Long-run Regressions: Theory and Application to US Asset Markets [1]

Charlotte S. Hansen[2]
Zicklin School of Business
Baruch College, New York, NY

Bjorn E. Tuypens
Oak Hill Platinum Partners, L.L.C.
Rye Brook, NY

October 27, 2004

[1]We would like to thank Tim Bollerslev, Clive Granger, Ayman Hindy, Chi-fu Huang, Stefan Krieger, Peter Phillips, Nagpurnanand Prabhala, Robert Shiller, Rossen Valkanov, and Jiang Wang as well as seminar participants at Yale, Notre Dame, Baruch College, Oak Hill Platinum Partners, Norwegian School of Management, Copenhagen Business School, Aarhus Econometrics meeting 2004, the CEPR summer meeting in Gerzensee 2004, the FMA meetings in Zurich and in New Orleans 2004, and the Econometric Society's European summer meeting in Madrid 2004 for valuable comments. We also benefited from discussions with Carol Boyer, Erik Hjalmarsson, George Korniotis, Gary Sanger, Makram Talih, Sergio Turner, and Avi Wohl. The views, thoughts and opinions expressed in this paper are those of the authors in their individual capacity and should not in any way be attributed to Oak Hill Platinum Partners or to Bjorn Tuypens as a representative, officer, or employee of Oak Hill Platinum Partners. This paper previously circulated under the title "Predicting the S&P 500 - Simple Efficient Methods."

[2]Corresponding author: Charlotte S. Hansen, Mailing address: Department of Economics and Finance, Zicklin School of Business, Baruch College, One Baruch Way, Box B10-225, New York, NY 10010, Tel: (646) 312-3519, Fax: (646) 312-3451, Email: charlotte_hansen@baruch.cuny.edu.

Abstract

The question of long-run predictability in the aggregate US stock market is still unsettled. This is due to the lack of a robust method to judge the statistical significance of long-run regressions under the maintained hypothesis. By developing a spectral theory of long-run regressions with both long-run dependent and independent variables, we demonstrate a version of Engle's (1974) conjecture that asymptotically correct standard errors can be computed by multiplying the ordinary least squares standard errors by the square root of 2/3 times the length of the forecast horizon. We generalize Stambaugh's (1999) bias formula to the long-run regression model proposed in this paper. In addition, we find, that for persistent predictive variables, the OLS estimator in our regression model is more efficient than the estimator in the predictive regressions suggested by Campbell and Shiller (1988) and Hodrick (1992). Application of our method shows that the long-run earnings yield significantly predicts up to 69% of the variation in the 10-year S&P 500 real return, and up to 49% of long-run bond returns.

The question of whether the aggregate US stock market is predictable is still unsettled. On the one hand, Campbell and Shiller (1998; 2001) and Campbell and Yogo (2003) argue that dividend yields predict aggregate stock returns, especially in the long run. On the other hand, recent studies by Wolf (2000), Lanne (2002), Goyal and Welch (2003), Valkanov (2003), Ferson, Sarkissian and Simin (2003), and Torous, Valkanov and Yan (2005) suggest that dividend yields have hardly any forecasting power for returns, and certainly not in the long run. Other (macro)economic variables have been suggested as well, such as price earnings ratios (Lamont (1998) and Campbell and Shiller (1988*a*)), book-to-market ratios (Pontif and Schall (1998)), or the ratio of stock market wealth to aggregate wealth (Lettau and Ludvigson (2001)), with various degrees of success, where success most often is measured by significant Newey-West corrected standard errors.

The absence of consensus on the question of long-run forecastability is due to the lack of a satisfactory econometric theory for dealing with regressions of long-run variables. Indeed, to analyze the long-run behavior of (stationary) time series, and especially to test the significance of their comovement, the most convenient tool available to the researcher is the Hansen and Hodrick (1980) approach, which is the basis for the Newey and West (1987) correction for autocorrelation in the error term.

However, the Hansen and Hodrick approach is often criticized on the grounds of its poor small sample performance, with severe size distortions when analyzing long-run returns (cf. Hodrick (1992) and Ang and Bekaert (2001)). The reason is that the Hansen and Hodrick approach requires the calculation of the long-run covariance matrix of the estimator. This is particularly hard when there is strong dependence in the error term, as in the case of overlapping data, combined with the typical small sample sizes characterizing macroeconomic data sets. To overcome these small sample problems, the bootstrap (Goetzmann and Jorion (1993)), the vector autoregression (VAR) approach (Campbell and Shiller (1988*b*), Hodrick (1992), Nelson and Kim (1993)), and the Richardson and Stock (1989) approach have been proposed. Politis, Romano and Wolf (1999, chapter 13) criticize the bootstrap and VAR approaches for being heavily dependent on the underlying model used to describe the data generating processes and propose using subsampling instead. However, when facing long-run regressions, they admit that their method faces some undercoverage of subsampling intervals at long horizons, due to the very strong correlation of the residuals in the regression.

The Richardson and Stock (1989) approach and subsequent approaches along the lines of Valkanov (2003) and Torous et al. (2005) are valid methods under the null hypothesis of no predictability, but provide no general theory to do inference under the alternative hypothesis of predictable stock returns. In any case, any method to conduct inference in long-run regressions faces at least two challenges: to adequately capture the dependence structure of the data, both under the null and the alternative, and secondly, whenever the forecast horizon is larger than one, to deal with the increased variability of the estimators.[1]

The objectives of the present paper are threefold. Firstly, we develop an econometric theory to analyze the regression of long-run returns on long-run predictive variables, under the maintained hypothesis, i.e. which covers both the hypothesis of predictable returns as well as the case where returns are unpredictable. In particular, we show that it suffices to multiply the standard errors obtained from the ordinary least squares regression by $\sqrt{2/3q}$ to obtain asymptotically correct standard errors, both under the null and the alternative hypotheses, where q is the overlap in the data, or equivalently, the forecast horizon considered. Accordingly, our procedure is very robust to small sample problems as it does not involve estimation of an additional quantity, viz. the (long-run) variance-covariance matrix of the estimators as in the Newey-West estimator. This is confirmed in our Monte-Carlo study. Inference for the long-run models proposed by Campbell and Shiller (1988*a*) and Hodrick (1992), which involves regressing the long-run (short-run) return on a vector of short-run (long-run) predictive variables, can also be simplified considerably.

Secondly, we analyze the asymptotic efficiency of the different regression models and the potential small sample bias of the regression coefficient estimator for the regression model proposed in this paper. We derive a sufficient condition under which a regression of a long-run variable on a single period predictive variable is less efficient than a regression involving long-run variables on both sides of the equation. The sufficient condition states that the explanatory variable needs to exhibit enough variation in the long-run relative to the variation in the overall time series; this condition can be tested using a variance ratio test. For the small sample bias analysis we follow Stambaugh (1999). We show that the bias in the long-run regression is smaller in relative terms than that in Stambaugh's predictive regressions and more so for persistent predictive variables. In addition, we show how to impose the null hypothesis when constructing the tests. These tests

[1]See Andrews (2004) for an improved block bootstrap method that addresses both of these issues.

involve scaling down the t-test by $\sqrt{1-R^2}$. Such a test is more conservative than tests derived under the maintained hypothesis.

The third objective of this paper consists of applying this methodology to forecast movements in US asset markets. It is shown that for the S&P 500, the long-run component of the earnings price ratio predicts up to 69% of the variation in 10-year real returns in the post-World War II period, and around 45% over the full sample 1871-2003, thereby providing strong evidence that the US stock market has been extremely predictable over its recent history. For an immediate overview of the results, we refer to Figure 3 to illustrate the performance of the forecasting model. If this historical evidence is of any guidance for the future of the stock market, the outlook is rather grim, since the model predicts that by 2013, the stock market should be at about 67% of its 2003 level of in real terms. Furthermore, our empirical findings suggest that the long-run earnings price ratio is a better predictor of long-run stock returns than the dividend yield.

We also apply our method to forecasting US investment grade corporate bond returns and returns on US sovereign bonds. Our results show that both bond markets exhibit long-run predictability, although this evidence is not as strong as for the equity market. This seems to suggest that the predictability of long-run stock returns is not just a consequence of time variation in the equity premium.

The remainder of this paper is organized as follows. Section 1 introduces the type of forecasting regressions considered in this paper and Section 2 develops the appropriate econometric theory to analyze this type of long-run forecasting model. The analysis is contained in Sections 3 through 5. Section 6 implements these methods to forecast the US stock and bond market returns. Section 7 presents the conclusions. A Monte-Carlo study, technical lemmas and proofs are contained in the appendix.

1 Forecasting Regressions

The type of forecasting regressions we analyze in this paper are as follows

$$r_{t+q}(q) = x_t'(q)\beta + u_{t+q}(q), \tag{1}$$

$$r_{t+q}(q) = x_t'\beta + v_{t+q}, \tag{2}$$

$$r_{t+1} = x_t'(q)\beta + v_{t+1}, \tag{2*}$$

where $r_t(q) = \sum_{j=1}^{q} r_{t-q+j}$, and $r_t(1) = r_t$ denotes the single period log return. Similarly, $x_t(q) = \sum_{j=1}^{q} x_{t-q+j}$, where x_t is a vector of $m \geq 1$ forecasting variables such as the log dividend yield or the log earnings yield.[2] A constant in the regression can be handled by using demeaned variables. Model (1) was first analyzed in the time domain by Hansen and Hodrick (1980), and Fama and French (1989) implemented this regression to test for mean reversion in stock returns. However, Model (2) is currently the most widely used model to forecast long-run returns based on financial ratios or other macroeconomic variables that have been suggested as predictive variables. We also analyze Model (2*), in which the next period's return is predicted from a long-run variable. This model was suggested by Hodrick (1992) and Bollerslev and Hodrick (1996) as a way to avoid the large degree of overlap in the data and thus autocorrelation in the error term.

We will argue that Model (1) has a very natural interpretation in the frequency domain. In effect, the moving sum on both sides of Model (1) transforms both sides of the equation into long horizon variables, so that it effectively defines a long-run relation between the (single period) variables. This results in a balanced regression model, with, under certain conditions, better efficiency properties than Models (2) and (2*). The right hand side of Model (2) is a long-run return, while the right hand side involves a vector of short-run variables, such as the one-period dividend yield and the one-period earnings yield. In this respect, Model (2) could be considered unbalanced, or misspecified.[3,4,5]

Finally, it should be noted that the long-run component of the predictive variable, as measured by $x_t(q)$, will, in most cases, have a natural interpretation in terms of past average yields. As

[2]We will always work with natural logarithmic transformations of both dependent and independent variables, and therefore will omit the log qualifier in the remainder of the paper.

[3]Interestingly, Friedman and Schwartz (1982, p.358) argue that the longer the forecasting horizon, the longer the time span should be over which anticipations are formed. This gives an intuitive appeal to Model (1), not shared by Models (2) and (2*).

[4]To understand the difference between a balanced and an unbalanced regression, it is useful to compare it with the theory of non-stationary time-series, where a balanced regression model necessarily involves a non-stationary time series on both sides of the equation.

[5]We take a moving average of the earnings yield, as opposed to taking a moving average of the earnings and then dividing by the price. From Modigliani and Miller (1958) we know that the earnings yield serves as a proxy for expected return under certain conditions. Moreover, we show in Hansen and Tuypens (2004) that a moving average of the earnings divided by the most recent price is a downward biased proxy for expected returns.

a simple example, when trying to forecast long-run expected returns based on its hypothesized property of mean reversion, it is natural to use a past long-run average return, as in Fama and French (1989). When considering predictive variables such as earnings or dividend yields, the averaging of the logarithm of 1 plus the yield corresponds in the same way to calculating the past geometric average yield.

Regression equations (1) and (2) give rise to autocorrelation in the error terms stemming from the use of overlapping data. In the presence of autocorrelation, General Least Squares (GLS) is the standard technique to conduct inference. However, GLS requires a strict econometric exogeneity condition $\mathbb{E}[u(q)_{t+q}|x(q)_t, x(q)_{t-1}, x(q)_{t+1}, ..] = 0$, which means that the forecast error is independent of all the data that we have available. This is not appropriate when considering a forecasting regression: while past information should be uncorrelated with the current forecast error if the forecast is efficient, this is not true for future information (Hansen and Hodrick (1980)).

Hansen and Hodrick (1980) estimate the long-run covariance matrix of the explanatory variable and the error term, and use a variant of the usual sandwich formula as the covariance matrix of the regression coefficient vector, which for Model (1) corresponds to

$$V_{HH} = \mathbb{E}\left(x_t(q)x_t(q)'\right)^{-1} \left(\sum_{j=-q+1}^{q-1} \mathbb{E}\left(u_{t+q}(q)u_{t+q+j}(q)'\right) \mathbb{E}\left(x_t(q)x_{t+j}(q)'\right) \right) \mathbb{E}\left(x_t(q)x_t(q)'\right)^{-1}.$$

Unfortunately, estimating the long-run variance of the explanatory variable and the error term gives rise to small sample problems when the autocorrelation is large as in the case of overlapping data, combined with a small sample size (Hodrick (1992)). Another drawback of this approach is that the covariance matrix is not guaranteed to be positive semi-definite. To avoid the latter problem, Newey and West (1987) suggest a slightly modified estimation that ensures that the covariance matrix is always positive semi-definite.

A more direct way to do inference for regressions involving long-run variables is to use frequency domain techniques. Indeed, a moving sum of the variable corresponds in effect to its long-run or zero-frequency component.

2 Setup and Assumptions

To set up our model it is helpful to consider the regression of the one period return, r_{t+q}, on a vector of explanatory variables, x_t, lagged q periods:

$$r_{t+q} = x_t'\beta + u_{t+q}. \tag{3}$$

Note that this model encompasses the hypothesis of unpredictable returns if $\beta = 0$ and u_t is white noise. The parameter $q \geq 1$ determines the forecast horizon; for $q = 1$ the traditional predictive model as in Stambaugh (1999) is obtained.

To analyze model (3) in the spectral domain would require an exogeneity condition, similar to that when applying GLS: $\mathbb{E}[x_{t-j}u_{t+q}] = 0$, for all j, or for all leads and lags (see e.g., Corbae, Ouliaris and Phillips (2002) for a recent reference). To circumvent this exogeneity condition we suggest using a block regression model, similar to that used in Phillips, Moon and Xiao (2001), where the time series of returns and predictive variables are divided into blocks of equal length q. We introduce a new notation where the single period return $r_{(k+1)\cdot q+\tau}$ is written as r_τ^{k+1}, which is observation τ in block $k+1$. Similarly, we let the predictive variable, $x_{k\cdot q+\tau}$, be written as x_τ^k to emphasize that it is observation τ in the previous block, k. Using this new notation, equation (3) can be rewritten as

$$r_\tau^{k+1} = x_\tau^{k\prime}\beta + u_\tau^{k+1}, \tag{4}$$

$$r_0^k = r_q^{k-1}, \tag{5}$$

$$x_0^k = x_q^{k-1}, \tag{6}$$

where $\tau = \{1, ..., q\}$ is the time index within each block $k = \{1, 2, ..., M\}$. The block model (4) - (6) has a natural interpretation in terms of information. To forecast the return in block $k+1$, only information contained in block k shall be used.

[Include Figure 1 about here]

This system defines a sequence of M blocks with q observations $\{\{r_\tau^k, x_\tau^k\} : \tau = 1, ..., q\}$ in each block $k = 1, ..., M$. The initial conditions, (5) and (6), in each block are set so that they correspond to the last observation in the previous block. In this sense, the model is articulated to capture the evolution of a single multi-variate time series. The observable series is then $\left\{\left\{r_\tau^k, x_\tau^k\right\} : \tau = 1, ..., q; k = 1, ..., M\right\}$.

Furthermore, let x_τ^{k-1} be pre-determined; i.e., x_τ^{k-1} is uncorrelated with the error term in block k, or more generally, $\mathbb{E}\left[x_\tau^{k-j} u_\tau^k\right] = 0$, for all $j > 0$. The justification for this assumption is as in Hansen and Hodrick (1980): assuming that the market is informationally efficient, $x_\tau^{k-1\prime}\beta$ provides an optimal forecast of the return in block k, with a forecast error that is necessarily uncorrelated with the forecasting variables. In addition, we assume that the one-period variables $\left\{r_\tau^k, x_\tau^k\right\}$ in each block are jointly linear processes, with *iid* innovations and finite fourth moments. This means that the series are jointly covariance stationary and therefore have a representation in the frequency domain (cf. Hamilton (1994, p. 165)).

For estimation purposes, we will use overlapping blocks to improve the efficiency of our estimators. In terms of model (4)-(6), this implies that we take an average over q block structures of length q, denoted by $\left(M^1, M^2, ..., M^q\right)$, where M^1 is the block structure starting with the first observation, M^2 is the block structure starting in the second observation, etc. Figure 1 visualizes these block structures. These block structures result in $T = T^* - 2q + 1$ overlapping blocks, where T^* is the number of one-period return observations. For the asymptotic results to hold we will assume that $q/T \to 0$ as $q \to \infty$ and $T \to \infty$.

Before we proceed, we will summarize the assumptions so that we can refer to them throughout the paper:

Assumption A

A1: Single period returns, r_t, are described by model (3).

A2: x_τ^{k-1} is pre-determined; i.e., x_τ^{k-1} is uncorrelated with the error term in block k, or more generally, $\mathbb{E}\left[x_\tau^{k-j} u_\tau^k\right] = 0$, for all $j > 0$.

A3: $\left\{r_\tau^k, x_\tau^k\right\}$ in each block are jointly linear processes, with *iid* innovations and finite fourth moments.

A4: $q/T \to 0$ as $q \to \infty$ and $T \to \infty$.

To translate the model (4)-(6) into the frequency domain, we first write (4) in vector form as follows:

$$r^{k+1} = x^k \beta + u^{k+1}, \tag{7}$$

where $r^k = (r_1^k, r_2^k, ..., r_q^k)'$ and $x^k = (x_1^{k'}, x_2^{k'}, ..., x_q^{k'})'$. We then premultiply (7) by the q dimensional column vector U_q, with the j'th entry given by $\left(e^{2\pi jki/q}/\sqrt{2\pi q}\right)$, where $i = \sqrt{-1}$. This gives us the discrete Fourier transform (DFT) of (4):

$$w_{r^{k+1}}(\lambda_s) = w_{x^k}(\lambda_s)\beta + w_{u^{k+1}}(\lambda_s), \ \lambda_s = 2\pi s/q, \ s = -[q/2]+1, ..., [q/2], \tag{8}$$

where

$$w_x(\lambda_s) = \frac{1}{\sqrt{2\pi q}} \sum_{\tau=1}^{q} x_\tau e^{i\tau\lambda_s}$$

is the DFT of the time series x at the fundamental frequency λ_s, and $[\cdot]$ denotes the greatest lesser integer. The OLS estimator of β in the frequency domain can be constructed as follows:

$$\hat{\beta} = \left(\sum_{s=-[q/2]+1}^{[q/2]} |w_{x^k}(\lambda_s)|^2 \right)^{-1} \sum_{s=-[q/2]+1}^{[q/2]} w_{x^k}(\lambda_s) w_{r^{k+1}}(\lambda_s)^*,$$

where w^* denotes the complex conjugate of w, and $|w_x|^2 = w_x w_x^*$. As was suggested by Hannan (1970), the frequency domain regression technique is particularly suited when the relation (8) holds only for a particular frequency, say ω, resulting in

$$w_{r^{k+1}}(\omega) = w_{x^k}(\omega)\beta_\omega + w_{u^{k+1}}(\omega),$$

with the slope coefficient defined as follows:

$$\beta_\omega = f_{x^k x^k}(\omega)^{-1} f_{x^k r^{k+1}}(\omega).$$

Here, $f_{x^k x^k}(\omega)$ is the spectrum at frequency ω of x^k, and $f_{x^k r^{k+1}}(\omega)$ denotes the cross spectrum between x^k and r^{k+1}. Within our block model, $f_{x^k x^k}(\omega)$ and $f_{x^k r^{k+1}}(\omega)$ are defined as follows:

$$f_{x^k x^k}(\omega) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \mathbb{E}\left[x_\tau^k x_{\tau+j}^{k\prime}\right] e^{-i\omega j} \tag{9}$$

$$\begin{aligned} f_{x^k r^{k+1}}(\omega) &= \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \mathbb{E}\left[x_\tau^k r_{\tau+j}^{k+1}\right] e^{-i\omega j} \\ &= \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \mathbb{E}\left[x_\tau^k r_{\tau+q+j}^{k}\right] e^{-i\omega j}. \end{aligned} \tag{10}$$

Similarly, we define

$$f_{u^k u^k}(\omega) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \mathbb{E}\left[u_\tau^k u_{\tau+j}^{k\prime}\right] e^{-i\omega j} \tag{11}$$

to be the block definition of the spectrum at frequency ω of the error term. From our linearity assumption, it follows that x_t and r_t are covariance stationary; therefore $f_{x^k x^k}(\omega)$, $f_{x^k r^{k+1}}(\omega)$ and $f_{u^k u^k}(\omega)$ are identical across all blocks, allowing us to write $f_{xx}(\omega)$, $f_{xr}(\omega)$ and $f_{uu}(\omega)$ for simplicity.

3 Long-run Regressions Using Ordinary Least Squares

To focus on the long-run only, we will restrict the estimation to the zero frequency, i.e., $\omega = 0$. For notational simplicity, we will write $f_\bullet$ for $f_\bullet(0)$, and β for β_0. We estimate β by estimating f_{xx} and f_{xr} across all sub-samples and average all sub-sample estimators to achieve consistency.[6] This corresponds to applying ordinary least squares (OLS) to estimate Model (1), resulting in the following estimator:

$$\begin{aligned} \hat{\beta}_{OLS} &= \left(\frac{1}{T} \sum_{t=1}^{T} \left(\sum_{j=1}^{q} x_{t+j} \sum_{j=1}^{q} x'_{t+j} \right) \right)^{-1} \frac{1}{T} \sum_{t=1}^{T} \left(\sum_{j=1}^{q} x_{t+j} \sum_{j=1}^{q} r_{t+q+j} \right) \\ &= \widetilde{f}_{xx}^{-1} \widetilde{f}_{xr}, \end{aligned} \tag{12}$$

where

$$\begin{aligned} \widetilde{f}_{xx} &= \frac{1}{2\pi q} \frac{1}{T} \sum_{t=1}^{T} \left(\sum_{j=1}^{q} x_{t+j} \sum_{j=1}^{q} x'_{t+j} \right) \\ \widetilde{f}_{xr} &= \frac{1}{2\pi q} \frac{1}{T} \sum_{t=1}^{T} \left(\sum_{j=1}^{q} x_{t+j} \sum_{j=1}^{q} r_{t+q+j} \right). \end{aligned}$$

The intuition behind this approach is straightforward. Every block regression provides an (asymptotically) unbiased, albeit inconsistent, estimator of f_{xx} and f_{xr}. The unbiasedness stems from the assumption that the (summed) error term is uncorrelated with the (summed) forecasting variable. The inconsistency is due to the fact that for each block, we estimate a variance and a covariance from only one observation. It is by averaging all these individual covariance and variance estimates that consistency is achieved.

An asymptotically equivalent approach is to estimate β using a Bartlett kernel estimator which

[6] The idea of averaging across sub-samples goes back to Bartlett (1950).

is defined by

$$\hat{\beta}_{Bartlett} = \left(\sum_{h=-q+1}^{q-1} \left(1 - \frac{|h|}{q}\right) \hat{\Gamma}_{xx}(h) \right)^{-1} \sum_{h=-q+1}^{q-1} \left(1 - \frac{|h|}{q}\right) \hat{\Gamma}_{xr}(q+h),$$

where the covariances are given by

$$\hat{\Gamma}_{xx}(h) = \frac{1}{T} \sum_{t=1}^{T} x_t x'_{t+h} \tag{13}$$

$$\hat{\Gamma}_{xr}(q+h) = \frac{1}{T} \sum_{t=1}^{T} x_t r_{t+h+q} \tag{14}$$

and the term $\left(1 - \frac{|h|}{q}\right)$ is called the Bartlett kernel or the triangular kernel. The slope coefficient in equation (1) can be estimated using either OLS or a Bartlett estimator; the methods are asymptotically equivalent (see Appendix B for a proof of the equivalence between the OLS and the Bartlett estimators). This is the basis for deriving the limiting distribution of the coefficient estimate of β, which is described in the following proposition. The proof of this proposition also appears in Appendix B.

Proposition 1 *Under Assumption A, the slope coefficient in Model* (1) *is asymptotically distributed as follows*

$$\sqrt{\frac{T}{q}} \left(\hat{\beta} - \beta \right) \xrightarrow{d} N\left(0, (2/3)\, f_{xx}^{-1} f_{uu}\right), \tag{15}$$

where f_{xx} and f_{uu} are defined in equations (9) *and* (11)*, respectively.*

Let us compare this result to the standard OLS theory in the time domain. Eq. (15) looks very similar to the standard OLS, where the scaled coefficient estimate converges in distribution to a normal distribution centered in the true value of the coefficient. Standard OLS theory would suggest a convergence rate of $\sqrt{T}$; we see that the actual convergence rate is only $\sqrt{T/q}$ under the assumption that $q/T \to 0$ as $q \to \infty, T \to \infty$. This is a consequence of using long-run returns. The long-run returns are constructed as sums of the q most recent short-run returns. Hence, long-run returns $r_t(q) = r_{t-q+1} + ... + r_t$ and $r_{t+s}(q) = r_{t+s-q+1} + ... + r_{t+s}$ with less than q periods apart will share $q - s$ short-run returns and will, per construction, be correlated. Accordingly, only T/q long-run returns are non-overlapping and thus independent. This results in a loss of degrees of freedom when analyzing the long-run.

We proceed by taking a closer look at the variance expression and how to estimate it, which is important when doing inference.

3.1 Practical Inference

Proposition 1 gives us an explicit expression for the variance of the coefficient estimate and we may use this variance to form a t-test for the significance of the coefficient estimate. In this section we show that the variance of the estimator is closely linked to that obtained by OLS. We use the asymptotic theory developed above to analyze tests based on $\hat{\beta}$ with the regular standard errors, i.e., not corrected for autocorrelation. Consider the residuals from Model (1),

$$\widehat{u}_{t+q}(q) = r_{t+q}\left(q\right) - x_t\left(q\right)'\hat{\beta}.$$

Assume, furthermore, that T long-run variables are used. Then the traditional estimate of the variance of $\sqrt{T}\left(\hat{\beta}-\beta\right)$ is equal to

$$\begin{aligned}
& \left(\frac{1}{T}\sum_{t=1}^{T} x_t\left(q\right) x_t\left(q\right)'\right)^{-1} \frac{1}{T}\sum_{t=1}^{T} \widehat{u}_{t+q}(q)^2 \\
= \; & \left(\frac{1}{T}\sum_{t=1}^{T} x_t\left(q\right) x_t\left(q\right)'\right)^{-1} \frac{1}{T}\sum_{t=1}^{T} \left(r_{t+q}\left(q\right) - x_t\left(q\right)'\hat{\beta}\right)^2 \\
= \; & \left(\frac{1}{T}\sum_{t=1}^{T} x_t\left(q\right) x_t\left(q\right)'\right)^{-1} \frac{1}{T}\sum_{t=1}^{T} \left(\sum_{j=1}^{q} r_{t+j} - \sum_{j=1}^{q} x'_{t-q+j}\hat{\beta}\right)^2 \\
= \; & \left(\frac{1}{T}\sum_{t=1}^{T} x_t\left(q\right) x_t\left(q\right)'\right)^{-1} \frac{1}{T}\sum_{t=1}^{T} \left(\sum_{j=1}^{q} \left(r_{t+j} - x'_{t-q+j}\hat{\beta}\right)\right)^2 \\
= \; & \left(\frac{1}{T}\sum_{t=1}^{T} x_t\left(q\right) x_t\left(q\right)'\right)^{-1} \frac{1}{T}\sum_{t=1}^{T} \left(\sum_{j=1}^{q} \hat{u}_{t+j}\right)^2 \\
& \xrightarrow{p} f_{xx}^{-1} f_{uu},
\end{aligned}$$

where the last line follows since $\hat{\beta} \xrightarrow{p} \beta$, and from our assumption of stationarity (A3). As mentioned above, the latter assumption results in an identical spectrum across blocks, defined as f_{xx} and f_{uu} in (9) and (11).

From comparing this last line with the asymptotically correct variance in Proposition 1, we see that the OLS variance is biased and needs to be multiplied by $\frac{2}{3}q$. This means that for practical inference, one can obtain asymptotically correct standard errors simply by multiplying the OLS standard errors by $\sqrt{2/3q}$. Equivalently, it suffices to divide the OLS t-statistic by $\sqrt{2/3q}$ in order to judge the significance of the coefficient estimate. We state these results in the following proposition.

Proposition 2 *Under Assumptions A asymptotically correct standard errors of the coefficient estimate in Model* (1) *are computed as*

$$\hat{SE}_{HT} = \hat{SE}_{OLS} \cdot \sqrt{\frac{2}{3}q}, \tag{16}$$

where $\hat{SE}_{OLS}$ denotes the OLS standard errors and q is the length of the forecasting horizon. Asymptotically correct t-statistics are computed as

$$t\text{-}stat_{HT} = \frac{t\text{-}stat_{OLS}}{\sqrt{\frac{2}{3}q}}, \tag{17}$$

where $t\text{-}stat_{OLS}$ denotes the t-statistic based on OLS standard errors and q is the length of the forecasting horizon.

One may ask whether this differs from the standard errors suggested in the previous literature. Engle (1974) conjectured that an approximate correction for overlapping data could be obtained by multiplying the standard errors by $\sqrt{q}$. Our results show that an exact correction involves a further rescaling of the standard errors by $\sqrt{2/3}$. More recently, Valkanov (2003) has proposed a t-statistic rescaled by $1/\sqrt{T}$ and shows that this statistic converges to a non-standard distribution. This result holds under the Richardson and Stock (1989) framework, where it is assumed that the forecasting horizon remains a fixed fraction of the sample size, even in the limit: $q/T \to \delta$ as $q \to \infty$, $T \to \infty$, where δ is a real number between 0 and 1. It therefore follows that the t-statistic can be divided by either by either $\sqrt{T}$ or $\sqrt{q}$; in both cases this rescaled t-statistic converges to a proper limiting distribution.

The Hansen and Hodrick (1980) variance estimator is another often used method to obtain confidence intervals in the presence of autocorrelation in the error terms. In the following section we will analyze this estimator using the current framework and compare it to ours.

3.2 The Hansen-Hodrick and Newey-West Correction for Autocorrelation

Hansen and Hodrick (1980) correct for the autocorrelation in the error term by calculating an estimator of the variance of the regression coefficient in the presence of overlapping data. They assume that $\mathbb{E}\left[x_t\left(q\right)x_{t-v}\left(q\right)u_t(q)u_{t-v}(q)\right] = \mathbb{E}\left[u_t(q)u_{t-v}(q)\right]\mathbb{E}\left[x_t\left(q\right)x_{t-v}\left(q\right)\right]$ and propose the following estimator for the variance:

$$\left(\frac{1}{T}\sum_{t=1}^{T}x_t\left(q\right)x_t\left(q\right)'\right)^{-1}\sum_{v=-q^*+1}^{q^*-1}\left(\frac{1}{T}\sum_{t=1}^{T}x_t\left(q\right)x_{t-v}\left(q\right)'\right)\left(\frac{1}{T}\sum_{t=1}^{T}\widehat{u}_{t+q}(q)\widehat{u}_{t+q-v}(q)\right)$$
$$\times\left(\frac{1}{T}\sum_{t=1}^{T}x_t\left(q\right)x_t\left(q\right)'\right)^{-1},$$

with $q = q^*$. This variance estimator is intended to capture the autocorrelation of the errors and the explanatory variables. We analyze the behavior of this expression using sequential limits, meaning that we first let q go to infinity while keeping q^* fixed, and then we let q^* go to infinity. (See Phillips and Moon (2001) for a discussion on sequential asymptotics). Let $\hat{u}_{t+q+j} = r_{t+q+j} - x'_{t+j}\hat{\beta}$, i.e., the residual from (3). Then, we obtain, using similar reasoning as for the regular variance of the OLS estimator,

$$\begin{aligned}\widehat{V}_{HH} &= \left(\frac{1}{T}\sum_{t=1}^{T}\sum_{j=-q+1}^{q-1}\left(1-\frac{|j|}{q}\right)x_t x'_{t-j}\right)^{-1}\\
&\times\sum_{v=-q^*+1}^{q^*-1}\left[\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{j=-q+1}^{q-1}\left(1-\frac{|j|}{q}\right)x_t x'_{t-v-j}\right)\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{j=-q+1}^{q-1}\left(1-\frac{|j|}{q}\right)\hat{u}_{t+q}\hat{u}_{t+q-v-j}\right)\right]\\
&\times\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{j=-q+1}^{q-1}\left(1-\frac{|j|}{q}\right)x_t x'_{t-j}\right)^{-1}\\
&\xrightarrow{p} f_{xx}^{-1}\left(\sum_{v=-q^*+1}^{q^*-1}f_{xx_v}f_{uu_v}\right)f_{xx}^{-1} = V_{HH_{q^*}}.\end{aligned}$$

Note that

$$f_{xx_v} = (2\pi)^{-1}\lim_{q\to\infty}\sum_{h=-q+1}^{q-1}\Gamma_{xx_v}\left(h\right)\cos\left(0h\right) = (2\pi)^{-1}\lim_{q\to\infty}\sum_{h=-q+1}^{q-1}\Gamma_{xx}\left(h\right) = f_{xx}, \tag{18}$$

for $v \leq q-1$, where

$$\Gamma_{xx_v}\left(k\right) = \mathbb{E}[x_t x'_{t-v}].$$

Using a similar argument for the error process, we confirm that $f_{uu_v} = f_{uu}$. Now taking limits when q^* goes to infinity yields

$$\frac{V_{HH_{q^*}}}{q^*} = \frac{1}{q^*} f_{xx}^{-1} \left(\sum_{v=-q^*+1}^{q^*-1} f_{xx} f_{uu} \right) f_{xx}^{-1} \to f_{xx}^{-1} f_{uu} \int_{-1}^{1} dr = 2 f_{xx}^{-1} f_{uu},$$

so that

$$\frac{V_{HH_{q^*}}}{q^*} \to 2 f_{xx}^{-1} f_{uu},$$

resulting in twice the OLS variance estimate. When applying the Newey and West (1987) method,[7] the same reasoning applies, such that

$$\hat{V}_{NW} \xrightarrow{p} f_{xx}^{-1} \left(\sum_{v=-q^*+1}^{q^*-1} \left(1 - \frac{|v|}{q} \right) f_{xx_v} f_{uu_v} \right) f_{xx}^{-1} = V_{NW_{q^*}},$$

and

$$\frac{V_{NW_{q^*}}}{q^*} \to f_{xx}^{-1} f_{uu}.$$

Two observations are in place here. First, from this analysis, it follows that the order of the correction is asymptotically correct, in that both Hansen-Hodrick and Newey-West scale up the variance by a factor of q, the length of the forecasting horizon. However, under the assumption that $q, T \to \infty$ with $q/T \to 0$, neither method achieves the correct scaling which is $2/3q$ as in Proposition 1. From the analysis presented here, it appears that the Newey-West method comes closer to the asymptotically correct variance in Proposition 1 and accordingly it should be more appropriate. Moreover, as is well known, the Newey-West method has the property of guaranteeing that the variance is positive semi-definite, contrary to the Hansen-Hodrick approach.

Second, the asymptotic analyses of both the Hansen-Hodrick and the Newey-West variances rely on (18) being valid. In practice, $\hat{f}_{xx_v} \neq \hat{f}_{xx}$, due to the use of the Bartlett weighting scheme, which is naturally present when using overlapping data to calculate the sample estimate of (18). This means that the exact properties of the Hansen-Hodrick and the Newey-West approach are to be examined in Monte-Carlo studies. Andrews (1991, Tables IV-VI) shows that in small sample sizes, with large overlap (or equivalently, with strong dependence in the error term), the size distortion from using the Newey-West method is substantial.

Note that, in general, the estimation of the long-run variance-covariance matrix through the use of Heteroscedasticity and Autocorrelation Consistent (HAC) estimators such as in Newey and West

[7] The Newey-West method is essentially the Hansen and Hodrick covariance matrix with a Bartlett kernel.

(1987) or Andrews (1991) is (asymptotically) robust to conditional heteroscedasticity of unknown form. The central limit theorem used in this paper (Lemma 2 in Appendix B) has been proved for linear processes, which do not necessarily allow for heteroscedasticity. Hannan (1970, p. 288) states, however, that this central limit theorem can without doubt be extended to more general processes. This conjecture is worth investigating, but is beyond the scope of this paper. The intuition behind Hannan's conjecture is fairly straightforward, at least for long-run frequency analysis: when considering long-run regressions, all variables are averages, so that conditional heteroscedasticity is likely to be averaged out in the long run, and therefore, of little influence for the asymptotic distribution of the estimators. Furthermore, the covariance matrix of the estimator is valid for fourth order stationary processes, a class which allows for conditional heteroscedasticity. We assess the performance of our standard errors in the presence of GARCH effects, in a Monte-Carlo study included in Appendix A.

3.3 Bias in Predictive Regressions

In a seminal paper, Stambaugh (1986), later published as Stambaugh (1999), demonstrated that predictive regressions, such as a regression of the return on the dividend yield, give rise to a potentially important small sample bias in the regression coefficient estimator. The bias stems from the presence of correlation between the innovations in the two time series. This result has recently been extended by Lewellen (2004) to the case where the predictive variable is a near unit root process. Stambaugh's predictive regression is set up as follows:

$$r_{t+1} = \beta x_t + u_{t+1}, \tag{19}$$

$$x_t = \rho x_{t-1} + z_t, \tag{20}$$

where the first equation (19) is a one-period ahead forecasting or predictive regression, and equation (20) is an autoregressive model of order 1 for the predictive variable. It is assumed that the vector (u, z) of innovations is normally independently distributed with mean zero. In this setup, Stambaugh proves that the correlation between u_{t+1} and v_{t+1} induces a bias in the ordinary least squares regression, which equals:

$$\begin{aligned} \mathbb{E}\left[\widehat{\beta} - \beta\right] &= \mathbb{E}\left[\widehat{\rho} - \rho\right] \Gamma_{zz}(0)^{-1} \Gamma_{zu}(0) \\ &\approx \left(\frac{-(1+3\rho)}{T}\right) \Gamma_{zz}(0)^{-1} \Gamma_{zu}(0), \end{aligned} \tag{21}$$

where $\Gamma_{zu}(h)$ denotes the covariance between z_t and u_{t+h} as defined in (14).[8] Under the normality assumption the bias in the autoregressive coefficient is approximately equal to $-\left(\frac{1+3\rho}{T}\right)$. Within the context of earnings or the dividend yield as predictive variable, the innovation in the yield is typically negatively related to an innovation in the return process, so that $\Gamma_{zu}(0) < 0$. It therefore follows that the larger the autoregressive coefficient in (20), the larger the (upward) bias in the regression coefficient in (19).

Model (1) involves taking a moving average of the predictive variable, and consequently leads naturally to a large degree of persistence (i.e., a large ρ) in the predictive variable. A back-of-the-envelope calculation might therefore conclude that Model (1) suffers from substantial bias in the regression coefficient.

It turns out, however, as we will demonstrate in this section, that this argument is misguided. Although the bias in Model (1) is similar in magnitude to the Stambaugh bias for the short-run regression model (19)-(20), the bias is a much smaller fraction of the coefficient value. The following proposition, proved in the appendix, states the expression for the bias in Model (1), assuming the predictive variable follows (20).

Proposition 3 *For Model (1) the small sample bias is given by:*

$$\mathbb{E}\left[\widehat{\beta}-\beta\right] = \mathbb{E}\left[\widehat{f}_{xx}^{-1}\widehat{f}_{xz_q}\right]\Gamma_{zz}(0)^{-1}\Gamma_{zu}(0). \tag{22}$$

The bias equals a ratio of the covariance between the innovations in the predictive variable and the predicted variable, and the variance of the innovations in the predicted variable, multiplied by the expected value of $\widehat{f}_{xx}^{-1}\widehat{f}_{xz_q}$. This last expression is the long-run covariance of the predictive variable and its q period leaded innovations divided by the long-run variance of the predictive variable. Note that if $q = 1$, the short-run and the long-run covariances are identical: $\widehat{\Gamma}_{xx}(0) = 2\pi\widehat{f}_{xx}$ and $\widehat{\Gamma}_{xz}(1) = 2\pi\widehat{f}_{xz_1}$. Since $(\hat{\rho}-\rho) = \hat{\Gamma}_{zz}(0)^{-1}\hat{\Gamma}_{zu}(1)$, formula (22) coincides with Stambaugh's formula (21) in this case.

The sign of this bias is unknown; the weighted sum of the covariances, $\widehat{\Gamma}_{xz_q}(h)$, will depend on the particular data process under consideration. As argued above, the ratio $\Gamma_{zz}(0)^{-1}\Gamma_{zu}(0)$ is negative in the case of the earnings yield. For the S&P 500 index over the period 1871-2003,

[8]Note that this bias expression is valid in the multivariate case when ρ is a diagonal matrix with the AR(1) coefficients on the diagonal.

the ratio is estimated to -5.7 using the log earnings yield and -14.9 using the dividend yield.[9] We use Monte-Carlo simulations to estimate the value of the term $\mathbb{E}\left[\widehat{f}_{xx}^{-1}\widehat{f}_{xz_q}\right]$, the part of the bias that stems from the autocorrelation in the predictive variable. The results are reported in Table 1 using 10,000 replications. We report the results for different horizons, $q = 1, 2, 5, 8, 10$, and for comparison we also report Stambaugh's approximation for the bias. It turns out that for more persistent predictive variables the bias approximation works particularly well.

[**Include Table 1 about here**]

We can use Table 1 to compute the bias in the coefficient estimate for Model (1). For regression Model (1) with 130 observations, $q = 10$, and an autocorrelation of 0.70 in the predictive variable, the term $\mathbb{E}\left[\widehat{f}_{xx}^{-1}\widehat{f}_{xz_q}\right]$ equals -0.041. With a ratio $\Gamma_{zz}(0)^{-1}\Gamma_{zu}(0)$ of -5.7 this gives rise to a bias of 0.234 in the coefficient estimate.

The bias should be compared with the coefficient estimate in order to judge if this is substantial. Table 2 reports the bias, $\mathbb{E}\left[\widehat{\beta} - \beta\right]$, relative to the bias-adjusted coefficient estimate, $\hat{\beta} - \mathbb{E}\left[\widehat{\beta} - \beta\right]$, for regressions of q-period returns on q-period earnings yields. The coefficient estimate from the regression of the 10-year long-run S&P 500 real return on the 10-year log earnings yield amounts to 2.27, or a bias-adjusted coefficient of 2.036. The bias in the coefficient estimate thus amounts to 12% of this value.

[**Include Table 2 about here**]

The bias is a larger fraction of the coefficient estimate for shorter horizons and for smaller samples. For samples of 100 observations and a horizon of 1, the bias exceeds 100% of the coefficient value. As expected, the bias decreases as the sample size increases.

We also include bias relative to the bias-adjusted coefficient estimate for autocorrelation in the predictive variables of 0.80, 0.90, and 0.99. We confirm the findings by Stambaugh (1999) that the bias in predictive regression, with a horizon of $q = 1$, increases when we increase the autocorrelation in the predictive variable.

[9] Stambaugh (1999) estimates this ratio for the dividend yield for a value weighted portfolio of NYSE stocks to be in the range -13 to -22 for different sub-samples over the period 1926-1996.

3.4 The R-squared

When judging the significance of long-run regressions, a common method is to look at the R^2 as an indicator of the fit of the regression model. However, since Granger and Newbold (1974) it is well known that level-on-level regressions can give rise to the so-called spurious regression problem, a situation where both the regression coefficient and the R^2 become random due to the presence of strong autocorrelation in the error term of the regression. Richardson and Stock (1989) and, more recently, Ferson et al. (2003), extend the Granger and Newbold argument to long-run regressions of the type considered in this paper, and argue that the high R^2 and t-statistics of these regressions are similarly spurious. As in Granger and Newbold, the spurious nature of the regressions arguably stems from the strong autocorrelation in the error term caused by the use of overlapping data.

In this section, we will demonstrate that if the forecast horizon is small compared to the sample size, i.e., if q/T goes to zero, then the spurious regression problem is of little concern, and the R^2 is a relevant indicator of the fit of the regression model. In addition, when the R^2 is large and approaches unity, the long-run regression becomes similar to a cointegrating regression: both time series share the same long-run component. Note that just taking non-overlapping data to estimate any long-run regression model does not remove the spurious regression problem; in any case, as long as q/T is large, there is simply not enough information in the data to estimate consistently the regression coefficient and the R^2, whether one uses overlapping data or not.

For the analysis of the t-statistics, we refer to the previous sections, where we derived the asymptotic theory to find asymptotically correct confidence intervals for the $t-$statistics under the assumption that q/T goes to zero.

To derive the theory for the R^2, we start from the definition of R^2, from which it follows that

$$R^2 = \frac{\hat{\beta}' x(q)' x(q) \hat{\beta}}{r'(q) r(q)} \xrightarrow{p} \frac{\beta' f_{xx} \beta}{f_{rr}} = f_{rr}^{-1} f_{xr}' f_{xx}^{-1} f_{xr}; \tag{23}$$

so that R^2 is a measure of how much of the long-run variability in returns can be explained by the long-run component of the predictive variable, or equivalently, R^2 is the squared long-run correlation. Consistency is guaranteed under the assumption that q/T goes to zero, as opposed to the spurious regression case where R^2 is inconsistent and is distributed in the limit as a random variable.

To further understand the relevance of the R^2, consider the expression for the variance in the asymptotic distribution of the least squares estimator. Ignoring the scaling factor of 2/3, note that:

$$f_{xx}^{-1} f_{uu} = f_{xx}^{-1} \left(f_{rr} - \beta' f_{xx} \beta, \right).$$

Therefore, it follows that

$$\begin{aligned} f_{xx}^{-1} f_{uu} &= f_{xx}^{-1} \left(f_{rr}^{-1} f_{rr} - \frac{\beta' f_{xx} \beta}{f_{rr}} \right) f_{rr} \\ &= f_{xx}^{-1} \left(1 - R^2 \right) f_{rr}. \end{aligned} \tag{24}$$

This means that the higher the R^2, the lower the variance of the estimator. Consequently, the R^2 is a relevant measure to consider both in the time and in the frequency domain, and we can use it to compare different predictive variables.

When R^2 goes to 1, the regression falls in the class of cointegrating relations, where all the long-run variance of the regressor can be explained by the long-run variance of the regressand. To see this, start off from Model (1),

$$\sum_{j=1}^{q} r_{t+1+q-j} = \sum_{j=1}^{q} x'_{t-j} \beta + \sum_{j=1}^{q} u_{t+1+q-j}. \tag{25}$$

Take squares on both side of (25)

$$\hat{f}_{rr} = \beta' \hat{f}_{xx} \beta + \hat{f}_{uu} + 2 \hat{f}_{xu},$$

which yields in the limit

$$f_{rr} = \beta' f_{xx} \beta + f_{uu} + 2 f_{xu}. \tag{26}$$

Under the assumption of cointegration, f_{uu}/f_{rr} equals zero,[10] and f_{xu} equals zero under the assumption of informational efficiency (see Section 2). Dividing both sides of (26) by f_{rr}, it follows therefore that

$$R^2 = \frac{\beta' f_{xx} \beta}{f_{rr}} = 1.$$

Note, however, that when $R^2 = 1$, the asymptotic theory of the previous sections fails to hold. Indeed, in this case the variance of the estimator equals zero; the scaling of the estimator $\hat{\beta}$ by $\sqrt{T/q}$ is insufficient to derive the asymptotic distribution. A different asymptotic theory applies,

[10] To see this, note that cointegration involves a non-stationary dependent variable ($f_{rr} \to \infty$), and implies a stationary error term ($f_{uu} \to 0$).

as derived in Phillips (1987) in the time domain and by Phillips (1991) in the frequency domain. In this case, the estimator converges at speed T; i.e., the estimator becomes super-consistent, irrespectively of whether we take a moving average or not (Phillips 1991). Of course, since all our regression theory is derived under the assumption that both regressor and regressand are stationary, this is of little importance here; it serves only as a limiting case to understand the significance of obtaining a high R^2 when performing a long-run regression.

To summarize this section, we have found that under the assumption that q/T goes to zero, R^2 is a consistent indicator of the explanatory power of a regression model. The higher the R^2, the more precisely we can estimate the regression coefficient. In the limiting case where the R^2 approaches 1, the regression becomes of the cointegrating type and the regression coefficient becomes super-consistent.

3.5 Inference Under the Null Hypothesis

So far this paper has developed a theory of inference in long-run regressions under the maintained hypothesis, i.e., a theory that is both valid under the null of no predictability and the alternative hypothesis of predictability of returns.

In this section, we use our framework to construct tests that explicitly impose the null hypothesis. The idea of deriving test statistics under the null to test for predictability in long-run regressions goes back to Richardson and Stock (1989), Richardson and Smith (1991), and more recently, Torous et al. (2005). The advantage of such tests is that the limiting distribution is free from any estimation error in the regression coefficient, since the regression coefficient is set to zero.

It turns out that to construct such a test in our framework simply involves rescaling our proposed t-tests from Proposition 2 by $\sqrt{1-R^2}$. To see this, recall (23), the expression for the R^2 derived in the previous section. Imposing the null means $\beta = 0$, and therefore, $R^2 = 0$.

Recall now the variance of the coefficient estimator (24), also derived in the previous section. Using the fact that R^2 equals zero, we obtain the expression for the variance of the estimator under the null hypothesis:

$$\frac{2}{3} f_{xx}^{-1} f_{rr}.$$

Also for the Hansen-Hodrick or Newey-West statistic, it suffices to multiply the corresponding t-statistic by $\sqrt{1-R^2}$. To see this, recall from Section 3.2 that using sequential asymptotics, the

Newey-West variance converges to

$$f_{xx}^{-1} f_{uu} = f_{xx}^{-1}(1 - R^2) f_{rr}.$$

We summarize these results in the following proposition.

Proposition 4 *Under Assumption A, asymptotically correct standard errors of the coefficient estimate in Model* (1) *under the null hypothesis of no predictability are computed as*

$$\hat{SE}_{HTH_0} = \frac{\hat{SE}_{OLS} \cdot \sqrt{\frac{2}{3} q}}{\sqrt{1 - R^2}}, \tag{27}$$

where $\hat{SE}_{OLS}$ *denotes the OLS standard errors and* q *is the length of the forecasting horizon. Asymptotically correct t-statistics under the null hypothesis are computed as*

$$t\text{-}stat_{HTH_0} = \frac{t\text{-}stat_{OLS} \cdot \sqrt{1 - R^2}}{\sqrt{\frac{2}{3} q}}, \tag{28}$$

where $t\text{-}stat_{OLS}$ *denotes the t-statistic based on OLS standard errors and* q *is the length of the forecasting horizon.*

The argument holds in general, and one can calculate standard errors, OLS or Newey-West, under the null hypothesis, simply by dividing the standard error by $\sqrt{1 - R^2}$. Equivalently, one can form the t-statistic under the null by scaling down the t-statistic by $\sqrt{1 - R^2}$.

4 General Kernels

Estimating the coefficient in the long-run regression Model (1) corresponds to using a Bartlett kernel estimate. However, a wide spectrum of other kernels can be used. We generalize the estimation procedure above to kernel-based estimators of the following type:

$$\hat{\beta} = \left(\sum_{h=-q+1}^{q-1} k\left(\frac{h}{q}\right) \hat{\Gamma}_{xx}(h) \right)^{-1} \sum_{h=-q+1}^{q-1} k\left(\frac{h}{q}\right) \hat{\Gamma}_{xr}(q+h), \tag{29}$$

where $\hat{\Gamma}_{xx}(h)$ and $\hat{\Gamma}_{xr}(q+h)$ are defined in (13) and (14), and $k(h/q)$ is the kernel. The kernel must belong to the following class

$$\mathcal{K} = \left\{ \begin{array}{c} k(\cdot) : \mathbb{R} \to [-1, 1] \mid k(0) = 1,\ k(x) = k(-x),\ \text{for all } x \in \mathbb{R}, \\ \int_{-\infty}^{+\infty} k^2(x)\, dx < \infty,\ k(\cdot)\ \text{is a continuous function and } K(\lambda) \geq 0 \text{ for all } \lambda \in \mathbb{R} \end{array} \right\},$$

where $K(\lambda) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} k(x) e^{-ix\lambda} dx$ is referred to as the spectral window generator. This class of kernels guarantees that the denominator of $\hat{\beta}$ is positive semi-definite. The restriction to kernels subject to a truncation lag reflects the block structure in (4). This is to prevent covariances of the form $\mathbb{E}[r_t r_{t+q+j}]$ for $j > q-1$ or $j < -q+1$ from entering the estimator, thereby leading to bias in the estimator of the cross spectrum as defined by (10). The class $\mathcal{K}$ includes widely used kernels such as the Bartlett and the Parzen kernels (Andrews, 1991), and can be extended to truncation type of estimators with a modified spectral window generator defined as $K^*(\lambda) = \max\{0, K(\lambda)\}$ as in Politis and Romano (1995). The asymptotic distribution of this type of kernel based estimators is as follows:

$$\sqrt{\frac{T}{q}}\left(\hat{\beta}-\beta\right) \xrightarrow{d} N\left(0, \left(\int_{-1}^{1} k^2(x)\, dx\right) f_{xx}^{-1} f_{uu}\right),$$

where in the case of the Bartlett estimator $\int_{-1}^{1} k^2(x)\, dx = 2/3$, and in case of the Parzen kernel $\int_{-1}^{1} k^2(x)\, dx = 0.539$.

5 Forecasting the Long-run with a Short-Run Variable

As argued before, current practice consists of regressing long-run returns on a single period predictive variable such as the dividend yield or the price earnings ratio (see e.g., Campbell and Shiller, 1998). We will now provide a formal analysis of this type of model in terms of spectral quantities, with the objective of simplifying the inference procedure and comparing its efficiency with Model (1).

5.1 Inference

As before, it can be shown that the Hansen and Hodrick approach of correcting the standard errors is an asymptotically valid method, but might suffer from small sample deficiencies. We now show that in the traditional long-run regressions as well, a computationally simpler and therefore more robust method exists to obtain asymptotically correct standard errors. Recall Model (2) defined above:

$$\sum_{j=1}^{q} r_{t+j} = x_t'\beta + v_{t+q}.$$

The OLS estimator is defined in the time domain as follows:

$$\hat{\beta} = \left(\frac{1}{T} \sum_{t=1}^{T} x_t x_t' \right)^{-1} \frac{1}{T} \sum_{t=1}^{T} \sum_{j=1}^{q} x_t r_{t+j}.$$

We now analyze the numerator and denominator of $\hat{\beta}$ separately. The numerator can be rewritten as follows:

$$\frac{1}{T} \sum_{t=1}^{T} \sum_{j=1}^{q} x_t r_{t+j} = \sum_{h=-[q/2]+1}^{[q/2]-1} \hat{\Gamma}_{xr} \left([q/2] + h \right) \xrightarrow{p} 2\pi f_{xr}, \tag{30}$$

as $q \to \infty$ and $T \to \infty$, where $\hat{\Gamma}_{xr}$ is defined in (14). The $[\cdot]$ denotes the greatest lesser integer. The denominator is the sample estimator of the variance of the predictive variable and satisfies a law of large numbers with convergence rate $\sqrt{T}$:

$$\frac{1}{T} \sum_{t=1}^{T} x_t x_t' \xrightarrow{p} \Gamma_{xx}(0).$$

From (30) we see that this estimate may be interpreted as a truncated kernel estimator of the long-run covariance between returns and the predictive variable. The truncated kernel is defined by $k(x) = 1$ for $|x| \leq 1$ and equal to zero elsewhere. Hence, the coefficient β may be estimated either from OLS or from a truncated kernel method; these methods are asymptotically equivalent. Having identified the kernel, we can use the same line of reasoning as we used for Proposition 1 to show that $\hat{f}_{xr}$ has the following asymptotic distribution:

$$\sqrt{\frac{T}{q/2}} \left(\hat{f}_{xr} - f_{xr} \right) \xrightarrow{d} N \left(0, 2 \left(f_{xx} f_{rr} + f_{xr} f_{xr}' \right) \right),$$

where the factor of 2 in the variance stems from $\int k^2(x)\, dx = 2$ for the truncated kernel. Then by the continuous mapping theorem, we obtain the following result.

Proposition 5 *Under the conditions stated in Assumption A, the slope coefficient in Model* (2) *is asymptotically distributed as follows:*

$$\sqrt{\frac{T}{q}} \left(\hat{\beta} - \beta \right) \xrightarrow{d} N \left(0, \Gamma_{xx}(0)^{-2} (2\pi)^2 \left(f_{xx} f_{rr} + f_{xr} f_{xr}' \right) \right), \tag{31}$$

or equivalently,

$$\sqrt{\frac{T}{q}} \left(\hat{\beta} - \beta \right) \xrightarrow{d} N \left(0, \Gamma_{xx}(0)^{-2} \left(\Omega_{xx} \omega_{rr} + \omega_{xr} \omega_{xr}' \right) \right). \tag{32}$$

The equivalence between (31) and (32) stems from the link between the spectral matrix (left hand side) and the long-run covariance matrix (right hand side):

$$2\pi \begin{pmatrix} f_{rr} & f'_{xr} \\ f_{xr} & f_{xx} \end{pmatrix} = \begin{pmatrix} \omega_{rr} & \omega'_{rx} \\ \omega_{xr} & \Omega_{xx} \end{pmatrix}. \tag{33}$$

Therefore, under these conditions, asymptotically correct confidence intervals are obtained by estimating the long-run covariance matrix of the one-period return and the predictive variable, together with the variance of the predictive variable. This procedure is more robust than estimating the covariance matrix of the estimator, using Newey-West or Hansen-Hodrick, which involve estimating the long-run covariance matrix of the overlapping error and the predictive variable. Under the null hypothesis of no predictability, the asymptotic theory simplifies to

$$\sqrt{\frac{T}{q}}\left(\hat{\beta}-\beta\right) \xrightarrow{d} N\left(0, \Gamma_{xx}(0)^{-2}\Omega_{xx}\omega_{rr}\right).$$

To circumvent estimating this covariance matrix, Hodrick (1992) and Bollerslev and Hodrick (1996) have proposed estimating regression Model (2*) instead of Model (2). In their model the short-run return is regressed on the long-run predictive variable; accordingly, this removes the overlap in the left hand side variable and thereby the large degree of autocorrelation in the error term. As shown by Bollerslev and Hodrick (1996), the numerator of this regression coefficient estimator is similar to that for Model (2), under the assumption of covariance stationarity. The denominator of the regression coefficient estimator is similar to that of Model (1). We use our setup to analyze this regression model and obtain the asymptotic distribution of the regression coefficient estimate of Model (2*). For simplicity we focus on the univariate case. We state this result as a proposition (the proof is in Appendix B).

Proposition 6 *Under the conditions stated in Assumption A, the slope coefficient in Model* (2*) *with* $m = 1$ *is asymptotically distributed as follows:*

$$\sqrt{\frac{T}{q}}\left(\hat{\beta}-\beta\right) \xrightarrow{d} N\left(0, f_{xx}^{-2}\left(f_{xx}f_{rr} + \frac{1}{3}f_{xr}^2\right)\right).$$

or equivalently,

$$\sqrt{\frac{T}{q}}\left(\hat{\beta}-\beta\right) \xrightarrow{d} N\left(0, \Omega_{xx}^{-2}\ \left(\Omega_{xx}\omega_{rr} + \frac{1}{3}\omega_{xr}^2\right)\right).$$

For Model (2*) we can obtain asymptotically correct confidence intervals by estimating the long-run covariance matrix of the short-run return and the short-run predictive variable, just as for Model (2). Instead of the variance of the short-run predictive variable, we now need the variance of the long-run predictive variable. This also corresponds to estimating the long-run variance of the predictive variable. Under the null, the asymptotic theory reduces to

$$\sqrt{\frac{T}{q}}\left(\hat{\beta}-\beta\right)\xrightarrow{d}N\left(0,\Omega_{xx}^{-1}\omega_{rr}\right).$$

A relevant question now is which of these models is the most efficient. We examine this question in the next subsection.

5.2 Efficiency Comparisons

When estimating Model (1) we implicitly use a Bartlett weighting scheme, which reduces the variability of the estimator compared to the rectangular weighting scheme implicitly used in Model (2). Moreover, Model (2) establishes a long-run relationship over a period of q only; the balanced approach, Model (1), defines a long-run relationship over a period with length $2q$.

We now derive the condition under which the coefficient estimates from Model (1) are more efficiently estimated than those estimated from Model (2). We need to find a condition such that the difference between the variance-covariance matrix from Model (1) and that from Model (2) is a positive semi-definite matrix (Hamilton (1994, p. 741)). Using the fact that $f_{uu}=f_{rr}-\beta' f_{xx}\beta=f_{rr}-f_{xr}'f_{xx}^{-1}f_{xr}$, and using relation (33), we can rewrite the variance-covariance in Proposition 1 as $\frac{2}{3}\Omega_{xx}^{-2}(\Omega_{xx}\omega_{rr}-\Omega_{xx}\omega_{xr}'\Omega_{xx}^{-1}\omega_{xr})$. This leads to a sufficient condition for Model (1) to be more efficient than Model (2), which we state in the following proposition.

Proposition 7 *Under the conditions stated in Assumption A, the regression coefficient is more efficiently estimated using Model* (1) *than using Model* (2) *if*

$$\Gamma_{xx}(0)^{-1}\Omega_{xx}\geq\sqrt{\frac{2}{3}}I_m,$$

where Ω_{xx} is the long-run variance of x_t and $\Gamma_{xx}(0)$ is the (overall) variance of x_t, and I_m is the m dimensional identity matrix.

Proof: See Appendix B.

In the univariate case this is a variance ratio test. Intuitively, this condition stipulates that the long-run component of the explanatory variable exhibits enough variation when compared with the overall variation in the time series. As noted by Campbell and Mankiw (1987), the variance ratio test has the following asymptotic distribution: $\sqrt{\frac{T}{q}}(\hat{VR}-1) \xrightarrow{d} N(0, \frac{4}{3})$, where $\hat{VR}$ denotes the estimate of the variance ratio $\frac{\Omega_{xx}}{\Gamma_{xx}(0)}$. This allows us to construct a t-test= $\frac{\sqrt{\frac{T}{q}}(\hat{VR}-1)}{\sqrt{\frac{4}{3}}}$ to judge whether the variance ratio is significantly different from unity.

It is also interesting to see how Model (1) fares against Model (2*). To simplify, we focus on the univariate case.

Proposition 8 *Under the conditions stated in Assumption A, the regression coefficient is always at least as efficiently estimated using Model* (1) *as when using Model* (2*).

Proposition 9 *Under the conditions stated in Assumption A, the regression coefficient is more efficiently estimated using Model* (2*) *than using Model* (2) *if*

$$\Gamma_{xx}(0)^{-1}\Omega_{xx} \geq 1,$$

where Ω_{xx} *is the long-run variance of* x_t *and* $\Gamma_{xx}(0)$ *is the (overall) variance of* x_t.

Before setting up a regression framework for the data at hand, it is therefore advisable to examine the ratio between the long-run variance and the (overall) variance of the predictive variable, to select the most efficient estimation method.

6 Forecasting the S&P 500

The debate is still unsettled on whether the stock market is forecastable or not. Earlier research suggested that there was strong long-run predictability based on highly significant t-statistics on the regression coefficient for traditional long-run regressions (Campbell, Lo and MacKinlay, 1997). This evidence has been criticized as merely proving evidence that the forecast horizon is too long with respect to the sample size; see Wolf (2000), Torous et al. (2005) and Ferson et al. (2003) for recent references. Traditionally, there has been little evidence of short-run predictability of stock returns.

In this section, we examine the question of whether long-run stock returns are predictable by estimating the forecasting regression Models (1) and (2) and conducting inference using our new asymptotic techniques. We focus on two forecasting variables that figure prominently in the literature: the earnings yield and the dividend yield.

6.1 The Data

We focus on long-run forecasting horizons of 10 years. We have chosen this horizon for two reasons. First, Campbell and Shiller (2001) argue in a recent paper that stock markets exhibit strong forecastability at the 10-year horizon. Second, we experimented with various horizons, between 3 and 15 years, and the best fit was found for the 10-year horizon. The results for Model (1) are reported in Table 3. Figure 2 supports this observation. We see that the actual returns (solid lines) and the predicted returns follow each other most closely at the 10 year horizon. While looking at figures to determine the "optimal" forecast horizon might be considered data-snooping, from a technical point of view this is equivalent to choosing the bandwidth in the spectral estimator. We could also have implemented a data-based bandwidth selection method; see Andrews (1991) or Newey and West (1994) for details.

[**Include Table 3 about here**]

[**Include Figure 2 about here**]

The (annual) data on prices, earnings, and dividends were available from Professor Shiller's web site (http://www.robertshiller.com), and cover the period from 1871 through 2003. The time series are corrected for inflation using the consumer price index. We compute the annual log-returns as $\ln(\frac{P_t+D_t}{P_{t-1}})$, where P_t denotes the index level in year t and D_t represents the dividends in year t. Both P_t and D_t are in real terms. The earnings price ratio, e/p_t, is computed as the natural logarithm of 1 plus the annual earnings divided by the current index level, $\ln(1+E_t/P_t)$. Likewise, the dividend price ratio, d/p_t, is computed as the natural logarithm of 1 plus the total annual dividends divided by the current index level, $\ln(1+D_t/P_t)$. We refer to these ratios as yields. The mean, the standard deviation, and the first order autocorrelation of the time series are summarized in Table 4. We also report the Phillip-Perron unit root test and the Kwiatkowski-Phillips-Schmidt-Shin stationarity test. The Phillips-Perron test rejects that any of the time series contain a unit root. The Kwiatkowski-Phillips-Schmidt-Shin test cannot reject that one-year returns and earnings yields are stationary, but it rejects that the one year dividend yield is stationary at the 5% confidence level.

[Include Table 4 about here]

Besides the total sample period we also examine two sub-periods: 1920-1990, which covers the depressions in the 30s and 70s but excludes the dot-com bubble in the late 1990s, and a post-World War II period covering 1950 trough 2003.

6.2 Predicting Returns with long-run Variables

We use the forecasting relation (1) with a long-run forecasting variable on the right hand side:

$$\sum_{j=1}^{q} r_{t+j} = \beta \sum_{j=1}^{q} x_{t-q+j} + v_{t+q}. \tag{34}$$

To analyze the long-run, we forecast 10-year returns, which corresponds to an analysis over a period of 20 years (10 year return, plus the earnings yield over the past 10 years). Table 5 shows the ability of the long-run component of the earnings yield to forecast long-run returns.

[Include Table 5 about here]

The forecasting performance at the 10 year horizon is quite remarkable for each sample period. If we limit the sample to the period 1950-2003 the model is able to explain up to 69% of the variation in long-run returns. We report both OLS and Bartlett estimates and see that for all sample periods the two methods yield almost identical coefficient estimates. This is what we would expect, given the simulation study in Appendix A and the proof in Appendix B showing that the two methods are asymptotically equivalent. The earnings yield has a positive coefficient ranging between 2.27 for the overall sample period to 2.86 in the post-World War II period. The coefficient is significant as judged by all of the reported t-statistics. We report three t-statistics. The $t\text{-stat}_{HT}$ is calculated as equation (17); $t\text{-stat}_{Bartlett}$ is based on Bartlett estimates of the coefficient and of the standard errors as presented in Proposition 1. For comparison we also report the Newey-West (NW) corrected t-statistics. We see that the Newey-West t-statistics are even more significant, reflecting the size distortion problem mentioned in Section 3.2.

As argued in Section 3.3, long-run predictive regressions might suffer from a biased regression coefficient. Therefore, we report the bias-adjusted coefficient estimates along with t-tests (see Panel

B). We use Table 1 to calculate the bias. The results show that even with the bias-adjustment we find that earnings yields predict real long-run returns.

In Panel C we additionally impose the null hypothesis of no predictability; this decreases the probability of rejecting the null when the null is in fact true. We compute the t-statistics using Proposition 4. Even when imposing the null of no predictability, we reject it.

These findings are illustrated graphically in Figure 3. The bold line represents the actual returns and the line marked with a "+" presents the predicted returns using the OLS coefficient estimate from the period 1920-1990.[11] The two lines move closely together and they follow a cycle with a period of approximately 30 years, in which the stock market goes successively through a period of low returns and corresponding low earnings ratios, followed by a period of high returns and subsequently high price earnings ratios. Incidentally, the graphical evidence suggests that expected real returns are negative for the last part of the sample, suggesting that an irrational exuberance explanation along the lines of Shiller (2000) cannot be rejected statistically. Indeed, as pointed out by Fama and French (1989), negative expected returns are incompatible with a rational expectations model.[12]

Furthermore, if the performance of this simple model is any guideline for the expected return in the stock market for the next ten years, the outlook is very bleak. The prediction for the next ten years is approximately -4% annual real return for the next ten years, so that by 2013, the stock market should be at about 67% of its current level in real terms. The intuition behind this rather grim result is as follows. Historically the real return on the S&P 500 stock market has been around 7%, about the same as the earnings yield, suggesting that historically, the average earnings yield has been a reasonable proxy for the expected average return; see Modigliani and Miller (1961) for exact conditions under which the earnings yield proxies for expected returns, and Hansen and Tuypens (2004) for a discussion of the trailing earnings over price ratio. Considering that the earnings yield hovers around 5% in 2003; prices have to fall, in real terms, by another 30% to bring the expected return back to its historical average.

[**Include Figure 3 about here**]

[11] The picture is almost identical using Bartlett coefficient estimates.

[12] To be fully precise, net returns (gross returns minus the riskfree) interest rate should be considered.

The nature of the long-run predictability suggests that an explanation in terms of business cycle fluctuations might be rather misguiding. Indeed, major business cycles have a period of around 8 years; minor business cycles have a period of around 4 years (Sargent, 1979). The cycle observed here has a period of 30 years - well beyond any period implied by the business cycle.

The dividend yield has frequently been used in the previous literature to predict equity returns starting with Scholes and Black (1974). It is interesting to compare the predictive power of the dividend yield with that of the earnings yield. Table 6 contains the estimates of regression (34) with $x_t = \ln(1 + E_t/P_t)$. It is immediately apparent that both t-statistics and R^2s for the long-run component of the dividend yield are much lower than for the long-run component of the earnings yield.

[**Include Table 6 about here**]

Figure 4 shows the actual returns (bold line) and the predicted returns (line marked with a "+") using dividend yields and an estimation period 1920-1990. Before World War II, the model does reasonably well in predicting stock returns: the two lines move closely together. However, it fails to predict the downturn during World War II and the boom in the 1960s. From the 1980s onward, the dividend yield predicts that returns ought to have been lower than actually realized. This might be a result of the diminished dividend payout ratio after 1978 as reported by Fama and French (2001).

[**Include Figure 4 about here**]

Our results therefore suggest that long-run returns may be forecasted better by using the earnings yield instead of the dividend yield. This can be confirmed by the mean squared error (MSE) for the out-of-sample prediction period 1991-2003. The MSE for this period amounts to 0.026 using the earnings yield and 0.087 using the dividend yield.

6.3 Predicting Returns with Single Period Variables

According to Campbell et al. (1997), the (single-period) dividend yield has been the best forecasting variable for the stock market, especially after World War II. The theory developed in this paper argues against the use of single period variables to forecast long-run returns when the predictive variable is persistent. We apply our asymptotic theory to test whether single-period dividend yields have forecasting power for long-run stock returns. Table 7, Panel A presents the results for Model (2) with the past dividend yield as the explanatory variable. Immediately apparent are the lower R^2 and lower t-statistics, compared to the long-run regressions of Model (1). The slope coefficients range between 1.4 and 2.3; all but the 1920-1990 period's estimates are deemed insignificant at the 95% confidence level using the asymptotic theory developed in Proposition 2. Note that if we based our conclusions on the t-statistic based on Newey-West, we would not have been able to reject that the single period dividend yield predicts long-run returns. Similar results emerge when replacing the single-period dividend yield with the single-period earnings yield as reported in Table 7, Panel B.

[**Include Table 7 about here**]

Figures 5 and 6 confirm these findings. The predicted time series (lines marked with a "+") seem a lot noisier than the actual returns (bold lines), reflecting that the forecasting model is unbalanced, since we are using a short-run variable to predict a long-run return. While Figure 5 seems to indicate that the dividend yield has some forecasting power in various sub-samples, it predicts a boom in the late 1930s, but misses almost entirely the boom in the late 1990s. The earnings yield does a better job in predicting the recent boom.

[**Include Figure 5 about here**]

[**Include Figure 6 about here**]

If we compare the R^2 of regressions using the dividend yield with the regression using the earnings yield, we see that there are no major differences. The earnings yield explains a greater portion of the total variation in long-run stock returns in the overall period, but the dividend yield seems to explain slightly more of the variation in the long-run returns in the post-World War II period. It is more interesting to examine the performance of the dividend yield and the earnings yield out-of-sample. We compute the MSE over the out-of sample period 1991-2003. Using the earnings yield as a predictive variable results in a MSE of 0.075, compared with a MSE of 0.094 when using the dividend yield.

We now come back to the question of whether we should use a short-run variable or a long-run variable to predict the long-run return. First, the R^2 is higher when using a long-run variable versus a short-run variable. Second, the variance ratio from Proposition 7 equals 5.0 for the dividend yield and 4.2 for the earnings yield. These ratios are much larger than $\sqrt{\frac{2}{3}}$, telling us that Model (1) is more efficient when using either the dividend yield or the earnings yield as predictive variables.[13]

Summing up, the empirical findings suggest that when forecasting stock returns with earnings or dividend yields one obtains more efficient estimates using a balanced regression like Model (1). In addition, for the S&P 500 index, the earnings yield seems to be the most informative predictive variable.

6.4 Earnings Growth

It is widely believed that all long-run predictability stems from changes in the equity premium. If predictability is a result of time variation in the equity premium, then earnings growth should not be predictable (Cochrane (2001, p. 414)). For the earnings yield to go back to the historical average, it therefore must be that prices adjust, not earnings. All reversion to the mean should be through the price component of the earnings yield. This can be tested using our forecasting methodology to predict long-run changes in earnings using the earnings yield as the explanatory variable. Table 8 contains the result.

[Include Table 8 about here]

At first, the results suggest that earnings growth is unpredictable. However, after bias-adjusting the coefficient estimates, using Table 1 and $\Gamma_{zz}(0)^{-1}\Gamma_{zu}(0) = 6.1$, there is evidence of predictability in the sub-sample 1950-2003. If we take a look at Figure 7 we see that the actual and predicted earnings growth move closer together for the post-World War II period. Indeed, a cycle has appeared in the last part of the sample, which shows up as significant when adjusting for bias, even when testing under the null. The regression coefficient shows that a high (low) earnings yield is followed by high (low) earnings growth. This suggests that the earnings yield predicts real activity, not only returns on stocks.

[Include Figure 7 about here]

[13] The Campbell and Mankiw (1987) t-test$=\frac{\sqrt{\frac{T}{q}}(\hat{VR}-1)}{\sqrt{\frac{4}{3}}}$ takes the values 12.3 and 9.1 for the dividend yield and the earnings yield, respectively. Based on these test statistics, we conclude that the variance ratio for the earnings yield and for the dividend yield are significantly larger than 1.

6.5 Forecasting Bond Returns

We can further disentangle whether the predictability stems from time varying risk premia or from time variation in the real returns of the economy by looking at short and long duration bond returns. If short duration bond returns are also predictable from the earnings yield it suggests that predictability in the stock and bond market is driven by time varying risk free rates. If long duration bond returns are predictable using the earnings yield, this suggests that long duration bonds and stocks share a common risk factor; or equivalently, that the equity risk premium affects long bonds and equities in the same way. We can use our regression methodology to judge if real long term bond returns are predictable.

6.5.1 Aaa Corporate Bond Returns

We consider Moody's time series of yields on Aaa rated corporate bonds with an average maturity of 30 years over the period 1919-2003. We construct the bond returns and adjust for inflation. The mean and standard deviation are presented in Table 4. We estimate Model (1) with the 10 year bond return on the right hand side and past 10 year earnings yield of the S&P 500 index on the left hand side. The results are shown in Table 9.

[Include Table 9 about here]

While the estimation results are not as strong as for stocks, over the full sample, there is evidence of significant predictability using the earnings yield as a forecasting variable over the full period. Around 35% of the long-run variation can be explained using past earnings yields.

Figure 8 illustrates the forecasting properties of the earnings yield for future 10 year corporate bond returns. The predicted return, based on an estimation period from 1920-1990, does a fairly good job with a MSE of 0.06 over the period 1991-2003. As in the corresponding figure for stock market returns, most of these results seem to be driven by a 30-year long cycle; although this is less apparent for bond returns than for stock returns.

[Include Figure 8 about here]

6.5.2 Returns on Long and Short Duration US Treasuries

Lastly, we apply our forecasting model to the 10-year return on government bonds. We consider a basket of long term US treasury bonds, neither due nor callable in less than 10 years. For the short term bonds, we compute the 10-year return from holding the most recently issued 3-month T-bill.[14] The estimation results are provided in Tables 10 and 11, respectively.

[Include Table 10 about here]

[Include Table 11 about here]

The time series and forecasting model behave very similarly to the investment grade corporate bonds considered above. Past earnings yields are significant over the whole sample period, even after adjusting for bias and imposing the null. The earnings yield is able to explain up to 49% of the long-run variation in long term treasury returns and up to 30% of the long-run variation in the T-bill returns. As before, all long term predictability seems to stem from a long-run cycle of around 30 years, as suggested by Figure 9.

[Include Figure 9 about here]

[Include Figure 10 about here]

The presence of predictability in the US treasury bond market suggests that explanations based on time varying risk premia for stock returns might be incomplete, since both markets seem to have time variation in their expected returns that comove around a common long-run cycle. Instead, our results point partially towards time varying real returns and growth rates as the driver for the presence of long-run predictability in the US equity market.

7 Conclusion

The contributions of this paper are threefold. First, we derive asymptotically correct standard errors of the coefficient in long-run forecasting regressions under the maintained hypothesis. Concretely, in the regression of long-run returns on a vector of long-run predictive variables, asymptotically

[14] Data is from NBER.

correct standard errors are obtained by multiplying the OLS standard errors by $\sqrt{2/3q}$, where q is the length of the forecasting horizon considered. This procedure does not involve the estimation of an additional estimator, viz. the long-run covariance matrix of the predictive variables and the error term. Accordingly, this method is much more robust to small sample problems than the Hansen and Hodrick (1980) and Newey and West (1987) methods. We establish a simple device to construct the t-statistics under the null hypothesis of no predictability, namely by scaling down the t-statistic by $\sqrt{1-R^2}$. We also demonstrate that standard errors from the typical long-run regression of a long-run return on a short-run predictive variable can be simplified considerably.

A second contribution consists of addressing the efficiency of the different long-run regression models and the potential small sample biases of the regression coefficient estimator for Model (1). We compare the efficiency of a regression of a long-run variable on a short-run predictive variable with that of using a long-run variable on both sides of the equation. A sufficient condition for the first model being more efficient than the latter is that the predictive variable exhibits enough variation in the long-run relative to the variation in the overall time series. We also analyze the small sample bias for long-run regressions along the lines of Stambaugh (1999). The bias for long-run regression with horizons longer than one period is in the same range as that for the one period predictive regressions. However, the bias is a much smaller fraction of the regression coefficient. Interestingly, the bias shrinks the more persistent variables.

Third, we apply this methodology to forecast the movements in the US asset market. Our results show that the long-run component of the earnings yield predicts up to 69% of the variation in 10-year S&P 500 index returns in the post-World War II period. This spectacular forecasting performance is due to a cycle of approximately 30 years that has repeated itself two and a half times since the 1920s. The long-run component of the dividend yield has much less forecasting ability, which may be a consequence of the declining payout ratio for US companies since the 1970s as documented by Fama and French (2001). We find some evidence that earnings growth is predictable in the post-World War II subperiod. In addition, Aaa corporate bond returns and treasury returns are heavily influenced by the same 30-year cycle that drives the equity returns, and their long-run variation is predictable up to 49%.

Overall, these findings suggest that current macroeconomic and asset-pricing explanations of stock market predictability in terms of business cycle fluctuations of the pricing kernel might be incomplete, as business cycles are generally believed to last at most 8 years. The long-run predictability of bond returns and stock returns suggest explanations in terms of demographic dynamics. Alternatively, this result could be interpreted as random effects of long-run time series, possibly caused by waves of irrational exuberance that might be of little or no use in forecasting the future. An analysis using international data could shed light on what effect is at work to generate this type of cycle and determine whether the long-run predictability as observed in the US asset markets has any economic significance. We leave this for future research.

A Simulation Study

The OLS coefficient estimate is asymptotically equivalent to the Bartlett coefficient estimate (cf. Appendix B). Furthermore, we showed in Section 3.1 that asymptotically correct standard errors can be obtained by multiplying the OLS standard errors by $\sqrt{\frac{2}{3}q}$ (cf. Proposition 2). We shall refer to these standard errors as HT standard errors. The question that remains is how well these methods perform for the sample sizes used in practice. A Monte-Carlo study allows us to assess the finite sample properties of the Bartlett and the OLS coefficient estimates as well as the t-statistics based on various standard errors. We focus on t-statistics based on the Newey-West standard error using $q = 10$ lags, the Bartlett standard error, the HT standard error estimate, and lastly, we examine the properties of a t-statistic based on the OLS standard error multiplied by $\sqrt{q}$, as suggested by Engle (1974).

We consider the following data generating process for returns:

$$r_t = \beta x_{t-q} + u_t,$$

where u_t is the error term. We consider two situations: one where the error is white noise with variance σ_u^2, and another where u_t evolves in according to a $GARCH(1,1)$ model:

$$\begin{aligned} u_t &= \epsilon_t \sqrt{h_{t-1}}, \text{ where } \epsilon_t \sim nid(0,1) \\ h_t &= \omega + \theta h_{t-1} + \phi u_t^2. \end{aligned}$$

We assume the predictive variable, x_t, follows an $AR(1)$ process

$$x_t = \rho x_{t-1} + z_t,$$

where z_t is white noise with variance σ_z^2. We regress the long-run variable, $r_t(q)$, on the past long-run predictive variable $x_{t-q}(q)$, where the long-run returns are computed using $q = 10$. This method automatically builds in a high degree of serial correlation in the error term of this regression. We consider two hypotheses: the null of no predictability where the true coefficient $\beta = 0$, and the alternative of return predictability where we set $\beta = 2.5$.

[Include Table 12 about here]

Table 12 contains the results from 5,000 Monte-Carlo replications of $T = 50, 100, 200,$ and $1,000$ observations in each replication. In Panel A we let innovations in the one-period predictive variable and the one-period return be white noise with σ_z equal to 0.026 and σ_u equal to 0.2. Under the null of no predictability the average OLS coefficient estimate and the average Bartlett coefficient estimate are both close to the true value of zero. Hence, either method seems to be able to estimate the coefficient without bias. We consider an alternative hypothesis where $\beta = 2.5$. Again, the Bartlett and the OLS coefficient estimates come close to the true value. For 50 observations in each simulation the coefficient estimates seem to be slightly downward biased, with average coefficients of 2.39 and 2.33 for the OLS and the Bartlett estimator, respectively. In larger samples with 100 observations or more, both methods seem to do very well.

Now, take a look at the standard deviation of the 5,000 coefficient estimates. Both the OLS and the Bartlett coefficient estimates becomes less volatile as the sample size increases, and they are about equally volatile. In each replication we compute the Newey-West standard error based on $q = 10$ lags. We observe that on average the Newey-West standard error does not capture the actual variation in the coefficient estimates, especially for sample sizes of less than 1,000 observations. This may lead to overestimation of the t-statistics based on Newey-West standard errors, and accordingly, the coefficient will be judged too often as significantly different from zero.

It is more appropriate to examine the size of the tests, defined as the probability that the test will reject the null hypothesis although it is true. Typically, we base our conclusion on a size of 5%. We report the size of the test in the left hand side of the table. The size of the Newey-West t-test ranges from 32% for samples of 50 observations to 10% in samples of 1,000 observations. The size of the t-tests based on the Bartlett estimates and the t-test based on HT are quite similar, and the size is about 6% for the sample sizes under consideration. This is quite close to the conventional size of 5%. Lastly, we examine the properties of a t-test based on OLS standard errors multiplied by $\sqrt{q}$ as suggested by Engle (1974). The size of this test is small and ranges between 2.5%-3.7%, making it a conservative test.

Under the alternative hypothesis we examine the size adjusted power of the tests. This measures the probability that the test will reject the null hypothesis when it in fact is false, while correcting for the actual size of the test. We would like the size-adjusted power of a test to be as high as possible. The simulation results show that the Newey-West t-statistic has slightly less power than the other tests. Note that the size adjusted power of the HT and Engle tests are identical, which is due to the fact that these test statistics are proportional by construction.

In Panel B we allow for autocorrelation in the predictive variable of $\rho = 0.7$. Interestingly, both the Bartlett and the HT t-tests maintain good size and power properties. Autocorrelation in the one-period predicting variable does not alter the conclusions from Panel A.

In Panel C we furthermore allow for GARCH(1,1) in the innovations of the short-run return. We use the parameters $\hat{\omega} = 0.002757, \hat{\theta} = 0.094634, \hat{\phi} = 0.845135$, which are obtained from fitting a GARCH(1,1) model to the annual S&P 500 index returns over the period 1871 through 2002. Interestingly, Engle's t-test has a size very close to 5% for sample sizes between 50 and 200, whereas the Bartlett and the HT t-tests have a larger size, ranging between 7% and 10%. For larger sample sizes it seems as if the HT and the Bartlett t-tests would have better size properties, while Engle's t-test would be more conservative with a size smaller than the conventional 5%. Under the alternative of predictability, the Bartlett, the HT, and Engle's t-tests all have better power properties than the Newey-West t-test, with perhaps a slight advantage to the Bartlett test. As a result of the Monte-Carlo study, we report both the OLS and the Bartlett coefficients in the empirical section. In addition, we report the size and size adjusted power properties of the t-statistics based on Newey-West, HT, and Bartlett standard errors.

Panel D reports the simulations for $\rho = 0.70$, allowing for correlation between the innovations in returns and predictive variables. We set $\Gamma_{uz}(0) = -0.018$, which is the empirical covariance between return and earnings yield innovations over the period 1871-2003. As expected, the coefficient estimates are biased upwards. We bias-adjust the coefficient estimates using Table 1 and the estimated value -5.7 for $\Gamma_{xx}(0)^{-1}\Gamma_{xz}(0)$. The bias-adjustment works well in the sense that the average bias-adjusted coefficient estimate is close to its true value. We use the bias-adjusted value of the coefficient when constructing the t-statistics. This method has similar size and power properties to those of the previous Panel.

In addition to the other statistics, we also report the size and power from conducting inference while imposing the null of no predictability by imposing $R^2 = 0$. The t_{HT}-statistics are constructed using Proposition 4 with the bias-adjusted coefficient in the numerator. The t_{HT}-statistics are constructed similarly, but with the variance under the maintained hypothesis divided by $(1 - R^2_{Bartlett})$ (see Proposition 4). As expected, imposing the null improves the size of the test. We also report the power without adjusting for the size, and find that the power properties of the tests proposed in this paper deteriorate in particular for smaller samples. This suggests that tests which impose the null might be too conservative when the hypothesis in questions is the alternative of predictability.

B Proofs

Lemma 1 *(Hannan, 1970, Theorem 9, p. 280) Let x_t be a random vector, fourth order stationary with finite fourth-order cumulants, and satisfying an absolute summability condition. If $\hat{f}_{x_i x_j}$ and $\hat{f}_{x_k x_l}$ are kernel based narrow band estimators around the zero frequency and subject to a truncation lag, then for $q \to \infty$, $T \to \infty$, with $q/T \to 0$,*

$$\lim_{T,q\to\infty} \frac{T}{q} cov\left(\hat{f}_{x_i x_j}, \hat{f}_{x_k x_l}\right) = \int_{-1}^{+1} k^2(x)\, dx \left(f_{x_i x_k} f_{x_j x_l} + f_{x_i x_l} f_{x_j x_k}\right).$$

Lemma 2 *(Hannan, 1970, Theorem 11, p. 289) Let x_t be a vector of square summable linear processes with iid innovations having finite fourth moments, and assume that*

$$\lim_{T\to\infty} \sqrt{\frac{T}{q}} \left\| f_{xx} - \mathbb{E}\hat{f}_{xx} \right\| = 0, \tag{35}$$

where $\|\cdot\|$ is the spectral norm, then $\hat{f}_{xx}$ is asymptotically distributed as follows

$$\sqrt{\frac{T}{q}} \left(\hat{f}_{xx} - f_{xx}\right) \underset{d}{\to} N(0, V),$$

where V is calculated as in Lemma 1.

Proof of Proposition 1. To derive the asymptotic distribution, first let

$$\hat{\Gamma}_{xx}(h) = \frac{1}{T} \sum_{t=1}^{T} x_t x'_{t+h}$$

and $k\left(\frac{h}{q}\right)$ be a kernel. We work with the Bartlett kernel estimate, since it is asymptotically equivalent to the OLS estimate (a proof of the equivalence is shown below). Note that

$$\begin{aligned}
\hat{\beta} &= \left(\sum_{h=-q+1}^{q-1} k\left(\frac{h}{q}\right)\hat{\Gamma}_{xx}(h)\right)^{-1} \sum_{h=-q+1}^{q-1} k\left(\frac{h}{q}\right)\hat{\Gamma}_{xr}(q+h) \\
&= \left(\sum_{h=-q+1}^{q-1} k\left(\frac{h}{q}\right)\left(\sum_{t=1}^{T} x_t x'_{t+h}\right)\right)^{-1} \sum_{h=-q+1}^{q-1} k\left(\frac{h}{q}\right)\left(\sum_{t=1}^{T} x_t r_{t+q+h}\right) \\
&= \left(\sum_{h=-q+1}^{q-1} k\left(\frac{h}{q}\right)\left(\sum_{t=1}^{T} x_t x'_{t+h}\right)\right)^{-1} \sum_{h=-q+1}^{q-1} k\left(\frac{h}{q}\right)\left(\sum_{t=1}^{T} x_t \left(x'_{t+h}\beta + u_{t+q+h}\right)\right) \\
&= \beta + \left(\sum_{h=-q+1}^{q-1} k\left(\frac{h}{q}\right)\left(\sum_{t=1}^{T} x_t x'_{t+h}\right)\right)^{-1} \sum_{h=-q+1}^{q-1} k\left(\frac{h}{q}\right)\left(\sum_{t=1}^{T} x_t u_{t+q+h}\right).
\end{aligned}$$

And therefore,

$$\begin{aligned}
\hat{\beta} - \beta &= \left(\sum_{h=-q+1}^{q-1} k\left(\frac{h}{q}\right)\left(\sum_{t=1}^{T} x_t x'_{t+h}\right)\right)^{-1} \sum_{h=-q+1}^{q-1} k\left(\frac{h}{q}\right)\left(\sum_{t=1}^{T} x_t u_{t+q+h}\right) \\
&= \hat{f}_{xx}^{-1}\hat{f}_{xu},
\end{aligned}$$

where

$$f_{xu} = \frac{1}{2\pi}\sum_{j=-\infty}^{\infty} \mathbb{E}\left[x_t^k u_{t+j}^{k+1}\right] = 0.$$

Let x_i be a vector containing the ith predicting variable. The condition (35) is verified for $q/T \to 0$ under the assumptions stated in Andrews (1991). By Lemma 2, we can find the distribution of $\widehat{f}_{x_i u}$,

$$\sqrt{\frac{T}{q}}\left(\hat{f}_{x_i u}\right) \to N\left(0, \left(\int_{-1}^{1} k^2(x)\,dx\right) V_{ii}\right),$$

for $i = 1, ..., m$, where

$$\begin{aligned}
V_{ii} &= \left(f_{uu} f_{x_i x_i} + f_{u x_i}^2\right) \\
&= f_{uu} f_{x_i x_i}.
\end{aligned}$$

We can also use Lemma 2 for any linear combination of $\hat{f}_{x_i u}$: for any sequence $\{a_i\}_{i=1}^m \sum_{i=1}^m a_i \sqrt{\frac{T}{q}}\left(\hat{f}_{x_i u}(\lambda)\right) \xrightarrow{d} N$. Application of the Cramèr-Wold device (see e.g., Brockwell and Davis (1991, p. 204)) yields:

$$\sqrt{\frac{T}{q}}\left(\hat{f}_{xu}\right) \to N\left(0, \left(\int_{-1}^{1} k^2(x)\,dx\right) V\right),$$

where the diagonal elements are given by V_{ii}. Off-diagonal elements can be found from further application of Lemma 1,

$$\begin{aligned}\left(\int_{-1}^{1} k(x)^2\,dx\right) V_{ij} &= \lim_{T,q\to\infty} \frac{T}{q} Cov\left(\hat{f}_{x_i u}, \hat{f}_{x_j u}\right) \\ &= \left(\int_{-1}^{1} k(x)^2\,dx\right)\left(f_{uu} f_{x_i x_j} + f_{u x_i} f_{u x_j}\right) \\ &= \left(\int_{-1}^{1} k(x)^2\,dx\right) f_{uu} f_{x_i x_j}.\end{aligned}$$

Combining we get

$$V = f_{uu} f_{xx}.$$

Application of the continuous mapping theorem yields

$$\sqrt{\frac{T}{q}}\left(\hat{\beta} - \beta\right) \xrightarrow{d} N\left(0, \left(\int_{-1}^{1} k^2(x)\,dx\right) f_{xx}^{-1} f_{uu}\right).$$

When using the Bartlett kernel $\left(\int_{-1}^{1} k^2(x)\,dx\right) = 2/3$, for the Parzen kernel $\left(\int_{-1}^{1} k^2(x)\,dx\right) = 0.539$, and $\left(\int_{-1}^{1} k^2(x)\,dx\right) = 2$ for the truncated kernel. ■

Proof of equivalence between the OLS and the Bartlett estimator. In Proposition 1 we show that the Bartlett estimator converges in distribution to a normal with mean β and variance given in Proposition 1. In order to show that the two estimators are equivalent we need to show that the OLS estimator converges to the same distribution function.

The OLS estimator is given by

$$\widetilde{\beta} = \widetilde{f}_{xx}^{-1} \widetilde{f}_{xr},$$

where

$$\begin{aligned}\widetilde{f}_{xr} &= \frac{1}{2\pi} \sum_{h=-q+1}^{q-1} \left(1 - \frac{|h|}{q}\right) \widetilde{\Gamma}_{xr}(h+q), \\ \widetilde{f}_{xx} &= \frac{1}{2\pi} \sum_{h=-q+1}^{q-1} \left(1 - \frac{|h|}{q}\right) \widetilde{\Gamma}_{xx}(h)\end{aligned}$$

$$\begin{aligned}\widetilde{\Gamma}_{xr}(h+q) &= \frac{1}{q-h} \sum_{\tau=1}^{q-h} \left(\frac{1}{T} \sum_{s=\tau}^{T+\tau-1} x_s r_{s+h+q}\right), \text{ for } h \geq 0 \\ &= \frac{1}{q-|h|} \sum_{\tau=|h|+1}^{q} \left(\frac{1}{T} \sum_{s=\tau}^{T+\tau-1} x_s r_{s+h+q}\right), \text{ for } h < 0\end{aligned}$$

$$\widetilde{\Gamma}_{xx}(h) = \frac{1}{q-h}\sum_{\tau=1}^{q-h}\left(\frac{1}{T}\sum_{s=\tau}^{T+\tau-1} x_s x'_{s+h}\right), \text{ for } h \geq 0$$
$$= \frac{1}{q-|h|}\sum_{\tau=|h|+1}^{q}\left(\frac{1}{T}\sum_{s=\tau}^{T+\tau-1} x_s x'_{s+h}\right), \text{ for } h < 0.$$

The Bartlett estimator (kernel estimator) is given by

$$\widehat{\beta} = \widehat{f}_{xx}^{-1}\widehat{f}_{xr},$$

where

$$\widehat{f}_{xr} = \frac{1}{2\pi}\sum_{h=-q+1}^{q-1}\left(1-\frac{|h|}{q}\right)\widehat{\Gamma}_{xr}(h+q),$$
$$\widehat{f}_{xx} = \frac{1}{2\pi}\sum_{h=-q+1}^{q-1}\left(1-\frac{|h|}{q}\right)\widehat{\Gamma}_{xx}(h),$$
$$\widehat{\Gamma}_{xr}(h+q) = \frac{1}{T}\sum_{s=1}^{T-h}\left(x_s r_{s+h+q}\right), \text{ for } h \geq 0$$
$$= \frac{1}{T}\sum_{s=|h|+1}^{T}\left(x_s r_{s+h+q}\right), \text{ for } h < 0$$
$$\widehat{\Gamma}_{xx}(h) = \frac{1}{T}\sum_{s=1}^{T-h}\left(x_s x'_{s+h}\right), \text{ for } h \geq 0$$
$$= \frac{1}{T}\sum_{s=|h|+1}^{T}\left(x_s x'_{s+h}\right), \text{ for } h < 0.$$

First, for $h \geq 0$ consider the difference between the two estimators for the covariance, where $|\cdot|$ stands for absolute value, element-by-element:

$$\left|\widetilde{\Gamma}_{xx}(h) - \widehat{\Gamma}_{xx}(h)\right|$$
$$= \left|\frac{1}{q-h}\sum_{\tau=1}^{q-h}\left(\frac{1}{T}\sum_{s=\tau}^{T+\tau-1} x_s x'_{s+h}\right) - \frac{1}{T}\sum_{s=1}^{T-h}\left(x_s x'_{s+h}\right)\right|$$
$$\leq \frac{1}{q-h}\sum_{\tau=1}^{q-h}\left|\left(\frac{1}{T}\sum_{s=\tau}^{T+\tau-1} x_s x'_{s+h}\right) - \frac{1}{T}\sum_{u=1}^{T-h} x_u x'_{u+h}\right|$$
$$= \frac{1}{q-h}\sum_{\tau=1}^{q-h}\left|-\frac{1}{T}\sum_{u=1}^{\tau-1} x_u x'_{u+h} + \left(\frac{1}{T}\sum_{s=\tau}^{T-h} x_s x'_{s+h}\right) - \frac{1}{T}\sum_{s=\tau}^{T-h} x_u x'_{u+h} + \frac{1}{T}\sum_{u=T-h+1}^{T+\tau-1} x_u x'_{u+h}\right|$$

$$
\begin{aligned}
&= \frac{1}{q-h}\sum_{\tau=1}^{q-h}\left|-\frac{1}{T}\sum_{u=1}^{\tau-1}x_u x'_{u+h} \ +\frac{1}{T}\sum_{u=T-h+1}^{T+\tau-1}x_u x'_{u+h}\right| \\
&\leq \frac{1}{q-h}\sum_{\tau=1}^{q-h}\left(\frac{1}{T}\sum_{u=1}^{\tau-1}|x_u|\left|x'_{u+h}\right| + \frac{1}{T}\sum_{u=T-h+1}^{T+\tau-1}|x_u|\left|x'_{u+h}\right|\right) \\
&\leq \frac{1}{q-h}\sum_{\tau=1}^{q-h}\left(\frac{1}{T}\sum_{u=1}^{q-h}|x_u|\left|x'_{u+h}\right| + \frac{1}{T}\sum_{u=T-h+1}^{T+q-h}|x_u|\left|x'_{u+h}\right|\right) \\
&\leq \frac{1}{T}\sum_{u=1}^{q-h}|x_u|\left|x'_{u+h}\right| + \frac{1}{T}\sum_{u=T-h+1}^{T+q-h}|x_u|\left|x'_{u+h}\right|.
\end{aligned}
$$

In the second and the fourth line we use the triangle inequality. In the third line we split up the last sum in three parts - the first runs from 1 to $\tau-1$, the second from τ to $T-h$, and the last from $T-h+1$ to $T+\tau-1$. In the sixth line, the sums are all over positive terms, so the sum is larger if we sum over more terms. Hence, we let the first sum run to the maximum value of τ, and we let the last one run to $T+q-h$. Note that in doing so, we do not have to take the average any longer, which gets us to the last equation.

Similarly, for $h<0$

$$
\begin{aligned}
\left|\widetilde{\Gamma}_{xx}(h)-\widehat{\Gamma}_{xx}(h)\right| &= \left|\frac{1}{q-|h|}\sum_{\tau=|h|+1}^{q}\left(\frac{1}{T}\sum_{s=\tau}^{T+\tau-1}x_s x'_{s+h}\right)-\frac{1}{T}\sum_{s=|h|+1}^{T}\left(x_s x'_{s+h}\right)\right| \\
&\leq \frac{1}{q-|h|}\sum_{\tau=|h|+1}^{q}\left|\left(\frac{1}{T}\sum_{s=\tau}^{T+\tau-1}x_s x'_{s+h}\right)-\frac{1}{T}\sum_{s=|h|+1}^{T}\left(x_s x'_{s+h}\right)\right| \\
&\leq \frac{1}{q-|h|}\sum_{\tau=|h|+1}^{q}\left|-\frac{1}{T}\sum_{s=|h|+1}^{\tau-1}x_s x'_{s+h}+\frac{1}{T}\sum_{s=T+1}^{T+\tau-1}x_s x'_{s+h}\right| \\
&\leq \frac{1}{q-|h|}\sum_{\tau=|h|+1}^{q}\left(\frac{1}{T}\sum_{s=|h|+1}^{\tau-1}|x_s|\left|x'_{s+h}\right|+\frac{1}{T}\sum_{s=T+1}^{T+\tau-1}|x_s|\left|x'_{s+h}\right|\right) \\
&\leq \frac{1}{q-|h|}\sum_{\tau=|h|+1}^{q}\left(\frac{1}{T}\sum_{s=|h|+1}^{q}|x_s|\left|x'_{s+h}\right|+\frac{1}{T}\sum_{s=T+1}^{T+q}|x_s|\left|x'_{s+h}\right|\right) \\
&\leq \frac{1}{T}\sum_{u=1+|h|}^{q}|x_u|\left|x'_{u+h}\right|+\frac{1}{T}\sum_{u=T+1}^{T+q}|x_u|\left|x'_{u+h}\right|.
\end{aligned}
$$

Now consider,

$$\begin{aligned}
\left|\widetilde{f}_{xx}-\widehat{f}_{xx}\right| &= \left|\frac{1}{2\pi}\sum_{h=-q+1}^{q-1}\left(1-\frac{|h|}{q}\right)\widetilde{\Gamma}_{xx}(h)-\frac{1}{2\pi}\sum_{h=-q+1}^{q-1}\left(1-\frac{|h|}{q}\right)\widehat{\Gamma}_{xx}(h)\right| \\
&\leq \frac{1}{2\pi}\sum_{h=-q+1}^{q-1}\left(1-\frac{|h|}{q}\right)\left|\widetilde{\Gamma}_{xx}(h)-\widehat{\Gamma}_{xx}(h)\right| \\
&\leq \frac{1}{2\pi}\sum_{h=-q+1}^{q-1}\left(1-\frac{|h|}{q}\right)\left|\frac{1}{T}\sum_{u=1+|h\wedge 0|}^{(q-h)\wedge q}|x_u|\left|x'_{u+h}\right|+\frac{1}{T}\sum_{u=(T-h+1)\wedge(T+1)}^{(T+q-h)\wedge(T+q)}|x_u|\left|x'_{u+h}\right|\right|,
\end{aligned}$$

where $\wedge$ denotes the minimum operator. We multiply the right hand side by $\frac{q}{q-|h|}\geq 1$,

$$\begin{aligned}
\left|\widetilde{f}_{xx}-\widehat{f}_{xx}\right| \leq\ & \frac{q}{T}\frac{1}{2\pi}\sum_{h=-q+1}^{q-1}\left(1-\frac{|h|}{q}\right)\frac{1}{q-|h|}\sum_{u=1+|h\wedge 0|}^{(q-h)\wedge q}|x_u|\left|x'_{u+h}\right| \\
&+\frac{q}{T}\frac{1}{2\pi}\sum_{h=-q+1}^{q-1}\left(1-\frac{|h|}{q}\right)\frac{1}{q-|h|}\sum_{u=(T-h+1)\wedge(T+1)}^{(T+q-h)\wedge(T+q)}|x_u|\left|x'_{u+h}\right|.
\end{aligned} \tag{36}$$

We use the fact that $\frac{1}{\sqrt{2\pi q}}\sum_{s=1}^{q}|x_s|$ is the discrete fourier transform (dft) of $|x_s|$ at the zero frequency. The process $|x_s|$ has mean $\mu>0$. From Hannan (1973, Theorem 3) we know that the dft $\omega_x(\lambda_s)=\frac{1}{\sqrt{2\pi q}}\left(\sum_{s=1}^{q}|x_s|e^{is\lambda_s}\right)$ satisfies a Central Limit Theorem for dfts of stationary processes with limit

$$\begin{aligned}
N_c(0,f_{xx}(\lambda_s))\text{ for }\lambda_s &\neq 0\text{ and} \\
N(\frac{\mu}{\sqrt{2\pi}},f_{xx}(\lambda_s))\text{ for }\lambda_s &= 0
\end{aligned}$$

and are independently distributed as $q\to\infty$.

Consequently,

$$\begin{aligned}
|\omega_x(0)|^2 &= \omega_x(0)\omega_x(0)^* = \\
&= \frac{1}{\sqrt{2\pi q}}\sum_{s=1}^{q}|x_s|\frac{1}{\sqrt{2\pi q}}\sum_{u=1}^{q}\left|x'_u\right| \\
&= \frac{1}{2\pi q}\sum_{s,u=1}^{q}|x_u|\left|x'_{u+h}\right|
\end{aligned}$$

$$= \frac{1}{2\pi q} \sum_{h=-q+1}^{q-1} \sum_{u=1}^{q} |x_u| \left|x'_{u+h}\right| \text{ where } 1 \leq u+h \leq q$$

$$= \frac{1}{2\pi} \sum_{h=-q+1}^{q-1} \left(1 - \frac{|h|}{q}\right) \frac{1}{q-|h|} \sum_{u=1+|h\wedge 0|}^{(q-h)\wedge q} |x_u| \left|x'_{u+h}\right|.$$

From probability theory we know that the square of a normal distribution with non-zero mean is a non-central Chi-square distribution, which is $O_p(1)$ (Davidson (1994, p. 187)). Therefore $\frac{q}{T}$ times $|\widetilde{\omega}_x(0)|^2$is $O_p(\frac{q}{T})$.[15] Combining these observations we obtain:

$$\left|\widetilde{f}_{xx} - \widehat{f}_{xx}\right| \leq \frac{q}{T} O_p(1) + \frac{q}{T} O_p(1) = O_p(\frac{q}{T}) = o_p(1).$$

Hence, $\sqrt{\frac{T}{q}} |\widetilde{f}_{xx} - \widehat{f}_{xx}| = \sqrt{\frac{T}{q}} O_p(\frac{q}{T}) = O_p(\sqrt{\frac{q}{T}}) = o_p(1)$. Consequently, $\sqrt{\frac{T}{q}} \widetilde{f}$ converges in distribution to the same distribution function as$\sqrt{\frac{T}{q}} \widehat{f}$ (see e.g., Hamilton (1994, Proposition 7.3)).

We can use arguments similar to the above for $\left|\widetilde{\Gamma}_{xr}(h+q) - \widehat{\Gamma}_{xr}(h+q)\right|$ and show that $\left|\widetilde{f}_{xr} - \widehat{f}_{xr}\right| = O_p(\frac{q}{T}) = o_p(1)$. Finally, the Continuous Mapping Theorem implies that $\sqrt{\frac{T}{q}} \widehat{\beta}$ and $\sqrt{\frac{T}{q}} \widetilde{\beta}$ converge to the same distribution function. ∎

Proof of Proposition 3. The difference between the coefficient estimate and the true coefficient for Model (1) is given by

$$\widehat{\beta} - \beta = \left(\sum_{h=-q+1}^{q-1} \left(1 - \frac{|h|}{q}\right) \widehat{\Gamma}_{xx}(h) \right)^{-1} \sum_{h=-q+1}^{q-1} \left(1 - \frac{|h|}{q}\right) \widehat{\Gamma}_{xu}(q+h), \tag{37}$$

$$\text{where } \widehat{\Gamma}_{xu}(q+h) = \frac{1}{T} \sum_{t=1}^{T} x_t u_{t+q+h}. \tag{38}$$

If u_{t+q+h} and z_{t+q+h} are correlated then

$$u_{t+q+h} = z'_{t+q+h} \Gamma_{zz}(0)^{-1} \Gamma_{zu}(0) + \varepsilon_{t+q+h},$$

where ε_{t+q+h} is white noise. Now substitute this into equation (38)

$$\widehat{\Gamma}_{xu}(q+h) = \frac{1}{T} \sum_{t=1}^{T} x_t z'_{t+q+h} \Gamma_{zz}(0)^{-1} \Gamma_{zu}(0) + \frac{1}{T} \sum_{t=1}^{T} x_t \varepsilon_{t+q+h}$$

and substitute this expression into (37)

[15]See Davidson (1994, p. 187) for an explanation of the O_p notation.

$$
\begin{aligned}
\widehat{\beta}-\beta &= \left(\underbrace{\sum_{h=-q+1}^{q-1}\left(1-\frac{|h|}{q}\right)\widehat{\Gamma}_{xx}(h)}_{2\pi\widehat{f}_{xx}}\right)^{-1} \times \\
&\quad \sum_{h=-q+1}^{q-1}\left(1-\frac{|h|}{q}\right)\left(\underbrace{\frac{1}{T}\sum_{t=1}^{T}x_t z'_{t+q+h}}_{\widehat{\Gamma}_{xz}(q+h)}\Gamma_{zz}(0)^{-1}\Gamma_{zu}(0)+\underbrace{\frac{1}{T}\sum_{t=1}^{T}x_t\varepsilon_{t+q+h}}_{\widehat{\Gamma}_{x\varepsilon}(q+h)}\right) \\
&= \left(2\pi\widehat{f}_{xx}\right)^{-1}2\pi\widehat{f}_{xz_q}\Gamma_{zz}(0)^{-1}\Gamma_{zu}(0)+\left(2\pi\widehat{f}_{xx}\right)^{-1}2\pi\widehat{f}_{x\varepsilon_q} \\
&= \widehat{f}_{xx}^{-1}\widehat{f}_{xz_q}\Gamma_{zz}(0)^{-1}\Gamma_{zu}(0)+\widehat{f}_{xx}^{-1}\widehat{f}_{x\varepsilon_q}.
\end{aligned}
$$

Using arguments similar to Stambaugh (1999, p. 412),

$$
\begin{aligned}
\mathbb{E}\left[\varepsilon_t|z\right] &= 0 \text{ (orthogonality)} \\
\mathbb{E}\left[\varepsilon_t|x_0,x_1,...,x_{T-1}\right] &= \mathbb{E}\left[\varepsilon_t|x_{0,}z_1,...,z_{T-1}\right]=0
\end{aligned}
$$

implies $E\left[\widehat{f}_{xx}^{-1}\widehat{f}_{x\varepsilon_q}\right]=0$ and therefore

$$
\begin{aligned}
\mathbb{E}\left[\ \widehat{\beta}-\beta\right] &= \mathbb{E}\left[\widehat{f}_{xx}^{-1}\widehat{f}_{xz_q}\Gamma_{zz}(0)^{-1}\Gamma_{zu}(0)\right] \\
&= \mathbb{E}\left[\widehat{f}_{xx}^{-1}\widehat{f}_{xz_q}\right]\Gamma_{zz}(0)^{-1}\Gamma_{zu}(0).
\end{aligned}
$$

■

Proof of Proposition 6. First consider

$$
\begin{aligned}
\widehat{\beta}-\beta &= \left(\sum_{h=-q+1}^{q-1}\left(1-\frac{|h|}{q}\right)\widehat{\Gamma}_{xx}(h)\right)^{-1}\sum_{h=-[q/2]+1}^{[q/2]-1}\widehat{\Gamma}_{xu}([q/2]+h) \\
&= \frac{\widehat{f}_{xr}}{\widehat{f}_{xx}}-\frac{f_{xr}}{f_{xx}} \\
&= \frac{\widehat{f}_{xr}f_{xx}-\widehat{f}_{xx}f_{xr}}{\widehat{f}_{xx}f_{xx}}.
\end{aligned}
$$

Now

$$
V\left(\widehat{\beta}-\beta\right)=\frac{V(\widehat{f}_{xr})f_{xx}^2+V(\widehat{f}_{xx})f_{xr}^2-2Cov(\widehat{f}_{xx},\widehat{f}_{xr})f_{xx}f_{xr}}{f_{xx}^4}.
$$

From Lemma 2, we obtain

$$\begin{aligned}
V(\widehat{f}_{xx}) &= \int_{-1}^{1} k_B^2(x)dx2f_{xx}^2 \\
&= \int_{-1}^{1} (1-|x|)^2 dx2f_{xx}^2 \\
&= \frac{4}{3}f_{xx}^2 \\
V(\widehat{f}_{xr}) &= \int_{-.5}^{.5} k_T^2(x)dx\left(f_{xx}f_{rr}+f_{xr}^2\right) \\
&= \int_{-.5}^{.5} 1^2 dx\left(f_{xx}f_{rr}+f_{xr}^2\right) \\
&= \left(f_{xx}f_{rr}+f_{xr}^2\right) \\
Cov(\widehat{f}_{xx},\widehat{f}_{xr}) &= \int k_T(x)k_B(x)dx\left(f_{xx}f_{xr}+f_{xr}f_{xx}\right) \\
&= \int_{0}^{1} (1-|x|)dx2f_{xx}f_{xr} \\
&= \frac{1}{2}2f_{xx}f_{xr} = f_{xx}f_{xr}.
\end{aligned}$$

Now insert these variances and covariances into the expression above and rearrange

$$\begin{aligned}
V\left(\widehat{\beta}-\beta\right) &= \frac{\left(f_{xx}f_{rr}+f_{xr}^2\right)f_{xx}^2+\frac{4}{3}f_{xx}^2f_{xr}^2-2\left(f_{xx}f_{xr}\right)f_{xx}f_{xr}}{f_{xx}^4} \\
&= f_{xx}^{-2}\left(f_{xx}f_{rr}+\frac{1}{3}f_{xr}^2\right).
\end{aligned}$$

■

Proof of Proposition 7. Model (1) is more efficient than Model (2) if the difference between the variance covariance matrices of Model (1) and Model (2)

$$\Gamma_{xx}(0)^{-2}\left(\Omega_{xx}\omega_{rr}+\omega_{xr}\omega_{xr}'\right)-\frac{2}{3}\Omega_{xx}^{-2}\left(\Omega_{xx}\omega_{rr}-\Omega_{xx}\omega_{xr}'\Omega_{xx}^{-1}\omega_{xr}\right)$$

is positive semi-definite (PSD) (cf. Hamilton (1994, p. 741)). If

$$\frac{2}{3}\Omega_{xx}^{-2}\leq\Gamma_{xx}(0)^{-2}$$

then we can show that the difference between the variance-covariance matrix for Model (1) and

Model (2) is a positive semi-definite matrix:

$$\begin{aligned} & \left(\Omega_{xx}\omega_{rr} + \omega_{xr}\omega'_{xr}\right) - \left(\Omega_{xx}\omega_{rr} - \Omega_{xx}\omega'_{xr}\Omega_{xx}^{-1}\omega_{xr}\right) \\ = \ & \omega_{xr}\omega'_{xr} + \Omega_{xx}\omega'_{xr}\Omega_{xx}^{-1}\omega_{xr} \\ = \ & \underbrace{\Omega_{xx}}_{\text{PSD}}\underbrace{\omega'_{xr}\Omega_{xx}^{-1}\omega_{xr}}_{>0} + \underbrace{\omega_{xr}\omega'_{xr}}_{\text{PSD}}. \end{aligned}$$

The last line follows from the fact that Ω_{xx} and its inverse are symmetric and positive semi-definite matrices: $\omega'_{xr}\Omega_{xx}^{-1}\omega_{xr} \geq 0$ for any vector ω_{xr}. Note that when the inverse matrix exists (determinant is non-zero) then $\omega'_{xr}\Omega_{xx}^{-1}\omega_{xr} > 0$, i.e., Ω_{xx}^{-1} and Ω_{xx} are positive definite. ■

Proof of Proposition 9. Model (2*) is more efficient than Model (2) if

$$\Gamma_{xx}(0)^{-2}\left(\Omega_{xx}\omega_{rr} + \omega_{xr}\omega'_{xr}\right) - \Omega_{xx}^{-2}\left(\Omega_{xx}\omega_{rr} + \frac{1}{3}\omega_{xr}\omega'_{xr}\right) \geq 0.$$

Note that the terms, $(\Omega_{xx}\omega_{rr} + \omega_{xr}\omega'_{xr})$ and $\left(\Omega_{xx}\omega_{rr} + \frac{1}{3}\omega_{xr}\omega'_{xr}\right)$ are positive since they are sums of variances. Note furthermore that the first term is larger than the second term:

$$\begin{aligned} & \left(\Omega_{xx}\omega_{rr} + \omega_{xr}\omega'_{xr}\right) - \left(\Omega_{xx}\omega_{rr} + \frac{1}{3}\omega_{xr}\omega'_{xr}\right) \\ = \ & \omega_{xr}\omega'_{xr} - \frac{1}{3}\omega_{xr}\omega'_{xr} \\ = \ & \frac{2}{3}\omega_{xr}\omega'_{xr} \geq 0. \end{aligned}$$

If $\Gamma_{xx}(0)^{-1}\Omega_{xx} \geq 1$ or equivalently $\Gamma_{xx}(0)^{-2} \geq \Omega_{xx}^{-2}$ then it follows that

$$\Gamma_{xx}(0)^{-2}\left(\Omega_{xx}\omega_{rr} + \omega_{xr}\omega'_{xr}\right) - \Omega_{xx}^{-2}\left(\Omega_{xx}\omega_{rr} + \frac{1}{3}\omega_{xr}\omega'_{xr}\right) \geq 0.$$

■

References

Andrews, D. W. (1991), 'Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation', *Econometrica* **59**(3), 817–858.

Andrews, D. W. (2004), 'The Block-Block Bootstrap: Improved Asymptotic Refinements', *Econometrica* **72**(3), 673–700.

Ang, A. and Bekaert, G. (2001), 'Stock Return Predictability: Is it there?', NBER Working Paper No. 8207.

Bartlett, M. S. (1950), 'Periodogram Analysis and Continuous Spectra', *Biometrika* **37**, 1–16.

Bollerslev, T. and Hodrick, R. J. (1996), *Handbook of Applied Econometrics, Macroeconomics*, Vol. 1, Basil Blackwell, pp. 415–458.

Brockwell, P. J. and Davis, R. A. (1991), *Time Series: Theory and Methods*, 2 edn, Springer-verlag.

Campbell, J., Lo, A. and MacKinlay, A. (1997), *The Econometrics of Financial Markets*, Princeton University Press.

Campbell, J. and Mankiw, N. (1987), 'Are Output Fluctuations Transitory?', *The Quarterly Journal of Economics* **102**(4), 857–880.

Campbell, J. and Shiller, R. (1988*a*), 'Stock Prices, Earnings and Expected Dividends', *Journal of Finance* **43**(3), 661–676.

Campbell, J. and Shiller, R. (1988*b*), 'The Dividend-Price Ration and Small Sample Bias', *Economics Letters* **29**, 325–331.

Campbell, J. and Shiller, R. (1998), 'Valuation Ratios and the Long-Run Stock Market Outlook', *Journal of Portfolio Management* **24**(2), 11–26.

Campbell, J. and Shiller, R. (2001), 'Valuation Ratios and the Long-Run Stock Market Outlook: An Update', Cowles Foundation Discussion Paper No. 1295, Yale University.

Campbell, J. Y. and Yogo, M. (2003), 'Efficient Tests of Stock Return Predictability', NBER Working paper No. 10026.

Cochrane, J. (2001), *Asset Pricing*, Princeton University Press.

Corbae, D., Ouliaris, S. and Phillips, P. C. (2002), 'Band Spectral Regression with Trending Data', *Econometrica* **70**(3), 1067–1109.

Davidson, J. (1994), *Stochastic Limit Theory*, Oxford University Press.

Engle, R. F. (1974), 'Band Spectrum Regression', *International Economic Review* **15**(1), 1–11.

Fama, E. F. and French, K. R. (1989), 'Dividend Yields and Expected Stock Returns', *Journal of Financial Economics* **22**(1), 3–25.

Fama, E. F. and French, K. R. (2001), 'Disappearing dividends: changing firm characteristics or lower propensity to pay', *Journal of Financial Economics* **60**, 3–43.

Ferson, W. E., Sarkissian, S. and Simin, T. T. (2003), 'Spurious Regressions in Financial Economics?', *Journal of Finance* **58**(4), 1393–1413.

Friedman, M. and Schwartz, A. J. (1982), *Monetary Trends in the United States and the United Kingdom*, National Bureau of Economic Research, University of Chicago Press.

Goetzmann, W. N. and Jorion, P. (1993), 'Testing the Predictive Power of Dividend Yields', *Journal of Finance* **48**(2), 663–679.

Goyal, A. and Welch, I. (2003), 'Predicting the Equity Premium With Dividend Ratios', *Management Science* **49**(5), 639–654.

Granger, C. and Newbold, P. (1974), 'Spurious Regressions in Econometrics', *Journal of Econometrics* **74**, 111–120.

Hamilton, J. D. (1994), *Time Series Analysis*, Princeton University Press.

Hannan, E. (1970), *Multiple Time Series*, John Wiley and Sons, Inc.

Hannan, E. (1973), 'Central Limit Therorems for Times Series Regressions', *Zeitschrift fur Wahrscheinlichkeitstheorie und verwandte Gebiete* **26**, 157–170.

Hansen, C. S. and Tuypens, B. (2004), 'Proxying Expected Returns with Price Earnings Ratios', Working Paper.

Hansen, L. P. and Hodrick, R. J. (1980), 'Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis', *The Journal of Political Economy* **88**(5), 829–853.

Hodrick, R. (1992), 'Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement', *Review of Financial Studies* **5**(3), 357–386.

Lamont, O. (1998), 'Earnings and Expected Returns', *Journal of Finance* **53**(5), 1563–1587.

Lanne, M. (2002), 'Testing the Predictability of Stock Returns', *Review of Economics and Statistics* **84**, 407 – 415.

Lettau, M. and Ludvigson, S. (2001), 'Consumption, Aggregate Wealth, and Expected Stock Returns', *Journal of Finance* **56**(3), 815–849.

Lewellen, J. (2004), 'Predicting Returns with Financial Ratios', *Forthcoming in Journal of Financial Economics* .

Modigliani, F. and Miller, M. H. (1958), 'The Cost of Capital, Corporation Finance and the Theory of Investment', *The American Economic Review* **48**(3), 261–297.

Modigliani, F. and Miller, M. H. (1961), 'Dividend Policy, Growth and the Valuation of Shares', *The Journal of Business* **34**(4), 411–33.

Nelson, C. and Kim, M. (1993), 'Predictable Stock Returns: The Role of Small Sample Bias', *Journal of Finance* **48**(2), 641–661.

Newey, W. K. and West, K. D. (1987), 'A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix', *Econometrica* **55**(3), 703–708.

Newey, W. K. and West, K. D. (1994), 'Automatic Lag Selection in Covariance Matrix Estimation', *The Review of Economic Studies* **61**(4), 631–653.

Phillips, P. C. (1987), 'Time Series with a Unit Root', *Econometrica* **55**, 277–301.

Phillips, P. C. (1991), 'Spectral Regression for cointegrated time series', *Nonparametric and semiparametric methods in econometrics and statistics: Proceedings of the fifth international symposium in economic theory and econometrics* pp. 413–435.

Phillips, P. C. and Moon, H. (2001), 'Linear Regression Limit Theory for Nonstationary Panel Data', *Econometrica* **67**(5).

Phillips, P. C., Moon, H. and Xiao, Z. (2001), 'How to Estimate Autoregressive Roots Near Unity', *Econometric Theory* **17**, 29–69.

Politis, D. and Romano, J. (1995), 'Bias-Corrected Non-Parametric Spectral Estimation', *J. Time Series Analysis* **16**, 67–104.

Politis, D., Romano, J. and Wolf, M. (1999), *Subsampling*, Springer-Verlag.

Pontif, J. and Schall, L. D. (1998), 'Book-to-Market Ratios as Predictors of Market Returns', *Journal of Financial Economics* **49**(2), 141–160.

Richardson, M. and Smith, T. (1991), 'Tests of Financial Models in the Presence of Overlapping Observations', *The Review of Financial Studies* **4**(2), 227–254.

Richardson, M. and Stock, J. (1989), 'Drawing Inferences from Statistics Based on Multiyear Asset Returns', *Journal of Financial Economics* **25**, 323–348.

Sargent, T. J. (1979), *Macroeconomic Theory*, Academic Press.

Scholes, M. and Black, F. (1974), 'The Effects of Dividend Yield and Dividend Policy on Common Stock Prices and Returns', *Journal of Financial Economics* **1**, 1–22.

Shiller, R. (2000), *Irrational Exuberance*, Princeton University Press.

Stambaugh, R. F. (1986), 'Bias in Regressions with Lagged Stochastic Regressors', Working Paper, University of Chicago, Chicago, IL.

Stambaugh, R. F. (1999), 'Predictive Regressions', *Journal of Financial Economics* **54**.

Torous, W., Valkanov, R. and Yan, S. (2005), 'On Predicting Stock Returns with Nearly Integrated Explanatory Variables', *Journal of Business (forthcoming)* **78**(1).

Valkanov, R. (2003), 'Long-Run Regressions: Theoretical Results and Applications', *Jorunal of Financial Economics* **68**, 201–232.

Wolf, M. (2000), 'Stock Returns and Dividend Yields Revisited: A New Look at an Old Problem', *Journal of Business Economics and Statistics* **18**(1), 18–30.

Table 1: Simulated moment of $f_{xx}^{-1} f_{xz_q}$.

$T =$	50	100	130	200	1000
$\rho = 0.70$					
$-(1+3\rho)/T$	-0.062	-0.031	-0.024	-0.016	-0.003
$\mathbb{E}[f_{xx}^{-1} f_{xz_1}]$	-0.054	-0.029	-0.022	-0.015	-0.003
$\mathbb{E}[f_{xx}^{-1} f_{xz_2}]$	-0.060	-0.032	-0.024	-0.017	-0.003
$\mathbb{E}[f_{xx}^{-1} f_{xz_5}]$	-0.071	-0.039	-0.029	-0.020	-0.004
$\mathbb{E}[f_{xx}^{-1} f_{xz_8}]$	-0.085	-0.046	-0.036	-0.024	-0.005
$\mathbb{E}[f_{xx}^{-1} f_{xz_{10}}]$	-0.095	-0.052	-0.041	-0.027	-0.005
$\rho = 0.80$					
$-(1+3\rho)/T$	-0.068	-0.034	-0.026	-0.017	-0.003
$\mathbb{E}[f_{xx}^{-1} f_{xz_1}]$	-0.061	-0.031	-0.025	-0.016	-0.003
$\mathbb{E}[f_{xx}^{-1} f_{xz_2}]$	-0.065	-0.034	-0.027	-0.017	-0.004
$\mathbb{E}[f_{xx}^{-1} f_{xz_5}]$	-0.070	-0.038	-0.029	-0.020	-0.004
$\mathbb{E}[f_{xx}^{-1} f_{xz_8}]$	-0.075	-0.042	-0.033	-0.022	-0.005
$\mathbb{E}[f_{xx}^{-1} f_{xz_{10}}]$	-0.079	-0.045	-0.036	-0.024	-0.005
$\rho = 0.90$					
$-(1+3\rho)/T$	-0.074	-0.037	-0.028	-0.019	-0.004
$\mathbb{E}[f_{xx}^{-1} f_{xz_1}]$	-0.069	-0.036	-0.028	-0.018	-0.004
$\mathbb{E}[f_{xx}^{-1} f_{xz_2}]$	-0.071	-0.037	-0.029	-0.019	-0.004
$\mathbb{E}[f_{xx}^{-1} f_{xz_5}]$	-0.070	-0.039	-0.030	-0.019	-0.004
$\mathbb{E}[f_{xx}^{-1} f_{xz_8}]$	-0.069	-0.039	-0.030	-0.020	-0.004
$\mathbb{E}[f_{xx}^{-1} f_{xz_{10}}]$	-0.068	-0.040	-0.031	-0.020	-0.005

Table 1 continues on the next page

Table 1 continued

$T =$	50	100	130	200	1000
$\rho = 0.99$					
$-(1+3\rho)/T$	-0.079	-0.040	-0.031	-0.020	-0.004
$\mathbb{E}[f_{xx}^{-1} f_{xz_1}]$	-0.083	-0.045	-0.036	-0.023	-0.004
$\mathbb{E}[f_{xx}^{-1} f_{xz_2}]$	-0.083	-0.045	-0.036	-0.023	-0.004
$\mathbb{E}[f_{xx}^{-1} f_{xz_5}]$	-0.076	-0.044	-0.035	-0.023	-0.004
$\mathbb{E}[f_{xx}^{-1} f_{xz_8}]$	-0.068	-0.041	-0.034	-0.022	-0.004
$\mathbb{E}[f_{xx}^{-1} f_{xz_{10}}]$	-0.062	-0.040	-0.033	-0.022	-0.004

This table reports the simulated value of $\mathbb{E}[f_{xx}^{-1} f_{xz_q}]$ for $q = 1, 2, 5, 8, 10$. This term is needed in order to calculate the bias in the regression coefficient estimates for the predictive regression Model (1). The expression for the bias is given in eq. (22). For $q = 1$ the formula coincides with Stambaugh's bias formula eq. (21). The results are from 10,000 Monte-Carlo replications of $T = 50, 100, 130, 200, 1000$ observations in each replication. We report the simulations for autocorrelation of $\rho = 0.70, 80, 90, 99$. For comparison we also report the approximation $-(1+3\rho)/T$.

Table 2: Bias in regression coefficients relative to β (in percentage).

	$\hat{\beta}$	Sample size				
		50	100	130	200	1000
$\rho = 0.70$						
Stambaugh $q = 1$	0.29	-1152.0	138.0	77.3	44.7	6.9
Model (1) $q = 2$	1.12	46.3	20.7	14.9	9.8	1.8
Model (1) $q = 5$	1.63	34.7	16.3	11.9	7.8	1.5
Model (1) $q = 8$	2.18	30.0	14.3	10.7	6.8	1.3
Model (1) $q = 10$	2.27	33.0	15.6	12.0	7.5	1.4
$\rho = 0.80$						
Stambaugh $q = 1$	0.29	-542.2	167.8	105.8	48.7	7.5
Model (1) $q = 2$	1.12	52.2	21.6	16.8	10.1	2.1
Model (1) $q = 5$	1.63	34.1	15.8	12.0	7.7	1.5
Model (1) $q = 8$	2.18	25.5	12.8	9.8	6.4	1.3
Model (1) $q = 10$	2.27	25.7	13.3	10.2	6.5	1.3
$\rho = 0.90$						
Stambaugh $q = 1$	0.29	-356.2	275.7	130.8	59.8	7.8
Model (1) $q = 2$	1.12	59.7	24.7	18.0	11.1	2.0
Model (1) $q = 5$	1.63	34.3	16.3	12.0	7.6	1.5
Model (1) $q = 8$	2.18	22.8	11.9	9.0	5.7	1.2
Model (1) $q = 10$	2.27	21.3	11.5	8.8	5.6	1.2

Table 2 continues on the next page

Table 2 continued

	$\hat{\beta}$	Sample size				
		50	100	130	200	1000
$\rho = 0.99$						
Stambaugh $q = 1$	0.29	-245.6	1098.1	262.9	85.6	9.6
Model (1) $q = 2$	1.12	78.2	31.6	23.4	13.8	2.3
Model (1) $q = 5$	1.63	38.2	18.9	14.5	9.0	1.6
Model (1) $q = 8$	2.18	22.4	12.7	10.0	6.4	1.2
Model (1) $q = 10$	2.27	19.2	11.5	9.3	6.1	1.1

This table reports the biases (in percentage of the bias-adjusted β coefficient, $\mathbb{E}[\hat{\beta} - \beta]/(\hat{\beta} - \mathbb{E}[\hat{\beta} - \beta])$) in the regression coefficient estimates for the predictive regression Model (1) and Stambaugh's predictive regressions. The biases involve simulations of the moments. The simulations are based on the data for S&P 500 index returns and log earnings yield over the period 1871-2003 with $\Gamma_{zz}(0)^{-1}\Gamma_{uz}(0) = -5.7$. The results are from 10,000 Monte-Carlo replications of $T = 50, 100, 130, 200, 1000$ observations in each replication. We report the bias for autocorrelation of $\rho = 0.70, 80, 90, 99$. For comparison we also report the OLS coefficients, $\hat{\beta}$, for the two regression models.

Table 3: Regression of the S&P 500 q-year real return on the long-run component of the earnings yield for different horizons.

Horizon q	$\hat{\beta}_{OLS}$	t-stat$_{HT}$	t-stat$_{NW}$	R^2
3-years	1.198	2.005*	1.928*	0.060
5-years	1.625	2.570*	3.237**	0.153
7-years	2.070	3.437**	4.404**	0.318
10-years	2.271	3.694**	4.875**	0.448
12-years	2.054	2.902**	4.278**	0.384
15-years	1.288	1.353	2.842**	0.152

** means significance at the 1% level; * significance at the 5% level.

This table reports estimates of parameters for the predictive regression Model (1), where the S&P 500 q-period long-run real log-return regressed on the q-period long-run component of the earnings yield for q=3, 5, 7, 10, 12, and 15 years. All data are in real terms. The sample period covers 1871-2003. $\hat{\beta}_{OLS}$ is the OLS estimate of the slope coefficient. The t-stat$_{HT}$ is defined in (17) and t-stat$_{NW}$ denotes t-statistic based on the Newey-West (1987) standard errors with q lags.

Table 4: Descriptive Statistics.

		Mean	Stdev	ρ	PP	KPSS
S&P 500 index	r_t	0.068	0.200	0.010	-11.280	0.025
	e/p_t	0.074	0.026	0.721	-4.523	0.090
	d/p_t	0.045	0.015	0.746	-4.052	0.492*
US Aaa Corporate bond index	r_t	0.042	0.125	0.212	-7.440	0.294
US Treasury bond index	r_t	0.027	0.087	0.243	-7.279	0.175
US T-bill bond index	r_t	0.002	0.065	0.387	-5.789	0.702*

Summary statistics of one-year real log returns, r_t, one-year log earnings yield, $e/p_t = \ln(1 + E_t/P_t)$, and one year dividend-price ratio, $d/p_t = \ln(1 + D_t/P_t)$. E_t, D_t, and P_t are the earnings, the dividends, and the price levels, all variables are adjusted for inflation. We let ρ denote the first order autocorrelation coefficient. PP is the Phillips-Perron unit root test and KPSS is the Kwiatkowski-Phillips-Schmidt-Shin stationarity test. The * indicates that we reject stationarity at the 5% confidence level. The sample period for the S&P 500 index covers 1871 to 2003. The sample period for corporate bonds and long term Treasury bonds is from 1920-2003, the 3-months Treasury bills sample period is from 1934-2003.

Table 5: Regression of the S&P 500 10-year real return on the long-run component of the earnings yield.

Panel A: Results under the maintained hypothesis, not bias-adjusted

Sample	$\hat{\beta}_{OLS}$	$\hat{\beta}_{Bartlett}$	t-stat$_{HT}$	t-stat$_{Bartlett}$	t-stat$_{NW}$	R^2
1871-2003	2.271	2.278	3.694**	3.855**	4.875**	0.448
1920-1990	2.882	2.797	3.967**	4.129**	7.585**	0.603
1950-2003	2.858	2.850	4.122**	4.448**	5.656**	0.685

Panel B: Bias-adjusted coefficients, under the maintained hypothesis

Sample	$\hat{\beta}_{OLS}$	$\hat{\beta}_{Bartlett}$	t-stat$_{HT}$	t-stat$_{Bartlett}$	t-stat$_{NW}$	R^2
1871-2003	2.035	2.043	3.312**	3.457**	4.370**	0.448
1920-1990	2.587	2.502	3.560**	3.694**	6.809**	0.603
1950-2003	2.314	2.306	3.338**	3.599**	4.580**	0.685

Panel C: Bias-adjusted coefficients, imposing the null hypothesis

Sample	$\hat{\beta}_{OLS}$	$\hat{\beta}_{Bartlett}$	t-stat$_{HTH_0}$	t-stat$_{BartlettH_0}$	t-stat$_{NWH_0}$	R^2
1871-2003	2.035	2.043	2.460*	2.568*	3.246**	0.448
1920-1990	2.587	2.502	2.243*	2.327*	4.289**	0.603
1950-2003	2.314	2.306	1.872*	2.019*	2.569*	0.685

** means significance at the 1% level; * significance at the 5% level.

This table reports estimates of parameters for the predictive regression Model (1), where the S&P 500 10-year real log-return regressed on the 10-year long-run component of the earnings yield. All data are in real terms. $\hat{\beta}_{OLS}$ is the OLS estimate of the slope coefficient and $\hat{\beta}_{Bartlett}$ is the Bartlett estimate of the slope coefficient. In Panels B and C we report the bias-adjusted slope coefficient estimates using the bias formula (22). The t-stat$_{Bartlett}$ are t-statistics computed from the asymptotical correct standard errors derived in Proposition 1 and t-stat$_{HT}$ is defined in (17). t-stat$_{NW}$ denotes t-statistic based on the Newey-West (1987) standard errors with 10 lags. t-stat$_{HTH_0}$ and (27) denote the t-test while imposing the null.

Table 6: Regression of the 10-year real return on the long-run component of the dividend price ratio.

Panel A: Results under the maintained hypothesis, not bias-adjusted

Sample	$\hat{\beta}_{OLS}$	$\hat{\beta}_{Bartlett}$	t-stat$_{HT}$	t-stat$_{Bartlett}$	t-stat$_{NW}$	R^2
1871-2003	2.526	2.394	1.887*	1.955*	2.362**	0.175
1920-1990	3.379	3.373	2.274*	2.298*	3.587**	0.333
1950-2003	4.409	3.664	2.695**	2.227*	3.848**	0.482

Panel B: Bias-adjusted coefficients, under the maintained hypothesis

Sample	$\hat{\beta}_{OLS}$	$\hat{\beta}_{Bartlett}$	t-stat$_{HT}$	t-stat$_{Bartlett}$	t-stat$_{NW}$	R^2
1871-2003	1.912	1.779	1.428	1.453	1.787*	0.175
1920-1990	2.607	2.601	1.755*	1.772*	2.768**	0.333
1950-2003	2.988	2.243	1.826*	1.363	2.607**	0.482

Panel C: Bias-adjusted coefficients, imposing the null hypothesis

Sample	$\hat{\beta}_{OLS}$	$\hat{\beta}_{Bartlett}$	t-stat$_{HTH_0}$	t-stat$_{BartlettH_0}$	t-stat$_{NWH_0}$	R^2
1871-2003	1.912	1.779	1.297	1.320	1.623	0.175
1920-1990	2.607	2.601	1.433	1.447	2.261**	0.333
1950-2003	2.988	2.243	1.314	0.981	1.876*	0.482

** means significance at the 1% level; * significance at the 5% level.

This table reports estimates of parameters for the predictive regression Model (1), where the S&P 500 10-year real log-return regressed on the 10-year long-run component of the log-dividend-price ratio (dividend yield). $\hat{\beta}_{OLS}$ is the OLS estimate of the slope coefficient and $\hat{\beta}_{Bartlett}$ is the Bartlett estimate of the slope coefficient. In Panels B and C we report the bias-adjusted slope coefficient estimates using the bias formula (22). The t-stat$_{Bartlett}$ are t-statistics computed from the asymptotical correct standard errors derived in Proposition 1 and t-stat$_{HT}$ is defined in (17). t-stat$_{NW}$ denotes t-statistic based on the Newey-West (1987) standard errors with 10 lags. t-stat$_{HTH_0}$ and (27) denote the t-test while imposing the null.

Table 7: Regression of 10-year real returns on the past earnings yield or the past dividend yield.

Panel A: Past dividend yield

Sample	$\hat{\beta}_{OLS}$	$\hat{\beta}_{Trunc}$	t-stat$_{Trunc}$	t-stat$_{NW}$	R^2
1871-2003	1.429	1.176	1.497	2.252*	0.14
1920-1990	2.088	2.187	1.984*	3.026**	0.29
1950-2003	2.320	2.589	1.321	2.162*	0.25

Panel B: Past earnings yield

Sample	$\hat{\beta}_{OLS}$	$\hat{\beta}_{Trunc}$	t-stat$_{Trunc}$	t-stat$_{NW}$	R^2
1871-2003	1.022	0.974	2.122*	5.167**	0.23
1920-1990	1.288	1.310	2.379*	6.856**	0.40
1950-2003	1.036	1.024	1.222	2.616**	0.24

** means significance at the 1% level; * significance at the 5% level.

This table reports estimates of parameters for the predictive regression Model (2), where the S&P 500 10-year real log-return is regressed on the past dividend yield in Panel A and on the past earnings yield in Panel B. All data are in real terms. $\hat{\beta}_{OLS}$ is the OLS estimate of the slope coefficient and $\hat{\beta}_{Trunc}$ is the Truncated kernel estimate of the slope coefficient. The t-stat$_{Trunc}$ are computed from the asymptotical correct standard errors derived in Proposition 2. t-stat$_{NW}$ denotes t-statistic based on the Newey-West (1987) standard errors with 10 lags.

Table 8: Regression of 10-year earnings growth on the long-run component of the earnings yield.

Panel A: Results under the maintained hypothesis, not bias-adjusted

Sample	$\hat{\beta}_{OLS}$	$\hat{\beta}_{Bartlett}$	t-stat$_{HT}$	t-stat$_{Bartlett}$	t-stat$_{NW}$	R^2
1871-2003	0.173	0.076	0.287	0.129	0.506	0.005
1920-1990	0.491	0.526	0.638	0.748	1.748*	0.038
1950-2003	0.593	0.632	1.049	1.123	2.676**	0.126

Panel B: Bias-adjusted coefficients, under the maintained hypothesis

Sample	$\hat{\beta}_{OLS}$	$\hat{\beta}_{Bartlett}$	t-stat$_{HT}$	t-stat$_{Bartlett}$	t-stat$_{NW}$	R^2
1871-2003	0.458	0.361	0.758	0.610	1.336	0.005
1920-1990	0.848	0.883	1.102	1.257	3.020**	0.038
1950-2003	1.251	1.291	2.215*	2.293*	5.648**	0.126

Panel C: Bias-adjusted coefficients, imposing the null

Sample	$\hat{\beta}_{OLS}$	$\hat{\beta}_{Bartlett}$	t-stat$_{HTH_0}$	t-stat$_{BartlettH_0}$	t-stat$_{NW}$	R^2
1871-2003	0.458	0.361	0.757	0.608	1.332	0.005
1920-1990	0.848	0.883	1.081	1.233	2.962**	0.038
1950-2003	1.251	1.291	2.071*	2.144*	5.280**	0.126

** means significance at the 1% level; * significance at the 5% level.

This table reports estimates of parameters for the predictive regression Model (1), where the 10-year earnings growth is regressed on the long-run component of the earnings yield. $\hat{\beta}_{OLS}$ is the OLS estimate of the slope coefficient and $\hat{\beta}_{Bartlett}$ is the Bartlett estimate of the slope coefficient. In Panels B and C we report the bias-adjusted slope coefficient estimates using the bias formula (22). The t-stat$_{Bartlett}$ are t-statistics computed from the asymptotical correct standard errors derived in Proposition 1 and t-stat$_{HT}$ is defined in (17). t-stat$_{NW}$ denotes t-statistic based on the Newey-West (1987) standard errors with 10 lags. t-stat$_{HTH_0}$ and (27) denote the t-test while imposing the null.

Table 9: Regression of 10-year real Aaa bond returns on the long-run component of S&P 500 earnings yield.

Panel A: Results under the maintained hypothesis, not bias-adjusted

Sample	$\hat{\beta}_{OLS}$	$\hat{\beta}_{Bartlett}$	t-stat$_{HT}$	t-stat$_{Bartlett}$	t-stat$_{NW}$	R^2
1920-2003	2.027	1.814	2.450*	2.135*	3.217**	0.357
1920-1990	1.723	2.081	2.099*	2.439*	2.740**	0.332
1950-2003	1.790	1.529	1.727*	1.595	2.129*	0.277

Panel B: Bias-adjusted coefficients, under the maintained hypothesis

Sample	$\hat{\beta}_{OLS}$	$\hat{\beta}_{Bartlett}$	t-stat$_{HT}$	t-stat$_{Bartlett}$	t-stat$_{NW}$	R^2
1920-2003	1.965	1.752	2.375*	2.062*	3.119**	0.357
1920-1990	1.646	2.004	2.005*	2.348*	2.617**	0.332
1950-2003	1.647	1.386	1.589*	1.445	1.959*	0.277

Panel C: Bias-adjusted coefficients, imposing the null hypothesis

Sample	$\hat{\beta}_{OLS}$	$\hat{\beta}_{Bartlett}$	t-stat$_{HTH_0}$	t-stat$_{BartlettH_0}$	t-stat$_{NWH_0}$	R^2
1920-2003	1.965	1.752	1.904*	1.653*	2.501*	0.357
1920-1990	1.646	2.004	1.638*	1.918*	2.138*	0.332
1950-2003	1.647	1.386	1.351	1.229	1.666*	0.277

** means significance at the 1% level; * significance at the 5% level.

This table reports estimates of parameters for the predictive regression Model (1), where the 10-year bond return is regressed on the long-run component of the S&P 500 earnings yield. All data are in real terms. $\hat{\beta}_{Bartlett}$ is the Bartlett estimate of the slope coefficient. In Panels B and C we report the bias-adjusted slope coefficient estimates using the bias formula (22). The t-stat$_{Bartlett}$ are t-statistics computed from the asymptotical correct standard errors derived in Proposition 1 and t-stat$_{HT}$ is defined in (17). t-stat$_{NW}$ denotes t-statistic based on the Newey-West (1987) standard errors with 10 lags. t-stat$_{HTH_0}$ and (27) denote the t-test while imposing the null.

Table 10: Regression of long duration treasury note 10-year real returns on the long-run component of S&P 500 earnings yield.

Panel A: Results under the maintained hypothesis, not bias-adjusted

Sample	$\hat{\beta}_{OLS}$	$\hat{\beta}_{Bartlett}$	$t\text{-stat}_{HT}$	$t\text{-stat}_{Bartlett}$	$t\text{-stat}_{NW}$	R^2
1920-2003	1.531	1.441	3.238*	2.701**	4.850**	0.489
1920-1990	1.487	1.578	2.790**	2.669**	4.122**	0.464
1950-2003	1.214	1.224	2.127*	2.263*	3.508**	0.367

Panel B: Bias-adjusted coefficients, under the maintained hypothesis

Sample	$\hat{\beta}_{OLS}$	$\hat{\beta}_{Bartlett}$	$t\text{-stat}_{HT}$	$t\text{-stat}_{Bartlett}$	$t\text{-stat}_{NW}$	R^2
1920-2003	1.453	1.363	3.072**	2.554*	4.602**	0.489
1920-1990	1.389	1.480	2.605**	2.503*	3.849**	0.464
1950-2003	1.033	1.043	1.809*	1.928 *	2.985**	0.367

Panel C: Bias-adjusted coefficients, imposing the null hypothesis

Sample	$\hat{\beta}_{OLS}$	$\hat{\beta}_{Bartlett}$	$t\text{-stat}_{HTH_0}$	$t\text{-stat}_{BartlettH_0}$	$t\text{-stat}_{NWH_0}$	R^2
1871-2003	1.453	1.363	2.196*	1.826*	3.289**	0.489
1920-1990	1.389	1.480	1.908*	1.833*	2.819**	0.464
1950-2003	1.033	1.043	1.439	1.534	2.375*	0.367

** means significance at the 1% level; * significance at the 5% level.

This table reports estimates of parameters for the predictive regression Model (1), where the 10-year treasury bond real return is regressed on the long-run component of the S&P 500 earnings yield. All data are in real terms. $\hat{\beta}_{Bartlett}$ is the Bartlett estimate of the slope coefficient. In Panels B and C we report the bias-adjusted slope coefficient estimates using the bias formula (22). The $t\text{-stat}_{Bartlett}$ are t-statistics computed from the asymptotical correct standard errors derived in Proposition 1 and $t\text{-stat}_{HT}$ is defined in (17). $t\text{-stat}_{NW}$ denotes t-statistic based on the Newey-West (1987) standard errors with 10 lags. $t\text{-stat}_{HTH_0}$ and (27) denote the t-test while imposing the null.

Table 11: Regression of short duration treasury bill 10-year real returns on the long-run component of S&P 500 earnings yield.

Panel A: Results under the maintained hypothesis, not bias-adjusted

Sample	$\hat{\beta}_{OLS}$	$\hat{\beta}_{Bartlett}$	$t\text{-stat}_{HT}$	$t\text{-stat}_{Bartlett}$	$t\text{-stat}_{NW}$	R^2
1934-2003	1.145	1.016	1.938*	1.747*	3.159**	0.298
1934-1990	0.960	1.095	1.332	1.709*	2.886**	0.205
1950-2003	0.883	0.978	1.512	1.704*	3.416**	0.227

Panel B: Bias-adjusted coefficients, under the maintained hypothesis

Sample	$\hat{\beta}_{OLS}$	$\hat{\beta}_{Bartlett}$	$t\text{-stat}_{HT}$	$t\text{-stat}_{Bartlett}$	$t\text{-stat}_{NW}$	R^2
1934-2003	1.108	0.979	1.875*	1.683*	3.057**	0.298
1934-1990	0.913	1.048	1.268	1.637	2.745**	0.205
1950-2003	0.797	0.892	1.365	1.554	3.084**	0.227

Panel C: Bias-adjusted coefficients, imposing the null hypothesis

Sample	$\hat{\beta}_{OLS}$	$\hat{\beta}_{Bartlett}$	$t\text{-stat}_{HTH_0}$	$t\text{-stat}_{BartlettH_0}$	$t\text{-stat}_{NW}$	R^2
1934-2003	1.108	0.979	1.571	1.410	2.561**	0.298
1934-1990	0.913	1.048	1.131	1.460	2.449*	0.205
1950-2003	0.797	0.892	1.200	1.367	2.712**	0.227

** means significance at the 1% level; * significance at the 5% level.

This table reports estimates of parameters for the predictive regression Model (1), where the 10-year treasury bond real return is regressed on the long-run component of the S&P 500 earnings yield. All data are in real terms. $\hat{\beta}_{Bartlett}$ is the Bartlett estimate of the slope coefficient. In Panels B and C we report the bias-adjusted slope coefficient estimates using the bias formula (22). The $t\text{-stat}_{Bartlett}$ are t-statistics computed from the asymptotical correct standard errors derived in Proposition 1 and $t\text{-stat}_{HT}$ is defined in (17). $t\text{-stat}_{NW}$ denotes t-statistic based on the Newey-West (1987) standard errors with 10 lags. $t\text{-stat}_{HTH_0}$ and (27) denote the t-test while imposing the null.

Table 12: Monte-Carlo Simulation Results

Panel A: $\rho = 0$, $u_t = \epsilon_t$								
	$\beta = 0$				$\beta = 2.5$			
$T =$	50	100	200	1000	50	100	200	1000
Av. $\hat{\beta}_{OLS}$	0.053	-0.011	0.012	0.012	2.392	2.495	2.457	2.511
Av. $\hat{\beta}_{Bartlett}$	0.020	-0.008	0.006	0.013	2.330	2.478	2.440	2.510
Stdev($\hat{\beta}_{OLS}$)	3.167	2.140	1.899	0.645	3.236	2.189	1.848	0.638
Stdev($\hat{\beta}_{Bartlett}$)	2.805	2.027	1.817	0.642	2.893	2.065	1.761	0.634
Av. Stdev$_{OLS}$	1.241	0.817	0.710	0.245	1.254	0.819	0.708	0.245
Av. Stdev$_{NW}$	1.723	1.462	1.333	0.544	1.758	1.457	1.332	0.544
Av. Stdev$_{HT}$	3.203	2.108	1.834	0.632	3.239	2.116	1.828	0.632
Av. R^2_{OLS}	0.131	0.068	0.054	0.007	0.194	0.148	0.132	0.101
95% CI	0.440	0.246	0.197	0.026	0.587	0.435	0.377	0.182
Av. $R^2_{Bartlett}$	0.108	0.062	0.050	0.007	0.171	0.142	0.127	0.101
95% CI	0.380	0.224	0.181	0.026	0.535	0.407	0.362	0.181
Size/Power$_{NW}$	0.332	0.211	0.194	0.107	0.114	0.196	0.216	1.000
Size/Power$_{HT}$	0.065	0.065	0.063	0.055	0.143	0.232	0.266	1.000
Size/Power$_{Bartlett}$	0.071	0.061	0.063	0.056	0.153	0.248	0.285	1.000
Size/Power$_{Engle}$	0.031	0.026	0.027	0.020	0.143	0.232	0.266	1.000

	Panel B: $\rho = 0.7$, $u_t = \epsilon_t$							
	$\beta = 0$				$\beta = 2.5$			
$T =$	50	100	200	1000	50	100	200	1000
Av. $\hat{\beta}_{OLS}$	-0.008	-0.002	-0.019	0.007	2.384	2.471	2.481	2.503
Av. $\hat{\beta}_{Bartlett}$	-0.011	-0.002	-0.016	0.007	2.292	2.441	2.464	2.503
Stdev($\hat{\beta}_{OLS}$)	1.391	0.888	0.741	0.242	1.403	0.882	0.750	0.243
Stdev($\hat{\beta}_{Bartlett}$)	1.250	0.843	0.712	0.241	1.296	0.845	0.726	0.243
Av. Stdev$_{OLS}$	0.473	0.297	0.256	0.086	0.487	0.299	0.256	0.086
Av. Stdev$_{NW}$	0.717	0.565	0.517	0.205	0.754	0.574	0.521	0.205
Av. Stdev$_{HT}$	1.222	0.767	0.662	0.222	1.257	0.772	0.662	0.222
Av. R^2_{OLS}	0.159	0.083	0.063	0.008	0.431	0.448	0.450	0.461
95% CI	0.524	0.297	0.223	0.030	0.832	0.752	0.724	0.566
Av. $R^2_{Bartlett}$	0.131	0.075	0.057	0.008	0.395	0.434	0.441	0.461
95% CI	0.450	0.269	0.212	0.030	0.794	0.726	0.707	0.566
Size/Power$_{NW}$	0.354	0.236	0.205	0.104	0.298	0.658	0.803	1.000
Size/Power$_{HT}$	0.111	0.099	0.089	0.078	0.416	0.756	0.881	1.000
Size/Power$_{Bartlett}$	0.111	0.098	0.088	0.076	0.439	0.788	0.896	1.000
Size/Power$_{Engle}$	0.058	0.047	0.040	0.027	0.416	0.756	0.881	1.000

Panel C: $\rho = 0.7$, $u_t \sim GARCH(1,1)$								
	$\beta = 0$				$\beta = 2.5$			
$T =$	50	100	200	1000	50	100	200	1000
Av. $\hat{\beta}_{OLS}$	0.024	0.027	-0.014	0.002	2.416	2.467	2.496	2.496
Av. $\hat{\beta}_{Bartlett}$	0.022	0.026	-0.015	0.001	2.324	2.438	2.476	2.496
Stdev($\hat{\beta}_{OLS}$)	1.408	0.866	0.741	0.241	1.424	0.891	0.740	0.238
Stdev($\hat{\beta}_{Bartlett}$)	1.256	0.825	0.716	0.239	1.297	0.846	0.717	0.237
Av. Stdev$_{OLS}$	0.473	0.299	0.256	0.086	0.483	0.299	0.255	0.087
Av. Stdev$_{NW}$	0.717	0.574	0.514	0.205	0.742	0.580	0.518	0.206
Av. Stdev$_{HT}$	1.222	0.771	0.660	0.223	1.248	0.773	0.659	0.224
Av. R^2_{OLS}	0.155	0.079	0.062	0.008	0.442	0.451	0.457	0.458
95% CI	0.512	0.280	0.225	0.029	0.842	0.758	0.728	0.565
Av. $R^2_{Bartlett}$	0.129	0.072	0.057	0.008	0.407	0.437	0.448	0.457
95% CI	0.435	0.258	0.211	0.029	0.801	0.742	0.714	0.565
Size/Power$_{NW}$	0.356	0.226	0.197	0.101	0.316	0.700	0.825	1.000
Size/Power$_{HT}$	0.099	0.089	0.090	0.068	0.434	0.772	0.882	1.000
Size/Power$_{Bartlett}$	0.101	0.091	0.093	0.069	0.445	0.793	0.902	1.000
Size/Power$_{Engle}$	0.053	0.037	0.038	0.026	0.434	0.772	0.882	1.000

Panel D: $\rho = 0.7$, Correlation between innovations								
Bias-adjustment of coefficient								
	$\beta = 0$				$\beta = 2.5$			
$T =$	50	100	200	1000	50	100	200	1000
Av. $\hat{\beta}_{OLS}$	0.569	0.292	0.147	0.030	3.001	2.788	2.645	2.529
Av. $\hat{\beta}_{Bartlett}$	0.529	0.283	0.143	0.030	2.871	2.753	2.635	2.528
Av. $\hat{\beta}_{OLS}$ bias adj.	0.006	-0.014	-0.011	-0.003	2.438	2.482	2.487	2.497
Av. $\hat{\beta}_{Bartlett}$ bias adj.	-0.035	-0.023	-0.015	-0.002	2.308	2.447	2.477	2.496
Stdev($\hat{\beta}_{OLS}$)	1.766	1.163	0.797	0.342	1.836	1.168	0.786	0.344
Stdev($\hat{\beta}_{Bartlett}$)	1.603	1.111	0.781	0.341	1.670	1.119	0.769	0.342
Av. Stdev$_{OLS}$	0.653	0.418	0.280	0.121	0.668	0.420	0.281	0.122
Av. Stdev$_{NW}$	0.992	0.794	0.596	0.287	1.021	0.800	0.601	0.288
Av. Stdev$_{HT}$	1.685	1.079	0.724	0.313	1.724	1.085	0.727	0.314
Av. R^2_{OLS}	0.155	0.080	0.041	0.008	0.384	0.351	0.327	0.305
95% CI	0.506	0.293	0.155	0.030	0.788	0.661	0.554	0.408
97.5% CI	0.588	0.356	0.198	0.039	0.832	0.709	0.591	0.427
Av. $R^2_{Bartlett}$	0.130	0.073	0.039	0.008	0.353	0.339	0.323	0.305
95% CI	0.442	0.267	0.149	0.029	0.740	0.638	0.545	0.407
97.5% CI	0.521	0.332	0.188	0.038	0.787	0.692	0.577	0.427
Size/Power$_{NW}$	0.321	0.217	0.160	0.107	0.199	0.436	0.800	1.000
Size/Power$_{HT}$	0.090	0.084	0.080	0.075	0.280	0.548	0.860	1.000
Size/Power$_{Bartlett}$	0.089	0.084	0.083	0.075	0.307	0.564	0.876	1.000
Size/Power$_{Engle}$	0.044	0.037	0.036	0.026	0.280	0.548	0.860	1.000

Panel D: $\rho = 0.7$, Correlation between innovations								
Bias-adjustment of coefficient, imposing the null								
	$\beta = 0$				$\beta = 2.5$			
$T =$	50	100	200	1000	50	100	200	1000
Size/Power$_{NW}$	0.264	0.191	0.142	0.097	0.161	0.373	0.773	1.000
Size/Power$_{HT}$	0.015	0.045	0.060	0.065	0.298	0.545	0.861	1.000
Size/Power$_{Bartlett}$	0.029	0.049	0.060	0.066	0.317	0.580	0.874	1.000
Size/Power$_{Engle}$	0.001	0.009	0.018	0.024	0.298	0.545	0.861	1.000
Power$_{NW}$ (not adj. for size)					0.500	0.700	0.924	1.000
Power$_{HT}$ (not adj. for size)					0.018	0.474	0.876	1.000
Power$_{Bartlett}$ (not adj. for size)					0.092	0.509	0.890	1.000
Power$_{Engle}$ (not adj. for size)					0.000	0.189	0.732	1.000

This table reports the average (Av.) and the standard deviation (Stdev) of the estimated OLS slope coefficient estimate, $\hat{\beta}_{OLS}$, and of the Bartlett slope coefficient estimate, $\hat{\beta}_{Bartlett}$. We also report the average of the standard deviations of the coefficient estimates in each simulation, the average of the t-statistics, and the average R^2 along with their 95% confidence intervals (CI). Lastly, we report the power of each of the t-statistics under the null hypothesis and the size adjusted power of each t-statistic under the alternative hypothesis. We report the results based on Newey-West variances with q lags (NW), based $\frac{2}{3}q$ times the OLS variance (HT) as in Proposition 2, and based on variances calculated using Bartlett kernel estimates as in Proposition 1. We use 5000 Monte-Carlo simulations of $T = 50, 100, 200, 1000$ observations in each Monte-Carlo replication. The true coefficient estimate is β=0 in the left side of the table and β=2.5 in the right side of the table. We generate the short-run returns as follows $r_t = \beta x_{t-q} + u_t$, where $x_t = \rho x_{t-1} + z_t$ and the error terms $u_t = \epsilon_t \sqrt{h_{t-1}}$ and $\epsilon_t \sim nid(0, .2)$, $h_t = \omega + \theta h_{t-1} + \phi u_t^2$ the OLS estimates are obtained regressing the long-run return $r_t(q)$ on the long-run predictive variable $x_{t-q}(q)$. Long-run variables are computed using a horizon of $q = 10$ periods. NW denotes Newey-West. Panel A contains the results when $\rho = 0$ and $u_t = \epsilon_t$. Panel B contains the results for $\rho = 0.7$ and $u_t = \epsilon_t$. Panel C contains the results for $\rho = 0.7$ and u_t evolves in according to a $GARCH(1,1)$ process. In Panel D $\rho = 0.7$ and the innovations in the short-run returns and in the predictive variables are correlated with correlation of -0.0018. The statistics are adjusted for bias using the values in Table 1 and $\Gamma_{zz}^{-1}\Gamma_{zu} = -5.7$. In the last rows we additionally report the size and size adjusted power and power of the t-statistics where we impose the null of $\beta = 0$.

Figure 1: Time series divided into different block structures.

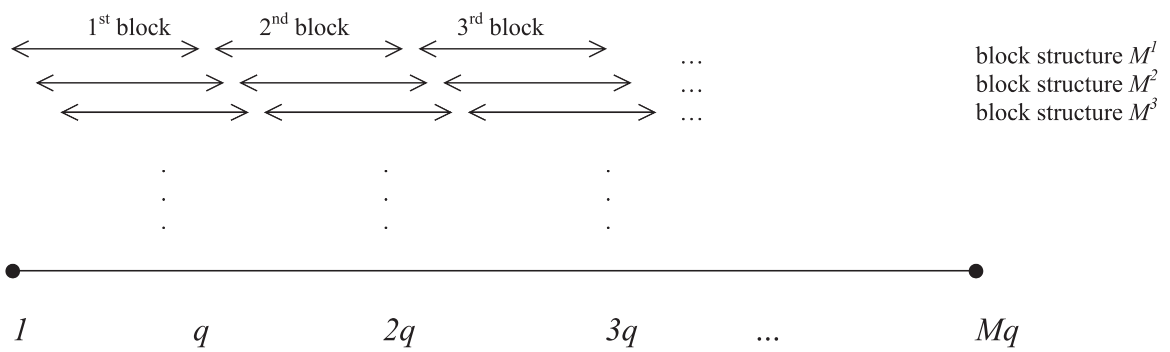

Figure 2: S&P 500 q-year real returns (bold lines) and prediction (line marked with a "+") based on the q-year long-run component of past earnings yield, estimation periods: 1920-1990. Different horizons $q = 3, 5, 7, 10, 12, 15$.

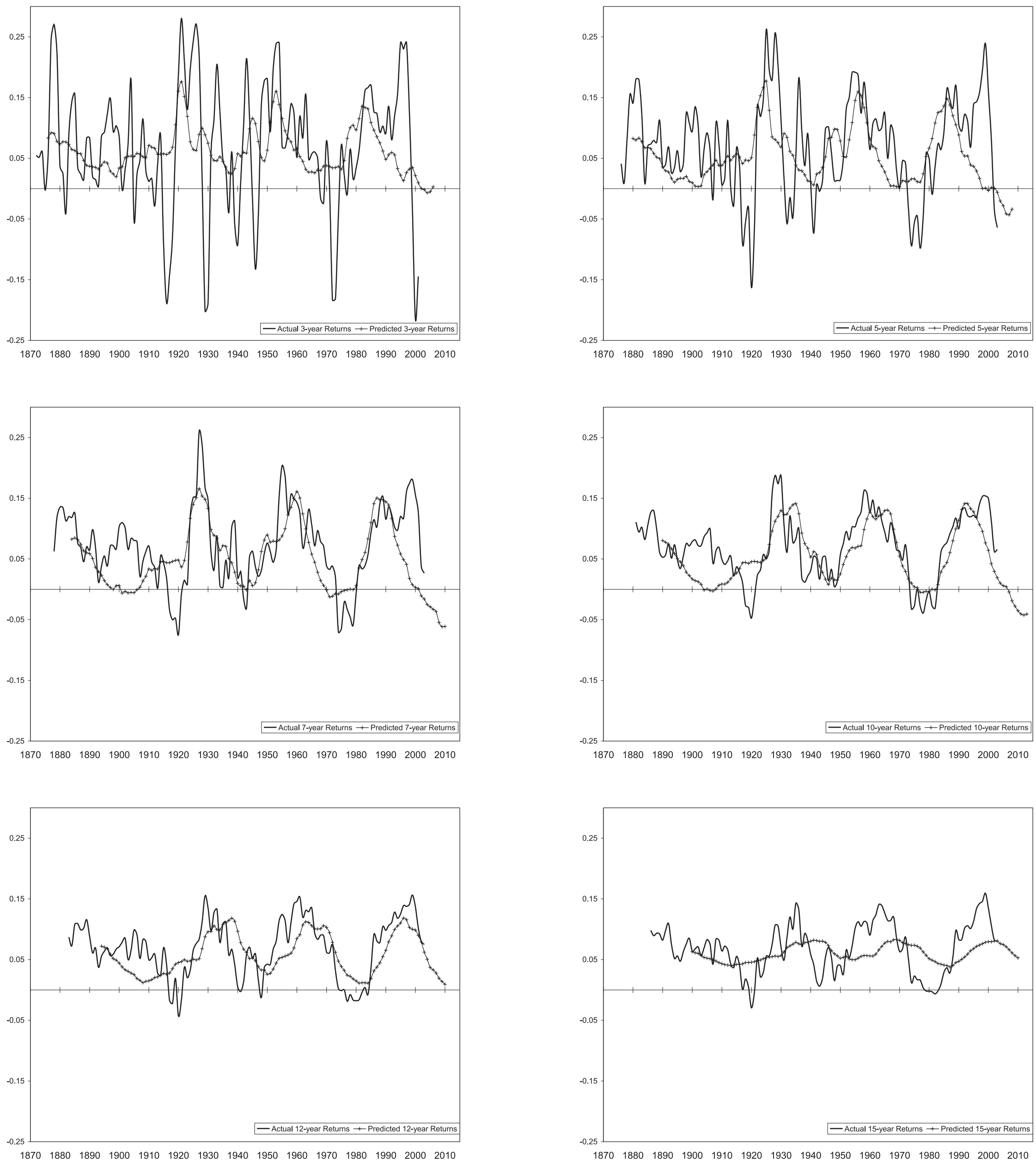

Figure 3: S&P 500 10-year real returns (bold line) and prediction (line marked with a "+") based on the long-run component of past earnings yield, estimation period: 1920-1990.

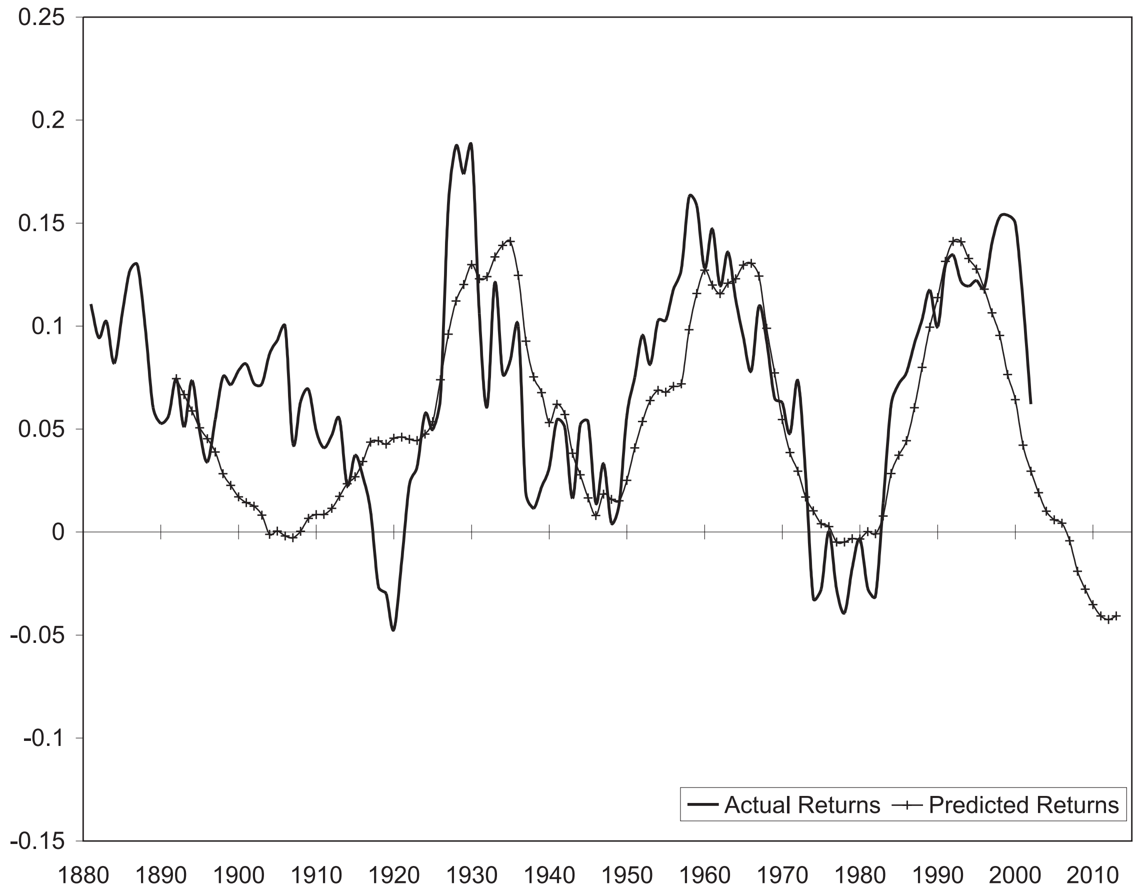

Figure 4: S&P 500 10-year real returns (bold line) and prediction (line marked with a "+") based on the long-run component of the past dividend yield, estimation period: 1920-1990.

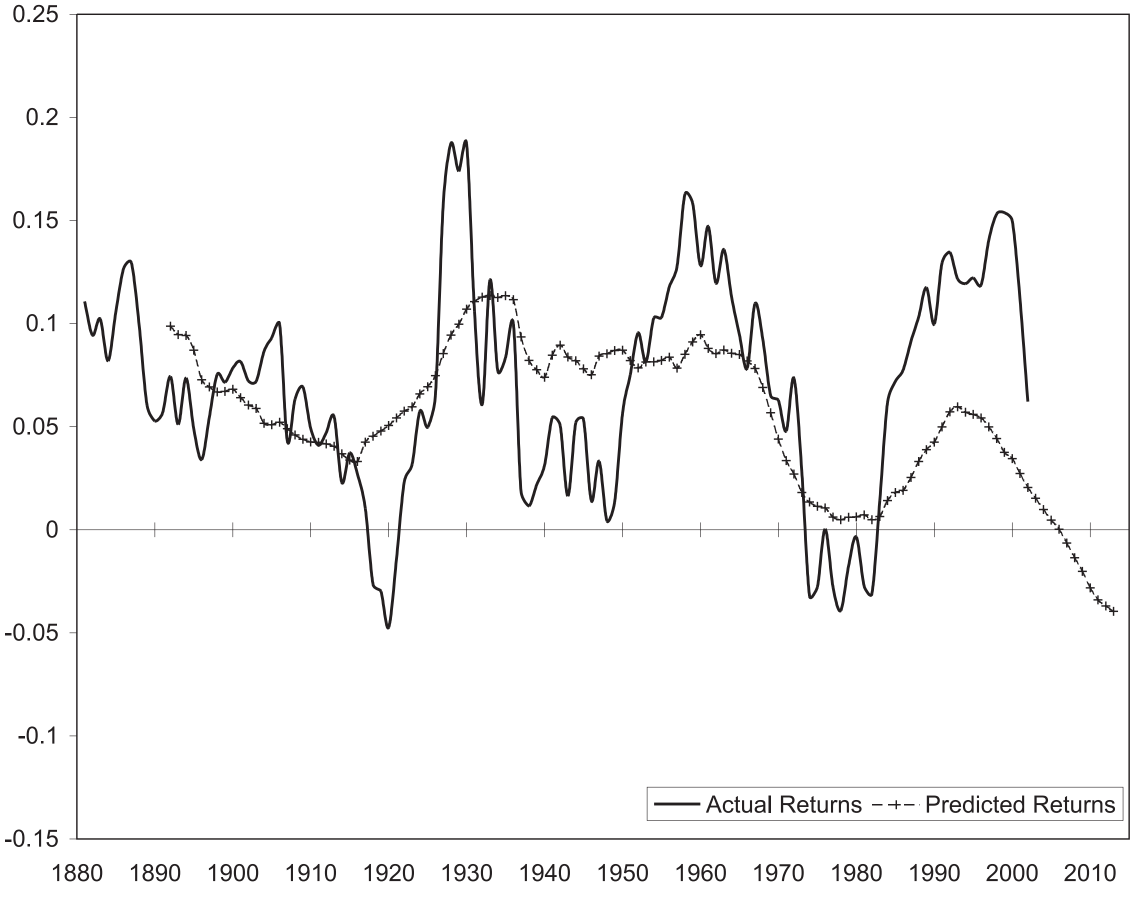

Figure 5: S&P 500 10-year real returns (bold line) and prediction (line marked with a "+") based on the past dividend yield, estimation period: 1920-1990.

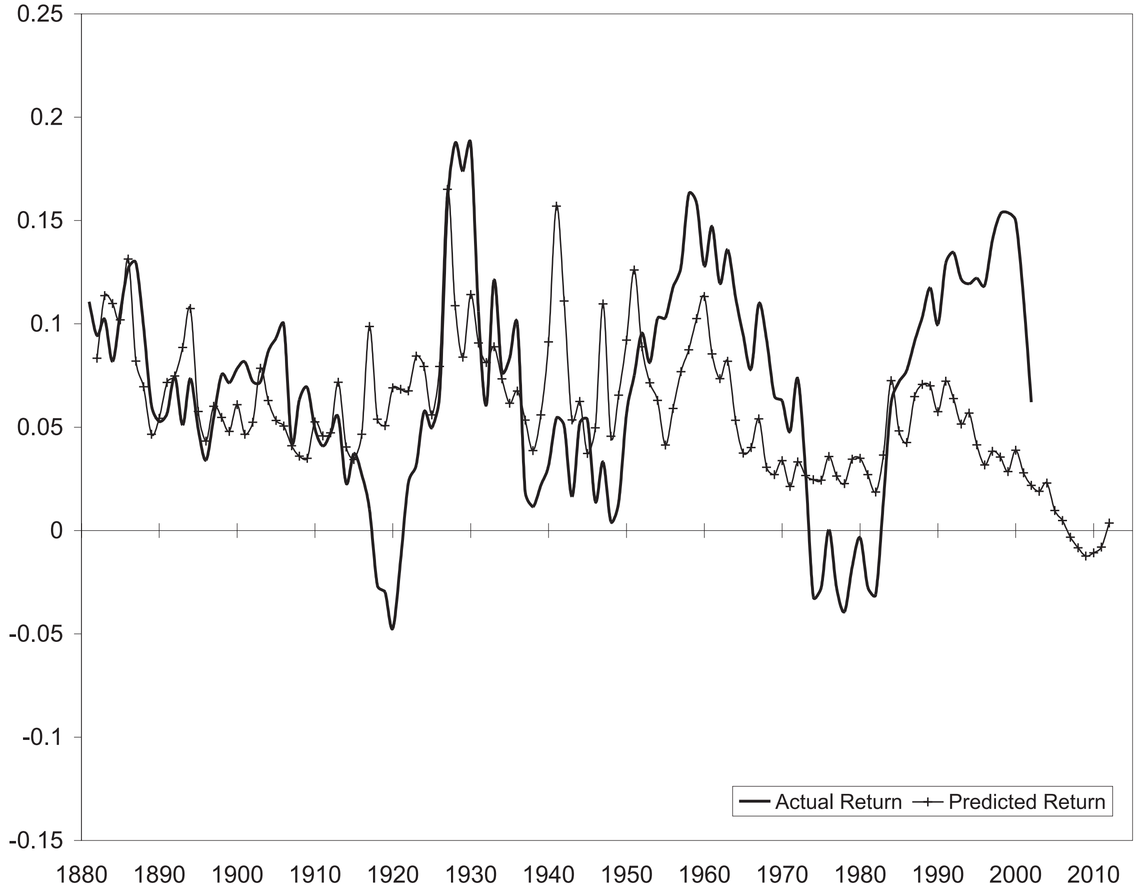

Figure 6: S&P 500 10-year real returns (bold line) and prediction (line marked with a "+") based on the earnings yield, estimation period: 1920-1990.

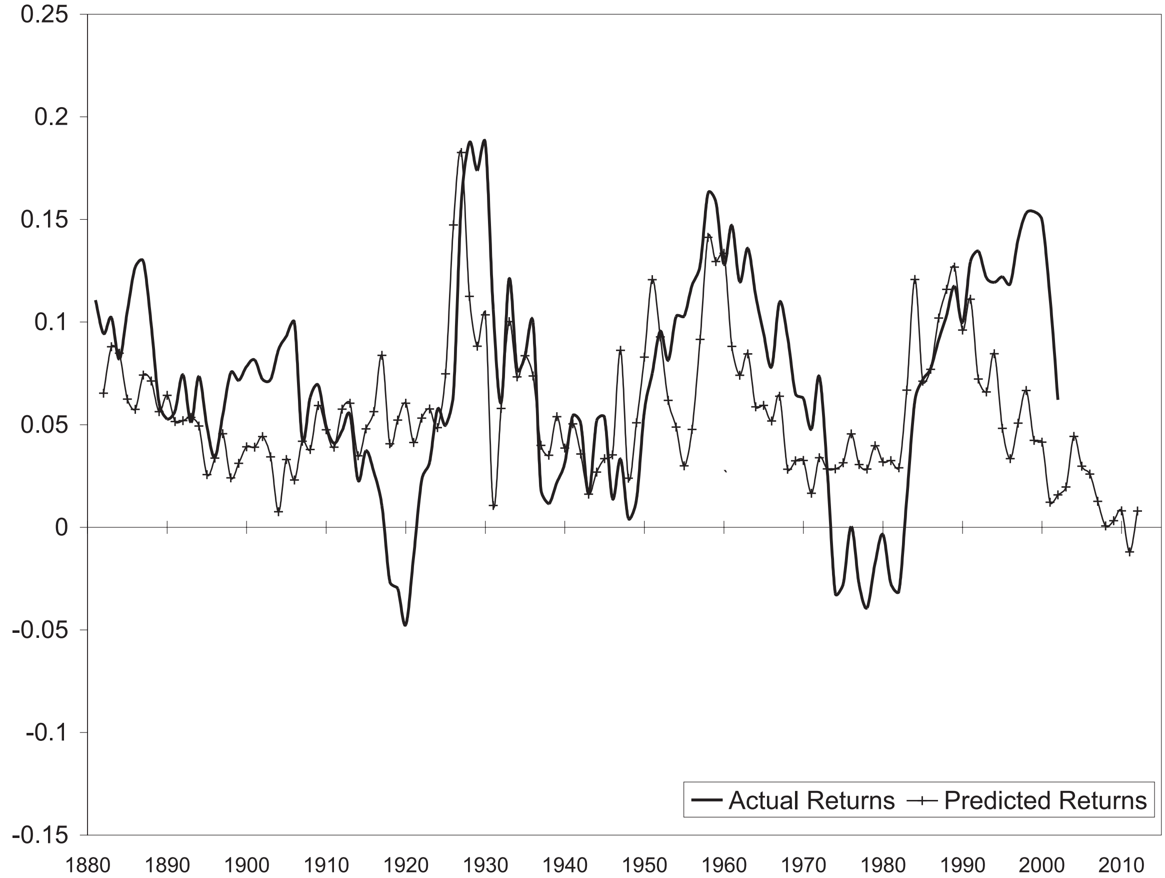

Figure 7: S&P 500 10-year earnings growth (bold line) and prediction (line marked with a "+") based on the long-run component of the earnings yield, estimation period: 1920-1990.

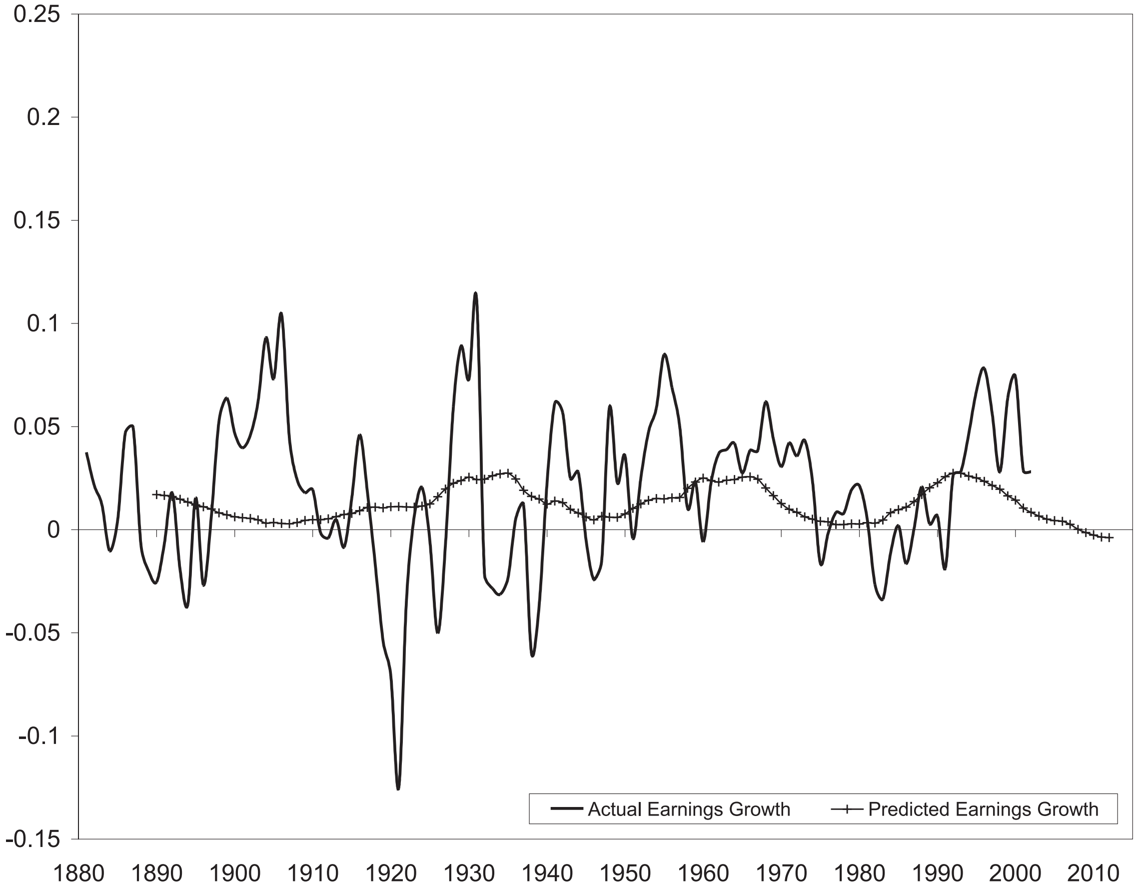

Figure 8: Investment grade corporate bond real 10-year returns (bold line) and predicted returns (line marked with a "+") based on past S&P 500 earnings yields, estimation period: 1920-1990.

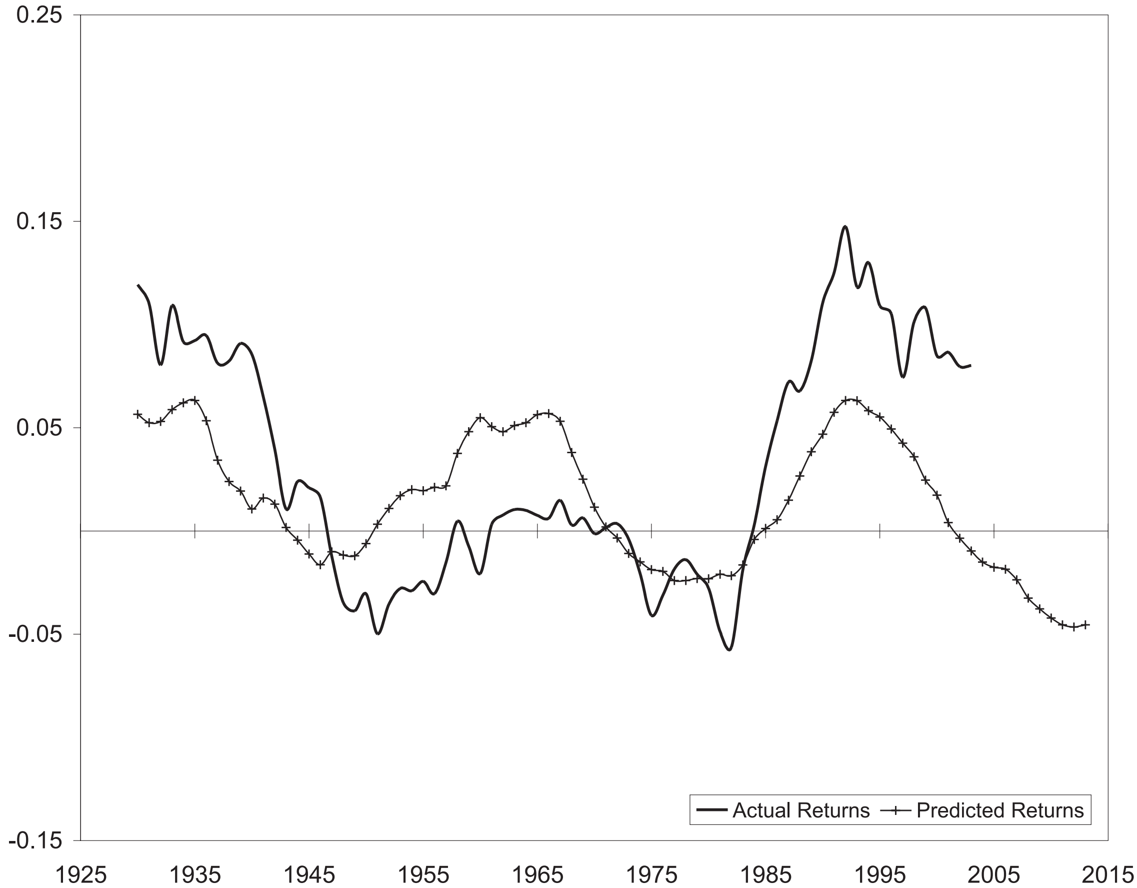

Figure 9: Long duration treasury bond real 10-year returns (bold line) and predicted returns (line marked with a “+”) based on past earnings yields, estimation period: 1920-1990.

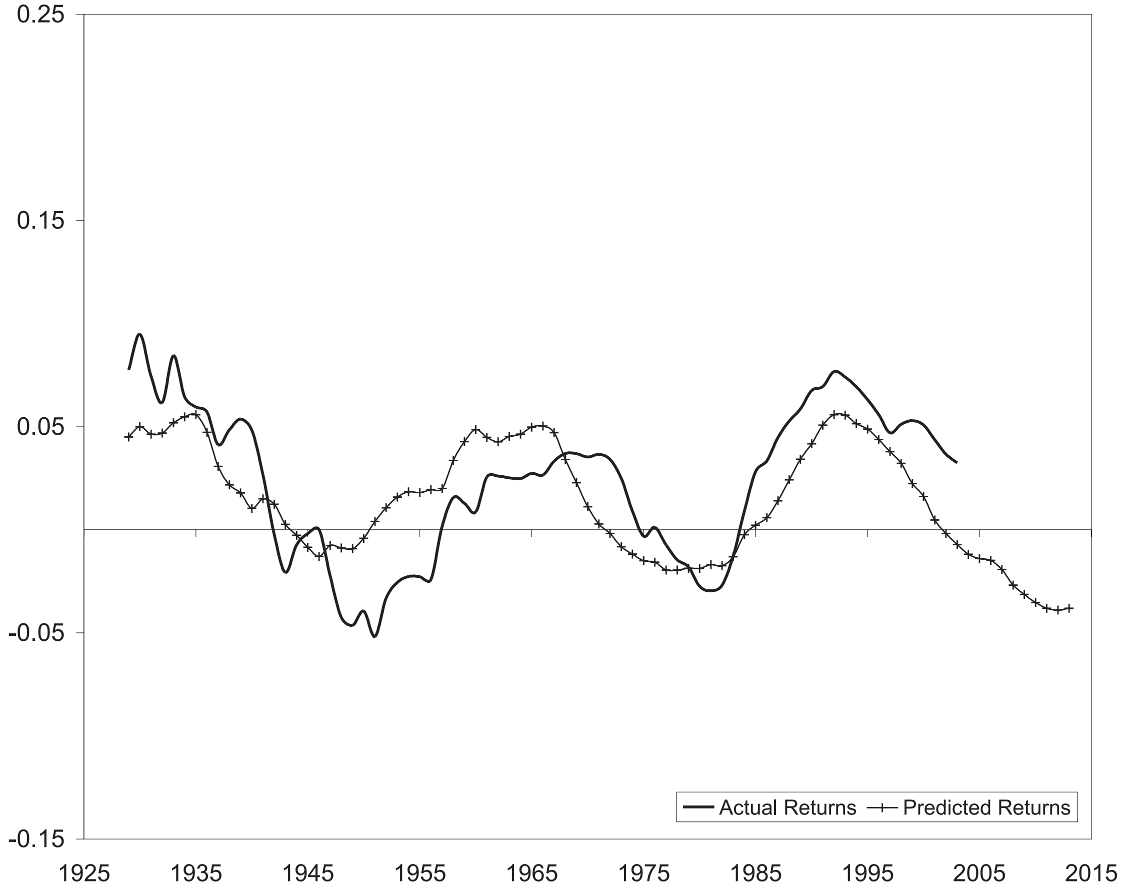

Figure 10: Short duration treasury bill real 10-year returns (bold line) and predicted returns (line marked with a "+") based on past earnings yields, estimation period: 1934-1990.