JOURNAL OF FINANCIAL AND QUANTITATIVE ANALYSIS Vol. 46, No. 3, June 2011, pp. 815–839
COPYRIGHT 2011, MICHAEL G. FOSTER SCHOOL OF BUSINESS, UNIVERSITY OF WASHINGTON, SEATTLE, WA 98195
doi:10.1017/S0022109011000135

New Methods for Inference in Long-Horizon Regressions

Erik Hjalmarsson*

Abstract

I develop new results for long-horizon predictive regressions with overlapping observations. I show that rather than using autocorrelation robust standard errors, the standard *t*-statistic can simply be divided by the square root of the forecasting horizon to correct for the effects of the overlap in the data. Further, when the regressors are persistent and endogenous, the long-run ordinary least squares (OLS) estimator suffers from the same problems as the short-run OLS estimator, and it is shown how similar corrections and test procedures as those proposed for the short-run case can also be implemented in the long run. An empirical application to stock return predictability shows that, contrary to many popular beliefs, evidence of predictability does not typically become stronger at longer forecasting horizons.

I. Introduction

Predictive regressions are used frequently in empirical finance and economics. The underlying economic motivation is often the test of a rational expectations model, which implies that the innovations to the dependent variable should be orthogonal to all past information (i.e., the dependent variable should not be predictable using any lagged regressors). Although this orthogonality condition should hold at any time horizon, it is popular to test for predictability by regressing sums of future values of the dependent variable onto the current value of the regressor. A leading example is the question of stock return predictability, where regressions with 5- or 10-year returns are often used (e.g., Campbell and Shiller (1988), Fama and French (1988)). While stock return predictability also serves as

*Hjalmarsson, e.hjalmarsson@qmul.ac.uk, Queen Mary, University of London, School of Economics and Finance, Mile End Road, London E1 4NS, UK. The paper has greatly benefited from advice by an anonymous referee. Other helpful comments have also been provided by Don Andrews, Hendrik Bessembinder (the editor), John Campbell, Dobrislav Dobrev, Ray Fair, Jon Faust, Lennart Hjalmarsson, Randi Hjalmarsson, Yuichi Kitamura, Taisuke Otsu, Peter Phillips, Robert Shiller, as well as participants in the econometrics seminar and workshop at Yale University, the international finance seminar at the Federal Reserve Board, the finance seminar at Göteborg University, the 2005 World meeting of the Econometric Society in London, and the 2007 Copenhagen Conference on Stock Return Predictability.

https://doi.org/10.1017/S0022109011000135 Published online by Cambridge University Press

the motivating example in this paper, the results are applicable to a much wider range of empirical questions.[1]

The main inferential issue in long-horizon regressions has been the proper calculation of standard errors. Since overlapping observations are typically used, the regression residuals will exhibit strong serial correlation; standard errors failing to account for this fact will lead to biased inference. Therefore, autocorrelation robust estimation of the standard errors (e.g., Hansen and Hodrick (1980), Newey and West (1987)) is typically used. However, these robust estimators tend to perform poorly in finite samples, since the serial correlation induced in the error terms by overlapping data is often very strong.

The main contribution of this paper is the development of new results and methods for long-run regressions with overlapping observations. Using a framework where the predictors are highly persistent variables, as in Stambaugh (1999) and Campbell and Yogo (2006), I show how to obtain asymptotically correct test statistics, with good small sample properties, for the null hypothesis of no predictability. Rather than using robust standard errors, I find that the standard t-statistic can simply be divided by the square root of the forecasting horizon to correct for the effects of the overlap in the data. Further, when the regressor is persistent and endogenous, the long-run ordinary least squares (OLS) estimator suffers from the same problems as the short-run OLS estimator, and similar corrections and test procedures as those proposed by Campbell and Yogo for the short-run case should also be used in the long run; again, the resulting test statistics should be scaled due to the overlap.[2] Thus, these results lead to simple and more efficient inference in long-run regressions by obviating the need for robust standard error estimation methods and controlling for the endogeneity and persistence of the regressor.

The asymptotic results in this paper are derived under the assumption that the forecasting horizon remains fixed as the sample size increases and thus complement the results of Valkanov (2003), who relies on the assumption that the forecasting horizon grows with the sample size. Interestingly, both asymptotic approaches lead to similar scaling results for the t-statistic. In relation to Valkanov's study, the current paper makes two important contributions. First, I show that with exogenous regressors, the scaled standard t-statistic will be normally distributed and standard inference can thus be performed. Second, when the regressors are endogenous, the inferential methods can be suitably modified to correct for the biasing endogeneity effects; this can be seen as an analogue of the inferential procedures developed by Campbell and Yogo (2006) for short-run, 1-period-horizon

[1]Other applications of long-horizon regressions include tests of exchange rate predictability (Mark (1995), Berkowitz and Giorgianni (2001), and Rossi (2005)), the Fisher effect (Mishkin (1990), (1992), Boudoukh and Richardson (1993)), and the neutrality of money (Fisher and Seater (1993)).

[2]A predictive regressor is generally referred to as endogenous if the innovations to the returns are contemporaneously correlated with the innovations to the regressor. When the regressor is strictly stationary, such endogeneity has no impact on the properties of the estimator, but when the regressor is persistent in some manner, the properties of the estimator will be affected (e.g., Stambaugh (1999)). Nelson and Kim (1993) may be the first to raise the biasing problems of endogenous regressors in the long-horizon case.

https://doi.org/10.1017/S0022109011000135 Published online by Cambridge University Press

https://doi.org/10.1017/S0022109011000135 Published online by Cambridge University Press

regressions. Importantly, the modified test statistic in the endogenous case is again normally distributed. In contrast, Valkanov's test statistics have highly nonstandard distributions, both for exogenous and endogenous regressors, which require simulation of the critical values for each specific case.

Asymptotic results are, of course, only useful to the extent that they provide relevant information regarding the finite sample properties of an econometric procedure. As shown in Monte Carlo simulations, both the asymptotic results derived under the assumptions in this paper and those derived under the assumptions in Valkanov's (2003) paper provide good approximations of finite sample behavior. The simulations show that the asymptotic normal distribution of the test statistics derived in this paper provides a good approximation in finite samples, even when the forecasting horizon is large compared to the sample size, resulting in rejection rates that are very close to the nominal size of the test under the null hypothesis. In terms of power properties, the Monte Carlo simulations show that the tests proposed here are fairly similar to those of Valkanov, although there are typically some power advantages to the current procedures. This is especially true when the regressors are endogenous, in which case the test procedures derived here can be substantially more powerful than the test proposed by Valkanov.

The rest of the paper is organized as follows. Section II sets up the model and derives the theoretical results, and Section III discusses the practical implementation of the methods in the paper. Section IV describes the Monte Carlo simulations that illustrate the finite sample properties of the methods. An empirical application to stock return predictability is given in Section V, where it is shown that, contrary to many popular beliefs, evidence of predictability does not typically become stronger at longer forecasting horizons. Section VI concludes, and technical proofs are found in the Appendix.

II. Long-Run Estimation

A. Model and Assumptions

Although the results derived in this paper are of general applicability, it is helpful to discuss the model and derivations in light of the specific question of stock return predictability. Thus, let the dependent variable be denoted r_t, which would typically represent excess stock returns when analyzing return predictability, and the corresponding regressor, x_t.[3] In long-run regressions, the focus of interest is the fitted regression,

$$r_{t+q}(q) = \alpha_q + \beta_q x_t + u_{t+q}(q), \tag{1}$$

where $r_t(q) = \sum_{j=1}^{q} r_{t-q+j}$, and long-run future returns are regressed onto a 1-period predictor. The OLS estimator of β_q in equation (1), using overlapping observations, is denoted by $\hat{\beta}_q$.

[3]The asymptotic results presented in Section II all generalize immediately to the case of multiple regressors. However, the Bonferroni methods described in Section III are currently only developed for the case of a single regressor.

It is important to stress that equation (1) is a *fitted* regression and should not be interpreted as the data-generating process (DGP). That is, the DGP more naturally describes the true model for r_t, rather than $r_t(q)$, since the latter is just a sum of the former. In particular, the behavior of r_t and x_t are assumed to satisfy

$$r_{t+1} = \alpha + \beta x_t + u_{t+1}, \tag{2}$$

$$x_{t+1} = \gamma + \rho x_t + v_{t+1}, \tag{3}$$

where $\rho = 1 + c/T$, $t = 1, \ldots, T$, and T is the sample size; the demeaned versions of r_t and x_t will be denoted as $\underline{r}_t$ and $\underline{x}_t$, respectively. Throughout the paper, the results are derived under the null hypothesis of no predictability, in which case equation (2) simplifies to $r_{t+1} = \alpha + u_{t+1}$ and immediately implies that $\beta_q = 0$ for all q. However, it will sometimes be useful to refer to a DGP that also allows for return predictability, such as in the power simulations, and equation (2) therefore allows for deviations from the null.

The joint error process, $w_t = (u_t, v_t)'$, is assumed to satisfy the usual martingale difference assumption and can be conditionally heteroskedastic, as long as it is covariance stationary; Assumption 1 in Appendix B states the formal conditions. The covariance matrix is denoted $\Omega = \mathrm{E}[w_t w_t'] = [(\omega_{11}, \omega_{12}), (\omega_{12}, \omega_{22})]$. The error terms u_t and v_t are often highly correlated, and the regressor will be referred to as endogenous whenever this correlation, which will be labeled $\delta \equiv \omega_{12}/\sqrt{\omega_{11}\omega_{22}}$, is nonzero.

The autoregressive root of the regressor is parameterized as being local-to-unity, which captures the near unit-root, or highly persistent, behavior of many predictor variables, but is less restrictive than a pure unit-root assumption. The near unit-root construction, where the autoregressive root drifts closer to unity as the sample size increases, is used as a tool to enable an asymptotic analysis where the persistence in the data remains large relative to the sample size, also as the sample size increases to infinity. That is, if ρ is treated as fixed and strictly less than unity, then as the sample size grows, the process x_t will behave as a strictly stationary process asymptotically, and the standard 1st-order asymptotic results will not provide a good guide to the actual small sample properties of the model. For $\rho = 1$, the usual unit-root asymptotics apply to the model, but this is clearly a restrictive assumption for most potential predictor variables. Instead, by letting $\rho = 1 + c/T$, the effects from the high persistence in the regressor will appear also in the asymptotic results, but without imposing the strict assumption of a unit root.

The greatest problem in dealing with regressors that are near unit-root processes is the nuisance parameter c, which is generally unknown and not consistently estimable.[4] It is nevertheless useful to first derive inferential methods under the assumption that c is known, and then use the arguments of Cavanagh, Elliot, and Stock (1995) to construct feasible tests. The remainder of this section derives and outlines the inferential methods used for estimating and performing tests on β_q in equation (1), treating c as known. Section III discusses how the methods of

[4]That is, ρ can be estimated consistently, but not with enough precision to identify $c = T(\rho - 1)$.

https://doi.org/10.1017/S0022109011000135 Published online by Cambridge University Press

Cavanagh et al. and Campbell and Yogo (2006) can be used to construct feasible tests with c unknown.

B. The Limiting Distribution of the Long-Run OLS Estimator

The following theorem states the asymptotic distribution of the long-run OLS estimator of equation (1) and provides the key building block for the rest of the analysis.

Theorem 1. Under the null hypothesis of no predictability, for a fixed q as $T \to \infty$,

$$\frac{T}{q}\left(\hat{\beta}_q - 0\right) \quad \Rightarrow \quad \left(\int_0^1 dB_1 \underline{J}_c\right)\left(\int_0^1 \underline{J}_c^2\right)^{-1}, \tag{4}$$

where $B(\cdot) = (B_1(\cdot), B_2(\cdot))'$ denotes a 2-dimensional Brownian motion with variance-covariance matrix Ω, $J_c(r) = \int_0^r e^{(r-s)c} dB_2(s)$, and $\underline{J}_c = J_c - \int_0^1 J_c$.

Theorem 1 shows that under the null of no predictability, the limiting distribution of $\hat{\beta}_q$ is identical to that of the standard short-run, 1-period, OLS estimator $\hat{\beta}$ in equation (2), which is easily shown to converge to this distribution at a rate T (Cavanagh et al. (1995)), although $\hat{\beta}_q$ needs to be standardized by q^{-1}. This additional standardization follows since the estimated parameter β_q is of an order q times larger than the short-run parameter β, as discussed at length in Boudoukh, Richardson, and Whitelaw (2008).

The equality between the long-run asymptotic distribution under the null hypothesis, shown in Theorem 1, and that of the short-run OLS estimator may seem puzzling. The intuition behind this result stems from the persistent nature of the regressors. In a (near) unit-root process, the long-run movements dominate the behavior of the process. Therefore, regardless of whether one focuses on the long-run behavior, as is done in a long-horizon regression, or includes both the short-run and long-run information, as is done in a standard 1-period OLS estimation, the asymptotic result is the same since, asymptotically, the long-run movements are all that matter.

The limiting distribution of $\hat{\beta}_q$ is nonstandard and a function of the local-to-unity parameter c. Since c is not known, and not consistently estimable, the exact limiting distribution is not known in practice, which makes valid inference difficult. Cavanagh et al. (1995) suggest putting bounds on c in some manner and then find the most conservative value of the limiting distribution for some value of c within these bounds. Campbell and Yogo (2006) suggest first modifying the estimator or, ultimately, the resulting test statistic, in an optimal manner for a known value of c, which results in more powerful tests. Again using a bounds procedure, the most conservative value of the modified test statistic can be chosen for a value of c within these bounds. I will pursue a long-run analogue of this latter approach here, since it leads to more efficient tests and because the relevant limiting distribution is standard normal, which greatly simplifies practical inference. Before deriving the modified estimator and test statistic, however, it is instructive to consider the special case of exogenous regressors where no modifications are needed.

https://doi.org/10.1017/S0022109011000135 Published online by Cambridge University Press

C. The Special Case of Exogenous Regressors

Suppose the regressor x_t is exogenous in the sense that u_t is uncorrelated with v_t and thus $\delta = \omega_{12} = 0$.[5] In this case, the limiting processes B_1 and J_c are orthogonal to each other, and the limiting distribution in (4) simplifies. In particular, it follows that

$$\frac{T}{q}\left(\hat{\beta}_q - 0\right) \Rightarrow \left(\int_0^1 dB_1 \underline{J}_c\right)\left(\int_0^1 \underline{J}_c^2\right)^{-1} \equiv \text{MN}\left(0, \omega_{11}\left(\int_0^1 \underline{J}_c^2\right)^{-1}\right), \tag{5}$$

where $\text{MN}(\cdot)$ denotes a mixed normal distribution. That is, $\hat{\beta}_q$ is asymptotically distributed as a normal distribution with a random variance. Thus, conditional on this variance, $\hat{\beta}_q$ is asymptotically normally distributed. The practical implication of this result is that regular test statistics will have standard distributions. In fact, the following convenient result for the standard t-statistic now holds.

Corollary 1. Let t_q denote the standard t-statistic corresponding to $\hat{\beta}_q$. That is,

$$t_q = \frac{\hat{\beta}_q}{\sqrt{\left(\frac{1}{T-q}\sum_{t=1}^{T-q}\hat{u}_t(q)^2\right)\left(\sum_{t=1}^{T-q}\underline{x}_t^2\right)^{-1}}}, \tag{6}$$

where $\hat{u}_{t+q}(q) = r_{t+q}(q) - \hat{\alpha}_q - \hat{\beta}_q x_t$ are the estimated residuals. Then, under the null hypothesis of no predictability and an exogenous regressor with $\delta = 0$, for a fixed q as $T \to \infty$,

$$\frac{t_q}{\sqrt{q}} \Rightarrow N(0,1). \tag{7}$$

Thus, by standardizing the t-statistic for $\hat{\beta}_q$ by the square root of the forecasting horizon, the effects of the overlap in the data are controlled for and a standard normal distribution is obtained.

D. Endogeneity Corrections

As discussed previously, the long-run OLS estimator suffers from the same endogeneity problems as the short-run estimator; that is, when the regressor is endogenous, the limiting distribution is nonstandard and a function of the unknown parameter c. To address this issue, I consider a version of the augmented regression of Phillips (1991), together with the Bonferroni methods of Campbell and

[5] Although hardly any regressors in a typical macro or finance regression will be completely exogenous, some are "sufficiently" exogenous for the bias in OLS inference to be of negligible importance. For instance, Campbell and Yogo (2006) show that OLS inference with interest-rate-based predictors in stock return regressions tends to be fairly accurate.

https://doi.org/10.1017/S0022109011000135 Published online by Cambridge University Press

Yogo (2006). For now, I assume that ρ, or equivalently c, is known and derive an estimator and test statistic under this assumption.

Note that, for a given ρ, the innovations v_t can be obtained from $v_t = x_t - \rho x_{t-1}$. Consider first the 1-period regression. Once the innovations v_t are obtained, an implementation of the augmented regression equation of Phillips (1991), which he proposed for the pure unit-root case, is now possible:

$$r_{t+1} = \alpha + \beta x_t + \gamma v_{t+1} + u_{t+1\cdot v}. \tag{8}$$

Here $u_{t\cdot v} = u_t - \omega_{12}\omega_{22}^{-1} v_t$ and $\gamma = \omega_{12}\omega_{22}^{-1}$ (Phillips (1991)), and denote the variance of $u_{t\cdot v}$ as $\omega_{11\cdot 2} = \omega_{11} - \omega_{12}^2\omega_{22}^{-1}$. The idea behind equation (8) is that by including the innovations v_t as a regressor, the part of u_t that is correlated with v_t is removed from the regression residuals, which are now denoted $u_{t\cdot v}$ to emphasize this fact. The regressor x_t therefore behaves as if it were exogenous. It follows that the OLS estimator of β in equation (8) will have an asymptotic mixed normal distribution, with the same implications as discussed previously in the case of exogenous regressors.[6]

As discussed in Hjalmarsson (2007), there is a close relationship between inference based on the augmented regression equation (8) and the inferential procedures proposed by Campbell and Yogo (2006). To see this, suppose first that the covariance matrix for the innovation process, Ω, is known, and hence also $\gamma = \omega_{12}\omega_{22}^{-1}$ and $\omega_{11\cdot 2}$. The t-test for $\beta = 0$ in (8) is then asymptotically equivalent to

$$t_{\mathrm{AUG}} = \frac{\sum_{t=1}^{T-1} \left(\underline{r}_{t+1} - \gamma v_{t+1}\right) \underline{x}_t}{\sqrt{\omega_{11\cdot 2}\left(\sum_{t=1}^{T-1} \underline{x}_t^2\right)}} = \frac{\sum_{t=1}^{T-1} \left(\underline{r}_{t+1} - \omega_{12}\omega_{22}^{-1} v_{t+1}\right) \underline{x}_t}{\sqrt{\omega_{11\cdot 2}\left(\sum_{t=1}^{T-1} \underline{x}_t^2\right)}}, \tag{9}$$

which is, in fact, identical to Campbell and Yogo's Q-statistic. In practice, Ω is not known, but γ will be consistently estimated by OLS estimation of equation (8) and $\omega_{11\cdot 2}$ is estimated as the sample variance of the residuals.[7]

In the current context, the augmented regression equation is attractive, since it can easily be generalized to the long-horizon case. Thus, consider the augmented long-run regression equation

$$r_{t+q}(q) = \alpha_q + \beta_q x_t + \gamma_q v_{t+q}(q) + u_{t+q\cdot v}(q), \tag{10}$$

where $v_t(q) = \sum_{j=1}^{q} v_{t-q+j}$. The idea is the same as in the 1-period case, only now the corresponding long-run innovations $v_{t+q}(q)$ are included as an additional

[6] Amihud and Hurvich (2004) also consider augmented predictive regressions in the 1-period case, although they do not explicitly analyze the case where the autoregressive root is local to unity.

[7] Campbell and Yogo (2006) derive their Q-statistic as the uniformly most powerful test (conditional on an ancillary statistic) in a Gaussian framework. Phillips (1991) shows that OLS estimation of β in equation (8) is identical to estimation of β via Gaussian full information maximum likelihood of the system described by equations (2) and (3). The OLS estimator of equation (8) is thus effectively Fisher efficient in the sense that it asymptotically achieves the maximum concentration. These optimality properties in the short-run case motivate the use of the corresponding long-run extensions later. However, this does not imply any optimality for the long-run procedures.

https://doi.org/10.1017/S0022109011000135 Published online by Cambridge University Press

regressor. Under the null hypothesis of no predictability, equation (10) follows immediately from the specification of the short-run returns in equation (2). That is, under the null hypothesis,

$$r_{t+1} = \alpha + u_{t+1} = \alpha + \gamma v_{t+1} + u_{t+1\cdot v}, \tag{11}$$

with γ and $u_{t\cdot v}$ as previously specified. Equation (10) now follows from summing up across q-periods on each side of equation (11), with $\alpha_q = q\alpha$ and $\gamma_q = \gamma$.[8] Let $\hat{\beta}_q^+$ be the OLS estimator of β_q in equation (10), using overlapping observations. The following result now holds.

Theorem 2. Under the null hypothesis of no predictability, for a fixed q as $T \to \infty$,

$$\frac{T}{q}\left(\hat{\beta}_q^+ - 0\right) \Rightarrow \text{MN}\left(0, \omega_{11\cdot 2}\left(\int_0^1 \underline{J}_c^2\right)^{-1}\right). \tag{12}$$

The only difference from the result for the exogenous regressor case is the variance $\omega_{11\cdot 2}$, which reflects the fact that the variation in u_t that is correlated with v_t has been removed. As in the exogenous case, given the asymptotically mixed normal distribution of $\hat{\beta}_q^+$, standard test procedures can now be applied to test the null of no predictability. In particular, the scaled t-statistic corresponding to $\hat{\beta}_q^+$ will be normally distributed, as shown in the following corollary.

Corollary 2. Let t_q^+ denote the standard t-statistic corresponding to $\hat{\beta}_q^+$. That is,

$$t_q^+ = \frac{\hat{\beta}_q^+}{\sqrt{\left(\frac{1}{T-q}\sum_{t=1}^{T-q}\hat{u}_{t\cdot v}^+(q)^2\right) a'\left(\sum_{t=1}^{T-q}\underline{z}_t\underline{z}_t'\right)^{-1} a}}, \tag{13}$$

where $\hat{u}_{t\cdot v}^+(q)$ are the estimated residuals, $\underline{z}_t = (\underline{x}_t, v_{t+q}(q))$, and $a = (1,0)'$. Then, under the null hypothesis of no predictability, for a fixed q as $T \to \infty$,

$$\frac{t_q^+}{\sqrt{q}} \Rightarrow N(0,1). \tag{14}$$

Figure 1 illustrates the results of Theorem 2 and Corollary 2. In particular, Figure 1 shows the density of the standardized long-run estimates $\hat{\beta}_q^+/q$ for

[8] Under the alternative of predictability given by equation (2), the long-run augmented regression equation does not strictly hold in the sense that one cannot write $r_{t+q}(q) = \alpha_q + \beta_q x_t + \gamma_q v_{t+q}(q) + u_{t+q\cdot v}(q)$, where $v_{t+q}(q)$ and $u_{t+q\cdot v}(q)$ are orthogonal to each other. In order to do so, one would need to replace $\gamma_q v_{t+q}(q) \equiv \gamma_q \sum_{j=1}^q v_{t+j}$ by $\sum_{j=1}^q \xi_j v_{t+j}$ for some ξ_j that are not identical for each j. Equation (10) is, however, still valid under the alternative in the sense that including $v_{t+q}(q)$ will control for some of the noise in the error term. The long-horizon regression is a fitted regression, and complete endogeneity corrections will in general be functions of "short-run" parameters (e.g., β) in the DGP. To the extent that one does not wish to first impose and estimate some short-run predictive relationship in order to estimate the long-run relationship, full endogeneity corrections under the alternative will therefore typically not be possible.

https://doi.org/10.1017/S0022109011000135 Published online by Cambridge University Press

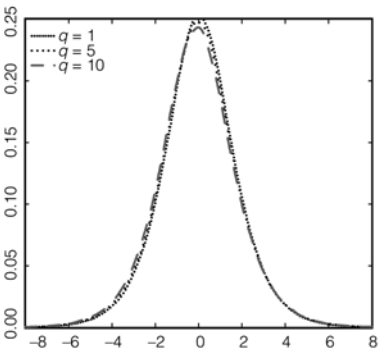
$q = 1$, 5, and 10, as well as the density of the corresponding scaled t-statistics, $t_q^+/\sqrt{q}$, which are plotted along with the standard normal density. The densities are obtained from simulating data generated by equations (2) and (3) with $T=100$, under the null of no predictability, and subsequently estimating equation (10). The normally distributed error terms u_t and v_t both have unit variance, and the correlation (δ) between them is set equal to -0.9. The local-to-unity parameter c is set equal to -2.5. One million repetitions are performed and the kernel density estimates of the resulting $\hat{\beta}_q^+/q$ and $t_q^+/\sqrt{q}$ are shown in Figure 1. The graphs clearly illustrate the previous results: The standardized long-run estimates, $\hat{\beta}_q^+/q$, all converge to approximately the same distribution for different values of q, and the scaled t-statistics from the augmented regression equation are very close to standard normally distributed.

FIGURE 1

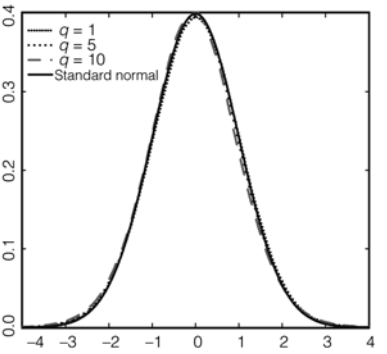
Density Plots of the Scaled Estimates of the Long-Run Slope Coefficient, $\hat{\beta}_q^+/q$, and the Scaled t-Statistics, $t_q^+/\sqrt{q}$, for $q = 1, 5$, and 10, with $T = 100$

The standard normal density is plotted as a comparison to the densities of the scaled t-statistics in Graph B of Figure 1. The densities are obtained from simulating data generated by equations (2) and (3) under the null of no predictability and subsequently estimating equation (10). The error terms u_t and v_t both have unit variance and correlation $\delta = -0.9$. The local-to-unity parameter c is set to -2.5. One million repetitions are performed, and the kernel density estimates of the resulting $\hat{\beta}_q^+/q$ and $t_q^+/\sqrt{q}$ are shown in the graphs.

Graph A. Density of Scaled Slope Coefficient



Graph B. Density of Scaled t-Statistic



Thus, for a given ρ, inference becomes trivial also in the case with endogenous regressors, since the scaled t-statistic corresponding to the estimate of β_q from the augmented regression equation (10) is normally distributed. In practice, ρ is typically unknown, and the next section outlines methods for implementing a feasible test.

Before moving on, it should also be stressed that although the arguments leading up to equation (10) were motivated by a Gaussian setup (see footnote 7), the results in Theorem 2 and Corollary 2 hold under much more general conditions on the error terms (see Assumption 1 in Appendix B). This is an important point because, under the assumption of normality, one might conjecture that there are few gains from considering long-horizon returns, since these will also be normally distributed and the likelihood of these aggregate long-run returns will not

https://doi.org/10.1017/S0022109011000135 Published online by Cambridge University Press

contain any additional information over the 1-period short-run returns. In contrast, long-run returns might be more useful in nonnormal cases. Although a formal analysis of the potential gains from long-horizon methods is outside the scope of this paper, I briefly return to this question in the simulation section, where results for t-distributed innovations are presented.[9]

III. Feasible Methods

To implement the methods for endogenous regressors described in the previous section, knowledge of the parameter c (or equivalently, for a given sample size, ρ) is required. Since c is typically unknown and not estimable in general, the bounds procedures of Cavanagh et al. (1995) and Campbell and Yogo (2006) can be used to obtain feasible tests.

Although c is not estimable, a confidence interval for c can be obtained, as described by Stock (1991). By evaluating the estimator and corresponding test statistic for each value of c in that confidence interval, a range of possible estimates and values of the test statistic is obtained. A conservative test can then be formed by choosing the most conservative value of the test statistic, given the alternative hypothesis. If the confidence interval for c has a coverage rate of $100(1-\alpha_1)\%$ and the nominal size of the test is $\alpha_2\%$, then by Bonferroni's inequality, the final conservative test will have a size no greater than $\alpha = \alpha_1 + \alpha_2\%$.

Thus, suppose that one wants to test $H_0 : \beta_q = 0$ versus $H_1 : \beta_q > 0$. The first step is to obtain a confidence interval for c, with confidence level $100(1-\alpha_1)\%$, which is denoted $[\underline{c}, \overline{c}]$. For all values of $\tilde{c} \in [\underline{c}, \overline{c}]$, $\hat{\beta}_q^+(\tilde{c})$ and the corresponding $t_q^+(\tilde{c})$ are calculated, where the estimator and test statistic are written as functions of $\tilde{c}$ to emphasize the fact that a different value is obtained for each $\tilde{c} \in [\underline{c}, \overline{c}]$. Let $t_{q,\min}^+ \equiv \min_{\tilde{c}\in[\underline{c},\overline{c}]} t_q^+(\tilde{c})$ be the minimum value of $t_q^+(\tilde{c})$ that is obtained for $\tilde{c} \in [\underline{c}, \overline{c}]$ and $t_{q,\max}^+ \equiv \max_{\tilde{c}\in[\underline{c},\overline{c}]} t_q^+(\tilde{c})$ be the maximum value. A conservative test of the null hypothesis of no predictability, against a positive alternative, is then given by evaluating $t_{q,\min}^+/\sqrt{q}$ against the critical values of the standard normal distribution; the null is rejected if $t_{q,\min}^+ \geq z_{\alpha_2}$, where z_{α_2} denotes the

[9] The previous results are derived under the assumption that the predictor variable has an autoregressive root ρ that is local to unity. In ongoing research, the author has shown that a similar scaling result for the test statistics also applies in the strictly stationary case with ρ fixed and strictly less than unity. In fact, it turns out that for a fixed ρ strictly less than unity, the $1/\sqrt{q}$ scaling rule provides a conservative bound on the scaling factor, and the exact scaling factor in the strictly stationary case is a function of ρ. That is, for a fixed ρ strictly less than unity, the t-statistics should be scaled by $1/\sqrt{f(q,\rho)}$ where $f(q,\rho) < q$ for $\rho < 1$ and $\lim_{\rho\uparrow 1} f(q,\rho) = q$; the proof is available from the author. The $1/\sqrt{q}$ scaling is thus valid outside the near unit-root framework analyzed here, although it will in general be conservative for strictly stationary predictors. Although the $1/\sqrt{f(q,\rho)}$ scaling potentially offers better-sized tests when ρ is strictly less than unity, its implementation requires an estimate of the autoregressive root ρ, which can lead to size distortions in the resulting scaled test; this is particularly true for ρ close to unity where the $f(q,\rho)$ function is most sensitive to the exact value of ρ. Thus, in practice, for predictors with autoregressive roots fairly close to 1, the best feasible approach appears to be to use the $1/\sqrt{q}$ scaling derived previously. A more exhaustive treatment of the testing problem in the stationary case is left for future research.

https://doi.org/10.1017/S0022109011000135 Published online by Cambridge University Press

$1 - \alpha_2$ quantile of the standard normal distribution. The resulting test of the null hypothesis will have a size no greater than $\alpha = \alpha_1 + \alpha_2$. An analogous procedure can be used to test against a negative alternative.[10]

Campbell and Yogo (2006) use the Dickey–Fuller generalized least squares (DF-GLS) unit-root test of Elliot, Rothenberg, and Stock (1996) to obtain a confidence interval for c. Here, I instead rely on a new unit-root test developed by Chen and Deo (2009a), (2009b). The Chen and Deo test, which is based on the idea of restricted likelihood inference (Kalbfleisch and Sprott (1970)), is convenient to use because the test statistic is (approximately) χ^2 distributed both under the null of a unit root as well as for roots close to unity. It is therefore trivial to invert the test statistic to obtain confidence intervals for c. The implementation of this unit-root test is described in Appendix A.

In general, Bonferroni's inequality will be strict, and the overall size of the test just outlined will be less than α. By shrinking the coverage rates of the confidence interval for c, a less conservative test can be achieved. Such procedures are discussed at length in Campbell and Yogo (2006), and I follow a similar approach here. Start with fixing α_2 at 10%, so that the nominal size of the tests evaluated for each $\tilde{c}$ is equal to 10%. Further, set the desired size of the overall Bonferroni test, labeled $\tilde{\alpha}$, to 10% as well. The degree of endogeneity, and hence the size of the biasing effects, is a function of the correlation δ between the innovations u_t and v_t. Thus, for each value of δ, one then searches over a grid of negative cs for values of α_1 such that the overall size of the test will be no greater than $\tilde{\alpha}$ for any value of c.[11] Since the asymptotic properties of the scaled estimators and corresponding test statistics derived here are identical for any q, one only needs to find values α_1 for $q = 1$. Table 1 gives the values of α_1, which result in 1-sided Bonferroni tests with size no greater than $\tilde{\alpha}/2$ equal to 5%, or, alternatively, 2-sided tests with size no greater than 10%.

As discussed later in conjunction with the Monte Carlo results, using these α_1 values, obtained for $q = 1$, in the long-run case appears to work well, although there is a tendency to underreject when the forecasting horizon is large relative to the sample size. Thus, there may be some scope for improving the procedure by size-adjusting the confidence interval for c differently for different combinations of q and T, but at the expense of much simplicity. Since the potential gains do not appear large, I do not pursue that here, although it would be relatively easy to implement on a case-by-case basis in applied work.

The practical implementation of the methods in this paper can be summarized as follows:

[10] An alternative approach is to invert the test statistics and form conservative confidence intervals instead. This approach will deliver qualitatively identical results, in terms of whether the null hypothesis is rejected or not. However, the distribution of the long-run estimator under the alternative hypothesis need not be the same as under the null hypothesis, in which case the confidence intervals are only valid under the null hypothesis. Presenting confidence intervals based on the distribution under the null hypothesis may therefore be misleading.

[11] Nonnegative cs are ruled out since, from an economic perspective, it is very unlikely that the predictor variables are truly nonstationary, even though they are highly persistent (see, e.g., Lewellen (2004), Baker, Taliaferro, and Wurgler (2006)).

https://doi.org/10.1017/S0022109011000135 Published online by Cambridge University Press

TABLE 1

Significance Levels for the Confidence Interval for c in the Bonferroni Test

Table 1 reports the values of α_1, for a given value of the correlation δ between the innovations u_t and v_t, which result in a 1-sided Bonferroni test with size no greater than 5% when the confidence interval for c, obtained from inverting the Chen and Deo (2009a), (2009b) test, has a coverage rate of $100(1 - \alpha_1)\%$. Two separate α_1s are reported: The first should be used when testing against a negative alternative, and the second when testing against a positive alternative. Only values for $\delta < 0$ are reported, since a regression with a negative δ can always be achieved by changing the sign of the regressor.

	Test against H_1:				
	$\beta < 0$	$\beta > 0$		$\beta < 0$	$\beta > 0$
δ	α_1	α_1	δ	α_1	α_1
–0.999	0.11	0.04	–0.500	0.68	0.20
–0.975	0.19	0.06	–0.475	0.70	0.21
–0.950	0.24	0.08	–0.450	0.72	0.21
–0.925	0.28	0.09	–0.425	0.74	0.22
–0.900	0.31	0.10	–0.400	0.75	0.23
–0.875	0.34	0.11	–0.375	0.77	0.24
–0.850	0.37	0.11	–0.350	0.79	0.24
–0.825	0.40	0.12	–0.325	0.81	0.25
–0.800	0.43	0.13	–0.300	0.82	0.26
–0.775	0.45	0.13	–0.275	0.84	0.27
–0.750	0.48	0.14	–0.250	0.86	0.28
–0.725	0.50	0.14	–0.225	0.87	0.29
–0.700	0.52	0.15	–0.200	0.89	0.29
–0.675	0.54	0.16	–0.175	0.91	0.30
–0.650	0.56	0.16	–0.150	0.92	0.31
–0.625	0.58	0.17	–0.125	0.94	0.32
–0.600	0.60	0.18	–0.100	0.96	0.33
–0.575	0.62	0.18	–0.075	0.97	0.34
–0.550	0.64	0.19	–0.050	0.99	0.35
–0.525	0.66	0.19	–0.025	0.99	0.37

i) Using OLS estimation for each equation, obtain the estimated residuals from equations (2) and (3). Calculate the correlation δ from these residuals.[12]

ii) Calculate the restricted likelihood ratio test (RLRT) statistic of Chen and Deo (2009a), (2009b), described in Appendix A, and obtain $\underline{c}$ and $\overline{c}$ by inverting the test statistic, using the confidence level α_1 in Table 1 corresponding to the estimated value of δ. Note that a different α_1 is used depending on whether one is testing against a positive or a negative alternative.

iii) For a grid of values $\tilde{c} \in [\underline{c}, \overline{c}]$, calculate $\hat{\beta}_q^+(\tilde{c})$ and $t_q^+(\tilde{c})$, and find $t_{q,\min}^+ \equiv \min_{\tilde{c}\in[\underline{c},\overline{c}]} t_q^+(\tilde{c})$ and $t_{q,\max}^+ \equiv \max_{\tilde{c}\in[\underline{c},\overline{c}]} t_q^+(\tilde{c})$.

iv) If the alternative hypothesis is $\beta_q > 0$, compare $t_{q,\min}^+/\sqrt{q}$ to the 95% critical values of the standard normal distribution (i.e., 1.645), and if the alternative hypothesis is $\beta_q < 0$, compare $t_{q,\max}^+/\sqrt{q}$ to the 5% critical values of the standard normal distribution.

The previous procedure results in a 1-sided test at the 5% level, or alternatively a 2-sided test at the 10% level.

[12] Alternatively, if one does not wish to use the short-run equation (2) to obtain estimates of the residuals u_t, one could impose the null hypothesis of no predictability, in which case the residuals u_t are simply equal to the demeaned returns.

https://doi.org/10.1017/S0022109011000135 Published online by Cambridge University Press

IV. Monte Carlo Results

All of the previous asymptotic results are derived under the assumption that the forecasting horizon is fixed. Valkanov (2003) also studies long-run regressions with near-integrated regressors, but he derives his asymptotic results under the assumption that $q/T \to \lambda \in (0,1)$ as $q, T \to \infty$. That is, he assumes that the forecasting horizon grows with the sample size. Under such conditions, the asymptotic results are, at least at first glance, quite different from those derived in this paper. There is, of course, no right or wrong way to perform the asymptotic analysis; what matters in the end is how well the asymptotic distributions capture the actual finite sample properties of the test statistics. To this end, Monte Carlo simulations are therefore conducted. Since Valkanov's methods are known to have good size properties, I merely present power results for his tests.

A. Size Properties

I start by analyzing the size properties of the scaled t-statistics proposed earlier in the paper. Equations (2) and (3) are simulated, with u_t and v_t drawn from an independent and identically distributed (IID) bivariate normal distribution with mean 0, unit variance and correlations $\delta = 0, -0.7, -0.90, -0.95$, and -0.99. The sample size is either $T = 100$ or $T = 500$. The intercept α is set to 0, and the local-to-unity parameter c is set to either $-2.5, -10$, or -20 for $T = 100$ and to $-1, -5$, or -20 for $T = 500$. Since the sizes of the tests are evaluated, the slope coefficient β is set to 0, which implies that $\beta_q = 0$ as well. All results are based on 10,000 repetitions.

Two different test statistics are considered: the scaled t-statistic corresponding to the long-run OLS estimate $\hat{\beta}_q$ (i.e., $t_q/\sqrt{q}$), and the scaled Bonferroni t-statistic described previously (i.e., $t^+_{q,\min}/\sqrt{q}$). All tests are evaluated against a positive 1-sided alternative at the 5% level (i.e., the null is rejected if the scaled test statistic exceeds 1.645).

The results are given in Table 2. The 1st set of columns shows the rejection rates for the scaled OLS t-statistic under the null hypothesis of no predictability. When the regressors are exogenous, such that $\delta = 0$, this test statistic should be asymptotically normally distributed. The normal distribution appears to work well in finite samples, with rejection rates close to the nominal 5% size. For larger negative cs, and for large q relative to T, the size drops and the test becomes somewhat conservative; this is primarily true for forecasting horizons that span more than 10% of the sample size.[13] Overall, however, the scaling by $1/\sqrt{q}$ of the standard t-test appears to work well in practice for exogenous regressors. As is expected from the previous asymptotic analysis, the scaled OLS t-test tends to overreject for endogenous regressors with $\delta < 0$, which highlights that the

[13] As mentioned in footnote 9, for larger negative cs, the scale factor $1/\sqrt{q}$ will overcontrol for the overlap in the data, and the difference between the scale factor $1/\sqrt{f(q,\rho)}$ derived for the stationary case and the factor $1/\sqrt{q}$ used here increases as q increases. As pointed out previously, the practical implementation of the $1/\sqrt{f(q,\rho)}$ scaling is made difficult by the need to estimate ρ, and the $1/\sqrt{q}$ scaling therefore may still be the best feasible approach, despite being slightly conservative.

https://doi.org/10.1017/S0022109011000135 Published online by Cambridge University Press

biasing effects of endogenous regressors are a great problem also in long-horizon regressions.

TABLE 2

Finite Sample Sizes for the Scaled Long-Run OLS *t*-Test and the Scaled Bonferroni Test

The first column of Table 2 gives the forecasting horizon q, and the top row below the labels gives the value of the parameter δ, the correlation between the innovation processes. The remaining entries present, for each combination of q and δ, the average rejection rates under the null hypothesis of no predictability for the corresponding test. The results are based on the Monte Carlo simulation described in the main text, and the average rejection rates are calculated over 10,000 repetitions. Results for sample sizes T equal to 100 and 500 and for local-to-unity parameters $c = -2.5, -10, -20$, and $c = -1, -5, -20$, respectively, are given; for $T = 100$, these values of c correspond to autoregressive roots $\rho = 0.975$, 0.9, and 0.8, respectively, and for $T = 500$, they correspond to $\rho = 0.998$, 0.99, and $= 0.96$, respectively.

	Long-Run OLS *t*-Test					Bonferroni Test				
δ	0.00	−0.7	−0.90	−0.95	−0.99	0.00	−0.7	−0.90	−0.95	−0.99
q	$T = 100$, $c = -2.5$									
1	0.052	0.190	0.245	0.264	0.279	0.045	0.035	0.033	0.035	0.029
5	0.048	0.187	0.253	0.271	0.285	0.041	0.032	0.026	0.028	0.025
10	0.046	0.185	0.254	0.274	0.293	0.040	0.029	0.021	0.019	0.018
20	0.034	0.163	0.230	0.254	0.270	0.038	0.020	0.011	0.007	0.004
q	$T = 500$, $c = -1$									
1	0.049	0.224	0.311	0.337	0.358	0.047	0.043	0.040	0.041	0.029
25	0.050	0.230	0.326	0.353	0.380	0.044	0.044	0.036	0.035	0.025
50	0.051	0.235	0.337	0.372	0.401	0.046	0.042	0.030	0.025	0.018
100	0.046	0.216	0.322	0.363	0.399	0.049	0.035	0.022	0.015	0.007
q	$T = 100$, $c = -10$									
1	0.054	0.114	0.139	0.143	0.147	0.047	0.024	0.023	0.026	0.021
5	0.041	0.100	0.123	0.127	0.129	0.035	0.017	0.016	0.016	0.016
10	0.032	0.085	0.105	0.112	0.115	0.032	0.012	0.008	0.008	0.009
20	0.015	0.052	0.068	0.071	0.072	0.020	0.007	0.003	0.002	0.002
q	$T = 500$, $c = -5$									
1	0.048	0.142	0.176	0.187	0.195	0.045	0.029	0.034	0.038	0.031
25	0.040	0.133	0.174	0.185	0.199	0.035	0.023	0.025	0.028	0.025
50	0.037	0.126	0.172	0.186	0.196	0.034	0.018	0.016	0.017	0.015
100	0.029	0.103	0.142	0.153	0.160	0.031	0.013	0.007	0.005	0.004
q	$T = 100$, $c = -20$									
1	0.053	0.091	0.104	0.106	0.108	0.046	0.019	0.019	0.019	0.017
5	0.030	0.060	0.069	0.072	0.074	0.026	0.011	0.008	0.010	0.011
10	0.017	0.037	0.045	0.047	0.048	0.018	0.007	0.003	0.003	0.004
20	0.003	0.014	0.014	0.015	0.015	0.007	0.002	0.001	0.001	0.000
q	$T = 500$, $c = -20$									
1	0.050	0.088	0.102	0.105	0.111	0.046	0.022	0.025	0.028	0.022
25	0.028	0.059	0.073	0.075	0.080	0.025	0.011	0.010	0.012	0.014
50	0.016	0.040	0.048	0.051	0.052	0.018	0.007	0.005	0.005	0.007
100	0.004	0.013	0.018	0.019	0.020	0.009	0.001	0.001	0.001	0.001

The next set of columns shows the results for the scaled Bonferroni test. The rejection rates for all δ are now typically close to and below 5%, indicating that the proposed correction in the augmented regression equation (10) works well in finite samples. When δ is close to -1, the Bonferroni test tends to become more conservative. Since the coverage rates for c that are used in the Bonferroni test (see Table 1) are obtained by finding the largest α_1 such that the Bonferroni test has size $\leq 5\%$ across a grid of cs, the test will in general be conservative for a given c. As $\delta \to -1$, the outcome of the test statistic becomes much more sensitive to the exact value of c at which it is evaluated, reflected by the greater coverage rates $(1-\alpha_1)$ in Table 1, and the conservative nature of the test is therefore exacerbated.

https://doi.org/10.1017/S0022109011000135 Published online by Cambridge University Press

https://doi.org/10.1017/S0022109011000135 Published online by Cambridge University Press

As in the OLS case, the Bonferroni test also has a tendency to underreject for large q relative to T.

To evaluate how the tests perform when the innovations are not normally distributed, the same simulation exercise as previously conducted is performed with u_t and v_t following t-distributions with 3 degrees of freedom, standardized to have unit population variances. Again, the innovations have correlations δ, and all other aspects of the simulations remain identical. The results are shown in Table 3 and are very similar to those for the normal case shown in Table 2. The proposed methods thus appear to work well also in nonnormal cases, and with fairly heavy-tailed distributions.[14]

In summary, the previous simulations confirm the main conclusions from the formal asymptotic analysis: i) When the regressor is exogenous, the standard t-statistic scaled by the square root of the forecasting horizon will be normally distributed, and ii) when the regressor is endogenous, the scaled t-statistic corresponding to the augmented regression equation will be normally distributed.

B. Power Properties

Since the test procedures proposed in this paper appear to have good size properties and, if anything, underreject rather than overreject the null, the second important consideration is their power to reject the null when the alternative is in fact true. Again, the same simulation design is used, with the data generated by equations (2) and (3). Unless otherwise noted, the innovations are normally distributed. In order to assess the power of the tests, however, the slope coefficient β in equation (2) now varies between 0 and 0.2.

In addition to the scaled t-statistics considered in the size simulations, I now also study an additional test statistic based on Valkanov (2003). Valkanov derives his asymptotic results under the assumption that $q/T \to \lambda \in (0, 1)$ as $q, T \to \infty$, and he shows that under this assumption, $t/\sqrt{T}$ will have a well-defined distribution. That is, he proposes to scale the standard OLS t-statistic by the square root of the sample size, rather than by the square root of the forecasting horizon, as suggested in this paper. The scaled t-statistic in Valkanov's analysis is not normally distributed. Its asymptotic distribution is a function of the parameters λ (the degree of overlap), the local-to-unity parameter c, and the degree of endogeneity δ; critical values must be obtained by simulation for a given combination of these 3 parameters. Since the critical values are a function of c, which is unknown, Valkanov's scaled t-test is generally infeasible. He therefore proposes a so-called sup-bound test, where the test is evaluated at some bound for c, outside of which it is assumed that c will not lie. Ruling out explosive processes, he suggests using $c = 0$ in the sup-bound test, which results in a conservative 1-sided test against $\beta > 0$ for $\delta < 0$.[15] To avoid confusion, I will continue to refer to the tests proposed in this paper as scaled tests, whereas I will refer to the test suggested by

[14] As pointed out before, no normality assumptions are needed for the previous derivations, although the t-distribution with 3 degrees of freedom does violate the condition of finite 4th-order moments in Assumption 1 in Appendix B. These results thus show that this is in general not a necessary condition for the proposed methods to work.

[15] Lewellen (2004) suggests a similar procedure in 1-period (short-run) regressions.

TABLE 3

Finite Sample Sizes for the Scaled Long-Run OLS *t*-Test and the Scaled Bonferroni Test with *t*-Distributed Innovations with 3 Degrees of Freedom

The first column of Table 3 gives the forecasting horizon q, and the top row below the labels gives the value of the parameter δ, the correlation between the innovation processes. The remaining entries present, for each combination of q and δ, the average rejection rates under the null hypothesis of no predictability for the corresponding test. The results are based on the Monte Carlo simulation described in the main text, and the average rejection rates are calculated over 10,000 repetitions. Results for sample sizes T equal to 100 and 500 and for local-to-unity parameters $c = -2.5, -10, -20$, and $c = -1, -5, -20$, respectively, are given; for $T = 100$, these values of c correspond to autoregressive roots $\rho = 0.975$, 0.9, and 0.8, respectively, and for $T = 500$, they correspond to $\rho = 0.998$, 0.99, and $= 0.96$, respectively.

	Long-Run OLS *t*-Test					Bonferroni Test				
δ	0.00	−0.7	−0.90	−0.95	−0.99	0.00	−0.7	−0.90	−0.95	−0.99
q	$T = 100, c = -2.5$									
1	0.051	0.181	0.231	0.248	0.255	0.043	0.032	0.031	0.029	0.027
5	0.049	0.184	0.239	0.256	0.266	0.041	0.029	0.025	0.024	0.023
10	0.048	0.182	0.247	0.259	0.270	0.042	0.033	0.020	0.017	0.016
20	0.036	0.160	0.221	0.243	0.256	0.033	0.021	0.010	0.007	0.005
q	$T = 500, c = -1$									
1	0.049	0.228	0.311	0.334	0.358	0.047	0.045	0.042	0.041	0.029
25	0.053	0.234	0.327	0.354	0.381	0.047	0.044	0.037	0.033	0.026
50	0.053	0.234	0.331	0.366	0.398	0.046	0.045	0.032	0.025	0.018
100	0.043	0.221	0.326	0.365	0.403	0.050	0.041	0.024	0.017	0.005
q	$T = 100, c = -10$									
1	0.052	0.107	0.124	0.130	0.134	0.044	0.024	0.024	0.023	0.023
5	0.040	0.094	0.111	0.118	0.120	0.033	0.020	0.016	0.016	0.018
10	0.035	0.078	0.096	0.101	0.104	0.031	0.014	0.010	0.009	0.011
20	0.016	0.051	0.059	0.060	0.061	0.021	0.008	0.004	0.003	0.003
q	$T = 500, c = -5$									
1	0.053	0.139	0.177	0.184	0.192	0.049	0.033	0.033	0.035	0.030
25	0.047	0.136	0.172	0.184	0.192	0.039	0.026	0.024	0.026	0.023
50	0.044	0.129	0.165	0.180	0.187	0.039	0.021	0.016	0.017	0.015
100	0.031	0.108	0.140	0.151	0.158	0.035	0.014	0.008	0.006	0.004
q	$T = 100, c = -20$									
1	0.052	0.082	0.090	0.094	0.099	0.044	0.018	0.017	0.019	0.018
5	0.030	0.059	0.065	0.070	0.072	0.027	0.012	0.008	0.009	0.012
10	0.018	0.035	0.040	0.042	0.043	0.018	0.008	0.004	0.004	0.005
20	0.003	0.012	0.015	0.016	0.015	0.007	0.003	0.001	0.000	0.001
q	$T = 500, c = -20$									
1	0.053	0.090	0.100	0.101	0.101	0.050	0.025	0.024	0.025	0.021
25	0.031	0.060	0.071	0.074	0.075	0.027	0.011	0.008	0.011	0.011
50	0.017	0.037	0.045	0.048	0.049	0.018	0.007	0.004	0.004	0.006
100	0.005	0.014	0.019	0.019	0.020	0.008	0.004	0.001	0.001	0.001

Valkanov explicitly as Valkanov's (sup-bound) test. Following Valkanov's exposition, I focus on the case of $q/T = 0.1$. For simplicity, I only consider the cases of $\delta = 0$ and $\delta = -0.9$, and $T = 100$.

Figure 2 shows the power curves for the scaled OLS t-test proposed in this paper and the test suggested by Valkanov (2003) for $\delta = 0$, $q = 10$, and $T = 100$. For both $c = -2.5$ and $c = -10$, the scaled OLS t-test is marginally more powerful than Valkanov's sup-bound test. Overall, for the case of exogenous regressors, there appears to be no loss of power from using the simple scaled and normally distributed t-test suggested here, but, in fact, some marginal power gains.

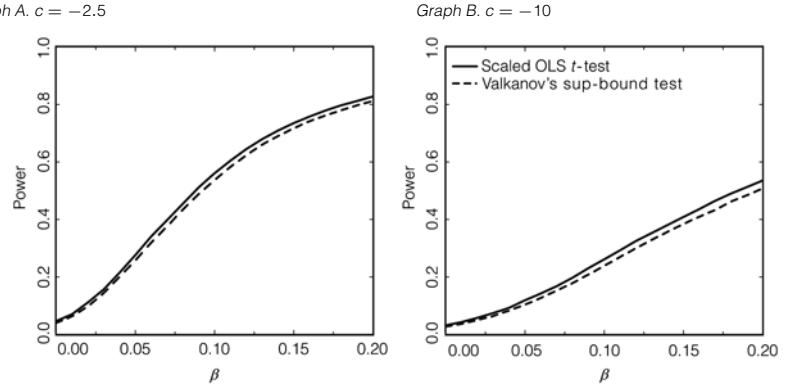
Figure 3 shows the results for endogenous regressors with $\delta = -0.9$, $q = 10$, and $T = 100$. Since the scaled t-test based on the OLS estimator is known to be biased in this case, I only show the results for the scaled Bonferroni test, along

https://doi.org/10.1017/S0022109011000135 Published online by Cambridge University Press

FIGURE 2

Power Curves for Exogenous Regressors with $T = 100$, $q = 10$, and $\delta = 0.0$

Graphs A and B of Figure 2 show the average rejection rates for a 1-sided 5% test of the null hypothesis of $\beta = 0$ against a positive alternative. The *x*-axis shows the true value of the parameter β, and the *y*-axis indicates the average rejection rate. Graph A shows the results for the case of $c = -2.5$ ($\rho = 0.975$), and Graph B shows the results for $c = -10$ ($\rho = 0.9$). The results for the scaled OLS *t*-test derived in this paper are indicated by the solid lines and the results for Valkanov's (2003) sup-bound test are indicated by the dashed lines. The results are based on the Monte Carlo simulations described in the main text, and the power is calculated as the average rejection rates over 10,000 repetitions.

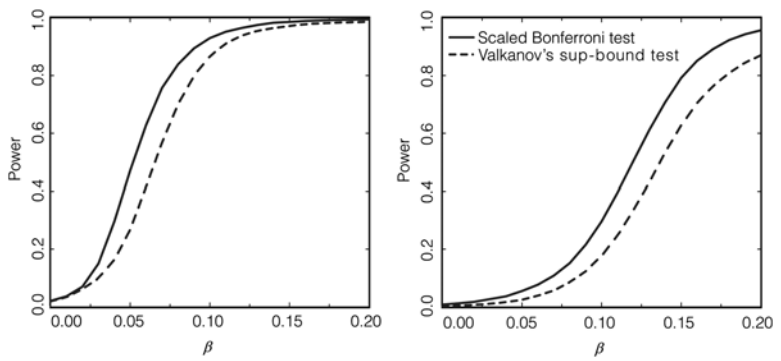


with Valkanov's (2003) test. In this endogenous case, Valkanov's sup-bound test is clearly dominated by the scaled Bonferroni test.

FIGURE 3

Power Curves for Endogenous Regressors with $T = 100$, $q = 10$, and $\delta = -0.9$

Graphs A and B of Figure 3 show the average rejection rates for a 1-sided 5% test of the null hypothesis of $\beta = 0$ against a positive alternative. The *x*-axis shows the true value of the parameter β, and the *y*-axis indicates the average rejection rate. Graph A shows the results for the case of $c = -2.5$ ($\rho = 0.975$), and Graph B shows the results for $c = -10$ ($\rho = 0.9$). The results for the scaled Bonferroni test are indicated by the solid lines and the results for Valkanov's (2003) sup-bound test are indicated by the dashed lines. The results are based on the Monte Carlo simulations described in the main text, and the power is calculated as the average rejection rates over 10,000 repetitions.



To sum up, the simulations show that both the scaled OLS *t*-test and the scaled Bonferroni test have good power properties when compared to the test proposed by Valkanov (2003). This is especially true for the Bonferroni test used with endogenous regressors, which tends to dominate Valkanov's test.

https://doi.org/10.1017/S0022109011000135 Published online by Cambridge University Press

C. Power across Different Horizons

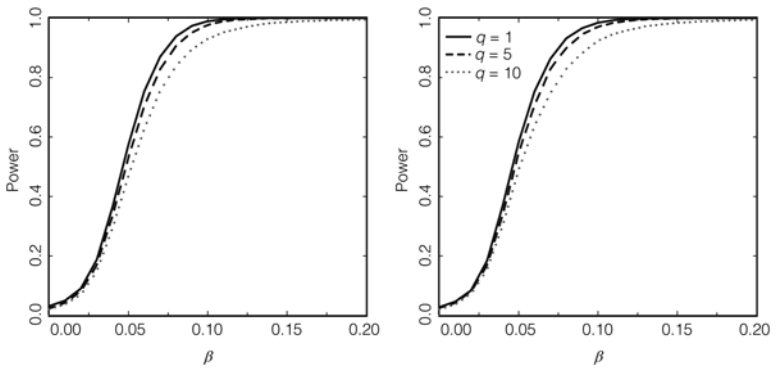
Figure 4 compares power for the scaled Bonferroni test across different forecasting horizons, including the 1-period horizon. Graph A shows the case with normal innovations, and Graph B shows the case with t-distributed innovations with 3 degrees of freedom; again, the t-distributed innovations are standardized to have unit population variances. The sample size is still $T = 100$ and $c = -2.5$. Only the case with endogenous regressors, letting $\delta = -0.9$, is shown.

FIGURE 4

Power Curves for Different Forecasting Horizons q with $T = 100$, $\delta = -0.9$, and $c = -2.5$ ($\rho = 0.975$)

Graphs A and B of Figure 4 show the average rejection rates for a 1-sided 5% test of the null hypothesis of $\beta = 0$ against a positive alternative. Only results for the scaled Bonferroni test are shown. The x-axis shows the true value of the parameter β, and the y-axis indicates the average rejection rate. Graph A shows the results for the case with normal innovations, and Graph B shows the case with t-distributed innovations with 3 degrees of freedom, standardized to have unit population variances. The results for the 1-period short-horizon tests ($q = 1$) are indicated by the solid lines, the results for the long-horizon tests with $q = 5$ are indicated by the dashed lines, and the results for the long-horizon tests with $q = 10$ are indicated by the dotted lines. The results are based on the Monte Carlo simulations described in the main text, and the power is calculated as the average rejection rates over 10,000 repetitions.

Graph A. Normal Innovations *Graph B. t-Distributed Innovations*



Two conclusions are immediate from the study of Figure 4. First, power is decreasing in the forecasting horizon: The 1-period test is marginally more powerful than the long-horizon test with $q = 5$, which in turn is somewhat more powerful than the long-horizon test with $q = 10$. Second, this conclusion holds irrespective of whether the innovations are normal or not. In fact, the power curves are very similar for the normal case and the t-distributed case.

Overall, it thus seems difficult to achieve power gains for longer horizons when the model under the alternative hypothesis is given by equation (2), which is the standard model for evaluating long-horizon regressions (e.g., Campbell (2001), Nelson and Kim (1993)), both with normal and nonnormal innovations.[16]

[16]In the Gaussian case, this is certainly not surprising. As mentioned in footnote 7, when the true model is given by equation (2), the 1-period ($q = 1$) t-test based on the augmented regression equation is optimal in the Gaussian case when the persistence ρ is known. Although this does not imply that the corresponding feasible Bonferroni test is optimal for the case of unknown ρ, it is not surprising if the Bonferroni test based on the short-run regression dominates similar feasible tests (e.g., the long-run Bonferroni tests).

https://doi.org/10.1017/S0022109011000135 Published online by Cambridge University Press

But, the reason for running long-horizon regressions is presumably that one believes the model is not as simple as that given by equation (2) under the alternative, and other alternatives may exist where the long-horizon tests outperform the short-horizon tests.[17]

A more exhaustive discussion of power properties across different horizons is outside the scope of the current paper, which focuses on obtaining correctly sized tests under the null hypothesis of no predictability, a necessary start for analyzing power properties.

V. Long-Run Stock Return Predictability

To illustrate the methods derived in this paper, I revisit the question of stock return predictability. I focus on the predictive ability of traditional valuation ratios, and use aggregate U.S. data for the dividend-price (DP) ratio, the earnings-price (EP) ratio, and the book-to-market (BM) ratio. The returns data consist of annual excess returns, over the T-Bill rate, on the Standard & Poor's (S&P) 500 index. These data are a subset of those used by Goyal and Welch (2008), although the series have been updated to include data up to and including 2007; a more detailed description of each series is provided in their paper.[18] The return series, as well as the DP and EP ratios, are available from 1871 onward, and the BM ratio is available from 1921 onward. All data are on an annual frequency, and I consider forecast horizons from 1 to 10 years. All regressions are run using log-transformed variables with log excess returns as the dependent variable.

The 2 key data characteristics that define the properties of the regression estimators analyzed in this paper are the near persistence and endogeneity of the regressors. Table 4 presents point estimates as well as confidence intervals for the autoregressive root ρ, and the analogue intervals for the local-to-unity parameter c, calculated by inverting the Chen and Deo (2009a), (2009b) unit-root test (see Appendix A). Estimates of the correlation δ between the innovations to returns and the innovations to the regressors are also provided. As is evident, there tends to be a large negative correlation between the innovations to the returns and the valuation ratios, although somewhat less so for the EP ratio. In addition, the predictor variables show signs of having autoregressive roots that are close to unity. Standard OLS inference is thus likely to be biased for all 3 predictor variables, but probably less so for the EP ratio.

The scaled t-statistics from the long-run predictive regressions are presented in Table 5 for each horizon $q = 1, \ldots, 10$. For each horizon and forecasting variable, 3 different t-statistics are presented: i) the scaled OLS t-statistic, ii) the scaled Bonferroni t-statistic, and iii) the nonscaled Newey and West (1987) t-statistic, calculated using q lags, which is often used instead of scaling. The Bonferroni test statistic is calculated in the same manner as described in

[17]For instance, Paye and Timmermann (2006) find evidence of coefficient instability in (short-run) predictive regressions for stock returns. In this case, equation (2) is no longer the correct model under the alternative, and it is possible that such DGPs may in some instances give power advantages to long-run tests in detecting deviations from the null of no predictability.

[18]The data were obtained from Amit Goyal's Web site (http://www.hec.unil.ch/agoyal/).

https://doi.org/10.1017/S0022109011000135 Published online by Cambridge University Press

TABLE 4

Characteristics of the Predictor Variables

Table 4 reports the key time-series characteristics of the dividend-price (DP) ratio, the earnings-price (EP) ratio, and the book-to-market (BM) ratio. The first 3 columns indicate the predictor variable being used, the sample period, and the number of observations in that sample. All data are on an annual frequency. The column labeled $\widehat{\delta}$ gives the estimated correlations between the innovations to the predictor variables and the innovations to the corresponding excess returns. The column labeled $\widehat{\rho}_{\text{OLS}}$ gives the OLS estimates of the autoregressive root ρ. The last 2 columns give the 95% confidence intervals for the autoregressive root ρ and the corresponding local-to-unity parameter c, obtained by inverting the Chen and Deo (2009a), (2009b) unit-root test statistic.

Variable	Sample Period	Obs.	$\widehat{\delta}$	$\widehat{\rho}_{\text{OLS}}$	95% CI for ρ	95% CI for c
DP	1872–2007	136	−0.836	0.882	[0.809, 0.986]	[−25.976, −1.904]
EP	1872–2007	136	−0.436	0.731	[0.622, 0.859]	[−51.408, −19.176]
BM	1921–2007	87	−0.864	0.906	[0.822, 1.000]	[−15.486, 0.000]

Section III and should be approximately normally distributed for all predictor variables, whereas for the scaled OLS t-test, the normal approximation is unlikely to hold given the persistence and endogeneity of the predictors.

TABLE 5

Empirical Results for Tests of Long-Run Predictability

Table 5 presents the outcomes of the long-run test statistics as functions of the forecasting horizon (q), given in years in the 1st column. Results for the scaled OLS t-test ($t_{q,\text{OLS}}/\sqrt{q}$), the scaled Bonferroni test ($t^{+}_{q,\text{min}}/\sqrt{q}$), and the (nonscaled) t-test using Newey and West (1987) standard errors ($t_{q,\text{NW}}$) are presented. Three different predictor variables are considered: the dividend-price (DP) ratio, the earnings-price (EP) ratio, and the book-to-market (BM) ratio. All data are on an annual frequency. The sample periods for the DP ratio and the EP ratio start in 1872, and the sample period for the BM ratio starts in 1921. All samples end in 2007.

	DP			EP			BM		
Forecast Horizon (q)	$t_{q,\text{OLS}}/\sqrt{q}$	$t^{+}_{q,\text{min}}/\sqrt{q}$	$t_{q,\text{NW}}$	$t_{q,\text{OLS}}/\sqrt{q}$	$t^{+}_{q,\text{min}}/\sqrt{q}$	$t_{q,\text{NW}}$	$t_{q,\text{OLS}}/\sqrt{q}$	$t^{+}_{q,\text{min}}/\sqrt{q}$	$t_{q,\text{NW}}$
1	1.395	−0.399	1.331	1.583	1.117	1.698	1.884	0.365	1.817
2	1.701	0.488	1.934	1.850	1.660	2.479	1.721	0.033	1.851
3	1.498	0.828	1.902	1.616	1.550	2.118	1.634	−0.029	1.829
4	1.578	1.096	2.246	1.351	1.309	1.889	1.584	−0.011	1.979
5	1.745	1.441	2.749	1.304	1.307	2.027	1.469	0.021	2.215
6	1.730	1.605	2.790	1.319	1.375	2.046	1.424	0.093	2.441
7	1.644	1.584	2.677	1.282	1.326	1.949	1.352	0.279	2.314
8	1.604	1.539	2.537	1.373	1.378	2.234	1.209	0.281	2.077
9	1.391	1.398	2.151	1.530	1.519	2.868	1.050	0.202	1.807
10	1.191	1.228	1.762	1.592	1.557	3.093	0.949	0.149	1.621

The test results in Table 5 show that there is no robust evidence of return predictability using any of the 3 valuation ratios. The robust scaled Bonferroni t-statistic is virtually never significant, with only a single outcome (for the EP ratio at the 2-year horizon) greater than 1.645, which indicates the cutoff for significance in the 1-sided 5% level test. Inference based on the scaled OLS t-statistics would lead to a somewhat more positive view of predictability, although the overall evidence is still weak, with no outcomes greater than 1.9. Of course, the OLS results are not robust to the endogeneity and persistence of the regressors, and they are thus likely to be biased.

In general, the scaled t-statistics show no signs of systematically increasing with the horizon. In contrast, using the (nonscaled) t-statistics based on Newey and West (1987) standard errors, the results would suggest that predictability is

https://doi.org/10.1017/S0022109011000135 Published online by Cambridge University Press

much stronger, and often significant, at longer horizons. It is apparent that Newey-West standard errors can fail substantially in controlling the size of long-horizon tests, and the Newey-West t-statistics also clearly illustrate why long-run predictability is often thought to be stronger than short-run predictability.

The results in Table 5 thus illustrate 2 key findings. First, and contrary to many popular beliefs, the evidence of predictability does not typically become stronger at longer forecasting horizons.[19] There are some exceptions, such as the results for the Bonferroni t-statistics for the DP ratio, but overall there is little tendency for the results to look stronger at longer horizons. Second, these results show that it is also important to control for the biasing effect of persistent and endogenous regressors in long-horizon regressions, as seen from the sometimes large differences between the OLS and the Bonferroni test statistics.

VI. Conclusion

I derive several new results for long-horizon regressions that use overlapping observations. In particular, I show how to properly correct for the overlap in the data in a simple manner that obviates the need for autocorrelation robust standard error methods in these regressions. Further, when the regressors are persistent and endogenous, I show how to correct the bias in the long-run test procedures, using methods similar to those previously proposed for the short-run case. Since the proposed test statistics are asymptotically normally distributed, these new methods lead to both more efficient and simpler inference in long-horizon regressions. Finally, the new procedures are shown to work well in finite samples, with average rejection rates for the long-horizon tests that are close to the nominal size, even when the forecasting horizon is large relative to the sample size.

Appendix A. Implementing the Chen and Deo Test

The so-called restricted log-likelihood of equation (3) is given by

$$L(\mathbf{x},\rho,\omega_{22}) = -\frac{T-1}{2}\log\omega_{22}+\frac{1}{2}\log\left\{\frac{1+\rho}{(T-2)(1-\rho)+2}\right\}-\frac{1}{2\omega_{22}}Q(\rho),$$

where

$$Q(\rho) = \left(1-\rho^2\right)x_1^2+\sum_{t=2}^{T}(x_t-\rho x_{t-1})^2 - \frac{1-\rho}{(T-2)(1-\rho)+2}\left\{x_1+x_T+(1-\rho)\sum_{t=2}^{T-1}x_t\right\}.$$

The RLRT for testing the null hypothesis $H_0:\rho=\rho_0$ against $H_1:\rho\neq\rho_0$ is given by

$$R_T(\rho_0) = 2L(\mathbf{x},\hat{\rho},\hat{\omega}_{22})-2L(\mathbf{x},\rho_0,\hat{\omega}_{22,0}),$$

[19] This paper is not the first to make this comment. Torous, Valkanov, and Yan (2004) and Ang and Bekaert (2007) find that the evidence of predictability tends to be strongest at shorter horizons. Boudoukh et al. (2008) explicitly question the prevailing view of long-horizon predictability.

https://doi.org/10.1017/S0022109011000135 Published online by Cambridge University Press

where $L(\cdot)$ is the restricted log-likelihood function given previously and $(\hat{\rho}, \hat{\omega}_{22})$ and $(\rho_0, \hat{\omega}_{22,0})$ are the unconstrained and constrained restricted maximum likelihood (REML) estimates, respectively. The unconstrained REML estimate $\hat{\rho}$ is given by

$$\hat{\rho} = \arg \min_{\rho \in (-1,1]} \left\{ (T-1) \log Q(\rho) - \log \left(\frac{1+\rho}{(T-2)(1-\rho)+2} \right) \right\},$$

and

$$\hat{\omega}_{22} = \frac{1}{T-1} Q(\hat{\rho}), \qquad \hat{\omega}_{22,0} = \frac{1}{T-1} Q(\rho_0).$$

Chen and Deo (2009b) show that, for $\rho \leq 1$, the distribution of $R_T(\rho)$ is approximately χ^2 with 1 degree of freedom, under the null hypothesis. A confidence interval for ρ can therefore be obtained by inverting the acceptance region of R_T, using the χ^2_1 distribution. A confidence interval for c is then given by the transformation $c = T(\rho - 1)$.

Appendix B. Proofs

For ease of notation, the case with no intercept is treated. The results generalize immediately to regressions with fitted intercepts by replacing all variables by their demeaned versions. Throughout the proofs, q is treated as fixed, and thus $qT^{-1} = o(1)$. The results below are derived under the following formal assumption for the error processes.

Assumption 1. Let $w_t = (u_t, v_t)'$ and $\mathcal{F}_t = \{w_s | s \leq t\}$ be the filtration generated by w_t. Then:

1. $\mathrm{E}[w_t | \mathcal{F}_{t-1}] = 0$.
2. $\mathrm{E}[w_t w_t'] = \Omega = [(\omega_{11}, \omega_{12}), (\omega_{12}, \omega_{22})]$.
3. $\sup_t \mathrm{E}[u_t^4] < \infty$, $\sup_t \mathrm{E}[v_t^4] < \infty$, and $\mathrm{E}[x_0^2] < \infty$.

Proof of Theorem 1. Under the null hypothesis,

$$\begin{aligned} \frac{T}{q}\left(\hat{\beta}_q - 0\right) &= \left(\frac{1}{qT} \sum_{t=1}^{T-q} u_{t+q}(q)\, x_t \right) \left(\frac{1}{T^2} \sum_{t=1}^{T-q} x_t^2 \right)^{-1} \\ &= \left(\frac{1}{qT} \sum_{t=1}^{T-q} \sum_{j=1}^{q} u_{t+j} x_t \right) \left(\frac{1}{T^2} \sum_{t=1}^{T-q} x_t^2 \right)^{-1}. \end{aligned}$$

By standard arguments, $(1/(qT)) \sum_{t=1}^{T-q} \sum_{j=1}^{q} u_{t+j} x_t = (1/(qT)) \sum_{t=1}^{T-q} (u_{t+1} x_t + \ldots + u_{t+q} x_t) \Rightarrow \int_0^1 dB_1 J_c$, as $T \to \infty$, since for any $h > 0$, $(1/T) \sum_{t=1}^{T-q} u_{t+h} x_t \Rightarrow \int_0^1 dB_1 J_c$ (Phillips (1987), (1988)). Therefore, $(T/q)(\hat{\beta}_q - 0) \Rightarrow (\int_0^1 dB_1 J_c)(\int_0^1 J_c^2)^{-1}$. □

Proof of Theorem 2. Let $\mathbf{r}_{+q}^q = (r_{1+q}(q), \ldots, r_{T-q}(q))'$ be the $(T-q) \times 1$ vector of observations, and define $\mathbf{x}$ and $\mathbf{v}_{+q}^q$ analogously. Also, let $Q_{\mathbf{v}^q} = \mathbf{I} - \mathbf{v}_{+q}^q (\mathbf{v}_{+q}^{q\prime} \mathbf{v}_{+q}^q)^{-1} \mathbf{v}_{+q}^{q\prime}$. The OLS estimator of β_q in equation (10) is now given by $\hat{\beta}_q^+ = (\mathbf{r}_{+q}^{q\prime} Q_{\mathbf{v}^q} \mathbf{x})(\mathbf{x}' Q_{\mathbf{v}^q} \mathbf{x})^{-1}$. Under the null hypothesis, $Q_{\mathbf{v}^q} \mathbf{r}_{+q}^q = Q_{\mathbf{v}^q} \mathbf{u}_{+q}^q$, and thus $(T/q)(\hat{\beta}_q^+ - 0) = (q^{-1} T^{-1} \mathbf{u}_{+q}^{q\prime} Q_{\mathbf{v}^q} \mathbf{x})(T^{-2} \mathbf{x}' Q_{\mathbf{v}^q} \mathbf{x})^{-1}$. First,

$$\begin{aligned} Q_{\mathbf{v}^q} \mathbf{u}_{+q}^q &= \mathbf{u}_{+q}^q - \mathbf{v}_{+q}^q \left(\mathbf{v}_{+q}^{q\prime} \mathbf{v}_{+q}^q \right)^{-1} \mathbf{v}_{+q}^{q\prime} \mathbf{u}_{+q}^q \\ &= \mathbf{u}_{+q}^q - \mathbf{v}_{+q}^q \left(\frac{1}{qT} \sum_{t=1}^{T-q} v_{t+q}(q)^2 \right)^{-1} \left(\frac{1}{qT} \sum_{t=1}^{T-q} u_{t+q}(q)\, v_{t+q}(q) \right). \end{aligned}$$

https://doi.org/10.1017/S0022109011000135 Published online by Cambridge University Press

Let $\hat{\gamma}_{vv}^{k}(h) = (1/T)\sum_{t=1}^{T-q} v_{t+k}v_{t+k+h}$ and denote $\hat{\gamma}_{vv}(h) = \hat{\gamma}_{vv}^{1}(h)$. By some algebraic manipulations,

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T-q} v_{t+q}(q)^2 &= \hat{\gamma}_{vv}^{1}(0) + \hat{\gamma}_{vv}^{2}(0) + \cdots + \hat{\gamma}_{vv}^{q-1}(0) + \hat{\gamma}_{vv}^{q}(0)\\ &\quad + \hat{\gamma}_{vv}^{2}(-1) + \hat{\gamma}_{vv}^{3}(-1) + \cdots + \hat{\gamma}_{vv}^{q}(-1)\\ &\quad \vdots\\ &\quad + \hat{\gamma}_{vv}^{q}(-(q-1))\\ &\quad + \hat{\gamma}_{vv}^{1}(q-1)\\ &\quad \vdots\\ &\quad + \hat{\gamma}_{vv}^{1}(1) + \hat{\gamma}_{vv}^{2}(1) + \cdots + \hat{\gamma}_{vv}^{q-1}(1).\end{aligned}$$

Now, observe that $\hat{\gamma}_{vv}(h) = \hat{\gamma}_{vv}^{1}(h) = (1/T)\sum_{t=1}^{T-q} v_{t+1}v_{t+h+1}$ and $\hat{\gamma}_{vv}^{k}(h) = (1/T)\sum_{t=1}^{T-q} v_{t+k}v_{t+k+h} = (1/T)\sum_{t=k}^{T-q+k-1} v_{t+1}v_{t+h+1}$. So $\hat{\gamma}_{vv}^{k}(h)$ is thus an identical estimator to $\hat{\gamma}_{vv}(h)$, but uses observations shifted k steps. Letting "$\stackrel{d}{=}$" denote distributional equivalence, it follows that

$$\begin{aligned}\frac{1}{qT}\sum_{t=1}^{T-q} v_{t+q}(q)^2 &\stackrel{d}{=} \frac{1}{q}\left[q\hat{\gamma}_{vv}(0) + (q-1)\hat{\gamma}_{vv}(-1) + \cdots + \hat{\gamma}_{vv}(-(q-1))\right.\\ &\quad \left. + \hat{\gamma}_{vv}(q-1) + \cdots + (q-1)\hat{\gamma}_{vv}(1)\right]\\ &= \sum_{h=-q+1}^{q-1}\left(1 - \frac{|h|}{q}\right)\hat{\gamma}_{vv}(h),\end{aligned}$$

and as $T \to \infty$, $(1/(qT))\sum_{t=1}^{T} v_{t+q}(q)^2 \stackrel{d}{=} \sum_{h=-q+1}^{q-1}(1 - (|h|/q))\hat{\gamma}_{vv}(h) \to_p \omega_{22}$, since $\hat{\gamma}_{vv}(h) \to_p 0$ for $h \neq 0$ by the martingale difference assumption. Similarly, as $T \to \infty$, $(1/(qT))\sum_{t=1}^{T-q} u_{t+q}(q)v_{t+q}(q) \to_p \omega_{12}$, and by the same arguments used in the previous proof, $(1/(qT))\mathbf{v}_{+q}^{q\prime}\mathbf{x} = (1/(qT))\sum_{t=1}^{T-q} v_{t+q}(q)x_t \Rightarrow \int_0^1 dB_2 J_c$. Thus,

$$\begin{aligned}q^{-1}T^{-1}\mathbf{u}_{+q}^{q\prime}Q_{\mathbf{v}^q}\mathbf{x} &= q^{-1}T^{-1}\mathbf{u}_{+q}^{q\prime}\mathbf{x}\\ &\quad - \left(q^{-1}T^{-1}\mathbf{u}_{+q}^{q\prime}\mathbf{v}_{+q}^{q}\right)\left(q^{-1}T^{-1}\mathbf{v}_{+q}^{q\prime}\mathbf{v}_{+q}^{q}\right)^{-1}\left(q^{-1}T^{-1}\mathbf{v}_{+q}^{q\prime}\mathbf{x}\right)\\ \Rightarrow \int_0^1 dB_1 J_c - \omega_{12}\omega_{22}^{-1}\int_0^1 dB_2 J_c &= \int_0^1 dB_{1\cdot 2} J_c.\end{aligned}$$

Finally, as $T \to \infty$, using the previous results, since $qT^{-1} = o(1)$, it follows that,

$$\begin{aligned}T^{-2}\mathbf{x}'Q_{\mathbf{v}^q}\mathbf{x} &= T^{-2}\mathbf{x}'\mathbf{x} - qT^{-1}\left(q^{-1}T^{-1}\mathbf{x}'\mathbf{v}_{+q}^{q}\right)\left(q^{-1}T^{-1}\mathbf{v}_{+q}^{q\prime}\mathbf{v}_{+q}^{q}\right)^{-1}\left(q^{-1}T^{-1}\mathbf{v}_{+q}^{q\prime}\mathbf{x}\right)\\ \Rightarrow \int_0^1 J_c^2. \ \square\end{aligned}$$

Proof of Corollary 1. Observe that under the null hypothesis, as $T \to \infty$,

$$\begin{aligned}\hat{\omega}_{11} &= \frac{1}{q(T-q)}\sum_{t=1}^{T-q}\hat{u}_{t+q}(q)^2 = \frac{1}{q(T-q)}\sum_{t=1}^{T-q}\left(u_{t+q}(q) + O_p\left(\frac{q}{T}\right)\right)^2\\ &= \frac{1}{q(T-q)}\sum_{t=1}^{T-q}u_{t+q}(q)^2 + O_p\left(\frac{q}{T}\right) \to_p \omega_{11},\end{aligned}$$

where the asymptotic limit follows by same argument as in the previous proof. $\square$

https://doi.org/10.1017/S0021090110001035 Published online by Cambridge University Press

Proof of Corollary 2. This follows in an identical manner, since, as $T \to \infty$,

$$\begin{aligned}\hat{\omega}_{11\cdot 2} &= \frac{1}{q(T-q)} \sum_{t=1}^{T-q} \hat{u}_{t+q\cdot v}^{+}(q)^2 \\ &= \frac{1}{q(T-q)} \sum_{t=1}^{T-q} \left(u_{t+q}(q) - \omega_{12}\omega_{22}^{-1} v_{t+q}(q)\right)^2 + O_p\left(\frac{q}{T}\right) \to_p \omega_{11\cdot 2}. \ \square\end{aligned}$$

References

Amihud, Y., and C. M. Hurvich. "Predictive Regressions: A Reduced-Bias Estimation Method." *Journal of Financial and Quantitative Analysis*, 39 (2004), 813–841.

Ang, A., and G. Bekaert. "Stock Return Predictability: Is It There?" *Review of Financial Studies*, 20 (2007), 651–707.

Baker, M.; R. Taliaferro; and J. Wurgler. "Predicting Returns with Managerial Decision Variables: Is There a Small-Sample Bias?" *Journal of Finance*, 61 (2006), 1711–1730.

Berkowitz, J., and L. Giorgianni. "Long-Horizon Exchange Rate Predictability?" *Review of Economics and Statistics*, 83 (2001), 81–91.

Boudoukh J., and M. Richardson. "Stock Returns and Inflation: A Long-Horizon Perspective." *American Economic Review*, 83 (1993), 1346–1355.

Boudoukh J.; M. Richardson; and R. F. Whitelaw. "The Myth of Long-Horizon Predictability." *Review of Financial Studies*, 21 (2008), 1577–1605.

Campbell, J. Y. "Why Long Horizons? A Study of Power Against Persistent Alternatives." *Journal of Empirical Finance*, 8 (2001), 459–491.

Campbell, J. Y., and R. Shiller. "Stock Prices, Earnings, and Expected Dividends." *Journal of Finance*, 43 (1988), 661–676.

Campbell, J. Y., and M. Yogo. "Efficient Tests of Stock Return Predictability." *Journal of Financial Economics*, 81 (2006), 27–60.

Cavanagh, C. L.; G. Elliot; and J. H. Stock. "Inference in Models with Nearly Integrated Regressors." *Econometric Theory*, 11 (1995), 1131–1147.

Chen, W. W., and R. S. Deo. "Bias Reduction and Likelihood-Based Almost Exactly Sized Hypothesis Testing in Predictive Regressions Using the Restricted Likelihood." *Econometric Theory*, 25 (2009a), 1143–1179.

Chen, W. W., and R. S. Deo. "The Restricted Likelihood Ratio Test at the Boundary in Autoregressive Series." *Journal of Time Series Analysis*, 30 (2009b), 618–630.

Elliot G.; T. J. Rothenberg; and J. H. Stock. "Efficient Tests for an Autoregressive Unit Root." *Econometrica*, 64 (1996), 813–836.

Fama, E. F., and K. R. French. "Dividend Yields and Expected Stock Returns." *Journal of Financial Economics*, 22 (1988), 3–25.

Fisher, M. E., and J. J. Seater. "Long-Run Neutrality and Superneutrality in an ARIMA Framework." *American Economic Review*, 83 (1993), 402–415.

Goyal, A., and I. Welch. "A Comprehensive Look at the Empirical Performance of Equity Premium Prediction." *Review of Financial Studies*, 21 (2008), 1455–1508.

Hansen, L. P., and R. J. Hodrick. "Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis." *Journal of Political Economy*, 88 (1980), 829–853.

Hjalmarsson, E. "Fully Modified Estimation with Nearly Integrated Regressors." *Finance Research Letters*, 4 (2007), 92–94.

Kalbfleisch, J. D., and D. A. Sprott. "Application of Likelihood Methods to Models Involving Large Numbers of Parameters." *Journal of the Royal Statistical Society B*, 32 (1970), 175–194.

Lewellen, J. "Predicting Returns with Financial Ratios." *Journal of Financial Economics*, 74 (2004), 209–235.

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

Mishkin, F. S. "What Does the Term Structure Tell Us About Future Inflation?" *Journal of Monetary Economics*, 25 (1990), 77–95.

Mishkin, F. S. "Is the Fisher Effect for Real? A Reexamination of the Relationship between Inflation and Interest Rates." *Journal of Monetary Economics*, 30 (1992), 195–215.

Nelson, C. R., and M. J. Kim. "Predictable Stock Returns: The Role of Small Sample Bias." *Journal of Finance*, 48 (1993), 641–661.

https://doi.org/10.1017/S0022109011000135 Published online by Cambridge University Press

Newey, W. K., and K. D. West. "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica*, 55 (1987), 703–708.
Paye, B. S., and A. Timmermann. "Instability of Return Prediction Models." *Journal of Empirical Finance*, 13 (2006), 274–315.
Phillips, P. C. B. "Towards a Unified Asymptotic Theory of Autoregression." *Biometrika*, 74 (1987), 535–547.
Phillips, P. C. B. "Regression Theory for Near-Integrated Time Series." *Econometrica*, 56 (1988), 1021–1043.
Phillips, P. C. B. "Optimal Inference in Cointegrated Systems." *Econometrica*, 59 (1991), 283–306.
Rossi, B. "Testing Long-Horizon Predictive Ability with High Persistence, and the Meese-Rogoff Puzzle." *International Economic Review*, 46 (2005), 61–92.
Stambaugh, R. F. "Predictive Regressions." *Journal of Financial Economics*, 54 (1999), 375–421.
Stock, J. H. "Confidence Intervals for the Largest Autoregressive Root in U.S. Economic Time Series." *Journal of Monetary Economics*, 28 (1991), 435–459.
Torous, W.; R. Valkanov; and S. Yan. "On Predicting Stock Returns with Nearly Integrated Explanatory Variables." *Journal of Business*, 77 (2004), 937–966.
Valkanov, R. "Long-Horizon Regressions: Theoretical Results and Applications." *Journal of Financial Economics*, 68 (2003), 201–232.

https://doi.org/10.1017/S0022109011000135 Published online by Cambridge University Press