Covariance-based orthogonality tests for regressors with unknown persistence

Alex Maynard and Katsumi Shimotsu*

Department of Economics, University of Toronto and
Department of Economics, Queen’s University

December 15, 2004

Abstract

This paper develops a new covariance-based test of orthogonality that may be attractive as a useful alternative to regression-based tests when conditioning variables have roots close or equal to unity. In this case standard regression-based orthogonality tests can suffer from both size distortion and possible mis-specification under the alternative hypothesis. The new test statistic has a standard normal limit distribution for both unit root and local to unity conditioning variables, without prior knowledge of the local to unity parameter. If the conditioning variable is stationary the test remains conservative and consistent. Thus the new test requires neither size correction nor unit root pre-test. Moreover, it is capable of detecting a wider range of alternatives to orthogonality than regression-based tests. Asymptotic results are derived, including the asymptotic normality of the one-sided long-run covariance estimator, for which frequency domain representations are inapplicable. Simulations suggest good small sample performance. As an empirical application we test for the predictability of stock returns using two persistent regressors, the dividend-price-ratio and short-term interest rate. The results suggest a dominant effect of size distortion in the case of the dividend-price ratio and regression mis-specification in the case of the interest rate.

JEL Classification: C12,C22
Keywords: orthogonality test, covariance estimation, local-to-unity, unit roots, market efficiency, predictive regression

*Both authors are presenting authors and correspondence may be sent to either amaynard@chass.utoronto.ca (A. Maynard) or shimotsu@qed.econ.queensu.ca (K. Shimotsu). We thank Don Andrews, Dietmar Bauer, John Chao, John Galbraith, Sainan Jin, John Maheu, Angelo Melino, John Rust, Sam Thompson, Tim Vogelsang, Yoon-Jae Whang, and participants at the 2004 North American Summer and Far Eastern Econometric Society meetings, the CESG, MESG, and the Columbia, Concordia, Cornell, University of Maryland, and Yale Econometrics seminars for useful discussion. Part of this work was completed while the authors were visiting the Cowles Foundation and we gratefully acknowledge their hospitality. Maynard thanks the SSHRC for research funding.
1 Introduction

This paper develops a new covariance-based method for testing orthogonality conditions that may be attractive when the conditioning variable has a root close or equal to unity. This new method provides a single asymptotic t-test that has correct size when the conditioning variable is modelled as either a unit root or local to unity process (with finite c). It has conservative size, but remains consistent, when the conditioning variable is stationary. It thus provides a sound basis for inference without reference to prior knowledge, estimates, or pre-tests regarding the size of the root.

Furthermore, we note that standard regression based tests are not designed to detect empirically relevant alternatives to orthogonality for which the conditioning variable (e.g. interest rates, dividend yields) is near nonstationary, but the dependent variable (e.g. stock returns) is presumed stationary. By contrast, the covariance based test is explicitly designed to maintain power against such unbalanced alternatives.

To fix ideas, consider a test of the orthogonality condition

\[ H_0 : E[y_t | I_{x,t-1}] = 0 \] (1)

where \( I_{x,t-1} = \sigma(x_{t-1}, x_{t-2}, x_{t-3}, \ldots) \), denotes the information contained in the past history of \( x_t \). Several common empirical applications may be cast in this form, including tests of stock return predictability, forward rate unbiasedness,\(^1\) the permanent income hypothesis, the expectations hypothesis of the term structure, and the constant real interest rate hypothesis.\(^2\)

Although the alternative \( H_A : E[y_t | I_{x,t-1}] \neq 0 \) implicit in (1) is quite general, it is common in empirical work to test this orthogonality condition via the parameter restriction \( \beta_1 = 0 \) in simple regressions of the form

\[ y_t = \beta_0 + \beta_1 x_{t-1} + \varepsilon_{1t}. \] (2)

Such tests were traditionally formulated with stationary regressors in mind. However, it has come to be understood that many of the regressors, such as interest rates, dividend-price ratios, and forward premia are highly serially correlated and may be well characterized by roots near unity (e.g. Mankiw and Shapiro (1986), Stambaugh (1999)).

The near nonstationarity of \( x_t \) poses two problems for the orthogonality test in (1) when formulated as in (2). The first is the well known size problem (Mankiw and Shapiro (1986), Cavanagh et al. (1995), and Stambaugh (1999)) that arises when \( x_{t-1} \) is predetermined, but not strictly exogenous, and has a root near unity. The size distortion is known to depend on the local to unity parameter and is not solved by two stage inference based on unit root pre-test (Cavanagh et al. (1995), Elliott (1998)).

The size problem in (2) has recently generated an active literature. Two basic approaches have been explored. The first is to maintain the regression specification

\(^1\)See Maynard (2004) for an application to the forward rate unbiasedness test.

\(^2\)See Mankiw and Shapiro (1986) and references therein.
in (2), but correct size. In a local-to-unity context, solutions of this type include two stage bounds procedures (Cavanagh et al. (1995), Torous et al. (2005), Valkanov (2003), Campbell and Yogo (2003)), reformulation of the problem as a stationarity test on \( y_t \) (Wright (2000), Lanne (2002)), and conditionally optimal inference employing sufficient statistics (Jansson and Moreira (2004)).

A second approach is to search for alternative estimators or testing methods for (1). In this vein, Campbell and Dufour (95 & 97) provide exact tests of (1) using non-parametric sign and sign rank tests, while Toda and Yamamoto (1995) (see also Saikkonen and Lütkepohl (1996)) have also shown that, by choosing the lag order sufficiently large, one can estimate VAR’s formulated in levels and test general parameter restrictions even if the the order of integration of the process is unknown.

Our test also falls into this second category, motivated in part by what we view as a potential for mis-specification under the alternative hypothesis (\( H_A : E[y_t|I_{x,t-1}] \neq 0 \)), that arises specifically when (2) is used as a means of testing the orthogonality condition in (1). In particular, it often occurs in empirical applications that \( y_t \) (e.g. stock or exchange rate returns) appears clearly stationary on both empirical and a priori theoretical grounds. For example, stock returns show little persistence, and an I(1) stock return process would imply an I(2) stock price, which may lack plausibility. By contrast, many of the variables used in \( x_t \) (e.g. dividend price ratio, interest rates, forward premia) exhibit roots near unity. This suggests the possibility of regression imbalance in (2). In this case, the regression specification (2) allows for only two possibilities: either \( y_t \) is unpredictable by \( I_{x,t-1} \) (\( \beta_1 = 0 \)) or \( y_t \) must be nonstationary and cointegrated with \( x_t \) (\( \beta_1 \neq 0 \)). This rules out a wide class of alternatives to the orthogonality condition in (1) in which \( y_t \) is stationary, yet predictable based on \( I_{x,t-1} \). For example, were it known a priori that \( y_t \sim I(0) \) and \( x_t \sim I(1) \) then a more appropriate alternative to (1) would be given by \( \gamma_1 \neq 0 \) in

\[
y_t = \gamma_0 + \gamma_1 (x_{t-1} - x_{t-2}) + \varepsilon_{1t},
\]

Likewise, if \( x_t \) is modelled local-to-unity as

\[
x_t = (1 + \frac{c}{n}) x_{t-1} + u_t, \quad t = 1, 2, \ldots, n, \quad n = 1, 2, \ldots \quad c < 0,
\]

with \( x_t \equiv 0 \) for \( t \leq 0 \), then one might consider the quasi-differenced alternative

\[
y_t = \gamma_0 + \gamma_1 \left( x_{t-1} - \left(1 + \frac{c}{n}\right) x_{t-2} \right) + \varepsilon_{1t},
\]

with \( \gamma_1 \neq 0 \). However, these parametric approaches require either a unit root pre-test, or a priori knowledge of \( c \).

---

3 Other solutions, often in more tightly parameterized models, include finite sample size corrections (Stambaugh (1999), Lewellen (2003)), Bayesian approaches (Elliott and Stock (1994), Stambaugh (1999), Lewellen (2003)), and resampling approaches (Nelson and Kim (1993), Goetzmann and Jorion (1993), Wolf (2000)).

4 This specification issue does not arise in the more general context when (2) itself is the primary object of interest, as in stationary or (near) cointegrating regression.

5 This latter point has also been recognized by Wright (2000) and Lanne (2002) who provide inference procedures for \( \beta_1 \) using the KKPS stationarity test.
Without pre-test, or prior knowledge on \(c\), we develop a covariance-based test of (1) that incorporates the same alternatives as (2) when \(x_t\) is stationary, but includes more general alternatives of the type described by (5), when \(x_t\) is modelled as local to unity or unit root nonstationary. Moreover, the test avoids the problem of over-rejection discussed above.

We begin with the following intuition. Consider first \(x_t\) stationary \((I(0))\). The orthogonality condition (1) then imposes \(\text{cov}(y_t, x_{t-1}) = 0\) under the null hypothesis. Next, rewrite \(x_{t-1}\) as an infinite sum of its past first-differences:

\[
x_{t-1} = (x_{t-1} - x_{t-2}) + (x_{t-2} - x_{t-3}) + \ldots = \Delta x_{t-1} + \Delta x_{t-2} + \ldots
\]

This purely algebraic decomposition then allows us to rewrite the contemporaneous covariance between \(y_t\) and \(x_{t-1}\) in terms of a (one-sided) long run covariance between \(y_t\) and the first-difference \(\Delta x_{t-1}\) as

\[
\text{cov}(y_t, x_{t-1}) = \sum_{h=1}^{\infty} \text{cov}(y_t, \Delta x_{t-h}) \quad (6)
\]

The next step is to extend this decomposition to the case where \(x_t\) follows a unit root \((I(1))\) process. In particular, we can define a contemporaneous covariance between \(y_t\) and \(x_{t-1}\) in analogous fashion, as the long-run covariance between \(y_t\) and the first-difference of \(x_{t-1}\), as

\[
\text{cov}(y_t, x_{t-1}) = \sum_{h=1}^{t-1} \text{cov}(y_t, \Delta x_{t-h})
\]

initializing \(x_t\) at \(t = 0\). Assume \(y_t\) and \(\Delta x_{t-1}\) are stationary. We then define a quasi-covariance between \(x_{t-1}\) and \(y_t\) as

\[
\lambda_{y, \Delta x} := \lim_{t \to \infty} \sum_{h=1}^{t-1} \text{cov}(y_t, \Delta x_{t-h}) = \sum_{h=1}^{\infty} \text{cov}(y_t, \Delta x_{t-h}) \quad (7)
\]

which is well-defined if \(\sum_{h=0}^{\infty} |\text{cov}(y_t, \Delta x_{t-h})| < \infty\).\(^6\) As seen from (6), when \(x_t\) is stationary, the quasi-covariance is written as

\[
\lambda_{y, \Delta x} = \text{cov}(y_t, x_{t-1}) \quad (8)
\]

Therefore, \(\lambda_{y, \Delta x}\) is well-defined for both \(x_t \sim I(1)\) and \(x_t \sim I(0)\) and provides either an exact \((x_t \sim I(0))\) or an asymptotic \((x_t \sim I(1))\) measure of the contemporaneous covariance between \(y_t\) and \(x_{t-1}\).

The quasi-covariance maintains the same definition under the local-to-unity model (4) and as shown in the Appendix, when \(\sum_{p=1}^{\infty} p|\text{cov}(y_t, u_{t-p})| < \infty\), it takes the form

\[
\lambda_{y, \Delta x} = \sum_{h=1}^{\infty} \text{cov}(y_t, u_{t-h}) + O\left(n^{-1}\right) \quad (9)
\]

\(^6\)When \(x_0 \neq 0\), \(\text{cov}(y_t, x_{t-1})\) is defined as \(\sum_{h=1}^{t-1} \text{cov}(y_t, \Delta x_{t-h}) + \text{cov}(y_t, x_0)\) and (7) continues to apply under the relatively weak and reasonable assumption that \(\lim_{t \to \infty} \text{cov}(y_t, x_0) \to 0\).
Note that the quasi-covariance is defined in the same way in all three cases.

We propose a test of the orthogonality condition (1) based on the restriction \( \lambda_y, \Delta x = 0 \). Regardless of the order of integration in \( x_t \), it is clear that this provides a valid basis for testing orthogonality, as this restriction is an implication of the null hypothesis in (1) since \( \Delta x_{t-h} \) belongs to \( I_{x,t-1} \) for \( h \geq 1 \). For \( x_t \) stationary the specification of our alternative is the same as in (2). This follows since \( \beta_1 = \text{cov}(y_t, x_{t-1}) / \text{var}(x_t) \) has a finite dominator, so that \( \beta_1 = 0 \) if and only if \( \lambda_y, \Delta x = \text{cov}(y_t, x_{t-1}) = 0 \).

The difference arises when \( x_t \) is modelled as nonstationary, or local to unity. As discussed above, (2) now allows only those alternatives for which \( y_t \) contains a unit (or near unit) root component, omitting alternatives such as (3) and (5). By contrast, \( \lambda_y, \Delta x \) is non-zero for a broader range of alternatives to (1), including unbalanced alternatives \( (y_t \sim I(0), x_t \sim I(1) \) or near \( I(1) \)).\(^7\) For example, under the alternatives specified by \( \gamma_1 \neq 0 \) in (5), assuming \( \varepsilon_{1t} \) is independent of past \( u_t \), it follows from (9) that

\[
\lambda_y, \Delta x = \gamma_1 \sum_{h=1}^{\infty} \text{cov}(u_{t-1}, u_{t-h}) + O(n^{-1}) \neq 0,
\]
in general. On the other hand, the regression coefficient \( \hat{\beta}_1 = n^{-1} \sum_{t=2}^{n} (y_t - \bar{y}) x_{t-1} / n^{-1} \sum_{t=2}^{n} (x_{t-1} - \bar{x})^2 \) tends to zero by standard argument. This is not because \( \lim_{t \to \infty} \text{cov}(y_t, x_{t-1}) = \lambda_y, \Delta x \) is zero, but rather because \( \text{var}(x_t) \) diverges.

The parameter \( \lambda_y, \Delta x \) is well defined and consistently estimated by the same standard kernel covariance estimator for both stationary and unit root nonstationary \( x_t \), without the necessity of pretesting or estimating the root of \( x_t \). The feature may be useful in applied work, as it is often difficult to distinguish with confidence between \( I(0) \) and \( I(1) \) alternatives. A second desirable property of the estimator is that it is shown to have a unique limit distribution for all values of the local to unity parameter \( c \). This allows us to avoid two-stage inference procedures, such as Bonferroni bounds, that are often necessitated by the lack of a consistent time-series estimator for \( c \). We construct a large sample test, based on a single \( t \)-statistic with a limiting standard normal distribution under both unit root and local to unity assumptions. No bias corrections or other adjustments are required. The test is shown to remain conservative and consistent when \( x_t \) is stationary. As a by product, we establish the asymptotic normality of the one-sided long-run covariance estimator, for which frequency domain representations are inapplicable.

As an empirical application, these methods are used to revisit well-known orthogonality tests involving the prediction of stock returns using dividend-yields and interest rates. Both variables are highly persistent leading much recent literature to explore size distortions. Moreover, while there is strong reason to believe that stock returns are stationary, standard regression-based tests, even if even-size corrected, may restrict power to alternatives that imply near-nonstationary components in stock returns. By using covariance-based tests we not only ensure valid inference,\(^7\) While not our primary focus, it can also be seen that cointegration alternatives, for which \( y_t \) and \( x_t \) are cointegrated unit roots lie in the alternative \( H_A : E[y_t | x_{t-1}] \neq 0 \) since \( \lambda_y, \Delta x \) is then infinite.
but also allow for alternatives that leave returns stationary, while still violating orthogonality. The results suggest a dominant effect of size distortion in the case of the dividend-price ratio and regression mis-specification in the case of the interest rate.

The remainder of the paper is organized as follows. Section 2 introduces the kernel-based estimator of $\lambda_y, \Delta x$ and demonstrates its asymptotic behavior when $x_t$ is $I(1), I(0),$ and local-to-unity. Section 3 discusses how to conduct inference based on the estimate of $\lambda_y, \Delta x,$ and Section 4 reports some simulation results. The empirical application is reported in Section 5, and Section 6 concludes. Proofs are given in the Appendix in Section 7, and Section 8 collects some technical results.

2 Estimation of quasi-covariance

In this section, we develop an estimator of the quasi-covariance and derive its asymptotic properties. First we consider the case when $x_t$ is $I(1)$.

Assumption A

$(y_t, \Delta x_t)$ are generated by

$$z_t = \left( \begin{array}{c} y_t \\ \Delta x_t \end{array} \right) = A(L) \varepsilon_t = \sum_{j=0}^{\infty} A_j \varepsilon_{t-j}, \quad \sum_{j=0}^{\infty} j \| A_j \| < \infty, \quad (10)$$

$$\varepsilon_t \sim \text{i.i.d.} (0, I_2), \quad \text{with finite fourth moment,}$$

$$\sum_{h=-\infty}^{\infty} |h|^\delta \| \Gamma(h) \| < \infty, \quad \delta > 1; \quad \Gamma(h) = \begin{bmatrix} \Gamma_{yy}(h) & \Gamma_{y\Delta x}(h) \\ \Gamma_{\Delta x y}(h) & \Gamma_{\Delta x \Delta x}(h) \end{bmatrix} = E z_t z_t',$$

where $\| A \|$ is the supremum norm of a matrix $A$.

The assumption $\text{var}(\varepsilon_t) = I_2$ is innocuous because we do not normalize the elements of $A_j$. We propose to estimate a quasi-covariance by

$$\hat{\lambda}_{y, \Delta x} = \sum_{h=1}^{n-1} k \left( \frac{h}{m} \right) \hat{\Gamma}_{\Delta x y}(h); \quad \hat{\Gamma}_{\Delta x y}(h) = \frac{1}{n} \sum_{t=h+1}^{n} y_t \Delta x_{t-h}, \quad (11)$$

where $m$ is the bandwidth and $k(x)$ is the kernel. We assume $k(x)$ and $m$ satisfy the following assumptions.

Assumption K

The kernel $k(x)$ is continuous and uniformly bounded with $k(0) = 1, \int_{0}^{\infty} |k(x)| x^{1/2} dx < \infty, \int_{0}^{\infty} k^2(x) dx < \infty$ and

$$\lim_{x \to 0} \frac{1 - k(x)}{|x|^q} = k_q < \infty, \quad \text{with } \delta > q.$$
Assumption M

\[
\frac{1}{m} + \frac{m^q}{n} \rightarrow 0 \quad \text{as } n \rightarrow \infty.
\]

Assumption K is satisfied by the Bartlett kernel with \( q = 1 \). Other kernels such as the Parzen kernel, Tukey-Hanning kernel, and Quadratic Spectral kernel satisfy Assumption K with \( q = 2 \).

Let \( f_{yy}(\lambda) \) denote the spectral density of \( y_t \) and \( f_{\Delta x y}(\lambda) \) denote the cross-spectral density between \( \Delta x_t \) and \( y_t \), and similarly for \( f_{y \Delta x}(\lambda) \) and \( f_{\Delta x \Delta x}(\lambda) \). The following lemma shows the asymptotic bias and variance of \( \hat{\lambda}_{y, \Delta x} \) and its consistency.

**Lemma 1** If Assumptions A, K and M hold, then

\begin{enumerate}[(a)]  
\item \( \lim_{n \rightarrow \infty} n^{-1} \operatorname{var} (\hat{\lambda}_{y, \Delta x} - \lambda_{y, \Delta x}) = -k_q \sum_{h=1}^{\infty} \Gamma_{\Delta x y}(h) q h^q \),
\item \( \lim_{n \rightarrow \infty} n^{-1} \operatorname{var} (\hat{\lambda}_{y, \Delta x}) = V \equiv 4\pi \int_{0}^{\infty} k^2(x) \, dx \left\{ f_{yy}(0) f_{\Delta x \Delta x}(0) + [f_{y \Delta x}(0)]^2 \right\} ,
\item \( \hat{\lambda}_{y, \Delta x} \rightarrow_p \lambda_{y, \Delta x} \) as \( n \rightarrow \infty \).
\end{enumerate}

The proof of part (a) is omitted because it is the same as that of Theorem 10 in Hannan (1970, p. 283). Part (b) is a one-sided version of Theorem 9 of Hannan (1970, p. 280).

**Remark 1** If \( k(x) \) is symmetric, we have \( 2V = \lim_{n \rightarrow \infty} \operatorname{var}(\hat{\omega}_{y, \Delta x}) \), where \( \hat{\omega}_{y, \Delta x} \) is the estimate of the long-run covariance between \( y_t \) and \( \Delta x_t \). So, the asymptotic variance of \( \hat{\lambda}_{y, \Delta x} \) is just half the limiting variance for the two-sided case.

**Remark 2** From Lemma 1, the asymptotic mean squared error is minimized by choosing \( m \) such that

\[
m^* = \left( \frac{2q k^2}{\pi} \left( \sum_{h=1}^{\infty} \Gamma_{\Delta x y}(h) q h^q \right)^2 \right) \left( \frac{n}{V} \right)^{1/(2q+1)}.
\]

Assuming \( k(x) \) is symmetric, we can rewrite \( m^* \) as

\[
m^* = \left( q k^2 \alpha(q) n \left/ \int_{-\infty}^{\infty} k^2(x) \, dx \right. \right)^{1/(2q+1)}, \tag{12}
\]

\[
\alpha(q) = 4 \left( \frac{(2\pi)^{-1} \sum_{h=1}^{\infty} \Gamma_{\Delta x y}(h) q h^q}{f_{yy}(0) f_{\Delta x \Delta x}(0) + [f_{y \Delta x}(0)]^2} \right)^2,
\]

giving expressions similar to those in Andrews (1991, pp. 825, 830). If \( m \) is chosen optimally, then the rate of convergence is \( n^{q/(2q+1)} \).
2.1 The limit distribution when $x_t$ is $I(1)$

It is well known that the estimator of the two-sided long-run covariance between $y_t$ and $\Delta x_t$ has normal limiting distribution (Hannan, 1970, Theorem 11, p. 289). However, currently there are no results that show the asymptotic normality of the one-sided long-run covariance estimator. One of the reasons is because the off-diagonal one-sided long-run covariance estimator does not admit a simple expression in terms of periodograms. To see why, let $I_z(\omega)$ be the periodogram of $z_t$, then it follows that

$$
\sum_{h=1}^{n-1} k \left( \frac{h}{m} \right) \frac{1}{n} \sum_{t=h+1}^{n} z_{t-h} z_t' = \sum_{h=1}^{n-1} k \left( \frac{h}{m} \right) \int_{-\pi}^{\pi} I_z(\omega) e^{i\omega h} d\omega
$$

$$
= \int_{-\pi}^{\pi} I_z(\omega) K_n(\omega) d\omega, \quad K_n(\omega) = \sum_{h=1}^{n-1} k \left( \frac{h}{m} \right) e^{i\omega h}.
$$

It is easy to see that $K_n(\omega)$ does not have a simple expression such as Fejér kernel, and indeed it has a nonnegligible imaginary part. In the present paper, we work directly with $\hat{\Gamma}_{y, \Delta x}$ by applying the martingale approximation a la Phillips and Solo (1992) and show the asymptotic normality of $\hat{\lambda}_{y, \Delta x}$. The following theorem establishes it.

**Theorem 2** If Assumptions A, K and M hold and $m^2/n + n/m^{2q+1} \to 0$, then

$$
\sqrt{\frac{n}{m}} \left( \hat{\lambda}_{y, \Delta x} - \lambda_{y, \Delta x} \right) \to_d N(0, V), \text{ as } n \to \infty.
$$

Unlike the regression-based tests, neither a non-zero intercept in $(y_t, x_t)$ nor a linear trend in $x_t$ affects the limiting distribution. Because $\Delta(x_t + \mu) = \Delta x_t$, $\hat{\lambda}_{y, \Delta x}$ is invariant to the presence of a non-zero intercept in $x_t$. For a non-zero intercept in $y_t$ and a linear trend in $x_t$, if we replace $(y_t, \Delta x_{t-h})$ with $(y_t - \bar{y} - \Delta \bar{x}, \Delta x_{t-h} - \Delta \bar{x})$, where $\bar{y}$ and $\Delta \bar{x}$ denote the sample average of $y_t$ and $\Delta x_t$, then $\hat{\lambda}_{y, \Delta x}$ has the stated asymptotic distribution.\[^9\]

The optimal bandwidth $m^*$ does not satisfy the rate condition on $m$ of Theorem 2, which is a standard result when the bandwidth is chosen to minimize the mean squared error. $m$ needs to grow faster than $m^*$ for Theorem 2 to hold. Since the optimal rate of increase of $m$ is $n^{1/(2q+1)}$ from Remark 2 (2), the upper bound on $m$, $m^2/n \to 0$, does not appear to pose a severe problem when $q$ is 1 or 2.

2.2 The limit distribution when $x_t$ is modelled as local to unity

Consider the case where $x_t$ is a local-to-unity process:

\[^9\] If $y_t$ is trend stationary, employing the detrended residual gives the same limiting distribution.
Assumption B

\[ x_t = (1 + c/n)x_{t-1} + u_t, \quad t = 1, 2, \ldots, n, \quad n = 1, 2, \ldots \quad c < 0, \]
\[ x_t = 0 \quad \text{for } t \leq 0, \]
\[ z_t^* = (y_t, u_t)' \] satisfies Assumption A.

Then \( \lambda_{y,\Delta x} = \sum_{h=1}^{\infty} \text{cov}(y_t, u_{t-h}) + O(n^{-1}) \) as seen in (9). The following Lemma establishes the first order equivalence of the limit theory for \( \hat{\lambda}_{y,\Delta x} \) under both \( I(1) \) and local to unity assumptions on \( x_t \).

**Lemma 3** Suppose Assumptions B, K and M hold. Then \( \hat{\lambda}_{y,\Delta x} = \sum_{h=1}^{n-1} k(h/m)\hat{\Gamma}_{xy}(h) + O_p((m/n)) \). If, in addition, \( m^2/n + n/m^{2q+1} \to 0 \), then \( \sqrt{n/m}(\hat{\lambda}_{y,\Delta x} - \lambda_{y,\Delta x}) \to_d N(0, V) \), where \( \hat{\Gamma}_{xy}(h) \) and \( V \) are defined in (11) and Lemma 1, respectively, with \( u_t \) replacing \( \Delta x_t \).

The fact that the limiting distribution is the same for all finite \( c \leq 0 \) has important practical implications, since it means that no prior knowledge on \( c \) is required in order to conduct inference. By contrast, many econometric procedures, including several common cointegration tests, that are valid for \( c = 0 \) may fail for \( c < 0 \).

### 2.3 The limit distribution when \( x_t \) is \( I(0) \)

The argument so far is based on the assumption that \( x_t \) is \( I(1) \). However, in practice often we do not have strong prior knowledge about whether \( x_t \) is \( I(1) \) or \( I(0) \).

With an additional Lipschitz continuity assumption on the kernel, \( \lambda_{y,\Delta x} \) converges to \( E y_t x_{t-1} = \lambda_{y,\Delta x} \) when \( x_t \) is an \( I(0) \) process. Let us first state the assumptions on \( x_t \) and \( y_t \).

Assumption C

\[ v_t = \begin{pmatrix} y_t \\ x_t \end{pmatrix} = B(L) \varepsilon_t = \sum_{j=0}^{\infty} B_j \varepsilon_{t-j}, \quad \sum_{j=0}^{\infty} j||B_j|| < \infty, \quad (13) \]
\[ \varepsilon_t \sim \text{i.i.d. } (0, I_2), \quad \text{with finite fourth moment} \]
\[ \sum_{-\infty}^{\infty} |h|^{\delta} ||\gamma(h)|| < \infty, \quad \delta > 1; \quad \gamma(h) = \begin{bmatrix} \gamma_{y}(h) & \gamma_{y}(h) \\ \gamma_{x}(h) & \gamma_{xx}(h) \end{bmatrix} = E v_t v_{t+h}', \]
and \( f_x(0), f_y(0) > 0 \), where \( f_x(\lambda) \) and \( f_y(\lambda) \) are the spectral density of \( x_t \) and \( y_t \).

We use \( \gamma(h) \) to denote the autocovariance of \( v_t \) to distinguish it from the autocovariance of \( z_t \) in Assumption A. Note that \( \lambda_{y,\Delta x} = E y_t x_{t-1} = \gamma_{xy}(1) \).

**Lemma 4** If Assumptions C, K and M hold and \( k(x) \) is Lipschitz(1), then
\[ \sqrt{n} \left( \hat{\lambda}_{y,\Delta x} - \lambda_{y,\Delta x} \right) = k(1/m)\sqrt{n} \left( \gamma_{xy}(1) - \gamma_{xy}(1) \right) + B_n + o_p(1), \quad (14) \]
where \( \hat{\gamma}_{xy} (1) = n^{-1} \sum_{t=2}^{n} y_t x_{t-1} \) and \( B_n \) is the bias term satisfying
\[
B_n = \begin{cases} 
0, & \text{if } E y_t x_{t-h} = 0 \text{ for all } h \geq 1, \\
O(n^{1/2} m^{-q}), & \text{otherwise.}
\end{cases}
\]

In addition, \( k(1/m) \sqrt{n}(\hat{\gamma}_{xy} (1) - \gamma_{xy} (1)) \rightarrow_d N (0, \Xi) \) as \( n \rightarrow \infty \), where
\[
\Xi = \sum_{u=-\infty}^{\infty} \{ \gamma_{xx} (u) \gamma_{yy} (u) + \gamma_{xy} (u + 1) \gamma_{yx} (u - 1) \} + \sum_{u=-\infty}^{\infty} k_{xyxy} (0, 1, u, u + 1),
\]
and \( k_{xyxy}(0, a, b, c) \) is the fourth cumulant of \( (x_t, y_{t+a}, x_{t+b}, y_{t+c})' \).

If you knew \( x_t \sim I(0) \), then you would estimate \( E y_t x_{t-1} \) by \( \hat{\gamma}_{xy} (1) \), and the limiting variance of \( \hat{\lambda}_{y, \Delta x} \) is the same as that of \( \hat{\gamma}_{xy} (1) \). Therefore, \( \hat{\lambda}_{y, \Delta x} \) is robust to misspecification of the integration of order, apart from the bias term in (14).

## 3 Possible ways to conduct inference

### 3.1 Estimation of the limiting variance of the estimator

Suppose \( x_t \) is \( I(1) \) and Lemma 2 gives the limiting distribution of \( \hat{\lambda}_{y, \Delta x} \). To conduct inference, we need to estimate \( V \). Of course, we can use \( \hat{V} = 4 \pi^2 \int_0^\infty k^2 (x) \, dx \{ \hat{f}_{yy} (0) \hat{f}_{\Delta x \Delta x} (0) + \hat{f}_{y \Delta x} (0) \hat{f}_{\Delta xy} (0) \} \), where \( \hat{f}_{ab} \) is a consistent periodogram-based estimator of \( f_{ab} \).

We may also consider another estimator of \( V, V' \), whose particularly good performance is suggested by simulations in Section 4. It is based on the exact finite sample variance of \( \hat{\lambda}_{y, \Delta x} \), which is given by (see equations (32)-(34) in the proof of Lemma 1)
\[
\frac{n}{m} \text{var} \left( \hat{\lambda}_{y, \Delta x} \right) = \frac{1}{m} \sum_{h' = 1}^{n-1} \sum_{h = 1}^{n-1} k \left( \frac{h'}{m} \right) k \left( \frac{h}{m} \right) \sum_{u = -\infty}^{\infty} \{ \Gamma_{\Delta x \Delta x} (u) \Gamma_{yy} (u + h - h') + \Gamma_{\Delta x y} (u + h) \Gamma_{y \Delta x} (u + h') + \Gamma_{xyxy} (0, h', u, u + 1) \} \phi_n (u, h', h),
\]
where \( \phi_n (u, h', h) \) is defined in the proof of Lemma 1. The terms involving the cumulants disappear in the limit. Define \( \tilde{V} \) by replacing \( \Gamma_{ab} \) with \( \tilde{\Gamma}_{ab} \), which reduces the error from the approximation of the discrete sum in (30) by an integral:
\[
\tilde{V} = \frac{1}{m} \sum_{h' = 1}^{n-1} \sum_{h = 1}^{n-1} k \left( \frac{h'}{m} \right) k \left( \frac{h}{m} \right) \sum_{u = -\infty}^{\infty} \left\{ \tilde{k} \left( \frac{u}{m} \right) \tilde{\Gamma}_{\Delta x \Delta x} (u) \tilde{k} \left( \frac{u + h - h'}{m} \right) \tilde{\Gamma}_{yy} (u + h - h') + \tilde{k} \left( \frac{u + h}{m} \right) \tilde{\Gamma}_{\Delta x y} (u + h) \tilde{k} \left( \frac{u - h'}{m} \right) \tilde{\Gamma}_{y \Delta x} (u - h') \right\} \phi_n (u, h', h),
\]
where \( \tilde{k}(x) \) and \( \tilde{m} \) are a kernel and bandwidth. \( \tilde{k}(x) \) and \( \tilde{m} \) can, but do not need to, be the same as \( k(x) \) and \( m \). Estimating \( V \) by \( \tilde{V} \) gives better finite sample performance than estimating \( V \) by \( \hat{V} \). (The results using \( \tilde{V} \) are not reported in the present paper).
Suppose \((y_t, \Delta x_t)\) satisfies Assumption A and hence \(x_t\) is \(I(1)\). Then, we may construct a t-type statistic
\[
t_\lambda = \frac{\sqrt{\frac{m}{n}} (\hat{\lambda}_{y,\Delta x} - \lambda_{y,\Delta x})}{\sqrt{\tilde{V}}}. \tag{16}
\]
The following Lemma shows that \(t_\lambda\) converges to a \(N(0,1)\) random variable.

**Lemma 5** If Assumptions A or B and K and M hold, the kernel \(\tilde{k}(x)\) satisfies Assumption K with \(\tilde{q} \) replacing \(q\), \(\tilde{k}(x) = 0\) if \(|x| > 1\), and \(1/\tilde{m} + \tilde{m}^2/n \to 0\), then \(\tilde{V} \to_p V\) as \(n \to \infty\).

**Corollary 6** If the assumptions of Theorem 2 or Lemma 3 and Lemma 5 hold, then \(t_\lambda \to_d N(0,1)\) as \(n \to \infty\).

### 3.2 Conservative inference: inference when \(x_t\) is modelled as \(I(1)\) but is actually \(I(0)\)

Consider the case when \((y_t, x_t)\) follows (13) and \(x_t\) is actually \(I(0)\). It is easy to show that the test based on \(t_\lambda\) is consistent. Suppose we test \(H_0 : \lambda_{y,\Delta x} = 0\) but \(\lambda_{y,\Delta x} \neq 0\). Then we have, from Lemma 4,
\[
t_\lambda = \frac{n^{1/2} \hat{\lambda}_{y,\Delta x}}{(\tilde{V})^{1/2} m^{1/2}} = \frac{n^{1/2} (\hat{\lambda}_{y,\Delta x} - \lambda_{y,\Delta x}) + n^{1/2} \lambda_{y,\Delta x}}{(\tilde{V})^{1/2} m^{1/2}} = \frac{n^{1/2} \lambda_{y,\Delta x} (1 + o_p(1))}{(\tilde{V})^{1/2} m^{1/2}}.
\]
Since \(\tilde{V} \to_p f_{\Delta x \Delta x}(0) = 0\) and \(n^{1/2} m^{-1/2} \to \infty\), it follows that \(|t_\lambda| \to \infty\) as \(n \to \infty\). Indeed, \(t_\lambda\) diverges at a faster rate than \(n^{1/2} m^{-1/2}\), the rate of divergence in the \(I(1)\) case.

Since \(t_\lambda\) is based on the autocovariance of \(y_t\) and \(\Delta x_t\), the inference based on \(t_\lambda\) might be misleading. However, if the Bartlett kernel \(\tilde{k}(x) = (1 - |x|)1\{|x| \leq 1\}\) is used in \(\tilde{V}\) in (15) and \(Ey_\Delta x_{t-h} = 0\) for all \(h \geq 1\) (which holds under the null hypothesis of orthogonality), then \(t_\lambda\) is \(O_p((\tilde{m}/m)^{1/2})\). Therefore, when \(\tilde{m}\) is chosen appropriately, \(\hat{\lambda}_{y,\Delta x}\) serves as a tool for conservative inference.

**Lemma 7** If Assumptions C, K and M hold, \(\tilde{k}(x)\) is the Bartlett kernel, \(1/\tilde{m} + \tilde{m}/n \to 0\), and \(Ey_\Delta x_{t-h} = 0\) for all \(h \geq 1\), then \(t_\lambda = O_p((\tilde{m}/m)^{1/2})\) as \(n \to \infty\).

In order to understand the convergence, rewrite \(t_\lambda\) as
\[
t_\lambda = \frac{n^{1/2} (\hat{\lambda}_{y,\Delta x} - \lambda_{y,\Delta x})}{(\tilde{V})^{1/2} m^{1/2}}.
\]
The numerator converges to a Gaussian random variable from Lemma 4. \(\tilde{V}\) in the denominator is an estimate of \(f_{\Delta x \Delta x}(0) = 0\) and hence converges to 0 as \(\tilde{m} \to \infty\). Because \(m\) tends to infinity, the asymptotic behavior of \(t_\lambda\) depends on the rate of
convergence of \( \tilde{V} \). Letting \( \tilde{m} \) tend to infinity but not too fast prevents \( \tilde{V} \) from converging to 0 too fast and makes \( t_\lambda \) converge to 0 in probability.

Therefore, by choosing \( \tilde{m} \) appropriately, the \( t_\lambda \) statistic provides a conservative inferential tool that converges to \( N(0,1) \) if \( x_t \) is \( I(1) \) or local to unity but converges to zero when \( x_t \) is \( I(0) \). Thus, the rejection rate will not exceed the nominal level. This is summarized in the following Lemma.

**Lemma 8** If either (i) Assumptions A or B and K and M hold, or (ii) Assumptions C, K and M hold with 
\[
E[y_t x_{t-h}] = 0 \quad \text{for all } h \geq 1, \quad \text{and, in addition, } k(x) \text{ is Lipshitz}(1),
\]
\[
k(x) \text{ is the Bartlett kernel, and } m^2/n + n/m^{2q+1} + 1/m + m/n + m/m \to 0,
\]
then 
\[
\Pr(|t_\lambda| \geq z_{1-\alpha/2}) \to \alpha' \leq \alpha \text{ as } n \to \infty,
\]
where \( z_{1-\alpha/2} \) is the \( 100(1-\alpha/2) \) percentile of the \( N(0,1) \) distribution.

## 4 Finite sample performance: simulation results

This section provides a modest simulation study to gauge the small sample accuracy of the proposed test. The results indicate reasonable (and often quite good) size and power in sample sizes as small as 100.

For the simulations below we have in mind a test of \( y_t \) orthogonal to \( I_{x,t-1} \), the information contained in past \( x_t \), as in (1). This is often tested in practice using a regression of \( y_t \) on \( x_{t-1} \) as in (2). Since size distortions rule out standard regression only for \( x_t \) highly serially correlated, it is this case that we focus on. In particular, we consider both first and second order autoregressive models for \( x_t \):

\[
x_t = \rho_0 + \rho_1 x_{t-1} + u_{2t}, \quad \text{AR}(1) \quad (17)
\]
\[
x_t = \rho_0 + \rho_1 x_{t-1} + \rho_2 x_{t-2} + u_{2t}. \quad \text{AR}(2) \quad (18)
\]

The AR(1) model may also be written as a unit root/local to unity process by letting
\[
\rho_1 = 1 + c/n, \quad c \leq 0. \quad (19)
\]

Often the primary economic interest centers on the relation between \( y_t \) and \( x_{t-1} \). Under the null hypothesis \( y_t \) is orthogonal to \( I_{x,t-1} \) and often an efficient market condition will also imply that \( y_t \) is orthogonal to its own past. In the simulations, the process for \( y_t \) under the null hypothesis is therefore specified by

\[
y_t = d_t + u_{1t}, \quad (20)
\]

where the innovation \( u_{1t} \) is discussed below and the deterministic component \( d_t \) consists of either an intercept or a trend:

\[
d_t = \delta_0 \quad \text{or} \quad (21)
\]
\[
d_t = \delta_0 + \delta_1 t. \quad (22)
\]

We employ two different specifications for \( y_t \) under the alternative hypothesis when investigating finite sample power. First we consider the standard regression specification

\[
y_t = d_t + \beta x_{t-1} + u_{1t}. \quad (23)
\]
In the unit root/local to unity context, this may be referred to as a balanced alternative, since for \( \beta \neq 0 \), both \( y_t \) and \( x_t \) contain an equally persistent component. In fact, when \( x_t \) has a unit root the two are cointegrated. While this has traditionally been the alternative on which the literature has focused, in certain applications there may be an unappealing aspect to it. For example, as discussed in the introduction, it is not clear that one would want to model near unit root components in stock or exchange rate returns on theoretical grounds and empirically they show little serial correlation.\(^{10}\) Thus, it also seems reasonable to consider test performance under unbalanced alternatives, in which \( x_t \) is persistent but \( y_t \) is not. A simple alternative of this type is given by a regression of \( y_t \) on prefiltered \( x_t \) as in

\[
y_t = d_t + \gamma \left( 1 - \left( 1 + \frac{c}{n} \right) L \right) x_{t-1} + u_{1t},
\]

where \( x_t \) is given by the AR(1) specification in (17). Of course, (24) is not itself unbalanced, but it implies an imbalance between \( y_t \) and \( x_{t-1} \). This may be rewritten as

\[
y_t = d_t + \gamma u_{2,t-1} + u_{1t},
\]

in which form it also makes sense for more general models of \( x_t \).

Finally, since the orthogonality between \( y_t \) and past \( x_t \) (i.e. \( x_{t-j}, j \geq 1 \)) does not rule out contemporaneous covariance between \( y_t \) and \( x_t \), we allow the two innovation processes to be correlated under both the null and alternative. They are specified by

\[
\begin{align*}
u &= \begin{pmatrix} u_{1t} & u_{2t} \end{pmatrix}' = \Sigma^{1/2} \varepsilon_t, \quad \varepsilon_t \sim \text{i.i.d. } N(0, I_2) \\
\Sigma &= \Sigma^{1/2}(\Sigma^{1/2})' = \begin{pmatrix} 1 & \sigma_{12} \\ \sigma_{21} & 1 \end{pmatrix}.
\end{align*}
\]

Our primary interest lies in the performance of the covariance-based t-statistic \( t_\lambda \) given in (16), which was estimated as follows. In the trend model (22), we first demeaned \( \Delta x_t \) (thereby removing the trend in \( x_t \)) and detrended \( y_t \) prior to estimation. In the intercept model (21) only \( y_t \) was demeaned. Using this detrended (or demeaned) data we then estimated the quasi-covariance \( \lambda_{y,\Delta x} \) defined in (7) using the standard kernel covariance estimator \( \hat{\lambda}_{y,\Delta x} \) given in (11). Likewise, we estimated its asymptotic variance \( V \) (see Lemma 1) using the kernel estimator \( \widetilde{V} \) following (15).

Both kernel estimation procedures require the choice of kernel and bandwidth. The theoretical results allow considerable flexibility in the choice of the kernel \( k(x) \) in the estimation of \( \lambda_{y,\Delta x} \). However, to ensure conservative inference for stationary \( x_t \), Lemma 7 mandates use of the Bartlett kernel for \( \tilde{k}(x) \) in the estimation of \( \widetilde{V} \). We therefore used the Bartlett kernel for both estimators. The bandwidth parameter \( m \) in the estimation of \( \lambda_{y,\Delta x} \) is selected using the optimal bandwidth formula given in (12). Implementation of this formula in practice requires the use of a first-stage parametric approximation model. As in Andrews (91) this is assumed only to

\(^{10}\)Conceivably, for fixed \( \rho_1 < 1 \) the persistent component may be "hidden" by noise from a second component, a conjecture which is not easily confirmed or refuted.
provide a parsimonious approximation, not a correct specification. Although separate univariate AR(1) models are typically employed, the optimal bandwidth in this case depends on the behavior of the cross auto-correlations and necessitates a joint model. Including a moving average component also seems desirable given possible over-differencing in $\Delta x_t$. A VARMA(1,1) was therefore used as the first stage model for $(y_t, \Delta x_t)$. Employing the asymptotically efficient three stage linear regression method of Dufour and Pelletier (2002) allowed us to avoid non-linear optimization, keeping estimation simple. The choice of the second bandwidth parameter $\tilde{m}$ used in estimation of $V$ is constrained by Lemma 7 which requires $\tilde{m} = o(m)$. While clearly arbitrary, our choice of $\tilde{m} = m^{0.9}$ appeared sufficient to insure conservative inference in the stationary case, with minimal cost in overall performance.

We also provide some comparisons to both the standard regression t-test and the size-adjusted regression-based approach, using the two stage Bonferroni-bounds test of Cavanagh et al. (1995) (hereafter CES). All results below are based on 2000 replications, with results reported for sample sizes of $n = 100$ and 400.

4.1 Size

We first simulate under the null hypothesis with $y_t$ given by (20) and $x_t$ given by the AR(1) process (17) with $\rho_1$ modelled local to unity as in (19). Results are provided for various values of both $c$ (and therefore $\rho_1$) and $\sigma_{12}$. In order to set a basis of comparison, Table 1 shows empirical rejection rates for the standard two-sided regression t-test ($y_t$ regressed on $x_{t-1}$) with a nominal level of 5 percent. The rejection rates are reasonable for small values of $\rho_1$ and/or $\sigma_{12}$ but grow highly unreliable as $\rho_1$ approaches one and the residual correlation increases. The size problem is particularly severe in the model with trend, for which rejection rates can exceed 50 percent.

By contrast, the rejection rates for the covariance-based t-test $t_\lambda$ shown in Table 2 are fairly accurate over the whole range of parameter values in the both the intercept and trend models. Furthermore, the test generally works well in sample sizes as small as one hundred and becomes quite reliable for $n = 400$. Consistent with the theory, the test can become slightly conservative for large (negative) values of $c$. However, with only a few exceptions, the empirical rejection rates remain within two percentage points of the nominal value. This good performance was anticipated from the previous section, given the asymptotic standard normal distribution of the covariance-based test statistic. Of course, good performance may also be obtained by properly size adjusting the regression-based tests, as in the bounds tests of CES (see their Table 4).

Following the literature, we also consider longer-horizon returns. We require no explicit corrections to handle the moving average components induced by the

---

11 We impose some constraints on the ARMA parameters to insure stationarity and invertibility and also impose $n^{0.9}$ as an upper bound on $m$.

12 Note that Hodrick or HAC standard errors are unnecessary here since $u_t$ is taken to be i.i.d.
overlapping returns; for fixed horizon (k) we simply have to define
\[ y_{t,k} = (y_t + y_{t-1} + \ldots + y_{t-k+1}) = \phi(L)y_t \quad \text{and} \quad \Delta x_{t,k} = \Delta x_{t-k} = L^k \Delta x_t \quad (26) \]
and apply the theoretical results to \((y_{t,k}, \Delta x_{t,k}) = \text{diag}(\phi(L), L^k)A(L)\varepsilon_t = A^*(L)\varepsilon_t.\)
In finite sample, the accuracy of long-horizon tests typically depends on the ratio \(k/n.\)
In Table 3, we match this ratio to the sample size \((n = 924)\) and longest horizon \((2 \text{ years, } k = 24)\) used in the empirical application, yielding \(k = 3\) for \(n = 100.\) This again yields reasonable size performance. Holding \(n = 100\) fixed, but increasing the horizon to \(k = 5\) leads to only a slight deterioration.

The model above is the baseline model most often used to evaluate size distortions in this context. However, our test is designed to work in a more general setting and it is also of interest to investigate finite sample performance under higher order autoregressive specifications for \(x_t,\) such as the AR(2) model (18), with roots on or close to the unit circle. Rudebusch (1992, Table 2) finds that an AR(2) with \(\rho_1 + \rho_2\) slightly below unity (with \(\rho_1 > 1\) and \(\rho_2 < 0\)) provides a good fit for a number of macroeconomic and financial time series. In order to roughly match these estimates we set
\[ \rho_1 = 1.5 \quad \text{and} \quad \rho_2 = -0.5 + c/n \quad (27) \]
for the same values of \(c\) considered above. Thus, like in the AR(1) model, \(x_t\) is unit-root nonstationary for \(c = 0\) \((\rho_2 = -0.5)\) and stationary but strongly correlated for \(c < 0\) \((\rho_2 < -0.5).\) The rejection rates for the covariance-based tests are shown in Table 4. In the demeaned case, the results remain fairly accurate even for \(n = 100.\) In the detrended case there is a tendency to over-reject in certain cases for \(n = 100\) but this improves considerably for \(n = 400.\) By contrast, finite sample rejection rates for least squares (available upon request) reach to above 50% and do not improve with sample size for fixed \(c.\)

In summary the size of the proposed covariance-based test seems generally to be reasonable, and is often quite accurate, even in sample sizes as small as \(n = 100.\) We next consider finite sample power.

4.2 Power

We first consider the power of the covariance-based test \(t_{\lambda}\) against the balanced regression alternative given in (23) with \(\beta \neq 0\) and local to unity \(x_t\) given by (17) and (19).\(^{13}\) For \(c = 0\) this alternative constitutes a cointegrating relation, while for \(c << 0\) the alternative is a stationary regression. The results are shown in Table 5. As expected, the power of the test is reasonable, increasing in both sample size and distance from the null.\(^{14}\)

One of the goals of the covariance-based test was to simultaneously maintain power against “unbalanced” alternatives which allow \(y_t\) (e.g. returns) to be stationary, despite near or even exact unit root behavior in \(x_t.\) This avoids, for example,
the requirement that stock prices or exchange rates contain an $I(2)$ or near-$I(2)$ component under the alternative hypothesis when predictor variables are persistent. More generally, it avoids the transformation of the orthogonality test into a unit root/cointegration test as the root in $x_t$ approaches one. The pre-filtered regression (25), together with (17), therefore provides a natural alternative in which to consider finite sample power in that it holds $y_t$ stationary (but not over-differenced) regardless of the persistence in $x_t$. In doing so, it incorporates both (2) and (3) as special cases for $\rho_1 = 0$ and $\rho_1 = 1$ respectively.

Finite sample power results for the covariance-based tests under the unbalanced alternative (24) with $x_t$ given by (17) and (19) are shown in Table 6. The test is calculated in the same way as before, again using $y_t$ and $x_{t-1}$ as inputs (i.e. we don’t make use of the knowledge that $y_t$ and $x_t$ are unbalanced). These rejection rates appear quite reasonable, again increasing in both sample size and distance from the null hypothesis.

Many existing tests are based on a size adjusted regression of the type shown in (2). These procedures may be expected to have good power against regression alternatives when $x_t$ is stationary (e.g. $\rho_1 << 1$ in (17) and $\beta \neq 0$ in (23)) and against cointegration or near-cointegration alternatives when $x_t$ is near $I(1)$, (e.g. $\rho_1 \approx 1$ in (17) and $\beta \neq 0$ in (23)). This is confirmed in Table 7, which shows finite sample power for the CES Bonferroni test procedure against $\beta \neq 0$ in (23) with local to unity $x_t$ given by (17) and (19). As expected, the test exhibits very good power against this alternative and is in this case more powerful than $t_\lambda$.

On the other hand, it is not clear that tests based on (2) should have much power against unbalanced alternatives, since the parameter restriction tested (i.e. $\beta_1 = 0$) is satisfied for all unbalanced relations and $\tilde{\beta} \rightarrow_p 0$. Table 8 provides rejection rates for the CES Bonferroni test against the same unbalanced alternative (and same DGP) used to assess the power of $t_\lambda$ in Table 6. Confirming the reasoning above, the regression based test does quite well for the larger values of $c$ when $x_t$ and $y_t$ behave in a stationary manner, but performance deteriorates rapidly as $x_t$ approaches nonstationarity (small $c$) and the alternative becomes unbalanced. Moreover, for small $c$ the power does not seem to improve as we move further into the alternative. Nor, for fixed values of $c$, do rejection rates increase much as the sample size increases. For example, in the worst case for $c = 0$ and $\sigma_{12} = 0.95$, the power remains under 10 percent even for a population $R^2$ of 0.5 and a sample size of four-hundred.

These simulations suggest that the covariance-based orthogonality test may provide power against a wider range of alternatives than do existing size-adjusted regression-based tests. In particular, they appear to provide reasonable power against both balanced and unbalanced alternatives whereas regression-based tests do particularly well against the balanced alternatives for which they were designed, but provide little reliable power against unbalanced alternatives. This added generality does of course come at some cost in terms of power against certain specific alternatives and, in this sense, the two testing approaches (regression and covariance-based) are properly seen as compliments rather than substitutes.
5 Application to tests of stock return predictability

We use the method developed above to test the orthogonality of stock returns to the information in past short-term interest rates and dividend yields. Under the market efficiency/constant risk premium hypothesis it should not be possible to systematically forecast stock returns. Early tests of this hypothesis found fairly substantial predictability and thus had a large impact on the finance literature (see Campbell and Shiller (1988a,b), Fama and French (1988), Hodrick (1992), Shiller (1984)).

Although theoretical considerations may rule out exact unit root behavior in dividend yields\(^1\) and interest rates, near unit roots in the local to unity sense can not be ruled out a priori. Empirically, both series are highly persistent, with confidence intervals on the largest root often containing one (Torous et al. (2005)). Moreover, although pre-determined, there is no reason to believe that these regressors are fully exogenous. For example, the stock price enters both the return and dividend yield. The combination of near unit root behavior and a failure of strict exogeneity is a recipe for size problems (Cavanagh et al. (1995)). Consequently, subsequent doubts have been raised regarding the evidence for predictability on account of the strong persistence in the regressors (Stambaugh (1986 & 1999) and Mankiw and Shapiro (1986)).\(^2\) This has spurred a large literature in an attempt to address this issue, and the degree to which evidence of predictability has been overstated remains a subject of ongoing debate.\(^3\)

Thus the literature to date has focused primarily on the issue of size distortion. However, as discussed at length in the introduction, near unit root regressors may also raise specification issues under the alternative, in the sense that a stationary variable, such as a stock return should not be linearly predictable by a unit (or near unit) root regression (Lanne (2002)). However, even if \(y_t \sim I(0)\) and \(x_t \sim I(1)\), \(y_t\) may still be predictable based on the past history of \(x_t\), as exemplified by (3) and (5), with \(\gamma_1 \neq 0\). Yet, as the simulations underscored, regression tests based on (2) have unreliable power against such alternatives, even if size adjusted. Therefore, while evidence of predictability may be overstated due to size distortion, it is also possible that it has been understated due to near unit root specification issues. Since the covariance-based tests address both issues simultaneously they may be useful in untangling these two effects.

Following Campbell et al. (1997, chapter 7), we use monthly returns from 1927 to 1994 and also consider separately the two subperiods: 1927-1951 and 1952-1994.\(^4\) Monthly log returns are calculated as \(r_{t+1} = \ln \left( \frac{P_{t+1} + D_{t+1}}{P_t} \right)\), where \(P_t\) and \(D_t\) are the stock price and dividend from the CRSP value-weighed index of NYSE.

15Campbell and Shiller (1988a,b), but see Tuybens (2002) for an alternative viewpoint.
16Also of concern have been the accuracy of the standard errors in long-horizon regressions (Richardson and Stock (1989), Valkanov (2003)).
17This literature includes resampling and simulation methods (Hodrick (1992), Nelson and Kim (1993), Goetzmann and Jorion (1993), Wolf (2000), and Ang and Bekaert (2001)), local to unity corrections along the lines of Cavanagh et al. (1995) (Viceira (1997), Valkanov (2003), Torous et al. (2005), and Campbell and Yogo (2003)), and finite sample or Bayesian approaches (Stambaugh (1999) and Lewellen (2003)).
18We thank John Campbell for kindly providing us with this data.
AMEX, and NASDAQ stocks. Real returns are formed by deflating nominal returns by the CPI.\footnote{Similar results were also found replacing real by excess returns.}

The dividend-price ratio is calculated in the standard way as the sum of dividends paid over the past twelve months, divided by the current level of the index: $d_t - p_t = \ln \left( \frac{D_t + \ldots + D_{t-11}}{P_t} \right)$. We denote the one month treasury bill rate by $i_t$. Following the literature, we also consider longer-horizon returns of the form $r_{t+1} + \ldots + r_{t+k}$ for $k = 1, 3, 12, \text{ and } 24$. HAC standard errors are employed for $k > 1$ in the regression analysis, but for fixed $k$ our covariance-based test requires no correction or adjustment, as discussed in the simulation section above.

Table 9 shows the standard regression results. The interest rate regressions show only modest evidence of predictability whereas evidence using dividend yields is quite strong. Intuition for the potential bias and size distortion in these regressions is provided by Lewellen (2003) who expresses the bias in $\hat{\beta}$ in (2) in terms of the bias in $\hat{\rho}$ in (17) and the residual covariance $\sigma_{12}$ in (26):

$$E \left[ \hat{\beta} - \beta \right] = \frac{\sigma_{12}}{\sigma_{22}} E \left[ \hat{\rho} - \rho \right]. \quad (28)$$

The two ingredients needed to produce bias are thus persistent regressors and residual serial correlation. Table 10 shows the Stock (1991) confidence interval on the largest root in $x_t$ together with the estimated residual correlation $\delta = \text{corr}(\varepsilon_{1t}, \varepsilon_{2t})$. The two series both show large roots, with confidence intervals on the largest root containing one, but display quite different residual correlation properties. Estimates of $\delta$ are small for the interest rate series, suggesting only modest size distortion, but are close to negative one for the dividend price ratio. Intuitively, an increase in the current stock price corresponds to a higher return but lower dividend yield. Since the AR(1) coefficient estimate $\hat{\rho}_1$ is downward biased (Hurwitz (1950)), negative residual correlation implies positive bias in $\hat{\beta}$ (see 28)). In other words, the bias runs in the same direction as the observed alternative, leaving the results difficult to interpret.

This preliminary analysis suggests that size distortion may matter more in the case of the dividend price ratio, whereas the near-unit root specification issues may be relatively more important for the interest rate regressions. This conjecture is supported by the results in Table 11. The table shows both the covariance-based test statistic $t_\lambda$, for which standard normal critical values apply, and the optimal bandwidth $m^*$, both of which are calculated in the same way as in the simulations (see Section 4 for details).\footnote{The detrended version of the estimator is employed because of its more general applicability. Similar results (available upon request) were obtained using the demeaned version.}

Our results tell two different tales: one for the dividend price ratio and a second for the interest rate. In the case of the dividend price ratio, the covariance-based tests show far weaker evidence of predictability than do the standard regression-based tests. This agrees with the conclusion in several (but not all) previous studies which size correct the regression in (2)\footnote{Viceira (1997), Wolf (2000), Torous et al. (2005), Valkanov (2003), but see also Lewellen (2003) and Campbell and Yogo (2003) who conclude more strongly in favor of predictability.} and based on the results shown it seems
difficult to make a strong case for predictability using the dividend yield. However, as always, one must exercise some caution in interpreting a failure to reject. In particular, while our test, being semi-parametric in nature, was found to have power against a wider range of alternatives (see Section 4) than the regression-based test, the latter naturally had better power when the regression in (2) was properly specified. Nevertheless it seems safe to conclude that the evidence and degree of predictability found in regressions using dividend yields is, at the least, somewhat overstated.

For the interest rate, the story is reversed, with the covariance-based tests in Table 11 providing stronger evidence of predictability than the regression tests of Table 9, particularly at the 3 and 12 month horizons during the 1952-1994 period, and the 3 month horizon in the 1927-1994 sample. This difference may be due to the fact that the regression test restricts the alternative to a direct linear relation between returns (which show little persistence) and highly persistent interest rates. We confirm this conjecture by showing in Table 12 that a stronger relation in fact exists between stock returns and an ad hoc stochastically detrended version of the interest rate, 

\[ x_t = i_t - \sum_{j=0}^{11} i_{t-j}, \]

sometimes employed in this literature (see Campbell (1991)). In conclusion, we find that standard regression tests based on (2) may overstate predictability using the dividend yield due to size distortion, but understate the predictive content in interest rates by restricting the nature of the alternative.

6 Conclusion

In regression-based orthogonality tests it is often the case that the regressor is highly serially correlated, with an autoregressive root close or possibly equal to unity. This is well known to cause size problems in standard tests, due to the nonstandard nature of the test statistic under both unit root and local to unity assumptions. Simple two-stage procedures employing unit root tests together with size correction can generally correct this problem in the I(1) case, but still produce size distortions under local to unity assumptions.

Roots near unity may also artificially restrict the allowable alternative hypothesis, leading to poor size-adjusted power under reasonable alternatives. For example, when the regressor has a unit root but the dependent variable does not, no linear relation between the two can exist, so that the true regression coefficient is forcibly equal to zero. A properly adjusted t-test based on this regression coefficient should therefore generally support the null of orthogonality. However, such a regression imbalance would not rule out a violation of orthogonality due to a linear relationship between the dependent variable and stationary transformations of the regressor.

The covariance-based t-test proposed here produces good size and power against reasonable alternatives regardless of whether the regressor is stationary, nonstationary, or local to unity. This comes without resort to unit root pre-tests or other forms of prior information. Furthermore, because nonstandard distributions are avoided,

\[ \text{During the 1930s and 1940s the short rates were pegged by the Federal Reserve (See Campbell et al. (1997, p. 268)) and thus neither test shows evidence of predictability during the earlier sample.} \]

\[ \text{This amounts to a gradual differencing over a 12 month period.} \]
size adjustments are unnecessary. Simulation results suggest reasonably good size and power in samples as small as one hundred, making this a practical tool for use in empirical applications.

7 Appendix A: Proofs

In the following sections, \( C \) denotes a generic constant such that \( C \in (0, \infty) \) unless specified otherwise, and it may take different values in different places.

7.1 Proof of (9)

From the definition of \( x_t \), we have

\[
\Delta x_t = u_t + \frac{c}{n} x_{t-1} = \begin{cases} 
  u_t + \frac{c}{n} \sum_{k=0}^{t-2} (1 + c/n)^k u_{t-1-k}, & t \geq 1, \\
  0, & t \leq 0,
\end{cases}
\]

with \( \sum_{k=0}^{t-1} 0 = 0 \). It follows that \( \text{cov}(y_t, \Delta x_{t-h}) = \text{cov}(y_t, u_{t-h}) + \frac{c}{n} \sum_{k=0}^{t-h-2} (1 + c/n)^k \text{cov}(y_t, u_{t-h-1-k}) \) for \( t \geq h + 1 \), and 0 for \( t \leq h \). Therefore,

\[
\lambda_{y, \Delta x} = \lim_{t \to \infty} \sum_{h=1}^{t-1} \text{cov}(y_t, \Delta x_{t-h})
= \lim_{t \to \infty} \sum_{h=1}^{t-1} \text{cov}(y_t, u_{t-h}) + \frac{c}{n} \lim_{t \to \infty} \sum_{h=1}^{t-1} \sum_{k=0}^{t-h-2} \left(1 + \frac{c}{n}\right)^k \text{cov}(y_t, u_{t-h-1-k}).
\]

The first term converges to \( \sum_{h=1}^{\infty} \text{cov}(y_t, u_{t-h}) \). The second term is bounded by \( \frac{c}{n} \sum_{p=0}^{\infty} (p+1) \left| \text{cov}(y_t, u_{t-p}) \right| = O(n^{-1}) \), and the stated result follows. \( \blacksquare \)

7.2 Proof of Lemma 1


\[
\frac{n}{m} \text{var} \left( \lambda_{y, \Delta x} \right) = \frac{n}{m} \sum_{h'=1}^{h-1} \sum_{h=1}^{h-1} k \left( \frac{h'}{m} \right) k \left( \frac{h}{m} \right) \text{cov} \left( \hat{\Gamma}_{\Delta xy}(h'), \hat{\Gamma}_{\Delta xy}(h) \right).
\]

Hannan (1970) p. 313 gives

\[
\text{cov} \left( \hat{\Gamma}_{\Delta xy}(h'), \hat{\Gamma}_{\Delta xy}(h) \right) = n^{-1} \sum_{u=-\infty}^{\infty} \{ \Gamma_{\Delta x\Delta x}(u) \Gamma_{yy}(u+h-h') + \Gamma_{\Delta x\Delta x}(u+h) \Gamma_{yy}\Delta x(u-h') + k_{\Delta x\Delta x}(0, h', u, u+h) \phi_{\Delta x}(u, h', h) \}.
\]
where $k_{\Delta y \Delta y}(0, h', u, u + h)$ is the fourth cumulant of $z_t$ (see Hannan, 1970, p.23 for the definition) and $\phi_n(u, h', h)$ is given by (the formula of $\phi_n(u, h', h)$ for $-n + h' \leq u \leq 0$ in Hannan has a typo)

$$
\phi_n(u, h', h) = \begin{cases} 
0, & u \leq -n + h'; \\
1 - h'/n, & 0 \leq u \leq h' - h; \\
1 - \frac{h' - u}{n}, & -n + h' \leq u \leq 0; \\
0, & u \geq n - h. 
\end{cases}
$$

It follows that (30) is comprised of

$$
\frac{1}{m} \sum_{h'=1}^{n-1} \sum_{h=1}^{n-1} k\left(\frac{h'}{m}\right) k\left(\frac{h}{m}\right) \sum_{u=-\infty}^{\infty} \Gamma_{\Delta x \Delta x}(u) \Gamma_{yy}(u + h - h') \phi_n(u, h', h) 
+ \frac{1}{m} \sum_{h'=1}^{n-1} \sum_{h=1}^{n-1} k\left(\frac{h'}{m}\right) k\left(\frac{h}{m}\right) \sum_{u=-\infty}^{\infty} \Gamma_{\Delta x y}(u + h) \Gamma_{\Delta x y}(u - h') \phi_n(u, h', h) 
+ \frac{1}{m} \sum_{h'=1}^{n-1} \sum_{h=1}^{n-1} k\left(\frac{h'}{m}\right) k\left(\frac{h}{m}\right) \sum_{u=-\infty}^{\infty} k_{\Delta x y \Delta x y}(0, h', u, u + h) \phi_n(u, h', h). \tag{32}
$$

Let $v = h' - h$, and we can rewrite (32) as

$$
\sum_{u=-\infty}^{\infty} \sum_{v=-n+2}^{n-2} \Gamma_{\Delta x \Delta x}(u) \Gamma_{yy}(u - v) \left\{ \frac{1}{m} \sum_{h} k\left(\frac{h + v}{m}\right) k\left(\frac{h}{m}\right) \right\}, \tag{35}
$$

where the summation $\sum_{h}$ runs only for $\{h : 1 \leq h \leq n-1$ and $1 \leq h + v \leq n-1\}$. The bracketed expression converges to $\int_{0}^{\infty} k^2(x) \, dx$ by the argument in Hannan (1970) pp. 314-15. Furthermore,

$$
\sum_{u=-\infty}^{\infty} \sum_{v=-n+2}^{n-2} \Gamma_{\Delta x \Delta x}(u) \Gamma_{yy}(u - v) \rightarrow 4\pi^2 f_{\Delta x \Delta x}(0) f_{yy}(0) \text{ as } n \rightarrow \infty,
$$

and hence (32) converges to $4\pi^2 f_{\Delta x \Delta x}(0) f_{yy}(0) \int_{0}^{\infty} k^2(x) \, dx$ as $n \rightarrow \infty$. Similarly, (33) converges to $4\pi^2 f_{\Delta x y}(0) f_{\Delta x y}(0) \int_{0}^{\infty} k^2(x) \, dx = 4\pi^2 [f_{\Delta x y}(0)]^2 \int_{0}^{\infty} k^2(x) \, dx$. For (34), from Hannan (1970, p. 211), the fourth cumulant of $z_t$ satisfies

$$
\sum_{q=-\infty}^{\infty} \sum_{r=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} |k_{ijkl}(0, q, r, s)| < \infty, \quad i, j, k, l = \{y, \Delta x\}, \tag{36}
$$

and hence (34) is $O(m^{-1})$, and the stated result follows. 

### 7.3 Proof of Theorem 2

In view of Lemma 1, it suffices to show that $\sqrt{n/m}(\hat{\lambda}_{y, \Delta x} - E\hat{\lambda}_{y, \Delta x}) \rightarrow_d N(0, V)$. First, observe that

$$
\sqrt{n/m} \left(\hat{\lambda}_{y, \Delta x} - E\hat{\lambda}_{y, \Delta x}\right)
= \frac{1}{\sqrt{m}} \sum_{h=1}^{n-1} k\left(\frac{h}{m}\right) \frac{1}{\sqrt{n}} \sum_{t=h+1}^{n} (y_t \Delta x_{t-h} - E y_t \Delta x_{t-h}) = I + II, \tag{37}
$$

21
where

\[ I = \frac{1}{\sqrt{m}} \sum_{h=1}^{n-1} k \left( \frac{h}{m} \right) \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (y_t \Delta x_{t-h} - Ey_t \Delta x_{t-h}) , \]

\[ II = -\frac{1}{\sqrt{m}} \sum_{h=1}^{n-1} k \left( \frac{h}{m} \right) \frac{1}{\sqrt{n}} \sum_{t=1}^{h} (y_t \Delta x_{t-h} - Ey_t \Delta x_{t-h}) . \]

From Lemma 9, Minkowski’s inequality, and \( \int_0^\infty |k(x)|x^{1/2}dx < \infty \), we have

\[ E(II)^2 = O \left( \frac{m}{n} \sum_{h=1}^{n-1} k \left( \frac{h}{m} \right) \right)^2 = O \left( \frac{m^2}{n} \right), \] (38)

Lemma 10 gives

\[ I = \sum_{t=1}^{n} Z_t + R_n; \quad Z_t = n^{-1/2} m^{-1/2} \sum_{h=1}^{n-1} k \left( \frac{h}{m} \right) \sum_{r=1}^\infty \varepsilon_{t-r} f^{hr}(1) \varepsilon_t , \] (39)

where \( ER_n^2 = o(1) \) and \( f^{hr}(1) \) is defined in the statement of Lemma 10. Therefore, \( \sqrt{n/m} (\lambda_t y_t \Delta x - E\lambda_t y_t \Delta x) \rightarrow_d N(0, V) \) follows if we show

\[ \sum_{t=1}^{n} Z_t \rightarrow_d N(0, V), \quad \text{as} \ n \rightarrow \infty. \] (40)

Let \( \mathcal{I}_t = \sigma(\varepsilon_t, \varepsilon_{t-1}, \ldots) \). Since \( Z_t \in \mathcal{I}_t \) and \( E(Z_t|\mathcal{I}_{t-1}) = 0 \), \( Z_t \) is a martingale difference sequence and (40) follows from the martingale CLT of Brown (1971) if

(i) \( \sum_{t=1}^{n} E(Z^2_t|\mathcal{I}_{t-1}) = \frac{1}{n} \sum_{t=1}^{n} E(nZ^2_t|\mathcal{I}_{t-1}) \rightarrow_p V, \)

(ii) \( \sum_{t=1}^{n} E(Z^2_t 1\{|Z_t| \geq \delta\} \rightarrow_p 0 \quad \text{for all} \ \delta > 0. \)

First we show (i). Observe that

\[ E(nZ^2_t|\mathcal{I}_{t-1}) = m^{-1} \sum_{h=1}^{n-1} \sum_{u=1}^{n-1} k \left( \frac{h}{m} \right) k \left( \frac{u}{m} \right) \sum_{r=1}^\infty \sum_{s=1}^\infty \varepsilon_{t-r} f^{hr}(1)(f^{us}(1))' \varepsilon_{t-s}. \]

\( E(nZ^2_t|\mathcal{I}_{t-1}) \) is stationary and ergodic because \( \varepsilon_t \) is i.i.d. Furthermore, from the law of iterated expectations we have

\[ E \left[ E(nZ^2_t|\mathcal{I}_{t-1}) \right] = nEZ^2_t. \]

Therefore, (i) follows from the ergodic theorem if

\[ nEZ^2_t \rightarrow V. \] (41)

22
From (37)-(39), we have
\[
\sqrt{\frac{n}{m}} \left( \hat{\lambda}_{y,\Delta x} - E\hat{\lambda}_{y,\Delta x} \right) = \sum_{t=1}^{n} Z_t + II + R_n, \quad E(II + R_n)^2 = o(1),
\]
or equivalently,
\[
\sum_{t=1}^{n} Z_t = \sqrt{\frac{n}{m}} \left( \hat{\lambda}_{y,\Delta x} - E\hat{\lambda}_{y,\Delta x} \right) - (II + R_n).
\]
Taking the second moment of the both sides gives
\[
E \left( \sum_{t=1}^{n} Z_t \right)^2 = E \left( \sqrt{\frac{n}{m}} \left( \hat{\lambda}_{y,\Delta x} - E\hat{\lambda}_{y,\Delta x} \right) - (II + R_n) \right)^2. \tag{42}
\]
The left hand side of (42) is \(\sum_{t=1}^{n} EZ_t^2 = nEZ_t^2\), since \(Z_t\) is a stationary martingale difference sequence. The right hand side of (42) is
\[
\text{var} \left( \sqrt{\frac{n}{m}} \left( \hat{\lambda}_{y,\Delta x} - E\hat{\lambda}_{y,\Delta x} \right) \right) = \text{var} \left( \sqrt{\frac{n}{m}} \left( \hat{\lambda}_{y,\Delta x} - E\hat{\lambda}_{y,\Delta x} \right) \right) \to V,
\]
because
\[
E \left( \sqrt{\frac{n}{m}} \left( \hat{\lambda}_{y,\Delta x} - E\hat{\lambda}_{y,\Delta x} \right) \right)^2 = \text{var} \left( \sqrt{\frac{n}{m}} \left( \hat{\lambda}_{y,\Delta x} - E\hat{\lambda}_{y,\Delta x} \right) \right) \to V,
\]
and \(E(II + R_n)^2 = o(1)\).
Therefore, we establish (41) and (i). For (ii), the stationarity of \(Z_t\) gives \(\sum_{t=1}^{n} E(Z_t^21\{|Z_t| \geq \delta\}) = E(nZ_t^21\{|nZ_t^2| \geq n\delta^2\}) \to 0\) follows from \(E(nZ_t^2) \to V < \infty\) and the dominated convergence theorem, giving (40) and the stated result follows. ■

7.4 Proof of Lemma 3

From (29), we have
\[
\frac{1}{n} \sum_{t=h+1}^{n} y_t \Delta x_t - h = \frac{1}{n} \sum_{t=h+1}^{n} y_t u_{t-h} - T_{nh}, \quad T_{nh} = \frac{c}{n^2} \sum_{t=h+1}^{n} \sum_{k=0}^{t-h-2} \left( 1 - \frac{c}{n} \right)^k y_t u_{t-h-1-k}.
\]
The required result follows because \(E|\sum_{h=1}^{n-1} k(h/m)T_{nh}| \) is bounded by \(n^{-2} \sum_{h=1}^{n-1} |k(h/m)| \times \sum_{t=h+1}^{n} \sum_{k=-\infty}^{\infty} |\Gamma_{uy}(k)| = O(n^{-1} \sum_{h=1}^{n-1} |k(h/m)|) = O(mn^{-1})\). ■
7.5 Proof of Lemma 4

Some simple algebra gives

\[ \hat{\lambda}_{y, \Delta x} = \sum_{h=1}^{n-1} k \left( \frac{h}{m} \right) \frac{1}{n} \sum_{t=h+1}^{n} y_t \Delta x_{t-h} \]

\[ = \sum_{h=1}^{n-1} k \left( \frac{h}{m} \right) \frac{1}{n} \sum_{t=h+1}^{n} y_t x_{t-h} - \sum_{h=1}^{n-1} k \left( \frac{h}{m} \right) \frac{1}{n} \sum_{t=h+1}^{n} y_t x_{t-h-1} \]

\[ = \sum_{h=1}^{n-1} k \left( \frac{h}{m} \right) \frac{1}{n} \sum_{t=h+1}^{n} y_t x_{t-h} - \frac{1}{n} \sum_{t=p}^{n} y_t x_{t-p} \quad (p = h + 1) \]

\[ = k \left( \frac{1}{m} \right) \frac{1}{n} \sum_{t=2}^{n} y_t x_{t-1} + \sum_{h=2}^{n-1} \left[ k \left( \frac{h}{m} \right) - k \left( \frac{h-1}{m} \right) \right] \left( \frac{1}{n} \sum_{t=h+1}^{n} y_t x_{t-h} \right) \]

\[ = T_{1n} + T_{2n} + T_{3n} + T_{4n}. \]

For \( T_{1n} \), we have (note that \( \lambda_{y, \Delta x} = E y_t x_{t-1} = \gamma_{xy}(1) \))

\[ \sqrt{n} (T_{1n} - \lambda_{y, \Delta x}) = k(1/m) \sqrt{n} (\hat{\gamma}_{xy}(1) - \gamma_{xy}(1)) + (k(1/m) - 1) \sqrt{n} E y_t x_{t-1}. \]

From Theorem 14 of Hannan (1970, page 228) and \( k(1/m) \to 1 \) as \( n \to \infty \), we have

\[ k(1/m) \sqrt{n} (\hat{\gamma}_{xy}(1) - \gamma_{xy}(1)) \to_d N(0, \Xi), \quad \text{as } n \to \infty, \]

where \( \Xi \) is given by Hannan (1970) in equation (3.3) on page 209 and line 5 on page 211. The second term is \( O(n^{1/2} m^{-q}) E y_t x_{t-1} \) from Assumption K.

For \( T_{2n} \), first observe that

\[ E(T_{2n}) = \sum_{h=2}^{n-1} \left[ k \left( \frac{h}{m} \right) - k \left( \frac{h-1}{m} \right) \right] \frac{n-h}{n} \gamma_{xy}(h). \]

\( ET_{2n} = 0 \) when \( E y_t x_{t-h} = \gamma_{xy}(h) = 0 \) for all \( h \geq 1 \). Otherwise, fix a small \( \varepsilon > 0 \), then

\[ E|T_{2n}| \leq \sum_{h=2}^{\varepsilon m} k \left( \frac{h}{m} \right) - k \left( \frac{h-1}{m} \right) |\gamma_{xy}(h)| + C \sum_{h=\varepsilon m+1}^{n-1} |\gamma_{xy}(h)|. \]

Since \( k(x) - 1 = O(x^q) \) as \( x \to 0 \), the first term on the right is, for \( \varepsilon \) sufficiently small, \( O(\sum_{h=2}^{\varepsilon m} h/m)^q |\gamma_{xy}(h)| = O(m^{-q}) \). The second term on the right is bounded by \( \sum_{h=\varepsilon m}^{n-1} |\gamma_{xy}(h)| \leq (\varepsilon m)^{-q} \sum_{h=\varepsilon m}^{n-1} h^q |\gamma_{xy}(h)| = O(m^{-q}) \). Therefore, defining \( B_n = (k(1/m) - 1) \sqrt{n} E y_t x_{t-1} + \sqrt{n} ET_{2n} \) gives the bias term \( B_n \) in (14).
It remains to show that \( \text{var}(\sqrt{n}T_{2n}) = o(1) \) and \( \sqrt{n}(T_{3n} + T_{4n}) = o_p(1) \). From Hannan (1970) (equation (3.3) on page 209 and line 5 on page 211), we have

\[
\text{cov} \left( \sqrt{n} \hat{\gamma}_{xy}(h), \sqrt{n} \hat{\gamma}_{xy}(h') \right) = 
\sum_{u=-n+1}^{n-1} \left( 1 - \frac{|u|}{n} \right) \{ \gamma_{xx}(u) \gamma_{yy}(u + h - h') + \gamma_{xy}(u + h) \gamma_{yx}(u - h') \}
+ \sum_{u=-n+1}^{n-1} \left( 1 - \frac{|u|}{n} \right) k_{xyxy}(0, h, u, u + h').
\]

Therefore, from the Lipschitz condition on \( k(\cdot) \), the terms composing the variance of \( \sqrt{n}T_{2n} \) that do not involve \( k_{xyxy} \) are bounded by

\[
\frac{1}{m^2} \sum_{h=1}^{m} \sum_{h'=1}^{m} \sum_{u=-n+1}^{n-1} |\gamma_{xx}(u) \gamma_{yy}(u + h - h') + \gamma_{xy}(u + h) \gamma_{yx}(u - h')| 
\leq \frac{1}{m} \left[ \sum_{u=-\infty}^{\infty} |\gamma_{xx}(u)| \sum_{h=-\infty}^{\infty} |\gamma_{yy}(h)| + \sum_{u=-\infty}^{\infty} |\gamma_{xy}(u)| \sum_{h'=-\infty}^{\infty} |\gamma_{yx}(h')| \right] = O(m^{-1}).
\]

The term in the variance of \( \sqrt{n}T_{2n} \) that involves \( k_{xyxy} \) is bounded by

\[
\frac{1}{m^2} \sum_{h=1}^{m} \sum_{h'=1}^{m} \sum_{u=-n+1}^{n-1} |k_{xyxy}(0, h, u, u + h')| = O(m^{-2}),
\]

because \( \sum_{q=-\infty}^{\infty} \sum_{r=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} |k_{xyxy}(0, q, r, s)| < \infty \) from Hannan (1970, p. 211).

Finally, \( \sqrt{n}(T_{3n} + T_{4n}) = o_p(1) \) follows from

\[
\sqrt{n}(T_{3n} + T_{4n}) = \sum_{p=2}^{n} k \left( \frac{p-1}{m} \right) \frac{1}{\sqrt{n}} y_p x_0,
\]

\( x_0 = O_p(1) \), and

\[
E \left( \sum_{p=2}^{n} k \left( \frac{p-1}{m} \right) \frac{1}{\sqrt{n}} y_p \right)^2 \leq \frac{1}{n} \sum_{p=2}^{\infty} k \left( \frac{p-1}{m} \right) \sum_{r=-\infty}^{\infty} |\gamma_{yy}(r)| = O \left( \frac{m}{n} \right),
\]

and the stated result follows.

\[\boxed{\text{7.6 Proof of Lemma 5}}\]

The term of \( \tilde{V} \) that involves \( \hat{\Gamma}_{\Delta x \Delta x}(u) \) and \( \hat{\Gamma}_{yy}(u + h - h') \) reduces to, in view of equations (32) and (35) in the proof of Lemma 1,

\[
\sum_{v=-n+2}^{n-2} \sum_{u=-\infty}^{\infty} \tilde{k} \left( \frac{u}{m} \right) \hat{\Gamma}_{\Delta x \Delta x}(u) \tilde{k} \left( \frac{u - v}{m} \right) \hat{\Gamma}_{yy}(u - v) \left\{ \int_{u-v}^{\infty} k^2(x) \, dx + o(1) \right\}.
\]

25
Because \( \tilde{k}(x) = 0 \) for \(|x| > 1\) and \( \tilde{m}/n \to 0\), this simplifies to

\[
\sum_{u=-\tilde{m}}^{\tilde{m}} \tilde{k} \left( \frac{u}{\tilde{m}} \right) \hat{\Gamma}_{\Delta x \Delta x} (u) \sum_{v=-\tilde{m}}^{\tilde{m}} \tilde{k} \left( \frac{u-v}{\tilde{m}} \right) \hat{\Gamma}_{yy}(u-v) \left\{ \int_0^\infty k^2(x) \, dx + o(1) \right\},
\]

which converges to \( 4\pi^2 f_{\Delta x \Delta x}(0) \int_0^\infty k^2(x) \, dx \) in probability by the standard argument. Similarly, the term of \( V \) that involves \( \hat{\Gamma}_{\Delta xy}(u+h) \) and \( \hat{\Gamma}_{\Delta y \Delta x}(u-h') \) converges to \( 4\pi^2 [f_{\Delta y \Delta x}(0)]^2 \int_0^\infty k^2(x) \, dx \) in probability, and the stated result follows. The result for the local-to-unity case follows from the proof of Lemma 3.

### 7.7 Proof of Lemma 7

The Lemma follows if we show that there exists \( \eta > 0 \) such that

\[
\Pr(\tilde{V} \geq \eta \tilde{m}^{-1}) \to 1, \quad \text{as } n \to \infty. \tag{43}
\]

From the arguments in the proof of Lemma 5, \( \tilde{V} \) is equal to

\[
\left[ \sum_{u=-\tilde{m}}^{\tilde{m}} \tilde{k} \left( \frac{u}{\tilde{m}} \right) \hat{\Gamma}_{\Delta x \Delta x} (u) \sum_{v=-\tilde{m}}^{\tilde{m}} \tilde{k} \left( \frac{u-v}{\tilde{m}} \right) \hat{\Gamma}_{yy}(v) \left\{ \int_0^\infty k^2(x) \, dx + o(1) \right\} \right] + \left[ \sum_{u=-\tilde{m}}^{\tilde{m}} \tilde{k} \left( \frac{u}{\tilde{m}} \right) \hat{\Gamma}_{\Delta xy} (u) \sum_{v=-\tilde{m}}^{\tilde{m}} \tilde{k} \left( \frac{u-v}{\tilde{m}} \right) \hat{\Gamma}_{\Delta y \Delta x}(v) \left\{ \int_0^\infty k^2(x) \, dx + o(1) \right\} \right] \tag{44}
\]

For sufficiently large \( n \), (45) is equal to

\[
\sum_{u=-\tilde{m}}^{\tilde{m}} \tilde{k} \left( \frac{u}{\tilde{m}} \right) \hat{\Gamma}_{\Delta x \Delta x} (u) \left\{ \int_0^\infty k^2(x) \, dx + o(1) \right\} \geq 0.
\]

Because \( \sum_{v=-\tilde{m}}^{\tilde{m}} \tilde{k}(v/\tilde{m}) \hat{\Gamma}_{yy}(v) \to_p f_y(0) > 0 \) by the standard argument, (43) follows if there exists \( \varepsilon > 0 \) such that

\[
\Pr \left( \sum_{v=-\tilde{m}}^{\tilde{m}} \tilde{k} \left( \frac{v}{\tilde{m}} \right) \hat{\Gamma}_{\Delta x \Delta x}(v) \geq \varepsilon \tilde{m}^{-1} \right) \to 1, \quad \text{as } n \to \infty, \tag{46}
\]

where (Priestley, 1981, p. 439)

\[
W_\tilde{m}(\lambda) = \frac{1}{2\pi} \sum_{h=-\tilde{m}}^{\tilde{m}} \tilde{k} \left( \frac{h}{\tilde{m}} \right) e^{i\lambda h} = \frac{1}{2\pi \tilde{m}} \frac{\sin^2(\tilde{m}\lambda/2)}{\sin^2(\lambda/2)} \geq 0,
\]

is the Fejer kernel. From Phillips (1999, Theorem 2.2 and Remark 2.4), we have

\[
w_{\Delta x}(\lambda) = \left( 1 - e^{i\lambda} \right) w_x(\lambda) + e^{i(n+1)\lambda} (2\pi n)^{-1/2} \chi_n.
\]
It follows that

\[
\int_{-\pi}^{\pi} W_m(\lambda) I_{\Delta x}(\lambda) d\lambda \\
= \int_{-\pi}^{\pi} W_m(\lambda) |1 - e^{i\lambda}|^2 I_x(\lambda) d\lambda \tag{47}
+ (2\pi n)^{-1/2} X_n \int_{-\pi}^{\pi} W_m(\lambda) 2 \text{Re} \left[ (1 - e^{i\lambda}) w_x(\lambda) e^{-i(n+1)\lambda} \right] d\lambda \tag{48}
+ \int_{-\pi}^{\pi} W_m(\lambda) d\lambda (2\pi n)^{-1} X_n^2. \tag{49}
\]

We can ignore (49) because it is nonnegative. For (48), it follows from the Cauchy-Schwartz inequality and Lemma 12 (b) that

\[
\int_{-\pi}^{\pi} W_m(\lambda) 2 \text{Re} \left[ (1 - e^{i\lambda}) w_x(\lambda) e^{-i(n+1)\lambda} \right] d\lambda \\
\leq \left( \int_{-\pi}^{\pi} W_m(\lambda) \left| 2 \text{Re} \left[ (1 - e^{i\lambda}) w_x(\lambda) e^{-i(n+1)\lambda} \right] \right|^2 d\lambda \right)^{1/2} \left( \int_{-\pi}^{\pi} W_m(\lambda) d\lambda \right)^{1/2} \\
= O_p \left( \left( \int_{-\pi}^{\pi} W_m(\lambda) \lambda^2 d\lambda \right)^{1/2} \right) = O_p \tilde{m}^{-1/2},
\]

and (48) = \( O_p(n^{-1/2}\tilde{m}^{-1/2}) = o_p(\tilde{m}^{-1}) \) follows. Rewrite (47) as

\[
\int_{-\pi}^{\pi} W_m(\lambda) |1 - e^{i\lambda}|^2 E I_x(\lambda) d\lambda \\
+ \int_{-\pi}^{\pi} W_m(\lambda) |1 - e^{i\lambda}|^2 (I_x(\lambda) - EI_x(\lambda)) d\lambda \\
= A_1 + A_2.
\]

For \( A_1 \), because \( f_x(0) > 0 \) and \( f_x(\lambda) \) is continuous in the neighborhood of the origin since \( \sum j||B_j|| < \infty \), there exist \( D \in (0, 1) \) and \( c_1, c_2 > 0 \) such that, for sufficiently large \( n \) (Hannan, Theorem 2, p. 248)

\[
\inf_{\lambda \in [-D\pi, D\pi]} |1 - e^{i\lambda}|^2 \lambda^{-2} \geq c_1, \quad \inf_{\lambda \in [-D\pi, D\pi]} E I_x(\lambda) \geq c_2.
\]

Therefore, in conjunction with Lemma 12 (a), we obtain

\[
A_1 \geq c_1 c_2 \int_{-D\pi}^{D\pi} W_m(\lambda) \lambda^2 d\lambda \geq c_1 c_2 \kappa \tilde{m}^{-1}, \quad \kappa > 0.
\]

For \( A_2 \), it follows from Theorem 2 and Corollary 1 of Hannan (1970, pp. 248-9) and their proof that

\[
\left\{ \begin{array}{l}
\sup_{\lambda, \lambda' \in [-\pi, \pi]} \left| \text{cov} \left( I_x(\lambda), I_x(\lambda') \right) \right| = O(1), \\
\text{cov} \left( I_x(\lambda), I_x(\lambda') \right) = o(1), \quad \lambda \neq \lambda'.
\end{array} \right. \tag{50}
\]

27
Therefore,

\[ E(A_2^2) = \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} W_m(\lambda) W_m(\lambda') |1-e^{i\lambda}|^2 |1-e^{i\lambda'}|^2 \text{cov}(I_x(\lambda), I_x(\lambda')) \, d\lambda d\lambda' \]

\[ \leq C \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} W_m(\lambda) W_m(\lambda') \lambda^2(\lambda')^2 |\text{cov}(I_x(\lambda), I_x(\lambda'))| \, d\lambda d\lambda' \]

\[ = o(\tilde{m}^{-2}) \]

where the interchange of expectation and integration in the first line is valid by (50) and Fubini’s Theorem, and the last line follows from Lemma 12 (b), (50), and the dominated convergence theorem. Therefore, there exists \( \eta' > 0 \) such that (47)+(48)+(49)\( \geq \eta' \tilde{m}^{-1} \) with probability approaching one, and (46) and the stated result follow. ■

8 Appendix B: technical results

Lemma 9 Under the assumptions of Theorem 2,

\[ E \left( \sum_{t=1}^{h} (y_t \Delta x_{t-h} - E y_t \Delta x_{t-h}) \right)^2 = O(h), \quad h = 1, \ldots, n - 1. \]

Proof Observe that

\[ E \left( \sum_{t=1}^{h} (y_t \Delta x_{t-h} - E y_t \Delta x_{t-h}) \right)^2 = \text{var} \left( \sum_{t=1}^{h} y_t \Delta x_{t-h} \right) \leq E \left( \sum_{t=1}^{h} y_t \Delta x_{t-h} \right)^2. \]

From the product theorem (e.g. Hannan, 1970, pp. 23, 209), \( E(\sum_{t=1}^{h} y_t \Delta x_{t-h})^2 \) is equal to (recall \( \Gamma_y \Delta x(h) = E y_t \Delta x_{t+h} \))

\[ E \left( \sum_{t=1}^{h} y_t \Delta x_{t-h} \sum_{s=1}^{h} y_s \Delta x_{s-h} \right) \]

\[ = \sum_{t=1}^{h} \sum_{s=1}^{h} \Gamma_y \Delta x(h) \Gamma_y \Delta x(h) + \sum_{t=1}^{h} \sum_{s=1}^{h} \Gamma_y (s-t) \Gamma \Delta x \Delta x(s-t) \]

\[ + \sum_{t=1}^{h} \sum_{s=1}^{h} \Gamma_y \Delta x(s-h-t) \Gamma \Delta x_y(s-t+h) + \sum_{t=1}^{h} \sum_{s=1}^{h} k_y \Delta x_y \Delta x(t, t-h, s, s-h) \]

\[ = h^2(\Gamma_y \Delta x(h))^2 + \sum_{l=-h+1}^{h-1} (h - |l|) \Gamma y_l (l) \Gamma \Delta x \Delta x(l) \]

\[ + \sum_{l=-h+1}^{h-1} (h - |l|) \Gamma y (l-h) \Gamma \Delta x (l+h) + \sum_{l=-h+1}^{h-1} (h - |l|) k_y \Delta x_y \Delta x(0, -h, l, l-h). \]
The first term on the right is bounded by \((\sup_s |\Gamma_y x(s)|^2 < \infty\). The second and third terms on the right are bounded by \(h \sup_s ||\Gamma(s)|| \sum_{t=-\infty}^{\infty} |k_t x y| \leq Ch\). From (36), the fourth term on the right is bounded by \(h \sum_{t=-\infty}^{\infty} \sum_{t=-\infty}^{\infty} |k_t x y| \leq Ch\), and the stated result follows.

\[\Box\]

**Lemma 10** Under the assumptions of Theorem 2,

\[
\frac{1}{\sqrt{n}} \sum_{h=1}^{n-1} k(h) \left( \frac{h}{m} \right) \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (y_t x_{t-h} - E y_t x_{t-h}) = \sum_{t=1}^{n} Z_t + R_n,
\]

where \(ER_n^2 = o(1)\) and

\[
Z_t = n^{-1/2} m^{-1/2} \sum_{h=1}^{n-1} k(h) \left( \frac{h}{m} \right) \sum_{r=1}^{\infty} \sum_{t-r \geq h} f^{hr}(1) \varepsilon_t,
\]

\[
f^{hr}(1) = \sum_{j=0}^{\infty} \left[ (A_j^2)_{j+r-h} \right] A_j^1 + (A_j^4)_{j+r} A_j^2_{j-r},
\]

and \(A_j^1\) and \(A_j^2\) denote the first and second row of \(A_j\), respectively.

**Proof** The proof follows from an argument similar to Remark 3.9 (i) of Phillips and Solo (1992, p. 980). First, we find an alternate expression of \(\sum_{t=1}^{n} y_t x_{t-h}\) so that it can be approximated by a martingale. Express \(y_t\) and \(x_t\) as

\[
\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} A^1(L) \varepsilon_t \\ A^2(L) \varepsilon_t \end{pmatrix} = \begin{pmatrix} \sum_{j=0}^{\infty} A^1_j \varepsilon_{t-j} \\ \sum_{j=0}^{\infty} A^2_j \varepsilon_{t-j} \end{pmatrix},
\]

where \(A^1_j\) and \(A^2_j\) are the first and second row of \(A_j\), respectively. Observe that

\[
y_t x_{t-h} = A^1(L) \varepsilon_t A^2(L) \varepsilon_{t-h} = \sum_{j=0}^{\infty} A^1_j \varepsilon_{t-j} \sum_{k=0}^{\infty} A^2_k \varepsilon_{t-h-k} = \sum_{j=0}^{\infty} A^1_j \varepsilon_{t-j} A^2_{s-h} \varepsilon_{t-s}, \quad (s = h + k).
\]

Since \(A^2_{s-h} \varepsilon_{t-s}\) is a scalar, the first term on the right is

\[
\text{tr} \left( \sum_{j=0}^{\infty} (A^2_{j-h})' A^1_j \varepsilon_{t-j} \varepsilon_{t-j}' \right) = \text{tr} \left( f^{h0}(L) \varepsilon_t \varepsilon_t' \right), \quad f^{h0}(L) = \sum_{j=0}^{\infty} (A^2_{j-h})' A^1_j = \sum_{j=0}^{\infty} f^{h0} L^j.
\]
The second term on the right is, since $A_{s}^{2} \equiv 0$ for $s < 0$, 

\[
\begin{align*}
\text{tr} \left( \sum_{j=0}^{\infty} \sum_{s=0, s \neq j}^{\infty} (A_{s-h}^{2})' A_{j}^1 \varepsilon_{t-j} \varepsilon'_{t-s} \right)
\end{align*}
\]

\[
= \text{tr} \left( \sum_{j=0}^{\infty} \sum_{s=j+1}^{\infty} (A_{s-h}^{2})' A_{j}^1 \varepsilon_{t-j} \varepsilon'_{t-s} \right) + \text{tr} \left( \sum_{j=0}^{\infty} \sum_{s=0}^{j-1} (A_{s-h}^{2})' A_{j}^1 \varepsilon_{t-j} \varepsilon'_{t-s} \right)
\]

\[
= \text{tr} \left( \sum_{j=0}^{\infty} \sum_{s=j+1}^{\infty} (A_{s-h}^{2})' A_{j}^1 \varepsilon_{t-j} \varepsilon'_{t-s} \right) + \text{tr} \left( \sum_{s=0}^{\infty} \sum_{j=s+1}^{\infty} (A_{j}^1)' A_{s-h}^{2} \varepsilon_{t-s} \varepsilon'_{t-j} \right)
\]

\[
= \text{tr} \left( \sum_{j=0}^{\infty} \sum_{s=j+1}^{\infty} [(A_{s-h}^{2})' A_{j}^1 + (A_{s}^1)' A_{j-h}^{2}] \varepsilon_{t-j} \varepsilon'_{t-s} \right)
\]

\[
= \text{tr} \left( \sum_{r=1}^{\infty} f^{hr}(L) \varepsilon_{t} \varepsilon'_{t-r} \right),
\]

where

\[
f^{hr}(L) = \sum_{j=0}^{\infty} f^{hr}_{j} L_{j}^2, \quad f^{hr}_{j} = (A_{j+1+r-h}^{2})' A_{j}^1 + (A_{j+r}^1)' A_{j-h}^{2}.
\]

Therefore, we may express $y_{t} \Delta x_{t-h}$ as

\[
y_{t} \Delta x_{t-h} = \text{tr} \left( f^{h0}(L) \varepsilon_{t} \varepsilon'_{t} + \sum_{r=1}^{\infty} f^{hr}(L) \varepsilon_{t} \varepsilon'_{t-r} \right).
\]

Apply the B/N decomposition (Phillips and Solo (1992)) to $f^{hr}(L)$ and rewrite it as

\[
f^{hr}(L) = f^{hr}(1) - (1 - L) \tilde{f}^{hr}(L), \quad r = 0, 1, \ldots,
\]

with

\[
\tilde{f}^{hr}(L) = \sum_{j=0}^{\infty} \tilde{f}^{hr}_{j} L_{j}^2, \quad \tilde{f}^{hr}_{j} = \sum_{s=j+1}^{\infty} f^{hr}_{s} = \sum_{s=j+1}^{\infty} [(A_{s}^2)_{r+s-r-h}^{2} + (A_{s}^1)_{r+s-r}^{1}] \quad (51)
\]

It follows that

\[
\frac{1}{\sqrt{n}} \sum_{t=1}^{n} y_{t} \Delta x_{t-h} = \text{tr} \left( f^{h0}(1) \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \varepsilon_{t} \varepsilon'_{t} + \sum_{r=1}^{\infty} f^{hr}(1) \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \varepsilon_{t} \varepsilon'_{t-r} \right) + r_{nh}, \quad (52)
\]

where

\[
r_{nh} = \frac{1}{\sqrt{n}} \text{tr} \left( \tilde{f}^{h0}(L)(\varepsilon_{0} \varepsilon'_{0} - \varepsilon_{n} \varepsilon'_{n}) \right) + \frac{1}{\sqrt{n}} \text{tr} \left( \sum_{r=1}^{\infty} \tilde{f}^{hr}(L)(\varepsilon_{0} \varepsilon'_{0} - \varepsilon_{n} \varepsilon'_{n-r}) \right).
\]
From Lemma 11, we have
\[ E|r_{nh}|^2 \leq Cn^{-1}, \quad h = 1, \ldots, n - 1. \] (53)

Furthermore, observe that
\[ Ey_t \Delta x_{t-h} = E \left( \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} A_j^1 \varepsilon_{t-j-k-h} (A_k^2)' \right) \]
\[ = \sum_{j=0}^{\infty} A_j^1 (A_{j-h})' = \text{tr} \left( \sum_{j=0}^{\infty} (A_{j-h}^2)' A_j^1 \right) = \text{tr} (f^{h_0}(1)). \]

In conjunction with (52), it follows that
\[ \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (y_t \Delta x_{t-h} - Ey_t \Delta x_{t-h}) = \text{tr} \left( f^{h_0}(1) \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (\varepsilon_t \varepsilon_t' - I_2) + \sum_{r=1}^{\infty} f^{hr}(1) \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \varepsilon_t \varepsilon_{t-r} \right) + r_{nh}, \]
and hence
\[ \frac{1}{\sqrt{m}} \sum_{h=1}^{m} k \left( \frac{h}{m} \right) \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (y_t \Delta x_{t-h} - Ey_t \Delta x_{t-h}) = I + II + III, \]

where \( III = m^{-1/2} \sum_{h=1}^{m} k(h/m) r_{nh} \) and

\[ I = \frac{1}{\sqrt{m}} \sum_{h=1}^{m} k \left( \frac{h}{m} \right) \text{tr} \left( f^{h_0}(1) \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (\varepsilon_t \varepsilon_t' - I_2) \right) \]
\[ = \text{tr} \left( \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (\varepsilon_t \varepsilon_t' - I_2) \right) \frac{1}{\sqrt{m}} \sum_{h=1}^{m} k \left( \frac{h}{m} \right) f^{h_0}(1) \]
\[ II = \frac{1}{\sqrt{m}} \sum_{h=1}^{m} k \left( \frac{h}{m} \right) \text{tr} \left( \sum_{r=1}^{\infty} f^{hr}(1) \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \varepsilon_t \varepsilon_{t-r} \right) \]
\[ = \sum_{t=1}^{n} Z_t, \quad Z_t = n^{-1/2} m^{-1/2} \sum_{h=1}^{m} k \left( \frac{h}{m} \right) \sum_{r=1}^{\infty} \varepsilon_{t-r} f^{hr}(1) \varepsilon_t. \]

From (53) and Minkowski’s inequality, we have \( E(III)^2 = O(m^{-1}(\sum_{h=1}^{m} n^{-1/2})^2) = O(mn^{-1}) \). For \( I \), first observe that, since \( A_j \equiv 0 \) for \( j < 0 \),
\[ \|f^{h_0}(1)\| = \left\| \sum_{j=0}^{\infty} (A_j^2)' A_j^1 \right\| = \left\| \sum_{j=0}^{\infty} (A_j^2)' A_j^1 \right\| \]
\[ \leq \sup_s \|A_s\| \sum_{j=0}^{\infty} \|A_j\| \leq Cn^{-\delta} \sum_{j=h}^{\infty} j^\delta \|A_j\| \leq Cn^{-\delta}, \quad h = 1, \ldots, n - 1. \]
Therefore, \( |m^{-1/2} \sum_{h=1}^{m} k(h/m) f^{h0}(1)| \leq C m^{-1/2} \), and it follows that \( E(I)^2 = O(m^{-1}) \), giving the stated result. ■

**Lemma 11** Under the assumptions of Theorem 2, for \( h = 1, \ldots, n-1 \),

\[
(a) \quad E \left( \text{tr} \left( \tilde{f}^{h0}(L) \varepsilon_t \varepsilon_t' \right) \right)^2 < \infty, \quad (b) \quad E \left( \text{tr} \left( \sum_{r=1}^{\infty} \tilde{f}^{hr}(L) \varepsilon_t \varepsilon_{t-r}' \right) \right)^2 < \infty.
\]

**Proof** We need to show the result only for \( t = n \), because \( \varepsilon_t \) is i.i.d. For part (a), since \( \text{tr}(\tilde{f}^{h0}(L) \varepsilon_n \varepsilon_n') = \sum_{j=0}^{\infty} \text{tr}(\tilde{f}^{h0}(L) \varepsilon_{n-j} \varepsilon_{n-j}') = \sum_{j=0}^{\infty} \varepsilon_{n-j}' \tilde{f}^{h0}(L) \varepsilon_{n-j} \), we have

\[
E \left( \text{tr} \left( \tilde{f}^{h0}(L) \varepsilon_n \varepsilon_n' \right) \right)^2 = \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} E \left( \varepsilon_{n-j}' \tilde{f}^{h0}(L) \varepsilon_{n-k} \right)^2 \leq C \left( \sum_{j=0}^{\infty} ||\tilde{f}^{h0}(L)|| \right)^2 + C \sum_{j=0}^{\infty} ||\tilde{f}^{h0}(L)||^2.
\]

This is finite because, uniformly in \( h = 1, \ldots, n-1 \),

\[
||\tilde{f}^{h0}(L)|| = \left| \sum_{s=j+1}^{\infty} f^{h0}_s \right| \leq \sum_{s=j+1}^{\infty} \|(A^2_{s-h})'A^1_s\| \leq \sup_r ||A_r^1|| (j+1)^{-\delta} \sum_{s=j+1}^{\infty} s^\delta ||A_s^1|| \leq C j^{-\delta},
\]

and \( \delta > 1 \).

For part (b), rewrite \( \text{tr}(\sum_{r=1}^{\infty} \tilde{f}^{hr}(L) \varepsilon_n \varepsilon_{n-r}') \) as

\[
\sum_{r=1}^{\infty} \sum_{j=0}^{\infty} \text{tr} \left( \tilde{f}^{hr}(L) \varepsilon_{n-j} \varepsilon_{n-r-j}' \right) = \sum_{j=0}^{\infty} \sum_{r=1}^{\infty} \varepsilon_{n-j}' \left( \tilde{f}^{hr}(L) \right)' \varepsilon_{n-r-j} = \sum_{j=0}^{\infty} \xi^h_{n-j},
\]

where \( \xi^h_{n-j} = \varepsilon_{n-j}' \sum_{r=1}^{\infty} (\tilde{f}^{hr}(L))' \varepsilon_{n-r-j} \). Since \( \xi^h_{n-j} \in I_{n-j} = \sigma(\varepsilon_{n-j}, \varepsilon_{n-j-1}, \ldots) \) and \( E(\xi^h_{n-j}|I_{n-j-1}) = 0 \), it follows that

\[
E \left( \sum_{j=0}^{\infty} \xi^h_{n-j} \right)^2 = \sum_{j=0}^{\infty} E(\xi^h_{n-j})^2 \leq C \sum_{j=0}^{\infty} \sum_{r=1}^{\infty} \|\tilde{f}^{hr}(L)\|^2 \leq C \left( \sup_{j,r} \|\tilde{f}^{hr}(L)\| \right) \sum_{j=0}^{\infty} \sum_{r=1}^{\infty} ||\tilde{f}^{hr}(L)||. \tag{54}
\]

Now

\[
||\tilde{f}^{hr}(L)|| = \left| \sum_{s=j+1}^{\infty} f^{hr}_s \right| = \left| \sum_{s=j+1}^{\infty} (A^2_{s+r-h})'A^1_s + \sum_{s=j+1}^{\infty} (A^1_{s+r})'A^2_{s-h} \right|.
\]
Hence $\sup_h \sup_{j,r} ||\tilde{f}_{j,r}^h|| \leq \sup_p ||A_p|| \sum_{s=0}^{\infty} ||A_s|| < \infty$. Furthermore, uniformly in $h = 1, \ldots, n - 1$,

$$
\sum_{j=0}^{\infty} \sum_{r=1}^{\infty} ||\tilde{f}_{j,r}^h|| \leq \sum_{j=0}^{\infty} \sum_{r=1}^{\infty} \sum_{s=j+1}^{\infty} ||A_{s+r-h}|| ||A_s|| + \sum_{j=0}^{\infty} \sum_{r=1}^{\infty} \sum_{s=j+1}^{\infty} ||A_{s+r}|| ||A_{s-h}|| \\
\leq \sum_{j=0}^{\infty} \sum_{s=j+1}^{\infty} ||A_s|| \sum_{r=0}^{\infty} ||A_r|| + \sum_{j=0}^{\infty} \sum_{s=j+1}^{\infty} ||A_{s-h}|| \sum_{r=0}^{\infty} ||A_r||. \quad (55)
$$

The first term in (55) is bounded by $\sum_{j=0}^{\infty} \sum_{s=j+1}^{\infty} ||A_s|| = \sum_{j=1}^{\infty} j ||A_j|| < \infty$. The second term in (55) is bounded by

$$
\sum_{j=0}^{\infty} \sum_{p=\max\{j-h+1,0\}}^{\infty} ||A_p|| = \sum_{j=h+1}^{\infty} \sum_{p=j-h+1}^{\infty} ||A_p|| = \sum_{j=1}^{\infty} \sum_{p=s+1}^{\infty} ||A_p|| = \sum_{s=1}^{\infty} s ||A_s|| < \infty.
$$

Therefore, the right hand side of (54) is finite, and part (b) follows. ■

**Lemma 12** For $W_m(\lambda) = (2\pi \tilde{m})^{-1}[\sin^2(\tilde{m}\lambda/2)/\sin^2(\lambda/2)]$, there exist $D \in (0,1)$ and $\kappa > 0$ such that

$$
(a) \quad \int_{-D\pi}^{D\pi} W_m(\lambda) \lambda^2 d\lambda \geq \kappa \tilde{m}^{-1}, \quad (b) \quad \sup_{\lambda \in [-\pi,\pi]} |W_m(\lambda)| \lambda^2 d\lambda \leq C\tilde{m}^{-1}.
$$

**Proof** We can find a constant $c \in (0,1)$ such that, for $\lambda \in [-\pi, \pi],

$$
c(\lambda/2)^2 \leq \sin^2(\lambda/2) \leq (\lambda/2)^2. \quad (56)
$$

Therefore, there exists $\kappa > 0$ such that

$$
\int_{-D\pi}^{D\pi} W_m(\lambda) \lambda^2 d\lambda \geq \kappa \tilde{m}^{-1} \int_{-D\pi}^{D\pi} \sin^2(\tilde{m}\lambda/2)d\lambda \\
= 2C\tilde{m}^{-2} \int_{-\pi/2}^{\pi/2} \sin^2(\theta)d\theta \geq 2C\tilde{m}^{-2}[\tilde{m}D] \int_{-\pi/2}^{\pi/2} \sin^2(\theta)d\theta \\
\sim 2C\tilde{m}^{-1} \int_{-\pi/2}^{\pi/2} \sin^2(\theta)d\theta \geq \kappa \tilde{m}^{-1},
$$

giving part (a). Part (b) follows from (56) and $|\sin x| \leq 1$. ■

**References**


Tuypens, B. (2002). Examining the statistical properties for financial ratios. Mimeo, Yale University.


Table 1: Regression t-statistic: finite sample size (AR(1))

<table>
<thead>
<tr>
<th>c</th>
<th>ρ₁</th>
<th>σ₁²</th>
<th>0.25</th>
<th>0.50</th>
<th>0.75</th>
<th>0.95</th>
<th>σ₁²</th>
<th>0.25</th>
<th>0.50</th>
<th>0.75</th>
<th>0.95</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Demeaned Case (n = 100)</td>
<td></td>
<td>Detrended Case (n = 100)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1.000</td>
<td>0.054</td>
<td>0.062</td>
<td>0.109</td>
<td>0.204</td>
<td>0.267</td>
<td>0.052</td>
<td>0.087</td>
<td>0.177</td>
<td>0.351</td>
<td>0.549</td>
</tr>
<tr>
<td>-1</td>
<td>0.990</td>
<td>0.057</td>
<td>0.071</td>
<td>0.104</td>
<td>0.161</td>
<td>0.220</td>
<td>0.059</td>
<td>0.087</td>
<td>0.159</td>
<td>0.300</td>
<td>0.463</td>
</tr>
<tr>
<td>-5</td>
<td>0.950</td>
<td>0.056</td>
<td>0.057</td>
<td>0.068</td>
<td>0.087</td>
<td>0.114</td>
<td>0.047</td>
<td>0.077</td>
<td>0.116</td>
<td>0.166</td>
<td>0.237</td>
</tr>
<tr>
<td>-10</td>
<td>0.900</td>
<td>0.054</td>
<td>0.052</td>
<td>0.066</td>
<td>0.078</td>
<td>0.080</td>
<td>0.057</td>
<td>0.058</td>
<td>0.083</td>
<td>0.117</td>
<td>0.162</td>
</tr>
<tr>
<td>-20</td>
<td>0.800</td>
<td>0.060</td>
<td>0.060</td>
<td>0.043</td>
<td>0.058</td>
<td>0.065</td>
<td>0.053</td>
<td>0.060</td>
<td>0.064</td>
<td>0.082</td>
<td>0.106</td>
</tr>
</tbody>
</table>

The table shows rejection rates under the null hypothesis for a nominal 5% test using $t_\beta$. $y_t$ is given by (20) and $x_t$ by (17) with $\rho_1$ given by (19), with local-to-unity parameter $c$. Details are given in the text.

Table 2: Covariance-based t-statistic: finite sample size (AR(1))

<table>
<thead>
<tr>
<th>c</th>
<th>ρ₁</th>
<th>σ₁²</th>
<th>0.25</th>
<th>0.50</th>
<th>0.75</th>
<th>0.95</th>
<th>σ₁²</th>
<th>0.25</th>
<th>0.50</th>
<th>0.75</th>
<th>0.95</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Demeaned Case (n = 100)</td>
<td></td>
<td>Detrended Case (n = 100)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1.000</td>
<td>0.034</td>
<td>0.033</td>
<td>0.042</td>
<td>0.052</td>
<td>0.067</td>
<td>0.037</td>
<td>0.037</td>
<td>0.055</td>
<td>0.067</td>
<td>0.087</td>
</tr>
<tr>
<td>-1</td>
<td>0.990</td>
<td>0.032</td>
<td>0.027</td>
<td>0.033</td>
<td>0.037</td>
<td>0.042</td>
<td>0.036</td>
<td>0.036</td>
<td>0.049</td>
<td>0.052</td>
<td>0.079</td>
</tr>
<tr>
<td>-5</td>
<td>0.950</td>
<td>0.033</td>
<td>0.032</td>
<td>0.037</td>
<td>0.034</td>
<td>0.040</td>
<td>0.032</td>
<td>0.041</td>
<td>0.040</td>
<td>0.050</td>
<td>0.061</td>
</tr>
<tr>
<td>-10</td>
<td>0.900</td>
<td>0.036</td>
<td>0.030</td>
<td>0.040</td>
<td>0.036</td>
<td>0.037</td>
<td>0.030</td>
<td>0.033</td>
<td>0.041</td>
<td>0.042</td>
<td>0.051</td>
</tr>
<tr>
<td>-20</td>
<td>0.800</td>
<td>0.028</td>
<td>0.026</td>
<td>0.032</td>
<td>0.036</td>
<td>0.033</td>
<td>0.026</td>
<td>0.028</td>
<td>0.033</td>
<td>0.033</td>
<td>0.043</td>
</tr>
</tbody>
</table>

The table shows rejection rates under the null hypothesis for a nominal 5% test using $t_\lambda$. $y_t$ is given by (20) and $x_t$ by (17) with $\rho_1$ given by (19), with local-to-unity parameter $c$. Details are given in the text.

36
Table 3: Covariance-based t-statistic: finite sample size (long-horizon returns $n = 100$)

<table>
<thead>
<tr>
<th>$c$</th>
<th>$\rho_1$</th>
<th>$\sigma_{12} = 0$</th>
<th>0.25</th>
<th>0.50</th>
<th>0.75</th>
<th>0.95</th>
<th>$\sigma_{12} = 0$</th>
<th>0.25</th>
<th>0.50</th>
<th>0.75</th>
<th>0.95</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Demeaned Case ($k = 3$)</td>
<td>Detrended Case ($k = 3$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.044 0.045 0.054 0.051 0.059</td>
<td>0.051 0.061 0.060 0.076 0.076</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.051 0.044 0.052 0.055 0.067</td>
<td>0.048 0.057 0.061 0.069 0.075</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.041 0.047 0.037 0.054 0.053</td>
<td>0.041 0.060 0.057 0.050 0.061</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.040 0.046 0.042 0.042 0.037</td>
<td>0.046 0.045 0.046 0.045 0.044</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.022 0.028 0.030 0.032 0.041</td>
<td>0.034 0.034 0.031 0.037 0.035</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Demeaned Case ($k = 5$)</td>
<td>Detrended Case ($k = 5$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.044 0.058 0.061 0.065 0.082</td>
<td>0.046 0.067 0.068 0.093 0.108</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.045 0.056 0.047 0.064 0.069</td>
<td>0.049 0.056 0.062 0.086 0.100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.041 0.048 0.051 0.042 0.055</td>
<td>0.044 0.056 0.054 0.061 0.058</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.036 0.034 0.026 0.049 0.043</td>
<td>0.037 0.036 0.045 0.048 0.057</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.016 0.021 0.023 0.028 0.024</td>
<td>0.029 0.025 0.030 0.038 0.031</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The table shows rejection rates under the null hypothesis for a nominal 5% test using $t_\lambda$. The long-horizon return $y_{t,k}$ and $x_{t,k}$ are given by (26) where $x_t$ and $y_t$ follow (17) and (20) respectively with $\rho_1$ given by (19), with local-to-unity parameter $c$. $k = 3$ is chosen to match the ratio of the sample size to the longest horizon in the empirical application for a simulation sample size of $n = 100$. Details are given in the text.

Table 4: Covariance-based t-statistic: finite sample size (AR(2))

<table>
<thead>
<tr>
<th>$c$</th>
<th>$\rho_1 + \rho_2$</th>
<th>$\sigma_{12} = 0$</th>
<th>0.25</th>
<th>0.50</th>
<th>0.75</th>
<th>0.95</th>
<th>$\sigma_{12} = 0$</th>
<th>0.25</th>
<th>0.50</th>
<th>0.75</th>
<th>0.95</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Demeaned Case ($n = 100$)</td>
<td>Detrended Case ($n = 100$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.046 0.050 0.054 0.048 0.061</td>
<td>0.051 0.061 0.072 0.073 0.092</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.044 0.060 0.067 0.060 0.066</td>
<td>0.061 0.046 0.062 0.060 0.089</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.053 0.048 0.060 0.060 0.056</td>
<td>0.042 0.061 0.054 0.056 0.057</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.051 0.056 0.058 0.056 0.043</td>
<td>0.060 0.057 0.057 0.074 0.045</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.047 0.060 0.051 0.060 0.052</td>
<td>0.053 0.058 0.067 0.064 0.065</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Demeaned Case ($n = 400$)</td>
<td>Detrended Case ($n = 400$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.064 0.058 0.059 0.059 0.061</td>
<td>0.050 0.058 0.057 0.065 0.068</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.056 0.053 0.056 0.060 0.070</td>
<td>0.060 0.061 0.059 0.063 0.064</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.067 0.060 0.052 0.055 0.055</td>
<td>0.058 0.060 0.062 0.064 0.059</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.062 0.056 0.052 0.062 0.060</td>
<td>0.054 0.057 0.065 0.052 0.060</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.052 0.063 0.064 0.058 0.056</td>
<td>0.074 0.058 0.054 0.060 0.065</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The table shows rejection rates under the null hypothesis for a nominal 5% test using $t_\lambda$. $y_t$ is given by (20) and $x_1$ by (18) with $\rho_1$ and $\rho_2$ given by (27), with local-to-unity parameter $c$. Details are given in the text.
Table 5: Covariance-based t-statistic: finite sample power \((y_t = \beta x_{t-1} + \varepsilon_{1,t})\)

<table>
<thead>
<tr>
<th>(c)</th>
<th>(\sigma_{12})</th>
<th>(\beta = 0.15)</th>
<th>(0.20)</th>
<th>(0.35)</th>
<th>(0.50)</th>
<th>(0.75)</th>
<th>(1.00)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.000</td>
<td>0.115</td>
<td>0.226</td>
<td>0.745</td>
<td>0.975</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>(\rho_1 = 1.000)</td>
<td>0.500</td>
<td>0.105</td>
<td>0.202</td>
<td>0.684</td>
<td>0.961</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>0.950</td>
<td>0.071</td>
<td>0.153</td>
<td>0.656</td>
<td>0.968</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>-1</td>
<td>0.000</td>
<td>0.210</td>
<td>0.349</td>
<td>0.636</td>
<td>0.781</td>
<td>0.805</td>
<td>0.831</td>
</tr>
<tr>
<td>(\rho_1 = 0.990)</td>
<td>0.500</td>
<td>0.135</td>
<td>0.268</td>
<td>0.564</td>
<td>0.671</td>
<td>0.759</td>
<td>0.760</td>
</tr>
<tr>
<td></td>
<td>0.950</td>
<td>0.060</td>
<td>0.118</td>
<td>0.341</td>
<td>0.536</td>
<td>0.678</td>
<td>0.730</td>
</tr>
<tr>
<td>-2.5</td>
<td>0.000</td>
<td>0.207</td>
<td>0.339</td>
<td>0.667</td>
<td>0.790</td>
<td>0.843</td>
<td>0.854</td>
</tr>
<tr>
<td>(\rho_1 = 0.975)</td>
<td>0.500</td>
<td>0.150</td>
<td>0.252</td>
<td>0.586</td>
<td>0.714</td>
<td>0.792</td>
<td>0.817</td>
</tr>
<tr>
<td></td>
<td>0.950</td>
<td>0.074</td>
<td>0.124</td>
<td>0.383</td>
<td>0.573</td>
<td>0.739</td>
<td>0.780</td>
</tr>
<tr>
<td>-7.5</td>
<td>0.000</td>
<td>0.226</td>
<td>0.349</td>
<td>0.750</td>
<td>0.910</td>
<td>0.936</td>
<td>0.948</td>
</tr>
<tr>
<td>(\rho_1 = 0.925)</td>
<td>0.500</td>
<td>0.171</td>
<td>0.267</td>
<td>0.671</td>
<td>0.837</td>
<td>0.914</td>
<td>0.915</td>
</tr>
<tr>
<td></td>
<td>0.950</td>
<td>0.077</td>
<td>0.144</td>
<td>0.439</td>
<td>0.661</td>
<td>0.850</td>
<td>0.887</td>
</tr>
<tr>
<td>-20</td>
<td>0.000</td>
<td>0.220</td>
<td>0.365</td>
<td>0.794</td>
<td>0.962</td>
<td>0.997</td>
<td>0.999</td>
</tr>
<tr>
<td>(\rho_1 = 0.800)</td>
<td>0.500</td>
<td>0.138</td>
<td>0.266</td>
<td>0.690</td>
<td>0.922</td>
<td>0.988</td>
<td>0.993</td>
</tr>
<tr>
<td></td>
<td>0.950</td>
<td>0.049</td>
<td>0.122</td>
<td>0.477</td>
<td>0.809</td>
<td>0.968</td>
<td>0.983</td>
</tr>
</tbody>
</table>

B. Detrended Case \((n = 400)\)

<table>
<thead>
<tr>
<th>(c)</th>
<th>(\sigma_{12})</th>
<th>(\beta = 0.15)</th>
<th>(0.20)</th>
<th>(0.35)</th>
<th>(0.50)</th>
<th>(0.75)</th>
<th>(1.00)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.000</td>
<td>0.594</td>
<td>0.881</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>(\rho_1 = 1.000)</td>
<td>0.500</td>
<td>0.559</td>
<td>0.863</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>0.950</td>
<td>0.539</td>
<td>0.863</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>-1</td>
<td>0.000</td>
<td>0.747</td>
<td>0.824</td>
<td>0.850</td>
<td>0.857</td>
<td>0.853</td>
<td>0.837</td>
</tr>
<tr>
<td>(\rho_1 = 0.998)</td>
<td>0.500</td>
<td>0.730</td>
<td>0.800</td>
<td>0.846</td>
<td>0.831</td>
<td>0.849</td>
<td>0.836</td>
</tr>
<tr>
<td></td>
<td>0.950</td>
<td>0.700</td>
<td>0.775</td>
<td>0.816</td>
<td>0.840</td>
<td>0.842</td>
<td>0.831</td>
</tr>
<tr>
<td>-2.5</td>
<td>0.000</td>
<td>0.774</td>
<td>0.867</td>
<td>0.896</td>
<td>0.884</td>
<td>0.895</td>
<td>0.887</td>
</tr>
<tr>
<td>(\rho_1 = 0.994)</td>
<td>0.500</td>
<td>0.754</td>
<td>0.828</td>
<td>0.864</td>
<td>0.879</td>
<td>0.882</td>
<td>0.870</td>
</tr>
<tr>
<td></td>
<td>0.950</td>
<td>0.735</td>
<td>0.811</td>
<td>0.831</td>
<td>0.860</td>
<td>0.879</td>
<td>0.871</td>
</tr>
<tr>
<td>-7.5</td>
<td>0.000</td>
<td>0.841</td>
<td>0.949</td>
<td>0.968</td>
<td>0.965</td>
<td>0.970</td>
<td>0.976</td>
</tr>
<tr>
<td>(\rho_1 = 0.981)</td>
<td>0.500</td>
<td>0.843</td>
<td>0.939</td>
<td>0.964</td>
<td>0.963</td>
<td>0.971</td>
<td>0.976</td>
</tr>
<tr>
<td></td>
<td>0.950</td>
<td>0.822</td>
<td>0.905</td>
<td>0.948</td>
<td>0.959</td>
<td>0.965</td>
<td>0.962</td>
</tr>
<tr>
<td>-20</td>
<td>0.000</td>
<td>0.831</td>
<td>0.966</td>
<td>0.998</td>
<td>0.999</td>
<td>0.999</td>
<td>0.998</td>
</tr>
<tr>
<td>(\rho_1 = 0.950)</td>
<td>0.500</td>
<td>0.859</td>
<td>0.976</td>
<td>0.996</td>
<td>0.997</td>
<td>0.998</td>
<td>0.997</td>
</tr>
<tr>
<td></td>
<td>0.950</td>
<td>0.864</td>
<td>0.959</td>
<td>0.992</td>
<td>0.991</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

The table shows rejection rates under the alternative hypothesis for a nominal 5% test using \(t_\lambda\). \(y_t\) is given by (23) and \(x_t\) by (17) with \(\rho_1\) given by (19), with local-to-unity parameter \(c\). Details are given in the text.
Table 6: Covariance-based t-statistic: finite sample power ($y_t = \beta (1 - \rho) x_{t-1} + \varepsilon_{1,t}$)

| $c$ | $\sigma_{12}$ | $\beta = 0.15$ | $0.20$ | $0.35$ | $0.50$ | $0.75$ | $1.00$ | $r^2 = 0.02$ | $0.04$ | $0.11$ | $0.20$ | $0.36$ | $0.50$ |
|-----|--------------|----------------|--------|--------|--------|--------|--------|-----|--------|--------|--------|--------|--------|--------|
| ($\rho_1$) | $A$. Detrended Case ($n = 100$) | | | | | | | | | | | | | |
| $c = 0$ | 0.000 | 0.161 | 0.275 | 0.689 | 0.923 | 0.999 | 1.000 | | | | | | | |
| ($\rho_1 = 1.000$) | 0.500 | 0.127 | 0.225 | 0.671 | 0.923 | 0.999 | 1.000 | | | | | | | |
| $c = -1$ | 0.000 | 0.173 | 0.263 | 0.696 | 0.931 | 0.999 | 1.000 | | | | | | | |
| ($\rho_1 = 0.990$) | 0.500 | 0.136 | 0.248 | 0.649 | 0.926 | 0.999 | 1.000 | | | | | | | |
| $c = -2.5$ | 0.000 | 0.164 | 0.271 | 0.703 | 0.936 | 0.999 | 1.000 | | | | | | | |
| ($\rho_1 = 0.975$) | 0.500 | 0.139 | 0.262 | 0.666 | 0.928 | 1.000 | 1.000 | | | | | | | |
| $c = -7.5$ | 0.000 | 0.166 | 0.305 | 0.722 | 0.944 | 1.000 | 1.000 | | | | | | | |
| ($\rho_1 = 0.925$) | 0.500 | 0.158 | 0.260 | 0.673 | 0.927 | 0.999 | 1.000 | | | | | | | |
| $c = -20$ | 0.000 | 0.159 | 0.242 | 0.691 | 0.939 | 1.000 | 1.000 | | | | | | | |
| ($\rho_1 = 0.800$) | 0.500 | 0.106 | 0.197 | 0.608 | 0.922 | 0.998 | 1.000 | | | | | | | |
| $c = -20$ | 0.000 | 0.159 | 0.242 | 0.691 | 0.939 | 1.000 | 1.000 | | | | | | | |
| ($\rho_1 = 0.800$) | 0.500 | 0.106 | 0.197 | 0.608 | 0.922 | 0.998 | 1.000 | | | | | | | |
| $c = -20$ | 0.000 | 0.159 | 0.242 | 0.691 | 0.939 | 1.000 | 1.000 | | | | | | | |
| ($\rho_1 = 0.800$) | 0.500 | 0.106 | 0.197 | 0.608 | 0.922 | 0.998 | 1.000 | | | | | | | |
| $c = -20$ | 0.000 | 0.159 | 0.242 | 0.691 | 0.939 | 1.000 | 1.000 | | | | | | | |
| ($\rho_1 = 0.800$) | 0.500 | 0.106 | 0.197 | 0.608 | 0.922 | 0.998 | 1.000 | | | | | | | |

$B$. Detrended Case ($n = 400$)

| $c$ | $\sigma_{12}$ | $\beta = 0.15$ | $0.20$ | $0.35$ | $0.50$ | $0.75$ | $1.00$ | $r^2 = 0.02$ | $0.04$ | $0.11$ | $0.20$ | $0.36$ | $0.50$ |
|-----|--------------|----------------|--------|--------|--------|--------|--------|-----|--------|--------|--------|--------|--------|--------|
| ($\rho_1$) | | | | | | | | | | | | | | |
| $c = 0$ | 0.000 | 0.584 | 0.834 | 1.000 | 1.000 | 1.000 | 1.000 | | | | | | | |
| ($\rho_1 = 1.000$) | 0.500 | 0.569 | 0.830 | 1.000 | 1.000 | 1.000 | 1.000 | | | | | | | |
| $c = -1$ | 0.000 | 0.602 | 0.824 | 0.998 | 1.000 | 1.000 | 1.000 | | | | | | | |
| ($\rho_1 = 0.998$) | 0.500 | 0.608 | 0.852 | 1.000 | 1.000 | 1.000 | 1.000 | | | | | | | |
| $c = -2.5$ | 0.000 | 0.609 | 0.842 | 1.000 | 1.000 | 1.000 | 1.000 | | | | | | | |
| ($\rho_1 = 0.994$) | 0.500 | 0.592 | 0.825 | 0.999 | 1.000 | 1.000 | 1.000 | | | | | | | |
| $c = -7.5$ | 0.000 | 0.583 | 0.832 | 0.999 | 1.000 | 1.000 | 1.000 | | | | | | | |
| ($\rho_1 = 0.981$) | 0.500 | 0.597 | 0.849 | 0.999 | 1.000 | 1.000 | 1.000 | | | | | | | |
| $c = -20$ | 0.000 | 0.582 | 0.836 | 0.998 | 1.000 | 1.000 | 1.000 | | | | | | | |
| ($\rho_1 = 0.950$) | 0.500 | 0.549 | 0.815 | 0.999 | 1.000 | 1.000 | 1.000 | | | | | | | |

The table shows rejection rates under the alternative hypothesis for a nominal 5% test using $t_\lambda$. $y_t$ is given by (24) and $x_t$ by (17) with $\rho_1$ given by (19), with local-to-unity parameter $c$. Details are given in the text.
Table 7: CES Bonferroni method: finite sample power ($y_t = \beta x_{t-1} + \epsilon_{1,t}$)

<table>
<thead>
<tr>
<th>$c$</th>
<th>$\sigma_{12}$</th>
<th>$\beta = 0.15$</th>
<th>0.20</th>
<th>0.35</th>
<th>0.50</th>
<th>0.75</th>
<th>1.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Demeaned Case ($n = 100$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$c = 0$</td>
<td>0.000</td>
<td>0.327</td>
<td>0.503</td>
<td>0.938</td>
<td>0.999</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>($\rho_1 = 1.000$)</td>
<td>0.500</td>
<td>0.387</td>
<td>0.600</td>
<td>0.945</td>
<td>0.999</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>$c = -1$</td>
<td>0.000</td>
<td>0.955</td>
<td>0.989</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>($\rho_1 = 0.990$)</td>
<td>0.500</td>
<td>0.913</td>
<td>0.967</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>$c = -2.5$</td>
<td>0.000</td>
<td>0.929</td>
<td>0.988</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>($\rho_1 = 0.975$)</td>
<td>0.500</td>
<td>0.883</td>
<td>0.974</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>$c = -7.5$</td>
<td>0.000</td>
<td>0.863</td>
<td>0.966</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>($\rho_1 = 0.925$)</td>
<td>0.500</td>
<td>0.814</td>
<td>0.936</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>$c = -20$</td>
<td>0.000</td>
<td>0.628</td>
<td>0.853</td>
<td>0.999</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>($\rho_1 = 0.800$)</td>
<td>0.500</td>
<td>0.636</td>
<td>0.816</td>
<td>0.994</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>B. Demeaned Case ($n = 400$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$c = 0$</td>
<td>0.000</td>
<td>0.840</td>
<td>0.984</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>($\rho_1 = 1.000$)</td>
<td>0.500</td>
<td>0.882</td>
<td>0.988</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>$c = -1$</td>
<td>0.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>($\rho_1 = 0.998$)</td>
<td>0.500</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>$c = -2.5$</td>
<td>0.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>($\rho_1 = 0.994$)</td>
<td>0.500</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>$c = -7.5$</td>
<td>0.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>($\rho_1 = 0.981$)</td>
<td>0.500</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>$c = -20$</td>
<td>0.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>($\rho_1 = 0.950$)</td>
<td>0.500</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

The table shows rejection rates under the alternative hypothesis for a nominal 5% test using the CES Bonferroni test. $y_t$ is given by (23) and $x_t$ by (17) with $\rho_1$ given by (19), with local-to-unity parameter $c$. Details are given in the text.
Table 8: CES Bonferroni method: finite sample power \((y_t = \beta(1 - \rho L)x_{t-1} + \varepsilon_{1,t})\)

<table>
<thead>
<tr>
<th>(c)</th>
<th>(\sigma_{12})</th>
<th>(\beta = 0.15)</th>
<th>(0.20)</th>
<th>(0.35)</th>
<th>(0.50)</th>
<th>(0.75)</th>
<th>(1.00)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(r^2 = 0.02)</td>
<td>(0.04)</td>
<td>(0.11)</td>
<td>(0.20)</td>
<td>(0.36)</td>
<td>(0.50)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. Demeaned Case ((n = 100))</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(c = 0)</td>
<td>0.000</td>
<td>0.075</td>
<td>0.087</td>
<td>0.138</td>
<td>0.213</td>
<td>0.291</td>
<td>0.369</td>
</tr>
<tr>
<td>((\rho_1 = 1.000))</td>
<td>0.500</td>
<td>0.039</td>
<td>0.042</td>
<td>0.067</td>
<td>0.098</td>
<td>0.137</td>
<td>0.201</td>
</tr>
<tr>
<td>(c = -1)</td>
<td>0.950</td>
<td>0.034</td>
<td>0.041</td>
<td>0.035</td>
<td>0.051</td>
<td>0.071</td>
<td>0.099</td>
</tr>
<tr>
<td>((\rho_1 = 0.990))</td>
<td>0.500</td>
<td>0.085</td>
<td>0.090</td>
<td>0.178</td>
<td>0.270</td>
<td>0.386</td>
<td>0.519</td>
</tr>
<tr>
<td>(c = -2.5)</td>
<td>0.950</td>
<td>0.047</td>
<td>0.059</td>
<td>0.081</td>
<td>0.137</td>
<td>0.238</td>
<td>0.343</td>
</tr>
<tr>
<td>((\rho_1 = 0.975))</td>
<td>0.000</td>
<td>0.078</td>
<td>0.137</td>
<td>0.214</td>
<td>0.334</td>
<td>0.537</td>
<td>0.696</td>
</tr>
<tr>
<td>(c = -7.5)</td>
<td>0.950</td>
<td>0.032</td>
<td>0.025</td>
<td>0.043</td>
<td>0.062</td>
<td>0.119</td>
<td>0.180</td>
</tr>
<tr>
<td>((\rho_1 = 0.925))</td>
<td>0.500</td>
<td>0.071</td>
<td>0.072</td>
<td>0.149</td>
<td>0.227</td>
<td>0.376</td>
<td>0.550</td>
</tr>
<tr>
<td>(c = -20)</td>
<td>0.950</td>
<td>0.038</td>
<td>0.044</td>
<td>0.074</td>
<td>0.149</td>
<td>0.249</td>
<td>0.362</td>
</tr>
<tr>
<td>((\rho_1 = 0.800))</td>
<td>0.500</td>
<td>0.101</td>
<td>0.156</td>
<td>0.343</td>
<td>0.544</td>
<td>0.800</td>
<td>0.934</td>
</tr>
<tr>
<td>(c = -20)</td>
<td>0.950</td>
<td>0.146</td>
<td>0.247</td>
<td>0.580</td>
<td>0.837</td>
<td>0.985</td>
<td>0.999</td>
</tr>
<tr>
<td>((\rho_1 = 0.800))</td>
<td>0.500</td>
<td>0.150</td>
<td>0.257</td>
<td>0.543</td>
<td>0.839</td>
<td>0.987</td>
<td>1.000</td>
</tr>
<tr>
<td>(c = -7.5)</td>
<td>0.950</td>
<td>0.146</td>
<td>0.234</td>
<td>0.519</td>
<td>0.787</td>
<td>0.994</td>
<td>1.000</td>
</tr>
<tr>
<td>B. Demeaned Case ((n = 400))</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(c = 0)</td>
<td>0.000</td>
<td>0.074</td>
<td>0.088</td>
<td>0.150</td>
<td>0.197</td>
<td>0.307</td>
<td>0.387</td>
</tr>
<tr>
<td>((\rho_1 = 1.00))</td>
<td>0.500</td>
<td>0.038</td>
<td>0.038</td>
<td>0.066</td>
<td>0.089</td>
<td>0.141</td>
<td>0.218</td>
</tr>
<tr>
<td>(c = -1)</td>
<td>0.950</td>
<td>0.034</td>
<td>0.038</td>
<td>0.030</td>
<td>0.044</td>
<td>0.060</td>
<td>0.098</td>
</tr>
<tr>
<td>((\rho_1 = 0.998))</td>
<td>0.500</td>
<td>0.073</td>
<td>0.096</td>
<td>0.184</td>
<td>0.299</td>
<td>0.399</td>
<td>0.534</td>
</tr>
<tr>
<td>(c = -2.5)</td>
<td>0.950</td>
<td>0.049</td>
<td>0.050</td>
<td>0.084</td>
<td>0.137</td>
<td>0.247</td>
<td>0.345</td>
</tr>
<tr>
<td>((\rho_1 = 0.994))</td>
<td>0.500</td>
<td>0.039</td>
<td>0.026</td>
<td>0.040</td>
<td>0.073</td>
<td>0.133</td>
<td>0.193</td>
</tr>
<tr>
<td>(c = -7.5)</td>
<td>0.950</td>
<td>0.058</td>
<td>0.073</td>
<td>0.141</td>
<td>0.232</td>
<td>0.420</td>
<td>0.568</td>
</tr>
<tr>
<td>((\rho_1 = 0.981))</td>
<td>0.500</td>
<td>0.040</td>
<td>0.048</td>
<td>0.098</td>
<td>0.145</td>
<td>0.257</td>
<td>0.393</td>
</tr>
<tr>
<td>(c = -20)</td>
<td>0.950</td>
<td>0.058</td>
<td>0.162</td>
<td>0.381</td>
<td>0.617</td>
<td>0.858</td>
<td>0.964</td>
</tr>
<tr>
<td>((\rho_1 = 0.950))</td>
<td>0.500</td>
<td>0.088</td>
<td>0.119</td>
<td>0.294</td>
<td>0.516</td>
<td>0.784</td>
<td>0.927</td>
</tr>
<tr>
<td>(c = -20)</td>
<td>0.950</td>
<td>0.097</td>
<td>0.103</td>
<td>0.224</td>
<td>0.388</td>
<td>0.688</td>
<td>0.915</td>
</tr>
<tr>
<td>((\rho_1 = 0.950))</td>
<td>0.500</td>
<td>0.150</td>
<td>0.253</td>
<td>0.583</td>
<td>0.863</td>
<td>0.992</td>
<td>1.000</td>
</tr>
<tr>
<td>(c = -20)</td>
<td>0.950</td>
<td>0.162</td>
<td>0.221</td>
<td>0.511</td>
<td>0.819</td>
<td>0.992</td>
<td>1.000</td>
</tr>
</tbody>
</table>

The table shows rejection rates under the alternative hypothesis for a nominal 5% test using the CES Bonferroni test. \(y_t\) is given by (24) and \(x_t\) by (17) with \(\rho_1\) given by (19), with local-to-unity parameter \(c\). Details are given in the text.
Table 9: Regressions of k-period long-horizon real stock returns on the treasury bill and dividend price ratio

<table>
<thead>
<tr>
<th></th>
<th>Treasury Bills</th>
<th>Dividend Price Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Forecast Horizon (k)</td>
<td></td>
</tr>
<tr>
<td>sample</td>
<td>k = 1.0 3.0 12.0 24.0</td>
<td>k = 1.0 3.0 12.0 24.0</td>
</tr>
<tr>
<td>1927-β</td>
<td>-0.702 -1.906 -3.014 -1.388</td>
<td>0.016 0.043 0.200 0.386</td>
</tr>
<tr>
<td>1994 R²</td>
<td>0.001 0.002 0.002 0.000</td>
<td>0.007 0.014 0.073 0.143</td>
</tr>
<tr>
<td>1994 tβ</td>
<td>-0.962 -0.964 -0.402 -0.136</td>
<td>2.389 1.598 2.658 4.221</td>
</tr>
<tr>
<td>1927-β</td>
<td>-1.224 -6.179 -27.898 -106.2</td>
<td>0.024 0.054 0.304 0.667</td>
</tr>
<tr>
<td>1951 R²</td>
<td>0.000 0.002 0.011 0.089</td>
<td>0.007 0.011 0.086 0.217</td>
</tr>
<tr>
<td>1951 tβ</td>
<td>-0.300 -0.528 -0.698 -1.277</td>
<td>1.472 0.886 2.134 3.796</td>
</tr>
<tr>
<td>1952-β</td>
<td>-1.343 -3.497 -6.114 -0.993</td>
<td>0.027 0.080 0.327 0.579</td>
</tr>
<tr>
<td>1994 R²</td>
<td>0.006 0.013 0.009 0.000</td>
<td>0.018 0.049 0.188 0.322</td>
</tr>
<tr>
<td>1994 tβ</td>
<td>-1.785 -1.669 -0.726 -0.082</td>
<td>3.098 3.728 3.845 3.589</td>
</tr>
</tbody>
</table>

Entries show results from a regression \( y_{t+k} = r_{t+1} + \ldots + r_{t+k} \) on \( x_t = i_t \) or \( x_t = d_t - p_t \). Regressions are estimated by OLS with HAC standard errors for \( k > 1 \), using the Bartlett (Newey-West) kernel with bandwidth set to \( k - 1 \).

Table 10: Confidence intervals on largest roots and residual correlation

<table>
<thead>
<tr>
<th></th>
<th>Treasury Bills</th>
<th>Dividend Price Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(( y_t = r_t ) ( x_t = i_t ))</td>
<td>(( y_t = r_t ) ( x_t = d_t - p_t ))</td>
</tr>
<tr>
<td>sample</td>
<td>95 % CI on</td>
<td>95 % CI on</td>
</tr>
<tr>
<td>period</td>
<td>largest root in ( x_t )</td>
<td>largest root in ( x_t )</td>
</tr>
<tr>
<td>1927 to 1994</td>
<td>(0.9836 1.004) -0.0768</td>
<td>(0.9623 0.998) -0.9615</td>
</tr>
<tr>
<td>1927 to 1951</td>
<td>(0.9421 1.010) 0.1169</td>
<td>(0.9273 1.007) -0.9601</td>
</tr>
<tr>
<td>1952 to 1994</td>
<td>(0.9680 1.006) -0.3036</td>
<td>(0.9543 1.003) -0.9754</td>
</tr>
</tbody>
</table>

Confidence intervals on the largest root are based on Stock (1991) using the Ng and Perron (2001) MIC criteria to select lag-length with a maximum of six lags.
Table 11: Covariance-based orthogonality tests on k-period long-horizon real stock returns using the treasury bill and dividend price ratio

<table>
<thead>
<tr>
<th></th>
<th>Treasury Bills</th>
<th>Dividend Price Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Forecast Horizon (k)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>k = 1.0</td>
<td>3.0</td>
</tr>
<tr>
<td>1927- $t_\lambda$</td>
<td>-0.3934</td>
<td>-2.2120</td>
</tr>
<tr>
<td>1927- $t_\lambda$</td>
<td>0.5946</td>
<td>-0.6085</td>
</tr>
<tr>
<td>1951 $m^*$</td>
<td>0.7821</td>
<td>3.9311</td>
</tr>
<tr>
<td>1952- $t_\lambda$</td>
<td>-1.1613</td>
<td>-2.4188</td>
</tr>
<tr>
<td>1994 $m^*$</td>
<td>1.7534</td>
<td>5.7159</td>
</tr>
</tbody>
</table>

Standard normal critical values apply. $t_\lambda$ is the test statistic and $m^*$ is the optimal bandwidth. The estimation and bandwidth procedures are described in detail in the text.

Table 12: Regression of long horizon real stock returns on stochastically detrended one-month treasury bill rates

<table>
<thead>
<tr>
<th></th>
<th>Forecast Horizon (k)</th>
<th></th>
<th>Forecast Horizon (k)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>k = 1.0</td>
<td>3.0</td>
<td>12.0</td>
<td>24.0</td>
</tr>
<tr>
<td>$\hat{\beta}$</td>
<td>1927-</td>
<td>-5.468</td>
<td>-17.181</td>
<td>-41.663</td>
</tr>
<tr>
<td>$R^2$</td>
<td>1994</td>
<td>0.005</td>
<td>0.016</td>
<td>0.023</td>
</tr>
<tr>
<td>$t_\beta$</td>
<td>-2.119</td>
<td>-2.888</td>
<td>-1.840</td>
<td>-0.156</td>
</tr>
<tr>
<td>$\hat{\beta}$</td>
<td>1952-</td>
<td>-6.547</td>
<td>-18.621</td>
<td>-56.406</td>
</tr>
<tr>
<td>$R^2$</td>
<td>1994</td>
<td>0.019</td>
<td>0.047</td>
<td>0.103</td>
</tr>
<tr>
<td>$t_{\hat{\beta}}$</td>
<td>-3.182</td>
<td>-3.597</td>
<td>-3.055</td>
<td>-1.245</td>
</tr>
</tbody>
</table>

Entries show results from a regression $y_{t+k} = r_{t+1} + \ldots + r_{t+k}$ on $x_t = i_t - \sum_{j=0}^{11} i_{t-j}$. Regressions are estimated by OLS with HAC standard errors, using the Bartlett (Newey-West) kernel with bandwidth set to $k - 1$. 

43