Queen's Economics Department Working Paper No. 1122

Covariance-based orthogonality tests for regressors with unknown persistence

Alex Maynard
Wilfrid Laurier University

Katsumi Shimotsu
Queen's University

Department of Economics
Queen's University
94 University Avenue
Kingston, Ontario, Canada
K7L 3N6

2-2007

Covariance-based orthogonality tests for regressors with unknown persistence

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

February 16, 2007

Abstract

This paper develops a new test of orthogonality based on a zero restriction on the covariance between the dependent variable and the predictor. The test provides a useful alternative to regression-based tests when conditioning variables have roots close or equal to unity. In this case standard predictive regression tests can suffer from well-documented size distortion. Moreover, under the alternative hypothesis, they force the dependent variable to share the same order of integration as the predictor, whereas in practice the dependent variable often appears stationary while the predictor may be near-nonstationary. By contrast, the new test does not enforce the same orders of integration and is therefore capable of detecting alternatives to orthogonality that are excluded by the standard predictive regression model. Moreover, the test statistic has a standard normal limit distribution for both unit root and local-to-unity conditioning variables, without prior knowledge of the local-to-unity parameter. If the conditioning variable is stationary, the test remains conservative and consistent. Thus the new test requires neither size correction nor unit root pre-test. Simulations suggest good small sample performance. As an empirical application, we test for the predictability of stock returns using two persistent predictors, the dividend-price-ratio and short-term interest rate.

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

---

*Both authors are presenting authors and correspondence may be sent to either amaynard@wlu.ca (A. Maynard) or shimotsu@econ.queensu.ca (K. Shimotsu). This paper has benefited substantially from comments by Peter Phillips, Bruce Hansen and three anonymous referees. We also thank Don Andrews, Dietmar Bauer, John Chao, John Galbraith, Chuan Goh, Sainan Jin, John Maheu, James MacKinnon, Angelo Melino, Babara Rossi, John Rust, Sam Thompson, Tim Vogelsang, Yoon-Jae Whang, and participants at the 2004 North American Summer and Far Eastern Econometric Society meetings, the CESG, MESG, and seminars at the Bank of Canada, the Bank of Japan, Columbia, Concordia, Cornell, Ohio State University, the University of Guelph, the University of Maryland, the University of Michigan, Wilfrid Laurier University and Yale for useful discussion. Part of this work was completed while the authors were visiting the Cowles Foundation and we gratefully acknowledge their hospitality. Maynard thanks the SSHRC for research funding.

1 Introduction

This paper develops a new test of orthogonality based on a zero restriction on the covariance between the dependent variable and the predictor. When the predictor is stationary, this zero-covariance restriction is identical to the restriction imposed in commonly employed predictive regressions, in which the dependent variable (y_t) is regressed on the predictor (x_{t-1}). However, when the predictor has a root close or equal to unity, the predictive regression forces the dependent variable to have the same order of integration as the predictor under the alternative hypothesis. In other words, the dependent variable has a (near) unit root component that cointegrates with the predictor. Since there is often good reason to think that the dependent variable is stationary, this regression model may not provide the only relevant alternative to orthogonality. The zero-covariance restriction proposed here allows for the detection of empirically relevant alternatives to orthogonality, in which the conditioning variable (e.g. interest rates, dividend yields) may be either stationary, near nonstationary, or $I(1)$, but the dependent variable (e.g. stock returns) is presumed stationary.

Predictive regression tests are also known to suffer from substantial size distortion when regressors have roots near unity and are predetermined but not strictly exogenous (e.g. Mankiw and Shapiro (1986), Cavanagh et al. (1995), and Stambaugh (1999)). This size distortion is known to depend on the local to unity parameter and is not solved by two stage inference based on unit root pre-test (Cavanagh et al. (1995), Elliott (1998)). Another advantage of the covariance based approach is that it yields a single asymptotic t-type test that has correct size when the conditioning variable is modelled as either a unit root or local to unity process (with finite local-to-unity parameter c). The test has conservative size, but remains consistent when the conditioning variable is stationary. It thus provides a sound basis for inference without reference to prior knowledge, estimates, or pre-tests regarding the size of the root.

The size distortion problem mentioned above has recently generated an active literature aimed at correcting inference in regression-based predictive tests. In a local-to-unity context, solutions of this type include bounds procedures (Cavanagh et al. (1995), Torous et al. (2005), Valkanov (2003), Campbell and Yogo (2006)), reformulation of the problem as a stationarity test on y_t (Wright (2000), Lanne (2002)), and conditionally optimal inference employing sufficient statistics (Jansson and Moreira (2006)). Other solutions, often in more tightly parametrized models, include finite sample size corrections (Stambaugh (1999), Lewellen (2004)), augmented regression methods (Amihud and Hurvich (2004), Amihud et al. (2004)), Bayesian approaches (Elliott and Stock (1994), Stambaugh (1999), Lewellen (2004)), and resampling approaches (Nelson and Kim (1993), Goetzmann and Jorion (1993), Wolf (2000)).

Our test shares the attractive feature of the procedures described above in that it maintains good size when the predictor is persistent. This is obtained without size correction, pretest, or information on the local-to-unity parameter c. On the other hand, our test differs from the size-corrected regressions in that it is based on a different parameter restriction. As explained in the paragraphs below, this allows us to consider alternatives not covered by the regression model, including unbalanced alternatives in which y_t remains stationary despite $I(1)$ or local-to-unity behavior in x_t.

In this way, our approach differs somewhat fundamentally from that taken in much of the previous literature. The traditional framework, in which x_t is stationary and, under the alternative, y_t is a constant linear function of x_{t-1}, has intuitive appeal in many applications and is often consistent with economic or financial theory. For example, if under a failure of rational expectations, asset prices temporarily deviate from their fundamental value, the levels of valuation measures such as the earnings or dividend price ratio may have some predictive power for future market corrections. Thus, even as the recent literature has relaxed the stationarity assumptions on x_t it has generally maintained alternative specifications in which y_t depends linearly on x_{t-1}. A major advantage of this approach is that it has allowed researchers to re-examine exactly the same alternatives that have been influential in the previous literature. Thus this branch of the econometrics literature has made very important and useful contributions in both econometrics and empirical finance.

On the other hand, once we allow the possibility that x_t is (near) nonstationary, there is also an argument in favor of broadening the alternative to allow for certain more general forms of predictability. The reason is that, under the fixed-coefficient linear alternative discussed above, a (near) permanent component to x_t implies a near permanent component to y_t. However, y_t is typically a financial return and empirically such returns generally show little serial correlation. This is apparent on comparing the behavior of the real stock return series shown in the top panel of Figure 1 with that of the two persistent predictors, the log-dividend price ratio and treasury bill rate, shown in the bottom two panels.[1] Likewise, a stationary return may arguably be more appealing from the perspective of economic or financial theory. For example, in a standard rational expectations model, the predictable component in the excess return ($E_{t-1}y_t$) reflects a time varying risk premium, so that a stationary risk premium would imply a stationary return series. By explicitly testing a covariance restriction rather than a regression coefficient restriction we widen the set of alternatives to include those in which y_t remains predictable yet stationary even when x_t has a (near) unit root. These alternatives are not permitted under the regression based tests that have traditionally been employed, even when size-corrected.

Naturally, this is not the only class of alternatives that is consistent with the empirical observation that x_t typically appears (near) nonstationary whereas y_t appears (near) stationary. For example, it may be argued that local $o(n^{-1/2})$ versions of the traditional linear regression alternative sufficiently dampen the nonstationary component so as to be consistent with near-stationary returns. While our covariance test is also shown to be consistent against such alternatives, this is not the primary alternative that it was designed to capture and, consequently, it does not match the regression based test in providing consistency against local regression alternatives of

[1] A full description of the data employed for these figures is provided in Section 6.

order n^{-1}. Thus the covariance based test proposed here must properly be considered as a complement rather than a substitute for existing regression based tests.

As in our approach, there are other tests of orthogonality that are based on alternative procedures to predictive regression. In this vein, Campbell and Dufour (1995, 1997) provide exact tests of orthogonality using non-parametric sign and sign rank tests, while Toda and Yamamoto (1995) (see also Saikkonen and Lütkepohl (1996)) have also shown that, by choosing the lag order sufficiently large, one can estimate VARs formulated in levels and test general parameter restrictions even if the order of integration of the process is unknown.

There is also a recent literature seeking to model the regression imbalances ($y_t \sim I(0)$, x_t near $I(1)$) commonly observed in practice. Marmer (2004) suggests a model in which y_t appears as $I(0)$ but has a predictable component formulated as a nonlinear function of a unit root process. By rescaling the innovation variance of x_t by $\sqrt{n}$, Moon et al. (2004) model y_t as the sum of a small but persistent regressor and a large white noise error term. In a more general context, Phillips (2005) suggests a coordinate cointegration approach in terms of $L_2[0,1]$ basis functions that allows for the modelling of relationships between stochastically imbalanced variables. Our framework is arguably simpler in that we allow for an imbalance between y_t (I(0)) and x_t ($I(1)$), in which y_t is still predictable based on the past history of x_t, within the standard linear process setting.

As an empirical application, we revisit well-known orthogonality tests involving the prediction of stock returns using dividend-yields and interest rates. We find little evidence for predictability in the case of the dividend-price ratio where regression tests may suffer substantial size distortion. However, the covariance-based tests confirm the modest evidence of predictability found using the interest rate.

The remainder of the paper is organized as follows. Section 2 outlines our basic approach. Section 3 introduces the kernel-based estimator of the covariance between y_t and x_{t-1} and demonstrates its asymptotic behavior when x_t is $I(1)$, $I(0)$, and local-to-unity. Section 4 discusses inference based on the covariance estimate, and Section 5 reports some simulation results. The empirical application is reported in Section 6, and Section 7 concludes. Proofs are given in Appendix A (Section 8) and Appendix B (Section 9) collects some technical results.

2 Covariance-based orthogonality testing

We consider a test of the orthogonality or non-predictability condition

$$H_0 : E\left[y_t \,|I_{x,t-1}\right] = 0, \tag{1}$$

where $I_{x,t-1} = \sigma(x_{t-1}, x_{t-2}, x_{t-3}, \ldots)$ denotes the information contained in the past history of x_t. Several common empirical applications may be cast in this form, including tests of stock return predictability, forward rate unbiasedness,[2] the permanent

[2] See Maynard (2006) for an application to the forward rate unbiasedness test.

income hypothesis, the expectations hypothesis of the term structure, and the constant real interest rate hypothesis.[3]

In practice one often tests the simpler restriction implied by (1):

$$\operatorname{cov}(y_t, x_{t-1}) = 0. \tag{2}$$

Typically this is accomplished by modeling y_t as a linear function of x_{t-1}

$$y_t = \beta_0 + \beta_1 x_{t-1} + \varepsilon_t, \quad E[\varepsilon_t | I_{x,t-1}] = 0, \tag{3}$$

estimating β_1 by OLS, and testing the restriction $\beta_1 = 0$. In doing so, one estimates

$$\beta_1 = \frac{\operatorname{cov}(y_t, x_{t-1})}{\operatorname{var}(x_{t-1})} \tag{4}$$

by its sample analog $\widehat{\beta}_1 = \widehat{\operatorname{cov}}(y_t, x_{t-1})/\widehat{\operatorname{var}}(x_{t-1})$ and tests $\beta_1 = 0$.

Such tests were traditionally formulated with stationary regressors in mind, in which case testing $\beta_1 = 0$ is equivalent to testing $\operatorname{cov}(y_t, x_{t-1}) = 0$ because $\operatorname{var}(x_{t-1})$ in (4) is finite. However, it has come to be understood that many of the regressors, such as interest rates, dividend-price ratios, and forward premia are highly persistent and may be well characterized by roots near unity (e.g. Mankiw and Shapiro (1986), Stambaugh (1999)). Moreover, in practice there is usually considerable uncertainty regarding the size of the largest root in the regressor, with confidence intervals frequently containing both one and values considerably below unity. On the other hand, it often occurs that y_t (e.g. a stock or exchange rate return) appears stationary on empirical and/or a priori theoretical grounds.

This suggests a potential imbalance between the stationary dependent variable and possibly near $I(1)$ regressor in (3), which may cause some difficulty when (3) is used to test for the orthogonality condition in (1). Suppose, for example, that x_t is $I(1)$ and y_t is $I(0)$. In this case, only one value of β_1, i.e. $\beta_1 = 0$, can satisfy the regression model in (3), because there cannot exist a simple linear relationship between $I(0)$ and $I(1)$ variables. Thus, were to take the model in (3) literally, by forcing $\beta_1 = 0$ this imbalance would itself imply that y_t was unpredictable.

Moreover, when y_t is $I(0)$ and x_t is $I(1)$ or near $I(1)$, the equivalence between a test of $\beta_1 = 0$ and a test of $\operatorname{cov}(y_t, x_{t-1}) = 0$ breaks down. This is due to the fact that $\widehat{\beta}_1 = \widehat{\operatorname{cov}}(y_t, x_{t-1})/\widehat{\operatorname{var}}(x_{t-1})$ converges to zero in probability regardless of whether the restriction $\operatorname{cov}(y_t, x_{t-1}) = 0$ holds, simply because $\widehat{\operatorname{cov}}(y_t, x_{t-1}) = O_p(1)$, while $\widehat{\operatorname{var}}(x_{t-1}) = O_p(n)$. In other words, tests based on $\beta_1 \neq 0$ are not designed to detect alternatives to (1) in which y_t is stationary and x_t is nonstationary. In fact, under the alternative implied by $\beta_1 \neq 0$ in (3), when x_t is $I(1)$, y_t is both nonstationary and cointegrated with x_t.

Outside the simple regression framework in (3), the fact that x_t and y_t have different orders of integration does not in itself imply that y_t is unpredictable by the

[3]See Mankiw and Shapiro (1986) and references therein.

past history of x_t. In fact, when y_t is $I(0)$ and x_t is $I(1)$ a more general DGP is given by the joint linear process

$$\begin{pmatrix} y_t \\ \Delta x_t \end{pmatrix} = A(L)\varepsilon_t = \sum_{j=0}^{\infty} A_j \varepsilon_{t-j}, \quad \sum_{j=0}^{\infty} j||A_j|| < \infty, \quad \varepsilon_t \sim \text{i.i.d.}(0, I_2), \tag{5}$$

in which (1) holds only for very specific parameter values.[4] One economic interpretation of this alternative is that the past history of x_t has explanatory power for the stationary risk premium ($E_{t-1}y_t$) and thus predictive power for the excess return y_t.

The covariance condition in (2) allows for a more flexible test than the restriction that $\beta_1 = 0$ in (3). Clearly, when x_t is stationary and $\text{var}(x_{t-1})$ is finite the two conditions are equivalent (see (4)). In the more interesting case, in which y_t is $I(0)$ but x_t is $I(1)$ and properly initialized, $\text{cov}(y_t, x_{t-1})$ is time varying, but still well-defined. Likewise, the orthogonality condition in (1) continues to imply that $\text{cov}(y_t, x_{t-1}) = 0$ for all t. Finally, when (1) fails, this restriction is typically violated, i.e. $\text{cov}(y_t, x_{t-1}) \neq 0$ in general. In other words, the restriction $\text{cov}(y_t, x_{t-1}) = 0$ admits a wide class of alternatives to (1), including alternatives not covered by a test of $\beta_1 = 0$, in which y_t is stationary but x_t is $I(1)$, as in (5).

Therefore, we propose to test directly the condition $\text{cov}(y_t, x_{t-1}) = 0$ instead of testing $\beta_1 = 0$. However, since $\text{cov}(y_t, x_{t-1})$ depends on t when x_t is $I(1)$, we base our test on the *limiting covariance*, which we define as $\lim_{t\to\infty} \text{cov}(y_t, x_{t-1})$. When y_t and x_{t-1} are stationary the limiting covariance is simply the covariance between y_t and x_{t-1}. When y_t is $I(0)$ and x_{t-1} is $I(1)$ it provides an asymptotic approximation to their covariance. Since the orthogonality condition (1) implies that $\text{cov}(y_t, x_{t-1}) = 0$ for all t it also imposes the testable condition that $\lim_{t\to\infty} \text{cov}(y_t, x_{t-1}) = 0$, regardless of whether x_t is $I(0)$ or $I(1)$. On the other hand, as explained below, the limiting covariance is generally non-zero when (1) fails, even when y_t is $I(0)$ and x_t is $I(1)$, as in (5).

The limiting covariance between y_t and x_{t-1} may also be usefully reformulated in terms of the one-sided long-run covariance between y_t and Δx_{t-1}. This interpretation is helpful in establishing its properties. It also motivates the kernel estimator that we employ later. Assume $x_0 \equiv 0$ for simplicity and that $(y_t, \Delta x_t)$ is covariance stationary. Because x_{t-1} may be written as the sum of its past differences as $x_{t-1} = \sum_{h=1}^{t-1} \Delta x_{t-h}$, the covariance takes the form

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

and the limiting covariance may be expressed as

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

[4] Since we impose no a priori restriction on the largest root of x_t, we also consider the DGPs in which (y_t, x_t) follows a joint linear process, as well as local-to-unity models for x_t.

which is well-defined so long as $\sum_{h=0}^{\infty} |\text{cov}(y_t, \Delta x_{t-h})| < \infty$.[5] When x_t is stationary, (6) simplifies to $\lambda_{y,\Delta x} = \lim_{t\to\infty} \text{cov}(y_t, x_{t-1}) = \text{cov}(y_t, x_{t-1})$. Finally, if x_t is modelled local-to-unity as

$$x_t = \left(1 + \frac{c}{n}\right) x_{t-1} + u_t, \quad t = 1, 2, \ldots, n, \quad n = 1, 2, \ldots \quad c < 0, \tag{7}$$

with $x_t \equiv 0$ for $t \leq 0$, then as shown in Appendix A, (6) takes the form

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

and is well-defined when $\sum_{h=1}^{\infty} h|\text{cov}(y_t, u_{t-h})| < \infty$.

The (limiting) covariance cannot be consistently estimated by the sample covariance of y_t and x_{t-1}, which converges to a random variable when x_t is $I(1)$. Instead, we estimate (6) by a standard kernel estimator, which is consistent for both stationary and nonstationary x_t, without the necessity of pretesting or estimating the root of x_t. This feature may be useful in applied work, as it is often difficult to distinguish with confidence between $I(0)$ and $I(1)$ alternatives. A second desirable property of the estimator is that it is shown to have the same limit distribution for all finite values of the local to unity parameter c. This allows us to avoid two-stage inference procedures, such as Bonferroni bounds, that are often necessitated by the lack of a consistent time-series estimator for c. We construct a large sample test, based on a single test statistic with a limiting standard normal distribution under both unit root and local to unity assumptions. No bias corrections or other adjustments are required. The test is shown to remain conservative and consistent when x_t is stationary.

Unlike the regression model in (3), our test is designed to have power against a general class of alternatives to (1) under which y_t is stationary and predictable based on the past history of x_{t-j} $j \geq 1$, regardless of whether x_t is $I(0)$, $I(1)$, or local to unity. For example, in the unbalanced case given by (5), the limiting covariance is given by $\lambda_{y,\Delta x} = \sum_{h=1}^{\infty} \sum_{j=0}^{\infty} A_j^1 (A_{j-h}^2)'$, where A_j^i denotes the ith row of A_j, and is generally non-zero when y_t may be predicted by past x_t. For instance, a simple example is

$$y_t = \gamma_0 + \gamma_1 (x_{t-1} - x_{t-2}) + \varepsilon_{1t}, \quad \gamma_1 \neq 0, \tag{9}$$

in which case $\lambda_{y,\Delta x} = \gamma_1 \sum_{h=1}^{\infty} E(\Delta x_{t-1} \Delta x_{t-h})$. Our test also remains generally consistent against the regression alternative specified in (3) with $\beta_1 \neq 0$. By contrast, tests based on $\widehat{\beta}_1$ are not designed to have power against alternatives such as (5) or (9), in which y_t is $I(0)$, but x_t may be $I(1)$.

On the other hand, tests based on $\widehat{\beta}_1$ naturally have better power if (3) is the correctly specified model. Likewise, if (9) is the correct model, then a regression of y_t on Δx_{t-1} will of course have better power. However, in practice it is often difficult to determine a priori which model is correct and confidence intervals on the largest

[5] When $x_0 \neq 0$, $\text{cov}(y_t, x_{t-1})$ is defined as $\sum_{h=1}^{t-1} \text{cov}(y_t, \Delta x_{t-h}) + \text{cov}(y_t, x_0)$ and (6) continues to apply under the relatively weak and reasonable assumption that $\lim_{t\to\infty} \text{cov}(y_t, x_0) \to 0$.

root of x_t are typically wide. Even if we restrict ourselves to the simplest parametric models in (3) and (9), the model selection exercise is non-trivial, as it involves a unit root pre-test that may complicate the second-stage inference. Two-step procedures that involve unit-root pretesting are known to suffer from size distortion when the roots may be close to unity (e.g. Cavanagh et al. (1995, Table 1)).

3 Estimation of limiting covariance

In this section, we develop an estimator of the limiting covariance between y_t and x_{t-1}, as defined in (6), and derive its asymptotic properties. First we consider the case when x_t is $I(1)$.

Assumption A

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

$$z_t = \begin{pmatrix} y_t \\ \Delta x_t \end{pmatrix} = A(L)\varepsilon_t = \sum_{j=0}^{\infty} A_j \varepsilon_{t-j}, \quad \sum_{j=0}^{\infty} j^{\delta} ||A_j|| < \infty, \quad \delta > 1; \tag{10}$$

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

$$\sum_{h=-\infty}^{\infty} |h|^{\delta} \|\Gamma(h)\| < \infty, \quad \Gamma(h) = \begin{bmatrix} \Gamma_{yy}(h) & \Gamma_{y\Delta x}(h) \\ \Gamma_{\Delta xy}(h) & \Gamma_{\Delta x \Delta x}(h) \end{bmatrix} = E z_t z_{t+h}',$$

where $||A|| = (\text{tr}(A'A))^{1/2}$ is the Euclidean norm of a matrix A.

The assumption that $\text{var}(\varepsilon_t) = I_2$ is innocuous because we do not normalize the elements of A_j. Note that under these assumptions y_t is stationary, although x_t may be $I(1)$. However, the orthogonality condition (1) holds only under particular parameter choices, for example, when the off-diagonal elements of A_j are zero for all $j > 0$. The limiting covariance between y_t and x_{t-1}, which, under the above assumptions, may be expressed as $\lambda_{y,\Delta x} = \lim_{t\to\infty} \text{cov}(y_t, x_{t-1}) = \sum_{h=1}^{\infty} \text{cov}(y_t, \Delta x_{t-h}) = \sum_{h=1}^{\infty} \sum_{j=0}^{\infty} A_j^1 (A_{j-h}^2)'$ is zero when (1) holds, but is generally non-zero otherwise.

It is well known that when x_t is $I(1)$ the sample covariance between y_t and x_{t-1} converges to a random variable. Thus we propose instead to estimate the limiting covariance between y_t and x_{t-1} based on its reformulation in (6). This expression suggests the following one-sided kernel covariance estimator

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

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

[6] We use $k((h-1)/m)$ instead of $k(h/m)$ in the definition $\widehat{\lambda}_{y,\Delta x}$ so that the leading term, $\widehat{\Gamma}_{\Delta xy}(1)$, is multiplied by 1, instead of $k(1/m)$, leading to a smaller bias, especially when m is small.

Assumption K

The kernel $k(x)$ is continuous at $x = 0$ and uniformly bounded with $k(0) = 1$, $\int_0^\infty \bar{k}(x)x^2 dx < \infty$, and $\lim_{x\to 0+} \frac{1-k(x)}{|x|^q} = k_q < \infty$ with $\delta \geq q$, where $\bar{k}(x) = \sup_{y\geq x}|k(y)|$.

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 condition $\int_0^\infty \bar{k}(x)x^{1/2}dx < \infty$ is necessary to rule out pathological cases where the convergence of $m^{-1}\sum_{h=1}^{n-1} k(h/m)$ to $\int_0^\infty k(x)dx$ fails, for instance $k(x) = 1$ for any integer x but $k(x)$ is the same as the QS kernel otherwise. See Lemma 1 of Jansson (2002) and the discussion therein.

Let $f_{yy}(\lambda)$ denote the spectral density of y_t, let $f_{\Delta xy}(\lambda)$ denote the cross-spectral density between Δx_t and y_t, and define $f_{y\Delta x}(\lambda)$ and $f_{\Delta x\Delta x}(\lambda)$ analogously. The following lemma provides the asymptotic bias and variance of $\widehat{\lambda}_{y,\Delta x}$ and shows its consistency.

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

(a) $\lim_{n\to\infty} m^q E(\widehat{\lambda}_{y,\Delta x} - \lambda_{y,\Delta x}) = -k_q \sum_{h=1}^\infty \Gamma_{\Delta xy}(h)h^q$,
(b) $\lim_{n\to\infty} nm^{-1} var(\widehat{\lambda}_{y,\Delta x}) = V \equiv 4\pi^2 f_{yy}(0) f_{\Delta x\Delta x}(0) \int_0^\infty k^2(x)\,dx$,
(c) $\widehat{\lambda}_{y,\Delta x} \to_p \lambda_{y,\Delta x}$ *as* $n \to \infty$.

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

Remark 1 $\lambda_{y,\Delta x} \neq 0$ *provided that* $\sum_{h=1}^\infty \sum_{j=0}^\infty A_j^1 (A_{j-h}^2)' \neq 0$.

Remark 2 *It is well known that the limiting variance of the two-sided long-run covariance estimate between* y_t *and* Δx_t *is given by* $4\pi^2 \int_{-\infty}^\infty k^2(x)dx\{f_{yy}(0)f_{\Delta x\Delta x}(0) + [f_{y\Delta x}(0)]^2\}$. *Hence, in the special case where* $y_t = \Delta x_t$, V *is 1/4 of the limiting variance for the two-sided case.*

Remark 3 *From Lemma 1, 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

$$\begin{aligned} m^* &= \left(qk_q^2 \alpha(q) n \Big/ \int_{-\infty}^{\infty} k^2(x)\, dx \right)^{1/(2q+1)}, \qquad (12) \\ \alpha(q) &= \frac{4\left((2\pi)^{-1} \sum_{h=1}^{\infty} \Gamma_{\Delta xy}(h) h^q\right)^2}{f_{yy}(0)\, f_{\Delta x \Delta x}(0)}, \end{aligned}$$

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

Interestingly, when the truncated kernel ($k(x) = 1, |x| \leq 1$, 0 otherwise) is employed, we obtain $\hat{\lambda}_{y,\Delta x} = \widehat{\mathrm{cov}}(y_t, x_{t-1} - x_{t-m-1})$, which equals the numerator of the coefficient in the regression of y_t on $x_{t-1} - x_{t-m-1}$. For fixed $m = 1$ this simplifies to a regression of y_t on Δx_{t-1}. A test based on such a regression might appear simpler than the one proposed here. However, this would test only the restriction $\mathrm{cov}(y_t, \Delta x_{t-1}) = 0$, which is quite distinct from $\lim_{t\to\infty} \mathrm{cov}(y_t, x_{t-1}) = 0$, the restriction tested by our procedure. The requirement that $m \to \infty$ in Assumption M is necessary for the consistent estimation of the limiting covariance between y_t and x_{t-1}, without which a valid test of the restriction in (2) is not possible.

On the other hand, one might also consider a test based on the regression of y_t on $x_{t-1} - x_{t-m}$, while allowing $m \to \infty$. Unfortunately, not only would the choice of m pose a difficulty, but more fundamentally, such a regression entails a drawback similar to that of the regression of y_t on x_{t-1}. In particular, when y_t is $I(0)$ and x_t is $I(1)$, it follows from Lemma 1 and the divergence of $\widehat{\mathrm{var}}(x_{t-1} - x_{t-m-1})$ that the regression coefficient converges to 0 in probability regardless of whether or not (1) holds.

3.1 The limit distribution when x_t is $I(1)$

It is well known that the estimator of the two-sided long-run covariance between y_t and Δx_t has a normal limiting distribution. However, currently there are no corresponding distributional results for the one-sided long-run covariance estimator, partly because its analysis is substantially more involved than that of its two-sided counterpart. To see why, let Σ and Λ denote the contemporaneous covariance and the one-sided long-run covariance matrix, so that

$$\begin{aligned} \Sigma &= E\left[z_t z_t'\right] = \Gamma(0), \quad \Lambda = \sum_{h=1}^{\infty} \Gamma(h) = \begin{bmatrix} \Lambda_{yy} & \Lambda_{y\Delta x} \\ \Lambda_{\Delta xy} & \Lambda_{\Delta x \Delta x} \end{bmatrix}, \\ \sum_{h=-\infty}^{\infty} \Gamma(h) &= 2\pi f_z(0) = \Sigma + \Lambda + \Lambda'. \end{aligned}$$

The limiting distribution of spectral density estimators has been studied widely in the statistics literature and it is well known that (e.g. Hannan, 1970, Theorem 11, p.289)

$$\sqrt{\frac{n}{m}} \left(\sum_{h=-n+1}^{n-1} k\left(\frac{h}{m}\right) \widehat{\Gamma}(h) - 2\pi f_z(0) \right) \to_d N(0, \Phi),$$

where Φ is finite. Let the one-sided long-run covariance estimate be $\widehat{\Lambda} = \sum_{h=1}^{n-1} k(h/m)\,\widehat{\Gamma}(h)$, where $\widehat{\Gamma}(h)$ is the hth sample autocovariance of z_t. Deriving the limiting distribution of the *diagonal* elements of $\widehat{\Lambda}$ is easy, because we can use the identities

$$\Lambda_{yy} = \sum_{h=1}^{\infty} E(y_t y_{t+h}) = \frac{1}{2}\left[\sum_{h=-\infty}^{\infty} E(y_t y_{t+h}) - E y_t^2\right], \quad \widehat{\Lambda}_{yy} = \frac{1}{2}\left[\sum_{h=-n+1}^{n-1} k\left(\frac{h}{m}\right)\widehat{\Gamma}_{yy}(h) - \widehat{\Gamma}_{yy}(0)\right], \tag{13}$$

where $\widehat{\Gamma}_{yy}(h)$ is the $(1,1)$th element of $\widehat{\Gamma}(h)$, and then $\sqrt{n/m}(\widehat{\Lambda}_{yy} - \Lambda_{yy}) \to_d N(0, \Phi_{11}/4)$. However, deriving the asymptotics of the *off-diagonal* elements of $\widehat{\Lambda}$ is not trivial, because $E(y_t \Delta x_{t+h}) \neq E(y_t \Delta x_{t-h})$ and hence the identity in (13) does generalize to the off-diagonal elements:

$$\Lambda_{y\Delta x} = \sum_{h=1}^{\infty} E(y_t \Delta x_{t+h}) \neq \frac{1}{2}\left[\sum_{h=-\infty}^{\infty} E(y_t \Delta x_{t+h}) - E y_t \Delta x_t\right].$$

Therefore, we need to return to the original derivation. The limiting distributions of spectral density estimators have traditionally been analyzed using their representation in terms of periodograms (e.g. Hannan, 1970, Theorem 11):

$$\begin{aligned}
\sum_{h=-n+1}^{n-1} k\left(\frac{h}{m}\right)\widehat{\Gamma}(h) &= \sum_{h=-n+1}^{n-1} k\left(\frac{h}{m}\right)\int_{-\pi}^{\pi} I_z(\omega) e^{i\omega h} d\omega \\
&= \int_{-\pi}^{\pi} K_n(\omega)\, I_z(\omega) d\omega \approx \sum_{j=-n}^{n} K_n(\omega_j)\, I_z(\omega_j)\,\frac{2\pi}{n},
\end{aligned}$$

where $K_n(\omega) = \sum_{h=-n+1}^{n-1} k(h/m)\, e^{i\omega h}$ (frequency window) and $\omega_j = 2\pi j/n$. Under standard regularity conditions, $K_n(\omega)$ approaches a delta-function as $n \to \infty$, and since $I_z(\omega_j)$ are asymptotically independent, we obtain the asymptotic normality of $\sum_{h=-n+1}^{n-1} k(h/m)\,\widehat{\Gamma}(h)$.

However, this approach does not work in the one-sided case when the summation of $\widehat{\Gamma}(h)$ is taken only for positive h. Specifically,

$$\begin{aligned}
\sum_{h=1}^{n-1} k\left(\frac{h}{m}\right)\widehat{\Gamma}(h) &= \sum_{h=1}^{n-1} k\left(\frac{h}{m}\right)\int_{-\pi}^{\pi} I_z(\omega) e^{i\omega h} d\omega \\
&= \int_{-\pi}^{\pi} \tilde{K}_n(\omega) I_z(\omega) d\omega, \quad \tilde{K}_n(\omega) = \sum_{h=1}^{n-1} k\left(\frac{h}{m}\right) e^{i\omega h}.
\end{aligned}$$

Unlike $K_n(\omega)$, the one-sided $\tilde{K}_n(\omega)$ does not have a simple expression such as a Fejér kernel. In particular, it has a nonnegligible imaginary part, because it involves only positive h. The fact that there is no published result on the limiting distribution of $\widehat{\Lambda}_{y\Delta x}$, despite its importance in econometrics, also suggests its difficulty.

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.

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

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

Unlike the regression-based tests, neither a non-zero intercept in (y_t, x_t) nor a linear trend in x_t affects the limiting distribution. Because $\Delta(x_t + \mu) = \Delta x_t$, $\widehat{\lambda}_{y,\Delta x}$ is invariant to the presence of a non-zero intercept in x_t. For a non-zero intercept in y_t and a linear trend in x_t, if we replace $(y_t, \Delta x_{t-h})$ with $(y_t - \overline{y}, \Delta x_{t-h} - \overline{\Delta x})$, where $\overline{y}$ and $\overline{\Delta x}$ denote the sample average of y_t and Δx_t, then $\widehat{\lambda}_{y,\Delta x}$ has the stated asymptotic distribution.[7]

The optimal bandwidth m^* in (12) does not satisfy the rate condition on m of Theorem 2, which is a standard result when the bandwidth is chosen to minimize the mean squared error. m needs to grow faster than m^* for Theorem 2 to hold. Since the optimal rate of increase of m is $n^{1/(2q+1)}$ from Remark 3, the upper bound on m, $m^2/n \to 0$, does not appear to pose a severe problem when q is 1 or 2.

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

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

Assumption B

$$\begin{aligned}
x_t &= (1 + c/n)\, x_{t-1} + u_t, \quad t = 1, 2, \ldots, n, \quad n = 1, 2, \ldots \quad c < 0, \\
x_t &\equiv 0 \quad \text{for } t \le 0, \\
z_t^* &= (y_t, u_t)' \text{ satisfies Assumption A.}
\end{aligned}$$

Then $\lambda_{y,\Delta x} = \sum_{h=1}^{\infty} \text{cov}(y_t, u_{t-h}) + O(n^{-1})$ as seen in (8). The following 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 on x_t.

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

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

[7] If y_t is trend stationary, employing the detrended residual gives the same limiting distribution.

3.3 The limit distribution when x_t is $I(0)$

The argument so far is based on the assumption that x_t is $I(1)$/local-to-unity. However, in practice often we do not have strong prior knowledge as to whether x_t is $I(1)$ or $I(0)$. With an additional Lipschitz continuity assumption on the kernel, $\widehat{\lambda}_{y,\Delta x}$ converges to $Ey_tx_{t-1} = \lambda_{y,\Delta x}$ when x_t is an $I(0)$ process. Let us first state the assumptions on x_t and y_t.

Assumption C

$$
\begin{aligned}
v_t &= \begin{pmatrix} y_t \\ x_t \end{pmatrix} = B(L)\,\varepsilon_t = \sum_{j=0}^{\infty} B_j \varepsilon_{t-j}, \quad \sum_{j=0}^{\infty} j||B_j|| < \infty, \qquad (14)\\
\varepsilon_t &\sim \text{ i.i.d.}\,(0, I_2)\,, \quad \text{with finite fourth moment}\\
\sum_{-\infty}^{\infty} |h|^{\delta}\,\|\gamma(h)\| &< \infty, \quad \delta > 1; \quad \gamma(h) = \begin{bmatrix} \gamma_{yy}(h) & \gamma_{yx}(h) \\ \gamma_{xy}(h) & \gamma_{xx}(h) \end{bmatrix} = Ev_t v_{t+h}',
\end{aligned}
$$

and $\tilde{f}_{xx}(0), \tilde{f}_{yy}(0) > 0$, where $\tilde{f}_{xx}(\lambda)$ and $\tilde{f}_{yy}(\lambda)$ are the spectral densities of x_t and y_t.

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

Lemma 4 *If Assumptions C, K and M hold and $k(x)$ is Lipschitz(1), then*

$$
\sqrt{n}\left(\widehat{\lambda}_{y,\Delta x} - \lambda_{y,\Delta x}\right) = \sqrt{n}\left(\widehat{\gamma}_{xy}(1) - \gamma_{xy}(1)\right) + B_n + o_p(1), \qquad (15)
$$

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

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

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

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

and $k_{xyxy}(0,a,b,c)$ is the fourth cumulant of $(x_t, y_{t+a}, x_{t+b}, y_{t+c})'$.

If one knew x_t were $I(0)$, then one would estimate Ey_tx_{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 order of integration, apart from the bias term in (15), which disappears when (1) holds.

4 Possible ways to conduct inference

4.1 Estimation of the limiting variance of the estimator

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

We may also consider another estimator of V, $\tilde{V}$, whose particularly good performance is suggested by the simulations in Section 5. It is based on the exact finite sample variance of $\widehat{\lambda}_{y,\Delta x}$, which is given by (see equations (34)-(36) in the proof of Lemma 1)

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

where $\phi_n(u,h',h)$ is defined in the proof of Lemma 1. Since the terms involving $\Gamma_{\Delta xy}(\cdot)\Gamma_{y\Delta x}(\cdot)$ and the cumulants disappear in the limit and $\phi_n(u,h',h) = 1$ for the relevant values of (u,h',h), define $\tilde{V}$ as:

$$
\begin{aligned}
\tilde{V} \; = \; & \frac{1}{m}\sum_{h'=1}^{n-1}\sum_{h=1}^{n-1} k\left(\frac{h'-1}{m}\right) k\left(\frac{h-1}{m}\right) \\
& \times \sum_{u=-\infty}^{\infty} \tilde{k}\left(\frac{u}{\tilde{m}}\right) \widehat{\Gamma}_{\Delta x \Delta x}(u)\, \tilde{k}\left(\frac{u+h-h'}{\tilde{m}}\right) \widehat{\Gamma}_{yy}(u+h-h'), \qquad (16)
\end{aligned}
$$

where $\tilde{k}(x)$ is a kernel and $\tilde{m}$ is a bandwidth. $\tilde{k}(x)$ and $\tilde{m}$ can, but do not need to, be the same as $k(x)$ and m.[8] Estimating V by $\tilde{V}$ reduces the error from the approximation of the discrete sum in (32) by an integral and gives better finite sample performance than estimating V by $\widehat{V}$. (The results using $\widehat{V}$ are not reported in the present paper).

Suppose that $(y_t, \Delta x_t)$ satisfies Assumption A and hence x_t is $I(1)$. Then, we may test hypotheses on $\lambda_{y,\Delta x}$ using a t-type statistic of the form

$$
t_\lambda = \frac{\sqrt{\frac{n}{m}}\left(\hat{\lambda}_{y,\Delta x} - \lambda_{y,\Delta x}\right)}{\sqrt{\tilde{V}}}, \qquad (17)
$$

[8] $\tilde{V}$ is not guaranteed to be always positive, although we encountered no case where $\tilde{V}$ is nonpositive in the simulations we conducted. One may use another estimate that is always positive:

$$
\widetilde{V}_2 = \frac{1}{m}\sum_{h=1}^{n-1} k^2\left(\frac{h-1}{m}\right) \sum_{u=-n+1}^{n-1} \widetilde{k}\left(\frac{u}{\widetilde{m}}\right) \widehat{\Gamma}_{\Delta x \Delta x}(u) \sum_{v=-n+1}^{n-1} \widetilde{k}\left(\frac{v}{\widetilde{m}}\right) \widehat{\Gamma}_{yy}(v).
$$

However, $\widetilde{V}_2$ did not perform as well as $\tilde{V}$ in simulations.

where $\hat{\lambda}_{y,\Delta x}$ and $\tilde{V}$ are defined in (11) and (16) respectively. Of course orthogonality (1) implies the null $\lambda_{y,\Delta x} = 0$.[9] The following Lemma and its corollary show that t_λ converges to a $N(0,1)$ random variable.

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

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

4.2 Conservative inference when x_t is $I(0)$

Consider the case in which (y_t, x_t) follows (14) and thus x_t is $I(0)$. It is easy to show that the test based on t_λ remains consistent. Suppose that we test $H_0 : \lambda_{y,\Delta x} = Ey_t x_{t-1} = 0$ but $\lambda_{y,\Delta x} \neq 0$. Then we have, from Lemma 4,

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

Since $\tilde{V} \to_p f_{\Delta x \Delta x}(0) = 0$ and $n^{1/2} m^{-1/2} \to \infty$, it follows that $|t_\lambda| \to \infty$ as $n \to \infty$. Indeed, t_λ diverges at a faster rate than $n^{1/2} m^{-1/2}$, the rate of divergence in the $I(1)$ case.

Although Corollary 6 does not hold when x_t is $I(0)$, we still have $t_\lambda \to_p 0$ when $\tilde{m}$ and $\tilde{k}(x)$ are chosen appropriately. It thus serves as a tool for conservative inference even when x_t is $I(0)$. In particular, if the Bartlett kernel $\tilde{k}(x) = (1 - |x|)\mathbf{1}\{|x| \leq 1\}$ is used in $\tilde{V}$ in (16) and $Ey_t x_{t-h} = 0$ for all $h \geq 2$ (which holds under the null hypothesis of orthogonality), then t_λ is $O_p((\tilde{m}/m)^{1/2})$.

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

In order to understand the convergence, rewrite t_λ as

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

The numerator converges to a Gaussian random variable from Lemma 4. $\tilde{V}$ in the denominator is an estimate of $f_{\Delta x \Delta x}(0) = 0$ and hence converges to 0 as $\tilde{m} \to \infty$. Because m tends to infinity, the asymptotic behavior of t_λ depends on the rate of convergence of $\tilde{V}$. Letting $\tilde{m}$ tend to infinity but not too quickly prevents $\tilde{V}$ from converging to 0 too fast and makes t_λ converge to 0 in probability.

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

[9] A detailed description of the implementation is given in Section 5.

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

4.3 Power against nonstationary regression alternatives

Based on the empirical observation that the dependent variables in predictive tests (e.g. returns) typically appear stationary, we have so far focused on the models where y_t is assumed to be $I(0)$. As discussed in Section 2, for fixed values of β_1, this rules out regression-type alternatives where x_t is $I(1)$ and y_t is linearly related to x_{t-1} as in (3). However, under local alternatives for which β_1 is of order $n^{-1/2}$ or smaller, the variance of the nonstationary component $\beta_1 x_{t-1}$ no longer dominates that of the stationary residual. Thus such local alternatives are potentially consistent with the observation that y_t appears essentially stationary in its behavior.

Thus we next examine the power of our covariance-based test against the following regression alternative, which includes both fixed and local alternatives.

Assumption D

(y_t, x_t) is generated by

$$y_t = \beta_n x_{t-1} + u_t,$$

where $(u_t, \Delta x_t)'$ is generated by a linear process $A(L)\varepsilon_t$ that satisfy Assumption A and $x_0 = O_p(1)$. $\Gamma_{uu}(h), \Gamma_{u\Delta x}(h), \Gamma_{\Delta x u}(h)$, and $\Gamma_{\Delta x \Delta x}(h)$ are defined analogously to Assumption A and they satisfy the summability condition in Assumption A.

We exclude the constant term in order to simplify the discussion. Under Assumption D, we have $\text{cov}(y_t, x_{t-1}) = \beta_n \text{var}(x_{t-1}) + \text{cov}(u_t, x_{t-1})$, which is not equal to zero in general. Indeed, if $1/\beta_n = o(n)$ then $\text{cov}(y_t, x_{t-1})$ diverges at the rate $n\beta_n$. The following lemma summarizes the behavior of $\hat{\lambda}_{y,\Delta x}$ and $\tilde{V}$.

Lemma 9 *If Assumptions D, K, and M hold, $\lambda_{u,\Delta x} = 0$ and $m^2/n \to 0$, then*

$$(a) \quad \hat{\lambda}_{y,\Delta x} = m\beta_n \left[(1/2)\left(B^2(1) + 2\pi f_{\Delta x \Delta x}(0)\right)\right] \int_0^1 k(x)dx\,(1 + o_p(1)).$$

If, in addition, $\tilde{m}^2/n \to 0$ and $n\beta_n \to \infty$, then

$$(b) \quad \tilde{V} = \left\{\left(\beta_n^2 n\tilde{m} \int_0^1 \tilde{k}(x)dx \int_0^1 B^2(r)dr(1 + o_p(1))\right) + 2\pi f_{uu}(0)\right\} \times 2\pi f_{\Delta x \Delta x}(0) \int_0^\infty k^2(x)\,dx(1 + o_p(1)).$$

$\hat{\lambda}_{y,\Delta x}$ converges to a non-degenerate limit if $m\beta_n$ is bounded away from 0. The behavior of $\tilde{V}$ is dominated either by the term with $\beta_n^2 n\tilde{m}$ or $2\pi f_{uu}(0)$. Assume $\tilde{m}/m \to 0$ as assumed in Lemma 8, and let ξ_n denote a generic random variable that has a non-degenerate limiting distribution. If $\beta_n^2 n\tilde{m} \to \infty$, then $\tilde{V}$ diverges and $t_\lambda = (n/m)^{1/2}\hat{\lambda}_{y,\Delta x}(\tilde{V})^{-1/2} = (m/\tilde{m})^{1/2}\xi_n$, which diverges to infinity. If β_n tends to 0 faster and $\beta_n^2 n\tilde{m} \to 0$, then $\tilde{V}$ is $O_p(1)$ and $t_\lambda = \beta_n n^{1/2} m^{1/2}\xi_n$.

Consequently, the covariance-based test has a nonnegligible power against local alternatives of the form $\beta_n \sim bn^{-1/2}m^{-1/2}$ with $b \neq 0$. Since regression-based tests have power against alternatives for which $\beta_n \sim bn^{-1}$, the covariance-based test loses power relative to regression-based tests by the order of $(n/m)^{1/2}$ as the price of its greater flexibility. This reflects the fact that t_λ uses m sample autocovariances whereas regression-based tests use effectively n sample autocovariances. Despite the loss of power, the t_λ-test is still consistent against the alternatives with $\beta_n \sim bn^{-1/2}m^{-c}$ for $c \in (0, 1/2)$, in which the variance of the $I(1)$ component is dominated by the variance of the residual u_t and y_t may appear to be $I(0)$. This result is summarized in the following Corollary, which follows immediately from Lemma 9.

Corollary 10 *If the assumptions of Lemma 9 hold and $\tilde{m}/m \to 0$, then the test statistic t_λ for the test of $H_0 : \lambda_{y,\Delta x} = 0$ has a nonnegligible power against alternatives in Assumption D of the form $\beta_n = b\,(nm)^{-1/2}$ with $b \neq 0$ and diverges against alternatives of the form $\beta_n = bn^{-1/2}m^{-c}$ for $c \in (0, 1/2)$.*

5 Finite sample performance: simulation results

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

For the simulations below we have in mind a test of y_t orthogonal to $I_{x,t-1}$, the information contained in past x_t, as in (1). This is often tested in practice using a regression of y_t on x_{t-1} as in (3). Since size distortions rule out standard regression only for x_t highly serially correlated, it is this case that we focus on. In particular, we consider both first and second order autoregressive models for x_t:

$$x_t = \rho_0 + \rho_1 x_{t-1} + u_{2t}, \quad \text{AR(1)} \tag{18}$$

$$x_t = \rho_0 + \rho_1 x_{t-1} + \rho_2 x_{t-2} + u_{2t}. \quad \text{AR(2)} \tag{19}$$

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

$$\rho_1 = 1 + c/n \ , \ c \leq 0. \tag{20}$$

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

$$y_t = d_t + u_{1t}, \tag{21}$$

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

$$d_t = \delta_0 \quad \text{or} \tag{22}$$
$$d_t = \delta_0 + \delta_1 t. \tag{23}$$

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

$$y_t = d_t + \beta x_{t-1} + u_{1t}. \tag{24}$$

In the unit root/local to unity context, (24) implies that y_t and x_t contain an equally persistent component and are cointegrated when $\beta \neq 0$. While (24) has traditionally been the alternative on which the literature has focused, in certain applications it may be overly restrictive. For example, as discussed in the introduction, it is not clear that one would want to model (near) unit root components in stock or exchange rate returns on theoretical grounds, and empirically they show little serial correlation. Thus, it also seems reasonable to consider test performance under alternatives that allow y_t to remain stationary even when x_t is highly persistent. A simple alternative of this type is given by a regression of y_t on quasi-differenced x_t as in

$$y_t = d_t + \gamma \left(1 - \left(1 + \frac{c}{n}\right) L\right) x_{t-1} + u_{1t}, \tag{25}$$

where x_t is given by the AR(1) specification in (18). While both sides of (25) are stationary, the data generating process permits y_t and x_t to have different orders of integration. This may be rewritten as

$$y_t = d_t + \gamma u_{2,t-1} + u_{1t}, \tag{26}$$

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

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

$$\begin{aligned} u &= \begin{pmatrix} u_{1t} & u_{2t} \end{pmatrix}' = \Sigma^{1/2}\varepsilon_t, \quad \varepsilon_t \sim \text{i.i.d. } N(0, I_2) \\ \Sigma &= \Sigma^{1/2}(\Sigma^{1/2})' = \begin{pmatrix} 1 & \sigma_{12} \\ \sigma_{12} & 1 \end{pmatrix}. \end{aligned} \tag{27}$$

Our primary interest lies in the performance of the covariance-based statistic t_λ given in (17), which was estimated as follows. In the trend model (23), we first demeaned Δx_t (thereby removing the trend in x_t) and detrended y_t prior to estimation. In the intercept model (22) only y_t was demeaned. Using this detrended (or demeaned) data we then estimated the limiting covariance $\lambda_{y,\Delta x}$ defined in (6) using the standard kernel covariance estimator $\widehat{\lambda}_{y,\Delta x}$ given in (11). Likewise, we estimated its asymptotic variance V (see Lemma 1) using the kernel estimator $\tilde{V}$ following (16).

Both kernel estimation procedures require the choice of kernel and bandwidth. The theoretical results allow considerable flexibility in the choice of the kernel $k(x)$

in the estimation of $\lambda_{y,\Delta x}$. However, to ensure conservative inference for stationary x_t, Lemma 7 mandates use of the Bartlett kernel for $\tilde{k}(x)$ in the estimation of V. We therefore used the Bartlett kernel for both estimators.[10] The bandwidth parameter m in the estimation of $\lambda_{y,\Delta x}$ is selected using the optimal bandwidth formula given in (12). Implementation of this formula in practice requires the use of a first-stage parametric approximation model. As in Andrews (1991), this is assumed only to provide a parsimonious approximation, not a correct specification. Although separate univariate AR(1) models are typically employed, the optimal bandwidth in this case depends on the behavior of the cross auto-correlations and necessitates a joint model. Including a moving average component also seems desirable given possible over-differencing in Δx_t. Therefore we estimated the first-stage VARMA(1,1) model $z_t = \widehat{A} z_{t-1} + \widehat{\varepsilon}_t + \widehat{B}\widehat{\varepsilon}_{t-1}$ for $z = (y_t, \Delta x_t)'$ employing the three-stage linear regression method of Dufour and Pelletier (2002) and then used the estimated parameters to calculate m^* in (12).[11] The choice of the second bandwidth parameter $\tilde{m}$ used in the estimation of V is constrained by Lemma 7, which requires $\tilde{m} = o(m)$. While clearly arbitrary, our choice of $\tilde{m} = m^{0.9}$ appeared sufficient to insure conservative inference in the stationary case, with minimal cost in overall performance.

As a basis of comparison we also provide some simulation results for both the standard regression t-test and the size-adjusted regression-based approach, using the two stage Bonferroni-bounds test of Cavanagh et al. (1995) (hereafter CES).[12] These tests are developed under the assumption that y_t is a martingale difference sequence (MDS) under the null, which is commonly imposed in the finance literature when one period returns are used. Our test was developed in a more general setting, allowing for serial correlation in y_t, and thus our use of V as defined in Lemma 1b and estimated by $\tilde{V}$ in (16) is analogous to the use of robust standard errors in regression. This generality can be quite useful, particularly when employing long-horizon returns for which the MDS assumption is necessarily violated and when analyzing test power, since the MDS assumption cannot be imposed under the alternative hypothesis. On the other hand, when the MDS assumption is appropriate it may come at some cost in terms of power. In order to allow for a clear comparison to predictive regression methods which do not employ HAC standard error estimates, we simplify V by imposing the MDS assumption on y_t unless stated otherwise. Specifically, $\Gamma_{yy}(h)$ and $\widehat{\Gamma}_{yy}(h)$ with $|h| > 0$ in V and $\tilde{V}$ are replaced by 0, giving $V_{MDS} \equiv 2\pi^2 \Gamma_{yy}(0) f_{\Delta x \Delta x}(0) \int_0^\infty k^2(x)\, dx$ and

$$\tilde{V}_{MDS} = \frac{1}{m} \sum_{h'=1}^{n-1} \sum_{h=1}^{n-1} k\left(\frac{h'-1}{m}\right) k\left(\frac{h-1}{m}\right) \tilde{k}\left(\frac{h'-h}{\tilde{m}}\right) \widehat{\Gamma}_{\Delta x \Delta x}(h'-h)\, \widehat{\Gamma}_{yy}(0).$$

[10] Alternate choices for $k(x)$ are explored in Section 5.3.

[11] In this model, $f(0) = \frac{1}{2\pi}(\Delta + \Delta' - \Gamma(0))$ where $\Delta = \Gamma(0)[I_2 - A']^{-1}$ and $vec(\Gamma(0)) = [I_2 - (A \otimes A)]^{-1} vec(A\Sigma B' + \Sigma + B\Sigma A' + B\Sigma B')$, and the term $\sum_{h=1}^{\infty} \Gamma_{\Delta x y}(h) h^q$ is given by the (2,1)th element of $(2\pi)^{-1}\Gamma(1)(I_2 - A')^{-2}$, where $\Gamma(1) = \Gamma(0)A' + \Sigma B'$. We impose some constraints on the ARMA parameters to insure stationarity and invertibility and also impose $n^{0.9}$ as an upper bound on m.

[12] We thank James Stock for use of his Gauss code. Following CES, we impose a finite sample size-adjustment, without which the test would be quite conservative. The reader is referred to CES for the details.

In order to preserve space we report only the results with $\tilde{V}_{MDS}$. The full set of results using the more general $\tilde{V}$ given by (16) are available upon request. The size of the tests are quite similar in both cases. As expected, although the overall pattern is similar, the power is generally better when the MDS assumption is imposed. All results below are based on 2000 replications, with results reported for sample sizes of $n = 100$ and 200.

5.1 Size

We first simulate under the null hypothesis with y_t given by (21) and x_t given by the AR(1) process (18) with ρ_1 modelled local-to-unity as in (20). Results are provided for various values of both c (and therefore ρ_1) and σ_{12}. In order to set a basis of comparison, Table 1 shows empirical rejection rates for the standard two-sided regression t-test (y_t regressed on x_{t-1}) with a nominal level of 5 percent.[13] The rejection rates are reasonable for small values of ρ_1 and/or σ_{12} but grow highly unreliable as ρ_1 approaches one and the residual correlation increases. The size problem is particularly severe in the model with trend, for which rejection rates can exceed 50 percent.

By contrast, the covariance-based test t_λ, whose rejection rates are shown in Table 2, is quite reliable in sample sizes as small as one hundred across the whole range of parameter values in the both the intercept and trend models. The test can become slightly conservative, particularly for large (negative) values of c. This is consistent with Lemma 8. However, the empirical rejection rates remain within a few percentage points of the nominal value. This good performance corroborates the theoretical results. Of course, good performance may also be obtained by properly size adjusting the regression-based tests, as in the bounds tests of CES (see their Table 4).

Following the empirical literature, we also consider longer-horizon returns. We may no longer impose an MDS assumption when estimating V, but otherwise require no explicit corrections to handle the moving average components induced by the overlapping returns; for fixed horizon (k) we simply have to define

$$y_{t,k} = (y_t + y_{t-1} + \ldots + y_{t-k+1}) = \phi(L)y_t \quad \text{and} \quad \Delta x_{t,k} = \Delta x_{t-k+1} = L^{(k-1)}\Delta x_t \tag{28}$$

and apply the theoretical results to $(y_{t,k}, \Delta x_{t,k}) = \text{diag}(\phi(L), L^{k-1})A(L)\varepsilon_t = A^*(L)\varepsilon_t$.[14] In finite sample, the accuracy of long-horizon tests typically depends on the ratio k/n. In Table 3, we match this ratio to the sample size ($n = 924$) and longest horizon (2 years, $k = 24$) used in the empirical application, yielding $k = 3$ for $n = 100$. This

[13] Results (available upon request) using a 10% nominal level also confirm the the good properties of the covariance-based test discussed below. Note that Hodrick or HAC standard errors are unnecessary here since u_t is taken to be i.i.d.

[14] In estimation of V, the MDS assumption in no longer tenable since $y_{t,k}$ follows an $MA(k)$ if y_t is an MDS. Thus in (16) we impose only that $\text{cov}(y_{t,k}, y_{t-h,k}) = 0$ for $h > k-1$. To choose m^* we estimate the same VARMA(1,1) described above using $z = (y_t, \Delta x_t)'$ and adjust the formula in (12) to account for the long-horizon returns. Specifically, we replace $f_{yy}(0)$ by $k^2 f_{yy}(0)$ and $(2\pi)^{-1}\Gamma(1)(I_2 - A')^{-2}$ by $(2\pi)^{-1}\Gamma(1)(I - (A')^k)(I - A')^{-3}$ in footnote 11.

again yields reasonable size performance. Holding $n = 100$ fixed, but increasing the horizon to $k = 5$ leads to only a slight deterioration.

The model above is the baseline model most often used to evaluate size distortions in this context. However, our test is designed to work in a more general setting and it is also of interest to investigate finite sample performance under higher order autoregressive specifications for x_t, such as the AR(2) model (19), with roots on or close to the unit circle. Rudebusch (1992, Table 2) finds that an AR(2) with $\rho_1 + \rho_2$ slightly below unity (with $\rho_1 > 1$ and $\rho_2 < 0$) provides a good fit for a number of macroeconomic and financial time series. In order to roughly match these estimates we set

$$\rho_1 = 1.5 \quad \text{and} \quad \rho_2 = -0.5 + c/n \tag{29}$$

for the same values of c considered above. Thus, as in the AR(1) model, x_t is unit-root nonstationary for $c = 0$ ($\rho_2 = -0.5$) and near-nonstationary for $c < 0$ ($\rho_2 < -0.5$). The rejection rates for the covariance-based tests are shown in Table 4. The results remain fairly accurate even for $n = 100$, particularly in the demeaned case. By contrast, finite sample rejection rates for least squares (available upon request) reach to above 50% and do not improve with sample size for fixed c.

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

5.2 Power

We first consider the power of the covariance-based test t_λ against the standard regression alternative given in (24) with $\beta \neq 0$ and local to unity x_t given by (18) and (20).[15] For $c = 0$ this alternative constitutes a cointegrating relation, while for $c << 0$ the alternative is a stationary regression. The results are shown in Table 5. To save space we show only the demeaned case. As expected, the power of the test is reasonable, increasing in both sample size and distance from the null.

One of the goals of the covariance-based test was to simultaneously maintain power against alternatives which allow y_t (e.g. returns) to be stationary, despite near or even exact unit root behavior in x_t. This avoids, for example, the implication that stock prices or exchange rates contain a (near) $I(2)$ component under the alternative hypothesis when predictor variables are persistent. The data generating process in (25), together with (18), therefore provides a natural alternative in which to consider finite sample power in that it holds y_t stationary (but not over-differenced) regardless of the persistence in x_t. In doing so, it incorporates both (3) and (9) as special cases for $\rho_1 = 0$ and $\rho_1 = 1$ respectively.

Finite sample power results for the covariance-based tests under the alternative in (25) with x_t given by (18) and (20) are shown in Table 6. The test is calculated in the same way as before, again using y_t and x_{t-1} as inputs. These rejection rates appear quite reasonable, again increasing in both sample size and distance from the null hypothesis.

[15] We do not report size adjusted power, since both tests considered in this section have good size.

Many existing tests are based on a size adjusted regression of the type shown in (3). These procedures may be expected to have good power against regression alternatives when x_t is stationary (e.g. $\rho_1 << 1$ in (18) and $\beta \neq 0$ in (24)) and against cointegration or near-cointegration alternatives when x_t is near $I(1)$ (e.g. $\rho_1 \approx 1$ in (18) and $\beta \neq 0$ in (24)). This is confirmed in Table 7, which shows finite sample power for the CES Bonferroni test procedure against $\beta \neq 0$ in (24) with local to unity x_t given by (18) and (20). As expected, the test exhibits very good power against this alternative and is in this case more powerful than t_λ.

On the other hand, it is not clear that regression tests based on (3) should have much power against alternatives in which y_t and x_{t-1} exhibit different orders of integration. Table 8 provides rejection rates for the CES Bonferroni test against the same DGP used to assess the power of t_λ in Table 6, i.e. equations (18), (20), and (25). Confirming the reasoning above, the regression-based test does quite well for the larger values of c when x_t and y_t behave in a stationary manner, but performance deteriorates rapidly as x_t approaches nonstationarity (small c) and the alternative becomes unbalanced. Moreover, for small c the power does not seem to improve as we move further into the alternative. Nor, for fixed values of c, do rejection rates increase much as the sample size increases. For example, in the worst case for $c = 0$ and $\sigma_{12} = 0.95$ (panel 2, row 2, last column), the power remains under 5 percent even for a population R^2 of 0.20 and a sample size of two-hundred. In fact, in results not shown, it remains under 10 percent for an R^2 of 0.50.

These simulations suggest that the covariance-based orthogonality test may provide power against a wider range of alternatives than do existing size-adjusted regression-based tests. In particular, they appear to provide reasonable power against both standard regression and unbalanced (i.e. $y_t \sim I(0)$, $x_{t-1} \sim I(1)$) alternatives, whereas regression-based tests do particularly well against the alternatives for which they were designed, but provide little reliable power against unbalanced alternatives. This added generality does of course come at some cost in terms of power against certain specific alternatives and, in this sense, the two testing approaches (regression and covariance-based) are properly seen as compliments rather than substitutes.

5.3 Kernel Comparison

In the preceding simulations we employed the Bartlett (Newey-West) kernel for both $k(x)$ and $\widetilde{k}(x)$. In the case of $\widetilde{k}$ this is needed to satisfy the conditions of Lemma 7. However, other kernel choices may be considered for k. We therefore replicated Tables 2, 5, and 6 using three alternative kernels, the Parzen, Quadratic Spectral, and Tukey-Hanning kernels, defined respectively by $k(x) = \left(1 - 6x^2 + 6|x|^3\right) I\left(|x| < 1/2\right) + 2\left(1 - |x|\right)^3 I\left(1/2 \leq |x| \leq 1\right)$, $k(x) = 25/\left(12\pi^2 x^2\right)\left(\sin\left(6\pi x/5\right) - \cos\left(6\pi x/5\right)\right)$, and $k(x) = 1/2\left(1 + cos\pi x\right) I(|x| \leq 1)$. Although the performance of the test varied slightly according to the choice of kernel, the differences did not seem large enough to be important in practice. In addition, the relative ordering of the test across kernel choices varied with the choice of the specification of the data generating process.

Table 9 provides a representative comparison across three of these four kernels, the

Bartlett (NW), Parzen (PZ), and Quadratic Spectral (QS).[16] The first two columns provide the values of c and σ_{12} respectively, the third column denotes the choice of kernel, the fourth column provides rejection rates under the null hypothesis, and the remaining columns provide rejection rates under the alternative. In the top panel the alternative is specified in levels as in Table 5, whereas in the bottom panel the alternative is rebalanced as in Table 6. As is apparent from the table, use of the Parzen kernel delivers modest power improvements against the levels alternative, whereas, depending on the parameters of the DGP, either the Bartlett (NW) or the Quadratic Spectral (QS) kernel tends to perform best against the rebalanced alternative.

6 Application to tests of stock return predictability

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

Although theoretical considerations may rule out exact unit root behavior in dividend yields[17] and interest rates, near unit roots in the local to unity sense cannot be ruled out a priori. Empirically, both series are highly persistent, with confidence intervals on the largest root often containing one (Torous et al. (2005)). Moreover, although pre-determined, there is no reason to believe that these regressors are fully exogenous. For example, the stock price enters both the return and dividend yield. The combination of near unit root behavior and a failure of strict exogeneity is a recipe for size problems (Cavanagh et al. (1995)). Consequently, subsequent doubts have been raised regarding the evidence for predictability on account of the strong persistence in the regressors (Stambaugh (1986 & 1999) and Mankiw and Shapiro (1986)).[18] This has spurred a large literature in an attempt to address this issue, and the degree to which evidence of predictability has been overstated remains a subject of ongoing debate.[19]

Thus the literature to date has focused primarily on the issue of size distortion. However, near unit root regressors may also raise specification issues under the alternative, in the sense that a stationary variable, such as a stock return should not be linearly predictable by a unit (or near unit) root regressor (Lanne (2002)). Even if y_t

[16] The full results are available upon request.

[17] Campbell and Shiller (1988a,b), but see Tuypens (2002) for an alternative viewpoint.

[18] Also of concern have been the accuracy of the standard errors in long-horizon regressions (Richardson and Stock (1989), Valkanov (2003)).

[19] This literature includes resampling and simulation methods (Hodrick (1992), Nelson and Kim (1993), Goetzmann and Jorion (1993), Wolf (2000), and Ang and Bekaert (2005)), local to unity corrections along the lines of Cavanagh et al. (1995) (Viceira (1997), Valkanov (2003), Torous et al. (2005), and Campbell and Yogo (2006)), and finite sample or Bayesian approaches (Stambaugh (1999) and Lewellen (2004)).

is $I(0)$ and x_t is $I(1)$, y_t may still be predictable based on the past history of x_t, as exemplified by (9) with $\gamma_1 \neq 0$. Yet, as the simulations underscored, regression tests based on (3) have unreliable power against such alternatives, even if size-adjusted. Therefore, while evidence of predictability may be overstated due to size distortion, it is also possible that it has been understated due to near unit root specification issues. Since the covariance-based tests address both issues simultaneously they may be useful in untangling these two effects.

We use monthly returns from 1927 to 2003. Following Campbell et al. (1997, chapter 7), we also consider separately the 1927-1951 and 1952-1994 subsamples, along with the more recent 1952-2003 subperiod.[20] Monthly log returns are calculated as $r_{t+1} = ln\left((P_{t+1} + D_{t+1})/P_t\right)$, where P_t and D_t are the stock price and dividend from the CRSP value-weighed index of NYSE, AMEX, and NASDAQ stocks. Real returns are formed by deflating nominal returns by the CPI.[21] The dividend-price ratio is calculated in the standard way as the sum of dividends paid over the past twelve months, divided by the current level of the index: $d_t - p_t = ln\left((D_t + \ldots + D_{t-11})/P_t\right)$. We denote the one-month treasury bill rate by i_t. Following the literature, we also consider longer-horizon returns of the form $r_{t+1} + \ldots + r_{t+k}$ for $k = 1, 3, 12$, and 24. HAC standard errors are employed for $k > 1$ in the regression analysis. Likewise, for the covariance-based test, the MDS assumption on y_t is relaxed in the estimation of V, as detailed in footnote 14. For fixed k our test requires no further correction or adjustment, as discussed in the simulation section above.

The first three rows in each panel of Table 10 show the standard regression coefficients, R^2s, and t-statistics. We defer discussion of the fourth row until later. The interest rate regressions show modest evidence of predictability whereas evidence using dividend yields is generally quite strong. Intuition for the potential bias and size distortion in these regressions is provided by Lewellen (2004) who expresses the bias in $\widehat{\beta}$ in (3) in terms of the bias in $\widehat{\rho}$ in (18) and the residual covariance σ_{12} in (27):

$$E\left[\hat{\beta} - \beta\right] = \frac{\sigma_{12}}{\sigma_{22}} E\left[\,\hat{\rho} - \rho\right]. \tag{30}$$

The two ingredients needed to produce bias are thus persistent regressors and residual serial correlation. Table 11 shows the Stock (1991) confidence interval on the largest root in x_t together with the estimated residual correlation $\delta = corr(\varepsilon_{1t}, \varepsilon_{2t})$. The two series both show large roots, with confidence intervals on the largest root containing one, but display quite different residual correlation properties. Estimates of δ are small for the interest rate series, suggesting only modest size distortion, but are close to negative one for the dividend price ratio. Intuitively, an increase in the current stock price corresponds to a higher return but lower dividend yield. Since the AR(1) coefficient estimate $\hat{\rho_1}$ is downward biased (Hurwitz (1950)), negative residual correlation implies positive bias in $\widehat{\beta}$ (see (30)). In other words, the bias runs in the same direction as the observed alternative, leaving the results difficult to interpret.

[20] We thank John Campbell for kindly providing us with the original data for this project, which was later updated from CRSP.

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

This preliminary analysis suggests that size distortion may be a more serious concern in the case of the dividend price ratio than it is for the interest rate regressions. This conjecture is supported by the results in Table 12. The table shows the covariance-based test statistic t_λ, for which standard normal critical values apply, the two-sided p-value, and the optimal bandwidth m^*, calculated in the same way as in the simulations (see Section 5 for details).[22]

For the interest rate, the covariance-based tests in Table 12 provide overall evidence of predictability similar to that found from the standard regression tests of Table 10. In both cases, evidence of predictability is confined to the later sample periods (1952 -1994 & 1952-2003). However, the strongest evidence from the covariance test come at medium horizons (3-12), whereas regression t-tests are largest at short horizons (1-3), showing little evidence of predictability at the 12 month horizon. On the other hand, in the case of the dividend price ratio, the covariance-based tests show far weaker evidence of predictability than do the standard regression-based tests. This agrees with the conclusion in several (but not all) previous studies that address the issue of size distortion in (3).[23]

It may also be instructive to compare the empirical results from the standard regression and covariance based tests to those from a properly size adjusted regression test, such as the CES tests discussed in the section above. Since the CES tests apply only at short-horizons (k=1), we employ instead a similar bounds test proposed by Valkanov (2003) for use in the long horizon context. The last row of each panel of Table 10 shows the p-value from the Valkanov (2003) sup-bound tests. These p-values confirm the modest evidence of predictability found from the interest rate regressions. This is expected in light of the small residual cross correlations in Table 11, which suggest only modest distortion in standard tests. On the other hand, evidence of predictability in the dividend yield regression, for which residual cross-correlations are large, disappears in all but one of four samples, the 1952-1994 sub-period.

Thus with a few exceptions, the covariance and size-adjusted regression tests tell similar stories regarding predictability. Both tests confirm the modest evidence of interest rate predictability found from standard regressions, with the covariance test finding evidence at medium to longer horizons. Similarly both tests suggest show substantially less evidence of predictability than implied by standard, but possibly size-distorted regression tests.

7 Conclusion

In regression-based orthogonality tests it is often the case that the regressor is highly serially correlated, with an autoregressive root close or possibly equal to unity. This

[22]We also employed fixed bandwidths of $m = 1$, 2, 5, and 10 (results available upon request). This produced similar qualitative results, but with larger values of m further increasing evidence of predictability using i_t and changing the sign of some insignificant statistics using $d_t - p_t$. The demeaned version of the estimator was employed. Similar results (available upon request) were obtained using the detrended version.

[23]Viceira (1997), Wolf (2000), Torous et al. (2005), Valkanov (2003) , but see also Lewellen (2004) and Campbell and Yogo (2006) who conclude more strongly in favor of predictability.

is well known to cause size problems in standard tests. Simple two-stage procedures employing unit root tests together with size correction can generally correct this problem in the $I(1)$ case, but still produce size distortions under local-to-unity assumptions.

Roots near unity may also artificially restrict the allowable alternative hypothesis, leading to poor size-adjusted power under certain 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 (i.e. $y_t \sim I(0)$, $x_t \sim I(1)$) would not rule out a violation of orthogonality due to a linear relationship between the dependent variable and stationary transformations of the regressor.

The covariance-based test proposed here produces good size and power against alternatives in which the dependent variable (e.g. stock returns) is stationary, regardless of whether the predictor 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.

Finally, we note that the formulation of the limiting covariance defined in (6) could be usefully modified in various ways in order to provide alternate orthogonality tests that also remain appropriate when x_t and y_t potentially exhibit different orders of integration. For example, one might consider a test based on the restriction that $\sum_{h=1}^{\infty} |\text{cov}\,(y_t, \Delta x_{t-h})| = 0$.[24] This departs from the standard covariance restriction in (2) that we generalize here, yet provides a stronger implication of (1) and thus could potentially be a promising avenue for future research. From a theoretical perspective, such an extension would be interesting and non-trivial, since deriving the joint properties of the absolute values of the sample covariances $\left|n^{-1} \sum_{t=h+1}^{n} y_t \Delta x_{t-h}\right|$ would be complicated considerably by the presence of the absolute value sign.

8 Appendix A: Proofs

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

8.1 Proof of (8)

From the definition of x_t, we have

$$\Delta x_t = u_t + \frac{c}{n} x_{t-1} = \begin{cases} u_t + \frac{c}{n} \sum_{k=0}^{t-2} (1 + c/n)^k \, u_{t-1-k}, & t \geq 1, \\ 0, & t \leq 0, \end{cases} \tag{31}$$

[24] We thank an anonymous referee for this valuable observation.

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

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

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

8.2 Proof of Lemma 1

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

$$\frac{n}{m}\text{var}\left(\widehat{\lambda}_{y,\Delta x}\right) = \frac{n}{m}\sum_{h'=1}^{n-1}\sum_{h=1}^{n-1} k\left(\frac{h'-1}{m}\right)k\left(\frac{h-1}{m}\right)\text{cov}\left(\widehat{\Gamma}_{\Delta xy}\left(h'\right), \widehat{\Gamma}_{\Delta xy}\left(h\right)\right). \quad (32)$$

Hannan (1970) p. 313 gives

$$\begin{aligned}
& \text{cov}\left(\widehat{\Gamma}_{\Delta xy}\left(h'\right), \widehat{\Gamma}_{\Delta xy}\left(h\right)\right) \qquad (33) \\
= \;& n^{-1}\sum_{u=-\infty}^{\infty}\{\Gamma_{\Delta x\Delta x}\left(u\right)\Gamma_{yy}(u+h-h') + \Gamma_{\Delta xy}(u+h)\Gamma_{y\Delta x}(u-h') \\
& + k_{\Delta xy\Delta xy}(0, h', u, u+h)\}\phi_n(u, h', h),
\end{aligned}$$

where $k_{\Delta xy\Delta xy}(0, h', u, u+h)$ is the fourth cumulant of z_t (see Hannan, 1970, p.23 for the definition) and $\phi_n(u, h', h)$ is given by (the formula of $\phi_n(u, h', h)$ for $-n+h' \leq u \leq 0$ in Hannan has a typo)

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

It follows that (32) is comprised of

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

$$+\frac{1}{m}\sum_{h'=1}^{n-1}\sum_{h=1}^{n-1} k\left(\frac{h'-1}{m}\right)k\left(\frac{h-1}{m}\right)\sum_{u=-\infty}^{\infty}\Gamma_{\Delta xy}(u+h)\Gamma_{y\Delta x}(u-h')\phi_n(u, h', h) \quad (35)$$

$$+\frac{1}{m}\sum_{h'=1}^{n-1}\sum_{h=1}^{n-1} k\left(\frac{h'-1}{m}\right)k\left(\frac{h-1}{m}\right)\sum_{u=-\infty}^{\infty}k_{\Delta xy\Delta xy}(0, h', u, u+h)\phi_n(u, h', h). \quad (36)$$

Let $v = h' - h$, and we can rewrite (34) as

$$\sum_{v=-n+2}^{n-2} \sum_{u=-\infty}^{\infty} \Gamma_{\Delta x \Delta x}(u) \Gamma_{yy}(u-v) \left\{ \frac{1}{m} \sum_h{}' \phi_n(u, h+v, h) k\left(\frac{h+v-1}{m}\right) k\left(\frac{h-1}{m}\right) \right\}, \tag{37}$$

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 $\int_0^\infty k^2(x)\, dx$ by the argument in Hannan (1970) pp. 314-15. Furthermore, $\sum_{v=-n+2}^{n-2} \sum_{u=-\infty}^{\infty} \Gamma_{\Delta x \Delta x}(u) \Gamma_{yy}(u-v) \to 4\pi^2 f_{\Delta x \Delta x}(0) f_{yy}(0)$ as $n \to \infty$, and hence (34) converges to V.

For (35), define $u' = u + h$ and rewrite (35) as

$$\sum_{u'=-\infty}^{\infty} \Gamma_{\Delta xy}(u') \frac{1}{m} \sum_{h'=1}^{n-1} \sum_{h=1}^{n-1} k\left(\frac{h'-1}{m}\right) k\left(\frac{h-1}{m}\right) \Gamma_{y\Delta x}(u' - h - h') \phi_n(u' - h, h', h). \tag{38}$$

Let $\beta \in (0, 1/2)$ and split the sum into three: (i) $|u'| \ge \log m$, (ii) $|u'| < \log m$ and $h + h' \le m^\beta$, and (iii) $|u'| < \log m$ and $h + h' > m^\beta$. For (i), the sum is bounded by

$$\sum_{|u'| \ge \log m} \left|\Gamma_{\Delta xy}(u')\right| \frac{1}{m} \sum_{h=1}^{n-1} \left| k\left(\frac{h-1}{m}\right) \right| \sum_{h'=1}^{n-1} \left|\Gamma_{y\Delta x}(u' - h - h')\right| \sup_x |k(x)| \to 0,$$

since $\sum_{u=-\infty}^{\infty} |\Gamma_{\Delta xy}(u)| < \infty$ and thus the tail sum is $o(1)$. For (ii), the sum is bounded by

$$\sum_{|u'| < \log m} \left|\Gamma_{\Delta xy}(u')\right| \frac{1}{m} \left[\sum_{h=1}^{m^\beta} \left| k\left(\frac{h-1}{m}\right) \right| \right]^2 \sup_u |\Gamma_{y\Delta x}(u)| = O\left(m^{2\beta - 1}\right) \to 0.$$

For (iii), define $v = u' - h - h'$ and note that $|v| \ge m^\beta/2$ for sufficiently large m. Thus, the sum is bounded by

$$\sum_{|u'| < \log m} \left|\Gamma_{\Delta xy}(u')\right| \frac{1}{m} \sum_{h=1}^{n-1} \left| k\left(\frac{h-1}{m}\right) \right| \sum_{v \ge m^\beta/2} |\Gamma_{y\Delta x}(v)|\, |v| (m^\beta/2)^{-1} \sup_x |k(x)| = O(m^{-\beta}),$$

since $\sum_{v=-\infty}^{\infty} |\Gamma_{y\Delta x}(v)|\, |v| < \infty$ by (10). Hence (35) converges to 0. (36) is $O(m^{-1})$ because the fourth cumulant of z_t satisfies (Hannan, 1970, p. 211)

$$\sum_{q=-\infty}^{\infty} \sum_{r=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} |k_{ijkl}(0, q, r, s)| < \infty, \quad i, j, k, l = \{y, \Delta x\}, \tag{39}$$

and the stated result follows. ■

8.3 Proof of Theorem 2

In view of Lemma 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

$$\begin{aligned} & \sqrt{\frac{n}{m}}\left(\widehat{\lambda}_{y,\Delta x} - E\widehat{\lambda}_{y,\Delta x}\right) \\ = & \frac{1}{\sqrt{m}}\sum_{h=1}^{n-1} k\left(\frac{h-1}{m}\right)\frac{1}{\sqrt{n}}\sum_{t=h+1}^{n}\left(y_t\Delta x_{t-h} - Ey_t\Delta x_{t-h}\right) = I + II, \end{aligned} \tag{40}$$

where

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

From Lemma 11, Minkowski's inequality, $\int_0^\infty |\bar{k}(x)|x^{1/2}dx < \infty$, and an argument similar to Lemma 1 of Jansson (2002), we have

$$E(II)^2 = O\left(\frac{1}{mn}\left(\sum_{h=1}^{n-1}\left|k\left(\frac{h-1}{m}\right)\right|h^{1/2}\right)^2\right) = O\left(\frac{m^2}{n}\right), \tag{41}$$

Lemma 13 gives

$$I = \sum_{t=1}^{n} Z_t + R_n; \quad Z_t = n^{-1/2}m^{-1/2}\sum_{h=1}^{n-1} k\left(\frac{h-1}{m}\right)\sum_{r=1}^{\infty}\varepsilon_{t-r}' f^{hr}(1)\varepsilon_t, \tag{42}$$

where $ER_n^2 = o(1)$ and $f^{hr}(1)$ is defined in the statement of Lemma 13. Therefore, $\sqrt{n/m}(\widehat{\lambda}_{y,\Delta x} - E\widehat{\lambda}_{y,\Delta x}) \to_d N(0,V)$ follows if we show that

$$\sum_{t=1}^{n} Z_t \to_d N(0,V), \quad \text{as } n \to \infty. \tag{43}$$

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 (43) follows from the martingale CLT of 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

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

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

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

Therefore, (i) follows from the ergodic theorem if

$$nEZ_t^2 \to V. \tag{44}$$

From (40)-(42), 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

$$E\left(\sum_{t=1}^{n} Z_t\right)^2 = E\left(\sqrt{\frac{n}{m}}\left(\widehat{\lambda}_{y,\Delta x} - E\widehat{\lambda}_{y,\Delta x}\right) - (II+R_n)\right)^2. \tag{45}$$

The left hand side of (45) is $\sum_{t=1}^{n} EZ_t^2 = nEZ_t^2$, since Z_t is a stationary martingale difference sequence. The right hand side of (45) is

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

because

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

and $E(II+R_n)^2 = o(1)$.

Therefore, we establish (44) 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, giving (43) and the stated result follows. ■

8.4 Proof of Lemma 3

From (31), we have

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

The required result follows because $E|\sum_{h=1}^{n-1} k((h-1)/m) T_{nh}|$ is bounded by $n^{-2}\sum_{h=1}^{n-1}|k((h-1)/m)|\sum_{t=h+1}^{n}\sum_{k=-\infty}^{\infty}|\Gamma_{uy}(k)| = O(n^{-1}\sum_{h=1}^{n-1}|k((h-1)/m)|) = O(mn^{-1})$. ■

8.5 Proof of Lemma 4

Some simple algebra gives

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

For T_{1n}, we have from Theorem 14 of Hannan (1970, page 228) (note that $\lambda_{y,\Delta x} = Ey_t x_{t-1} = \gamma_{xy}(1)$)

$$\sqrt{n}(T_{1n} - \lambda_{y,\Delta x}) = \sqrt{n}(\widehat{\gamma}_{xy}(1) - \gamma_{xy}(1)) \to_d N(0, \Xi),$$

as $n \to \infty$, where Ξ is given by Hannan (1970) in equation (3.3) on page 209 and line 5 on page 211. For T_{2n}, first observe that

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

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

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

Since $k(x) - 1 = O(x^q)$ as $x \to 0$, the first term on the right is, for ε sufficiently small, $O(\sum_{h=2}^{\varepsilon m}(h/m)^q|\gamma_{xy}(h)|) = O(m^{-q})$. The second term on the right is bounded by $\sum_{h=\varepsilon m}^{n-1}|\gamma_{xy}(h)| \leq (\varepsilon m)^{-q}\sum_{h=\varepsilon m}^{n-1} h^q|\gamma_{xy}(h)| = O(m^{-q})$. Therefore, defining $B_n = \sqrt{n}ET_{2n}$ gives the bias term B_n in (15).

It remains to show that $\text{var}(\sqrt{n}T_{2n}) = o(1)$ and $\sqrt{n}(T_{3n} + T_{4n}) = o_p(1)$. From

Hannan (1970) (equation (3.3) on page 209 and line 5 on page 211), we have

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

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

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

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

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

because $\sum_{q=-\infty}^{\infty}\sum_{r=-\infty}^{\infty}\sum_{s=-\infty}^{\infty}|k_{xyxy}(0,q,r,s)| < \infty$ from Hannan (1970, p. 211). Finally, $\sqrt{n}(T_{3n}+T_{4n}) = o_p(1)$ follows from

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

$x_0 = O_p(1)$, and

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

and the stated result follows. ■

8.6 Proof of Lemma 5

In view of equations (34) and (37) in the proof of Lemma 1, $\tilde{V}$ reduces to

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

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

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

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. The result for the local-to-unity case follows from the proof of Lemma 3. ■

8.7 Proof of Lemma 7

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

$$\Pr(\tilde{V} \geq \eta \tilde{m}^{-1}) \to 1, \quad \text{as } n \to \infty. \tag{46}$$

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

$$\sum_{u=-\tilde{m}}^{\tilde{m}} \tilde{k}\left(\frac{u}{\tilde{m}}\right)\widehat{\Gamma}_{\Delta x \Delta x}\,(u) \sum_{v=-\tilde{m}}^{\tilde{m}} \tilde{k}\left(\frac{v}{\tilde{m}}\right)\widehat{\Gamma}_{yy}(v)\left\{\int_0^\infty k^2\,(x)\,dx + o\,(1)\right\} \tag{47}$$

Because $\sum_{v=-\tilde{m}}^{\tilde{m}} \tilde{k}(v/\tilde{m})\widehat{\Gamma}_{yy}(v) \to_p 2\pi \tilde{f}_{yy}(0) > 0$ by the standard argument, (46) follows if there exists $\varepsilon > 0$ such that

$$\begin{aligned}
& \Pr\left(\sum_{v=-\tilde{m}}^{\tilde{m}} \tilde{k}\left(\frac{v}{\tilde{m}}\right)\widehat{\Gamma}_{\Delta x \Delta x}(v) \geq \varepsilon \tilde{m}^{-1}\right) \\
= \; & \Pr\left(2\pi \int_{-\pi}^{\pi} W_{\tilde{m}}\,(\lambda)\, I_{\Delta x}(\lambda) d\lambda \geq \varepsilon \tilde{m}^{-1}\right) \to 1, \quad \text{as } n \to \infty,
\end{aligned} \tag{48}$$

where (Priestley, 1981, p. 439)

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

is the Fejér 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

$$\begin{aligned}
& \int_{-\pi}^{\pi} W_{\tilde{m}}\,(\lambda)\, I_{\Delta x}(\lambda) d\lambda \\
= \; & \int_{-\pi}^{\pi} W_{\tilde{m}}\,(\lambda)\, |1 - e^{i\lambda}|^2 I_x(\lambda) d\lambda && (49) \\
& + (2\pi n)^{-1/2} X_n \int_{-\pi}^{\pi} W_{\tilde{m}}\,(\lambda)\, 2\mathrm{Re}\left[(1 - e^{i\lambda}) w_x(\lambda) e^{-i(n+1)\lambda}\right] d\lambda && (50) \\
& + \int_{-\pi}^{\pi} W_{\tilde{m}}\,(\lambda)\, d\lambda (2\pi n)^{-1} X_n^2. && (51)
\end{aligned}$$

We can ignore (51) because it is nonnegative. For (50), it follows from the Cauchy-Schwartz inequality and Lemma 14 (b) that

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

and (50)$= O_p(n^{-1/2}\tilde{m}^{-1/2}) = o_p(\tilde{m}^{-1})$ follows. Rewrite (49) as

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

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

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

Therefore, in conjunction with Lemma 14 (a), we obtain

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

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

$$\begin{cases} \sup_{\lambda,\lambda'\in[-\pi,\pi]} \left|\mathrm{cov}\left(I_x(\lambda), I_x(\lambda')\right)\right| = O(1), \\ \mathrm{cov}\left(I_x(\lambda), I_x(\lambda')\right) = o(1), \quad \lambda \neq \lambda'. \end{cases} \tag{52}$$

Therefore,

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

where the interchange of expectation and integration in the first line is valid by (52) and Fubini's Theorem, and the last line follows from Lemma 14 (b), (52), and the dominated convergence theorem. Therefore, there exists $\eta' > 0$ such that (49)+(50)+(51)$\geq \eta'\tilde{m}^{-1}$ with probability approaching one, and (48) and the stated result follow. ■

8.8 Proof of Lemma 9

Assume $x_0 = 0$ without the loss of generality. We show part (a) first. Substituting $y_t = \beta_n x_{t-1} + u_t$ into $\hat{\lambda}_{y,\Delta x}$, we have

$$\hat{\lambda}_{y,\Delta x} = \beta_n \sum_{h=1}^{n-1} k\left(\frac{h-1}{m}\right) \frac{1}{n} \sum_{t=h+1}^{n} x_{t-1}\Delta x_{t-h} + \hat{\lambda}_{u,\Delta x}, \tag{53}$$

where the second term is $O_p((m/n)^{1/2})$ from Lemma 1. If we show

$$\frac{1}{n}\sum_{t=h+1}^{n} x_{t-1}\Delta x_{t-h} = \frac{1}{n}\sum_{t=1}^{n} x_{t-1}\Delta x_t + \sum_{j=0}^{\infty} \Gamma_{\Delta x \Delta x}(j) + R_{nh}, \tag{54}$$

with $E|R_{nh}| = O(n^{-1}h^2 + n^{-1/2}h)$, then the stated result follows from $n^{-1}\sum_{t=1}^{n} x_{t-1}\Delta x_t \to_d (1/2)(B^2(1) - \Gamma_{\Delta x\Delta x}(0))$.

To show (54), first rewrite the left hand side of (54) as

$$\sum_{t=h+1}^{n} x_{t-1}\Delta x_{t-h} = \sum_{t=1}^{n-h} x_{t+h-1}\Delta x_t = \frac{1}{n}\sum_{t=1}^{n-h}\left(\Delta x_{t+h-1} + \cdots + \Delta x_t\right)\Delta x_t + \frac{1}{n}\sum_{t=1}^{n-h} x_{t-1}\Delta x_t. \tag{55}$$

The first term on the right of (55) equals $\sum_{j=0}^{h-1}(\hat{\Gamma}_{\Delta x\Delta x}(j) + r_{nhj})$ with $E|r_{nhj}| = O(n^{-1}(h-j))$. Then $E|\sum_{j=0}^{h-1} r_{nhj}| = O(n^{-1}h^2)$. Now $\sum_{j=0}^{h-1}\hat{\Gamma}_{\Delta x\Delta x}(j)$ is an estimate of $\sum_{j=0}^{\infty}\Gamma_{\Delta x\Delta x}(j)$ with a rectangular kernel (i.e., $q = \infty$) and hence $\sum_{j=0}^{\infty}\Gamma_{\Delta x\Delta x}(j) + r_{nh}^a$ where $E|r_{nh}^a| = O(n^{-1/2}h^{1/2})$ from Lemma 1. The second term on the right of (55) equals $n^{-1}\sum_{t=1}^{n} x_{t-1}\Delta x_t + r_{nh}$ with $E|r_{nh}| \le Cn^{-1/2}h$ from the Cauchy-Schwartz inequality. Therefore, we show (54) and complete the proof of part (a).

We proceed to prove part (b). From the arguments in the proof of Lemma 5, $\tilde{V}$ is equal to $\sum_{u=-\tilde{m}}^{\tilde{m}} \tilde{k}(u/\tilde{m})\hat{\Gamma}_{\Delta x\Delta x}(u) \sum_{v=-\tilde{m}}^{\tilde{m}} \tilde{k}(v/\tilde{m})\hat{\Gamma}_{yy}(v)\left\{\int_0^\infty k^2(x)\,dx + o(1)\right\}$. A standard argument gives $\sum_{u=-\tilde{m}}^{\tilde{m}} \tilde{k}(u/\tilde{m})\hat{\Gamma}_{\Delta x\Delta x}(u) \to_p 2\pi f_{\Delta x\Delta x}(0)$. Rewrite $\hat{\Gamma}_{yy}(v)$ as $\hat{\Gamma}_{yy}(v) = n^{-1}\sum_{t=1}^{n-v} y_t y_{t+v} = G_{1v} + G_{2v} + G_{3v} + G_{4v}$, where

$$\begin{array}{ll} G_{1v} = \beta_n^2 n^{-1}\sum_{t=1}^{n-v} x_{t-1}x_{t-1+v}, & G_{2v} = \beta_n n^{-1}\sum_{t=1}^{n-v} x_{t-1}u_{t+v}, \\ G_{3v} = \beta_n n^{-1}\sum_{t=1}^{n-v} u_t x_{t-1+v}, & G_{4v} = n^{-1}\sum_{t=1}^{n-v} u_t u_{t+v}. \end{array}$$

We consider G_{2v} first. Note that, in view of $x_0 = 0$, with $h = t - s$,

$$\frac{G_{2v}}{\beta_n} = \frac{1}{n}\sum_{t=2}^{n-v}\sum_{s=1}^{t-1}\Delta x_s u_{t+v} = \frac{1}{n}\sum_{h=1}^{n-v-1}\sum_{s=1}^{n-v-h}\Delta x_s u_{s+h+v} = \sum_{h=1}^{n-v-1}\hat{\Gamma}_{\Delta xu}(h+v).$$

$E[\sum_{h=0}^{n-v-1}\hat{\Gamma}_{\Delta xu}(h+v)]^2 = O(1)$ easily follows from the covariance between $\hat{\Gamma}_{\Delta xu}(h+v)$ and $\hat{\Gamma}_{\Delta xu}(h'+v)$ given by (33), and hence we have $\sum_{v=-\tilde{m}}^{\tilde{m}}\tilde{k}(v/\tilde{m})G_{2v} = O_p(\beta_n\tilde{m})$. A similar argument gives $\sum_{v=-\tilde{m}}^{\tilde{m}}\tilde{k}(v/\tilde{m})G_{3v} = O_p(\beta_n\tilde{m})$. For G_{1v}, it follows from the Cauchy-Schwartz inequality that

$$\frac{1}{n}\sum_{t=1}^{n-v} x_{t-1}x_{t-1+v} - \frac{1}{n}\sum_{t=1}^{n} x_{t-1}^2 = r_{nv}, \quad E|r_{nv}| = O(v + n^{1/2}v^{1/2}),$$

and hence $(\beta_n^2 n\tilde{m})^{-1}\sum_{v=-\tilde{m}}^{\tilde{m}}\tilde{k}(v/\tilde{m})G_{1v}\to_d\int_0^1\tilde{k}(x)dx\int_0^1 B^2(r)dr$. Finally, the stated result follows because $\sum_{v=-\tilde{m}}^{\tilde{m}}\tilde{k}(v/\tilde{m})G_{4v}\to_p 2\pi f_{uu}(0)$. ∎

9 Appendix B: technical results

Lemma 11 *Under the assumptions of Theorem 2,*

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

Proof Observe that

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

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

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

The first term on the right is bounded by $(\sup_s s|\Gamma_{y\Delta x}(s)|)^2<\infty$. The second and third terms on the right are bounded by $h\sup_s||\Gamma(s)||\sum_{l=-\infty}^{\infty}||\Gamma(l)||\le Ch$. From (39), the fourth term on the right is bounded by $h\sum_{l=-\infty}^{\infty}\sum_{r=-\infty}^{\infty}|k_{y\Delta xy\Delta x}(0,-r,l,l-r)|\le Ch$, and the stated result follows. ∎

Lemma 12 *Define $\tilde{f}^{hr}(L)=\sum_{j=0}^{\infty}\tilde{f}_j^{hr}L^j$ with $\tilde{f}_j^{hr}=\sum_{s=j+1}^{\infty}[(A_{s+r-h}^2)'A_s^1+(A_{s+r}^1)'A_{s-h}^2]$, where A_j^1 and A_j^2 are the first and second rows of A_j, respectively. Under the assumptions of Theorem 2, for $h=1,\dots,n-1$,*

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

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

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

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

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

and $\delta > 1$.

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

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

where $\xi_{n-j}^h = \varepsilon_{n-j}'\sum_{r=1}^{\infty}(\tilde{f}_j^{hr})'\varepsilon_{n-r-j}$. 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

$$E\left(\sum_{j=0}^{\infty}\xi_{n-j}^h\right)^2 = \sum_{j=0}^{\infty}E(\xi_{n-j}^h)^2 \leq C\sum_{j=0}^{\infty}\sum_{r=1}^{\infty}||\tilde{f}_j^{hr}||^2 \leq C\left(\sup_{j,r}||\tilde{f}_j^{hr}||\right)\sum_{j=0}^{\infty}\sum_{r=1}^{\infty}||\tilde{f}_j^{hr}||. \tag{56}$$

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

$$\begin{aligned} \sum_{j=0}^{\infty}\sum_{r=1}^{\infty}||\tilde{f}_j^{hr}|| &\leq \sum_{j=0}^{\infty}\sum_{r=1}^{\infty}\sum_{s=j+1}^{\infty}||A_{s+r-h}||||A_s|| + \sum_{j=0}^{\infty}\sum_{r=1}^{\infty}\sum_{s=j+1}^{\infty}||A_{s+r}||||A_{s-h}|| \\ &\leq \sum_{j=0}^{\infty}\sum_{s=j+1}^{\infty}||A_s||\sum_{r=0}^{\infty}||A_r|| + \sum_{j=0}^{\infty}\sum_{s=j+1}^{\infty}||A_{s-h}||\sum_{r=0}^{\infty}||A_r||. \end{aligned} \tag{57}$$

The first term in (57) 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 (57) 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 (56) is finite, and part (b) follows. ■

Lemma 13 *Under the assumptions of Theorem 2,*

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

where $ER_n^2 = o(1)$ *and* $Z_t = n^{-1/2}m^{-1/2}\sum_{h=1}^{n-1} k((h-1)/m)\sum_{r=1}^{\infty}\varepsilon'_{t-r}f^{hr}(1)\varepsilon_t$, $f^{hr}(1) = \sum_{j=0}^{\infty}[(A^2_{j+r-h})'A^1_j + (A^1_{j+r})'A^2_{j-h}]$, *and* A^1_j *and* A^2_j *denote the first and second row of* A_j, *respectively.*

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

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

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

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

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

$$\mathrm{tr}\left(\sum_{j=0}^{\infty}(A^2_{j-h})'A^1_j\varepsilon_{t-j}\varepsilon'_{t-j}\right) = \mathrm{tr}\ \left(f^{h0}(L)\varepsilon_t\varepsilon'_t\right), \quad f^{h0}(L) = \sum_{j=0}^{\infty}(A^2_{j-h})'A^1_jL^j = \sum_{j=0}^{\infty} f^{h0}_jL^j.$$

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

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

where $f^{hr}(L) = \sum_{j=0}^{\infty} f_j^{hr} L^j$ with $f_j^{hr} = (A_{j+r-h}^2)' A_j^1 + (A_{j+r}^1)' A_{j-h}^2$. Therefore, we may express $y_t \Delta x_{t-h}$ as

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

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

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

with $\tilde{f}^{hr}(L) = \sum_{j=0}^{\infty} \tilde{f}_j^{hr} L^j$ and $\tilde{f}_j^{hr} = \sum_{s=j+1}^{\infty} f_s^{hr} = \sum_{s=j+1}^{\infty} [(A_{s+r-h}^2)' A_s^1 + (A_{s+r}^1)' A_{s-h}^2]$. It follows that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} y_t \Delta x_{t-h} = \operatorname{tr}\left(f^{h0}(1) \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \varepsilon_t \varepsilon_t' + \sum_{r=1}^{\infty} f^{hr}(1) \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \varepsilon_t \varepsilon_{t-r}' \right) + r_{nh}, \quad (58)$$

where

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

From Lemma 12, we have

$$E|r_{nh}|^2 \leq C n^{-1}, \quad h = 1, \ldots, n-1. \quad (59)$$

Furthermore, observe that

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

In conjunction with (58), it follows that

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

and hence

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

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

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

From (59) and Minkowski's inequality, we have $E(III)^2 = O(m^{-1}n^{-1}(\sum_{h=1}^{n-1}|k(h/m)|)^2) = O(mn^{-1})$. For I, first observe that, since $A_j \equiv 0$ for $j < 0$,

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

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

Lemma 14 *For $W_{\tilde{m}}(\lambda) = (2\pi\tilde{m})^{-1}[\sin^2(\tilde{m}\lambda/2)/\sin^2(\lambda/2)]$, there exist $D \in (0,1)$ and $\kappa > 0$ such that*

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

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

$$c(\lambda/2)^2 \le \sin^2(\lambda/2) \le (\lambda/2)^2. \tag{60}$$

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

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

giving part (a). Part (b) follows from (60) and $|\sin x| \le 1$. ■

References

Amihud, Y. and C. M. Hurvich (2004). Predictive regressions: A reduced-bias estimation method. *Journal of Financial and Quantitative Analysis 39*, 813–841.

Amihud, Y., C. M. Hurvich, and Y. Wang (2004). Hypothesis testing in predictive regressions. Mimeographed, New York University.

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

Ang, A. and G. Bekaert (2005). Stock return predictability: Is it there? *Review of Financial Studies*. forthcoming.

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

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

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

Campbell, J. and M. Yogo (2006). Efficient tests of stock return predictability. *Journal of Financial Economics 81*, 27–60.

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

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

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

Dufour, J.-M. and D. Pelletier (2002). Linear estimation of weak VARMA models with a macroeconomic application. In *Proceedings of the Business and Economic Statistics Section of the American Statistical Association*, Washington, DC, 2659-2664.

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

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

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

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

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

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

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

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

Jansson, M. and M. J. Moreira (2006). Optimal inference in regression models with nearly integrated regressors. *Econometrica 74*(3), 681–714.

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

Lewellen, J. (2004). Predicting returns with financial ratios. *Journal of Financial Economics 74*, 209–235. forthcoming.

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

Marmer, V. (2004). Nonlinearity, nonstationarity and spurious forecasts. Mimeographed, Yale University.

Maynard, A. (2006). The forward premium anomaly: statistical artifact or economic puzzle? New evidence from robust tests. *Canadian Journal of Economics 39*(4), 1244–1281.

Moon, R., A. Rubia, and R. Valkanov (2004). Long-horizon regressions when the predictor is slowly variying. Mimeographed.

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

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

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

Phillips, P. C. B. (2005). Challenges of trending time series econometrics. *Mathematics and Computers in Simulation 68*, 401–416.

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

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

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

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

Saikkonen, P. and H. Lütkepohl (1996). Infinite-order cointegrated vector autoregressive processes. *Econometric Theory 12*, 814–844.

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

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

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

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

Toda, H. Y. and T. Yamamoto (1995). Statistical inference in vector autoregressions with possibly integrated processes. *Journal of Econometrics 66*, 225–250.

Torous, W., R. Valkanov, and S. Yan (2005). On predicting stock returns with nearly integrated explanatory variables. *Journal of Business 77*(4), 937–966.

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

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

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

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

Wright, J. (2000). Confidence sets for cointegrating coefficients based on stationarity tests. *Journal of Business and Economic Statistics 18*(2), 211–222.

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

c	ρ_1	σ_{12} =0	0.25	0.50	0.75	0.95	σ_{12} =0	0.25	0.50	0.75	0.95
		Demeaned Case ($n = 100$)					Detrended Case ($n = 100$)				
0	1.000	0.056	0.071	0.108	0.178	0.275	0.053	0.091	0.182	0.351	0.559
-1	0.990	0.056	0.069	0.097	0.155	0.226	0.059	0.083	0.153	0.296	0.446
-5	0.950	0.051	0.054	0.067	0.091	0.115	0.054	0.068	0.113	0.182	0.254
-10	0.900	0.056	0.058	0.057	0.079	0.087	0.052	0.069	0.085	0.120	0.158
-20	0.800	0.045	0.056	0.046	0.066	0.072	0.049	0.058	0.061	0.079	0.108
		Demeaned Case ($n = 200$)					Detrended Case ($n = 200$)				
0	1.000	0.057	0.065	0.110	0.178	0.287	0.050	0.091	0.171	0.355	0.549
-1	0.995	0.053	0.060	0.098	0.154	0.220	0.049	0.073	0.152	0.293	0.461
-5	0.975	0.059	0.072	0.073	0.094	0.109	0.049	0.070	0.115	0.179	0.246
-10	0.950	0.049	0.056	0.061	0.080	0.080	0.053	0.065	0.086	0.118	0.164
-20	0.900	0.050	0.058	0.057	0.068	0.065	0.067	0.051	0.072	0.096	0.115

The table shows rejection rates under the null hypothesis for a nominal 5% test using the standard t-statistic from a regression of y_t on x_{t-1}. y_t is given by (21) and x_t by (18) with ρ_1 given by (20), with local-to-unity parameter c. Details are given in the text.

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

c	ρ_1	σ_{12} =0	0.25	0.50	0.75	0.95	σ_{12} =0	0.25	0.50	0.75	0.95
		Demeaned Case ($n = 100$)					Detrended Case ($n = 100$)				
0	1.000	0.032	0.036	0.035	0.038	0.034	0.034	0.028	0.034	0.042	0.045
-1	0.990	0.033	0.036	0.037	0.031	0.030	0.029	0.041	0.031	0.040	0.043
-5	0.950	0.038	0.034	0.040	0.036	0.030	0.032	0.036	0.033	0.040	0.034
-10	0.900	0.036	0.033	0.030	0.033	0.026	0.034	0.034	0.032	0.030	0.024
-20	0.800	0.028	0.022	0.020	0.022	0.022	0.025	0.025	0.021	0.022	0.015
		Demeaned Case ($n = 200$)					Detrended Case ($n = 200$)				
0	1.000	0.036	0.036	0.030	0.035	0.040	0.036	0.036	0.042	0.034	0.039
-1	0.995	0.035	0.030	0.040	0.032	0.035	0.033	0.035	0.037	0.041	0.043
-5	0.975	0.034	0.030	0.037	0.041	0.041	0.037	0.029	0.035	0.029	0.036
-10	0.950	0.038	0.030	0.037	0.032	0.036	0.030	0.034	0.034	0.032	0.026
-20	0.900	0.029	0.025	0.033	0.029	0.022	0.035	0.035	0.030	0.032	0.027

The table shows rejection rates under the null hypothesis for a nominal 5% test using t_λ. y_t is given by (21) and x_t by (18) with ρ_1 given by (20), with local-to-unity parameter c. Details are given in the text.

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

c	ρ_1	$\sigma_{12}=0$	0.25	0.50	0.75	0.95	$\sigma_{12}=0$	0.25	0.50	0.75	0.95
		Demeaned Case ($k=3$)					Detrended Case ($k=3$)				
0	1.000	0.046	0.043	0.050	0.057	0.044	0.040	0.055	0.050	0.062	0.065
-1	0.990	0.044	0.041	0.044	0.044	0.048	0.050	0.048	0.048	0.060	0.059
-5	0.950	0.044	0.046	0.037	0.036	0.040	0.046	0.039	0.041	0.054	0.052
-10	0.900	0.035	0.037	0.043	0.039	0.032	0.040	0.038	0.042	0.036	0.037
-20	0.800	0.028	0.032	0.024	0.024	0.027	0.030	0.030	0.024	0.032	0.035
		Demeaned Case ($k=5$)					Detrended Case ($k=5$)				
0	1.000	0.055	0.060	0.058	0.059	0.064	0.048	0.049	0.049	0.074	0.089
-1	0.990	0.046	0.052	0.050	0.054	0.053	0.055	0.049	0.057	0.062	0.072
-5	0.950	0.046	0.036	0.042	0.050	0.049	0.043	0.050	0.038	0.057	0.057
-10	0.900	0.034	0.036	0.033	0.037	0.036	0.037	0.036	0.046	0.041	0.052
-20	0.800	0.020	0.017	0.024	0.026	0.032	0.026	0.017	0.030	0.028	0.041

The table shows rejection rates under the null hypothesis for a nominal 5% test using t_λ. The long-horizon return $y_{t,k}$ and $x_{t,k}$ are given by (28) where x_t and y_t follow (18) and (21) respectively with ρ_1 given by (20), with local-to-unity parameter c. $k = 3$ is chosen to match the ratio of the sample size to the longest horizon in the empirical application for a simulation sample size of $n = 100$. Details are given in the text.

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

c	$\rho_1+\rho_2$	$\sigma_{12}=0$	0.25	0.50	0.75	0.95	$\sigma_{12}=0$	0.25	0.50	0.75	0.95
		Demeaned Case ($n=100$)					Detrended Case ($n=100$)				
0	1.000	0.048	0.057	0.053	0.052	0.054	0.048	0.050	0.060	0.063	0.081
-1	0.990	0.053	0.056	0.049	0.058	0.057	0.049	0.055	0.050	0.065	0.062
-5	0.950	0.062	0.048	0.057	0.063	0.060	0.051	0.056	0.053	0.067	0.060
-10	0.900	0.046	0.056	0.058	0.060	0.054	0.061	0.051	0.049	0.052	0.056
-20	0.800	0.052	0.046	0.043	0.056	0.068	0.053	0.058	0.050	0.058	0.062
		Demeaned Case ($n=200$)					Detrended Case ($n=200$)				
0	1.000	0.059	0.055	0.061	0.061	0.057	0.054	0.056	0.062	0.067	0.066
-1	0.995	0.056	0.054	0.062	0.052	0.061	0.051	0.054	0.062	0.067	0.070
-5	0.975	0.057	0.047	0.056	0.050	0.059	0.058	0.044	0.060	0.054	0.064
-10	0.950	0.066	0.058	0.058	0.059	0.066	0.052	0.057	0.066	0.057	0.049
-20	0.900	0.053	0.060	0.051	0.063	0.053	0.052	0.061	0.057	0.064	0.065

The table shows rejection rates under the null hypothesis for a nominal 5% test using t_λ. y_t is given by (21) and x_t by (19) with ρ_1 and ρ_2 given by (29), with local-to-unity parameter c. Details are given in the text.

Table 5: Covariance-based t-statistic: finite sample power ($y_t = \beta x_{t-1} + \varepsilon_{1,t}$)

c	σ_{12}	$\beta = 0.10$	0.15	0.20	0.35	0.50
A. Demeaned Case ($n = 100$)						
$c = 0$	0.500	0.037	0.103	0.208	0.709	0.968
($\rho_1 = 1.000$)	0.950	0.036	0.099	0.199	0.694	0.975
$c = -1$	0.500	0.155	0.320	0.537	0.861	0.931
($\rho_1 = 0.990$)	0.950	0.264	0.465	0.643	0.869	0.926
$c = -2.5$	0.500	0.155	0.340	0.536	0.878	0.953
($\rho_1 = 0.975$)	0.950	0.270	0.504	0.655	0.897	0.945
$c = -7.5$	0.500	0.155	0.303	0.498	0.897	0.973
($\rho_1 = 0.925$)	0.950	0.220	0.444	0.621	0.912	0.972
$c = -20$	0.500	0.105	0.251	0.408	0.870	0.987
($\rho_1 = 0.800$)	0.950	0.141	0.308	0.470	0.882	0.985
B. Demeaned Case ($n = 200$)						
$c = 0$	0.500	0.088	0.242	0.506	0.973	1.000
($\rho_1 = 1.000$)	0.950	0.071	0.221	0.488	0.972	1.000
$c = -1$	0.500	0.410	0.720	0.867	0.959	0.971
($\rho_1 = 0.995$)	0.950	0.597	0.806	0.905	0.959	0.966
$c = -2.5$	0.500	0.416	0.717	0.891	0.983	0.985
($\rho_1 = 0.988$)	0.950	0.620	0.851	0.936	0.983	0.982
$c = -7.5$	0.500	0.345	0.663	0.870	0.992	0.998
($\rho_1 = 0.963$)	0.950	0.572	0.827	0.942	0.991	0.993
$c = -20$	0.500	0.302	0.602	0.826	0.995	1.000
($\rho_1 = 0.900$)	0.950	0.444	0.744	0.909	0.995	1.000

The table shows rejection rates under the alternative hypothesis for a nominal 5% test using t_λ. y_t is given by (24) and x_t by (18) with ρ_1 given by (20), with local-to-unity parameter c. Details are given in the text.

Table 6: Covariance-based t-statistic: finite sample power ($y_t = \gamma(1-\rho L)x_{t-1} + \varepsilon_{1,t}$)

c	σ_{12}	$\gamma = 0.10$	0.15	0.20	0.35	0.50
(ρ_1)		$r^2 = 0.01$	0.02	0.04	0.11	0.20
A. Demeaned Case ($n = 100$)						
$c = 0$	0.500	0.085	0.172	0.264	0.664	0.919
$(\rho_1 = 1.000)$	0.950	0.080	0.156	0.252	0.639	0.884
$c = -1$	0.500	0.087	0.153	0.273	0.668	0.914
$(\rho_1 = 0.990)$	0.950	0.089	0.173	0.277	0.660	0.903
$c = -2.5$	0.500	0.088	0.175	0.281	0.691	0.931
$(\rho_1 = 0.975)$	0.950	0.089	0.162	0.276	0.666	0.901
$c = -7.5$	0.500	0.088	0.179	0.267	0.683	0.918
$(\rho_1 = 0.925)$	0.950	0.076	0.133	0.249	0.623	0.893
$c = -20$	0.500	0.078	0.149	0.229	0.646	0.910
$(\rho_1 = 0.800)$	0.950	0.044	0.106	0.154	0.531	0.873
B. Demeaned Case ($n = 200$)						
$c = 0$	0.500	0.152	0.317	0.520	0.925	0.997
$(\rho_1 = 1.000)$	0.950	0.154	0.314	0.500	0.907	0.992
$c = -1$	0.500	0.156	0.318	0.532	0.930	0.995
$(\rho_1 = 0.995)$	0.950	0.150	0.314	0.507	0.913	0.994
$c = -2.5$	0.500	0.161	0.338	0.518	0.937	0.996
$(\rho_1 = 0.988)$	0.950	0.166	0.322	0.512	0.912	0.998
$c = -7.5$	0.500	0.148	0.320	0.539	0.922	0.998
$(\rho_1 = 0.963)$	0.950	0.136	0.295	0.497	0.900	0.994
$c = -20$	0.500	0.134	0.292	0.481	0.921	0.997
$(\rho_1 = 0.900)$	0.950	0.088	0.192	0.332	0.858	0.993

The table shows rejection rates under the alternative hypothesis for a nominal 5% test using t_λ. y_t is given by (25) and x_t by (18) with ρ_1 given by (20), with local-to-unity parameter c. Details are given in the text.

Table 7: CES Bonferroni method: finite sample power ($y_t = \beta x_{t-1} + \varepsilon_{1,t}$)

c	σ_{12}	$\beta = 0.10$	0.15	0.20	0.35	0.50
		A. Demeaned Case ($n = 100$)				
$c = 0$	0.500	0.193	0.378	0.568	0.945	0.999
($\rho_1 = 1.000$)	0.950	0.226	0.397	0.599	0.961	1.000
$c = -1$	0.500	0.727	0.904	0.970	1.000	1.000
($\rho_1 = 0.990$)	0.950	0.642	0.829	0.940	0.997	1.000
$c = -2.5$	0.500	0.686	0.887	0.969	1.000	1.000
($\rho_1 = 0.975$)	0.950	0.604	0.832	0.927	0.998	1.000
$c = -7.5$	0.500	0.582	0.826	0.939	0.999	1.000
($\rho_1 = 0.925$)	0.950	0.514	0.784	0.903	0.996	1.000
$c = -20$	0.500	0.365	0.642	0.823	0.994	1.000
($\rho_1 = 0.800$)	0.950	0.351	0.625	0.782	0.989	1.000
		B. Demeaned Case ($n = 200$)				
$c = 0$	0.500	0.333	0.617	0.830	0.998	1.000
($\rho_1 = 1.000$)	0.950	0.373	0.643	0.871	1.000	1.000
$c = -1$	0.500	0.961	0.999	1.000	1.000	1.000
($\rho_1 = 0.995$)	0.950	0.929	0.991	1.000	1.000	1.000
$c = -2.5$	0.500	0.964	0.995	1.000	1.000	1.000
($\rho_1 = 0.988$)	0.950	0.922	0.992	1.000	1.000	1.000
$c = -7.5$	0.500	0.908	0.993	1.000	1.000	1.000
($\rho_1 = 0.963$)	0.950	0.891	0.976	0.999	1.000	1.000
$c = -20$	0.500	0.803	0.968	0.996	1.000	1.000
($\rho_1 = 0.900$)	0.950	0.769	0.950	0.993	1.000	1.000

The table shows rejection rates under the alternative hypothesis for a nominal 5% test using the CES Bonferroni test. y_t is given by (24) and x_t by (18) with ρ_1 given by (20), with local-to-unity parameter c. Details are given in the text.

Table 8: CES Bonferroni method: finite sample power ($y_t = \gamma(1-\rho L)x_{t-1} + \varepsilon_{1,t}$)

c	σ_{12}	$\gamma = 0.10$	0.15	0.20	0.35	0.50
(ρ_1)		$r^2 = 0.01$	0.02	0.04	0.11	0.20
A. Demeaned Case ($n = 100$)						
$c = 0$	0.500	0.044	0.041	0.044	0.060	0.097
$(\rho_1 = 1.000)$	0.950	0.036	0.036	0.030	0.034	0.048
$c = -1$	0.500	0.051	0.039	0.059	0.088	0.135
$(\rho_1 = 0.990)$	0.950	0.028	0.032	0.030	0.041	0.060
$c = -2.5$	0.500	0.045	0.069	0.084	0.143	0.234
$(\rho_1 = 0.975)$	0.950	0.030	0.046	0.044	0.081	0.146
$c = -7.5$	0.500	0.082	0.098	0.146	0.297	0.468
$(\rho_1 = 0.925)$	0.950	0.059	0.087	0.111	0.213	0.365
$c = -20$	0.500	0.093	0.170	0.232	0.546	0.829
$(\rho_1 = 0.800)$	0.950	0.092	0.149	0.208	0.512	0.810
B. Demeaned Case ($n = 200$)						
$c = 0$	0.500	0.034	0.041	0.046	0.061	0.100
$(\rho_1 = 1.000)$	0.950	0.037	0.038	0.024	0.031	0.043
$c = -1$	0.500	0.043	0.043	0.066	0.080	0.141
$(\rho_1 = 0.995)$	0.950	0.032	0.030	0.032	0.038	0.064
$c = -2.5$	0.500	0.064	0.059	0.072	0.145	0.235
$(\rho_1 = 0.988)$	0.950	0.040	0.044	0.047	0.077	0.149
$c = -7.5$	0.500	0.076	0.103	0.137	0.277	0.472
$(\rho_1 = 0.963)$	0.950	0.068	0.092	0.109	0.238	0.382
$c = -20$	0.500	0.106	0.175	0.239	0.574	0.844
$(\rho_1 = 0.900)$	0.950	0.098	0.137	0.200	0.501	0.813

The table shows rejection rates under the alternative hypothesis for a nominal 5% test using the CES Bonferroni test. y_t is given by (25) and x_t by (18) with ρ_1 given by (20), with local-to-unity parameter c. Details are given in the text.

Table 9: Covariance-based t-statistic: Comparison of kernels

A.		$y_t = \beta x_{t-1} + \varepsilon_{1,t}$	(Demeaned Case, $n = 100$)					
c	σ_{12}	kernel	$\beta = 0$	0.10	0.15	0.20	0.35	0.50
$c = -1$	0.500	PZ	0.038	0.183	0.381	0.621	0.889	0.936
		QS	0.041	0.161	0.325	0.546	0.940	0.998
		NW	0.037	0.155	0.320	0.537	0.861	0.931
$c = -1$	0.950	PZ	0.031	0.297	0.505	0.682	0.888	0.939
		QS	0.028	0.142	0.303	0.487	0.920	0.995
		NW	0.030	0.264	0.465	0.643	0.869	0.926
$c = -20$	0.500	PZ	0.022	0.119	0.275	0.461	0.895	0.990
		QS	0.019	0.130	0.273	0.453	0.906	0.996
		NW	0.020	0.105	0.251	0.408	0.870	0.987
$c = -20$	0.950	PZ	0.024	0.148	0.338	0.507	0.903	0.988
		QS	0.022	0.074	0.160	0.281	0.816	0.980
		NW	0.022	0.141	0.308	0.470	0.882	0.985
B.		$y_t = \gamma(1 - \rho L)x_{t-1} + \varepsilon_{1,t}$	(Demeaned Case, $n = 100$)					
c	σ_{12}	kernel	$\gamma = 0$	0.10	0.15	0.20	0.35	0.50
$c = -1$	0.500	PZ	0.038	0.078	0.141	0.248	0.630	0.909
		QS	0.041	0.089	0.166	0.278	0.673	0.925
		NW	0.037	0.087	0.153	0.273	0.668	0.914
$c = -1$	0.950	PZ	0.031	0.071	0.143	0.221	0.609	0.879
		QS	0.028	0.085	0.162	0.257	0.643	0.901
		NW	0.030	0.089	0.173	0.277	0.660	0.903
$c = -20$	0.500	PZ	0.022	0.066	0.129	0.209	0.600	0.893
		QS	0.019	0.065	0.128	0.207	0.596	0.882
		NW	0.020	0.078	0.149	0.229	0.646	0.910
$c = -20$	0.950	PZ	0.024	0.038	0.081	0.116	0.461	0.817
		QS	0.022	0.039	0.080	0.117	0.435	0.783
		NW	0.022	0.044	0.106	0.154	0.531	0.873

The table shows rejection rates for a 5% test using t_λ under both the null (Column 4) and the alternative hypothesis (Columns 5-9). The two forms of the alternative hypothesis shown in Panels A and B match those in tables 5 and 6 respectively. Column 3 gives the kernel choice for $k(x)$ in (11), with PZ, QZ, and NW denoting the Parzen, Quadratic Spectral and Newey-West (Bartlett) kernels respectively. x_t is specified by (18) with ρ_1 given by (20), with local-to-unity parameter c. Details are given in the text.

Table 10: Regressions of k-period long-horizon real stock returns on the treasury bill and dividend price ratio

		Treasury Bills				Dividend Price Ratio			
		Forecast Horizon (k)				Forecast Horizon (k)			
sample		k = 1.0	3.0	12.0	24.0	k = 1.0	3.0	12.0	24.0
1927-	$\widehat{\beta}$	-0.581	-1.676	-3.236	-2.168	0.006	0.021	0.100	0.202
2003	R^2	0.001	0.002	0.002	0.000	0.002	0.008	0.039	0.080
	t_β	-0.830	-0.881	-0.433	-0.209	1.395	1.402	1.950	2.341
	$p - V_{sup}$	0.368	0.326	0.650	0.842	0.972	0.577	0.600	0.605
1927-	$\widehat{\beta}$	0.615	-0.876	-22.184	-94.145	0.014	0.057	0.267	0.620
1951	R^2	0.000	0.000	0.007	0.071	0.003	0.013	0.071	0.200
	t_β	0.153	-0.081	-0.564	-1.125	0.920	0.913	1.834	3.883
	$p - V_{sup}$	0.865	0.974	0.700	0.363	0.901	0.900	0.993	0.792
1952-	$\widehat{\beta}$	-1.410	-3.910	-8.098	-3.027	0.025	0.079	0.332	0.598
1994	R^2	0.007	0.017	0.017	0.001	0.016	0.049	0.194	0.342
	t_β	-1.916	-2.023	-0.985	-0.260	2.867	3.752	3.949	3.775
	$p - V_{sup}$	0.038	0.025	0.299	0.745	0.080	0.010	0.042	0.048
1952-	$\widehat{\beta}$	-1.290	-3.537	-6.764	-3.507	0.008	0.028	0.125	0.238
2003	R^2	0.005	0.011	0.009	0.001	0.005	0.018	0.081	0.146
	t_β	-1.772	-1.820	-0.827	-0.314	1.849	2.162	2.358	2.167
	$p - V_{sup}$	0.057	0.035	0.386	0.752	0.626	0.218	0.375	0.470

Entries show results from a regression of $y_{t+k} = r_{t+1} + \ldots + r_{t+k}$ on $x_t = i_t$ or $x_t = d_t - p_t$. Regressions are estimated by OLS with HAC standard errors for $k > 1$, using the Bartlett (Newey-West) kernel with bandwidth set to $k - 1$. $p - V_{sup}$ is the two-sided p-value from the Valkanov (2003) sup-bound test.

Table 11: Confidence intervals on largest roots and residual correlation

	$x_t = \mu_x + v_t$ $y_t = \beta_0 + \beta_1 x_{t-1} + \varepsilon_{1,t}$			$(1-\alpha L)b(L)v_t = \varepsilon_{2,t}$ $\delta = corr\,(\varepsilon_{1,t}, \varepsilon_{2,t})$		
	Treasury Bills ($y_t = r_t \quad x_t = i_t$)			Dividend Price Ratio ($y_t = r_t \quad x_t = d_t - p_t$)		
sample period	95 % CI on largest root in x_t		$\widehat{\delta}$	95 % CI on largest root in x_t		$\widehat{\delta}$
1927 to 2003	(0.9841	1.004)	-0.1066	(0.9839	1.004)	-0.9506
1927 to 1951	(0.9507	1.011)	0.0245	(0.9148	1.004)	-0.9328
1952 to 1994	(0.9670	1.006)	-0.2965	(0.9587	1.004)	-0.9807
1952 to 2003	(0.9733	1.005)	-0.1890	(0.9871	1.007)	-0.9732

Confidence intervals on the largest root are based on Stock (1991) using the Ng and Perron (2001) MIC criteria to select lag-length with a maximum of six lags.

Table 12: Covariance-based orthogonality tests on k-period long-horizon real stock returns using the treasury bill and dividend price ratio

		Treasury Bills				Dividend Price Ratio			
		Forecast Horizon (k)				Forecast Horizon (k)			
sample		k = 1.0	3.0	12.0	24.0	k = 1.0	3.0	12.0	24.0
1927-2003	t_λ	0.073	-0.305	-1.050	-0.529	-0.974	1.052	-1.672	-0.016
	p-value	0.942	0.760	0.294	0.597	0.330	0.293	0.095	0.987
	$[m^*]$	[0.56]	[0.35]	[0.15]	[0.13]	[2.90]	[1.33]	[0.50]	[0.30]
1927-1951	t_λ	0.696	1.086	0.507	0.612	-0.696	0.996	-1.408	-0.438
	p-value	0.486	0.277	0.612	0.541	0.486	0.319	0.159	0.661
	$[m^*]$	[0.66]	[0.30]	[0.11]	[0.07]	[2.62]	[1.10]	[0.31]	[0.41]
1952-1994	t_λ	-1.385	-1.883	-2.088	-1.392	-1.384	-0.629	-0.145	1.442
	p-value	0.166	0.060	0.037	0.164	0.166	0.529	0.885	0.149
	$[m^*]$	[3.61]	[2.26]	[0.90]	[0.55]	[1.60]	[0.65]	[0.25]	[0.16]
1952-2003	t_λ	-0.864	-1.491	-1.718	-1.099	-1.269	-0.487	-0.364	0.964
	p-value	0.388	0.136	0.086	0.272	0.204	0.626	0.716	0.335
	$[m^*]$	[2.69]	[1.64]	[0.68]	[0.48]	[1.54]	[0.66]	[0.24]	[0.14]

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

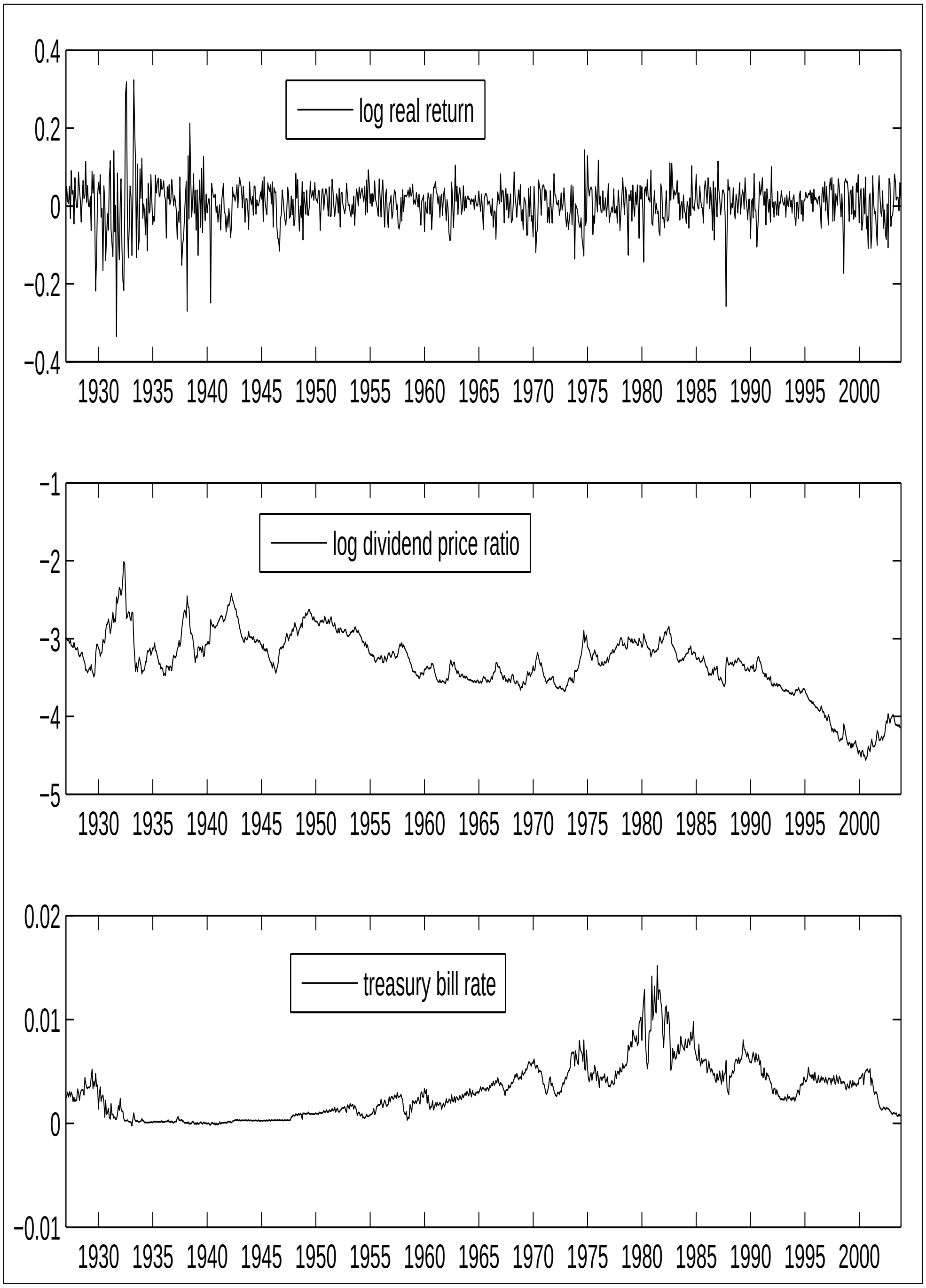

Figure 1: Log real return, log dividend price ratio, and treasury bill rate