IHS Economics Series
Working Paper 263
January 2011

Integrated Modified OLS Estimation and Fixed-b Inference for Cointegrating Regressions

Timothy J. Vogelsang
Martin Wagner

Impressum

Author(s):
Timothy J. Vogelsang, Martin Wagner

Title:
Integrated Modified OLS Estimation and Fixed-b Inference for Cointegrating Regressions

ISSN: Unspecified

2011 Institut für Höhere Studien - Institute for Advanced Studies (IHS)
Josefstädter Straße 39, A-1080 Wien
E-Mail: office@ihs.ac.at
Web: www.ihs.ac.at

All IHS Working Papers are available online:
http://irihs.ihs.ac.at/view/ihs_series/
This paper is available for download without charge at:
https://irihs.ihs.ac.at/id/eprint/2037/

263

Reihe Ökonomie

Economics Series

Integrated Modified OLS Estimation and Fixed-*b* Inference for Cointegrating Regressions

Timothy J. Vogelsang, Martin Wagner

Institut für Höhere Studien
Institute for Advanced Studies
Vienna

263

Reihe Ökonomie

Economics Series

Integrated Modified OLS Estimation and Fixed-*b* Inference for Cointegrating Regressions

Timothy J. Vogelsang, Martin Wagner

January 2011

Institut für Höhere Studien (IHS), Wien
Institute for Advanced Studies, Vienna

Contact:

Timothy J. Vogelsang
Department of Economics
Michigan State University
310 Marshall-Adams Hall
East Lansing, MI 48824-1038, USA
email: tjv@msu.edu

Martin Wagner
Department of Economics and Finance
Institute for Advanced Studies
Stumpergasse 56
1060 Vienna, Austria
☎: +43/1/599 91-150
email: martin.wagnger@ihs.ac.at
and
Frisch Centre for Economic Research, Oslo

Founded in 1963 by two prominent Austrians living in exile – the sociologist Paul F. Lazarsfeld and the economist Oskar Morgenstern – with the financial support from the Ford Foundation, the Austrian Federal Ministry of Education and the City of Vienna, the Institute for Advanced Studies (IHS) is the first institution for postgraduate education and research in economics and the social sciences in Austria. The **Economics Series** presents research done at the Department of Economics and Finance and aims to share “work in progress” in a timely way before formal publication. As usual, authors bear full responsibility for the content of their contributions.

Das Institut für Höhere Studien (IHS) wurde im Jahr 1963 von zwei prominenten Exilösterreichern – dem Soziologen Paul F. Lazarsfeld und dem Ökonomen Oskar Morgenstern – mit Hilfe der Ford-Stiftung, des Österreichischen Bundesministeriums für Unterricht und der Stadt Wien gegründet und ist somit die erste nachuniversitäre Lehr- und Forschungsstätte für die Sozial- und Wirtschafts-wissenschaften in Österreich. Die **Reihe Ökonomie** bietet Einblick in die Forschungsarbeit der Abteilung für Ökonomie und Finanzwirtschaft und verfolgt das Ziel, abteilungsinterne Diskussionsbeiträge einer breiteren fachinternen Öffentlichkeit zugänglich zu machen. Die inhaltliche Verantwortung für die veröffentlichten Beiträge liegt bei den Autoren und Autorinnen.

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 approach of Phillips and Hansen (1990) 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.

Keywords
Bandwidth, cointegration, fixed-*b* asymptotics, Fully Modified OLS, IM-OLS, kernel

JEL Classification
C31, C32

Comments

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 No. 13398. The comments of conference participants at the Econometrics Study Group in Rauischholzhausen, the 4th CIREQ Time Series Conference in Montreal, an ETSERN Meeting in Pamplona, the 37th Macromodels Conference in Pultusk, the 4th International Conference on Computational and Financial Econometrics in London, and of seminar participants at the Institute for Advanced Studies in Vienna and Michigan State University are gratefully acknowledged.

Contents

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) estimators of the parameters are super-consistent, i.e. converge 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 have been proposed that lead to zero mean Gaussian mixture limiting distributions, which in turn makes 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. 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, i.e. a 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 the development of fixed-b inference. 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 results are 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. The only other paper in the literature that develops fixed-b theory for inference in cointegration regression is Bunzel (2006), who analyzes tests based on the DOLS estimator.

Developing useful fixed-b results for tests based on IM-OLS leads to some new challenges compared to tests based on DOLS or tests in stationary regressions. Specifically, the residuals of the IM-

OLS regression cannot be used to obtain asymptotically pivotal fixed-b test statistics. Fixed-b inference instead has to be based on the residuals of a particularly further augmented regression, as discussed in detail in Section 5. A similar complication also arises in Vogelsang and Wagner (2010), 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 necessitate 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. 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. 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 for the IM-OLS parameter estimates is discussed, both with standard and fixed-b asymptotic theory, and 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 available upon request 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.

2 FM-OLS Estimation and Inference in Cointegrating Regressions

Consider the following regression model for $t = 1, 2, ..., T$

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

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

where y_t is a scalar time series and x_t is a $k \times 1$ vector of time series. For notational brevity we here 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 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 with

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

where clearly $\Omega_{vu} = \Omega_{uv}'$. 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 $\sigma_{u\cdot v}^2 = \Omega_{uu} - \Omega_{uv}\Omega_{vv}^{-1}\Omega_{uv}'$ 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 $\Lambda = \sum_{j=1}^{\infty}\mathbb{E}(\eta_t\eta_{t-j}')$, 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 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

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

Using this notation, the OLS estimator is defined as

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

To state asymptotic results the following scaling matrix is needed:

$$A = \left[\begin{array}{cc} T^{-1/2} & \mathbf{0} \\ \mathbf{0} & T^{-1} I_k \end{array} \right].$$

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

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

$$\Rightarrow \left(\begin{array}{cc} 1 & \int B_v(r)'dr \\ \int B_v(r)dr & \int B_v(r)B_v(r)'dr \end{array} \right)^{-1} \left(\begin{array}{c} \int dB_u(r) \\ \int B_v(r)dB_u(r) + \Delta_{vu} \end{array} \right) = \Theta = \left[\begin{array}{c} \Theta_\mu \\ \Theta_\beta \end{array} \right],$$

where $\Delta_{vu} = \Sigma_{vu} + \Lambda_{vu}$. Unless otherwise stated, 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)** $\lambda_{12} = \mathbf{0}$, $\Delta_{vu} = \mathbf{0}$, 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 robust standard errors.

When the regressors are endogenous, the limiting distribution of $T(\widehat{\beta} - \beta)$ is obviously more complicated because of 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 very 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 $B_v(r) = \Omega_{vv}^{1/2} W_2(r)$. Using $B_{u \cdot v}(r)$, one can write

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

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

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

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

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

where $\Delta_{vu}^+ = \Delta_{vu} - \Delta_{vv}\Omega_{vv}^{-1}\Omega_{vu}$. 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

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

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

$$\mathcal{M}^* = T \begin{pmatrix} 0 \\ \widehat{\Delta}_{vu}^+ \end{pmatrix}.$$

It is shown in Phillips and Hansen (1990) that

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

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 $T(\widehat{\beta}^+ - \beta)$ is such that conditional on $B_v(r)$, a zero mean normal limit is obtained. 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

$$\widehat{\sigma}_{u \cdot v}^2 = \widehat{\Omega}_{uu} - \widehat{\Omega}_{uv} \widehat{\Omega}_{vv}^{-1} \widehat{\Omega}_{vu}. \tag{7}$$

In practice FM-OLS requires the choice of bandwidth and kernel. While the bandwidth and kernel play no role asymptotically when appealing to consistency results for $\widehat{\Omega}$ and $\widehat{\Delta}$, in finite samples the kernel and bandwidth affect the sampling distributions of $\widehat{\theta}^+$ and of t and *Wald* statistics based on $\widehat{\theta}^+$. To obtain an approximation 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, Phillips and Jin (2008). The theory there has been developed only for models with stationary regressions, 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 integrated regressors and the specification of the deterministic components.

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. 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 b and $k(\cdot)$ (fixed-b). It becomes a question as to which limit provides a more useful approximation. If one wants to capture the impact of kernel and bandwidth 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 $\widehat{\Delta}$ is more challenging because the literature does not yet provide blueprints. We derive the corresponding result, which may itself be of independent interest, in detail in the appendix in the proof of Theorem 1.

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

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

where $\widehat{B}_u(r) = B_u(r) - r\Theta_\mu - \int_0^r B_v(s)' ds \Theta_\beta$ and $d\widehat{B}_u(r) = dB_u(r) - \Theta_\mu - B_v(r)' dr \Theta_\beta$. 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 generic 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

$$Q_b(P_1, P_2) = -\frac{1}{b^2}\int_0^1\int_0^1 k''(\frac{|r-s|}{b})P_1(s)P_2(r)'dsdr \qquad (8)$$
$$+\frac{1}{b}\int_0^1 k'(\frac{|1-s|}{b})\left(P_1(1)P_2(s)' + P_1(s)P_2(1)'\right)ds + P_1(1)P_2(1)',$$

$$Q_b^{\Delta}(P_1, P_2) = -\frac{1}{b^2}\int_0^1\int_r^1 k''(\frac{|r-s|}{b})P_1(s)P_2(r)'drds + \frac{1}{b}\int_0^1 k'(\frac{|1-s|}{b})P_1(1)P_2(s)'ds \qquad (9)$$
$$+\frac{1}{b}k'_+(0)\int_0^1 P_1(s)P_2(s)'ds + P_1(1)P_2(1)' - \int_0^1 P_1(s)dP_2(s)' - \Lambda'_{12}.$$

The second case considered refers to the Bartlett kernel, in which case the stochastic processes $Q_b(P_1, P_2)$ and $Q_b^{\Delta}(P_1, P_2)$ become

$$Q_b(P_1, P_2) = \frac{2}{b}\int_0^1 P_1(s)P_2(s)'ds - \frac{1}{b}\int_0^{1-b}\left(P_1(s)P_2(s+b)' + P_1(s+b)P_2(s)'\right)ds \qquad (10)$$
$$-\frac{1}{b}\int_{1-b}^1\left(P_1(1)P_2(s)' + P_1(s)P_2(1)'\right)ds + P_1(1)P_2(1)',$$
$$Q_b^{\Delta}(P_1, P_2) = \frac{1}{b}\int_0^1 P_1(s)P_2(s)'ds - \frac{1}{b}\int_0^{1-b} P_1(s+b)P_2(s)'ds \qquad (11)$$
$$-\frac{1}{b}\int_{1-b}^1 P_1(1)P_2(s)'ds + P_1(1)P_2(1)' - \int_0^1 P_1(s)dP_2(s)' - \Lambda'_{12}.$$

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}.

Theorem 1 *Assume that the FCLT (3) holds. Let $M = bT$, where $b \in (0, 1]$ is held fixed as $T \to \infty$, then as $T \to \infty$*

$$\widehat{\Omega} \Rightarrow Q_b(P_{\widehat{\eta}}, P_{\widehat{\eta}}), \quad \widehat{\Delta} \Rightarrow Q_b^{\Delta}(P_{\widehat{\eta}}, P_{\widehat{\eta}}) \qquad (12)$$

and in particular

$$\widehat{\Omega}_{vv} \Rightarrow Q_b(B_v, B_v), \quad \widehat{\Omega}_{vu} \Rightarrow Q_b(B_v, \widehat{B}_u),$$
$$\widehat{\Delta}_{vv} \Rightarrow Q_b^{\Delta}(B_v, B_v), \quad \widehat{\Delta}_{vu} \Rightarrow Q_b^{\Delta}(B_v, \widehat{B}_u).$$

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

$$A^{-1}\left(\widehat{\theta}^+ - \theta\right) = \left(A\widetilde{X}'\widetilde{X}A\right)^{-1}\left(A\widetilde{X}'y^+ - A\mathcal{M}^*\right) \qquad (13)$$
$$\Rightarrow \begin{pmatrix} 1 & \int B_v(r)'dr \\ \int B_v(r)dr & \int B_v(r)B_v(r)'dr \end{pmatrix}^{-1}\begin{pmatrix} \int dB_{uv}^b(r) \\ \int B_v(r)dB_{uv}^b(r) + \mathcal{B}_1 - \mathcal{B}_2 \end{pmatrix},$$

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

$$\mathcal{B}_1 = \Delta_{vu} - Q^{\Delta}_b(B_v, \widehat{B}_u),$$
$$\mathcal{B}_2 = \left(\Delta_{vv} - Q^{\Delta}_b(B_v, B_v)\right) Q_b(B_v, B_v)^{-1}Q_b(B_v, \widehat{B}_u).$$

Theorem 1 shows that under the fixed-b asymptotic approximation, the limit 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 e.g. the orthogonalization step of FM-OLS. The term $\int B_v(r)dB^b_{uv}(r)$ is close to a zero mean Gaussian mixture only if in $B^b_{uv}(r) = B_u(r) - B_v(r)'Q_b(B_v, B_v)^{-1}Q_b(B_v, \widehat{B}_u)$ the Q_b terms are close to the population quantities Ω^{-1}_{vv} 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 $Q^{\Delta}_b(B_v, \widehat{B}_u)$ and $Q^{\Delta}_b(B_v, B_v)$ are close to Δ_{vu} and Δ_{vv}. If these approximations are not accurate an additive bias is present. Thus, the result of Theorem 1 shows that FM-OLS relies critically on the consistency approximation being accurate and the result also shows how moving around kernel and bandwidth impacts the sampling behavior of FM-OLS.

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 of β with a zero mean Gaussian mixture limiting distribution. Like FM-OLS, the transformation has two steps but neither step requires estimators 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

$$y_t = f_t'\delta + x_t'\beta + u_t, \tag{14}$$

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

$$T^{-1}\tau_F^{-1}\sum\nolimits_{t=1}^{[rT]} f_t \to \int_0^r f(s)ds \text{ with } \int_0^1 f(s)f(s)'ds > 0. \tag{15}$$

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

Computing the partial sum of both sides of (14) gives the model

$$S^y_t = S^{f\prime}_t\delta + S^{x\prime}_t\beta + S^u_t, \tag{16}$$

where $S^y_t = \sum_{j=1}^t y_j$, $S^f_t = \sum_{j=1}^t f_j$ and S^x_t and S^u_t are defined analogously. In vector notation, using similar notation as in the discussion of the OLS estimator, we have

$$S^y = S^{\widetilde{x}\prime}\theta + S^u, \tag{17}$$

with $S^{\widetilde{x}}$ stacking S_t^f and S_t^x. Define the OLS estimator in the partial sum regression as

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

which leads to

$$\widetilde{\theta} - \theta = \left(S^{\widetilde{x}\prime} S^{\widetilde{x}}\right)^{-1} \left(S^{\widetilde{x}\prime} S^u\right). \tag{19}$$

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

$$\sum_{t=1}^{T} x_t u_t \tag{20}$$

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

$$\sum_{t=1}^{T} S_t^x S_t^u \tag{21}$$

in $\widetilde{\theta}$. Sums of the form of (20) 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, sums of the form of (21) 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 (16):

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

Redefine $S^{\widetilde{x}}$ so that it stacks S_t^f, S_t^x, x_t and redefine θ so that it stacks δ, β, γ. With this economical use of notation, the matrix form of (22) is still given by (17) and the OLS estimator is still formally given by (18) and (19). 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 the OLS estimator of (22).

Theorem 2 *Suppose that (3) and (15) hold. Then as $T \to \infty$*

$$\begin{aligned} \begin{pmatrix} T^{1/2}\tau_F(\widetilde{\delta} - \delta) \\ T(\widetilde{\beta} - \beta) \\ (\widetilde{\gamma} - \Omega_{vv}^{-1}\Omega_{vu}) \end{pmatrix} &= A_{IM}^{-1}\left(\widetilde{\theta} - \theta\right) = \left(T^{-2}A_{IM}S^{\widetilde{x}\prime}S^{\widetilde{x}}A_{IM}\right)^{-1}\left(T^{-2}A_{IM}S^{\widetilde{x}\prime}S^u\right) \\ &\Rightarrow \sigma_{u\cdot v}\left(\Pi \int g(s)g(s)'ds\Pi'\right)^{-1} \Pi \int g(s)w_{u\cdot v}(s)ds \\ &= \sigma_{u\cdot v}(\Pi')^{-1}\left(\int g(s)g(s)'ds\right)^{-1} \int [G(1) - G(s)]dw_{u\cdot v}(s) = \Psi, \end{aligned} \tag{23}$$

where

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

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 $\Omega_{vv}^{-1}\Omega_{vu}$. This is therefore the population parameter vector for $\widetilde{\gamma}$ that is non-zero in case of regressor endogeneity.

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

$$\begin{aligned} V_{IM} = {} & \sigma_{u \cdot v}^2 (\Pi')^{-1} \left(\int g(s)g(s)'ds \right)^{-1} \left(\int [G(1) - G(s)][G(1) - G(s)]'ds \right) \\ & \times \left(\int g(s)g(s)'ds \right)^{-1} \Pi^{-1}. \end{aligned} \tag{24}$$

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 $\Pi_{FM} = \text{diag}(I_p, \Omega_{vv}^{1/2})$ the latter is given by

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

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

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

where

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

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 Kejrival and Perron (2008), where we use the more flexible version discussed in Choi and Kurozumi (2008) 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 5,000.

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. 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 RMSE of DOLS and IM-OLS tends to be larger than the RMSE of OLS and FM-OLS, although 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 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 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 (24). 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, Stock and Watson (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

$$\lim_{T\to\infty} A_R^{-1} R A_{IM} = R^*, \tag{25}$$

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

The expression (24) immediately suggests estimators, $\check{V}_{IM}$, for V_{IM} of the form

$$\check{V}_{IM} = \check{\sigma}^2_{u\cdot v}\left(T^{-2}A_{IM}S^{\widetilde{x}\prime}S^{\widetilde{x}}A_{IM}\right)^{-1}\left(T^{-4}A_{IM}C'CA_{IM}\right)\left(T^{-2}A_{IM}S^{\widetilde{x}\prime}S^{\widetilde{x}}A_{IM}\right)^{-1},$$

where $\check{\sigma}^2_{u\cdot v}$ is an estimator of $\sigma^2_{u\cdot v}$ and C is the matrix formed by stacking the vector

$$c_t = S_T^{S^{\widetilde{x}}} - S_{t-1}^{S^{\widetilde{x}}},$$

with $S_t^{S^{\widetilde{x}}} = \sum_{j=1}^t S_j^{\widetilde{x}}$.

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 the first differences of the OLS residuals of the IM-OLS regression (22), $\Delta\widetilde{S}_t^u$ to directly estimate $\sigma^2_{u\cdot v}$ by

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

It turns out (see Theorem 3 below) that $\widetilde{\sigma}^2_{u\cdot v}$ is not consistent under standard assumptions on bandwidth and kernel as discussed e.g. in Jansson (2002). 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.

Having now discussed all necessary quantities we can define the *Wald* statistics for different estimators of $\sigma^2_{u\cdot v}$ as

$$\check{W} = (R\widetilde{\theta} - r)'[RA_{IM}\check{V}_{IM}A_{IM}R']^{-1}(R\widetilde{\theta} - r), \tag{26}$$

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 below.

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 ignores the impact of kernel and bandwidth choices. In order to capture the effects of these choices fixed-b asymptotic theory needs to be developed. Clearly, 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, resulting *Wald* statistics have pivotal asymptotic distributions that only depend upon kernel and bandwidth (as well as the number of integrated regressors and the specification of the deterministic component) 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, since 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 will be shown in Lemma 2 below, 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, it is correlated with the limit of $\widetilde{\theta}$, with this correlation itself being a complicated function of 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 challenge in cointegrating regressions for fixed-b theory that does not arise in stationary regression settings.

In order to construct asymptotically pivotal test statistics under fixed-b asymptotics it turns out that the OLS residuals of a particularly augmented version of (22) can be considered. This further augmented regression is given by

$$S_t^y = S_t^{f\prime}\delta + S_t^{x\prime}\beta + x_t'\gamma + z_t'\kappa + S_t^{u*}, \tag{27}$$

where

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

The asymptotic distribution of the OLS estimator of the parameters in (27) is given in Lemma 1 in the appendix.

Let $\widetilde{S}_t^{u*}$ denote the OLS residuals from (27) and define

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

which is used to define a third estimator of V_{IM} given by

$$\widetilde{V}_{IM}^{*} = \widetilde{\sigma}_{u\cdot v}^{2*}\left(T^{-2}A_{IM}S^{\widetilde{x}\prime}S^{\widetilde{x}}A_{IM}\right)^{-1}(T^{-4}A_{IM}C'CA_{IM})\left(T^{-2}A_{IM}S^{\widetilde{x}\prime}S^{\widetilde{x}}A_{IM}\right)^{-1}.$$

The following lemma characterizes the asymptotic behavior of the partial sum processes of the first differenced OLS residuals of the IM-OLS regression (22) and of the further augmented regression (27), which is needed to subsequently discuss fixed-b asymptotics for test statistics.

Lemma 2 *Let $\widetilde{S}_t^{u}$ and $\widetilde{S}_t^{u*}$ denote the residuals of regressions* (16) *and* (22)*. The asymptotic behavior of the corresponding partial sum processes is given by*

$$T^{-1/2}\sum_{t=2}^{[rT]}\Delta\widetilde{S}_t^{u} \Rightarrow$$

$$\sigma_{u\cdot v}\left[\int_0^r dw_{u\cdot v}(s) - g(r)'\left(\int_0^1 g(s)g(s)'ds\right)^{-1}\int_0^1 (G(1)-G(s))\,dw_{u\cdot v}(s)\right] = \sigma_{u\cdot v}\widetilde{P}(r), \qquad (28)$$

$$T^{-1/2}\sum_{t=2}^{[rT]}\Delta\widetilde{S}_t^{u*} \Rightarrow$$

$$\sigma_{u\cdot v}\left[\int_0^r dw_{u\cdot v}(s) - h(r)'\left(\int_0^1 h(s)h(s)'ds\right)^{-1}\int_0^1 (H(1)-H(s))\,dw_{u\cdot v}(s)\right] = \sigma_{u\cdot v}\widetilde{P}^{*}(r), \qquad (29)$$

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, it holds that Ψ, the limit of $\widetilde{\theta}$, and $\widetilde{P}^{*}(r)$ are, conditional upon $W_v(r)$, independent.

It follows from (23) that, conditional on $W_v(r)$, the random part of Ψ is the Gaussian random variable $\int_0^1[G(1)-G(s)]dw_{u\cdot v}(s)$. Straightforward calculations show that, conditional on $W_v(r)$, the random process $\widetilde{P}(r)$ as defined in (28) is correlated with this random variable, which implies that the fixed-b limit of $\widetilde{\sigma}_{u\cdot v}^{2}$ is correlated with Ψ and this correlation depends on nuisance parameters through Π. The important result of Lemma 2 is that the random process $\widetilde{P}^{*}(r)$ defined in (29) is uncorrelated with Ψ. Given that, conditional upon $W_v(r)$, Ψ and $\widetilde{P}^{*}(r)$ are both Gaussian, it follows that they are independent. This result forms the basis for pivotal test statistics using fixed-b asymptotics, defined by

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

The asymptotic behavior of the *Wald* statistics 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 $\widetilde{W}^*$.

Theorem 3 *Assume that the FCLT (3) holds, that the deterministic components satisfy (15) and that R satisfies (25). Suppose that the bandwidth, M, and kernel, $k(\cdot)$, satisfy conditions such that $\widehat{\sigma}_{u\cdot v}^2$ is consistent. Then as $T \to \infty$*

$$\widehat{W} \Rightarrow \chi_q^2,$$

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. When $q = 1$,

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

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

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 independent of $Q_b(\widetilde{P}^*, \widetilde{P}^*)$. When $q = 1$,

$$\widetilde{t}^* = \frac{R\widetilde{\theta} - r}{\sqrt{RA_{IM}\widetilde{V}_{IM}^* A_{IM}R'}} \Rightarrow \frac{Z}{\sqrt{Q_b(\widetilde{P}^*, \widetilde{P}^*)}}, \tag{30}$$

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

Because of consistency of $\widehat{\sigma}^2_{u\cdot v}$ inference using $\widehat{W}$ is standard. Because $d'_\gamma d_\gamma > 0$, $\widetilde{\sigma}^2_{u\cdot v}$ is upwardly biased and the critical values of $\widetilde{W}$ are smaller than those of the χ^2_q distribution. Therefore, using χ^2_q critical values for $\widetilde{W}$ leads to a conservative test from the perspective of standard bandwidth rule asymptotics. The fixed-b limit distribution of $\widetilde{W}^*$ is similar to what is obtained for *Wald* tests 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, also observed in stationary regressions, the process depends upon $W_v(r)$, i.e. upon the number of integrated regressors included in the cointegrating regression (a similar finding was made by Bunzel (2006) for fixed-b inference using DOLS). Thus, critical values need to be simulated taking into account the specification of the deterministic components, the number of integrated regressors, the kernel function and the bandwidth choice. In a supplementary appendix we tabulate 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.

6 Finite Sample Performance of Test Statistics

In this section we provide some finite sample results 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 $\widehat{\sigma}^2_{u\cdot v}$. The IM-OLS statistics were implemented in three ways: The first uses $\widehat{\sigma}^2_{u\cdot v}$ and is labeled IM(O), the second uses $\widetilde{\sigma}^2_{u\cdot v}$ and is labeled IM(D) and the third uses $\widetilde{\sigma}^{2*}_{u\cdot v}$ 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 $M = 1, 2, ..., T$. 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 before and the bandwidth for the long run variance estimation is chosen according to Andrews (1991). Rejections for the OLS, FM and DOLS statistics are carried out using $N(0, 1)$ critical values in all cases and also the rejections for IM(O) and IM(D) 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. In contrast, rejections for 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 IM(fb) t and *Wald* statistics using the limiting random variable given by Theorem 3. Therefore, a different critical value is used for FM(fb) for each value of the bandwidth. The empirical rejection probabilities were computed using 5,000 replications, and the nominal level is 0.05 in all cases.

Tables 2 and 3 report empirical null rejection probabilities using the data dependent bandwidth choice 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 each of these 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 closer to FM-OLS. The IM(O) and IM(D) tests also show some over-rejections, which are however less severe than for FM-OLS. IM(D) usually has slightly lower rejection rates than IM(O) and this is to be 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. 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 (even larger than IM(O) and IM(D)) when $\rho_1, \rho_2 = 0.9$.

In order to obtain a better understanding of the role that the bandwidth choice plays in the over-rejection problem we plot in Figures 1–8 null rejection probabilities as a function of $b \in (0, 1]$. The first four figures show the results for the Bartlett kernel for $T = 100$ for increasing values of ρ_1, ρ_2. In Figure 1, with $\rho_1, \rho_2 = 0$, for small bandwidths all tests have rejection probabilities close to 0.05. As the bandwidth increases, with the exception of IM(fb) all rejection probabilities increase substantially. This figure shows that with $\rho_1, \rho_2 = 0$, the fixed-b approximation performs well for all bandwidths. As the values of ρ_1, ρ_2 increase, we see in Figures 2–4 that the rejections take a J-curve shape for FM, IM(O) and IM(D) and over-rejection becomes a serious problem regardless of bandwidth for these tests whereas in contrast the rejection probabilities of the IM(fb) test stay essentially constant as the bandwidth increases. For bandwidths larger than about 10% of the sample size the rejection probabilities of IM(D) are typically larger than those of IM(O), whereas they are similar for smaller bandwidths.

Only when $\rho_1, \rho_2 = 0.9$, is the IM(fb) test using the Bartlett kernel severely affected by over-rejections. The kernel choice is important for this last finding because the severe over-rejections for the IM(fb) tests occur to a much smaller extent when instead of the Bartlett the QS kernel is employed. This is shown in Figures 5 ($T = 100$) and 6 ($T = 200$), where the results for $\rho_1, \rho_2 = 0.9$ are displayed. The results are similar as for the Bartlett kernel for the FM, IM(O) and IM(D) tests, but are much better for the IM(fb) test, which now over-rejects much less severely across a wide range of bandwidths. Looking at the results for $T = 200$ shows that, compared to $T = 100$, the over-rejection problems of FM, IM(O) and IM(D) are somewhat reduced when small bandwidths are used. The IM(fb) test has, for $T = 200$, rejections just above 0.05 as long as b is not too small.

In Figures 7 and 8 we show results for the *Wald* test for $T = 100$ for $\rho_1, \rho_2 = 0.9$. Again the kernel choice turns out to be important, since when using the Bartlett kernel (Figure 7) all tests have rejection probabilities of at least 0.55 regardless of the bandwidth. Using the QS kernel (Figure 8) can lead to much smaller rejection probabilities, especially when the bandwidth is large. The discussed results make clear that in case of strong serial correlation and endogeneity, the QS kernel is preferred over the Bartlett kernel for IM(fb).

The overall picture is that the IM(fb) test is the most robust statistic in terms of over-rejection problems. Increasing the values of ρ_1, ρ_2 causes over-rejections to emerge, but the tendency to over-reject is much smaller than for the other statistics. 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 the results for the case $\rho_1, \rho_2 = 0.6$ for the *Wald* test for $T = 100$. Patterns are similar for other values of ρ_1, ρ_2, for the t-tests and for other sample sizes. Starting from the null values of β_1 and β_2 equal to 1, we consider under the alternative $\beta_1 = \beta_2 \in (1, 2]$, using – including the null value – in total 21 values on a grid with mesh 0.05. We focus on size-corrected power because of the over-rejection problems under the null hypothesis. This allows to see power differences 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 Figures 9–12 we depict the power of the FM and IM *Wald* tests using the QS kernel. The figures are ordered according to increasing bandwidth $b = 0.06, 0.3, 0.7, 1.0$. The first thing to note is that, by and large, size-corrected power of the FM and the IM tests is similar. FM has highest power for $b = 0.06$, whereas for the other values of b highest power is achieved by IM(O). These figures show that partial summing before estimation (using the IM-OLS estimator) implies only minimal power losses.

Figures 13 (Bartlett) and 14 (QS) show the impact of the bandwidth on the power of the IM(fb) test by displaying the power curves for eight values of $b = 0.02, 0.06, 0.1, 0.3, 0.5, 0.7, 0.9, 1.0$. The figures make it obvious that power depends on the bandwidth and tends to decrease with increasing bandwidth. Again, this is an observation made in a stationary regression setting by Kiefer and Vogelsang (2005). The most striking feature when comparing these two figures is that power when using the QS kernel is much more sensitive to the bandwidth choice than power when using the Bartlett kernel. With large bandwidths, power with the QS kernel is much lower than when small bandwidths are used. This has 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. 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 forms the basis of the data dependent bandwidth rule developed by Sun et al. (2008).

Finally, Figures 15 (Bartlett) and 16 (QS) allow for power comparisons across the various tests (OLS, FM-OLS, DOLS, IM-OLS). In these figures power of the IM(fb) test is shown for $b = 0.06, 0.1, 1.0$ and using the Andrews (1991) data dependent bandwidth. When using the Bartlett kernel, OLS and FM tests have the highest size-corrected power, with the power of all IM(fb) tests being slightly lower and the power of the DOLS test being substantially lower. For the QS kernel the results are as expected, given the results of Figures 13 and 14, because the power of the IM(fb) tests is sensitive to the bandwidth when using the QS kernel. In this case the power of the IM(fb) test with $b = 1$ is lower than 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 kernel and bandwidth 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 kernel and bandwidth choices. The limiting distribution shows that the accuracy of long run variance estimation is crucial for the properties of the FM-OLS estimator.

The derivation of the fixed-b limit distribution of the FM-OLS estimator presents some challenges for fixed-b theory. The first challenge is that fixed-b theory has to be extended from the stationary regression framework to the world of unit roots and cointegration as in Bunzel (2006). In cointegrating regressions fixed-b limits of long run variance estimators depend not only upon the specification of the deterministic components, but also upon the number of integrated regressors. The second challenge 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 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 results for FM-OLS suggest that the various long run and half long run variance estimators used to construct FM-OLS 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 simple 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 kernel, bandwidth or numbers of leads and lags). When the IM-OLS estimates are to be used for inference, still a long run variance needs to be estimated. 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 but constructing tests is more complicated than in stationary regressions because the OLS residuals of the IM-OLS regression cannot themselves be used for pivotal fixed-b inference. Instead the residuals of a specifically further augmented regression can form the basis for fixed-b inference. 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 kernel as well as the bandwidth 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 bias and RMSE. Typically, the bias is smaller than the bias of FM-OLS and the 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/endogeneity is not too strong. When serial correlation/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.

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 estimated $\widetilde{\gamma}$ can serve as a basis for endogeneity testing.

Acknowledgements

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 No. 13398. The comments of conference participants at the Econometrics Study Group in Rauischholzhausen, the 4th CIREQ Time Series Conference in Montreal, an ETSERN Meeting in Pamplona, the 37th Macromodels Conference in Pultusk, the 4th International Conference on Computational and Financial Econometrics in London, and of seminar participants at the Institute for Advanced Studies in Vienna and Michigan State University are gratefully acknowledged.

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. and Kurozumi, E.: (2008), Model selection criteria for the leads-and-lags cointegrating regression. Mimeo.

De Jong, R. and Davidson, J.: (2000a), The functional central limit theorem and weak convergence to stochastic integrals I, *Econometric Theory* **16**, 621–642.

De Jong, R. and 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. and Vogelsang, T. J.: (2008), Fixed-b asymptotic approximation of the sampling behavior of nonparametric spectral density estimators, *Journal of Time Series Analysis* **29**, 142–162.

Jansson, M.: (2002), Consistent covariance matrix estimation for linear processes, *Econometric Theory* **18**, 1449–1459.

Kejriwal, M. and 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. and 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. and Durlauf, S. N.: (1986), Multiple regression with integrated processes, *Review of Economic Studies* **53**, 473–496.

Phillips, P. C. B. and Hansen, B. E.: (1990), Statistical inference in instrumental variables regression with I(1) processes, *Review of Economic Studies* **57**, 99–125.

Phillips, P. C. B. and Loretan, M.: (1991), Estimating long run economic equilibria, *Review of Economic Studies* **58**, 407–436.

Phillips, P. C. B and 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. and 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. and 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. and Jin, S.: (2008), Optimal bandwidth selection in heteroskedasticity-autocorrelation robust testing, *Econometrica* **76**, 175–194.

Vogelsang, T.J. and Wagner, M.: (2010), A fixed-b perspective on the Phillips-Perron tests. Mimeo.

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

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

with $\Theta = [\Theta_\mu, \Theta_\beta']'$ as defined in the main text in the discussion of the OLS estimator. Using the definition of $\widehat{\eta}_t = [\widehat{u}_t, v_t']'$ and stacking now leads to

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

We define correspondingly

$$dP_{\widehat{\eta}}(r) = \left[\begin{array}{c} dB_u(r) \\ dB_v(r) \end{array}\right] - \left[\begin{array}{c} \left[\begin{array}{cc} 1 & B_v(r)' \end{array}\right]\Theta \\ 0 \end{array}\right] dr.$$

Remark 1 *For the IM-OLS estimator we consider more general deterministic components so it is useful for clarity sake 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* (14)*, 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 replace $\left[\begin{array}{cc} r & \int_0^r B_v(s)' \end{array}\right]$with $\left[\begin{array}{cc} \int_0^r f(s)' & \int_0^r B_v(s)' \end{array}\right]$ and replace $\left[\begin{array}{cc} 1 & B_v(r)' \end{array}\right]$ with $\left[\begin{array}{cc} f(r)' & B_v(r)' \end{array}\right]$.*

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

$$\begin{aligned} \widehat{\eta}_t = \left[\begin{array}{c} \widehat{u}_t \\ v_t \end{array}\right] &= \left[\begin{array}{c} u_t \\ v_t \end{array}\right] - \left[\begin{array}{c} (\widehat{\mu} - \mu) + x_t'(\widehat{\beta} - \beta) \\ 0 \end{array}\right] \\ &= \eta_t - \lambda_t \end{aligned}$$

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

$$T^{-1/2}S_{[rt]}^\lambda \Rightarrow \left[\begin{array}{c} \left[\begin{array}{cc} r & \int_0^r B_v(s)' \end{array}\right]\Theta \\ 0 \end{array}\right].$$

Now consider

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

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

$$T^{-1}\sum_{t=2}^{T}S_{t-1}^{\eta}\lambda_t' = T^{-1}\sum_{t=2}^{T}S_{t-1}^{\eta}\left[(\widehat{\mu}-\mu)+x_t'(\widehat{\beta}-\beta),\ 0\right]$$
$$= \left[T^{-1}\sum_{t=2}^{T}T^{-1/2}S_{t-1}^{\eta}T^{1/2}(\widehat{\mu}-\mu)+T^{-1}\sum_{t=2}^{T}T^{-1/2}S_{t-1}^{\eta}T^{-1/2}x_t'T(\widehat{\beta}-\beta),\ 0\right]$$
$$\Rightarrow \left[\int B(r)dr\Theta_\mu+\int B(r)B_v(r)'dr\Theta_\beta,\ 0\right]. \qquad (32)$$

The third term in (31) can be rewritten as

$$T^{-1}\sum_{t=2}^{T}S_{t-1}^{\eta}\lambda_t' = T^{-1}\sum_{t=2}^{T-1}\left(S_{t-1}^{\lambda}-S_t^{\lambda}\right)S_t^{\eta\prime}+T^{-1}S_{T-1}^{\lambda}S_T^{\eta\prime}-T^{-1}\lambda_1\eta_1'$$
$$= T^{-1}S_{T-1}^{\lambda}S_T^{\eta\prime}-T^{-1}\sum_{t=2}^{T-1}\lambda_tS_t^{\eta\prime}-T^{-1}\lambda_1\eta_1' \qquad (33)$$

For the first term in (33) it holds that

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

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

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

It remains to consider the fourth term in (31)

$$T^{-1}\sum_{t=2}^{T}S_{t-1}^{\lambda}\lambda_t' = T^{-1}\sum_{t=2}^{T}\begin{bmatrix}(t-1)(\widehat{\mu}-\mu)+S_{t-1}^{x\prime}(\widehat{\beta}-\beta)\\ 0\end{bmatrix}\begin{bmatrix}(\widehat{\mu}-\mu)+x_t'(\widehat{\beta}-\beta)\\ 0\end{bmatrix}'$$
$$= \begin{bmatrix}T^{-1}\sum_{t=2}^{T}\left[(t-1)(\widehat{\mu}-\mu)+S_{t-1}^{x\prime}\right]\left[(\widehat{\mu}-\mu)+(\widehat{\beta}-\beta)'x_t\right] & 0\\ 0 & 0\end{bmatrix}. \qquad (34)$$

Upon multiplication, the non-zero term in (34) can be written as

$$T\left(\widehat{\mu}-\mu\right)^2 T^{-2}\sum_{t=2}^{T}(t-1)+T^{1/2}\left(\widehat{\mu}-\mu\right)T(\widehat{\beta}-\beta)'T^{-5/2}\sum_{t=2}^{T}(t-1)x_t+$$
$$T^{-5/2}\sum_{t=2}^{T}S_{t-1}^{x\prime}T(\widehat{\beta}-\beta)T^{1/2}\left(\widehat{\mu}-\mu\right)+T(\widehat{\beta}-\beta)'T^{-3}\sum_{t=2}^{T}S_{t-1}^{x}x_t'T(\widehat{\beta}-\beta),$$

from which the limit can immediately be deduced

$$\frac{1}{2}\Theta_\mu^2+\Theta_\mu\Theta_\beta'\int rB_v(r)dr+\int\left(\int_0^r B_v(s)ds\right)dr\Theta_\beta\Theta\mu+\Theta_\beta'\int\left(\int_0^r B_v(s)ds\right)B_v(r)'dr\Theta_\beta. \quad (35)$$

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

$$T^{-1}\sum_{t=2}^{T}\widehat{S}_{t-1}\widehat{\eta}_t' \Rightarrow \int P_{\widehat{\eta}}(r)dP_{\widehat{\eta}}(r)'+\Lambda'. \quad (36)$$

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

$$\begin{aligned}\widehat{\Delta}&=T^{-1}\sum_{i=1}^{T}\sum_{j=i}^{T}k_{ij}\widehat{\eta}_j\widehat{\eta}_i'\\&=\widehat{\Omega}-T^{-1}\sum_{i=2}^{T}\sum_{j=i}^{i-1}k_{ij}\widehat{\eta}_j\widehat{\eta}_i'\\&=\widehat{\Omega}-T^{-1}\sum_{i=2}^{T}A_i\widehat{\eta}_i'-T^{-1}\sum_{i=2}^{T}k_{i,i-1}\widehat{S}_{i-1}\widehat{\eta}_i',\end{aligned}$$

using

$$\sum_{j=1}^{i-1}k_{ij}\widehat{\eta}_j=\underbrace{\sum_{j=1}^{i-2}\left(k_{ij}-k_{i,j+1}\right)\widehat{S}_j}_{=A_i}+k_{i,i-1}\widehat{S}_{i-1}$$

and $\widehat{S}_j=\sum_{i=1}^{j}\widehat{\eta}_j$. Continuing we get

$$\begin{aligned}\widehat{\Delta}&=\widehat{\Omega}-T^{-1}\sum_{i=2}^{T}\left(A_i+k_{i,i-1}\widehat{S}_{i-1}\right)\widehat{\eta}_i'\\&=\widehat{\Omega}-T^{-1}\sum_{i=2}^{T}A_i\widehat{\eta}_i'-T^{-1}\sum_{i=2}^{T}k_{i,i-1}\widehat{S}_{i-1}\widehat{\eta}_i'\\&=\widehat{\Omega}-T^{-1}\sum_{i=2}^{T}\left(A_i-A_{i+1}\right)\widehat{S}_i'-T^{-1}A_T\widehat{S}_T'-T^{-1}\sum_{i=2}^{T}k_{i,i-1}\widehat{S}_{i-1}\widehat{\eta}_i'.\end{aligned}$$

Next insert

$$
\begin{aligned}
A_i - A_{i+1} &= \sum_{j=1}^{i-2} \left(k_{ij} - k_{i,j+1}\right) \widehat{S}_j - \sum_{j=1}^{i-1} \left(k_{i+1,j} - k_{i+1,j+1}\right) \widehat{S}_j \\
&= \sum_{j=1}^{i-1} \left[\left(k_{ij} - k_{i,j+1}\right) - \left(k_{i+1,j} - k_{i+1,j+1}\right)\right] \widehat{S}_j - \left(k_{i+1,i-1} - k_{i+1,i}\right) \widehat{S}_{i-1}
\end{aligned}
$$

in the above expression for $\widehat{\Delta}$ to get

$$
\begin{aligned}
\widehat{\Delta} = \widehat{\Omega} &- T^{-1} \sum_{i=2}^{T-1} \sum_{j=1}^{i-2} \left[\left(k_{ij} - k_{i,j+1}\right) - \left(k_{i+1,j} - k_{i+1,j+1}\right)\right] \widehat{S}_j \widehat{S}_i' + T^{-1} \sum_{i=2}^{T-1} \left(k_{i+1,i-1} - k_{i+1,i}\right) \widehat{S}_{i-1} \widehat{S}_i' \\
&- T^{-1} \sum_{j=1}^{T-2} \left(k_{Tj} - k_{T,j+1}\right) \widehat{S}_j \widehat{S}_T' - T^{-1} \sum_{i=2}^{T} k_{i,i-1} \widehat{S}_{i-1} \widehat{\eta}_i' \\
= \widehat{\Omega} &- T^{-1} \sum_{i=2}^{T-1} \sum_{j=1}^{i-2} \left[\left(k_{ij} - k_{i,j+1}\right) - \left(k_{i+1,j} - k_{i+1,j+1}\right)\right] \widehat{S}_j \widehat{S}_i' + \left[k\left(\frac{2}{M}\right) - k\left(\frac{1}{M}\right)\right] T^{-1} \sum_{i=2}^{T-1} \widehat{S}_{i-1} \widehat{S}_i' \\
&- T^{-1} \sum_{j=1}^{T-2} \left(k_{Tj} - k_{T,j+1}\right) \widehat{S}_j \widehat{S}_T' - k\left(\frac{1}{M}\right) T^{-1} \sum_{i=2}^{T} \widehat{S}_{i-1} \widehat{\eta}_i', \qquad (37)
\end{aligned}
$$

making the dependence upon M explicit again 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 $Q_b\left(P_1, P_2\right)$ and $Q_b^{\Delta}\left(P_1, P_2\right)$ for both types of kernels.

For twice continuously differentiable kernels with $k(0) = 1$ we thus obtain from (37)

$$\begin{aligned}
\widehat{\Delta} &\Rightarrow Q_b\left(P_{\widehat{\eta}}(r), P_{\widehat{\eta}}(r)\right) + \frac{1}{b^2}\int\int_0^r k''\left(\frac{|r-s|}{b}\right)P_{\widehat{\eta}}(s)P_{\widehat{\eta}}(r)'dsdr + \frac{1}{b}k'_+(0)\int P_{\widehat{\eta}}(s)P_{\widehat{\eta}}(s)'ds-\\
&-\frac{1}{b}\int k'\left(\frac{1-s}{b}\right)P_{\widehat{\eta}}(s)P_{\widehat{\eta}}(1)'ds - \int P_{\widehat{\eta}}(s)dP_{\widehat{\eta}}(s)' - \Lambda'\\
&= -\frac{1}{b^2}\int\int k''\left(\frac{|r-s|}{b}\right)P_{\widehat{\eta}}(r)P_{\widehat{\eta}}(s)'dsdr + \frac{1}{b}\int k'\left(\frac{1-s}{b}\right)\left(P_{\widehat{\eta}}(1)P_{\widehat{\eta}}(s)' + P_{\widehat{\eta}}(s)P_{\widehat{\eta}}(1)'\right)ds+\\
&+ P_{\widehat{\eta}}(1)P_{\widehat{\eta}}(1)' + \frac{1}{b^2}\int\int_0^r k''\left(\frac{|r-s|}{b}\right)P_{\widehat{\eta}}(s)P_{\widehat{\eta}}(r)'dsdr + \frac{1}{b}k'_+(0)\int P_{\widehat{\eta}}(s)P_{\widehat{\eta}}(s)'ds-\\
&-\frac{1}{b}\int k'\left(\frac{1-s}{b}\right)P_{\widehat{\eta}}(s)P_{\widehat{\eta}}(1)'ds - \int P_{\widehat{\eta}}(s)dP_{\widehat{\eta}}(s)' - \Lambda'\\
&= -\frac{1}{b^2}\int\int_r^1 k''\left(\frac{|r-s|}{b}\right)P_{\widehat{\eta}}(r)P_{\widehat{\eta}}(s)'dsdr + \frac{1}{b}\int k'\left(\frac{1-s}{b}\right)P_{\widehat{\eta}}(1)P_{\widehat{\eta}}(s)'ds+\\
&+\frac{1}{b}k'_+(0)\int P_{\widehat{\eta}}(s)P_{\widehat{\eta}}(s)'ds + P_{\widehat{\eta}}(1)P_{\widehat{\eta}}(1)' - \int P_{\widehat{\eta}}(s)dP_{\widehat{\eta}}(s)' - \Lambda' = Q_b^{\Delta}\left(P_{\widehat{\eta}}, P_{\widehat{\eta}}\right). \qquad (38)
\end{aligned}$$

For the Bartlett kernel we obtain similarly as above

$$\begin{aligned}
\widehat{\Delta} &\Rightarrow Q_b\left(P_{\widehat{\eta}}(r), P_{\widehat{\eta}}(r)\right) + \frac{1}{b}\int_0^{1-b} P_{\widehat{\eta}}(s)P_{\widehat{\eta}}(s+b)'ds - \frac{1}{b}\int P_{\widehat{\eta}}(s)P_{\widehat{\eta}}(s)'ds+\\
&+\int_{1-b}^1 P_{\widehat{\eta}}(s)P_{\widehat{\eta}}(1)'ds - \int P_{\widehat{\eta}}(s)dP_{\widehat{\eta}}(s)' - \Lambda'\\
&= \frac{2}{b}\int P_{\widehat{\eta}}(s)P_{\widehat{\eta}}(s)'ds - \frac{1}{b}\int_0^{1-b}\left(P_{\widehat{\eta}}(s)P_{\widehat{\eta}}(s+b)' + P_{\widehat{\eta}}(s+b)P_{\widehat{\eta}}(s)'\right)ds-\\
&-\frac{1}{b}\int_{1-b}^1\left(P_{\widehat{\eta}}(1)P_{\widehat{\eta}}(s)' + P_{\widehat{\eta}}(s)P_{\widehat{\eta}}(1)'\right)ds + P_{\widehat{\eta}}(1)P_{\widehat{\eta}}(1)' + \frac{1}{b}\int_0^{1-b} P_{\widehat{\eta}}(s)P_{\widehat{\eta}}(s+b)'ds-\\
&-\frac{1}{b}\int P_{\widehat{\eta}}(s)P_{\widehat{\eta}}(s)'ds + \int_{1-b}^1 P_{\widehat{\eta}}(s)P_{\widehat{\eta}}(1)'ds - \int P_{\widehat{\eta}}(s)dP_{\widehat{\eta}}(s)' - \Lambda'\\
&= \frac{1}{b}\int P_{\widehat{\eta}}(s)P_{\widehat{\eta}}(s)'ds - \frac{1}{b}\int_0^{1-b} P_{\widehat{\eta}}(s+b)P_{\widehat{\eta}}(s)'ds - \frac{1}{b}\int_{1-b}^1 P_{\widehat{\eta}}(1)P_{\widehat{\eta}}(s)'ds+\\
&+ P_{\widehat{\eta}}(1)P_{\widehat{\eta}}(1)' - \int P_{\widehat{\eta}}(s)dP_{\widehat{\eta}}(s)' - \Lambda' = Q_b^{\Delta}\left(P_{\widehat{\eta}}, P_{\widehat{\eta}}\right). \qquad (39)
\end{aligned}$$

The results in (38) and (39) establish (12). The remaining claims of the theorem follow by simply inserting the corresponding sub-matrices 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 $Q_b^+ = Q_b^{\Delta}(B_v, \widehat{B}_u) - Q_b^{\Delta}(B_v, B_v)Q_b(B_v, B_v)^{-1}Q_b(B_v, \widehat{B}_u)$ 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 Theorem 2

We consider the asymptotic behavior of the OLS estimator $\widetilde{\theta} = \left(\widetilde{\delta}', \widetilde{\beta}', \widetilde{\gamma}'\right)'$ of $\theta = (\delta', \beta', 0)'$ in equation (22), i.e. we consider

$$A_{IM}^{-1}\left(\widetilde{\theta} - \theta\right) = \left(T^{-2} A_{IM} S^{\widetilde{x}\prime} S^{\widetilde{x}} A_{IM}\right)^{-1} \left(T^{-2} A_{IM} S^{\widetilde{x}\prime} S^{u}\right), \tag{40}$$

using the notation of the main text. We consider the two terms on the right hand side of (40) 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

$$\left(T^{-2} A_{IM} S^{\widetilde{x}\prime} S^{\widetilde{x}} A_{IM}\right)^{-1} \Rightarrow (\Pi')^{-1}\left(\int g(s)g(s)'ds\right)^{-1}\Pi^{-1}. \tag{41}$$

Analogously, for a typical entry of the second term in (40) 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

$$\begin{aligned} T^{-2} A_{IM} S^{\widetilde{x}\prime} S^{u} &\Rightarrow \Pi \int g(r) B_u(r) dr \\ &= \sigma_{u\cdot v}\Pi\int g(r) w_{u\cdot v} dr + \Pi\int g(r) W_v(r)' dr \lambda_{uv}', \end{aligned} \tag{42}$$

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

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

$$\begin{aligned} (\Pi)^{-1}\left(\int g(r)g(r)'dr\right)^{-1}\int g(r)W_v(r)'dr\lambda_{uv}' &= (\Pi)^{-1}\begin{pmatrix} 0 \\ 0 \\ I_k \end{pmatrix}\lambda_{uv}' \\ &= \begin{pmatrix} 0 \\ 0 \\ \left((\Omega_{vv}^{1/2})'\right)^{-1}\lambda_{uv}' \end{pmatrix}. \end{aligned} \tag{43}$$

From the definition of the respective quantities it follows that the non-zero block at the end of (43) is equal to $\Omega_{vv}^{-1}\Omega_{vu}$.

Altogether, upon subtracting $\Omega_{vv}^{-1}\Omega_{vu}$ this establishes the asymptotic behavior of the OLS estimator $\widetilde{\theta}$. The representation given in (23) then follows using integration by parts and the definition of $G(r)$.

Lemma 1 *Consider the augmented regression* (27) *as given in Section 5 and denote with* $\widetilde{\theta}^* = \left(\widetilde{\delta}^{*\prime}, \widetilde{\beta}^{*\prime}, \widetilde{\gamma}^{*\prime}, \widetilde{\kappa}^{*\prime}\right)'$ *the OLS estimator of the parameter vector* $\theta^* = (\delta', \beta', 0, 0)'$. *It holds that*

$$A_M^{-1}\left(\widetilde{\theta}^* - \theta^*\right) \Rightarrow \sigma_{u\cdot v}\left(\Pi_M'\right)^{-1}\left(\int h(s)h(s)'ds\right)^{-1}\int h(r)w_{u\cdot v}(r)dr + \begin{pmatrix} 0 \\ 0 \\ \Omega_{vv}^{-1}\Omega_{vu} \\ 0 \end{pmatrix} \tag{44}$$

$$= \sigma_{u\cdot v}\left(\Pi_M'\right)^{-1}\left(\int h(s)h(s)'ds\right)^{-1}\int \left[H(1) - H(s)\right]dw_{u\cdot v}(s) + \begin{pmatrix} 0 \\ 0 \\ \Omega_{vv}^{-1}\Omega_{vu} \\ 0 \end{pmatrix},$$

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 of Lemma 1

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. Thus, partition the additional regressors as $z_t = (z_t^{f\prime}, z_t^{S^x\prime}, z_t^{x\prime})'$. The limit, when appropriately scaled (with the appropriate scaling being $T^{-5/2}A_{IM}$), is given by

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

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

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

Combined, this can be written as

$$
\begin{aligned}
T^{-5/2} A_{IM} z_{[rt]} &\Rightarrow \Pi \left(r \int g(s) ds - \int_0^r \left(\int_0^s g(m) dm \right) ds \right) \\
&= \Pi \left(r G(1) - \int_0^r G(s) ds \right) \\
&= \Pi \left(\int_0^r [G(1) - G(s)] ds \right). \qquad (45)
\end{aligned}
$$

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

$$
T^{-5/2} A_{IM} z_{[rT]} T^{-1/2} S^{u}_{[rT]} \Rightarrow \Pi \left(\int_0^r [G(1) - G(s)] ds \right) B_u(r). \qquad (46)
$$

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

$$
\begin{aligned}
A_M^{-1} \left(\widetilde{\theta}^{*} - \theta \right) &\Rightarrow \left(\Pi_M' \right)^{-1} \left(\int h(s) h(s)' ds \right)^{-1} \left(\int h(s) B_u(s) ds \right) \qquad (47) \\
&= \sigma_{u \cdot v} \left(\Pi_M' \right)^{-1} \left(\int h(s) h(s)' ds \right)^{-1} \left(\int h(s) w_{u \cdot v}(s) ds \right) + \begin{pmatrix} 0 \\ 0 \\ \Omega_{vv}^{-1} \Omega_{vu} \\ 0 \end{pmatrix} \\
&= \sigma_{u \cdot v} \left(\Pi_M' \right)^{-1} \left(\int h(s) h(s)' ds \right)^{-1} \left(\int [H(1) - H(s)] d w_{u \cdot v}(s) \right) + \begin{pmatrix} 0 \\ 0 \\ \Omega_{vv}^{-1} \Omega_{vu} \\ 0 \end{pmatrix},
\end{aligned}
$$

where the second line follows from the first using the same argument as in Theorem 1 is used and the third line follows via integration by parts and the definition of $H(r)$.

Proof of Lemma 2

We consider the OLS residuals from (22),

$$
\begin{aligned}
\widetilde{S}^{u}_t &= S^{y}_t - S^{\widetilde{x}\prime}_t \widetilde{\theta} \\
&= S^{u}_t - S^{\widetilde{x}\prime}_t \left(\widetilde{\theta} - \theta \right)
\end{aligned}
$$

and their first differences,

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

Consequently,

$$\begin{aligned}
T^{-1/2}\sum_{t=2}^{[rT]}\Delta\widetilde{S}_t^u &= T^{-1/2}\sum_{t=2}^{[rT]}u_t - T^{-1/2}\sum_{t=2}^{[rT]}\widetilde{x}_t'A_{IM}A_{IM}^{-1}\left(\widetilde{\theta}-\theta\right) \\
&\Rightarrow \sigma_{u\cdot v}w_{u\cdot v}(r) + \lambda_{uv}W_v(r) - g(r)'\Pi'\left\{\sigma_{u\cdot v}\left(\Pi'\right)^{-1}\left(\int g(s)g(s)'ds\right)^{-1}\times\right. \\
&\quad\left.\times\int[G(1)-G(s)]dw_{u\cdot v}(s) + \begin{pmatrix} 0 \\ 0 \\ \Omega_{vv}^{-1}\Omega_{vu}\end{pmatrix}\right\} \\
&= \sigma_{u\cdot v}\left[\int_0^r dw_{u\cdot v}(s) - g(r)'\left(\int g(s)g(s)'ds\right)^{-1}\int[G(1)-G(s)]dw_{u\cdot v}(s)\right] \\
&= \sigma_{u\cdot v}\widetilde{P}(r),
\end{aligned}$$

where the limit given in the second and third line follows from results already discussed in the proof of Theorem 2. In this respect note that $T^{-1/2}\sum_{t=2}^{[rT]}\widetilde{x}_t'A_{IM} = T^{-1/2}S_{[rT]}^{\widetilde{x}\prime}A_{IM} - T^{-1/2}\widetilde{x}_1'A_{IM}$, with the last term vanishing asymptotically. The fourth line follows from the fact that $\lambda_{uv}W_v(r) - g(r)\Pi'\begin{pmatrix} 0 \\ 0 \\ \Omega_{vv}^{-1}\Omega_{vu}\end{pmatrix} = 0$, which follows, using straightforward algebra, from the definition of the involved quantities.

Now consider the residuals from regression (27),

$$\begin{aligned}
\widetilde{S}_t^{u*} &= S_t^y - S_t^{\widetilde{x}\prime}\widetilde{\theta}^* \\
&= S_t^u - S_t^{\widetilde{x}\prime}\left(\widetilde{\theta}^*-\theta^*\right),
\end{aligned}$$

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 above. The remaining steps are exactly the same as before for $\Delta\widetilde{S}_t^u$ before, i.e. we get

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

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 $\widetilde{\theta}$. 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

$$\begin{aligned}\text{Cov}\left(\int [G(1)-G(s)]dw_{u\cdot v}(s), \widetilde{P}^*(r)\right) &= \int_0^r [G(1)-G(s)]'ds - \\ &- h(r)'\left(\int h(s)h(s)'\right)^{-1}\int [H(1)-H(s)][G(1)-G(s)]'ds\end{aligned} \tag{48}$$

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 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 (48) is indeed equal to $h_2(r)'$.

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, $\widehat{\sigma}^2_{u\cdot v}$ or $\widetilde{\sigma}^2_{u\cdot v}$. As a difference to Theorem 2 and Lemma 1 we now include in the notation in the true parameter vector the population parameter for $\gamma = \Omega_{vv}^{-1}\Omega_{vu}$, thus $\theta = (\delta', \beta', \gamma')'$, with $\gamma = \Omega_{vv}^{-1}\Omega_{vu}$ and $\widetilde{\theta}$ denotes the OLS estimator of θ discussed in Theorem 2.

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

$$\left(T^{-2}A_{IM}S^{\widetilde{x}\prime}S^{\widetilde{x}}A_{IM}\right)^{-1}\left(T^{-4}A_{IM}C'CA_{IM}\right)\left(T^{-2}A_{IM}S^{\widetilde{x}\prime}S^{\widetilde{x}}A_{IM}\right)^{-1}, \tag{49}$$

with C defined in the main text. For the outer terms (that are inverted) the limit has already been established in the proof of Theorem 2 in equation (41) 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 (49) converges – obviously up to the scalar $\sigma^2_{u\cdot v}$ – to V_{IM} as given in (24) in the main text.

Under the null hypothesis both of the two defined statistics can be – as is usual in a linear regression model – written as

$$\begin{aligned}\check{W} &= (R\widetilde{\theta}-r)'\left[RA_{IM}\check{V}_{IM}A_{IM}R'\right]^{-1}(R\widetilde{\theta}-r) \\ &= \left(R(\widetilde{\theta}-\theta)\right)'\left[RA_{IM}\check{V}_{IM}A_{IM}R'\right]^{-1}\left(R(\widetilde{\theta}-\theta)\right) \\ &= \left(A_R^{-1}RA_{IM}A_{IM}^{-1}(\widetilde{\theta}-\theta)\right)'\left[A_R^{-1}RA_{IM}\check{V}_{IM}A_{IM}R'\left(A_R^{-1}\right)'\right]^{-1}\left(A_R^{-1}RA_{IM}A_{IM}^{-1}(\widetilde{\theta}-\theta)\right)\end{aligned} \tag{50}$$

Now, by assumption the restriction matrix asymptotically fulfills $\lim_{T\to\infty} A_R^{-1} R A_{IM} = R^*$, 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 $\sigma^2_{u\cdot v}$ using $\widehat{V}_{IM}$ it follows that

$$\widehat{W} \Rightarrow (R^*\Phi(V_{IM}))' \left(R^* V_{IM} R^{*\prime}\right)^{-1} (R^*\Phi(V_{IM})) \sim \chi^2_q. \tag{51}$$

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

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

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 $\Delta\widetilde{S}^u_t$, $\widetilde{\sigma}^2_{u\cdot v}$, asymptotically coincides with the long run variance estimator of $u^+_t - v'_t(\widetilde{\gamma} - \gamma)$. Define $\eta^{+\prime}_t = [u^+_t, v'_t]$ and its long run variance

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

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

Next note that

$$u^+_t - v'_t(\widetilde{\gamma} - \gamma) = \eta^{+\prime}_t \begin{bmatrix} 1 \\ -(\widetilde{\gamma} - \gamma) \end{bmatrix},$$

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

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

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

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

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

$$\widetilde{W}^* \Rightarrow (R^*\Phi(V_{IM}))' \left(Q_b(\widetilde{P}^*, \widetilde{P}^*)R^*V_{IM}R^{*\prime}\right)^{-1} (R^*\Phi(V_{IM})) \sim \frac{\chi_q^2}{Q_b(\widetilde{P}^*, \widetilde{P}^*)}, \tag{52}$$

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 (52) – 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.

Table 1: Finite Sample Bias and RMSE of the Various Estimators of β_1, $T = 100$

Panel A: Bias

					FM-OLS, Bartlett Kernel							
ρ_1	ρ_2	OLS	IM-OLS	DOLS	M=6	10	30	50	70	90	100	AND
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

					FM-OLS, Bartlett Kernel							
ρ_1	ρ_2	OLS	IM-OLS	DOLS	M=6	10	30	50	70	90	100	AND
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

Table 2: Empirical Null Rejection Probabilities, 0.05 Level, t-tests for $H_0 : \beta_1 = 1$
Data Dependent Bandwidths and Lag Lengths

Panel A: T=100

		Bartlett kernel					QS kernel				
ρ_1, ρ_2	OLS	DOLS	FM	IM(O)	IM(D)	IM(fb)	DOLS	FM	IM(O)	IM(D)	IM(fb)
0.0	.0544	.2610	.1110	.0802	.0736	.0570	.2568	.1322	.0926	.0856	.0450
0.3	.1608	.3836	.1378	.1038	.1004	.0652	.3760	.1414	.1020	.0986	.0836
0.6	.4126	.5390	.2182	.1444	.1518	.1198	.5296	.2094	.1284	.1378	.0556
0.9	.7700	.7074	.5592	.4412	.4256	.5480	.7038	.5392	.4194	.4082	.3936

Panel B: T=200

		Bartlett kernel					QS Kernel				
ρ_1, ρ_2	OLS	DOLS	FM	IM(O)	IM(D)	IM(fb)	DOLS	FM	IM(O)	IM(D)	IM(fb)
0.0	.0484	.0796	.0822	.0722	.0628	.0392	.0794	.0878	.0766	.0672	.0324
0.3	.1592	.1280	.0996	.0892	.0812	.0776	.1168	.0968	.0816	.0736	.0582
0.6	.4204	.2240	.1640	.1092	.1070	.0920	.2084	.1484	.0964	.0920	.0552
0.9	.7712	.4256	.4890	.3212	.2942	.4280	.4058	.4768	.3066	.2780	.4564

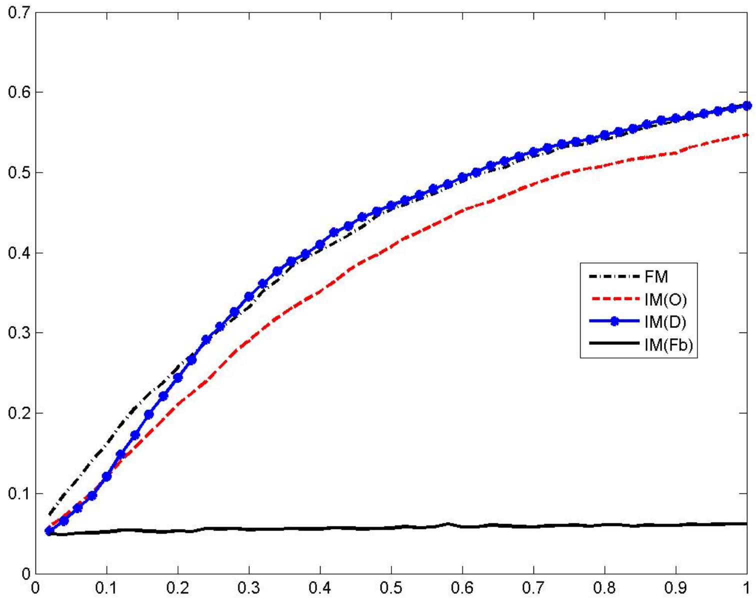
Table 3: Empirical Null Rejection Probabilities, 0.05 Level, *Wald* tests for $H_0 : \beta_1 = 1, \beta_2 = 1$
Data Dependent Bandwidths and Lag Lengths

Panel A: T=100

		Bartlett kernel					QS kernel				
ρ_1, ρ_2	OLS	DOLS	FM	IM(O)	IM(D)	IM(fb)	DOLS	FM	IM(O)	IM(D)	IM(fb)
0.0	.0578	.3442	.1340	.0972	.0890	.0612	.3394	.1668	.1172	.1074	.0426
0.3	.2158	.5180	.1786	.1330	.1258	.0692	.5054	.1868	.1300	.1256	.1024
0.6	.5772	.7250	.3002	.1970	.2070	.1538	.7126	.2840	.1702	.1900	.0568
0.9	.9372	.8990	.7624	.6378	.6114	.7498	.8948	.7322	.6080	.5856	.5418

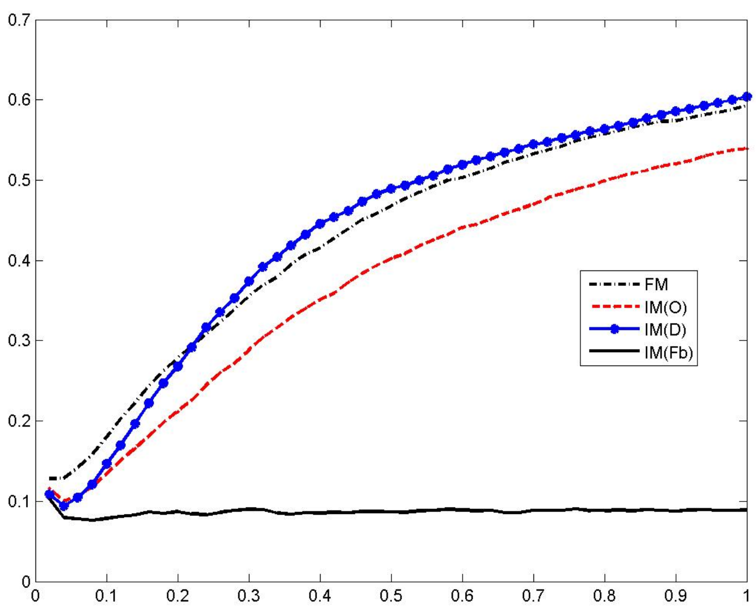
Panel B: T=200

		Bartlett kernel					QS Kernel				
ρ_1, ρ_2	OLS	DOLS	FM	IM(O)	IM(D)	IM(fb)	DOLS	FM	IM(O)	IM(D)	IM(fb)
0.0	.0512	.0838	.0950	.0748	.0666	.0356	.0836	.1052	.0808	.0724	.0248
0.3	.2028	.1620	.1218	.1004	.0920	.0846	.1470	.1194	.0906	.0832	.0608
0.6	.5752	.3000	.1964	.1324	.1314	.1074	.2746	.1776	.1130	.1140	.0540
0.9	.9432	.6002	.6686	.4524	.4140	.5950	.5758	.6536	.4330	.3920	.6276



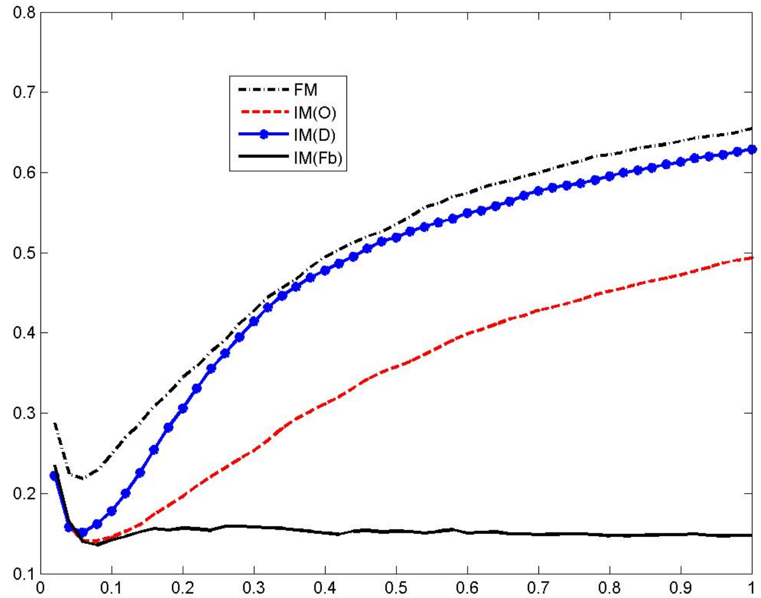
b

Figure 1: Empirical Null Rejections, t-test, $T = 100$, $\rho_1 = \rho_2 = 0.0$, Bartlett Kernel



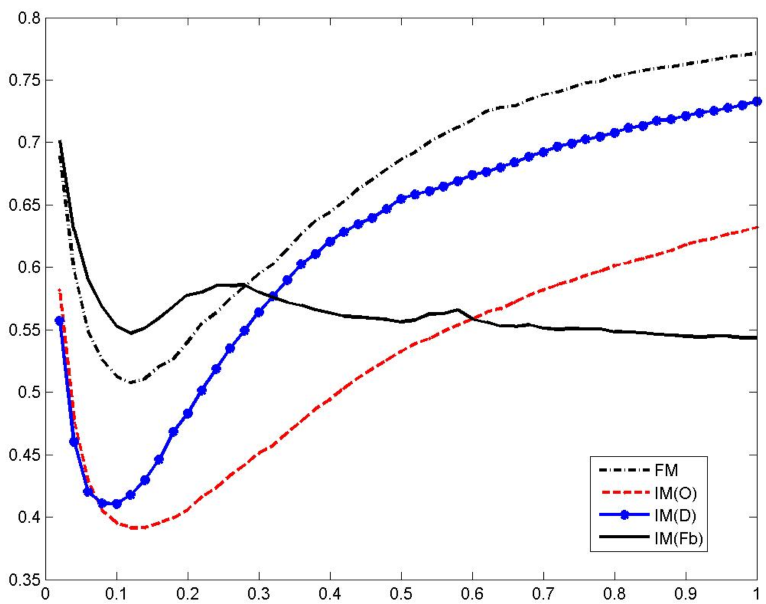
b

Figure 2: Empirical Null Rejections, t-test, $T = 100$, $\rho_1 = \rho_2 = 0.3$, Bartlett Kernel



b

Figure 3: Empirical Null Rejections, t-test, $T = 100$, $\rho_1 = \rho_2 = 0.6$, Bartlett Kernel



b

Figure 4: Empirical Null Rejections, t-test, $T = 100$, $\rho_1 = \rho_2 = 0.9$, Bartlett Kernel

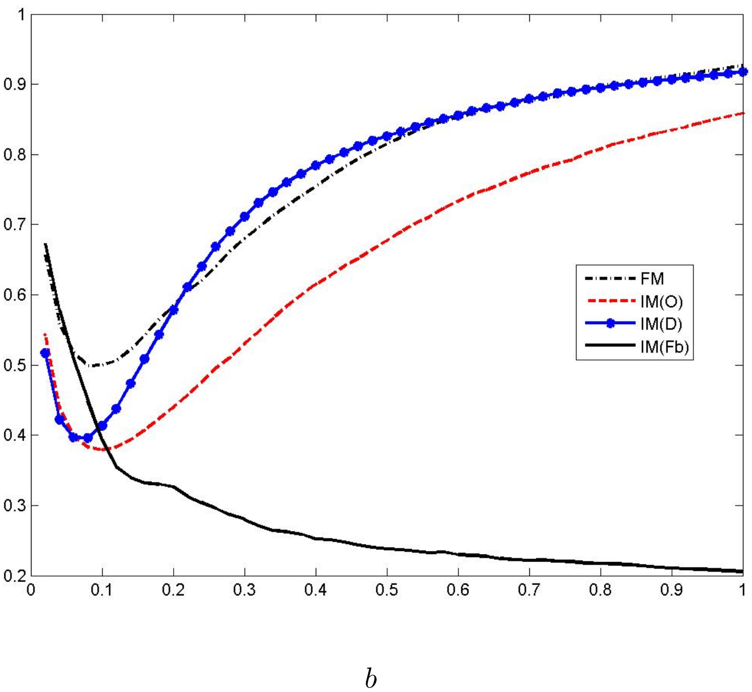


Figure 5: Empirical Null Rejections, t-test, $T = 100$, $\rho_1 = \rho_2 = 0.9$, QS Kernel

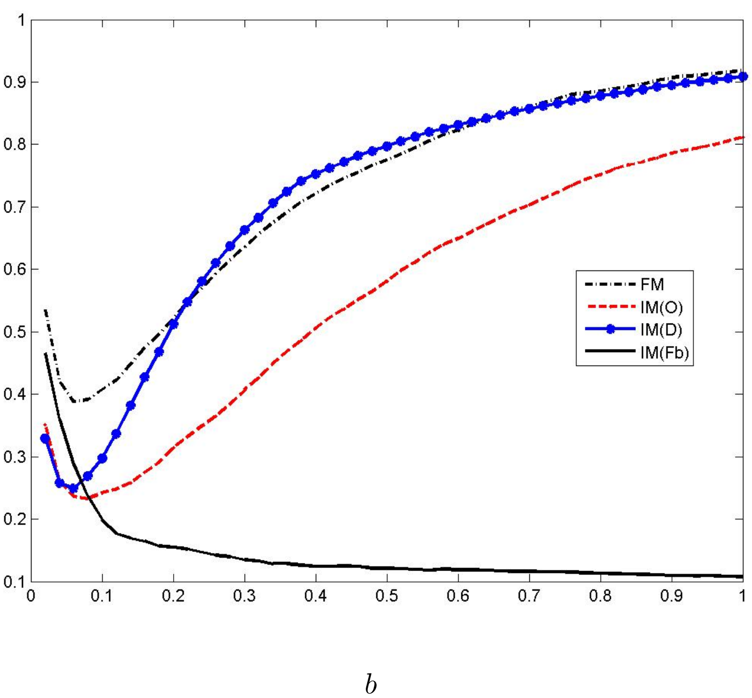
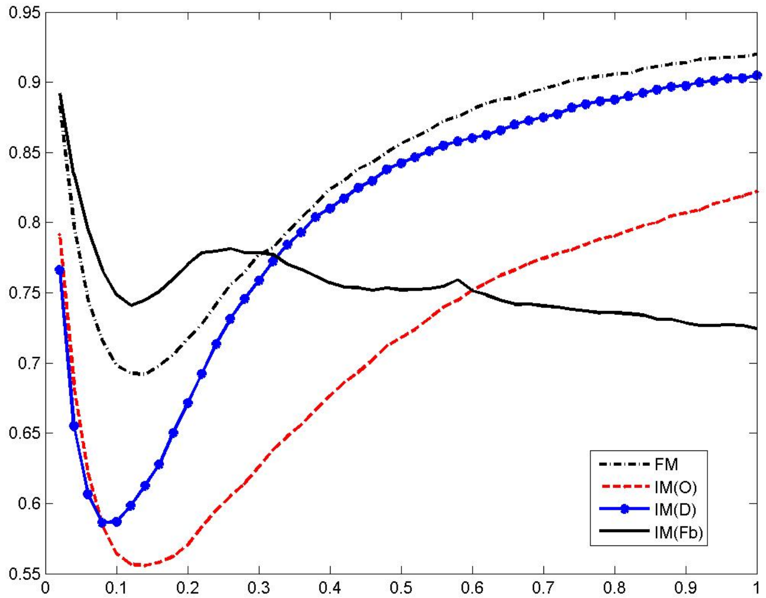
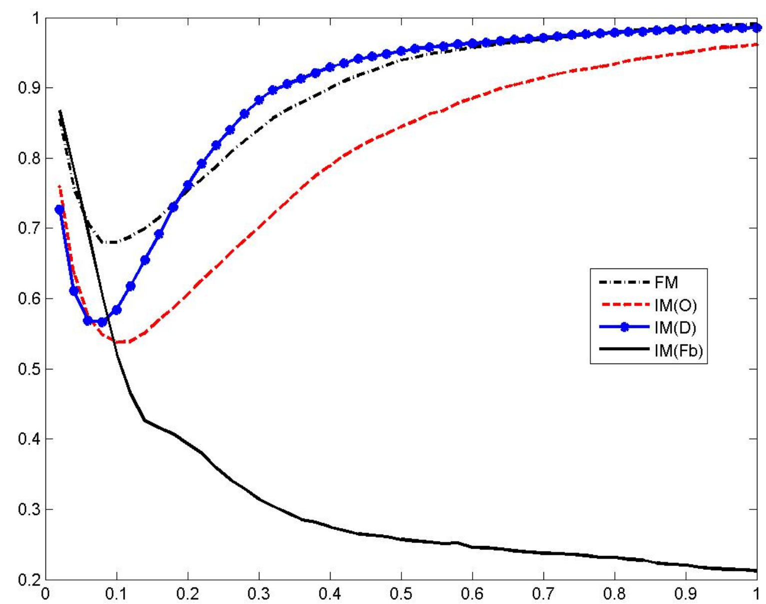


Figure 6: Empirical Null Rejections, t-test, $T = 200$, $\rho_1 = \rho_2 = 0.9$, QS Kernel



b

Figure 7: Empirical Null Rejections, *Wald* test, $T = 100$, $\rho_1 = \rho_2 = 0.9$, Bartlett Kernel



b

Figure 8: Empirical Null Rejections, *Wald* test, $T = 100$, $\rho_1 = \rho_2 = 0.9$, QS Kernel

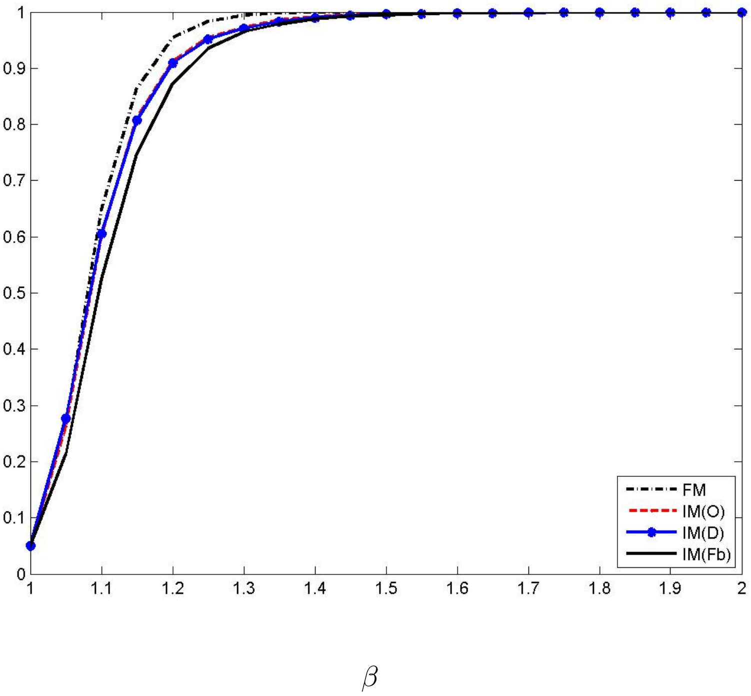


Figure 9: Size Adjusted Power, *Wald* test, $T = 100$, $\rho_1 = \rho_2 = 0.6$, QS Kernel, $b = 0.06$

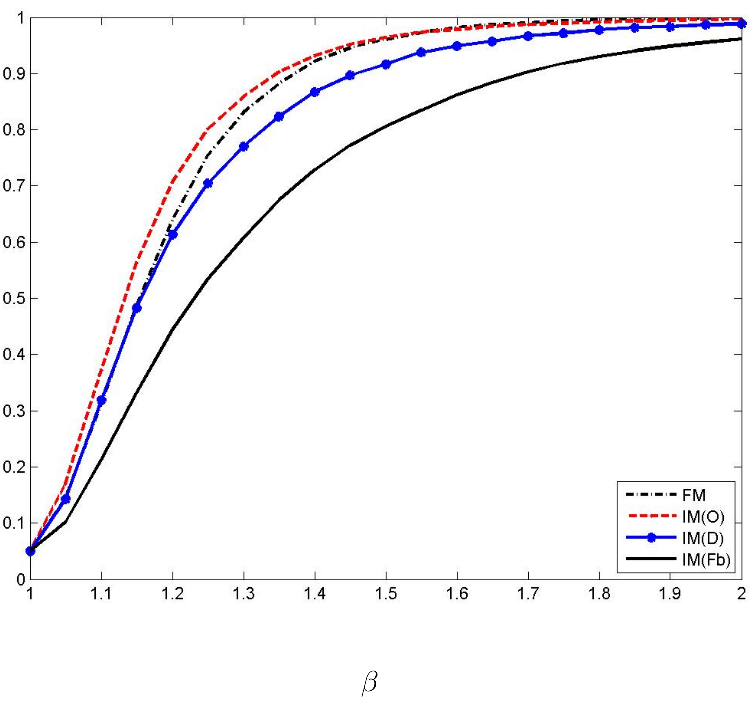
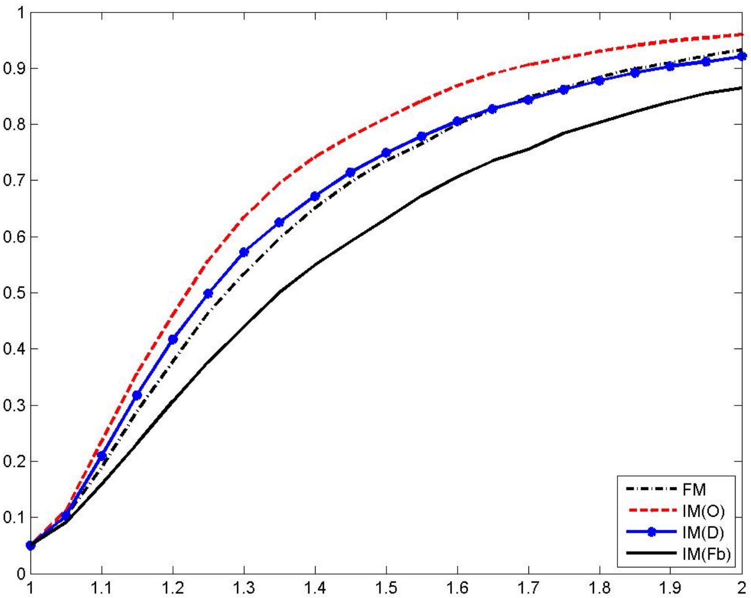
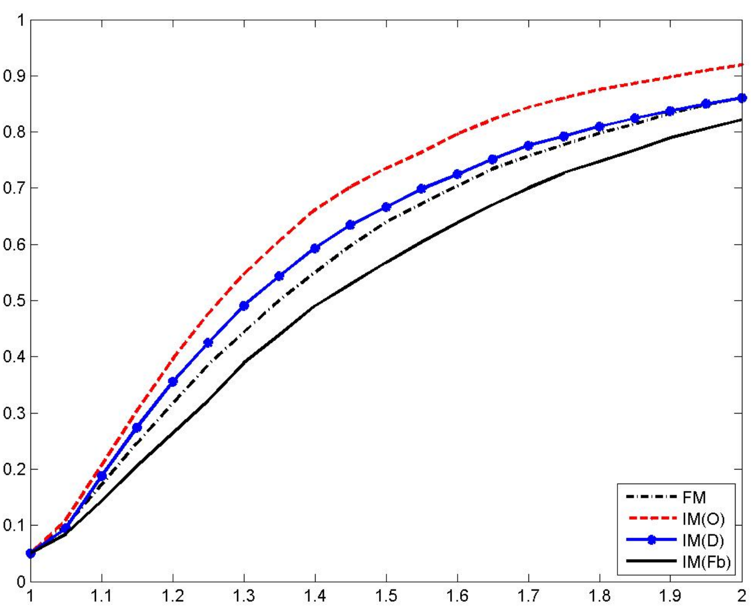


Figure 10: Size Adjusted Power, *Wald* test, $T = 100$, $\rho_1 = \rho_2 = 0.6$, QS Kernel, $b = 0.3$



β

Figure 11: Size Adjusted Power, *Wald* test, $T = 100$, $\rho_1 = \rho_2 = 0.6$, QS Kernel, $b = 0.7$



β

Figure 12: Size Adjusted Power, *Wald* test, $T = 100$, $\rho_1 = \rho_2 = 0.6$, QS Kernel, $b = 1.0$

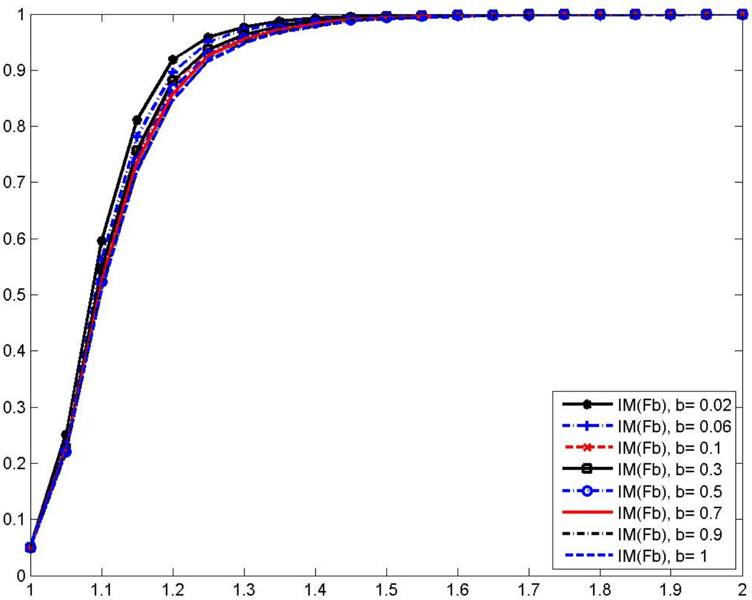


Figure 13: Size Adjusted Power, *Wald* test, $T = 100$, $\rho_1 = \rho_2 = 0.6$, Bartlett Kernel

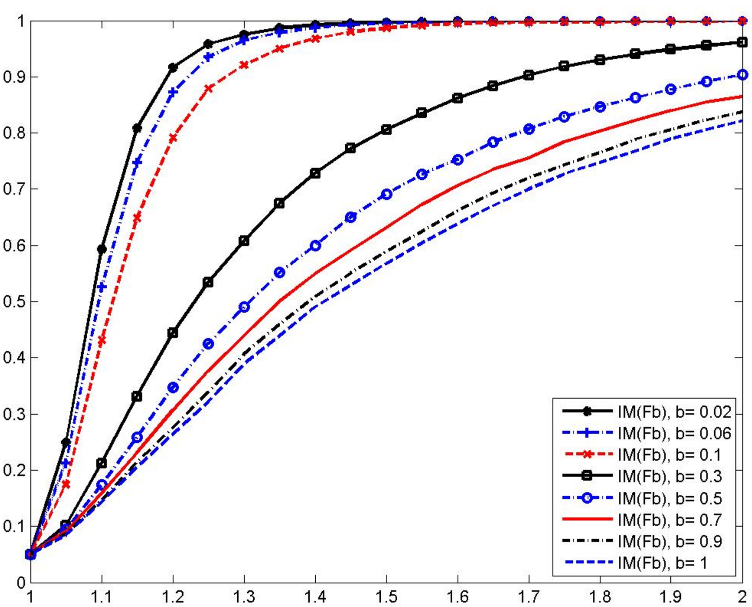
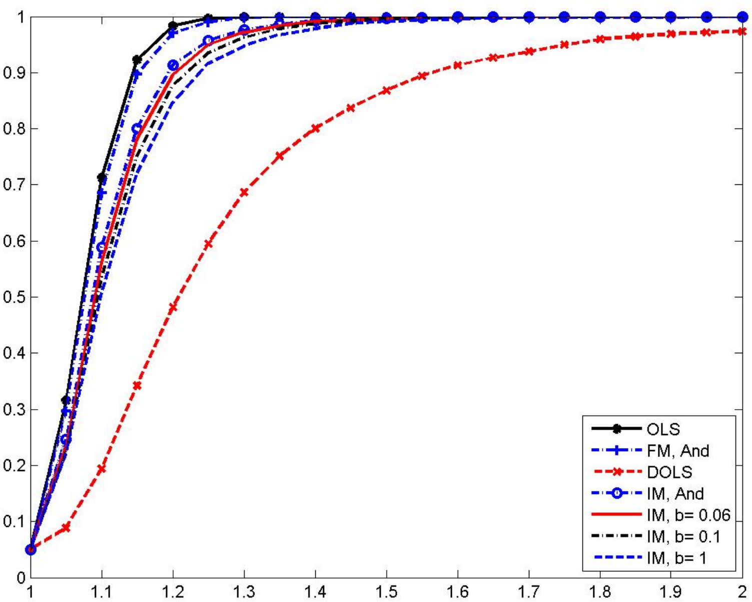
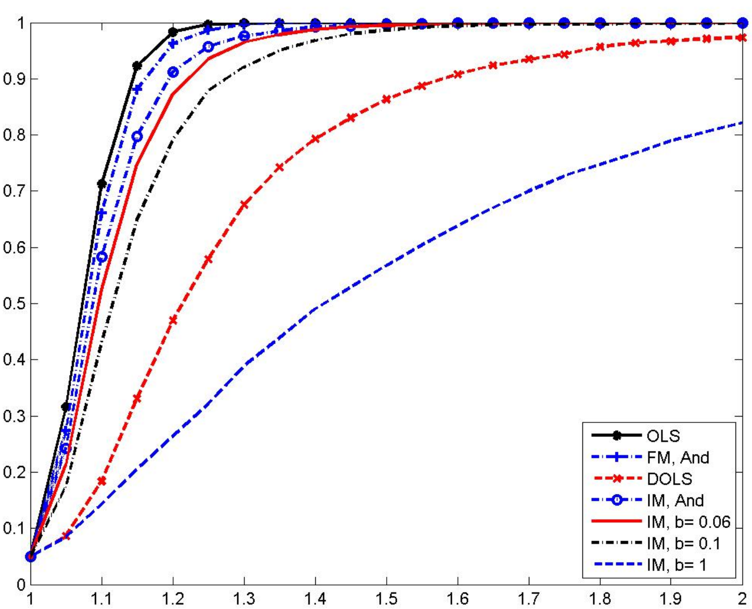


Figure 14: Size Adjusted Power, *Wald* test, $T = 100$, $\rho_1 = \rho_2 = 0.6$, QS Kernel



β

Figure 15: Size Adjusted Power, *Wald* test, $T = 100$, $\rho_1 = \rho_2 = 0.6$, Bartlett Kernel



β

Figure 16: Size Adjusted Power, *Wald* test, $T = 100$, $\rho_1 = \rho_2 = 0.6$, QS Kernel

Authors: Timothy J. Vogelsang, Martin Wagner

Title: Integrated Modified OLS Estimation and Fixed-b Inference for Cointegrating Regressions

Reihe Ökonomie / Economics Series 263

Editor: Robert M. Kunst (Econometrics)
Associate Editors: Walter Fisher (Macroeconomics), Klaus Ritzberger (Microeconomics)

ISSN: 1605-7996

ISSN: 1605-7996