Covariance-based orthogonality tests for regressors with unknown persistence
(Preliminary and Incomplete)

Alex Maynard and Katsumi Shimotsu *
Department of Economics, University of Toronto and
Department of Economics, Queen's University

February 20, 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. On the other hand, while often overturning these rejections, recently employed size-corrected regression-based tests may restrict power to alternatives that become less plausible the more persistent the regressor. The covariance-based tests, which have correct size without restricting power, also show considerably weaker evidence against orthogonality than do traditional regressions. Nevertheless, even allowing for near-unit root behavior, in many cases we still reject orthogonality at long horizons using the dividend yield and at short to medium horizons using the one-month treasury bill rate.

1 Introduction

This paper develops a new covariance-based method for testing orthogonality when the conditioning variable has a root close or equal to unity. This new method provides

*Both authors are presenting authors and correspondence may be sent to either. We thank John Chao, Jonas Fisher, John Galbraith, John Maheu, Angelo Melino, Sam Thompson, and participants at the CESG, MESG, and Yale Econometrics Prospectus Workshop 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.

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 non-trivial 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}

$$y_t = \beta_0 + \beta_1 x_{t-1} + \varepsilon_{2t}, \tag{1}$$

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

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

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

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

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 rates and the regression of excess foreign currency returns on the lagged forward premium. Although the returns in such regressions generally show little persistence, the regressors are 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 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 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 Mankiw and Shapiro (1986), Cavanagh et al. (1995), and Stambaugh (1999). It results from a Dickey-Fuller type bias in the nonstandard distribution of the estimator $\hat{\beta}_1$ when $\sigma_{12} \neq 0$ and ρ_1 is close to one.[1] 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

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

one (Cavanagh et al. (1995), Elliot (1998)).[2] 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 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, $\beta_1 \neq 0$ in (1) provides a reasonable alternative hypothesis. However, if x_t is local to unity or a unit root process, this will only constitute an appropriate alternative if the process for y_t (e.g. returns) is thought to be equally persistent, a possibility which may often lack empirical support. Note for example, that if y_t is stationary, but x_t has a unit root, then β_1 must equal 0 and y_t must be i.i.d. in order for (1) to hold true. Yet, this merely implies that the regression specification is "unbalanced" and does not imply orthogonality in the more general sense. And it is not clear if the tests of orthogonality based on β_1 in (1) have any reasonable power, because setting $\beta_1 = 0$ is not sufficient to provide a correct dgp under the null hypothesis. A more reasonable test of orthogonality to consider in this context would be a regression of y_t on the pre-filtered version of x_{t-1} as in

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

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 this may come at the expense of complicating the size problem. 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, 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

$$\text{cov}(y_t, x_{t-1}) = \sum_{h=1}^{\infty} \text{cov}(y_t, \Delta x_{t-h}). \tag{3}$$

[2] more appropriate corrections based on first stage confidence intervals for the local to unity parameter.

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[3]

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

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

$$\lambda_{y,\Delta x} := \lim_{t\to\infty} \sum_{h=1}^{t-1} \text{cov}\left(y_t, \Delta x_{t-h}\right) = \sum_{h=1}^{\infty} \text{cov}\left(y_t, \Delta x_{t-h}\right), \tag{4}$$

which is well-defined if $\sum_{h=0}^{\infty} |\text{cov}(y_t, \Delta x_{t-h})| < \infty$. As seen from (3), when x_t is stationary, the pseudo-covariance is written as

$$\lambda_{y,\Delta x} = \text{cov}\left(y_t, x_{t-1}\right). \tag{5}$$

Therefore, $\lambda_{y,\Delta x}$ is well-defined both when x_t is $I(1)$ and $I(0)$ and provides an 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 $\sum_{p=1}^{\infty} p|\text{cov}(y_t, u_{t-p})| < \infty$, the pseudo-covariance takes the form

$$\lambda_{y,\Delta x} = \sum_{h=1}^{\infty} \text{cov}(y_t, u_{t-h}) + O\left(n^{-1}\right), \tag{6}$$

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

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 $\beta_1 \neq 0$ both y_t and x_t

[3]This concept, expanded upon here, was first given in Maynard (2002).

have 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. $\gamma \neq 0$) for which tests based on β_1 can not provide power. For example, although no longer infinite, $\lambda_{y,\Delta x} \neq 0$ also holds for $\gamma \neq 0$.

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 a provide single estimator for $\lambda_{y,\Delta x}$ without the necessity of pretesting or estimating ρ_1. The feature appears to be particularly useful when empirical researchers have to conduct econometric analysis when the variables in question could be $I(0)$ or $I(1)$, for instance testing the efficient market hypothesis based on the current period stock return and past dividend or labor-productivity issues (Fisher, 2002, Christiano, Eichenbaum and Vigfussen, 2002).

A second useful feature 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 conservative 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 in dividend-yields have often been 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 are able to allow for alternatives that leave returns stationary, while still violating orthogonality. The covariance-based tests also show considerably weaker evidence against orthogonality than do traditional regressions. Nevertheless, even allowing for near-unit root behavior, in many cases we still reject orthogonality at long horizons using the dividend yield and at short to medium horizons using the one-month treasury bill rate.

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 are all differ substantially approach and most address only the issue of size. Beginning with Cavanagh et al. (1995), several papers (Torous, Valkanov, and Yan (01), Valkanov (2003), and Campbell and Yogo (2003)) employ local to unity asymptotics to provide size corrections for regression based tests. Cavanagh et al (1995) for example, provide critical vales using two-stage bounds Bonferroni and Scheffe type bounds procedures. Stambaugh (1999) and Lewellen (2002) give finite sample corrections to regression based tests under more restrictive assumptions, while Elliott and Stock (1994) and Stambaugh (1999) consider Bayesian approaches. With suitable (strictly exogenous) instruments, the FM-IV estimator of Kitamura and Phillips (1997) 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. However, they find proper specification of the process for x_t still has important power implications, and independence assumptions on y_t may restrict their use in test with long-horizon returns. Bootstrap and subsampling approaches have also been employed under the assumption of a fixed root less than unity (Nelson and Kim (1993), Goetzmann and Jorion (1993) and Wolf (2000)).

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. 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 pseudo-covariance

In this section, we develop an estimator of the pseudo-covariance and derive its asymptotic properties. First we state the assumptions.

Assumption A

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

$$\begin{aligned} z_t &= \begin{pmatrix} y_t \\ \Delta x_t \end{pmatrix} = A(L)\,\varepsilon_t = \sum_{j=0}^{\infty} A_j \varepsilon_{t-j}, \quad \sum_{j=0}^{\infty} j\|A_j\| < \infty, \qquad (7) \\ \varepsilon_t &\sim iid\,(0, I_2), \quad \text{with finite fourth moment,} \\ \sum_{h=-\infty}^{\infty} |h|^q \,\|\Gamma(h)\| &< \infty, \quad q \geq 1; \quad \Gamma(h) = \begin{bmatrix} \Gamma_{yy}(h) & \Gamma_{y\Delta x}(h) \\ \Gamma_{\Delta xy}(h) & \Gamma_{\Delta x\Delta x}(h) \end{bmatrix} = E z_t z_{t+h}', \end{aligned}$$

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

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

$$\widehat{\lambda}_{y,\Delta x} = \sum_{h=1}^{n-1} k\left(\frac{h}{m}\right) \widehat{\Gamma}_{\Delta xy}(h); \quad \widehat{\Gamma}_{\Delta xy}(h) = \frac{1}{n} \sum_{t=h+1}^{n} y_t \Delta x_{t-h}, \qquad (8)$$

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

Assumption K

The kernel $k(x)$ is continuous and uniformly bounded with $k(0) = 1$, $\int_0^\infty |k(x)|x^{1/2}dx < \infty$, $\int_0^\infty k^2(x)dx < \infty$ and

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

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

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

2.3 Corollary

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

2.4 Remarks

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

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

   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

$$\begin{aligned} m^* &= \left(qk_q^2 \alpha(q) n \Big/ \int_{-\infty}^{\infty} k^2(x)\, dx \right)^{1/(2q+1)}, \qquad (9) \\ \alpha(q) &= \frac{4\left((2\pi)^{-1} \sum_{h=1}^{\infty} \Gamma_{\Delta xy}(h) h^q\right)^2}{f_{yy}(0) f_{\Delta x \Delta x}(0) + [f_{y\Delta x}(0)]^2}, \end{aligned}$$

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 Maynard (2002) shows that $\sum_{h=0}^{\infty} h^2 |\Gamma_{y\Delta x}(h)| < \infty$, 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 is 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* Phillips and Solo (1992) 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 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

$$\begin{aligned} v_t &= \begin{pmatrix} y_t \\ x_t \end{pmatrix} = B(L)\varepsilon_t = \sum_{j=0}^{\infty} B_j \varepsilon_{t-j}, \quad \sum_{j=0}^{\infty} j||B_j|| < \infty, \qquad (10) \\ \varepsilon_t &\sim iid(0, I_2), \quad \text{with finite fourth moment} \\ \sum_{-\infty}^{\infty} |h|^q \|\gamma(h)\| &< \infty, \quad \gamma(h) = \begin{bmatrix} \gamma_{yy}(h) & \gamma_{yx}(h) \\ \gamma_{xy}(h) & \gamma_{xx}(h) \end{bmatrix} = E v_t v_{t+h}', \end{aligned}$$

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*

$$\sqrt{n}\left(\widehat{\lambda}_{y,\Delta x} - \lambda_{y,\Delta x}\right) = k(1/m)\sqrt{n}\left(\widehat{\gamma}_{xy}(1) - \gamma_{xy}(1)\right) + B_n + o_p(1), \qquad (11)$$

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 orthogonarity between y_t and $I_{x,t-1}$, then m needs to sastisfy 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 (11).

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

Let x_t be a local-to-unity process:

$$x_t = \left(1 - \frac{c}{n}\right) x_{t-1} + u_t, \quad t = 1, 2, \ldots, \quad c > 0, \tag{12}$$

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 (6), 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 (12) 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 (8) 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, positive c) on x_t. The fact that the limiting distribution is the same for all finite $c \geq 0$ has the 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 $c > 0$ (Elliott, 1998).

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 the limiting variance of $(n/m)^{1/2}\widehat{\lambda}_{y,\Delta x}$, V. 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 the equations (24)-(26) in the proof of Lemma 2.2)

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

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 (22) by the integral in (29):

$$
\begin{aligned}
\widetilde{V} \;=\;& \frac{1}{m}\sum_{h'=1}^{n-1}\sum_{h=1}^{n-1} k\left(\frac{h'}{m}\right)k\left(\frac{h}{m}\right)\sum_{u=-\infty}^{\infty}\left\{\begin{array}{l}\widetilde{k}\left(\frac{u}{\widetilde{m}}\right)\widehat{\Gamma}_{\Delta x\Delta x}(u)\,\widetilde{k}\left(\frac{u+h-h'}{\widetilde{m}}\right)\widehat{\Gamma}_{yy}(u+h-h')\\ +\widetilde{k}\left(\frac{u+h}{\widetilde{m}}\right)\widehat{\Gamma}_{\Delta xy}(u+h)\,\widetilde{k}\left(\frac{u-h'}{\widetilde{m}}\right)\widehat{\Gamma}_{y\Delta x}(u-h')\end{array}\right\}\\
&\times\phi_n(u,h',h), \qquad (13)
\end{aligned}
$$

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 superior finite sample performance than estimating V by $\widehat{V}$ (the results with using $\widehat{V}$ is not reported in the present paper).

Suppose $(y_t, \Delta x_t)$ satisfies Assumption A and hence x_t is $I(1)$. Then the statistic

$$
t_\lambda = \frac{\sqrt{\frac{n}{m}}\left(\hat{\lambda}_{y,\Delta x} - \lambda_{y,\Delta x}\right)}{\sqrt{\widetilde{V}}} \to_d N(0,1),
$$

from Lemma 2.6 if $\widetilde{V} \to_p V$. 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 (10) 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 (13) and $Ey_t x_{t-h} = 0$ for all $h \geq 1$, 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 orthogonarity 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 $f_x(0) = 0$ and hence converges to 0 as $\widetilde{m} \to \infty$. Because m tends to infinity, the asymptotic behavior of the denominator and 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. The t_λ statistic also works well by itself in the local to unity and unit root cases. Since these tests are of primary interest when roots are close or equal to unity, t_λ may be simpler to compute and equally useful 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 which we focus on. In particular, we consider three models for x_t:

$$x_t = \rho_0 + \rho_1 x_{t-1} + u_{1,t}, \quad \text{AR(1)} \tag{14}$$
$$x_t = \rho_0 + \rho_1 x_{t-1} + \rho_2 x_{t-2} + u_{1,t}, \quad \text{AR(2) and} \tag{15}$$
$$(1-L)^d x_t = \varepsilon_{1,t} \ , \ 0 < d < 1 \ . \quad \text{ARFIMA(0,d,0)}. \tag{16}$$

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

$$\rho_1 = 1 + \frac{c}{n} \ , \ c \leq 0. \tag{17}$$

Often the primary economic interest centers on the relation between y_t and x_{t-1}. Two possible processes for y_t are considered here. First we consider the standard regression specification

$$y_t = d_t + \beta x_{t-1} + u_{2,t}. \tag{18}$$

The deterministic component d_t consists of either an intercept or a trend:

$$d_t = \delta_0 \quad \text{or}$$
$$d_t = \delta_0 + \delta_1 t.$$

However, as discussed above, this implies (perhaps unrealistically) that y_t is I(1) when x_t is I(1) and $\beta \neq 0$. Therefore, we also simulate from a regression of y_t on pre-filtered values of x_{t-1}, as given by

$$y_t = d_t + \gamma u_{1,t-1} + u_{2,t}. \tag{19}$$

With an AR(1) specification this data generating process is equivalent to

$$y_t = d_t + \gamma \left(1 - \rho_1 L\right) x_{t-1} + u_{2,t},$$

which simply becomes a regression of y_t on Δx_{t-1} when $\rho_1 = 1$. Note also that the distinction between these two data generating processes is relevant only when considering power. Under the null hypothesis y_t is orthogonal to $I_{x,t-1}$, implying that $\beta = \gamma = 0$. Thus under the null hypothesis y_t is simply given by

$$y_t = d_t + u_{2,t}. \tag{20}$$

Finally, since y_t orthogonal to past 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

$$u = \begin{pmatrix} u_{1t} & u_{2t} \end{pmatrix}' = \Sigma^{1/2}\varepsilon_t, \quad \varepsilon_t \sim N\left(0, I\right)$$
$$\Sigma = \Sigma^{1/2}(\Sigma^{1/2})' = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}.$$

Our primary interest lies in the performance of the covariance-based t-statistic, which is estimated as follows. In the trend model, Δx_t was de-meaned and y_t de-trended prior to estimation, where as in the intercept model on y_t was demeaned.[4]

[4]Note that removing the mean from Δx_t removes the trend in x_t

Since the regressor x_{t-1} is lagged we define $z_t = (y_t, \Delta x_{t-1})'$, where Δx_t and y_t are now de-meaned. The pseudo-covariance

$$\lambda_{y,\Delta x} = \sum_{h=1}^{\infty} E\left(y_t, \Delta x_{t-h}\right)$$

is next estimated using the standard kernel covariance estimator (see Newey-West (87), Andrews (91)) given by (8), using the Bartlett (Newey-West) kernel: $k(x) = 1 - |x|$ for $|x| \leq 1$ and zero otherwise (see Hannan (70, p. 278)).

Implementation of the optimal bandwidth selection procedure requires the use of a fist-stage parametric approximation. As in Andrews (91) this is assumed to only 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. Allowing for a moving average component also seems desirable given possible over-differencing in Δx_t. A VARMA(1,1) is therefore employed. Using the asymptotically efficient three stage linear regression method of Dufour and Pelletier (2002) avoids non-linear optimization, keeping estimation simple.[5]

Next, we estimate $\widetilde{V}$, the approximate finite sample variance of $\hat{\lambda}_{y,\Delta x}$ given in (13) using the same kernel and lag-length specifications.[6] This is the variance estimator for the I(1) and local to unity cases and is used to form the t-statistic

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

It is this covariance-based t-test, that is employed in the tables discussed below, together with standard normal critical values.

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 (14) with ρ_1 modelled local to unity as in (17). Both ρ_1 and contemporaneous innovation covariance σ_{12} are allowed to vary, while the other parameters are held fixed. In order to set a basis of comparison, in Table 1 we first provide empirical rejection rates for a standard two-sided t-test on the coefficient in the regression of y_t on x_{t-1}. All tests are conducted at a 5 percent nominal level. The regressions shown in the top half of the table, include only an intercept, those in the bottom half contain both an intercept and time trend. The size is approximately correct for low values of ρ_1 and/or σ_{12} but grows unreliable as ρ_1 approaches one and the residual correlation increases. The size problem for the regression with intercept becomes even more serious when a trend term is included.

[5]Simulation and empirical results using the optimal bandwidth formula derived earlier are still underway. The current results rely on the Andrews (91) formula using a diagonal weight matrix with diagonal elements $(2, 1, 1, 2)$. This appears to work reasonably well.

[6]Estimation and simulation results with $\widetilde{m}/m \to 0$, needed to guarantee conservative inference for x_t, are currently underway. The reported results hold $\widetilde{m} = m$, but nevertheless show reasonably good finite sample performance.

The covariance-based test developed above was shown to provide correct large sample size for $\rho = 1$ and for all local to unity alternatives with finite $cleq0$, as well as conservative inference for all fixed values of $\rho < 1$. As shown by the rejection rates in Table 2, the tests also work quite well in sample sizes as small $n = 100$. In contrast with the rejection rates for the standard t-test discussed above, the finite sample size is quite accurate for all values of ρ and σ_{12}, including even $\rho = 1$ and $\sigma_{12} = 0.95$, lying within 2 percentage points of the nominal level in all cases. This results from the fact the long-run covariance estimator upon which the estimator is based is asymptotically normal even for $\rho_1 = 1$, and as a result is not effected by the same unit root biases. Note that behavior the behavior in the "de-trended" and "de-meaned" cases are quite similar. To economize on space here on in we report only the more realistic "de-trended" case, in which Δx_t is de-meaned prior to estimation.

It is also of interest to investigate finite sample performance under higher order autoregressive specifications for x_t, such as the AR(2) model (15), with roots on, or outside, but close to the unit circle. Such a specification seems to be of practical relevance. Rudebusch (92, Table 2), for example, estimates autoregressive trend stationary models for fourteen macroeconomic and financial series (including GNP, industrial production, employment, prices, money stock, velocity, bond yields, and stock prices). For a majority of the series a second (or higher) lag shows up significant. In fact, the autoregressive estimates are fairly similar across the series, with $\hat{\rho}_1 > 1$, $\hat{\rho}_2 < 0$ and the sum of the coefficients $\hat{\rho}_1 + \hat{\rho}_2$, close to, but slightly below one (a unit root could not be rejected). Tables 4 shows finite sample rejection rates for the covariance-based t-tests, with x_t simulated from an AR(2) roughly matching the characteristics of the economic series described above. The first root ρ_1 is set equal to 1.5 in all cases while the second root ρ_2 is taken both equal to -0.5 (unit root case) and slightly less than -0.5 (near nonstationary case). The rejection rates, while still reasonable, are less accurate than in the single lag case. In particular they show a moderate tendency to over-reject, with a rejection rate just over 10 percent in the worst case. However, as shown in Table 3 the standard t-tests also fare worse than before, with rejection rates as high as 30 percent, so that the net improvement that comes from using the covariance-based t-test in place of the standard t-test may actually be larger than in the single lag case.

Finally, we investigate the size of the test under the assumption of long-memory or fractional integration, simulating x_t from (16). Like the local to unity specification, the long-memory processes also provide an intermediate model between standard short-memory stationary I(0) and unit root processes. Depending on the degree of long memory, such a process may be either stationary ($d < 0.5$) or non-stationary ($d > 0.5$). Stock and exchange rate volatilities, as well interest rates, foreign exchange forward premium, and unemployment rates are arguably well modelled by such a process. Although covariance t-tests have not been theoretically justified under long-memory assumptions ($\hat{\lambda}_{y,\Delta x}$ was however shown to be consistent) it is still of some practical interest to know how such a test would fare were x_t, for example, mistakenly modelled as local to unity when it was actually a long memory process. Table 5 shows rejection for the covariance when x_t is given by fractionally integrated model (16) for various values of $0 < d \leq 1$. The size of the covariance-based test appears quite accurate for all values of d, ranging from 0.04 to no more than 0.07.

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

4.2 Power

Finite sample power is examined using the autoregressive model (14) for x_t together with either the standard (18) or pre-filtered (19) regression with $\beta \neq 0$ or $\gamma \neq 0$ respectively. In both cases, rejection rates are reported for various values of β or γ, ρ_1, and σ_{12} for $n = 100$, with the other parameters held constant. The innovation variances are set to one, $\sigma_{22} = \sigma_{11} = 1$, and the true trend d_t is set equal to zero, although the data is de-meaned or de-trended prior to estimation.

Table 6 shows finite sample power of the covariance-based test against $\beta \neq 0$ in (18). Since for $\beta \neq 0$ a linear regression can be used to predict y_t using x_{t-1} this represents a violation of orthogonality. When x_t (and hence y_t) is stationary this a standard stationary regression relation. When x_t (and hence y_t) has a unit root the two are cointegrated. The rejection rates indicate reasonable power in both cases, with power increasing as β is taken further away from zero.

Table 7 next shows the power of the covariance-based t-test for $\gamma \neq 0$ in (19). Since for $\gamma \neq 0$ y_t can be predicted using a linear regression of y_t on $(1 - \rho_1 L)x_{t-1}$ this also constitutes a violation of orthogonality. This second alternative is arguably more realistic for a returns process when ρ_1 is close or equal to one, since y_t remains stationary (actually serially uncorrelated) even when the root in x_t is close or equal to unity. A second advantage is that the correlation between y_t and the regressor $(1 - \rho_1 L)x_{t-1}$and therefore 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}. This makes the rejection rates easier to interpret since they correspond not just to a particular value of γ but also to a particular R^2. As can be seen in the table, the rejection rates appear reasonable, although they vary considerably across parameter values. For an R^2 of 0.02 the rejection rates range from a low just above 20 percent to rejection rates well above 90 percent. As the R^2 increases the rejection rates become uniformly higher with minimum rejection rates of 34 percent for an R^2 of 0.04, 74 percent for an R^2 of 0.11, and 90 percent for an R^2 of 0.20.

In summary, the covariance t-tests not only provide fairly accurate size in small samples but also reasonable power against various alternatives.

5 Empirical Application

We apply the covariance-based tests to two well known orthogonality tests from the finance literature. These test the orthogonality of stock returns (r_{t+1}) relative to past information inherent in the log dividend-price ratio $(d_t - p_t)$ and the short term interest rate (i_t).

Standard regression results imply that the dividend-price-ratio contains substantial predictive power for stock returns, particularly for long-horizon returns [Campbell and Shiller (1988a,b), Fama and French (1988b), Hodrick(1992), Shiller (1984)]. This result has been interpreted either as a failure of market efficiency under risk-neutral assumptions, or alternatively, as evidence of a time varying risk premium. Theoretical considerations are generally argued to imply stationarity of the dividend-price-ratio [Campbell and Shiller (1988)], although this has recently been disputed [Tuypens (2002)]. Nevertheless, empirically, the dividend-price ratio is highly persistent. Confidence intervals for the largest root are typically wide, containing the unit root together with stationary values [Torous, Valkanov, and Yan (01)]. Likewise, it can be difficult to reject unit roots during important sample periods [Campbell (92), Campbell, Lo, and MacKinlay (1997, chapter 7, footnote 20)]. Furthermore, although the dividend yield is a pre-determined regressor, it seems unlikely to be strictly exogenous, since the stock price, which is closely related to past returns, enters into the denominator. The combination of near unit root behavior and a failure of strict exogeneity is a recipe for size problems [Cavanagh, Elliot, and Stock (1995)]. As we argue, it may also have important implications for power. Consequently, there has been an ongoing debate as to the robustness of these findings. This issue was originally raised Stambaugh (1986), Mankiw and Shapiro (1992), but a large literature has since developed employing sophisticated econometric tools.

One branch of this literature employs simulation and resampling techniques, assuming a fixed value of the autoregressive coefficient, close to, but less than one. Hodrick (1992) provides a careful simulation study suggesting possible over-rejection, yet finds that even after accounting for size problems, the dividend yield still has some predictive power for stock returns. Nelson and Kim (1993) draw similar conclusions using randomized (bootstrapped without replacement) null distributions. On the other hand, Goetzmann and Jorion (1993) and Wolf (2000) fail to find substantial evidence of predictability at all using bootstrap and subsampling procedures, respectively. Most recently Ang and Bekaert (2001) conclude (based on a Monte Carlo study) that there is evidence for short (less than a year) but not long-horizon predictability.

Under local to unity or unit root assumptions many common resampling techniques lack theoretical justification, leading to a second stream of the literature using local to unity based asymptotic techniques of the type proposed by Cavanagh, Elliot, and Stock (1995). Viceira (1997) finds no evidence of predictability for one-month returns under local to unity assumptions. Applying the long-run restrictions of the dynamic Gordon growth model, Valkanov (2003) finds no evidence of predictability at long-horizons (short-horizons are not tested). Torous, Valkanov, and Yan (2001) extend the Cavanagh, Elliot, and Stock (1995) methodology, to long-horizon regressions. They find evidence for short-horizon (a year or less), but not long-horizon predictability. Lanne (2002) takes an indirect approach. Failing to reject stationarity in the return series, he argues that stock returns are therefore unpredictable by any highly persistent (i.e., local to unity) regressor. Tuypens (2002) makes a similar argument.

Earlier in this paper, we argued that persistent regressors create two potential problems: size distortions and conceptual difficulties regarding the framing of the alternative hypothesis. The existing literature concentrates almost exclusively on fixing the size of regression-based tests. In this sense it addresses one part of the problem, but not the other. By appropriately correcting size, they are capable of overturning previously spurious rejections. However, at the same time, the regression-based nature of the tests restricts power to alternatives that become less plausible the more persistent the regressor. In this case, the reversal of previous results may instead be due to a lack of power against relevant alternatives. In principle, either of these two effects, or more likely, a combination of both may be present.

To be concrete, suppose that the dividend yield is modelled as a local to unity or unit root process. A non-zero population regression coefficient in the return-dividend-price-ratio regression would then imply local to unity or unit root behavior in the stock return. Therefore, unless we allow the stock return also to have a near unit root component, the population regression coefficient in stock-return-dividend-yield regression must be zero by definition. In this case, any properly sized ninety-five percent confidence interval based on this regression coefficient, should include zero ninety-five percent of the time. In fact, this is exactly the point made by Lanne (2002) in arguing that the dividend yield can not have predictive power. However, under the original null hypothesis of risk-neutral market efficiency, the returns are not only linearly unrelated to past dividend-yields, but are *fully orthogonal to all information* in past dividend-yields. This includes first-differences, high-frequency components, and deviations of the dividend yield from recent historical averages, all of which could contain potential predictive value for returns, even if the dividend yield contained a unit root. Therefore, previous studies using regression-based tests, while properly correcting for size, may still have inadvertently restricted the alternative hypothesis. The weakening and in many cases overturning of the previous stylized facts, could therefore be due either to the appropriate size correction, in which case earlier results were misleading, and/or to undue restrictions on allowable alternatives. By contrast, the covariance-based t-test has reasonable power against both the traditional alternatives considered in regression-based tests, as well alternatives that involve predictability based on prefiltered or first-differenced dividend yields.

Following much of the literature, we use monthly returns from 1927 to 1994 and also consider separately the two subperiods: 1927-1951 and 1952-1994. The data are the same as used by Campbell, Lo, and MacKinlay (1997, chapter 7).[7] The monthly log or continuously compounded nominal 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. The k period return is then given by $r_{t+1} + \ldots + r_{t+k}$ for a horizon of $k = 1, 3, 12, 24, 36$, and 48 months. The nominal returns are next deflated by the CPI to the create the real return series and the one month treasury bill is subtracted to create the excess return series. Following standard practice, in order to avoid seasonality, the dividend yield is calculated as the sum of dividends paid on the index over the previous year, divided

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

by the current level of the index: $d_t - p_t = ln\left((D_t + \ldots + D_{t-11})/P_t\right)$.

Standard regression results using robust standard errors are shown in Table 8. Results for excess returns are similar. Based on these results, the dividend-price ratio has some predictive power at all horizons, and particularly strong predictive power at longer horizons. The covariance-based t tests are shown in Table 9 using real returns and in Table 10 using excess returns. Since nearly all plausible alternative economic models imply a positive covariance between returns and dividend-yields, Hodrick(1992) advocates the use of one-sided critical values. Based on this criteria (one-sided 5 % critical value of 1.645) we reject orthogonality at long horizons (3-4 years) over the entire data set and quite strongly reject at the one year horizon during the latter period. As such, the results provide some ammunition to both sides of the debates. In accordance with the more recent literature discussed above, they confirm that evidence against predictability weakens considerably when the near-unit root nature of the regressors are taken into account. This suggests possible size problems in earlier regressions. On the other hand, even after allowing for near-root behavior, we can still reject orthogonality in several cases.

A second concern raised in the literature, comes from the reduction in effective sample size due to the overlapping nature of long-horizon returns, typically of the form of the form $r_{t+1} + r_{t+2} + ... + r_{t+k}$ for k as large as 48 in monthly data. Even for a well-behaved stationary regressor, HAC standard errors are therefore necessary in order for proper asymptotic inference, and tests may still over-reject in finite sample [Hodrick (1992), Nelson and Kim (1993), Richardson and Stock (1989), Valkanov (2003), Ang and Bekaert (2001)]. While there is no particular reason to think that the covariance-based test proposed here has any has any special finite sample advantage in this regard, it worth pointing out that no independence or white noise restrictions are required on y_{t+1}. With a few restrictions, it may follow any short-memory stationary process. The procedure therefore remains asymptotically valid for long-horizon regressions (fixed $k > 1$). By contrast, some other attractive non-parametric alternatives, such as sign and rank tests, do require independence and are not directly applicable in this context.

Jegadeesh (91) and Cochrane (91) suggest a useful way to bypass this problem using an alternative regression of the one period return r_{t+1} on the sum of the past k dividend yields $\sum_{j=0}^{k-1}(d_{t-k} - p_{t-k})$. This avoids long-horizon returns, and the resulting the resulting correlation in the residuals, while still testing the same null hypothesis. To see this note that

$$\text{cov}\left(r_{t+1}, \sum_{j=0}^{k-1}(d_{t-k} - p_{t-k})\right) = \text{cov}(r_{t+1} + r_{t+2} + ... + r_{t+k}, d_{t-k} - p_{t-k}),$$

with both covariance terms equal to zero under the null hypothesis. However, by adding a moving average structure to the dependent variable, this comes at the cost of increasing the autocorrelation in the already persistent regressor. This further increases the relative attractiveness of covariance based tests and other robust approaches relative to standard regression. [8]

[8] A final problem, common to nearly all long horizon tests, is that orthogonality at different

Table 11 shows results from this alternative regression. The results are generally similarly to those in Table 8. This has eliminated the overlapping data problem, but not the near unit root problem. Covariance based tests based on this same formulation are presented in Table 12. Similar using the excess return in place or real returns are shown in Table 13. Both results are qualitatively similar to the long-horizon covariance based test presented above.

Interest rate measures have also been found to have some predictive power for stock returns [Campbell (1987), Fama and French (1989), Hodrick (1992), Keim and Stambaugh (1986)]. Here, we focus on the short-term interest rates. Unlike the dividend-yield, the short rate has been found to help primarily in forecasting short-horizon returns. Thus overlapping data becomes less of a concern. On the other hand, short-term interest rates are if anything, more persistent than dividend yields, giving rise to similar concerns. [9]

Covariance based tests for real returns using the treasury bill rate are provided in Tables 15 and 17. Similar results using excess returns are shown in Tables 16 and 18. Previous evidence of predictability in short horizon returns (particularly 3-12 months) is strongly confirmed in the latter period (1952-1994) and somewhat more weakly so over the entire sample. The negative sign of the estimates is again consistent with previous studies and implies that a rise in interest rates leads to an expected decline in stock market value.

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

---

horizons are tested separately, where in principle joint tests should be conducted to avoid data snooping. [Ang and Bekaert (2001)]. We are equally guilty of this sin.

[9] Stationary transformations are sometimes applied to the treasury bill rate. For example, in their textbook treatment, Campbell, Lo, and MacKinlay (1997, chapter 7) stochastically detrend the interest rate, replacing the level by a triangularly weighted moving average of past changes, using $i_t - \sum_{j=0}^{11} i_{t-j}$. It is interesting to note the rather close resemblance between this add hoc procedure and the Bartlett (Newey-West) kernel weighting procedure employed in the covariance-based tests developed here. Both allow power in directions that maintain the stationarity of the return series.

between the dependent variable and the 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.

7.1 Proof of (6)

From the definition of x_t, we have

$$\Delta x_t = u_t - \frac{c}{n} x_{t-1} = \begin{cases} u_t - \frac{c}{n} \sum_{k=0}^{t-2} \left(1 - \frac{c}{n}\right)^k u_{t-1-k}, & t \geq 1, \\ 0, & t \leq 0. \end{cases} \tag{21}$$

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

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

Therefore,

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

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

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

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

$$\frac{n}{m}\mathrm{var}\left(\widehat{\lambda}_{y,\Delta x}\right)=\frac{n}{m}\sum_{h'=1}^{n-1}\sum_{h=1}^{n-1}k\left(\frac{h'}{m}\right)k\left(\frac{h}{m}\right)\mathrm{cov}\left(\widehat{\Gamma}_{\Delta xy}\left(h'\right),\widehat{\Gamma}_{\Delta xy}\left(h\right)\right). \tag{22}$$

Hannan (1970) p. 313 gives

$$\begin{aligned}&\mathrm{cov}\left(\widehat{\Gamma}_{\Delta xy}\left(h'\right),\widehat{\Gamma}_{\Delta xy}\left(h\right)\right) \qquad (23)\\ =\;& n^{-1}\sum_{u=-\infty}^{\infty}\{\Gamma_{\Delta x\Delta x}\left(u\right)\Gamma_{yy}(u+h-h')+\Gamma_{\Delta xy}(u+h)\Gamma_{y\Delta x}(u-h')\\ &+k_{\Delta xy\Delta xy}(0,h',u,u+h)\}\phi_n(u,h',h),\end{aligned}$$

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

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

It follows that (22) is comprised of

$$\frac{1}{m}\sum_{h'=1}^{n-1}\sum_{h=1}^{n-1}k\left(\frac{h'}{m}\right)k\left(\frac{h}{m}\right)\sum_{u=-\infty}^{\infty}\Gamma_{\Delta x\Delta x}\left(u\right)\Gamma_{yy}(u+h-h')\phi_n(u,h',h) \tag{24}$$

$$+\frac{1}{m}\sum_{h'=1}^{n-1}\sum_{h=1}^{n-1}k\left(\frac{h'}{m}\right)k\left(\frac{h}{m}\right)\sum_{u=-\infty}^{\infty}\Gamma_{\Delta xy}(u+h)\Gamma_{y\Delta x}(u-h')\phi_n(u,h',h) \tag{25}$$

$$+\frac{1}{m}\sum_{h'=1}^{n-1}\sum_{h=1}^{n-1}k\left(\frac{h'}{m}\right)k\left(\frac{h}{m}\right)\sum_{u=-\infty}^{\infty}k_{\Delta xy\Delta xy}(0,h',u,u+h)\phi_n(u,h',h). \tag{26}$$

Let $v=h'-h$, and we can rewrite (24) as

$$\sum_{v=-n+2}^{n-2}\sum_{u=-\infty}^{\infty}\Gamma_{\Delta x\Delta x}\left(u\right)\Gamma_{yy}(u-v)\left\{\frac{1}{m}\sum_{h}{}'\phi_n(u,h+v,h)k\left(\frac{h+v}{m}\right)k\left(\frac{h}{m}\right)\right\}, \tag{27}$$

where the summation $\sum_h'$ runs only for $\{h:1\le h\le n-1 \text{ and } 1\le h+v\le n-1\}$. The bracketed expression converges to $\int_0^\infty k^2\left(x\right)dx$ by the argument in Hannan (1970) 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 (24) converges to $4\pi^2 f_{\Delta x\Delta x}(0)\, f_{yy}(0) \int_0^\infty k^2(x)\,dx$ as $n \to \infty$. Similarly, (25) converges to $4\pi^2 f_{\Delta xy}(0)\, f_{y\Delta x}(0) \int_0^\infty k^2(x)\,dx = 4\pi^2 [f_{y\Delta x}(0)]^2 \int_0^\infty k^2(x)\,dx$. For (26), from Hannan (1970, p. 211), the fourth cumulant of z_t satisfies

$$\sum_{q=-\infty}^{\infty} \sum_{r=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} |k_{ijkl}(0,q,r,s)| < \infty, \quad i,j,k,l = \{y, \Delta x\}. \tag{28}$$

Therefore, (26) 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

$$\frac{n}{m}\text{var}\left(\widehat{\lambda}_{y,\Delta x}\right) \to 4\pi^2 \int_0^\infty k^2(x)\,dx \left\{f_{\Delta x\Delta x}(0)\, f_{yy}(0) + [f_{y\Delta x}(0)]^2\right\}, \tag{29}$$

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 $\sqrt{n/m}(\widehat{\lambda}_{y,\Delta x} - E\widehat{\lambda}_{y,\Delta x}) \to_d N(0,V)$. First, observe that

$$\begin{aligned} & \sqrt{\frac{n}{m}}\left(\widehat{\lambda}_{y,\Delta x} - E\widehat{\lambda}_{y,\Delta x}\right) \\ = \; & \frac{1}{\sqrt{m}} \sum_{h=1}^{n-1} k\left(\frac{h}{m}\right) \frac{1}{\sqrt{n}} \sum_{t=h+1}^{n} \left(y_t\Delta x_{t-h} - Ey_t\Delta x_{t-h}\right) = I + II, \end{aligned} \tag{30}$$

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

$$\begin{aligned} E(II)^2 &= O\left(\frac{1}{mn}\left(\sum_{h=1}^{n-1}\left|k\left(\frac{h}{m}\right)\right| h^{1/2}\right)^2\right) \\ &= O\left(\frac{m^2}{n}\left(\sum_{h=1}^{n-1}\left|k\left(\frac{h}{m}\right)\right|\left(\frac{h}{m}\right)^{1/2}\frac{1}{m}\right)^2\right) = O\left(\frac{m^2}{n}\right), \end{aligned} \tag{31}$$

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

$$I = \sum_{t=1}^{n} Z_t + R_n; \quad Z_t = n^{-1/2} m^{-1/2} \sum_{h=1}^{n-1} k\left(\frac{h}{m}\right) \sum_{r=1}^{\infty} \varepsilon_{t-r}' f^{hr}(1) \varepsilon_t, \tag{32}$$

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

$$\sum_{t=1}^{n} Z_t \to_d N(0,V), \quad \text{as } n \to \infty. \tag{33}$$

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 (33) follows from martingale CLT (Brown, 1971) if

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

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 iid. 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

$$nEZ_t^2 \to V. \tag{34}$$

From (30)-(32), 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

$$E\left(\sum_{t=1}^{n} Z_t\right)^2 = E\left(\sqrt{\frac{n}{m}}\left(\widehat{\lambda}_{y,\Delta x} - E\widehat{\lambda}_{y,\Delta x}\right) - (II + R_n)\right)^2. \tag{35}$$

The left hand side of (35) 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 (35) 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 (34) 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, (33) and the stated result follow. ■

7.4 Proof of Lemma 2.8

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

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

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 Hannan (1970) in equation (3.3) on page 209 and line 5 on page 211. The second term is $O(n^{1/2}m^{-q})Ey_t x_{t-1}$ 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 $Ey_t x_{t-h} = \gamma_{xy}(h) = 0$ 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 $B_{1n} = O(\sum_{h=2}^{m} (h/m)^q |\gamma_{xy}(h)|) = O(m^{-q})$. B_{2n} is bounded by, since $k(x)$ is Lipschitz(1),

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

Therefore, defining $B_n = (k(1/m) - 1)\sqrt{n} E y_t x_{t-1} + ET_{2n}$ gives the bias term B_n in (11).

It remains to show $\text{var}(\sqrt{n} T_{2n}) = o(1)$ and $\sqrt{n}(T_{3n} + T_{4n}) = o_p(1)$. From Hannan (1970) (equation (3.3) on page 209 and line 5 on page 211), we have

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

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

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

The 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

$$\sqrt{n}(T_{3n}+T_{4n})=\sum_{p=2}^{n}k\left(\frac{p-1}{m}\right)\frac{1}{\sqrt{n}}y_px_0,$$

$x_0=O_p(1)$, and

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

and the stated result follows. ■

7.5 Proof of Lemma 2.11

From (21), we have

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

The stated result follows if we show

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

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

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

giving the stated result. ■

7.6 Proof of Lemma 3.2

From equations (24), (25) and (27) in the proof of Lemma 2.2, $\widetilde{V}$ comprises of two parts, the first of which is

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

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 $\eta > 0$ that

$$\Pr(\widetilde{V} \geq \eta \widetilde{m}^{-1}) \to 1, \quad \text{as } n \to \infty. \tag{36}$$

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

$$\left[\sum_{u=-\widetilde{m}}^{\widetilde{m}} \widetilde{k}\left(\frac{u}{\widetilde{m}}\right) \widehat{\Gamma}_{\Delta x \Delta x}(u) \sum_{v=-\widetilde{m}}^{\widetilde{m}} \widetilde{k}\left(\frac{v}{\widetilde{m}}\right) \widehat{\Gamma}_{yy}(v) \right] \left\{ \int_0^\infty k^2(x)\, dx + o(1) \right\} \tag{37}$$

$$+ \left[\sum_{u=-\widetilde{m}}^{\widetilde{m}} \widetilde{k}\left(\frac{u}{\widetilde{m}}\right) \widehat{\Gamma}_{\Delta xy}(u) \sum_{v=-\widetilde{m}}^{\widetilde{m}} \widetilde{k}\left(\frac{v}{\widetilde{m}}\right) \widehat{\Gamma}_{y\Delta x}(v) \right] \left\{ \int_0^\infty k^2(x)\, dx + o(1) \right\} \tag{38}$$

(38) 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 (37), 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, (36) follows if there exists $\varepsilon > 0$ such that

$$\begin{aligned} & \Pr\left(\sum_{v=-\widetilde{m}}^{\widetilde{m}} \widetilde{k}\left(\frac{v}{\widetilde{m}}\right) \widehat{\Gamma}_{\Delta x \Delta x}(v) \geq \varepsilon \widetilde{m}^{-1} \right) \\ = & \Pr\left(2\pi \int_{-\pi}^{\pi} W_{\widetilde{m}}(\lambda)\, I_{\Delta x}(\lambda) d\lambda \geq \varepsilon \widetilde{m}^{-1} \right) \to 1, \quad \text{as } n \to \infty, \end{aligned} \tag{39}$$

where (Priestley, 1981, p. 439)

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

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

$$\begin{aligned}
& \int_{-\pi}^{\pi} W_{\widetilde{m}}(\lambda) I_{\Delta x}(\lambda) d\lambda \\
= & \int_{-\pi}^{\pi} W_{\widetilde{m}}(\lambda) |1-e^{i\lambda}|^2 I_x(\lambda) d\lambda \quad (40)\\
& +(2\pi n)^{-1/2} X_n \int_{-\pi}^{\pi} W_{\widetilde{m}}(\lambda) 2\mathrm{Re}\left[(1-e^{i\lambda}) w_x(\lambda) e^{-i(n+1)\lambda}\right] d\lambda \quad (41)\\
& + \int_{-\pi}^{\pi} W_{\widetilde{m}}(\lambda) d\lambda (2\pi n)^{-1} X_n^2. \quad (42)
\end{aligned}$$

We can ignore (42) because it is nonnegative. For (41), 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 (41)$= O_p(n^{-1/2}\widetilde{m}^{-1/2}) = o_p(\widetilde{m}^{-1})$ follows. Rewrite (40) as

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

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)

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

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

$$\begin{cases} \sup_{\lambda,\lambda'\in[-\pi,\pi]} \left|\mathrm{cov}\left(I_x(\lambda), I_x(\lambda')\right)\right| = O(1), \\ \mathrm{cov}\left(I_x(\lambda), I_x(\lambda')\right) = o(1), \quad \lambda \neq \lambda'. \end{cases} \quad (43)$$

Therefore,

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

where the interchange of expectation and integration in the first line is valid by (43) and Fubini's Theorem, and the last line follows from Lemma 8.7 (b), (43), and the dominated convergence theorem. Therefore, there exists $\eta' > 0$ such that (40)+(41)+(42)$\geq \eta' \widetilde{m}^{-1}$ with probability approaching one, and (39) and the stated result follow. ■

8 Appendix B: technical results

8.1 Lemma

Under the assumptions of Theorem 2.6,

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

8.2 Proof

Observe that

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

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

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

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 $h \sup_s ||\Gamma(s)|| \sum_{l=-\infty}^{\infty} ||\Gamma(l)|| \leq Ch$. From (28), the fourth term on the right is bounded by $h \sum_{l=-\infty}^{\infty} \sum_{r=-\infty}^{\infty} |k_{y\Delta xy\Delta x}(0,-r,l,l-r)| \leq Ch$, and the stated result follows. ■

8.3 Lemma

Under the assumptions of Theorem 2.6,

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

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

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

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

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

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

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

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

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

with

$$\widetilde{f}^{hr}(L) = \sum_{j=0}^{\infty} \widetilde{f}_j^{hr} L^j, \quad \widetilde{f}_j^{hr} = \sum_{s=j+1}^{\infty} f_s^{hr} = \sum_{s=j+1}^{\infty} \left[(A_{s+r-h}^2)' A_s^1 + (A_{s+r}^1)' A_{s-h}^2\right]. \quad (44)$$

It follows that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} y_t \Delta x_{t-h} = \mathrm{tr}\left(f^{h0}(1) \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \varepsilon_t \varepsilon_t' + \sum_{r=1}^{\infty} f^{hr}(1) \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \varepsilon_t \varepsilon_{t-r}'\right) + r_{nh}, \quad (45)$$

where

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

From Lemma 8.5, we have

$$E|r_{nh}|^2 \leq C n^{-1}, \quad h = 1, \ldots, n-1. \quad (46)$$

Furthermore, observe that

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

In conjunction with (45), it follows that

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

and hence

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

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 (46) 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 $j < 0$,

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

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

8.5 Lemma

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

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

8.6 Proof

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

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

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

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

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

$$E\left(\sum_{j=0}^{\infty} \xi_{n-j}^h \right)^2 = \sum_{j=0}^{\infty} E(\xi_{n-j}^h)^2 \leq C \sum_{j=0}^{\infty} \sum_{r=1}^{\infty} ||\widetilde{f}_j^{hr}||^2 \leq C \left(\sup_{j,r} ||\widetilde{f}_j^{hr}|| \right) \sum_{j=0}^{\infty} \sum_{r=1}^{\infty} ||\widetilde{f}_j^{hr}||. \tag{47}$$

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 $h = 1, \ldots, n-1$,

$$\begin{aligned} \sum_{j=0}^{\infty} \sum_{r=1}^{\infty} ||\widetilde{f}_j^{hr}|| &\leq \sum_{j=0}^{\infty} \sum_{r=1}^{\infty} \sum_{s=j+1}^{\infty} ||A_{s+r-h}|| ||A_s|| + \sum_{j=0}^{\infty} \sum_{r=1}^{\infty} \sum_{s=j+1}^{\infty} ||A_{s+r}|| ||A_{s-h}|| \\ &\leq \sum_{j=0}^{\infty} \sum_{s=j+1}^{\infty} ||A_s|| \sum_{r=0}^{\infty} ||A_r|| + \sum_{j=0}^{\infty} \sum_{s=j+1}^{\infty} ||A_{s-h}|| \sum_{r=0}^{\infty} ||A_r||. \end{aligned} \tag{48}$$

The first term in (48) 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 (48) 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 (47) 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 $\lambda \in [-\pi, \pi]$,

$$c(\lambda/2)^2 \leq \sin^2(\lambda/2) \leq (\lambda/2)^2. \tag{49}$$

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 (49) and $|\sin x| \leq 1$. ■

References

[1] Andrews, D. W. K. (1991) Heteroskedasticity and autocorrelation consistent covariance estimation. *Econometrica* **59**: 817-858.

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

[3] Brown, B. M. (1971). Martingale central limit theorems. *Annals of Mathematical Statistics* **42**: 59-66.

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

[5] 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.

[6] Campbell, John Y. (1987). Stock returns and the term structure. *Journal of Financial Economics*, June, 18. pp 373-400.

[7] Campbell, J. (1991). A Variance Decomposition for Stock Returns. *Economic Journal*, 101, 157-179.

[8] Campbell, John Y. and Robert J. Shiller (1988). Stock Prices, Earnings, and Expected Dividends. *Journal of Finance* 43:3 pp. 661-76.

[9] Cambell,J., A. Lo and C. MacKinlay (1997). *The Econometrics of Financial Markets*. Princeton University Press.

[10] Cambell,J. and M. Yogo (2003). Efficient tests of stock return predictability. *Mimeo, Department of Economics, Harvard University.*

[11] Cavanagh, C. L., G Elliot, and J.H. Stock (1995). Inference in models with nearly integrated regressors. *Econometric Theory* 11: 1131-1147.

[12] Christiano, Eichenbaum and R. Vigfussen (2002). What Happens After a Technology Shock? mimeo, Northwestern University.

[13] Cochrane, (1991). Volatility Tests and Efficient Markets: A Review Essay. *Journal of Monetary Economics*, 27 463-487.

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

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

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

[17] Fama, Eugene and Kenneth French (1988). Permanent and temporary components of stock prices. *Journal of Political Economy.* 96:2,

[18] Fama, Eugene and Kenneth French (1989). Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics*,25, 23-49.

[19] Fisher, J. (2002). Technology Shocks Matter. Federal reserve bank of Chicago.

[20] Goetzmann, W. and P. Jorion (1993). Testing the predictive power of dividend yields. *Journal of Finance* 48:2, 663-679.

[21] Hannan, E. J. (1970). *Multiple time series.* New York, Wiley.

[22] Hansen, L. and R. Hodrick (1980). Forward exchange rates as optimal predictors of future spot rates: an econometric analysis. *Journal of Political Economy* 88, 829-853.

[23] Hodrick, R. (1992). Dividend yields and expected stock returns: Alternative procedures for inference and measurement. *Review of Financial Studies*, 5, 357-386.

[24] Jagadeesh, N. (1991). Seasonality in stock price mean reversion: Evidence from the U.S. and the U.K. *Journal of Finance*, 46, 1427-1444.

[25] Keim, Donald B. and Robert T. Stambaugh. (1986). Predicting returns in stock and bond markets.*Journal of Financial Economics.* 17, pp357-90.

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

[27] Kitamura, Yuichi and Peter C. B. Phillips (1997). Fully Modified IV, GIVE and GMM estimation with possibly nonstationary regressors and instruments. *Journal of Econometrics,* **80:** 85-123.

[28] Mankiw, N.G. and M. Shapiro, (1986). Do we reject too often? Small sample Properties of tests of rational expectations models. *Economics Letters*, 20 139-145..

[29] Mankiw, N.G., D. Romer, and M. Shapiro, (1985). An Unbiased Reexamination of Stock Market Volatility. *Journal of Finance*, 40 667-687.

[30] Maynard, A. (2002). Persistence, bias, and the forward premium anomaly: How bad is it? Mimeo, University of Toronto.

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

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

[33] Phillips, P.C.P. (1995). Fully modified least squares and vector autoregression. *Econometrica* 63 (5), 1023-1078.

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

[35] Politis and Romano (1994) Large sample confidence regions based on subsamples under minimal assumptions. *Annals of Statistics* 22: 2031-2050.

[36] Priestley, M. B. (1981). *Spectral Analysis and Time Series.* New York, Academic Press.

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

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

[39] Shiller, R. (1984). Stock prices and Social Dynamics. Brookings papers on Economic Activity, 2, 457-498.

[40] Stambaugh, R.F., (1982). Bias in regressions with lagged stochastic regressors. CRSP Working Paper 156, University of Chicago.

[41] Stambaugh, R.F. (1999). Predictive regressions. *Journal of Financial Economics* 54, 375-421.

[42] Torous, W., R. Valkanov, and S. Yan, (2001). On predicting stock returns with nearly integrated explanatory variables. *Journal of Business*, forthcoming.

[43] Tuypens, B. (2002). Examining the statistical properties for financial ratios. mimeo, Yale University.

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

[45] Viceira, L. Testing for Structural Change in the Predictability of Asset Returns. manuscript, Harvard University Working Paper, 1997.

[46] 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.

9 Tables

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

c	ρ	$\sigma_{1,2} = 0$	0.25	0.50	0.75	0.95
		A. Demeaned Case				
0.000	1.000	0.048	0.056	0.099	0.136	0.173
-1.000	0.990	0.045	0.055	0.089	0.111	0.129
-5.000	0.950	0.054	0.061	0.056	0.085	0.085
-10.000	0.900	0.051	0.059	0.058	0.070	0.069
-20.000	0.800	0.058	0.045	0.062	0.063	0.059
		B. Detrended Case				
0.000	1.000	0.060	0.090	0.165	0.236	0.291
-1.000	0.990	0.058	0.072	0.139	0.212	0.264
-5.000	0.950	0.050	0.064	0.099	0.129	0.150
-10.000	0.900	0.057	0.059	0.084	0.086	0.118
-20.000	0.800	0.050	0.052	0.069	0.070	0.074

Results are based on 2000 replications with N = 0100. The process is given by $x_t = \rho x_{t-1} + \epsilon_{1,t}$ and $y_t = \epsilon_{2,t}$ for $\rho = 1 + c/N$ and $\epsilon_t \sim i.i.d.\ N(0, \Sigma)$. Tests are conducted at the 0.05 level. Sizes are reported for a standard t-statistic based on the regession of y_t on x_{t-1}

Table 2: Finite Sample size (local to unity)

c	ρ	$\sigma_{1,2} = 0$	0.25	0.50	0.75	0.95
		A. Demeaned Case				
0.000	1.000	0.045	0.047	0.037	0.052	0.053
-1.000	0.990	0.054	0.048	0.043	0.050	0.056
-5.000	0.950	0.052	0.039	0.052	0.048	0.047
-10.000	0.900	0.046	0.053	0.039	0.038	0.044
-20.000	0.800	0.046	0.049	0.045	0.041	0.056
		B. Detrended Case				
0.000	1.000	0.048	0.042	0.054	0.054	0.070
-1.000	0.990	0.051	0.060	0.057	0.057	0.066
-5.000	0.950	0.059	0.049	0.060	0.059	0.052
-10.000	0.900	0.061	0.051	0.041	0.040	0.049
-20.000	0.800	0.067	0.044	0.052	0.051	0.056

Results are based on 2000 replications with N = 0100. The process is given by $x_t = \rho x_{t-1} + \epsilon_{1,t}$ and $y_t = \epsilon_{2,t}$ for $\rho = 1 + c/N$ and $\epsilon_t \sim i.i.d.\ N(0, \Sigma)$. Tests are conducted at the 0.05 level. Rejections rates are calculated for long-run covariance-based t-test.

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

$\rho_1 + \rho_2$	ρ_2	$\sigma_{1,2} = 0$	0.25	0.50	0.75	0.95
		A. Demeaned Case				
1.000	-0.500	0.050	0.067	0.099	0.162	0.170
0.990	-0.510	0.045	0.059	0.070	0.091	0.120
0.975	-0.525	0.059	0.052	0.067	0.087	0.103
0.950	-0.550	0.057	0.064	0.060	0.059	0.075
0.925	-0.575	0.059	0.055	0.068	0.059	0.069
0.900	-0.600	0.055	0.064	0.049	0.056	0.064
0.800	-0.700	0.063	0.048	0.058	0.046	0.062
		B. Detrended Case				
1.000	-0.500	0.051	0.091	0.159	0.252	0.317
0.990	-0.510	0.060	0.071	0.120	0.180	0.203
0.975	-0.525	0.051	0.073	0.099	0.140	0.149
0.950	-0.550	0.061	0.052	0.079	0.096	0.109
0.925	-0.575	0.068	0.059	0.061	0.093	0.104
0.900	-0.600	0.046	0.060	0.072	0.080	0.084
0.800	-0.700	0.057	0.066	0.055	0.065	0.066

Results are based on 2000 replications with N = 0100. The process is given by $x_t = \rho_1 x_{t-1} + \rho_2 x_{t-1} \epsilon_{1,t}$ and $y_t = \epsilon_{2,t}$ for $\rho_1 = 1.5$ and $\epsilon_t \sim i.i.d.\ N(0, \Sigma)$. Tests are conducted at the 0.05 level. Rejections rates are calculated for long-run covariance-based t-test.

Table 4: Finite Sample size (AR(2))

$\rho_1 + \rho_2$	ρ_2	$\sigma_{1,2} = 0$	0.25	0.50	0.75	0.95
1.000	-0.500	0.064	0.061	0.091	0.110	0.104
0.990	-0.510	0.062	0.065	0.070	0.091	0.094
0.975	-0.525	0.067	0.079	0.080	0.086	0.085
0.950	-0.550	0.070	0.061	0.061	0.076	0.083
0.925	-0.575	0.064	0.073	0.066	0.077	0.082
0.900	-0.600	0.064	0.057	0.062	0.070	0.071
0.800	-0.700	0.074	0.070	0.075	0.085	0.079

Results are based on 2000 replications with N = 100. The process is given by $x_t = \rho_1 x_{t-1} + \rho_2 x_{t-1} \epsilon_{1,t}$ and $y_t = \epsilon_{2,t}$ for $\rho_1 = 1.5$ and $\epsilon_t \sim i.i.d.\ N(0, \Sigma)$. Tests are conducted at the 0.05 level. Rejections rates are calculated for long-run covariance-based t-test.

Table 5: Finite Sample size (ARFIMA)

d	$\sigma_{1,2} = 0$	0.25	0.50	0.75	0.95
1.000	0.051	0.048	0.061	0.052	0.062
0.800	0.070	0.061	0.061	0.051	0.049
0.600	0.070	0.047	0.052	0.052	0.048
0.400	0.056	0.067	0.061	0.056	0.056
0.200	0.072	0.062	0.052	0.057	0.069

Results are based on 2000 replications with N = 0100. The process is given by $(1 - L)^d x_t = \epsilon_{1,t}$ and $y_t = \epsilon_{2,t}$ and $\epsilon_t i.i.d. \sim\ N(0, \Sigma)$. Tests are conducted at the 0.05 level. Rejections rates are calculated for long-run covariance-based t-test.

Table 6: Finite Sample power ($y_t = \beta x_{t-1} + \epsilon_{2,t}$)

ρ_1	$\sigma_{1,2}$	β =0.15	0.20	0.35	0.50	0.75	1.00
1.000	0.000	0.143	0.254	0.738	0.970	1.000	1.000
	0.500	0.120	0.219	0.688	0.962	1.000	1.000
	0.950	0.091	0.200	0.680	0.963	0.999	1.000
0.990	0.000	0.247	0.374	0.671	0.754	0.792	0.806
	0.500	0.130	0.195	0.464	0.683	0.753	0.740
	0.950	0.080	0.118	0.307	0.507	0.640	0.692
0.975	0.000	0.253	0.380	0.718	0.799	0.828	0.836
	0.500	0.143	0.218	0.527	0.724	0.782	0.776
	0.950	0.086	0.150	0.331	0.540	0.691	0.734
0.925	0.000	0.257	0.435	0.791	0.902	0.939	0.950
	0.500	0.155	0.248	0.585	0.815	0.886	0.893
	0.950	0.099	0.154	0.405	0.651	0.829	0.875
0.800	0.000	0.251	0.406	0.833	0.969	0.992	0.995
	0.500	0.166	0.281	0.674	0.905	0.981	0.980
	0.950	0.109	0.176	0.484	0.743	0.952	0.982

Results are based on 2000 replications with N = 100. The process is given by $x_t = 1+\rho * x_{t-1}\epsilon_{1,t}$ and $y_t = \beta xt - 1+\epsilon_{2,t}$, and $\rho = 1+c/n$. and $\epsilon_t \sim i.i.d.\ N(0, \Sigma)$. Tests are conducted at the 0.05 level. Rejections rates are calculated for long-run covariance-based t-test. The Bartlett (Newey-West) kernel is employed. Optimal bandwidths are employed based on initial estimates from a VARMA(1,1).

Table 7: Finite Sample power ($y_t = \beta(1 - \rho L)x_{t-1} + \epsilon_{2,t}$)

ρ_1	$\sigma_{1,2}$	$\beta = 0.15$	0.20	0.35	0.50	0.75	1.00
		r^2 =0.02	0.04	0.11	0.20	0.36	0.50
1.000	0.000	0.240	0.393	0.756	0.891	0.954	0.972
	0.500	0.275	0.458	0.916	0.992	0.999	1.000
	0.950	0.634	0.748	0.996	1.000	1.000	0.999
0.990	0.000	0.241	0.372	0.745	0.905	0.956	0.985
	0.500	0.289	0.486	0.919	0.996	1.000	1.000
	0.950	0.684	0.821	0.998	1.000	1.000	1.000
0.975	0.000	0.249	0.363	0.763	0.901	0.970	0.986
	0.500	0.281	0.511	0.931	0.998	1.000	1.000
	0.950	0.751	0.906	1.000	1.000	1.000	0.999
0.925	0.000	0.246	0.353	0.761	0.927	0.989	0.999
	0.500	0.339	0.523	0.947	0.998	1.000	1.000
	0.950	0.896	0.984	1.000	1.000	1.000	1.000
0.800	0.000	0.210	0.350	0.742	0.942	0.998	1.000
	0.500	0.345	0.539	0.936	0.998	1.000	1.000
	0.950	0.980	0.999	1.000	1.000	1.000	1.000

Results are based on 2000 replications with N = 100. The process is given by $x_t = 1 + \rho * x_{t-1}\epsilon_{1,t}$ and $y_t = \beta\epsilon_{1,t-1} + \epsilon_{2,t}$, and $\rho = 1 + c/n$ and $\epsilon_t \sim i.i.d.\ N(0, \Sigma)$. Tests are conducted at the 0.05 level. Rejections rates are calculated for long-run covariance-based t-test. The Bartlett (Newey-West) kernel is employed. Optimal bandwidths are calculated using Andrews(91) based on initial estimates from a VARMA(1,1).

Table 8: Regression of long horizon real stock returns on log dividend price ratios

$$y_{t+k} = r_{t+1} + \ldots + r_{t+k}$$
$$x_t = ln\left((D_t + \ldots + D_{t-12})/P_t\right)$$

Forecast Horizon (k)

$k =$	1.000	3.000	12.000	24.000	36.000	48.000
			1927 to 1994			
$\hat{\beta}$	0.016	0.043	0.200	0.386	0.533	0.654
R^2	0.007	0.014	0.073	0.143	0.207	0.261
t_β	2.389	1.598	2.658	4.221	4.843	4.707
			1927 to 1951			
$\hat{\beta}$	0.024	0.054	0.304	0.667	0.925	1.085
R^2	0.007	0.011	0.086	0.217	0.330	0.419
t_β	1.472	0.886	2.134	3.796	3.180	4.226
			1952 to 1994			
$\hat{\beta}$	0.027	0.080	0.327	0.579	0.757	0.843
R^2	0.018	0.049	0.188	0.322	0.411	0.417
t_β	3.098	3.728	3.845	3.589	3.775	4.051

Regressions are estimated by OLS with Newey-West standard errors, setting the bandwidth equal to $k-1$.

Table 9: Covariance-based orthogonality tests on long horizon real stock returns using log dividend price ratios

$$y_{t+k} = r_{t+1} + \ldots + r_{t+k}$$
$$x_t = ln\left((D_t + \ldots + D_{t-12})/P_t\right)$$

Forecast Horizon (k)

$k =$	1.000	3.000	12.000	24.000	36.000	48.000
			1927 to 1994			
t_λ	1.5568	-0.8273	1.0232	1.6360	1.7722	1.8415
M	1.6979	8.5535	50.8592	60.0334	57.9444	61.1089
			1927 to 1951			
t_λ	1.5358	0.0336	-0.1144	0.6480	1.2349	1.6439
M	3.0416	5.6921	32.0844	37.9558	42.0417	42.0936
			1952 to 1994			
t_λ	0.0684	-0.8066	1.7281	0.5334	0.4550	-0.3376
M	3.1168	4.9193	49.1232	79.0276	45.8764	53.6242

The Bartlett (Newey-West) kernel is employed. Optimal bandwidths are calculated using Andrews (91) based on initial estimates from a VARMA(1,1).

Table 10: Covariance-based orthogonality tests on long horizon excess stock returns using log dividend price ratios

$$y_{t+k} = r_{t+1} + \ldots + r_{t+k}$$
$$x_t = ln\left((D_t + \ldots + D_{t-12})/P_t\right)$$

Forecast Horizon (k)

$k =$	1.000	3.000	12.000	24.000	36.000	48.000
			1927 to 1994			
t_λ	1.3939	-0.8743	0.8447	1.5824	1.7727	1.9412
M	1.8233	8.4450	55.5214	64.0799	61.2324	64.3484
			1927 to 1951			
t_λ	1.2946	-0.0970	-0.3579	0.3777	1.0294	1.2946
M	2.9226	5.7579	35.7451	40.2412	49.6852	45.8701
			1952 to 1994			
t_λ	-0.0218	-0.8390	2.2099	1.0010	0.8855	-0.0260
M	3.4792	3.6139	46.1167	59.2861	41.0933	46.8382

The Bartlett (Newey-West) kernel is employed. Optimal bandwidths are calculated using Andrews (91) based on initial estimates from a VARMA(1,1).

Table 11: Regression of one-period real stock returns on the sum of the past k log dividend price ratios

$$y_{t+1} = r_{t+1}$$
$$x_t = ln\left((D_t + \ldots + D_{t-12})/P_t\right)$$
$$x_{t,k} = \sum_{j=0}^{k-1} x_{t-j}$$

Forecast Horizon (k)

$k =$	1.000	3.000	12.000	24.000	36.000	48.000
			1927 to 1994			
$\hat{\beta}$	0.013	0.005	0.001	0.001	0.001	0.000
R^2	0.004	0.006	0.007	0.008	0.007	0.007
t_β	1.899	2.184	2.395	2.464	2.346	2.280
			1927 to 1951			
$\hat{\beta}$	0.016	0.007	0.002	0.002	0.002	0.002
R^2	0.003	0.006	0.009	0.019	0.022	0.026
t_β	1.016	1.320	1.627	2.310	2.426	2.605
			1952 to 1994			
$\hat{\beta}$	0.024	0.009	0.003	0.001	0.001	0.000
R^2	0.015	0.018	0.021	0.020	0.010	0.006
t_β	2.792	3.026	3.258	3.144	2.245	1.623

Regressions are estimated by OLS

Table 12: Covariance-based orthogonality tests on one-period real stock returns using the sum of the past k log dividend price ratios

$$y_{t+1} = r_{t+1}$$
$$x_t = ln\left((D_t + \ldots + D_{t-12})/P_t\right)$$
$$x_{t,k} = \sum_{j=0}^{k-1} x_{t-j}$$

Forecast Horizon (k)

$k =$	1.000	3.000	12.000	24.000	36.000	48.000
			1927 to 1994			
t_λ	-0.7603	-0.4709	0.9442	1.6167	1.4943	1.8470
M	7.9767	19.6421	52.2798	67.3831	73.3721	77.8931
			1927 to 1951			
t_λ	0.0131	-0.4910	-0.0587	0.8633	1.3027	1.9363
M	5.3690	13.3450	38.1420	45.0255	52.7361	54.7142
			1952 to 1994			
t_λ	-1.1611	0.4537	1.9966	1.1060	0.5152	-0.1549
M	6.6822	20.5155	40.9108	50.5702	53.8202	56.2971

The Bartlett (Newey-West) kernel is employed. Optimal bandwidths are calculated using Andrews (91) based on initial estimates from a VARMA(1,1).

Table 13: Covariance-based orthogonality tests on one-period excess stock returns using the sum of the past k log dividend price ratios

$$y_{t+1} = r_{t+1}$$
$$x_t = ln\left((D_t + \ldots + D_{t-12})/P_t\right)$$
$$x_{t,k} = \sum_{j=0}^{k-1} x_{t-j}$$

Forecast Horizon (k)

$k =$	1.000	3.000	12.000	24.000	36.000	48.000
			1927 to 1994			
t_λ	-0.8736	-0.6610	0.6238	1.3907	1.4338	1.9937
M	7.9008	19.8360	53.4400	66.4861	74.4763	80.7502
			1927 to 1951			
t_λ	-0.1985	-0.7069	-0.4273	0.4111	0.7884	1.6089
M	5.2661	13.2929	38.1117	46.0110	53.2981	55.1735
			1952 to 1994			
t_λ	-1.0454	0.6150	2.3565	1.3829	0.8519	0.1502
M	6.6214	19.6746	41.1088	50.6446	53.9344	56.2212

The Bartlett (Newey-West) kernel is employed. Optimal bandwidths are calculated using Andrews (91) based on initial estimates from a VARMA(1,1).

Table 14: Covariance-based orthogonality tests on one-period real stock returns using the sum of the past k one-month treasury bill rates

$$y_{t+1} = r_{t+1}$$
$$x_t = i_t$$
$$x_{t,k} = \sum_{j=0}^{k-1} x_{t-j}$$

Forecast Horizon (k)

$k =$	1.000	3.000	12.000	24.000	36.000	48.000
			1927 to 1994			
t_λ	-0.9259	-1.9595	-2.1294	0.0952	-0.0851	-0.1499
M	5.5045	8.6988	14.3506	29.8385	79.6510	61.8166
			1927 to 1951			
t_λ	0.7627	-0.2085	-0.1921	-0.1539	-0.4502	-0.6276
M	2.3860	7.7249	97.9758	40.7324	178.5452	42.9660
			1952 to 1994			
t_λ	-1.2465	-2.5412	-3.3329	-0.0605	0.2308	0.1637
M	3.9029	6.4884	14.7001	34.7366	88.8239	64.8568

The Bartlett (Newey-West) kernel is employed. Optimal bandwidths are calculated using Andrews (91) based on initial estimates from a VARMA(1,1).

Table 15: Covariance-based orthogonality tests on long horizon real stock returns using one-month treasury bill rates

$$y_{t+k} = r_{t+1} + \ldots + r_{t+k}$$
$$x_t = i_t$$

Forecast Horizon (k)

$k =$	1.000	3.000	12.000	24.000	36.000	48.000
			1927 to 1994			
t_λ	-1.5117	-2.0220	-0.3204	-0.3652	-0.9623	-0.9220
M	5.4735	14.1007	47.7010	61.7078	71.6007	80.0715
			1927 to 1951			
t_λ	0.3851	0.2580	1.4957	-0.1954	-1.9794	-2.1226
M	2.4431	9.7259	31.2871	43.0105	53.2809	54.4925
			1952 to 1994			
t_λ	-2.1945	-3.0077	-1.1404	0.1190	0.0247	0.0854
M	5.3864	13.4467	43.4885	54.0247	63.8107	83.6159

The Bartlett (Newey-West) kernel is employed. Optimal bandwidths are calculated using Andrews (91) based on initial estimates from a VARMA(1,1).

Table 16: Covariance-based orthogonality tests on long horizon excess stock returns using one-month treasury bill rates

$$y_{t+k} = r_{t+1} + \ldots + r_{t+k}$$
$$x_t = i_t$$

Forecast Horizon (k)

$k =$	1.000	3.000	12.000	24.000	36.000	48.000
			1927 to 1994			
t_λ	-1.3171	-1.7708	-0.3147	-0.6396	-1.2892	-1.3521
M	5.5695	14.1678	49.7583	64.9060	78.7877	85.1121
			1927 to 1951			
t_λ	0.5312	0.5332	1.3079	-0.6559	-2.0636	-2.0170
M	2.3619	9.8670	33.7999	48.7350	62.4440	67.1765
			1952 to 1994			
t_λ	-2.0799	-2.9289	-1.1101	0.0759	-0.1781	-0.2163
M	5.1226	13.3057	43.3002	50.7068	55.3688	61.1098

The Bartlett (Newey-West) kernel is employed. Optimal bandwidths are calculated using Andrews (91) based on initial estimates from a VARMA(1,1).

Table 17: Covariance-based orthogonality tests on one-period real stock returns using the sum of the past k one-month treasury bill rates

$$y_{t+1} = r_{t+1}$$
$$x_t = i_t$$
$$x_{t,k} = \sum_{j=0}^{k-1} x_{t-j}$$

Forecast Horizon (k)

$k =$	1.000	3.000	12.000	24.000	36.000	48.000
			1927 to 1994			
t_λ	-0.9259	-1.9595	-2.1294	0.0952	-0.0851	-0.1499
M	5.5045	8.6988	14.3506	29.8385	79.6510	61.8166
			1927 to 1951			
t_λ	0.7627	-0.2085	-0.1921	-0.1539	-0.4502	-0.6276
M	2.3860	7.7249	97.9758	40.7324	178.5452	42.9660
			1952 to 1994			
t_λ	-1.2465	-2.5412	-3.3329	-0.0605	0.2308	0.1637
M	3.9029	6.4884	14.7001	34.7366	88.8239	64.8568

The Bartlett (Newey-West) kernel is employed. Optimal bandwidths are calculated using Andrews (91) based on initial estimates from a VARMA(1,1).

Table 18: Covariance-based orthogonality tests on one-period excess stock returns using the sum of the past k one-month treasury bill rates

$$y_{t+1} = r_{t+1}$$
$$x_t = i_t$$
$$x_{t,k} = \sum_{j=0}^{k-1} x_{t-j}$$

Forecast Horizon (k)

$k =$	1.000	3.000	12.000	24.000	36.000	48.000
			1927 to 1994			
t_λ	-0.8219	-1.6810	-1.6398	0.2551	-0.2636	-0.3875
M	5.5282	8.8894	15.5341	32.5941	65.9793	62.0156
			1927 to 1951			
t_λ	0.7217	0.1783	0.1343	0.2250	-0.3759	-0.2789
M	2.4175	7.7554	98.0647	41.3458	69.0598	42.7828
			1952 to 1994			
t_λ	-1.0129	-2.4480	-3.2474	0.0023	0.0190	-0.0822
M	3.5796	6.0396	14.4741	35.9146	94.9259	68.4612

The Bartlett (Newey-West) kernel is employed. Optimal bandwidths are calculated using Andrews (91) based on initial estimates from a VARMA(1,1).