections for the OLS, IM(O) and IM(D) statistics are carried out using $N(0, 1)$ critical values for all values of M. Given the results of Theorem 3 this implies that the test statistic IM(D) is asymptotically conservative under traditional asymptotic theory. In contrast, rejections for DOLS, FM(Fb) and IM(Fb) are carried out using fixed-*b* asymptotic critical values. For each value of b, i.e. given M/T, asymptotic critical values were simulated for the DOLS, FM(Fb) and IM(Fb) statistics using the limiting random variables given by Bunzel (2006) (DOLS), Jin et al. (2006) (FM(Fb)) and our Theorem 3 (IM(Fb)). The critical values for DOLS, FM(Fb) and IM(Fb) depend on both kernel and bandwidth. The empirical rejection probabilities were computed using 5000 replications, and the nominal level is 0.05 in all cases.

Tables 2 and 3 report empirical null rejection probabilities using data dependent bandwidth choices for the Bartlett and the QS kernel. Table 2 contains the results for the t-tests and Table 3 contains the results for the *Wald* tests. In both tables Panel A corresponds to $T = 100$ and Panel B to $T = 200$. We only briefly summarize some main findings related to both tables. When $\rho_1, \rho_2 = 0$, as expected OLS tests work well with rejections close to 0.05 and, as also expected, increasing the values of ρ_1, ρ_2 leads to very large over-rejections. For $\rho_1, \rho_2 = 0$ the IM(Fb) test has rejections that tend to be below 0.05 whilst the other tests show some over-rejections. DOLS tests exhibit substantial over-rejections when $T = 100$, even in the case of no serial correlation and no endogeneity. For $T = 200$ the over-rejection problems of DOLS are substantially less severe. The IM(O) and IM(D) tests also show some over-rejections, which are however less severe than for FM(Fb). IM(D) usually has slightly lower rejection rates than IM(O) and this is as expected, given the conservative nature of the IM(D) test under standard asymptotics. With increasing values of ρ_1, ρ_2, all tests' over-rejection problems become more pronounced. Overall, the test least affected by the over-rejection problem, when using the Andrews (1991) data dependent bandwidth, is the IM(Fb) test, which only suffers from large over-rejections (sometimes larger than IM(O), IM(D) and FM(Fb)) when $\rho_1, \rho_2 = 0.9$.

Some of these over-rejection problems in the $\rho_1, \rho_2 = 0.9$ case can be attributed to the bandwidths chosen by the Andrews method. For example, for IM(Fb) the average values of b across the replications are .0648 ($T = 100$) and .0515 ($T = 200$) for the Bartlett kernel. The corresponding average values of b for $\widehat{\sigma}_{u^+}^2$ used by FM(Fb) are .0796 and .0473. The average b values for the QS kernel are slightly smaller in all cases. These are relatively small bandwidths and it is well known in the fixed-*b* literature that

Table 2
Empirical null rejection probabilities, 0.05 level, t-tests for $h_0 : \beta_1 = 1$ data dependent bandwidths and lag lengths.

ρ_1, ρ_2	OLS	Bartlett kernel					QS kernel				
		DOLS	FM	IM(O)	IM(D)	IM(Fb)	DOLS	FM	IM(O)	IM(D)	IM(Fb)
Panel A: $T = 100$											
0.0	.0544	.2408	.0708	.0802	.0736	.0570	.2298	.0904	.0926	.0856	.0450
0.3	.1608	.3432	.0960	.1038	.1004	.0652	.3488	.0986	.1020	.0986	.0836
0.6	.4126	.5196	.1774	.1444	.1518	.1198	.5030	.1544	.1284	.1378	.0556
0.9	.7700	.6816	.5142	.4412	.4256	.5480	.6624	.4464	.4194	.4082	.3936
Panel B: $T = 200$											
0.0	.0484	.0604	.0700	.0722	.0628	.0392	.0600	.0720	.0766	.0672	.0324
0.3	.1592	.1094	.0880	.0892	.0812	.0776	.0970	.0808	.0816	.0736	.0582
0.6	.4204	.1856	.1554	.1092	.1070	.0920	.1508	.1320	.0964	.0920	.0552
0.9	.7712	.3178	.4752	.3212	.2942	.4280	.2530	.4536	.3066	.2780	.4564

Table 3
Empirical null rejection probabilities, 0.05 level, *Wald* tests for $H_0 : \beta_1 = 1, \beta_2 = 1$ data dependent bandwidths and lag lengths.

ρ_1, ρ_2	OLS	Bartlett kernel					QS kernel				
		DOLS	FM	IM(O)	IM(D)	IM(Fb)	DOLS	FM	IM(O)	IM(D)	IM(Fb)
Panel A: $T = 100$											
0.0	.0578	.3144	.0708	.0972	.0890	.0612	.3028	.0980	.1172	.1074	.0426
0.3	.2158	.4566	.1144	.1330	.1258	.0692	.4650	.1126	.1300	.1256	.1024
0.6	.5772	.7052	.2284	.1970	.2070	.1538	.6816	.1888	.1702	.1900	.0568
0.9	.9372	.8812	.7048	.6378	.6114	.7498	.8680	.6076	.6080	.5856	.5418
Panel B: $T = 200$											
0.0	.0512	.0654	.0776	.0748	.0666	.0356	.0598	.0782	.0808	.0724	.0248
0.3	.2028	.1320	.1030	.1004	.0920	.0846	.1116	.0944	.0906	.0832	.0608
0.6	.5752	.2324	.1834	.1324	.1314	.1074	.1892	.1556	.1130	.1140	.0540
0.9	.9432	.4472	.6492	.4524	.4140	.5950	.3482	.6206	.4330	.3920	.6276

small bandwidths lead to larger over-rejection problems than bigger bandwidths when serial correlation is strong.

In order to illustrate the role that the bandwidth and kernel choices play in the over-rejection problem we plot in Figs. 1–5 null rejection probabilities of the t-tests as a function of $b \in (0, 1]$. The first two figures show the results for the Bartlett kernel for $T = 100$ for increasing values of ρ_1, ρ_2. In Fig. 1, with $\rho_1, \rho_2 = 0.3$, for small bandwidths all tests have rejection probabilities close to 0.09 so there are some minor over-rejection problems. As the bandwidth increases, with the exception of FM(Fb) and IM(Fb), all rejection probabilities increase substantially. FM(Fb) shows some over-rejection problems that are initially increasing in b but then decline as b increases further. Rejections for IM(Fb) are close to 9% for all values of b indicating that the fixed-b approximation performs reasonably well for IM(Fb). In the working paper, Vogelsang and Wagner (2013b), rejections for IM(fb) are close to 5% for all values of b when there is no serial correlation or endogeneity ($\rho_1, \rho_2 = 0.0$).

As the values of ρ_1, ρ_2 increase to 0.9, we see in Fig. 2 that the rejections take a J-curve shape for IM(O) and IM(D) and over-rejection becomes a serious problem regardless of bandwidth for these tests with IM(O) having larger over-rejection problems than IM(D). The IM(Fb) test is now severely affected by over-rejections. Interestingly, in this case FM(Fb) has less of an over-rejection problem than IM(Fb) although all four tests are severely size distorted. If we increase T, the over-rejection problems of IM(Fb) become less problematic. Fig. 3 shows results for $T = 500$ whereas Fig. 4 shows results for $T = 1000$. Increasing T to 500 substantially reduces the over-rejections of IM(Fb) with rejections close to 12%–13% for non-small bandwidths. FM(Fb) continues to have substantial over-rejection problems. We see in Fig. 4 that increasing T further to 1000 yields rejections for IM(Fb) that are approximately 6%–7%. Again, FM(Fb) continues to exhibit nontrivial over-rejection problems. Figs. 3 and 4 clearly show that when serial correlation and endogeneity are strong, a larger sample size is needed for IM(Fb) to have reasonable size whereas the other statistics continue to have over-rejection problems.

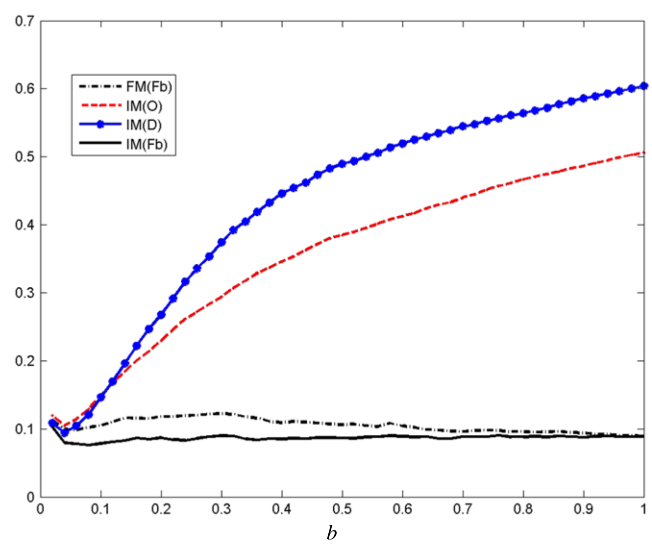

Fig. 1. Empirical null rejections, t-test, $T = 100$, $\rho_1 = \rho_2 = 0.3$, Bartlett kernel.

Fig. 5 provides some results for the QS kernel. For the sake of brevity we only report results for $\rho_1, \rho_2 = 0.9$ and for $T = 1000$. The working paper provides additional results for the QS kernel. Similar to the Bartlett kernel case, we see in Fig. 5 that IM(Fb) tends to have rejections close to 5% whereas FM(Fb) can over-reject. Not surprisingly, IM(O) and IM(D) show substantial over-rejection problems. If we compare Fig. 5 to 4, we see that IM(Fb) has less over-rejection problems when using the QS kernel compared to the Bartlett kernel. Clearly, the QS kernel leads to less size distorted tests than the Bartlett kernel.

The overall picture is that the IM(Fb) test is the most robust statistic in terms of over-rejection problems although for a given

Table 4
Fixed-b asymptotic critical values for t-test of β in regression with intercept and one regressor.

Panel A: Bartlett kernel

b	0.02	0.04	0.06	0.08	0.10	0.12	0.14	0.16	0.18	0.20
95%	1.6932	1.8255	1.9707	2.1396	2.3210	2.5175	2.7227	2.9298	3.1351	3.3396
97.5%	2.0285	2.2063	2.3979	2.6170	2.8474	3.0972	3.3545	3.6098	3.8663	4.1107
99%	2.4460	2.6685	2.9201	3.1932	3.4966	3.8131	4.1413	4.4832	4.8197	5.1422
99.5%	2.7455	2.9923	3.2870	3.6355	3.9600	4.3330	4.7332	5.1388	5.4687	5.8420
b	0.22	0.24	0.26	0.28	0.30	0.32	0.34	0.36	0.38	0.40
95%	3.5348	3.7171	3.8848	4.0360	4.1949	4.3322	4.4521	4.5517	4.6500	4.7492
97.5%	4.3616	4.5772	4.7941	4.9891	5.1695	5.3217	5.4714	5.6035	5.7264	5.8454
99%	5.4265	5.6797	5.9238	6.1835	6.4076	6.6029	6.7954	6.9543	7.1223	7.2694
99.5%	6.1353	6.4441	6.7049	6.9934	7.3124	7.5098	7.6841	7.9199	8.0373	8.1602
b	0.42	0.44	0.46	0.48	0.50	0.52	0.54	0.56	0.58	0.60
95%	4.8455	4.9227	4.9890	5.0645	5.1309	5.2114	5.2894	5.3637	5.4394	5.5033
97.5%	5.9590	6.0499	6.1588	6.2430	6.3396	6.4336	6.5389	6.6441	6.7322	6.8134
99%	7.3677	7.5084	7.5941	7.7294	7.8500	8.0009	8.1297	8.2313	8.3557	8.5135
99.5%	8.3116	8.4798	8.6175	8.7959	8.9384	9.1105	9.2329	9.4176	9.5586	9.7002
b	0.62	0.64	0.66	0.68	0.70	0.72	0.74	0.76	0.78	0.80
95%	5.5858	5.6636	5.7333	5.7873	5.8568	5.9161	5.9785	6.0525	6.1182	6.1826
97.5%	6.8984	7.0026	7.0860	7.1694	7.2608	7.3434	7.4407	7.5176	7.5911	7.6673
99%	8.6067	8.6957	8.8061	8.8890	9.0184	9.1032	9.2298	9.3602	9.4847	9.5579
99.5%	9.7874	9.9185	10.0433	10.2004	10.3577	10.4819	10.5370	10.6463	10.7998	10.8832
b	0.82	0.84	0.86	0.88	0.90	0.92	0.94	0.96	0.98	1.00
95%	6.2477	6.3023	6.3524	6.4244	6.4760	6.5296	6.5847	6.6304	6.6853	6.7365
97.5%	7.7403	7.8267	7.8905	7.9613	8.0327	8.0984	8.1740	8.2414	8.3106	8.3767
99%	9.6555	9.7774	9.8532	9.9286	10.0028	10.1165	10.2018	10.2871	10.3599	10.4436
99.5%	10.9947	11.1332	11.2467	11.3489	11.4627	11.5894	11.6811	11.7972	11.9234	12.0290

Panel B: Quadratic spectral kernel

b	0.02	0.04	0.06	0.08	0.10	0.12	0.14	0.16	0.18	0.20
95%	1.7338	1.9209	2.1699	2.4856	2.8936	3.4082	4.0151	4.8014	5.7308	6.8305
97.5%	2.0886	2.3448	2.6656	3.0945	3.6345	4.3537	5.2460	6.3794	7.7178	9.2666
99%	2.5234	2.8663	3.3071	3.8767	4.6549	5.7091	7.0310	8.6937	10.7070	13.1511
99.5%	2.8407	3.2338	3.8202	4.5328	5.5177	6.8401	8.4881	10.8076	13.4736	16.4481
b	0.22	0.24	0.26	0.28	0.30	0.32	0.34	0.36	0.38	0.40
95%	8.0140	9.2449	10.5013	11.7472	12.9608	14.0871	15.0642	16.0279	16.8018	17.5136
97.5%	11.0222	12.8987	14.7963	16.8181	18.7732	20.7165	22.4850	24.0161	25.5221	26.8964
99%	15.8087	18.8405	21.8897	25.3523	28.6677	31.5962	34.5779	37.4654	40.2910	43.0964
99.5%	19.9960	24.0370	28.0318	32.5506	36.7802	42.0681	46.7491	50.7974	55.0350	59.9235
b	0.42	0.44	0.46	0.48	0.50	0.52	0.54	0.56	0.58	0.60
95%	18.1475	18.7095	19.2702	19.7309	20.1433	20.5939	20.9625	21.3219	21.6411	21.9317
97.5%	28.1398	29.2209	30.1876	31.0764	31.9833	32.7286	33.3458	34.0529	34.6470	35.2694
99%	45.7145	47.7651	50.4107	52.2371	53.9473	55.5799	57.1530	58.7989	60.3068	61.2087
99.5%	63.5465	67.1811	70.8659	74.7957	77.7241	80.2765	82.1976	84.2852	86.3021	88.9058
b	0.62	0.64	0.66	0.68	0.70	0.72	0.74	0.76	0.78	0.80
95%	22.1648	22.4990	22.7303	22.9469	23.1772	23.4815	23.7279	23.9809	24.1831	24.4107
97.5%	35.8575	36.3833	36.7923	37.3439	37.7349	38.3346	38.8993	39.3871	39.9879	40.3649
99%	62.7834	63.5792	64.7326	65.6511	66.8285	68.1831	68.6392	69.7165	70.7337	72.4124
99.5%	91.1713	92.8712	95.3349	97.2121	98.6277	100.872	103.718	105.910	107.660	109.374
b	0.82	0.84	0.86	0.88	0.90	0.92	0.94	0.96	0.98	1.00
95%	24.6562	24.9178	25.1008	25.3928	25.5731	25.8380	26.0156	26.2197	26.3767	26.5555
97.5%	41.0117	41.5149	42.0449	42.7282	43.3055	43.7696	44.3047	44.7995	45.2353	45.7322
99%	73.8060	74.9712	76.3487	77.8025	78.8081	80.1633	81.2284	82.3518	83.6382	84.7245
99.5%	111.604	112.292	114.338	115.999	117.609	120.408	122.528	124.114	126.265	127.964

Note: Left tail critical values follow by symmetry around zero.

sample size, increasing the values of ρ_1, ρ_2 causes over-rejections to emerge. Larger bandwidths in conjunction with the QS kernel lead to test statistics with the least over-rejection problems. Similar over-rejection patterns have been observed by Kiefer and Vogelsang (2005) in stationary regression settings.

We now turn to the analysis of the power properties of the tests. For the sake of brevity we only display results for the case $\rho_1, \rho_2 = 0.6$ for the *Wald* test for $T = 100$ and using the QS kernel. Patterns are similar for other values of ρ_1, ρ_2, for the t-tests, for $T = 200$ and for the Bartlett kernel. Starting from the null values of β_1 and β_2 equal to 1, we consider under the alternative $\beta_1 = \beta_2 = \beta \in (1, 2]$, using (including the null value) a total of 21 values on a grid with mesh 0.05. We focus on size-corrected power because of the potential over-rejection problems under the null hypothesis. This allows us to see power differences across tests while holding null rejection probabilities constant at 0.05. Clearly, this is useful for theoretical power comparisons, but it has to be kept in mind that such size-corrections are not feasible in practice.

In Fig. 6 we depict the power of the FM(Fb) and IM *Wald* tests using the QS kernel with $b = 0.3$. Patterns are qualitatively similar

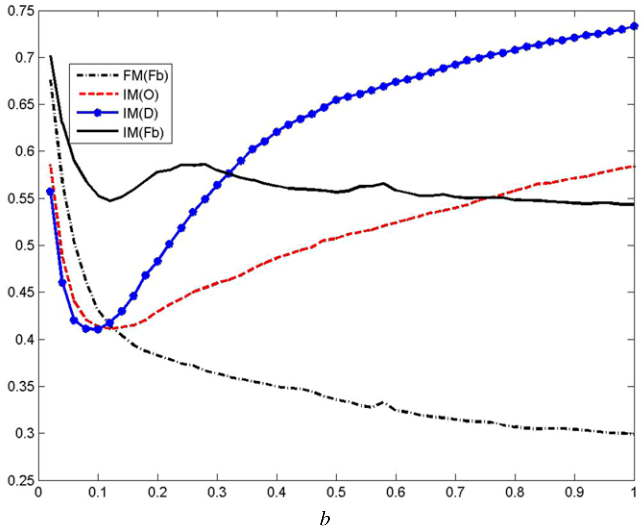

Fig. 2. Empirical null rejections, t-test, $T = 100$, $\rho_1 = \rho_2 = 0.9$, Bartlett kernel.

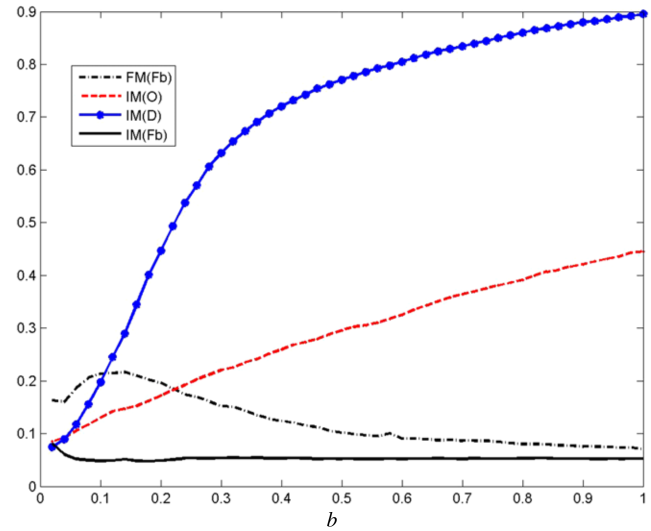

Fig. 5. Empirical null rejections, t-test, $T = 1000$, $\rho_1 = \rho_2 = 0.9$, QS kernel.

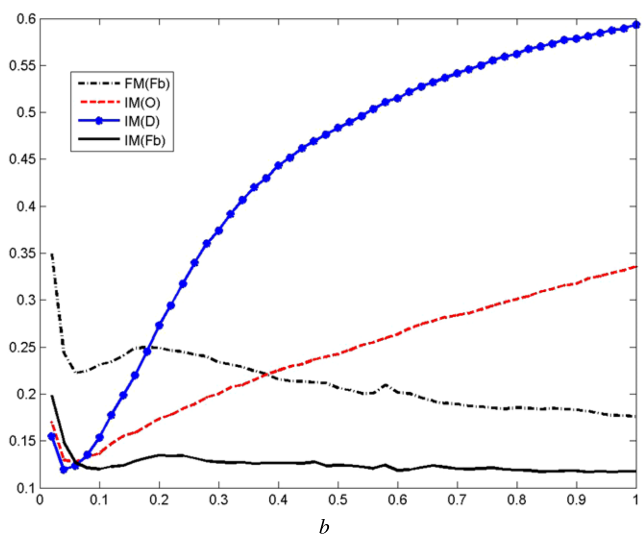

Fig. 3. Empirical null rejections, t-test, $T = 500$, $\rho_1 = \rho_2 = 0.9$, Bartlett kernel.

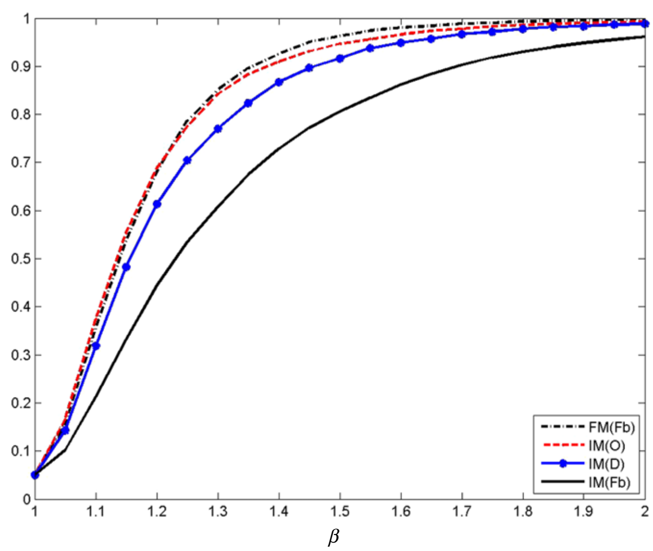

Fig. 6. Size adjusted power, *Wald* test, $T = 100$, $\rho_1 = \rho_2 = 0.6$, QS kernel, $b = 0.3$.

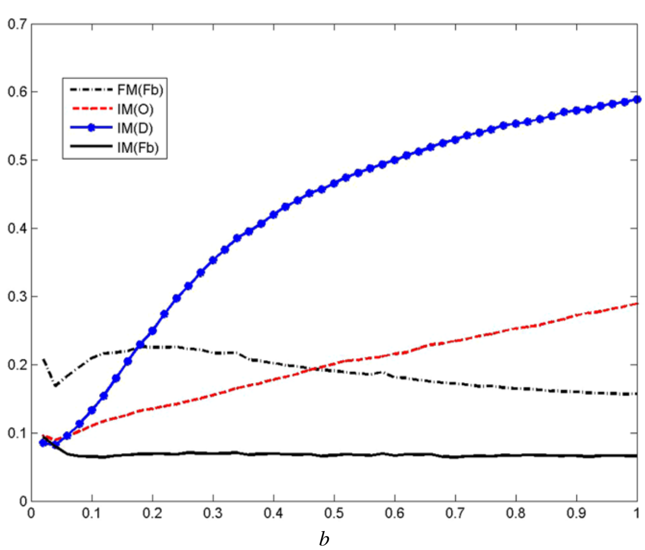

Fig. 4. Empirical null rejections, t-test, $T = 1000$, $\rho_1 = \rho_2 = 0.9$, Bartlett kernel.

for other values of b. The first thing to note is that size-corrected power of the FM(Fb) and IM(O) are similar. This suggests that partial summing before estimation (when using the IM-OLS estimator) implies only minimal power losses. The second thing to note is that IM(Fb) has the least power across the four tests. The use of $\tilde{\sigma}^{2*}_{u\cdot v}$ to obtain asymptotically fixed-b inference and less finite sample size distortions comes at the price of a small but nontrivial reduction in power.

Fig. 7 shows the impact of the bandwidth on the power of the IM(Fb) test by displaying the power curves for eight values of $b = 0.02, 0.06, 0.1, 0.3, 0.5, 0.7, 0.9, 1.0$. It is obvious that power depends on the bandwidth and tends to decrease when the bandwidth is increased, and power is highly sensitive to the bandwidth. In contrast, results reported in the working paper for the Bartlett kernel show power with the Bartlett kernel is much less sensitive to b although power does decrease as b increases. Similar patterns in power were found in stationary regression setting by Kiefer and Vogelsang (2005). That power decreases quickly with b for the QS kernel needs to be seen in conjunction with the observation made earlier that tests using the QS kernel suffer much less from over-rejection problems than those using the Bartlett kernel especially when large bandwidths are used. Thus, the price of robustness to over-rejections is lower power, or alternatively, higher power can be obtained at the cost of greater size distortions. A similar size-power trade-off with respect to kernel and bandwidth choice has been found in Kiefer and Vogelsang (2005) for stationary regressions and it is this trade-off that forms the basis of the data dependent bandwidth rule developed by Sun et al. (2008).

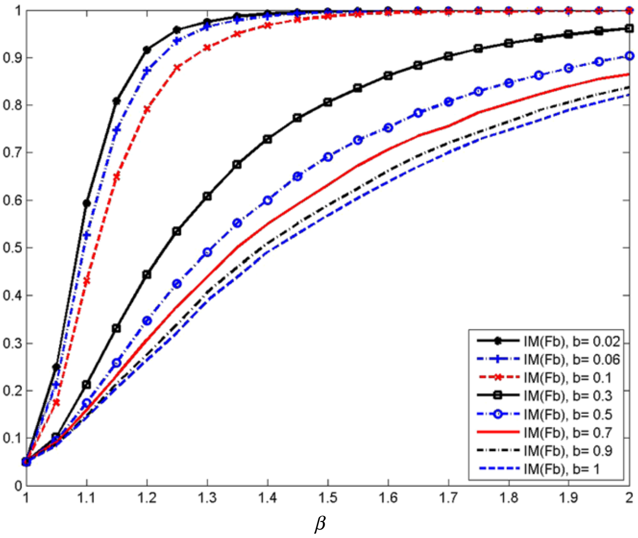

Fig. 7. Size adjusted power, *Wald* test, $T = 100$, $\rho_1 = \rho_2 = 0.6$, QS kernel.

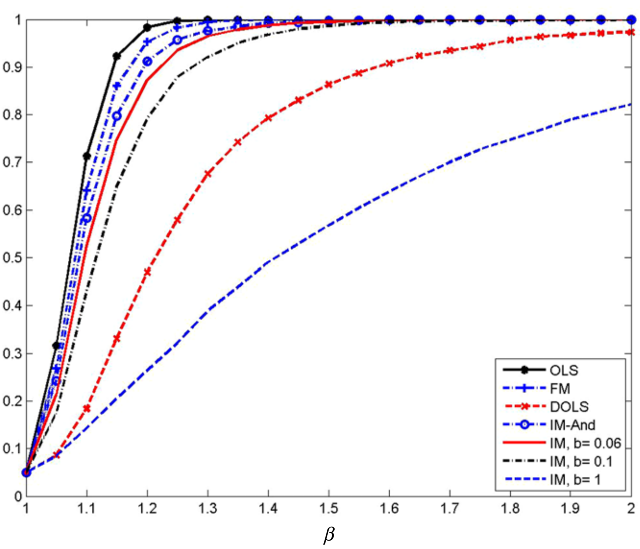

Fig. 8. Size adjusted power, *Wald* test, $T = 100$, $\rho_1 = \rho_2 = 0.6$, QS kernel.

Finally, Fig. 8 allows for power comparisons across the various tests (OLS, FM-OLS, DOLS, IM-OLS). In Fig. 8 power of the IM(Fb) test is shown for $b = 0.06, 0.1, 1.0$ and using the Andrews data dependent bandwidth. FM(Fb) is implemented using the Andrews bandwidth formula. The OLS and FM(Fb) tests have the highest size-corrected power, with the power of the IM(Fb) tests with $b = 0.06, 0.1$ being slightly lower and the power of the DOLS test being substantially lower. Power of the IM(Fb) test with $b = 1$ is lower than even the power of the DOLS test.

7. Summary and conclusions

The paper begins by deriving the fixed-b limit distribution of the FM-OLS estimator of Phillips and Hansen (1990). Fixed-b asymptotic theory has been developed in Kiefer and Vogelsang (2005) to capture the impact of bandwidth and kernel choices on tests in stationary HAC regressions, whose effects are not captured by standard asymptotic theory. Clearly, such choices in long run variance estimation are necessary when implementing the FM-OLS estimator. The fixed-b asymptotic distribution of the FM-OLS estimator features complicated dependencies upon bandwidth and kernel choices and is not asymptotically unbiased. The fixed-b limiting distribution shows that the accuracy of long run variance estimation is crucial for the properties of the FM-OLS estimator. We also show that the fixed-b limit distribution converges to the traditional limit distribution of the FM-OLS estimator when $b \to 0$.

The derivation of the fixed-b limit distribution of the FM-OLS estimator presents some hurdles for fixed-b theory. The first hurdle is that fixed-b theory has to be extended from the stationary regression framework to the world of unit roots and cointegration, an endeavor undertaken to a certain extent also in Bunzel (2006) and Jin et al. (2006). In cointegrating regressions fixed-b limits of long run variance estimators depend not only upon the deterministic components, but also upon the number of integrated regressors. The second hurdle is the need to derive the fixed-b limit of half long run variance matrix estimators which turn out to have complicated forms including additive nuisance parameters. This results in a fixed-b limit distribution of the FM-OLS estimator of very complicated form that does not lend itself to the construction of test statistics that are asymptotically free of nuisance parameters. Our fixed-b result for the FM-OLS estimator shows that the various long run and half long run variance estimators used to construct the FM-OLS estimator need to be close to their population values for FM-OLS to work in practice.

Consequently, we propose a new simple tuning parameter free estimator that is based on a partial sum transformed regression augmented by the original integrated regressors themselves, referred to as IM-OLS estimator. The advantage of the partial sum transformation is that it results in a zero mean mixed Gaussian limiting distribution without the need to choose any tuning parameter (like bandwidth, kernel or numbers of leads and lags). When the IM-OLS estimates are to be used for inference, a long run variance still needs to be estimated. Pivotal inference can be done in two ways. In a straightforward way one can use a consistent estimator of the required long run variance and this leads to tests having standard asymptotic distributions. Alternatively, fixed-b inference is possible for the IM-OLS estimator. The construction of fixed-b test statistics requires to further orthogonalize the IM-OLS residuals with respect to a set of specifically constructed additional regressors. These further modified residuals form the basis for pivotal fixed-b inference. Hence, critical values for the resultant fixed-b, t and *Wald* statistics can be tabulated. Similar to what Bunzel (2006) found for DOLS tests, these critical values depend upon the deterministic components included, the number of integrated regressors and, of course, the bandwidth as well as the kernel chosen.

The theoretical analysis is complemented by a simulation study, in which the performance of the new estimator and test statistics based upon it is compared with OLS, FM-OLS and DOLS. The IM-OLS estimator shows good performance in terms of both bias and RMSE. Typically, the bias of IM-OLS is smaller than the bias of FM-OLS and its RMSE is typically a bit larger than the RMSE of FM-OLS. The size and power analysis of the tests shows that the fixed-b approach is very useful also in the context of cointegrating regressions. It leads to test statistics that are more robust, in terms of having lower size distortions than all other test statistics, at the expense of only very minor power losses provided serial correlation and/or endogeneity is not too strong. When serial correlation and/or endogeneity is strong, the tests based on all estimators examined have severe null over-rejection problems although IM-OLS with the QS kernel and a large bandwidth has the least over-rejection problems in this case. Thus, in case of strong serial correlation or endogeneity the QS kernel is preferred over the Bartlett kernel.

Future research will study IM-OLS type estimators for panels of cointegrated time series, for higher order cointegrating regressions and for nonlinear cointegration relationships (that are linear in parameters). Furthermore, we will investigate whether and how the estimator $\widetilde{\gamma}$ can serve as a basis for endogeneity testing.

Acknowledgments

The research for this paper was initiated while Vogelsang was a visiting professor at the Institute for Advanced Studies in June

2009. The authors gratefully acknowledge financial support from the Jubiläumsfonds of the Oesterreichische Nationalbank under Grant Nos. 13398 and 15334. We thank Peter Robinson, an associate editor, and two anonymous referees for helpful and constructive comments. The comments of conference participants at the Econometrics Study Group in Rauischholzhausen, the 4th CIREQ Time Series Conference in Montreal, the 2011 NSF/NBER Time Series Conference at MSU, an ETSERN Meeting in Pamplona, the 37th Macromodels Conference in Pultusk, the 4th International Conference on Computational and Financial Econometrics in London, the 4th Italian Congress of Econometrics and Empirical Economics in Pisa, the Swiss Society of Economics and Statistics Meeting in Lucerne, the Statistical Week in Leipzig, the 66th Econometric Society European Meeting in Malaga, and of seminar participants at the Institute for Advanced Studies in Vienna, University of Ljubljana, Michigan State University, CORE Louvain, Erasmus University Rotterdam, University of Michigan, Emory University and Boston University are gratefully acknowledged.

Appendix. Proofs

Proof of Theorem 1. In line with the formulation in the main text we consider here the case with the intercept as the only deterministic component in the regression. Define the partial sum process $\widehat{S}_t = \sum_{j=1}^{t} \widehat{\eta}_j$. We start by establishing functional central limit theorems for $T^{-1/2}\widehat{S}_{[rT]}$ and $T^{-1}\sum_{t=2}^{T} \widehat{S}_{t-1}\widehat{\eta}_t'$. Consider first

$$\begin{aligned} T^{-1/2}\sum_{t=1}^{[rT]}\widehat{u}_t &= T^{-1/2}\sum_{t=1}^{[rT]}u_t - \frac{[rT]}{T}T^{1/2}\left(\widehat{\mu}-\mu\right) \\ &\quad - T^{-3/2}\sum_{t=1}^{[rT]}x_t'T(\widehat{\beta}-\beta) \\ &\Rightarrow \int_0^r dB_u(s) - r\int_0^r B_v^*(s)'ds\Theta. \end{aligned}$$

Using the definition of $\widehat{\eta}_t = [\widehat{u}_t, v_t']'$ and stacking now leads to

$$\begin{aligned} T^{-1/2}\widehat{S}_{[rT]} &= T^{-1/2}\sum_{t=1}^{[rT]}\widehat{\eta}_t \Rightarrow \begin{bmatrix} \int_0^r dB_u(s) - \int_0^r B_v^*(s)'ds\Theta \\ B_v(r) \end{bmatrix} \\ &= \begin{bmatrix} B_u(r) \\ B_v(r) \end{bmatrix} - \begin{bmatrix} \int_0^r B_v^*(s)'ds\Theta \\ 0 \end{bmatrix} = P_{\widehat{\eta}}(r). \end{aligned}$$

We define correspondingly

$$dP_{\widehat{\eta}}(r) = \begin{bmatrix} dB_u(r) \\ dB_v(r) \end{bmatrix} - \begin{bmatrix} B_v^*(r)'\Theta \\ 0 \end{bmatrix} dr.$$

Remark 1. For the IM-OLS estimator we consider more general deterministic components so it is useful for the sake of clarity to point out the obvious changes this implies for $\widehat{u}_t$ and its limit processes. In this more general case if we let Θ denote the limit of the OLS estimator in the cointegrating regression (13), i.e. the regression including f_t as deterministic components, then in the definitions of $P_{\widehat{\eta}}(r)$ and $dP_{\widehat{\eta}}(r)$ we would redefine $B_v^*(r)'$ as $B_v^*(r)' = \left[f(r)' \quad B_v(r)'\right]$.

Let us now return to the specific case considered for the OLS estimator, i.e. the intercept only case. Consider $T^{-1}\sum_{t=2}^{T}\widehat{S}_{t-1}\widehat{\eta}_t'$, using

$$\begin{aligned} \widehat{\eta}_t = \begin{bmatrix} \widehat{u}_t \\ v_t \end{bmatrix} &= \begin{bmatrix} u_t \\ v_t \end{bmatrix} - \begin{bmatrix} (\widehat{\mu}-\mu) + x_t'(\widehat{\beta}-\beta) \\ 0 \end{bmatrix} \\ &= \eta_t - \lambda_t \end{aligned}$$

and thus for the partial sums $\widehat{S}_t = S_t^{\eta} - S_t^{\lambda}$. By assumption it holds that $T^{-1/2}S_{[rT]}^{\eta} \Rightarrow B(r) = \Omega^{1/2}W(r)$ and using the results for the limits of the OLS estimators we have

$$T^{-1/2}S_{[rT]}^{\lambda} \Rightarrow \begin{bmatrix} \int_0^r B_v^*(s)'ds\Theta \\ 0 \end{bmatrix}.$$

Now consider

$$\begin{aligned} T^{-1}\sum_{t=2}^{T}\widehat{S}_{t-1}\widehat{\eta}_t' &= T^{-1}\sum_{t=2}^{T}\left(S_{t-1}^{\eta} - S_{t-1}^{\lambda}\right)(\eta_t - \lambda_t)' \\ &= T^{-1}\sum_{t=2}^{T}S_{t-1}^{\eta}\eta_t' - T^{-1}\sum_{t=2}^{T}S_{t-1}^{\eta}\lambda_t' \\ &\quad - T^{-1}\sum_{t=2}^{T}S_{t-1}^{\lambda}\eta_t' + T^{-1}\sum_{t=2}^{T}S_{t-1}^{\lambda}\lambda_t'. \end{aligned} \tag{31}$$

We consider each of the four terms in (31) in turn. Under the stated assumption (3) it holds that

$$T^{-1}\sum_{t=2}^{T}S_{t-1}^{\eta}\eta_t' \Rightarrow \int B(r)dB(r) + \Lambda.$$

For the second term in (31) we get

$$\begin{aligned} T^{-1}\sum_{t=2}^{T}S_{t-1}^{\eta}\lambda_t' &= T^{-1}\sum_{t=2}^{T}S_{t-1}^{\eta}\left[(\widehat{\mu}-\mu) + x_t'(\widehat{\beta}-\beta),\ 0\right] \\ &= \Bigg[T^{-1}\sum_{t=2}^{T}T^{-1/2}S_{t-1}^{\eta}T^{1/2}\left(\widehat{\mu}-\mu\right) \\ &\quad + T^{-1}\sum_{t=2}^{T}T^{-1/2}S_{t-1}^{\eta}T^{-1/2}x_t'T(\widehat{\beta}-\beta),\ 0\Bigg] \\ &\Rightarrow \left[\int B(r)B_v^*(r)'dr\Theta,\ 0\right]. \end{aligned} \tag{32}$$

The third term in (31) can be rewritten as

$$\begin{aligned} T^{-1}\sum_{t=2}^{T}S_{t-1}^{\lambda}\eta_t' &= T^{-1}\sum_{t=2}^{T-1}\left(S_{t-1}^{\lambda} - S_t^{\lambda}\right)S_t^{\eta\prime} \\ &\quad + T^{-1}S_{T-1}^{\lambda}S_T^{\eta\prime} - T^{-1}\lambda_1\eta_1' \\ &= T^{-1}S_{T-1}^{\lambda}S_T^{\eta\prime} - T^{-1}\sum_{t=2}^{T-1}\lambda_t S_t^{\eta\prime} - T^{-1}\lambda_1\eta_1'. \end{aligned} \tag{33}$$

For the first term in (33) it holds that

$$T^{-1}S_{T-1}^{\lambda}S_T^{\eta\prime} \Rightarrow \begin{bmatrix} \int B_v^*(r)'dr\Theta \\ 0 \end{bmatrix} B(1)',$$

for the second term in (33) it can be shown that it has up to transposition the same limit as given in (32) and the third term converges to 0. Combining this we get

$$T^{-1}\sum_{t=2}^{T}S_{t-1}^{\lambda}\eta_t' \Rightarrow \begin{bmatrix} \int B_v^*(r)'dr\Theta \\ 0 \end{bmatrix} B(1)' - \begin{bmatrix} \Theta'\int B_v^*(r)B(r)'dr \\ 0 \end{bmatrix}.$$

It remains to consider the fourth term in (31)

$$\begin{aligned} &T^{-1}\sum_{t=2}^{T}S_{t-1}^{\lambda}\lambda_t' \\ &= T^{-1}\sum_{t=2}^{T}\begin{bmatrix} (t-1)\left(\widehat{\mu}-\mu\right) + S_{t-1}^{x\prime}(\widehat{\beta}-\beta) \\ 0 \end{bmatrix}\begin{bmatrix} (\widehat{\mu}-\mu) + x_t'(\widehat{\beta}-\beta) \\ 0 \end{bmatrix}' \end{aligned}$$

$$= \begin{bmatrix} T^{-1}\sum_{t=2}^{T}\left[(t-1)(\widehat{\mu}-\mu)+S_{t-1}^{x\prime}(\widehat{\beta}-\beta)\right]\left[(\widehat{\mu}-\mu)+(\widehat{\beta}-\beta)'x_t\right] & 0 \\ 0 & 0 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} \int\int_0^r B_v^*(s)'ds\Theta\Theta' B_v^*(r)dr & 0 \\ 0 & 0 \end{bmatrix}.$$

Combining the above results leads by appropriate rearranging of the terms to

$$T^{-1}\sum_{t=2}^{T}\widehat{S}_{t-1}\widehat{\eta}_t' \Rightarrow \int P_{\widehat{\eta}}(r)dP_{\widehat{\eta}}(r)' + \Lambda.$$

We now turn to $\widehat{\Delta}$ itself, using the shorthand notation $k_{ij} = k\left(\frac{|i-j|}{M}\right)$, suppressing the dependence upon M, given by

$$\widehat{\Delta} = T^{-1}\sum_{i=1}^{T}\sum_{j=i}^{T}k_{ij}\widehat{\eta}_i\widehat{\eta}_j' = \widehat{\Omega} - T^{-1}\sum_{i=2}^{T}\sum_{j=1}^{i-1}k_{ij}\widehat{\eta}_i\widehat{\eta}_j'.$$

Using $\widehat{S}_j = \sum_{s=1}^{j}\widehat{\eta}_s$, routine algebraic manipulations (see Vogelsang and Wagner (2013b)) give

$$\begin{aligned}
\widehat{\Delta} &= \widehat{\Omega} - T^{-1}\sum_{i=2}^{T-1}\sum_{j=1}^{i-2}\left[(k_{ij}-k_{i,j+1})-(k_{i+1,j}-k_{i+1,j+1})\right]\widehat{S}_i\widehat{S}_j' \\
&\quad + T^{-1}\sum_{i=2}^{T-1}\left(k_{i-1,i+1}-k_{i+1,i}\right)\widehat{S}_i\widehat{S}_{i-1}' \\
&\quad - T^{-1}\sum_{j=1}^{T-2}\left(k_{Tj}-k_{T,j+1}\right)\widehat{S}_T\widehat{S}_j' - T^{-1}\sum_{i=2}^{T}k_{i-1,i}\widehat{\eta}_i\widehat{S}_{i-1}' \\
&= \widehat{\Omega} - T^{-1}\sum_{i=2}^{T-1}\sum_{j=1}^{i-2}\left[(k_{ij}-k_{i,j+1})-(k_{i+1,j}-k_{i+1,j+1})\right]\widehat{S}_i\widehat{S}_j' \\
&\quad + \left[k\left(\frac{2}{M}\right)-k\left(\frac{1}{M}\right)\right]T^{-1}\sum_{i=2}^{T-1}\widehat{S}_i\widehat{S}_{i-1}' \\
&\quad - T^{-1}\sum_{j=1}^{T-2}\left(k_{Tj}-k_{T,j+1}\right)\widehat{S}_T\widehat{S}_j' - k\left(\frac{1}{M}\right)T^{-1}\sum_{i=2}^{T}\widehat{\eta}_i\widehat{S}_{i-1}', \qquad (34)
\end{aligned}$$

making the dependence upon M explicit again for several terms in the last line.

As is common in fixed-b asymptotic theory, compare Hashimzade and Vogelsang (2008), the limits depend upon the properties of the kernel function used. We first derive the result for twice differentiable kernels with $k(0) = 1$ and afterwards derive the result for the Bartlett kernel. The results follow by using the above derived limits and the asymptotic properties (under fixed-b limits, i.e. $M = bT$) of the kernel functions as developed in a univariate setting in Hashimzade and Vogelsang (2008). The result that $\widehat{\Omega} \Rightarrow Q_b\left(P_{\widehat{\eta}}, P_{\widehat{\eta}}\right)$ follows directly from algebraic expressions given by Hashimzade and Vogelsang (2008), extended in obvious ways to our multivariate setting, that allow to write $\widehat{\Omega}$ as a continuous function of $T^{-1/2}\widehat{S}_{[rT]}$ and the kernel. Also note that we use, as in the text, the same shorthand notation $Q_b(P_1, P_2)$ and $Q_b^{\Delta}(P_1, P_2)$ for both types of kernels.

For twice continuously differentiable kernels with $k(0) = 1$ we thus obtain from (34)

$$\begin{aligned}
\widehat{\Delta} &\Rightarrow Q_b\left(P_{\widehat{\eta}}(r), P_{\widehat{\eta}}(r)\right) + \frac{1}{b^2}\int\int_0^r k''\left(\frac{|r-s|}{b}\right)P_{\widehat{\eta}}(r)P_{\widehat{\eta}}(s)'dsdr \\
&\quad + \frac{1}{b}k_+'(0)\int P_{\widehat{\eta}}(s)P_{\widehat{\eta}}(s)'ds \\
&\quad - \frac{1}{b}\int k'\left(\frac{1-s}{b}\right)P_{\widehat{\eta}}(1)P_{\widehat{\eta}}(s)'ds - \int dP_{\widehat{\eta}}(s)P_{\widehat{\eta}}(s)' - \Lambda' \\
&= -\frac{1}{b^2}\int\int_r^1 k''\left(\frac{|r-s|}{b}\right)P_{\widehat{\eta}}(r)P_{\widehat{\eta}}(s)'dsdr \\
&\quad + \frac{1}{b}\int k'\left(\frac{1-s}{b}\right)P_{\widehat{\eta}}(s)P_{\widehat{\eta}}(1)'ds \\
&\quad + \frac{1}{b}k_+'(0)\int P_{\widehat{\eta}}(s)P_{\widehat{\eta}}(s)'ds + P_{\widehat{\eta}}(1)P_{\widehat{\eta}}(1)' \\
&\quad - \int dP_{\widehat{\eta}}(s)P_{\widehat{\eta}}(s)' - \Lambda' = Q_b^{\Delta}\left(P_{\widehat{\eta}}, P_{\widehat{\eta}}\right) - \Lambda'. \qquad (35)
\end{aligned}$$

For the Bartlett kernel we obtain similarly as above

$$\begin{aligned}
\widehat{\Delta} &\Rightarrow Q_b\left(P_{\widehat{\eta}}(r), P_{\widehat{\eta}}(r)\right) + \frac{1}{b}\int_0^{1-b}P_{\widehat{\eta}}(s+b)P_{\widehat{\eta}}(s)'ds \\
&\quad - \frac{1}{b}\int P_{\widehat{\eta}}(s)P_{\widehat{\eta}}(s)'ds \\
&\quad + \int_{1-b}^{1}P_{\widehat{\eta}}(1)P_{\widehat{\eta}}(s)'ds - \int dP_{\widehat{\eta}}(s)P_{\widehat{\eta}}(s)' - \Lambda' \\
&= \frac{1}{b}\int P_{\widehat{\eta}}(s)P_{\widehat{\eta}}(s)'ds - \frac{1}{b}\int_0^{1-b}P_{\widehat{\eta}}(s)P_{\widehat{\eta}}(s+b)'ds \\
&\quad - \frac{1}{b}\int_{1-b}^{1}P_{\widehat{\eta}}(s)P_{\widehat{\eta}}(1)'ds + P_{\widehat{\eta}}(1)P_{\widehat{\eta}}(1)' \\
&\quad - \int dP_{\widehat{\eta}}(s)P_{\widehat{\eta}}(s)' - \Lambda' = Q_b^{\Delta}\left(P_{\widehat{\eta}}, P_{\widehat{\eta}}\right) - \Lambda'. \qquad (36)
\end{aligned}$$

The results in (35) and (36) establish (11). The remaining claims of the theorem follow by simply inserting the corresponding submatrices of the fixed-b limits of $\widehat{\Omega}$ and $\widehat{\Delta}$ in the expressions for the FM-OLS estimator. In particular it holds that under fixed-b asymptotics

$$A\widetilde{X}'u^+ \Rightarrow \begin{pmatrix} \int dB_u(r) - \int dB_v(r)'Q_b(B_v, B_v)^{-1}Q_b(B_v, \widehat{B}_u) \\ \int B_v dB_u(r) + \Delta_{vu} - \left(\int B_v dB_v(r)' + \Delta_{vv}\right)Q_b(B_v, B_v)^{-1}Q_b(B_v, \widehat{B}_u) - Q_b^{+} \end{pmatrix},$$

with $Q_b^{+} = Q_b^{\Delta}(B_v, \widehat{B}_u) - \Lambda_{uv}' - \left(Q_b^{\Delta}(B_v, B_v) - \Lambda_{vv}'\right)Q_b(B_v, B_v)^{-1}Q_b(B_v, \widehat{B}_u)$ denoting the fixed-b limit of $\widehat{\Delta}_{vu}^{+}$. The result then follows by rearranging terms and using the definition of $B_{uv}^b(r)$.

Proof of Corollary 1. We start by observing that in $B_{uv}^b(r) = B_u(r) - B_v(r)Q_b(B_v, B_v)^{-1}Q_b(B_v, \widehat{B}_u)$, $\mathcal{B}_1$ and $\mathcal{B}_2$ the terms $Q_b(B_v, \widehat{B}_u)$ and $Q_b^{\Delta}(B_v, \widehat{B}_u)$ appear. Thus, consider

$$\begin{aligned}
\widehat{B}_u(r) &= B_u(r) - \int_0^r B_v^*(s)'ds\Theta \\
&= B_u(r) - \int_0^r B_v^*(s)'ds\left(\int_0^1 B_v^*(s)B_v^*(s)'ds\right)^{-1} \\
&\quad \times \left(\int_0^1 B_v^*(s)dB_u(s) + \Delta_{vu}^*\right) \\
&= B_u(r) - \int_0^r B_v^*(s)'ds\left(\int_0^1 B_v^*(s)B_v^*(s)'ds\right)^{-1} \\
&\quad \times \left(\int_0^1 B_v^*(s)dB_{u\cdot v}(s) + \int_0^1 B_v^*(s)dB_v(s)'\Omega_{vv}^{-1}\Omega_{vu} + \Delta_{vu}^*\right).
\end{aligned}$$

Upon regrouping terms we obtain

$$\widehat{B}_u(r) = \widetilde{B}_{u\cdot v}(r) + B_v(r)'\Omega_{vv}^{-1}\Omega_{vu} - F\left(B_v^*(r)\right),$$

with

$$\begin{aligned}
\widetilde{B}_{u\cdot v}(r) &= B_{u\cdot v}(r) - \int_0^r B_v^*(s)ds\left(\int_0^1 B_v^*(s)B_v^*(s)'ds\right)^{-1} \\
&\quad \times \int_0^1 B_v^*(s)dB_{u\cdot v}(s),
\end{aligned}$$

$$F\left(B_v^*(r)\right) = \int_0^r B_v^*(s)'ds \left(\int_0^1 B_v^*(s)B_v^*(s)'ds\right)^{-1} \times \left(\int_0^1 B_v^*(s)dB_v(s)'\Omega_{vv}^{-1}\Omega_{vu} + \Delta_{vu}^*\right),$$

as given already in the formulation of the corollary. This immediately leads to

$$\begin{aligned} Q_b(B_v, \widehat{B}_u) &= Q_b\left(B_v, \widetilde{B}_{u\cdot v} - F(B_v^*) + B_v'\Omega_{vv}^{-1}\Omega_{vu}\right) \\ &= Q_b(B_v, \widetilde{B}_{u\cdot v}) - Q_b\left(B_v, F(B_v^*)\right) + Q_b(B_v, B_v)\Omega_{vv}^{-1}\Omega_{vu}, \end{aligned}$$

and the same decomposition for $Q_b^\Delta(B_v, \widehat{B}_u)$. Given these decompositions we can write $\int_0^1 B_v^*(r)dB_{uv}^b(r)$, $\mathcal{B}_1$ and $\mathcal{B}_2$ in terms of

$$\begin{aligned} \int_0^1 B_v^*(r)dB_{uv}^b(r) &= \int_0^1 B_v^*(r)dB_{u\cdot v}(r) \\ &- \int_0^1 B_v^*(r)dB_v(r)'Q_b(B_v, B_v)^{-1}Q_b(B_v, \widetilde{B}_{u\cdot v}) \\ &+ \int_0^1 B_v^*(r)dB_v(r)'Q_b(B_v, B_v)^{-1}Q_b(B_v, F(B_v^*)), \end{aligned}$$

$$\begin{aligned} \mathcal{B}_1 &= \begin{pmatrix} 0 \\ \Delta_{vu} - \left(Q_b^\Delta(B_v, \widehat{B}_u) - \Lambda_{vu}'\right) \end{pmatrix} \\ &= \begin{pmatrix} 0 \\ \Omega_{vu} \end{pmatrix} - \begin{pmatrix} 0 \\ Q_b^\Delta(B_v, \widetilde{B}_{u\cdot v}) \end{pmatrix} \\ &+ \begin{pmatrix} 0 \\ Q_b^\Delta\left(B_v, F(B_v^*)\right) \end{pmatrix} - \begin{pmatrix} 0 \\ Q_b^\Delta(B_v, B_v)\Omega_{vv}^{-1}\Omega_{vu} \end{pmatrix}, \end{aligned}$$

$$\begin{aligned} \mathcal{B}_2 &= \begin{pmatrix} 0 \\ \left(\Omega_{vv} - Q_b^\Delta(B_v, B_v)\right) Q_b(B_v, B_v)^{-1}Q_b(B_v, \widehat{B}_u) \end{pmatrix} \\ &= \begin{pmatrix} 0 \\ \Omega_{vv}Q_b(B_v, B_v)^{-1}Q_b(B_v, \widetilde{B}_{u\cdot v}) \end{pmatrix} \\ &- \begin{pmatrix} 0 \\ \Omega_{vv}Q_b(B_v, B_v)^{-1}Q_b\left(B_v, F(B_v^*)\right) \end{pmatrix} + \begin{pmatrix} 0 \\ \Omega_{vu} \end{pmatrix} \\ &- \begin{pmatrix} 0 \\ Q_b^\Delta(B_v, B_v)Q_b(B_v, B_v)^{-1}Q_b(B_v, \widetilde{B}_{u\cdot v}) \end{pmatrix} \\ &+ \begin{pmatrix} 0 \\ Q_b^\Delta(B_v, B_v)Q_b(B_v, B_v)^{-1}Q_b\left(B_v, F(B_v^*)\right) \end{pmatrix} \\ &- \begin{pmatrix} 0 \\ Q_b^\Delta(B_v, B_v)\Omega_{vv}^{-1}\Omega_{vu} \end{pmatrix}. \end{aligned}$$

The result now follows by regrouping of terms into the term $\int B_v^*(r)dB_{u\cdot v}(r)$, the terms involving $\widetilde{B}_{u\cdot v}(r)$ and the terms involving $F(B_v^*(r))$, the latter two compressed by using the operator $\mathcal{F}(\cdot)$ as defined in the formulation of the corollary.

Proof of Proposition 1. We sketch the result for the Bartlett kernel only. Calculations are similar, but more cumbersome, for the other kernels considered in the paper. Using Corollary 1, the proposition is established by showing that

$$\operatorname*{plim}_{b\to 0} \mathcal{F}(\widetilde{B}_{u\cdot v}) = \operatorname*{plim}_{b\to 0} \mathcal{F}\left(F(B_v^*)\right) = 0.$$

Given the forms of $\mathcal{F}(\widetilde{B}_{u\cdot v})$ and $\mathcal{F}\left(F(B_v^*)\right)$ it is sufficient to show that

$$\operatorname*{plim}_{b\to 0} Q_b\left(B_v, \int_0^r B_v^*(s)ds\right) = 0,$$

$$\operatorname*{plim}_{b\to 0} Q_b^\Delta\left(B_v, \int_0^r B_v^*(s)ds\right) = 0,$$

$$\operatorname*{plim}_{b\to 0} Q_b(B_v, \widetilde{B}_{u\cdot v}) = 0, \qquad \operatorname*{plim}_{b\to 0} Q_b^\Delta(B_v, \widetilde{B}_{u\cdot v}) = 0.$$

Note that establishing the results for $Q_b(B_v, \widetilde{B}_{u\cdot v})$ and $Q_b^\Delta(B_v, \widetilde{B}_{u\cdot v})$ require showing that

$$\operatorname*{plim}_{b\to 0} Q_b\left(B_v, \int_0^r B_v^*(s)ds\right) = 0,$$

$$\operatorname*{plim}_{b\to 0} Q_b^\Delta\left(B_v, \int_0^r B_v^*(s)ds\right) = 0,$$

$$\operatorname*{plim}_{b\to 0} Q_b(B_v, B_{u\cdot v}) = 0, \qquad \operatorname*{plim}_{b\to 0} Q_b^\Delta(B_v, B_{u\cdot v}) = 0,$$

and so these four limits are what need to be shown. Showing that $\operatorname{plim}_{b\to 0} Q_b(B_v, B_{u\cdot v}) = 0$ is trivial. It is well known in the fixed-b literature that as $b \to 0$, fixed-b limiting random variables converge to the long run variance being estimated. The long run covariance between B_v and $B_{u\cdot v}$ is zero given that B_v and $B_{u\cdot v}$ are independent and so it follows that $\operatorname{plim}_{b\to 0} Q_b(B_v, B_{u\cdot v}) = 0$. The result for $Q_b^\Delta(B_v, B_{u\cdot v})$ follows naturally. For the other two limits it is sufficient to show that the mean and variance of the given quantities converge to zero as $b \to 0$. Explicit computations establishing the zero limits for the means are given in the working paper Vogelsang and Wagner (2013b) and computations for the variances are similar but more tedious.

Proof of Theorem 2. We consider the asymptotic behavior of the OLS estimator $\widetilde{\theta} = \left(\widetilde{\delta}', \widetilde{\beta}', \widetilde{\gamma}'\right)'$ in Eq. (21) and define $\theta = \left(\delta', \beta', \Omega_{vu}'\Omega_{vv}^{-1}\right)'$. It follows directly that

$$A_{IM}^{-1}\left(\widetilde{\theta} - \theta\right) = \left(T^{-2}A_{IM}S^{\widetilde{x}\prime}S^{\widetilde{x}}A_{IM}\right)^{-1}\left(T^{-2}A_{IM}S^{\widetilde{x}\prime}S^u\right) - (0, 0, \Omega_{vu}'\Omega_{vv}^{-1})' \tag{37}$$

using the notation of the main text. We consider the first two terms on the right hand side of (37) separately and start with the first one. In order to establish the limit for $T \to \infty$ we first consider $T^{-1/2}A_{IM}S_{[rT]}^{\widetilde{x}}$,

$$\begin{bmatrix} T^{-1}\tau_F^{-1}\sum_{t=1}^{[rT]} f_t \\ T^{-3/2}\sum_{t=1}^{[rT]} x_t \\ T^{-1/2}x_{[rT]} \end{bmatrix} \Rightarrow \begin{bmatrix} \int_0^r f(s)ds \\ \Omega_{vv}^{1/2}\int_0^r W_v(s)ds \\ \Omega_{vv}^{1/2}W_v(r) \end{bmatrix} = \Pi g(r).$$

This immediately implies that

$$\left(T^{-2}A_{IM}S^{\widetilde{x}\prime}S^{\widetilde{x}}A_{IM}\right)^{-1} \Rightarrow (\Pi')^{-1}\left(\int g(s)g(s)'ds\right)^{-1}\Pi^{-1}. \tag{38}$$

Analogously, for a typical entry of the second term in (37) it holds that

$$T^{-1/2}A_{IM}S_{[rT]}^{\widetilde{x}}T^{-1/2}S_{[rT]}^u \Rightarrow \Pi g(r)B_u(r)$$

and hence

$$\begin{aligned} T^{-2}A_{IM}S^{\widetilde{x}\prime}S^u &\Rightarrow \Pi\int g(r)B_u(r)dr \\ &= \sigma_{u\cdot v}\Pi\int g(r)w_{u\cdot v}dr + \Pi\int g(r)W_v(r)'dr\lambda_{uv}', \end{aligned}$$

using $B_u(r) = \sigma_{u\cdot v}w_{u\cdot v}(r) + \lambda_{uv}W_v(r)$.

Next note that $W_v(r)$ is the last block-component in $g(r)$, therefore

$$\begin{aligned} &(\Pi')^{-1}\left(\int g(r)g(r)'dr\right)^{-1}\int g(r)W_v(r)'dr\lambda_{uv}' \\ &= (\Pi')^{-1}\begin{pmatrix} 0 \\ 0 \\ I_k \end{pmatrix}\lambda_{uv}' = \begin{pmatrix} 0 \\ 0 \\ \left((\Omega_{vv}^{1/2})'\right)^{-1}\lambda_{uv}' \end{pmatrix}. \end{aligned} \tag{39}$$

From the definition of the respective quantities it follows that the non-zero block at the end of (39) is equal to $\Omega_{vv}^{-1}\Omega_{vu}$. Upon canceling $\Omega_{vv}^{-1}\Omega_{vu}$ this establishes the asymptotic behavior of the OLS estimator $\widetilde{\theta}$. The representation given in (25) then follows using integration by parts and the definition of $G(r)$.

Proof of Proposition 2. A detailed proof of the proposition is given in the working paper, Vogelsang and Wagner (2013b). The proof is essentially a proof of the Gauss–Markov theorem in continuous time.

Proof of Lemma 2. We next provide a proof of Lemma 2. To establish the result for $T^{-1/2}\sum_{t=2}^{[rT]}\Delta\widetilde{S}_t^{u*}$ it is convenient to note that using standard projection arguments it follows that the adjusted residuals, $\widetilde{S}_t^{u*}$, given by (24) are *exactly* the same as the OLS residuals from the regression

$$S_t^y = S_t^{f\prime}\delta + S_t^{x\prime}\beta + x_t'\gamma + z_t'\kappa + S_t^u. \tag{40}$$

The following lemma establishes the limit distribution of the OLS estimators from (40) which is needed to obtain the limit of $T^{-1/2}\sum_{t=2}^{[rT]}\Delta\widetilde{S}_t^{u*}$.

Lemma 1. *Consider regression (40) and denote by $\widetilde{\theta}^* = \left(\widetilde{\delta}^{*\prime}, \widetilde{\beta}^{*\prime}, \widetilde{\gamma}^{*\prime}, \widetilde{\kappa}^{*\prime}\right)'$ the vector of OLS estimators. Define the vector $\theta^* = \left(\delta', \beta', \Omega_{vu}'\Omega_{vv}^{-1}, 0\right)'$. It holds that*

$$\begin{aligned}A_M^{-1}\left(\widetilde{\theta}^* - \theta^*\right) &\Rightarrow \sigma_{u\cdot v}\left(\Pi_M'\right)^{-1}\left(\int h(s)h(s)'ds\right)^{-1}\\ &\quad\times\int h(r)w_{u\cdot v}(r)dr\\ &= \sigma_{u\cdot v}\left(\Pi_M'\right)^{-1}\left(\int h(s)h(s)'ds\right)^{-1}\\ &\quad\times\int [H(1)-H(s)]\,dw_{u\cdot v}(s),\end{aligned}$$

where $A_M = diag(A_{IM}, T^{-2}A_{IM})$, $\Pi_M = diag(\Pi, \Pi)$,

$$h(r) = \begin{pmatrix} g(r) \\ \int_0^r [G(1)-G(s)]ds \end{pmatrix}, \qquad H(r) = \int_0^r h(s)ds.$$

Proof. The proof builds on the results already obtained in Theorem 1 and essentially only the asymptotic behavior of the additional regressors and their cross-products with the error process has to be established. Recall that z_t is defined by (23) and partition z_t as $z_t = (z_t^{f\prime}, z_t^{S^x\prime}, z_t^{x\prime})'$. The limit, when appropriately scaled with $T^{-5/2}A_{IM}$, is given by

$$\begin{aligned}T^{-3}\tau_F^{-1}z_{[rT]}^f &= T^{-3}\tau_F^{-1}[rT]\sum_{t=1}^T S_t^f - T^{-3}\tau_F^{-1}\sum_{t=1}^{[rT]}\sum_{j=1}^t S_j^f\\ &= \frac{[rT]}{T}T^{-1}\sum_{t=1}^T T^{-1}\tau_F^{-1}S_t^f\\ &\quad - T^{-1}\sum_{t=1}^{[rT]}T^{-1}\sum_{j=1}^t T^{-1}\tau_F^{-1}S_j^f\\ &\to r\int\left(\int_0^s f(m)dm\right)ds\\ &\quad - \int_0^r\left(\int_0^s\left(\int_0^n f(m)dm\right)dn\right)ds,\end{aligned}$$

$$\begin{aligned}T^{-7/2}z_{[rT]}^{S^x} &= T^{-7/2}[rT]\sum_{t=1}^T S_t^x - T^{-7/2}\sum_{t=1}^{[rT]}\sum_{j=1}^t S_j^x\\ &= \frac{[rT]}{T}T^{-1}\sum_{t=1}^T T^{-3/2}S_t^x - T^{-1}\sum_{t=1}^{[rT]}T^{-1}\sum_{j=1}^t T^{-3/2}S_j^x\\ &\Rightarrow r\Omega_{vv}^{1/2}\int\left(\int_0^s W_v(m)dm\right)ds\\ &\quad - \Omega_{vv}^{1/2}\int_0^r\left(\int_0^s\left(\int_0^n W_v(m)dm\right)dn\right)ds,\end{aligned}$$

$$\begin{aligned}T^{-5/2}z_{[rT]}^x &= T^{-5/2}[rT]\sum_{t=1}^T x_t - T^{-5/2}\sum_{t=1}^{[rT]}\sum_{j=1}^t x_j\\ &= \frac{[rT]}{T}T^{-1}\sum_{t=1}^T T^{-1/2}x_t - T^{-1}\sum_{t=1}^{[rT]}T^{-1}\sum_{j=1}^t T^{-1/2}x_j\\ &\Rightarrow r\Omega_{vv}^{1/2}\int W_v(r)dr\\ &\quad - \Omega_{vv}^{1/2}\int_0^r\left(\int_0^s W_v(m)dm\right)ds.\end{aligned}$$

Combined, this can be written as

$$\begin{aligned}T^{-5/2}A_{IM}z_{[rt]} &\Rightarrow \Pi\left(r\int g(s)ds - \int_0^r\left(\int_0^s g(m)dm\right)ds\right)\\ &= \Pi\left(rG(1) - \int_0^r G(s)ds\right)\\ &= \Pi\left(\int_0^r [G(1)-G(s)]ds\right).\end{aligned}$$

For the cross-product of regressors and errors it holds that

$$T^{-5/2}A_{IM}z_{[rT]}T^{-1/2}S_{[rT]}^u \Rightarrow \Pi\left(\int_0^r [G(1)-G(s)]ds\right)B_u(r).$$

These preliminary results, combined with the results from Theorem 1 imply that

$$\begin{aligned}A_M^{-1}\left(\widetilde{\theta}^* - \theta^*\right) &\Rightarrow \left(\Pi_M'\right)^{-1}\left(\int h(s)h(s)'ds\right)^{-1}\\ &\quad\times\left(\int h(s)B_u(s)ds\right) - \begin{pmatrix}0\\0\\ \Omega_{vv}^{-1}\Omega_{vu}\\ 0\end{pmatrix}\\ &= \sigma_{u\cdot v}\left(\Pi_M'\right)^{-1}\left(\int h(s)h(s)'ds\right)^{-1}\\ &\quad\times\left(\int h(s)w_{u\cdot v}(s)ds\right)\\ &= \sigma_{u\cdot v}\left(\Pi_M'\right)^{-1}\left(\int h(s)h(s)'ds\right)^{-1}\\ &\quad\times\left(\int [H(1)-H(s)]dw_{u\cdot v}(s)\right),\end{aligned}$$

where the second line follows from the same argument as in Theorem 1 and the third line follows via integration by parts and the definition of $H(r)$. This completes the proof of Lemma 1.

We are now ready to prove Lemma 2. Consider the OLS residuals from (21),

$\widetilde{S}_t^u = S_t^y - S_t^{\widetilde{x}\prime}\widetilde{\theta} = S_t^u - x_t'\Omega_{vv}^{-1}\Omega_{vu} - S_t^{\widetilde{x}\prime}\left(\widetilde{\theta} - \theta\right)$

and their first differences,

$\Delta\widetilde{S}_t^u = u_t - \Delta x_t'\Omega_{vv}^{-1}\Omega_{vu} - \widetilde{x}_t'\left(\widetilde{\theta} - \theta\right)$.

Consequently,

$$T^{-1/2}\sum_{t=2}^{[rT]}\Delta\widetilde{S}_t^u = T^{-1/2}\sum_{t=2}^{[rT]}u_t - T^{-1/2}x_{[rT]}'\Omega_{vv}^{-1}\Omega_{vu}$$
$$-T^{-1/2}\sum_{t=2}^{[rT]}\widetilde{x}_t'A_{IM}A_{IM}^{-1}\left(\widetilde{\theta}-\theta\right)$$
$$\Rightarrow \sigma_{u\cdot v}w_{u\cdot v}(r) - g(r)'\Pi'\left\{\sigma_{u\cdot v}\left(\Pi'\right)^{-1}\left(\int g(s)g(s)'ds\right)^{-1}\right.$$
$$\left.\times\int[G(1)-G(s)]dw_{u\cdot v}(s)\right\}$$
$$=\sigma_{u\cdot v}\left[\int_0^r dw_{u\cdot v}(s) - g(r)'\left(\int g(s)g(s)'ds\right)^{-1}\right.$$
$$\left.\times\int[G(1)-G(s)]dw_{u\cdot v}(s)\right] = \sigma_{u\cdot v}\widetilde{P}(r),$$

where the limit follows from results already discussed in the proof of Theorem 2. In this respect note that $T^{-1/2}\sum_{t=2}^{[rT]}\widetilde{x}_t'A_{IM} = T^{-1/2}S_{[rT]}^{\widetilde{x}'}A_{IM} - T^{-1/2}\widetilde{x}_1'A_{IM}$, with the last term vanishing asymptotically.

Now consider the residuals from regression (40) which are, as discussed, exactly identical to the adjusted residuals given by (24), $\widetilde{S}_t^{u*} = S_t^y - S_t^{\widetilde{x}}\widetilde{\theta}^* = S_t^u - x_t'\Omega_{vv}^{-1}\Omega_{vu} - S_t^{\widetilde{x}'}\left(\widetilde{\theta}^*-\theta^*\right)$,

where here $S_t^{\widetilde{x}} = \left(S_t^{f\prime}, S_t^{x\prime}, x_t', z_t'\right)'$ and $\widetilde{\theta}^*$ and θ^* are as given in Lemma 1. The remaining steps are exactly the same as for $\Delta\widetilde{S}_t^u$ before, i.e. we get

$$T^{-1/2}\sum_{t=2}^{[rT]}\Delta\widetilde{S}_t^{u*} = T^{-1/2}\sum_{t=2}^{[rT]}u_t - \Delta x_{[rT]}'\Omega_{vv}^{-1}\Omega_{vu}$$
$$\times - T^{-1/2}\sum_{t=2}^{[rT]}\widetilde{x}_t'A_M A_M^{-1}\left(\widetilde{\theta}^*-\theta^*\right)$$
$$\Rightarrow \sigma_{u\cdot v}\left[\int_0^r dw_{u\cdot v}(s) - h(r)'\left(\int h(s)h(s)'ds\right)^{-1}\right.$$
$$\left.\times\int[H(1)-H(s)]dw_{u\cdot v}(s)\right] = \sigma_{u\cdot v}\widetilde{P}^*(r),$$

with $h(r)$ and $H(r)$ as defined in the formulation of the lemma.

To finish the proof of the lemma it remains to establish (conditional) independence of $\widetilde{P}^*(r)$ and the limiting distribution of $\widetilde{\theta}$. Conditional upon $W_v(r)$ the two quantities are Gaussian processes defined in terms of the Gaussian process $w_{u\cdot v}(r)$. Since they are conditionally Gaussian, conditional independence is established by showing that they are conditionally uncorrelated. With respect to $\widetilde{\theta}$ the relevant quantity is given by $\int[G(1)-G(s)]dw_{u\cdot v}(s)$, since the other components in the limiting distribution are non-random conditional upon $W_v(r)$. By definition of the quantities it holds that

$$\mathrm{Cov}\left(\widetilde{P}^*(r), \int[G(1)-G(s)]dw_{u\cdot v}(s)\right)$$
$$=\int_0^r[G(1)-G(s)]'ds - h(r)'\left(\int h(s)h(s)'\right)^{-1}$$
$$\times\int[H(1)-H(s)][G(1)-G(s)]'ds. \quad (41)$$

The first term is equal to (the transpose of) the second block of $h(r)$, $h_2(r)$ say, and the proof is completed by showing that also the second term in (41) is equal to $h_2(r)'$.

Using once again integration by parts it follows that

$$\int[H(1)-H(s)][G(1)-G(s)]'ds = \int h(s)h_2(s)'ds.$$

This in turn implies that the product of the two integrals is equal to $\begin{bmatrix}0\\I\end{bmatrix}$, which finally shows that the second term in (41) is indeed equal to $h_2(r)'$. Exactly analogous calculations as above in (41) show that correlation is present between $\widetilde{P}(r)$ and the limit distribution of $\widetilde{\theta}$.

Proof of Theorem 3. Use as in the main text as shorthand notation for the two *Wald* statistics considered $\check{W}$, with $\check{W} \in \{\widehat{W}, \widetilde{W}\}$. The test statistics only differ with respect to the used estimator of the long run variance parameter, $\widehat{\sigma}_{u\cdot v}^2$ or $\widetilde{\sigma}_{u\cdot v}^2$. As in the proof of Theorem 2, $\widetilde{\theta}$ denotes the vector of OLS estimators $\left(\widetilde{\delta}', \widetilde{\beta}', \widetilde{\gamma}'\right)'$ and θ is the vector $\left(\delta', \beta', \Omega_{vu}'\Omega_{vv}^{-1}\right)'$.

Before we turn to the test statistics themselves we consider the covariance matrices. Up to the different estimators of the scalar quantity $\sigma_{u\cdot v}^2$ both estimators of the covariance matrix are given by

$$\left(T^{-2}A_{IM}S^{\widetilde{x}\prime}S^{\widetilde{x}}A_{IM}\right)^{-1}\left(T^{-4}A_{IM}C'CA_{IM}\right)\left(T^{-2}A_{IM}S^{\widetilde{x}\prime}S^{\widetilde{x}}A_{IM}\right)^{-1}, \quad (42)$$

with C defined in the main text. For the outer terms the limit has already been established in the proof of Theorem 2 in Eq. (38) and thus it only remains to consider the expression in the middle. Straightforward calculations show that $T^{-3/2}A_{IM}c_{[rT]} \Rightarrow G(1)-G(r)$ and this implies that the central expression converges to $\int[G(1)-G(s)][G(1)-G(s)]'ds$. Consequently, the expression (42) converges, up to the scalar $\sigma_{u\cdot v}^2$, to V_{IM} as given in (26).

Under the null hypothesis both of the two defined statistics can be – as is usual in a linear regression model – written as (with $\check{V}_{IM}$ as given in the main text):

$$\check{W} = (R\widetilde{\theta}-r)'\left[RA_{IM}\check{V}_{IM}A_{IM}R'\right]^{-1}(R\widetilde{\theta}-r)$$
$$=\left(R(\widetilde{\theta}-\theta)\right)'\left[RA_{IM}\check{V}_{IM}A_{IM}R'\right]^{-1}\left(R(\widetilde{\theta}-\theta)\right)$$
$$=\left(A_R^{-1}RA_{IM}A_{IM}^{-1}(\widetilde{\theta}-\theta)\right)'\left[A_R^{-1}RA_{IM}\check{V}_{IM}A_{IM}R'\left(A_R^{-1}\right)'\right]^{-1}$$
$$\times\left(A_R^{-1}RA_{IM}A_{IM}^{-1}(\widetilde{\theta}-\theta)\right).$$

Now, by assumption the restriction matrix fulfills $\lim_{T\to\infty}A_R^{-1}RA_{IM} = R^*$, and $A_{IM}^{-1}(\widetilde{\theta}-\theta) \Rightarrow \Phi(V_{IM})$ under the null hypothesis. Therefore, in case of consistent estimation of the conditional long run variance $\sigma_{u\cdot v}^2$ using $\widehat{V}_{IM}$ it follows that

$$\widehat{W} \Rightarrow \left(R^*\Phi(V_{IM})\right)'\left(R^*V_{IM}R^{*\prime}\right)^{-1}\left(R^*\Phi(V_{IM})\right) \sim \chi_q^2.$$

We now consider the asymptotic behavior of the test statistic $\widetilde{W}$ using $\widetilde{\sigma}_{u\cdot v}^2$. It follows from the definition of $\widetilde{S}_t^u$ that

$$\Delta\widetilde{S}_t^u = u_t^+ - v_t'(\widetilde{\gamma}-\gamma) - f_t'(\widetilde{\delta}-\delta) - v_t'(\widetilde{\beta}-\beta),$$

with $\gamma = \Omega_{vv}^{-1}\Omega_{vu}$ and $u_t^+ = u_t - v_t'\gamma$. As discussed in Jansson (2002), the terms $f_t'(\widetilde{\delta}-\delta)$ and $v_t'(\widetilde{\beta}-\beta)$ can be neglected for long run variance estimation. Thus, the long run variance estimator based on $\Delta\widetilde{S}_t^u$, $\widetilde{\sigma}_{u\cdot v}^2$, asymptotically coincides with the long run variance estimator of $u_t^+ - v_t'(\widetilde{\gamma}-\gamma)$. Define $\eta_t^{+\prime} = [u_t^+, v_t']$ and its long run variance

$$\Omega^+ = \begin{bmatrix}\sigma_{u\cdot v}^2 & 0\\ 0 & \Omega_{vv}\end{bmatrix}.$$

An infeasible long run variance estimator, $\widehat{\Omega}^+$, using the unobserved η_t^+ is under the assumptions of Jansson (2002) consistent, i.e. $\widehat{\Omega}^+ \xrightarrow{p} \Omega^+$.

Next note that

$$u_t^+ - v_t'(\widetilde{\gamma} - \gamma) = \eta_t^{+\prime} \begin{bmatrix} 1 \\ -(\widetilde{\gamma} - \gamma) \end{bmatrix},$$

which implies that the HAC estimator, $\widetilde{\Omega}$ say, for $u_t^+ - v_t'(\widetilde{\gamma} - \gamma)$ can be written as

$$[1 \ -(\widetilde{\gamma} - \gamma)']\widehat{\Omega}^+ \begin{bmatrix} 1 \\ -(\widetilde{\gamma} - \gamma) \end{bmatrix}.$$

From Theorem 2 we know that

$$\begin{aligned} \widetilde{\gamma} - \gamma \ &\Rightarrow [0_{k\times p} \ 0_{k\times k} \ I_k]\sigma_{u.v}(\Pi')^{-1} \\ &\quad \times \left(\int g(s)g(s)'ds \right)^{-1} \int [G(1) - G(s)]dw_{u\cdot v} \\ &= \sigma_{u.v}(\Omega_{vv}^{-1/2})'d_\gamma, \end{aligned}$$

with d_γ as defined in the main text. This implies that

$$\begin{aligned} \widetilde{\Omega} &= [1 \ -(\widetilde{\gamma} - \gamma)']\widehat{\Omega}^+ \begin{bmatrix} 1 \\ -(\widetilde{\gamma} - \gamma) \end{bmatrix} \\ &\Rightarrow [1 \ -\sigma_{u\cdot v}d_\gamma'\Omega_{vv}^{-1/2}] \begin{bmatrix} \sigma_{u\cdot v}^2 & 0 \\ 0 & \Omega_{vv} \end{bmatrix} \begin{bmatrix} 1 \\ -\sigma_{u.v}(\Omega_{vv}^{-1/2})'d_\gamma \end{bmatrix} \\ &= \sigma_{u\cdot v}^2(1 + d_\gamma' d_\gamma). \end{aligned}$$

Thus, we have shown that $\widetilde{\sigma}_{u\cdot v}^2 \Rightarrow \sigma_{u\cdot v}^2(1 + d_\gamma' d_\gamma)$, which in turn implies the result for $\widetilde{W}$ as given in the formulation of the theorem using otherwise the same arguments as for $\widehat{W}$.

The result for the fixed-b test statistic $\widetilde{W}^*$ is slightly different because the fixed-b limit of the covariance matrix is such that $\widetilde{V}^* \Rightarrow Q_b(\widetilde{P}^*, \widetilde{P}^*)V_{IM}$. This implies that

$$\begin{aligned} \widetilde{W}^* &\Rightarrow \left(R^*\Phi(V_{IM})\right)' \left(Q_b(\widetilde{P}^*, \widetilde{P}^*)R^*V_{IM}R^{*\prime}\right)^{-1} \left(R^*\Phi(V_{IM})\right) \\ &\sim \frac{\chi_q^2}{Q_b(\widetilde{P}^*, \widetilde{P}^*)}, \end{aligned} \tag{43}$$

with numerator and denominator independent of each other. In Lemma 2 it has been shown that Ψ and $\widetilde{P}^*(r)$ are independent of each other conditional upon $W_v(r)$. Furthermore, the numerator of (43) – being a chi-square distribution – is independent of $W_v(r)$, which implies that the numerator and denominator are also independent of each other unconditionally. This in turn allows for the simulation of fixed-b critical values.

The stated results for the t-tests follow, obviously, as special cases of the *Wald* test results.

References

Andrews, D.W.K., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59, 817–854.
Bunzel, H., 2006. Fixed-*b* asymptotics in single-equation cointegration models with endogenous regressors. Econometric Theory 22, 743–755.
Choi, I., Kurozumi, E., 2012. Model selection criteria for the leads-and-lags cointegrating regression. J. Econometrics 169, 224–238.
De Jong, R., Davidson, J., 2000a. The functional central limit theorem and weak convergence to stochastic integrals I. Econometric Theory 16, 621–642.
De Jong, R., Davidson, J., 2000b. The functional central limit theorem and weak convergence to stochastic integrals II. Econometric Theory 16, 643–666.
Hansen, B., 1992. Convergence to stochastic integrals for dependent heterogeneous processes. Econometric Theory 8, 489–500.
Hashimzade, N., Vogelsang, T.J., 2008. Fixed-*b* asymptotic approximation of the sampling behavior of nonparametric spectral density estimators. J. Time Ser. Anal. 29, 142–162.
Jansson, M., 2002. Consistent covariance matrix estimation for linear processes. Econometric Theory 18, 1449–1459.
Jin, S., Phillips, P.C.B., Sun, Y., 2006. A new approach to robust inference in cointegration. Econom. Lett. 91, 300–306.
Kejriwal, M., Perron, P., 2008. Data dependent rules for selection of the number of leads and lags in the dynamic OLS cointegrating regression. Econometric Theory 24, 1425–1441.
Kiefer, N.M., Vogelsang, T.J., 2005. A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. Econometric Theory 21, 1130–1164.
Phillips, P.C.B., 1988. Weak convergence of sample covariance matrices to stochastic integrals via martingale approximations. Econometric Theory 4, 528–533.
Phillips, P.C.B., 1995. Fully modified least squares and vector autoregression. Econometrica 59, 1023–1078.
Phillips, P.C.B., Durlauf, S.N., 1986. Multiple regression with integrated processes. Rev. Econom. Stud. 53, 473–496.
Phillips, P.C.B., Hansen, B.E., 1990. Statistical inference in instrumental variables regression with I(1) processes. Rev. Econom. Stud. 57, 99–125.
Phillips, P.C.B., Loretan, M., 1991. Estimating long run economic equilibria. Rev. Econom. Stud. 58, 407–436.
Phillips, P.C.B, Perron, P., 1988. Testing for a unit root in time series regression. Biometrika 75, 335–346.
Saikkonen, P., 1991. Asymptotically efficient estimation of cointegrating regressions. Econometric Theory 7, 1–21.
Sims, C., Stock, J.H., Watson, M.W., 1990. Inference in linear time series models with some unit roots. Econometrica 58, 113–144.
Stock, J.H., 1987. Asymptotic properties of least squares estimators of cointegrating vectors. Econometrica 55, 1035–1056.
Stock, J.H., Watson, M.W., 1993. A simple estimator of cointegrating vectors in higher order integrated systems. Econometrica 61, 783–820.
Sun, Y., Phillips, P.C.B., Jin, S., 2008. Optimal bandwidth selection in heteroskedasticity-autocorrelation robust testing. Econometrica 76, 175–194.
Vogelsang, T.J., Wagner, M., 2013a. A fixed-*b* perspective on the Phillips–Perron unit root tests. Econometric Theory 29, 609–628.
Vogelsang, T.J., Wagner, M., 2013b. Integrated Modified OLS Estimation and Fixed-*b* Inference for Cointegrating Regressions. Working Paper, Department of Economics, Michigan State University.