Asymptotic Power Advantages of Long-Horizon Regressions

Nelson Mark* Donggyu Sul†

*University of Auckland, NMARK@ND.EDU
†University of Auckland, d.sul@auckland.ac.nz

Asymptotic Power Advantages of Long-Horizon Regressions

Nelson C. Mark † and Donggyu Sul‡

April 2002

Abstract

Local asymptotic power advantages are available for testing the hypothesis that the slope coefficient is zero in regressions of $y_{t+k} - y_t$ on x_t for $k > 1$, when $\{\Delta y_t\} \sim I(0)$ and $\{x_t\} \sim I(0)$. The advantages of these long-horizon regression tests accrue in empirically relevant regions of the admissible parameter space. In Monte Carlo experiments, small sample power advantages to long-horizon regression tests accrue in a region of the parameter space that is larger than that predicted by the asymptotic analysis.

Comments welcomed.

† Department of Economics, Ohio State University and NBER. ‡ Department of Economics, University of Auckland. Sul gratefully acknowledges thanks to University of Auckland for support under grant no. 3601013/9254.

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 hypothesized predictor of future returns known at time t. 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 exchange rate and x_t the deviation of the exchange rate from a set of macroeconomic fundamentals.[1] To test the predictability of the return, one can perform a short-horizon regression test by regressing the one-period ahead return r_{t+1} on x_t and doing a t-test on the slope coefficient. However, empirical research in finance and economics frequently goes beyond this to employ a long-horizon regression strategy in which a multi-period future return on the asset, $y_{t,k} = \sum_{j=1}^{k} r_{t+j}$, is regressed on x_t,

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

and the null hypothesis $H_0 : \beta_k = 0$ tested using a t-statistic constructed with a heteroskedastic and autocorrelation consistent (HAC) standard error. Typically, researchers find that there is a range over $k > 1$ in which the marginal significance level of a test of no predictability is declining in k. The short-horizon regression test may fail to reject the hypothesis of no predictability when the long-horizon test does reject. Not only do the asymptotic t-ratios tend to increase with horizon but so do point estimates of the slope coefficient and the regression R^2.

The underlying basis for these results are not fully understood and they are puzzling because the long-horizon regression is built up by addition of the intervening short-horizon regressions. 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." There are two aspects to this question. The first is whether long-horizon regression tests can be justified on the basis of asymptotic theory. The second question concerns small sample bias of OLS in the presence of a predetermined but endogenous regressor and potential small sample size distortions

[1] This line of research includes Fama and French (1988) 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 (2001) who regressed long-horizon exchange rate returns on the deviation of the exchange rate from its fundamental value. Alexius (2001) and Chinn and Merideth (2001) regress long-horizon exchange rate returns on long-term bond yield differentials.

of the tests. This paper is primarily concerned with the first question concerning asymptotic justification.

Using local-to-zero asymptotic analysis, we show that there exist nontrivial regions of the admissible parameter space under which long-horizon regression tests have asymptotic power advantages over short-horizon regression tests. When the regressor is exogenous, long-horizon regressions can have substantial local asymptotic power advantages over short-horizon regressions but the power advantages occur in regions either where the regression error or the predictor $\{x_t\}$ (or both) exhibit negative serial correlation. While noteworthy, this does not provide the asymptotic justification for empirical findings that returns are predictable. Negative serial correlation of the regressor is not a prominent characteristic of the data used empirical applications of long-horizon regressions nor is strict exogeneity is a realistic assumption in applied work.

For example, endogeneity arises in the case of stock returns because both the one-period ahead return r_{t+1} and the current dividend yield x_t depend on the stock price at time t so that innovations to the time $t+1$ dividend yield will in general be correlated with the regression error in (1) even though x_t is not. In general, endogeneity might arise simply because the short-horizon predictive regression is not a structural equation but is a linear least squares projection of the future return r_{t+1} onto x_t. When we relax the assumption of exogeneity in favor of a data generating process that exhibits local-to-zero endogeneity, we find that asymptotic power advantages associated with long-horizon regression accrue in the empirically relevant region of the parameter space—where $\{x_t\}$ is positively autocorrelated and persistent, where the short-horizon regression error exhibits low to moderate serial correlation, and where the innovations to the regressor and the regression error are negatively contemporaneously correlated.

These theoretical power comparisons are valid asymptotically and for local alternative hypotheses. This leads to the question as to whether there are any practical power advantages associated with long-horizon regression tests in samples of small to moderate size. We investigate this issue by examining finite sample size-adjusted power comparisons of long-and short-horizon regressions in a set of Monte Carlo experiments. This analysis confirms that size-adjusted power advantages accrue to long-horizon regressions even in sample sizes of 100. The power advantages are obtained for persistent regressors in a similar but larger region as found in the asymptotic analysis–that is where the regression error exhibits low to moderate serial correlation and its innovation is negatively correlated with the regressor's innovation.

We now mention related issues and papers in the literature. 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 constrained to be small relative to the sample size with $k/T \to 0$ as $T \to \infty$. There is a different long-horizon regression that has been employed in the literature in which the future k-period return (from t to $t+k$) is regressed on the past k-period return (from $t-k$ to t) [Fama and French (1988b)]. An issue that arises in this work is that the return horizon k can be large relative to the size of the sample T. Richardson and Stock (1989) employ an alternative asymptotic theory in which both $k \to \infty$ and $T \to \infty$ but $k/T \to \delta$, $(0 < \delta < 1)$ and show that the test statistics converge to functions of Brownian motions. Daniel (2001) studies optimal tests of this kind. Valkanov (1999) employs the Richardson and Stock asymptotic distribution theory to the long-horizon regressions of the type that we study when the regressor $x_t \sim I(1)$.

A paper closely related to ours is Campbell (1993), who studied an environment where the regressor $\{x_t\}$ follows an AR(1) 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) corroborate by Monte Carlo analysis. Mankiw and Shapiro (1986), Hodrick (1992), Kim and Nelson (1993), and Goetzmann and Jorion (1993), Mark (1995), and Kilian (1999a) study small-sample inference issues and Stambaugh (1999) proposes a Bayesian analysis to deal with small sample bias. Kilian and Taylor (2002) examine finite sample properties under nonlinearity of the data generation process and Clark and McCraken (2001) study the predictive power of long-horizon out-of-sample forecasts.

The remainder of the paper is as follows. The next section reviews two canonical examples of the use of long-horizon regression tests in the empirical finance and international economics literature which motivate our study. Section 2 presents our local-to-zero asymptotic power analysis when the regressor $\{x_t\}$ is econometrically exogenous. In section 3 we relax the exogeneity assumption in favor of a sequence of data generating processes that exhibit local-to-zero endogeneity. We include here as well, the results of a Monte Carlo experiment to assess finite sample size-adjusted relative power comparisons of the long- and short-horizon regression tests. Section 4 concludes. Derivations are relegated to the appendix.

1 Canonical empirical examples

We illustrate and motivate the econometric issues with two canonical empirical examples. The first example begins with Fama and French (1988b) and Campbell and Shiller (1988) who study the ability of the log-dividend yield to predict future stock returns. We revisit this work with an examination of dividend yields and returns on the Standard and Poors (S&P) index of equities. Returns from month t to $t+1$ on the index from 1871.01 to 1995.12 are $r_{t+1} = \ln\left((P_{t+1} + D_t)/P_t\right)$ where P_t is the price of the S&P index and D_t is the annual flow of dividends from $t-11$ through month t.[2] Here, the short-horizon regression is formed by annual ($k = 12$) returns since dividends are an annual flow. Campbell et. al. (1997) 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 is a natural choice for prediction of future returns.

We run the equity return regressions at horizons of 1, 2, 4, and 8 years and compute HAC standard errors using the automatic lag selection method of Newey and West (1994). As can be seen from panel A of Table 1, the evidence for return predictability appears to strengthen as the horizon is lengthened. Slope coefficient point estimates, HAC asymptotic t-ratios, and regression R^2s for the stock return regression all increase with return horizon.

In our second empirical example [see Mark (1995) and Chinn and Meese (1995)] the long-horizon regression is used to test whether standard monetary fundamentals have predictive power for future exchange rate returns. Here, the return is the depreciation rate of the exchange rate $r_{t+1} = \ln(S_{t+1}/S_t)$ where S_t is the nominal exchange rate. The regressor is $x_t = \ln(F_t/\ln S_t)$, the fundamental value is $F_t = (M_t/M_t^*)(Y_t^*/Y_t)$ where M_t and Y_t are the domestic money supply and domestic income respectively, and asterisks refer to foreign country variables. According to the monetary model

[2] These data were used in Robert J. Shiller (2001) and were obtained from his web site.

of exchange rate determination, the exchange rate is the expected present value of future values of the fundamental F_t. Assuming that the pricing relationship holds in the long run and noting that the fundamentals evolve more smoothly than the exchange rate, suggests that the current deviation of the log exchange rate from the log fundamental $\ln(F_t/S_t)$ should predict future exchange rate returns.

We revisit the long-horizon predictability of exchange rate returns with an examination of US–UK data set.[3] These data are 100 quarterly observations spanning from 1973.1 to 1997.3. Here, 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). Exchange rate regression estimates at horizons of 1, 2, 3, and 4 years are shown in panel B of Table 1. The familiar pattern of t-ratios and regression R^2s increasing with horizon are present here as well.

We note that in both examples, the regressor $\{x_t\}$ is highly persistent. The augmented Dickey–Fuller and Phillips–Perron unit root tests reported in Table 2 gives a sense of this persistence. An analysis of the entire sample of 1500 observations of the log dividend yield allows the unit root to be rejected at the 5 percent level but if one were to analyze the first 288 monthly observations (or 24 years) the unit root would not be rejected. Similarly, the third column of the table shows that a unit root in the deviation of the log exchange rate from the log fundamentals cannot be rejected at standard significance levels. Failure to reject the null hypothesis does not require us to accept it and such a decision can be guided by the well known low power properties in small samples of unit root tests. Evidence against a unit root is potentially stronger in an analysis of a long historical record, as in Rapach and Wohar (2001). In the ensuing analysis, we pay close attention to environments in which $\{x_t\}$ is persistent but $I(0)$.

2 Asymptotic power under exogeneity

The analysis in this section is based on a sequence of data generating processes with an exogenous regressor given by

Assumption 1 *(Exogeneity.) The observations obey*

$$\Delta y_{t+1} = \beta_1(T)x_t + e_{t+1}, \tag{2}$$

[3]These data are from Mark and Sul (2001).

where T is the sample size, and $\{x_t\}$ and $\{e_t\}$ are independent zero mean covariance stationary sequences. The slope coefficient is given by the sequence of local alternatives $\beta_1(T) = b_1/\sqrt{T}$ where b_1 is a fixed constant.

For analytical convenience, the constant in the regression is suppressed although a constant is included in all of our Monte Carlo simulations. The short horizon regression is the linear least squares projection of Δy_{t+1} onto x_t. It is used to estimate functions of the underlying moments of the distribution between $\{y_t\}$ and $\{x_t\}$. By construction, $\mathrm{E}(e_{t+1}x_t) = 0$ but because (2) is not a structural equation, we do not require the error sequence $\{e_t\}$ to be serially uncorrelated.

We use the following notation. $C_j(x) = \mathrm{E}(x_t x_{t-j})$ is the autocovariance function for $\{x_t\}$, and $\rho_j(x) = C_j(x)/C_0(x)$ is its autocorrelation function. Note that $\rho_j(x)x_t$ is the linear least squares projection of x_{t+j} onto x_t. Analogously, the autocovariance and autocorrelation function for $\{e_t\}$ are denoted $C_j(e) = \mathrm{E}(e_t e_{t-j})$ and $\rho_j(e) = C_j(e)/C_0(e)$, respectively. Let γ be the parameter vector of the data generating process. Although the above defined moments depend on γ, we suppress the notational dependence when when no confusion will arise.

Using the projection representation, $x_{t+j} = \rho_j(x)x_t + u_{t+j,j}$ where $u_{t+j,j}$ is the least squares projection error, the long-horizon regression ($k > 1$) is obtained by addition of short-horizon regressions,

$$y_{t+k} - y_t = \beta_k(T)x_t + \epsilon_{t+k,k}, \tag{3}$$

where

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

The dependence of $\epsilon_{t+k,k}$ on the projection errors $u_{t+j,j}$ vanish asymptotically. As a result, the asymptotic variance of the OLS estimator is calculated under the null. The asymptotic distribution for the OLS estimator of the slope coefficient $\hat{\beta}_k$ in the

$k-$horizon regression is[4]

$$\sqrt{T}\left(\hat{\beta}_k\right) \xrightarrow{D} N\left[b_1\left(1+\sum_{j=1}^{k-1}\rho_j(x)\right), V\left(\hat{\beta}_k\right)\right], \tag{4}$$

where

$$\begin{aligned} V\left(\hat{\beta}_k\right) &= \frac{\Omega_{0k}+2\sum_{j=1}^{\infty}\Omega_{jk}}{C_0^2(x)}, & (5)\\ \Omega_{jk} &= \lim_{T\to\infty} \mathrm{E}\left(x_{t-k}x_{t-j-k}\epsilon_{t,k}\epsilon_{t-j-k,k}\right) = C_j(x)G_{j,k}(e), & (6)\\ G_{j,k}(e) &= kC_j(e)+\sum_{s=1}^{k-1}(k-s)\left(C_{j-s}(e)+C_{j+s}(e)\right). & (7) \end{aligned}$$

Under the sequence of local alternatives, the squared t-ratio for a test of the null hypothesis $H_0 : \beta_k = 0$ has the asymptotic non central chi-square distribution

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

with noncentrality parameter

$$\lambda_k = \frac{b_1^2\left[1+\sum_{j=1}^{k-1}\rho_j(x)\right]^2}{V(\hat{\beta}_k)}. \tag{8}$$

We are now ready to state the criterion under which a long-horizon regression test has local asymptotic power advantage over the short-horizon regression test.

Proposition 1 *Let γ be the denote the parameter vector for the data generating process. The long-horizon regression ($k > 1$) test of $H_0 : \beta_k = 0$ has an asymptotic local power advantage over the short-horizon regression ($k = 1$) test if*

$$\theta(k;\gamma) = \frac{\lambda_k(\gamma)}{\lambda_1(\gamma)} = \left(\frac{\beta_k(T)}{\beta_1(T)}\right)^2 \frac{V(\hat{\beta}_1)}{V(\hat{\beta}_k)} > 1. \tag{9}$$

We take $\theta(k,\gamma)$ to be a measure of relative local asymptotic power.[5] Under as-

[4]See the appendix.

[5]We assume a local alternative hypothesis because the t-test is a consistent test under a fixed alternative. That is, under a fixed alternative hypothesis, the power of both the short-horizon

sumption 1, the ratio of the slope coefficients is $\beta_k(T)/\beta_1(T,) = \left[1 + \sum_{j=1}^{k-1} \rho_j(x)\right]$.

In the remainder of this section, we explore whether there exist regions of the admissible parameter space under which long-horizon regressions satisfy (9). We evaluate relative local asymptotic power of long-horizon regression tests under various assumptions concerning the dynamics governing the regressor $\{x_t\}$ and the short-horizon regression error $\{e_t\}$. The regions of the parameter space over which there are no power advantages to long-horizon regression hold little interest for us. Accordingly, in the analysis to follow, we focus on parameter values under which long-horizon regression tests do have power advantages.

We begin with the environment considered by Berben (2000) in which the regressor and the regression error each follow independent AR(1) processes.

Case I. Let $\{x_t\}$ and $\{e_t\}$ evolve according to

$$e_t = \mu e_{t-1} + m_t \tag{10}$$

$$x_t, = \phi x_{t-1} + v_t, \tag{11}$$

where $(m_t, v_t)' \overset{iid}{\sim} (0, I_2)$. Then the parameter vector of the DGP is $\gamma = (\phi, \mu)$ and $\rho_j(x) = \phi^j$, $\rho_j(e) = \mu^j$. Let $g_{j,k}(e) = G_{j,k}(e)/C_0(e) = k\mu^j + \sum_{s=1}^{k-1}(k-s)\left[\mu^{|j-s|} + \mu^{j+s}\right]$. The measure of relative asymptotic power is

$$\theta(k;\gamma) = \lim_{p\to\infty}\left[\left(\frac{1-\phi^k}{1-\phi}\right)^2 \left(\frac{\sum_{j=-p}^{p} \phi^j g_{j,1}(e)}{\sum_{j=-p}^{p} \phi^j g_{j,k}(e)}\right)\right]. \tag{12}$$

For given values of $|\mu| < 1$ and $|\phi| < 1$, we evaluate (12) over horizons $1 \leq k \leq 20$. The summations in forming the long-run variances are truncated at $p = 1000$. Table 3 reports maximized values of $\theta(k;\gamma)$ for selected values of $\gamma = (\phi,\ \mu)$. Table entries of $\theta(k;\gamma) = 1$ indicate local asymptotic power is maximized at $k = 1$. The longest horizon for which long-horizon regression tests have local asymptotic power advantages $(\theta(k;\gamma) > 1)$ is $k = 2$. As can be seen, asymptotic power advantages accrue to the long-horizon test when the regression error $\{e_t\}$ is negatively serially

regression and the long-horizon t-tests are asymptotically 1. Because both tests are consistent, it is becomes difficult to compare their asymptotic power. The analysis of power under local alternatives lets the alternative get close to the null at the same rate as the accumulation of new information leads to improved precision in estimation and inference, $\sqrt{T}$. This serves to offset the power gains one would observe under a fixed alternative. Power under local alternative remains modest (less than 1) asymptotically thus facilitating an asymptotic comparison.

correlated although the regressor $\{x_t\}$ may exhibit either positive or negative serial correlation. Values for which $\theta(k;\gamma) > 1$ are plotted in Figure 1 for $k = 2$ in the range $-0.99 \leq \mu \leq -0.38$ and $0 \leq \phi \leq 0.7$. The figure delineates the region of the parameter space under which the regression test at horizon $k = 2$ has a local asymptotic power advantage over the short-horizon regression.

Case II. In this case, we allow $\{e_t\}$ to follow an AR(2) and $\{x_t\}$ to follow an AR(1),

$$e_t = \mu_1 e_{t-1} + \mu_2 e_{t-2} + m_t, \quad (13)$$

$$x_t = \phi x_{t-1} + v_t, \quad (14)$$

where $\gamma = (\phi, \mu_1, \mu_2)$, $(m_t, v_t)' \overset{iid}{\sim} (0, I_2)$, and $\rho_j(x) = \phi^j$. The first-order autocorrelation for $\{e_t\}$ is $\rho_1(e) = \mu_1/(1-\mu_2)$. For $j \geq 2$, the autocorrelation function is obtained recursively by the Yule-Walker equations, $\rho_j(e) = \mu_1 \rho_{j-1}(e) + \mu_2 \rho_{j-2}(e)$. It follows that $\theta(k,\gamma)$ is given by (12) with $g_{j,k}(e,\gamma) = k\rho_j(e) + \sum_{s=1}^{k-1}(k-s)\left[\rho_{j-s}(e) + \rho_{j+s}(e)\right]$. The admissible region of the parameter space is $|\phi| < 1$ and the triangular region for (μ_1, μ_2) that ensures that $\{e_t\}$ is stationary.

Table 4 displays selected values of $\theta(k;\gamma)$ in the region of positive serial correlation $(0 < \phi < 1)$ of the regressor along with the horizon under which the measure of relative asymptotic power is maximized. Summations for the asymptotic variances are truncated at $p = 1000$.[6] The table also shows the first two autocorrelations for $\{e_t\}$ and the variance ratio statistic for $\{e_t\}$ at horizon 10 as a summary of the autocorrelation function of the error term. From the results given in the top half of the table, it can be seen that for persistent regressors (large ϕ), somewhat modest power gains are available when both $\{x_t\}$ and $\{e_t\}$ are persistent. The dramatic asymptotic power advantages, however, accrue to the long-horizon regression test when the error term exhibits negative serial correlation. Figure 2 displays subregions of the parameter space under which power advantages are obtained with plots of $\theta(k;\gamma) > 1$ in regions of persistent $\{x_t\}$. Each figure corresponds to a given value of ϕ. Power advantages of long-horizon regression test are concentrated in the region of complex roots in which the autocorrelation function of $\{e_t\}$ fluctuates in sign.

Case III. We now assume that the error term follows an AR(1) and the regressor

[6] Local asymptotic power advantages were also found to accrue to long-horizon regression in the region of $(-1 < \phi < 0)$ but these results are not shown as this is not empirically relevant.

follows an AR(2).

$$e_t = \mu e_{t-1} + m_t, \tag{15}$$

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + v_t, \tag{16}$$

where $\gamma = (\phi_1, \phi_2, \mu)$, $(m_t, v_t) \overset{iid}{\sim} (0, I_2)$, $\rho_j(e) = \mu^j$, and $g_{j,k}(e,\gamma) = k\mu^j + \sum_{s=1}^{k-1}\left(\mu^{|j-s|} + \mu^{j+s}\right)$. The autocorrelation function for $\{x_t\}$ is obtained recursively for $j > 1$ by $\rho_j(x) = \phi_1\rho_{j-1}(x) + \phi_2\rho_{j-2}(x)$ with $\rho_1(x) = \phi_1/(1-\phi_2)$. The measure of relative asymptotic power of the long-horizon regression test is

$$\theta(k;\gamma) = \lim_{p\to\infty}\left(\left[1 + \sum_{j=1}^{k-1}\rho_j(x)\right]^2 \frac{\sum_{j=-p}^{p}\rho_j(x)g_{j,1}(e,\gamma)}{\sum_{j=-p}^{p}\rho_j(x)g_{j,k}(e,\gamma)}\right).$$

Table 5 reports $\theta(k;\gamma)$ evaluated at selected parameter values. As in case I and case II, local asymptotic power advantages are available to long-horizon regression when the regression error $\{e_t\}$ is negatively serially correlated. Given $-1 < \mu < 0$, sizable relative local power accrues to the long-horizon test when $\{x_t\}$ is persistent but in regions where the sign of the autocorrelations oscillate. For example, a measure of relative power of 97.0 is obtained under high persistence of the regressor—the variance ratio of $\{x_t\}$ at horizon 10 is 4.7. In regions of modest negative serial correlation of $\{e_t\}$ (e.g., $\mu = -0.42$) and $\rho_j(x) > 0$ for all j, long-horizon regression tests have much smaller power advantages ($\theta = 1.097$). Under case III, we find that relative power is maximized at horizons ranging from $k = 1, \ldots 10$.

Figure 3 plots $\theta(k;\gamma)$ for selected parameter values under case III. Each figure corresponds to a fixed value of μ. As can be seen, relative asymptotic power is most sensitive to the properties of the regression error $\{e_t\}$. The more negatively serially correlated is $\{e_t\}$, the larger the statistical advantage accruing to the long-horizon test.

To summarize, when the regressor is econometrically exogenous, potential asymptotic power advantages are available to long-horizon regression tests. Power advantages accrue to long-horizon tests in the empirically relevant case where the regressor is persistent. These power advantages tend to be quite modest when the short-horizon regression error term exhibits low or positive serial correlation and can be dramatic when the error is negatively serially correlated. Large negative serial correlation of the regression error, however, is not a feature of either stock return or foreign exchange return data so the cases that we have studied in this section probably is not relevant to the empirical work. Moreover, because the short-horizon regression is not a structural equation the assumption of exogeneity is typically violated in applications. In the next section, we relax the exogeneity assumption.

3 Asymptotic power under endogeneity

In the short-horizon regression for stock returns discussed in section 1, we regressed $\Delta y_{t+1} = \ln(P_{t+1} + D_t) - \ln P_t$ on $x_t = \ln D_{t-1} - \ln P_t$. While regression error is uncorrelated with the regressor by construction, the exogeneity of $\{x_t\}$ in this case is an untenable assumption. This is because both y_{t+1} and x_{t+1} depend on $\ln P_{t+1}$ and we would thus expect that the regression error and the innovation to $\{x_t\}$ to be negatively correlated, $\mathrm{E}(v_{t+1}e_{t+1}) < 0$. Similarly, in the short-horizon regression for exchange rates, we regress $\Delta y_{t+1} = \ln S_{t+1} - \ln S_t$ on $x_{t+1} = \ln F_{t+1} - \ln S_{t+1}$ and expect the innovation to $\{x_{t+1}\}$ and the short-horizon regression error to be negatively correlated. The expected negative correlation in the innovations to the regression error and to the regressor are in fact present in the data. Fitting a first-order vector autoregression to $(e_t, v_t)'$, gives an estimated innovation correlation of -0.948 for stocks and -0.786 for exchange rates.

A simple vector error correction model (VECM) makes a similar point. As an example, suppose that the bi-variate sequence $\{(y_t, z_t)'\}$ obeys the first-order VECM with cointegration vector $(-1, 1)$ and equilibrium error $x_t = z_t - y_t$,

$$\begin{pmatrix} \Delta y_t \\ \Delta z_t \end{pmatrix} = \begin{pmatrix} \delta_1 x_{t-1} \\ \delta_2 x_{t-1} \end{pmatrix} + \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} \Delta y_{t-1} \\ \Delta z_{t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_t \\ u_t \end{pmatrix}. \tag{17}$$

To relate the VECM to the empirical examples, in the case of equity returns, y_t is the log price of the equity portfolio and z_t is the log dividend. From Campbell et. al. (1997), the return on equity has the approximate representation $r_{t+1} \simeq \rho \Delta y_{t+1} + (1-\rho)x_t$, where ρ is the implied discount factor using the average dividend yield as the discount rate. In the analysis of exchange rates, Δy_t is the exchange rate return and z_t is the log of the fundamentals.

The VECM (17) has the equivalent restricted vector autoregressive (VAR) representation for $(\Delta y_t, x_t)$,

$$\begin{pmatrix} \Delta y_t \\ x_t \end{pmatrix} = \begin{pmatrix} (a_{11}+a_{12}) & (\delta_1+a_{12}) \\ (a_{22}-a_{12}+a_{21}-a_{11}) & (1+\delta_2-\delta_1+a_{22}-a_{12}) \end{pmatrix} \begin{pmatrix} \Delta y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} 0 & -a_{12} \\ 0 & (a_{12}-a_{22}) \end{pmatrix} \begin{pmatrix} \Delta y_{t-2} \\ x_{t-2} \end{pmatrix} + \begin{pmatrix} \epsilon_t \\ u_t - \epsilon_t \end{pmatrix}. \tag{18}$$

By inspection of (18), $\{x_t\}$ and $\{\Delta y_t\}$ are correlated both contemporaneously and dynamically (at leads and lags). Writing out the first equation of (18) and advancing the time index gives,

$$\Delta y_{t+1} = (\delta_1 + a_{12})x_t + ((a_{11}+a_{12})\Delta y_t - a_{12}x_{t-1} + \epsilon_{t+1}). \tag{19}$$

The short-horizon predictive regression regresses Δy_{t+1} on x_t. The resulting slope coefficient in such a regression is $\delta_1 + a_{12}$ and the regression error is $(a_{11}+a_{12})\Delta y_t - a_{12}x_{t-1}+\epsilon_{t+1}$ which is serially correlated and is also correlated with x_t. This latter correlation is innocuous, however, because the objective of the short-horizon regression is not to estimate this $\delta_1 + a_{12}$ per se, but it is 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.

3.1 Local-to-zero endogeneity

The VECM example motivates the presumption of endogeneity of the regressor in the short-horizon predictive regression. We will investigate local asymptotic power properties of short- and long-horizon regression in a less cumbersome representation given by

Assumption 2 *(Local endogeneity.) The observations obey*

$$\Delta y_{t+1} = b_1(T)x_t + e_{t+1}, \tag{20}$$

$$x_{t+1} = \rho_1(x)x_t + u_{t+1,1}, \tag{21}$$

where $\rho_1(x)x_t$ is the linear least squares projection of x_{t+1} onto x_t and $u_{t+1,1}$ is the associated projection error. The errors $(e_t, u_{t,1})$ are covariance stationary and have *the Wold representation*

$$\begin{pmatrix} e_t \\ u_{t,1} \end{pmatrix} = \Psi(L,T) \begin{pmatrix} m_{t-j} \\ n_{t-j} \end{pmatrix}, \tag{22}$$

where $\Psi(L,T) = \sum_{j=0}^{\infty} \begin{pmatrix} \psi_{11,j} & \psi_{12,j}(T) \\ \psi_{21,j}(T) & \psi_{22,j} \end{pmatrix}$, $\quad \psi_{12,j}(T) = \frac{\psi_{12,j}}{\sqrt{T}} \quad \psi_{21,j}(T) = \frac{\psi_{21,j}}{\sqrt{T}}$, $(m_t, n_t)' \overset{iid}{\sim} [0, \Sigma(T)]$, $\Sigma(T) = \begin{pmatrix} 1 & \rho_{mn}(T) \\ \rho_{mn}(T) & 1 \end{pmatrix}$, $\rho_{mn}(T) = \frac{\rho_{mn}}{\sqrt{T}}$. *The* $\psi_{rs,j}$, $r, s = 1, 2$, *and for all* $j > 0$ *and* ρ_{mn} *are fixed constants.*

Endogeneity is regulated through $\psi_{21,j}(T), \psi_{12,j}(T)$ and $\rho_{mn}(T)$ and is local-to-zero in the sense that $\mathrm{E}\left(e_t u_{t-j,1}\right) \to 0$ as $T \to \infty$ for all j. Representing e_{t+1} as a projection onto x_t plus a projection error, $e_{t+1} = c_1(T)x_t + a_{t+1}$ gives the short-horizon regression

$$\Delta y_{t+1} = \beta_1(T)x_t + \epsilon_{t+1,1}, \tag{23}$$

where $\epsilon_{t+1,1} = a_{t+1}$, $\quad \beta_1(T) = b_1(T) + c_1(T), b_1(T) = b_1/\sqrt{T}$, and $c_1(T) = c_1/\sqrt{T}$. The projection error $a_{t+1} = \epsilon_{t+1,1}$ is constructed to be uncorrelated with x_t but will in general exhibit local-to-zero dependence on x_{t-j} for $j \neq 0$.

To obtain the local-to-zero two-period horizon regression, we add together the short-horizon regression at $t+1$ and $t+2$,

$$y_{t+2} - y_t = \beta_1(T)\left[1 + \rho_1(x)\right] x_t + \left(a_{t+2} + a_{t+1} + u_{t+1,1}\right).$$

Note that due to the local-to-zero dependence of a_{t+2} on x_t, the long-horizon slope coefficient $\beta_2(T)$ is not $\beta_1(T)\left[1 + \rho_1(x)\right]$ as was the case when x_t is exogenous. In general, we write the long-horizon regression as

$$y_{t+k} - y_t = \beta_k(T)x_t + \epsilon_{t+k,k}, \tag{24}$$

where $\beta_k(T) = (b_k + c_k)/\sqrt{T}$, but b_k and c_k depend not only on b_1 and c_1 but on $\Sigma(T)$ and $\Psi(L,T)$. Under local-to-zero endogeneity, potentially large power advantages for long-horizon regression exist if $\beta_k(T)/\beta_1(T)$ grows (locally) at a faster rate with k than it does under exogeneity and this will be the case if $(b_k/b_1) > (c_k/c_1)$. Indeed, the relative power advantage is arbitrarily large as $b_1 + c_1 \to 0$. Because the endogeneity is local-to-zero, the asymptotic variance of the OLS estimator is obtained under the null hypothesis.

Determination of relative asymptotic power relies on Proposition 1 which continues to apply. We begin our investigation in this section with

Case IV. Let the observations be generated by

$$\begin{aligned} \Delta y_{t+1} &= b_1(T)x_t + e_{t+1}, && (25)\\ x_{t+1} &= \phi x_t + v_{t+1}, && (26)\\ \begin{pmatrix} e_t \\ v_t \end{pmatrix} &= \begin{pmatrix} a_{11} & a_{12}(T) \\ 0 & 0 \end{pmatrix} \begin{pmatrix} e_{t-1} \\ v_{t-1} \end{pmatrix} + \begin{pmatrix} m_t \\ n_t \end{pmatrix}, && (27) \end{aligned}$$

where

$$\begin{pmatrix} m_t \\ n_t \end{pmatrix} \overset{iid}{\sim} [0, \Sigma(T)], \qquad \Sigma(T) = \begin{pmatrix} 1 & \rho_{mn}(T) \\ \rho_{mn}(T) & 1 \end{pmatrix},$$
$$b_1(T) = \frac{b_1}{\sqrt{T}}, \quad a_{12}(T) = \frac{a_{12}}{\sqrt{T}}, \quad \rho_{mn}(T) = \frac{\rho_{mn}}{\sqrt{T}},$$

$\gamma = (\phi, b_1, a_{11}, a_{12}, \rho_{mn},)$, $|a_{11}| < 1$, $\rho_1(x) = \phi$, and $v_{t+1} = u_{t+1,1}$ is the projection error. In the appendix, we show that the short-horizon regression is $\Delta y_{t+1} = \beta_1(T)x_t + \epsilon_{t+1,1}$ where $\beta_1(T) = b_1(T) + c_1(T,\gamma)$, $b_1(T) = b_1/\sqrt{T}$, $c_1(T,\gamma) = \mathrm{E}(e_{t+1}x_t)/\mathrm{E}(x_t)^2 = c_1(\gamma)/\sqrt{T}$, and

$$c_1(\gamma) = \frac{a_{12} + a_{11}\rho_{mn}}{1 - a_{11}\phi}\left(1 - \phi^2\right). \tag{28}$$

The long-horizon regression is $y_{t+k} - y_t = \beta_k(T)x_t + \epsilon_{t+k,k}$ where

$$\beta_k(T) = b_1(T)\frac{1-\phi^k}{1-\phi} + c_1(T,\gamma)\frac{1-a_{11}^k}{1-a_{11}}. \tag{29}$$

Figure 4 gives a comparison of potential asymptotic power advantages for a particular specification of endogeneity by plotting the ratio of the slope coefficients

$$B_k(\gamma) = \frac{\beta_k(T)}{\beta_1(T)} = \frac{\frac{1-\phi^k}{1-\phi}b_1 + \frac{1-a_{11}^k}{1-a_{11}}c_1(\gamma)}{b_1 + c_1(\gamma)}, \tag{30}$$

where $c_1(\gamma)$ is given in (28) for $a_{11} = 0.5, a_{12} = -0.1, \rho_{mn} = -0.1, b_1 = 0.1, \phi = 0.9$. The figure also displays the ratio $B_k(\gamma) = (1-\phi^k)/(1-\phi)$ with $\phi = 0.9$ that obtains under exogeneity. The ratio of the slopes under endogeneity is increasing in k at a faster rate and lies above the ratio under exogeneity over the range of $k = 1, \ldots, 16$ considered.[7]

Table 6 reports values of the measure of relative asymptotic power

$$\theta(k;\gamma) = \lim_{p\to\infty}\left(B_k^2(\gamma)\frac{\sum_{j=-p}^{p}\phi^j g_{j,1}(e)}{\sum_{j=-p}^{p}\phi^j g_{j,k}(e)}\right), \tag{31}$$

where $B_k(\gamma)$ is given in (30), and $g_{j,k}(e)$ follows analogously from case I for a persistent regressor ($\phi = 0.95$), moderate asymptotic serial correlation for the regression error ($a_{11} = 0.5$) and varying degrees of endogeneity ($-0.9 \leq \rho_{mn} \leq 0.9$, $-0.9 \leq a_{12} \leq 0.3$). There is a diagonal band along (ρ_{mn}, a_{12}) pairs for which long-horizon regression tests have local asymptotic power advantages. For values of $\rho_{mn} \geq 0.1$, displayed in the top portion of the table, values of $\theta(k;\gamma) > 1$ are obtained either when the regressor and the regression error are negatively correlated at all leads and lags $\mathrm{E}(x_{t-j}e_t) < 0$ for all j and finite T (which occurs for relatively low a_{12} values), or when the error is negatively correlated with past values of x_t ($\mathrm{E}(x_{t-j}e_t) < 0$ for $j > 1$), and positively correlated ($\mathrm{E}(x_{t-j}e_t > 0$, for $j \leq 1$) with future x_t (relative large a_{12} values). In the second two panels ($-0.9 \leq \rho_{mn} \leq 0.1$ and $-0.8 \leq a_{12} \leq 0.3$), $\mathrm{E}(x_{t-j}e_t) < 0$ for all j.

Notice how the same values of $\theta(k;\gamma)$ recur for alternative values of γ. We find $\theta(13) = 890.50$ with $(\rho_{mn}, a_{12}) = (0.7, -0.9)$, $(0.5, -0.8)$, $(0.3, -0.0)$, $(7.1, -0.6)$, $(-0.1, 0.5)$, $(-0.3, -0.4)$, $(-0.5, -0.4)$, $(-0.8, -0.6)$ and $(-0.9, -0.5)$. Note also that given k, a_{11}, ϕ, the asymptotic variance $V(\hat{\beta}_k)$ is invariant to ρ_{mn} or a_{12}. The long-horizon regression test has local asymptotic power advantages in empirically relevant regions of the parameter space as well as in regions that do not conform well to our canonical empirical examples ($\rho_{mn} > 0$).

Table 7 reports the analogous local asymptotic power comparisons over $-0.9 \leq \rho_{mn} \leq 0.9$, $-0.9 \leq a_{12} \leq 0$, for a persistent regressor $\phi = 0.95$, and low asymptotic serial correlation in the regression error $a_{11} = 0.1$. Long-horizon regression has local asymptotic power advantages in a larger region of (ρ_{mn}, a_{12}) than obtained for $a_{11} = 0.5$. The largest long-horizon regression power gains occur in the region $\rho_{mn} < 0$ and $a_{12} < 0$ and $\mathrm{E}(x_{t-j}e_t) \leq 0$ for all j and finite T.

[7]The ratio of the long-horizon to short-horizon regression slope coefficients has a limiting value and is not forever increasing in k. Local power also is not forever increasing in k since $V(\hat{\beta}_k)$ is forever increasing in k.

Asymptotic serial correlation in $\{e_t\}$ is not necessary (nor, as we have seen sufficient) to give rise to asymptotic power advantages in the long-horizon regression test. In table 8, we set $a_{11} = 0$. Local power advantage are seen to accrue in the region $a_{12} < 0$.

Before concluding this section, we note that Campbell (1993) studied asymptotic power of long- and short-horizon regressions in a model with endogeneity in which the short-horizon regression error is serially uncorrelated and negatively correlated with the innovation to x_{t+1}. He showed that long-horizon regression tests had approximate slope advantages over short-horizon regression tests but did not find finite sample power advantages in his Monte Carlo experiments. We cannot make a direct comparison to his work because his approximate slope analysis was done under a fixed alternative. The closest approximation that we can make to Campbell's environment is by setting $a_{11} = a_{12} = 0$. But under local-to-zero endogeneity, when $a_{11} = 0$ neither the slope coefficients nor the asymptotic OLS variances depend on ρ_{mn} and this brings us back to case I with $\mu = 0$ which is a configuration under which long-horizon regression tests have no local power advantages over short-horizon regression tests.

3.2 Monte Carlo Experiments

While our primary focus lies in understanding whether there are conditions under which long-horizon regression tests have local asymptotic power advantages, it is the finite sample properties of the tests are of ultimate interest. A potential pitfall of local asymptotic analysis is that the effect of critical nuisance parameters (e.g., a_{12} and ρ_{mn}) are eliminated from the asymptotic variances, although not from evaluation of the ratio of the slope coefficients.

This section reports the results of a small Monte Carlo experiment that corresponds to case IV. The experiment should shed light on two questions. The first question is whether the power advantages of long-horizon regression predicted by the local asymptotic analysis is present in samples of small to moderate size. If so, then the second question is whether the small sample power advantages accrue in roughly the same region of the parameter space as predicted by the asymptotic analysis.

The DGP for our Monte Carlo experiment is modeled after case IV which exhibits endogeneity. We consider a sample size of $T = 100$ and performed 2000 replications for each experiment. The DGP under the null hypothesis is given by $b_1 = a_{12} = \rho_{mn}$. Under the alternative hypothesis, $b_1 = 0.1$ and a range of a_{12} and ρ_{mn} are considered. HAC standard errors are given by Newey–West (1987) with 20 lags.

Table 9 reports the maximum size-adjusted relative power of a one-sided long-horizon regression test at the 5 percent level over horizons 1 through 20. Under both the null and alternative hypotheses we set $a_{11} = 0.5, \phi = 0.95$. Finite sample power advantages are seen to accrue to long-horizon regression tests. The region of the parameter space that predicts local asymptotic power advantages for long-horizon regression tests is evidently a subset of the region that gives finite sample power advantage.

Table 10 reports the results of an analogous experiment with $a_{11} = 0.0$ under both the null and the alternative. Long-horizon regression tests continue to provide finite sample power advantages over short-horizon regressions under a linear data generating process and over a larger region of the parameter space than that predicted by the asymptotic analysis.

4 Conclusion

In this paper we provide asymptotic justification for employing long-horizon predictive regressions to test the null hypothesis of no predictability. Local asymptotic power advantages accrue to long-horizon regression tests whether the regressor is exogenous or endogenous although the assumption of exogeneity is often untenable in applied work. Under an endogenous regressor, we find that both local asymptotic power advantages as well as finite sample size-adjusted power advantages accrue to long-horizon regression tests in empirically relevant regions of the parameter space. The finite sample power advantages to long-horizon regression obtained in our Monte Carlo experiments are not the artifact of small sample bias or size distortion.

Our results lend support to empirical findings that equity returns and exchange rate returns are predictable but do not obviate the need to improve on the asymptotic distribution as an approximation to the exact sampling distribution in applied work, say via the bootstrap. This seems to be relevant for exchange rate prediction since the time-series available over the modern flexible exchange rate experience begins in 1973 but is perhaps less of an issue for equities since reasonably long time series on equity returns and dividend yields are available.

References

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

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

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.

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

Campbell, John Y. 1993. "Why Long Horizons: A Study of Power Against Persistent Alternatives," NBER Technical Working Paper No. 142.

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

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

Chinn, Menzie D. and Guy Meredith. "Long-Horizon Uncovered Interest Rate Parity," National Bureau of Economic Research Working Paper: 6797, November 1998

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.

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

Daniel, Kent, 2001. "The Power and Size of Mean Reversion Tests," *Journal of Empirical Finance, forthcoming.*

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

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

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

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

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

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

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.

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.

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

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).

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.

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.

Newey, W. and Kenneth D. West, (1987). "A Simple, Positive Semi-definite Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica* 55: pp. 703-8.

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

Rapach, David E. and Mark E. Wohar, 2001. "Testing the Monetary Model of Exchange Rate Determination: New Evidence from a Century of Data," *mimeo*, Department of Economics, University of Nebraska at Omaha.

Richardson, Matthew and James Stock, 1989, "Drawing Inferences from Statistics Based on Multi-Year Asset Returns," *Journal of Financial Economics,* 25, 323–48.

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

Valkanov, Rossen, 1999. "Long-Horizon Regressions: Theoretical Results and Applications to the Expected Returns/Dividend Yields and Fisher Effect Relations," *mimeo*, Anderson GSM, UCLA.

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

Notes: Stock return data are monthly observations from 1871.01 to 1995.12. Foreign exchange return data are quarterly observations from 1973.1 to 1997.3.

Table 2: Persistence of $\{x_t\}$ in the data.

		Dividend yield T=1500	Dividend yield T=288	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. AC is the autocorrelation.

Table 3: Local asymptotic power under case I. Maximized $\theta(k;\gamma)$ for selected values of $\gamma = (\phi, \mu)$.

	μ						
ϕ	-0.95	-0.85	-0.75	-0.65	-0.55	-0.45	-0.35
-0.81	3.362	1.069	1.000	1.000	1.000	1.000	1.000
-0.61	6.160	1.974	1.137	1.000	1.000	1.000	1.000
-0.41	8.198	2.652	1.543	1.067	1.000	1.000	1.000
-0.21	9.476	3.103	1.829	1.283	1.000	1.000	1.000
-0.01	9.994	3.328	1.995	1.423	1.106	1.000	1.000
0.19	9.752	3.326	2.041	1.490	1.184	1.000	1.000
0.39	8.750	3.097	1.967	1.482	1.213	1.042	1.000
0.59	6.988	2.642	1.773	1.400	1.193	1.062	1.000
0.79	4.466	1.960	1.459	1.244	1.125	1.049	1.000

Note: Values of $\theta(k;\gamma) > 1$ obtained only for $k = 2$. Values of $\theta(k;\gamma) = 1$ occur when $k = 1$, in which case long horizon regression tests have no asymptotic power advantage.

Table 4: Local asymptotic power for case II. Selected $\theta(k;\gamma)$, $\gamma = (\phi, \mu_1, \mu_2)$.

$\theta(k;\gamma)$	k	ϕ	μ_1	μ_2	$\mathrm{VR}_e[10]$	$\rho_1(e)$	$\rho_2(e)$
1.099	19	0.980	1.880	-0.960	5.145	0.959	0.843
1.197	14	0.880	1.800	-0.960	2.406	0.918	0.693
1.169	16	0.980	1.840	-0.960	3.565	0.939	0.767
1.197	14	0.880	1.800	-0.960	2.406	0.918	0.693
1.216	14	0.980	1.800	-0.960	2.406	0.918	0.693
1.197	14	0.880	1.800	-0.960	2.406	0.918	0.693
1.216	14	0.980	1.800	-0.960	2.406	0.918	0.693
88.997	15	0.880	-1.920	-0.960	0.065	-0.980	0.921
57.359	2	0.780	-1.880	-0.920	0.059	-0.979	0.921
45.520	2	0.680	-1.800	-0.840	0.050	-0.978	0.921
29.953	2	0.480	-1.600	-0.640	0.036	-0.976	0.921
20.967	7	0.880	-1.760	-0.920	0.076	-0.917	0.693
11.011	9	0.980	-1.840	-0.920	0.083	-0.958	0.843
10.107	5	0.680	-1.640	-0.840	0.063	-0.891	0.622

Notes: $\mathrm{VR}_e[10]$ is the variance ratio statistic at horizon 10 for the error term, $\{e_t\}$.

Table 5: Local asymptotic power for case III. Selected $\theta(k;\gamma)$, $\gamma = (\phi_1, \phi_2, \mu)$.

$\theta(k,\gamma)$	k	μ	ϕ_1	ϕ_2	$VR_x[10]$	$\rho_1(x)$	$\rho_2(x)$
96.990	2	-0.920	0.000	0.960	4.689	0.000	0.960
72.375	2	-0.920	0.000	0.920	4.397	0.000	0.920
44.031	2	-0.920	-0.040	0.840	2.734	-0.250	0.850
25.291	2	-0.920	-0.120	0.800	1.230	-0.600	0.872
15.611	2	-0.920	-0.040	0.480	1.898	-0.077	0.483
9.519	2	-0.920	0.320	0.360	4.078	0.500	0.520
5.640	2	-0.920	0.480	0.080	3.033	0.522	0.330
3.742	2	-0.720	0.160	0.600	4.601	0.400	0.664
1.169	10	-0.920	-1.720	-0.800	0.067	-0.956	0.844
1.407	8	-0.920	-1.600	-0.760	0.070	-0.909	0.695
1.588	6	-0.920	-1.440	-0.640	0.071	-0.878	0.624
1.002	3	-0.220	-0.120	0.800	1.230	-0.600	0.872
1.010	3	-0.220	0.240	0.720	8.639	0.857	0.926
2.385	2	-0.920	0.600	0.360	8.878	0.938	0.923
1.142	2	-0.620	0.600	0.360	8.878	0.938	0.923
1.097	2	-0.420	0.360	0.600	8.799	0.900	0.924

Notes: $VR_x[10]$ is the variance ratio statistic at horizon 10 for the regressor, $\{x_t\}$.

Table 6: Local asymptotic power for case IV. $\theta(k;\gamma)$ with $a_{11} = 0.5, b_1 = 0.1, \phi = 0.95$. Optimal horizon in parentheses.

	a_{12}									
ρ_{mn}	-0.9	-0.8	-0.7	-0.6	-0.5	-0.4	-0.3	-0.2	-0.1	0.0
0.9	16.805	4.133	1.996	1.274	1.005	1	1	1	-0.1	0.0
	(11)	(9)	(8)	(5)	(2)	(1)	(1)	(1)	(1)	(1)
0.8	84.361	7.24	2.741	1.554	1.098	1	1	1	(1)	(1)
	(12)	(10)	(9)	(7)	(4)	(1)	(1)	(1)	(1)	(1)
0.7	890.50	16.805	4.133	1.996	1.274	1.005	1	1	1	1
	(13)	(11)	(9)	(8)	(5)	(2)	(1)	(1)	(1)	(1)
0.6	29.813	84.361	7.24	2.741	1.554	1.098	1	1	1	1
	(13)	(12)	(10)	(9)	(7)	(4)	(1)	(1)	(1)	(1)
0.5	8.647	890.50	16.805	4.133	1.996	1.274	1.005	1	1	1
	(14)	(13)	(11)	(9)	(8)	(5)	(2)	(1)	(1)	(1)
0.4	3.929	29.813	84.361	7.24	2.741	1.554	1.098	1	1	1
	(15)	(13)	(12)	(10)	(9)	(7)	(4)	(1)	(1)	(1)
0.3	2.188	8.647	890.50	16.805	4.133	1.996	1.274	1.005	1	1
	(15)	(14)	(13)	(11)	(9)	(8)	(5)	(2)	(1)	(1)
0.2	1.367	3.929	29.813	84.361	7.24	2.741	1.554	1.098	1	1
	(16)	(15)	(13)	(12)	(10)	(9)	(7)	(4)	(1)	(1)

	a_{12}									
ρ_{mn}	-0.8	-0.7	-0.6	-0.5	-0.4	-0.3	-0.2	-0.1	0	0.1
0.1	2.188	8.647	890.50	16.805	4.133	1.996	1.274	1.005	1	1
	(15)	(14)	(13)	(11)	(9)	(8)	(5)	(2)	(1)	(1)
0	1.367	3.929	29.813	84.361	7.24	2.741	1.554	1.098	1	1
	(16)	(15)	(13)	(12)	(10)	(9)	(7)	(4)	(1)	(1)
-0.1	1	2.188	8.647	890.50	16.805	4.133	1.996	1.274	1.005	1
	(1)	(15)	(14)	(13)	(11)	(9)	(8)	(5)	(2)	(1)
-0.2	1	1.367	3.929	29.813	84.361	7.24	2.741	1.554	1.098	1
	(1)	(16)	(15)	(13)	(12)	(10)	(9)	(7)	(4)	(1)
-0.3	1	1	2.188	8.647	890.50	16.805	4.133	1.996	1.274	1.005
	(1)	(1)	(15)	(14)	(13)	(11)	(9)	(8)	(5)	(2)
-0.4	1	1	1.367	3.929	29.813	84.361	7.24	2.741	1.554	1.098
	(1)	(1)	(16)	(15)	(13)	(12)	(10)	(9)	(7)	(4)

	a_{12}									
ρ_{mn}	-0.6	-0.5	-0.4	-0.3	-0.2	-0.1	0	0.1	0.2	0.3
-0.5	2.188	8.647	890.50	16.805	4.133	1.996	1.274	1.005	1	1
	(15)	(14)	(13)	(11)	(9)	(8)	(5)	(2)	(1)	(1)
-0.6	1.367	3.929	29.813	84.361	7.24	2.741	1.554	1.098	1	1
	(16)	(15)	(13)	(12)	(10)	(9)	(7)	(4)	(1)	(1)
-0.7	1	2.188	8.647	890.50	16.805	4.133	1.996	1.274	1.005	1
	(1)	(15)	(14)	(13)	(11)	(9)	(8)	(5)	(2)	(1)
-0.8	1	1.367	3.929	29.813	84.361	7.24	2.741	1.554	1.098	1
	(1)	(16)	(15)	(13)	(12)	(10)	(9)	(7)	(4)	(1)
-0.9	1	1	2.188	8.647	890.50	16.805	4.133	1.996	1.274	1.005
	(1)	(1)	(15)	(14)	(13)	(11)	(9)	(8)	(5)	(2)

Table 7: Local asymptotic power for case IV. $\theta(k;\gamma)$ with $a_{11} = 0.10, b_1 = 0.1, \phi = 0.95$. Optimal horizon in parentheses.

	a_{12}									
ρ_{mn}	-0.9	-0.8	-0.7	-0.6	-0.5	-0.4	-0.3	-0.2	-0.1	0.0
0.9	36.113	11.042	5.425	3.286	2.251	1.672	1.32	1.1	1	1
	(8)	(8)	(7)	(6)	(6)	(5)	(4)	(3)	(1)	(1)
0.8	42.918	12.079	5.757	3.432	2.329	1.718	1.348	1.118	1	1
	(8)	(8)	(7)	(7)	(6)	(5)	(4)	(3)	(1)	(1)
0.7	51.886	13.275	6.122	3.591	2.411	1.765	1.378	1.136	1	1
	(8)	(8)	(7)	(7)	(6)	(5)	(4)	(3)	(1)	(1)
0.6	64.05	14.663	6.524	3.761	2.499	1.815	1.408	1.155	1.009	1
	(8)	(8)	(7)	(7)	(6)	(5)	(4)	(3)	(2)	(1)
0.5	81.144	16.288	6.97	3.945	2.591	1.868	1.44	1.174	1.019	1
	(9)	(8)	(7)	(7)	(6)	(5)	(4)	(3)	(2)	(1)
0.4	106.282	18.207	7.465	4.144	2.689	1.923	1.473	1.194	1.03	1
	(9)	(8)	(7)	(7)	(6)	(5)	(5)	(4)	(2)	(1)
0.3	145.39	20.497	8.018	4.359	2.794	1.981	1.509	1.218	1.041	1
	(9)	(8)	(7)	(7)	(6)	(5)	(5)	(4)	(2)	(1)
0.2	211.17	23.26	8.639	4.592	2.905	2.042	1.548	1.242	1.052	1
	(9)	(8)	(8)	(7)	(6)	(5)	(5)	(4)	(2)	(1)
0.1	334.915	26.637	9.342	4.846	3.024	2.108	1.587	1.267	1.067	1
	(9)	(8)	(8)	(7)	(6)	(6)	(5)	(4)	(3)	(1)
0	612.179	30.827	10.137	5.123	3.151	2.178	1.629	1.293	1.083	1
	(9)	(8)	(8)	(7)	(6)	(6)	(5)	(4)	(3)	(1)
-0.1	1463.86	36.113	11.042	5.425	3.286	2.251	1.672	1.32	1.1	1
	(9)	(8)	(8)	(7)	(6)	(6)	(5)	(4)	(3)	(1)
-0.2	7179.029	42.918	12.079	5.757	3.432	2.329	1.718	1.348	1.118	1
	(9)	(8)	(8)	(7)	(7)	(6)	(5)	(4)	(3)	(1)
-0.3	149456.8	51.886	13.275	6.122	3.591	2.411	1.765	1.378	1.136	1
	(9)	(8)	(8)	(7)	(7)	(6)	(5)	(4)	(3)	(1)
-0.4	3447.781	64.05	14.663	6.524	3.761	2.499	1.815	1.408	1.155	1.009
	(9)	(8)	(8)	(7)	(7)	(6)	(5)	(4)	(3)	(2)
-0.5	1005.906	81.144	16.288	6.97	3.945	2.591	1.868	1.44	1.174	1.019
	(9)	(9)	(8)	(7)	(7)	(6)	(5)	(4)	(3)	(2)
-0.6	470.857	106.282	18.207	7.465	4.144	2.689	1.923	1.473	1.194	1.03
	(9)	(9)	(8)	(7)	(7)	(6)	(5)	(5)	(4)	(2)
-0.7	271.451	145.39	20.497	8.018	4.359	2.794	1.981	1.509	1.218	1.041
	(9)	(9)	(8)	(7)	(7)	(6)	(5)	(5)	(4)	(2)
-0.8	176.073	211.17	23.26	8.639	4.592	2.905	2.042	1.548	1.242	1.052
	(9)	(9)	(8)	(8)	(7)	(6)	(5)	(5)	(4)	(2)
-0.9	123.222	334.915	26.637	9.342	4.846	3.024	2.108	1.587	1.267	1.067
	(9)	(9)	(8)	(8)	(7)	(6)	(6)	(5)	(4)	(3)

Table 8: Local asymptotic power for case IV. $\theta(k;\gamma)$, $a_{11} = 0.0, b_1 = 0.1, \phi = 0.95$. Optimal horizon in parentheses.

a_{12}	$\theta(k;\gamma)$	k	a_{12}	$\theta(k;\gamma)$	k
-0.9	40.636	8	0.1	1.000	1
-0.8	13.037	7	0.2	1.000	1
-0.7	6.494	7	0.3	1.000	1
-0.6	3.948	6	0.4	1.000	1
-0.5	2.697	6	0.5	1.000	1
-0.4	1.991	5	0.6	1.000	1
-0.3	1.553	4	0.7	1.000	1
-0.2	1.269	4	0.8	1.000	1
-0.1	1.083	3	0.9	1.000	1
0.0	1.000	1			

Note: $\theta(k,\gamma)$ is invariant to ρ_{mn} when $a_{11} = 0$.

Table 9: Monte Carlo experiment for case IV. Relative size-adjusted power $a_{11} = 0.5, b_1 = 0.1, \phi = 0.95$, optimal horizon in parentheses. $T = 100$.

	a_{12}									
ρ_{mn}	-0.9	-0.8	-0.7	-0.6	-0.5	-0.4	-0.3	-0.2	-0.1	0.0
0.9	5.4	2.8	1.8	1.3	1.1	1.0	1.0	1.0	1.0	1.0
	(12)	(9)	(6)	(4)	(2)	(1)	(1)	(1)	(1)	(1)
0.8	6.4	3.5	2.1	1.4	1.1	1.0	1.0	1.0	1.0	1.0
	(10)	(10)	(10)	(5)	(3)	(1)	(1)	(1)	(1)	(1)
0.7	8.9	5.1	2.8	1.7	1.3	1.0	1.0	1.0	1.0	1.0
	(12)	(10)	(10)	(5)	(4)	(2)	(1)	(1)	(1)	(1)
0.6	10.3	6.2	3.5	2.1	1.4	1.0	1.0	1.0	1.0	1.0
	(12)	(11)	(9)	(9)	(7)	(4)	(1)	(1)	(1)	(1)
0.5	13.2	7.7	4.2	2.6	1.5	1.1	1.0	1.0	1.0	1.0
	(12)	(12)	(12)	(9)	(7)	(7)	(1)	(1)	(1)	(1)
0.4	20.9	9.8	5.7	3.0	1.9	1.3	1.0	1.0	1.0	1.0
	(14)	(14)	(13)	(11)	(9)	(6)	(1)	(1)	(1)	(1)
0.3	30.9	14.8	7.9	4.2	2.4	1.5	1.0	1.0	1.0	1.0
	(14)	(13)	(13)	(13)	(13)	(9)	(3)	(1)	(1)	(1)
0.2	67.5	22.9	10.5	5.6	3.0	1.8	1.3	1.0	1.0	1.0
	(14)	(13)	(14)	(13)	(13)	(9)	(6)	(1)	(1)	(1)
0.1	80.0	41.3	15.6	8.1	3.9	2.3	1.5	1.1	1.0	1.0
	(13)	(13)	(13)	(13)	(13)	(10)	(6)	(4)	(1)	(1)
0	109.5	74.7	22.9	11.2	6.3	3.0	1.9	1.3	1.0	1.0
	(13)	(13)	(13)	(13)	(13)	(13)	(6)	(6)	(1)	(1)
-0.1	188.0	62.7	39.2	16.6	8.4	4.2	2.3	1.5	1.1	1.0
	(19)	(13)	(13)	(11)	(11)	(11)	(7)	(6)	(6)	(1)
-0.2	410.0	170.0	57.3	25.4	12.9	6.6	3.1	1.9	1.3	1.0
	(16)	(16)	(13)	(16)	(12)	(8)	(10)	(8)	(6)	(2)
-0.3	372.5	382.5	156.0	53.7	21.3	10.4	4.9	2.5	1.6	1.1
	(17)	(17)	(12)	(12)	(12)	(12)	(10)	(6)	(6)	(2)
-0.4	335.0	347.5	357.5	72.5	36.8	15.6	7.5	3.4	2.0	1.3
	(17)	(18)	(14)	(14)	(14)	(11)	(10)	(10)	(6)	(6)
-0.5	305.0	327.5	340.0	340.0	70.0	29.0	11.1	5.2	2.6	1.5
	(19)	(19)	(19)	(19)	(14)	(10)	(10)	(10)	(1)	(6)
-0.6	272.5	290.0	307.5	327.5	134.0	45.0	18.3	7.5	3.2	1.8
	(17)	(19)	(20)	(20)	(19)	(12)	(12)	(12)	(9)	(5)
-0.7	235.0	245.0	260.0	277.5	300.0	126.0	43.0	14.3	5.5	2.9
	(20)	(20)	(20)	(20)	(20)	(20)	(12)	(12)	(11)	(6)
-0.8	190.0	207.5	222.5	240.0	262.5	295.0	126.0	44.7	11.5	4.5
	(18)	(19)	(18)	(18)	(19)	(19)	(19)	(11)	(11)	(7)
-0.9	160.0	172.5	197.5	212.5	237.5	265.0	302.5	136.0	39.0	9.6
	(20)	(20)	(20)	(20)	(20)	(20)	(20)	(20)	(20)	(20)

Table 10: Monte Carlo experiment for case IV. Relative size-adjusted power $a_{11} = 0.1, b_1 = 0.1, \phi = 0.95$, optimal horizon in parentheses. $T = 100$.

	a_{12}									
ρ_{mn}	-0.9	-0.8	-0.7	-0.6	-0.5	-0.4	-0.3	-0.2	-0.1	0.0
0.9	8.1	5.5	3.9	2.9	2.2	1.7	1.3	1.1	1.0	1.0
	(9)	(9)	(6)	(6)	(5)	(5)	(4)	(3)	(2)	(1)
0.8	8.5	5.6	3.9	2.8	2.2	1.7	1.3	1.1	1.0	1.0
	(10)	(9)	(8)	(5)	(5)	(4)	(4)	(3)	(1)	(1)
0.7	10.0	6.6	4.4	3.1	2.2	1.7	1.4	1.1	1.0	1.0
	(9)	(9)	(7)	(7)	(5)	(4)	(4)	(3)	(2)	(1)
0.6	10.6	7.0	4.8	3.1	2.3	1.7	1.4	1.1	1.0	1.0
	(9)	(9)	(7)	(7)	(7)	(5)	(4)	(3)	(1)	(1)
0.5	11.5	7.3	5.0	3.3	2.3	1.7	1.3	1.1	1.0	1.0
	(7)	(7)	(7)	(7)	(7)	(7)	(4)	(2)	(1)	(1)
0.4	12.3	8.5	5.2	3.5	2.4	1.8	1.4	1.1	1.0	1.0
	(9)	(6)	(6)	(6)	(6)	(6)	(4)	(2)	(1)	(1)
0.3	14.3	9.2	5.7	3.9	2.5	1.8	1.4	1.1	1.0	1.0
	(9)	(9)	(7)	(7)	(6)	(4)	(3)	(3)	(2)	(1)
0.2	18.4	10.9	6.6	4.3	2.8	1.9	1.5	1.2	1.0	1.0
	(9)	(9)	(6)	(6)	(6)	(6)	(3)	(3)	(2)	(1)
0.1	22.7	11.8	7.6	4.7	3.0	2.1	1.6	1.2	1.0	1.0
	(10)	(7)	(6)	(6)	(6)	(6)	(4)	(4)	(2)	(1)
0	29.0	15.9	9.0	5.3	3.3	2.2	1.6	1.2	1.0	1.0
	(11)	(6)	(6)	(6)	(6)	(6)	(5)	(5)	(2)	(1)
-0.1	46.2	18.5	10.0	5.9	3.5	2.3	1.7	1.3	1.1	1.0
	(7)	(7)	(6)	(6)	(6)	(6)	(6)	(4)	(2)	(1)
-0.2	61.4	26.6	12.9	6.6	3.9	2.5	1.8	1.3	1.1	1.0
	(8)	(8)	(8)	(6)	(6)	(6)	(6)	(4)	(2)	(1)
-0.3	76.4	36.7	17.1	7.6	4.4	2.7	1.9	1.4	1.1	1.0
	(10)	(10)	(6)	(6)	(6)	(6)	(6)	(6)	(2)	(1)
-0.4	86.5	40.7	20.5	8.6	4.6	2.9	2.0	1.4	1.1	1.0
	(10)	(10)	(6)	(6)	(6)	(6)	(6)	(6)	(3)	(1)
-0.5	81.5	57.0	26.1	10.8	5.0	3.0	1.9	1.4	1.1	1.0
	(10)	(10)	(10)	(10)	(6)	(6)	(6)	(3)	(3)	(1)
-0.6	143.5	76.5	27.3	10.9	5.0	3.0	1.9	1.4	1.1	1.0
	(10)	(10)	(10)	(10)	(5)	(5)	(5)	(5)	(2)	(1)
-0.7	126.0	91.7	42.3	15.0	6.0	3.7	2.1	1.5	1.1	1.0
	(12)	(12)	(10)	(6)	(6)	(6)	(6)	(4)	(4)	(2)
-0.8	239.0	128.5	40.6	18.4	7.2	3.7	2.1	1.5	1.2	1.0
	(11)	(11)	(11)	(11)	(7)	(6)	(4)	(4)	(4)	(2)
-0.9	205.0	114.0	51.6	22.1	8.6	3.7	2.1	1.4	1.1	1.0
	(8)	(8)	(8)	(8)	(8)	(8)	(8)	(3)	(3)	(1)

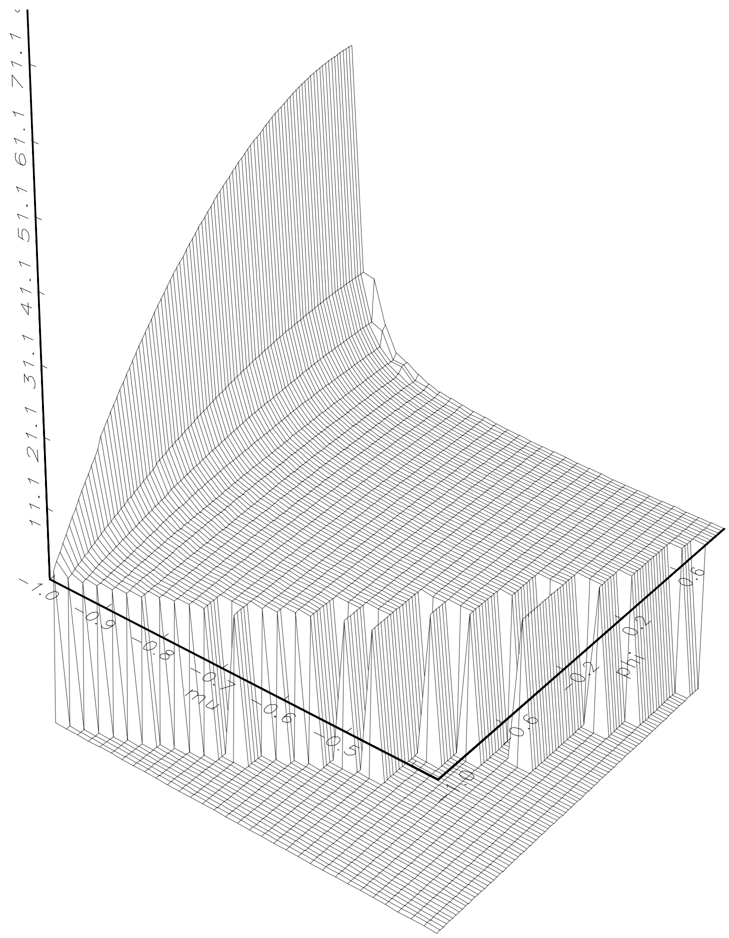

Figure 1: Relative asymptotic power for Berben's case.

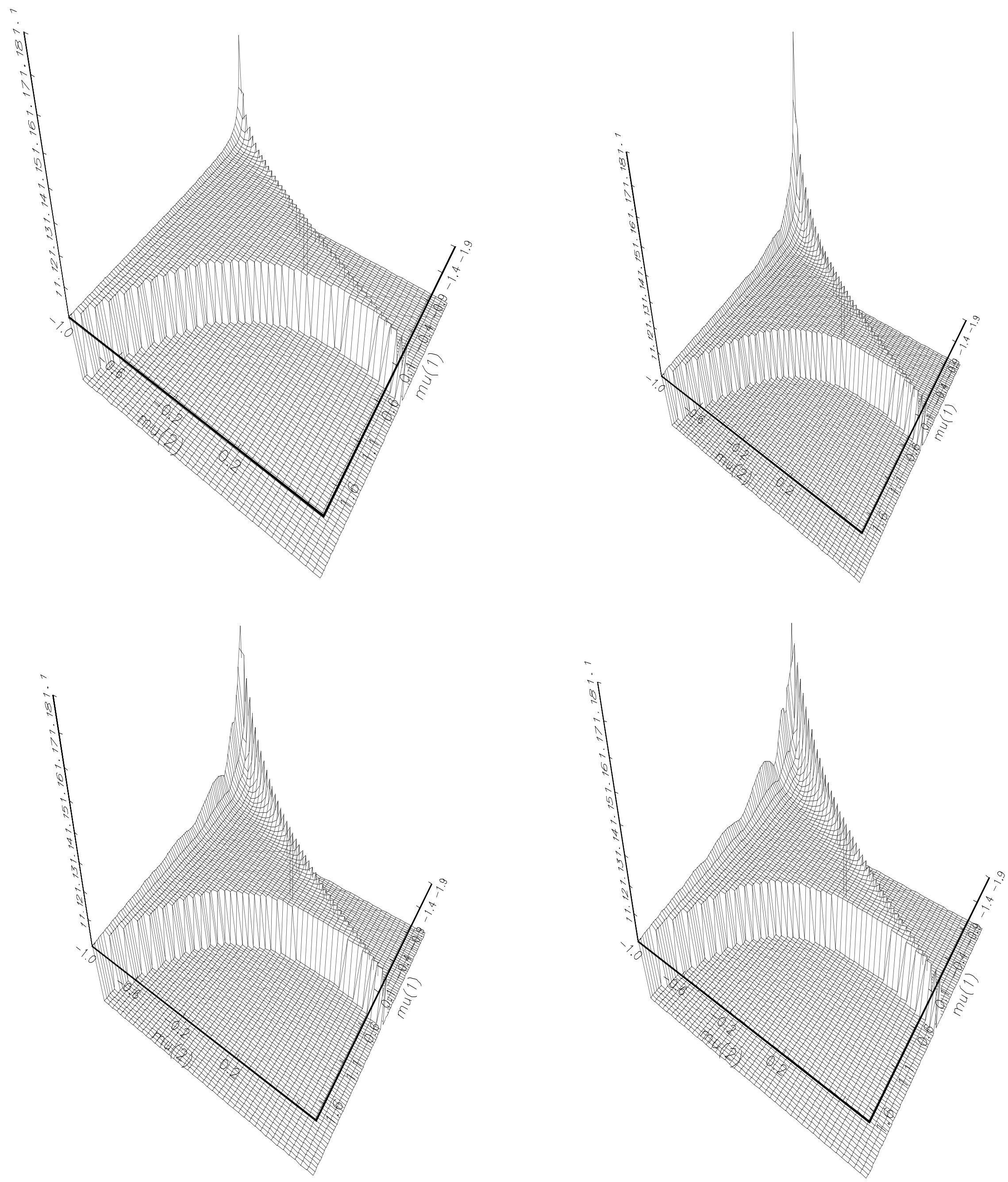

Figure 2: Relative asymptotic power $x_t = \phi x_{t-1} + v_t, e_t = \mu_1 e_{t-1} + \mu_2 e_{t-2} + m_t$. Clockwise from upper left, $\phi = 0.98, 0.88, 0.68, 0.78$

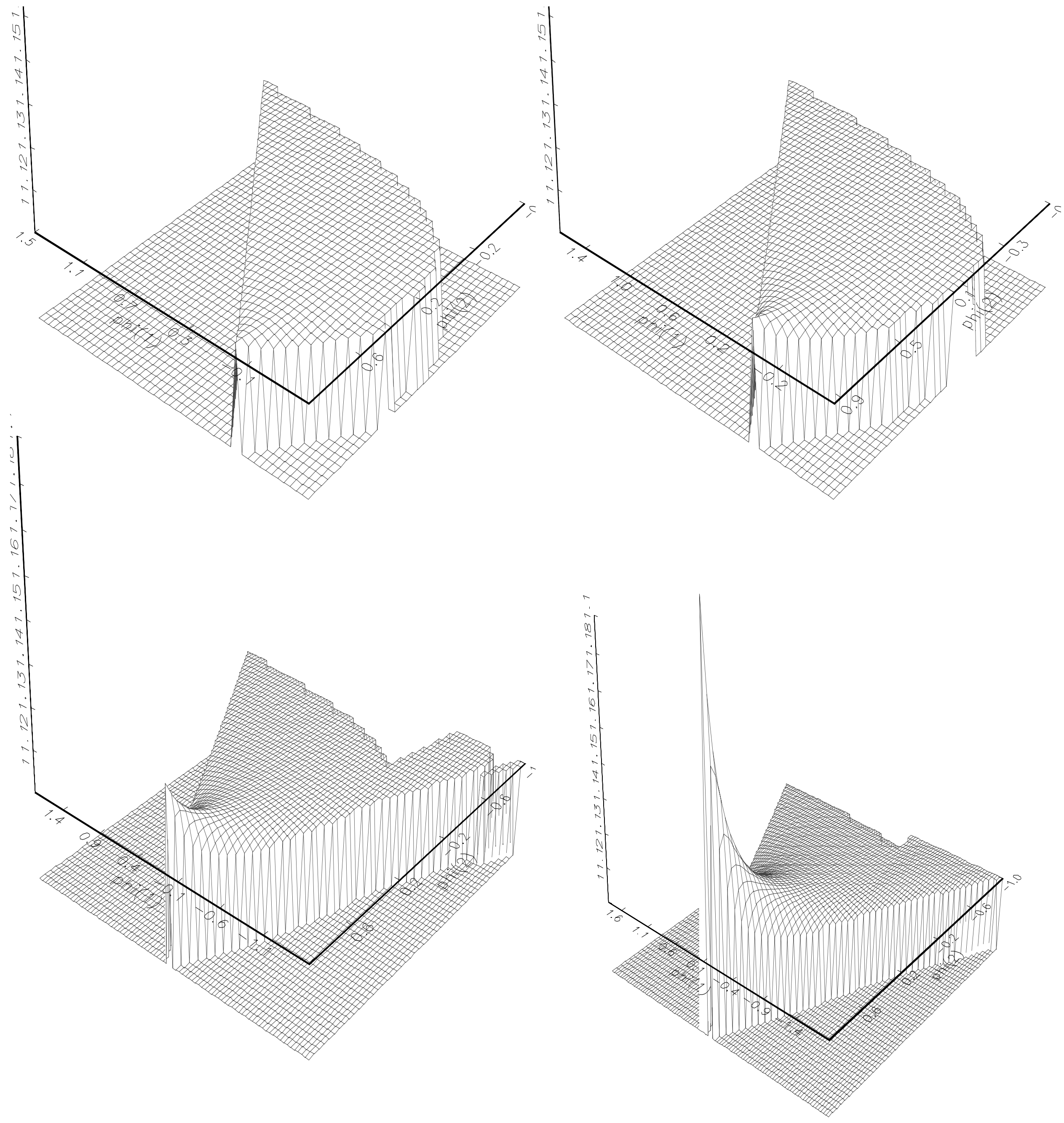

Figure 3: Relative asymptotic power $x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + v_t, e_t = \mu e_{t-1} + m_t$. Clockwise from upper left, $\mu = -0.62, -0.72, -0.92, -0.82$.

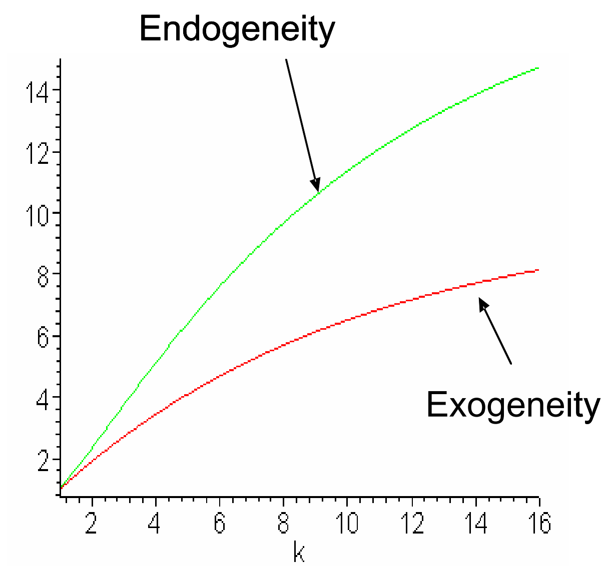

Figure 4: $\beta_k(T)/\beta_1(T)$ under endogeneity and exogeneity. $a_{11} = 0.5, a_{12} = -0.1, \rho_{mn} = -0.1, b_1 = 0.1, \phi = 0.90$.

Appendix

Derivation of eq.(3). Begin with Assumption 1, use the projection representation for x_{t+1} and advance the time subscript in (2) to obtain $\Delta y_{t+2} = \beta_1(T)x_{t+1} + e_{t+2} = \beta_1(T)\rho_1(x)x_t + e_{t+1} + \beta_1(T)u_{t+1,1}$. Add this result to (2) get for $k = 2$,

$$y_{t+2} - y_t = \beta_1(T)\left[1 + \rho_1(x)\right] x_t + (e_{t+1} + e_{t+2} + \beta_1(T)u_{t+1,1})$$

Continuing on for arbitrary $k > 1$ gives (3).

Derivation of eqs. (5)–(7) The asymptotic variance of $\hat{\beta}_k$ is $V(\hat{\beta}_k) = W_k/(\mathrm{E}(x_t)^2)^2 = W_k/C_0^2(x)$, where $W_k = \Omega_{0,k} + 2\sum_{j=1}^{\infty}\Omega_{j,k}$, and $\Omega_{jk} = \lim_{T\to\infty} \mathrm{E}\left(x_{t-k}x_{t-k-j}\epsilon_{t,k}\epsilon_{t-k-j,k}\right)$. Since $\epsilon_{t,k}$ is asymptotically independent of $u_{t+j,j}$, it follows that

$$\begin{aligned}
\Omega_{jk} &= \lim_{T\to\infty} \mathrm{E}\left(x_{t-k}x_{t-k-j}\epsilon_{t,k}\epsilon_{t-k-j,k}\right) \\
&= \mathrm{E}\left(x_{t-k}x_{t-k-j}\right) \mathrm{E}\left(\sum_{s=0}^{k-1} e_{t-j} \sum_{s=0}^{k-1} e_{t-j-s}\right) \\
&= C_j(x)G_{j,k}(e)
\end{aligned}$$

where

$$G_{j,k}(e) \equiv \mathrm{E}\left(\sum_{s=0}^{k-1} e_{t-j} \sum_{s=0}^{k-1} e_{t-j-s}\right) = kC_j(e) + \sum_{s=1}^{k-1}(k-s)\left[C_{j-s}(e) + C_{j+s}(e)\right]$$

Derivation of (28). Let $a_{11}(L) \equiv (1 - a_{11}L)^{-1} = \sum_{j=0}^{\infty} a_{11}^j L^j$ and $\phi(L) \equiv (1-\phi L)^{-1} = \sum_{j=0}^{\infty}\phi^j L^j$. From (**??**) we obtain,

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

It follows that

$$\begin{aligned}
\mathrm{E}(e_{t+1}x_t) &= \mathrm{E}\left(\left[a_{12}(T)a_{11}(L)v_t + a_{11}(L)m_{t+1}\right]\left[\phi(L)v_t\right]\right) \\
&= \frac{a_{12}(T) + a_{11}\rho_{mn}(T)}{1 - a_{11}\phi} = \frac{a_{12} + a_{11}\rho_{mn}}{(1 - a_{11}\phi)\sqrt{T}} \qquad \text{(A.32)}
\end{aligned}$$

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

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

Derivation of (30) Note that for $k = 2$,

$$\begin{aligned} y_{t+2} - y_t &= b_2(T)(1+\phi)x_t + a_{t+2,2} \\ a_{t+2,2} &= e_{t+2} + e_{t+1} + \beta_1(T)v_{t+1} \end{aligned}$$

Therefore, $b_2(T) = b_1(T)(1+\phi) = \frac{b_1}{\sqrt{T}}(1+\phi)$. As before, we can write

$$\begin{aligned} e_{t+2} &= a_{12}(T)v_{t+1} + a_{12}(T)a_{11}a_{11}(L)v_t + m_{t+2} + a_{11}m_{t+1} + a_{11}^2 a_{11}(L)m_t \\ x_t &= \phi(L)v_t \end{aligned}$$

from which we obtain,

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

It follows that

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

Continuing on in this way, it can be seen that for any k, $b_k(T) = b_1(T)\left(\sum_{j=0}^{k-1}\phi^j\right) = b_1(T)\left(\frac{1-\phi^k}{1-\phi}\right)$, and

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

Finally, divide by $\mathrm{E}\left(x_t^2\right) = C_0(x) = \left(1-\phi^2\right)^{-1}$ to get

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