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

April 18, 2004

Abstract

This paper develops a new covariance-based test of orthogonality that may be attractive when regressors have roots close or equal to unity. In this case standard regression-based orthogonality tests can suffer from (i) size distortions and (ii) uncertainty regarding the appropriate model in which to frame the alternative hypothesis. The new test has good size and power against a wide range of reasonable alternatives for stationary, non-stationary, and local to unity regressors, while avoiding non-standard limiting distributions, size correction, and unit root pre-tests. Asymptotic results are derived and 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 recent literature highlights the role of size distortions in traditional tests using these predictors. However, we argue that even size corrected regression tests still restrict attention to "balanced" alternatives that become less plausible the more persistent the regressor. Thus the net impact of persistent regressions on orthogonality tests remains uncertain. Covariance based tests maintain correct size without restricting the alternatives, allowing us to sort out the two effects. Using this test we find weaker evidence of predictability using dividend price ratios but stronger evidence using interest rates. This suggests a dominant effect of size distortion in the first case and restrictions on the alternative in the second.

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, Jonas Fisher, John Galbraith, John Maheu, Angelo Melino, John Rust, Sam Thompson, Tim Vogelsang, and participants at the CESG, MESG, and the Columbia, 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 when the conditioning variable has a root close or equal to unity. This new method provides a t-test with good size properties without reference to prior knowledge, estimates, or pre-tests regarding the size of the root. Furthermore it has nontrivial power against a broad range of alternatives. In this sense the alternative hypothesis is defined more broadly than in common regression specifications.

To fix ideas, consider the following simple orthogonality regression of y_t on lagged x_{t-1}

$$y_t = \beta_0 + \beta_1 x_{t-1} + \varepsilon_{1t}, \tag{1}$$

together with a first order autoregressive specification for the marginal distribution of x_t

$$x_t = \rho_0 + \rho_1 x_{t-1} + \varepsilon_{2t}, \tag{2}$$

with a value of ρ_1 close or possibly equal to unity and innovations given by

$$\varepsilon_t = \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \sim \text{i.i.d.} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \right). \tag{3}$$

This would appear to roughly characterize several common empirical applications, including, for example, orthogonality tests involving the regression of log returns on the lagged dividend yield or interest rate and the regression of excess foreign currency returns on the lagged forward premium.[1] Several other applications, including tests of the permanent income hypothesis, the expectations hypothesis of the term structure, and the constant real interest rate hypothesis may also be cast in the form of an orthogonality test.[2] Although the dependent variables, such as returns, in these regressions may show little persistence, the regressors are often highly serially correlated and may be well characterized by roots near unity.

Typically it is the case that a risk-neutral market efficiency condition implies orthogonality of y_t with respect to $I_{x,t-1} = \sigma(x_{t-1}, x_{t-2}, x_{t-3}, \ldots)$, the information contained in all past values of x_t. The null hypothesis of orthogonality implies $\beta_1 = 0$, and a common test of market efficiency is provided by a standard t-test on this parameter. This null hypothesis carries no implications regarding the root ρ_1 of x_t and often this parameter may not be of direct economic interest. However, if the value of ρ_1 is close or equal to one, it can become a difficult nuisance parameter. In particular, as discussed below, a large value of ρ_1 can impact a standard orthogonality t-test in two ways: (i) by causing size distortions and (ii) by leading to trivial power under certain reasonable alternatives. Roots in x_t equal to one may sometimes be ruled out on a priori grounds (nominal interest rates, for example, should not be negative), but roots close to one in a local to unity sense generally can not be.

The size problem is well documented in **?** (**?**), **?** (**?**), and **?** (**?**). It results from a Dickey-Fuller type bias in the nonstandard distribution of the estimator $\widehat{\beta}_1$ when

[1] See **?** (**?**) for an application to the forward rate unbiasedness test.

[2] See **?** (**?**) and references therein.

$\sigma_{12} \neq 0$ and ρ_1 is close to one.[3] For a known value of $\rho_1 = 1$ this bias may be corrected using cointegration type adjustments in order to obtain correct asymptotic size. More realistically, for fixed values of $\rho_1 \leq 1$, two stage procedures based on a unit root pre-test can also provide correct large sample size. Unfortunately, such procedures have been found to overcorrect under a local to unity specification, again creating size distortions in the empirically relevant case of moderate sample sizes and roots just slightly below one (**?** (**?**), **?** (**?**)).[4] The covariance based t-test proposed here is fundamentally different than the regression t-test, and thus avoids this size distortion problem altogether.

The second issue that arises for ρ_1 close or equal to one is that of power under reasonable alternatives. This appears to be less widely discussed. Under the null hypothesis of orthogonality, we have $\beta_1 = 0$, allowing y_t (e.g. returns) to be stationary even if x_t (e.g. dividend yields) follows a local to unity or unit root process. The null hypothesis is thus reasonably incorporated into a regression such as (1). However, the exact form of the alternative is usually left unspecified by economic theory and just what constitutes a sensible alternative may depend on the persistence properties of the data. If x_t is clearly stationary, then $\beta_1 \neq 0$ in (1) provides a reasonable alternative hypothesis. However, if x_t has a near unit root, this will only provide an appropriate alternative if y_t is thought to contain an equally persistent component, a possibility that may often lack empirical support. Note, for example, that if x_t has a unit root then the model above allows just two possibilities: orthogonality ($\beta_1 = 0$) or cointegration ($\beta_1 \neq 0$). However, this seems unduly restrictive: orthogonality is not synonymous with a lack of cointegration. In particular, this omits a large class of "unbalanced" alternatives for which y_t is stationary, x_t has a near (or exact) unit root, and yet, despite this imbalance, the past history of x_t contains predictive content for y_t. For instance, if x_t had an exact unit root, but y_t was stationary, then a more reasonable test of orthogonality would be a regression of y_t on Δx_{t-1} as in

$$y_t = \gamma_0 + \gamma_1 (x_{t-1} - x_{t-2}) + \varepsilon_{1t}. \tag{4}$$

But this requires a unit root pre-test, whose problem was discussed above. Of course, more elaborate parametric regressions can also nest both alternatives, but are not commonly employed in this context and may come at the expense of complicating the size problem, particularly in the near unit root case (see **?** (**?**)). The new test we propose has reasonable power against both sets of alternatives: $\beta_1 \neq 0$ when x_t is stationary, and $\gamma_1 \neq 0$ when x_t is a unit root process, without requiring size correction or pre-test.

The covariance-based testing approach we develop begins with the following intuition. Consider first x_t stationary ($I(0)$). The regression coefficient $\beta_1 = \text{cov}(y_t, x_{t-1})/\text{var}(x_{t-1})$ is equal to 0 if and only if the numerator $\text{cov}(y_t, x_{t-1}) = 0$. For stationary x_t the orthogonality test may therefore be restated as a test of $\text{cov}(y_t, x_{t-1}) = 0$. Next,

[3] σ_{12} is unrestricted under the null hypothesis similarly to ρ_1 and must also be considered a nuisance parameter.

[4] **?** (**?**) provide more appropriate corrections based on first stage confidence intervals for the local to unity parameter.

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 Δx_{t-1} as

$$\operatorname{cov}(y_t, x_{t-1}) = \sum_{h=1}^{\infty} \operatorname{cov}(y_t, \Delta x_{t-h}). \tag{5}$$

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[5]

$$\operatorname{cov}(y_t, x_{t-1}) = \sum_{h=1}^{t-1} \operatorname{cov}(y_t, \Delta x_{t-h}),$$

initializing x_t at $t = 0$. Assume y_t and Δx_{t-1} are stationary, and define a *quasi-covariance* between x_{t-1} and y_t as

$$\lambda_{y,\Delta x} := \lim_{t\to\infty} \sum_{h=1}^{t-1} \operatorname{cov}(y_t, \Delta x_{t-h}) = \sum_{h=1}^{\infty} \operatorname{cov}(y_t, \Delta x_{t-h}), \tag{6}$$

which is well-defined if $\sum_{h=0}^{\infty} |\operatorname{cov}(y_t, \Delta x_{t-h})| < \infty$. As seen from (5), when x_t is stationary, the quasi-covariance is written as

$$\lambda_{y,\Delta x} = \operatorname{cov}(y_t, x_{t-1}). \tag{7}$$

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

When x_t has a root close to unity, a useful model of x_t is the so-called local-to-unity process

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

with $x_t \equiv 0$ for $t \leq 0$. As shown in the Appendix, when $\sum_{p=1}^{\infty} p|\operatorname{cov}(y_t, u_{t-p})| < \infty$, the quasi-covariance takes the form

$$\lambda_{y,\Delta x} = \sum_{h=1}^{\infty} \operatorname{cov}(y_t, u_{t-h}) + O\left(n^{-1}\right), \tag{8}$$

and it bears the same meaning as when x_t is $I(1)$.

[5]This concept, expanded upon here, was first given in **?** (**?**).

The proposed orthogonality test is then based on a test of the null hypothesis that $\lambda_{y,\Delta x} = 0$, a parameter which is well defined for both stationary and unit root nonstationary x_t. To see the relationship between this parameter restriction and more common tests of orthogonality, note first that y_t orthogonal to $I_{x,t-1}$ implies $\lambda_{y,\Delta x} = 0$. This follows from the fact that Δx_{t-h} belongs to $I_{x,t-1}$ for $h \geq 1$ and is therefore orthogonal to y_t, implying $\text{cov}(y_t, \Delta x_{t-h}) = 0$ for all $h \geq 1$. Thus a rejection of $\lambda_{y,\Delta x} = 0$ constitutes a valid rejection of orthogonality.

Next consider the power of the test against various alternatives. When x_t is stationary, $\beta_1 = \frac{\text{cov}(y_t, x_{t-1})}{\text{var}(x_t)}$ has a finite dominator, so that $\beta_1 = 0$ if and only if $\lambda_{y,\Delta x} = \text{cov}(y_t, x_{t-1}) = 0$. So for stationary x_t the test has power against the same alternatives as the standard t-test. For $\rho_1 = 1$ (unit root) and $\beta_1 \neq 0$ y_t and x_t are cointegrated unit roots, implying an infinite (and hence non-zero) value of $\lambda_{y,\Delta x}$. Therefore the test still maintains power against $\beta_1 \neq 0$ for nonstationary x_t. In addition, it also provides power against other reasonable alternatives (e.g. $\gamma \neq 0$ in (4) when $\rho_1 = 1$), which are not properly included in the alternative hypothesis when testing the restriction that $\beta_1 = 0$. For example, although no longer infinite, $\lambda_{y,\Delta x} \neq 0$ also holds for $\rho_1 = 1$ and $\gamma \neq 0$.

Estimation follows from the fact that 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. Thus we can provide a single estimator for $\lambda_{y,\Delta x}$ without the necessity of pretesting or estimating ρ_1. 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 is shown to have a unique limit distribution for all (finite) values of the local to unity parameter c. This allows us to avoid two-stage inference procedures, such as Bonferroni bounds, that are generally necessitated by the lack of a consistent estimator for the local to unity parameter. We provide an asymptotically exact test, based on a single t-statistic with a limiting standard normal distribution that is equally valid under both unit root and local to unity (finite c) assumptions. No bias corrections or other adjustments are required. This test is suggested primarily when roots are close to unity so that a local to unity model is appropriate. However, it also shown to provide conservative inference when x_t is stationary.

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. In fact, using size-corrected regression-based tests, original results suggesting strong predictive content have often been weakened and sometimes overturned. However, while properly correcting for size, such regressions may also restrict power to alternatives that imply near-nonstationarity in stock returns (i.e. near I(2) behavior in stock prices). By using covariance-based tests we not only correct size, but also allow for alternatives that leave returns stationary, while still violating orthogonality. Our tests agree with past literature in indicating substantially weaker evidence of return predictability using dividend-yields. However, the results are reversed for the short-term interest rate, in which we find even stronger evidence of predictability using our tests. Thus the size distortion associated with near unit roots appears to be the overriding factor in the dividend yield regression, whereas the restriction on the alternative matters more in the case of the interest rate. This difference in outcomes makes sense in the light of the far stronger residual correlation found using the dividend yield.

The size problem inherent in these regressions has generated an active area of research and a number of alternative techniques have been proposed. However, they all differ substantially in approach and most address primarily the issue of size. Beginning with **?** (**?**), several papers (**?** (**?**), **?** (**?**), and **?** (**?**)) employ local to unity asymptotics to provide size corrections for regression based tests. **?** (**?**) for example, provide critical values using two-stage Bonferroni and Scheffe type bounds procedures. **?** (**?**) and **?** (**?**) give finite sample corrections to regression based tests under more restrictive assumptions, while **?** (**?**) and **?** (**?**) consider Bayesian approaches. With suitable (strictly exogenous) instruments, the FM-IV estimator of **?** (**?**) can also eliminate size problems, even under local to unity assumptions and without prior testing on ρ_1. Finally, sign and rank tests (Campbell and Dufour (1995, 1997)) provide exact finite sample size without restrictions on x_t under the null hypothesis, though proper specification of the mean process for x_t still matters for power and white noise assumptions on y_t may complicate their use in tests with long-horizon returns. Bootstrap and subsampling approaches have also been employed under the assumption of a fixed root less than unity (**?** (**?**), **?** (**?**), and **?** (**?**)).

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 state the assumptions.

Assumption A

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

$$
\begin{aligned}
z_t &= \begin{pmatrix} y_t \\ \Delta x_t \end{pmatrix} = A(L)\,\varepsilon_t = \sum_{j=0}^{\infty} A_j \varepsilon_{t-j}, \quad \sum_{j=0}^{\infty} j||A_j|| < \infty, \qquad (9) \\
\varepsilon_t &\sim \text{i.i.d.}\,(0, I_2)\,, \quad \text{with finite fourth moment}, \\
\sum_{h=-\infty}^{\infty} |h|^q \,\|\Gamma(h)\| &< \infty, \quad q \geq 1; \quad \Gamma(h) = \begin{bmatrix} \Gamma_{yy}(h) & \Gamma_{y\Delta x}(h) \\ \Gamma_{\Delta xy}(h) & \Gamma_{\Delta x\Delta x}(h) \end{bmatrix} = E z_t z_{t+h}',
\end{aligned}
$$

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

$$\widehat{\lambda}_{y,\Delta x} = \sum_{h=1}^{n-1} k\left(\frac{h}{m}\right)\widehat{\Gamma}_{\Delta xy}(h); \quad \widehat{\Gamma}_{\Delta xy}(h) = \frac{1}{n}\sum_{t=h+1}^{n} y_t \Delta x_{t-h}, \tag{10}$$

where m is the bandwidth and $k(x)$ is the kernel.[6] 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

$$\lim_{x\to 0} \frac{1 - k(x)}{|x|^q} = k_q < \infty.$$

Assumption M

$$\frac{1}{m} + \frac{m^q}{n} \to 0 \quad \text{as } n \to \infty.$$

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 $q = 2$. The following two lemmas show the asymptotic bias and variance of $\widehat{\lambda}_{y,\Delta x}$ and its consistency.

2.1 Lemma

If Assumptions A, K and M hold, then

$$\lim_{n\to\infty} m^q E\left(\widehat{\lambda}_{y,\Delta x} - \lambda_{y,\Delta x}\right) = -k_q \sum_{h=1}^{\infty} \Gamma_{\Delta xy}(h)h^q.$$

The proof of this Lemma is omitted because it is the same as that of Theorem 10 in Hannan (1970, p. 283). Let $f_{yy}(\lambda)$ denote the spectral density of y_t and $f_{\Delta xy}(\lambda)$ denote the cross-spectral density between Δx_t and y_t, and similarly for $f_{y\Delta x}(\lambda)$ and $f_{\Delta x\Delta x}(\lambda)$. The following Lemma is a one-sided version of Theorem 9 of Hannan (1970, p. 280).

2.2 Lemma

$$\lim_{n\to\infty} \frac{n}{m}\text{var}\left(\widehat{\lambda}_{y,\Delta x}\right) = V \equiv 4\pi^2 \int_0^\infty k^2(x)\,dx \left\{f_{yy}(0)\, f_{\Delta x\Delta x}(0) + [f_{y\Delta x}(0)]^2\right\}.$$

[6] Note that this is simply the off-diagonal element of a one-sided long-run covariance estimator of the type employed in the HAC literature (e.g. **?** (**?**), **?** (**?**)) applied to $(y_t, \Delta x_t)$.

2.3 Corollary

If Assumptions A, K and M hold, then $\widehat{\lambda}_{y,\Delta x} \rightarrow_p \lambda_{y,\Delta x}$ *as* $n \rightarrow \infty$.

2.4 Remarks

1. If $k(x)$ is symmetric, we have

$$V = \left(\tfrac{1}{2}\right) 4\pi^2 \int_{-\infty}^{\infty} k^2(x)\,dx \left\{ f_{yy}(0) f_{\Delta x \Delta x}(0) + \left[f_{y\Delta x}(0)\right]^2 \right\} = \left(\tfrac{1}{2}\right) \lim_{n\rightarrow\infty} \operatorname{var}\left(\hat{\omega}_{y,\Delta x}\right),$$

   where $\hat{\omega}_{y,\Delta x}$ is the estimate of the long-run covariance between y_t and Δx_t. So, the asymptotic variance of $\widehat{\lambda}_{y,\Delta x}$ is just half the limiting variance for the two-sided case.

2. From Lemmas 2.1 and 2.2, the asymptotic mean squared error is minimized by choosing m such that

$$m^* = \left(2qk_q^2 \left(\sum_{h=1}^{\infty} \Gamma_{\Delta xy}(h) h^q \right)^2 n \Big/ V \right)^{1/(2q+1)}.$$

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

$$\begin{aligned} m^* &= \left(qk_q^2 \alpha(q) n \Big/ \int_{-\infty}^{\infty} k^2(x)\,dx \right)^{1/(2q+1)}, \qquad (11) \\ \alpha(q) &= \frac{4\left((2\pi)^{-1}\sum_{h=1}^{\infty}\Gamma_{\Delta xy}(h)h^q\right)^2}{f_{yy}(0) f_{\Delta x\Delta x}(0) + \left[f_{y\Delta x}(0)\right]^2}, \end{aligned}$$

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

3. When Δx_t follows ARFIMA(p,d,q) with $-1 < d < 0$, Lemma 2.2 still holds, but $f_{\Delta x\Delta x}(0) = f_{y\Delta x}(0) = f_{\Delta xy}(0) = 0$ and the limiting variance is 0. This suggests that the rate of convergence is faster when Δx_t is overdifferenced and will depend on d.

4. Lemma 1 (11) (p. 12) of **?** (**?**) shows that $\sum_{h=0}^{\infty} h^2 \left|\Gamma_{y\Delta x}(h)\right| < \infty$, so Lemma 2.2 holds if Δx_t is $I(d)$ with $-1 < d < 0$. Intuition for this result is given on p. 16-17 in the paragraph "Why then are the biases in Table 5 so reasonable..." and the proof is in Appendix A3 (p. 29).

2.5 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 Δx_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 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

$$\begin{aligned}\sum_{h=1}^{n-1} k\left(\frac{h}{m}\right)\frac{1}{n}\sum_{t=h+1}^{n} z_{t-h}z_t' &= \sum_{h=1}^{n-1} k\left(\frac{h}{m}\right)\int_{-\pi}^{\pi} I_z(\omega)e^{i\omega h}d\omega \\ &= \int_{-\pi}^{\pi} I_z(\omega)K_n(\omega)d\omega, \quad K_n(\omega)=\sum_{h=1}^{n-1} k\left(\frac{h}{m}\right)e^{i\omega h}.\end{aligned}$$

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 $\widehat{\Gamma}_{y\Delta x}$ by applying the martingale approximation *a la* **?** (**?**) and show the asymptotic normality of $\widehat{\lambda}_{y,\Delta x}$. The following theorem establishes it.

2.6 Theorem

If Assumptions A, K and M hold, $\exists\, \delta > 1$ such that $\sum_{h=-\infty}^{\infty} |h|^{\delta} ||\Gamma(h)|| < \infty$, and $m^2/n + n/m^{2q+1} \to 0$, then

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

The optimal bandwidth m^* does not satisfy the rate condition on m of Theorem 2.6, 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 2.6 to hold. Since the optimal rate of increase of m is $n^{1/(2q+1)}$ from Remark 2.4 (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.

2.7 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, $\widehat{\lambda}_{y,\Delta x}$ converges to $Ey_t x_{t-1} = \lambda_{y,\Delta x}$ when x_t is an $I(0)$ process. Let us first state the assumptions on x_t and y_t.

Assumption B

$$\begin{aligned} v_t &= \begin{pmatrix} y_t \\ x_t \end{pmatrix} = B(L)\varepsilon_t = \sum_{j=0}^{\infty} B_j \varepsilon_{t-j}, \quad \sum_{j=0}^{\infty} j||B_j|| < \infty, \quad (12) \\ \varepsilon_t &\sim \text{i.i.d.}\,(0, I_2), \quad \text{with finite fourth moment} \\ \sum_{-\infty}^{\infty} |h|^q \|\gamma(h)\| &< \infty, \quad \gamma(h) = \begin{bmatrix} \gamma_{yy}(h) & \gamma_{yx}(h) \\ \gamma_{xy}(h) & \gamma_{xx}(h) \end{bmatrix} = E v_t v_{t+h}', \end{aligned}$$

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 $\gamma_{xy}(1) = Ey_t x_{t-1} = \lambda_{y,\Delta x}$.

2.8 Lemma

If Assumptions B, K and M hold and $k(x)$ *is Lipschitz(1), then*

$$\sqrt{n}\left(\widehat{\lambda}_{y,\Delta x} - \lambda_{y,\Delta x}\right) = k(1/m)\sqrt{n}\left(\widehat{\gamma}_{xy}(1) - \gamma_{xy}(1)\right) + B_n + o_p(1), \tag{13}$$

where $\widehat{\gamma}_{xy}(1) = n^{-1}\sum_{t=2}^{n} y_t x_{t-1}$ and B_n is the bias term satisfying

$$B_n = \begin{cases} 0, & \text{if } Ey_t x_{t-h} = 0 \text{ for all } h \geq 1, \\ O(n^{1/2}m^{-q}), & \text{otherwise.} \end{cases}$$

In addition, $k(1/m)\sqrt{n}(\widehat{\gamma}_{xy}(1) - \gamma_{xy}(1)) \to_d N(0, \Xi)$ *as* $n \to \infty$, *where*

$$\Xi = \sum_{u=-\infty}^{\infty}\left\{\gamma_{xx}(u)\gamma_{yy}(u) + \gamma_{xy}(u+1)\gamma_{yx}(u-1)\right\} + \sum_{u=-\infty}^{\infty} k_{xyxy}(0,1,u,u+1).$$

2.9 Remarks

1. When x_t is $I(1)$, Theorem 2.6 requires the rate condition $m^2/n + n/m^{2q+1} \to 0$. Therefore, if $Ey_t x_{t-h} \neq 0$ for some h, then we need to use a kernel with $q = 2$ and choose m so that $m^2/n + n/m^4 \to 0$ for $\widehat{\lambda}_{y,\Delta x}$ to have a Gaussian limiting distribution centered around $\lambda_{y,\Delta x}$ both when x_t is $I(1)$ and $I(0)$. However, when the hypothesis of interest is the orthogonality between y_t and $I_{x,t-1}$, then m needs to satisfy only $m^2/n + n/m^{(2q+1)} \to 0$.

2. If you knew $x_t = I(0)$, then you would estimate $Ey_t x_{t-1}$ by $\widehat{\gamma}_{xy}(1)$, and the limiting variance of $\widehat{\lambda}_{y,\Delta x}$ is the same as that of $\widehat{\gamma}_{xy}(1)$. Therefore, $\widehat{\lambda}_{y,\Delta x}$ is robust to misspecification of the integration of order, apart from the bias term in (13).

2.10 The limit distribution when x_t is modelled as local to unity

Let x_t be a local-to-unity process:

$$x_t = \left(1 + \frac{c}{n}\right)x_{t-1} + u_t, \quad t = 1, 2, \ldots, \quad c < 0, \tag{14}$$

with $x_t \equiv 0$ for $t \leq 0$. Then $\lambda_{y,\Delta x} = \sum_{h=1}^{\infty}\mathrm{cov}(y_t, u_{t-h}) + O(n^{-1})$ as seen in (8), and also $\widehat{\lambda}_{y,\Delta x}$ behaves very similarly as when x_t is $I(1)$. The following Lemma establishes the limiting behavior of $\widehat{\lambda}_{y,\Delta x}$.

2.11 Lemma

Suppose x_t *is generated by (14) with* (y_t, u_t) *satisfying Assumption A. Then* $\widehat{\lambda}_{y,\Delta x} = \sum_{h=1}^{n-1} k(h/m)\widehat{\Gamma}_{uy}(h) + O_p((m/n))$, *where* $\widehat{\Gamma}_{uy}(h)$ *is defined in (10) with* u_t *replacing* x_t.

This Lemma establishes the first order equivalence of the limit theory for $\widehat{\lambda}_{y,\Delta x}$ under both $I(1)$ and local to unity assumptions (finite, negative c) on 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. This would seem to be a desirable property. By contrast, many econometric procedures, including several common cointegration tests, that are valid for $c = 0$ may fail for $c < 0$ (**?** (**?**)).

3 Possible ways to conduct inference

3.1 Estimation of the limiting variance of the estimator

Suppose x_t is $I(1)$ and Lemma 2.6 gives the limiting distribution of $\widehat{\lambda}_{y,\Delta x}$. To conduct inference, we need to estimate V, the limiting variance of $(n/m)^{1/2}\widehat{\lambda}_{y,\Delta x}$. Of course, we can use $\widehat{V} = 4\pi^2 \int_0^\infty k^2(x)\,dx\{\widehat{f}_{yy}(0)\,\widehat{f}_{\Delta x\Delta x}(0) + \widehat{f}_{y\Delta x}(0)\,\widehat{f}_{\Delta xy}(0)\}$, where $\hat{f}_{ab}$ is a standard periodogram-based estimator of f_{ab}. By standard arguments, this is a consistent estimator of V.

We may consider another estimator of V, $\widetilde{V}$, whose particularly good performance is suggested by simulations in Section 4. It is based on the exact finite sample variance of $\widehat{\lambda}_{y,\Delta x}$, which is given by (see equations (31)-(33) in the proof of Lemma 2.2)

$$
\begin{aligned}
&\frac{n}{m}\mathrm{var}\left(\widehat{\lambda}_{y,\Delta x}\right) \\
= \; & \frac{1}{m}\sum_{h'=1}^{n-1}\sum_{h=1}^{n-1} k\left(\frac{h'}{m}\right)k\left(\frac{h}{m}\right)\sum_{u=-\infty}^{\infty}\big\{\Gamma_{\Delta x\Delta x}(u)\,\Gamma_{yy}(u+h-h') \\
& + \Gamma_{\Delta xy}(u+h)\,\Gamma_{y\Delta x}(u-h') + k_{\Delta xy\Delta xy}(0,h',u,u+h)\big\}\,\phi_n(u,h',h),
\end{aligned}
$$

where $\phi_n(u,h',h)$ is defined in the proof of Lemma 2.2 . The terms involving the cumulants disappear in the limit. Define $\widetilde{V}$ by replacing Γ_{ab} with $\widehat{\Gamma}_{ab}$, which reduces the error from the approximation of the discrete sum in (29) by the integral in (36):

$$
\begin{aligned}
\widetilde{V} \; = \; & \frac{1}{m}\sum_{h'=1}^{n-1}\sum_{h=1}^{n-1} k\left(\frac{h'}{m}\right)k\left(\frac{h}{m}\right)\sum_{u=-\infty}^{\infty}\left\{\begin{array}{l}\widetilde{k}\left(\frac{u}{\widetilde{m}}\right)\widehat{\Gamma}_{\Delta x\Delta x}(u)\,\widetilde{k}\left(\frac{u+h-h'}{\widetilde{m}}\right)\widehat{\Gamma}_{yy}(u+h-h') \\ +\widetilde{k}\left(\frac{u+h}{\widetilde{m}}\right)\widehat{\Gamma}_{\Delta xy}(u+h)\,\widetilde{k}\left(\frac{u-h'}{\widetilde{m}}\right)\widehat{\Gamma}_{y\Delta x}(u-h')\end{array}\right\} \\
& \times\phi_n(u,h',h), \qquad (15)
\end{aligned}
$$

where $\widetilde{k}(x)$ and $\widetilde{m}$ are kernel and bandwidth. $\widetilde{k}(x)$ and $\widetilde{m}$ can, but do not need to, be the same as $k(x)$ and m. Estimating V by $\widetilde{V}$ gives better finite sample performance than estimating V by $\widehat{V}$. (The results using $\widehat{V}$ are not reported in the present paper).

Suppose $(y_t, \Delta x_t)$ satisfies Assumption A and hence x_t is $I(1)$. Then, if $\widetilde{V} \to_p V$, we may construct a t-type statistic

$$t_\lambda = \frac{\sqrt{\frac{n}{m}}\left(\hat{\lambda}_{y,\Delta x} - \lambda_{y,\Delta x}\right)}{\sqrt{\widetilde{V}}}, \tag{16}$$

that converges to a $N(0,1)$ random variable in distribution from Lemma 2.6. The following Lemma shows that it is indeed the case.

3.2 Lemma

If Assumptions A, K, and M hold, the kernel $\widetilde{k}(x)$ satisfies Assumption K with $\widetilde{k}(x) = 0$ if $|x| > 1$, and $1/\widetilde{m} + \widetilde{m}^q/n \to 0$, then $\widetilde{V} \to_p V$ as $n \to \infty$.

3.3 Corollary

If the assumptions of Theorem 2.6 and Lemma 3.2 hold, then $t_\lambda \to_d N(0,1)$ as $n \to \infty$.

3.4 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 (12) and x_t is actually $I(0)$. Since t_λ is based on the autocovariance of y_t and Δx_t, the inference based on t_λ might be misleading. However, if the Bartlett kernel $\widetilde{k}(x) = (1 - |x|)\mathbf{1}\{|x| \leq 1\}$ is used in $\widetilde{V}$ in (15) and $Ey_t x_{t-h} = 0$ for all $h \geq 1$ (which holds under the null hypothesis of orthogonality), then t_λ is $O_p((\widetilde{m}/m)^{1/2})$. Therefore, when $\widetilde{m}$ is chosen appropriately, large values of $\hat{\lambda}_{y,\Delta x}$ suggest the rejection of the orthogonality between y_t and $I_{x,t-1}$, and $\hat{\lambda}_{y,\Delta x}$ serves as a tool for conservative inference. The following Lemma establishes it. The power property of t_λ when x_t is $I(0)$ can be checked by simulation.

3.5 Lemma

If Assumptions B, K and M hold, $\widetilde{k}(x) = (1 - |x|)\mathbf{1}\{|x| \leq 1\}$, $1/\widetilde{m} + \widetilde{m}^q/n \to 0$, and $Ey_t x_{t-h} = 0$ for all $h \geq 1$, then $t_\lambda = O_p((\widetilde{m}/m)^{1/2})$ as $n \to \infty$.

In order to understand the convergence, rewrite t_λ as

$$t_\lambda = \frac{n^{1/2}(\hat{\lambda}_{y,\Delta x} - \lambda_{y,\Delta x})}{(\widetilde{V})^{1/2} m^{1/2}}.$$

The numerator converges to a Gaussian random variable from Lemma 2.8. $\widetilde{V}$ in the denominator is an estimate of $f_{\Delta x \Delta x}(0) = 0$ and hence converges to 0 as $\widetilde{m} \to \infty$. Because m tends to infinity, the asymptotic behavior of t_λ depends on the rate of convergence of $\widetilde{V}$. Letting $\widetilde{m}$ tend to infinity but not too fast prevents $\widetilde{V}$ from converging to 0 too fast and makes t_λ converge to 0 in probability.

In summary, by choosing $\widetilde{m}$ appropriately, the t_λ statistic provides a standard inferential tool if x_t is $I(1)$ or local to unity but converges to zero when x_t is $I(0)$. Simulation results reported below suggest that it works well in practice.

4 Finite sample performance: simulation results

This section provides a modest simulation study to gage 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 $\mathcal{I}_{x,t-1}$, the information contained in past x_t, as is often tested in practice using a regression of y_t on x_{t-1}. 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:

$$\begin{aligned} x_t &= \rho_0 + \rho_1 x_{t-1} + u_{2t}, \quad \text{AR(1)} && (17)\\ x_t &= \rho_0 + \rho_1 x_{t-1} + \rho_2 x_{t-2} + u_{2t}. \quad \text{AR(2)} && (18)\end{aligned}$$

The AR(1) model may also be written as a unit root/local to unity process by letting

$$\rho_1 = 1 + \frac{c}{n}\ , \ c \leq 0. \quad (19)$$

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:

$$\begin{aligned} d_t &= \delta_0 \quad \text{or} && (21)\\ d_t &= \delta_0 + \delta_1 t. && (22)\end{aligned}$$

We employ two different specifications for y_t under the alternative 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 is 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, given the choice, it is not clear that one would want to model near unit root components in stock or exchange rate returns, especially as this implies near I(2) components in the levels. Moreover, empirically, returns show little serial correlation.[7] 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

$$y_t = d_t + \gamma (1 - \rho_1 L) x_{t-1} + u_{1t}, \tag{24}$$

where x_t is given by the AR(1) specification in (17). This may be rewritten as

$$y_t = d_t + \gamma u_{2,t-1} + u_{1t}, \tag{25}$$

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

$$\begin{aligned} u &= \begin{pmatrix} u_{1t} & u_{2t} \end{pmatrix}' = \Sigma^{1/2}\varepsilon_t, \quad \varepsilon_t \sim N(0, I) \\ \Sigma &= \Sigma^{1/2}(\Sigma^{1/2})' = \begin{pmatrix} 1 & \sigma_{12} \\ \sigma_{21} & 1 \end{pmatrix}. \end{aligned}$$

Our primary interest lies in the performance of the covariance-based t-statistic t_λ given in (16), which was estimated as follows. In the trend model (22), we first demeaned Δx_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 (6) using the standard kernel covariance estimator $\widehat{\lambda}_{y,\Delta x}$ given in (10). Likewise, we estimated its asymptotic variance V (see Lemma 2.2) 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 3.5 mandates use of the Bartlett (**?** (**?**)) kernel for $\widetilde{k}(x)$ in the estimation of V. We therefore used the Bartlett kernel for both estimators, setting

$$k(x) = \widetilde{k}(x) = 1 - |x| \quad \text{for} \quad |x| \leq 1$$

and zero otherwise. The bandwidth parameter m in the estimation of $\lambda_{y,\Delta x}$ is chosen to minimize the asymptotic mean squared error in the spirit of Andrews (91) using the optimal bandwidth formula given in (11). Implementation of this formula in practice requires the use of a fist-stage parametric approximation model. As in Andrews (91) this is assumed only to 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

[7]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.

given possible over-differencing in Δx_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 **?** (?) allowed us to avoid non-linear optimization, keeping estimation simple.[8] The choice of the second bandwidth parameter $\widetilde{m}$ used in estimation of V is constrained by Lemma 3.5 which requires $\widetilde{m} = o(m)$. While this offers many possibilities, our choice of $\widetilde{m} = m^{0.9}$ 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 **?** (?) (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 ρ_1 modelled local to unity as in (19). Results are provided for various values of both c (and therefore ρ_1) and σ_{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 ρ_1 and/or σ_{12} but grow highly unreliable as ρ_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_λ 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 results from the fact that the covariance estimator upon which the test statistic is based is asymptotically normal even for $\rho_1 = 1$, and as a result is not effected by the same unit root biases as the regression based tests. Good performance may also be obtained by properly size adjusting the regression based tests, as in the bounds tests of CES, together with the finite sample adjustments detailed therein (see their Table 4).

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. Such a specification may be of practical relevance. 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

[8] 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.

series. In order to roughly match these estimates we set

$$\rho_1 = 1.5 \quad \text{and} \quad \rho_2 = -0.5 + \frac{c}{n} \tag{26}$$

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 3. 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_λ against the balanced regression alternative given in (23) with $\beta \neq 0$ and local to unity x_t given by (17) and (19). 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 4. As expected the power of the test is reasonable, increasing in both sample size and distance from the null.[9]

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 (1) and (4) as special cases for $\rho_1 = 0$ and $\rho_1 = 1$ respectively. A second advantage of this model is that the population R^2 for a regression of y_t on $(1-\rho_1 L)x_{t-1}$ remains constant across different values of ρ_1 and σ_{12}, allowing for a clearer interpretation of rejection rates.

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 5. These rejection rates appear quite reasonable, again increasing in both sample size and distance from the null hypothesis. With a sample size of one hundred rejection rates become sizable

[9]To save space we show only the results for the detrended case. Results for the demeaned case (available upon request) are generally similar, except that the power appears a bit weaker for $c = -1$ and $c = -2.5$.

at an R^2 of about 0.10. With 400 observations rejection rates exceed fifty percent even for an R^2 of 0.02.

Many existing tests are based on a size adjusted regression of the type shown in (1). 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 has a root close or equal to unity (e.g. $\rho_1 \approx 1$ in (17) and $\beta \neq 0$ in (23)). This is confirmed in Table 6, 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_λ.

On the other hand, it is not clear that tests based on (1) should have much power against unbalanced alternatives, since the parameter restriction tested (i.e. $\beta_1 = 0$) is satisfied for *all* unbalanced relations. In fact, if the size adjustments were made using the fully-modified approach (**?** (**?**)) one would expect a rejection rate equal to size when $c = 0$ for any unbalanced alternative. Since most existing tests are not fully modified, exact rejection rates are less clear, but may be examined by simulation. Table 7 provides rejection rates for the CES Bonferonni test against the same unbalanced alternative (and same DGP) used to assess the power of t_λ in Table 5. 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.[10] This added generality does of course come at some real 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

[10] While one may conceivably provide size adjustments for larger more flexible regressions, which nest both alternatives, we are not aware of any existing results, and note also that size corrections, particular those based on bounds procedures, may become increasingly complicated and conservative as additional regressors are added.

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), **?** (?), **?** (?), **?** (?)).

Although theoretical considerations may rule out exact unit root behavior in dividend yields[11] 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 (**?** (?)). Moreover, although predetermined, 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 (**?** (?)). Consequently, beginning with **?** (?) and **?** (?) subsequent doubts have been raised regarding the evidence for predictability on account of the strong persistence in the regressors. Also of concern have been the accuracy of the standard errors in long-horizon regressions (**?** (?), **?** (?)).

A large literature has since developed using various techniques to evaluate and/or correct for the size distortion in these regressions. These include resampling and simulation methods,[12] local to unity corrections along the lines of **?** (?),[13] and finite sample or Bayesian approaches.[14] A reading of the literature suggests some consensus that size problems may have led the original regressions to overstate the degree of predictability. However, the extent of this overstatement remains a subject of ongoing debate, with some studies finding little or no predictability and other finding significant predictability even after adjusting for size distortion.[15]

? (?) and **?** (?) offer a second perspective. They argue that if returns are linearly related to interest rates or dividend yields and if these show near unit-root behavior, then one must expect a near-unit component to returns as well. Since persistent components to stock returns are neither theoretically appealing nor empirically observed, they take this as evidence against predictability. Our interpretation is somewhat different. Rather than viewing this as evidence against predictability, we view it instead as a shortcoming of tests based on (1) that may in fact make predictability harder to detect. Under the null hypothesis of risk-neutral market efficiency, the returns are not only linearly unrelated to the level of past interest rates or dividend-yields, but are *fully orthogonal to all information* in the past history of these regressors. This includes first-differences, high-frequency components, and deviations of the regressors from recent historical averages, all of which could contain potential predictive value for a stationary return series, even if the regressor itself displays near unit root behavior. By insisting on a balanced alternative, in which both sides of the equation share the same persistent component, a test based on (1), even if size-adjusted,

[11] Campbell and Shiller (1988a,b), but see **?** (?) for an alternative viewpoint.

[12] See **?** (?), **?** (?), **?** (?), **?** (?), and **?** (?).

[13] See **?** (?), **?** (?), **?** (?), and **?** (?).

[14] See **?** (?) and **?** (?).

[15] It is hard to categorize past work neatly into these two groups, but very roughly speaking **?** (?), **?** (?), **?** (?), **?** (?), and **?** (?) fall into the first group, while **?** (?), **?** (?), **?** (?), and **?** (?) belong to the second.

may potentially overlook predictive content of this type, leaving us to consider only alternatives that arguably become less attractive the more persistent the regressor.

In principle, the net impact of persistent regressors on predictability tests is therefore uncertain as they may lead regression tests either to overstate predictability as a result of bias and size distortion or to understate predictability by restricting the nature of the alternative. 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.[16] Monthly log returns are calculated as $r_{t+1} = ln\left((P_{t+1} + D_{t+1})/P_t\right)$, where P_t and D_t are the stock price and dividend from the CRSP value-weighed index of NYSE, AMEX, and NASDAQ stocks. Real returns are formed by deflating nominal returns by the CPI.[17] 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: $d_t - p_t = ln\left((D_t + \ldots + D_{t-11})/P_t)\right)$. 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 $k = 1, 3, 12$, and 24. The relatively weak assumptions that we place on the univariate behavior of y_t, even under the null hypothesis, becomes an advantage in the context of long-horizon approaches. Thus we require no explicit corrections or extensions to handle the moving average components induced by the overlapping returns.[18] The test statistic t_λ remains asymptotically standard normal, as both the estimator and variance estimate adjust automatically to the properties of the data. Nevertheless, the critique of **?** (**?**)[19] may still apply when the horizon is large relative to sample size and for this reason we limit ourselves to a maximum horizon of two years.[20]

Table 8 shows the standard results from separate regressions of the k period return on interest rates and dividend yields using HAC standard errors for $k > 1$. The interest rate regressions show only very 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 **?** (**?**) who expresses the bias in $\widehat{\beta}$ in (1) in terms of the bias in $\widehat{\rho}$ in (2) and the residual covariance σ_{12} in (3):

$$E\left[\hat{\beta} - \beta\right] = \frac{\sigma_{12}}{\sigma_{22}} E\left[\hat{\rho} - \rho\right]. \tag{27}$$

The two ingredients needed to produce bias are thus persistent regressors and residual serial correlation. Table 9 shows the **?** (**?**) confidence interval on the largest root in

[16]We thank John Campbell for kindly providing us with this data.

[17]Similar results were also found replacing real by excess returns.

[18]**?** (**?**) and **?** (**?**) extend the local to unity approach of CES to apply to long-horizon returns. It may be more complicated to extend exact finite sample approaches (**?** (**?**), **?** (**?**), Campbell and Dufour, (95 & 97)) to the long-horizon case.

[19]See also **?** (**?**).

[20]Following much of the literature, the tests are conducted separately at each horizon and must be interpreted accordingly. A joint covariance based test is conceivable but well beyond the scope of the current paper.

x_t together with the estimated residual correlation $\delta = corr(\varepsilon_{1t}, \varepsilon_{2t})$. 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 δ 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 (**?** (**?**)), negative residual correlation implies positive bias in $\widehat{\beta}$ (see 27). In other words, the bias runs in the same direction as the observed alternative, leaving the results difficult to interpret.

Thus the preliminary analysis suggests a priori that size distortion likely plays a central role in the case of the dividend price ratio, whereas the specification/power issues associated with near unit roots may be of greater importance for the interest rate regressions. Roughly speaking, this is what we find in Table 10 by applying our covariance based method to test the orthogonality of the returns to the information in past interest rates and dividend price ratios. The tests are conducted in the same way as in the simulations and the reader is referred to Section 4 for details.[21] The table shows both the optimal bandwidth m^* and the value of the test statistic t_λ based on this bandwidth. Standard normal critical values apply in all cases.

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 previous studies which size correct these regressions (**?** (**?**), **?** (**?**), **?** (**?**), **?** (**?**)) and based on the results shown it would be difficult to make a strong case for predictability using the dividend yield. However, before drawing too strong a conclusion, we note a few marginally significant t-statics in additional robustness results (available upon request), as well as some disagreement on this in the literature.[22] As always, one must exercise caution in interpreting a failure to reject. Nevertheless it seems safe to conclude that the evidence and degree of predictability found in regressions using dividend yields is, at the least, overstated.

Thus, in the case of the dividend yield, the near unit root problem seems to arise primarily in the context of size distortion. For the interest rate, the specification of the alternative appears instead to be more important. This becomes evident upon comparison of the covariance based tests in Table 10 with the regression tests in Table 8. During the 1930s and 1940s the short rates were pegged by the Federal Reserve[23] and thus neither test shows evidence of predictability during the earlier sample. However, during the later sample, while the regressions in Table 8 show just a hint of predictability at short-horizons, the covariance based t statistics in Table 10 show clear evidence of predictability at the three and, to a lesser extent, the twelve month horizon.[24] Moreover, this suggests predictability in the full sample

[21] The detrended version of the estimator is employed both because of its more general applicability and because of relative power advantages revealed in simulation even in the absence of any trend. Quite similar results (available upon request) were obtained using the demeaned version.

[22] See, for example, **?** (**?**) and **?** (**?**).

[23] See Campbell et al. (1997, p. 268).

[24] Additional results available upon request also show this significance to be quite robust.

as well, an implication which is detected quite clearly at the three month horizon by the covariance based test, but not at all by the regression t-test. The inability of the regression tests to fully detect this predictability may be on account of their restricting the alternative to a direct linear relation between returns (which show little persistence) and highly persistent interest rates. We confirm this conjecture in Table 11 by showing 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 **?** (**?**)).[25] In conclusion, we find that standard regression tests based on (1) 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 t and F 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 alternatives 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, 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: 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.

[25] This amounts to a gradual differencing over a 12 month period.

7.1 Proof of (8)

From the definition of x_t, we have

$$\Delta x_t = u_t + \frac{c}{n} x_{t-1} = \begin{cases} u_t + \frac{c}{n} \sum_{k=0}^{t-2} \left(1 + \frac{c}{n}\right)^k u_{t-1-k}, & t \geq 1, \\ 0, & t \leq 0, \end{cases} \tag{28}$$

with $\sum_{k=0}^{-1} \equiv 0$. It follows that

$$\mathrm{cov}(y_t, \Delta x_{t-h}) = \begin{cases} \mathrm{cov}(y_t, u_{t-h}) + \frac{c}{n} \sum_{k=0}^{t-h-2} \left(1 + \frac{c}{n}\right)^k \mathrm{cov}(y_t, u_{t-h-1-k}), & t \geq h+1, \\ 0, & t \leq h. \end{cases}$$

Therefore,

$$\begin{aligned} \lambda_{y,\Delta x} &= \lim_{t\to\infty} \sum_{h=1}^{t-1} \mathrm{cov}\ (y_t, \Delta x_{t-h}) \\ &= \lim_{t\to\infty} \sum_{h=1}^{t-1} \mathrm{cov}(y_t, u_{t-h}) + \frac{c}{n} \lim_{t\to\infty} \sum_{h=1}^{t-1} \sum_{k=0}^{t-h-2} \left(1 + \frac{c}{n}\right)^k \mathrm{cov}(y_t, u_{t-h-1-k}). \end{aligned}$$

The first term converges to $\sum_{h=1}^{\infty} \mathrm{cov}(y_t, u_{t-h})$. The second term is bounded by (by letting $p = k + h$)

$$\begin{aligned} \frac{c}{n} \lim_{t\to\infty} \sum_{h=1}^{t-1} \sum_{k=0}^{t-h-2} \left|\mathrm{cov}(y_t, u_{t-h-1-k})\right| &= \frac{c}{n} \lim_{t\to\infty} \sum_{p=1}^{t-2} \sum_{k=0}^{p-1} \left|\mathrm{cov}(y_t, u_{t-1-p})\right| \\ &= O\left(\frac{1}{n} \sum_{p=1}^{\infty} p \left|\mathrm{cov}(y_t, u_{t-1-p})\right|\right) = O\left(\frac{1}{n}\right), \end{aligned}$$

giving the stated result. ■

7.2 Proof of Lemma 2.2

The proof closely follows that of Theorem 9 of Hannan (1970, p. 280). See Hannan (1970) pp. 313-316 for details. Observe that

$$\frac{n}{m} \mathrm{var}\left(\widehat{\lambda}_{y,\Delta x}\right) = \frac{n}{m} \sum_{h'=1}^{n-1} \sum_{h=1}^{n-1} k\left(\frac{h'}{m}\right) k\left(\frac{h}{m}\right) \mathrm{cov}\left(\widehat{\Gamma}_{\Delta xy}\left(h'\right), \widehat{\Gamma}_{\Delta xy}\left(h\right)\right). \tag{29}$$

Hannan (1970) p. 313 gives

$$\begin{aligned} & \mathrm{cov}\left(\widehat{\Gamma}_{\Delta xy}\left(h'\right), \widehat{\Gamma}_{\Delta xy}\left(h\right)\right) \qquad (30) \\ = \ & n^{-1} \sum_{u=-\infty}^{\infty} \{\Gamma_{\Delta x \Delta x}\left(u\right) \Gamma_{yy}(u + h - h') + \Gamma_{\Delta xy}(u+h) \Gamma_{y\Delta x}(u - h') \\ & + k_{\Delta xy \Delta xy}(0, h', u, u+h)\} \phi_n(u, h', h), \end{aligned}$$

where $k_{\Delta xy\Delta xy}(0,h',u,u+h)$ 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)

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

It follows that (29) is comprised of

$$\frac{1}{m}\sum_{h'=1}^{n-1}\sum_{h=1}^{n-1} k\left(\frac{h'}{m}\right) k\left(\frac{h}{m}\right) \sum_{u=-\infty}^{\infty} \Gamma_{\Delta x\Delta x}\left(u\right)\Gamma_{yy}(u+h-h')\phi_n(u,h',h) \quad (31)$$

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

$$+\frac{1}{m}\sum_{h'=1}^{n-1}\sum_{h=1}^{n-1} k\left(\frac{h'}{m}\right) k\left(\frac{h}{m}\right) \sum_{u=-\infty}^{\infty} k_{\Delta xy\Delta xy}(0,h',u,u+h)\phi_n(u,h',h). \quad (33)$$

Let $v = h' - h$, and we can rewrite (31) as

$$\sum_{v=-n+2}^{n-2}\sum_{u=-\infty}^{\infty} \Gamma_{\Delta x\Delta x}\left(u\right)\Gamma_{yy}(u-v)\left\{\frac{1}{m}\sum_{h}{}'\phi_n(u,h+v,h)k\left(\frac{h+v}{m}\right)k\left(\frac{h}{m}\right)\right\}, \quad (34)$$

where the summation $\sum_h'$ runs only for $\{h : 1 \leq h \leq n-1 \text{ and } 1 \leq h+v \leq n-1\}$. The bracketed expression converges to $\int_0^\infty k^2\left(x\right)dx$ by the argument in **?** (?) pp. 314-15. Furthermore,

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

and hence (31) converges to $4\pi^2 f_{\Delta x\Delta x}\left(0\right)f_{yy}\left(0\right)\int_0^\infty k^2\left(x\right)dx$ as $n\to\infty$. Similarly, (32) converges to $4\pi^2 f_{\Delta xy}\left(0\right)f_{y\Delta x}\left(0\right)\int_0^\infty k^2\left(x\right)dx = 4\pi^2[f_{y\Delta x}\left(0\right)]^2\int_0^\infty k^2\left(x\right)dx$. For (33), from Hannan (1970, p. 211), the fourth cumulant of z_t satisfies

$$\sum_{q=-\infty}^{\infty}\sum_{r=-\infty}^{\infty}\sum_{s=-\infty}^{\infty}|k_{ijkl}(0,q,r,s)| < \infty, \quad i,j,k,l = \{y,\Delta x\}. \quad (35)$$

Therefore, (33) is bounded by

$$C\frac{1}{m}\sum_{h'=-\infty}^{\infty}\sum_{h=-\infty}^{\infty}\sum_{u=-\infty}^{\infty}\left|k_{\Delta xy\Delta xy}(0,h',u,u+h)\right| = O\left(\frac{1}{m}\right),$$

and it follows that

$$\frac{n}{m}\text{var}\left(\widehat{\lambda}_{y,\Delta x}\right) \to 4\pi^2\int_0^\infty k^2\left(x\right)dx\left\{f_{\Delta x\Delta x}\left(0\right)f_{yy}\left(0\right)+[f_{y\Delta x}\left(0\right)]^2\right\}, \quad (36)$$

as $n\to\infty$, giving the stated result. ■

7.3 Proof of Theorem 2.6

In view of Lemma 2.1, it suffices to show that $\sqrt{n/m}(\widehat{\lambda}_{y,\Delta x} - E\widehat{\lambda}_{y,\Delta x}) \rightarrow_d N(0,V)$. First, observe that

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

where

$$\begin{aligned} I &= \frac{1}{\sqrt{m}}\sum_{h=1}^{n-1} k\left(\frac{h}{m}\right)\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(y_t\Delta x_{t-h} - Ey_t\Delta x_{t-h}\right), \\ II &= -\frac{1}{\sqrt{m}}\sum_{h=1}^{n-1} k\left(\frac{h}{m}\right)\frac{1}{\sqrt{n}}\sum_{t=1}^{h}\left(y_t\Delta x_{t-h} - Ey_t\Delta x_{t-h}\right). \end{aligned}$$

From Lemma 8.1 and Minkowski's inequality, we have

$$\begin{aligned} E(II)^2 &= O\left(\frac{1}{mn}\left(\sum_{h=1}^{n-1}\left|k\left(\frac{h}{m}\right)\right|h^{1/2}\right)^2\right) \\ &= O\left(\frac{m^2}{n}\left(\sum_{h=1}^{n-1}\left|k\left(\frac{h}{m}\right)\right|\left(\frac{h}{m}\right)^{1/2}\frac{1}{m}\right)^2\right) = O\left(\frac{m^2}{n}\right), \end{aligned} \tag{38}$$

because $\sum_{h=1}^{n-1}\left|k\left(h/m\right)\right|\left(h/m\right)^{1/2}m^{-1} \sim \int_0^\infty |k(x)|x^{1/2}dx < \infty$. Lemma 8.3 gives

$$I = \sum_{t=1}^{n} Z_t + R_n; \quad Z_t = n^{-1/2}m^{-1/2}\sum_{h=1}^{n-1} k\left(\frac{h}{m}\right)\sum_{r=1}^{\infty}\varepsilon_{t-r}' f^{hr}(1)\varepsilon_t, \tag{39}$$

where $ER_n^2 = o(1)$ and $f^{hr}(1)$ is defined in the statement of Lemma 8.3. Therefore, $\sqrt{n/m}(\widehat{\lambda}_{y,\Delta x} - E\widehat{\lambda}_{y,\Delta x}) \rightarrow_d N(0,V)$ follows if we show

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

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 martingale CLT **?** (**?**) if

$$\begin{aligned} \text{(i)} \quad \sum_{t=1}^{n} E(Z_t^2|\mathcal{I}_{t-1}) &= \frac{1}{n}\sum_{t=1}^{n} E(nZ_t^2|\mathcal{I}_{t-1}) \rightarrow_p V, \\ \text{(ii)} \quad \sum_{t=1}^{n} E(Z_t^2\mathbf{1}\{|Z_t| &\geq \delta) \rightarrow_p 0 \quad \text{for all } \delta > 0. \end{aligned}$$

First we show (i). Observe that

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

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

$$E\left[E(nZ_t^2|\mathcal{I}_{t-1})\right] = nEZ_t^2.$$

Therefore, (i) follows from the ergodic theorem if

$$nEZ_t^2 \to V. \tag{41}$$

From (37)-(39), we have

$$\sqrt{\frac{n}{m}}\left(\widehat{\lambda}_{y,\Delta x} - E\widehat{\lambda}_{y,\Delta x}\right) = \sum_{t=1}^{n} Z_t + II + R_n, \quad E(II+R_n)^2 = o(1),$$

or equivalently,

$$\sum_{t=1}^{n} Z_t = \sqrt{\frac{n}{m}}\left(\widehat{\lambda}_{y,\Delta x} - E\widehat{\lambda}_{y,\Delta x}\right) - (II+R_n).$$

Taking the second moment of the both sides gives

$$E\left(\sum_{t=1}^{n} Z_t\right)^2 = E\left(\sqrt{\frac{n}{m}}\left(\widehat{\lambda}_{y,\Delta x} - E\widehat{\lambda}_{y,\Delta x}\right) - (II+R_n)\right)^2. \tag{42}$$

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. From

$$E\left(\sqrt{\frac{n}{m}}\left(\widehat{\lambda}_{y,\Delta x} - E\widehat{\lambda}_{y,\Delta x}\right)\right)^2 = \mathrm{var}\left(\sqrt{\frac{n}{m}}\left(\widehat{\lambda}_{y,\Delta x} - E\widehat{\lambda}_{y,\Delta x}\right)\right) \to V,$$

$E(II+R_n)^2 = o(1)$, and Cauchy-Schwartz inequality, the right hand side of (42) is

$$\mathrm{var}\left(\sqrt{\frac{n}{m}}\left(\widehat{\lambda}_{y,\Delta x} - E\widehat{\lambda}_{y,\Delta x}\right)\right)^2 + o(1) \to V.$$

Therefore, we establish (41) and (i). For (ii), the stationarity of Z_t gives $\sum_{t=1}^{n} E(Z_t^2\mathbf{1}\{|Z_t| \geq \delta) = E(nZ_t^2\mathbf{1}\{|nZ_t^2| \geq n\delta^2)$, and $E(nZ_t^2\mathbf{1}\{|nZ_t^2| \geq n\delta^2) \to 0$ follows from $E(nZ_t^2) \to V < \infty$ and the dominated convergence theorem. Therefore, (40) and the stated result follow. ■

7.4 Proof of Lemma 2.8

Some simple algebra gives

$$\begin{aligned}
\widehat{\lambda}_{y,\Delta x} &= \sum_{h=1}^{n-1} k\left(\frac{h}{m}\right)\frac{1}{n}\sum_{t=h+1}^{n} y_t \Delta x_{t-h} \\
&= \sum_{h=1}^{n-1} k\left(\frac{h}{m}\right)\frac{1}{n}\sum_{t=h+1}^{n} y_t x_{t-h} - \sum_{h=1}^{n-1} k\left(\frac{h}{m}\right)\frac{1}{n}\sum_{t=h+1}^{n} y_t x_{t-h-1} \\
&= \sum_{h=1}^{n-1} k\left(\frac{h}{m}\right)\frac{1}{n}\sum_{t=h+1}^{n} y_t x_{t-h} - \sum_{p=2}^{n} k\left(\frac{p-1}{m}\right)\frac{1}{n}\sum_{t=p}^{n} y_t x_{t-p} \quad (p = h+1) \\
&= k\left(\frac{1}{m}\right)\frac{1}{n}\sum_{t=2}^{n} y_t x_{t-1} + \sum_{h=2}^{n-1}\left[k\left(\frac{h}{m}\right) - k\left(\frac{h-1}{m}\right)\right]\frac{1}{n}\sum_{t=h+1}^{n} y_t x_{t-h} \\
&\quad - \sum_{p=2}^{n-1} k\left(\frac{p-1}{m}\right)\frac{1}{n} y_p x_0 - k\left(\frac{n-1}{m}\right)\frac{1}{n} y_n x_0 \\
&= T_{1n} + T_{2n} + T_{3n} + T_{4n}.
\end{aligned}$$

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

$$\sqrt{n}(T_{n1} - \lambda_{y,\Delta x}) = k(1/m)\sqrt{n}(\widehat{\gamma}_{xy}(1) - \gamma_{xy}(1)) + (k(1/m) - 1)\sqrt{n} E y_t x_{t-1}.$$

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

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

where Ξ is given by **?** (**?**) 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 $\varepsilon > 0$ so that

$$\begin{aligned}
E(T_{2n}) &= \sum_{h=2}^{\varepsilon m}\left[k\left(\frac{h}{m}\right) - k\left(\frac{h-1}{m}\right)\right]\frac{n-h}{n}\gamma_{xy}(h) \\
&\quad + \sum_{h=\varepsilon m+1}^{n-1}\left[k\left(\frac{h}{m}\right) - k\left(\frac{h-1}{m}\right)\right]\frac{n-h}{n}\gamma_{xy}(h) \\
&= B_{1n} + B_{2n}.
\end{aligned}$$

Since $k(x) - 1 = O(x^q)$ as $x \to 0$ from Assumption K, choosing ε sufficiently small gives $B_{1n} = O(\sum_{h=2}^{m}(h/m)^q |\gamma_{xy}(h)|) = O(m^{-q})$. Since $k(x)$ is Lipschitz(1), B_{2n} is bounded by

$$C\frac{1}{m}\sum_{h=\varepsilon m}^{n-1}|\gamma_{xy}(h)| \leq C\frac{1}{m}\left(\frac{1}{\varepsilon m}\right)^{q}\sum_{h=\varepsilon m}^{n-1}h^{q}|\gamma_{xy}(h)| = O(m^{-q}).$$

Therefore, defining $B_n = (k(1/m)-1)\sqrt{n}Ey_t x_{t-1} + ET_{2n}$ gives the bias term B_n in (13).

It remains to show that $\mathrm{var}(\sqrt{n}T_{2n}) = o(1)$ and $\sqrt{n}(T_{3n}+T_{4n}) = o_p(1)$. From **?** (**?**) (equation (3.3) on page 209 and line 5 on page 211), we have

$$\begin{aligned}
&\mathrm{cov}\left(\sqrt{n}\widehat{\gamma}_{xy}(h), \sqrt{n}\widehat{\gamma}_{xy}(h')\right)\\
=\ &\sum_{u=-n+1}^{n-1}\left(1-\frac{|u|}{n}\right)\left\{\gamma_{xx}(u)\gamma_{yy}(u+h-h') + \gamma_{xy}(u+h)\gamma_{yx}(u-h')\right\}\\
&+\sum_{u=-n+1}^{n-1}\left(1-\frac{|u|}{n}\right)k_{xyxy}(0,h,u,u+h').
\end{aligned}$$

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

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

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

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

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

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

and the stated result follows. ■

7.5 Proof of Lemma 2.11

From (28), we have

$$\frac{1}{n}\sum_{t=h+1}^{n} y_t \Delta x_{t-h} = \frac{1}{n}\sum_{t=h+1}^{n} y_t u_{t-h} - \frac{c}{n^2}\sum_{t=h+1}^{n}\sum_{k=0}^{t-h-2}\left(1-\frac{c}{n}\right)^k y_t u_{t-h-1-k}.$$

The stated result follows if we show

$$T_n = \frac{c}{n^2}\sum_{h=1}^{n-1} k\left(\frac{h}{m}\right)\sum_{t=h+1}^{n}\sum_{k=0}^{t-h-2}\left(1-\frac{c}{n}\right)^k y_t u_{t-h-1-k} = O_p\left(\frac{m}{n}\right).$$

Since $(1-\frac{c}{n})^{\,k} = O(1)$, $E|T_n|$ is bounded by

$$\begin{aligned} & \frac{1}{n^2}\sum_{h=1}^{n-1} k\left(\frac{h}{m}\right)\sum_{t=h+1}^{n}\sum_{k=0}^{t-h-2}\left|E y_t u_{t-h-1-k}\right| \\ \leq\ & \frac{1}{n^2}\sum_{h=1}^{n-1} k\left(\frac{h}{m}\right)\sum_{t=h+1}^{n}\sum_{k=-\infty}^{\infty}\left|\Gamma_{uy}(k)\right| = O\left(\frac{1}{n}\sum_{h=1}^{n-1} k\left(\frac{h}{m}\right)\right) = O\left(\frac{m}{n}\right), \end{aligned}$$

giving the stated result. ■

7.6 Proof of Lemma 3.2

From equations (31), (32) and (34) in the proof of Lemma 2.2, $\widetilde{V}$ comprises of two parts, the first of which is

$$\sum_{v=-n+2}^{n-2}\sum_{u=-\infty}^{\infty}\widetilde{k}\left(\frac{u}{\widetilde{m}}\right)\widehat{\Gamma}_{\Delta x \Delta x}(u)\widetilde{k}\left(\frac{u-v}{\widetilde{m}}\right)\widehat{\Gamma}_{yy}(u-v)\left\{\int_0^\infty k^2(x)\,dx + o(1)\right\}.$$

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

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

which converges to $4\pi^2 f_{\Delta x\Delta x}(0)\, f_{yy}(0)\int_0^\infty k^2(x)\,dx$ in probability by the standard argument. A similar argument gives $\sum_{v=-n+2}^{n-2}\sum_{u=-\infty}^{\infty}\widetilde{k}((u+h)/\widetilde{m})\widehat{\Gamma}_{\Delta xy}(u+h)\widetilde{k}((u-h')/\widetilde{m})\widehat{\Gamma}_{y\Delta x}(u-h') \to_p 4\pi^2[f_{y\Delta x}(0)]^2\int_0^\infty k^2(x)\,dx$, and the stated result follows. ■

7.7 Proof of Lemma 3.5

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

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

From the arguments in the proof of Lemma 3.2, $\widetilde{V}$ is equal to

$$\left[\sum_{u=-\widetilde{m}}^{\widetilde{m}} \widetilde{k}\left(\frac{u}{\widetilde{m}}\right) \widehat{\Gamma}_{\Delta x \Delta x}(u) \sum_{v=-\widetilde{m}}^{\widetilde{m}} \widetilde{k}\left(\frac{v}{\widetilde{m}}\right) \widehat{\Gamma}_{yy}(v)\right] \left\{\int_0^\infty k^2(x)\, dx + o(1)\right\} \quad (44)$$

$$+\left[\sum_{u=-\widetilde{m}}^{\widetilde{m}} \widetilde{k}\left(\frac{u}{\widetilde{m}}\right) \widehat{\Gamma}_{\Delta x y}(u) \sum_{v=-\widetilde{m}}^{\widetilde{m}} \widetilde{k}\left(\frac{v}{\widetilde{m}}\right) \widehat{\Gamma}_{y \Delta x}(v)\right] \left\{\int_0^\infty k^2(x)\, dx + o(1)\right\} \quad (45)$$

and (45) is equal to

$$\left[\sum_{u=-\widetilde{m}}^{\widetilde{m}} \widetilde{k}\left(\frac{u}{\widetilde{m}}\right) \widehat{\Gamma}_{y \Delta x}(u)\right]^2 \left\{\int_0^\infty k^2(x)\, dx + o(1)\right\} \geq 0 \quad a.s.,$$

for sufficiently large n. For (44), because $\sum_{v=-\widetilde{m}}^{\widetilde{m}} \widetilde{k}(v/\widetilde{m})\widehat{\Gamma}_{yy}(v) \to_p f_y(0) > 0$ by the standard argument, (43) follows if there exists $\varepsilon > 0$ such that

$$\begin{aligned} & \Pr\left(\sum_{v=-\widetilde{m}}^{\widetilde{m}} \widetilde{k}\left(\frac{v}{\widetilde{m}}\right) \widehat{\Gamma}_{\Delta x \Delta x}(v) \geq \varepsilon \widetilde{m}^{-1}\right) \\ = \; & \Pr\left(2\pi \int_{-\pi}^{\pi} W_{\widetilde{m}}(\lambda)\, I_{\Delta x}(\lambda) d\lambda \geq \varepsilon \widetilde{m}^{-1}\right) \to 1, \quad \text{as } n \to \infty, \end{aligned} \quad (46)$$

where (Priestley, 1981, p. 439)

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

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

$$w_{\Delta x}(\lambda) = \left(1 - e^{i\lambda}\right) w_x(\lambda) + e^{i(n+1)\lambda} (2\pi n)^{-1/2} X_n.$$

It follows that

$$\begin{aligned} & \int_{-\pi}^{\pi} W_{\widetilde{m}}(\lambda)\, I_{\Delta x}(\lambda) d\lambda \\ = \; & \int_{-\pi}^{\pi} W_{\widetilde{m}}(\lambda)\, |1 - e^{i\lambda}|^2 I_x(\lambda) d\lambda & (47) \\ & + (2\pi n)^{-1/2} X_n \int_{-\pi}^{\pi} W_{\widetilde{m}}(\lambda)\, 2\mathrm{Re}\left[(1 - e^{i\lambda}) w_x(\lambda) e^{-i(n+1)\lambda}\right] d\lambda & (48) \\ & + \int_{-\pi}^{\pi} W_{\widetilde{m}}(\lambda)\, d\lambda (2\pi n)^{-1} X_n^2. & (49) \end{aligned}$$

We can ignore (49) because it is nonnegative. For (48), it follows from the Cauchy-Schwartz inequality and Lemma 8.7 (b) that

$$
\begin{aligned}
& \int_{-\pi}^{\pi} W_{\widetilde{m}}(\lambda) \, 2\mathrm{Re}\left[(1-e^{i\lambda})w_x(\lambda)e^{-i(n+1)\lambda}\right] d\lambda \\
\leq & \left(\int_{-\pi}^{\pi} W_{\widetilde{m}}(\lambda) \left|2\mathrm{Re}\left[(1-e^{i\lambda})w_x(\lambda)e^{-i(n+1)\lambda}\right]\right|^2 d\lambda\right)^{1/2} \left(\int_{-\pi}^{\pi} W_{\widetilde{m}}(\lambda)\, d\lambda\right)^{1/2} \\
= & O_p\left(\left(\int_{-\pi}^{\pi} W_{\widetilde{m}}(\lambda)\, \lambda^2 d\lambda\right)^{1/2}\right) = O_p(\widetilde{m}^{-1/2}),
\end{aligned}
$$

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

$$
\begin{aligned}
& \int_{-\pi}^{\pi} W_{\widetilde{m}}(\lambda) \, |1-e^{i\lambda}|^2 EI_x(\lambda) d\lambda \\
& + \int_{-\pi}^{\pi} W_{\widetilde{m}}(\lambda) \, |1-e^{i\lambda}|^2 \left(I_x(\lambda) - EI_x(\lambda)\right) d\lambda \\
= & A_1 + A_2.
\end{aligned}
$$

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, sufficiently large n (Hannan, Theorem 2, p. 248)

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

Therefore, in conjunction with Lemma 8.7 (a), we obtain

$$
A_1 \geq c_1 c_2 \int_{-D\pi}^{D\pi} W_{\widetilde{m}}(\lambda)\, \lambda^2 d\lambda \geq c_1 c_2 \kappa \widetilde{m}^{-1}, \quad \kappa > 0.
$$

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

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

Therefore,

$$
\begin{aligned}
E(A_2)^2 &= \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} W_{\widetilde{m}}(\lambda)\, W_{\widetilde{m}}(\lambda')\, |1-e^{i\lambda}|^2 |1-e^{i\lambda'}|^2 \mathrm{cov}\left(I_x(\lambda), I_x(\lambda')\right) d\lambda d\lambda' \\
&\leq C \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} W_{\widetilde{m}}(\lambda)\, W_{\widetilde{m}}(\lambda')\, \lambda^2 (\lambda')^2 \left|\mathrm{cov}\left(I_x(\lambda), I_x(\lambda')\right)\right| d\lambda d\lambda' \\
&= o(\widetilde{m}^{-2})
\end{aligned}
$$

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 8.7 (b), (50), and the dominated convergence theorem. Therefore, there exists $\eta' > 0$ such that (47)+(48)+(49)$\geq \eta' \widetilde{m}^{-1}$ with probability approaching one, and (46) and the stated result follow. ■

8 Appendix B: technical results

8.1 Lemma

Under the assumptions of Theorem 2.6,

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

8.2 Proof

Observe that

$$E\left(\sum_{t=1}^{h}(y_t\Delta x_{t-h}-Ey_t\Delta x_{t-h})\right)^2=\mathrm{var}\left(\sum_{t=1}^{h}y_t\Delta x_{t-h}\right)\le E\left(\sum_{t=1}^{h}y_t\Delta x_{t-h}\right)^2.$$

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 $\Gamma_{y\Delta x}(h)=Ey_t\Delta x_{t+h}$)

$$\begin{aligned}
& E\left(\sum_{t=1}^{h}y_t\Delta x_{t-h}\sum_{s=1}^{h}y_s\Delta x_{s-h}\right)\\
= & \sum_{t=1}^{h}\sum_{s=1}^{h}\Gamma_{y\Delta x}(h)\Gamma_{y\Delta x}(h)+\sum_{t=1}^{h}\sum_{s=1}^{h}\Gamma_{yy}(s-t)\Gamma_{\Delta x\Delta x}(s-t)\\
& +\sum_{t=1}^{h}\sum_{s=1}^{h}\Gamma_{y\Delta x}(s-h-t)\Gamma_{\Delta xy}(s-t+h)+\sum_{t=1}^{h}\sum_{s=1}^{h}k_{y\Delta xy\Delta x}(t,t-h,s,s-h)\\
= & h^2(\Gamma_{y\Delta x}(h))^2+\sum_{l=-h+1}^{h-1}(h-|l|)\Gamma_{yy}(l)\Gamma_{\Delta x\Delta x}(l)\\
& +\sum_{l=-h+1}^{h-1}(h-|l|)\Gamma_{y\Delta x}(l-h)\Gamma_{\Delta xy}(l+h)+\sum_{l=-h+1}^{h-1}(h-|l|)k_{y\Delta xy\Delta x}(0,-h,l,l-h).
\end{aligned}$$

The first term on the right is bounded by $(\sup_s s|\Gamma_{y\Delta x}(s)|)^2<\infty$. The second and third terms on the right are bounded by $h\sup_s||\Gamma(s)||\sum_{l=-\infty}^{\infty}||\Gamma(l)||\le Ch$. From (35), the fourth term on the right is bounded by $h\sum_{l=-\infty}^{\infty}\sum_{r=-\infty}^{\infty}|k_{y\Delta xy\Delta x}(0,-r,l,l-r)|\le Ch$, and the stated result follows. ■

8.3 Lemma

Under the assumptions of Theorem 2.6,

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

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

$$\begin{aligned} Z_t &= n^{-1/2}m^{-1/2}\sum_{h=1}^{n-1} k\left(\frac{h}{m}\right)\sum_{r=1}^{\infty}\varepsilon_{t-r}' f^{hr}(1)\varepsilon_t, \\ f^{hr}(1) &= \sum_{j=0}^{\infty}\left[(A_{j+r-h}^2)'A_j^1 + (A_{j+r}^1)'A_{j-h}^2\right], \end{aligned}$$

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

8.4 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 $\sum_{t=1}^n y_t\Delta x_{t-h}$ so that it can be approximated by a martingale. Express y_t and Δx_t as

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

where A_j^1 and A_j^2 are the first and second row of A_j, respectively. Observe that

$$\begin{aligned} y_t\Delta x_{t-h} &= A^1(L)\varepsilon_t A^2(L)\varepsilon_{t-h} \\ &= \sum_{j=0}^{\infty} A_j^1\varepsilon_{t-j}\sum_{k=0}^{\infty} A_k^2\varepsilon_{t-h-k} \\ &= \sum_{j=0}^{\infty} A_j^1\varepsilon_{t-j}A_{j-h}^2\varepsilon_{t-j} + \sum_{j=0}^{\infty} A_j^1\varepsilon_{t-j}\sum_{s=h,\neq j}^{\infty} A_{s-h}^2\varepsilon_{t-s}, \quad (s = h+k). \end{aligned}$$

Since $A_{j-h}^2\varepsilon_{t-j}$ is a scalar, the first term on the right is

$$\mathrm{tr}\left(\sum_{j=0}^{\infty}(A_{j-h}^2)'A_j^1\varepsilon_{t-j}\varepsilon_{t-j}'\right) = \mathrm{tr}\ \left(f^{h0}(L)\varepsilon_t\varepsilon_t'\right), \quad f^{h0}(L) = \sum_{j=0}^{\infty}(A_{j-h}^2)'A_j^1L^j = \sum_{j=0}^{\infty} f_j^{h0}L^j.$$

The second term on the right is, since $A_s^2 \equiv 0$ for $s < 0$,

$$\begin{aligned} &\sum_{j=0}^{\infty} A_j^1\varepsilon_{t-j}\sum_{s=0,\neq j}^{\infty} A_{s-h}^2\varepsilon_{t-s} \\ &= \mathrm{tr}\left(\sum_{j=0}^{\infty}\sum_{s=0,\neq j}^{\infty}(A_{s-h}^2)'A_j^1\varepsilon_{t-j}\varepsilon_{t-s}'\right) \end{aligned}$$

$$
\begin{aligned}
&= \operatorname{tr}\left(\sum_{j=0}^{\infty}\sum_{s=j+1}^{\infty}(A_{s-h}^2)'A_j^1\varepsilon_{t-j}\varepsilon_{t-s}'\right)+\operatorname{tr}\left(\sum_{j=0}^{\infty}\sum_{s=0}^{j-1}(A_{s-h}^2)'A_j^1\varepsilon_{t-j}\varepsilon_{t-s}'\right)\\
&= \operatorname{tr}\left(\sum_{j=0}^{\infty}\sum_{s=j+1}^{\infty}(A_{s-h}^2)'A_j^1\varepsilon_{t-j}\varepsilon_{t-s}'\right)+\operatorname{tr}\left(\sum_{s=0}^{\infty}\sum_{j=s+1}^{\infty}(A_j^1)'A_{s-h}^2\varepsilon_{t-s}\varepsilon_{t-j}'\right)\\
&= \operatorname{tr}\left(\sum_{j=0}^{\infty}\sum_{s=j+1}^{\infty}\left[(A_{s-h}^2)'A_j^1+(A_s^1)'A_{j-h}^2\right]\varepsilon_{t-j}\varepsilon_{t-s}'\right)\\
&= \operatorname{tr}\left(\sum_{j=0}^{\infty}\sum_{r=1}^{\infty}\left[(A_{j+r-h}^2)'A_j^1+(A_{j+r}^1)'A_{j-h}^2\right]\varepsilon_{t-j}\varepsilon_{t-j-r}'\right)\quad (r=s-j)\\
&= \operatorname{tr}\left(\sum_{r=1}^{\infty}f^{hr}(L)\varepsilon_t\varepsilon_{t-r}'\right),
\end{aligned}
$$

where

$$
f^{hr}(L)=\sum_{j=0}^{\infty}f_j^{hr}L^j,\quad f_j^{hr}=(A_{j+r-h}^2)'A_j^1+(A_{j+r}^1)'A_{j-h}^2.
$$

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

$$
y_t\Delta x_{t-h}=\operatorname{tr}\left(f^{h0}(L)\varepsilon_t\varepsilon_t'+\sum_{r=1}^{\infty}f^{hr}(L)\varepsilon_t\varepsilon_{t-r}'\right).
$$

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

$$
f^{hr}(L)=f^{hr}(1)-(1-L)\widetilde{f}^{hr}(L),\quad r=0,1,\dots,
$$

with

$$
\widetilde{f}^{hr}(L)=\sum_{j=0}^{\infty}\widetilde{f}_j^{hr}L^j,\quad \widetilde{f}_j^{hr}=\sum_{s=j+1}^{\infty}f_s^{hr}=\sum_{s=j+1}^{\infty}\left[(A_{s+r-h}^2)'A_s^1+(A_{s+r}^1)'A_{s-h}^2\right].\tag{51}
$$

It follows that

$$
\frac{1}{\sqrt{n}}\sum_{t=1}^{n}y_t\Delta x_{t-h}=\operatorname{tr}\left(f^{h0}(1)\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\varepsilon_t\varepsilon_t'+\sum_{r=1}^{\infty}f^{hr}(1)\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\varepsilon_t\varepsilon_{t-r}'\right)+r_{nh},\tag{52}
$$

where

$$
r_{nh}=\frac{1}{\sqrt{n}}\operatorname{tr}\left(\widetilde{f}^{h0}(L)(\varepsilon_0\varepsilon_0'-\varepsilon_n\varepsilon_n')\right)+\frac{1}{\sqrt{n}}\operatorname{tr}\left(\sum_{r=1}^{\infty}\widetilde{f}^{hr}(L)(\varepsilon_0\varepsilon_{-r}'-\varepsilon_n\varepsilon_{n-r}')\right).
$$

From Lemma 8.5, we have

$$
E|r_{nh}|^2\le Cn^{-1},\quad h=1,\dots,n-1.\tag{53}
$$

Furthermore, observe that

$$\begin{aligned}
Ey_t\Delta x_{t-h} &= E\left(\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}A_j^1\varepsilon_{t-j}\varepsilon_{t-k-h}'(A_k^2)'\right) \\
&= \sum_{j=0}^{\infty}A_j^1(A_{j-h}^2)' = \operatorname{tr}\left(\sum_{j=0}^{\infty}(A_{j-h}^2)'A_j^1\right) = \operatorname{tr}\left(f^{h0}(1)\right).
\end{aligned}$$

In conjunction with (52), it follows that

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(y_t\Delta x_{t-h} - Ey_t\Delta x_{t-h}\right) \\
=\ &\operatorname{tr}\left(f^{h0}(1)\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\varepsilon_t\varepsilon_t' - I_2\right) + \sum_{r=1}^{\infty}f^{hr}(1)\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\varepsilon_t\varepsilon_{t-r}'\right) + r_{nh},
\end{aligned}$$

and hence

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

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

$$\begin{aligned}
I &= \frac{1}{\sqrt{m}}\sum_{h=1}^{m}k\left(\frac{h}{m}\right)\operatorname{tr}\left(f^{h0}(1)\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\varepsilon_t\varepsilon_t' - I_2\right)\right) \\
&= \operatorname{tr}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\varepsilon_t\varepsilon_t' - I_2\right)\frac{1}{\sqrt{m}}\sum_{h=1}^{m}k\left(\frac{h}{m}\right)f^{h0}(1)\right) \\
II &= \frac{1}{\sqrt{m}}\sum_{h=1}^{m}k\left(\frac{h}{m}\right)\operatorname{tr}\left(\sum_{r=1}^{\infty}f^{hr}(1)\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\varepsilon_t\varepsilon_{t-r}'\right) \\
&= \sum_{t=1}^{n}Z_t; \quad Z_t = n^{-1/2}m^{-1/2}\sum_{h=1}^{n-1}k\left(\frac{h}{m}\right)\sum_{r=1}^{\infty}\varepsilon_{t-r}'f^{hr}(1)\varepsilon_t.
\end{aligned}$$

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 $j < 0$,

$$\begin{aligned}
||f^{h0}(1)|| &= \left\|\sum_{j=0}^{\infty}(A_{j-h}^2)'A_j^1\right\| = \left\|\sum_{j=h}^{\infty}(A_{j-h}^2)'A_j^1\right\| \\
&\le \sup_s ||A_s||\sum_{j=h}^{\infty}||A_j|| \le Ch^{-\delta}\sum_{j=h}^{\infty}j^{\delta}||A_j|| \le Ch^{-\delta}, \quad h = 1,\ldots,n-1.
\end{aligned}$$

Therefore, $||m^{-1/2}\sum_{h=1}^{m}k(h/m)f^{h0}(1)|| \le Cm^{-1/2}$, and it follows that $E(I)^2 = O(m^{-1})$, giving the stated result. ■

8.5 Lemma

Under the assumptions of Theorem 2.6, for $t = 0$, n and $h = 1, \ldots, n-1$,

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

8.6 Proof

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

$$\begin{aligned} E\left(\mathrm{tr}\left(\widetilde{f}^{h0}(L)\varepsilon_n\varepsilon_n'\right)\right)^2 &= \sum_{j=0}^{\infty}\sum_{k=0}^{\infty}E\left(\varepsilon_{n-j}'\widetilde{f}_j^{h0}\varepsilon_{n-j}\varepsilon_{n-k}'\widetilde{f}_k^{h0}\varepsilon_{n-k}\right) \\ &\leq C\left(\sum_{j=0}^{\infty}||\widetilde{f}_j^{h0}||\right)^2 + C\sum_{j=0}^{\infty}||\widetilde{f}_j^{h0}||^2. \end{aligned}$$

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

$$||\widetilde{f}_j^{h0}|| = \left\|\sum_{s=j+1}^{\infty}f_s^{h0}\right\| \leq \sum_{s=j+1}^{\infty}||(A_{s-h}^2)'A_s^1|| \leq \sup_r ||A_r||(j+1)^{-\delta}\sum_{s=j+1}^{\infty}s^{\delta}||A_s|| \leq Cj^{-\delta},$$

and $\delta > 1$.

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

$$\sum_{r=1}^{\infty}\sum_{j=0}^{\infty}\mathrm{tr}\left(\widetilde{f}_j^{hr}\varepsilon_{n-j}\varepsilon_{n-r-j}'\right) = \sum_{r=1}^{\infty}\sum_{j=0}^{\infty}\varepsilon_{n-j}'\left(\widetilde{f}_j^{hr}\right)'\varepsilon_{n-r-j} = \sum_{j=0}^{\infty}\xi_{n-j}^h,$$

where $\xi_{n-j}^h = \varepsilon_{n-j}'\sum_{r=1}^{\infty}(\widetilde{f}_j^{hr})'\varepsilon_{n-r-j}$. Since $\xi_{n-j}^h \in \mathcal{I}_{n-j} = \sigma(\varepsilon_{n-j}, \varepsilon_{n-j-1}, \ldots)$ and $E(\xi_{n-j}^h|\mathcal{I}_{n-j-1}) = 0$, it follows that

$$E\left(\sum_{j=0}^{\infty}\xi_{n-j}^h\right)^2 = \sum_{j=0}^{\infty}E(\xi_{n-j}^h)^2 \leq C\sum_{j=0}^{\infty}\sum_{r=1}^{\infty}||\widetilde{f}_j^{hr}||^2 \leq C\left(\sup_{j,r}||\widetilde{f}_j^{hr}||\right)\sum_{j=0}^{\infty}\sum_{r=1}^{\infty}||\widetilde{f}_j^{hr}||. \tag{54}$$

Now

$$||\widetilde{f}_j^{hr}|| = \left\|\sum_{s=j+1}^{\infty}f_s^{hr}\right\| = \left\|\sum_{s=j+1}^{\infty}(A_{s+r-h}^2)'A_s^1 + \sum_{s=j+1}^{\infty}(A_{s+r}^1)'A_{s-h}^2\right\|.$$

Hence $\sup_h\sup_{j,r}||\widetilde{f}_j^{hr}|| \leq \sup_p||A_p||\sum_{s=0}^{\infty}||A_s|| < \infty$. Furthermore, uniformly in $h = 1, \ldots, n-1$,

$$\begin{aligned} \sum_{j=0}^{\infty}\sum_{r=1}^{\infty}||\widetilde{f}_j^{hr}|| &\leq \sum_{j=0}^{\infty}\sum_{r=1}^{\infty}\sum_{s=j+1}^{\infty}||A_{s+r-h}||||A_s|| + \sum_{j=0}^{\infty}\sum_{r=1}^{\infty}\sum_{s=j+1}^{\infty}||A_{s+r}||||A_{s-h}|| \\ &\leq \sum_{j=0}^{\infty}\sum_{s=j+1}^{\infty}||A_s||\sum_{r=0}^{\infty}||A_r|| + \sum_{j=0}^{\infty}\sum_{s=j+1}^{\infty}||A_{s-h}||\sum_{r=0}^{\infty}||A_r||. \end{aligned} \tag{55}$$

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

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

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

8.7 Lemma

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

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

8.8 Proof

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

$$c(\lambda/2)^2 \leq \sin^2(\lambda/2) \leq (\lambda/2)^2. \tag{56}$$

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

$$\begin{aligned} \int_{-D\pi}^{D\pi} W_{\widetilde{m}}(\lambda)\,\lambda^2 d\lambda &\geq C\widetilde{m}^{-1}\int_{-D\pi}^{D\pi}\sin^2(\widetilde{m}\lambda/2)d\lambda \\ = 2C\widetilde{m}^{-2}\int_{-\widetilde{m}D\pi/2}^{\widetilde{m}D\pi/2}\sin^2(\theta)d\theta &\geq 2C\widetilde{m}^{-2}[\widetilde{m}D]\int_{-\pi/2}^{\pi/2}\sin^2(\theta)d\theta \\ &\sim 2CD\widetilde{m}^{-1}\int_{-\pi/2}^{\pi/2}\sin^2(\theta)d\theta \geq \kappa\widetilde{m}^{-1}, \end{aligned}$$

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

References

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance estimation. *Econometrica 59*, 817–858.

Ang, A. and G. Bekaert (2001). Stock return predictability: Is it there? NBER Working paper 8207.

Brown, B. M. (1971). Martingale central limit theorems. *Annals of Mathematical Statistics 42*, 59–66.

Campbell, B. and J.-M. Dufour (1995). Exact nonparametric orthogonality and random walk tests. *The Review of Economics and Statistics 77*(1), 1–16.

Campbell, B. and J.-M. Dufour (1997). Exact nonparametric tests of orthogonality and random walk in the presence of a drift parameter. *International Economic Review 38*(1), 151–173.

Campbell, J. (1991). A variance decomposition for stock returns. *Economic Journal 101*, 157–179.

Campbell, J., A. Lo, and C. MacKinlay (1997). *The Econometrics of Financial Markets*. Princeton, N.J.: Princeton University Press.

Campbell, J. and M. Yogo (2003). Efficient tests of stock return predictability. Mimeo, Department of Economics, Harvard University.

Campbell, J. Y. and R. J. Shiller (1988a). The dividend price ratio and expectations of future dividends and discount factors. *Review of Financial Studies 1*, 195–227.

Campbell, J. Y. and R. J. Shiller (1988b). Stock prices, earnings, and expected dividends. *Journal of Finance 43*(3), 661–76.

Cavanagh, C. L., G. Elliott, and J. Stock (1995). Inference in models with nearly integrated regressors. *Econometric Theory 11*, 1131–1147.

Dufour, J.-M. and D. Pelletier (2002). Linear estimation of weak VARMA models with a macroeconomic application. Mimeo, Universite de Montreal.

Elliott, G. (1998). On the robustness of cointegration methods when regressors have almost unit roots. *Econometrica 66*(1), 149–158.

Elliott, G. and J. Stock (1994). Inference in time series regression when the order of integration of a regressor is unknown. *Econometric Theory 10*, 672–700.

Fama, E. and K. French (1988). Permanent and temporary components of stock prices. *Journal of Political Economy 96*(2), 246–273.

Goetzmann, W. and P. Jorion (1993). Testing the predictive power of dividend yields. *Journal of Finance 48*(2), 663–670.

Hannan, E. J. (1970). *Multiple time series. New York, Wiley.* New York, N.Y.: Wiley.

Hodrick, R. (1992). Dividend yields and expected stock returns: Alternative procedures for inference and measurement. *Review of Financial Studies 5*, 357–368.

Hurwitz, L. (1950). Least squares bias in time series. In T. Koopmans (Ed.), *Statistical Inference in Dynamic Economic Models, Cowles Commission Monograph Number 10.* Wiley.

Kitamura, Y. and P. C. B. Phillips (1997). Fully modified IV, GIVE and GMM estimation with possibly nonstationary regressors and instruments. *Journal of Econometrics 80*, 85–123.

Lanne, M. (2002). Testing the predictability of stock returns. *The Review of Economics and Statistics 84*(3), 407–415.

Lewellen, J. (2003). Predicting returns with financial ratios. *Journal of Financial Economics*. forthcoming.

Mankiw, N. and M. Shapiro (1986). Do we reject too often? Small sample properties of tests of rational expectations models. *Economics Letters 20*, 139–145.

Maynard, A. (2004). Persistence, bias, and the forward premium anomaly: How bad is it? Mimeo, University of Toronto.

Nelson, C. and M. Kim (1993). Predictable stock returns: The role of small sample bias. *Journal of Finance 48*, 641–661.

Newey, W. K. and K. D. West (1987). A simple positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*, 703–708.

Ng, S. and P. Perron (2001). Lag length selection and the construction of unit root tests with good size and power. *Econometrica 69*(6), 1519–1554.

Phillips, P. C. B. (1999). Discrete fourier transforms of fractional processes. Cowles Foundation Discussion Paper #1243, Yale University.

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

Phillips, P. C. B. and V. Solo (1992). Asymptotics for linear processes. *Annals of Statistics 20(2)*, 971–1001.

Priestley, M. B. (1981). *Spectral Analysis and Time Series*. New York, N.Y.: New York, Academic Press.

Richardson, M. and J. Stock (1989). Drawing inferences from statistics based on multi-year asset returns. *Journal of Financial Economics 25*, 323–348.

Rudebusch, G. D. (1992). Trends and random walks in macroeconomic time series: A reexamination. *International Economic Review 33*, 661–680.

Shiller, R. (1984). Stock prices and social dynamics. *Brookings papers on Economic Activity 2*, 457–498.

Stambaugh, R. (1986). Bias in regressions with lagged stochastic regressors. CRSP Working Paper 156, University of Chicago.

Stambaugh, R. (1999). Predictive regressions. *Journal of Financial Economics 54*, 375–421.

Stock, J. (1991). Confidence intervals for the largest autoregressive root in U.S. economic time series. *Journal of Monetary Economics 28*(3), 435–460.

Torous, W., R. Valkanov, and S. Yan (2005). On predicting stock returns with nearly integrated explanatory variables. *Journal of Business 78*(1). forthcoming.

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

Valkanov, R. (2003). Long-horizon regressions: theoretical results and applications. *Journal of Financial Economics 68*, 201–232.

Viceira, L. (1997). Testing for structural change in the predictability of asset returns. Mimeo, Harvard University.

Wolf, M. (2000). Stock returns and dividend yields revisited: A new way to look at an old problem. *Journal of Business and Economic Statistics 18*(1), 18–30.

Table 1: Regression t-statistic: Finite Sample size (local to unity)

c	$\rho_1+\rho_2$	$\sigma_{12}=0$	0.25	0.50	0.75	0.95	$\sigma_{12}=0$	0.25	0.50	0.75	0.95
		Demeaned Case ($n=100$)					Detrended Case ($n=100$)				
0	1.000	0.054	0.062	0.109	0.204	0.267	0.052	0.087	0.177	0.351	0.549
-1	0.990	0.057	0.071	0.104	0.161	0.220	0.059	0.087	0.159	0.300	0.463
-5	0.950	0.056	0.057	0.068	0.087	0.114	0.047	0.077	0.116	0.166	0.237
-10	0.900	0.054	0.052	0.066	0.078	0.080	0.057	0.058	0.083	0.117	0.162
-20	0.800	0.060	0.060	0.043	0.058	0.065	0.053	0.060	0.064	0.082	0.106
		Demeaned Case ($n=400$)					Detrended Case ($n=400$)				
0	1.000	0.052	0.070	0.109	0.187	0.270	0.046	0.087	0.172	0.367	0.565
-1	0.998	0.050	0.059	0.087	0.158	0.216	0.052	0.080	0.145	0.281	0.466
-5	0.988	0.049	0.049	0.066	0.092	0.126	0.058	0.063	0.093	0.165	0.247
-10	0.975	0.049	0.051	0.062	0.073	0.088	0.055	0.066	0.087	0.117	0.158
-20	0.950	0.057	0.056	0.065	0.060	0.072	0.052	0.054	0.066	0.088	0.093

The table shows rejection rates under the null hypothesis for a nominal 5% test using t_β. y_t is given by (20) and x_t by (17) with ρ_1 given by (19). Details are given in the text.

Table 2: Covariance based t-statistic: Finite Sample size (local to unity)

c	$\rho_1+\rho_2$	$\sigma_{12}=0$	0.25	0.50	0.75	0.95	$\sigma_{12}=0$	0.25	0.50	0.75	0.95
		Demeaned Case ($n=100$)					Detrended Case ($n=100$)				
0	1.000	0.034	0.033	0.042	0.052	0.067	0.037	0.037	0.055	0.067	0.087
-1	0.990	0.032	0.027	0.033	0.037	0.042	0.036	0.036	0.049	0.052	0.079
-5	0.950	0.033	0.032	0.037	0.034	0.040	0.032	0.041	0.040	0.050	0.061
-10	0.900	0.036	0.030	0.040	0.036	0.037	0.030	0.033	0.041	0.042	0.051
-20	0.800	0.028	0.026	0.032	0.036	0.033	0.026	0.028	0.033	0.033	0.043
		Demeaned Case ($n=400$)					Detrended Case ($n=400$)				
0	1.000	0.034	0.036	0.036	0.044	0.048	0.041	0.038	0.036	0.052	0.053
-1	0.998	0.034	0.034	0.044	0.047	0.050	0.038	0.036	0.039	0.050	0.054
-5	0.988	0.037	0.033	0.046	0.042	0.044	0.031	0.034	0.041	0.045	0.053
-10	0.975	0.041	0.032	0.044	0.043	0.038	0.037	0.035	0.034	0.049	0.048
-20	0.950	0.035	0.036	0.036	0.041	0.032	0.032	0.033	0.042	0.042	0.035

The table shows rejection rates under the null hypothesis for a nominal 5% test using t_λ. y_t is given by (20) and x_t by (17) with ρ_1 given by (19). Details are given in the text.

Table 3: Covariance based t-statistic: Finite Sample size (AR(2))

c	$\rho_1 + \rho_2$	σ_{12} =0	0.25	0.50	0.75	0.95	σ_{12} =0	0.25	0.50	0.75	0.95
		Demeaned Case ($n = 100$)					Detrended Case ($n = 100$)				
0	1.000	0.046	0.050	0.054	0.048	0.061	0.051	0.061	0.072	0.073	0.092
-1	0.990	0.044	0.060	0.067	0.060	0.066	0.061	0.046	0.062	0.060	0.089
-5	0.950	0.053	0.048	0.060	0.060	0.056	0.042	0.061	0.054	0.056	0.057
-10	0.900	0.051	0.056	0.058	0.056	0.043	0.060	0.057	0.057	0.074	0.045
-20	0.800	0.047	0.060	0.051	0.060	0.052	0.053	0.058	0.067	0.064	0.065
		Demeaned Case ($n = 400$)					Detrended Case ($n = 400$)				
0	1.000	0.064	0.058	0.059	0.059	0.061	0.050	0.058	0.057	0.065	0.068
-1	0.998	0.056	0.053	0.056	0.060	0.070	0.060	0.061	0.059	0.063	0.064
-5	0.988	0.067	0.060	0.052	0.055	0.055	0.058	0.060	0.062	0.064	0.059
-10	0.975	0.062	0.056	0.052	0.062	0.060	0.054	0.057	0.065	0.052	0.060
-20	0.950	0.052	0.063	0.064	0.058	0.056	0.074	0.058	0.054	0.060	0.065

The table shows rejection rates under the null hypothesis for a nominal 5% test using t_λ. y_t is given by (20) and x_t by (18) with ρ_1 and ρ_2 given by (26). Details are given in the text.

Table 4: Covariance based t-statistic: Finite Sample power ($y_t = \beta x_{t-1} + \varepsilon_{1,t}$)

c	σ_{12}	β =0.15	0.20	0.35	0.50	0.75	1.00
		A. Detrended Case ($n = 100$)					
$c = 0$	0.000	0.115	0.226	0.745	0.975	1.000	1.000
($\rho_1 = 1.000$)	0.500	0.105	0.202	0.684	0.961	1.000	1.000
	0.950	0.071	0.153	0.656	0.968	1.000	1.000
$c = -1$	0.000	0.210	0.349	0.636	0.781	0.805	0.831
($\rho_1 = 0.990$)	0.500	0.135	0.268	0.564	0.671	0.759	0.760
	0.950	0.060	0.118	0.341	0.536	0.678	0.730
$c = -2.5$	0.000	0.207	0.339	0.667	0.790	0.843	0.854
($\rho_1 = 0.975$)	0.500	0.150	0.252	0.586	0.714	0.792	0.817
	0.950	0.074	0.124	0.383	0.573	0.739	0.780
$c = -7.5$	0.000	0.226	0.349	0.750	0.910	0.936	0.948
($\rho_1 = 0.925$)	0.500	0.171	0.267	0.671	0.837	0.914	0.915
	0.950	0.077	0.144	0.439	0.661	0.850	0.887
$c = -20$	0.000	0.220	0.365	0.794	0.962	0.997	0.999
($\rho_1 = 0.800$)	0.500	0.138	0.266	0.690	0.922	0.988	0.993
	0.950	0.049	0.122	0.477	0.809	0.968	0.983
		B. Detrended Case ($n = 400$)					
$c = 0$	0.000	0.594	0.881	1.000	1.000	1.000	1.000
($\rho_1 = 1.000$)	0.500	0.559	0.863	1.000	1.000	1.000	1.000
	0.950	0.539	0.863	1.000	1.000	1.000	1.000
$c = -1$	0.000	0.747	0.824	0.850	0.857	0.853	0.837
($\rho_1 = 0.998$)	0.500	0.730	0.800	0.846	0.831	0.849	0.836
	0.950	0.700	0.775	0.816	0.840	0.842	0.831
$c = -2.5$	0.000	0.774	0.867	0.896	0.884	0.895	0.887
($\rho_1 = 0.994$)	0.500	0.754	0.828	0.864	0.879	0.882	0.870
	0.950	0.735	0.811	0.831	0.860	0.879	0.871
$c = -7.5$	0.000	0.841	0.949	0.968	0.965	0.970	0.976
($\rho_1 = 0.981$)	0.500	0.843	0.939	0.964	0.963	0.971	0.976
	0.950	0.822	0.905	0.948	0.959	0.965	0.962
$c = -20$	0.000	0.831	0.966	0.998	0.999	0.999	0.998
$\rho_1 = 0.950$)	0.500	0.859	0.976	0.996	0.997	0.998	0.997
	0.950	0.864	0.959	0.992	0.991	1.000	1.000

The table shows rejection rates under the alternative hypothesis for a nominal 5% test using t_λ. y_t is given by (23) and x_t by (17) with ρ_1 given by (19). Details are given in the text.

Table 5: Covariance based t-statistic: Finite Sample power ($y_t = \beta(1-\rho L)x_{t-1}+\varepsilon_{1,t}$)

c	σ_{12}	$\beta = 0.15$	0.20	0.35	0.50	0.75	1.00
(ρ_1)		r^2 =0.02	0.04	0.11	0.20	0.36	0.50
		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
	0.950	0.109	0.191	0.634	0.913	0.998	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
	0.950	0.106	0.204	0.640	0.921	0.998	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
	0.950	0.128	0.208	0.642	0.932	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
	0.950	0.074	0.168	0.639	0.931	1.000	0.999
$c = -20$	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
	0.950	0.025	0.074	0.442	0.870	0.998	1.000
		B. Detrended Case ($n = 400$)					
$c = 0$	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
	0.950	0.581	0.845	0.998	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
	0.950	0.606	0.832	0.998	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
	0.950	0.598	0.843	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
	0.950	0.597	0.833	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
	0.950	0.490	0.771	0.999	1.000	1.000	1.000

The table shows rejection rates under the alternative hypothesis for a nominal 5% test using t_λ. y_t is given by (24) and x_t by (17) with ρ_1 given by (19). Details are given in the text.

Table 6: CES Bonferroni Method: Finite Sample power ($y_t = \beta x_{t-1} + \varepsilon_{1,t}$)

c	σ_{12}	β =0.15	0.20	0.35	0.50	0.75	1.00
		A. Demeaned Case ($n = 100$)					
$c = 0$	0.000	0.327	0.503	0.938	0.999	1.000	1.000
($\rho_1 = 1.000$)	0.500	0.387	0.600	0.945	0.999	1.000	1.000
	0.950	0.404	0.587	0.963	1.000	1.000	1.000
$c = -1$	0.000	0.955	0.989	1.000	1.000	1.000	1.000
($\rho_1 = 0.990$)	0.500	0.913	0.967	1.000	1.000	1.000	1.000
	0.950	0.848	0.928	0.998	1.000	1.000	1.000
$c = -2.5$	0.000	0.929	0.988	1.000	1.000	1.000	1.000
($\rho_1 = 0.975$)	0.500	0.883	0.974	1.000	1.000	1.000	1.000
	0.950	0.829	0.928	0.996	1.000	1.000	1.000
$c = -7.5$	0.000	0.863	0.966	1.000	1.000	1.000	1.000
($\rho_1 = 0.925$)	0.500	0.814	0.936	1.000	1.000	1.000	1.000
	0.950	0.763	0.893	0.997	1.000	1.000	1.000
$c = -20$	0.000	0.628	0.853	0.999	1.000	1.000	1.000
($\rho_1 = 0.800$)	0.500	0.636	0.816	0.994	1.000	1.000	1.000
	0.950	0.598	0.807	0.987	1.000	1.000	1.000
		B. Demeaned Case ($n = 400$)					
$c = 0$	0.000	0.840	0.984	1.000	1.000	1.000	1.000
($\rho_1 = 1.000$)	0.500	0.882	0.988	1.000	1.000	1.000	1.000
	0.950	0.891	0.990	1.000	1.000	1.000	1.000
$c = -1$	0.000	1.000	1.000	1.000	1.000	1.000	1.000
($\rho_1 = 0.998$)	0.500	1.000	1.000	1.000	1.000	1.000	1.000
	0.950	1.000	1.000	1.000	1.000	1.000	1.000
$c = -2.5$	0.000	1.000	1.000	1.000	1.000	1.000	1.000
($\rho_1 = 0.994$)	0.500	1.000	1.000	1.000	1.000	1.000	1.000
	0.950	1.000	1.000	1.000	1.000	1.000	1.000
$c = -7.5$	0.000	1.000	1.000	1.000	1.000	1.000	1.000
($\rho_1 = 0.981$)	0.500	1.000	1.000	1.000	1.000	1.000	1.000
	0.950	1.000	1.000	1.000	1.000	1.000	1.000
$c = -20$	0.000	1.000	1.000	1.000	1.000	1.000	1.000
($\rho_1 = 0.950$)	0.500	1.000	1.000	1.000	1.000	1.000	1.000
	0.950	1.000	1.000	1.000	1.000	1.000	1.000

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 ρ_1 given by (19). Details are given in the text.

Table 7: CES Bonferroni Method: Finite Sample power ($y_t = \beta(1-\rho L)x_{t-1} + \varepsilon_{1,t}$)

c (ρ_1)	σ_{12}	$\beta = 0.15$ r^2 =0.02	0.20 0.04	0.35 0.11	0.50 0.20	0.75 0.36	1.00 0.50
		A. Demeaned Case ($n = 100$)					
$c = 0$	0.000	0.075	0.087	0.138	0.213	0.291	0.369
($\rho_1 = 1.000$)	0.500	0.039	0.042	0.067	0.098	0.137	0.201
	0.950	0.034	0.041	0.035	0.051	0.071	0.099
$c = -1$	0.000	0.085	0.090	0.178	0.270	0.386	0.519
($\rho_1 = 0.990$)	0.500	0.047	0.059	0.081	0.137	0.238	0.343
	0.950	0.032	0.025	0.043	0.062	0.119	0.180
$c = -2.5$	0.000	0.078	0.137	0.214	0.334	0.537	0.696
($\rho_1 = 0.975$)	0.500	0.071	0.072	0.149	0.227	0.376	0.550
	0.950	0.038	0.044	0.074	0.149	0.249	0.362
$c = -7.5$	0.000	0.101	0.156	0.343	0.544	0.800	0.934
($\rho_1 = 0.925$)	0.500	0.103	0.138	0.308	0.461	0.755	0.911
	0.950	0.082	0.113	0.235	0.381	0.653	0.857
$c = -20$	0.000	0.146	0.247	0.580	0.837	0.985	0.999
($\rho_1 = 0.800$)	0.500	0.150	0.257	0.543	0.839	0.987	1.000
	0.950	0.146	0.234	0.519	0.787	0.994	1.000
		B. Demeaned Case ($n = 400$)					
$c = 0$	0.000	0.074	0.088	0.150	0.197	0.307	0.387
($\rho_1 = 1.0$)	0.500	0.038	0.038	0.066	0.089	0.141	0.218
	0.950	0.034	0.038	0.030	0.044	0.060	0.098
$c = -1$	0.000	0.073	0.096	0.184	0.299	0.399	0.534
($\rho_1 = 0.998$)	0.500	0.049	0.050	0.084	0.137	0.247	0.345
	0.950	0.029	0.026	0.040	0.073	0.133	0.193
$c = -2.5$	0.000	0.067	0.115	0.240	0.362	0.577	0.738
($\rho_1 = 0.994$)	0.500	0.058	0.073	0.141	0.232	0.420	0.568
	0.950	0.040	0.048	0.098	0.145	0.257	0.393
$c = -7.5$	0.000	0.123	0.162	0.381	0.617	0.858	0.964
($\rho_1 = 0.981$)	0.500	0.088	0.119	0.294	0.516	0.784	0.927
	0.950	0.097	0.103	0.224	0.388	0.688	0.915
$c = -20$	0.000	0.182	0.273	0.640	0.886	0.994	1.000
($\rho_1 = 0.950$)	0.500	0.150	0.253	0.583	0.863	0.992	1.000
	0.950	0.162	0.221	0.511	0.819	0.992	1.000

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 ρ_1 given by (19). Details are given in the text.

Table 8: Regressions of k-period long-horizon real stock returns on the treasury bill and dividend price ratio

		Treasury Bills				Dividend Price Ratio			
		Forecast Horizon (k)				Forecast Horizon (k)			
sample		k = 1.0	3.0	12.0	24.0	k = 1.0	3.0	12.0	24.0
1927-	$\widehat{\beta}$	-0.702	-1.906	-3.014	-1.388	0.016	0.043	0.200	0.386
1994	R^2	0.001	0.002	0.002	0.000	0.007	0.014	0.073	0.143
	t_β	-0.962	-0.964	-0.402	-0.136	2.389	1.598	2.658	4.221
1927-	$\widehat{\beta}$	-1.224	-6.179	-27.898	-106.2	0.024	0.054	0.304	0.667
1951	R^2	0.000	0.002	0.011	0.089	0.007	0.011	0.086	0.217
	t_β	-0.300	-0.528	-0.698	-1.277	1.472	0.886	2.134	3.796
1952-	$\widehat{\beta}$	-1.343	-3.497	-6.114	-0.993	0.027	0.080	0.327	0.579
1994	R^2	0.006	0.013	0.009	0.000	0.018	0.049	0.188	0.322
	t_β	-1.785	-1.669	-0.726	-0.082	3.098	3.728	3.845	3.589

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, using the Bartlett (Newey-West) kernel with bandwidth set to $k - 1$.

Table 9: Confidence intervals on largest roots and residual correlation

	$x_t = \mu_x + v_t$			$(1-\alpha L)b(L)v_t = \varepsilon_{2,t}$		
	$y_t = \beta_0 + \beta_1 x_{t-1} + \varepsilon_{1,t}$			$\delta = corr\,(\varepsilon_{1,t}, \varepsilon_{2,t})$		
	Treasury Bills ($y_t = r_t \quad x_t = i_t$)			Dividend Price Ratio ($y_t = r_t \quad x_t = d_t - p_t$)		
sample period	95 % CI on largest root in x_t		$\widehat{\delta}$	95 % CI on largest root in x_t		$\widehat{\delta}$
1927 to 1994	(0.9836	1.004)	-0.0768	(0.9623	0.998)	-0.9615
1927 to 1951	(0.9421	1.010)	0.1169	(0.9273	1.007)	-0.9601
1952 to 1994	(0.9680	1.006)	-0.3036	(0.9543	1.003)	-0.9754

Confidence intervals on the largest root are based on **?** (**?**) using the **?** (**?**) MIC criteria to select lag-length with a maximum of six lags.

Table 10: Covariance-based orthogonality tests on k-period long-horizon real stock returns using the treasury bill and dividend price ratio

		Treasury Bills				Dividend Price Ratio			
		Forecast Horizon (k)				Forecast Horizon (k)			
sample		k = 1.0	3.0	12.0	24.0	k = 1.0	3.0	12.0	24.0
1927-	t_λ	-0.3934	-2.2120	-0.5063	0.4190	0.5381	-0.1317	1.6114	0.6324
1994	m^*	1.9371	3.2009	21.3161	12.2515	0.7306	5.2664	101.5913	11.7834
1927-	t_λ	0.5946	-0.6085	0.4691	1.7341	0.1987	0.5993	0.2207	-0.0952
1951	m^*	0.7821	3.9311	3.2306	17.8799	0.1565	4.1292	48.7965	15.9576
1952-	t_λ	-1.1613	-2.4188	-1.8863	-0.8217	0.7235	-0.9962	0.7878	0.9637
1994	m^*	1.7534	5.7159	25.6198	16.1262	0.3661	3.8794	5.4009	52.7930

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

Table 11: Regression of long horizon real stock returns on stochastically detrended one-month treasury bill rates

		Forecast Horizon (k)					Forecast Horizon (k)			
		$k = 1.0$	3.0	12.0	24.0		$k = 1.0$	3.0	12.0	24.0
$\hat{\beta}$	1927-	-5.468	-17.181	-41.663	-4.492	1927-	3.144	-6.183	73.712	158.98
R^2	1994	0.005	0.016	0.023	0.000	1951	0.000	0.000	0.012	0.031
t_β		-2.119	-2.888	-1.840	-0.156		0.304	-0.183	0.618	1.073
$\hat{\beta}$	1952-	-6.547	-18.621	-56.406	-26.115					
R^2	1994	0.019	0.047	0.103	0.013					
t_β		-3.182	-3.597	-3.055	-1.245					

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 $k - 1$.