HIER

Harvard Institute of Economic Research

Discussion Paper Number 2084

Predicting the Equity Premium Out of Sample:
Can Anything Beat the Historical Average?

by

John Y. Campbell
and
Samuel B. Thompson

July 2005

Harvard University
Cambridge, Massachusetts

Predicting the Equity Premium Out of Sample:

Can Anything Beat the Historical Average?

John Y. Campbell and Samuel B. Thompson[1]

First draft: October 8, 2004
This version: June 20, 2005

Abstract

A number of variables are correlated with subsequent returns on the aggregate US stock market in the 20th Century. Some of these variables are stock market valuation ratios, others reflect patterns in corporate finance or the levels of short- and long-term interest rates. Amit Goyal and Ivo Welch (2004) have argued that in-sample correlations conceal a systematic failure of these variables out of sample: None are able to beat a simple forecast based on the historical average stock return. In this note we show that forecasting variables with significant forecasting power in-sample generally have a better out-of-sample performance than a forecast based on the historical average return, once sensible restrictions are imposed on the signs of coefficients and return forecasts. The out-of-sample predictive power is small, but we find that it is economically meaningful. We also show that a variable is quite likely to have poor out-of-sample performance for an extended period of time even when the variable genuinely predicts returns with a stable coefficient.

1 Introduction

Towards the end of the last century, academic finance economists came to take seriously the view that aggregate stock returns are predictable. During the 1980's a number of papers studied valuation ratios, such as the dividend-price ratio, earnings-price ratio, or smoothed earnings-price ratio. Value-oriented investors in the tradition of Graham and Dodd (1934) had always asserted that high valuation ratios are an indication of an undervalued stock market and should predict high subsequent returns, but these ideas did not carry much weight in the academic literature until authors such as Rozeff (1984), Fama and French (1988), and Campbell and Shiller (1988a,b) found that valuation ratios are positively correlated with subsequent returns and that the implied predictability of returns is substantial at longer horizons. Around the same time, several papers pointed out that yields on short- and long-term Treasury and corporate bonds are correlated with subsequent stock returns (Fama and Schwert 1977, Keim and Stambaugh 1986, Campbell 1987, Fama and French 1989).

During the 1990's and early 2000's, research continued on the prediction of stock returns from valuation ratios (Kothari and Shanken 1997, Pontiff and Schall 1998) and interest rates (Hodrick 1992). Several papers suggested new predictor variables exploiting information in corporate payout and financing activity (Lamont 1998, Baker and Wurgler 2000), the level of consumption in relation to wealth (Lettau and Ludvigson 2001), and the relative valuations of high- and low-beta stocks (Polk, Thompson, and Vuolteenaho 2003). At the same time, several authors expressed concern that the apparent predictability of stock returns might be spurious. Many of the predictor variables in the literature are highly persistent, and Stambaugh (1999) pointed out that persistence leads to biased coefficients in predictive regressions if innovations in the predictor variable are correlated with returns (as is strongly the case for valuation ratios, although not for interest rates). Under the same conditions the standard t-test for predictability has incorrect size (Cavanagh, Elliott, and Stock 1995). These problems are exacerbated if researchers are data mining, considering large numbers of variables and reporting only those results that are apparently statistically significant (Ferson, Sarkissian, and Simin 2003). An active recent literature discusses alternative econometric methods for correcting the Stambaugh bias and conducting valid inference (Cavanagh, Elliott, and Stock 1995, Mark 1995, Kilian 1999, Ang and Bekaert 2003, Jansson and Moreira 2003, Lewellen 2004, Torous, Valkanov, and Yan 2004, Campbell and Yogo 2005, Polk, Thompson, and Vuolteenaho 2005).

A somewhat different critique emphasizes that predictive regressions have often performed poorly out-of-sample (Goyal and Welch 2003, 2004, Butler, Grullon, and Weston 2004). This critique had particular force during the bull market of the late 1990's, when low valuation ratios predicted extraordinarily low stock returns that did not materialize until the early 2000's (Campbell and Shiller 1998). Goyal and Welch (2004) argue that the poor out-of-sample performance of predictive regressions is a systemic problem, not confined to any one decade. They compare predictive regressions with historical average returns and find that historical average returns almost always generate superior return forecasts. They write: "Our paper has systematically investigated the empirical real-world out-of-sample performance of plain linear regressions to predict the equity premium. We find that none of the popular variables has worked—and not only post-1990... Our profession has yet to find a variable that has had meaningful robust empirical equity premium forecasting power, at least from the perspective of a real-world investor."

In this note we evaluate the out-of-sample performance of a wide variety of forecasting variables and argue that the case for stock return predictability is much stronger than Goyal and Welch admit. We review the empirical evidence in Section 2. We first discuss the mundane but important issues of return measurement and sample selection, arguing that it is important to evaluate predictive power using high-quality data on total returns: Price returns, or estimated total returns based on interpolated dividends, are not an acceptable alternative. For this reason we use the period since 1927, when CRSP monthly total returns are available, as our out-of-sample forecast evaluation period.

Next we compare in-sample and out-of-sample forecast performance. We use the in-sample t statistic as a measure of the apparent in-sample predictability from a given variable. We show that many of the variables with particularly poor out-of-sample performance have low t statistics, so in-sample and out-of-sample forecast evaluation methods often deliver similar results. We also calculate an out-of-sample R^2 statistic that can be compared with the usual in-sample R^2 statistic. Like Goyal and Welch, we find poor out-of-sample performance for many of the usual linear regressions.

Goyal and Welch recommend that one should adopt "the perspective of a real-world investor". Our next contribution is to argue that a real-world investor would not mechanically forecast using a linear regression, but would impose some restrictions on the regression coefficients. We consider two alternative restrictions: first, that the regression coefficient has the theoretically expected sign; and second, that the fitted

2

value of the equity premium is positive. We impose these restrictions sequentially and then together, and find that they substantially improve the out-of-sample evidence for predictability.

Section 2 shows that several commonly used forecasting variables do have some ability to predict stock returns out-of-sample. The out-of-sample R^2 statistics are positive, but very small. This raises the question of whether the predictive power is economically meaningful. In Section 3 we show that even very small R^2 statistics are relevant for investors because they can generate large improvements in portfolio performance. In a related exercise, we calculate the fees that investors would be willing to pay to exploit the information in each of our forecasting variables.

In Section 4 we discuss the interpretation of out-of-sample forecasting results. Goyal and Welch write as if poor out-of-sample performance is strong evidence against the view that stock returns are predictable. We show to the contrary, that a predictive model may have a stable coefficient equal to its in-sample OLS estimate, and with high probability the model will not beat the historical average return out of sample. A similar point has been made by Inoue and Kilian (2004), who argue that in-sample tests of predictability are often more powerful than out-of-sample tests. Section 5 briefly concludes.

2 Empirical results

In this section we conduct an out-of-sample forecasting exercise inspired by Goyal and Welch (2004), with modifications that reveal the sensitivity of their conclusions. We use a monthly time horizon and predict simple monthly stock returns. This immediately creates a tradeoff between the length of the data sample and the quality of the available data. High-quality total return data are available monthly from CRSP since 1927, while total monthly returns before that time are constructed by interpolation of lower-frequency dividend payments and therefore may be suspect. Accordingly we use the CRSP data period as our out-of-sample forecast evaluation period, but use earlier data to estimate an initial regression.

Table 1, whose format is based on the tables in Goyal and Welch, reports the results. We begin by discussing panel A, and then discuss modifications to the basic method reported in panels B, C, and D. Each row of the table considers

3

a different forecasting variable. The first four rows consider valuation ratios: the dividend price ratio, earnings price ratio, smoothed earnings price ratio, and book to market ratio. Each of these ratios has some accounting measure of corporate value in the numerator, and market value in the denominator. The smoothed earnings price ratio, proposed by Campbell and Shiller (1988b, 1998) is the ratio of a 10-year moving average of earnings to current prices. Campbell and Shiller argue that this ratio should have better forecasting power than the current earnings price ratio because aggregate corporate earnings display short-run cyclical noise; in particular earnings drop close to zero in recession years such as 1934 and 1992, creating spikes in the current earnings price ratio that have nothing to do with stock market valuation levels.[2]

The next seven rows consider nominal interest rates and inflation: the short-term interest rate, long-term bond yield, lagged long-term bond return, the term spread between long- and short-term Treasury yields, the default spread between corporate and Treasury bond yields, the lagged excess return on corporate over Treasury bonds, and the lagged rate of inflation.

The last four rows of the table evaluate forecasting variables that have been proposed more recently: the cross-sectional beta premium of Polk, Thompson, and Vuolteenaho (2003), the dividend payout ratio proposed by Lamont (1998), the equity share of new issues proposed by Baker and Wurgler (2000), and the consumption-wealth ratio of Lettau and Ludvigson (2001). This last variable is based on a cointegrating relationship between consumption, aggregate labor income, and aggregate financial wealth. Rather than estimate a separate cointegrating regression, we simply include the three variables directly in the forecasting equation for stock returns.

The first column of Table 1 reports the first date at which we have data to run the forecasting regression. For dividends, earnings, and stock returns we have data, originally assembled by Robert Shiller, back to 1871. Other data series typically begin shortly after the end of World War I. All data series continue to the end of 2003 as reported in the second column of the table. The third column reports the date at which we begin the out-of-sample forecast evaluation. This is the beginning of 1927, when accurate data on total monthly stock returns become available from

---

[2]Goyal and Welch consider these variables, and also the ratio of lagged dividends to lagged prices (the "dividend yield" in Goyal and Welch's terminology). We drop this variable as there is no reason to believe that it should predict better than the ratio of lagged dividends to current prices (the "dividend price ratio").

CRSP, or 20 years after the date in column 1, whichever comes later.

The fourth and fifth columns of Table 1 report the full-sample t statistic for the significance of each variable in forecasting stock returns, and the adjusted R^2 statistic of the full-sample regression.[3] It is immediately obvious from the column of t statistics that many of the valuation ratios and interest-rate variables are statistically insignificant in predicting stock returns over this long sample period. The most successful variables are the two variants of the earnings price ratio, the Treasury bill rate, and the inflation rate. It may not be surprising that interest-rate variables are weak predictors over the sample periods used here, as interest-rate behavior changed radically in the early 1950's when the modern era of monetary policy began. The three recently proposed variables are much more successful return predictors in-sample, with t statistics of at least 2.4.

The remaining columns of the table evaluate the out-of-sample performance of these forecasts. The sixth column, labelled "Delta RMSE", reports the difference in the root mean squared error between the predictive regression and a forecast equal to the historical average return measured at each date (equivalent to a regression of stock returns onto a constant). When this difference is negative, the historical average return beats the predictive regression out of sample.

The seventh and eighth columns report the mean out-of-sample residual for the predictive regression and the historical average return forecast. In the first few rows of the table, which have initial data from the late 19th and early 20th centuries and an out-of-sample period starting in 1927, these residuals are typically positive. This reflects the strong performance of the US stock market in the later 20th Century. In rows corresponding to slowly moving valuation ratios such as the dividend-price and earnings-price ratios, and also in the cross-sectional premium row, the residuals are more positive for the predictive regression than for the historical average return, reflecting the tendency of these variables to generate pessimistic return forecasts towards the end of the 20th Century.

The last column reports an out-of-sample R^2 statistic that can be compared with

---

[3]The adjustment of the R^2 statistic for degrees of freedom makes only a very small difference in samples of the size used here. For a regression from 1871 through 2003, the adjustment is -0.06%, and it is -0.11% for a regression from 1927 through 2003.

the in-sample R^2 statistic. This is computed as

$$R_{OS}^2 = 1 - \frac{\sum_{t=1}^{T}(r_t - \widehat{r}_t)^2}{\sum_{t=1}^{T}(r_t - \overline{r}_t)^2}, \tag{1}$$

where $\widehat{r}_t$ is the fitted value from a predictive regression estimated through period $t - 1$, and $\overline{r}_t$ is the historical average return estimated through period $t - 1$. The out-of-sample R^2 has the same sign as the change in the root mean squared error reported in column 6, but it is measured in the same units as the in-sample R^2 in column 5.

The out-of-sample performance of the predictor variables is quite mixed. Panel A of Table 1 shows that only two out of four valuation ratios, three out of seven interest-rate variables, and two out of four recently proposed variables deliver positive out-of-sample R^2 statistics.

It is premature, however, to conclude with Goyal and Welch that predictive regressions cannot profitably be used by investors in real time. A regression estimated over a short sample period can easily generate perverse results, such as a negative coefficient when theory suggests that the coefficient should be positive. Since out-of-sample forecast evaluation begins as little as 20 years after the start of the data set, this can be an important problem in practice. For example, in the early 1930s the earnings-price ratio was very high, but the coefficient on the predictor was estimated to be negative. This led to a negative forecast of the equity premium in the early 1930s and subsequent poor forecast performance. In practice, an investor would not use a perverse coefficient but would likely conclude that the coefficient is zero, in effect imposing prior knowledge on the output of the regression.

In panels B, C, and D we explore the impact of imposing sensible restrictions on the out-of-sample forecasting exercise. In panel B we set the regression coefficient to zero whenever it has the "wrong" sign (different from the theoretically expected sign estimated over the full sample). In panel C we assume that investors rule out a negative equity premium, and set the forecast to zero whenever it is negative. We follow the same procedure for the historical mean forecast, setting it to zero whenever it is negative. In panel D we impose first the sign restriction on the coefficient, and then the sign restriction on the forecast.

These restrictions improve the out-of-sample performance of predictive regressions. In panel A, as we noted above, only 2 out of 4 valuation ratios have a positive out-of-sample R^2. In panels B and D this improves to 3 out of 4. Similarly, in panel A only

6

3 out of 7 interest-rate variables have a positive out-of-sample R^2 statistic, but this improves to 4 out of 7 in panels B and C, and 5 out of 7 in panel D. The restriction that the equity premium be positive helps the performance of the cross-sectional equity premium and the dividend payout ratio, so that all four recently proposed variables have positive out-of-sample R^2 in panels C and D. Importantly, once we impose these restrictions the regressions that perform well out-of-sample now tend to be the ones that also work well in-sample. In fact the out-of-sample R^2 in panel D sometimes exceeds the in-sample R^2. The main exception is the book-to-market ratio, which generates a negative out-of-sample R^2 statistic in all four panels of the table.

Figure 1 illustrates the effect of the restrictions for the smoothed earnings-price ratio. The coefficient restriction significantly improves the forecasts in the 1930s, when the coefficient was estimated to be negative. The forecast restrictions are binding during the 1960s and 1990s, and improve the forecast performance during the 1990s. Valuation ratios were unusually low during these periods, leading to unprecedentedly low forecasts. Campbell and Shiller (2001) also noted the unusually low earnings-price ratios of the 1990s, and wrote "We do not find this extreme forecast credible; when the independent variable has moved so far from the historically observed range, we cannot trust a linear regression line." Our forecast restriction offers a simple way to correct for this incredible forecast.

Looking at the performance of individual variables, it is striking how much better the earnings-based valuation ratios perform than the dividend-price ratio. This may well be due to changes in payout policy as firms have shifted from paying dividends to repurchasing shares. Boudoukh, Michaely, Richardson, and Roberts (2004) emphasize that in recent years the total payout to price ratio, including share repurchases, has much stronger predictive power than the dividend-price ratio. Also, short- and long-term Treasury yields perform reasonably well both in-sample and out-of-sample, consistent with the conclusion of Ang and Bekaert (2003) that these variables are robust return predictors. The performance of these variables would be stronger if we started the sample period later, because the interest-rate process changed dramatically at the time of the Federal Reserve-Treasury Accord in 1951. The recently proposed variables tend to perform well, particularly Lettau and Ludvigson's combination of consumption, income, and wealth; however it should be remembered that these variables may have been selected with the aid of a specification search based on almost all the data used here, and this may give them an artificial advantage over predictor variables that were proposed in the late 1980's or before.

7

All the regressions we have reported predict simple stock returns rather than log stock returns. The use of simple returns makes little difference to the comparison of predictive regressions with historical mean forecasts, but all forecasts tend to generate higher mean residuals when log returns are used. The reason for this is that high stock market volatility in the 1920's and 1930's depressed log returns relative to simple returns in this period. Thus the gap between average stock returns in the late 20th Century and the early 20th Century is greater when log returns are used.

3 How large an R^2 should we expect?

In the previous section we showed that many of the forecasting variables that have been discussed in the literature do have positive out-of-sample predictive power for aggregate US stock returns, when reasonable restrictions are imposed on the predictive regression. However the R^2 statistics are very small in magnitude. This raises the important question of whether they are economically meaningful.

To explore this issue, consider the following example:

$$r_{t+1} = \mu + x_t + \varepsilon_{t+1}, \tag{2}$$

where r_{t+1} is the excess simple return on a risky asset over the riskless interest rate, μ is the unconditional average excess return, x_t is a predictor variable with mean zero, and ε_{t+1} is a random shock with mean zero. For tractability, consider an investor with a single-period horizon and mean-variance preferences. The investor's objective function is expected portfolio return less $(\gamma/2)$ times portfolio variance, where γ can be interpreted as the coefficient of relative risk aversion.[4] If the investor does not observe x_t, she chooses a portfolio weight in the risky asset

$$\alpha_t = \alpha = \left(\frac{1}{\gamma}\right)\left(\frac{\mu}{\sigma_x^2 + \sigma_\varepsilon^2}\right) \tag{3}$$

and earns an average excess return of

$$\left(\frac{1}{\gamma}\right)\left(\frac{\mu^2}{\sigma_x^2 + \sigma_\varepsilon^2}\right) = \frac{S^2}{\gamma}, \tag{4}$$

---

[4]Merton (1969) presents the analogous portfolio solution for the case where the investor has power utility with relative risk aversion γ, asset returns are lognormally distributed, and the portfolio can be continuously rebalanced. Campbell and Viceira (2002, Chapter 2) use a discrete-time approximate version of Merton's solution.

where S is the unconditional Sharpe ratio of the risky asset.

If the investor observes x_t, she sets

$$\alpha_t = \left(\frac{1}{\gamma}\right)\left(\frac{\mu + x_t}{\sigma_\varepsilon^2}\right), \tag{5}$$

where the denominator is now σ_ε^2 rather than $\sigma_x^2 + \sigma_\varepsilon^2$ because the variation in the predictor variable x_t is now expected and does not contribute to risk. The investor earns an average excess return of

$$\left(\frac{1}{\gamma}\right)\left(\frac{\mu^2 + \sigma_x^2}{\sigma_\varepsilon^2}\right) = \left(\frac{1}{\gamma}\right)\left(\frac{S^2 + R^2}{1 - R^2}\right), \tag{6}$$

where

$$R^2 = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_\varepsilon^2} \tag{7}$$

is the R^2 statistic for the regression of excess return on the predictor variable x_t.

The difference between the two expected returns is

$$\left(\frac{1}{\gamma}\right)\left(\frac{R^2}{1 - R^2}\right)(1 + S^2) \tag{8}$$

which is always larger than R^2/γ, and is close to R^2/γ when the time interval is short and R^2 and S^2 are both small. The *proportional* increase in the expected return from observing x_t is

$$\left(\frac{R^2}{1 - R^2}\right)\left(\frac{1 + S^2}{S^2}\right) \tag{9}$$

which is always larger than R^2/S^2 and is close to R^2/S^2 when the time interval is short and R^2 and S^2 are both small.

This analysis shows that the right way to judge the magnitude of R^2 is to compare it with the squared Sharpe ratio S^2. If R^2 is large relative to S^2, then an investor can use the information in the predictive regression to obtain a large proportional increase in portfolio return. In our monthly data since 1871, the monthly Sharpe ratio for stocks is 0.108, corresponding to an annual Sharpe ratio of 0.374. The squared monthly Sharpe ratio $S^2 = 0.012 = 1.2\%$. This can be compared with the out-of-sample R^2 statistic for, say, the earnings-price ratio of 0.25% in Panel D of Table 1. A mean-variance investor can use the earnings-price ratio to increase her

9

average monthly portfolio return by a proportional factor of $0.25/1.2 = 21\%$. The absolute increase in portfolio return depends on risk aversion, but is about 25 basis points per month or 3% per year for an investor with unit risk aversion, and about 1% per year for an investor with a risk aversion coefficient of three. Predictor variables in Table 1 with higher out-of-sample R^2 statistics imply correspondingly larger increases in return.

The investor who observes x_t gets a higher portfolio return in part by taking on greater risk. Thus the increase in the average return is not pure welfare gain for a risk-averse investor. To take account of this, in Table 2 we calculate the welfare benefits generated by optimally trading on each predictor variable for an investor with relative risk aversion of three. We impose realistic portfolio constraints, preventing the investor from shorting stocks or taking more than 50% leverage, that is, confining the portfolio weight on stocks to lie between 0 and 150%. The investor's optimal portfolio depends on her estimate of stock return variance at each point in time, and we consider two alternative assumptions about how the investor forms this estimate. In the central panel of the table, the investor estimates variance using all data available up to the time her investment is made, while in the right hand panel, the investor estimates variance using a rolling five-year window of monthly data. The latter approach may be more appropriate if the investor believes that there are short-term fluctuations in variance. We report the utility level from investing with the historical mean forecast of the equity premium, and the changes in utility caused by investing with the unrestricted linear regression of Table 1, Panel A, or the doubly restricted linear regression of Table 1, Panel D. These utility differences have the units of expected return, so they can also be interpreted as the transactions costs or portfolio management fees that investors would be prepared to pay each month to exploit the information in the predictor variable.

In Table 2 the imposition of forecast restrictions makes less difference than it did in Table 1. The reason is that we rule out short sales, so investors are unable to act on negative forecasts of the equity premium even if their predictive regressions generate such forecasts. The use of time-varying variance forecasts makes little difference for valuation ratios or the recently proposed predictor variables, but it generally increases the utility gain from predicting stock returns with interest rates. Three out of seven interest rates generate positive utility gains when a constant variance assumption is used, while six out of seven do so when a time-varying variance estimate is used.

The predictor variables that generate the highest out-of-sample R^2 statistics gen-

erally deliver positive utility gains to investors. The main exception to this is the smoothed earnings-price ratio, which has a positive out-of-sample R^2 statistic but a negative utility gain. This result is driven by volatile monthly returns in the early 1930's, together with the diminishing marginal utility of risk-averse investors that downweights portfolio profits relative to losses. The smoothed earnings-price ratio increases utility if the evaluation period excludes the early 1930's, or if a longer investment horizon is used as we discuss below.

The utility gains reported in Table 2 are limited by the leverage constraint, together with the high average equity premium. Predictable variations in stock returns do not generate portfolio gains when there is a binding upper limit on equity investment. Utility gains would be larger if we relaxed the portfolio constraint or included additional assets, with higher average returns than Treasury bills, in the portfolio choice problem. On the other hand, Table 2 does not take any account of transactions costs. Modest gains from market timing strategies could be offset by the additional costs implied by those strategies. Optimal trading strategies in the presence of transactions costs are complex, and so we do not explore this issue further here. We note that even the baseline strategy based on a historical-mean forecast incurs rebalancing costs and that utility gains of 10 basis points per month, or 1.2% per year, as reported in several rows of Table 2, are sufficient to cover substantial additional costs.

Since small R^2 statistics can generate large benefits for investors, we should expect predictive regressions to have only modest explanatory power. Regressions with large R^2 statistics would be too profitable to believe. The saying "If you're so smart, why aren't you rich?" applies with great force here, and should lead investors to suspect that highly successful predictive regressions are spurious. Note, however, that the squared Sharpe ratio and average real interest rate increase in proportion with the investment horizon; thus much larger R^2 statistics are believable at longer horizons. Authors such as Fama and French (1988) have found that R^2 statistics increase strongly with the horizon when the predictor variable is persistent, a finding that is analyzed in Campbell, Lo, and MacKinlay (1997, Chapter 7) and Campbell (2001). This behavior is completely consistent with our analysis here.

Table 3 illustrates the effect of increasing the investment horizon on the performance of our four valuation ratios. These predictor variables are highly persistent, so we would expect their explanatory power to increase as we change the horizon from one month in the top panel, to one quarter in the middle panel, to one year

in the bottom panel. Indeed the out-of-sample R^2 statistics increase dramatically for the two earnings-based ratios and become positive for the book-to-market ratio. However the forecasting performance of the dividend-price ratio does not improve at longer horizons. The right hand part of Table 3 reports utility gains for these four ratios implemented at different horizons. The earnings-based ratios deliver solid improvements when used to invest over one year, as does the dividend-price ratio when the investor allows for changing volatility of stock returns, but the book-to-market ratio does not improve utility despite its positive out-of-sample R^2 at a one year horizon.

4 Reconciling in-sample with out-of-sample results

How do we reconcile apparent differences between in-sample and out-of-sample performance of the predictors? We emphasize that the differences are not typically large: In Table 1 small in-sample t statistics generally correspond to poor out-of-sample performance. However there are predictors, such as the book-to-market ratio, that display statistically significant t statistics along with inferior out-of-sample performance.

There are many reasons why out-of-sample evidence can contradict in-sample results. These include the effects of data mining, structural change, parameter uncertainty, and bad luck. Data mining occurs when a researcher evaluates several different predictors, but only reports those that are significant. This specification search can lead to spurious evidence of predictability. In principle out-of-sample analysis can counter data mining, since a spurious predictor should not work out-of-sample. Of course, the RMSE comparisons reported here and in Goyal and Welch (2004) are not true out-of-sample statistics. Just like the in-sample t statistics, they are functions of historical data. To see this point more clearly, suppose that finance academics decided to evaluate predictive relationships using RMSE comparisons instead of t statistics. A researcher could evaluate several different predictors, then choose to report only those that show large improvements in out-of-sample RMSE. So it seems unlikely that data mining explains the different results.

Certain kinds of structural change can lead to spurious findings of in-sample predictability (Clark and McCracken 2003, Giacomini and White 2003, Paye and Timmermann 2003). Suppose the predictive relationship is strong at the beginning of the sample but weakens over time. In this case the in-sample t statistic may detect

the strong relationship on average. However a market participant might not have been able to use the forecasting relationship, since at the beginning of the sample it would be hard to estimate, and toward the end of the sample it would have disappeared. While this represents a weakness of in-sample statistics, Inoue and Kilian (2004) have shown that out-of-sample statistics have the same problems. To continue the example, if the predictive relationship declines slowly the out-of-sample statistics will also detect past predictability. Inoue and Kilian (2004) argue that the relative performance of in-sample and out-of-sample statistics has a great deal to do with the form of structural change. Without a specific model of structural change, it is difficult to conclude which statistics are more useful to forecasters.

Next we turn to the effects of parameter uncertainty. To fix ideas, consider the predictive regression

$$r_{t+1} = \mu + \theta x_t + \varepsilon_{t+1}, \tag{10}$$

with $\mathrm{E}\,\varepsilon_t = 0$. r_t is the excess return on the S&P 500 and x_t is a predictor variable like the dividend-price ratio. Should a forecaster use the historical mean or the fitted value $\widehat{r}_{t+1} = \widehat{\mu} + \widehat{\theta} x_t$? If $\theta \neq 0$ the historical mean of r_{t+1} will be a biased predictor, while $\widehat{r}_{t+1}$ will be unbiased so long as the regression estimates are unbiased. On the other hand, estimation error for θ will increase the variance of $\widehat{r}_{t+1}$ over the historical mean. Thus there is a choice between the unbiased but noisy predictor $\widehat{r}_{t+1}$ and the possibly biased and less-noisy historical mean. If $\theta = 0$, we expect the historical mean to forecast more accurately since both predictors are unbiased while estimation error for θ will add noise to the forecast. As θ increases the bias of the historical mean becomes more important. However, if θ and the sample size are both small, then the noise from estimating θ may dominate the gains from eliminating the bias. In light of the bias-variance tradeoff, consider Inoue and Kilian's (2004) result that in-sample predictability tests are generally more powerful than out-of-sample tests. In some situations a powerful in-sample test could correctly detect predictability, but that predictability could be useless to a market participant who cannot estimate the predictive coefficient accurately enough to improve her forecast.

The bias-variance tradeoff becomes more complicated when the regression estimates are biased. This will occur when the predictor variable is persistent and shocks to the predictor are correlated with shocks to the market return. Following Stambaugh (1999), we model the predictor with

$$x_{t+1} = \nu + \rho x_t + u_{t+1} \tag{11}$$

13

with $\mathrm{E}\, u_{t+1} = 0$. Classical asymptotic theory states that in a large sample the ordinary least squares estimator for θ is unbiased and has the smallest sampling variability among unbiased estimators. It is well known, however, that when ρ is close to 1 and $\mathrm{Corr}\,(e_{t+1}, u_{t+1})$ is nonzero, classical asymptotics offers a poor approximation to the true sampling distribution of the OLS estimator in small samples. For example, Stambaugh (1999) shows that when x is the dividend yield the OLS estimator and the in-sample t statistic are biased upward, leading to findings of spurious predictability.

These concerns about parameter uncertainty led us to impose the forecast restrictions in Table 1. We required θ to have the theoretically correct sign, and we also required the forecasted equity premium to be positive. This solution is not optimal, but there does not appear to exist an optimal forecaster when the predictor is persistent. An unbiased estimator for θ does not exist, although Stambaugh (1999) and Amihud and Hurvich (2004) have developed bias corrections. Litterman (1986) encountered these issues when forecasting macroeconomic series. His solution was to impose Bayesian prior information on θ. Our forecast restrictions can loosely be thought of as a uniform prior in a restricted part of the parameter space.

Finally, we turn to the problem of bad luck. Consider the unrealistic scenario where θ is known and nonzero. In this case a market participant will use the forecasting variable x_t since it improves her forecast in expectation. However, in a finite sample it may still be the case that the historical mean beats the forecasting variable with known θ. This does not mean that the market participant should not use the forecasting variable. Rather she made the right decision but was unlucky.

We carried out a Monte Carlo experiment to assess the effects of parameter uncertainty and bad luck. For each predictor variable, we estimated equations (10) and (11) by ordinary least squares, and fit a constant-correlation GARCH model (Bollerslev 1990) to the residuals.[5] We then simulated 5000 data sets from the estimated mod-

---

[5]The error variances follow

$$\sigma_{e,t}^2 = \omega_1 + \alpha_1 e_{t-1}^2 + \beta_1 \sigma_{e,t-1}^2; \qquad \sigma_{u,t}^2 = \omega_2 + \alpha_2 u_{t-1}^2 + \beta_2 \sigma_{u,t-1}^2$$

with $\sigma_{e,t}^2 \equiv \mathrm{Var}_{t-1}(e_t)$ and $\sigma_{u,t}^2 \equiv \mathrm{Var}_{t-1}(u_t)$. The conditional correlations $\mathrm{Corr}_{t-1}(e_t, u_t)$ are constant. Let $(\widehat{e}_t, \widehat{u}_t)$ denote residuals from regressions (10) and (11). We estimate the CC GARCH model by maximum likelihood, assuming the errors (e_t, u_t) are bivariate normal. Bollerslev and Wooldridge (1992) have shown that these parameter estimates will be consistent even if the true errors are not distributed bivariate normal. Our Monte Carlo simulations do not impose normality; rather they match the empirical distribution of the residuals $(\widehat{e}_t, \widehat{u}_t)$. Our algorithm for simulating the jth pair (e_j, u_j) follows. We normalize the residuals by their implied variances to obtain the

els, and for each simulated data set compared the out-of-sample forecasting results from using the historical mean to the results from using the predictor variable. Since none of the estimated values of θ are zero, all of the data generating processes imply some degree of predictability. The central panel of Table 4 reports the percentage of simulated data sets where the predictor variable leads to an inferior mean-squared error forecast than the historical mean.

Table 4 shows that parameter uncertainty can have a large effect on out-of-sample performance. For example, the in-sample R^2 statistic for the book-to-market ratio is 0.677%, yet at those in-sample parameter values the historical mean beats the unrestricted regression forecast in 35.5% of the simulations. The table also shows that forecast restrictions lead to significant improvements on average. For prediction with the inflation rate, the unrestricted regression forecast is inferior to the mean in 54.4% of the draws, but once the restrictions are imposed that drops to 36.3%.

The table demonstrates that poor out-of-sample performance can also come from bad luck. The last column of the central panel of Table 4, labelled "known θ", reports the infeasible results from using the true value of θ. Even when θ is known, the historical mean beats the forecasting variable in a surprisingly large number of cases. For the book-to-market ratio, the historical mean is superior in almost 10% of the cases. Thus a strong "true" forecasting relationship can be associated with poor out-of-sample performance.

The right hand panel of Table 4 conducts a similar exercise for the utility gains reported in Table 2. The first column in this panel reports the fraction of cases in which a portfolio, optimally constructed given a restricted linear regression, delivers lower average utility than a portfolio constructed using the historical equity-premium forecast. Both portfolios are based on a 5-year rolling variance estimate for stock returns. The second column reports the fraction of cases in which the investor underperforms even though she knows the true coefficient of the linear regression. The results make it clear that prolonged underperformance is quite likely even when stock returns can be stably predicted using the forecast variables discussed in the recent finance literature.

In unreported results we also calculated the percent of simulations where the out-of-sample R^2 was smaller than the actual value in Table 1. Recall that the out-

---

empirical distribution of draws $(\widehat{e}_t/\sigma_{e,t}, \widehat{u}_t/\sigma_{u,t})$. We randomly draw a pair from this unit-variance distribution, then multiply the elements by the variances implied by the simulated Garch model.

of-sample R^2 compares the forecast based on the predictive regression to the forecast based on the historical mean, so a negative value implies that the historical mean is superior. If the simulation percentages are very low, then the simulated data tend to generate out-of-sample statistics which are larger than the ones we see in the data. This would suggest that the out-of-sample R^2 values we see in the data are lower than what we would expect based on the full sample parameter estimates. The simulation percentages can be interpreted as p-values for the null hypothesis that the in-sample and out-of-sample results are consistent with one another.

We computed 60 p-values, based on 15 predictors and four possible forecast restrictions. The median p-value was 46.4%. P-values for the book-to-market ratio and default yield predictors range from 0.9% to 5.5%. Therefore these two predictors generate unusually small out-of-sample R^2 statistics given the full sample results. For the rest of the predictors the smallest p-value is 3.7%, for the long-term return with the forecast restricted to be positive. All of the rest of the p-values are greater than 5%. We conclude that for nearly all the predictors and various forecast restrictions, the out-of-sample R^2 statistics in Table 1 are consistent with the in-sample t statistics and R^2 values.

These results shed light on an interesting question. Suppose a forecasting variable has a significant t statistic, but has underperformed the historical mean out-of-sample. Should an investor use that variable in the future? On the one hand, poor out-of-sample performance could indicate the presence of structural breaks in the data, or deleterious effects of parameter uncertainty. These factors suggest that investment decisions should not be based on this predictor variable. On the other hand, poor out-of-sample results could be due to bad luck. Also, since more data are available now, the effects of parameter uncertainty are less serious now than in the past. These considerations suggest that the investor should use the forecasting variable, relying on its in-sample predictive power. Without further theory or information this question does not have a clear answer, but our results in Table 4 suggest that one should not exaggerate the significance of poor out-of-sample statistics.

5 Conclusion

A number of variables are correlated with subsequent returns on the aggregate US stock market in the 20th Century. Some of these variables are stock market valuation ratios, others reflect the levels of short- and long-term interest rates, patterns in corporate finance or the cross-sectional pricing of individual stocks, or the level of consumption in relation to wealth. Amit Goyal and Ivo Welch (2004) have argued that in-sample correlations conceal a systematic failure of these variables out of sample: None are able to beat a simple forecast based on the historical average stock return.

In this note we have shown that most of these predictor variables, and almost all that are statistically significant in-sample, perform better out-of-sample than the historical average return forecast, once sensible restrictions are imposed on the signs of coefficients and return forecasts. The out-of-sample explanatory power is small, but nonetheless is economically meaningful for investors. We have also shown that a variable is quite likely to have poor out-of-sample performance for an extended period of time even when the variable genuinely predicts returns with a stable coefficient.

References

Amihud, Yakov and Clifford Hurvich, 2004, Predictive regressions: a reduced-bias estimation method, forthcoming *Journal of Financial and Quantitative Analysis.*

Ang, Andrew and Geert Bekaert, 2003, Stock return predictability: Is it there?, unpublished paper, Columbia University.

Baker, Malcolm and Jeffrey Wurgler, 2000, The equity share in new issues and aggregate stock returns, *Journal of Finance* 55, 2219–2257.

Bollerslev, Tim, 1990, Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model, *Review of Economics and Statistics* 72, 498–505.

Bollerslev, Tim and Jeff Wooldridge 1992, Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances, *Econometric Reviews* 11, 143–172.

Boudoukh, Jacob, Roni Michaely, Matthew Richardson, and Michael Roberts, 2004, On the importance of measuring payout yield: Implications for empirical asset pricing, NBER Working Paper 10651.

Butler, Alexander W., Gustavo Grullon, and James P. Weston, 2004, Can managers forecast aggregate market returns?, unpublished paper, University of South Florida and Rice University.

Campbell, John Y., 1987, Stock returns and the term structure, *Journal of Financial Economics* 18, 373–399.

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

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

Campbell, John Y. and Robert J. Shiller, 1988a, The dividend-price ratio and expectations of future dividends and discount factors, *Review of Financial Studies* 1, 195–228.

Campbell, John Y. and Robert J. Shiller, 1988b, Stock prices, earnings, and expected dividends, *Journal of Finance* 43, 661–676.

Campbell, John Y. and Robert J. Shiller, 1998, Valuation ratios and the long-run stock market outlook, *Journal of Portfolio Management.*

Campbell, John Y. and Robert J. Shiller, 2001, Valuation ratios and the long-run stock market outlook: an update, NBER Working Paper 8221.

Campbell, John Y. and Luis M. Viceira, 2002, *Strategic Asset Allocation: Portfolio Choice for Long-Term Investors*, Oxford University Press, New York, NY.

Campbell, John Y. and Motohiro Yogo, 2003, Efficient tests of stock return predictability, unpublished paper, Harvard University.

Cavanagh, Christopher L., Graham Elliott, and James H. Stock, 1995, Inference in models with nearly integrated regressors, *Econometric Theory* 11, 1131 –1147.

Clark, Todd E. and Michael W. McCracken, 2003, The power of tests of predictive ability in the presence of structural breaks, unpublished paper, Federal Reserve Bank of Kansas City and University of Missouri-Columbia.

Fama, Eugene F. and Kenneth R. French, 1988, Dividend yields and expected stock returns, *Journal of Financial Economics* 22, 3–25.

Fama, Eugene F. and Kenneth R. French, 1989, Business conditions and expected returns on stocks and bonds, *Journal of Financial Economics* 25, 23–49.

Fama, Eugene F. and G. William Schwert, 1977, Asset returns and inflation, *Journal of Financial Economics* 5, 115–146.

Ferson, Wayne E., Sergei Sarkissian, and Timothy T. Simin, 2003, Spurious regressions in financial economics?, *Journal of Finance* 58, 1393–1413.

Giacomini, Raffaella and Halbert White, Tests of conditional predictive ability, unpublished paper, University of California at Los Angeles and University of California at San Diego.

Goyal, Amit and Ivo Welch, 2003, Predicting the equity premium with dividend ratios, *Management Science* 49, 639–654.

Goyal, Amit and Ivo Welch, 2004, A comprehensive look at the empirical performance of equity premium prediction, NBER Working Paper 10483.

Graham, Benjamin and David L. Dodd, 1934, *Security Analysis*, first edition, McGraw Hill, New York, NY.

Hodrick, Robert J., 1992, Dividend yields and expected stock returns: Alternative procedures for inference and measurement, *Review of Financial Studies* 5, 257–286.

Inoue, Atsushi and Lutz Kilian, 2004, In-sample or out-of-sample tests of predictability: which one should we use?, forthcoming *Econometric Reviews*.

Jansson, Michael, and Marcelo J. Moreira, 2003, Optimal inference in regression models with nearly integrated regressors, unpublished paper, Harvard University.

Keim, Donald B. and Robert F. Stambaugh, 1986, Predicting returns in the stock and bond markets, *Journal of Financial Economics* 17, 357–390.

Kilian, Lutz, 1999, Exchange rates and monetary fundamentals: What do we learn from long-horizon regressions?, Journal of Applied Econometrics 14, 491–510.

Kothari, S.P. and Jay Shanken, 1997, Book-to-market, dividend yield, and expected market returns: A time-series analysis, *Journal of Financial Economics* 44, 169–203.

Lamont, Owen, 1998, Earnings and expected returns, *Journal of Finance* 53, 1563–1587.

Lettau, Martin and Sydney Ludvigson, 2001, Consumption, aggregate wealth, and expected stock returns, *Journal of Finance* 56, 815–849.

Lewellen, Jonathan, 2004, Predicting returns with financial ratios, *Journal of Financial Economics* 74, 209–235.

Litterman, Robert, 1986, Forecasting with Bayesian vector autoregressions: Five years of experience, *Journal of Business & Economic Statistics* 4, 25–38.

Mark, Nelson C., 1995, Exchange rates and fundamentals: Evidence on long-horizon predictability, *American Economic Review* 85, 201–218.

Merton, Robert C., 1969, Lifetime portfolio selection under uncertainty: the continuous time case, *Review of Economics and Statistics* 51, 247–257.

Paye, Bradley and Alan Timmermann, 2003, Instability of return prediction models, unpublished paper, University of California at San Diego.

Polk, Christopher, Samuel Thompson, and Tuomo Vuolteenaho, 2005, Cross-sectional forecasts of the equity premium, forthcoming, *Journal of Financial Economics.*

Pontiff, Jeffrey and Lawrence D. Schall, 1998, Book-to-market ratios as predictors of market returns, *Journal of Financial Economics* 49, 141–160.

Rozeff, Michael S., 1984, Dividend yields are equity risk premiums, *Journal of Portfolio Management* 11(1), 68–75.

Stambaugh, Robert F., 1999, Predictive regressions, *Journal of Financial Economics* 54, 375–421.

Torous, Walter, Rossen Valkanov, and Shu Yan, 2004, On predicting stock returns with nearly integrated explanatory variables, *Journal of Business* 77, 937–966.

Table 1 - Total returns

This table presents statistics on forecast errors for stock returns at a monthly frequency. We use S&P500 total returns (dividend included) where data prior to January 1927 was obtained from Robert Shiller's website. Panel A compares forecast error of historical mean vs. the error of the chosen model. Panel B introduces a restriction that the coefficient on model variable must be of the right sign, otherwise historical mean is used as a predictor instead. Panel C requires that the prediction be positive, otherwise it uses zero as its forecast for that period. Finally, Panel D uses both restrictions. We report differences in RMSE between the model and historical mean, mean residual over all prediction periods, and the out-of-sample R squared. Our in-sample statistics are the (heteroskedasticity-robust) t-statistic and adjusted R squared, both coming from the last regression (i.e. over the whole sample period prior to the last forecast in December 2003). For the consumption-wealth ratio, we report the robust F-statistic that all the slope coefficients are zero.

Panel A

	Sample Begin	End	Forecast Begin	In-Sample t statistic	R squared adj.	Out-of-Sample Delta RMSE	Mean	Hist. mean	R squared
Dividend Price ratio	1871m2	2003m12	1927m1	0.98	0.054%	-0.009%	0.344%	0.194%	-0.331%
Earnings Price ratio	1871m2	2003m12	1927m1	2.29	0.271%	0.007%	0.264%	0.194%	0.258%
Smooth Earnings Price ratio	1881m1	2003m12	1927m1	1.78	0.518%	0.003%	0.461%	0.189%	0.098%
Book-to-Market	1921m3	2003m12	1941m3	1.39	0.677%	-0.028%	-0.038%	-0.109%	-1.328%
T-Bill rate	1920m2	2003m12	1940m2	-1.82	0.273%	0.007%	0.255%	-0.084%	0.328%
Long Term yield	1919m1	2003m12	1939m1	-1.38	0.073%	-0.013%	0.280%	-0.088%	-0.573%
Long Term return	1926m1	2003m12	1946m1	1.28	0.071%	-0.015%	-0.180%	-0.127%	-0.693%
Term spread	1920m2	2003m12	1940m2	1.67	0.186%	0.008%	-0.156%	-0.084%	0.394%
Default Yield spread	1919m1	2003m12	1939m1	0.67	0.194%	-0.013%	0.109%	-0.088%	-0.595%
Default Return spread	1926m1	2003m12	1946m1	0.41	-0.056%	-0.006%	-0.124%	-0.127%	-0.281%
Inflation	1919m2	2003m12	1939m2	-1.64	0.305%	0.015%	-0.016%	-0.079%	0.701%
Cross-sectional premium	1937m5	2003m1	1957m5	2.64	1.013%	-0.013%	0.348%	-0.273%	-0.584%
Dividend Payout ratio	1871m12	2003m12	1927m1	-0.65	0.006%	-0.008%	0.096%	0.196%	-0.287%
Equity share of new issues	1927m12	2003m12	1947m12	2.43	0.833%	0.008%	-0.142%	-0.058%	0.392%
Consumption-wealth ratio	1952m1	2003m3	1972m1	9.58*	3.479%	0.027%	0.078%	-0.130%	1.152%

Panel B

	Sample Begin	End	Forecast Begin	In-Sample t statistic	R squared adj.	Out-of-Sample Delta RMSE	Mean	Hist. mean	R squared
Dividend Price ratio	1871m2	2003m12	1927m1	0.98	0.054%	0.000%	0.341%	0.194%	0.001%
Earnings Price ratio	1871m2	2003m12	1927m1	2.29	0.271%	0.008%	0.264%	0.194%	0.264%
Smooth Earnings Price ratio	1881m1	2003m12	1927m1	1.78	0.518%	0.007%	0.460%	0.189%	0.236%
Book-to-Market	1921m3	2003m12	1941m3	1.39	0.677%	-0.028%	-0.038%	-0.109%	-1.328%
T-Bill rate	1920m2	2003m12	1940m2	-1.82	0.273%	0.007%	0.255%	-0.084%	0.328%
Long Term yield	1919m1	2003m12	1939m1	-1.38	0.073%	-0.013%	0.280%	-0.088%	-0.573%
Long Term return	1926m1	2003m12	1946m1	1.28	0.071%	0.003%	-0.164%	-0.127%	0.149%
Term spread	1920m2	2003m12	1940m2	1.67	0.186%	0.008%	-0.156%	-0.084%	0.394%
Default Yield spread	1919m1	2003m12	1939m1	0.67	0.194%	-0.013%	0.109%	-0.088%	-0.595%
Default Return spread	1926m1	2003m12	1946m1	0.41	-0.056%	-0.005%	-0.124%	-0.127%	-0.226%
Inflation	1919m2	2003m12	1939m2	-1.64	0.305%	0.015%	-0.016%	-0.079%	0.701%
Cross-sectional premium	1937m5	2003m1	1957m5	2.64	1.013%	-0.013%	0.348%	-0.273%	-0.584%
Dividend Payout ratio	1871m12	2003m12	1927m1	-0.65	0.006%	-0.008%	0.096%	0.196%	-0.287%
Equity share of new issues	1927m12	2003m12	1947m12	2.43	0.833%	0.008%	-0.142%	-0.058%	0.392%
Consumption-wealth ratio	1952m1	2003m3	1972m1	9.58*	3.479%	0.027%	0.078%	-0.130%	1.152%

Panel C

	Sample Begin	End	Forecast Begin	In-Sample t statistic	R squared adj.	Out-of-Sample Delta RMSE	Mean	Hist. mean	R squared
Dividend Price ratio	1871m2	2003m12	1927m1	0.98	0.054%	-0.001%	0.339%	0.194%	-0.043%
Earnings Price ratio	1871m2	2003m12	1927m1	2.29	0.271%	0.007%	0.251%	0.194%	0.243%
Smooth Earnings Price ratio	1881m1	2003m12	1927m1	1.78	0.518%	0.014%	0.370%	0.189%	0.501%
Book-to-Market	1921m3	2003m12	1941m3	1.39	0.677%	-0.021%	-0.072%	-0.109%	-0.988%
T-Bill rate	1920m2	2003m12	1940m2	-1.82	0.273%	0.011%	0.165%	-0.084%	0.521%
Long Term yield	1919m1	2003m12	1939m1	-1.38	0.073%	0.005%	0.155%	-0.088%	0.211%
Long Term return	1926m1	2003m12	1946m1	1.28	0.071%	-0.013%	-0.186%	-0.127%	-0.607%
Term spread	1920m2	2003m12	1940m2	1.67	0.186%	0.009%	-0.163%	-0.084%	0.422%
Default Yield spread	1919m1	2003m12	1939m1	0.67	0.194%	-0.012%	0.107%	-0.088%	-0.547%
Default Return spread	1926m1	2003m12	1946m1	0.41	-0.056%	-0.006%	-0.124%	-0.127%	-0.281%
Inflation	1919m2	2003m12	1939m2	-1.64	0.305%	0.016%	-0.016%	-0.079%	0.709%
Cross-sectional premium	1937m5	2003m1	1957m5	2.64	1.013%	0.008%	0.227%	-0.273%	0.381%
Dividend Payout ratio	1871m12	2003m12	1927m1	-0.65	0.006%	0.000%	0.079%	0.196%	0.008%
Equity share of new issues	1927m12	2003m12	1947m12	2.43	0.833%	0.012%	-0.165%	-0.058%	0.555%
Consumption-wealth ratio	1952m1	2003m3	1972m1	9.58*	3.479%	0.036%	-0.107%	-0.130%	1.549%

Panel D

	Sample Begin	End	Forecast Begin	In-Sample t statistic	R squared adj.	Out-of-Sample Delta RMSE	Mean	Hist. mean	R squared
Dividend Price ratio	1871m2	2003m12	1927m1	0.98	0.054%	0.000%	0.341%	0.194%	0.001%
Earnings Price ratio	1871m2	2003m12	1927m1	2.29	0.271%	0.007%	0.252%	0.194%	0.249%
Smooth Earnings Price ratio	1881m1	2003m12	1927m1	1.78	0.518%	0.015%	0.369%	0.189%	0.538%
Book-to-Market	1921m3	2003m12	1941m3	1.39	0.677%	-0.021%	-0.072%	-0.109%	-0.988%
T-Bill rate	1920m2	2003m12	1940m2	-1.82	0.273%	0.011%	0.165%	-0.084%	0.521%
Long Term yield	1919m1	2003m12	1939m1	-1.38	0.073%	0.005%	0.155%	-0.088%	0.211%
Long Term return	1926m1	2003m12	1946m1	1.28	0.071%	0.005%	-0.170%	-0.127%	0.236%
Term spread	1920m2	2003m12	1940m2	1.67	0.186%	0.009%	-0.163%	-0.084%	0.422%
Default Yield spread	1919m1	2003m12	1939m1	0.67	0.194%	-0.012%	0.107%	-0.088%	-0.547%
Default Return spread	1926m1	2003m12	1946m1	0.41	-0.056%	-0.005%	-0.124%	-0.127%	-0.226%
Inflation	1919m2	2003m12	1939m2	-1.64	0.305%	0.016%	-0.016%	-0.079%	0.709%
Cross-sectional premium	1937m5	2003m1	1957m5	2.64	1.013%	0.008%	0.227%	-0.273%	0.381%
Dividend Payout ratio	1871m12	2003m12	1927m1	-0.65	0.006%	0.000%	0.079%	0.196%	0.008%
Equity share of new issues	1927m12	2003m12	1947m12	2.43	0.833%	0.012%	-0.165%	-0.058%	0.555%
Consumption-wealth ratio	1952m1	2003m3	1972m1	9.58*	3.479%	0.036%	-0.107%	-0.130%	1.549%

Table 2 - Portfolio choice

This table presents out-of-sample portfolio choice results at a monthly frequency. "Out-of-Sample R-squared" comes from Panel D of Table 1. "Assume constant variance" results calculate the variance from the entire available history of excess returns. "Allow changing variance" results calculate the variance from a rolling window of the last 5 years of data. "Utility level, mean fc" is the average value of the utility from forecasting the market with the historical mean. "Utility change" is the change in average utility from forecasting the market with the predictor instead of the historical mean. "Free parms" indicates that we used the unconstrained ols predictor of the equity premium (from Panel A of Table 1). "Constrained" indicates that we constrain the coefficient on the predictor to be the right sign, and we require that the prediction be non-negative (from Panel D of Table 1). The utility function is E(Rp) - (γ/2)Var(Rp), where Rp is the portfolio return and γ=3.

	Sample Begin	End	Forecast Begin	Out-of-Sample R squared	Assume constant variance			Allow changing variance		
					Utility level Mean fc	Utility change Free parms	Constrained	Utility level Mean fc	Utility change Free parms	Constrained
Dividend Price ratio	1871m2	2003m12	1927m1	0.001%	0.501%	-0.016%	-0.024%	0.660%	-0.037%	-0.024%
Earnings Price ratio	1871m2	2003m12	1927m1	0.249%	0.501%	0.048%	0.053%	0.660%	0.017%	0.020%
Smooth Earnings Price ratio	1881m1	2003m12	1927m1	0.538%	0.502%	-0.095%	-0.085%	0.660%	-0.053%	-0.050%
Book-to-Market	1921m3	2003m12	1941m3	-0.988%	0.667%	-0.063%	-0.063%	0.793%	-0.038%	-0.038%
T-Bill rate	1920m2	2003m12	1940m2	0.521%	0.639%	-0.008%	-0.008%	0.771%	0.117%	0.117%
Long Term yield	1919m1	2003m12	1939m1	0.211%	0.626%	-0.021%	-0.021%	0.753%	0.086%	0.086%
Long Term return	1926m1	2003m12	1946m1	0.236%	0.641%	0.007%	0.040%	0.767%	0.015%	0.060%
Term spread	1920m2	2003m12	1940m2	0.422%	0.639%	0.088%	0.088%	0.771%	0.131%	0.131%
Default Yield spread	1919m1	2003m12	1939m1	-0.547%	0.626%	-0.073%	-0.073%	0.753%	-0.075%	-0.075%
Default Return spread	1926m1	2003m12	1946m1	-0.226%	0.641%	-0.014%	-0.009%	0.767%	0.010%	0.013%
Inflation	1919m2	2003m12	1939m2	0.709%	0.630%	0.051%	0.051%	0.759%	0.108%	0.108%
Cross-sectional premium	1937m5	2003m1	1957m5	0.381%	0.611%	0.010%	0.010%	0.621%	0.057%	0.057%
Dividend Payout ratio	1871m12	2003m12	1927m1	0.008%	0.502%	0.094%	0.094%	0.659%	0.012%	0.012%
Equity share of new issues	1927m12	2003m12	1947m12	0.555%	0.649%	0.102%	0.102%	0.826%	0.128%	0.128%
Consumption-wealth ratio	1952m1	2003m3	1972m1	1.549%	0.533%	0.379%	0.379%	0.655%	0.300%	0.300%

Table 3 - Portfolio choice for the value-predictors only, various forecasting horizons

This table presents out-of-sample portfolio choice results at monthly, quarterly, and annual frequencies. "Out-of-Sample R-squared" imposes the restriction that the coefficient on the predictor has the right sign, also that the forecast is nonnegative. For further table notes, see Table 2.

	Sample Begin	End	Forecast Begin	Out-of-Sample R squared	Assume constant variance			Allow changing variance		
					Utility level Mean fc	Utility change Free parms	Utility change Constrained	Utility level Mean fc	Utility change Free parms	Utility change Constrained
Monthly results										
Dividend Price ratio	1871m2	2003m12	1927m1	0.001%	0.501%	-0.016%	-0.024%	0.660%	-0.037%	-0.024%
Earnings Price ratio	1871m2	2003m12	1927m1	0.249%	0.501%	0.048%	0.053%	0.660%	0.017%	0.020%
Smooth Earnings Price ratio	1881m1	2003m12	1927m1	0.538%	0.502%	-0.095%	-0.085%	0.660%	-0.053%	-0.050%
Book-to-Market	1921m3	2003m12	1941m3	-0.988%	0.667%	-0.063%	-0.063%	0.793%	-0.038%	-0.038%
Quarterly results										
Dividend Price ratio	1871q1	2003q4	1927q1	-0.004%	1.301%	-0.143%	-0.104%	1.786%	-0.145%	-0.092%
Earnings Price ratio	1871q1	2003q4	1927q1	0.763%	1.301%	0.140%	0.147%	1.786%	-0.007%	0.000%
Smooth Earnings Price ratio	1871q1	2003q4	1927q1	0.613%	1.271%	-0.418%	-0.401%	1.767%	-0.213%	-0.208%
Book-to-Market	1921q2	2003q4	1941q2	-1.630%	1.622%	-0.035%	-0.035%	2.221%	-0.161%	-0.161%
Yearly results										
Dividend Price ratio	1871	2003	1927	-0.159%	6.323%	-0.474%	-0.474%	5.947%	0.385%	0.385%
Earnings Price ratio	1871	2003	1927	1.731%	6.323%	0.089%	0.122%	5.947%	0.290%	0.298%
Smooth Earnings Price ratio	1881	2003	1927	6.643%	6.280%	0.170%	0.170%	5.883%	0.648%	0.648%
Book-to-Market	1922	2003	1942	0.640%	8.155%	-0.256%	-0.256%	7.802%	-0.462%	-0.462%

	In-sample R squared	Out-of-sample R squared underperformance					Portfolio underperformance	
		none	coef	premium	both	known θ	both	known θ
Dividend Price ratio	0.054%	46.1%	45.3%	43.0%	42.2%	14.7%	38.9%	16.4%
Earnings Price ratio	0.271%	23.9%	23.4%	19.7%	19.4%	9.4%	19.0%	10.4%
Smooth Earnings Price ratio	0.518%	19.7%	19.3%	13.6%	13.4%	4.9%	11.9%	6.1%
Book-to-Market	0.677%	35.5%	34.1%	28.5%	27.4%	9.5%	22.6%	10.9%
T-Bill rate	0.273%	69.8%	59.9%	61.9%	52.7%	27.8%	44.9%	30.7%
Long Term yield	0.073%	76.5%	65.3%	69.2%	58.3%	32.6%	49.3%	33.7%
Long Term return	0.071%	58.6%	51.2%	47.2%	40.2%	22.5%	38.7%	26.2%
Term spread	0.186%	60.1%	52.7%	48.1%	41.0%	23.8%	39.9%	27.4%
Default Yield spread	0.194%	66.8%	59.5%	57.7%	49.7%	28.0%	46.9%	30.9%
Default Return spread	-0.056%	76.2%	65.4%	68.2%	57.8%	34.4%	53.4%	36.2%
Inflation	0.305%	54.4%	47.5%	43.2%	36.3%	21.5%	35.6%	25.6%
Cross-sectional premium	1.013%	38.9%	36.5%	27.3%	24.7%	17.7%	27.6%	20.1%
Dividend Payout ratio	0.006%	76.6%	65.5%	72.9%	62.5%	38.4%	54.9%	40.2%
Equity share of new issues	0.833%	25.1%	22.3%	18.7%	16.1%	11.8%	17.6%	13.9%
Consumption-wealth ratio	3.479%	39.6%	38.5%	22.2%	21.5%	7.0%	53.6%	54.4%

Figure 1, Panel A: Forecasts based on smoothed earnings-price ratio
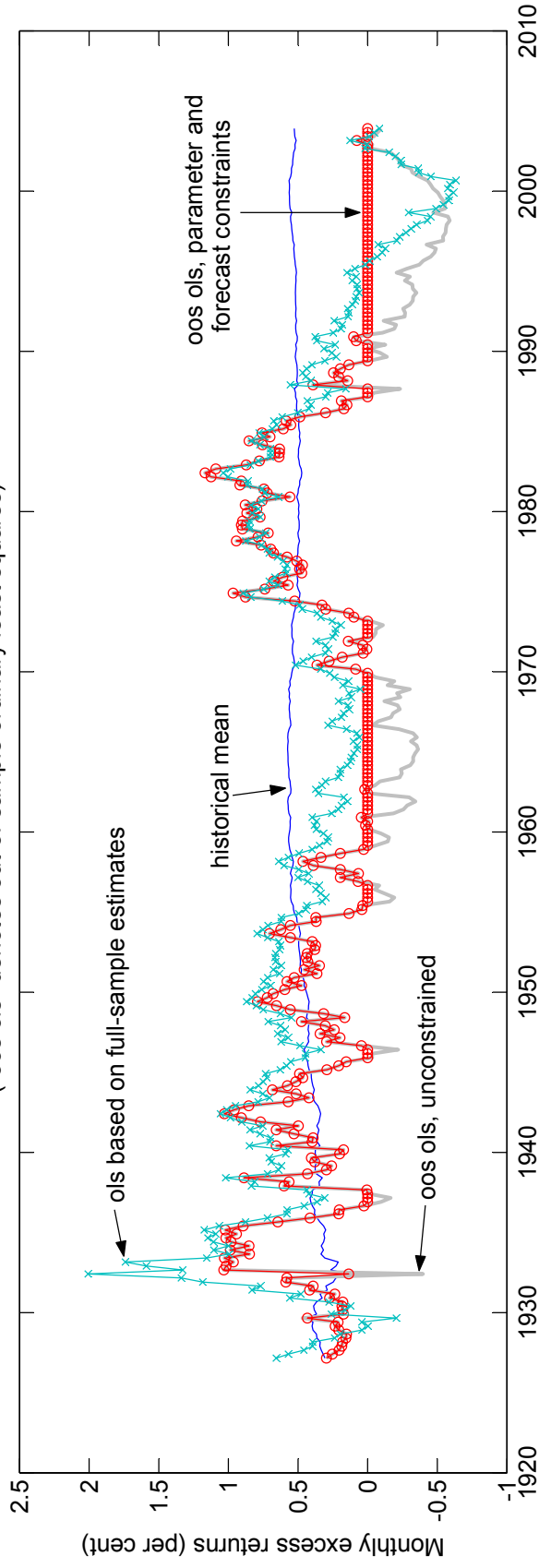("oos ols" denotes out-of-sample ordinary least squares)

Figure 1, Panel B: Cumulative R-squared based on smoothed earnings-price
(relative to historical mean)