THE USE OF PREDICTIVE REGRESSIONS
AT ALTERNATIVE HORIZONS IN FINANCE AND ECONOMICS

Nelson C. Mark
Donggyu Sul

Technical Working Paper **298**

TECHNICAL WORKING PAPER SERIES

THE USE OF PREDICTIVE REGRESSIONS
AT ALTERNATIVE HORIZONS IN FINANCE AND ECONOMICS

Nelson C. Mark
Donggyu Sul

Technical Working Paper 298
http://www.nber.org/papers/T0298

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2004

The Use of Predictive Regressions at Alternative Horizons in Finance and Economics
Nelson C. Mark and Donggyu Sul
NBER Technical Working Paper No. 298
July 2004
JEL No. C12, C22, G12, E47

ABSTRACT

When a k period future return is regressed on a current variable such as the log dividend yield, the marginal significance level of the t-test that the return is unpredictable typically increases over some range of future return horizons, k. Local asymptotic power analysis shows that the power of the long-horizon predictive regression test dominates that of the short-horizon test over a nontrivial region of the admissible parameter space. In practice, small sample OLS bias, which differs under the null and the alternative, can distort the size and reduce the power gains of long-horizon tests. To overcome these problems, we suggest a moving block recursive Jackknife estimator of the predictive regression slope coefficient and test statistics that is appropriate under both the null and the alternative. The methods are applied to testing whether future stock returns are predictable. Consistent evidence in favor of return predictability shows up at the 5 year horizon.

Nelson C. Mark
Department of Economics and Econometrics
University of Notre Dame
Notre Dame, IN 46556
and NBER
nmark@nd.edu

Donggyu Sul
Department of Economics
University of Auckland
dsul013@comc.auckland.ac.nz

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 let $x_t \sim I(0)$ be a time t 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 economics r_t might be the return on the log exchange rate and x_t the deviation of the exchange rate from a set of macroeconomic fundamentals.[1] A test of the predictability of the return can be conducted by regressing the one-period ahead return r_{t+1} on x_t and performing a t-test on the slope coefficient. Empirical research in finance and economics 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,

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

and conducting a t-test of the null hypothesis that the return is unpredictable $H_0 : \beta_k = 0$, where the t-statistic is constructed with a heteroskedastic and autocorrelation consistent (HAC) standard error. Frequently, it is found that the OLS slope estimates, asymptotic t-ratios, and R^2s are increasing over a range of horizons $k > 1$. Because the marginal significance level of the test of no predictability declines over this range of k, the test using the long horizon regression may reject the null hypothesis whereas the test using the short horizon regression may not. Because the long-horizon regression is built by adding up the intervening short-horizon regressions, these results present a puzzle and the underlying basis for them are not fully understood. As stated by Campbell, Lo, and MacKinlay (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."

This paper addresses the power question posed by Campbell et. al. There are two aspects to our study. The first concerns the asymptotic properties of the tests. To determine whether there exists an asymptotic theoretical motivation for using long-horizon regression, we compare the local asymptotic power of long- and short-horizon regression tests in two ways. Our first examination assumes that the regressor has a local-to-unity dominant autoregressive root. This approach is motivated by the high persistence of the predictive variables used in empirical applications. But because this also has the unattractive implication that the predictor is asymptotically unit root nonstationary, it is useful also to evaluate local asymptotic power when the observations are covariance stationary. Thus, our second examination of local asymptotic power is conducted under this alternative scenario.

[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, 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) regress long-horizon exchange rate returns on long-term bond yield differentials.

We find that the local asymptotic power of long-horizon regression tests dominate that of the short-horizon test over a nontrivial subset of the admissible parameter space. When the regressor is econometrically exogenous with local-to-unity autoregressive coefficient, there are no local power advantages to long-horizon regression tests. When the regressor is covariance stationary, local asymptotic power advantages accrue to long-horizon regression tests when the regression error or the regressor (or both) exhibit negative serial correlation. This noteworthy result may not be particularly useful in guiding empirical work both because the required correlation structure is not a common feature of the data used empirical applications of long-horizon regressions and because strict exogeneity is an unrealistic assumption for many applications.

Asymptotic results that speak directly to extant empirical work are obtained when the regressor is endogenous. Endogeneity arises not in the sense of a misspecification of the structural model because the predictive regressions we study are employed simply as a projection of the $k-$period future return r_{t+k} on x_t to estimate functions of the underlying moments of the distribution between $\{r_t\}$ and $\{x_t\}$. The motivation for using predictive regressions when the regressor is endogenous is analogous to the motivation for HAC estimation of covariance matrices in exchange for modeling the exact autocorrelation and conditional heteroskedasticity dynamics of the regression error. Here, whether the regressor has a local-to-unity autoregressive root or is covariance stationary, asymptotic power advantages accrue to long-horizon regression tests in empirically relevant regions of the parameter space—where $\{x_t\}$ is positively autocorrelated and persistent, the short-horizon regression error exhibits low to moderate serial correlation, and the innovations to the regressor and the regression error are contemporaneously correlated.

The second aspect of our study concerns small sample performance of long-horizon regression tests. While theoretical power comparisons are valid asymptotically for local alternative hypotheses, power gains of the long-horizon regression test can be attenuated in practice on account of small sample bias. We characterize the bias, the power loss attributable to the bias and the associated size distortion of the tests. The small sample bias under the null hypothesis is different than it is under the alternative so that a simple bias adjustment under the null does not result in a proper sized test. To obtain a test with better size, we suggest a moving block recursive jackknife strategy to reduce bias in the slope coefficient estimator and to obtain a test with the correct size. The jackknife strategy is appropriate both under the null and under the alternative, has good size properties and restores the power advantages of long-horizon regression tests in sample sizes likely to be available to applied researchers.

We illustrate the use of these methods in an examination of the predictability of returns on the Standard and Poors index. Using annual time series that begin in 1871 and recursively updating the sample from 1971 to 2002 gives stable recursive jackknifed test results that consistently reject the hypothesis of no predictability at return horizons of 10 years or more.

Studies of the econometrics of predictive regressions include Campbell (2001) who studied an environment where the regressor $\{x_t\}$ follows an AR(1) process and where the short-horizon regression error is serially uncorrelated. 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. 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 Monte Carlo analysis. Mankiw and Shapiro (1986), Hodrick (1992), Kim and Nelson (1993), and Goetzmann and Jorion (1993), Mark (1995), and Kilian (1999) study small-sample inference issues. Stambaugh (1999) proposes a Bayesian analysis to deal with small sample bias and Campbell and Yogo (2002) study point optimal tests in the short-horizon predictive regression.. Kilian and Taylor (2002) examine finite sample properties under nonlinearity of the data generation process and Clark and McCracken (2001) study the predictive power of long-horizon out-of-sample forecasts.

The long-horizon 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 with $k/T \to 0$ 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 $k/T \to \delta \in (0,1)$ and show that the test statistics converge to functions of Brownian motions. Daniel (2001) studies optimal tests of this kind. Valkanov (2003) employs the Richardson and Stock asymptotic distribution theory to the long-horizon regressions of the type that we study when the regressor $x_t \sim I(1)$.

The remainder of the paper is as follows. The next section sets the stage for our inquiry by presenting estimation results from two canonical examples of the use of predictive regression in finance in economics. Section 3 presents the local asymptotic power analysis and the small sample properties of the predictive regression tests are discussed in Section 4 discusses its small sample properties. Section 5 presents the moving block recursive jackknife method to adjust the OLS bias and to correct the size distortion of the tests. The jackknife methods are applied to study stock return predictability in Section 7 and Section 8 concludes. Proofs of propositions are contained in the appendix.

2 Canonical empirical examples

We illustrate the issues with two canonical empirical examples. The first example uses the log dividend yield as a predictor of future stock returns [Fama and French (1988b), Campbell and Shiller (1988)] using annual observations from 1871 to 1995.[2] 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.

Letting $r_{t+1} = \ln\left(\left(P_{t+1} + D_{t+1}\right)/P_t\right)$ and $x_t = \ln\left(D_t/P_t\right)$, where P and D are the price and dividend on the Standard and Poors index, we run the equity return regressions at horizons of 1, 2, 4, and 8 years. HAC standard errors are computed using the automatic lag selection method of Newey and West (1994). The results, displayed in panel A of Table 1, show that the evidence for return predictability appears to strengthen as the return horizon is lengthened. OLS slope coefficient point estimates, HAC asymptotic t-ratios, and regression R^2s all increase with return horizon.[3]

Our second example uses the deviation of the exchange rate from its monetary fundamentals value as a predictor of future exchange rate returns [Mark (1995) and Chinn and Meese (1995)]. Letting $r_{t+1} = \ln(S_{t+1}/S_t)$ be the exchange rate return where S is the nominal exchange rate, $x = \ln(F/S)$, $F = (M/M^*)(Y^*/Y)$, $M(M^*)$ and $Y(Y^*)$ are domestic (foreign) money and income respectively, and F is the monetary fundamentals value of the exchange rate. The monetary approach to exchange rate determination, which has a present value interpretation, gives a motivation for the predictive regression for exchange rate returns analogous to that for equity returns. Results are presented for horizons of 1,2,3, and 4 years for a US–UK data set consisting of 100 quarterly observations spanning from 1973.1 to 1997.3.[4] The results, displayed in panel B of Table 1 exhibit a similar pattern for slope coefficient point estimates, HAC asymptotic t-ratios and regression R^2s that increase with the return horizon.

In both examples, the regressor $\{x_t\}$ is highly persistent. Augmented Dickey–Fuller and Phillips–Perron unit root tests statistics for the regressors are shown in Table 2. The unit root can be rejected at the 5 percent level if the entire sample of 1500 monthly log dividend yield observations is used but if one analyzed only the first 288 monthly observations (or 24 years) the unit root would not be rejected. The third column of the table shows that with 24 years of data, a unit root in the deviation of the log exchange rate from the log fundamentals also cannot be rejected at standard significance levels. Failure to reject the null hypothesis does not require us to accept it especially in light of the well known low power in small samples of unit root tests. For the exchange rate predictor, evidence against a unit root is stronger in a long historical record, as found by Rapach and Wohar (2002). In the ensuing analysis, we pay close attention to

[2] These data were used in Robert J. Shiller (2000) and were obtained from his web site. Annual observatinos were constructed from these monthly data. Returns are $r_{t+1} = \ln\left((P_{t+1} + D_t)/P_t\right)$ 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.

[3] Because the dependent variable changes with k, the R^2s are not directly comparable across horizons.

[4] These data are from Mark and Sul (2001). S_t is the end-of-quarter dollar price of the pound, industrial production is used to proxy for income, US money is M2 and UK money is M0 (due to availability).

environments in which $\{x_t\}$ is persistent.

Table 1: Illustrative Long-Horizon Regressions

A. Returns on S&P index				
	Horizon in years			
	1	2	4	8
$\hat{\beta}$	0.131	0.263	0.483	0.833
t-ratio	2.827	3.333	3.993	5.445
R^2	0.151	0.285	0.492	0.701

B. Returns on \$/£ exchange rate				
	Horizon in years			
	1	2	3	4
$\hat{\beta}$	0.201	0.420	0.627	0.729
t-ratio	2.288	3.518	5.706	5.317
R^2	0.172	0.344	0.503	0.606

Table 2: Regressor persistence in the data.

		Monthly Dividend yield T=1500	Monthly Dividend yield T=288	Quarterly Deviation from Fundamentals T=100
ADF	τ_c	-3.58	-2.02	-1.66
	τ_t	-4.29	-2.66	-1.31
PP	τ_c	-3.45	-1.87	-1.78
	τ_t	-4.09	-2.25	-1.63
AC	1	0.986	0.985	0.940
	6	0.883	0.859	0.648
	12	0.732	0.670	0.273
	24	0.544	0.367	0.094
	36	0.474	0.161	-0.170

Notes: τ_c (τ_t) is the studentized coefficient for the unit root test with a constant (trend). ADF is the augmented Dickey–Fuller test and PP is the Phillips–Perron test. Approximate critical values for τ_c for $T = 1500, 288, 100$ 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 τ_t for $T = 1500, 288, 100$ 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. AC is the first-order autocorrelation coefficient.

3 Local asymptotic power

We study local asymptotic power from two perspectives. First, as in Campbell and Yogo (2002) and Valkanov (2003), we assume that the regressor has a local-to-unity autoregressive root to account for a highly persistent regressor. However, for problems such as in predicting stock returns, this formulation implies that in large samples the dividend yield and therefore the equity return is $I(1)$. To avoid the unattractive assumption that that equity returns are unit-root nonstationary, in subsection 3.2 we conduct an alternative local-to-zero asymptotic analysis where both the return sequence $\{y_t\}$ and the dividend-yield $\{x_t\}$ are covariance stationary.

For notational convenience, we work with predictive regressions of the form

$$\Delta y_{t+1} = \beta_1 x_t + e_{t+1}. \tag{2}$$

We suppress the regression constant since its inclusion has no effect on the asymptotic properties of the predictive regression tests. We reintroduce the constant below in our analysis of the small sample properties of the tests.[5]

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-to-unity asymptotic power

For our local-to-unity asymptotic analysis, the observations are generated according to

Assumption 1 *(Local-to-unity autoregressive root.) For sample size T, the observations have the representation*

$$\Delta y_{t+1} = \beta_1(T) x_t + e_{t+1} \tag{3}$$

$$x_{t+1} = \rho(T) x_t + u_{t+1} \tag{4}$$

where $\{e_{t+1}\}$ and $\{u_{t+1}\}$ are zero mean covariance stationary sequences. $\rho(T) = 1+c_1/T$ and $\beta_1(T) = b_1/T$ give the sequence of local alternatives where c_1 and b_1 are constants. For the long-horizon regression, the sequence of local alternatives at horizon k is $\beta_k(T) = (kb_1)/T$.

[5] This reformulation of the dependent variable maps exactly into the returns formulation for exchange rates ($y_t = \ln(S_t)$) and is an approximate representation of stock returns. The approximation follows from Campbell et. al. (1997), by letting y_t be the log stock price, x_t 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.

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_c^* be the diffusion process defined by $dJ_c^*(r) = c_1 J_c^*(r) + dB_1(r)$, with initial condition $J_c^*(0) = 0$, and $J_c = J_c^*(r) - \int_0^1 J_c^*(r)dr$. The slope coefficient from the $k-$horizon regression is $\hat{\beta}_k = \left(\sum_t x_t \Delta y_{t+k}\right)\left(\sum x_t^2\right)^{-1}$ with asymptotic t-ratio $t_\beta(k) = \hat{\beta}_k / \sqrt{V(\hat{\beta}_k)}$, where $V(\hat{\beta}_k) = \widehat{\Omega}_{ee}\left(\sum_t x_t^2\right)^{-1}$. Then following Phillips (1988), Cavanagh, Elliot and Stock (1995) and Berben and Van Dijk (1998), we have

Proposition 1 *(Local-to-unity asymptotic distribution) Under Assumption 1, the OLS estimator of the kth horizon regression slope coefficient is asymptotically distributed as,*

$$\begin{aligned} T\left(\hat{\beta}_k - \beta\right) \implies & kR\left\{\delta\left(\int J_c^2\right)^{-1} \int J_c dB_1 + \left(1-\delta^2\right)^{1/2} \left(\int J_c^2\right)^{-1} \int J_c dB_2^*\right\} \\ & + \frac{\Lambda_{xe} - \Lambda_{xe,k-1}}{\Omega_{xx}} \left(\int J_c^2\right)^{-1} + kb_1. \end{aligned} \tag{5}$$

Its corresponding $t-$statistic has asymptotic distribution,

$$t_\beta(k) \Longrightarrow \delta\tau_{1c} + \left(1-\delta^2\right)^{1/2} N(0,1) + \left(\frac{\Lambda_{xe} - \Lambda_{xe,k-1} + b_1}{\sqrt{\Omega_{xx}\Omega_{ee}}}\right)\theta_c \tag{6}$$

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

When the regressor is strictly exogenous [Campbell and Yogo (2002), Valkanov (2003)], then $\Lambda_{xe} = \Lambda_{xe,k-1} = 0$. This gives $t_\beta(k) \Longrightarrow \delta\tau_{1c} + \left(1-\delta^2\right)^{1/2} N(0,1) + \left(b_1/\sqrt{\Omega_{xx}\Omega_{ee}}\right)\theta_c$ which does not depend on k and

Corollary 1 *(Local-to-unity and exogeneity) If the regressor is econometrically exogenous then under Assumption 1, the long-horizon regression test has no asymptotic power advantage over the short-horizon regression test.*

Strict exogeneity, however, is unlikely to hold in many empirical applications. 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, $E(u_{t+1}e_{t+1}) < 0$. Analogously, when the exchange rate return $r_{t+1} = \ln S_{t+1} - \ln S_t$ is regressed on the deviation of the log fundamentals from the exchange rate, $x_{t+1} = \ln F_{t+1} - \ln S_{t+1}$ the dependence of both r_{t+1} and x_{t+1} on $\ln S_{t+1}$ suggests that the innovation to x_{t+1} and the short-horizon regression error will be negatively correlated.[6]

To make the point more formally, suppose that the bivariate sequence $\{(y_t, z_t)'\}$ can be represented as a first-order VECM with cointegration vector $(-1, 1)$,

$$\begin{pmatrix} \Delta y_t \\ \Delta z_t \end{pmatrix} = \begin{pmatrix} h_1 x_{t-1} \\ h_2 x_{t-1} \end{pmatrix} + \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} \Delta y_{t-1} \\ \Delta z_{t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_t \\ u_t \end{pmatrix}, \tag{7}$$

where the equilibrium error is $x_t = z_t - y_t$.[7] Eq.(7) has the equivalent restricted vector autoregressive (VAR) representation for $(\Delta y_t, x_t)$,

$$\begin{aligned} \begin{pmatrix} \Delta y_t \\ x_t \end{pmatrix} &= \begin{pmatrix} (a_{11} + a_{12}) & (h_1 + a_{12}) \\ (a_{22} - a_{12} + a_{21} - a_{11}) & (1 + h_2 - h_1 + a_{22} - a_{12}) \end{pmatrix} \begin{pmatrix} \Delta y_{t-1} \\ x_{t-1} \end{pmatrix} \\ &\quad + \begin{pmatrix} 0 & -a_{12} \\ 0 & (a_{12} - a_{22}) \end{pmatrix} \begin{pmatrix} \Delta y_{t-2} \\ x_{t-2} \end{pmatrix} + \begin{pmatrix} \epsilon_t \\ u_t - \epsilon_t \end{pmatrix}. \end{aligned} \tag{8}$$

By inspection of (8), $\{x_t\}$ and $\{\Delta y_t\}$ are seen to be correlated both contemporaneously and dynamically (at leads and lags). Writing out the first equation of (8) and advancing the time index gives the short-horizon regression

$$\Delta y_{t+1} = (h_1 + a_{12})x_t + \left[(a_{11} + a_{12})\Delta y_t - a_{12}x_{t-1} + \epsilon_{t+1}\right], \tag{9}$$

where slope coefficient is $h_1 + a_{12}$ and the regression error $(a_{11} + a_{12})\Delta y_t - a_{12}x_{t-1} + \epsilon_{t+1}$, is serially correlated and also correlated with x_t. The objective of the short-horizon regression is not to estimate this $h_1 + a_{12}$ 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 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 covariance estimator to avoid explicitly modeling the serial correlation and conditional heteroskedasticity of the regression error.

By (6), the limiting behavior of the difference between $t-$statistics at the horizons k and horizon 1 is $t^a_\beta(k) - t^a_\beta(1) \Longrightarrow -\left(\Lambda_{xe,k-1}/\sqrt{\Omega_{xx}\Omega_{ee}}\right)\theta_c$. For a one-tail test with $\beta_k > 0$ under the alternative, this difference will be positive if innovations to the regressor and

[6] 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.

[7] In exchange rate analysis, Δy_{t+1} is the exchange rate return and z_t is the log fundamentals. Equity returns and dividend yields do not have an exact VECM representation.

the regression error are negatively correlated in the sense that $\Lambda_{xe,k-1} < 0$. Thus, when the regressor is endogenous, we have

Corollary 2 *(Local-to-unity with endogenous regressor) Under Assumption 1, asymptotic power advantages accrue to long-horizon regression tests if $\Lambda_{xe,k-1} < 0$ for $k > 1$.*

The local-to-unity assumption for the autoregressive root implies that the predictor is asymptotically unit-root nonstationary. This is not an appropriate characterization for time-series such as the dividend yield (or earnings-to-price ratios). Therefore, it is useful to have an alternative statement about local power of long- and short-horizon regressions in a stationary environment.

3.2 Local asymptotic power under covariance stationarity

For the local asymptotic analysis under covariance stationarity, the observations are generated according to

Assumption 2 *(Covariance stationarity) For sample size T, the observations have the representation*

$$\Delta y_{t+1} = \beta_1(T)x_t + e_{t+1} \tag{10}$$

$$x_{t+1} = \rho x_t + u_{t+1} \tag{11}$$

where $\{e_{t+1}\}$ and $\{u_{t+1}\}$ are zero mean covariance stationary sequences, $-1 < \rho < 1$, and $\beta_1(T) = b_1/\sqrt{T}$ where b_1 is a constant. Local-to-zero endogeneity obeys $c_1(T) = c_1/\sqrt{T} = E\left(\sum_{t=1}^{T} x_t e_{t+1}\right)\left(\sum_{t=1}^{T} x_t^2\right)^{-1}$.

The long-horizon regression ($k > 1$) obtained by addition of short-horizon regressions is

$$y_{t+k} - y_t = \beta_k(T)x_t + \epsilon_{t,k}$$

where

$$\begin{aligned}
\beta_k(T) &= \beta_1(T)\left[1 + \sum_{j=1}^{k-1}\rho^j\right] = \frac{b_1}{\sqrt{T}}\left(\frac{1-\rho^k}{1-\rho}\right) = \frac{b_k}{\sqrt{T}} \\
\epsilon_{t,k} &= \sum_{j=1}^{k} e_{t+j} + \beta_1(T)\left(\sum_{j=1}^{k-1} u_{t+j}\right). \qquad (12)
\end{aligned}$$

Under the sequence of local alternatives, the OLS estimator at horizon $k > 1$ has probability limit $(b_k + c_k)/\sqrt{T}$, where $c_k/\sqrt{T} = E\left(\sum_{t=1}^{T} x_t \epsilon_{t,k}\right)\left(\sum_{t=1}^{T} x_t^2\right)^{-1}$. The direct dependence of $\epsilon_{t,k}$ on the projection errors u_{t+j} vanish asymptotically so 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

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

with noncentrality parameter

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

Local asymptotic power depends on the DGP's parameter values. We denote this dependence by writing the parameter vector that characterizes the DGP as γ and measure local asymptotic power between long-and short-horizon regression tests by $\theta(k, \gamma) = \lambda_k/\lambda_1$. We can now state

Proposition 2 *Under Assumption 2, the long-horizon regression* ($k > 1$) *test of the hypothesis that* x_t *does not predict future changes in* y_t *has asymptotic local power advantage over the short-horizon regression* ($k = 1$) *test if*

$$\theta(k,\gamma) = \frac{\lambda_k}{\lambda_1} = \underset{T\to\infty}{plim}\left[\frac{\hat{\beta}_k}{\hat{\beta}_1}\right]^2\left[\frac{V(\hat{\beta}_1)}{V(\hat{\beta}_k)}\right] = \left[\frac{b_k + c_k}{b_1 + c_1}\right]^2\left[\frac{\Omega_{ee}}{\Omega_{\epsilon\epsilon}(k)}\right] > 1$$

where Ω_{ee} *and* $\Omega_{\epsilon\epsilon}(k)$ *are the long run variances of* e_{t+1} *in (10) and* $\epsilon_{t,k}$ *in (12), respectively.*

If the regressor is exogenous, then $c_k = 0$ for all $k > 0$. For direct comparison with Corollary 1, we have,

Corollary 3 *(Exogenous regressor) If the regressor is exogenous, then under Assumption 2, the long-horizon regression has asymptotic local power advantages over the short-horizon regression if*

$$\left[\frac{1-\rho^k}{1-\rho}\right]^2\left[\frac{\Omega_{ee}}{\Omega_{\epsilon\epsilon}(k)}\right] > 1.$$

There obviously will be no power advantages to long horizon regression tests if $\{e_t\}$ is *iid* and $\rho = 0$ since in this case, $E\left(\epsilon_{t+k}\epsilon'_{t+k}/T\right) = E\left(\sum_{j=1}^k e_{t+j}\right)\left(\sum_{j=1}^k e_{t+j}\right)'/T + o(1)$ which gives $\Omega_{\epsilon\epsilon}(k)\Omega_{ee}^{-1} = k$.[8] With a persistent regressor, $\beta_k(T)/\beta_1(T) = \sum_{j=0}^{k-1}\rho^j$ $= \left(1-\rho^k\right)/\left(1-\rho\right)$, which is increasing in k and approaches k as ρ approaches unity.

[8] See Mark and Sul (2002) who compute the local-asymptotic power advantages for several parametric specifications under the exogeneity assumption.

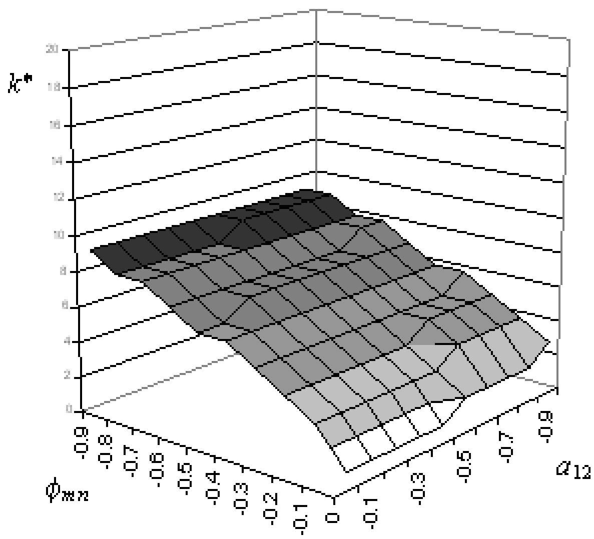

Figure 1: $k^* = \arg\max\left\{\theta\left(k,\gamma\right)\right\}, \; \gamma = (a_{11}, b_{1,}\rho) = (0.1, 0.1, 0.95)$.

Long-horizon regression tests will have local-to-zero asymptotic power advantages if $\Omega_{\epsilon\epsilon}(k)$ increases at a rate less than k. This can happen if the regression error is negatively serially correlated. Large negative serial correlation of the regression error is not a common feature of either stock return or foreign exchange return data and the exogeneity assumption is untenable.

When the regressor is endogenous, local power advantages can accrue to long-horizon regression if $b_1\left(1-\rho^k\right)/\left(1-\rho\right)+c_k$ is increasing at a faster rate than $\Omega_{\epsilon\epsilon}(k)$. If in addition, $c_1 < 0$ such that $b_1 + c_1 \simeq 0$ while $b_k + c_k > 0$ and is increasing in k for $k > 1$ the power advantage can be substantial. While it is difficult to construct simple parametric examples where the short-run and long-run correlations have different signs, such behavior can be observed in the data.[9] The following example lays out a parametric structure in which local-asymptotic power advantages are attained by long-horizon regression tests.

Example 1. *The observations are generated by*

$$\begin{aligned}
\Delta y_{t+1} &= \beta_1(T)x_t + e_{t+1}, \\
x_{t+1} &= \rho x_t + u_{t+1}, \\
\begin{pmatrix} e_t \\ u_t \end{pmatrix} &= \begin{pmatrix} a_{11} & a_{12}(T) \\ 0 & 0 \end{pmatrix}\begin{pmatrix} e_{t-1} \\ u_{t-1} \end{pmatrix} + \begin{pmatrix} m_t \\ n_t \end{pmatrix},
\end{aligned}$$

[9]In a related context, Lo (1991) finds that stock returns are positively serially correlated at daily horizons but become negatively serially correlated at annual frequencies.

where $(m_t, n_t)' \overset{iid}{\sim} [0, \Sigma(T)]\,, \Sigma(T) = \begin{pmatrix} 1 & \phi_{mn}(T) \\ \phi_{mn}(T) & 1 \end{pmatrix}$, $b_1(T) = b_1/\sqrt{T}$, $a_{12}(T) = a_{12}/\sqrt{T}$, and $\phi_{mn}(T) = \phi_{mn}/\sqrt{T}$. *Then*

$$c_1(T) = \frac{E(e_{t+1}x_t)}{E(x_t)^2} = \left(\frac{a_{12} + a_{11}\phi_{mn}}{\sqrt{T}\,(1 - a_{11}\rho)}\right)\left(1 - \rho^2\right),$$

with $c_1 = \left[\left(a_{12} + a_{11\phi_{mn}}\right)(1 - a_{11}\rho)^{-1}\right](1 - \rho^2)$, and $c_k = c_1\left(1 - a_{11}^k\right)(1 - a_{11})^{-1}$. *The unconditional mean of the long-horizon regression slope coefficient is*

$$E\left(\hat{\beta}_k\right) = \frac{b_1}{\sqrt{T}}\left(\frac{1 - \rho^k}{1 - \rho}\right) + \frac{c_1}{\sqrt{T}}\left(\frac{1 - a_{11}^k}{1 - a_{11}}\right)$$

The parameter vector for this DGP is $\gamma = (b_1, \rho, a_{11}, \phi_{mn}, a_{12})$ *and our measure of relative asymptotic power is*

$$\theta(k, \gamma) = \left(\frac{b_1 \sum_{j=0}^{k-1} \rho^j + c_1 \sum_{j=0}^{k-1} a_{11}^j}{b_1 + c_1}\right)^2 \frac{\Omega_{ee}}{\Omega_{\epsilon\epsilon}(k)}.$$

the asymptotic variances are computed under the null hypothesis. Figure 1 shows the horizon $k^* = \arg\max\{\theta(k, \gamma)\}$ *obtained by searching over* $k \in [0, 20]\,, \phi_{mn} \in [-0.9, 0]\,,$ $a_{12} \in [-0.9, 0]$ *with* $a_{11} = 0.1, b_1 = 0.1, \rho = 0.95$. *If the long-horizon regression test has no local power advantage* $k^* = 1$. *As can be seen from the figure, long-horizon regression consistently exhibits local asymptotic power advantages in this region of the parameter space.*

3.3 Monte Carlo analysis of local asymptotic predictions under covariance stationarity

In this subsection, we report the results of a Monte Carlo experiment that uses the DGP of example 1. We perform 2000 replications with $T = 500$. In this finite sample analog to the asymptotic calculations of example 1, we compute the size adjusted power of the predictive regression tests for every horizon $k \in [1, 20]$. Figure 2 plots the horizon k^* for which the relative size adjusted power is maximized.[10] The Monte Carlo experiment confirms that power gains can be achieved in finite samples with long-horizon tests when the regression error and innovations to the regressor are negatively correlated.

The surface generated by the Monte Carlo experiment this figure differs somewhat from that of Figure 1. Here, the greatest power advantage generally occurs at a longer horizon than the asymptotic calculations predicted and the region over which long-horizon test has power advantages are different. There are two reasons for the discrepancy. First, a potential pitfall of the local-to-zero asymptotic analysis in the stationary

[10] Asymptotic standard errors were calculated using the method of Andrews (1991).

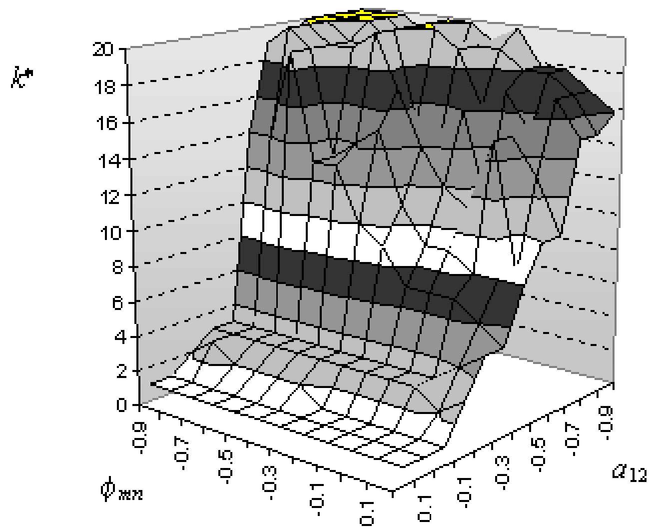

Figure 2: k^* which maximizes relative size adjusted power from Monte Carlo experiment. No constant in regression.$T = 500$, $\gamma = (a_{11}, b_{1,}\rho) = (0.1, 0.1, 0.95)$.

environment is that the effect of critical nuisance parameters (e.g., a_{12} and ϕ_{mn}) are eliminated from the asymptotic variance but remains important in small samples. The second possible reason is that the finite sample surface is sensitive to the particular HAC estimator of the asymptotic covariance matrix employed.

These power advantages are perhaps less evident in simulation studies [Berkowitz and Giorgianni (2001), Kilian (1999)]. One possibility for the discrepancy is that a nontrivial amount of small sample bias is introduced when the predictive regression is estimated with a constant. The above experiment did not include a constant.

To get an idea of the extent of the power loss that might be encountered in applied work, we include a constant in the regression and computed the size adjusted power for $T = 100$ for each of the optimal horizons k^* determined by the previous experiment. We adjust the parameters that characterize endogeneity along with local alternative true values Let $\left(a_{12}^T, \phi_{mn}^T, b_1^T\right)$ be the parameter values in the experiment for sample size T. We set $a_{12}^{100} = a_{12}^{500} \times \sqrt{500}, \phi_{mn}^{100} = \phi_{mn}^{500} \times \sqrt{500}$ and $b_1^{100} = b_1^{500} \times \sqrt{500}$. For each configuration of parameter values at the optimal horizon k^*, Figure 3 plots the absolute difference between the size adjusted power of the k^* horizon regression test generated from the $T = 500$ DGP without constant and the size adjusted power from the $T = 100$ experiment with constant. As is seen from the figure, finite sample power loss tends to be most pronounced in the same region where local asymptotic power gains are the largest (i.e., as ϕ_{mn} and a_{12} become increasingly negative).

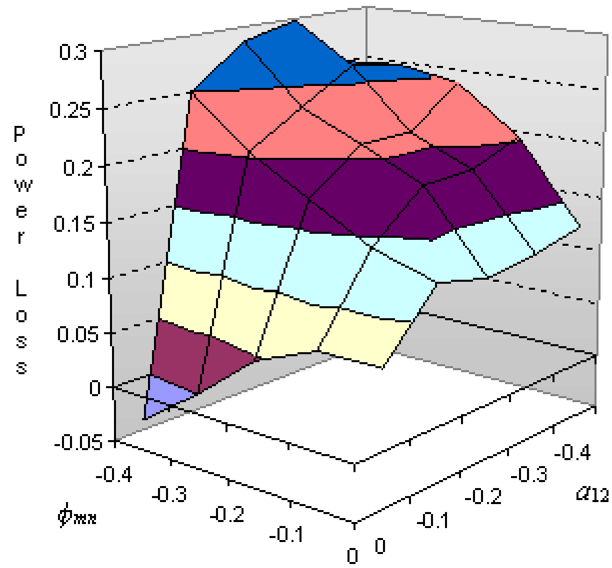

Figure 3: Power loss with endogenous regressor due to small sample bias.

4 Exact finite sample properties of long-horizon regressions

We have seen that small sample bias causes size distortion in the test and creates complications for inference. In section 4.1 we characterize the OLS small sample bias.[11] Here, we show that the magnitude and even the direction of the bias under the null hypothesis is not the same as it is under the alternative. Since a simple size adjustment under the null will not result in a test with the correct size, in Section 5 we propose a new bias reduction methods that are applicable both under the null and alternative hypotheses.

For our small sample analysis, the observations are generated according to

Assumption 3 *The observations obey*

$$\Delta y_{t+1} = \mu + \beta_1 x_t + e_{t+1}, \tag{13}$$

$$x_{t+1} = \rho x_t + u_{t+1}, \tag{14}$$

where $\{e_{t+1}\}$ *and* $\{u_{t+1}\}$ *are zero mean covariance stationary sequences.*

Note that the representation in Assumption 3 allows both $\{x_{t+1}\}$ and $\{e_{t+1}\}$ to follow general ARMA(p,q) processes. Our focus is on issues of bias, size distortion and power loss consequences from including the constant μ.

[11] The small sample bias in the short-horizon predictive regression was studied by Stambaugh (1999).

4.1 Small Sample OLS Bias

We first consider the case of an exogenous regressor. Endogeneity is considered in subsection 4.1.2.

4.1.1 Exogenous Regressor

Let a '$\sim$'denote the deviation of a variable from its sample mean. Writing (13) in deviations from the mean form gives,

$$\Delta\tilde{y}_{t+1} = \beta_1\tilde{x}_t + \tilde{e}_{t+1},$$

where $\hat{\beta}_1 = \beta_1 + \left(\sum \tilde{x}_t\tilde{e}_{t+1}\right)\left(\sum \tilde{x}_t^2\right)^{-1}$. We represent the contemporaneous correlation between $\tilde{e}_{t+1}$ and $\widetilde{u}_{t+1}$ as the projection $\tilde{e}_{t+1} = \delta\widetilde{u}_{t+1} + \widetilde{\eta}_{t+1}$, where $\delta = Cov\left(\tilde{e}_{t+1}, \widetilde{u}_{t+1}\right)/Var\left(\widetilde{u}_{t+1}\right)$. Let $\varphi_i = E\left[\sum_{t=1}^{T}\widetilde{x}_t\widetilde{u}_{t+1+i}\right]\left[\sum_{t=1}^{T}\widetilde{x}_t\right]^{-1}$. With an exogenous regressor, e_{t+1}, is not correlated with current and past values of x_t but it may be correlated with future values of x_t. In this case, φ_i is nonzero through the correlation between e_t and the sample mean of x_t. If $\delta < 0$, then probably $\varphi_i < 0$. We assume that this is the case in

Proposition 3 *(Small Sample Bias under Exogeneity) Let $\varphi_i < 0$ and the regressor be exogenous such that $E\left(e_{t+1}x_{t-s}\right) = 0, s \geq 0$. Under Assumption 3 and the null hypothesis of no predictability, the small sample bias in the OLS slope estimator in the $k-th$ horizon regression is*

$$E\left(\hat{\beta}_k^o\right) \equiv Bias\left(H_0(k)\right) = \delta\sum_{i=0}^{k}\varphi_{i,}.$$

and under the alternative hypothesis it is

$$E\left(\hat{\beta}_k^a - \beta_k^a\right) \equiv Bias\left(H_a(k)\right) = \begin{cases} \delta\varphi_0 & for\ k=1, \\ \delta\sum_{i=0}^{k}\varphi_i + \sum_{i=1}^{k-1}\beta_i^a\varphi_i & for\ k>1. \end{cases}$$

By Proposition (3), it can be seen that

$$\text{Bias}\left(H_a(k)\right) - \text{Bias}\left(H_0(k)\right) = \begin{cases} 0 & \text{for } k=1, \\ \sum_{i=1}^{k-1}\beta_i^a\varphi_i < 0 & \text{for } k>1. \end{cases}$$

In the short-horizon regression $(k=1)$, the small sample bias is identical both under the null and the alternative hypotheses. In the long-horizon $(k>1)$ regressions the bias is more pronounced under the null hypothesis than under the alternative.

The following Monte Carlo experiment verifies and quantifies these results. The DGP is as in Assumption 3 with $e_{t+1} = m_{t+1}$, $x_{t+1} = \rho x_t + u_{t+1}$, $\rho = 0.99$, $u_{t+1} = n_{t+1}$,

$(m_{t+1}, n_{t+1})' \overset{iid}{\sim} (0, \Sigma)$ where Σ is the 2×2 matrix with 1s on the diagonal and δ as the off-diagonal elements. Sample size is $T = 100$. The results, reported in Table 3, shows that larger upward biases are present under the null than under the alternative for horizons exceeding 1.

Table 3: Small-sample bias under the null and the alternative with exogenous regressor

The DGP is given in Assumption 3 with $e_{t+1} = m_{t+1}$, $x_{t+1} = \rho x_t + u_{t+1}$, $u_{t+} = n_{t+1}$, $(m_{t+1}, n_{t+1})' \overset{iid}{\sim} (0, \Sigma), \Sigma_{11} = \Sigma_{22} = 1, \Sigma_{12} = \Sigma_{21} = \delta$, $\beta_1^0 = 0, \beta_1^a = 0.1, \rho = 0.99, T = 100$.

	Null					Alternative				
δ	k=1	k=5	k=10	k=15	k=20	k=1	k=5	k=10	k=15	k=20
-0.1	0.005	0.025	0.046	0.067	0.080	0.005	-0.025	-0.161	-0.386	-0.691
-0.2	0.010	0.048	0.089	0.125	0.150	0.010	-0.001	-0.119	-0.328	-0.620
-0.3	0.015	0.071	0.131	0.182	0.221	0.015	0.022	-0.076	-0.270	-0.550
-0.4	0.020	0.095	0.174	0.240	0.291	0.020	0.046	-0.034	-0.213	-0.479
-0.5	0.025	0.118	0.216	0.297	0.362	0.025	0.069	0.009	-0.155	-0.409
-0.6	0.030	0.142	0.258	0.355	0.432	0.030	0.092	0.051	-0.098	-0.339
-0.7	0.036	0.165	0.301	0.412	0.502	0.036	0.116	0.093	-0.041	-0.269
-0.8	0.041	0.188	0.343	0.469	0.572	0.041	0.139	0.135	0.016	-0.199

4.1.2 Endogeneous Regressor

Unless sharper assumptions about the DGP are made, analytically characterizing the small sample bias when the regressor is endogenous becomes intractable. Accordingly, we adopt

Assumption 4 *The regressor $\{x_t\}$ follows the AR(p) process $x_t = \sum_{j=1}^{p} \rho_j x_{t-j} + u_t$ and the regression error $\{e_{t+1}\}$ follows the MA(1) process $e_{t+1} = m_{t+1} + \pi m_t$, where $m_{t+1} = \zeta u_{t+1} + \eta_{t+1}$ and where $\{u_t\}$ and $\{\eta_t\}$ are mutually and serially uncorrelated sequences.*

To see more explicitly the nature of the endogeneity, notice that the regression error also has the alternative representation $e_{t+1} = \zeta u_{t+1} + \pi \zeta u_t + \eta_{t+1} + \pi \eta_t$, in which the error process consists of an MA(1) component in the innovations in x_t and an MA(1) component in an *iid* random shock that is independent of u_t. The moving average component of e_t can be generalized to an MA(q) or to an ARMA(p, q) process but since the conclusions of this subsection are invariant to these generalizations, we stick to the simpler formulation.

Here, the null hypothesis of no predictability $(H_0 : \pi = 0, \beta_k = 0, k \geq 1)$, is as the null hypothesis under exogeneity of the regressor, which is characterized in Proposition 3. Under the alternative hypothesis, however, we have

Proposition 4 *(Small Sample Bias under Endogeneity) Let the observations be generated according to Assumption 4. Under the alternative hypothesis of predictability, the small sample bias of the OLS slope estimator in the kth horizon regression is*

$$E\left(\hat{\beta}_k^a - \beta_k^a\right) = \begin{cases} c + \zeta\varphi_0 - \pi\zeta \sum_{i=1}^{p} \rho_i \varphi_{i-1} & \text{for } k = 1, \\ c + \zeta \sum_{i=0}^{k} \varphi_i - \pi\zeta \sum_{i=1}^{p} \rho_i \varphi_{i-1} + \sum_{i=1}^{k-1} \beta_i^a \varphi_i & \text{for } k > 1, \end{cases}$$

where c is the asymptotic bias due to endogeneity.

The small sample bias under the null hypothesis is not the same as the OLS bias under the alternative as we note that

$$\text{Bias}(H_a(k))\text{-Bias}(H_0(k)) = \begin{cases} -\pi\zeta \sum_{i=1}^{p} \rho_i \varphi_{i-1} < 0 & \text{for } k = 1, \\ -\pi\zeta \sum_{i=1}^{p} \rho_i \varphi_{i-1} + \sum_{i=1}^{k-1} \beta_i^a \varphi_i & \text{for } k > 1. \end{cases}$$

The following Monte Carlo experiment verifies and quantifies the proposition. Here, $e_{t+1} = a_{12} u_t + m_{t+1}$. Table 4 shows that the upward bias is generally more pronounced under the null hypothesis than it is under the alternative. The essential point, however, is that the bias is not the same under the null and under the alternative.

Table 4: Small-sample bias under the null and the alternative when the Regressor is Endogenous

The DGP is given in Assumption 3 with $e_{t+1} = a_{12} u_t + m_{t+1}$, $x_{t+1} = \rho x_t + u_{t+1}$, $u_{t+1} = n_{t+1}$, $(m_{t+1}, n_{t+1})' \overset{iid}{\sim} (0, \Sigma), \Sigma_{11} = \Sigma_{22} = 1, \Sigma_{12} = \Sigma_{21} = \zeta = -0.7$, $\beta_1^0 = 0, \beta_1^a = 0.1,\ \rho = 0.99, T = 100$.

	a_{12}	k=1	k=5	k=10	k=15	k=20
Null	–	0.036	0.165	0.301	0.412	0.502
Alternative	-0.9	-0.010	0.242	0.397	0.406	0.295
	-0.8	-0.005	0.228	0.364	0.357	0.233
	-0.7	0.000	0.214	0.330	0.307	0.170
	-0.5	0.010	0.186	0.262	0.208	0.045
	-0.1	0.030	0.130	0.127	0.009	-0.206
	0.0	0.036	0.116	0.093	-0.041	-0.269

To see the opposing directions of bias under the null and alternative in graphical form, Figure 4 displays empirical distributions of $\hat{\beta}_1$ under the null and the alternative hypotheses with $T = 100$. The empirical distribution of $\hat{\beta}_1$ under the null hypothesis is seen to shift to right while that under the alternative shifts to the left.

To summarize, when the regressor is exogenous, the small sample bias at the first horizon under the null hypothesis is the same as that under the alternative. For $k > 1$,

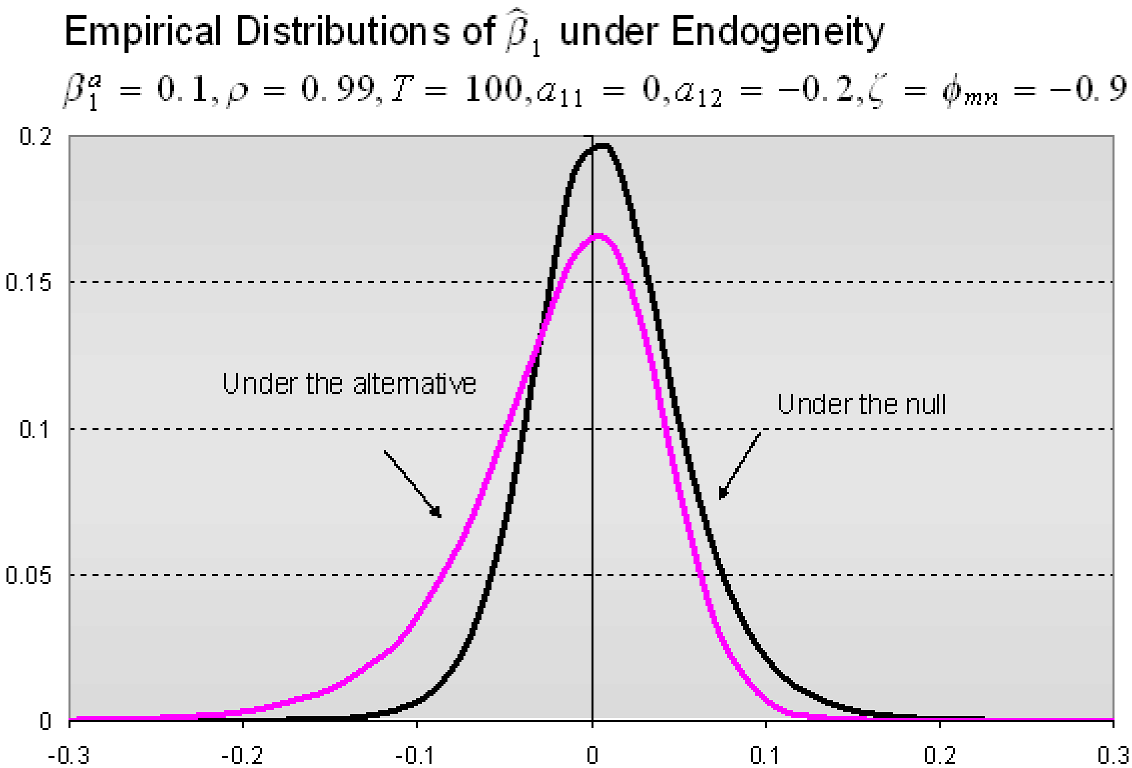

Figure 4: Upward bias under the null and downward bias under the alternative in the short-horizon predictive regression.

however, bias under the null is more pronounced than it is under the alternative and when the regressor is endogenous, the bias under the null hypothesis is different from that under the alternative at all horizons. If the bias were the same under the null and under the alternative, then the bias calculated under the null could be used as an adjustment factor to obtain tests with the correct size. Since this is not the case, we require a bias adjustment that applies both under the null and under the alternative hypotheses.

5 Bias and size adjustment

The OLS bias naturally distorts the size of long-horizon tests. For $x_t \sim AR(1)$, the OLS bias in the short-horizon regression is $E(\widehat{\beta}_1 - \beta_1) = -\delta\left(\frac{1+3\rho}{T}\right) + O(1/T^2)$, [Stambaugh (1999)] and the bias in the associated t-statistic is $E\left(t_{\hat{\beta}_1}\right) = -\delta\left(\frac{2\rho+1}{\sqrt{1-\rho^2}}\right) + O(T^{-1})$ where $\delta = \frac{Cov(x_t e_{t+1})}{Var(x_t)}$.[12] The OLS $t-$statistic is therefore biased upward for $\delta < 0$, which distorts the size of the test. Moreover, because the bias is not the same under the null and under the alternative, the (size adjusted under the null) power of the test will be affected.

In the long-horizon regression, bias, power loss, and size distortion can be magnified through cumulation of intervening short-horizon bias terms. This can be a particularly nettlesome problem when the regressor and regression error is only slightly correlated in the direction to give long-horizon regression a moderate local asymptotic power advantage. Since the small sample bias in the t-statistics can lead to such small sample power loss so as to lead one to erroneously conclude that long-horizon regressions are not useful, even asymptotically.

[12] Suppose $x_t = a + \rho x_{t-1} + u_t$ and u_t is iid. Tanaka (1984) shows that

$$P(t_{\widehat{\rho}} < z) = I(z) + \frac{i(z)}{\sqrt{T}}\left(\frac{2\rho+1}{\sqrt{1-\rho^2}}\right) + O(T^{-1})$$

where $I(z)$ is the standard normal cumulative distribution function and $i(z)$ is the standard normal density function. Obviously, the $t-$statistic is asymptotically standard normal. When x_t follows AR(1) with unknown mean like the above example, the OLS estimate of β_1 is given by $\sqrt{T}(\hat{\beta}_1 - \beta_1) = \delta\sqrt{T}(\widehat{\rho} - \rho) + \sqrt{T}(\hat{\xi} - \xi)$ where $\hat{\xi} = \left(\sum x_t \eta_{t+1}\right)\left(\sum x_t^2\right)^{-1}$. If we treat δ as a known parameter as in Cavanagh, Elliott and Stock (1995), then we have

$$P(t_{\hat{\beta}_1} < z) = \delta P(t_{\hat{\phi}} < z) + P(t_{\hat{\xi}} < z) = (\delta + 1)\, I(z) + \delta\frac{i(z)}{\sqrt{T}}\left(\frac{2\rho+1}{\sqrt{1-\rho^2}}\right) + O(T^{-1}).$$

The last term is unbiased. It follows that the unconditional mean of $t-$statistic for $\hat{\beta}_1$ is as stated in the text.

5.1 A recursive moving block Jackknife procedure

Quenouille (1956) suggested the Jackknife as a method to attenuate small sample bias. We draw on that idea and propose a recursive moving block Jackknife (henceforth RJK) procedure to obtain bias reduction in the long-horizon slope coefficient estimates as well as in the t-tests of the null hypothesis of no predictability. We begin with a discussion for obtaining bias reduction in the slope coefficient. Subsection 5.1.2 discusses the bias-reduction strategy for the test statistic.

5.1.1 OLS Bias Correction

The procedure is motivated as follows. Let Y be a random vector and $\mu = E(Y)$. Let θ_T be an estimator of the true parameter value $\theta_0 = g(\mu)$ where g is a known continuous function. Under regularity conditions assumed in Sargan (1976), the Edgeworth asymptotic expansion of $P(\sqrt{T}(\theta_T - \theta_o) < x)$ is

$$I\left(\frac{x}{\omega}\right) + i(\frac{x}{\omega})\left\{\psi_0 + \psi_2\left(\frac{x}{\omega}\right)^2\right\} + O\left(T^{-2}\right) \tag{15}$$

where ω^2 is the asymptotic variance of $\sqrt{T}(\theta_T - \theta_o)$, $I(z)$ is the standard normal cumulative distribution function (evaluated at z), $i(z)$ is the standard normal density, and ψ_0, ψ_2 are the Edgeworth expansion terms. The small sample bias of θ_T is

$$E(\theta_T - \theta_o) = \frac{\alpha}{T} + O\left(T^{-2}\right) \tag{16}$$

where $a_1 = \frac{1}{2}(\psi_0 + \psi_2)\omega$.

The bias formula (16) suggests the following procedure to correct for the bias in the predictive regression slope. To reduce clutter, we suppress the notational dependence on the horizon k. Let $\xi_s = (y_{t-1+s+k} - y_{t-1+s}, x_{t-1+s}), s = 1, ..., T_1, T_1 = T - k + 1$ be the 2-dimensional vector comprising of the dependent and independent variables of the long-horizon regression. Construct a moving block sample of size B from the original set of observations, $\left\{\xi_1, \cdots, \xi_B\right\}, \left\{\xi_2, \cdots, \xi_{B+1}\right\}, \cdots, \left\{\xi_{T_1-B+1}, \cdots, \xi_{T_1}\right\}$ .Using the data from each block, construct an estimate of the $k-$horizon slope coefficient β_0. Call the resulting estimate β_{Bj},where $j = 1, ..., T_1 - B + 1$ indexes the block of B observations. For each j, the analog to (16) is

$$\beta_{Bj} = \beta_0 + \frac{\alpha}{B} + O_p\left(B^{-2}\right). \tag{17}$$

Multiply both sides of (17) by B and taking the sample average gives

$$BE_B^*(\beta) = \alpha + B\beta_0 + O\left(B^{-2}\right), \tag{18}$$

where $E_B^*(\beta) = \frac{1}{T-B+1}\sum_{j=1}^{T-B+1}\beta_{Bj}$. Now repeat using block size $B + 1$, then block size $B + 2$, and so on through block size $B + (T_1 - B) = T_1$. This gives the sequence

$\left\{BE_B^*(\beta) = \alpha + B\beta_0, (B+1)\, E_{B+1}^*(\beta) = \alpha + (B+1)\,\beta_0, ..., T_1 E_T^*(\beta) = \alpha + \ T_1\beta_0\right\}$. Let $t = B, B+1, ..., T_1$, and define $z_t = tE_t^*(\beta)$. Write z_t as a regression on a constant and trend, $z_t = \alpha + \beta_0 t$. The coefficient estimate on the trend β_{RJK} is the RJK estimate of β_0.and it is is accurate in the following sense.

Proposition 5 *(Moving block recursive Jackknife)*

$$E(\beta_{RJK} - \beta_o) = O(T^{-2}).$$

For the choice of B we draw on Hall, Horowitz and Jing (1995) who provide the blocking rule on the bootstrap with dependent data. In the cases of bias estimation, the optimal block size suggested is by $T^{1/3}$ while for one or two sided distribution functions, the suggested size is $T^{1/4}$ and $T^{1/5}$, respectively.

5.1.2 Test statistic correction

Rothenberg (1984), Phillips and Park (1988) and Cribari-Neto and Ferrari (1995) provide Edgeworth expansions of the Wald, likelihood ratio and Lagrange multiplier tests under general conditions.[13] Here, we follow Cribari-Neto and Ferrari who showed that the Wald statistic has the asymptotic expansion,

$$W = W_T - \frac{\alpha_1}{T} W_T - \frac{\alpha_2}{T} W_T^2 \tag{19}$$

where W_T is the Wald statistic computed from a sample of size T and W is its 'true' value, α_1 and α_2 are Bartlett coefficients which are derived from the asymptotic expansion of the Wald statistic. While the explicit formulae for α_1 and α_2 depend on the specific DGP, our recursive moving-block Jackknife method does not require those formulae as it is designed to estimate the Bartlett coefficients.

The correction for the Wald statistic proceeds as follows. Construct a moving block sample of size B from the original set of observations, $\{\xi_1, \cdots, \xi_B\}$, $\{\xi_2, \cdots, \xi_{B+1}\}$, $\cdots$, $\{\xi_{T_1-B+1}, \cdots, \xi_{T_1}\}$. Using the data from each block, construct the Wald statistic (the squared t-ratio) $W_{B,j}, j = 1, ..., T_1 - B + 1$. From each block, the analog to (19) is

$$BW = -\alpha_1 W_{B,j} - \alpha_2 W_{B,j}^2 + BW_{B,j}.$$

Taking the average over j gives $BW = -\alpha_1 E_B^*(W) - \alpha_2 E_B^*(W^2) + BE_B^*(W)$, where $E_B^*(W) = \frac{1}{T_1-B+1}\sum_{j=1}^{T_1-k+1} W_{B,j}$, and $E_B^*(W^2) = \frac{1}{T_1-B+1}\sum_{j=1}^{T_1-k+1} W_{B,j}^2$. Repeat using block size $B+1$, then block size $B+2$, and so on through block size $B + (T_1 - B) = T_1$. Then for $t = B, B+1, ..., T_1$ we have $tW = -\alpha_1 E_t^*(W) - \alpha_2 E_t^*(W^2) + tE_t^*(W)$. Let $z_t = tE_t^*(W)$, and $q_t = z_t/t$. Then the size corrected test statistic is given by the estimated slope coefficient in the regression $z_t = \alpha_1 q_t + \alpha_2 q_t^2 + Wt$.

[13] Our strategy builds on Bartlett (1937) who suggested a correction technique to account for size distortion in test statistics.

6 Monte Carlo Experiments

We conduct a Monte Carlo experiment to assess the bias reduction achieved under the null hypothesis using the RJK procedure for short-and long-horizon regression coefficients and associated test statistics.[14] The DGP used is

$$\Delta y_{t+1} = \mu + \beta_1 x_t + e_{t+1} \tag{20}$$
$$x_{t+1} = \rho x_t + u_{t+1} \tag{21}$$

where $(e_{t+1}, u_{t+1})' = (m_{t+1}, n_{t+1})' \overset{iid}{\sim} (0, \Sigma), \Sigma$ is as in example 1. We set $\mu = \beta_1 = 0$ in constructing the pseudo data but estimate the regression (20) with a constant. The regressor is specified to be exogenous to focus directly on small sample bias induced by estimating the constant. We set $\rho = 0.9$ and consider various values of the correlation ϕ_{mn} between m_{t+1} and n_{t+1}. The value of ρ is suggested by the empirical example regarding stock return predictability. The results are shown in Table 5.

The OLS bias worsens as ϕ_{mn} becomes increasingly negative. The RJK procedure eliminates nearly all of the bias in the short-horizon predictive regression. The percentage of bias reduction for $\phi_{mn} = -0.9$ is $0.97, 0.62, 0.48, 0.43, 0.39$ for $k = 1, 5, 10, 15$, and 20 respectively. The relative bias reduction is fairly stable for alternative values of ϕ_{mn}.

The second two panels of the table report the size of the OLS t-test and the RJK Wald test of the hypothesis of no predictability. While the OLS t-test is oversized at $k = 1$, the RJK Wald test remains reasonably sized up through horizon 10, and becomes only moderately oversized for $k = 15$ and 20.

[14]The recursive jackknife test statistics are squared t-ratios. HAC covariance matrix estimation is done using the method of Sul, Phillips and Choi (2003).

Table 5: **Bias and Size under the Null Hypothesis**

ϕ_{mn}	$k=1$	$k=5$	$k=10$	$k=15$	$k=20$	$k=1$	$k=5$	$k=10$	$k=15$	$k=20$
	OLS Bias $\times 100$					RJK Bias $\times 100$				
-0.9	3.53	14.74	25.00	32.23	37.68	0.12	5.63	12.91	18.38	22.83
-0.7	2.75	11.44	19.35	25.03	29.33	0.09	4.34	9.96	14.26	17.72
-0.5	1.97	8.15	13.72	17.84	20.98	0.06	3.06	7.04	10.15	12.62
-0.3	1.19	4.86	8.12	10.66	12.62	0.03	1.79	4.13	6.05	7.53
-0.1	0.40	1.57	2.53	3.48	4.25	0.01	0.52	1.23	1.95	2.44
	OLS t-test 10% Size					RJK Wald test 10% Size				
-0.9	0.17	0.18	0.28	0.33	0.37	0.14	0.11	0.14	0.18	0.20
-0.7	0.15	0.16	0.25	0.30	0.34	0.12	0.09	0.14	0.17	0.18
-0.5	0.14	0.14	0.23	0.29	0.32	0.11	0.08	0.13	0.16	0.17
-0.3	0.14	0.13	0.22	0.27	0.30	0.10	0.08	0.12	0.15	0.16
-0.1	0.13	0.13	0.21	0.26	0.29	0.10	0.08	0.12	0.14	0.15
	OLS t-test 5% Size					RJK Wald test 5% Size				
-0.9	0.11	0.12	0.20	0.26	0.30	0.08	0.07	0.10	0.13	0.15
-0.7	0.09	0.10	0.18	0.23	0.27	0.06	0.06	0.10	0.12	0.14
-0.5	0.08	0.09	0.17	0.22	0.25	0.06	0.05	0.09	0.12	0.13
-0.3	0.08	0.08	0.15	0.20	0.23	0.05	0.05	0.08	0.11	0.11
-0.1	0.08	0.08	0.15	0.20	0.22	0.06	0.05	0.08	0.09	0.11

7 Equity return predictability revisited

In this section, we apply the RJK method to revisit the question of equity returns predictability. We begin by estimating the predictive regressions at horizons 1 through 20 beginning with 1971 as the end of the sample then recursively updating the sample through 2002. We begin by examining estimates of unadjusted (squared) robust OLS t-ratios. Figure 5 summarizes the essential features of these regressions by presenting the p-values of the test of no predictability at each horizon for each sample. A researcher who was estimating these regressions in real time would find that the hypothesis of no predictability could not be rejected at the 10 percent level at the one-period horizon. Small marginal significance levels for the test are consistently found at horizons 4 and above up through 1997. As the sample is updated through 2002, p-values for the test become large and exceed 0.1 at the shorter (less than 11) horizons. A change in the relation between the dividend yield and future returns appears to have taken place in 1997.

Figure 6 presents analogous information for the RJK estimates of the Wald statistic for the test of no predictability. As the sample is updated, the hypothesis of no predictability is consistently rejected at the 5-year horizon and at horizons of 10 and above. The deterioration in the p-values for samples ending in 1997 through 2002 for the OLS tests does not occur with the RJK test statistics.

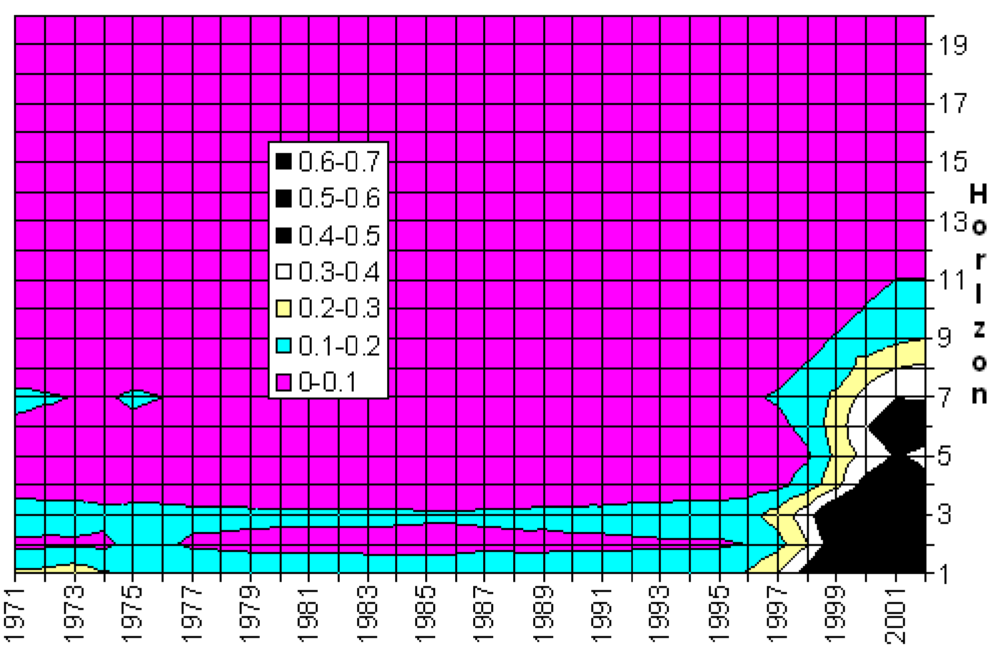

Figure 5: Significance level for standard OLS test of hypothesis of no predictability in the short-horizon regression.

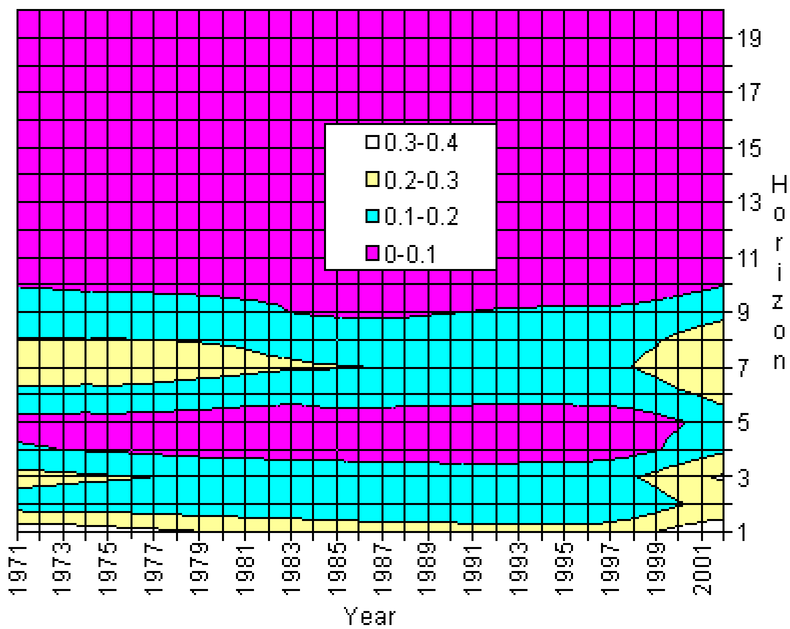

Figure 6: Significance level of test of hypothesis of no predictability in the short-horizon regression using the moving block recursive Jackknife method.

Returning to the changing dynamic character of the recursive OLS tests, we examine the recursive estimates for the short-horizon predictive regression. The RJK coefficient estimator provides an estimate of the true slope (β_1) combined with any asymptotic correlation due to endogeneity (c). It also removes small sample bias of the OLS estimator at the first horizon. We thus have the decompositions

$$\widehat{\beta}_{1,RJK} = \widehat{\beta}_1 + \widehat{c}, \tag{22}$$

$$\widehat{\beta}_{1,OLS} = \widehat{\beta}_{1,RJK} + \text{Bias}. \tag{23}$$

A direct estimate of the asymptotic endogeneity factor is provided by $\hat{c} = \hat{\Omega}_{ue}\hat{\Omega}_{xx}$ where $\hat{\Omega}_{ue}$ is the sample long run covariance between $\hat{u}_t$ and $\hat{e}_{t+1}$ and $\hat{\Omega}_{xx}$ is the sample long run variance of x_t. Figure 7 displays recursive estimates of the various pieces of the decompositions. Notice that the OLS estimates exhibit more instability than the RJK estimates. OLS lies above the RJK estimates until 1991. Both estimates decline somewhat when the sample extends to 1999. but OLS estimates drop by substantially more. The reason is that the endogeneity factor becomes much more negative during the bull market of the 1990s. The solid black line, which is the estimate of the 'true' slope β_1 declines during the 90s as well, but by a lesser amount than the endogeneity factor. As the OLS estimates decline towards zero, the power advantage tilts increasingly towards the longer horizons.

8 Conclusion

This paper provides asymptotic justification for long-horizon predictive regressions. Local asymptotic power was evaluated with a persistent (local-to-unity) regressor and for a covariance stationary regressor. Power advantages generally accrue to long-horizon regression when there is endogeneity between the regressor that is used to predict future values of the dependent variable and the regression error. The endogeneity is not necessarily the result of misspecification of a structural model since the predictive regressions are linear least squares projections used to evaluate the predictive content of a variable. Use of these regressions can be justified in the same way that a researcher chooses to employ HAC estimators instead of specifying a complete parametric model of autocorrelation and conditional heteroskedasticity.

The local asymptotic power advantages of long-horizon regression are not so obvious in small samples. The problem is that estimation of the constant term in the predictive regression causes the OLS estimator to be biased in small samples. A further complication is that the bias is not the same under the null hypothesis as it is under the alternative. Thus in small samples, the OLS t-test not only tends to be over sized, but the long-horizon t-test also suffers from power loss. A bias correction done under the null hypothesis will not result in tests of the proper size.

To address these small sample problems, we suggest a moving block recursive Jackknife method to estimate the coefficients in the asymptotic expansion of the OLS es-

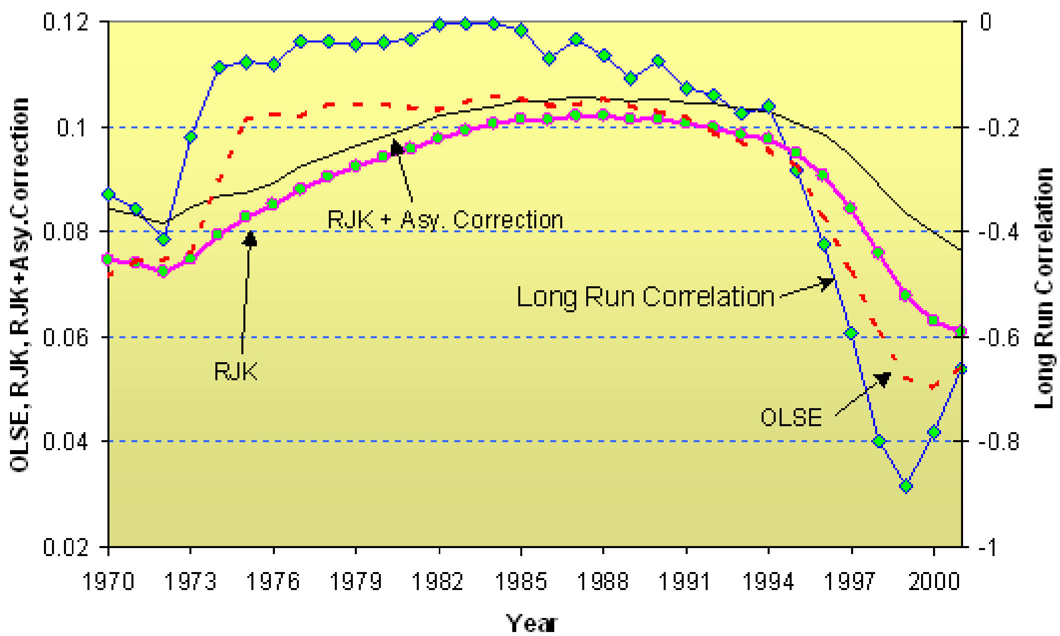

Figure 7: Recovering bias decompositions in the short-horizon predictive regression coefficient β_1.

timator and for the Wald statistic. The recursive Jackknife successfully provides bias correction in the short-horizon predictive regression. At long horizons, it is able to remove about half of the small sample bias and the Jackknifed Wald statistic is not unreasonably sized at either short or long horizons.

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

[8] CAMPBELL, JOHN Y., ANDREW W. LO, AND A. CRAIG MACKINLAY, 1997. *The Econometrics of Financial Markets*, Princeton University Press.

[9] CAMPBELL, JOHN Y. AND ROBERT J. SHILLER, 1988, 'Stock Prices, Earnings, and Expected Dividends,' *Journal of Finance*, July, 43, pp. 661-76.

[10] CAMPBELL, JOHN Y.AND MOTOHIRO YOGO, 2002. "Efficient Tests of Stock Return Predictability," mimeo, Harvard University.

[11] CAVANAGH, CHRISTOPHER L, GRAHAM ELLIOTT AND JAMES H. STOCK, 1995. "Inference in Models with Nearly Integrated Regressors," *Econometric Theory,* 11, 1131-47.

[12] 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.

[13] 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.

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

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

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

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

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

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

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

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

[22] Kim, Myung J. and Charles R. Nelson, 1993. "Predictable Stock Returns: The Role of Small Sample Bias," *Journal of Finance*, 48, 641–61.

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

[24] Kilian, Lutz and Mark P. Taylor, 2002. "Why Is It So Difficult To Beat The Random Walk Forecast of Exchange Rates?" *Journal of International Economics*, forthcoming.

[25] Lo, Andrew, 1991. "Long-Term Memory in Stock Market Prices," *Econometrica,*59, 1279-1313.2

[26] 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.

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

[28] 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).

[29] MARK, NELSON C. AND DONGGYU SUL, 2001. "Nominal Exchange Rates and Monetary Fundamentals: Evidence from a Small Post-Bretton Woods Panel," *Journal of International Economics*, 53, 29–52.

[30] MARK, NELSON C. AND DONGGYU SUL, 2002. "Asymptotic Power Advantages of Long-Horizon Regression," *mimeo,*University of Auckland.

[31] 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.

[32] NEWEY, WHITNEY K., AND KENNETH D. WEST, "Automatic Lag Selection in Covariance Matrix Estimation," *Review of Economic Studies*, October 1994, *61*(4), pp. 631–53.

[33] PHILLIPS, P.C.B., 1987. "Towards a Unified Asymptotic Theory for Autoregression," *Biometrika*, 74, 535-47.

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

[35] PHILLIPS, P.C.B., AND PARK, 1988. "Statistical Inference in Regressions with INtegrated Processes: Part 1," *Econometric Theory*, 4, 468-497.

[36] RAPACH, DAVID E. AND MARK E. WOHAR, 2002. "Testing the Monetary Model of Exchange Rate Determination: New Evidence from a Century of Data," .Journal of International Economics, 58, 359-385.

[37] RICHARDSON, MATTHEW AND JAMES STOCK, 1989. "Drawing Inferences from Statistics Based on Multiyear Asset Returns," *Journal of Financial Economics*, 25, 343-48.

[38] RICHARDSON, MATTHEW AND JAMES STOCK, 1989, "Drawing Inferences from Statistics Based on Multi-Year Asset Returns," *Journal of Financial Economics,* 25, 323–48.

[39] ROTHENBERG, THOMAS, 1984. "Approximate Normality of Generalized Least Squares Estimates," *Econometrica*, 52, 911-825.

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

[41] Sargan, J.D., 1976. "Econometric Estimators and teh Edgeworth Approximation," *Econometrica,* 44, 421-448.

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

[43] Stambaugh, Robert F. 1999. "Predictive Regressions," *mimeo*, Wharton School, University of Pennsylvania.

[44] Sul, Donggyu, Peter.C.B. Phillips, and Chi-Young Choi, 2003. "Prewhitening Bias in HAC Estimation," *mimeo*, University of Auckland.

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

Appendix

Proof of Proposition 1

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

$$\hat{\beta}_k - \beta_k = \frac{\sum_{t=1}^{T-k} x_t \epsilon_{t+k,k}}{\sum_{t=1}^{T-k} x_t^2}.$$

By Assumption 1, we have

$$T\left(\hat{\beta}_k - \beta_k\right) = kb_1 + \frac{T^{-1}\sum_{t=1}^{T-k} x_t \epsilon_{t+k,k}}{T^{-2}\sum_{t=1}^{T-k} x_t^2}.$$

By Lemmas 3.1 and Theorem 4.1 of Phillips (1988) and Cavanagh, Elliot and Stock (1995), it follows that

$$\begin{aligned} T\left(\hat{\beta}_k - \beta_k\right) \;\; &\Longrightarrow \;\; kR\left\{\delta\left(\int J_c^2\right)^{-1}\int J_c dB_1 + \left(1-\delta^2\right)^{1/2}\left(\int J_c^2\right)^{-1}\int J_c dB_2^*\right\} \\ &\quad + \frac{\Lambda_{xe} - \Lambda_{xe,k-1}}{\Omega_{xx}}\left(\int J_c^2\right)^{-1} + kb_1. \end{aligned}$$

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 $\Omega_{\epsilon\epsilon}(1) = \Omega_{ee}$, $t_\beta\left(k\right)$ can be rewritten as

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

From Phillips (1987) and Cavanagh, Elliot and Stock (1995), it is straightforward to show that

$$t_\beta(k) \Longrightarrow \delta\tau_{1c} + \left(1-\delta^2\right)^{1/2} N(0,1) + \left(\frac{\Lambda_{xe} - \Lambda_{xe,k-1} + b_1}{\sqrt{\Omega_{xx}\Omega_{ee}}}\right)\theta_c,$$

where $\Lambda_{xe,k-1} = \lim_{T\to\infty}\frac{1}{T}\sum_{t=k+1}^{T}\sum_{l=1}^{k-1} E(\Delta x_{t-l} u_t)$, $\tau_{1c} = \left(\int J_c^2\right)^{-1/2}\int J_c dB_1$, and $\theta_c = \left(\int J_c^2\right)^{-1/2}$. ■

Proof of Proposition 2

Let $\epsilon\left(k\right) = \left(\epsilon_{1,k}, ..., \epsilon_{T,k}\right)$ and $x = \left(x_1, ..., x_T\right)$. The proof of Proposition 2 is aided by the following lemmas.

Lemma 1 *Under Assumption 2, 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 2, 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

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

■

Lemma 3 *Under Assumption 2, 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

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

It holds because by Assumption 2 we get $\beta_k - \hat{\beta}_k \to^p 0$ and $T^{-1}x'e = T^{-1}x'\epsilon(k) = o_p(1)$.

■

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

Derivation of Example 1 formulae

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

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

Thus, we have determined that $\beta_1(T) \xrightarrow{p} b_1(T) + c_1(T)$ where

$$c_1(T) = \frac{E(e_{t+1}x_t)}{E(x_t)^2} = \frac{(a_{12} + a_{11}\phi_{mn})}{(1-a_{11}\rho)} \frac{1}{\sqrt{T}} \left(1-\rho^2\right),$$

Now, for $k=2$,

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

Therefore, $b_2(T) = b_1(T)(1+\rho) = \frac{b_1}{\sqrt{T}}(1+\rho)$. 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, $b_k(T) = b_1(T)\left(\sum_{j=0}^{k-1} \rho^j\right) = b_1(T)\left(\frac{1-\rho^k}{1-\rho}\right)$, 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

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

Proof of Proposition 3

For an exogenous regressor under the null hypothesis, we have at the first horizon

$$\begin{aligned}
\hat{\beta}_1^o &= \delta\left(\frac{\sum \tilde{x}_t \widetilde{u}_{t+1}}{\sum \tilde{x}_t^2}\right) + \left(\frac{\sum \tilde{x}_t \widetilde{\eta}_{t+1}}{\sum \tilde{x}_t^2}\right), \\
E\left(\hat{\beta}_1^o\right) &= \delta \varphi_0,
\end{aligned}$$

where $\varphi_j = E\left[\sum \tilde{x}_t \widetilde{u}_{t+j}\right]\left[\sum \tilde{x}_t\right]^{-1}$, which is nonzero due to the correlation between sample means. For the kth horizon, we have

$$E\hat{\beta}_{kT}^o = E\left(\frac{\sum_t \tilde{x}_t \tilde{e}_{t+1}}{\sum_t \tilde{x}_t^2}\right) + \cdots + E\left(\frac{\sum_t \tilde{x}_t \tilde{e}_{t+k}}{\sum_t \tilde{x}_t^2}\right). \tag{A.1}$$

Note that in general,

$$E\left(\frac{\sum_t \tilde{x}_t \tilde{e}_{t+1}}{\sum_t \tilde{x}_t^2}\right) \neq E\left(\frac{\sum_t \tilde{x}_t \tilde{e}_{t+k}}{\sum_t \tilde{x}_t^2}\right).$$

The correlation between $\tilde{e}_{t+1}$ and $\tilde{e}_{t+2}$ arises because both are deviations from the same sample mean. To see this,

$$\begin{aligned}
& E\left(e_{t+1} - \frac{1}{T-2}\sum_{t=2}^{T-1} e_{t+1}\right)\left(e_{t+2} - \frac{1}{T-2}\sum_{t=1}^{T-2} e_{t+2}\right) \\
= {} & E e_{t+1}e_{t+2} - \left(\frac{1}{T-2}\right) E e_{t+2}\sum_{t=2}^{T-1} e_{t+1} - \left(\frac{1}{T-2}\right) E e_{t+1}\sum_{t=1}^{T-2} e_{t+2} \\
& + \left(\frac{1}{T-2}\right)^2 E\sum_{t=2}^{T-1} e_{t+1}\sum_{t=1}^{T-2} e_{t+2} \\
= {} & 0 - \left(\frac{2}{T-2}\right)\sigma_e^2 + \left(\frac{1}{T-2}\right)^2 \sigma_e^2 (T-3) \simeq -\frac{\sigma_e^2}{T-2}.
\end{aligned}$$

This relation holds similarly between $\widetilde{u}_{t+1}$ and $\widetilde{u}_{t+k}$. Hence $(A.1)$ can be rewritten as

$$E\hat{\beta}_k = \delta \sum_{i=0}^{k-1} \varphi_i.$$

Under the alternative hypothesis,

$$\begin{aligned} E\hat{\beta}^a_{kT} - \beta^a_k &= E\left(\frac{\sum_t \tilde{x}_t\tilde{e}_{t+1}}{\sum_t \tilde{x}_t^2}\right) + \cdots + E\left(\frac{\sum_t \tilde{x}_t\tilde{e}_{t+k}}{\sum_t \tilde{x}_t^2}\right) \\ &\quad +\beta^a_1 E\left(\frac{\sum_t \tilde{x}_t\widetilde{u}_{t+1}}{\sum_t \tilde{x}_t^2}\right) + \cdots + \beta^a_{k-1} E\left(\frac{\sum_t \tilde{x}_t\widetilde{u}_{t+k-1}}{\sum_t \tilde{x}_t^2}\right) \\ &= \delta\sum_{i=0}^{k-1}\varphi_i + \sum_{i=1}^{k-1}\beta^a_i\varphi_i, \end{aligned}$$

for $k > 0$. Let $\text{Bias}(H_0(k)) = E\left(\hat{\beta}^o_{kT} - \beta^o_k\right)$ and $\text{Bias}(H_a(k)) = E\left(\hat{\beta}^a_{kT} - \beta^a_k\right)$. Then we have

$$Bias\left(H_a(k)\right) - Bias\left(H_0(k)\right) = \begin{cases} 0 & \text{for } k = 1, \\ \sum_{i=1}^{k-1}\beta^a_i\varphi_i < 0 & \text{for } k > 1 \end{cases}.$$

■

Proof of Proposition 4

Under Assumption 4, we have

$$\begin{aligned} E\left(\frac{\sum_t \tilde{x}_t\tilde{e}_{t+1}}{\sum_t \tilde{x}_t^2}\right) &= \zeta E\left(\frac{\sum_t \tilde{x}_t\widetilde{u}_{t+1}}{\sum_t \tilde{x}_t^2}\right) + \pi\zeta E\left(\frac{\sum_t\left(\rho_1\tilde{x}_{t-1} + \cdots + \rho_p\tilde{x}_{t-p} + \widetilde{u}_t\right)\widetilde{u}_t}{\sum_t \tilde{x}_t^2}\right) + O(T^{-2}) \\ &= \zeta K_0 + \pi\zeta\sum_{i=1}^{p}\rho_i\varphi_{i-1} + \pi\zeta E\left(\frac{\sum_t \widetilde{u}_t^2}{\sum_t \tilde{x}_t^2}\right) + O(T^{-2}) \end{aligned}$$

The third term can be rewritten as

$$\begin{aligned} E\left(\frac{\sum_t \widetilde{u}_t^2}{\sum_t \tilde{x}_t^2}\right) &= E\left(\frac{\sum_t\left(\tilde{x}_t - \sum_{j=1}^p\rho_j\tilde{x}_{t-j}\right)^2}{\sum_t \tilde{x}_t^2}\right) \\ &= E\left(\frac{\sum_t \tilde{x}_t^2 - 2\sum_t \tilde{x}_t\left(\sum_{j=1}^p\rho_j\tilde{x}_{t-j}\right) + \sum_t\left(\sum_{j=1}^p\rho_j\tilde{x}_{t-j}\right)^2}{\sum_t \tilde{x}_t^2}\right) \\ &= 1 - \left(\sum_{j=1}^p\rho_j\right)^2 - 2\left(\frac{\sum_t\left(\sum_{j=1}^p\rho_j\tilde{x}_{t-j}\right)\widetilde{u}_t}{\sum_t \tilde{x}_t^2}\right) \\ &= 1 - \left(\sum_{j=1}^p\rho_j\right)^2 - 2\sum_{i=1}^p\rho_i\varphi_{i-1}. \end{aligned}$$

It follows that

$$E\left(\frac{\sum_t \tilde{x}_t \tilde{e}_{t+1}}{\sum_t \tilde{x}_t^2}\right) = \zeta\varphi_0 - \pi\zeta \sum_{i=1}^{p} \rho_i \varphi_{i-1} + \pi\zeta \underbrace{\left[1 - \left(\sum_{j=1}^{p} \rho_j\right)^2\right]}_{c} + O(T^{-2}),$$

where $c = \left[1 - \left(\sum_{j=1}^{p} \rho_j\right)^2\right]$ is the asymptotic bias term due to the endogenous regressor. The small sample bias of $\hat{\beta}_1^a$ is

$$E\left(\hat{\beta}_1^a - \beta_1^a - c\right) = \zeta\varphi_0 - \pi\zeta \sum_{i=1}^{p} \rho_i \varphi_{i-1} + O(T^{-2}),$$

where c is the asymptotic bias term.

For the kth horizon, the asymptotic bias term does not change since e_{t+1} is not serially correlated. Hence the small sample bias of $\hat{\beta}_k^a$ is given by

$$E\left(\hat{\beta}_k^a - \beta_k^a - c\right) = \zeta \sum_{i=1}^{k} \varphi_i + \sum_{i=1}^{k-1} \beta_i^a \varphi_i - \pi\zeta \sum_{i=1}^{p} \rho_i \varphi_{i-1} + O(T^{-2})$$

■

Proof. (of Proposition 5) The regression is set up as $tE_t^*(\beta) = \alpha + \beta t + v_t$. Let a $\widetilde{\ }$ denote the deviation from the sample mean and note that $\frac{1}{T}\sum_t \widetilde{t}\widetilde{v}_t = O_p(1)$. Then

$$E\left(\beta_{RJK} - \beta_0\right) = E\left(\frac{\sum_t \widetilde{t}\widetilde{v}_t}{\sum_t \widetilde{t}^2}\right) = \frac{E\left(\sum_t \widetilde{t}\widetilde{v}_t\right)}{O\left(T^3\right)} = \frac{O(T)}{O\left(T^3\right)} = O\left(T^{-2}\right).$$

. ■