Board of Governors of the Federal Reserve System

International Finance Discussion Papers

Number 853

January 2006

Revised October 2008

Inference in Long-Horizon Regressions

Erik Hjalmarsson

Inference in Long-Horizon Regressions

Erik Hjalmarsson*

Division of International Finance

Federal Reserve Board, Mail Stop 20, Washington, DC 20551, USA

First draft: December 16, 2004

This draft: October 16, 2008

Abstract

I develop new results for long-horizon predictive regressions with overlapping observations. I show that rather than using auto-correlation 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; this is asymptotically an exact correction and not an approximate result. Further, when the regressors are persistent and endogenous, the long-run OLS estimator suffers from the same problems as does 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. New results for the power properties of long-horizon tests are also developed. The theoretical results are illustrated with an application to long-run stock-return predictability, where it is shown that once correctly sized tests are used, the evidence of predictability is generally much stronger at short rather than long horizons.

JEL classification: C22, G1.

Keywords: Predictive regressions; Long-horizon regressions; Stock return predictability.

*I have greatly benefitted from advice by Peter Phillips and Robert Shiller. Other helpful comments have also been provided by Don Andrews, John Campbell, Dobrislav Dobrev, Ray Fair, Jon Faust, Lennart Hjalmarsson, Randi Hjalmarsson, Yuichi Kitamura, Taisuke Otsu, 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 World meeting of the Econometric Society in London, 2005, and the Copenhagen Conference on Stock Return Predictability, 2007. Excellent research assistance has been provided by Benjamin Chiquoine.

Tel.: +1-202-452-2426; fax: +1-202-263-4850; email: erik.hjalmarsson@frb.gov.

1 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, and Fama and French, 1988). While stock return predictability will also serve as the motivating example in this paper, the results derived are applicable to a much wider range of empirical questions.[1]

The main inferential issue in long-horizon regressions has been the uncertainty regarding 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. Typically, auto-correlation robust estimation of the standard errors (e.g. Newey and West, 1987) is therefore 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.[2]

The main contribution of this paper is the development of new asymptotic results 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.[3] 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. This is not an approximation, but rather an exact asymptotic result. Further, when the regressor is persistent and endogenous, the long-run OLS estimator suffers from the same problems as does the short-run OLS estimator, and similar corrections and test procedures as those proposed by Campbell and Yogo (2006) for the short-run case should also be

[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, and Boudoukh and Richardson, 1993), and the neutrality of money (Fisher and Seater, 1993).

[2] Ang and Bekaert (2007) suggest using Hodrick (1992) auto-correlation robust standard errors, which they argue have good finite sample properties. However, these rely on the regressors being covariance stationary, which is a restrictive assumption for most forecasting variables as evidenced by the results in the empirical analysis in this paper.

[3] There is now a large literature on regressions with overlapping observations. Additional references to those mentioned previously include Hansen and Hodrick (1980), Richardson and Stock (1989), Richardson and Smith (1991), Nelson and Kim (1993), Goetzman and Jorion (1993), Campbell (2001), Daniel (2001), Mark and Sul (2004), Moon et al. (2004), Torous et al. (2004), Boudoukh et al. (2005), and Rapach and Wohar (2005). The study by Valkanov (2003) is the most closely related to this paper and is discussed in more detail below.

Studies on (short-run) predictive regressions in the context of persistent regressors include Mankiw and Shapiro (1986), Cavanagh et al. (1995), Stambaugh (1999), Lewellen (2004), Campbell and Yogo (2006), Janson and Moreira (2006), and Polk et al. (2006).

used in the long-run; again, the resulting test statistics should be scaled due to the overlap.[4] 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 results in this paper are derived under the assumption that the forecasting horizon increases with the sample size, but at a slower pace. Most previous work, e.g. Richardson and Stock (1989) and Valkanov (2003), rely on the assumption that the forecasting horizon grows at the same pace as the sample size so that the forecasting horizon remains a fraction of the sample size asymptotically. In some related work, Moon et al. (2004) consider both asymptotic approaches and find that although the asymptotic distributions are seemingly quite different under the two assumptions, they both tend to provide good approximations for the finite sample properties. Indeed, Valkanov (2003), who studies a similar econometric model to the one analyzed in this paper, derives a similar scaling result to the one found here. Asymptotic results are, of course, only useful to the extent that they provide us with 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 paper provide good approximations of finite sample behavior.

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

Monte Carlo simulations show that the asymptotic normal distribution of the test statistics derived in this paper provides a good approximation in finite samples, resulting in rejection rates that are very close to the nominal size of the test under the null hypothesis. This is also true when the overlap in the data is large. This shows that although the asymptotic results are derived under an assumption that the forecasting horizon is small compared to the sample size, the normal distribution of the scaled test statistics is not very sensitive to this restriction. In fact, the tests tend to become somewhat conservative and under reject, rather than over reject, as the forecasting horizon becomes large.

[4] 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.

Since the size properties of both the tests proposed here and those of Valkanov (2003) are good, it becomes interesting to compare the power properties. Using Monte Carlo simulations, it is evident that for exogenous regressors the power properties of the test proposed here are quite similar to that of Valkanov, although there are typically some slight power advantages to the current procedure. When the regressors are endogenous, however, the test procedure derived here is often much more powerful than the test proposed by Valkanov. This stems partly from the fact that the test here explicitly takes into account, and controls for, the biasing effects of the endogenous regressors, whereas Valkanov's test only adjusts the critical values of the test statistic. Part of the power gains are also achieved by using a Bonferroni method, as in Campbell and Yogo (2006), to control for the unknown persistence in the regressors, whereas Valkanov relies on a sup-bound method, which is typically less efficient; Campbell and Yogo (2006) find the same result in the one-period case when they compare their method to the sup-bound method proposed by Lewellen (2004).

In fact, the power simulations, and additional asymptotic results, reveal three interesting facts about the properties of long-run predictive tests. First, the power of long-run tests increases only with the sample size relative to the forecasting horizon. Keeping this ratio fixed as the sample size increases does not lead to any power gains for the larger sample size. This result also suggests that for a given sample size, the power of a test will generally decrease as the forecasting horizon increases; additional simulations also support this conjecture and find that in general the one-period test will be the most powerful test. Second, when the regressors are endogenous, tests that are based on the standard long-run OLS estimator will result in power curves that are sometimes *decreasing* in the magnitude of the slope coefficient. That is, as the model drifts further away from the null hypothesis, the power may decrease. This is true both for Valkanov's test, but also if one uses, for instance, Newey-West standard errors in a normal $t-$statistic. The test proposed here for the case of endogenous regressors does not suffer from this problem. The third finding is related to the second one, and shows that although the power of the long-horizon tests increases with the magnitude of the slope coefficient for alternatives close to the null hypothesis, there are no gains in power as the slope coefficient grows large. That is, the power curve is asymptotically horizontal when viewed as a function of the slope coefficient. Both the second and third findings arise from the fact that when forecasting over multiple horizons, there is uncertainty not just regarding the future path of the outcome variable (e.g. future excess stock returns), but also about the future path of the forecasting variable over these multiple horizons. These results therefore add a further note of caution to attempts at forecasting at very long-horizons relative to the sample size: even though correctly sized tests are available, the power properties of the test can be very poor. The sometimes decreasing power curves for endogenous regressors also makes the case stronger for using test of the type proposed here, which attempts to correct for the bias and inefficiency induced in the estimation procedure by the endogeneity, and not just correct the critical values.

The theoretical results in the paper are illustrated with an application to stock-return predictability. I use a U.S. data set with excess returns on the S&P 500, as well as the value weighted CRSP index as dependent variables. The dividend price ratio, the smoothed earnings price ratio suggested by Campbell and Shiller (1988), the short interest rate and the yield spread are used as predictor variables. In addition, I also analyze an international data set with nine additional countries with monthly data spanning at least fifty years for each country. The predictor variables in the international data include the dividend-price ratio and measures of both the short interest rate and the term spread.

The evidence of predictability using the dividend- and earnings-price ratios is overall fairly weak, both in the U.S. and the international data, once the endogeneity and persistence in the regressors have been controlled for. The empirical results are more favorable of predictability when using either of the interest rate variables as predictors. This is particularly true in the U.S. data, but also to some extent in other countries. Contrary to some popular beliefs, however, the case for predictability does not increase with the forecast horizon. In fact, the near opposite is true, with generally declining $t-$statistics as the forecasting horizon increases (similar results are also found by Torous et al., 2004, and Ang and Bekaert, 2007). Given the fairly weak evidence of predictability at the short horizon, these results are consistent with a loss of power as the forecasting horizon increases, which is in line with the theoretical results derived in this paper.

The rest of the paper is organized as follows. Section 2 sets up the model and derives the theoretical results and Section 3 discusses the practical implementation of the methods in the paper. Section 4 describes the Monte-Carlo simulations that illustrate the finite sample properties of the methods and Section 5 provides further discussion and analysis of the power properties of long-horizon tests under an alternative of predictability. The empirical application is given in Section 6 and Section 7 concludes. Technical proofs are found in the Appendix.

2 Long-run estimation

2.1 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.[5] The behavior of r_t and x_t are assumed to satisfy,

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

$$x_{t+1} = \gamma + \rho x_t + v_{t+1}, \tag{2}$$

where $\rho = 1 + c/T$, $t = 1, ..., T$, and T is the sample size. The error processes are assumed to satisfy the following conditions.

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

The model described by equations (1) and (2) and Assumption 1 captures the essential features of a predictive regression with a nearly persistent regressor. It states the usual martingale difference assumption for the error terms and allows the innovations to be conditionally heteroskedastic, as long as they are covariance stationary. The error terms u_t and v_t are also often highly correlated; the regressor will be referred to as endogenous whenever this correlation, which will be labelled $\delta \equiv \omega_{12}/\sqrt{\omega_{11}\omega_{22}}$, is non-zero.

The auto-regressive 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 first 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. Cavanagh et al. (1995), Lanne (2002), Valkanov (2003), Torous et al. (2004), and Campbell and Yogo (2006) all use similar models, with a near unit-root construct, to analyze the predictability of stock returns.

The greatest problem in dealing with regressors that are near unit-root processes is the nuisance parameter c; c is generally unknown and not consistently estimable.[6] It is nevertheless useful to first derive inferential

[5] The asymptotic results presented in Section 2 all generalize immediately to the case of multiple regressors. However, the Bonferroni methods described in Section 3 are currently only developed for the case of a single regressor.

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

methods under the assumption that c is known, and then use the arguments of Cavanagh et al. (1995) to construct feasible tests. The remainder of this section derives and outlines the inferential methods used for estimating and performing tests on β in equation (1), treating c as known. Section 3 discusses how the methods of Cavanagh et al. (1995), and Campbell and Yogo (2006), can be used to construct feasible tests with c unknown.

2.2 The fitted regression

In long-run regressions, the focus of interest is the fitted regression,

$$r_{t+q}(q) = \alpha_q + \beta_q x_t + u_{t+q}(q), \tag{3}$$

where $r_t(q) = \sum_{j=1}^{q} r_{t-q+j}$, and long-run future returns are regressed onto a one period predictor.

Let the OLS estimator of β_q in equation (3), using overlapping observations, be denoted by $\hat{\beta}_q$. A long-standing issue is the calculation of correct standard errors for $\hat{\beta}_q$. Since overlapping observations are used to form the estimates, the residuals $u_t(q)$ will exhibit serial correlation; standard errors failing to account for this fact will lead to biased inference. The common solution to this problem has been to calculate auto-correlation robust standard errors, using methods described by Hansen and Hodrick (1980) and Newey and West (1987). However, these robust estimators tend to have rather poor finite sample properties.

In this section, I derive the asymptotic properties of $\hat{\beta}_q$ under the assumption that the forecasting horizon q grows with the sample size but at a slower pace. The results complement those of Valkanov (2003), who treats the case where the forecasting horizon grows at the same rate as the sample size. Simulation results in Valkanov (2003) and later on in this paper show that both asymptotic approaches provide limiting distributions that are good proxies for the finite sample behavior of the long-run estimators. The asymptotic results derived here also provide additional understanding of the properties of the long-run estimators. In particular, the results here show the strong connection between the limiting distributions of the short- and long-run estimators. This finding has important implications for the construction of more efficient estimators and test-statistics that control for the endogeneity and persistence in the regressors. Unlike Valkanov (2003), the procedures in this paper avoid the need for simulation methods; the proposed test-statistics have limiting normal distributions, although in the case of endogenous regressors with unknown persistence, Bonferroni type methods need to be used to construct feasible tests.

2.3 The limiting distribution of the long-run OLS estimator

The following theorem states the asymptotic distribution of the long-run OLS estimator of equation (3), and provides the key building block for the rest of the analysis. The result is derived under the null hypothesis of no predictability, in which case the one period data generating process is simply $r_t = u_t$, and the long-run coefficient β_q will also be equal to zero.

Theorem 1 *Suppose the data is generated by equations (1) and (2), and that Assumption 1 holds. Under the null hypothesis of no predictability such that $\beta = 0$, as $q, T \to \infty$, with $q/T \to 0$,*

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

where $B(\cdot) = (B_1(\cdot), B_2(\cdot))'$ denotes a two 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, one-period, OLS estimator $\hat{\beta}$ in equation (1), 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 original short-run parameter β, as discussed at length in Boudoukh et al. (2005).

The convergence rate of T/q for the long-run estimator also confirms the conjecture made by Nelson and Kim (1993), regarding the size of the bias in a long-run regression with endogenous regressors. They conjecture, based on simulation results, that the size of the bias is consistent with Stambaugh's (1999) approximation of the one-period bias, if one takes the total number of *non-overlapping* observations as the relevant sample size. In the near unit-root framework analyzed here, the Stambaugh bias is revealed in the non-standard asymptotic distribution of $\hat{\beta}_q$, which has a non-zero mean whenever the correlation between u_t and v_t differs from zero. Thus, since the rate of convergence is T/q, the size of the bias in a given finite sample will be proportional to the number of non-overlapping observations.

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 one-period OLS estimation, the asymptotic result is the same since, asymptotically, the long-run movements are all that matter.[7]

The limiting distribution of $\hat{\beta}_q$ is non-standard and a function of the local-to-unity parameter c. Since c is not known, and not consistently estimable, the exact limiting distribution is therefore not known in practice, which makes valid inference difficult. Cavanagh et al. (1995) suggest putting bounds on c in some manner, and 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.

2.4 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 $\delta = \omega_{12} = 0$. 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

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

where $MN(\cdot)$ 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 follows easily.

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

$$t_q = \frac{\hat{\beta}_q}{\sqrt{\left(\frac{1}{T-q}\sum_{t=1}^{T-q}\hat{u}_t^+(q)^2\right)\left(\sum_{t=1}^{T-q}\underline{x}_t\underline{x}_t'\right)^{-1}}}, \tag{6}$$

where $\hat{u}_{t+q}^+(q) = r_{t+q}(q) - \hat{\alpha}_q - \hat{\beta}_q x_t$ are the estimated residuals. Then, under Assumption 1 and the null

[7] A similar point is made by Phillips (1991b) and Corbae et al. (2002) in regards to frequency domain estimation with persistent variables. They show that the asympototic distribution of the narrow band least squares estimator, which only uses frequencies close to zero and thus captures the long-run relationship in the data, is identical to the asymptotic distribution of the full frequency estimator (which is identical to standard OLS).

hypothesis of $\beta = 0$, as $q, T \to \infty$, such that $q/T \to 0$,

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

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. Although the mechanics behind this result are spelled out in the proof in the Appendix, it is useful to outline the intuition. Note that the result in (5) implies that

$$t = \frac{\frac{T}{q}\hat{\beta}_q}{\sqrt{\hat{\omega}_{11}\left(\frac{1}{T^2}\sum_{t=1}^{T} \underline{x}_t \underline{x}_t'\right)^{-1}}} = \frac{\hat{\beta}_q}{\sqrt{q^2 \hat{\omega}_{11}\left(\sum_{t=1}^{T} \underline{x}_t \underline{x}_t'\right)^{-1}}} \Rightarrow N(0,1) \tag{8}$$

for some consistent estimator $\hat{\omega}_{11}$, since $\frac{1}{T^2}\sum_{t=1}^{T} \underline{x}_t \underline{x}_t' \Rightarrow \int_0^1 \underline{J}_c^2$. Now, as discussed in the Appendix, a consistent estimator of ω_{11} is given by $\frac{1}{qT}\sum_{t=1}^{T-q} \hat{u}_t^+(q)^2$ where the extra division by q is required given the overlapping nature of the residuals. The result now follows immediately from the definition of t_q above.

2.5 Endogeneity corrections

As discussed above, the long-run OLS estimator suffers from the same endogeneity problems as the short-run estimator; that is, when the regressors are endogenous, the limiting distribution is non-standard and a function of the unknown parameter c. To address this issue, I consider a version of the augmented regression of Phillips (1991a), together with the Bonferroni methods of Campbell and 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 one-period regression. Once the innovations v_t are obtained, an implementation of the augmented regression equation of Phillips (1991a), which he proposed for the pure unit-root case, is now possible:

$$r_{t+1} = \alpha + \beta x_t + \gamma v_{t+1} + u_{t+1 \cdot v}. \tag{9}$$

Here $u_{t\cdot v} = u_t - \omega_{12}\omega_{22}^{-1} v_t$ and $\gamma = \omega_{12}\omega_{22}^{-1}$ (Phillips, 1991a), 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 (9) 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 under Assumption 1, the OLS estimator of β in equation (9) will have an asymptotic mixed normal distribution, with the same implications as discussed above in the case of exogenous regressors.

As discussed in Hjalmarsson (2007), there is a close relationship between inference based on the augmented regression equation (9) 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 (9) is then asymptotically equivalent to

$$t_{aug} = \frac{\sum_{t=1}^{T-1}\left(\underline{r}_{t+1} - \gamma v_{t+1}\right)\underline{x}_t}{\sqrt{\omega_{11\cdot 2}\left(\sum_{t=1}^{T-1}\underline{x}_t^2\right)}} = \frac{\sum_{t=1}^{T-1}\left(\underline{r}_{t+1} - \omega_{12}\omega_{22}^{-1}v_{t+1}\right)\underline{x}_t}{\sqrt{\omega_{11\cdot 2}\left(\sum_{t=1}^{T-1}\underline{x}_t^2\right)}}, \tag{10}$$

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 (9) and $\omega_{11\cdot 2}$ is estimated as the sample variance of the residuals. Campbell and Yogo derive their $Q-$statistic as the optimal test in a Gaussian framework. The optimality of the $t-$test in the augmented regression equation thus follows from their analysis, but also directly from the analysis of Phillips (1991a). He shows that OLS estimation of (9) is identical to Gaussian full information maximum likelihood of the system described by equations (1) and (2), which thus immediately leads to the optimality result.

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

$$r_{t+q}(q) = \alpha_q + \beta_q x_t + \gamma_q v_{t+q}(q) + u_{t+q\cdot 2}(q), \tag{11}$$

where $v_t(q) = \sum_{j=1}^{q} v_{t-q+j}$. The idea is the same as in the one-period case, only now the corresponding long-run innovations $v_{t+q}(q)$ are included as an additional regressor. Let $\hat{\beta}_q^+$ be the OLS estimator of β_q in equation (11), using overlapping observations. The following result now holds.

Theorem 2 *Suppose the data is generated by equations (1) and (2), and that Assumption 1 holds. Under the null hypothesis that $\beta = 0$, as $q, T \to \infty$, such that $q/T \to 0$,*

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

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 distributions 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 Assumption 1 and the null-hypothesis of $\beta = 0$, as $q, T \to \infty$, such that $q/T \to 0$,

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

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 (11) is normally distributed. In practice, ρ is typically unknown and the next section outlines methods for implementing a feasible test.

3 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 are 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 α_2 percent, then by Bonferroni's inequality, the final conservative test will have a size no greater than $\alpha = \alpha_1 + \alpha_2$ percent.

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 $t^{+}_{q,\min} \big/ \sqrt{q}$ against the critical values of a standard normal distribution; the null is rejected if $t^{+}_{q,\min} \geq z_{\alpha_2}$, where z_{α_2} denotes the $1-\alpha_2$ 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.[8]

Unlike in the short-run methods in Campbell and Yogo (2006), there is no guarantee that $\arg\min_{\tilde{c} \in [\underline{c}, \overline{c}]} t^{+}_{q}(\tilde{c})$ and $\arg\max t^{+}_{q}(\tilde{c})$ will be the endpoints of the confidence interval for c, although for most values of q they typically are; in fact, it is easy to show that asymptotically the minimum and maximum will always be at the endpoints, but this does not hold in finite samples for $q > 1$. The test-statistic should thus be evaluated for all values in $[\underline{c}, \overline{c}]$ in order to find $t^{+}_{q,\min}$ and $t^{+}_{q,\max}$; for $q = 1$, the same result as in Campbell and Yogo (2006) holds and the extreme values of the test statistic will always be obtained at the endpoints.

In general, Bonferroni's inequality will be strict and the overall size of the test just outlined will be less than α. A test with a pre-specified size can be achieved by fixing α_2 and adjusting α_1. That is, by shrinking the size of the confidence interval for c, a test of a desired size can be achieved. Such procedures are discussed at length in Campbell and Yogo (2006) and I rely on their results here. That is, since, for all values of q, the asymptotic properties of the estimators and corresponding test-statistics derived here are identical to those in Campbell and Yogo, it is reasonable to test if their adjustments to the confidence level of the interval for c also work in the long-run case considered here. Since the Campbell and Yogo methods are frequently used in one-period regressions, this allows the use of very similar procedures in long-run regressions. As discussed below in conjunction with the Monte Carlo results, using the Campbell and Yogo adjustments in the long-run case appear to work well, although there is a tendency to under reject when the forecasting horizon is large relative to the sample size. The power properties of the test still remain good, however. 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.

Campbell and Yogo fix α_2 at ten percent, so that the nominal size of the tests evaluated for each $\tilde{c}$ is equal to ten percent. They also set the desired size of the overall Bonferroni test, which they label $\tilde{\alpha}$, to ten percent. Since 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, they then search for each value of δ, for values of α_1, such

[8] 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 is not the same as under the null hypothesis (see the proofs of Theorems 3 and 4 in the Appendix), 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.

that the overall size of the test will be no greater than $\tilde{\alpha}$.[9]

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

(i) Using OLS estimation for each equation, obtain the estimated residuals from equations (1) and (2). Calculate the correlation δ from these residuals.

(ii) Calculate the DF-GLS unit-root test statistic of Elliot et al. (1996), and obtain $\underline{c}$ and $\overline{c}$ from Tables 2-11 in Campbell and Yogo (2005) corresponding to the estimated value of δ.

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

The above procedure results in a one-sided test at the five percent level, or alternatively a two-sided test at the ten percent level. Note that, although the analysis in Section 2.5 proposes an improved point estimator, $\hat{\beta}_q^+$, for a given ρ, in practice it is merely used as a device to deliver an improved and feasible test statistic. That is, since ρ is not known in practice, the scope for improving upon the standard (long-run) OLS estimator is limited, even though improved test statistics are obtained.

The estimate of δ and the confidence interval for c can be made more robust by allowing the regressors to follow an autoregressive process with p lags ($AR(p)$), rather than an $AR(1)$ process. That is, an $AR(p)$ process can be estimated for the regressor, and the DF-GLS statistic can be calculated using p lags. Since the outcome of $\hat{\beta}_q^+$ and t_q^+ can be quite sensitive to the choice of c, when δ is large in absolute terms, it can be important to pin down the confidence interval $[\underline{c}, \overline{c}]$ as well as possible. Although the augmented regression equation is only formally justified for the $AR(1)$ case, the outcome of $\hat{\beta}_q^+$ and t_q^+ will in general be much more sensitive to the choice of c than the effects of a, typically small, serially correlated component for higher order lags in the regressor. Thus, the main benefits from allowing for a richer auto-correlation structure in the regressor come from the proper calculation of $[\underline{c}, \overline{c}]$; the effects of using the augmented regression equation rather than a method that explicitly controls for higher order auto-correlations should be small on the other hand. In practice, as evidenced in Campbell and Yogo (2006), the difference between results based on an $AR(1)$ and an $AR(p)$ assumption seems to be fairly small. However, in order to keep the analysis as robust as possible, the empirical results in Section 6 are obtained using the $AR(p)$ specification; the implementation

[9] In practice, the confidence levels of the lower and upper bounds in the shrunk confidence interval are not symmetrical, and Campbell and Yogo (2006) find separate confidence levels $\underline{\alpha}_1$ and $\overline{\alpha}_1$ that correspond to the lower and upper bounds.

follows the methods described in Campbell and Yogo (2005), using the Bayesian information criterion (BIC) to choose the appropriate lag length.

4 Monte Carlo results

All of the above asymptotic results are derived under the assumption that the forecasting horizon grows with the sample size, but at a slower rate. Valkanov (2003) also studies long-run regressions with near-integrated regressors, but 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 at the same pace as 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.

4.1 Size properties

I start with analyzing the size properties of the scaled $t-$statistics proposed earlier in the paper. Equations (1) and (2) are simulated, with u_t and v_t drawn from an *iid* bivariate normal distribution with mean zero, unit variance and correlations $\delta = 0, -0.25, -0.50, -0.75,$ and -0.95. The intercept α is set to zero and the local-to-unity parameter c is set to either 0 or -10. The sample size is either equal to $T = 100$ or $T = 500$. Since the size of the tests are evaluated, the slope coefficient β is set to zero, which implies that $\beta_q = 0$ as well. All results are based on $10,000$ repetitions.

Three different test statistics are considered: the scaled $t-$statistic corresponding to the long-run OLS estimate $\hat{\beta}_q$, the scaled Bonferroni $t-$statistic described above $\left(\text{i.e. } t^{+}_{q,\min} \middle/ \sqrt{q}\right)$, and the scaled infeasible $t-$statistic corresponding to the infeasible estimate $\hat{\beta}^{+}_q$ for a known value of c. In practice, of course, the infeasible test is not feasible but in the Monte Carlo simulations the parameter c is known, and the test based on the infeasible estimate thus provides a benchmark. All tests are evaluated against a positive one-sided alternative at the five percent level; i.e. the null is rejected if the scaled test statistic exceeds 1.645.

The results are shown in Table 1. The first 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 five percent size. For $c = -10$ 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 percent of the sample size. 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 asymptotic analysis previously, the scaled OLS $t-$test tends to over reject for endogenous regressors with $\delta < 0$, which highlights that the biasing effects of endogenous regressors are a great problem also in long-horizon regressions.

The next set of columns shows the results for the scaled Bonferroni test. The rejection rates for all δ are now typically close to, or below, five percent, indicating that the proposed correction in the augmented regressions equation (9) works well in finite samples. Only for $T = 100$ and $c = 0$ is there a slight tendency to over reject when q is small, but the average rejection rates are still well within the acceptable range; Campbell and Yogo (2006) find similar rejection rates in their one-period test, for $T = 100$. Again, as in the OLS case, there is a tendency to under reject for large q relative to the sample size T. Since the Bonferroni test is formed based on the shrunk confidence intervals for c with the confidence levels provided in Table 2 of Campbell and Yogo (2006), this could perhaps be somewhat remedied by adjusting these confidence levels for large q.[10] However, as seen in the power simulations below, the Bonferroni test is not dramatically less powerful than the infeasible test, and there seems to be little need for complicating the procedure by requiring different tables for the confidence level for c, for different combinations of q and T.

Finally, the last set of columns in Table 1 shows the rejection rates for the scaled infeasible $t-$test $t_q^+/\sqrt{q}$, resulting from the infeasible estimate $\hat{\beta}_q^+$, which uses knowledge of the true value of c. As in the case of the Bonferroni test, the rejection rates are all close to the nominal five percent level, although there is still a tendency to under reject when q/T is large.

In summary, the above 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. The simulations also show that these scaled tests tend to be somewhat conservative when q is large relative to T; this observation is further discussed in the context of the power properties of the tests, analyzed below.

The size simulations were also performed under the assumption that the innovation processes were drawn from $t-$distributions with five degrees of freedom, to proxy for the fat tails that are observed in returns data. The results were very similar to those presented here and are available upon request.

[10] Table 2 in Campbell and Yogo (2006) gives the confidence levels for the confidence interval for c that is used in the Bonferroni test, for a given δ. Tables 2-11 in Campbell and Yogo (2005) give the actual confidence intervals for c, for a given δ and value of the DF-GLS unit-root test statistic. That is, for a given value of δ and the DF-GLS statistic, Tables 2-11 in Campbell and Yogo (2005) present the confidence intervals for c with confidence levels corresponding to those in Table 2 in Campbell and Yogo (2006).

4.2 Power properties

Since the test procedures proposed in this paper appear to have good size properties and, if anything, under reject rather than over reject the null, the second important consideration is their power to reject the null when the alternative is in fact true. The same simulation design as above is used, with the data generated by equations (1) and (2). In order to assess the power of the tests, however, the slope coefficient β in equation (1) now varies between 0 and 0.5. For simplicity, I only consider the cases of $\delta = 0$ and $\delta = -0.9$.

In addition to the three scaled $t-$statistics considered in the size simulations – i.e. the scaled OLS $t-$test, the scaled Bonferroni test, and the scaled infeasible test – I now also study two additional test-statistics 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 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. It's 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 three parameters.[11] Since the critical values are a function of c, Valkanov's scaled $t-$test is generally infeasible since this parameter is unknown. 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 one-sided test against $\beta > 0$ for $\delta < 0$.[12] In the results below, I report the power curves for both the infeasible test and the sup-bound test; for $c = 0$, they are identical. To avoid confusion, I will continue to refer to the tests proposed in this paper as scaled tests, whereas I will refer to the tests suggested by Valkanov explicitly as Valkanov's infeasible and sup-bound tests. Following Valkanov's exposition, I focus on the case of $q/T = 0.1$, but given the apparently conservative nature of the tests proposed here for large q, I also consider some results for $q/T = 0.2$.

Figure 1 shows the power curves for the scaled OLS $t-$test proposed in this paper and the two tests suggested by Valkanov, for $\delta = 0$, $q = 10$, and $T = 100$. For $c = 0$, the power curves are virtually identical, whereas for $c = -10$, Valkanov's infeasible test has some power advantages. The scaled OLS $t-$test is, however, marginally more powerful than Valkanov's sup-bound test for $c = -10$. 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.

[11] Valkanov (2003) provides critical values for $q/T = 0.1$, for different combinations of c and δ, and I use these values when applicable. In the power simulations below where $q/T = 0.2$, I simulate critical values in the same manner as in the original paper, with $T = 750$, and using $100,000$ repetitions.

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

Figure 2 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 and the scaled infeasible test based on the augmented regression, along with Valkanov's two tests. The results are qualitatively similar to those for exogenous regressors with $\delta = 0$. For $c = 0$, the power curves for the three tests are nearly identical, although the scaled infeasible test proposed in this paper tends to slightly dominate Valkanov's infeasible test. For $c = -10$, the scaled infeasible test is still the most powerful, and Valkanov's infeasible test is somewhat more powerful than the scaled Bonferroni test. The least powerful test is Valkanov's (feasible) sup-bound test. Note that one would expect the scaled infeasible test proposed here to be more powerful than Valkanov's infeasible test, since the test proposed here attempts to correct the bias in the estimation procedure and not just adjust the critical values of the test; this comparison is thus the analogue of the comparison between the infeasible (short-run) $Q-$test proposed by Campbell and Yogo and the infeasible $t-$test proposed by Cavanagh et al. (1995). Finally, it is noteworthy that the power of Valkanov's sup-bound test appears to *decrease* for large values of β when $c = -10$. A similar pattern is also hinted at for Valkanov's test with $c = 0$. These patterns become clearer as the forecasting horizon increases and will be analyzed at length below.

Given that the scaled Bonferroni test, in particular, seemed to be under sized for large values of q relative to T, it is interesting to see if this also translates into poor power properties. Figure 3 shows the results for $\delta = -0.9$, $T = 100$, and $q = 20$. Two observations are immediately obvious from studying the plots. First, the scaled Bonferroni test is reasonably powerful when compared to the infeasible scaled test, and very powerful compared to Valkanov's sup-bound test. Second, the declining pattern in the power curves for Valkanov's two tests that were hinted at in Figure 2 are now evident; as β becomes larger, the power of these two tests decline. This result is of course perplexing, since Valkanov's tests were explicitly derived under the assumption that q is large relative to the sample size T. In the following section, additional analytical results are derived that will shed some light on these findings. However, before turning to the formal analysis, results shown in Figure 4 provide further confirmation of the results in Figure 3, as well as highlight some additional findings.

Figure 4 confirms and elaborates on the findings in Figure 3. The right hand graph shows the power curves for $\delta = -0.9$, $c = -10$, $T = 500$, and $q = 100$. Using $T = 500$ confirms that the previous findings were not just a small sample artefact. The same pattern as in Figure 3 emerges for Valkanov's two tests: after an initial increase in power as β becomes larger, the power starts to decrease. Further results for larger values of β, which are not shown, indicate that the power curves do not converge to zero as β grows large; rather, they seem to level out after the initial decrease. Furthermore, the power curves for the scaled Bonferroni test and the scaled infeasible test do not seem to converge to one as β increases, although they do not decrease either, and stabilize at a much higher level than the power curves for Valkanov's tests. In addition, the power curve for the scaled OLS $t-$test is also shown. This test is biased for $\delta \neq 0$ but provides an interesting comparison to the power curves of Valkanov's test. As is seen, the scaled OLS $t-$test behaves in a very similar manner to Valkanov's infeasible test. It is thus apparent that the difference in behavior between the Bonferroni test and Valkanov's tests stems primarily from the endogeneity correction and not the manner in which they are scaled. Finally, the right hand graph in Figure 4 also shows that the patterns established for the power curves of the tests proposed both in this paper and in Valkanov (2003) are not a result of scaling the test statistic by either the sample size or the forecasting horizon. As shown, if one uses Newey-West standard errors to calculate the (non scaled) $t-$statistic from the long-run OLS regression, a similar pattern emerges; note that the test based on Newey-West standard errors will be biased both for the well established reason that the standard errors do not properly control for the overlap in the data, but also because the $t-$statistic from the long-run OLS regression does not control for the endogeneity in the regressors. The Newey-West standard errors were calculated using q lags.

It is worth pointing out that Valkanov (2003) also performs a Monte Carlo experiment of the power properties of his proposed test-statistics, without finding the sometimes decreasing patterns in the power curves reported here. However, Valkanov (2003) only considers the case with $\delta = -0.9$, $q = 10$, and $T = 100$, for values of β between 0 and 0.1. As seen in Figure 2 here, the power curves of all the tests are strictly increasing in β for these parameter values.

The left hand graph in Figure 4 further illustrates the above observations for the scaled OLS $t-$statistic in the case of $\delta = 0$. Here, with $T = 500$, $q = 100$ and $c = 0$, the power of the scaled OLS $t-$statistic and Valkanov's (infeasible) test statistic are almost identical and again seem to converge to some fixed level less than one. The results also suggest that the decrease in power seen for Valkanov's test in the previous plots does not occur when the regressors are exogenous. The $t-$statistic based on Newey-West standard errors is also shown to exhibit the same pattern; here, the bias in this test resulting from the overlap in the data alone is evident, with a rejection rate around 20 percent under the null.

To sum up, the simulations show that both the scaled OLS $t-$test and the scaled Bonferroni test have good (local) power properties when compared to the tests proposed by Valkanov (2003). This is especially true for the Bonferroni test used with endogenous regressors, which tends to dominate Valkanov's sup test for all values of β, and also dominates Valkanov's infeasible test for large values of β.

However, all of the tests discussed here, including the standard $t-$test based on Newey-West standard errors, seem to behave in a non-standard way as the value of the slope coefficient drift further away from the null hypothesis: rather than converging to one as β grows large, the power of the tests seem to converge to some value less than unity. In the next section, I provide an analytical explanation of these findings and discuss its implications.

5 Long-run inference under the alternative of predictability

5.1 Asymptotic power properties of long-run tests

The simulation evidence in the previous section raises questions about the properties of long-run tests under the alternative of predictability. In particular, the power of the tests does not seem to converge to one as the slope coefficient increases and, in addition, the power curves appear to sometimes *decrease* as the slope coefficient drifts away from the null hypothesis. In this section, I therefore derive some analytical results for the power properties of long-run tests. I first start by considering a fixed alternative, which provides the answer to why the power does not converge to unity when the slope coefficient increases. In the following sub-section, I consider the power against a local alternative, which helps explain the hump shaped pattern in the power curves. These analytical results also reveal some interesting features about the consistency of long-run tests. I focus on the standard (scaled) OLS $t-$statistic, since the behavior of the t_q^+-statistic is similar to the former with exogenous regressors.

The following theorem provides the asymptotic results for the distribution of the $t-$statistic under a fixed alternative of predictability. Results are given both for the asymptotics considered so far in this paper, i.e. $q/T = o(1)$, as well as the type of asymptotics considered by Valkanov (2003).

Theorem 3 *Suppose the data is generated by equations (1) and (2), and that Assumption 1 holds. Under the alternative hypothesis that $\beta \neq 0$:*

(i) As $q, T \to \infty$, such that $q/T \to 0$,

$$\frac{q}{T}\frac{t}{\sqrt{q}} \Rightarrow \sqrt{\frac{3}{\omega_{22}}}\left(\int_0^1 J_c^2\right)^{1/2} \text{ and thus } \frac{t}{\sqrt{q}} = O_p\left(\frac{T}{q}\right). \tag{15}$$

(ii) As $q, T \to \infty$, such that $q/T \to \lambda$,

$$\frac{t}{\sqrt{T}} \Rightarrow \frac{\int_0^{1-\lambda} J_c(r;\lambda)}{\sqrt{\left(\int_0^{1-\lambda} J_c^2(r;\lambda)\right)\left(\int_0^{1-\lambda} J_c^2\right)}} = O_p(1). \tag{16}$$

where $J_c(r;\lambda) = \int_r^{r+\lambda} J_c(r)$.

The asymptotic results in Theorem 3 help shed light on the general patterns seen in the figures above. Part *(i)* of the theorem, which provides the limiting distribution of the scaled $t-$test analyzed in this paper, shows that the power of this test will increase with the relative size of the sample to the forecasting horizon; thus, as long as the ratio between q and T is fixed, there are no asymptotic power gains. The power is also independent of the value of the slope coefficient β, as long as it is different from zero. This explains the leveling out of the power curves as β grows large, and their failure to converge to one for large values of β. The intuition behind the independence of β in the limiting distribution is best understood by explicitly writing out the equation for the long-run returns under the alternative of predictability. That is, since the true model is given by equations (1) and (2), the long-run regression equation is a fitted regression, rather than the data generating process. As shown in the proof of Theorem 3 in the Appendix, under the alternative of predictability, the long-run returns $r_{t+q}(q)$ actually satisfy the following relationship when ignoring the constant, derived from equations (1) and (2):

$$r_{t+q}(q) = \beta_q x_t + \beta \sum_{h=1}^{q-1} \left(\sum_{p=h}^{q-1} \rho^{p-h} \right) v_{t+h} + u_{t+q}(q). \tag{17}$$

There are now, in effect, two error terms, the usual $u_{t+q}(q)$ plus the additional term $\beta \sum_{h=1}^{q-1} \left(\sum_{p=h}^{q-1} \rho^{p-h} \right) v_{t+h}$, which stems from the fact that at time t there is uncertainty regarding the path of x_{t+j} for $j = 1, ..., q-1$. That is, since the true model is given by equations (1) and (2), there is uncertainty regarding both the future realizations of the returns, as well as of the predictor variable, when forming $q-$period ahead forecasts. Since the first error term $\beta \sum_{h=1}^{q-1} \left(\sum_{p=h}^{q-1} \rho^{p-h} \right) v_{t+h}$ is of an order of magnitude larger than $u_{t+q}(q)$, it will dominate the asymptotic behavior of the least squares estimator of β_q. As seen in the proof, the multiplication of this error term by β ultimately explains why a larger β will also lead to a larger error term, cancelling out any power gains that might otherwise have occurred as β drifts further away from zero.

Part *(ii)* of the theorem states that Valkanov's scaled $t-$statistic converges to a well defined limiting distribution that is independent of β and T, although it is a function of $\lambda = q/T$. Thus, under the assumptions on q and T maintained by Valkanov, the $t-$statistic scaled by $\sqrt{T}$ does not diverge and hence the power of the test does not converge to one. Of course, for a fixed $q/T = \lambda$, the same heuristic result follows from part *(i)*, since as long as q/T does not change, there are no power gains. Thus, although some caution is required when comparing the results in parts *(i)* and *(ii)* of the theorem, since they are derived under different assumptions, they lead to the same heuristic result. Indeed, for a fixed $q/T = \lambda$, it follows that $\sqrt{q} = \sqrt{\lambda T}$ and that the results for the scaled tests in this paper should be similar to those of Valkanov's tests.

The main message of Theorem 3 is thus that the only way to achieve power gains in long-run regressions is by increasing the sample size relative to the forecasting horizon; as long as this ratio is fixed, there are no asymptotic power gains as the sample size increases. The results in Theorem 3 also provide some intuition to a somewhat counter intuitive result in Valkanov (2003). As shown there, under the assumption that $q/T = \lambda$ asymptotically, the estimator of the long-run coefficient β_q is not consistent; however, a scaled version of the $t-$statistic has a well defined distribution. That is, even though the coefficient is not estimated consistently, valid tests can still be performed. Theorem 3 shows the limitation of this result: like the estimator, the test is not consistent since there are no asymptotic power gains for a fixed q/T.

Part *(i)* of Theorem 3 also suggests that for a fixed sample size, more powerful tests of predictability are achieved by setting the forecasting horizon as small as possible. That is, in general, one might expect power to be decreasing with the forecasting horizon. This is merely a heuristic argument, since the result in part *(i)* of Theorem 3 is an asymptotic result based on the assumption that $q/T = o(1)$. Nevertheless, it is interesting to briefly compare the finite sample power properties between tests at different horizons. The simulation results in the previous section already support the conjecture that power is decreasing with the horizon, in finite samples, as evidenced by the rather poor power properties for the really long horizons studied in Figures 3 and 4. The simulations in Figure 5 make these results even clearer. The simulation setup is the same as before, with $T = 100$ and $c = 0$. The left hand graph shows the power curves for the scaled OLS $t-$test, when $\delta = 0$, for three different forecasting horizons, $q = 1, 10$, and 20. It is evident that as q increases, the power uniformly decreases. The right hand graph shows the case of $\delta = -0.9$, and illustrates the power curves for the scaled Bonferroni test, for $q = 1, 10$, and 20. Again, there is a clear ranking of the power curves from short to long horizon. Qualitatively identical results, which are not shown, are obtained for $c = -10$, and for $T = 500$.

Overall, the results here are thus supportive of the notion that tests of predictability generally lose power as the forecasting horizon increases. This is in line with what one might expect based on classical statistical and econometric theory. In the case of exogenous regressors, the OLS estimates of the single period ($q = 1$) regression in equation (1) are identical to the full information maximum likelihood estimates and in the endogenous regressor case, OLS estimation of the one-period augmented regression equation (9) is likewise efficient. Standard analysis of power against a sequence of local alternatives then implies that a one-period Wald test (or, equivalently, a $t-$test) is asymptotically optimal (Engle, 1984). Campbell (2001) makes this point, but also finds that some alternative ways of comparing asymptotic power across horizons suggest that there may be power gains from using longer horizons; however, he finds little support for this in his Monte Carlo simulations.

5.2 Local asymptotic power

The asymptotic power properties in the previous section were derived under the assumption of a fixed alternative $\beta \neq 0$. As seen in the power curves in the figures above, it is clear that for small values of β, the power of the long-run tests is a function of β. And, in particular, there appears to be regions of the parameter space where the power of the tests are decreasing in the magnitude of the slope coefficient. These facts are not reflected in the results in Theorem 3, however, and the power properties in these regions of the parameter space are therefore likely better analyzed with a local alternative for β, as is common in the literature on evaluating the power of statistical tests. The following theorem provides a guide to the local power properties of the scaled OLS $t-$test proposed in this paper.

Theorem 4 *Suppose the data is generated by equations (1) and (2), and that Assumption 1 holds. Under the local alternative of $\beta = b/q$,*

$$\frac{t}{\sqrt{q}} \sim \frac{T}{q} \frac{b + \frac{q}{T}\left(\int_0^1 dB_1 J_c + \frac{b}{2}\int_0^1 dB_2 J_c\right)\left(\int_0^1 J_c^2\right)^{-1}}{\sqrt{\left(\omega_{11} + b\omega_{12} + \frac{b^2\omega_{22}}{3}\right)\left(\int_0^1 J_c^2\right)^{-1}}}, \tag{18}$$

where '$\sim$' denotes an approximate distributional equivalence.

This theorem heuristically shows the approximate distribution of the scaled OLS $t-$statistic for alternatives that are close to the null hypothesis, in the sense that the slope coefficient β shrinks towards zero with the forecasting horizon. For small to moderate values of b, it is evident that the $t-$statistic, and hence the power of the test, will depend on the value of b. For large b, and small q relative to T, it follows that

$$\frac{t}{\sqrt{q}} \sim \frac{T}{q}\sqrt{\frac{3}{\omega_{22}}}\left(\int_0^1 J_c^2\right)^{1/2} = O_p\left(\frac{T}{q}\right), \tag{19}$$

which is independent of b and identical to the result under the fixed alternative. For $b = 0$, the distribution under the null hypothesis is recovered. In fact, it is useful to separate the numerator of the $t-$statistic as follows:

$$\frac{T}{q}b + \left(\int_0^1 dB_1 J_c\right)\left(\int_0^1 J_c^2\right)^{-1} + \frac{b}{2}\left(\int_0^1 dB_2 J_c\right)\left(\int_0^1 J_c^2\right)^{-1}. \tag{20}$$

Here the first term is the pure drift part, which will dominate asymptotically provided that $q/T \to 0$, the second term is the usual variance term under the null hypothesis, and the third term reflects the uncertainty regarding the future path of the regressor x_t in a long-run regression, as discussed above in conjunction with the representation in equation (17). Obviously, the pure drift term is increasing in b, and the second term does not change with b. The third term is on average decreasing in b, since the outcomes of the random variable $\int_0^1 dB_2 J_c$ tend to be negative. That is, by effectively omitting the term $\beta \sum_{h=1}^{q-1} \left(\sum_{p=h}^{q-1} \rho^{p-h} \right) v_{t+h}$ in equation (17), a downward bias is induced in the estimator and the subsequent $t-$statistic because the relevant asymptotic 'covariance' measure between x_t and v_t is negative. However, these terms are all linear in the coefficient b, and can therefore not explain the non-monotonicity in the power curves that were found in the Monte Carlo simulations. Instead, the answer must be in the denominator.

For large b, the denominator is increasing in b but, in the case of $\omega_{12} < 0$, there is a range of b for which the denominator is decreasing in b; this explains the hump-shaped pattern in the power curves that was documented in the Monte Carlo study. To form an intuition behind these results, consider again the representation of the long-run regression in equation (17). Under the assumption that $\beta = b/q$, the usual error term $u_{t+q}(q)$ and the additional term $\beta \sum_{h=1}^{q-1} \left(\sum_{p=h}^{q-1} \rho^{p-h} \right) v_{t+h}$ will both be of the same order of magnitude. When calculating the variance of the fitted residual, which enters into the denominator of the $t-$statistic, the variance of both of these terms as well as their covariance will thus enter. The covariance ω_{12}, when it is negative, will induce the non-monotonicity in the $t-$statistic as a function of b. Initially, as the slope coefficient drifts away from zero, the first term will dominate and the power of the test is increasing in b, since the variance of $u_{t+q}(q)$ is independent of b. In a middle stage, the covariance term becomes important as well and the $t-$statistic decreases with the slope coefficient. Finally, as b grows large, the last term dominates and will exactly cancel out the dependence on b in the numerator and denominator.

Figure 6 shows the average power curves that result from direct simulations of the limiting random variables in equation (18). As in the previous simulations, the variances ω_{11} and ω_{22} are both set equal to one. I let $T/q = 5$ so that the results correspond to the finite sample power curves shown in Figures 3 and 4, where the forecasting horizon is equal to 20 percent of the sample size. The local-to-unity parameter c is set equal to zero and $10,000$ repetitions are used.

The left hand graph in Figure 6 shows the case of exogenous regressors ($\delta = \omega_{12} = 0$). The local power curve is weakly increasing and looks very similar to the finite sample results seen in Figure 4. For endogenous regressors with $\delta = -0.9$, shown in the right hand graph in Figure 6, the same hump shaped pattern as in Figures 3 and 4 is evident; the biased nature of the OLS $t-$test with endogenous regressors is also clearly evident with a rejection rate around 40 percent under the null. The power curves based directly on the asymptotic results in Theorem 4 thus seem to correspond well to the finite sample ones.

5.3 Practical implications

The results in this section help shed more light on the properties of long-run tests under the alternative of predictability. The main lesson is that the power of long-horizon tests only grows with the size of the sample relative to the forecasting horizon; keeping q/T fixed as T increases does not result in any power gains. The practical implications and recommendations must therefore be that inference on forecasts at very long horizons will be imprecise, and caution should be used in extending the forecasting horizon as larger samples become available. The results here also show that the asymptotic device used in this paper, where $q/T = o(1)$, provides an important benchmark comparison to the commonly used framework with $q/T = \lambda$, since the test statistics are only consistent under the former assumption. The theoretical results here also help explain the puzzling non-monotonicity in the power curves for long-run regressors, a finding which adds an additional note of caution to the use of long forecasting horizons. Note that the turning point of the power curve is not outside the relevant parameter region. As seen in Figure 3, for $c = 0$, the power is already declining for $\beta = 0.1$; the results in Campbell and Yogo (2006) show that in annual data, which the 100 observations in each simulated sample used to generate Figure 3 might represent, the estimates of β are between 0.1 and 0.3 for the dividend and earnings-price ratios. This also provides a strong case for the test based on the long-run augmented regression equation suggested in this paper, since it does not suffer from non-monotone power.

The results here also suggest that the power of predictive tests may be decreasing with the forecasting horizon, which would seem to imply that using one period tests is the best approach. The simulation results are supportive of this conjecture and the empirical results presented in the next section can also be interpreted as favorable of this view. However, the power comparisons across different forecasting horizons conducted in this paper are all informal and heuristic; a more thorough analysis, which is outside the scope of the current study, is required before any definitive statements can be made. Finally, one should recall one important caveat. The power results are all derived under the assumption that the true model is the one given by equations (1) and (2). This is a standard assumption used by, for instance, Nelson and Kim (1993) and Campbell (2001), but clearly other potential data generating processes that might lead to different results are possible. The results under the model analyzed here, however, can be considered a point of reference against which to compare other specifications.

6 Long-run stock return predictability

To illustrate the theoretical results derived in this paper, I revisit the question of stock return predictability. There have been many conflicting results regarding the existence of a predictable component in stock returns. However, recent work by Lewellen (2004) and Campbell and Yogo (2006), which rely on both more robust as well as more efficient methods of inference than previous research, do find evidence that stock returns are predictable to some degree. In this section, I extend their empirical analysis to the long-horizon case. Since the scaled long-run Bonferroni test, which controls for the endogeneity and persistence in the regressors, is effectively a long-run version of the methods developed in Campbell and Yogo (2006), the empirical results presented here provide a direct comparison with previous work. In the first part of the empirical analysis, I therefore analyze the same data as those used by Campbell and Yogo. I then consider the evidence in an international data set from nine additional countries. The section ends with a discussion of the results.

6.1 The data

6.1.1 The U.S. data

The data on U.S. stock returns and predictor variables are the same as those used by Campbell and Yogo (2006).[13] The returns data consist of the monthly and annual excess returns on the CRSP NYSE/AMEX value-weighted index over the period 1926-2002, as well as annual returns on the S&P 500 index over the period 1880-2002. The excess returns are calculated as the stock returns over the risk free rate, measured by the return on the one-month T-bill for the monthly data, and by the return on the three-month T-bill rolled over quarterly for the annual data. The predictor variables are the dividend-price ratio $(d-p)$, the smoothed earnings-price ratio $(e-p)$ suggested by Campbell and Shiller (1988), the 3-month T-bill rate (r_3), and the long-short yield spread $(y-r_1)$, which is defined as the difference between Moody's seasoned Aaa corporate bond yield and the one month T-bill rate. The dividend-price ratio is calculated as dividends over the past year divided by the current price and the (smoothed) earnings-price ratio as the average earnings of the past 10 years divided by the current price. Since earnings are not available for the CRSP data, the corresponding S&P 500 earnings are used. All regressions are run using log-transformed variables with the log excess returns as the dependent variable. The regressions involving the short-rate and the yield-spread as predictors are estimated over the period 1952-2002, since prior to this time period the interest rate was pegged by the Federal Reserve. The regressions with the CRSP data, using the dividend- and earnings-price ratios as predictors, are also analyzed over this period as a comparison to the full sample results.

6.1.2 The international data

The international data used in this paper come from Global Financial Data. Total returns, including direct returns from dividends, on market-wide indices in nine countries with at least 50 years of data were obtained, as well as the corresponding dividend-price ratios. Earnings data were typically only available over much shorter time periods and long-run regressions with the earnings-price ratio as a predictor are therefore not included in the international analysis. In addition, for each country, measures of the short and long interest

[13] These data were downloaded from Professor Yogo's website: http://finance.wharton.upenn.edu/~yogo/.

rates were obtained, from which measures of the term spread were constructed. The variable definitions follow the usual conventions in the literature. The dividend-price ratio is defined as the sum of dividends during the past year, divided by the current price. The measure of the short interest rate comes from the interest rate series constructed by Global Financial Data and uses rates on 3-month T-bills when available or, otherwise, private discount rates or interbank rates. The long rate is measured by the yield on long-term government bonds. When available, a 10 year bond is used; otherwise, I use that with the closest maturity to 10 years. The term spread is defined as the log difference between the long and the short rate. Excess stock returns are defined as the return on stocks, in the local currency, over the local short rate, which provides the international analogue of the typical forecasting regressions estimated for U.S. data.

The predictor variables used in the international sample are therefore the dividend-price ratio $(d-p)$, the short interest rate (r_s) and the term spread $(r_l - r_s)$, where the latter two are meant to capture similar features of stock return predictability as the corresponding interest rate variables in the U.S. sample, even though they are not defined in an identical manner.

The countries in the data are: Australia, Belgium, Canada, France, Germany, Italy, Japan, Sweden, and the U.K. The end date for each series is March 2004, although the starting date varies between the countries. The longest series is for Australia, which dates back to 1882, and the shortest for Germany, which goes back to 1953. All returns and interest rate data are on a monthly frequency. For a few of the older observations, the dividend-price ratios are given on an annual basis; these are transformed to monthly data by filling in the monthly dividends over the year with the previous year's values.[14]

All regressions are run using log-transformed variables with the log excess returns over the domestic short rate as the dependent variable. Following the convention used in the U.S. data, the data used in all interest rate regressions are restricted to start in 1952 or after.[15] Again, as a comparison, the predictive regression with the dividend price ratios are also run over this restricted sample period; in the international data, this is particularly useful, since the starting points of the series vary from country to country and imposing a common starting date allows for easier cross-country comparison.

[14] Certain countries had longer periods over which, for instance, data on dividends were missing (e.g. the U.K.), or where no data at all were available during some periods, such as the years around the two world wars (e.g. Germany). In these cases, I restrict myself to the longest available consecutive time-series, and do not attempt to parse together the discontinuous series.

[15] In the U.S., the interest rate was pegged by the Federal Reserve before this date. Of course, in other countries, deregulation of the interest rate markets occurred at different times, most of which are later than 1952. As seen in the international finance literature (e.g. Kaminsky and Schmukler, 2002), however, it is often difficult to determine the exact date of deregulation. And, if one follows classification schemes, such as those in Kaminsky and Schmukler (2002), then most markets are not considered to be fully deregulated until the 1980s, resulting in a very small sample period to study. Thus, the extent to which observed interest rates reflect actual market rates is hard to determine and one should keep this caveat in mind when interpreting the results.

6.2 Characteristics of the predictor variables

The two key data characteristics that define the properties of the regression estimators analyzed in this paper are the near persistence and endogeneity of the regressors. For the U.S. data, Table 2 shows confidence intervals for the autoregressive root ρ, and the analogue intervals for the local-to-unity parameter c, calculated by inverting the DF-GLS unit-root test, as well as estimates of the correlation between the innovations to returns and the innovations to the regressors (δ). The results are shown both for the full sample period, as well as for the post 1952 sample. As is evident, there is a large negative correlation between the innovations to the returns and the valuation-ratios. The short interest rate is nearly exogenous, however. The yield spread is also almost exogenous in the monthly data, although it exhibits a somewhat larger correlation in the annual data. Standard OLS inference might thus be expected to work fairly well when using the short rate or the yield spread as predictor variables. In addition, all variables, except perhaps the annual yield spread, show signs of having autoregressive roots that are close to unity.

The corresponding results for the international data are given in Table 3, where the sample period available for each country is also given. Overall, the international predictor variables are similar to the corresponding U.S. ones. The dividend-price ratio is highly persistent in all countries, and the null hypothesis of a unit root can typically not be rejected based on the DF-GLS test statistic. Furthermore, the dividend-price ratio is generally fairly endogenous, in the sense that the estimates of δ, the correlation between the innovations to the returns and the predictor process, are large in absolute value. Compared to the U.S. data, however, the estimates of δ for the dividend-price ratio are generally somewhat smaller in absolute value, typically ranging from -0.5 to -0.8, whereas in the U.S. data absolute values above 0.9 are common. The short interest rate and the term spread also behave similar to the U.S. counterparts. They are mostly exogenous but still highly persistent.

Both the U.S. and the international data thus seem to fit well the assumptions under which the results in this paper are derived. In addition, at least for the valuation ratios, there is a strong case for using test statistics that take into account the bias induced by the endogeneity and persistence in the regressors. For the interest rate variables, OLS inference should be fairly accurate.

6.3 Long-run regression results for the U.S. data

The results from the long-run regressions are presented graphically as plots of the scaled $t-$statistics against the forecasting horizon q. Although the results in previous sections suggest that using very long forecasting horizons are generally not advisable, I will show results for forecasting horizons out to 20 years in the annual data and 10 years in the monthly data, to illustrate the properties of the test statistics across most potential forecasting horizons that may be used in applied work.

In each plot, the values of the scaled OLS $t-$statistics along with the scaled Bonferroni $t-$statistics are plotted against the forecasting horizon; as a point of reference, the five percent significance level in a one sided test is also shown, i.e. a flat line equal to 1.645. The Bonferroni test statistic is calculated in the same manner as described in Section 3. Given the asymptotic results developed previously, the scaled Bonferroni $t-$statistic will be approximately normally distributed for all predictor variables, whereas for the scaled OLS $t-$test, the normal approximation will only be satisfied for exogenous variables and might thus be expected to work well with the interest rate variables. In addition to the scaled Bonferroni test statistic, I also show the value of the scaled t_q^+ statistic evaluated for $c = 0$ (i.e. $\rho = 1$). The *maximum* of this test statistic and the Bonferroni test statistic can be seen as the value of the Bonferroni test when explosive ($\rho > 1$) roots are ruled out *a priori*.[16] This additional statistic is not shown for the interest rate variables where the Bonferroni and OLS statistics are already very close to each other.[17]

The first set of results are displayed in Figure 7, which shows the scaled OLS and Bonferroni $t-$statistics from the regressions with the dividend- and earnings-price ratios in the annual full sample U.S. data. As is to be expected, the results for the one period forecasting horizon are qualitatively identical to those in Campbell and Yogo (2006). Thus, at the shorter horizons, there is some mixed evidence of predictability, with the null rejected for both the S&P 500 and the CRSP returns when using the earnings-price ratio, but only for the CRSP returns when using the dividend-price ratio. It is interesting to note that although the Bonferroni test is more robust than the OLS test, the numerical outcome of the Bonferroni test need not always be smaller than the biased OLS $t-$statistic. In addition, in Figure 7, the $t-$statistics based on Newey-West standard errors are also shown, calculated using q lags. Comparing the plots of these against the properly scaled $t-$statistics, it is apparent that Newey-West errors can fail substantially in controlling the size of long-horizon test. They also illustrate why long-run predictability is often thought to be stronger than short-run predictability. Given the well known biases in the Newey-West statistics, in the subsequent figures they are not shown in order to keep the graphs more easily readable.

Similar results to those in Figure 7 are also found in Figure 8, which shows the results for monthly CRSP returns, both for the full sample from 1926 and in the post 1952 sample, using the dividend- and earnings-price ratios as predictors. Again, there is mixed evidence of predictability. The results in Figure 8 also

[16] For $\delta < 0$, the t_q^+ statistic is asymptotically decreasing in ρ, and restricting ρ to be less than or equal to one (or equivalently $c \leq 0$) thus results in a more powerful test. In finite samples, the t_q^+ statistic is not always monotonically decreasing in c and the statement in the text is thus not always necesarrily true. That is, there could be some value for $\tilde{c} \in [\underline{c}, \bar{c}]$, with $\tilde{c} < 0$ such that $t_q^+(\tilde{c}) < t_q^+(0)$, in which case even though explosive roots are ruled out, the most conservative value is not obtained for $\rho = 1$. In this case, the procedure described in the text is not correct since it could lead to a test that over rejects. However, in all but one of the cases considered here, the statement in the text holds true and it is only marginally wrong once, with no qualitative impact at all. In fact, interior optima seems very rare in practice.

[17] Note also that it is only for $\delta < 0$ and an alternative of $\beta > 0$ for which $\rho = 1$ will provide a conservative upper bound. For other parameter combinations, the conservative bound might be achieved for small values of ρ.

illustrates that ruling out explosive processes, i.e. restricting ρ to be less than or equal to one, can have a substantial impact on the results. In the sub sample from 1952-2002, the evidence in favour of predictability is substantially greater when ruling out explosive processes. This great sensitivity stems from the extreme endogeneity of the dividend- and earnings-price ratios in the U.S. data, with absolute values of δ upwards of 0.99.

From the perspective of the theoretical analysis in the current paper, the results in Figures 7 and 8 illustrate two key findings. First, and contrary to many popular beliefs, the evidence of predictability does not typically become stronger at longer forecasting horizons. There are some exceptions, such as the results for the dividend-price ratio in the full CRSP sample in Figure 8, but overall there is little tendency for the results to look stronger at longer horizons. If anything, there is a tendency for the properly scaled $t-$statistics to become smaller as the horizon increases, which would be consistent with a loss of power. Second, these results show that it is important to control for the biasing effect of persistent and endogenous regressors also in long-horizon regressions, as seen from the often large difference between the OLS and the Bonferroni test statistics.

Figure 9 shows the results for the short rate and the yield spread, both for the annual and the monthly data. As expected, the OLS and Bonferroni results are now very close to each other, reflecting the nearly exogenous nature of the interest rate variables. For the short rate, the one-sided alternative is now a negative coefficient. In order to achieve easy comparison with the rest of the results in general, and the yield-spread in particular, the negative of the test statistics are plotted for the short rate. As seen, there is evidence of predictability at very short horizons, which disappears very fast as the horizon increases. In fact, the evidence is already gone in the annual data at the one-period horizon. A similar result is found for the yield spread, where the expected coefficient under the alternative of predictability is again positive.

The one-period, or short-run, empirical findings for the U.S. data are qualitatively identical to those of Campbell and Yogo (2006). The bottom line is that there is fairly robust evidence of predictability in U.S. data in the short run when using the two interest rate variables as predictors, whereas the evidence for the valuation ratios is more mixed. The results from the regressions with the dividend- and earnings-price ratios are made more difficult to interpret given the large endogeneity of the regressors. As is seen, for instance, restricting the autoregressive root to be less than or equal to unity can change the results rather dramatically, a point which is discussed in detail in Campbell and Yogo (2006); these results thus illustrate well the power gains that can be made with additional knowledge regarding the true autoregressive root in the process. Although restricting the regressor to be a non-explosive process seems like a fairly sensible restriction in most cases, it should also be stressed that imposing a non-explosive condition on the dividend-price ratio, for instance, is not necessarily completely innocuous. Lettau and Van Nieuwerburgh (2007) show that there is evidence of structural breaks in the valuation ratios in U.S. data and that if one takes into account these breaks, the predictive ability of these ratios improves. A structural break process is inherently non-stationary and is indeed very hard to distinguish from a highly persistent process of the kinds analyzed in this paper, especially if one allows for explosive roots. Some caution is therefore required in ruling out explosive processes, a point also made by Campbell and Yogo (2006).

6.4 Long-run regression results for the international data

The results for the international data are shown in Figures 10-13. The results for the dividend-price ratio are shown in Figure 10 for the full sample and in Figure 11 for the post 1952 sample. Given the somewhat mixed and overall fairly weak results in the U.S. data, the international results are close to what one might expect. There is some weak evidence of predictability in the full sample for Canada, as well as for Japan. In both the Canadian and Japanese cases, however, these results are no longer significant in the post 1952 sample, which is particularly striking for Japan since the full sample only stretches back to 1949. The results for both Canada and Japan are also sensitive to the exclusion of explosive roots. The only country for which there is consistently strong evidence is the U.K. Again, there is little evidence of stronger predictability in the long-run. The only significant result in this direction is for the full Canadian sample where there is no predictability at the first few horizons; the results are far from striking, however.

The results for the interest rate variables, shown in Figures 12 and 13, are somewhat more favorable of predictability. For the short-rate, shown in Figure 12, where again the alternative hypothesis is a negative coefficient and the negative of the test statistic is plotted, significance is found at short horizons in Canada and Germany, and close to significance in Australia, France, and Italy. The corrections for endogeneity have little effect, and the OLS and Bonferroni results are very close to each other. The only exception is at long horizons for Japan where there is some discrepancy, although not enough to change any conclusions if one were to rely on the OLS analysis.

For the term spread, shown in Figure 13, the results look similar but somewhat stronger, with significant short-run coefficients found for Canada, France, Germany, and Italy, and for a few horizons for Australia. Again, with the exception of Australia, the evidence of predictability disappears very fast as the horizon increases.

The results from the international data support the U.S. conclusions to some extent. The evidence in favour of predictability using the dividend-price ratio in international data is overall weak, with the only solid evidence coming from the U.K. data. The evidence from Canada and Japan is weaker and more sensitive to the sampling period. Although the scaled Bonferroni statistic is generally much smaller than the scaled OLS

$t-$statistic in the international data as well, the evidence based on the OLS results themselves is not that supportive of a predictive relationship either. Thus, although some power gains would still be had from a more precise knowledge of the autoregressive root in the data, the international results may be somewhat less susceptible to this critique than the U.S. results. The international results for the interest rate variables are again similar to those of the U.S. data, but do not fully support any generic statements about the predictive ability of these variables. However, there is some commonality across the country results for these variables. This is particularly true for Australia, Canada, France, Germany, and Italy, where the significant results are found.

6.5 Discussion of the empirical findings

The empirical findings can broadly be summed up as follows: (i) The evidence of predictability using the valuation ratios is overall fairly weak, both in the U.S. and the international data. (ii) The predictive ability of the interest rate variables appears fairly robust in the U.S. data and extends to some degree to the international data. (iii) With few exceptions, all evidence of predictability is found for the shortest horizons and any evidence that does exist tends to disappear as the forecasting horizon increases; this is particularly true for the interest rate variables where the test statistics are often almost monotonically declining in value with the forecasting horizons.

Points (i) and (ii) are discussed at some length in Campbell and Yogo (2006) and Ang and Bekaert (2007), although the international sample used by the latter is somewhat smaller than the one used here. Instead, I will focus on the third point regarding the long-run results. Contrary to many popular beliefs, the results here show that evidence of predictability in the long-run is not stronger than in the short-run. In fact, in most cases the opposite appears true.

If the data are generated by the standard model in equations (1) and (2), predictability in the short-run also implies predictability in the long-run. However, the analytical results in this paper also show that tests lose power as the horizon increases, which could explain the findings presented here. That is, even if the results from the one-period regressions are correct, and there is predictability in some cases, there is no guarantee that this predictability will be evident at longer horizons, given a decrease in the power to detect it. In practice, the evidence of predictability is weak also at short horizons, and it should therefore not be surprising that the null of no predictability cannot be rejected for longer horizons.[18] The empirical results are thus consistent with the model in equations (1) and (2), under which the analytical results were derived.

[18] In making statements about multiple horizons, there is always the issue of multiple testing, which is not adressed here. However, since the baseline assumption that would arise from the theoretical analysis is that the short-run results should be stronger than the long-run results, and these predictions are reflected in the actual results, the multiple testing issue seems less of a concern in the current discussion.

Consistent empirical findings of long-run, but not short-run, predictability, on the other hand, would suggest that equations (1) and (2) are not adequate tools for modelling return predictability.

Torous et al. (2004) and Ang and Bekaert (2007) also find that the evidence of predictability tends to be strongest at shorter horizons, although they do not suggest the possibility that this may be due to a lack of power in long horizon tests. Boudoukh et al. (2005) explicitly question the prevailing view of long-horizon predictability and reach similar conclusions to those presented here, although their focus is on the joint properties of the regression estimators across different horizons. Taken together, there is thus mounting evidence against the previously prevailing view that stock return predictability is more apparent in the long-run than in the short-run.

7 Conclusion

I derive several new results for long-horizon regressions that use overlapping observations when the regressors are endogenous and highly persistent. I show how to properly correct for the overlap in the data in a simple manner that obviates the need for auto-correlation robust standard error methods in these regressions. Further, when the regressors are persistent and endogenous, I show how to correct the long-run OLS estimators and test procedures in a manner similar to that proposed by Campbell and Yogo (2006) for the short-run case.

The analysis also highlights the boundaries of long-horizon regressions. Analytical results, supported by Monte Carlo simulations, show that there are no power gains to long-run tests as long as the ratio between the forecasting horizon and the sample size is fixed. Thus, increasing the forecasting horizon as more data becomes available is not a good strategy.

An empirical application to stock-return predictability illustrates these results and shows that, in line with the theoretical results of this paper, the evidence for predictability is typically weaker as the forecasting horizon gets longer, reflecting at least to some extent the loss of power in long-run tests.

A 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. Unless otherwise noted, all limits as $q, T \to \infty$ are under the condition that $q/T \to 0$.

Proof of Theorem 1. Under the null hypothesis,

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

By standard arguments, $\frac{1}{qT}\sum_{t=1}^{T-q}\sum_{j=1}^{q}u_{t+j}x_t=\frac{1}{qT}\sum_{t=1}^{T-q}\left(u_{t+1}x_t+...+u_{t+q}x_t\right)\Rightarrow\int_0^1 dB_1J_c$, as $q,T\to\infty$, such that $q/T\to 0$, since for any $h>0$, $\frac{1}{T}\sum_{t=1}^{T-q}u_{t+h}x_t\Rightarrow\int_0^1 dB_1J_c$ (Phillips, 1987 and 1988). Therefore, $\frac{T}{q}\left(\hat{\beta}_q-0\right)\Rightarrow\left(\int_0^1 dB_1J_c\right)\left(\int_0^1 J_c^2\right)^{-1}$. ■

Proof of Theorem 2. Let $\mathbf{r}_{+q}^{q}=\left(r_{1+q}\left(q\right),...,r_{T-q}\left(q\right)\right)'$ be the $(T-q)\times 1$ vector of observations, and define $\mathbf{x}$ and $\mathbf{v}_{+q}^{q}$ analogously . Also, let $Q_{\mathbf{v}^q}=\mathbf{I}-\mathbf{v}_{+q}^{q}\left(\mathbf{v}_{+q}^{q\prime}\mathbf{v}_{+q}^{q}\right)^{-1}\mathbf{v}_{+q}^{q\prime}$. The OLS estimator of β_q in (11) is now given by $\hat{\beta}_q^{+}=\left(\mathbf{r}_{+q}^{q\prime}Q_{\mathbf{v}^q}\mathbf{x}\right)\left(\mathbf{x}'Q_{\mathbf{v}^q}\mathbf{x}\right)^{-1}$. Under the null hypothesis, $Q_{\mathbf{v}^q}\mathbf{r}_{+q}^{q}=Q_{\mathbf{v}^q}\mathbf{u}_{+q}^{q}$, and thus $\frac{T}{q}\left(\hat{\beta}_q^{+}-0\right)=\left(q^{-1}T^{-1}\mathbf{u}_{+q}^{q\prime}Q_{\mathbf{v}^q}\mathbf{x}\right)\left(T^{-2}\mathbf{x}'Q_{\mathbf{v}^q}\mathbf{x}\right)^{-1}$. First,

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

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

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

Now, observe that $\hat{\gamma}_{vv}\left(h\right)=\hat{\gamma}_{vv}^{1}\left(h\right)=\frac{1}{T}\sum_{t=1}^{T-q}v_{t+1}v_{t+h+1}$ and $\hat{\gamma}_{vv}^{k}\left(h\right)=\frac{1}{T}\sum_{t=1}^{T-q}v_{t+k}v_{t+k+h}=\frac{1}{T}\sum_{t=k}^{T-q+k-1}v_{t+1}v_{t+h+1}$. $\hat{\gamma}_{vv}^{k}\left(h\right)$ is thus an identical estimator to $\hat{\gamma}_{vv}\left(h\right)$, but uses observations shifted k steps. Letting '$\overset{d}{=}$' denote distributional equivalence, it follows that,

$$\frac{1}{qT}\sum_{t=1}^{T-q} v_{t+q}\left(q\right)^2 \overset{d}{=} \frac{1}{q}\left[q\hat{\gamma}_{vv}\left(0\right)+\left(q-1\right)\hat{\gamma}_{vv}\left(-1\right)+...+\hat{\gamma}_{vv}\left(-\left(q-1\right)\right)+\hat{\gamma}_{vv}\left(q-1\right)+...+\left(q-1\right)\hat{\gamma}_{vv}\left(1\right)\right]$$
$$= \sum_{h=-q+1}^{q-1}\left(1-\frac{|h|}{q}\right)\hat{\gamma}_{vv}\left(h\right)$$

and as $q,T\to\infty$, $\frac{1}{qT}\sum_{t=1}^{T} v_{t+q}\left(q\right)^2 \overset{d}{=} \sum_{h=-q+1}^{q-1}\left(1-\frac{|h|}{q}\right)\hat{\gamma}_{vv}\left(h\right)\to_p \Omega_{22}$, by the results in Andrews (1991), since $qT^{-1}=o\left(1\right)$. Similarly, as $q,T\to\infty$, $\frac{1}{qT}\sum_{t=1}^{T-q} u_{t+q}\left(q\right)v_{t+q}\left(q\right)\to_p \omega_{12}$, and by the same arguments used in the previous proof, $\frac{1}{qT}\mathbf{v}_{+q}^{q\prime}\mathbf{x}=\frac{1}{qT}\sum_{t=1}^{T-q} v_{t+q}\left(q\right)x_t \Rightarrow \int_0^1 dB_2 J_c$. Thus,

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

Finally, as $q,T\to\infty$, using the above results, since $q/T\to 0$ it follows that,

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

■

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

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

where the asymptotic limit follows by same argument as in the previous proof. ■

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

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

■

Proof of Theorem 3. *(i)* Consider first the case when $q/T=o\left(1\right)$ as $q,T\to\infty$. Under the alternative of predictability, by summing up on both sides in equation (1), it follows that

$$\begin{aligned} & r_{t+q}(q) = \beta\left(x_t + x_{t+1} + ... + x_{t+q-1}\right) + u_{t+q}(q) \\ = & \ \beta\left(\left(x_t + \rho x_t + ... + \rho^{q-1}x_t\right) + v_{t+1} + \left(\rho v_{t+1} + v_{t+2}\right) + ... + \sum_{p=2}^{q}\rho^{q-p}v_{t+p-1}\right) + u_{t+q}(q) \\ = & \ \beta_q x_t + \beta\sum_{h=1}^{q-1}\left(\sum_{p=h}^{q-1}\rho^{p-h}\right)v_{t+h} + u_{t+q}(q), \end{aligned}$$

where $\beta_q = \beta\left(1+\rho+...+\rho^{q-1}\right) = \beta\nu(\rho, q) = q + o(1)$, since $\rho = 1 + c/T$. Using the results in Hjalmarsson (2008), it follows easily that, as $q, T \to \infty$, with $q/T = o(1)$, $\frac{T}{q}\left(\frac{\hat{\beta}_q}{q} - \frac{\beta_q}{q}\right) \Rightarrow \frac{\beta}{2}\left(\int_0^1 dB_2 J_c\right)\left(\int_0^1 J_c^2\right)^{-1}$, and thus $\hat{\beta}_q/q = \beta + o_p(1)$. Using the expression above, $\hat{u}_{t+q}(q) = r_{t+q}(q) - \hat{\beta}_q x_t = \left(\beta_q - \hat{\beta}_q\right)x_t + \beta\sum_{h=1}^{q-1}\left(\sum_{p=h}^{q-1}\rho^{p-h}\right)v_{t+h} + u_{t+q}(q)$. Clearly, $\beta\sum_{h=1}^{q-1}\left(\sum_{p=h}^{q-1}\rho^{p-h}\right)v_{t+h}$ is the dominant term. Now, using the fact that $\rho = 1 + c/T$,

$$\begin{aligned} & \frac{1}{q^3 T}\beta^2\sum_{t=1}^{T-q}\left(\sum_{h=1}^{q-1}\left(\sum_{p=h}^{q-1}\rho^{p-h}\right)v_{t+h}\right)^2 = \frac{1}{q^3 T}\beta^2\sum_{t=1}^{T-q}\left[\sum_{h=1}^{q-1}\sum_{k=1}^{q-1}(q-h)(q-k)v_{t+h}v_{t+k}\right] + o_p(1) \\ = & \ \frac{1}{q^3}\beta^2\sum_{h=1}^{q-1}(q-h)^2\frac{1}{T}\sum_{t=1}^{T-q}v_{t+h}^2 + o_p(1) = \beta^2\frac{1}{q^3}\left(\frac{1}{6}q - \frac{1}{2}q^2 + \frac{1}{3}q^3\right)\frac{1}{T}\sum_{t=1}^{T-q}v_{t+h}^2 + o_p(1) \to_p \beta^2\frac{1}{3}\omega_{22}, \end{aligned}$$

where the first equality follows from the local-to-unity nature of $\rho = 1 + c/T$, and the second equality from the martingale difference assumption on v_t. The scaled $t-$statistic, for $H_0 : \beta = 0$, thus satisfies

$$\frac{q}{T}\frac{t}{\sqrt{q}} = \frac{\frac{\hat{\beta}_q}{q}}{\sqrt{\left(\frac{1}{q^3 T}\sum_{t=1}^{T-q}\hat{u}_t(q)^2\right)\left(\frac{1}{T^2}\sum_{t=1}^{T-q}x_t^2\right)^{-1}}} \Rightarrow \frac{\beta}{\sqrt{\beta^2\frac{1}{3}\omega_{22}\left(\int_0^1 J_c^2\right)^{-1}}} = \frac{1}{\sqrt{\frac{1}{3}\omega_{22}\left(\int_0^1 J_c^2\right)^{-1}}}.$$

(ii) Consider next the case when $q/T = \lambda$ as $q, T \to \infty$. By summing up on both sides in equation (1), $r_{t+q}(q) = \beta\left(x_t + x_{t+1} + ... + x_{t+q-1}\right) + u_{t+q}(q) \equiv \beta x_{t+q-1}(q) + u_{t+q}(q)$, and the fitted regression is $r_{t+q}(q) = \beta x_t + u_{t+q}(q)$. It follows that,

$$\hat{\beta}_q = \left(\sum_{t=1}^{T-q} r_{t+q}(q)\,x_t\right)\left(\sum_{t=1}^{T-q}x_t^2\right)^{-1} = \beta\left(\sum_{t=1}^{T-q}x_{t+q-1}(q)\,x_t\right)\left(\sum_{t=1}^{T-q}x_t^2\right)^{-1} + \left(\sum_{t=1}^{T-q}u_{t+q}(q)\,x_t\right)\left(\sum_{t=1}^{T-q}x_t^2\right)^{-1}.$$

Now, when $q/T = \lambda$ as $q, T \to \infty$, it follows that $\frac{u_{t+q}(q)}{\sqrt{T}} = \frac{1}{\sqrt{T}}\sum_{j=1}^{q}u_{t+j} \Rightarrow B_1(r+\lambda) - B_1(r) \equiv B_1(r;\lambda)$, and thus,

$$\left(\frac{1}{T}\sum_{t=1}^{T-q}\frac{u_{t+q}(q)}{\sqrt{T}}\frac{x_t}{\sqrt{T}}\right)\left(\frac{1}{T}\sum_{t=1}^{T-q}\frac{x_t^2}{T}\right)^{-1} \Rightarrow \left(\int_0^{1-\lambda}B_1(r;\lambda)\,J_c(r)\,dr\right)\left(\int_0^{1-\lambda}J_c^2\right)^{-1} = O_p(1).$$

Similarly, $\frac{1}{T^{3/2}} x_{t+q-1}(q) = \frac{1}{T}\sum_{j=1}^{q-1} \frac{x_{t+j}}{\sqrt{T}} \Rightarrow \int_r^{r+\lambda} J_c(r) \equiv J_c(r;\lambda)$, and by the CMT,

$$\left(\frac{1}{T^3}\sum_{t=1}^{T-q} x_{t+q-1}(q)\, x_t\right)\left(\frac{1}{T^2}\sum_{t=1}^{T-q} x_t^2\right)^{-1} \Rightarrow \left(\int_0^{1-\lambda} J_c(r;\lambda)\, J_c(r)\right)\left(\int_0^{1-\lambda} J_c^2\right)^{-1}.$$

It follows that $\frac{\hat{\beta}_q}{T} \Rightarrow \beta\left(\int_0^{1-\lambda} J_c(r;\lambda)\, J_c(r)\right)\left(\int_0^{1-\lambda} J_c^2\right)^{-1}$. The fitted residuals satisfy $\hat{u}_{t+q}(q) = \beta x_{t+q-1}(q) + u_{t+q}(q) - \hat{\beta}_q x_t$, and

$$\frac{1}{T^4}\sum_{t=1}^{T-q} \hat{u}_t(q)^2 = \frac{1}{T^4}\sum_{t=1}^{T-q}\left(\beta x_{t+q-1}(q) + u_{t+q}(q) - \hat{\beta}_q x_t\right)^2 = \beta^2 \frac{1}{T}\sum_{t=1}^{T-q}\left(\frac{x_{t+q-1}(q)}{T^{3/2}}\right)^2 + O_p\left(T^{-1}\right) \Rightarrow \beta^2 \int_0^{1-\lambda} J_c(r;\lambda)^2.$$

The scaled $t-$statistic of Valkanov (2003), for testing the null hypothesis of $H_0 : \beta = 0$, therefore satisfy, as $q, T \to \infty$, with $q/T = \lambda$,

$$\frac{t}{\sqrt{T}} = \frac{\frac{\hat{\beta}_q}{T}\sqrt{\left(\frac{1}{T^2}\sum_{t=1}^{T-q} x_t^2\right)}}{\sqrt{\left(\frac{1}{T^4}\sum_{t=1}^{T-q}\hat{u}_t(q)^2\right)}} \Rightarrow \frac{\beta\left(\int_0^{1-\lambda} J_c(r;\lambda)\, J_c(r)\right)\left(\int_0^{1-\lambda} J_c^2\right)^{-1}\left(\int_0^{1-\lambda} J_c^2\right)^{1/2}}{\sqrt{\beta^2\int_0^{1-\lambda} J_c(r;\lambda)^2}} = \frac{\int_0^{1-\lambda} J_c(r;\lambda)\, J_c(r)}{\sqrt{\left(\int_0^{1-\lambda} J_c^2(r;\lambda)\right)\left(\int_0^{1-\lambda} J_c^2\right)}}.$$

■

Proof of Theorem 4. Using the results in Hjalmarsson (2008) again, it follows that for $\beta = b/q$, $\frac{T}{q}\left(\hat{\beta}_q - \beta_q\right) \Rightarrow \left(\int_0^1 dB_1 J_c + \frac{b}{2}\int_0^1 dB_2 J_c\right)\left(\int_0^1 J_c^2\right)^{-1}$, where $\beta_q = b\frac{\nu(\rho,q)}{q} = b + o(1)$, and thus $\hat{\beta}_q \sim b + \frac{q}{T}\left(\int_0^1 dB_1 J_c + \frac{b}{2}\int_0^1 dB_2 J_c\right)\left(\int_0^1 J_c^2\right)^{-1}$. As before,

$$r_{t+q}(q) = \beta_q x_t + \beta\sum_{h=1}^{q-1}\left(\sum_{p=h}^{q-1}\rho^{p-h}\right) v_{t+h} + u_{t+q}(q) = b x_t + \frac{b}{q}\sum_{h=1}^{q-1}\left(\sum_{p=h}^{q-1}\rho^{p-h}\right) v_{t+h} + u_{t+q}(q) + o(1),$$

and, $\hat{u}_{t+q}(q) = r_{t+q}(q) - \hat{\beta}_q x_t = \left(b - \hat{\beta}_q\right) x_t + \frac{b}{q}\sum_{h=1}^{q-1}\left(\sum_{p=h}^{q-1}\rho^{p-h}\right) v_{t+h} + u_{t+q}(q) + o(1)$. Now, consider

$$\begin{aligned}
&\frac{1}{qT}\sum_{t=1}^{T-q}\left(\frac{b}{q}\sum_{h=1}^{q-1}\left(\sum_{p=h}^{q-1}\rho^{p-h}\right) v_{t+h} + u_{t+q}(q)\right)^2 \\
= \; &\frac{1}{qT}\sum_{t=1}^{T-q}\left(\left(\frac{b}{q}\sum_{h=1}^{q-1}\left(\sum_{p=h}^{q-1}\rho^{p-h}\right) v_{t+h}\right)^2 + u_{t+q}(q)^2 + 2\frac{b}{q}\sum_{h=1}^{q-1}\left(\sum_{p=h}^{q-1}\rho^{p-h}\right) v_{t+h} u_{t+q}(q)\right).
\end{aligned}$$

By previous results, $\frac{1}{qT}\sum_{t=1}^{T-q}\left(\frac{b}{q}\sum_{h=1}^{q-1}\left(\sum_{p=h}^{q-1}\rho^{p-h}\right) v_{t+h}\right)^2 \to_p \frac{b^2\omega_{22}}{3}$ and $\frac{1}{qT}\sum_{t=1}^{T-q} u_{t+q}(q)^2 \to_p \omega_{11}$. Further, by similar arguments as before,

$$
\begin{aligned}
\frac{1}{qT}\sum_{t=1}^{T-q}\frac{b}{q}\sum_{h=1}^{q-1}\left(\sum_{p=h}^{q-1}\rho^{p-h}\right)v_{t+h}u_{t+q}\left(q\right) &= b\frac{1}{qT}\sum_{t=1}^{T-q}\sum_{j=1}^{q}\sum_{h=1}^{q-1}\left(1-\frac{h}{q}\right)v_{t+h}u_{t+j}+o_p\left(1\right) \\
= \quad b\frac{1}{T}\sum_{t=1}^{T-q}\frac{1}{q}\sum_{h=1}^{q-1}\left(1-\frac{h}{q}\right)v_{t+h}u_{t+h}+o_p\left(1\right)\rightarrow_p b\frac{1}{2}\omega_{12},
\end{aligned}
$$

since $\frac{1}{q}\sum_{h=1}^{q}\left(1-\frac{h}{q}\right)\rightarrow\frac{1}{2}$ as $q\rightarrow\infty$. Thus, $\frac{1}{qT}\sum_{t=1}^{T-q}\hat{u}_{t+q}\left(q\right)^2\rightarrow_p \omega_{11}+b\omega_{12}+\frac{b^2\omega_{22}}{3}$ and

$$
\frac{t}{\sqrt{q}}=\frac{\hat{\beta}_q}{\frac{q}{T}\sqrt{\left(\frac{1}{qT}\sum_{t=1}^{T-q}\hat{u}_t\left(q\right)^2\right)\left(\frac{1}{T^2}\sum_{t=1}^{T-q}x_t^2\right)^{-1}}}\sim\frac{T}{q}\frac{b+\frac{q}{T}\left(\int_0^1 dB_1J_c+\frac{b}{2}\int_0^1 dB_2J_c\right)\left(\int_0^1 J_c^2\right)^{-1}}{\sqrt{\left(\omega_{11}+b\omega_{12}+\frac{b^2\omega_{22}}{3}\right)\left(\int_0^1 J_c^2\right)^{-1}}}+o_p\left(1\right).
$$

■

References

Andrews, D.W.K., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation, Econometrica 59, 817-858.

Ang, A., and G. Bekaert, 2007. Stock return predictability: is it there? Review of Financial Studies 20, 651-707.

Boudoukh J., and M. Richardson, 1993. Stock returns and inflation: a long-horizon perspective, American Economic Review 83, 1346-1355.

Boudoukh J., M. Richardson, and R.F. Whitelaw, 2005. The myth of long-horizon predictability, forthcoming Review of Financial Studies.

Berkowitz, J., and L. Giorgianni, 2001. Long-horizon exchange rate predictability?, Review of Economics and Statistics 83, 81-91.

Campbell, J.Y., 2001. Why long horizons? A study of power against persistent alternatives, Journal of Empirical Finance 8, 459-491.

Campbell, J.Y., and R. Shiller, 1988. Stock prices, earnings, and expected dividends, Journal of Finance 43, 661-676.

Campbell, J.Y., and M. Yogo, 2005. Implementing the econometric methods in "Efficient tests of stock return predictability". Working paper, University of Pennsylvania.

Campbell, J.Y., and M. Yogo, 2006. Efficient tests of stock return predictability, Journal of Financial Economics 81, 27-60.

Cavanagh, C., G. Elliot, and J. Stock, 1995. Inference in models with nearly integrated regressors, Econometric Theory 11, 1131-1147.

Corbae D., S. Ouliaris, and P.C.B. Phillips, 2002. Band spectral regression with trending data, Econometrica 70, 1067-1109.

Daniel, K., 2001. The power and size of mean reversion tests, Journal of Empirical Finance 8, 493-535.

Elliot G., T.J. Rothenberg, and J.H. Stock, 1996. Efficient tests for an autoregressive unit root, Econometrica 64, 813-836.

Engle, R.F., 1984. Wald, likelihood ratio and lagrange multiplier tests in econometrics, in Handbook of Econometrics, vol II., edited by Z. Griliches, and M.D. Intriligator. Amsterdam, North Holland.

Fama, E.F., and K.R. French, 1988. Dividend yields and expected stock returns, Journal of Financial Economics 22, 3-25.

Fisher, M.E., and J.J. Seater, 1993. Long-run neutrality and superneutrality in an ARIMA Framework, American Economic Review 83, 402-415.

Goetzman W.N., and P. Jorion, 1993. Testing the predictive power of dividend yields, Journal of Finance 48, 663-679.

Hansen, L.P., and R.J. Hodrick, 1980. Forward exchange rates as optimal predictors of future spot rates: An Econometric Analysis, Journal of Political Economy 88, 829-853.

Hjalmarsson, E., 2007. Fully modified estimation with nearly integrated regressors, Finance Research Letters 4, 92-94.

Hjalmarsson, E., 2008. Interpreting long-horizon estimates in predictive regressions, Finance Research Letters 5, 104-117.

Hodrick, R.J., 1992. Dividend yields and expected stock returns: alternative procedures for inference and measurement, Review of Financial Studies 5, 357-386.

Jansson, M., and M.J. Moreira, 2006. Optimal inference in regression models with nearly integrated regressors, Econometrica 74, 681-714.

Kaminsky G.L., and S.L. Schmukler, 2002. Short-run pain, long-run gain: the effects of financial liberalization, Working Paper, George Washington University.

Lanne, M., 2002. Testing the predictability of stock returns, Review of Economics and Statistics 84, 407-415.

Lettau, M., and S. Van Nieuwerburgh, 2007. Reconciling the return predictability evidence, forthcoming Review of Financial Studies.

Lewellen, J., 2004. Predicting returns with financial ratios, Journal of Financial Economics, 74, 209-235.

Mankiw, N.G., and M.D. Shapiro, 1986. Do we reject too often? Small sample properties of tests of rational expectations models, Economics Letters 20, 139-145.

Mark, N.C., 1995. Exchange rates and fundamentals: evidence on long-horizon predictability, American Economic Review 85, 201-218.

Mark, N.C., and D. Sul, 2004. The use of predictive regressions at alternative horizons in finance and economics, NBER Technical Working Paper 298.

Mishkin, F.S., 1990. What does the term structure tell us about future inflation?, Journal of Monetary Economics 25, 77-95.

Mishkin, F.S., 1992. Is the Fisher effect for real?, Journal of Monetary Economics 30, 195-215.

Moon, R., A. Rubia, and R. Valkanov, 2004. Long-horizon regressions when the predictor is slowly varying, Working Paper, UCLA, Anderson School of Management.

Nelson, C.R., and M.J. Kim, 1993. Predictable stock returns: the role of small sample bias, Journal of Finance 48, 641-661.

Newey, W., and K. West, 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, Econometrica 55, 703-708.

Phillips, P.C.B, 1987. Towards a unified asymptotic theory of autoregression, Biometrika 74, 535-547.

Phillips, P.C.B, 1988. Regression theory for near-integrated time series, Econometrica 56, 1021-1043.

Phillips, P.C.B, 1991a. Optimal inference in cointegrated systems, Econometrica 59, 283-306.

Phillips, P.C.B, 1991b. Spectral regression for cointegrated time series, in Nonparametric and Semiparametric Methods in Economics and Statistics, edited by W. Barnett, J. Powell, and G. Tauchen. Cambridge, Cambridge University Press.

Polk, C., S. Thompson, and T. Vuolteenaho, 2006. Cross-sectional forecasts of the equity premium, Journal of Financial Economics 81, 101-141.

Rapach D.E., and M.E. Wohar, 2005. Valuation ratios and long-horizon stock price predictability, Journal of Applied Econometrics 20, 327-344.

Richardson, M., and T. Smith, 1991. Tests of financial models in the presence of overlapping observations, Review of Financial Studies 4, 227-254.

Richardson, M., and J.H. Stock, 1989. Drawing inferences from statistics based on multiyear asset returns, Journal of Financial Economics 25, 323-348.

Rossi, B., 2005. Testing long-horizon predictive ability with high persistence, and the Meese-Rogoff puzzle, International Economic Review 46, 61-92.

Stambaugh, R., 1999. Predictive regressions, Journal of Financial Economics 54, 375-421.

Stock, J.H., 1991. Confidence intervals for the largest autoregressive root in U.S. economic time-series. Journal of Monetary Economics 28, 435-460.

Torous, W., R. Valkanov, and S. Yan, 2004. On predicting stock returns with nearly integrated explanatory variables, Journal of Business 77, 937-966.

Valkanov, R., 2003. Long-horizon regressions: theoretical results and applications, Journal of Financial Economics 68, 201-232.

Table 1: Finite sample sizes for the scaled long-run OLS $t-$test, the scaled Bonferroni test and the scaled infeasible test. The first column 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 show, 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 equal to 0 and -10 are shown; for $T = 100$, these values of c correspond to autoregressive roots $\rho = 1$ and $\rho = 0.9$, respectively, and for $T = 500$, they correspond to $\rho = 1$ and $\rho = 0.98$.

	Long-Run OLS $t-$Test						Bonferroni Test						Infeasible Test (using true value of c)					
δ	0.00	-0.25	-0.50	-0.75	-0.90	-0.95	0.00	-0.25	-0.50	-0.75	-0.90	-0.95	0.00	-0.25	-0.50	-0.75	-0.90	-0.95
q									$T = 100, c = 0$									
1	0.058	0.102	0.188	0.295	0.385	0.419	0.044	0.045	0.042	0.046	0.053	0.068	0.053	0.056	0.051	0.057	0.057	0.052
5	0.052	0.110	0.186	0.303	0.399	0.434	0.043	0.046	0.042	0.043	0.053	0.057	0.054	0.065	0.061	0.057	0.047	0.041
10	0.048	0.102	0.185	0.306	0.421	0.458	0.041	0.048	0.046	0.042	0.044	0.048	0.065	0.063	0.064	0.060	0.050	0.043
15	0.048	0.101	0.171	0.294	0.405	0.451	0.042	0.047	0.044	0.035	0.032	0.034	0.066	0.066	0.064	0.060	0.049	0.041
20	0.043	0.088	0.166	0.289	0.392	0.435	0.041	0.042	0.043	0.031	0.025	0.023	0.064	0.063	0.060	0.059	0.048	0.044
25	0.040	0.082	0.154	0.261	0.365	0.403	0.033	0.036	0.034	0.025	0.016	0.013	0.060	0.059	0.061	0.053	0.046	0.040
									$T = 500, c = 0$									
1	0.051	0.099	0.176	0.298	0.381	0.415	0.045	0.043	0.034	0.034	0.040	0.043	0.050	0.054	0.052	0.048	0.051	0.048
25	0.051	0.100	0.179	0.305	0.407	0.438	0.043	0.042	0.036	0.029	0.034	0.041	0.059	0.058	0.064	0.049	0.047	0.039
50	0.053	0.108	0.181	0.301	0.410	0.458	0.045	0.042	0.039	0.033	0.028	0.035	0.067	0.066	0.063	0.055	0.048	0.037
75	0.051	0.099	0.177	0.287	0.398	0.450	0.042	0.045	0.036	0.030	0.024	0.022	0.064	0.067	0.068	0.058	0.050	0.038
100	0.045	0.092	0.161	0.280	0.382	0.436	0.036	0.039	0.035	0.023	0.017	0.018	0.067	0.062	0.067	0.059	0.047	0.036
125	0.039	0.080	0.146	0.258	0.363	0.406	0.032	0.029	0.025	0.017	0.014	0.012	0.058	0.055	0.056	0.054	0.044	0.042
									$T = 100, c = -10$									
1	0.052	0.071	0.097	0.123	0.137	0.143	0.044	0.04	0.032	0.034	0.043	0.051	0.054	0.053	0.054	0.053	0.057	0.059
5	0.044	0.058	0.083	0.102	0.117	0.130	0.030	0.033	0.026	0.025	0.034	0.037	0.049	0.047	0.047	0.049	0.046	0.050
10	0.033	0.045	0.063	0.082	0.098	0.108	0.027	0.024	0.021	0.016	0.017	0.020	0.044	0.043	0.044	0.046	0.041	0.038
15	0.022	0.036	0.048	0.067	0.082	0.089	0.021	0.022	0.017	0.011	0.009	0.011	0.041	0.037	0.040	0.040	0.034	0.031
20	0.018	0.027	0.039	0.056	0.064	0.069	0.017	0.014	0.012	0.007	0.006	0.007	0.035	0.035	0.030	0.031	0.028	0.029
25	0.011	0.019	0.028	0.039	0.046	0.048	0.013	0.011	0.009	0.005	0.005	0.003	0.026	0.029	0.028	0.029	0.025	0.023
									$T = 500, c = -10$									
1	0.050	0.069	0.091	0.122	0.136	0.140	0.047	0.033	0.027	0.023	0.033	0.039	0.049	0.056	0.052	0.052	0.053	0.055
25	0.038	0.059	0.078	0.106	0.123	0.128	0.033	0.031	0.023	0.020	0.021	0.027	0.046	0.049	0.043	0.048	0.044	0.041
50	0.033	0.045	0.070	0.087	0.104	0.104	0.022	0.030	0.022	0.014	0.014	0.016	0.044	0.045	0.044	0.043	0.037	0.039
75	0.024	0.036	0.054	0.076	0.086	0.091	0.024	0.023	0.016	0.010	0.009	0.008	0.039	0.040	0.039	0.041	0.038	0.030
100	0.018	0.029	0.040	0.057	0.064	0.071	0.018	0.018	0.015	0.008	0.005	0.006	0.034	0.035	0.034	0.031	0.029	0.027
125	0.010	0.020	0.031	0.038	0.051	0.053	0.015	0.015	0.01	0.005	0.004	0.003	0.029	0.025	0.028	0.026	0.022	0.022

Table 2: Characteristics of the predictor variables in the U.S. data. This table reports the key time-series characteristics of the dividend-price ratio $(d - p)$, the earnings-price ratio $(e - p)$, the short interest rate (r_3), and the yield spread $(y - r_1)$. The S&P 500 variables are on an annual frequency, whereas results for both the annual and monthly CRSP data are reported. All series end in 2002. The first two columns indicate the data set and predictor variable being used. The following three columns show the sampling frequency, the start date of the sample period, and the number of observations in that sample. The column labeled DF-GLS gives the value of the DF-GLS unit-root test statistic, and the column labeled δ gives the estimated correlations between the innovations to the predictor variables and the innovations to the corresponding excess returns. The last two columns give the 95% confidence intervals for the autoregressive root ρ and the corresponding local-to-unity parameter c, obtained by inverting the DF-GLS unit-root test statistic.

Series	Variable	Sample Freq.	Sample Begins	Obs.	DF-GLS	δ	95% CI for ρ	95% CI for c
S&P 500	$d - p$	Annual	1880	123	-0.855	-0.845	$[0.949, 1.033]$	$[-6.107, 4.020]$
S&P 500	$e - p$	Annual	1880	123	-2.888	-0.962	$[0.768, 0.965]$	$[-28.262, -4.232]$
CRSP	$d - p$	Annual	1926	77	-1.033	-0.721	$[0.903, 1.050]$	$[-7.343, 3.781]$
CRSP	$e - p$	Annual	1926	77	-2.229	-0.957	$[0.748, 1.000]$	$[-19.132, -0.027]$
CSRP	$d - p$	Monthly	1926.12	913	-1.657	-0.950	$[0.986, 1.003]$	$[-12.683, 2.377]$
CRSP	$e - p$	Monthly	1926.12	913	-1.859	-0.987	$[0.984, 1.002]$	$[-14.797, 1.711]$
CRSP	$d - p$	Monthly	1952.1	612	-0.072	-0.965	$[0.996, 1.007]$	$[-2.736, 4.555]$
CRSP	$e - p$	Monthly	1952.1	612	-0.953	-0.982	$[0.989, 1.006]$	$[-6.722, 3.893]$
CSRP	r_3	Annual	1952	51	-2.078	0.039	$[0.647, 1.014]$	$[-17.309, 0.703]$
CRSP	$y - r_1$	Annual	1952	51	-3.121	-0.243	$[0.363, 0.878]$	$[-31.870, -6.100]$
CSRP	r_3	Monthly	1952.1	612	-1.557	-0.070	$[0.981, 1.004]$	$[-11.683, 2.714]$
CRSP	$y - r_1$	Monthly	1952.1	612	-4.085	-0.067	$[0.921, 0.974]$	$[-48.847, -15.890]$

Table 3: Characteristics of the predictor variables in the international data. This table reports the key time-series characteristics of the dividend-price ratio ($d-p$), the short interest rate (r_s), and the term spread ($r_l - r_s$). All data are on a monthly frequency, and all series end in 2004. The first two columns indicate the country and predictor variable being used, and the next two columns show the start date of the sample period and the number of observations in that sample. The column labeled DF-GLS gives the value of the DF-GLS unit-root test statistic, and the column labeled δ gives the estimated correlations between the innovations to the predictor variables and the innovations to the corresponding excess returns. The last two columns give the 95% confidence intervals for the autoregressive root ρ and the corresponding local-to-unity parameter c, obtained by inverting the DF-GLS unit-root test statistic.

Country	Variable	Sample Begins	Obs.	DF-GLS	δ	95% CI for ρ	95% CI for c
Australia	$d-p$	1882.12	1456	−2.054	−0.567	[0.988, 1.001]	[−17.019, 0.829]
Belgium	$d-p$	1952.1	627	−1.927	−0.761	[0.975, 1.002]	[−15.540, 1.376]
Canada	$d-p$	1934.3	841	−0.646	−0.746	[0.994, 1.005]	[−4.960, 4.217]
France	$d-p$	1941.5	755	−1.468	−0.618	[0.986, 1.004]	[−10.848, 2.954]
Germany	$d-p$	1953.2	614	−1.725	−0.761	[0.978, 1.003]	[−13.377, 2.112]
Italy	$d-p$	1925.3	949	−1.766	−0.150	[0.985, 1.002]	[−13.802, 1.972]
Japan	$d-p$	1949.7	657	−0.232	−0.533	[0.995, 1.007]	[−3.209, 4.476]
Sweden	$d-p$	1919.2	1022	−2.403	−0.587	[0.979, 0.999]	[−21.373, −1.040]
UK	$d-p$	1924.2	962	−3.250	−0.637	[0.965, 0.992]	[−33.865, −7.243]
Australia	$d-p$	1952.1	627	−1.335	−0.736	[0.985, 1.005]	[−9.666, 3.286]
Belgium	$d-p$	1952.1	627	−1.927	−0.761	[0.975, 1.002]	[−15.540, 1.376]
Canada	$d-p$	1952.1	627	−0.078	−0.900	[0.996, 1.007]	[−2.753, 4.552]
France	$d-p$	1952.1	627	−1.372	−0.760	[0.984, 1.005]	[−9.951, 3.204]
Germany	$d-p$	1953.2	614	−1.725	−0.761	[0.978, 1.003]	[−13.377, 2.112]
Italy	$d-p$	1952.1	627	−1.421	−0.592	[0.983, 1.005]	[−10.398, 3.078]
Japan	$d-p$	1952.1	627	0.253	−0.535	[0.997, 1.008]	[−1.876, 4.702]
Sweden	$d-p$	1952.1	627	−1.812	−0.762	[0.977, 1.003]	[−14.285, 1.853]
UK	$d-p$	1952.1	627	−1.576	−0.752	[0.981, 1.004]	[−11.865, 2.656]
Australia	r_s	1952.1	627	−0.882	−0.164	[0.990, 1.006]	[−6.295, 3.993]
Belgium	r_s	1952.1	627	−1.128	−0.139	[0.987, 1.006]	[−7.993, 3.641]
Canada	r_s	1952.1	627	−1.451	−0.176	[0.983, 1.005]	[−10.681, 3.001]
France	r_s	1952.1	627	−1.498	−0.159	[0.982, 1.005]	[−11.130, 2.874]
Germany	r_s	1953.1	615	−2.263	−0.051	[0.968, 1.000]	[−19.544, −0.197]
Italy	r_s	1952.1	627	−1.192	−0.108	[0.986, 1.006]	[−8.457, 3.544]
Japan	r_s	1952.1	627	−0.195	−0.096	[0.995, 1.007]	[−3.100, 4.499]
Sweden	r_s	1952.1	627	−1.547	−0.155	[0.981, 1.004]	[−11.590, 2.744]
UK	r_s	1952.1	627	−1.250	−0.242	[0.986, 1.006]	[−8.972, 3.447]
Australia	$r_l - r_s$	1952.1	627	−1.624	0.024	[0.980, 1.004]	[−12.343, 2.512]
Belgium	$r_l - r_s$	1952.1	627	−3.249	−0.027	[0.946, 0.988]	[−33.847, −7.235]
Canada	$r_l - r_s$	1952.1	627	−3.307	0.059	[0.944, 0.987]	[−34.840, −7.900]
France	$r_l - r_s$	1952.1	627	−2.424	0.068	[0.965, 0.998]	[−21.647, −1.166]
Germany	$r_l - r_s$	1953.1	615	−2.751	−0.032	[0.957, 0.995]	[−26.159, −3.251]
Italy	$r_l - r_s$	1952.1	627	−3.373	−0.002	[0.943, 0.987]	[−35.981, −8.409]
Japan	$r_l - r_s$	1952.1	627	−1.306	0.040	[0.985, 1.005]	[−9.402, 3.341]
Sweden	$r_l - r_s$	1952.1	627	−5.169	0.093	[0.885, 0.951]	[−71.688, −30.638]
UK	$r_l - r_s$	1952.1	627	−2.226	0.031	[0.969, 1.000]	[−19.092, −0.011]

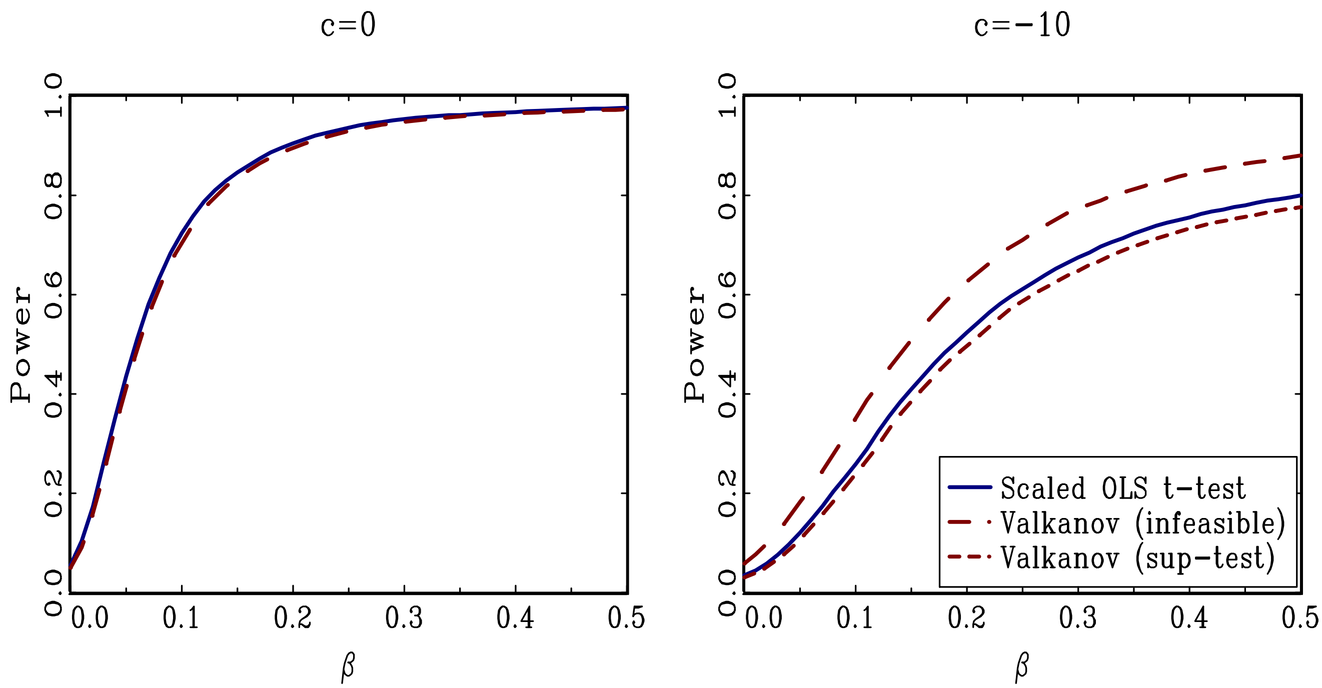

Figure 1: Power curves for exogenous regressors with $T = 100$, $q = 10$, and $\delta = 0.0$. The graphs show the average rejection rates for a one-sided 5 percent test of the null hypothesis of $\beta = 0$ against a positive alternative. The $x-$axis shows the true value of the parameter β, and the $y-$axis indicates the average rejection rate. The left-hand graph gives the results for the case of $c = 0$ ($\rho = 1$), and the right-hand graph gives the results for $c = -10$ ($\rho = 0.9$). The results for the scaled OLS $t-$test derived in this paper are given by the solid lines, the results for Valkanov's infeasible test are given by the long dashed lines, and the results for Valkanov's feasible sup-bound test are given by the short 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.

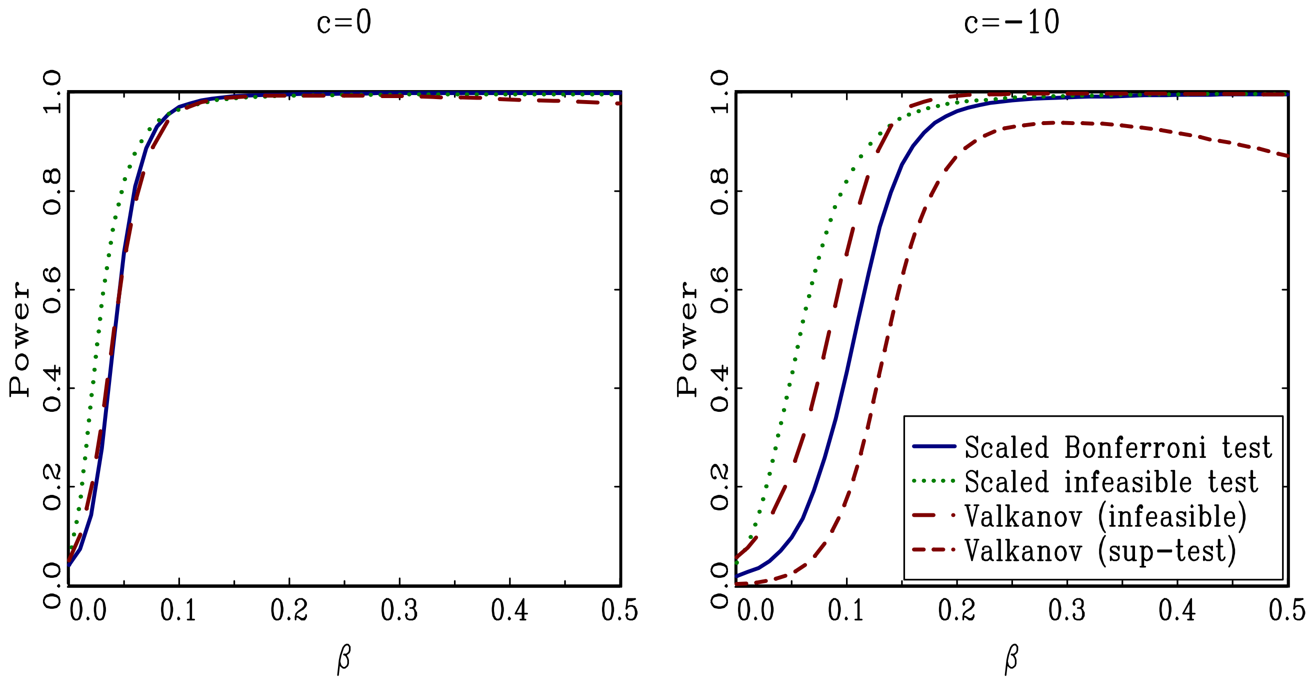

Figure 2: Power curves for endogenous regressors with $T = 100$, $q = 10$, and $\delta = -0.9$. The graphs show the average rejection rates for a one-sided 5 percent test of the null hypothesis of $\beta = 0$ against a positive alternative. The $x-$axis shows the true value of the parameter β, and the $y-$axis indicates the average rejection rate. The left-hand graph gives the results for the case of $c = 0$ ($\rho = 1$), and the right-hand graph gives the results for $c = -10$ ($\rho = 0.9$). The results for the scaled Bonferroni test are given by the solid lines, the results for the scaled infeasible $t-$test $\left(t_q^+ / \sqrt{q}\right)$ from the augmented regression equation, which uses knowledge of the true value of c, are given by the dotted line, the results for Valkanov's infeasible test are given by the long dashed lines, and the results for Valkanov's feasible sup-bound test are given by the short 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.

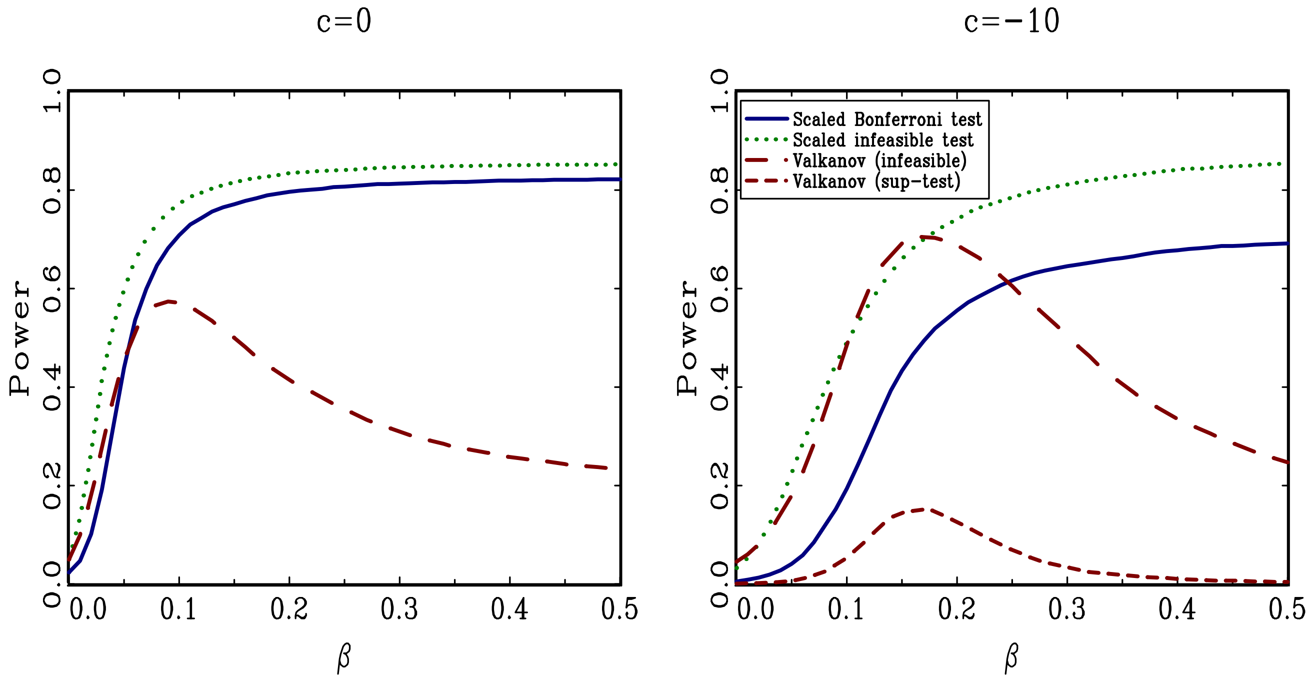

Figure 3: Power curves for endogenous regressors with $T = 100$, $q = 20$, and $\delta = -0.9$. The graphs show the average rejection rates for a one-sided 5 percent test of the null hypothesis of $\beta = 0$ against a positive alternative. The $x-$axis shows the true value of the parameter β, and the $y-$axis indicates the average rejection rate. The left-hand graph gives the results for the case of $c = 0$ ($\rho = 1$), and the right-hand graph gives the results for $c = -10$ ($\rho = 0.9$). The results for the scaled Bonferroni test are given by the solid lines, the results for the scaled infeasible $t-$test $\left(t_q^+ / \sqrt{q}\right)$ from the augmented regression equation, which uses knowledge of the true value of c, are given by the dotted line, the results for Valkanov's infeasible test are given by the long dashed lines, and the results for Valkanov's feasible sup-bound test are given by the short 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.

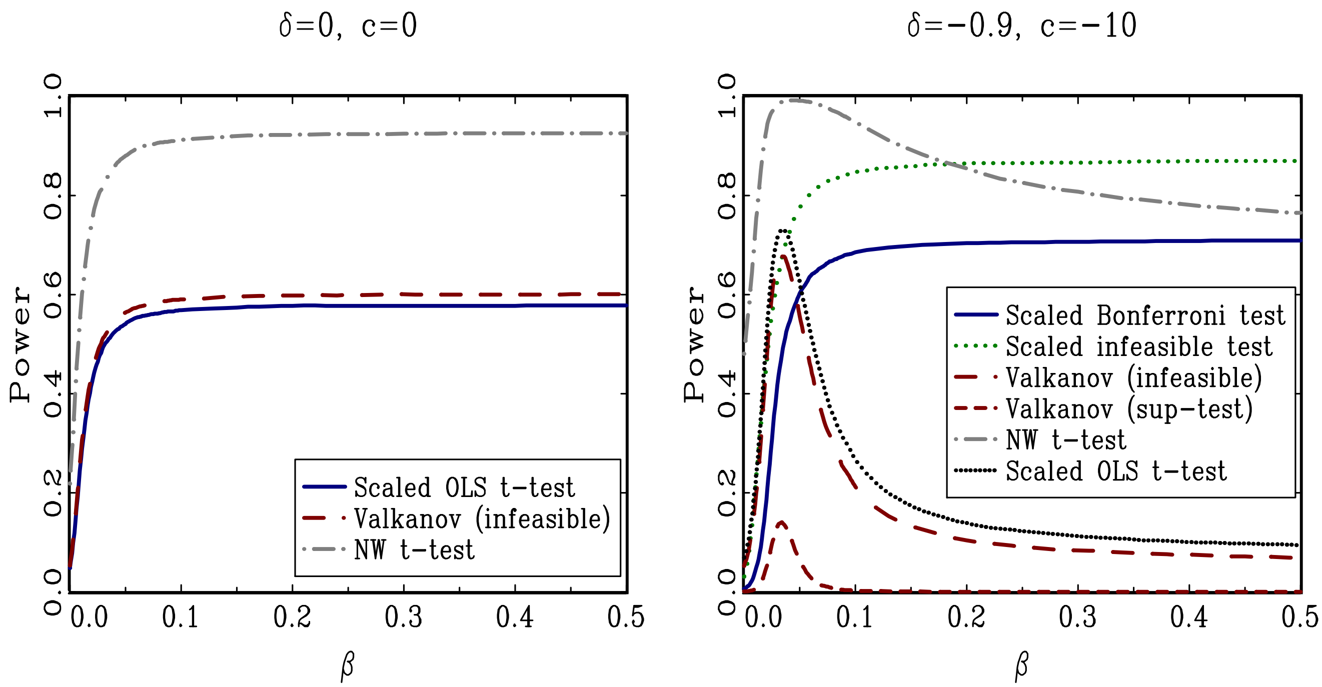

Figure 4: Power curves for $T = 500$, and $q = 100$. The graphs show the average rejection rates for a one-sided 5 percent test of the null hypothesis of $\beta = 0$ against a positive alternative. The $x-$axis shows the true value of the parameter β, and the $y-$axis indicates the average rejection rate. The left hand graph gives the results for the case of exogenous regressors with $\delta = 0$ and $c = 0$. The results for the scaled OLS $t-$test are given by the solid lines, the results for Valkanov's infeasible test, which coincides with Valkanov's sup-bound test for $c = 0$, are given by the long dashed lines, and the results for the (non-scaled) $t-$test using Newey-West standard errors are given by the dotted and dashed line. The right hand graph gives the results for the case of endogenous regressors with $\delta = -0.9$ and $c = -10$. The results for the scaled Bonferroni test are given by the solid lines, the results for the scaled infeasible $t-$test $\left(t_q^+ \big/ \sqrt{q}\right)$ from the augmented regression equation, which uses knowledge of the true value of c, are given by the dotted line, the results for Valkanov's infeasible test are given by the long dashed lines, the results for Valkanov's feasible sup-bound test are given by the short dashed lines, the results for the (non-scaled) $t-$test using Newey-West standard errors are given by the dotted and dashed line, and the results for the scaled OLS $t-$test are given by the finely dotted line. 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.

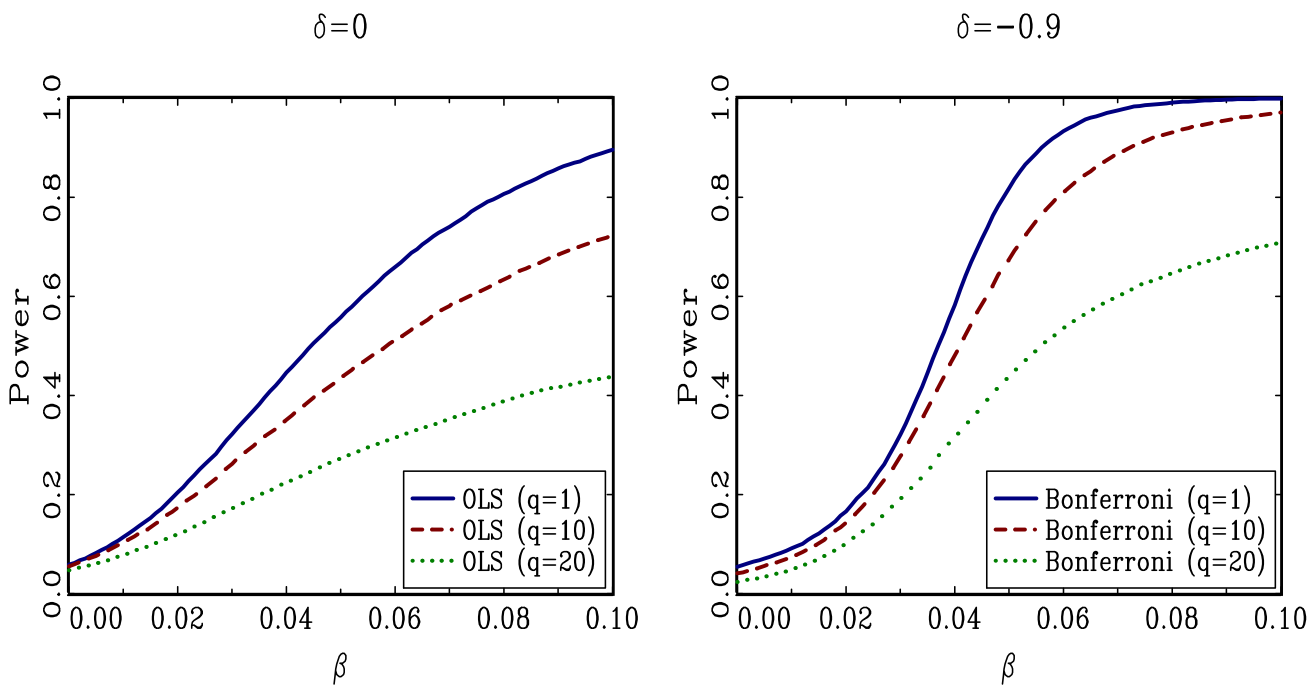

Figure 5: Comparison of power across horizons for $T = 100$ and $c = 0$. The graphs show the average rejection rates for a one-sided 5 percent test of the null hypothesis of $\beta = 0$ against a positive alternative. The $x-$axis shows the true value of the parameter β, and the $y-$axis indicates the average rejection rate. The left hand graph gives the results for the case of exogenous regressors with $\delta = 0$. The results for the one period OLS $t-$test are given by the solid line ($q = 1$), and the results for the scaled OLS $t-$tests for $q = 10$ and $q = 20$ are given by the the short dashed line and the dotted line, respectively. The right hand graph gives the results for the case of endogenous regressors with $\delta = -0.9$. The results for the one period ($q = 1$) Bonferroni test are given by the solid line, and the results for the scaled Bonferroni tests for $q = 10$ and $q = 20$ are given by the the short dashed line and the dotted line, respectively. 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.

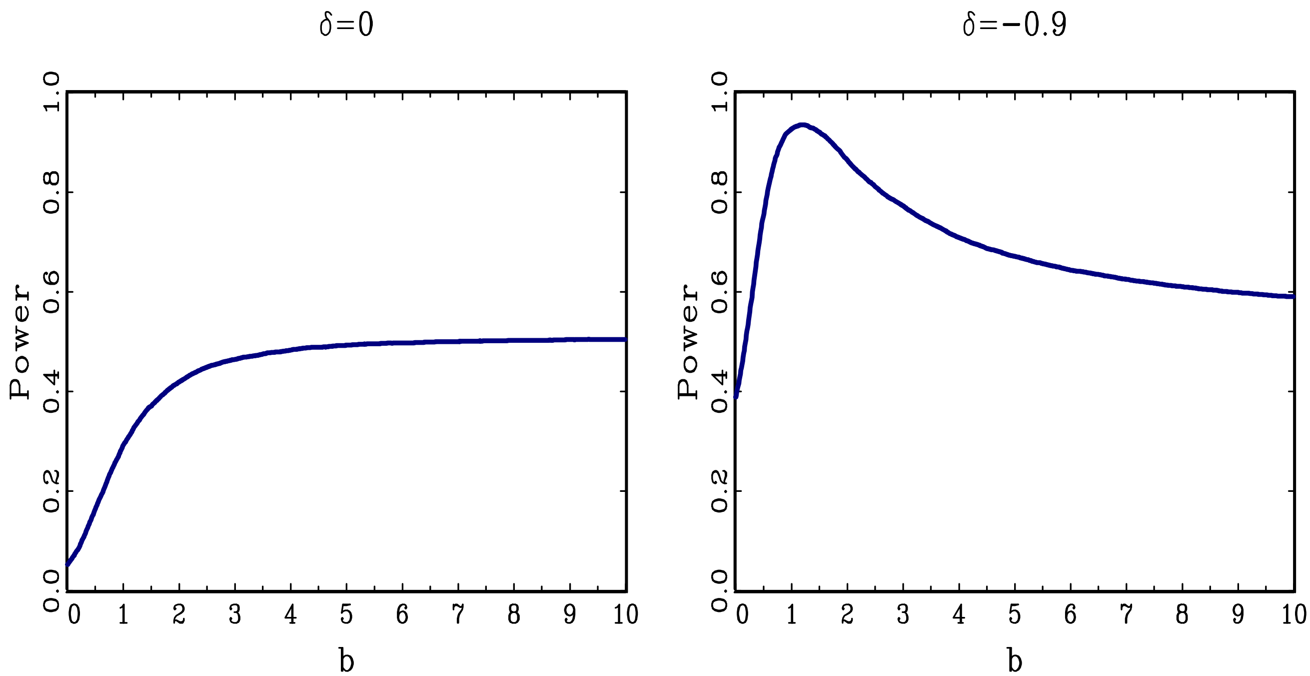

Figure 6: Local power curves for $T/q = 5$ and $c = 0$. The graphs show the average power curves for a one-sided 5 percent test of the null hypothesis against a positive local alternative, based on the distribution of the scaled OLS $t-$statistic derived in Theorem 4. The $x-$axis shows the true value of the parameter $b = \beta q$, and the $y-$axis indicates the average rejection rate. The left-hand graph gives the results for exogenous regressors with $\delta = 0$, and the right-hand graph gives the results for endogeneous regressors with $\delta = -0.9$. The results are obtained from direct simulation of the limiting random variables in equation (18), and the power is calculated as the average rejection rate over $10,000$ repetitions.

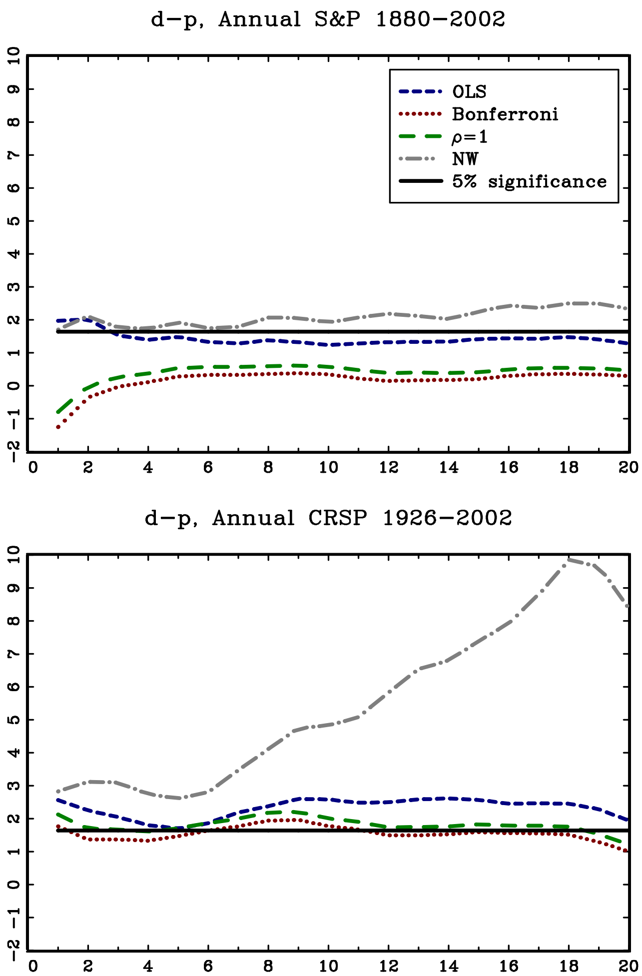

Figure 7: Empirical results for the annual U.S. data with valuation ratios as predictors. The graphs show the outcomes of the long-run test statistics as functions of the forecasting horizon. The $x-$axis shows the forecasting horizon q, and the $y-$axis shows the value of the test statistic. The left-hand graphs give the results for the dividend price ratio $(d-p)$, and the right-hand graphs give the results for the earnings-price ratio $(e-p)$. Results for the S&P 500 data are shown in the top graphs and results for the CRSP data in the bottom graphs. The results for the scaled OLS $t-$test are given by the short dashed lines, the results for the scaled Bonferroni test are given by the dotted lines, the results for the scaled $t-$test from the augmented regression equation under the assumption of $\rho = 1$ are given by the long dashed lines, and the results for the (non-scaled) $t-$test using Newey-West standard errors are given by the dotted and dashed line. The flat solid line shows the 5% significance level, equal to 1.645 based on the normal distribution, for the one sided test.

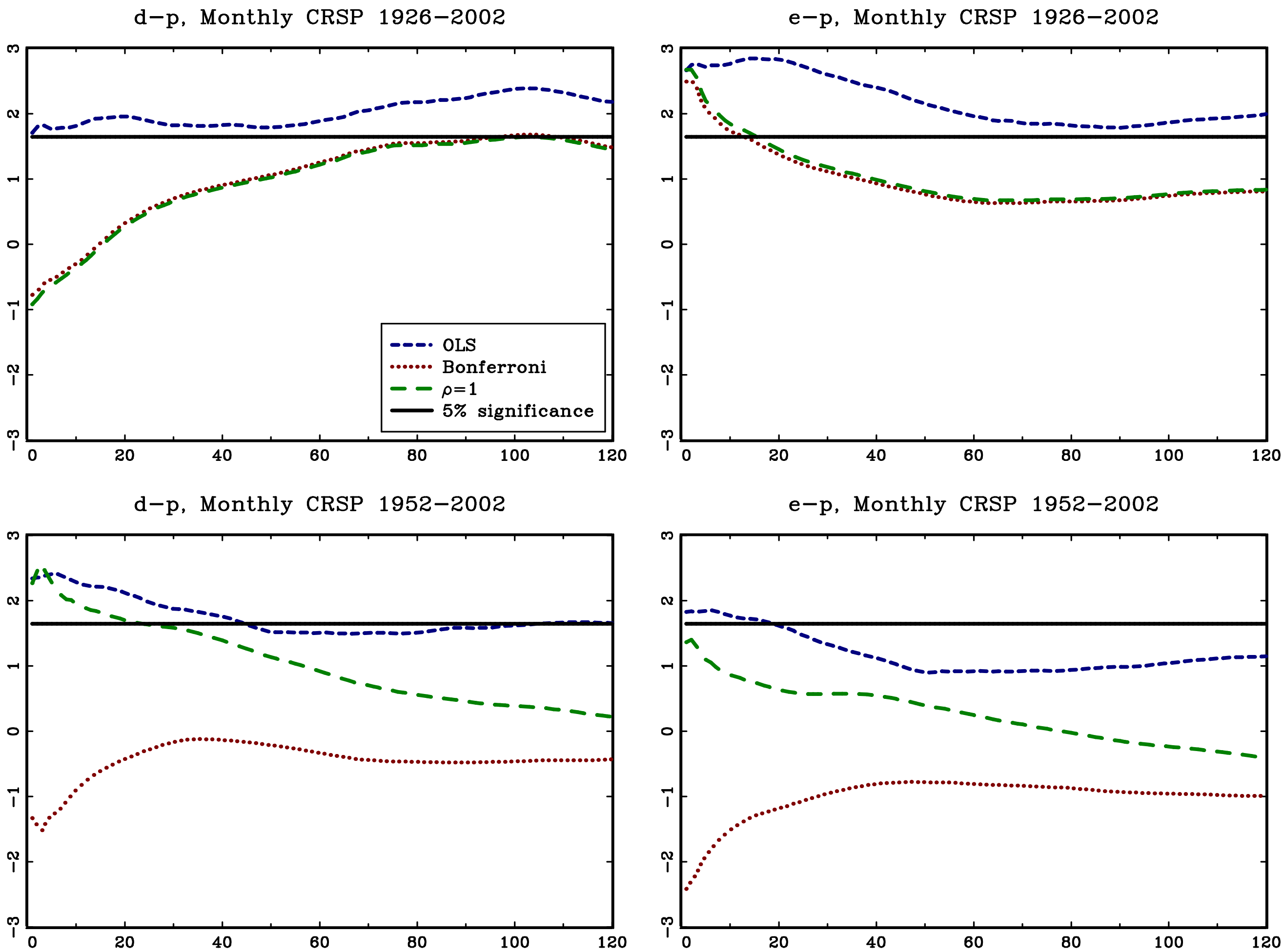

Figure 8: Empirical results for the monthly U.S. data with valuation ratios as predictors. The graphs show the outcomes of the long-run test statistics as functions of the forecasting horizon. The $x-$axis shows the forecasting horizon q, and the $y-$axis shows the value of the test statistic. The left-hand graphs give the results for the dividend price ratio $(d-p)$, and the right-hand graphs give the results for the earnings-price ratio $(e-p)$. Results for the full CRSP sample from 1926-2002 are shown in the top graphs and results for the restricted CRSP sample from 1952-2002 in the bottom graphs. The results for the scaled OLS $t-$test are given by the short dashed lines, the results for the scaled Bonferroni test are given by the dotted lines, and the results for the scaled $t-$test from the augmented regression equation under the assumption of $\rho = 1$ are given by the long dashed lines. The flat solid line shows the 5% significance level, equal to 1.645 based on the normal distribution, for the one sided test.

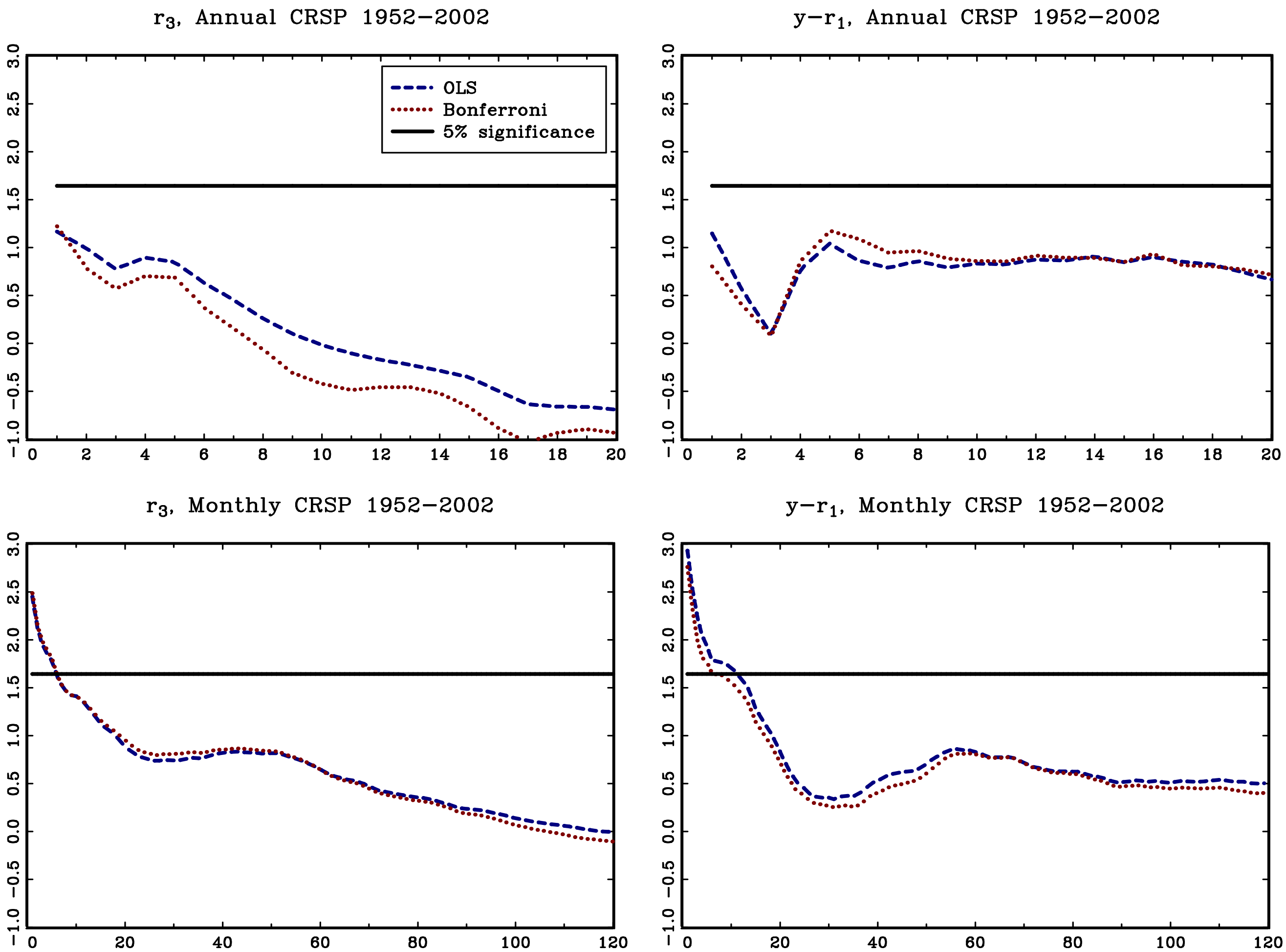

Figure 9: Empirical results for the U.S. data with interest rate variables as predictors. The graphs show the outcomes of the long-run test statistics as functions of the forecasting horizon. The $x-$axis shows the forecasting horizon q, and the $y-$axis shows the value of the test statistic. The left-hand graphs give the results for the short interest rate (r_3), and the right-hand graphs give the results for the yield spread ($y - r_1$). Results for the annual data are shown in the top graphs and results for the monthly data in the bottom graphs. The results for the scaled OLS $t-$test are given by the short dashed lines and the results for the scaled Bonferroni test are given by the dotted lines. The flat solid line shows the 5% significance level, equal to 1.645 based on the normal distribution, for the one sided test.

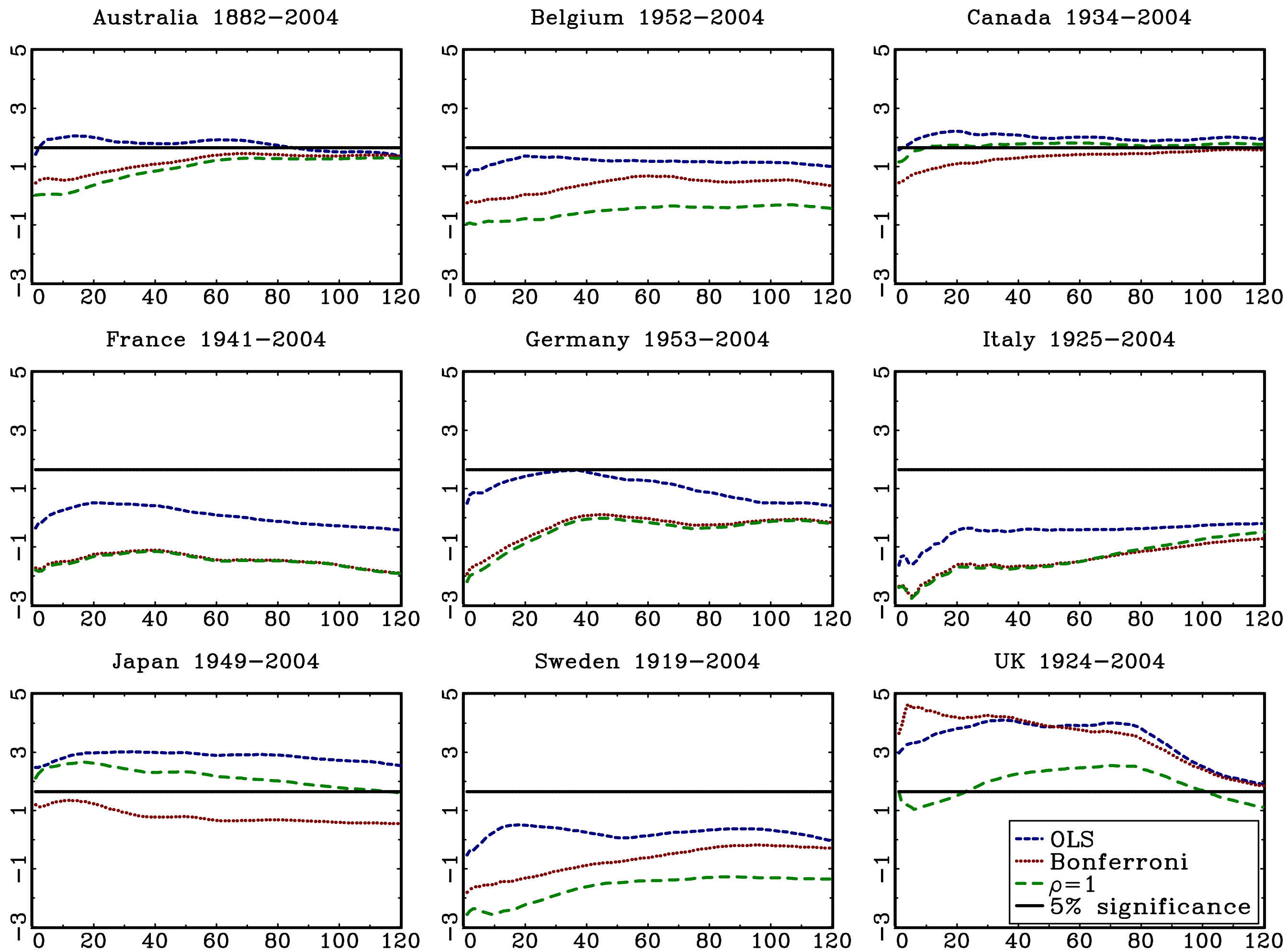

Figure 10: Empirical results for the international data with the dividend-price ratio as predictor, using the full sample for each country. The graphs show the outcomes of the long-run test statistics as functions of the forecasting horizon. The $x-$axis shows the forecasting horizon q, and the $y-$axis shows the value of the test statistic. The title of each graph indicates the country and sample period to which the results correspond. The results for the scaled OLS $t-$test are given by the short dashed lines, the results for the scaled Bonferroni test are given by the dotted lines, and the results for the scaled $t-$test from the augmented regression equation under the assumption of $\rho = 1$ are given by the long dashed lines. The flat solid line shows the 5% significance level, equal to 1.645 based on the normal distribution, for the one sided test.

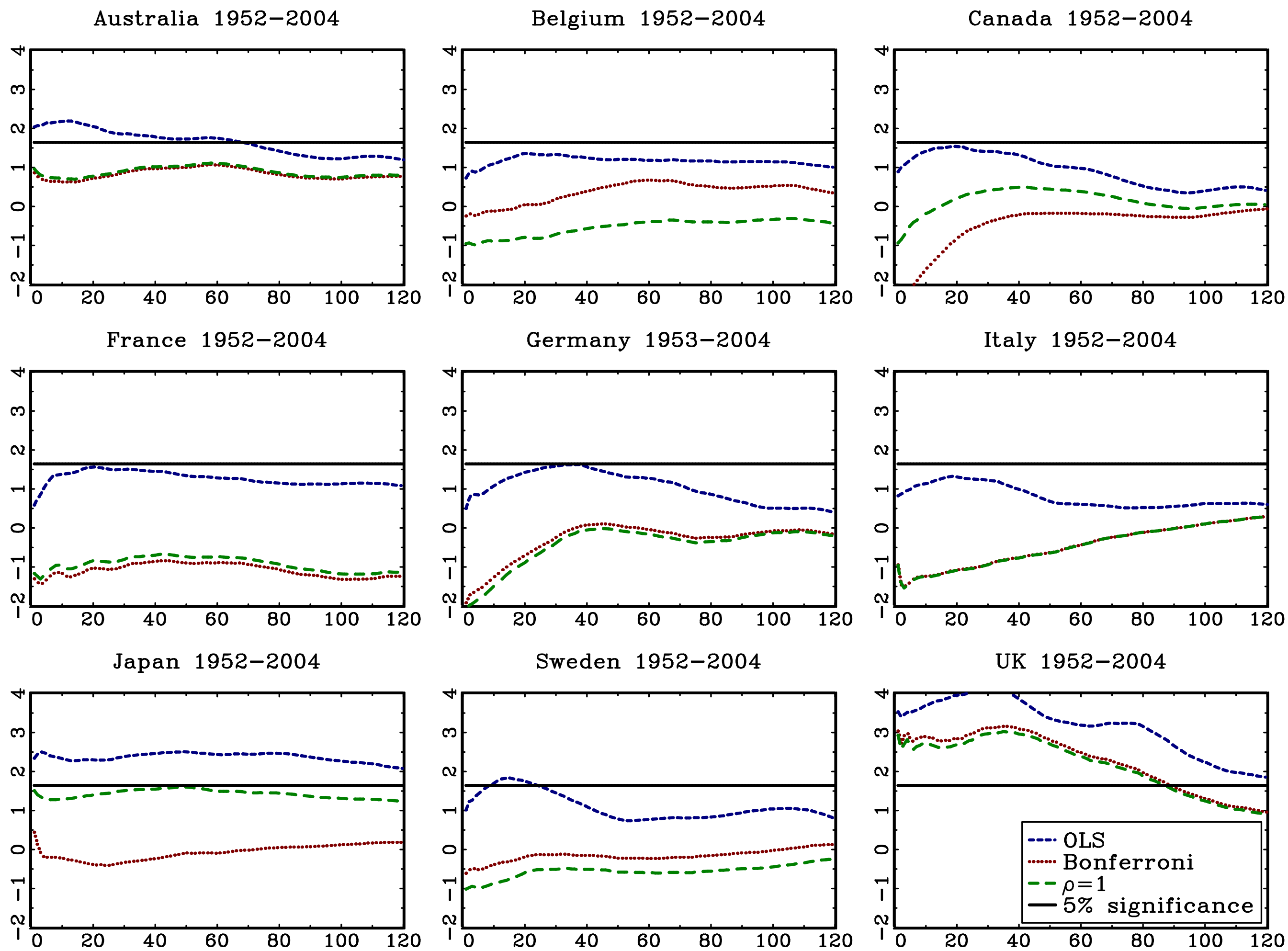

Figure 11: Empirical results for the international data with the dividend-price ratio as predictor, using data after 1952. The graphs show the outcomes of the long-run test statistics as functions of the forecasting horizon. The $x-$axis shows the forecasting horizon q, and the $y-$axis shows the value of the test statistic. The title of each graph indicates the country and sample period to which the results correspond. The results for the scaled OLS $t-$test are given by the short dashed lines, the results for the scaled Bonferroni test are given by the dotted lines, and the results for the scaled $t-$test from the augmented regression equation under the assumption of $\rho = 1$ are given by the long dashed lines. The flat solid line shows the 5% significance level, equal to 1.645 based on the normal distribution, for the one sided test.

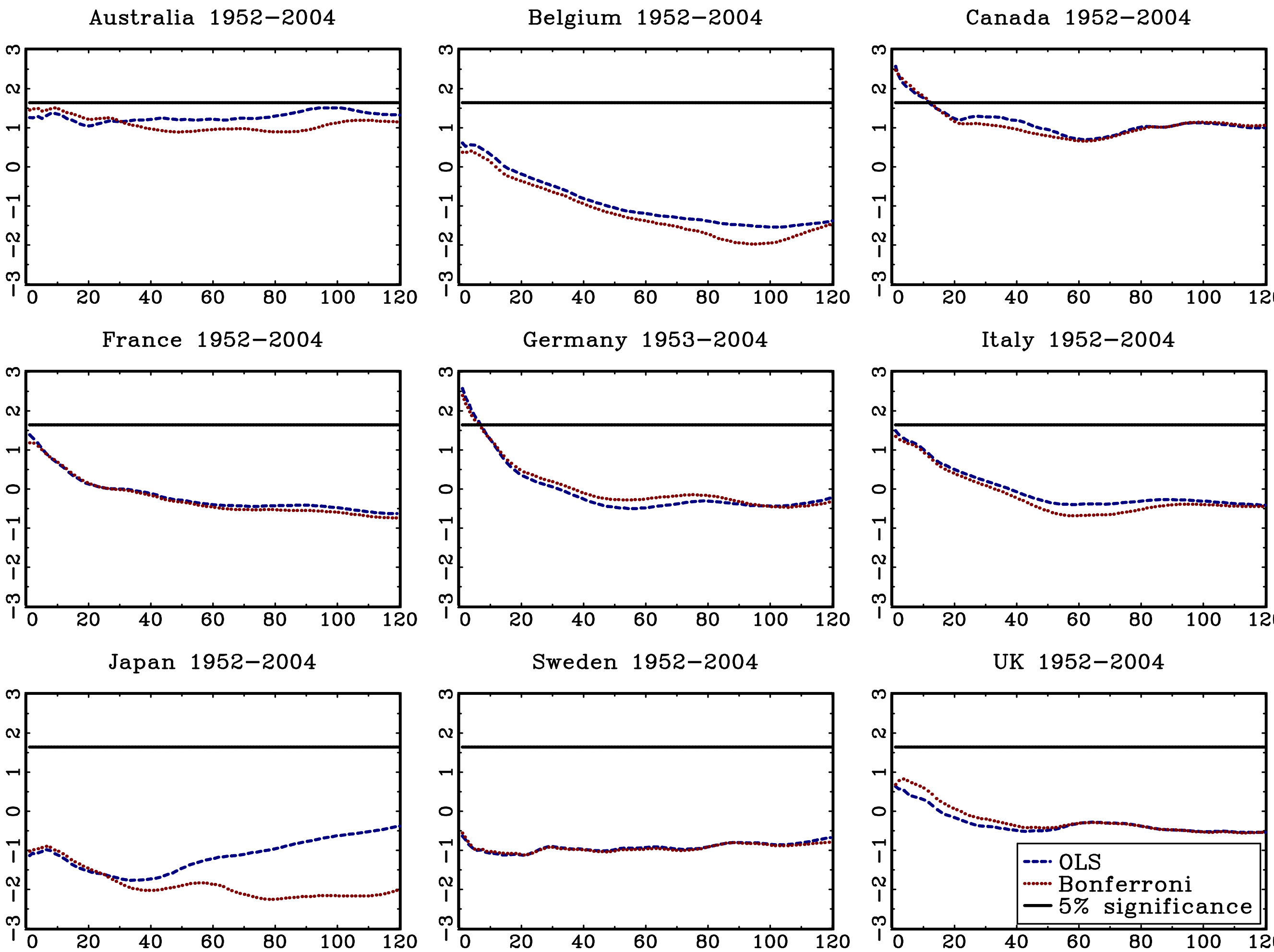

Figure 12: Empirical results for the international data with the short interest rate as predictor. The graphs show the outcomes of the long-run test statistics as functions of the forecasting horizon. The $x-$axis shows the forecasting horizon q, and the $y-$axis shows the value of the test statistic. The title of each graph indicates the country and sample period to which the results correspond. The results for the scaled OLS $t-$test are given by the short dashed lines and the results for the scaled Bonferroni test are given by the dotted lines. The flat solid line shows the 5% significance level, equal to 1.645 based on the normal distribution, for the one sided test.

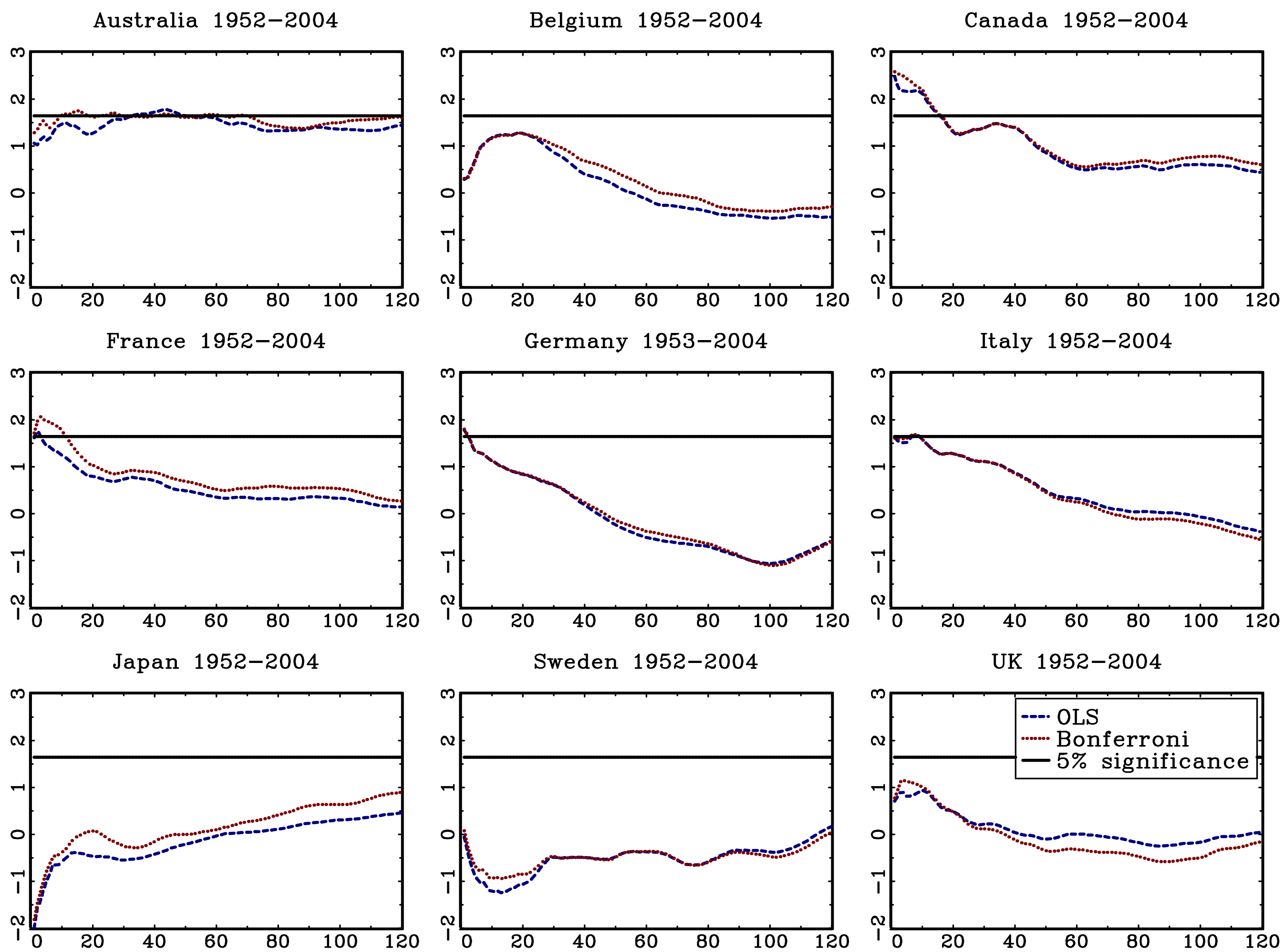

Figure 13: Empirical results for the international data with the term spread as predictor. The graphs show the outcomes of the long-run test statistics as functions of the forecasting horizon. The $x-$axis shows the forecasting horizon q, and the $y-$axis shows the value of the test statistic. The title of each graph indicates the country and sample period to which the results correspond. The results for the scaled OLS $t-$test are given by the short dashed lines and the results for the scaled Bonferroni test are given by the dotted lines. The flat solid line shows the 5% significance level, equal to 1.645 based on the normal distribution, for the one sided test.