Discussion Paper Series

No.178

Subsampling-Based Tests of Stock-Return Predictability

In Choi
Timothy K. Chue

July 2006

Hitotsubashi University Research Unit
for Statistical Analysis in Social Sciences
A 21st-Century COE Program

Institute of Economic Research
Hitotsubashi University
Kunitachi, Tokyo, 186-8603 Japan
http://hi-stat.ier.hit-u.ac.jp/

Subsampling-Based Tests of Stock-Return Predictability*

In Choi† and Timothy K. Chue‡

First Draft: March 2006
This Draft: June 2006

Abstract

We develop subsampling-based tests of stock-return predictability and apply them to U.S. data. These tests allow for multiple predictor variables with local-to-unit roots. By contrast, previous methods that model the predictor variables as nearly integrated are only applicable to univariate predictive regressions. Simulation results demonstrate that our subsampling-based tests have desirable size and power properties. Using stock-market valuation ratios and the risk-free rate as predictors, our univariate tests show that the evidence of predictability is more concentrated in the 1926–1994 subperiod. In bivariate tests, we find support for predictability in the full sample period 1926–2004 and the 1952–2004 subperiod as well. For the subperiod 1952–2004, we also consider a number of consumption-based variables as predictors for stock returns and find that they tend to perform better than the dividend–price ratio. Among the variables we consider, the predictive power of the consumption–wealth ratio proposed by Lettau and Ludvigson (2001a, 2001b) seems to be the most robust. Among variables based on habit persistence, Campbell and Cochrane's (1999) nonlinear specification tends to outperform a more traditional, linear specification.

Keywords: Subsampling, local-to-unit roots, predictive regression, stock-return predictability, consumption-based models.

1 Introduction

The finance profession has a long-standing interest in the study of stock-market predictability. For practitioners, having the ability to forecast future stock returns is clearly valuable for asset-allocation decisions. For academics, whether or not stock

*This paper was presented at the Hitotsubashi University. We thank the seminar participants, especially Taku Yamamoto, for helpful comments. The first author acknowledges financial support from the RGC Competitive Earmarked Research Grant 2003–2004 under Project No. HKUST6223/03H.

†Department of Economics, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong. E-mail: inchoi@ust.hk. Home page: http://ihome.ust.hk/~inchoi.

‡Department of Economics, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong. E-mail: timchue@ust.hk.

returns are predictable and by which variables they can be predicted affect how the stock market should be modelled theoretically. For example, the theories proposed by Campbell and Cochrane (1999), Lettau and Ludvigson (2001a, 2001b), Lustig and Van Nieuwerburgh (2005), Piazzesi, Schneider, and Tuzel (2006), Santos and Veronesi (2006), and Yogo (2006) all have testable implications regarding stock-return predictability.

Early studies of predictability rely on standard asymptotic distribution theory to draw inference. Examples of such studies include the works by Fama and Schwert (1977), Keim and Stambaugh (1986), Campbell (1987), Campbell and Shiller (1988), Fama and French (1988, 1989), and Hodrick (1992). However, as more recent studies have pointed out (see Elliott and Stock, 1994, and Stambaugh, 1999, for example), standard asymptotic distribution theory works poorly when the predictor variable is persistent and its innovations are highly correlated with stock returns.

To evaluate the evidence of predictability in this setting, new tests that model the predictor variable as nearly integrated have been developed. In particular, Torous, Valkanov, and Yan (2004) and Campbell and Yogo (2005) both use Bonferroni methods to derive tests that allow the predictor variable to contain a local-to-unit root.[1] Although these tests perform much better than the conventional t-tests, it is not clear how the Bonferroni methods can be extended to a multiple-predictive regression.

In practice, however, the need to carry out tests for multiple-predictive regressions is pressing because the theoretical models mentioned in the first paragraph suggest different variables that could be used to forecast returns. To examine the marginal and/or joint predictive power of these variables, we need to conduct statistical tests in a multivariate setting. Since many of these variables are highly persistent, using standard asymptotics for inference can be misleading. In the current literature, there is not yet any procedure that can test for predictability in the presence of multiple, nearly integrated regressors.

In this paper, we fill this void by developing subsampling-based predictability tests of that allow for multiple regressors with local-to-unit roots. The subsampling approach computes the statistic of interest for subsamples of the data (consecutive sample points in the case of time-series data) and the statistic's subsampled values are used to estimate its finite-sample distribution.[2] Romano and Wolf (2001) and Choi (2005b) examine the performance of subsampling when it is used to analyze time series with exact unit roots. In this study, we prove the validity of subsampling for time series with local-to-unit roots.

Since subsampling does not require the estimation of nuisance parameters, applying the procedure to a multiple-regression setting is no more difficult than applying it to a simple regression. By contrast, previous tests proposed by Torous et al. (2004) and Campbell and Yogo (2005) require the estimation of the degree of persistence of the predictor variables. These studies use Bonferroni methods to carry out this

[1] Valkanov (2003) also uses the local-to-unit root setup to examine stock-return predictability. Valkanov's methodology relies on a long-run restriction between the dividend–price ratio and stock returns implied by the dynamic Gordon growth model. This methodology is not applicable to predictor variables that do not have such a long-run relationship with stock returns.

[2] Politis, Romano, and Wolf (1999) provide a rigorous introduction to the theory of subsampling.

estimation in univariate tests, but multivariate extensions of their approaches seem infeasible. Wolf (2000) also uses subsampling methods to study predictive regressions, but he only examines a model with a single stationary regressor. Lanne (2002) makes use of stationarity tests to carry out inference on stock-return predictability. He allows the predictor variables to be nearly integrated, but bases his inference on stock-return data alone and ignores data on the predictor variables altogether. As Campbell and Yogo (2005) argue, such an approach tends to have poor power when the predictor variable is persistent but remains sufficiently far from being integrated. Finally, the bootstrap may seem to be a feasible alternative to subsampling, but it can be shown to be inconsistent for regressions with nearly integrated regressors. Basawa et al. (1991), Datta (1996), and Choi (2005a) demonstrate the failure of the bootstrap for the AR(1) and VAR models. One major strength of subsampling is that it can work even when the bootstrap method fails.

Our subsampling-based tests suggest that the evidence for stock-return predictability using stoc-market valuation ratios and the risk-free rate is quite strong. Our univariate tests show that the evidence is more concentrated in the subperiod from 1926–1994. In bivariate tests, we find evidence for predictability in the full sample period, 1926–2004, and the subperiods 1926–1994 and 1952–2004. We also demonstrate the value of being able to carry out joint tests—there are numerous cases where univariate tests are insignificant, but joint tests are not negligible.

We also show that a number of consumption-based variables have predictive power for stock returns in the subperiod 1952–2004. During this period, these variables tend to be better predictors for stock returns than the dividend–price ratio. Among the variables we consider, the predictive power of the consumption–wealth ratio (*cay*) proposed by Lettau and Ludvigson (2001a, 2001b) seems to be the most robust. Among variables that are based on habit persistence, Campbell and Cochrane's (1999) nonlinear specification tends to outperform a more traditional linear specification.

The rest of the paper is organized as follows. Section 2 introduces the model and test statistics for predictive regressions. Section 3 proposes subsampling-based methods for predictive regressions with one regressor. Section 4 extends the subsampling method of Section 3 to multiple regressions. Section 5 reports simulation results. Section 6 presents our empirical findings on stock-market predictability. Section 7 concludes. We relegate technical results to the appendices.

2 The model and test statistics for predictive regressions

Consider the simple linear regression model

$$y_t = \alpha + \beta x_{t-1} + u_{yt}, \ (t = 2, \ldots, T), \tag{1}$$

where

$$\begin{aligned} x_t &= \mu + v_t; \\ v_t &= \rho v_{t-1} + u_{vt}; \\ \rho &= e^{c/T}, \ c \in R. \end{aligned} \tag{2}$$

Model (1) is the prototypical predictive regression model that has been widely used in the finance literature. For instance, y_t denotes the excess stock return in period t and x_{t-1} is a variable observed at time period $t-1$ that may be able to predict y_t. In order to predict the excess stock return, such variables as interest rates, default spreads, dividend yield, the book-to-market and earnings–price ratios have been used.

Modelling x_t as a nearly integrated process[3] as in (2) reflects the fact that many predictors used in the finance literature are quite persistent. Campbell and Yogo (2005) report that the respective 95% confidence intervals for ρ are [0.957, 1.007] and [0.939,1.000] for the dividend–price and earnings–price ratios they studied. The modelling has also been used in the finance literature including Valkanov (2003), Torous et al. (2004), and Campbell and Yogo (2005). These articles use the representation $\rho = 1 + c/T$, but this is equivalent to our specification (2) for asymptotic analysis. We prefer using representation (2) because it simplifies the proofs in Appendixes I and II.

In model (1), it is reasonable to assume that u_{yt} and u_{vt} are correlated. For example, if x_t and y_t denote the dividend yield and the excess stock return, respectively, an increase in stock price will decrease the dividend yield and increase the stock return. More specifically, we assume

Assumption 1 Let $\|a\|_p = (E|a|^p)^{1/p}$. Suppose

(i) $u_{yt} = \delta u_{vt} + e_t$ where u_{vs} is independent of e_t for every s and t.

(ii) $u_t = \{u_{vt}, e_t\}$ is strictly stationary with $E(u_1) = 0$ and $E\|u_1\|^{2+\xi} < \infty$ for $\xi > 0$;

(iii) $\{u_t\}$ is strong mixing with its mixing coefficients $\alpha_{u,m}$ satisfying, for $\epsilon > 0$,

$$\sum_{m=1}^{\infty} \alpha_{u,m}^{\epsilon/(2+\epsilon)} < \infty.$$

Assumption 1 allows serial correlations in $\{u_{yt}\}$ and $\{u_{vt}\}$ and cross-sectional correlations between u_{yt} and u_{vs}. Most previous studies have assumed white noise processes for $\{u_{vt}\}$ and $\{u_{yt}\}$. Assumption 1 generalizes this, though in most financial applications it suffices to assume $\{u_{vt}\}$ and $\{u_{yt}\}$ are uncorrelated. In addition, it is not necessary to model the form of the serial correlation in $\{u_t\}$ in this study. Under Assumption 1, the functional central-limit theorem for $\{u_t\}$ also holds (cf. Phillips and Durlauf, 1986).

The null hypothesis we are interested in is

$$H_0 : \beta = \beta_0. \tag{3}$$

In most cases, we will set $\beta_0 = 0$, which corresponds to the unpredictability of y_t.

[3] See Bobkoski (1983), Chan and Wei (1987), and Phillips (1987) for nearly integrated processes.

For null hypothesis (3), we may consider the usual t-test:

$$t(\beta_0) = \frac{\hat{\beta} - \beta_0}{\sqrt{\hat{\sigma}_y^2 \left(\sum_{t=2}^{T} (x_{t-1} - \bar{x}_{-1})^2 \right)^{-1}}}, \tag{4}$$

where $\hat{\beta}$ is the OLS estimator of β, $\bar{x}_{-1} = \frac{1}{T-1} \sum_{t=2}^{T} x_{t-1}$, and $\hat{\sigma}_y^2$ is the usual estimator of $\sigma_y^2 = E(u_{yt}^2)$. The asymptotic distribution of $t(\beta_0)$ for the case of serially uncorrelated $\{u_{yt}\}$ is given (cf. Elliott and Stock, 1994) in the relation

$$t(\beta_0) \Rightarrow \gamma \frac{\int_0^1 \bar{J}_c(r) \mathrm{d}W(r)}{\sqrt{\int_0^1 \bar{J}_c(r)^2 \mathrm{d}r}} + \sqrt{1-\gamma^2} Z, \text{ as } T \to \infty,$$

where $\Rightarrow$ denotes weak convergence, $\bar{J}_c(r) = J_c(r) - \int_0^1 J_c(s) \mathrm{d}s$, $J_c(r)$ is an Ornstein–Uhlenbeck process generated by the stochastic differential equation $\mathrm{d}J_c(r) = cJ_c(r)\mathrm{d}r + \mathrm{d}W(r)$ with the initial condition $J_c(0) = 0$ and the standard Brownian motion $W(r)$, $\gamma = Corr(u_{yt}, u_{vt})$, and $Z \overset{d}{=} N(0,1)$, is independent of $(W(r), J_c(r))$. Unless $\gamma = 0$, the distribution of the t-test depends on the nuisance parameters c and γ, which makes it difficult to use it for statistical inference.

Under Assumption 1, model (1) can be rewritten as

$$y_t = \alpha' + \beta x_{t-1} + \delta(x_t - \rho x_{t-1}) + e_t, \tag{5}$$

where $\alpha' = \alpha - (1-\rho)\mu$. Note that the regressors are totally exogenous in model (5) such that the OLS estimator of β has a mixture normal distribution in the limit. Letting x_{t-1}^* be the residual obtained by regressing x_{t-1} on $\{1, x_t - \rho x_{t-1}\}$, the t-test for null hypothesis (3) is defined by

$$Q(\beta_0, \rho) = \frac{\tilde{\beta} - \beta_0}{\sqrt{\hat{\sigma}_e^2 \left(\sum_{t=2}^{T} x_{t-1}^{*2} \right)^{-1}}}, \tag{6}$$

where $\tilde{\beta}$ is the OLS estimator of β using model (5) and $\hat{\sigma}_e^2$ is the usual estimator of $\sigma_e^2 = E(e_t^2)$. The $Q(\beta_0, \rho)$ test is designed for serially uncorrelated $\{e_t\}$ and has some optimal properties as discussed in Campbell and Yogo (2005). If $\{e_t\}$ are serially correlated, $\hat{\sigma}_e^2$ should be replaced with the long-run variance estimator (see, e.g., Andrews, 1991). The $Q(\beta_0, \rho)$ weakly converges to a standard normal distribution.

In practice, however, the $Q(\beta_0, \rho)$ test is not feasible since the value of ρ is unknown. If we choose $\rho = 1$, it is asymptotically equivalent to Lewellen's (2004) bias-adjusted test, though the functional forms of $Q(\beta_0, 1)$ and Lewellen's test are different. When $\rho = 1$,

$$Q(\beta_0, 1) \Rightarrow Z - \frac{c\delta}{\sqrt{\sigma_e^2 \left(\sigma_v^2 \int_0^1 \bar{J}_c(r)^2 \mathrm{d}r \right)^{-1}}}, \text{ as } T \to \infty. \tag{7}$$

Again, the limiting distribution given in (7) involves nuisance parameters in a complicated way. If $\{e_t\}$ is serially correlated as in Assumption 1, the limiting distribution will contain additional parameters resulting from the serial correlation.

3 Subsampling test statistics

3.1 Subsampling

It was shown in Section 2 that $t(\beta_0)$ and $Q(\beta_0, 1)$ have limiting distributions that depend on inestimable parameters. Conventional asymptotic methods cannot be used for this reason. To remedy the situation, this section proposes using subsampling as a way to find approximations to the limiting distributions of the test statistics $t(\beta_0)$ and $Q(\beta_0, 1)$. Using smaller blocks of consecutively observed time series, the subsampling method computes $t(\beta_0)$ and $Q(\beta_0, 1)$, and then formulates empirical cumulative distribution functions using the computed values of the statistics. Subsample critical values are obtained from the empirical distribution functions. We use consecutively observed time series to retain the serial correlation structure present in the data. In addition, blocks may share common sample points. Figure 1 illustrates the scheme of formulating blocks for the subsampling method. The method explained so far is called *uncentered* subsampling, the meaning of which will become obvious shortly.

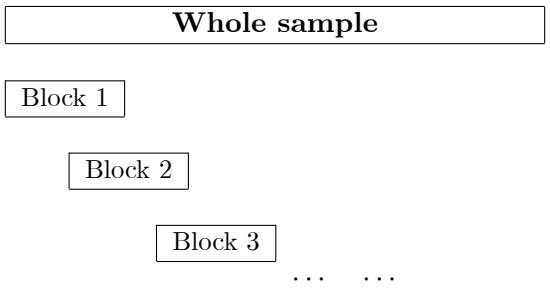

Figure 1: Blocks for Subsampling Method

To be more specific, let $t_{b,s}(\beta_0)$ be the t-test that uses the subsample $\{(y_s, x_s)$, $\ldots$, $(y_{s+b-1}, x_{s+b-1})\}$. We define $Q_{b,s}(\beta_0, 1)$ in the same way. The number of sample points in the subsample is b, which is called the block size. Index s denotes the starting point of the subsample. In this subsampling scheme, there will be $T - b + 1$ blocks with size b. Now, consider the empirical distribution functions using $t_{b,s}(\beta_0)$ and $Q_{b,s}(\beta_0, 1)$

$$L_T^t(x) = \frac{1}{T-b+1} \sum_{s=1}^{T-b+1} 1\{t_{b,s}(\beta_0) \leq x\}; \tag{8}$$

$$L_T^Q(x) = \frac{1}{T-b+1} \sum_{s=1}^{T-b+1} 1\{Q_{b,s}(\beta_0, 1) \leq x\}. \tag{9}$$

These are step functions of x.

Under Assumption 1, it is shown in part (i) of Theorem A.1 in Appendix I that $L_T^t(x)$ and $L_T^Q(x)$ become closer to their respective limiting distributions uniformly

in x and with probability approaching one as $T \to \infty$ if $b = O(T^{\eta})$ with $\frac{1}{2} < \eta \leq \frac{2}{3}$. The intuition for this result comes from the Glivenko–Cantelli lemma—the empirical distribution function of an iid random variable approximates the distribution function of the random variable. In (8), $t_{b,s}(\beta_0)$ and $Q_{b,s}(\beta_0, 1)$ are neither independent nor identically distributed, but they are asymptotically independent in the sense that blocks far apart are independent. They are also identically distributed in large samples. Thus, the empirical distributions (8) and (9) will mimic the limiting distributions of $t_{b,s}(\beta_0)$ and $Q_{b,s}(\beta_0, 1)$, respectively, in large samples.

Once b is chosen properly, approximations to the critical values of the limiting distributions of $t(\beta_0)$ and $Q(\beta_0, 1)$ can be obtained from (8) and (9). The test statistics $t(\beta_0)$ and $Q(\beta_0, 1)$ will have correct asymptotic sizes when the subsample critical values from (8) and (9) are used as proven in part (ii) of Theorem A.1. In practice, values of the test statistics $t(\beta_0)$ and $Q(\beta_0, 1)$ that use the full sample are compared with those of the subsample critical values in order to reach a statistical conclusion on the given null hypothesis.

When the null hypothesis is not true, the subsample critical values diverge in probability, but at lower rates than the corresponding test statistics using the full sample. Thus, the probability of rejecting the null hypothesis when it is not true converges to one as $T \to \infty$. This is formally proven in Choi (2005b) and Choi and Chue (2004).

Another way of subsampling is to center the test statistics at the coefficient estimator using the full sample. That is, we use for subsampling

$$t_{b,s}(\hat{\beta}) = \frac{\hat{\beta}_{b,s} - \hat{\beta}}{\sqrt{\hat{\sigma}^2_{b,s,y} \left(\sum_{t=s+1}^{s+b-1} (x_{t-1} - \bar{x}_{-1,b,s})^2 \right)^{-1}}} \tag{10}$$

and

$$Q_{b,s}(\tilde{\beta}, 1) = \frac{\tilde{\beta}_{b,s} - \tilde{\beta}}{\sqrt{\hat{\sigma}^2_{b,s,e} \left(\sum_{t=s+1}^{s+b-1} x^{*2}_{t-1} \right)^{-1}}}, \tag{11}$$

where $\hat{\beta}$ and $\tilde{\beta}$ are the estimators of β using the full sample and the estimators with subscripts b and s are those using the subsample $\{(y_s, x_s), \ldots, (y_{s+b-1}, x_{s+b-1})\}$. Subsampling using these statistics is called the centered subsampling.

Under the null hypothesis, the centering has no effect on $t_{b,s}(\hat{\beta})$ in the large sample because

$$t_{b,s}(\hat{\beta}) = \frac{\hat{\beta}_{b,s} - \beta_0}{\sqrt{\hat{\sigma}^2_{b,s,y} \left(\sum_{t=s+1}^{s+b-1} (x_{t-1} - \bar{x}_{-1,b,s})^2 \right)^{-1}}} - \frac{b}{T} \frac{T(\hat{\beta} - \beta_0)}{\sqrt{\hat{\sigma}^2_{b,s,y} \left(\sum_{t=s+1}^{s+b-1} (x_{t-1} - \bar{x}_{-1,b,s})^2 / b^2 \right)^{-1}}} \tag{12}$$

and the second term on the right-hand side of this relation is asymptotically negligible as long as $\frac{b}{T} \to 0$. The same analysis applies to $Q_{b,s}(\hat{\beta}, 1)$. The validity of the centered subsampling is formally proven in Theorem A.2 of Appendix I.

Under the alternative hypothesis $H_1 : \beta \neq \beta_0$, however, the first term on the right-hand side of relation (12) is stochastically bounded while the second term is still asymptotically negligible. This implies that the subsample critical values are stochastically bounded in contrast with the uncentered subsampling where critical values diverge in probability under the alternative. An important implication of this is that the tests using the centered subsampling are likely to have higher power than those using the uncentered subsampling. In practice, however, centered subsampling sometimes brings unacceptably high size distortions which discourages its use. For our problem, it works reasonably well as we will see in Section 5.

3.2 Choice of block sizes

The validity of subsampling requires that b grow as T does but at a slower rate. This requirement is too rough to use in choosing b in practice. However, there are a few methods known to work reasonably well in finite samples. Romano and Wolf (2001) suggest the minimum-volatility method, which is shown to work well for the confidence intervals of an $AR(1)$ coefficient. This method also works well for tests based on vector autoregressions and panel regressions (cf. Choi, 2005b, and Choi and Chue, 2004). The algorithm for the minimum-volatility method is:

> *Step 1:* From $b_i = b_{small}$ to $b_i = b_{big}$, calculate the subsample critical values c_i.
>
> *Step 2:* For each b_i ($i = small + l, \ldots, big - l$), calculate the standard deviation of the critical values $c_{i-l}, \ldots, c_{i+l}$, denoted SC_i. Here, l is a small positive integer.
>
> *Step 3:* Choose the block size that gives the minimum of SC_i over i.

Romano and Wolf recommend a small number for l (2 or 3) in Step 2 and also note that the results are insensitive to this choice.

Simulation results in Section 5 indicate that the minimum-volatility method works well for the uncentered subsampling. However, unreported simulation results reveal size distortions for the centered subsampling. Thus, we consider calibration rules for the centered subsampling. Assume that an adequate approximation to an optimal block size at each nominal size, λ, is related to the sample size by

$$b^{opt,\lambda} = T^{\zeta}. \tag{13}$$

In order to estimate parameter ζ of relation (13), we ran simulations for various sample sizes and data-generating processes, and related them to optimal block sizes. Data were generated by (1) and (2) with $\alpha = 0$ and $\mu = 0$, which do not have any effects on the finite-sample values of the test statistics. We also set $\beta = 0$ in (1). Independent standard normal numbers were used for $\{u_{vt}\}$ and $\{e_t\}$. For the

calibration rule, we used $T = 100, 250, 500$; $c = -5, -10, -15$; $\delta = -0.1, -0.5, -0.9, -1, -2, -3$. The calibration rules are devised separately for the 5% and 10% significance levels.

We used the following algorithm for the calibration rules at each significance level.

> *Step 1:* For each set of parameter values, generate the data 200 times and calculate the subsample critical values for every block size from 5 to $0.8\times T$.[4] In addition, record the critical values of the full-sample predictability tests from the 200 iterations.
>
> *Step 2:* For each set of parameter values and for each block size, record the median of the 200 subsample critical values from Step 1.
>
> *Step 3:* For each set of parameter values, record the block size whose median critical value from Step 2 is closest to the critical value of the full-sample predictability test of Step 1 in terms of absolute discrepancy.
>
> *Step 4:* Regress the natural logarithm of the optimal block size from Step 3 on $\ln(T)$ to estimate the parameter ζ.

Steps 1–3 provide the block sizes that produce subsample critical values closest to the corresponding finite-sample critical values of the tests. The calibration rules obtained from the above algorithm are reported in Table 1. We report only the rules for the t-test because the $Q(\beta, 1)$ test is not recommended when used together with the calibration rules. The calibration rules turn out to satisfy the condition $b = O(T^{\eta})$ with $\frac{1}{2} < \eta \leq \frac{2}{3}$, as required for the subsampling validity, although we experiment with a wider band for b than is allowed for by the theoretical restriction.

Table 1: Calibration Rules for the t-Test Centered Subsampling

Significance level	ζ
5%	0.61
10%	0.67

4 Predictive regressions with multiple regressors

This section considers the subsampling methods of the previous section for the multiple-regression model

$$y_t = \alpha + \beta' x_{t-1} + u_{yt}, \ (t = 2, \ldots, T), \tag{14}$$

where x_t is a $k \times 1$ vector modelled by

$$\begin{aligned} x_t &= \mu + v_t; \\ v_t &= \rho v_{t-1} + u_{vt}; \\ \rho &= e^{C/T}, \ C = diag[c_1, \ldots, c_k], \ c_i \in R \ (i = 1, \ldots, k). \end{aligned} \tag{15}$$

[4] This choice includes $b = T^{\eta}$ with $\frac{1}{2} < \eta \leq \frac{2}{3}$.

Here, every element of v_t is nearly integrated. We could consider other possibilities for matrix C without complicating our subsampling approach, but specification (15) seems most relevant in applications.

We assume that Assumption 1 in Section 2 holds with $u_{yt} = \delta' u_{vt} + e_t$, allowing correlation between $\{u_{yt}\}$ and $\{u_{vt}\}$. For the null hypothesis, $H_0 : \beta = \beta_0$, we consider the Wald test defined by

$$W(\beta^0) = \left(\hat{\beta} - \beta_0\right)' \left(\hat{\sigma}_y^2 \left(\sum_{t=2}^{T} (x_{t-1} - \bar{x}_{-1}) (x_{t-1} - \bar{x}_{-1})' \right)^{-1} \right)^{-1} \left(\hat{\beta} - \beta_0\right),$$

where $\hat{\beta}$ is the OLS estimator of β. The asymptotic distribution of $W(\beta^0)$ obviously depends on nuisance parameters that make it difficult to tabulate its distribution. For the null hypothesis on individual coefficients $H_0 : \beta_i = \beta_{i0}$ $(i = 1, \ldots, k)$, we may use a t-test based on model (14).

As in Section 2, we rewrite model (14) such that

$$y_t = \alpha' + \beta x_{t-1} + \delta'(x_t - \rho x_{t-1}) + e_t \tag{16}$$

and consider the Wald test,

$$MQ(\beta_0, \rho) = \left(\tilde{\beta} - \beta_0\right)' \left(\hat{\sigma}_e^2 \left(\sum_{t=2}^{T} x_{t-1}^* x_{t-1}^{*\prime} \right)^{-1} \right)^{-1} \left(\tilde{\beta} - \beta_0\right),$$

where $\tilde{\beta}$ is the OLS estimator of β using model (16) and x_{t-1}^* is the residual vector obtained by regressing x_{t-1} on $\{1, x_t - \rho x_{t-1}\}$. $MQ(\beta_0, \rho)$ weakly converges to a chi-square distribution. With $\rho = I$, $MQ(\beta_0, \rho)$ can be considered an extension of the $Q(\beta_0, 1)$ test. In general, the limiting distribution of $MQ(\beta_0, I)$ depends on nuisance parameters in a complicated way. It is a chi-square distribution only when $\delta = 0$ and/or $\rho = I$. For the null hypothesis on individual coefficients $H_0 : \beta_i = \beta_{i0}$ $(i = 1, \ldots, k)$, we can use t-test based on model (16).

The methods of subsampling these test statistics, $W(\beta^0)$, $MQ(\beta_0, I)$ and the t-ratios, are no different from those in Section 3. We construct relevant empirical distribution functions and use them to select critical values. Though Theorems A.1 and A.2 in Appendix I are for the case where x_t is scalar, it is straightforward to extend them to the case of multiple regressors.[5] Thus, the results in Theorems A.1 and A.2 can be used to justify the use of the uncentered and centered subsamplings. The minimum-volatility methods discussed in Section 3 can also be used without changes for the choice of block sizes.

In order to devise calibration rules for the centered subsampling of $W(\beta^0)$, we

[5] The only change we need is to extend Lemma A.5 to the case of multiple regressors. This can be done using the same method as for Lemma A.5 with more complex notation and is not deemed to deserve separate treatments.

used the same method[6] as in Section 3. The rules for the t-ratio and $W(\beta^0)$ based on model (14) with $k = 2$ are reported in Table 2 below. We report only the rules for the t-ratio and $W(\beta^0)$ because the $MQ(\beta, I)$ test and the corresponding t-test do not work satisfactorily along with the calibration rules.

Table 2: Calibration Rules for the Centered Subsampling of the t-ratio and Wald test ($k = 2$)

	ζ	
Significance level	t-ratio	Wald
5%	0.67	0.65
10%	0.67	0.66

5 Simulation

This section reports empirical size and power of the t-, Wald, $Q(\beta_0, 1)$, and $MQ(\beta_0, I)$ tests using subsample critical values. We consider the cases of one and two regressors. The alternative hypothesis is $H_1 : \beta > 0$ for the case of a single regressor. That for the case of two regressors is $H_1 : \beta \neq 0$ or $H_1 : \beta_i > 0$. Data were generated by (1) and (2) for the univariate case and (14) and (15) for the bivariate case. We set $\alpha = 0$ and $\mu = 0$ in the data generation because they have no effects on the finite-sample values of the test statistics. In addition, the elements of β in (14) have the same value ψ. For $\{u_{vt}\}$,

$$u_{vt} \sim iid\ N(0, 1)$$

was used for the univariate case; and

$$u_{vt} \sim iid\ N\left(0,\ \begin{matrix} 1 & 0.85 \\ 0.85 & 1 \end{matrix}\right)$$

for the bivariate case. Note that the elements of u_{vt} are cross-sectionally correlated in the bivariate case. In the data-generating process (15), the diagonal elements of matrix C were set to have the same value denoted by c in subsequent tables. The error terms $\{u_{yt}\}$ were generated by

$$u_{yt} = \phi' u_{vt} + e_t,$$

with $e_t \sim iid\ N(0, 1)$ and $\phi = [\delta, \delta]'$. Note that $\{u_{vt}\}$ and $\{e_t\}$ are independent. The parameter δ measures the degree of dependence between $\{u_{vt}\}$ and $\{u_{yt}\}$.

In the following tables, we considered the cases $T = 100$, 250, 500; $c = -5$, -10, -15; $\psi = 0$, 0.01, 0.05, 0.1; and $\delta = -0.5$, -1.5, -3.[7] We ran 2,000 replications

[6] We used the data-generating process (DGP) where the covariance between the two innovation processes for x_t is 0.85 and the bivariate vector ρ has the same element c/T. Using more general DGPs may yield different rules, but the difference was not noticeable according to our limited experimentation with different DGPs.

[7] Parameter δ denotes the covariance between $\{u_{yt}\}$ and $\{u_{vt}\}$. The corresponding correlations are -0.45, -0.83, and -0.95.

for these tests. Tables 3–5 report the empirical size and power of the t-, Wald, $Q(\beta_0, 1)$, and $MQ(\beta_0, I)$ tests using critical values from the uncentered subsampling with the minimum-volatility rule. Table 3 reports the empirical size and power of the tests for the case of a single regressor. Table 4 reports those of the joint tests for the two-regressor case, while Table 5 reports those of the individual tests. We also report in Tables 3 and 4 the size of the t-, Wald, $Q(\beta_0, 1)$, and $MQ(\beta_0, I)$ tests using standard distributions for the purpose of comparison. The results in Tables 3–5 are summarized as follows.

- The subsampling-based t- and Wald tests keep nominal size quite well across all values of T, δ, and c, though the tests tend to underreject as the value of c decreases.

- The $Q(\beta_0, 1)$ and $MQ(\beta_0, I)$ tests along with subsampling work reasonably well under the null hypothesis when $c = 0$. However, as c and δ decrease, their performance deteriorates. Especially when $c = -15$ and $\delta = -3$, all the tests are subject to size distortions. An explanation for the poor performance of the tests under the null is provided in Appendix III.

- The t- and Wald tests using standard distributions perform poorer as δ takes smaller values and c is closer to zero. This is well expected from standard theory. Overall, we observe significant advantage of using the subsampling critical values over those from standard distributions.

- The $Q(\beta_0, 1)$ and $MQ(\beta_0, I)$ tests using standard distributions perform poorly under the null hypothesis unless $c = 0$. The $MQ(\beta_0, I)$ test shows serious overrejections for $c \neq 0$.

- As expected, the power of the tests improves as the value of ψ increases and as T increases. When $\psi = 0.1$ and $T = 500$, the power is close to one in all the cases.

- As c takes smaller values, the power of the t- and $Q(\beta_0, 1)$ tests in Tables 3–5 tends to decrease. However, that of the $MQ(\beta_0, 1)$ tests in Table 4 tends to increase, most likely due to size distortions.

- As δ takes smaller values, the power of the t- and $Q(\beta_0, 1)$ tests in Tables 3–5 tends to decrease. By contrast, that of the $MQ(\beta_0, 1)$ tests in Table 4 tends to increase, again most likely due to size distortions.

Tables 6–8 report the size and power of the t- and Wald tests using critical values from the centered subsampling with the calibration rules from Tables 1 and 2. The size and power of the $Q(\beta_0, 1)$ and $MQ(\beta_0, I)$ tests are not reported because they are unsatisfactory and do not deserve the space. We do not report the results of centered subsampling using the minimum-volatility rule either because these are also quite poor. Tables 6–8 are summarized as follows.

- The t-test in Tables 6 and 8 and the Wald test in Table 7 keep nominal size reasonably well when subsampling critical values are used across all values of T, δ, and c, though some overrejections are observed when $c = 0$.
- The power of the tests improves as the value of ψ increases and as T increases. In particular, we observe significant power gain over uncentered subsampling.
- As c takes smaller values, the power of the tests decreases.
- As δ takes smaller values, the power of the tests tends to decrease.

For comparison, we report the finite-sample size and power of the Bonferroni t- and Q-tests examined by Campbell and Yogo (2005) in Table 8. Since the tests are designed for the case of a single regressor, the results should be compared to those in Tables 3 and 6. The results in Table 9 are summarized as follows.

- Panel A of Table 9 shows that the size properties for both tests are very good. Unlike the subsampling-based tests we consider above, there is no evidence that these Bonferroni tests overreject even when $c = 0$.
- Panel B shows that the Bonferroni Q-test tends to dominate the Bonferroni t-test in terms of power which is consistent with the results that Campbell and Yogo report.
- Overall, the Bonferroni Q-test appears to be more powerful than univariate centered subsampling reported in Table 6, except when T and ψ are both small ($T = 100$, $\psi = 0.01$).

The results above indicate that the t- and Wald tests along with centered subsampling work reasonably well in the case of two regressors and are recommended for empirical applications. When there is only one regressor, the Bonferroni Q-test appears to be the best choice.

6 Testing for stock-market predictability

In this section, we apply the subsampling-based tests developed above to investigate the predictability of stock returns. We use the t- and Wald tests along with centered subsampling in light of the simulation results of the previous section. We first consider such predictor variables as the dividend–price ratio, the earnings–price ratio, and the short-term interest rate, which have been widely used in previous studies. In addition to these popular variables, we also consider predictors proposed by more recent theories. In particular, we examine: (1) the consumption–wealth ratio (cay) proposed by Lettau and Ludvigson (2001a, 2001b), (2) the labor-income–consumption ratio (s^w) proposed by Santos and Veronesi (2006), (3) the surplus-consumption ratio based on the linear habit specification examined by Li (2001) and others, and (4) the surplus-consumption ratio based on the nonlinear habit specification proposed by Campbell and Cochrane (1999).

The fact that our approach can handle nearly integrated regressors and can be used in a multiple-regression setting allows us to address important economic questions. First, by including both the dividend–price ratio and a second variable as predictors in the predictive regression, we can investigate if the second predictor has predictive power for stock returns beyond that already contained in the dividend–price ratio. In some of the recent studies mentioned above, the authors use standard asymptotics to demonstrate that the predictors suggested by their theories have incremental predictive power for stock returns relative to the dividend–price ratio. Since these new predictors and the dividend–price ratio are typically highly persistent, the use of standard asymptotics may not be valid.[8] By contrast, since our approach performs well even in these cases, we can use it to examine if the new predictors indeed contain incremental information.

Second, some of the predictors mentioned above are closely related and one may be interested in knowing their joint predictive power. For example, we may be interested in knowing the joint predictive power of the consumption–wealth ratio *cay* and the labor-income–consumption ratio s^w as both *cay* and s^w depend on aggregate consumption and labor income. We can also test for the joint predictive power of different habit specifications. Using the subsampling approach, we can carry out joint tests that are robust to the presence of nearly integrated predictors.

Third, some of the predictors can be viewed as being in competition with each other. For example, Campbell and Cochrane (1999) propose a form of nonlinear habits to overcome certain deficiencies of the more traditional, linear habit specification. Since higher surplus consumption forecasts lower future returns in both of these models, we can use our subsampling-based tests to see which specification has more predictive power when both predictors are included in the predictive regression.[9]

Among recent studies, our analysis here is most closely related to the works of Ang and Bekaert (2005) and Campbell and Yogo (2005). Ang and Bekaert examine stock-market predictability in both the univariate and multivariate settings, but their tests are developed for stationary regressors. Campbell and Yogo model regressors as nearly integrated, but their Bonferroni procedure is only applicable to univariate tests. Our tests allow for near unit roots in the regressors and can be used in a multivariate setting.

6.1 Data description

We use the annual, quarterly, and monthly NYSE/AMEX value-weighted index data from the Center for Research in Security Prices (CRSP) to construct series of stock returns and dividend–price ratios at the corresponding frequencies. Since earnings data are not available from CRSP, we construct earnings–price ratios using Standard and Poor's (S&P) 500 data.

[8] Even though some theoretical models suggest that certain predictor variables (such as the dividend–price ratio) are stationary (see Ang and Bekaert 2005, for example, for one such model), empirical tests often cannot reject the hypothesis that those variables contain a unit root.

[9] Such a comparison is particularly interesting since the results reported by Li (2001) suggest that the two specifications have similar abilities in explaining expected return movement.

We follow Campbell and Shiller (1988) and Campbell and Yogo (2005) in our construction of the dividend–price and earnings–price ratios. We compute the dividend–price ratio as dividends over the past year divided by the current price, and the earnings–price ratio as a moving average of earnings over the past ten years divided by the current price.

In our predictability regressions, we forecast the excess returns on stocks over a risk-free rate. We use the one-month and three-month T-bill rates, respectively, for our monthly and quarterly analyses. For annual data, we construct the risk-free return by compounding the returns on the three-month T-bill. We also use the three-month T-bill rate as a predictor variable in some of our analyses. All T-bill data are obtained from CRSP.

The full sample period is from 1926 to 2004, although we also consider the subperiods 1926–1994 and 1952–2004. We examine the subperiod 1926–1994 because the valuation ratios we consider have dropped to their lowest levels in history since the late 1990's, and it is interesting to see if such changes affect our results. For regressions that include the three-month T-bill rate as a predictor variable, we only consider the subperiod 1952–2004. This is because the U.S. Federal Reserve pegged interest rates before the Treasury Accord of 1951, and interest-rate data prior to 1952 are difficult to interpret.

In addition to the dividend–price ratio, the earnings–price ratio, and the three-month T-bill rate, we consider a few consumption-based predictors proposed by recent research. We obtain real per capita consumption of nondurables and services from the Bureau of Economic Analysis (BEA), U.S. Department of Commerce. Our definition of labor income follows that in Lettau and Ludvigson (2001a) and Santos and Veronesi (2006). In particular, it is defined as wages and salaries plus transfer payments plus other labor income minus personal contributions for social insurance minus taxes. The source of all these series is the BEA. We use an updated series for *cay*, constructed by Lettau and Ludvigson (2001a, b), from Martin Lettau's website. The appendix to Lettau and Ludvigson (2001a) details the construction of this variable. Due to data limitation, tests involving the consumption-based predictors are carried out for the subperiod 1952–2004, and at quarterly and annual frequencies only.

6.2 Valuation ratios and the risk-free rate as predictors

We first consider tests of stock-return predictability using the dividend–price ratio, the earnings–price ratio, and the three-month T-bill rate as predictor variables. Table 11 reports results of our subsampling-based tests of stock-market predictability. We use centered subsampling where the block size is selected based on a calibration rule. We carry out one-sided tests at the 5% level. Panel A reports results of t-tests in univariate predictions and Panel B reports results of the individual t-tests and joint Wald tests in bivariate predictions.

From Panel A, we see that the strongest evidence for predictability comes from the 1926–1994 subsample. With the exception of the dividend–price ratio at the monthly frequency, the null of no predictability is rejected at the 5% level in all other setups. The evidence for predictability is much weaker for the full sample and for

the 1952–2004 subsample. For the full sample, we find that only the earnings–price ratio at the quarterly frequency has significant predictive power at the 5% level. For the 1952–2004 subperiod, only the short-term interest rate at the monthly frequency is significant at 5%.

To compare our results with those reported by Campbell and Yogo (2005), we also carry out our tests for 1926–2002, the period Campbell and Yogo consider. For this period, we find that the dividend–price ratio (at the annual frequency) and the earnings–price ratio (at both the quarterly and monthly frequencies) are significant at the 5% level. Overall, our univariate results regarding the predictive power of the two valuation ratios are almost identical to those obtained from the Bonferroni Q-test that Campbell and Yogo propose; with respect to the short-term interest rate, the Bonferroni Q-test finds evidence for predictability in both quarterly and monthly data, but we find such evidence only at the monthly frequency.

Turning to the bivariate results on Panel B, we see evidence of predictability that is more evenly spread out across subsamples. In particular, in the 1952–2004 subperiod, we can now reject the null of no joint predictability for the earnings–price ratio and the short-term interest rate at all three frequencies. In addition, the combination of the dividend–price ratio and the short-term interest rate is also rejected at the monthly frequency. For the other two samples, we find evidence of joint predictability for the two valuation ratios at both the monthly frequency (for the 1926–1994 subperiod) and the quarterly frequency (for 1926–1994 and 1926–2004).

Comparing the results on Panels A and B, we see the value of performing joint tests. For example, in the 1952–2004 subsample, even though univariate tests find predictability only at the monthly frequency, joint tests uncover predictability at all three frequencies. At the same time, predictor variables that are individually significant in univariate tests do not necessarily become jointly significant; looking at the 1926–1994 subperiod and using annual data, even though the valuation ratios are each individually significant, they are not jointly significant.

Lamont (1998) examines the joint predictive power of the dividend–price ratio and the earnings–price ratio for future stock returns. In particular, he finds that when both valuation ratios are used as predictors, the dividend–price ratio remains positive and significant, but the earnings–price ratio becomes *negative* and significant. Since Lamont uses standard asymptotics for inference, but the valuation ratios are highly persistent and strongly correlated with returns, we apply our subsampling-based tests on Lamont's sample (quarterly Standard & Poor's Composite Index data from 1947 to 1994) to test his conclusion. As expected, we find that the point estimates are the same as those found by Lamont (1998). More importantly, the coefficient on the dividend–price ratio remains positive and significant and that on the earnings–price ratio negative and significant at the 5% level.

However, from the results reported in Table 11, Panel B, we see that Lamont's (1998) finding seems to be present only in the most recent sample period. We find the same positive–negative pattern on the coefficient estimates for the dividend–price and earnings–price ratios in the 1952–2004 osubperiod only. By contrast, in the 1926–1994 subperiod and in the full sample, it is the dividend–price ratio (rather than the

earnings–price ratio) that turns negative.

6.3 Consumption-based variables as predictors

One testable implication of consumption-based asset pricing models is that certain state variables implied by theory have predictive power for stock returns. Thus, researchers often use the finding of predictability as evidence in support of their models. A popular approach is to include both the state variable under consideration and the dividend–price ratio in the predictive regression, and then examine if the state variable has any marginal predictive power beyond that of the dividend–price ratio. Since previous tests are based on standard asymptotic theories, we reexamine their results for robustness using our subsampling-based approach.

The first two consumption-based predictor variables we consider are the consumption–wealth ratio (*cay*) proposed by Lettau and Ludvigson (2001a, 2001b) and the labor-income–consumption ratio (s^w) proposed by Santos and Veronesi (2006). Both variables depend on aggregate consumption and aggregate labor income.

Table 12 shows that both variables have predictive power for stock returns. Panel A reports results for annual data and Panel B reports results for quarterly data. From rows 1 and 2 of Panel A, we see that both *cay* and s^w have predictive power for stock returns at the annual frequency beyond that contained in the dividend–price ratio. In row 3 of Panel A, we see that both variables are individually significant in a bivariate predictive regression. This finding suggests that even though both variables depend on aggregate consumption and labor income, they each contain independent predictive power for stock returns. By examining the corresponding rows on Panel B, we see that the predictive power of s^w weakens but that of *cay* remains. This result is consistent with the findings of Santos and Veronesi (2006), who show that the predictive power of s^w lies in annual and lower frequencies.

We next turn to two habit-based predictors. The first is the log surplus consumption based on a linear habit, $\ln(C_t - X_{lin,t})$, and the second is the log surplus consumption based on a nonlinear habit, $\ln(C_t - X_{nonlin,t})$. $X_{lin,t}$ and $X_{nonlin,t}$ denote the level of linear and nonlinear habit, respectively, and are discussed in detail below. As Campbell and Cochrane (1999) and Li (2001) argue, at times when surplus consumption is low, consumers become conditionally more risk averse and demand a higher expected return on stocks. Campbell and Cochrane rely on calibration results to demonstrate how this negative relationship between surplus consumption and expected stock returns arises. Li uses standard asymptotics to show that this relationship is there in the data as well. In particular, Li examines a wide range of linear habit specifications and finds that some have performance similar to that of Campbell and Cochrane's (1999) nonlinear specification.

In particular, Li (2001) considers linear habit $X_{lin,t}$ of the form

$$X_{lin,t} = \alpha \left(\frac{1 - \varphi^m}{1 - \varphi^{Jm}} \right) \sum_{j=1}^{J} \varphi^{(j-1)m} C_{t-j},$$

where C denotes the level of aggregate consumption. The parameter φ controls the

persistence of the habit, for example, how quickly the effects of past consumption die out. The integer m denotes the number of months elapsed between periods $t-1$ and t. The integer $J \geq 1$ measures the duration of the habit, that is, the truncation point beyond which past consumption has no effect. The constant $0 < \alpha < 1$ controls the level of the habit relative to current consumption. We follow Li (2001) in setting $\alpha = 0.98$. Since Li shows that specifications with $\varphi \geq 0.99$ and J equal to 5 years perform the best, we focus on the case of $\varphi = 0.99$ and $J = 20$ quarters.

In their nonlinear habit-persistence model, Campbell and Cochrane (1999) do not specify $X_{nonlin,t}$ directly. Instead, they define the surplus-consumption ratio $S_t \equiv \frac{C_t - X_{nonlin,t}}{C_t}$, and postulate that the log surplus-consumption ratio, s_t, follows a heteroskedastic AR(1) process,

$$s_{t+1} = (1-\phi)\overline{s} + \phi s_t + \lambda(s_t)(c_{t+1} - c_t - g),$$

where ϕ, g, and $\overline{s}$ are parameters, and c_t denotes $\ln C_t$. The function $\lambda(s_t)$ controls the sensitivity of s_{t+1} to contemporaneous consumption. We follow Campbell and Cochrane's (1999) choice of parameter values and the specification for $\lambda(s_t)$.[10] We assume that s_1 is equal its steady state value $\overline{s}$. To weaken the dependence of our results on this choice of s_1, we begin our simulation 20 quarters before the start date of our predictive regressions (i.e., we begin our simulation in 1947Q1), so that we can drop the first 20 quarters of observations when we examine the predictive power of $\ln(C_t - X_{nonlin,t})$ from 1952Q1 to 2004Q4.

When we include surplus consumption together with the dividend–price ratio in the predictive regression, the coefficients on both the linear and nonlinear specifications have the expected negative sign—higher surplus consumption forecasts lower future returns. However, the coefficients tend not to be significant, with the exception of the nonlinear specification $\ln(C_t - X_{nonlin,t})$ at the annual frequency.

The "head-to-head" comparison between $\ln(C_t - X_{lin,t})$ and $\ln(C_t - X_{nonlin,t})$ is more interesting. When we include both surplus-consumption variables in the predictive regression, the coefficient on the nonlinear habit remains negative and significant, but that on the linear habit turns significantly positive in annual data, and insignificantly positive in quarterly data. Campbell and Cochrane (1999) use simulations to show that their nonlinear habit specification can overcome many difficulties that the more traditional, linear specification faces in asset pricing. Here, we show that the nonlinear specification also has superior forecasting power for stock returns.

Finally, we investigate if the habit-based variables—the consumption–wealth ratio, *cay*, and the labor-income–consumption ratio, s^w—have the same information content in terms of stock-return prediction. When we include both *cay* and $\ln(C_t - X_{nonlin,t})$ in the predictive regression, we see that both variables remain significant with the expected signs. On the other hand, s^w and $\ln(C_t - X_{nonlin,t})$ become individually insignificant, although they are still jointly significant when we carry out

[10] Specifically, we assume that $\gamma = 2$, $\phi = 0.87$, $g = 1.89\%$, and $\sigma_c = 1.5\%$, where ϕ, g, and σ_c are annualized values. See Campbell and Cochrane (1999) for a detailed discussion on the specification of $\lambda(s_t)$.

the test at the annual frequency. These results indicate that *cay* has more incremental information than s^w does, relative to the information already contained in $\ln(C_t - X_{nonlin,t})$.

7 Conclusion

Many popular stock-return predictor variables are highly persistent. Even though there may be plausible theoretical grounds to assume that some of these variables are stationary, we often cannot empirically reject the notion that these variables contain a unit root. Thus, it is natural to model these predictor variables as being nearly integrated. At the same time, more than one such predictor variable is of interest to financial economists. Thus, in testing for stock-return predictability, it is important to be able to carry out tests in a multivariate setting.

Some previous studies on stock-return predictability has modelled the predictor variables as nearly integrated, but their methods are only applicable to univariate regressions. Other studies have examined multivariate forecasts, but the regressors are assumed to be stationary. We add to this literature by proposing a subsampling-based test that can incorporate nearly integrated regressors in a multivariate setting.

By avoiding the need to estimate various nuisance parameters (most notably, the degree of persistence of the predictor variables), our subsampling-based approach overcomes the main difficulty in extending univariate tests that allow for nearly integrated predictors (such as the Bonferroni tests) to a multivariate setting.

We carry out extensive simulations to demonstrate that our subsampling-based tests have desirable size and power properties. On the other hand, the performance of the conventional t-tests and Wald tests that use standard asymptotics deteriorate when the persistence of the predictor variables approaches unity. These results indicate that conclusions obtained using these tests (such as the findings in Ang and Bekaert, 2005) can be misleading if the regressors in question in fact contain near unit roots.

Our subsampling-based tests suggest that the evidence for stock-return predictability is quite strong when we use valuation ratios and the risk-free rate as predictors. In univariate tests, we find that the evidence is more concentrated in the subperiod from 1926–1994. In bivariate tests, we find evidence for predictability in the full sample period 1926–2004, and in the subperiods 1926–1994 and 1952–2004. We also demonstrate the value in being able to carry out joint tests—there are numerous cases where univariate tests are insignificant but joint tests are not.

We also consider the predictive power of a number of consumption-based variables. We find that both the consumption–wealth ratio (*cay*) proposed by Lettau and Ludvigson (2001a, 2001b) and the labor-income–consumption ratio (s^w) proposed by Santos and Veronesi (2006) have significant marginal information to forecast returns beyond that already contained in the dividend–price ratio. We also see that the information content of these two variables are not the same, as both variables are individually significant in tests using annual data. Between the surplus-consumption variables based on linear and nonlinear habits, we find that the nonlinear specification

has stronger predictive power.

There are several directions for further research. Although we have examined a number of consumption-based models in this study, we have not considered models that explicitly take housing or durables consumption into account. Lustig and Van Nieuwerburgh (2005), Piazzesi et al. (2006), and Yogo (2006) show that these models have desirable asset pricing implications. It would be interesting to compare these models' time-series predictive power for stock returns. Second, although it is easy to see that our theory applies to more than two regressors and to forecast horizons longer than the annual frequency, we have to carry out further simulations to investigate the performance of our tests in such cases.

Appendix I: The main theoretical results

This appendix proves that the subsampling approach provides valid approximations to the limiting distributions of $t(\beta_0)$ and $Q(\beta_0, 1)$ defined by (4) and (6), respectively. A major source of complication for the proof of the following theorem is that $t(\beta_0)$ and $Q(\beta_0, 1)$ are not identically distributed when they use blocks of data starting from s. This follows because the test statistics depend on v_{s-1} which is not identically distributed for each s. However, since the initial variable v_{s-1} does affect the limiting distributions of the test statistics as $b \to \infty$, they are identically distributed for large b which makes the following theorem hold true.

Theorem A.1 *Let $b = O(T^\eta)$ with $\frac{1}{2} < \eta \leq \frac{2}{3}$. Denote the limiting distribution functions of $t(\beta_0)$ and $Q(\beta_0, 1)$ as $J^t(\cdot)$ and $J^Q(\cdot)$, respectively. Suppose that $u_t = (u_{vt}, e_t)'$ satisfies Assumption 1. Then,*
(i) $\sup_{x \in R} \mid L_T^z(x) - J^z(x) \mid \xrightarrow{p} 0$ $(z = t,\ Q)$ *as* $T \to \infty$;
(ii) For $\lambda \in (0, 1)$, let $c^z_{T(1-\lambda)} = \inf\{x : L_T^z(x) \geq 1 - \lambda\}$ and $c^z_{(1-\lambda)} = \{x : J^z(x) = 1 - \lambda\}$. Then,

$$c^z_{T(1-\lambda)} \xrightarrow{p} c^z_{(1-\lambda)} \text{ as } T \to \infty.$$

Proof. (i) We will prove the results only for $Q(\beta_0, 1)$ because those for $t(\beta_0)$ can be proven in the same way. Consider $\Lambda_{b,s} = Q_{b,s}(\beta_0, 1)_{|v_{s-1}=0}$, which is the test statistic $Q_{b,s}(\beta_0, 1)$ conditional on the restriction $v_{s-1} = 0$. The empirical distribution function using $\Lambda_{b,s}$ is written as $M_T(x) = \frac{1}{T-b+1} \sum_{s=1}^{T-b+1} 1\{\Lambda_{b,s} \leq x\}$. Because $\{(u_{vt}, e_t)'\}$ is strictly stationary and because $\Lambda_{b,s}$ is a measurable function of only $\{e_t\}_{t=s+1}^{s+b-1}$ and $\{u_{vt}\}_{t=s}^{s+b-2}$ due to the conditioning $v_{s-1} = 0$, $\Lambda_{b,s}$ has the same distribution for each s and b. Thus, we may write

$$\begin{aligned} \left|M_T(x) - J^Q(x)\right| &\leq \left| \frac{1}{T-b+1} \sum_{s=1}^{T-b+1} \left(1\{\Lambda_{b,s} \leq x\} - E1\{\Lambda_{b,1} \leq x\}\right) \right| \\ &\quad + \left|E1\{\Lambda_{b,1} \leq x\} - J^Q(x)\right| \\ &= A_T(x) + B_T(x), \text{ say.} \end{aligned}$$

Since $1\{\Lambda_{b,s} \leq x\}$ is a measurable function of $\{e_t\}_{t=s+1}^{s+b-1}$ and $\{u_{vt}\}_{t=s}^{s+b-2}$, the law of large numbers for mixing processes stated in Lemma A.3 yields $A_T(x) \xrightarrow{a.s.} 0$ for each $x \in R$. From this follows $\sup_{x\in R} A_T(x) \xrightarrow{a.s.} 0$ due to Lemma A.4. In addition, $B_T(x)$ converges to zero uniformly in x since $E1\{\Lambda_{b,1} \leq x\} \to J^Q(x)$ for each x and $J^Q(\cdot)$ is a continuous function (see Lemma 3 of Chow and Teicher, 1988, p.265). Thus, we have proven that

$$\sup_{x\in R} \left|M_T(x) - J^Q(x)\right| \xrightarrow{a.s.} 0. \tag{A.1}$$

In order to prove a similar result for $L_T^Q(\cdot)$, consider an inequality

$$\sup_{x\in R} \left|L_T^Q(x) - J^Q(x)\right| \leq \sup_{x\in R} \left|L_T^Q(x) - M_T(x)\right| + \sup_{x\in R} \left|M_T(x) - J^Q(x)\right|.$$

The second term on the right-hand side of this inequality has been shown to be $o_{a.s.}(1)$. In order to show that the first term is negligible in probability, assume that

$$Q_{b,s}(\beta_0, 1) = \Lambda_{b,s} + \theta_{Tbs} \tag{A.2}$$

where

$$\max_s |\theta_{Tbs}| = o_p(1). \tag{A.3}$$

Then, letting $\varsigma_{Tb} = \max_s |\theta_{Tbs}|$,

$$\begin{aligned}
& \sup_{x\in R} \left|L_T^Q(x) - M_T(x)\right| \\
= & \sup_{x\in R} \left| \frac{1}{T-b+1} \sum_{s=1}^{T-b+1} \left(1\{Q_{b,s}(\beta_0,1) \leq x\} - 1\{\Lambda_{b,s} \leq x\}\right) \right| \\
\leq & \sup_{x\in R} \left| \frac{1}{T-b+1} \sum_{s=1}^{T-b+1} \left(1\{\Lambda_{b,s} \leq x+\varsigma_{Tb}\} - 1\{\Lambda_{b,s} \leq x\}\right) \right| \\
\leq & \sup_{x\in R} \left|M_T(x+\varsigma_{Tb}) - J^Q(x+\varsigma_{Tb})\right| + \sup_{x\in R} \left|J^Q(x+\varsigma_{Tb}) - J^Q(x)\right| \\
& + \sup_{x\in R} \left|M_T(x) - J^Q(x)\right|.
\end{aligned}$$

Using (A.1) and (A.3), we find that the first and third terms in the last inequality converge to zero in probability. The second term also converges to zero in probability since $J^Q(\cdot)$ is continuous. Thus, the proof will be complete if relations (A.2) and (A.3) are shown to hold. These are proven in Lemma A.5, so the stated result follows. ■

(ii) Since $L_T^z(\cdot)$ and $J^z(\cdot)$ are nondecreasing functions and $J^z(\cdot)$ is continuous, the result follows from part (i) and Lemma A.6 in Appendix II. ■

Let

$$\begin{aligned}
L_T^{t\bullet}(x) &= \frac{1}{T-b+1} \sum_{s=1}^{T-b+1} 1\{t_{b,s}(\hat{\beta}) \leq x\} \text{ and} \\
L_T^{Q\bullet}(x) &= \frac{1}{T-b+1} \sum_{s=1}^{T-b+1} 1\{Q_{b,s}(\tilde{\beta}, 1) \leq x\},
\end{aligned}$$

where $t_{b,s}(\hat{\beta})$ and $Q_{b,s}(\tilde{\beta},1)$ are defined by (10) and (11), respectively. The following theorem states that the empirical distribution functions, $L_T^{t\bullet}(x)$ and $L_T^{Q\bullet}(x)$, provide valid approximations to the limiting distributions of $t(\beta_0)$ and $Q(\beta_0,1)$, respectively.

Theorem A.2 *Suppose that the same assumptions for Theorem A.1 hold. Then,*
(i) $\sup_{x\in R} \mid L_T^{\bullet z}(x) - J^z(x) \mid \xrightarrow{p} 0$ $(z = t,\ Q)$ *as* $T \to \infty$;
(ii) For $\lambda \in (0,1)$, *let* $c_{T(1-\lambda)}^{\bullet z} = \inf\{x : L_T^{\bullet z}(x) \geq 1-\lambda\}$ *and* $c_{(1-\lambda)}^{z} = \{x : J^z(x) = 1-\lambda\}$. *Then,*

$$c_{T(1-\lambda)}^{\bullet z} \xrightarrow{p} c_{(1-\lambda)}^{z} \text{ as } T \to \infty.$$

Proof. (i) We prove the results only for $L_T^{t\bullet}(x)$. Using relation (12), write

$$L_T^{t\bullet}(x) = \frac{1}{T-b+1} \sum_{s=1}^{T-b+1} 1\{t_{b,s}(\beta_0) + \xi_{b,s,n} \leq x\},$$

where $\xi_{b,s,T} = -\frac{b}{T} \frac{T(\hat{\beta}-\beta_0)}{\sqrt{\hat{\sigma}_{b,s,y}^2 \left(\sum_{t=s+1}^{s+b-1} \left(x_{t-1} - \bar{x}_{-1,b,s}\right)^2 / b^2\right)^{-1}}}$. We deduce from this relation that for $\varepsilon > 0$

$$L_T^t(x-\varepsilon)1\{E_T\} \leq L_T^{t\bullet}(x)1\{E_T\} \leq L_T^t(x+\varepsilon), \tag{A.4}$$

where $1\{E_T\}$ is the indicator of the event $\{\mid \xi_{b,s,T} \mid \leq \varepsilon\}$. Since the event E_n holds with probability approaching one as T goes to infinity, (A.4) implies that the relation

$$L_T^t(x-\varepsilon) \leq L_T^{t\bullet}(x) \leq L_T^t(x+\varepsilon)$$

holds with probability tending to one. Because

$$L_T^t(x-\varepsilon) - J^t(x) \leq L_T^{t\bullet}(x) - J^t(x) \leq L_T^t(x+\varepsilon) - J^t(x),$$

we have

$$\sup_{x\in R} \mid L_T^{t\bullet}(x) - J^t(x) \mid \leq \max\left(\sup_{x\in R} \mid L_T^t(x+\varepsilon) - J^t(x) \mid,\ \sup_{x\in R} \mid L_T^t(x+\varepsilon) - J^t(x) \mid\right).$$

Thus, we obtain the stated result by sending $\varepsilon \to 0$ and using part (i) of Theorem A.1. ■
(ii) This follows as in the proof of part (ii) of Theorem A.1. ■

Appendix II: Technical lemmas

This appendix collects some technical lemmas used to prove Theorem A.1. The first result we require is the strong law of large numbers for the function of mixing processes with the number of arguments growing with the sample size.

Lemma A.3 *Let $Y_s = g(z_s, z_{s+1}, \ldots, z_{s+b})$, where b is an integer satisfying $b = O(T^\eta)$ with $0 < \eta < 1$. Suppose for some constants c and d that*

(a) $\sup_{s\geq 1} \|Y_s\|_{2+\varepsilon} < c < \infty$ for $\varepsilon > 0$;

(b) $\{z_s\}$ is strong mixing with its mixing coefficients $\alpha_{z,m}$ satisfying, for $\epsilon > 0$,

$$\sum_{m=1}^{\infty} \alpha_{z,m}^{\epsilon/(2+\epsilon)} < \infty.$$

Let $S_{T-b} = \sum_{s=1}^{T-b} (Y_s - E(Y_s))$. Then, as $T \to \infty$,

$$\frac{1}{T-b} S_{T-b} \xrightarrow{a.s.} 0.$$

Proof. See Lemma A.2 of Choi and Chue (2004). ■

The following lemma, taken from Davidson (1994, p. 332), states that the pointwise a.s. convergence of the empirical distribution function is sufficient for the Glivenko–Cantelli lemma. Davidson states the lemma for iid random variables, but its use for dependent random variables is easily verified as long as the pointwise strong law of large numbers holds for them.

Lemma A.4 *For a set of random variables $\{Y_1(\omega), \ldots, Y_T(\omega)\}$ defined on the probability space $(\Omega, \mathfrak{F}, P)$, define the empirical distribution function as*

$$F_T(x) = \frac{1}{T} \sum_{t=1}^{T} 1\{Y_t(\omega) \leq x\}.$$

If $F_T(x) \overset{a.s.}{\to} F(x)$ pointwise for $x \in R$ as $T \to \infty$, then

$$\sup_{x\in R} |F_T(x) - F(x)| \overset{a.s.}{\to} 0 \text{ as } T \to \infty.$$

In the following lemma, we prove relation (A.3) to support the proof of Theorem A.1 in Appendix I. The test statistic $Q_{b,s}(\beta_0, 1)$ can be written as

$$\begin{aligned} Q_{b,s}(\beta_0, 1) \;=\; g\Bigg(& \frac{\sum_{t=s+1}^{s+b-1} (v_{t-1} - \bar{v}_{-1,b,s})^2}{b^2}, \frac{\sum_{t=s+1}^{s+b-1} (v_{t-1} - \bar{v}_{-1,b,s})(\Delta v_t - \overline{\Delta v})}{b^{3/2}}, \\ & \frac{\sum_{t=s+1}^{s+b-1} (\Delta v_t - \overline{\Delta v})^2}{b}, \frac{\sum_{t=s+1}^{s+b-1} (v_{t-1} - \bar{v}_{-1,b,s}) e_t}{b}, \\ & \frac{\sum_{t=s+1}^{s+b-1} (\Delta v_t - \overline{\Delta v}) e_t}{b^{1/2}}, \frac{\sum_{t=s+1}^{s+b-1} e_t^2}{b} \Bigg), \end{aligned}$$

where $g(\cdot)$ is a continuous function, $\bar{v}_{-1,b,s} = \frac{\sum_{t=s+1}^{s+b-1} v_{t-1}}{b-1}$ and $\overline{\Delta v} = \frac{\sum_{t=s+1}^{s+b-1} \Delta v_t}{b-1}$. Let $v_{t-1} = \rho^{t-s} v_{s-1} + w_{t-1}$ with $w_{t-1} = u_{v,t-1} + \rho u_{v,t-2} + \cdots + \rho^{t-1-s} u_{v,s}$. Then, extracting

terms involving initial value v_{s-1}, we obtain

$$
\begin{aligned}
& Q_{b,s}(\beta_0, 1) \qquad\qquad (A.5)\\
= \; & h\Bigg(v_{s-1}^2 \frac{\sum_{t=s+1}^{s+b-1} \left(\rho^{t-s} - \bar{\rho}_b\right)^2}{b^2},\; v_{s-1} \frac{\sum_{t=s+1}^{s+b-1} \left(\rho^{t-s} - \bar{\rho}_b\right) w_{t-1}}{b^2},\\
& v_{s-1}^2 \frac{(\rho-1)\sum_{t=s+1}^{s+b-1} \left(\rho^{t-s} - \bar{\rho}_b\right)^2}{b^{3/2}},\; v_{s-1} \frac{\sum_{t=s+1}^{s+b-1} \left(\rho^{t-s} - \bar{\rho}_b\right) u_{vt}}{b^{3/2}},\\
& v_{s-1} \frac{(\rho-1)\sum_{t=s+1}^{s+b-1} \left(\rho^{t-s} - \bar{\rho}_b\right) w_{t-1}}{b^{3/2}}, v_{s-1}^2 \frac{(\rho-1)^2 \sum_{t=s+1}^{s+b-1} \left(\rho^{t-s} - \bar{\rho}_b\right)^2}{b},\\
& v_{s-1} \frac{(\rho-1)\sum_{t=s+1}^{s+b-1} \left(\rho^{t-s} - \bar{\rho}_b\right) u_{vt}}{b}, v_{s-1} \frac{(\rho-1)^2 \sum_{t=s+1}^{s+b-1} \left(\rho^{t-s} - \bar{\rho}_b\right) w_{t-1}}{b},\\
& v_{s-1} \frac{\sum_{t=s+1}^{s+b-1} \left(\rho^{t-s} - \bar{\rho}_b\right) e_t}{b}, v_{s-1} \frac{(\rho-1)\sum_{t=s+1}^{s+b-1} \left(\rho^{t-s} - \bar{\rho}_b\right) e_t}{b^{1/2}},\; r_{b,s} \Bigg),
\end{aligned}
$$

where $\bar{\rho}_b = \frac{\sum_{t=s+1}^{s+b-1} \rho^{t-s}}{b-1}$. In relation (A.5), $r_{b,s}$ denotes a vector of terms that do not involve v_{s-1}, These are functions of only $\{e_t\}_{t=s+1}^{s+b-1}$ and $\{u_{vt}\}_{t=s}^{s+b-2}$ and are $O_p(1)$.

Lemma A.5 *Let $b = O(T^\eta)$ with $\frac{1}{2} < \eta \leq \frac{2}{3}$. Then, $Q_{b,s}(\beta_0, 1) = \Lambda_{b,s} + \theta_{Tbs}$ and $\max_s |\theta_{Tbs}| = o_p(1)$.*

Proof. A Taylor expansion of (A.5) gives

$$
\begin{aligned}
\Psi_{b,s} \;=\;& h(0,\ldots,0,r_{b,s}) \\
&+\frac{\partial h}{\partial x_1}_{|x_1=x_1^*}\left(\frac{v_{s-1}}{T^{1/2}}\right)^2\frac{T}{b^2}\sum_{t=s+1}^{s+b-1}\left(\rho^{t-s}-\bar{\rho}_b\right)^2 \\
&+\frac{\partial h}{\partial x_2}_{|_{x_2=x_2^*}}\frac{v_{s-1}}{T^{1/2}}\frac{T^{1/2}}{b}\frac{\sum_{t=s+1}^{s+b-1}\left(\rho^{t-s}-\bar{\rho}_b\right)w_{t-1}}{b} \\
&+\frac{\partial h}{\partial x_3}_{|_{x_3=x_3^*}}\left(\frac{v_{s-1}}{T^{1/2}}\right)^2\frac{T}{b^{3/2}}(\rho-1)\sum_{t=s+1}^{s+b-1}\left(\rho^{t-s}-\bar{\rho}_b\right)^2 \\
&+\frac{\partial h}{\partial x_4}_{|x_4=x_4^*}\frac{v_{s-1}}{T^{1/2}}\frac{T^{1/2}}{b^{3/2}}\sum_{t=s+1}^{s+b-1}\left(\rho^{t-s}-\bar{\rho}_b\right)u_{vt} \\
&+\frac{\partial h}{\partial x_5}_{|x_5=x_5^*}\frac{v_{s-1}}{T^{1/2}}\frac{T^{1/2}}{b^{1/2}}(\rho-1)\frac{\sum_{t=s+1}^{s+b-1}\left(\rho^{t-s}-\bar{\rho}_b\right)w_{t-1}}{b} \\
&+\frac{\partial h}{\partial x_6}_{|x_6=x_6^*}\left(\frac{v_{s-1}}{T^{1/2}}\right)^2\frac{T}{b}(\rho-1)^2\sum_{t=s+1}^{s+b-1}\left(\rho^{t-s}-\bar{\rho}_b\right)^2 \\
&+\frac{\partial h}{\partial x_7}_{|x_7=x_7^*}\frac{v_{s-1}}{T^{1/2}}\frac{T^{1/2}}{b}(\rho-1)\sum_{t=s+1}^{s+b-1}\left(\rho^{t-s}-\bar{\rho}_b\right)u_{vt} \\
&+\frac{\partial h}{\partial x_8}_{|x_8=x_8^*}\frac{v_{s-1}}{T^{1/2}}T^{1/2}(\rho-1)^2\frac{\sum_{t=s+1}^{s+b-1}\left(\rho^{t-s}-\bar{\rho}_b\right)w_{t-1}}{b} \\
&+\frac{\partial h}{\partial x_8}_{|x_9=x_9^*}\frac{v_{s-1}}{T^{1/2}}\frac{T^{1/2}}{b}\sum_{t=s+1}^{s+b-1}\left(\rho^{t-s}-\bar{\rho}_b\right)e_t \\
&+\frac{\partial h}{\partial x_9}_{|x_{10}=x_{10}^*}\frac{v_{s-1}}{T^{1/2}}\frac{T^{1/2}}{b^{1/2}}(\rho-1)\sum_{t=s+1}^{s+b-1}\left(\rho^{t-s}-\bar{\rho}_b\right)e_t,
\end{aligned}
\tag{A.6}
$$

where x_i denotes the i-th argument of function $h(\cdot)$ and x_i^* lies on the line joining 0 and x_i. Obviously, $h(0,\ldots,0,z_{b,s})=\Lambda_{b,s}$. Thus, we need to show that the second to eleventh terms on the right-hand side of equation (A.6) are $o_p(1)$. Due to the weak convergence results for nearly $I(1)$ processes (cf. Phillips, 1988),

$$
\max_s\left|\frac{v_{s-1}}{T^{1/2}}\right| = O_p(1). \tag{A.7}
$$

In addition,

$$\begin{aligned}
&\sum_{t=s+1}^{s+b-1}\left(\rho^{t-s}-\bar{\rho}_b\right)^2 \\
&= \sum_{t=1}^{b-1}\left(e^{ct/T}-\frac{1}{b-1}\sum_{t=1}^{b-1}e^{ct/T}\right)^2 \\
&= \sum_{t=1}^{b-1}\left(1+\frac{ct}{T}+a_{tT}-\frac{1}{b-1}\sum_{t=1}^{b-1}\left(1+\frac{ct}{T}+a_{tT}\right)\right)^2 \\
&= \sum_{t=1}^{b-1}\left(\frac{ct}{T}-\frac{1}{b-1}\sum_{t=1}^{b-1}\frac{ct}{T}\right)^2+o(1) \\
&= \frac{c^2}{T^2}\sum_{t=1}^{b-1}\left(t-\frac{1}{b-1}\sum_{t=1}^{b-1}t\right)^2+o(1) \\
&= O(\frac{b^3}{T^2}),
\end{aligned} \tag{A.8}$$

which shows that $\sum_{t=s+1}^{s+b-1}\left(\rho^{t-s}-\bar{\rho}_b\right)^2$ is bounded in the limit due to the given assumption on b. Moreover, using the strong mixing inequality[11] for $\{e_t\}$, we obtain for a constant C

$$\begin{aligned}
&Var\left(\sum_{t=1}^{b-1}\left(\rho^t-\bar{\rho}_b\right)e_t\right) \\
&= \sigma_y^2\sum_{t=1}^{b-1}(\rho^t-\bar{\rho}_b)^2+2\sum_{m=1}^{b-2}\sum_{t=m+1}^{b-1}(\rho^t-\bar{\rho}_b)(\rho^{t-m}-\bar{\rho}_b)\sigma_m \\
&\leq \sigma_y^2\sum_{t=1}^{b-1}(\rho^t-\bar{\rho}_b)^2+2C\sum_{m=1}^{b-2}\alpha_m^{\epsilon/(2+\epsilon)}\sum_{t=m+1}^{b-1}\left|\rho^t-\bar{\rho}_b\right|\left|\rho^{t-m}-\bar{\rho}_b\right| \\
&\leq \sigma_y^2\sum_{t=1}^{b-1}(\rho^t-\bar{\rho}_b)^2+2C\sum_{m=1}^{b-2}\alpha_m^{\epsilon/(2+\epsilon)}\sqrt{\sum_{t=m+1}^{b-1}(\rho^t-\bar{\rho}_b)^2}\sqrt{\sum_{t=m+1}^{b-1}(\rho^{t-m}-\bar{\rho}_b)^2} \\
&\leq \sigma_y^2\sum_{t=1}^{b-1}(\rho^t-\bar{\rho}_b)^2+2C\left(\sum_{t=1}^{b-1}\left(\rho^t-\bar{\rho}_b\right)^2\right)\sum_{m=1}^{b-2}\alpha_m^{\epsilon/(2+\epsilon)}=O(1),
\end{aligned} \tag{A.9}$$

where $Cov(e_t, e_{t\pm m}) = \sigma_m$. The first inequality in (A.9) uses the strong mixing inequality, and the second uses the Cauchy–Schwarz inequality. The last equality is based on (A.8). Relation (A.9) implies that

$$\sum_{t=s+1}^{s+b-1}\left(\rho^{t-s}-\bar{\rho}_b\right)e_t=O_p(1) \tag{A.10}$$

[11] The inequality is $|Cov(u_{yt}, u_{y,t+m})| \leq 2(2^{1-1/(2+\epsilon)}+1)\alpha_{u,m}^{\epsilon/(2+\epsilon)}\left\|u_{yt}\right\|_{2+\epsilon}\left\|u_{y,t+m}\right\|_{2+\epsilon}$. For this, see Davidson (1994, p. 212).

Likewise,

$$\sum_{t=s+1}^{s+b-1} \left(\rho^{t-s} - \bar{\rho}_b\right) u_{vt} = O_p(1) \tag{A.11}$$

Furthermore, using $\max_{s+1\leq t\leq s+b-1} \left|\frac{w_{t-1}}{b^{1/2}}\right| = O_p(1)$ and relation (A.8), we obtain

$$\begin{aligned} & \frac{\sum_{t=s+1}^{s+b-1} \left(\rho^{t-s} - \bar{\rho}_b\right) w_{t-1}}{b} \\ \leq & \max_{s+1\leq t\leq s+b-1} \left|\frac{w_{t-1}}{b^{1/2}}\right| \frac{\sum_{t=s+1}^{s+b-1} \left|\rho^{t-s} - \bar{\rho}_b\right|}{b^{1/2}} \\ \leq & \max_{s+1\leq t\leq s+b-1} \left|\frac{w_{t-1}}{b^{1/2}}\right| \sqrt{\sum_{t=s+1}^{s+b-1} \left(\rho^{t-s} - \bar{\rho}_b\right)^2} = O_p(1). \end{aligned} \tag{A.12}$$

Note that the second inequality of (A.12) uses the Cauchy–Schwarz inequality. Since $\rho - 1 = O(\frac{1}{T})$ and $\frac{\partial h}{\partial x_i}|_{x_i=o_p(1)} = O_p(1)$ for $i = 1, \ldots, 9$, relations (A.7), (A.8), (A.10) and (A.11) imply that the second to eleventh terms on the right-hand side of equation (A.6) are $o_p(1)$ uniformly in s as required. This completes the proof. ■

The following lemma is used to prove part (ii) of Theorem A.1.

Lemma A.6 *Let* $H_n^{\leftarrow}(y) = \inf\{x : H_n(x) \geq y\}$ *and* $H^{\leftarrow}(y) = \inf\{x : H(x) \geq y\}$. *If (a)* $H_n(\cdot)$ *and* $H(\cdot)$ *are nondecreasing functions, (b)* $H(\cdot)$ *is continuous, and (c)* $\sup_{x\in R} |H_n(x) - H(x)| \xrightarrow{p} 0$ *as* $n \to \infty$, *then* $H_n^{\leftarrow}(y) \xrightarrow{p} H^{\leftarrow}(y)$ *for any* $y \in R$ *as* $n \to \infty$.

Proof. Let $g(x) = H(x) + \sup_{x\in R} |\delta_n(x)|$ and $f(x) = H(x) - \sup_{x\in R} |\delta_n(x)|$, where $\delta_n(x) = H_n(x) - H(x)$. Functions $g(\cdot)$ and $f(\cdot)$ are continuous in x and nondecreasing with probability one. Thus, for $y \in R$, we have with probability one

$$f^{\leftarrow}(y) \leq H^{\leftarrow}(y) \leq g^{\leftarrow}(y)$$

and

$$f^{\leftarrow}(y) \leq H_n^{\leftarrow}(y) \leq g^{\leftarrow}(y),$$

which imply

$$|H_n^{\leftarrow}(y) - H^{\leftarrow}(y)| \leq g^{\leftarrow}(y) - f^{\leftarrow}(y).$$

Since $\sup_{x\in R} |\delta_n(x)| \xrightarrow{p} 0$, $g^{\leftarrow}(y) - f^{\leftarrow}(y) \xrightarrow{p} 0$ as $n \to \infty$, from which follows the stated result. ■

Appendix III: Why does subsampling perform poorly for the $Q(\beta_0, 1)$ test?

When $c \neq 0$, the simulation results in Section 5 indicate that subsampling does not work well for the $Q(\beta_0, 1)$ test. This section tries to explain this. To this end, consider

the empirical distribution function

$$L_T^\tau(x) = \frac{1}{T-b+1}\sum_{s=1}^{T-b+1} 1\{\tau_{b,s} \leq x\},$$

where $\tau_{b,s}$ is either $t_{b,s}(\beta_0)$ or $Q_{b,s}(\beta_0, 1)$. The limiting distribution function of $\tau_{T,1}$ is denoted by $J^\tau(\cdot)$. The mean-squared error of $L_T(x)$ in estimating $J(x)$ is given by

$$\begin{aligned}
& E\left(L_T^\tau(x) - J^\tau(x)\right)^2 \\
= \; & E\left(\frac{1}{T-b+1}\sum_{s=1}^{T-b+1}\left(1\{\tau_{b,s} \leq x\} - E1\{\tau_{b,s} \leq x\}\right)\right. \\
& \left. + \frac{1}{T-b+1}\sum_{s=1}^{T-b+1} E1\{\tau_{b,s} \leq x\} - J^\tau(x)\right)^2 \\
= \; & E\left(\frac{1}{T-b+1}\sum_{s=1}^{T-b+1}\left(1\{\tau_{b,s} \leq x\} - E1\{\tau_{b,1} \leq x\}\right)\right)^2 \\
& + \left(\frac{1}{T-b+1}\sum_{s=1}^{T-b+1} E1\{\tau_{b,s} \leq x\} - J^\tau(x)\right)^2 \\
= \; & A_{T,b}(x) + B_{T,b}(x), \text{ say.}
\end{aligned}$$

If $b/T \to 0$ and Assumption 1 holds, it can be shown that $A_{T,b}(x) = o(1)$. Thus, it is the squared bias, $B_{T,b}(x)$, that is asymptotically equivalent to the mean-squared error. In Table 13, we calculated the mean-squared error and the squared bias using the same data-generating scheme as for Table 3 with $T = 250$, $c = -5$, $\delta = -3$ and $b = 15, 27$. We chose x such that $J^\tau(x) = 0.95$. Table 13 shows that the $Q(\beta_0, 1)$ test has much higher mean-squared error and squared bias than the t-test, which implies that the subsampling critical values for the $Q(\beta_0, 1)$ test are unlikely to be as accurate as those for the t-test in estimating its limiting critical values. This explains why the $Q(\beta_0, 1)$ test does not perform well in finite samples relative to the t-test. Obviously, the culprit is its bias. Since $\frac{1}{T-b+1}\sum_{s=1}^{T-b+1} E1\{\tau_{b,s} \leq x\} \simeq P\{\tau_{b,1} \leq x\}$ for large b, the substantial bias results from the slow convergence of the $Q(\beta_0, 1)$ test to its limiting distribution.[12]

Table 13: Mean squared error and squared biases of the t- and $Q(\beta_0, 1)$ tests

Notes: 1. Data were generated by $y_t = \beta x_{t-1} + u_{yt}$; $\beta = 0$; $x_t = \exp(-\frac{5}{T})x_{t-1} + u_{vt}$; $u_{yt} = -3u_{vt} + e_t$; $u_{vt} \sim iid\ N(0,1)$; $e_t \sim iid\ N(0,1)$; $T = 250$. 2. Block sizes correspond to $T^{0.5}$ and $T^{0.6}$. 3. uncentered subsampling was used. 4. Results are based on 2,000 replications.

	t-test		$Q(\beta_0, 1)$	
	MSE	Bias^2	MSE	Bias^2
$b = 15$	0.0166	0.0140	0.9022	0.9022
$b = 27$	0.0144	0.0099	0.9025	0.9025

[12] We performed the same experiment with larger sample sizes and $c = -15$. The $Q(\beta_0, 1)$ test statistic did not improve with the increased sample sizes and performed worse with $c = -15$.

References

Andrews, D.W.K. (1991): "Heteroskedasticity and autocorrelation consistent covariance matrix estimation," *Econometrica* 59, 817–858.

Ang, A. and G. Bekaert (2005): "Stock return predictability: is it there?" Mimeo, Columbia University.

Basawa, I.V., A.K. Mallik, W.P. McCormick, J.H. Reeves, and R.L. Taylor (1991): "Bootstrap test of significance and sequential bootstrap estimation for unstable first order autoregressive processes," *Communications in Statistical Theory and Methods* 20, 1015–1026.

Bobkoski, M.J. (1983): *Hypothesis Testing in Nonstationary Time Series.* Ph.D. Thesis, University of Wisconsin, Department of Statistics.

Campbell, J.Y. (1987): "Stock returns and the term structure," *Journal of Financial Economics* 18, 373–399.

Campbell, J.Y. and J.H. Cochrane (1999): "By force of habit: a consumption-based explanation of aggregate stock market behavior," *Journal of Political Economy* 107, 205–251.

Campbell, J.Y. and R.J. Shiller (1988): "Stock prices, earnings, and expected dividends," *Journal of Finance* 43, 661–676.

Campbell, J.Y. and M. Yogo (2005): "Efficient tests of stock return predictability," *Journal of Financial Economics*, forthcoming.

Chan, N.H. and C.Z. Wei (1987): "Asymptotic inference for nearly nonstationary AR(1) processes," *Annals of Statistics* 15, 1050–1063

Choi, I. (2005a): "Inconsistency of bootstrap for nonstationary, vector autoregressive processes," *Statistics and Probability Letters* 75, 39–48.

Choi, I. (2005b): "Subsampling vector autoregressive tests of linear constraints," *Journal of Econometrics* 124, 55–89.

Choi, I. and T.K. Chue (2004): "Subsampling hypothesis tests for nonstationary panels with applications to exchange rates and stock prices," *Journal of Applied Econometrics*, forthcoming.

Chow, Y.S. and H. Teicher (1988): *Probability Theory: Independence, Interchangeability and Martingales.* Springer–Verlag, New York.

Datta, S. (1996): "On asymptotic properties of bootstrap for AR(1) processes," *Journal of Statistical Planning and Inference* 53, 361–374.

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

Elliott, G. and J.H. Stock (1994): "Inference in time series regression when the order of integration of a regressor is unknown," *Econometric Theory* 10, 672–700.

Fama, E.F. and K.R. French (1988): "Dividend yields and expected stock returns," *Journal of Financial Economics* 22, 3–24.

Fama, E.F. and K.R. French (1989): "Business conditions and expected returns on stocks and bonds," *Journal of Financial Economics* 25, 23–49.

Fama, E.F. and G.W. Schwert (1977): "Asset returns and inflation," *Journal of Financial Economics* 5, 115–146.

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

Keim, D.B. and R.F. Stambaugh (1986): "Predicting returns in the stock and bond markets," *Journal of Financial Economics* 17, 357–390.

Lamont, O. (1998): "Earnings and expected returns," *Journal of Finance* 53, 1563–1587.

Lanne, M. (2002): "Testing the predictability of stock returns," *Review of Economics and Statistics* 84, 407–415.

Lettau, M. and S. Ludvigson (2001a): "Consumption, aggregate wealth and expected stock returns," *Journal of Finance* 56, 815–849.

Lettau, M. and S. Ludvigson (2001b): "Resurrecting the (C)CAPM: a cross-sectional test when risk premia are time varying," *Journal of Political Economy* 109, 1238–1287.

Lewellen, J. (2004): "Predicting returns with financial ratios," *Journal of Financial Economics* 74, 209–235.

Li, Y. (2001): "Expected returns and habit persistence," *Review of Financial Studies* 14, 861–899.

Lustig, H. and S. Van Nieuwerburgh (2005): "Housing collateral, consumption insurance and risk premia: An empirical perspective," *Journal of Finance* 60, 1167–1219.

Phillips, P.C.B. (1987): "Towards a unified asymptotic theory for autoregression," *Biometrika* 74, 535–547.

Phillips, P.C.B. (1988): "Regression theory for near–integrated time series," *Econometrica* 56, 1021–1043.

Phillips, P.C.B., and S. Durlauf (1986): "Multiple time series regression with integrated processes," *Review of Economic Studies* 53, 473–496.

Piazzesi, M., M. Schneider, and S. Tuzel (2006): "Housing, consumption, and asset pricing," *Journal of Financial Economics*, forthcoming.

Politis, D.N., J.P. Romano, and M. Wolf (1999): *Subsampling.* Springer–Verlag, New York.

Romano, J.P. and M. Wolf (2001): "Subsampling intervals in autoregressive models with linear trend," *Econometrica* 69, 1283–1314.

Santos, T. and P. Veronesi (2006): "Labor income and predictable stock returns," *Review of Financial Studies* 19, 1–44.

Stambaugh, R.F. (1999): "Predictive regressions," *Journal of Financial Economics* 54, 375–421.

Torous, W., R. Valkanov, and S. Yan (2004): "On predicting stock returns with nearly integrated explanatory variables," *Journal of Business* 77, 937–966.

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

Wolf, M. (2000): "Stock returns and dividend yields revisited: a new way to look at an old problem," *Journal of Business and Economic Statistics* 18, 18–30.

Yogo, M. (2006): "A consumption-based explanation of expected stock returns," *Journal of Finance*, forthcoming.

Table 3: Finite-Sample Size and Power of the t- and $Q(\beta_0, 1)$ Tests: The Case of One Regressor

Notes: 1. Data were generated by $y_t = \beta x_{t-1} + u_{yt}$; $\beta = 0$; $x_t = \exp(\frac{c}{T})x_{t-1} + u_{vt}$; $u_{yt} = \delta u_{vt} + e_t$; $u_{vt} \sim iidN(0,1)$; $e_t \sim iidN(0,1)$. 2. The minimum-volatility rule was used for subsampling. 3. Uncentered subsampling was used. 4. Results are based on 2,000 replications. 5. One-sided test with $H_1 : \beta > 0$.

Panel A: Size

			Subsampling		$N(0,1)$	
			t-test	$Q(\beta_0, 1)$	t-test	$Q(\beta_0, 1)$
$T = 100$	$\delta = -0.5$	$c = 0$	0.065	0.061	0.158	0.056
		$c = -5$	0.032	0.014	0.107	0.015
		$c = -15$	0.023	0.002	0.081	0.003
	$\delta = -1.5$	$c = 0$	0.074	0.061	0.346	0.056
		$c = -5$	0.035	0.001	0.153	0.001
		$c = -15$	0.030	0.000	0.109	0.000
	$\delta = -3$	$c = 0$	0.080	0.061	0.414	0.056
		$c = -5$	0.035	0.000	0.182	0.000
		$c = -15$	0.023	0.000	0.118	0.000
$T = 250$	$\delta = -0.5$	$c = 0$	0.062	0.056	0.163	0.054
		$c = -5$	0.034	0.014	0.100	0.009
		$c = -15$	0.030	0.003	0.080	0.003
	$\delta = -1.5$	$c = 0$	0.065	0.056	0.355	0.054
		$c = -5$	0.029	0.001	0.167	0.001
		$c = -15$	0.021	0.000	0.111	0.000
	$\delta = -3$	$c = 0$	0.066	0.055	0.430	0.054
		$c = -5$	0.034	0.000	0.199	0.000
		$c = -15$	0.027	0.000	0.119	0.000
$T = 500$	$\delta = -0.5$	$c = 0$	0.059	0.049	0.174	0.060
		$c = -5$	0.039	0.015	0.109	0.018
		$c = -15$	0.026	0.003	0.084	0.005
	$\delta = -1.5$	$c = 0$	0.061	0.049	0.358	0.060
		$c = -5$	0.023	0.001	0.184	0.001
		$c = -15$	0.019	0.000	0.113	0.000
	$\delta = -3$	$c = 0$	0.061	0.049	0.434	0.060
		$c = -5$	0.025	0.000	0.201	0.000
		$c = -15$	0.022	0.000	0.123	0.000

Panel B: Power

			$\psi = 0.01$		$\psi = 0.05$		$\psi = 0.1$	
T	δ	c	t-test	$Q(\beta_0, 1)$	t-test	$Q(\beta_0, 1)$	t-test	$Q(\beta_0, 1)$
100	-0.5	0	0.118	0.104	0.429	0.443	0.786	0.766
		-5	0.044	0.022	0.169	0.130	0.509	0.428
		-15	0.032	0.004	0.089	0.021	0.250	0.084
	-1.5	0	0.099	0.104	0.267	0.443	0.617	0.766
		-5	0.043	0.002	0.075	0.015	0.194	0.132
		-15	0.033	0.000	0.060	0.000	0.113	0.001
	-3	0	0.091	0.104	0.166	0.443	0.333	0.766
		-5	0.038	0.000	0.060	0.000	0.096	0.005
		-15	0.028	0.000	0.040	0.000	0.064	0.000
250	-0.5	0	0.206	0.232	0.905	0.885	0.997	0.987
		-5	0.080	0.505	0.689	0.605	0.986	0.950
		-15	0.059	0.010	0.366	0.152	0.882	0.686
	-1.5	0	0.146	0.232	0.792	0.885	0.994	0.987
		-5	0.044	0.005	0.304	0.258	0.923	0.894
		-15	0.035	0.000	0.153	0.002	0.523	0.093
	-3	0	0.101	0.232	0.463	0.885	0.917	0.987
		-5	0.041	0.000	0.111	0.015	0.412	0.612
		-15	0.030	0.000	0.069	0.000	0.174	0.000
500	-0.5	0	0.484	0.501	0.998	0.993	1.00	1.00
		-5	0.179	0.138	0.990	0.967	1.00	1.00
		-15	0.093	0.018	0.899	0.701	1.00	1.00
	-1.5	0	0.279	0.501	0.995	0.993	1.00	1.00
		-5	0.077	0.015	0.938	0.907	1.00	1.00
		-15	0.047	0.000	0.493	0.095	0.995	0.940
	-3	0	0.151	0.501	0.934	0.993	1.00	1.00
		-5	0.048	0.001	0.386	0.614	0.992	0.998
		-15	0.029	0.000	0.143	0.000	0.651	0.178

Table 4: Finite-Sample Size and Power of the Wald and $MQ(\beta_0, I)$ Tests: The Case of Two Regressors

Notes: 1. Data were generated by $y_t = \beta' x_{t-1} + u_{yt}$; $\beta = [0, 0]'$; $x_t = \begin{bmatrix} \exp(\frac{c}{T}) & 0 \\ 0 & \exp(\frac{c}{T}) \end{bmatrix} x_{t-1} + u_{vt}$; $u_{yt} = [\delta, \delta] u_{vt} + e_t$; $u_{vt} \sim iid\ N\left(0, \begin{bmatrix} 1 & 0.85 \\ 0.85 & 1 \end{bmatrix}\right)$; $e_t \sim iid\ N(0, 1)$. 2. The minimum-volatility rule was used for subsampling. 3. Uncentered subsampling is used. 4. Results are based on 2,000 replications.

Panel A: Size

			Subsampling		Chi-square	
			Wald	$MQ(\beta_0, I)$	Wald	$MQ(\beta_0, I)$
$T = 100$	$\delta = -0.5$	$c = 0$	0.087	0.074	0.164	0.050
		$c = -5$	0.056	0.184	0.010	0.189
		$c = -15$	0.053	0.404	0.071	0.530
	$\delta = -1.5$	$c = 0$	0.077	0.074	0.288	0.050
		$c = -5$	0.048	0.697	0.138	0.813
		$c = -15$	0.039	0.948	0.074	0.998
	$\delta = -3$	$c = 0$	0.080	0.074	0.319	0.050
		$c = -5$	0.043	0.947	0.140	0.994
		$c = -15$	0.042	0.987	0.078	1.00
$T = 250$	$\delta = -0.5$	$c = 0$	0.075	0.083	0.177	0.047
		$c = -5$	0.044	0.183	0.092	0.173
		$c = -15$	0.043	0.483	0.065	0.554
	$\delta = -1.5$	$c = 0$	0.064	0.083	0.291	0.047
		$c = -5$	0.038	0.755	0.118	0.812
		$c = -15$	0.032	0.987	0.072	1.00
	$\delta = -3$	$c = 0$	0.067	0.083	0.306	0.047
		$c = -5$	0.034	0.984	0.126	0.997
		$c = -15$	0.028	1.00	0.069	1.00
$T = 500$	$\delta = -0.5$	$c = 0$	0.078	0.076	0.179	0.043
		$c = -5$	0.042	0.212	0.095	0.183
		$c = -15$	0.032	0.490	0.067	0.558
	$\delta = -1.5$	$c = 0$	0.079	0.076	0.295	0.043
		$c = -5$	0.038	0.774	0.135	0.826
		$c = -15$	0.033	0.993	0.080	1.00
	$\delta = -3$	$c = 0$	0.073	0.076	0.317	0.043
		$c = -5$	0.036	0.989	0.134	0.996
		$c = -15$	0.033	1.00	0.080	1.00

Panel B: Power

			$\psi = 0.01$		$\psi = 0.05$		$\psi = 0.1$	
T	δ	c	Wald	$MQ(\beta_0, 1)$	Wald	$MQ(\beta_0, 1)$	Wald	$MQ(\beta_0, 1)$
		0	0.140	0.125	0.655	0.695	0.945	0.935
	-0.5	-5	0.073	0.111	0.273	0.206	0.772	0.716
		-15	0.053	0.316	0.141	0.097	0.417	0.142
		0	0.095	0.125	0.290	0.695	0.650	0.935
100	-1.5	-5	0.050	0.604	0.085	0.173	0.195	0.212
		-15	0.044	0.937	0.061	0.874	0.109	0.681
		0	0.094	0.125	0.154	0.695	0.298	0.935
	-3	-5	0.046	0.933	0.059	0.834	0.086	0.428
		-15	0.046	0.987	0.049	0.986	0.060	0.980
		0	0.321	0.398	0.996	0.992	1.00	1.00
	-0.5	-5	0.091	0.072	0.943	0.906	1.00	0.999
		-15	0.057	0.261	0.638	0.290	0.995	0.963
		0	0.141	0.398	0.868	0.992	1.00	1.00
250	-1.5	-5	0.045	0.501	0.303	0.504	0.953	0.994
		-15	0.034	0.976	0.109	0.636	0.514	0.146
		0	0.097	0.398	0.425	0.992	0.904	1.00
	-3	-5	0.042	0.956	0.087	0.180	0.313	0.913
		-15	0.030	1.00	0.054	0.999	0.128	0.943
		0	0.727	0.766	1.00	1.00	1.00	1.00
	-0.5	-5	0.255	0.188	1.00	1.00	1.00	1.00
		-15	0.106	0.114	0.998	0.977	1.00	1.00
		0	0.312	0.766	1.00	1.00	1.00	1.00
500	-1.5	-5	0.074	0.211	0.970	0.998	1.00	1.00
		-15	0.045	0.966	0.489	0.123	1.00	1.00
		0	0.156	0.766	0.924	1.00	1.00	1.00
	-3	-5	0.045	0.907	0.316	0.916	0.991	1.00
		-15	0.033	1.00	0.112	0.980	0.544	0.331

Table 5: Finite-Sample Size and Power of the t- and $Q(\beta_0, 1)$ Tests: The Case of Two Regressors

Notes: 1. Data were generated by $y_t = \beta' x_{t-1} + u_{yt}$; $\beta = [0, 0]'$; $x_t = \begin{bmatrix} \exp(\frac{c}{T}) & 0 \\ 0 & \exp(\frac{c}{T}) \end{bmatrix} x_{t-1} + u_{vt}$; $u_{yt} = [\delta, \delta] u_{vt} + e_t$; $u_{vt} \sim iid\ N\left(0, \begin{bmatrix} 1 & 0.85 \\ 0.85 & 1 \end{bmatrix}\right)$; $e_t \sim iid\ N(0, 1)$. 2. The minimum-volatility rule was used for subsampling. 3. Uncentered subsampling is used. 4. Results are based on 2,000 replications.

Panel A: Size

			t-test	$Q(\beta_0, 1)$
		$c = 0$	0.080	0.070
	$\delta = -0.5$	$c = -5$	0.053	0.041
		$c = -15$	0.040	0.019
		$c = 0$	0.078	0.070
$T = 100$	$\delta = -1.5$	$c = 5$	0.047	0.012
		$c = 15$	0.028	0.001
		$c = 0$	0.078	0.070
	$\delta = -3$	$c = -5$	0.042	0.004
		$c = -15$	0.025	0.000
		$c = 0$	0.070	0.063
	$\delta = -0.5$	$c = -5$	0.039	0.027
		$c = -15$	0.026	0.012
		$c = 0$	0.061	0.063
$T = 250$	$\delta = -1.5$	$c = -5$	0.034	0.007
		$c = -15$	0.022	0.001
		$c = 0$	0.058	0.063
	$\delta = -3$	$c = -5$	0.038	0.001
		$c = -15$	0.022	0.000
		$c = 0$	0.067	0.069
	$\delta = -0.5$	$c = -5$	0.040	0.033
		$c = -15$	0.025	0.012
		$c = 0$	0.062	0.069
$T = 500$	$\delta = -1.5$	$c = -5$	0.035	0.001
		$c = -15$	0.019	0.000
		$c = 0$	0.063	0.069
	$\delta = -3$	$c = -5$	0.028	0.001
		$c = -15$	0.021	0.000

Panel B: Power

			$\psi = 0.01$		$\psi = 0.05$		$\psi = 0.1$	
T	δ	c	t-test	$Q(\beta_0, 1)$	t-test	$Q(\beta_0, 1)$	t-test	$Q(\beta_0, 1)$
100	-0.5	0	0.097	0.090	0.165	0.217	0.284	0.424
		-5	0.062	0.047	0.091	0.102	0.159	0.208
		-15	0.043	0.021	0.063	0.040	0.095	0.080
	-1.5	0	0.083	0.090	0.108	0.217	0.145	0.424
		-5	0.050	0.016	0.062	0.042	0.076	0.104
		-15	0.030	0.002	0.041	0.005	0.050	0.010
	-3	0	0.080	0.090	0.091	0.217	0.113	0.424
		-5	0.044	0.004	0.050	0.007	0.056	0.024
		-15	0.026	0.000	0.028	0.000	0.033	0.000
250	-0.5	0	0.100	0.131	0.356	0.537	0.653	0.859
		-5	0.058	0.053	0.198	0.270	0.463	0.666
		-15	0.033	0.016	0.102	0.092	0.270	0.315
	-1.5	0	0.072	0.131	0.145	0.537	0.281	0.859
		-5	0.040	0.015	0.077	0.146	0.151	0.533
		-15	0.023	0.002	0.039	0.009	0.075	0.072
	-3	0	0.066	0.131	0.094	0.537	0.143	0.859
		-5	0.040	0.002	0.052	0.028	0.076	0.276
		-15	0.022	0.000	0.031	0.001	0.037	0.003
500	-0.5	0	0.139	0.207	0.676	0.870	0.952	0.995
		-5	0.070	0.089	0.449	0.672	0.868	0.973
		-15	0.043	0.029	0.235	0.322	0.673	0.835
	-1.5	0	0.086	0.207	0.275	0.870	0.567	0.995
		-5	0.047	0.033	0.132	0.525	0.336	0.957
		-15	0.030	0.001	0.066	0.061	0.163	0.565
	-3	0	0.076	0.207	0.147	0.870	0.284	0.995
		-5	0.038	0.003	0.067	0.277	0.133	0.901
		-15	0.024	0.000	0.039	0.001	0.070	0.088

Table 6: Finite-Sample Size and Power of the t-Test: The Case of One Regressor

Notes: 1. Data were generated by $y_t = \psi x_{t-1} + u_{yt}$; $x_t = \exp(\frac{c}{T})x_{t-1} + u_{vt}$; $u_{yt} = \delta u_{vt} + e_t$; $u_{vt} \sim iid\ N(0,1)$; $e_t \sim iid\ N(0,1)$. 2. The calibration rule was used for subsampling. 3. Centered subsampling was used. 4. Results are based on 2,000 replications. 5. One-sided test with $H_1 : \beta > 0$.

			Size		Power	
			$\psi = 0$	$\psi = 0.01$	$\psi = 0.05$	$\psi = 0.1$
$T = 100$	$\delta = -0.5$	$c = 0$	0.096	0.155	0.557	0.903
		$c = -5$	0.068	0.099	0.321	0.716
		$c = -15$	0.070	0.088	0.203	0.465
	$\delta = -1.5$	$c = 0$	0.110	0.148	0.400	0.788
		$c = -5$	0.071	0.085	0.164	0.365
		$c = -15$	0.062	0.072	0.118	0.222
	$\delta = -3$	$c = 0$	0.116	0.136	0.236	0.491
		$c = -5$	0.069	0.077	0.115	0.176
		$c = -15$	0.061	0.066	0.089	0.131
$T = 250$	$\delta = -0.5$	$c = 0$	0.088	0.288	0.970	1.00
		$c = -5$	0.059	0.132	0.822	0.999
		$c = -15$	0.055	0.103	0.571	0.978
	$\delta = -1.5$	$c = 0$	0.101	0.196	0.903	1.00
		$c = -5$	0.053	0.091	0.470	0.977
		$c = -15$	0.045	0.065	0.238	0.715
	$\delta = -3$	$c = 0$	0.099	0.146	0.619	0.972
		$c = -5$	0.053	0.069	0.191	0.604
		$c = -15$	0.040	0.053	0.110	0.268
$T = 500$	$\delta = -0.5$	$c = 0$	0.085	0.551	1.00	1.00
		$c = -5$	0.054	0.261	0.998	1.00
		$c = -15$	0.046	0.148	0.973	1.00
	$\delta = -1.5$	$c = 0$	0.088	0.365	1.00	1.00
		$c = -5$	0.037	0.115	0.987	1.00
		$c = -15$	0.039	0.066	0.691	1.00
	$\delta = -3$	$c = 0$	0.093	0.211	0.974	1.00
		$c = -5$	0.041	0.074	0.565	1.00
		$c = -15$	0.036	0.053	0.217	0.821

Table 7: Finite-Sample Size and Power of the Wald Test: The Case of Two Regressors

Notes: 1. Data were generated by $y_t = \beta' x_{t-1} + u_{yt}$; $\beta = [\psi, \psi]'$; $x_t = \begin{bmatrix} \exp(\frac{c}{T}) & 0 \\ 0 & \exp(\frac{c}{T}) \end{bmatrix} x_{t-1} + u_{vt}$; $u_{yt} = [\delta, \delta] u_{vt} + e_t$; $u_{vt} \sim iidN\left(0, \begin{bmatrix} 1 & 0.85 \\ 0.85 & 1 \end{bmatrix}\right)$; $e_t \sim iid\ N(0,1)$. 2. The calibration rule was used for subsampling. 3. Centered subsampling was used. 4. Results are based on 2,000 replications.

			Size		Power	
			$\psi = 0$	$\psi = 0.01$	$\psi = 0.05$	$\psi = 0.1$
$T = 100$	$\delta = -0.5$	$c = 0$	0.128	0.210	0.798	0.987
		$c = -5$	0.091	0.119	0.459	0.928
		$c = -15$	0.083	0.104	0.265	0.673
	$\delta = -1.5$	$c = 0$	0.142	0.171	0.452	0.838
		$c = -5$	0.082	0.098	0.168	0.375
		$c = -15$	0.070	0.075	0.117	0.210
	$\delta = -3$	$c = 0$	0.146	0.161	0.262	0.474
		$c = -5$	0.082	0.089	0.121	0.179
		$c = -15$	0.065	0.068	0.087	0.118
$T = 250$	$\delta = -0.5$	$c = 0$	0.100	0.418	1.00	1.00
		$c = -5$	0.058	0.160	0.989	1.00
		$c = -15$	0.053	0.093	0.821	1.00
	$\delta = -1.5$	$c = 0$	0.099	0.203	0.942	1.00
		$c = -5$	0.055	0.082	0.460	0.993
		$c = -15$	0.038	0.050	0.191	0.705
	$\delta = -3$	$c = 0$	0.087	0.134	0.541	0.960
		$c = -5$	0.053	0.064	0.142	0.484
		$c = -15$	0.036	0.043	0.080	0.205
$T = 500$	$\delta = -0.5$	$c = 0$	0.094	0.800	1.00	1.00
		$c = -5$	0.052	0.368	1.00	1.00
		$c = -15$	0.045	0.159	1.00	1.00
	$\delta = -1.5$	$c = 0$	0.100	0.389	1.00	1.00
		$c = -5$	0.044	0.100	0.994	1.00
		$c = -15$	0.036	0.059	0.661	1.00
	$\delta = -3$	$c = 0$	0.093	0.196	0.964	1.00
		$c = -5$	0.043	0.065	0.446	1.00
		$c = -15$	0.033	0.043	0.156	0.705

Table 8: Finite-Sample Size and Power of t-Test: The Case of Two Regressors

Notes: 1. Data were generated by $y_t = \beta' x_{t-1} + u_{yt}$; $\beta = [\psi, \psi]'$; $x_t = \begin{bmatrix} \exp(\frac{c}{T}) & 0 \\ 0 & \exp(\frac{c}{T}) \end{bmatrix} x_{t-1} + u_{vt}$; $u_{yt} = [\delta, \delta] u_{vt} + e_t$; $u_{vt} \sim iidN\left(0, \begin{bmatrix} 1 & 0.85 \\ 0.85 & 1 \end{bmatrix}\right)$; $e_t \sim iid\ N(0,1)/\ 2$. The calibration rule was used for subsampling. 3. Centered subsampling was used. 4. Results are based on 2,000 replications.

			Size		Power	
			$\psi = 0$	$\psi = 0.01$	$\psi = 0.05$	$\psi = 0.1$
$T = 100$	$\delta = -0.5$	$c = 0$	0.118	0.135	0.228	0.382
		$c = -5$	0.093	0.108	0.169	0.270
		$c = -15$	0.085	0.093	0.131	0.203
	$\delta = -1.5$	$c = 0$	0.133	0.138	0.172	0.223
		$c = -5$	0.094	0.097	0.114	0.139
		$c = -15$	0.067	0.071	0.084	0.106
	$\delta = -3$	$c = 0$	0.126	0.129	0.145	0.172
		$c = -5$	0.085	0.088	0.097	0.110
		$c = -15$	0.067	0.070	0.075	0.079
$T = 250$	$\delta = -0.5$	$c = 0$	0.107	0.161	0.461	0.788
		$c = -5$	0.072	0.097	0.298	0.641
		$c = -15$	0.057	0.071	0.190	0.453
	$\delta = -1.5$	$c = 0$	0.110	0.129	0.246	0.401
		$c = -5$	0.067	0.079	0.130	0.228
		$c = -15$	0.045	0.054	0.087	0.144
	$\delta = -3$	$c = 0$	0.111	0.124	0.178	0.247
		$c = -5$	0.064	0.071	0.093	0.140
		$c = -15$	0.048	0.054	0.064	0.084
$T = 500$	$\delta = -0.5$	$c = 0$	0.101	0.198	0.775	0.983
		$c = -5$	0.064	0.127	0.614	0.956
		$c = -15$	0.047	0.081	0.403	0.861
	$\delta = -1.5$	$c = 0$	0.104	0.148	0.379	0.673
		$c = -5$	0.063	0.087	0.216	0.486
		$c = -15$	0.044	0.054	0.134	0.294
	$\delta = -3$	$c = 0$	0.101	0.124	0.232	0.389
		$c = -5$	0.060	0.071	0.126	0.227
		$c = -15$	0.037	0.044	0.074	0.131

Table 9: Finite-Sample Size and Power of the Bonferroni t-Test for One Regressor

Notes: 1. Data were generated by $y_t = u_{yt}$; $x_t = \exp(\frac{c}{T})x_{t-1} + u_{vt}$; $u_{yt} = \delta u_{vt} + e_t$; $u_{vt} \sim iid\ N(0,1)$; $e_t \sim iid\ N(0,1)$. 2. Results are based on 2,000 replications.

			Size		Power	
T	δ	c	$\psi = 0$	$\psi = 0.01$	$\psi = 0.05$	$\psi = 0.1$
100	-0.5	0	0.057	0.116	0.604	0.933
		-5	0.042	0.055	0.241	0.721
		-15	0.037	0.052	0.142	0.400
	-1.5	0	0.053	0.090	0.672	0.987
		-5	0.050	0.058	0.180	0.788
		-15	0.043	0.054	0.132	0.326
	-3	0	0.060	0.092	0.751	0.998
		-5	0.050	0.060	0.144	0.839
		-15	0.052	0.064	0.154	0.366
250	-0.5	0	0.057	0.246	0.984	1.000
		-5	0.040	0.105	0.883	1.000
		-15	0.037	0.073	0.563	0.992
	-1.5	0	0.051	0.227	1.000	1.000
		-5	0.035	0.073	0.939	1.000
		-15	0.052	0.092	0.465	0.999
	-3	0	0.051	0.227	1.000	1.000
		-5	0.043	0.067	0.988	1.000
		-15	0.062	0.097	0.476	1.000
500	-0.5	0	0.050	0.604	1.000	1.000
		-5	0.030	0.219	1.000	1.000
		-15	0.037	0.134	0.992	1.000
	-1.5	0	0.061	0.678	1.000	1.000
		-5	0.036	0.150	1.000	1.000
		-15	0.051	0.141	0.999	1.000
	-3	0	0.055	0.734	1.000	1.000
		-5	0.047	0.132	1.000	1.000
		-15	0.050	0.146	1.000	1.000

Table 10: Finite-Sample Size and Power of the Bonferroni Q-Test for One Regressor

Notes: 1. Data were generated by $y_t = \psi x_{t-1} + u_{yt}$; $x_t = \exp(\frac{c}{T})x_{t-1} + u_{vt}$; $u_{yt} = \delta u_{vt} + e_t$; $u_{vt} \sim iid\ N(0,1)$; $e_t \sim iid\ N(0,1)$. 2. Results are based on 2,000 replications.

			Size		Power	
T	δ	c	$\psi = 0$	$\psi = 0.01$	$\psi = 0.05$	$\psi = 0.1$
		0	0.067	0.136	0.647	0.946
	-0.5	-5	0.063	0.096	0.339	0.796
		-15	0.052	0.068	0.191	0.466
		0	0.059	0.102	0.802	0.996
100	-1.5	-5	0.071	0.098	0.370	0.930
		-15	0.058	0.070	0.158	0.468
		0	0.072	0.120	0.932	1.000
	-3	-5	0.065	0.084	0.353	0.986
		-15	0.060	0.073	0.177	0.512
		0	0.055	0.247	0.983	1.000
	-0.5	-5	0.057	0.154	0.925	1.000
		-15	0.039	0.089	0.634	0.993
		0	0.043	0.219	1.000	1.000
250	-1.5	-5	0.044	0.119	0.982	1.000
		-15	0.047	0.088	0.663	1.000
		0	0.042	0.198	1.000	1.000
	-3	-5	0.047	0.122	0.999	1.000
		-15	0.047	0.086	0.718	1.000
		0	0.046	0.614	1.000	1.000
	-0.5	-5	0.049	0.307	1.000	1.000
		-15	0.041	0.171	0.992	1.000
		0	0.045	0.758	1.000	1.000
500	-1.5	-5	0.040	0.310	1.000	1.000
		-15	0.041	0.144	1.000	1.000
		0	0.045	0.898	1.000	1.000
	-3	-5	0.046	0.280	1.000	1.000
		-15	0.040	0.144	1.000	1.000

Table 11. Subsampling-Based Tests of Predictability: Valuation Ratios and the Risk-free Rate as Predictors

We use centered subsampling where the block size is selected based on the calibration rule discussed in Section 3. In all regressions, the dependent variable is log excess stock return on the NYSE/AMEX value-weighted index. The independent variables are the log dividend–price ratio (dp), the log earnings–price ratio (ep), and the log three-month T-bill rate (rf), all lagged one period. Panel A reports results of univariate one-sided t-tests and Panel B the results of bivariate Wald tests. * denote test statistics that are significant at the 5% level.

Panel A. Univariate Tests

Series	Regressor	$\widehat{\beta}$	t-test
1926–2004			
Annual	dp	0.126	2.124
	ep	0.147	2.572
Quarterly	dp	0.028	1.770
	ep	0.043	2.732*
Monthly	dp	0.006	1.399
	ep	0.012	2.462
1926–1994			
Annual	dp	0.269	3.016*
	ep	0.269	3.435*
Quarterly	dp	0.054	2.308*
	ep	0.077	3.513*
Monthly	dp	0.012	1.789
	ep	0.021	3.187*
1952–2004			
Annual	dp	0.114	1.768
	ep	0.098	1.534
	rf	-1.189	-1.457
Quarterly	dp	0.029	1.816
	ep	0.025	1.572
	rf	-0.407	-2.063
Monthly	dp	0.009	1.818
	ep	0.007	1.517
	rf	-0.159	-2.661*

Panel B. Bivariate Tests

Series	$\widehat{\beta}$ of regressors			Wald test
	dp	ep	rf	
1926–2004				
Annual	-0.034	0.176		6.607
Quarterly	-0.052*	0.090*		9.683*
Monthly	-0.019	0.029*		9.347
1926–1994				
Annual	0.082	0.211		12.00
Quarterly	-0.035	0.104*		13.11*
Monthly	-0.016	0.034*		12.01*
1952–2004				
Annual	0.163	-0.051		3.155
	0.168		1.928*	8.815
		0.164	2.052*	8.333*
Quarterly	0.056	-0.028		3.556
	0.046		-0.604	11.90
		0.046	-0.639*	11.57*
Monthly	0.021	0.012		3.869
	0.015		-0.221*	15.70*
		0.015	-0.233*	15.29*

Table 12. Subsampling-Based Tests of Predictability: Consumption-Based Variables as Predictors

We use centered subsampling where the block size is selected based on the calibration rule discussed in Section 3. In all regressions, the dependent variable is log excess stock return on the NYSE/AMEX value-weighted index. The independent variables are the log dividend–price ratio (*dp*), the consumption–wealth ratio (*cay*) proposed by Lettau and Ludvigson (2001a, 2001b), the labor-income–consumption ratio (s^w) proposed by Santos and Veronesi (2006), log surplus consumption based on linear habit, $\ln(C_t - X_{lin,t})$, and log surplus consumption based on nonlinear habit, $\ln(C_t - X_{nonlin,t})$, all lagged one period. All tests in this table are based on bivariate regressions. Columns 1–5 report estimates of the individual coefficients, with * denoting a coefficient that is significant at the 5% level in a one-sided t-test. The last column reports the Wald statistic of a joint test where both coefficients are different from zero, with * denoting a test statistic that is significant at the 5% level.

$\widehat{\beta}$ of regressors					Wald test
dp	*cay*	s^w	$\ln(C - X_{lin})$	$\ln(C - X_{nonlin})$	
Panel A. Annual Data (1952–2004)					
0.045	5.470*				15.99*
0.132		-1.654*			8.329
	5.411*	-0.887*			17.360*
0.059			-0.093		5.051
0.058				-0.128*	8.641*
			0.171*	-0.293*	9.332*
	5.230*			-0.116*	23.190*
		-0.598		-0.125	8.363*
Panel B. Quarterly Data (1952–2004)					
0.012	1.435*				15.050*
0.033		-0.427			8.579
	1.400*	-0.213			15.850*
0.019			-0.017		4.309
0.015				-0.029	7.377
			0.060	-0.085*	8.964*
	1.375*			-0.025*	18.420*
		-0.195		-0.027	7.402