The Power of Long-Horizon Predictive Regression Tests

Nelson C. Mark*
Department of Economics and Econometrics
University of Notre Dame and NBER

Donggyu Sul
Department of Economics
University of Auckland

October 10, 2006

Abstract

In a long-horizon regression, a $k-$period future return is regressed on a current variable such as the log dividend yield. The p-value of the t-test that the return is unpredictable typically declines over some range of return horizons, k. Local asymptotic analysis shows that the power of the long-horizon regression test dominates that of the short-horizon test over a nontrivial region of the admissible parameter space. However, OLS is biased in small samples and this bias distorts the size of asymptotic tests. We correct for test-size distortion with a recursive moving-block Bartlett correction. Application of the the Bartlett corrected test to historical equity returns yield evidence that the log dividend yield predicts returns at the 13 year horizon.

Keywords: Predictive regression, Long horizons, Stock returns, Local asymptotic power
JEL Classification: G12, C12, C22

*Corresponding author. Department of Economics and Econometrics, 437 Flanner Hall, University of Notre Dame, Notre Dame, IN 46556. Phone: 574-631-0518, Fax: 574-631-4783, e-mail: nmark@nd.edu

1 Introduction

Let $r_t \sim I(0)$ be the return on an asset or a portfolio of assets from time $t-1$ to t and x_t be a persistent hypothesized predictor of the asset's future returns. In finance r_t might be the return on equity and x_t the log dividend yield whereas in international finance r_t might be the return on the log exchange rate and x_t the deviation of the exchange rate from its fundamental value.[1] A test of return predictability can be conducted by regressing r_{t+1} on x_t and performing a t-test on the slope coefficient. Empirical research in finance frequently goes beyond this by regressing the asset's multi-period future return $y_{t,k} = \sum_{j=1}^{k} r_{t+j}$ on x_t,

$$y_{t,k} = \mu_k + \beta_k x_t + \epsilon_{t,k}, \tag{1}$$

and conducting a t-test of the null hypothesis $H_0 : \beta_k = 0$, where the t-statistic is constructed with a heteroskedastic and autocorrelation consistent (HAC) standard error. It is typically found that OLS slope estimates, asymptotic t-ratios, and R^2s increase over a range of horizons $k > 1$. Are the t-ratios increasing because the long-horizon test has more power to reject the null than the short-horizon test? Considering that the long-horizon regression is built by aggregation of intervening short-horizon regressions, the underlying basis for these results are not fully understood. As stated by Campbell et al. (1997), "An important unresolved question is whether there are circumstances under which long-horizon regressions have greater power to detect deviations from the null hypothesis than do short-horizon regressions."

In this paper, we address the power question posed by Campbell et al. We conduct the asymptotic power analysis both under the assumption that the regressor is covariance stationary and, because predictive variables used in empirical work are often very persistent, under the assumption that the regressor has a local-to-unity dominant autoregressive root. Under stationarity, testing is straightforward and can be done with the t-test. In the local-to-unity case, we approach testing with a variant of the sup-bound test discussed by Cavanaugh et al. (1995) and employ the sup–t^2 test which is asymptotically free from nuisance parameter dependencies.

[1]This line of research includes Fama and French (1988a) and Campbell and Shiller (1988) who regressed long-horizon equity returns on the log dividend yield. See also Mishkin (1992), who ran regressions of long-horizon inflation on long-term bond yields, Coe and Nason (2004) who regress long-horizon GDP growth on long-horizon money growth, Mark (1995), Mark and Choi (1997), Chinn and Meese (1995) and Rapach and Wohar (2002) who regressed long-horizon exchange rate returns on the deviation of the exchange rate from its fundamental value, Alexius (2001) and Chinn and Merideth (2002) who regress long-horizon exchange rate returns on long-term bond yield differentials.

Whether the regressor is stationary or is local-to-unity, we show that asymptotic power advantages can accrue to long-horizon regression tests in empirically relevant regions of the parameter space—specifically when $\{x_t\}$ is positively autocorrelated and persistent, when the short-horizon regression error exhibits low to moderate serial correlation, and when the regressor is endogenous.

While our asymptotic analysis answers the power question posed above in the affirmative, direct application of the asymptotic tests in small samples can result in misleading inference because they exhibit some size distortion. The cause of the size distortion is the (well-known) small sample OLS bias from the predictive regression [Stambaugh (1999)]. We confront the size-distortion issue by suggesting a recursive moving-block (RMB) Bartlett correction to the test statistic. The RMB Bartlett correction is a robust procedure that also preserves the power advantages for long-horizon tests.

Previous research on the econometrics of predictive regressions include Campbell (2001) who assumes an AR(1) regressor $\{x_t\}$ and a serially uncorrelated short-horizon regression error. Using the concept of approximate slope to measure its asymptotic power, he found that long-horizon regressions had approximate slope advantages over short-horizon regressions but his Monte Carlo experiments did not reveal systematic power advantages for long-horizon regressions in finite samples. Berben (2000) reported asymptotic power advantages for long-horizon regression when the exogenous predictor and the short-horizon regression error follow AR(1) processes whereas Berben and Van Dijk (1998) conclude that long-horizon tests do not have asymptotic power advantages when the regressor is unit-root nonstationary and is weakly exogenous—properties that Berkowitz and Giorgianni (2001) also find in simulation work. Mankiw and Shapiro (1986), Hodrick (1992), Kim and Nelson (1993), Goetzmann and Jorion (1993), Mark (1995), Kilian (1999), and Kilian and Taylor (2003) study small-sample inference issues using simulation methods. Stambaugh (1999) proposes a Bayesian analysis to deal with small-sample OLS bias and Campbell and Yogo (2002) study point optimal tests in the short-horizon predictive regression. Clark and McCracken (2001), Bhansali (1997), and Ing (2003) study the predictive power of long-horizon out-of-sample forecasts.

The long-horizon predictive regressions that we study regress returns at alternative horizons on the same explanatory variable. The regressions admit variations in k but the horizon is implicitly constrained to be small relative to the sample size in the sense that $k/T \to 0$ as $T \to \infty$. An alternative long-horizon regression employed in the literature regresses the future k-period return (from t to $t+k$) on the past k-period return (from $t-k$ to t) [Fama and French (1988b)]. In this alternative long-horizon regression, the return horizon k can be large relative to the size of the sample T. Richardson and Stock (1989) develop an alternative asymptotic theory where $k \to \infty$ and $T \to \infty$ but $k/T \to \kappa \in (0,1)$ and show that the test statistics converge to functions of Brownian motions. Daniel (2001) studies optimal tests of this kind, Kim et al. (1991) study the OLS sampling distribution with the bootstrap and randomization techniques, and Valkanov (2003) employs the Richardson and Stock asymptotic distribution theory to the long-horizon regressions when the regressor x_t follows a local-to-unity process.

The paper is organized as follows. To set the stage for our inquiry, the next section presents the canonical example of the use of long-horizon predictive regressions in finance–that of regressing future equity returns on the log dividend yield. Section 3 presents the local asymptotic power analysis and Section 4 discusses the RMB Bartlett correction of the test statistic. Results of simulation work to characterize the small sample properties of the long-horizon tests are presented in Section 5. Section 6 applies the RMB Bartlett corrected sup–t^2 test to test whether stock returns are predictable and Section 7 concludes. Proofs of the main results are contained in the appendix.

2 Canonical example

To motivate the issues, we revisit the question of whether the log dividend yield predicts future stock returns [Fama and French (1988b), Campbell and Shiller (1988)]. The predictive regression can be motivated as in Campbell et al. (1997) who show how the log dividend yield is the expected present value of future returns net of future dividend growth. If forecasts of future dividend growth are relatively smooth, this present-value relation suggests that the log dividend yield contains information useful for predicting future returns.

Let $r_{t+1} = \ln\left(\left(P_{t+1} + D_{t+1}\right)/P_t\right)$ and $x_t = \ln\left(D_t/P_t\right)$, where P_t is the beginning of year price of the S&P index and D_t is the annual flow of dividends in year t. We use annual observations from 1871 to 2002 and regress future returns at horizons of 1,5,10, and 15 years on x_t.[2] The results are shown in Table 1.

[2] These data were used in Robert J. Shiller (2000) and were obtained from his web site. Annual observations were constructed from these monthly data. Because the dependent variable changes with k, the R^2s are not directly comparable across horizons.

Table 1: Short and long-horizon equity return regressions

	$k=1$	$k=5$	$k=10$	$k=15$
$\widehat{\beta}_k$	0.072	0.250	0.716	1.206
$t_\beta(k)$	1.330	1.194	2.499	4.965
R^2	0.02	0.05	0.15	0.29

The OLS point estimates of the slope, conventional asymptotic t-ratios, and regression R^2s increase with return horizon which suggests that evidence for return predictability strengthens as the return horizon is lengthened. The conventional t-test cannot reject the null of no predictability at $k=1$ or 5 but does reject at $k=10$ and 15. The pattern exhibited between the point estimates and horizon is quite familiar in the literature and has come to be viewed as a stylized fact. To explain these features of the data, Campbell and Cochrane (1999) propose an asset pricing model in which the representative agent's preferences display habit persistence while Cecchetti et al. (2000) present a model in which the representative agent's beliefs are distorted.

How persistent is the regressor? The log dividend yield (the regressor) has a first-ordered autocorrelation of 0.843. Testing for a unit root in $\{x_t\}$ yields augmented Dickey–Fuller (ADF) test statistic values of -0.189 (with constant) and -1.106 (constant and trend) and Phillips–Perron (PP) test statistics of –0.850 (constant) and –2.01 (constant and trend).[3] The apparent nonstationarity of the dividend yield is driven in large part by the bull market of the late 1990s. When we end the sample in 1997, however, the ADF statistics become -2.965 (constant) and -3.758 (constant and trend). Corresponding PP-statistics are -2.640 (constant) and -3.656 (constant and trend). In summary, the log-dividend yield appears to be borderline stationary.

3 Local asymptotic power

We study local asymptotic power from two perspectives. First, we assume that both the return sequence $\{y_t\}$ and the dividend-yield $\{x_t\}$ are covariance stationary. Although regressors used in practice are often persistent, the $I(0)$ assumption turns out not to be

[3] Approximate critical values for the test (with constant) are -2.86, -2.86, and -2.89, respectively at the 5% level and -2.57, -2.57, and -2.58, respectively at the 10% level. Approximate critical values for the test (constant and trend) are -3.41, -3.43, and -3.45 respectively at the 5% level and -3.12, -3.13, and -3.15 respectively at the 10% level.

restrictive because the results that we obtain under stationarity can be applied directly in cases where the regressor follows a weak unit-root process in the sense of Park (2003a, 2006). But because weakly integrated time series are not widely discussed, we also study local asymptotic power under the more familiar assumption that the regressor has a local-to-unity autoregressive root.

Before proceeding with our analysis, however, we briefly discuss related work by Bhansali (1997) and Ing (2003) whose findings of long-horizon forecasting dominance under misspecification is similar in spirit to our findings. These authors compare predictive accuracy between direct (long-horizon) and plug-in (short-horizon) forecasts from an estimated AR(1) regression of a univariate series ARI(p, d) series. The direct $k-$period ahead forecast comes from the $k-$period regression $\hat{x}^d_{t+k|t} = \hat{\rho}_k x_t$ where $\hat{\rho}_k$ is the point estimate from the regression of x_{t+k} on x_t whereas the iterative plug-in forecast is $\hat{x}^p_{t+k} = \hat{\rho}^k_1 x_t,$ where $\hat{\rho}_1$ is the point estimate from regressing x_{t+1} on x_t. $\hat{x}^d_{t+k|t}$ is the analog to the long-horizon predictor and $\hat{x}^p_{t+k|t}$ is the analog to the short-horizon forecast. In the stationary case, Bhansali (1997) showed that the long-horizon forecasts dominate the short-horizon forecasts in asymptotic mean-square error under misspecification of the AR(1) regression (i.e., when the truth is $p > 1$ and $d = 0$). Not surprisingly, he finds that the forecasts are asymptotically equivalent when the AR(1) is correctly specified. Ing (2003) obtains a similar asymptotic equivalence result under correct specification in the nonstationary case ($p = 1$ and $d = 1$). The lesson from these studies is if the long-horizon prediction tests that we study are to have asymptotic power advantages over short-horizon tests, it must occur under endogeneity of the regressor (misspecification in our context). Thus a key issue is whether there is any correlation between the regressor and the regression error.

To set notation, we will work with predictive regressions of the form

$$r_{t+1} = \mu_r + \beta_1 x_t + e_{t+1}.$$

In the stationary case, we will suppress the regression constant since it has no effect on the asymptotic properties of the predictive regression tests. We will, however, reintroduce the constant in the local-to-unity analysis and in the simulation work on the small sample properties of the tests. Economic theory typically provides guidance on the appropriate sign of the slope coefficient under the alternative. Throughout the paper, we restrict out attention to the one-sided alternative for which $\beta_1 > 0$.

3.1 Local asymptotic power under a covariance stationary regressor

For our local asymptotic analysis under covariance stationarity, the observations will be generated according to

Assumption 1 *(Covariance stationarity) For sample size T, the observations have the representation*

$$
\begin{aligned}
r_{t+1} &= \beta_1(T)x_t + e_{t+1}, &(2)\\
x_{t+1} &= \rho x_t + u_{t+1}, &(3)
\end{aligned}
$$

where $-1 < \rho < 1$ and $\{e_{t+1}\}$ and $\{u_{t+1}\}$ are zero mean covariance stationary sequences. $\beta_1(T) = b/\sqrt{T}$ and $c(T) = c/\sqrt{T} = E\left(\sum_{t=1}^{T} x_t e_{t+1}\right)\left(\sum_{t=1}^{T} x_t^2\right)^{-1}$ give the sequence of local alternatives where b and c are constants.

We note that the endogeneity of the regressor, characterized by $c\left(T\right)$, is local-to-zero. The long-horizon regression ($k > 1$) obtained by addition of short-horizon regressions is

$$
y_{t,k} = \sum_{j=1}^{k} r_{t+j} = \beta_k(T)x_t + \epsilon_{t,k},
$$

where

$$
\begin{aligned}
\beta_k(T) &= \beta_1(T)\left[1 + \sum_{j=1}^{k-1} \rho^j\right] = \frac{b}{\sqrt{T}}\left(\frac{1-\rho^k}{1-\rho}\right),\\
\epsilon_{t,k} &= \sum_{j=1}^{k} e_{t+j} + \beta_1(T)\left(\sum_{j=1}^{k-1} u_{t+j}\right). &(4)
\end{aligned}
$$

Under the sequence of local alternatives, the OLS estimator at horizon $k > 1$ has probability limit $\left(b_k + c_k\right)/\sqrt{T}$, where $c_k(T) = c_k/\sqrt{T} = E\left(\sum_{t=1}^{T} x_t \epsilon_{t,k}\right)\left(\sum_{t=1}^{T} x_t^2\right)^{-1}$ and $b_k\left(T\right) = b_k/\sqrt{T} = \left(b/\sqrt{T}\right)\left(1-\rho^k\right)\left(1-\rho\right)^{-1}$. Because the direct dependence of $\epsilon_{t,k}$ on the projection errors u_{t+j} vanish asymptotically, the asymptotic variance of the OLS estimator may be calculated under the null hypothesis of no predictability ($c_k = b_k = 0$, $k > 0$). Under the sequence of local alternatives, the squared t-ratio for the test of the null hypothesis $H_0 : \beta_k = 0$ has the asymptotic noncentral chi-square distribution,

$$
t_{\beta}^2\left(k\right) = \frac{T\hat{\beta}_k^2}{V(\hat{\beta}_k)} \xrightarrow{D} \chi_1^2(\lambda_k),
$$

with noncentrality parameter,

$$\lambda_k = \frac{(b_k + c_k)^2}{V(\hat{\beta}_k)}.$$

The long-horizon test will have more power than the short-horizon test if the noncentrality parameter for horizon k exceeds the noncentrality parameter for horizon 1. To make the dependence of local asymptotic power on the DGP's parameter values explicit, we let γ be the parameter vector that characterizes the DGP and let $\theta(k,\gamma) = \lambda_k/\lambda_1$ be the measure of relative local asymptotic power between long-and short-horizon regression. We can now state

Proposition 1 *Under Assumption 1, the long-horizon* ($k > 1$) *regression test of the hypothesis that* x_t *does not predict future* r_t *has asymptotic local power advantage over the short-horizon* ($k = 1$) *regression test if*

$$\theta(k,\gamma) = \frac{\lambda_k}{\lambda_1} = \underset{T\to\infty}{plim}\left[\frac{\hat{\beta}_k}{\hat{\beta}_1}\right]^2 \left[\frac{V(\hat{\beta}_1)}{V(\hat{\beta}_k)}\right] = \left[\frac{b_k + c_k}{b + c}\right]^2 \left[\frac{\Omega_{ee}}{\Omega_{\epsilon\epsilon}(k)}\right] > 1,$$

where Ω_{ee} *and* $\Omega_{\epsilon\epsilon}(k)$ *are the long run variances of* e_{t+1} *in (2) and* $\epsilon_{t,k}$ *in (4), respectively.*

Exogenous regressor If the regressor is exogenous, then the power advantage condition of proposition 1 becomes

$$\theta(k,\gamma) = \frac{\lambda_k}{\lambda_1} = \left[\frac{1-\rho^k}{1-\rho}\right]^2 \left[\frac{\Omega_{ee}}{\Omega_{\epsilon\epsilon}(k)}\right] > 1.$$

There obviously will be no power advantages to long horizon regression tests if $\{e_t\}$ is *iid* and if there is no persistence in the regressor ($\rho = 0$) since in this case, $\Omega_{\epsilon\epsilon}(k)^{-1}\Omega_{ee} = 1/k$. When the regressor is persistent, $\beta_k(T)/\beta_1(T) = \sum_{j=0}^{k-1}\rho^j = \left(1-\rho^k\right)/(1-\rho)$ is increasing in k and approaches k as ρ approaches unity. Under the null, however, when e_t is either uncorrelated or is positively serially correlated, $\Omega_{\epsilon\epsilon}(k)$ also increases with k at a rate at least as great. Thus, it seems that power advantages will generally be unavailable to long-horizon regression tests when the regressor is exogenous.

Endogenous regressor Local asymptotic power advantages can accrue to long-horizon regression tests when the regressor is endogenous. To see why endogeneity of this sort may arise in applications, we consult the equity return example. When future equity returns $r_{t+1} = \ln(P_{t+1} + D_t) - \ln P_t$ are regressed on $x_t = \ln D_{t-1} - \ln P_t$, both r_{t+1} and

x_{t+1} depend on $\ln P_{t+1}$. It would not be surprising therefore, to find that the regression error and the innovation to x_t are negatively correlated, $E(u_{t+1}e_{t+1}) < 0$.[4]

The endogeneity can also be seen from with a parametric example with a slight reformulation of the dependent variable. Using an approximation from Campbell et al. (1997), let y_t be the log stock price, z_t be the log dividend and $x_t = z_t - y_t$ be the log dividend yield. Then $r_{t+1} \simeq \Phi\Delta y_{t+1} + (1-\Phi)\,x_t$ where Φ is the implied discount factor when the discount rate is the average dividend yield. Suppose that the bivariate sequence $\{(y_t, z_t)'\}$ can be represented as a first-order VECM with cointegration vector $(-1, 1)$,

$$\begin{pmatrix} \Delta y_t \\ \Delta z_t \end{pmatrix} = \begin{pmatrix} h_1 x_{t-1} \\ h_2 x_{t-1} \end{pmatrix} + \begin{pmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{pmatrix} \begin{pmatrix} \Delta y_{t-1} \\ \Delta z_{t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_t \\ v_t \end{pmatrix}, \tag{5}$$

where the equilibrium error is $x_t = z_t - y_t$. Eq.(5) has an equivalent restricted vector autoregressive (VAR) representation for $(\Delta y_t, x_t)$ where $\{x_t\}$ and $\{\Delta y_t\}$ are correlated both contemporaneously and dynamically (at leads and lags).[5] The first equation of the VAR representation gives the short-horizon regression

$$r_{t+1} = (1 + \Phi\left((h_1 + w_{12} - 1)\right) x_t + e_{t+1}, \tag{6}$$

with slope coefficient is $(1 + \Phi\left((h_1 + w_{12} - 1)\right)$ and regression error

$$e_{t+1} = \Phi\left[(w_{11} + w_{12})\Delta y_t - w_{12} x_{t-1} + \epsilon_{t+1}\right],$$

which is serially correlated and also correlated with x_t. The objective of the short-horizon regression is not to estimate $(1 + \Phi\left((h_1 + w_{12} - 1)\right)$ per se, but to estimate the projection coefficient of Δy_{t+1} on x_t which includes the correlation between the regressor x_t and $(\Delta y_t, x_{t-1})$ in the error term. A researcher presumably would use the predictive regression instead of estimating a complete specification of the dynamic correlation structure between Δy_{t+1} and x_t for the same reason that (s)he would use a HAC

[4] The predicted negative innovation correlation are in fact present in the data. Fitting a first-order vector autoregression to $(e_t, v_t)'$, we obtain an innovation correlation of -0.948 for stocks and -0.786 for exchange rates.

[5] The VAR representation is

$$\begin{aligned} \begin{pmatrix} \Delta y_t \\ x_t \end{pmatrix} &= \begin{pmatrix} (w_{11} + w_{12}) & (h_1 + w_{12}) \\ (w_{22} - w_{12} + w_{21} - w_{11}) & (1 + h_2 - h_1 + w_{22} - w_{12}) \end{pmatrix} \begin{pmatrix} \Delta y_{t-1} \\ x_{t-1} \end{pmatrix} \\ &+ \begin{pmatrix} 0 & -w_{12} \\ 0 & (w_{12} - w_{22}) \end{pmatrix} \begin{pmatrix} \Delta y_{t-2} \\ x_{t-2} \end{pmatrix} + \begin{pmatrix} \epsilon_t \\ v_t - \epsilon_t \end{pmatrix}. \end{aligned}$$

covariance estimator to avoid explicitly modeling the serial correlation and conditional heteroskedasticity of the regression error.

Next, we use a simple parametric example to show how the long-horizon test can have substantial local asymptotic power advantages. From proposition 1, long-horizon power advantages require $\left(b\left(1-\rho^k\right)(1-\rho)^{-1}+c_k\right)^2$ to increase with k at a faster rate than $\Omega_{\epsilon\epsilon}(k)$. The power advantage will be substantial if $c<0$ such that $b+c\simeq 0$ and $b_k+c_k>0$ is increasing in k for $k>1$. We won't use the VECM example because it presents a *fixed* alternative instead of a sequence of local alternatives. Also, the VECM becomes quite cumbersome due to its heavy parameterization. Instead, we consider an environment where

$$\begin{aligned} e_t &= a_{11}e_{t-1}+a_{12}(T)\,u_{t-1}+m_t, \\ u_t &= n_t, \end{aligned}$$

where $(m_t,n_t)' \overset{iid}{\sim} [0,\phi(T)]\,, \phi_{mm}=\phi_{nn}=1, \phi_{mn}(T)=\phi_{mn}/\sqrt{T}$, $b(T)=b/\sqrt{T}, a_{12}(T)= a_{12}/\sqrt{T}$. Then local-to-zero endogeneity is characterized by

$$c(T)=\frac{E(e_{t+1}x_t)}{E(x_t)^2}=\left(\frac{a_{12}+a_{11}\phi_{mn}}{\sqrt{T}\,(1-a_{11}\rho)}\right)\left(1-\rho^2\right),$$

with $c=\left[(a_{12}+a_{11}\phi_{mn})\,(1-a_{11}\rho)^{-1}\right](1-\rho^2)$, and $c_k=c\left(1-a_{11}^k\right)(1-a_{11})^{-1}$.

Figure 1 plots $[(b_k+c_k)/(b+c)]^2$ for various values of the DGP's parameter vector γ. As can be seen from the figure, $\left[(b_k+c_k)\,/\,(b+c)\right]^2$ increases at a rate much greater than k which results in local asymptotic power advantages for the long-horizon regression test over a substantial portion of the parameter space.

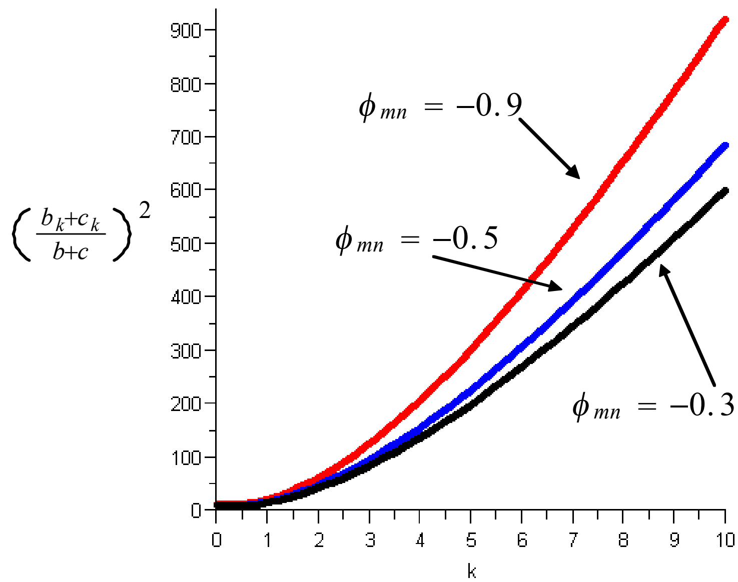

Figure 1. Plots of $\left[(b_k + c_k) / (b + c)\right]^2$ for parameter values $(a_{11}, a_{12}, b, \rho) = (0.1, -0.6, 0.1, 0.95)$. $\phi_{mn} = -0.9, -0.5, -0.3$ gives $c = -0.074, -0.070, -0.068$.

3.2 Local asymptotic power under a weak unit-root regressor

Our results for covariance stationary regressors can be applied to cases of persistent regressors that are weakly integrated in the sense of Park (2003a, 2006). These are series where $\rho(T) = 1 + \rho/T^{\alpha}, \rho < 0$ and $\alpha < 1$ ($\{x_t\}$ is local-to-unity if $\alpha = 1$). The essential difference between the local-to-unity and weakly integrated processes lies in their limit distributions. The weak unit-root process is normally distributed with convergence rate T^{α} while, the limiting distribution of $\{x_t\}$ with $\alpha = 1$ is a function of Ornstein-Uhlenbeck processes. Hence as long as $\alpha < 1$, Proposition 1 continues to apply.

3.3 Local asymptotic power under a local-to-unity regressor

For our local asymptotic analysis under a local-to-unity regressor, the observations will be generated according to

Assumption 2 *(Local-to-unity autoregressive root.) For sample size* T, *the observations have the representation,*

$$r_{t+1} = \mu_r + \beta_1(T)x_t + e_{t+1}, \tag{7}$$

$$x_{t+1} = \mu_x + \rho(T)x_t + u_{t+1}, \tag{8}$$

where $\{e_{t+1}\}$ *and* $\{u_{t+1}\}$ *are zero mean covariance stationary sequences.* $\rho(T) = 1+\alpha/T$ *and* $\beta_1(T) = b/T$ *give the sequence of local alternatives where* α *and* b *are constants. For the long-horizon regression, the sequence of local alternatives at horizon* k *is* $\beta_k(T) = (kb)/T$.

We will also require the following notation. 'Tildes' will refer to demeaned variables so that $\widetilde{x}_t = x_t - T^{-1}\sum_{t=1}^{T} x_t$. Let $\xi_t = (\Delta x_t', e_t')'$ and $\Omega = \Sigma + \Lambda + \Lambda'$ be it's long-run covariance matrix,

$$\Omega = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T}\sum_{l=-\infty}^{\infty} E(\xi_t\xi_{t-l}') = \begin{pmatrix} \Omega_{xx} & \Omega_{xe} \\ \Omega_{ex} & \Omega_{ee} \end{pmatrix},$$

where $\Sigma = \lim_{T\longrightarrow\infty} \sum_{t=1}^{T} E(\xi_t\xi_t') = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xe} \\ \Sigma_{ex} & \Sigma_{ee} \end{pmatrix}$, and $\Lambda = \lim_{T\longrightarrow\infty} \frac{1}{T}\sum_{t=1}^{T}\sum_{l=1}^{t-1} E(\xi_{t-l}\xi_t') = \begin{pmatrix} \Lambda_{xx} & \Lambda_{xe} \\ \Lambda_{ex} & \Lambda_{ee} \end{pmatrix}$. Next, let B_1 be a scaler Brownian motion with long run variance Ω_{xx}, J be the diffusion process defined by $dJ(r) = \alpha J(r) + dB_1(r)$, with initial condition $J(0) = 0$, and $\widetilde{J} = J(r) - \int_0^1 J(r)dr$. The slope coefficient from the $k-$horizon regression is $\hat{\beta}_k = \left(\sum_t \widetilde{x}_t \Delta\widetilde{y}_{t+k}\right)\left(\sum \widetilde{x}_t^2\right)^{-1}$ with asymptotic t-ratio $t_\beta(k) = \hat{\beta}_k/\sqrt{V(\hat{\beta}_k)}$, where $V(\hat{\beta}_k) = \widehat{\Omega}_{ee}\left(\sum_t \tilde{x}_t^2\right)^{-1}$. Then following Phillips (1988) and Cavanagh, Elliot and Stock (1995), we have

Proposition 2 *(Local-to-unity asymptotic distribution) Under Assumption 2, the OLS estimator of the* $k-th$ *horizon regression slope coefficient is asymptotically distributed as,*

$$T\hat{\beta}_k \Longrightarrow kG + k\frac{C_k}{\Omega_{xx}}\left(\int \tilde{J}^2\right)^{-1} + kb, \tag{9}$$

and its corresponding $t-$*statistic has asymptotic distribution,*

$$t_\beta(k) \Longrightarrow \delta\tau_{1c} + \left(1-\delta^2\right)^{1/2} N(0,1) + \frac{C_k}{\sqrt{\Omega_{xx}\Omega_{ee}}}\left(\int \tilde{J}^2\right)^{-1/2} + bR\left(\int \tilde{J}^2\right)^{1/2}, \tag{10}$$

where $G = R\left\{\delta\left(\int \tilde{J}^2\right)^{-1}\int \tilde{J}dB_1 + (1-\delta^2)^{1/2}\left(\int \tilde{J}^2\right)^{-1}\int \tilde{J}dB_2^*\right\}$, $R = \Omega_{xx}^{1/2}\Omega_{ee}^{-1/2}$, $\delta = \Omega_{xe}\left(\Omega_{xx}\Omega_{ee}\right)^{-1/2}$, $C_k = \Lambda_{xe}-\Lambda_{xe,k-1}$, $\Lambda_{xe,k-1} = \lim_{T\to\infty}\frac{1}{T}\frac{1}{k}\sum_{s=2}^{k}\sum_{t=k-1}^{T}\sum_{l=1}^{s-1}E(\Delta x_{t-l}e_t)$ *for* $k>1$, $\Lambda_{xe,0}=0$, $\tau_{1c} = \left(\int \tilde{J}^2\right)^{-1/2}\int \tilde{J}dB_1$,$B_2 = \delta B_1 + (1-\delta^2)^{1/2}B_2^*$, *and* B_2^* *is a standard Brownian motion distributed independently of* B_1.

3.3.1 Contrasting asymptotic analysis under stationarity and local-to-unity.

Whereas in the stationary case the endogeneity is not random, it is useful to note here that both the local-to-unity parameter α, the local *endogeneity parameter* c_k are random. As a result, under the local-to-unity assumption, it is not necessary to explicitly model the local to zero endogeneity in Assumption 2. To understand the randomness in c_k in the local-to-unity case, denote the endogeneity between x_t and e_{t+k} as $c_k(T) = c_k/T = \left(\sum_t \tilde{x}_t\tilde{\epsilon}_{t,k}\right)\left(\sum_t \tilde{x}_t^2\right)^{-1}$. Then using Proposition 2, the limit distribution of c_k is

$$T\left(c_k - kc\right) = T\left(\hat{\beta}_k - \beta_k\left(T\right)\right) = T\left(\hat{\beta}_k - kb\right) \Rightarrow kG + k\frac{\Lambda_{xe}-\Lambda_{xe,k-1}}{\Omega_{xx}}\left(\int \tilde{J}^2\right)^{-1}.$$

As can be seen, the limit distribution of c_k is random even under exogeneity ($c=0$), which is quite different from the stationary case.

3.3.2 Exogenous local-to-unity regressor

Campbell and Yogo (2002) and Valkanov (2003) study predictive regressions where the local-to-unity regressor is exogenous. In this case, $\Lambda_{xe} = \Lambda_{xe,k-1} = 0$ gives

$$t_\beta(k) \Longrightarrow \delta\tau_{1c} + \left(1-\delta^2\right)^{1/2}N(0,1) + bR\left(\int \tilde{J}^2\right)^{1/2},$$

which does not depend on k. Thus it follows that the long-horizon regression test has no asymptotic power advantages over the short horizon test when the regressor is exogenous.

3.3.3 Endogenous local-to-unity regressor

By (10), the limiting behavior of the difference between $t-$statistics at horizons k and horizon 1 is

$$t_\beta^a(k) - t_\beta^a(1) \Rightarrow -\left(\Lambda_{xe,k-1}/\sqrt{\Omega_{xx}\Omega_{ee}}\right)\left(\int \tilde{J}^2\right)^{-1/2}.$$

For a one-tail test with $\beta_k > 0$ under the alternative, $t^a_\beta(k)$ will be increasing in k if innovations to the regressor and the regression error are negatively correlated in the sense that $\Lambda_{xe,k-1} < 0$. Thus, we can say that under Assumption 2, asymptotic power advantages will accrue to long-horizon regression tests if

$$\Lambda_{xe,k-1} < 0 \quad \text{for} \quad k > 1. \tag{11}$$

A significant point to emphasize about (11) is the long-run nature of the endogeneity that gives rise to long-horizon power advantages. In other words, long-horizon power is obtained not just from correlation between Δx_{t-1} and e_t but from cross correlation between any lagged Δx_{t-l} and e_t.

3.3.4 Practical considerations for hypothesis testing

While we've seen that asymptotic power advantages can accrue to long-horizon regressions, the conventional t-test should not be used in practice. Under a local-to-unity regressor, the t-statistic depends on the regressor's local-to-unity parameter α, which cannot be consistently estimated from the time-series. For practical considerations, we approach testing using a variant of the sup-bound test discussed by Cavanaugh et al. (1995), which is an asymptotically valid test of predictability that does not depend on the nuisance parameter α. We use the squared t-ratio which allows two-sided tests. We refer to this as the sup–t^2 test.

To construct the test for given $\delta = \Omega_{xe}\left(\Omega_{xx}\Omega_{ee}\right)^{-1/2}$, let $q_{t_\beta,\alpha,\eta}$ be the 100η percentile of the distribution of $\delta^2\tau^2_{1c} + \left(1-\delta^2\right)N(0,1)^2$. Under the null, $t^2_\beta(k) \xrightarrow{L} \delta^2\tau^2_{1c} + \left(1-\delta^2\right)N(0,1)^2$. It follows that the most conservative sup–t^2 test with at most asymptotic level η is obtained by rejecting the null if $t^2_\beta(k) > q_{t_\beta,0,\eta}$. On the other hand, if $\alpha = \underline{\alpha} \ll 0$, then $t^2_\beta(k) \xrightarrow{L} N(0,1)^2 = \chi^2_1$ from which it follows that the most liberal test is obtained by rejecting the null if $t^2_\beta(k) > q_{t_\beta,\underline{\alpha},\eta}$, which is also equivalent to the conventional asymptotic chi-square test under stationarity of the regressor. Figure 2 displays the asymptotic critical values for the most conservative and the most liberal sup–t^2 tests.

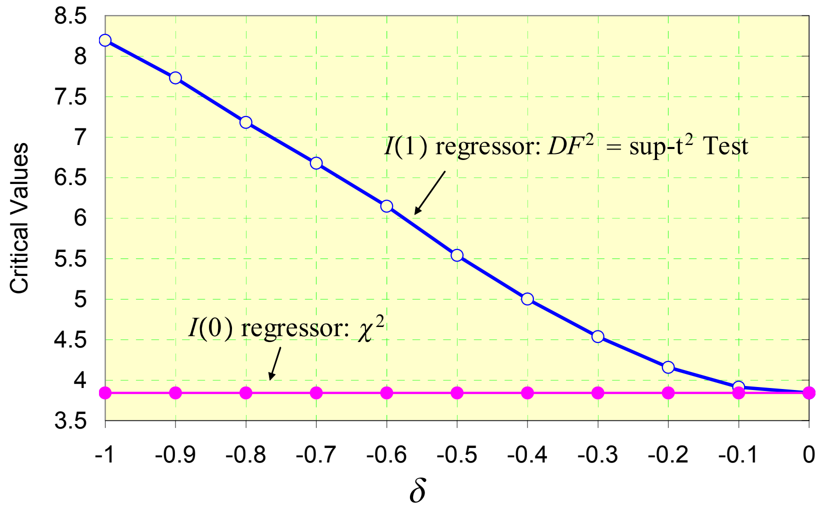

Figure 2. Critical values of asymptotic sup–t^2 test.

The direct application of the sup–t^2 test in small samples, however, may lead to misleading inference. The problem arises due to the well-known small-sample OLS bias in $\widehat{\beta}_k$ from the predictive regression.[6] This bias causes the asymptotic sup–t^2 test to be somewhat oversized at long horizons. Because direct size adjustments to the test statistic is not straightforward when the DGP is unknown, we discuss the RMB Bartlett correction for the test which is a nonparametric and robust resampling adjustment strategy.

4 Recursive moving block Bartlett correction for the sup t^2 test

In unit-root tests, Johansson (2004) and Nielson (1997) show that the Bartlett correction is very accurate for the likelihood ratio test while Larsson (1998) finds that the Bartlett correction of the t^2 test is highly accurate. The successful application of the Bartlett correction in the unit root context suggests that it will provide accurate adjustments in the local-to-unity context that we consider. Following Cribari et al. (1995), who obtained asymptotic expansions of test statistics for stationary processes and Park (2003b) who obtains asymptotic expansions of test statistics for unit-root processes, we obtain the

[6]See Stambaugh (1999) for the short-horizon OLS bias formula, Mark and Sul (2004) for long-horizon OLS bias formulae, and Kim and Nelson (1993) who estimate the bias for stock returns using randomization methods .

asymptotic expansion of the squared t-ratio[7]

$$W = W_T - \frac{a_1}{T} W_T - \frac{a_2}{T} W_T^2 + O_p\left(T^{-2}\right), \tag{12}$$

where W_T is the squared t-statistic computed from a sample of size T, W is its 'true' value, and a_1 and a_2 are 'Bartlett coefficients' which are derived from the asymptotic expansion of the statistic.

In a conventional Bartlett correction, one uses knowledge of the asymptotic DGP to calculate the Bartlett coefficients in the correction. In contrast to the conventional Bartlett correction, here we propose that one estimates the Bartlett coefficients a_1 and a_2 from the data. Since our strategy does not require precise knowledge of the underlying DGP, it is a robust procedure.

To apply the RMB Bartlett correction, proceed as follows. Construct a moving-block sample of size B from the original set of observations, $\{\xi_1, \cdots, \xi_B\}$, $\left\{\xi_2, \cdots, \xi_{B+1}\right\}$, $\cdots$, $\left\{\xi_{T_1-B+1}, \cdots, \xi_{T_1}\right\}$. Using the data from each block, construct the sup-t^2 statistic, $W_{B,j} = t_\beta^2$, $j = 1, ..., T_1 - B + 1$. From each block, form the analog to (12),

$$BW = -a_1 W_{B,j} - a_2 W_{B,j}^2 + BW_{B,j}.$$

Taking the average over j gives

$$BW = -a_1 E_B^*\left(W\right) - a_2 E_B^*\left(W^2\right) + BE_B^*\left(W\right),$$

where $E_B^*\left(W\right) = \frac{1}{T_1-B+1}\sum_{j=1}^{T_1-k+1} W_{B,j}$ and $E_B^*\left(W^2\right) = \frac{1}{T_1-B+1}\sum_{j=1}^{T_1-k+1} W_{B,j}^2$. Repeat using block size $B+1$, then block size $B+2$, and so on through block size $B+(T_1-B) = T_1$. For $t = B, B+1, ..., T_1$ we have

$$tW = -a_1 E_t^*\left(W\right) - a_2 E_t^*\left(W^2\right) + tE_t^*\left(W\right).$$

Now let $z_t = tE_t^*\left(W\right)$ and run the regression

$$z_t = a_1 E_t^*\left(W\right) + a_2 E_t^*\left(W^2\right) + Wt + \text{error}.$$

The estimated coefficient on the trend is the RMB Bartlett-corrected test statistic.

[7] Bartlett (1937) originally proposed this adjustment strategy to the log-likelihood ratio statistic to achieve a test with better size.

5 Simulation results

This section presents simulation results for assessing the performance of the RMB Bartlett corrected sup–t^2 test in small samples. All of the simulation work includes a constant in estimation.

The first set of results that we discuss are simulations that confirm the prediction that long-horizon asymptotic power advantages are present in small samples under regressor endogeneity. The DGP is as in Assumption 1 where

$$\begin{aligned} e_t &= a_{11}e_{t-1} + a_{12}u_{t-1} + m_t, \\ u_t &= n_t, \end{aligned}$$

and $(m_t, n_t)' \overset{iid}{\sim} [0, \phi], \phi_{mm} = \phi_{nn} = 1, -1 < \phi_{mn} < 0$. This DGP exhibits endogeneity of the regressor with a fixed alternative hypothesis. The endogeneity factor in the short-horizon regression is $c(T) = E\left(\sum x_t e_{t+1}\right)\left(\sum x_t^2\right)^{-1} = O(T^{-1})$ and for the long-horizon regression is $c_k(T) = O(T^{-1})$.[8] We generate 5000 samples of $T = 100$.

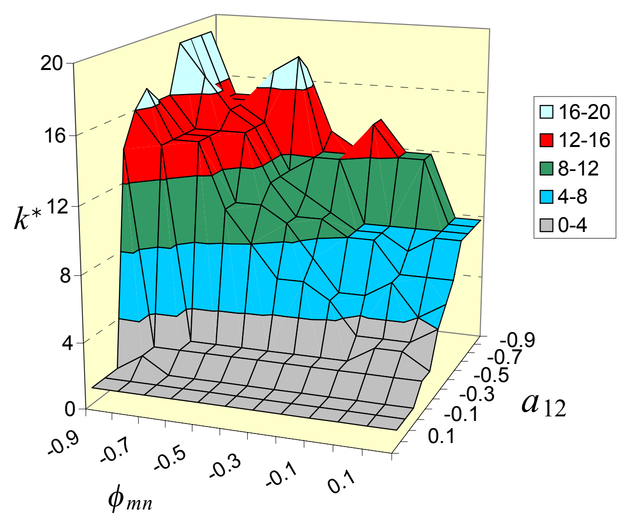

Figure 3. Horizons that maximize relative size-adjusted power between long- and short-horizon tests for $T = 100$.

[8] For this DGP, the endogeneity factor is $c(T) = (a_{12} + a_{11}\phi_{mn})\left(1 - (\rho(T))^2\right)(1 - a_{11}\rho(T))^{-1} = (a_{12} + a_{11}\phi_{mn})\frac{\alpha^2 - 2T\alpha}{T^2 a_{11} - T^2 - T\alpha a_{11}}$. Under the null hypothesis ($b = 0$), we set $a_{12} = a_{11} = 0$ but allow variations in ϕ_{mn}.

Figure 3 shows the horizon k^* that maximize the relative *size-adjusted power* of the conventional t-test. This is obtained by searching $k \in [1, 20]$, $\phi_{mn} \in [-0.9, 0.1]$, $a_{12} \in [-0.9, 0.1]$ with $(a_{11}, b, \alpha) = (0.1, 10, -5)$. $k^* = 1$ in cases where the long-horizon regression test does not have local power advantages. As can be seen from the figure, the size-adjusted power of long-horizon regression tests consistently dominate those of short-horizon tests in this region of the parameter space.[9]

Although we obtained these values of k^* using the conventional t-test, the results are not sensitive to this at all. In extensive simulation work, the k^* that maximizes the size-adjusted power of the sup$-t^2$ test, as well as the size-adjusted power of this test is very similar to that found for the conventional $t-$test. (These results are available upon request from the authors but are not reported in the paper to economize on space.)

Next, we examine the small sample performance of the sup–t^2 test. Table 2 displays simulations to examine the effective size of the asymptotic sup–t^2 test and the RMB Bartlett corrected test for $T = 100$. It can be seen that the asymptotic test is oversized at $k = 10, 15, 20$ whereas the RMB Bartlett-corrected test is reasonably sized at those horizons and is somewhat undersized for $k = 1, 5$. For $T = 100$, the Bartlett correction is seen to give tests that are better sized than the asymptotic test.

[9] Asymptotic standard errors computed by Andrews's (1991) method.

Table 2: $T = 100$, effective size of asymptotic and RMB-Bartlett corrected sup$-t^2$ test. $(a_{11}, a_{12}, \alpha) = (0, 0, -5)$

	Asymptotic					RMB-Bartlett corrected				
ϕ_{mn}	$k=1$	$k=5$	$k=10$	$k=15$	$k=20$	$k=1$	$k=5$	$k=10$	$k=15$	$k=20$
A. Nominal 5 % test										
-0.9	0.034	0.056	0.112	0.171	0.218	0.022	0.027	0.058	0.084	0.102
-0.7	0.035	0.049	0.114	0.158	0.207	0.026	0.023	0.064	0.085	0.104
-0.5	0.042	0.048	0.120	0.165	0.215	0.032	0.027	0.063	0.085	0.099
-0.3	0.054	0.062	0.128	0.180	0.217	0.035	0.029	0.072	0.090	0.105
-0.1	0.067	0.072	0.147	0.194	0.230	0.041	0.036	0.072	0.088	0.112
B. Nominal 10 % test										
-0.9	0.065	0.086	0.171	0.232	0.279	0.047	0.046	0.083	0.124	0.149
-0.7	0.061	0.076	0.155	0.199	0.251	0.041	0.041	0.083	0.118	0.129
-0.5	0.064	0.068	0.148	0.199	0.249	0.047	0.037	0.077	0.104	0.122
-0.3	0.067	0.068	0.140	0.192	0.239	0.042	0.035	0.078	0.102	0.114
-0.1	0.067	0.073	0.148	0.195	0.230	0.041	0.037	0.072	0.088	0.113

Next, we consider size and power performance of the asymptotic and RMB Bartlett corrected sup–t^2 test for sample sizes $T = 100, 200$, and 300 and report the results in Table 3. For longer horizons, say $k = 20$, a time-series length of $T = 300$ is required for the asymptotic sup–t^2 test to be correctly sized. The RMB Bartlett-corrected sup-t^2 test is for the most part undersized when $T = 200$ and $T = 300$. Local-to-unity power of the RMB Bartlett-corrected tests rival those of the size-adjusted asymptotic sup–t^2 tests. While the coarse grid of horizons that we report do not, in many cases, pick off the horizon that gives the test its maximal power, results for the horizons that we do report show that long-horizon power advantages hold up. These results indicate that the RMB Bartlett correction to the asymptotic sup–t^2 test should work well in practice.

Table 3: Local-to-Unity Effective Size and Power of asymptotic and RMB Bartlett corrected sup$-t^2$ test. $(a_{11}, a_{12}, \alpha, \phi_{mn}) = (0, 0, -5, -0.9)$ under the null. $\rho(T) = 1 + \alpha/T$, $b(T) = b/T$ with $(a_{11}, a_{12}, \alpha, b, \phi_{mn}) = (0.1, -0.3, -5, 20, -0.9)$ under the alternative.

	Asymptotic					RMB Bartlett corrected				
T	$k=1$	$k=5$	$k=10$	$k=15$	$k=20$	$k=1$	$k=5$	$k=10$	$k=15$	$k=20$
A.	Size of nominal 5% test					Size of nominal 5% test				
100	0.034	0.056	0.112	0.171	0.218	0.022	0.027	0.058	0.084	0.102
200	0.022	0.018	0.047	0.081	0.113	0.019	0.008	0.019	0.037	0.053
300	0.021	0.009	0.030	0.052	0.066	0.018	0.004	0.012	0.016	0.027
B.	Size of nominal 10% test					Size of nominal 10% test				
100	0.065	0.086	0.171	0.232	0.279	0.047	0.046	0.083	0.124	0.149
200	0.063	0.042	0.080	0.125	0.161	0.047	0.018	0.031	0.058	0.078
300	0.061	0.030	0.059	0.085	0.099	0.050	0.012	0.022	0.034	0.045
C.	Power of 5% size-adjusted test					Power of nominal 5% test				
100	0.880	0.883	0.930	0.843	0.649	0.636	0.581	0.896	0.844	0.675
200	0.827	0.808	0.881	0.900	0.928	0.546	0.246	0.576	0.857	0.924
300	0.794	0.804	0.803	0.883	0.937	0.412	0.131	0.330	0.670	0.886
D.	Power of 10% size-adjusted test					Power of nominal 10% test				
100	0.963	0.965	0.978	0.933	0.785	0.819	0.722	0.940	0.896	0.751
200	0.946	0.929	0.956	0.977	0.977	0.772	0.405	0.744	0.929	0.962
300	0.926	0.938	0.934	0.971	0.986	0.772	0.296	0.500	0.831	0.950

6 Predictability of long-horizon equity returns

We return to the empirical example and apply the RMB Bartlett correction to the sup–t^2 tests of whether the log dividend yield predicts future stock returns.

Since potential power advantages of long-horizon regressions hinge on the endogeneity of the regressor, we run a Hausman test to investigate whether this is the case. Lagged values of the dividend yield are evidently weak instruments since using three lags as instruments yields a χ^2_1 statistic value of 2.31 (p-value=0.128). Employing the real interest rate as an instrument yields a test statistic of 109.2 which rejects exogeneity of the dividend yield at any reasonable level. Employing the real interest rate and three lags of the dividend yield as instruments gives a test statistic of 9.69 (p-value=0.002). The weight of the evidence rejects the exogeneity of the dividend yield.

Because of the unusual behavior of stock prices associated with the bull market of the 90s and the subsequent decline in 2001-2002, the slope estimates are sensitive to the sample period. In recognition of this sensitivity, we run the regressions for horizons 1 through 20 initially using 1990 as the end of the sample and then recursively updating the sample through 2002. Since the true value of the local-to-unity parameter $\alpha < 0$ is unknown, the exact critical values for the test will be bounded between the critical values for the t^2 test and the sup–t^2 test.[10] To compare the inferences that one would draw from the most liberal and the most conservative tests, for each sample we conduct four tests of predictability: i) the conventional t^2 test, ii) the RMB corrected t^2 test, iii) the asymptotic sup–t^2 test, and iv) the RMB Bartlett-corrected sup–t^2 test.[11]

$k_{\min}$, shown in Table 4 is the shortest horizon for which the null is rejected at the 5-percent nominal level, whereas k^* is the horizon that gives the largest value of the test statistic. We make several remarks about the table. The first point to note is that the robustness of the RMB Bartlett-corrected test shows up in the stability of the results across the different samples. As observations from the 1990s are added to the sample, the asymptotic sup–t^2 test and the conventional t^2 test requires successively longer horizons to reject the null. When the sample ends in 1991, the asymptotic sup–t^2 test rejects the null with $k = 9$, but when the sample ends in 2002, the shortest horizon for which the test rejects is $k = 16$. In contrast, the $k_{\min}$ from both the RMB Bartlett corrected t^2 test and the corrected sup–t^2 test are comparatively stable.

Secondly, if the regressor is stationary or if it is a persistent and weak unit-root process, the RMB Bartlett corrected t^2 test will be appropriate. Application of this test is seen to consistently reject the null hypothesis at $k = 10$ in every sample with the strongest evidence against the null coming at $k^* = 19$ or 20.

If the regressor follows a local-to-unity process, then the RMB Bartlett-corrected sup–t^2 test is the appropriate choice. This test consistently rejects the null at $k = 13$. For samples ending in 1997 and 1998 it rejects the null at $k = 11$,and the maximal RMB Bartlett-corrected test statistics are obtained at horizon $k^* = 19$ for every sample.

[10] The critical values for the sup $-t^2$ test depend on the estimated value of δ. For the 1992 sample, the 5% critical value is 6.677. For all other samples, it is 7.1822.

[11] Although it is well-known that the asymptotic t^2 test suffers from substantial size distortion, the Bartlett-corrected version of the test is only modestly oversized. The small-sample performance of these tests are reported in the working paper [Mark and Sul (2004)].

Table 4: Stock Return Predictability

T	$k_{\min}$ t^2	$k_{\min}$ BC t^2	k^* BC t^2	$k_{\min}$ sup$-t^2$	$k_{\min}$ BC sup$-t^2$	k^* BC sup$-t^2$
1991	5	10	20	9	13	19
1992	5	10	20	9	13	19
1993	7	10	13	9	13	19
1994	8	10	14	9	13	19
1995	8	10	13	10	13	19
1996	8	10	14	10	13	19
1997	9	10	14	11	11	19
1998	9	10	14	12	11	19
1999	10	10	14	13	13	19
2000	12	10	19	14	13	19
2001	12	10	19	15	13	19
2002	13	10	19	16	13	19

Notes: $k_{\min}$ is the shortest horizon for which the test rejects the null hypothesis. k^* is the horizon that gives the maximal test statistic value. BC denotes RMB Bartlett correction.

7 Conclusion

Whether long-horizon regression tests have power advantages over short-horizon tests has been an open question for some time. This paper addressed this question and showed that long-horizon tests do have local asymptotic power advantages over short-horizon tests under endogeneity of the regressor. Power advantages can be found to accrue both under covariance stationarity of the regressor or under local-to unity. While asymptotic theoretical justification for using long horizons exists, small-sample OLS bias causes size distortion in the asymptotic tests. Because conventional bias adjustment may not be easily handled at long horizons when the DGP is unknown, we suggest resampling strategies to correct for test size distortion.

We examined the recursive moving block Bartlett correction to obtain sup–t^2 tests that are better sized in small samples and found it to be reasonably sized both at short and long horizons. The RMB Bartlett corrected test was also found to effectively maintain small-sample power advantages of long-horizon tests.

Application of the small-sample adjustments to U.S. stock market data finds that the hypothesis that the dividend yield does not predict returns is rejected with 13-year return horizons using the most conservative RMB Bartlett-corrected sup-t^2 test.

References

[1] Alexius, Annika. "Uncovered Interest Parity Revisited," *Review of International Economics*,3, August 2001, 505-517.

[2] Andrews, Donald W.K. 1991. "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation." *Econometrica,* 59,(3), pp. 817–858.

[3] Bartlett, M.S.,1937. "Properties of Sufficiency and Statistical Tests," *Proceedings of the Royal Society* A 160, 268-282.

[4] Berben, Robert-Paul, 2000. "Exchange Rates and Macroeconomic Fundamentals: A Long-Run Perspective." Ph.D. dissertation, University of Nijmegen, the Netherlands.

[5] Berben, Robert-Paul and Dick van Dijk, 1998. "Does the Absence of Cointegration Explain the Typical Findings in Long Horizon Regressions?" *mimeo*, Tinbergen Institute, Erasmus University of Rotterdam.

[6] Berkowitz, Jeremy and Lorenzo Giorgianni, 2001. "Long Horizon Predictability?" *Review of Economics and Statistics*, 83, 81-91.

[7] Bhansali, R.J., 1997. "Direct Autoregressive Predictors for Multistep Prediction: Order Selection and Performance Relative to the Plug In Predictors," *Statistica Sinica,* 7, 425–449.

[8] Campbell, John Y. 2001. "Why Long Horizons: A Study of Power Against Persistent Alternatives,"*Journal of Empirical Finance* 8, pp.459-491.

[9] Campbell, John Y., and John H. Cochrane, 1999. "By force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior," *Journal of Political Economy,* 107, 205–251.

[10] Campbell, John Y., Andrew W. Lo, and A. Craig MacKinlay, 1997. *The Econometrics of Financial Markets*, Princeton University Press.

[11] Campbell, John Y. and Robert J. Shiller, 1988, 'Stock Prices, Earnings, and Expected Dividends,' *Journal of Finance*, July, 43, pp. 661-76.

[12] Campbell, John Y. and Motohiro Yogo, 2002. "Efficient Tests of Stock Return Predictability," mimeo, Harvard University.

[13] Cavanagh, Christopher L, Graham Elliott and James H. Stock, 1995. "Inference in Models with Nearly Integrated Regressors," *Econometric Theory,* 11, 1131-47.

[14] Cecchetti, Stephen G., Pok-sang Lam, and Nelson C. Mark, 2000. "Asset Pricing with Distorted Beliefs: Are Equity Returns Too Good to be True?" *American Economic Review,* 90, 787-805.

[15] Chinn, Menzie D. and Guy Meredith, 2002. "Testing Uncovered Interest Parity at Short and Long Horizons during the Post-Bretton Woods Era," *mimeo* U.C. Santa Cruz.

[16] Chinn, Menzie D. and Richard A. Meese, 1995, 'Banking on Currency Forecasts: How Predictable is Change in Money?' *Journal of International Economics*, 38, 161–178.

[17] Clark, Todd and Michael L. McCracken, 2001. "Evaluating Long-Horizon Forecasts," *mimeo* Federal Reserve Bank of Kansas City.

[18] Coe, Patrick J. and James M. Nason, 2004. "Long-Run Monetary Neutrality and Long-Horizon Regressions," *Journal of Applied Econometrics,* 19, 355-373.

[19] Cribari-Neto and Ferrari, 1995. "Bartlett-Corrected Tests for Heteroskedastic Linear Models," *Economics Letters*, 48, 113-18.

[20] Daniel, Kent, 2001. "The Power and Size of Mean Reversion Tests," *Journal of Empirical Finance,* 8, 493-535.

[21] Fama, Eugene F. and Kenneth R. French, 1988a. 'Dividend Yields and Expected Stock Returns,' *Journal of Financial Economics*, October, 22, pp.3–25.

[22] Fama, Eugene F. and Kenneth R. French, 1988b. "Permanent and Temporary Components of Stock Prices," *Journal of Political Economy*, 96, 246–73.

[23] Goetzmann, William N. and Philippe Jorion, 1993. "Testing the Predictive Power of Dividend Yields," *Journal of Finance*, 48, 663-79.

[24] Hall, Peter, Joel L. Horowitz and Bing-Yi Jing, 1995. "On Blocking Rules for the Bootstrap with Dependent Data," *Biometrika,* 82, 561-74.

[25] Hodrick, Robert J., 1992, Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement, *Review of Financial Studies* 5, 357–86.

[26] Ing, Ching-Kang, 2003, "Multistep Prediction in Autoregressive Processes," *Econometric Theory*, 19, 254–279.

[27] Johansen, Soren, 2004."A Small Sample Correction of the Dickey-Fuller Test." In A. Welfe (ed): *New Directions in Macromodelling*, Chapter 3, 49-68, Elsevier.

[28] Kilian, Lutz, 1999. "Exchange Rates and Monetary Fundamentals: What Do We Learn from Long-Horizon Regressions?" *Journal of Applied Econometrics* 14, 491–510.

[29] Kilian, Lutz and Mark P. Taylor, 2003. "Why Is It So Difficult To Beat The Random Walk Forecast of Exchange Rates?" *Journal of International Economics*, 60, 85-107.

[30] Kim, Myung J. and Charles R. Nelson, 1993. "Predictable Stock Returns: The Role of Small Sample Bias," *Journal of Finance,*48, 641-661.

[31] Kim, Myung J., Charles R. Nelson, and Richard Startz, 1991. "Mean Reversion in Stock Prices? A Reappraisal of the Empirical Evidence," *The Review of Economic Studies,* 58, 505-528.

[32] Larsson, Rolf, 1998. "Bartlett Corrections fro Unit Root Test Statistics," *Journal of Time Series Analysis*, 19, 425–438.

[33] Mankiw, N. Gregory and Matthew Shapiro, 1986. "Do We Reject Too Often?: Small Sample Properties of Tests of Rational Expectations," *Economics Letters* 20, 139-45.

[34] Mark, Nelson C. (1995). 'Exchange Rates and Fundamentals: Evidence on Long-Horizon Prediction,' *American Economic Review*, 85, 201–218.

[35] Mark, Nelson C. and Doo-Yul Choi, 1997. "Real Exchange-Rate Prediction over Long Horizons," *Journal of International Economics*, 43,(1/2) 29–60. Reprinted in L.Sarno and M.P. Taylor eds., *New Developments in Exchange Rate Economics*, Edward Elgar, Cheltenham, UK (2001).

[36] Mark, Nelson C and Donggyu Sul, 2004. "The Use of Predictive Regressions at Alternative Horizons in Finance and Economics," NBER Technical Working paper No. 0298

[37] Mishkin, Frederic S. 1992. "Is the Fisher Effect for Real?: A Reexamination of the Relationship between Inflation and Interest Rates," *Journal of Monetary Economics*, 30, 195–216.

[38] Nielsen, B., 1997. "Bartlett Correction of the Unit Root Test in Autoregressive Models," *Biometrika*, 84, 500-504.

[39] Park, Joon, 2003a. "Weak Unit Roots," *mimeo,* Rice University.

[40] Park, Joon, 2003b. "Bootstrap Unit Root Tests," *Econometrica,* 71, 1845–1895.

[41] Park, Joon, 2006. "A Bootstrap Theory for Weakly Integrated Processes," *Journal of Econometrics*, 133, 639–672.

[42] Phillips, P.C.B., 1988. "Regression Theory for Near Integrated Time Series," *Econometrica*, 56, 1021-1043.

[43] Rapach, David and Mark Wohar, 2002. "Testing the Monetary Model of Exchange Rate Determination: New Evidence from A Century of Data," *Journal of International Economics*, 58, 359–385.

[44] Richardson, Matthew and James Stock, 1989. "Drawing Inferences from Statistics Based on Multiyear Asset Returns," *Journal of Financial Economics*, 25, 343-48.

[45] Rothenberg, Thomas, 1984. "Approximate Normality of Generalized Least Squares Estimates," *Econometrica*, 52, 911-825.

[46] Quenouille, M.H., 1956. "Notes on Bias in Estimation," *Biometrika,* 43, 353-360.

[47] Shiller, Robert J. *Irrational Exuberance*, Princeton University Press, Princeton New Jersey 2000.

[48] Stambaugh, Robert F. 1999. "Predictive Regressions," *Journal of Financial Economics*, 54, 375–421.

[49] Valkanov, Rossen, 2003."Long-Horizon Regressions: Theoretical Results and Applications," *Journal of Financial Economics,* 68, 201-232.

Appendix

This appendix provides proofs and details for the main results in the text.

A. Asymptotic Power Advantage when x_t is $I(0)$.

Let $\epsilon(k) = (\epsilon_{1,k}, ..., \epsilon_{T,k})$ and $x = (x_1, ..., x_T)$. The proof of Proposition 1 is aided by the following lemmas.

Lemma 1 *Under Assumption 1, the regression error is asymptotically orthogonal to the regressor,*

$$\underset{T\to\infty}{plim} T^{-1} x' \hat{\epsilon}(k) = 0.$$

Proof. $x'\epsilon(k) = O_p\left(T^{1/2}\right)$, which gives $T^{-1}x'\epsilon(k) = o_p(1)$. Since $\beta_k - \hat{\beta}_k \to^p 0$, it follows that $\underset{T\to\infty}{\text{plim}} T^{-1}x'\hat{\epsilon}(k) = \underset{T\to\infty}{\text{plim}} T^{-1}x'\left(\epsilon(k) + x\left(\beta_k - \hat{\beta}_k\right)\right) = 0$. ■

Lemma 2 *Under Assumption 1, the long run variance of $\hat{\epsilon}_{t,k}$ is*

$$\underset{T\to\infty}{plim} T^{-1}\hat{\epsilon}(k)\hat{\epsilon}'(k) = \Omega_{\epsilon\epsilon}(k) = E\left(T^{-1}\epsilon(k)\epsilon'(k)\right).$$

Proof. First note that $\beta_k - \hat{\beta}_k \to^p 0$ and $T^{-1}x'x$ is $O_p(1)$. Then from Lemma 1, we have

$$\begin{aligned}
\underset{T\to\infty}{\text{plim}} T^{-1}\hat{\epsilon}(k)\hat{\epsilon}'(k) &= \underset{T\to\infty}{\text{plim}} T^{-1}\left(\epsilon(k) + x'\left(\beta_k - \hat{\beta}_k\right)\right)\left(\epsilon(k) + x'\left(\beta_k - \hat{\beta}_k\right)\right)' \\
&= \underset{T\to\infty}{\text{plim}} T^{-1}\epsilon(k)\epsilon'(k) + \underset{T\to\infty}{\text{plim}}\left[\left(\beta_k - \hat{\beta}_k\right)^2\left(T^{-1}x'x\right)\right] \\
&= \Omega_{\epsilon\epsilon}(k).
\end{aligned}$$

■

Lemma 3 *Under Assumption 1, the ratio of the long run variance of the kth and 1st horizon regression coefficients is*

$$\underset{T\to\infty}{plim}\left[\frac{V\left(\hat{\beta}_1\right)}{V\left(\hat{\beta}_k\right)}\right] = \frac{\Omega_{ee}}{\Omega_{\epsilon\epsilon}(k)}.$$

Proof. From Lemma 1 and 2, we have

$$\begin{aligned}
\operatorname*{plim}_{T\to\infty}\left[\frac{V\left(\hat{\beta}_1\right)}{V\left(\hat{\beta}_k\right)}\right] &= \operatorname*{plim}_{T\to\infty}\left[\frac{(x'x/T)^{-1}\left(x'\hat{e}\hat{e}'x/T\right)(x'x/T)^{-1}}{(x'x/T)^{-1}\left(x'\hat{\epsilon}\left(k\right)\hat{\epsilon}'\left(k\right)x/T\right)(x'x/T)^{-1}}\right] \\
&= \operatorname*{plim}_{T\to\infty}\left[\frac{x'\hat{e}\hat{e}'x/T}{x'\hat{\epsilon}\left(k\right)\hat{\epsilon}'\left(k\right)x/T}\right] \\
&= \operatorname*{plim}_{T\to\infty}\left[\frac{x'ee'x/T+\left(\beta_1-\hat{\beta}_1\right)^2 x'x/T}{x'\epsilon\left(k\right)\epsilon'\left(k\right)x/T+\left(\beta_k-\hat{\beta}_k\right)^2 x'x/T}\right] \\
&= \operatorname*{plim}_{T\to\infty}\left[\frac{x'ee'x/T}{x'\epsilon\left(k\right)\epsilon'\left(k\right)x/T}\right]=\frac{\Omega_{ee}}{\Omega_{\epsilon\epsilon}\left(k\right)}
\end{aligned}$$

It holds because by Assumption 1 we get $\beta_k-\hat{\beta}_k \to^p 0$ and $T^{-1}x'e = T^{-1}x'\epsilon\left(k\right) = o_p\left(1\right)$.

■

Proof of Proposition 1 Proposition 1 follows directly from Lemmas 1,2 and 3.

B. Asymptotic Power Advantage when x_t is local-to-unity

Before we proceed the formal proof of Proposition 2, here we provide an intuitive explanation by using a simple example. Suppose that x_t is $I\left(1\right)$ and is correlated with e_{t+1}. Then we may split the regression error e_{t+1} into two components: $e_{t+1} = e^*_{t+1} + e^0_{t+1}$ where $\mathrm{E}\left(x_t e^*_{t+1}\right) \neq 0$ and $\mathrm{E}\left(x_t e^0_{t+1}\right) = 0$. Then we have

$$\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T-k}\mathrm{E}\left(x_t e^*_{t+1}\right) = \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T-k}\sum_{l=1}^{t-1}E\left(\Delta x_{t-l}e^*_{t+1}\right) = \Lambda_{xe}. \tag{A.1}$$

Note that correlation between e^0_{t+1} and u_{t+1} exists. From Hansen (1995), we have

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]}\begin{pmatrix} u_t \\ e^0_t \end{pmatrix} \Longrightarrow \begin{pmatrix} \Omega_{xx}B_1\left(r\right) \\ \Omega_{ee}\left\{\delta B_1\left(r\right)+\left(1-\delta^2\right)^{1/2}B^*_2\left(r\right)\right\} \end{pmatrix}, \tag{A.2}$$

where B_1 and B^*_2 are independent standard Brownian motions and $\Longrightarrow$ denotes weak convergence with respect to the uniform metric and $\delta = \Omega_{xe}\left(\Omega_{xx}\Omega_{ee}\right)^{-1/2}$.

Combining (A.1) with (A.2) provides the limit distribution of $T^{-1}\sum_{t=1}^{T-k}x_t e_{t+1}$,

$$T^{-1}\sum_{t=1}^{T-k}x_t e_{t+1} \Longrightarrow R\delta\int B_1 dB_1 + R\sqrt{1-\delta^2}B_1 dB^*_2 + \Lambda_{xe}, \tag{A.3}$$

where $R = \Omega_{xx}^{1/2}\Omega_{ee}^{-1/2}$. Under the general setting, it is not necessary to split the regression error term into two components. Replacing B by $\widetilde{J}$ in (A.3) yields the following lemmas.

Lemma 4 Under Assumption 2,

1. $T^{-2}\sum_{t=1}^{T-k}\tilde{x}_t^2 \Longrightarrow \Omega_{xx}\int \widetilde{J}^2 dr$
2. $T^{-1}\sum_{t=1}^{T-k}\tilde{x}_t\tilde{u}_{t+1} \Longrightarrow \Omega_{xx}^{1/2}\int \widetilde{J}dB_1 + \Lambda_{xx}$
3. $T^{-1}\sum_{t=1}^{T-k}\tilde{x}_t\tilde{e}_{t+1} \Longrightarrow R\delta\int \widetilde{J}dB_1 + R\sqrt{1-\delta^2}\widetilde{J}dB_2^* + \Lambda_{xe}$ where $R = \Omega_{xx}^{-1/2}\Omega_{ee}^{1/2}$, $\delta = \Omega_{xe}\left(\Omega_{xx}\Omega_{ee}\right)^{-1/2}$, and B_2 is the standard Brownian motion distributed independently of B_1.
4. $T^{-1}\sum_{t=1}^{T-k}\tilde{x}_t\tilde{e}_{t+k} \Longrightarrow R\delta\int \widetilde{J}dB_1 + R\sqrt{1-\delta^2}\widetilde{J}dB_2^* + \Lambda_{xe} - \Lambda_{xe,k-1}^*$ where $\Lambda_{xe,k-1}^* = \lim_{T\to\infty}\frac{1}{T}\sum_{t=k+1}^{T}\sum_{l=1}^{k-1}\mathrm{E}(\Delta x_{t-l}e_t)$.

Proof of Lemma 4: The proof of Part 1 and 2 are provided by Phillips (1988). Part 3 can be directly obtained from combining the results of Hansen (1995) and part (e) in lemma 3.1 in Phillips (1988). Note that if $\mathrm{E}(e_t u_{t-l}) \neq 0$ for $l \geq 0$ but $\mathrm{E}(x_t e_{t+k}) = 0$ for $k \geq 1$, then Part 3 becomes

$$T^{-1}\sum_{t=1}^{T-k}\tilde{x}_t\tilde{e}_{t+1} \Longrightarrow R\delta\int \widetilde{J}dB_1 + R\sqrt{1-\delta^2}\widetilde{J}dB_2$$

Further note that if $\mathrm{E}(e_t u_{t-l}) = 0$ for $l > 0$, $\mathrm{E}(e_t u_t) = \Sigma_{xe}$, and $\mathrm{E}(x_t e_{t+k}) = 0$ for $k \geq 1$, then Part 3 becomes

$$T^{-1}\sum_{t=1}^{T-k}\tilde{x}_t\tilde{e}_{t+1} \Longrightarrow R^*\delta^*\int \widetilde{J}dB_1 + R^*\sqrt{1-(\delta^*)^2}dB_2$$

where $R^* = \Sigma_{xx}^{-1/2}\Sigma_{ee}^{1/2}$, $\delta = \Sigma_{xe}\left(\Sigma_{xx}\Sigma_{ee}\right)^{-1/2}$.

Part 4: Observe that

$$\frac{1}{T}\sum_{t=1}^{T-k}x_t e_{t+k} = \frac{1}{T}\sum_{t=k-1}^{T}x_{t-k}e_t = \frac{1}{T}\sum_{t=k-1}^{T}x_{t-1}e_t - \frac{1}{T}\sum_{t=k-1}^{T}\sum_{l=1}^{k-1}\Delta x_{t-l}e_t + o_p(1), \quad \text{(A.4)}$$

where the last term is given by $-\frac{\alpha}{T^2}\sum_{l=1}^{k}x_{t-l}e_t$ which becomes $o_p(1)$. Hence the limit distribution in (A.4) can be directly obtained from Part 3 and by letting $\Lambda_{xe,k-1}^* = \lim_{T\to\infty}\frac{1}{T}\sum_{t=k+1}^{T}\sum_{l=1}^{k-1}\mathrm{E}(\Delta x_{t-l}e_t)$.

Proof of Proposition 2 Note that

$$\begin{aligned} T^{-1}\sum_{t=1}^{T-k} x_t \epsilon_{t+k,k} &= T^{-1}\sum_{t=1}^{T-k} x_t e_{t+1} + ... + T^{-1}\sum_{t=1}^{T-k} x_t e_{t+k} \\ &\quad + \frac{b_1}{T}\left(T^{-1}\sum_{t=1}^{T-k} x_t e_{t+1} + ... + T^{-1}\sum_{t=1}^{T-k} x_t e_{t+k}\right) \\ &= T^{-1}\sum_{t=1}^{T-k} x_t e_{t+1} + ... + T^{-1}\sum_{t=1}^{T-k} x_t e_{t+k} + o_p(1) \end{aligned}$$

The last term can be rewritten as

$$T^{-1}\sum_{t=1}^{T-k} x_t e_{t+1} + ... + T^{-1}\sum_{t=1}^{T-k} x_t e_{t+k} = \frac{k}{T}\sum_{t=k-1}^{T} x_{t-1} e_t - \frac{1}{T}\sum_{s=2}^{k}\sum_{t=k-1}^{T}\sum_{l=1}^{s-1} \Delta x_{t-l} e_t$$

For notational convenience, denote

$$\Lambda_{xe,k-1} = \lim_{T\to\infty} \frac{1}{T}\frac{1}{k}\sum_{s=2}^{k}\sum_{t=k-1}^{T}\sum_{l=1}^{s-1} \mathrm{E}\left(\Delta x_{t-l} e_t\right) = \frac{1}{k}\sum_{s=1}^{k} \Lambda^*_{xe,k-1}.$$

Note that $\Lambda^*_{xe,0} = 0$. Then directly from Lemma 4, we have

$$T^{-1}\sum_{t=1}^{T-k} \tilde{x}_t \tilde{\epsilon}_{t+k,k} \implies k\left(R\delta\int \widetilde{J} dB_1 + R\sqrt{1-\delta^2}\widetilde{J} dB_2 + \Lambda_{xe}\right) - k\Lambda_{xe,k-1}.$$

The OLS estimator of the slope coefficient for the kth horizon regression,

$$\hat{\beta}_k = \frac{\sum_{t=1}^{T-k} \tilde{x}_t \tilde{\epsilon}_{t+k,k}}{\sum_{t=1}^{T-k} \tilde{x}_t^2}.$$

By Assumption 2, we have

$$T\hat{\beta}_k = kb + \frac{T^{-1}\sum_{t=1}^{T-k} \tilde{x}_t \tilde{\epsilon}_{t+k,k}}{T^{-2}\sum_{t=1}^{T-k} \tilde{x}_t^2}.$$

From Lemma 4, it follows that

$$\begin{aligned} T\hat{\beta}_k &\implies kR\left\{\delta\int \widetilde{J} dB_1 + \left(1-\delta^2\right)^{1/2}\int \widetilde{J} dB_2\right\}\left(\int \widetilde{J}^2\right)^{-1} \\ &\quad + k\frac{\Lambda_{xe} - \Lambda_{xe,k-1}}{\Omega_{xx}}\left(\int \widetilde{J}^2\right)^{-1} + kb. \end{aligned}$$

Define $t_{\beta}(k) = \hat{\beta}_k / \sqrt{V(\hat{\beta}_k)}$, and $V(\hat{\beta}_k) = \hat{\Omega}_{\epsilon\epsilon}(k) \left[\sum x_t^2\right]^{-1}$. Since $\Omega_{\epsilon\epsilon}(1) = \Omega_{ee},\ t_{\beta}(k)$ can be rewritten as

$$t_{\beta}(k) = \frac{\hat{\beta}_k}{k\sqrt{\widehat{\Omega}_{ee}}} \left(\sum_{t=1}^{T-k} \tilde{x}_t^2 \right)^{1/2}.$$

Hence it is straightforward to show that

$$t_{\beta}(k) \Longrightarrow \delta\tau_{1c} + \left(1-\delta^2\right)^{1/2} N(0,1) + \left(\frac{\Lambda_{xe} - \Lambda_{xe,k-1}}{\sqrt{\Omega_{xx}\Omega_{ee}}} \right) \left(\int \tilde{J}^2 \right)^{-1/2} + bR \left(\int \tilde{J}^2 \right)^{1/2},$$

where $\tau_{1c} = \left(\int \tilde{J}^2 \right)^{-1/2} \int \widetilde{J} dB_1$.

C. Derivation of formulae in example 1

Let $a_{11}(L) \equiv (1 - a_{11}L)^{-1} = \sum_{j=0}^{\infty} a_{11}^j L^j$ and $\rho(L) \equiv (1-\rho L)^{-1} = \sum_{j=0}^{\infty} \rho^j L^j$, where L is the lag operator. Then rewrite the innovations process as,

$$\begin{aligned} e_{t+1} &= a_{12}(T) a_{11}(L) u_t + a_{11}(L) m_{t+1}, \\ x_t &= \rho(L) u_t. \end{aligned}$$

It follows that

$$\begin{aligned} E(e_{t+1}x_t) &= E\left(\left[a_{12}(T)a_{11}(L)u_t + a_{11}(L)m_{t+1}\right]\left[\rho(L)u_t\right] \right) \\ &= \frac{a_{12}(T) + a_{11}\phi_{mn}(T)}{1 - a_{11}\rho} = \frac{a_{12} + a_{11}\phi_{mn}}{(1-a_{11}\rho)\sqrt{T}}. \end{aligned}$$

Thus, we have determined that $\beta_1(T) \xrightarrow{p} b(T) + c(T)$ where

$$c(T) = \frac{E(e_{t+1}x_t)}{E(x_t)^2} = \frac{(a_{12} + a_{11}\phi_{mn})}{(1 - a_{11}\rho)} \frac{1}{\sqrt{T}} \left(1 - \rho^2\right),$$

Now, for $k = 2$,

$$\begin{aligned} r_{t+2} + r_{t+1} &= b_2(T)(1+\rho)x_t + \epsilon_{t,2}, \\ \epsilon_{t,2} &= e_{t+2} + e_{t+1} + \beta_1(T)u_{t+1}. \end{aligned}$$

Therefore, $b_2(T) = b(T)(1+\rho) = \frac{b}{\sqrt{T}}(1+\rho)$. As above, we can write

$$\begin{aligned} e_{t+2} &= a_{12}(T)u_{t+1} + a_{12}(T)a_{11}a_{11}(L)u_t + m_{t+2} + a_{11}m_{t+1} + a_{11}^2 a_{11}(L)m_t, \\ x_t &= \rho(L)u_t. \end{aligned}$$

from which we obtain,

$$E(e_{t+2}x_t) = \frac{a_{12}\,(T)\,a_{11}}{1-a_{11}\rho} + \frac{\phi_{mn}\,(T)\,{a_{11}}^2}{1-a_{11}\rho} = \frac{a_{11}\,(a_{12}+\phi_{mn}\,a_{11})}{(1-a_{11}\rho)\,\sqrt{T}} = a_{11}E(e_{t+1}x_t).$$

It follows that

$$E[\epsilon_{t,2}x_t] = E[(e_{t+2}+e_{t+1})x_t] = (1+a_{11})\left(\frac{a_{12}+\phi_{mn}a_{11}}{(1-a_{11}\rho)\sqrt{T}}\right) = c_2(T).$$

Continuing on in this way, it can be seen that for any k, $b_k(T) = b(T)\left(\sum_{j=0}^{k-1}\rho^j\right) = b(T)\left(\frac{1-\rho^k}{1-\rho}\right)$, and

$$E\left(\epsilon_{t,k}x_t\right) = \left(\frac{a_{12}+\phi_{mn}a_{11}}{(1-a_{11}\rho)\sqrt{T}}\right)\left(\sum_{j=0}^{k-1}a_{11}^j\right) = \left(\frac{a_{12}+\phi_{mn}a_{11}}{(1-a_{11}\rho)\sqrt{T}}\right)\left(\frac{1-a_{11}^k}{1-a_{11}}\right).$$

Finally, divide by $E\left(x_t^2\right) = \left(1-\rho^2\right)^{-1}$ to get

$$c_k(T) = \left(\frac{a_{12}+\phi_{mn}a_{11}}{(1-a_{11}\rho)\sqrt{T}}\right)\left(\frac{1-a_{11}^k}{1-a_{11}}\right)\left(1-\rho^2\right).$$