The Use of Predictive Regressions at Alternative Horizons in Finance and Economics

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

Donggyu Sul
Department of Economics
University of Auckland

May 12, 2005

Abstract

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

Keywords: Predictive regression, Long horizons, Stock returns, Small sample bias, Local-to-unity, Asymptotic power

JEL Classification: C12; C22; G12; E47

1 Introduction

Let $r_t \sim I(0)$ be the return on an asset or a portfolio of assets from time $t-1$ to t and x_t be a persistent but $I(0)$ 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 its fundamental value.[1] A test of return predictability can be conducted by regressing r_{t+1} on x_t and performing a t-test on the slope coefficient. Empirical research in finance 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 $H_0 : \beta_k = 0$, where the t-statistic is constructed with a heteroskedastic and autocorrelation consistent (HAC) standard error. It is typically found that OLS slope estimates, asymptotic t-ratios, and $R^2 s$ increase over a range of horizons $k > 1$. Because the asymptotic p-values of the test of no predictability decline over this range of k, the analyst may conclude that the long-horizon test rejects the null hypothesis when the short-horizon test does not. Considering that the long-horizon regression is built by aggregation of intervening short-horizon regressions, the underlying basis for these results are not fully understood. As stated by Campbell et. al. (1997), "An important unresolved question is whether there are circumstances under which long-horizon regressions have greater power to detect deviations from the null hypothesis than do short-horizon regressions."

In this paper, we address the power question posed by Campbell et. al. We show that long-horizon regression tests can have asymptotic power advantages over short-horizon tests when the regressor is endogenous and when the nature of the endogeneity occurs in an empirically relevant and nontrivial region of the parameter space. The endogeneity that we address does not arise in the sense of misspecification of a structural model because the predictive regressions we study are employed as projections of the future return on x_t to estimate functions of the underlying moments of the distribution between $\{r_t\}$ and $\{x_t\}$. The asymptotic power analysis is conducted under the assumption that the regressor has a local-to-unity dominant autoregressive root which is motivated by the high persistence of the predictive variables often used in empirical work. We approach testing with the sup–t^2 test, which is a variant of the sup-bound test discussed by Cavanaugh et. al. (1995) that is asymptotically valid and free from nuisance parameter dependencies.

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

While these results provide asymptotic theoretical justification for using long-horizon regressions, the sup-t^2 test suffers from moderate small-sample size distortion that is induced by small sample OLS bias. Implementing a bias adjustment at long horizons may be problematic when the underlying DGP is unknown. To obtain a test that is better sized and which retains the power advantages for long-horizon tests, we propose that a recursive moving-block Bartlett correction be applied to the test statistics. To address small-sample OLS bias, we propose a related recursive moving-block jackknife estimator.

Previous research on the econometrics of predictive regressions include Campbell (2001) who assumes an AR(1) regressor $\{x_t\}$ and a serially uncorrelated short-horizon regression error. Using the concept of approximate slope to measure its asymptotic power, he found that long-horizon regressions had approximate slope advantages over short-horizon regressions but his Monte Carlo experiments did not reveal systematic power advantages for long-horizon regressions in finite samples. Berben (2000) reported asymptotic power advantages for long-horizon regression when the exogenous predictor and the short-horizon regression error follow AR(1) processes. 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), 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 OLS bias and Campbell and Yogo (2002) study point optimal tests in the short-horizon predictive regression. Kilian and Taylor (2003) examine small-sample properties under nonlinearity of the data generation process (DGP) 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 in the sense that $k/T \to 0$ as $T \to \infty$. An alternative long-horizon regression employed in the literature regresses the future k-period return (from t to $t+k$) on the past k-period return (from $t-k$ to t) [Fama and French (1988b)]. In this alternative long-horizon regression, the return horizon k can be large relative to the size of the sample T. Richardson and Stock (1989) develop an alternative asymptotic theory where $k \to \infty$ and $T \to \infty$ but $k/T \to \delta \in (0,1)$ and show that the test statistics converge to functions of Brownian motions, Daniel (2001) studies optimal tests of this kind, and Kim et. al. (1991) study the OLS sampling distribution with the bootstrap and randomization techniques. 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 presents the local asymptotic power analysis and the small sample properties of the predictive regression tests. Section 3 presents the recursive moving-block jackknife estimator for attenuating the

OLS bias and discusses the Bartlett correction for the test statistics. Section 4 applies our methods to re-examine the dividend yield as a predictor of long-horizon returns on the Standard and Poors index. Employing annual time series that begin in 1871, recursive estimation from 1990 to 2002 gives stable Bartlett-corrected sup-t^2 tests that consistently reject the hypothesis of no predictability at horizons of 13 years or more. Proofs of propositions are contained in the appendix.

2 Power advantages of long-horizon regression tests

For notational convenience, we write the short-horizon predictive regression as

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

We suppress the regression constant in (2) since its has no effect on the asymptotic properties of the tests.[2] However, the constant does have consequences for small-sample properties of the tests and we will include it in our analysis of those issues below. 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.

2.1 Local-to-unity asymptotic power

The observations are generated by

Assumption 1 *For sample size T, the observations have the representation*

$$\begin{aligned} \Delta y_{t+1} &= \beta_1(T) x_t + e_{t+1}, & (3) \\ x_{t+1} &= \rho(T) x_t + u_{t+1}, & (4) \end{aligned}$$

where $\{e_{t+1}\}$ and $\{u_{t+1}\}$ are zero-mean covariance stationary sequences. $\rho(T) = 1 + c/T$ and $\beta_1(T) = b_1/T$ give the sequence of local alternatives where $c \leq 0$ and $b_1 \geq 0$ are constants. For the long-horizon regression, the sequence of local alternatives at horizon k is $\beta_k(T) = (kb_1)/T$.

Let $\xi_t = (\Delta x_t', e_t')'$ and its long-run covariance matrix be $\Omega = \Sigma + \Lambda + \Lambda' = \lim_{T \longrightarrow \infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{l=-\infty}^{\infty} E(\xi_t \xi_{t-l}') = \begin{pmatrix} \Omega_{xx} & \Omega_{xe} \\ \Omega_{ex} & \Omega_{ee} \end{pmatrix}$, 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 $\Lambda = \lim_{T \longrightarrow \infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{l=1}^{t-1} E(\xi_{t-l} \xi_t') = \begin{pmatrix} \Lambda_{xx} & \Lambda_{xe} \\ \Lambda_{ex} & \Lambda_{ee} \end{pmatrix}$. Endogeneity

[2] This reformulation of the dependent variable maps exactly into the returns formulation for exchange rates 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.

of the regressor may arise because the covariance between $\{e_{t+1}\}$ and $\{u_{t+1}\}$ has not been restricted. Any resulting endogeneity, however, will be local-to-zero in the sense that the dependence between the regressor and the regression error vanishes as $T \to \infty$.

Let B_1 be a scalar Brownian motion with long run variance Ω_{xx}, J_c^* be the diffusion process defined by $dJ_c^*(r) = cJ_c^*(r) + dB_1(r)$, 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 $\hat{\beta}_k = \left(\sum_t x_t\left(y_{t+k} - y_t\right)\right)\left(\sum_t x_t^2\right)^{-1}$ with asymptotic t-ratio $t_\beta(k) = \hat{\beta}_k/\sqrt{V(\hat{\beta}_k)}$, where $V(\hat{\beta}_k) = \widehat{\Omega}_{ee}\left(\sum_t x_t^2\right)^{-1}$. Following Phillips (1988) and Cavanagh et. al. (1995), we have

Proposition 1 *Under Assumption 1, the OLS estimator of the kth horizon regression slope coefficient is asymptotically distributed as,*

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

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, $R = \mathbf{\Omega}_{xx}^{-1/2}\mathbf{\Omega}_{ee}^{1/2}$, $\delta = \Omega_{xe}\left(\Omega_{xx}\Omega_{ee}\right)^{-1/2}$ *and* $\theta_c = \left(\int J_c^2\right)^{-1/2} > 0$.

When the regressor is exogenous, $\Lambda_{xe} = \Lambda_{xe,k-1} = 0$. It follows that $t_\beta(k) \Rightarrow \delta\tau_{1c} + \left(1-\delta^2\right)^{1/2} N(0,1) + \left(b_1/\sqrt{\Omega_{xx}\Omega_{ee}}\right)\theta_c$ which does not depend on k. This gives,

Corollary 1 *(Exogeneity) Under Assumption 1, if the regressor is econometrically exogenous, then the long-horizon regression test has no asymptotic power advantage over the short-horizon regression test.*

In empirical work, however, the regressor is unlikely to be econometrically exogenous. The return on equity $r_t = \ln(P_t + D_{t-1}) - \ln P_{t-1}$ and the log dividend yield $x_t = \ln D_{t-1} - \ln P_t$ both depend on P_t in a way to suggest that the regression error and the innovation to x_t will be negatively correlated, $E(u_{t+1}e_{t+1}) < 0$.[3]

[3] The predicted negative innovation correlation is in fact present in annual stock-return data. Fitting a first-order vector autoregression to $(e_t, v_t)'$, we obtain an innovation correlation of -0.948..

Endogeneity can also be seen to arise if the bivariate sequence $\{(y_t, z_t)'\}$ can be represented as a first-order vector-error correction model (VECM) with cointegration vector $(-1, 1)$,

$$\Delta Y_t = hx_{t-1} + A\Delta Y_{t-1} + \zeta_t,$$

where $Y_t' = (y_t, z_t), h' = (h_1, h_2), A = [a_{ij}]$ is a 2×2 matrix of coefficients, $\zeta_t' = (\epsilon_t, u_t)$, and $x_t = z_t - y_t$ is the equilibrium error.[4] The VECM has an equivalent restricted second-order vector autoregressive (VAR) representation for $X_t = (\Delta y_t, x_t)$, where $X_t = BX_{t-1} + CX_{t-2} + V_t,$ $B_{11} = (a_{11} + a_{12}), B_{12} = (h_1 + a_{12}), B_{21} = (a_{22} - a_{12} + a_{21} - a_{11}), B_{22} = (1 + h_2 - h_1 + a_{22} - a_{12}),$ $C_{11} = C_{21} = 0, C_{12} = -a_{12}, C_{22} = (a_{12} - a_{22}),$ and $V_t' = (\epsilon_t, u_t - \epsilon_t)$. From the VAR, it can be seen that $\{x_t\}$ and $\{\Delta y_t\}$ are correlated both contemporaneously and dynamically (at leads and lags). The first equation from the VAR representation is the short-horizon regression

$$\Delta y_{t+1} = (h_1 + a_{12})x_t + [(a_{11} + a_{12})\Delta y_t - a_{12}x_{t-1} + \epsilon_{t+1}], \tag{7}$$

with slope coefficient $h_1 + a_{12}$ and regression error $(a_{11} + a_{12})\Delta y_t - a_{12}x_{t-1} + \epsilon_{t+1,}$ that is both serially correlated and correlated with x_t. The objective of the short-horizon regression is not to estimate $h_1 + a_{12}$ 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.

With a local-to-unity regressor, eq.(2) resembles an unbalanced regression since Δy_{t+1} is nearly white noise and x_t is nearly integrated. Under the alternative $b_1 > 0$, there must be negative endogeneity because negative correlation between x_t and e_{t+1} is required for the linear combination $b_1 x_t + e_{t+1}$ to be nearly white noise (analogously, if $b_1 < 0$, positive endogeneity must be present). If the regressor is endogenous, it follows from eq.(6) that the limiting behavior of the difference between the $t-$statistics at horizons k and 1 is $t^a_\beta(k) - t^a_\beta(1) \Longrightarrow -\left(\Lambda_{xe,k-1}/\sqrt{\Omega_{xx}\Omega_{ee}}\right)\theta_c$, which is increasing in k. This gives,

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

2.2 Small-sample properties

All of the simulation work includes a constant in estimation. The first set of results that we discuss are simulations to confirm that the asymptotic predictions of long-horizon power advantages are present in small samples under regressor endogeneity. The DGP is as in Assumption 1 with $e_t = a_{11}e_{t-1} + a_{12}u_{t-1} + m_t,$ $u_t = n_t,$ where $(m_t, n_t)' \overset{iid}{\sim} [0, \phi], \phi_{mm} = \phi_{nn} = 1, -1 < \phi_{mn} < 0$. A property of this DGP is that the dependence between the regressor and regression error is local-to-zero in which the endogeneity factor

[4] 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.

in the short-horizon regression is $d_1(T) = E\left(\sum x_t e_{t+1}\right)\left(\sum x_t^2\right)^{-1} = O(T^{-1})$ and for the long-horizon regression is $d_k(T) = O(T^{-1})$.[5]

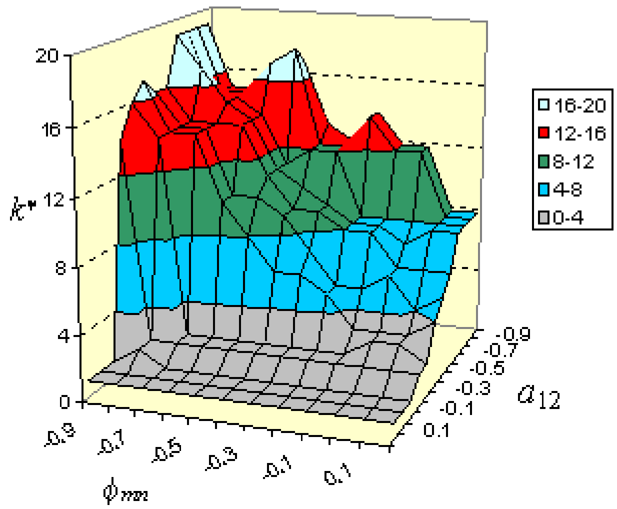

Figure 1. Horizons that maximize relative size-adjusted power.

From 5000 replications with $T = 100$, Figure 1 shows the horizon k^* that maximizes the relative size-adjusted power of the regression tests obtained by searching over $k \in [1, 20], \phi_{mn} \in [-0.9, 0], a_{12} \in [-0.9, 0]$ with $a_{11} = 0.1, b_1 = 10, c = -5$. In cases where the long-horizon regression test has no local power advantages, the result is $k^* = 1$. As can be seen from the figure, the size-adjusted power of long-horizon regression tests consistently dominate those of short-horizon tests in this region of the parameter space.[6] Table 1, displays the 5% size-adjusted power of the tests for alternative values of ϕ_{mn} at selected horizons.

Table 1: 5% size-adjusted power at selected horizons, $T = 100, a_{11} = 0.1, a_{12} = -0.9, b_1 = 10, c = -5$

$\phi_{mn} \backslash k$	1	5	10	15	20
-0.9	0.000	0.203	0.223	0.226	0.217
-0.7	0.001	0.210	0.221	0.245	0.252
-0.5	0.007	0.229	0.282	0.277	0.271
-0.3	0.021	0.251	0.324	0.307	0.269
-0.1	0.041	0.294	0.385	0.317	0.281

[5]For this DGP, the endogeneity factor is $d_1(T) = (a_{12} + a_{11}\phi_{mn})\left(1 - (\rho(T))^2\right)(1 - a_{11}\rho(T))^{-1} = (a_{12} + a_{11}\phi_{mn})\frac{c^2 - 2Tc}{T^2 a_{11} - T^2 - Tca_{11}}$. Under the null hypothesis ($b_1 = d_1 = 0$), we set $a_{12} = a_{11} = 0$ but allow variations in ϕ_{mn}.

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

While size-adjusted power advantages are seen to accrue to long-horizon regressions, the conventional t-test cannot be used in practice. This is because the conventional t-statistic depends on the regressor's local-to-unity parameter c, which cannot be consistently estimated from the time-series. For practical considerations, we approach testing using a variant of the sup-bound test discussed by Cavanaugh et. al. (1995), which is an asymptotically valid test of predictability that does not depend on the nuisance parameter c. A convenient variant that admits a Bartlett correction and allows two-sided tests is to use the squared t-ratio. We refer to this as the sup–t^2 test. To construct the test for given δ, let $q_{t_\beta,c,\eta}$ be the 100η percentile of the distribution of $\delta^2\tau_{1c}^2 + \left(1-\delta^2\right)N(0,1)^2$. Since under the null, $t_\beta^2(k) \xrightarrow{L} \delta^2\tau_{1c}^2 + \left(1-\delta^2\right)N(0,1)^2$, it follows that the most conservative sup–t^2 test with at most asymptotic level η is performed by rejecting the null if $t_\beta^2(k) > q_{t_\beta,0,\eta}$. On the other hand, if $c = \underline{c} \ll 0$, then $t_\beta^2(k) \xrightarrow{L} N(0,1)^2$ and it follows that the most liberal test rejects if $t_\beta^2(k) > q_{t_\beta,\underline{c},\eta}$ which is equivalent to the conventional asymptotic chi-square test under stationarity of the regressor.

We show below that in small samples, the asymptotic sup–t^2 test becomes somewhat oversized at long horizons on account small-sample bias in the OLS slope estimator.[7] Because direct bias and size adjustment at long horizons may not be straightforward when the DGP is unknown, we discuss a strategy to achieve small-sample adjustments that is based on resampling the data and which does not require knowledge of the DGP.

3 Small-sample OLS bias and test-statistic adjustments

Bias adjustment. Our small-sample adjustment for OLS bias draws on the jackknife method originally proposed by Quenouille (1956). Suppressing the notational dependence on the horizon k and letting the true slope value be β_0, the small-sample OLS bias is

$$E(\widehat{\beta} - \beta_o) = \frac{\alpha}{T} + O\left(T^{-2}\right), \tag{8}$$

where the constant α depends on parameters of the asymptotic distribution. Eq.(8) motivates the following procedure to estimate the first-order bias term α. 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 comprised of the dependent and independent variables of the long-horizon regression. Construct a moving-block sample with block size B from the original set of observations, $\{\xi_1, \cdots, \xi_B\}, \left\{\xi_2, \cdots, \xi_{B+1}\right\}, \cdots, \left\{\xi_{T_1-B+1}, \cdots, \xi_{T_1}\right\}$. Using the data from each block, estimate the $k-$horizon slope coefficient β_0. Call these estimates β_{Bj},where $j = 1, ..., T_1-$

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

$B+1$ indexes the block of B observations. For each j, the analog to (8) is

$$\beta_{Bj} = \beta_0 + \frac{\alpha}{B} + O_p\left(B^{-2}\right). \tag{9}$$

Multiplying both sides of (9) by B and taking the sample average gives

$$BE_B^*\left(\beta\right) = \alpha + B\beta_0 + O\left(B^{-1}\right), \tag{10}$$

where $E_B^*\left(\beta\right) = \frac{1}{T-B+1}\sum_{j=1}^{T-B+1}\beta_{Bj}$. Repeat with block size $B+1$, then block size $B+2$, and so on through block size $B+(T_1-B)=T_1$ to obtain the sequence $\left\{BE_B^*\left(\beta\right)=\alpha+B\beta_0, (B+1)E_{B+1}^*\left(\beta\right)=\alpha+(B+1)\beta_0, ..., T_1E_{T_1}^*\left(\beta\right)=\alpha+\ T_1\beta_0\right\}$. Let $t=B, B+1, ..., T_1$, define $z_t = tE_t^*\left(\beta\right)$, and write z_t as a regression on a constant and trend, $z_t = \alpha + \beta_0 t$. Call the estimated coefficient on the trend β_{RJK}. It is the recursive moving-block jackknife estimate of β_0 which is accurate in the following sense.

Proposition 2 *(Recursive moving-block jackknife)*

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

The true value of α under the null is different than it is under the alternative. But since the recursive moving-block jackknife is a method to estimate α, it provides a bias adjustment under the null as well as under the alternative. For the choice of B we draw on Hall, Horowitz and Jing (1995) who provide blocking rules on the bootstrap with dependent data. For bias estimation, the suggested optimal block size is $T^{1/3}$ while for one or two sided distribution functions, the suggested size is $T^{1/4}$ and $T^{1/5}$, respectively.

Table 2: OLS and Moving-Block Recursive Jackknife Bias under the Null. $c=-5, T=100, a_{11}=a_{12}=0$.

	OLS					Jackknife				
$\phi_{mn} \setminus k$	1	5	10	15	20	1	5	10	15	20
-0.9	0.038	0.169	0.308	0.420	0.521	0.001	0.068	0.172	0.265	0.357
-0.7	0.029	0.131	0.238	0.327	0.404	0.001	0.052	0.132	0.204	0.274
-0.5	0.021	0.094	0.170	0.233	0.287	0.000	0.037	0.093	0.144	0.191
-0.3	0.012	0.056	0.101	0.140	0.171	0.000	0.021	0.054	0.084	0.110
-0.1	0.004	0.019	0.032	0.046	0.055	0.000	0.006	0.016	0.024	0.029

The Monte Carlo work reported in Table 2 shows that the recursive moving-block jackknife estimator eliminates nearly all of the bias in the short-horizon regression. For $\phi_{mn} = -0.3$, it reduces the bias by 100% ($k=1$), 68% ($k=5$), 50% ($k=10$), 48% ($k=15$), and 47% ($k=20$). The relative reduction of bias is fairly stable for alternative values of ϕ_{mn}.

Recursive moving-block Bartlett correction. We show below that the sup–t^2 test is somewhat oversized in small samples. To obtain tests with better size, we apply a Bartlett correction.[8] We begin with the asymptotic expansion of the squared t-ratio,

$$W = W_T - \frac{\alpha_1}{T} W_T - \frac{\alpha_2}{T} W_T^2, \tag{11}$$

where W_T is the squared t-statistic computed from a sample of size T, W is its 'true' value, and α_1 and α_2 are 'Bartlett coefficients' which are derived from the asymptotic expansion of the statistic.[9] While α_1 and α_2 depend on parameters of the asymptotic DGP, our method does not require those formulae as it is designed to estimate the Bartlett coefficients. Moreover, because we are estimating the Bartlett coefficients whose true values are different under the null and the alternative, the recursive moving-block Bartlett correction should produce a test with correct size and preserve long-horizon power advantages.

Table 3: Effective size of asymptotic and moving-block recursive Bartlett Corrected sup$-t^2$ test. $a_{11} = a_{12} = 0, c = 5, T = 100$

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

To apply the Bartlett correction, proceed 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 sup-t^2 statistic, $W_{B,j} = t_\beta^2$, $j = 1, ..., T_1 - B + 1$. From each block, form the analog to (11)

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

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

[9] 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.

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 $B+(T_1-B) = T_1$. 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 rewrite as the regression $z_t = \alpha_1 E_t^*(W) + \alpha_2 E_t^*(W^2) + Wt$. The estimated coefficient on the trend is the recursive moving-block Bartlett-corrected test statistic.

Simulation results displayed in Table 3 show that the asymptotic sup–t^2 test is oversized at $k = 10, 15, 20$ whereas the Bartlett-corrected test is reasonably sized at those horizons and is somewhat undersized for $k = 1, 5$. For $T = 100$, the Bartlett correction is seen to give tests that are better sized than the asymptotic test.

Table 4: Local-to-Unity Effective Size and Power of asymptotic and Bartlett corrected sup$-t^2$ test. $a_{11} = a_{12} = 0, c = -5, \phi_{mn} = -0.90$ under the null $a_{11} = 0.1, a_{12} = -0.3, \rho = 1 - 5/T, b_1 = 20/T, \phi_{mn} = -0.9$ under the alternative

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

Table 4 reports size and power performance for various values of T. For longer horizons, say $k = 20$, a time-series length of $T = 300$ is required for the asymptotic sup–t^2 test to be correctly sized. The Bartlett-corrected sup-t^2 test is for the most part undersized when $T = 200$ and $T = 300$. Local-to-unity power of the Bartlett-corrected tests are seen to rival those of the size-adjusted asymptotic sup–t^2 tests. While the coarse grid of horizons that are reported do not, in many cases, pick off the horizon that gives the test its maximal power, results for the horizons that we do report show that long-horizon power advantages hold up. These results indicate that the Bartlett correction to the asymptotic sup–t^2 test should work well in practice.[10]

4 Predictability of long-horizon equity returns

We apply the Bartlett correction to long-horizon tests of whether the log dividend yield predicts future stock returns. 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, the present-value relation suggests that the log dividend yield should contain useful information for predicting future returns.

Regressing future returns at various horizons on the log dividend yield using annual observations from 1871 to 2002 produces the following customary results.[11]

Table 5: Short and long-horizon equity return regressions

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

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

The log dividend-yield (the regressor) is a persistent series. Its first-ordered autocorrelation is 0.843. We obtain augmented Dickey–Fuller (ADF) test statistic values of -0.189 (with constant) and -1.106 (constant and trend). The Phillips–Perron (PP) test

[10] Our results are in line with recent work by Nielsen (1997) and Johansen (2004) who show that the Bartlett correction for unit-root tests work well in practice.

[11] These data were used in Robert J. Shiller (2000) and were obtained from his web site. Annual observations were constructed from these monthly data. Returns are $r_{t+1} = \ln((P_{t+1} + D_t)/P_t)$ 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. Asymptotic t-ratios constructed using Newey-West (1994) automatic lag-length HAC standard errors. Because the dependent variable changes with k, the R^2s are not directly comparable across horizons.

statistics are -0.850 (constant) and -2.01 (constant and trend).[12] The apparent nonstationarity of the dividend yield is driven in large part by the bull market of the late 1990s. When we end the sample in 1997, however, the ADF statistics become -2.965 (constant) and -3.758 (constant and trend). Corresponding PP-statistics are -2.640 (constant) and -3.656 (constant and trend).

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

Because of the unusual behavior of stock prices associated with the bull market of the 90s and the subsequent decline in 2001-2002, the estimates are sensitive to the sample period. In recognition of this sensitivity, we run the regressions for horizons 1 through 20 initially using 1990 as the end of the sample and then recursively updating the sample through 2002. Since the true value of the local-to-unity parameter $c < 0$ is unknown, the exact critical values for the test are bounded by critical values for the conventional χ^2 test and the sup–t^2 test.[13] To compare the inferences that one would draw from the most liberal and the most conservative tests, for each sample we conduct four tests of predictability: the asymptotic sup–t^2 test, the Bartlett-corrected version of this test, the standard asymptotic chi-square test for stationary regressor and the Bartlett-corrected version of this test.[14]

$\underline{k}$, shown in Table 6 is the shortest horizon for which the null is rejected at the 5-percent nominal level. The Bartlett-corrected sup–t^2 test is consistently able to reject the null at $k = 13$ and for samples ending in 1997 and 1998 it rejects the null at $k = 11$. The maximal Bartlett-corrected test statistics are obtained at horizon $k^* = 19$ for every sample. The Bartlett-corrected asymptotic χ^2 test rejects the null at $k = 10$ in every sample.

The stability of the horizon for which the Bartlett-corrected tests reject the null contrasts sharply with the asymptotic test results. As observations from the 1990s are added to the sample, successively longer horizons are required to reject the null as the

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

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

[14] Although it is well-known that the asymptotic χ^2 test suffers from substantial size distortion, the Bartlett-corrected version of the test is only modestly oversized. The small-sample performance of these tests are reported in the working paper version [Mark and Sul (2004)] where it is shown that the main results in this paper hold when the regressor is covariance stationary.

$\underline{k}$ associated with the asymptotic sup–t^2 test and the conventional χ^2 test is increasing as the sample is lengthened. When the sample ends in 1991, the asymptotic sup–t^2 test rejects the null with $k = 9$, but when the sample ends in 2002, the shortest horizon for which the test rejects is $k = 16$.

The conventional t-statistic appears to exhibit sensitivity to the size and the direction of the bias. The bias of the short-horizon regression, which we estimate by the difference between OLS and the recursive moving-block jackknife estimate, is shown in the last column of Table 6. The estimated bias is positive through 1997 then turns negative. The negative bias evidently alters the size of the asymptotic tests whereas the Bartlett correction evidently does a reasonably good job of correcting size distortion.

Table 6: Stock Return Predictability

T	$\underline{k}$ asy. χ^2	$\underline{k}$ BC χ^2	$\underline{k}$ asy. sup$-t^2$	k^* conventional	$\underline{k}$ BC sup$-t^2$	k^* Bartlett	bias
1991	5	10	9	20	13	19	0.072
1992	5	10	9	20	13	19	0.077
1993	7	10	9	13	13	19	0.072
1994	8	10	9	14	13	19	0.063
1995	8	10	10	13	13	19	0.058
1996	8	10	10	14	13	19	0.043
1997	9	10	11	14	11	19	0.011
1998	9	10	12	14	11	19	-0.006
1999	10	10	13	14	13	19	-0.023
2000	12	10	14	19	13	19	-0.062
2001	12	10	15	19	13	19	-0.077
2002	13	10	16	19	13	19	-0.059

Notes: $\underline{k}$ is the shortest horizon for which the test rejects the null hypothesis. k^* is the horizon that gives the maximal test statistic value. BC denotes Bartlett correction. Bias is calculated at horizon $\underline{k}$ for BC sub$-t^2$.

5 Conclusion

Long-horizon regression tests have local asymptotic power advantages over short-horizon tests when the regresssor is persistent (local-to-unity) and endogenous. While asymptotic theoretical justification is available for using long horizons, small-sample bias of OLS causes size distortion in the asymptotic tests. Because conventional bias adjustment may not be easily handled at long horizons when the DGP is unknown, we suggest resampling strategies to achieve bias reduction and to correct for test size distortion.

Estimation bias is addressed by the recursive moving-block jackknife estimator, which successfully provides bias correction in the short-horizon predictive regression and controls for about half of the bias at long horizons. Small-sample size distortion of asymptotic tests are addressed by a recursive moving-block Bartlett correction. The Bartlett corrected sup–t^2 statistic is reasonably sized at short and long horizons and effectively maintains small-sample power advantages of long-horizon tests.

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

Acknowledgements

Comments from three anonymous referees and Whitney Newey (the editor) led to improvements in the paper. For useful comments on previous versions, we thank Steve Cecchetti, Paul Evans, Lutz Kilian, Lung-Fei Lee, Ryan Greenaway-McGrevy, Hu McCulloch, Charles Nelson, Masao Ogaki, Joon Park, Mark Wohar, and participants in the Econometrics Seminar at Ohio State University for thoughtful comments on earlier drafts. Sul thanks the University of Auckland for support under grant no. 3601013/9254. Mark thanks the National Science Foundation for support.

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

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

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

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

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

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

[14] 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.

[15] 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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

[30] 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.

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

[32] 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).

[33] 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.

[34] 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.

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

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

[37] Phillips, P.C.B., and Park, 1988. "Statistical Inference in Regressions with Integrated Processes: Part 1," *Econometric Theory*, 4, 468-497.

[38] 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.

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

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

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

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

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

[44] 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

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

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

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

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) 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

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

■