Forecast-Agnostic Portfolios*

Hongye Guo† Jessica A. Wachter‡

August 26, 2026

Abstract

We introduce forecast-agnostic (FA) portfolios that exhibit out-of-sample market-timing ability without relying on a quantitative return forecast. These portfolios go long or short the market based on the level of a predictor variable, analogous to the factors widely used to study the cross section stock market returns. Despite using predictor variables that typically deliver negative out-of-sample R^2 values (Goyal et al., 2024), FA portfolios deliver significantly positive alphas on average. This result shows that existing stock market predictors have been historically successful overall, contrary to what past research has shown. We explain these seemingly contradictory phenomena by interpreting regression coefficients as portfolio returns: genuine predictability is necessary for high portfolio returns, whereas achieving a positive out-of-sample R^2 additionally requires the ability to forecast the returns on the forecast-agnostic portfolios themselves. As these FA portfolio returns could not be too predictable, estimating them substantially penalizes the out-of-sample R^2 by the inverse of the estimation sample size and lowers its power. Simulations show that the statistic we propose has substantially higher power to detect predictability that extends beyond in-sample diagnostics and the out-of-sample R^2.

Keywords: Market timing, Efficient Market Hypothesis, Return Predictability

JEL codes: G12, G14, G40

*The authors do not have conflicts to disclose. We thank Patrick Adams, John Campbell, Andrei Gonçalves, Amit Goyal (discussant), Erik Hjalmarsson (discussant), Raymond Kan, Petra Sinagl (disucussant), Sebastian Stöckl, Tuomo Vuolteenaho, Yao Zeng, Guofu Zhou (discussant) and seminar participants in Alpine Finance Summit, NBER LTAM, WFA, and WSIR for helpful comments.

†University of Hong Kong. Email: hoguo@hku.hk.

‡Department of Finance, The Wharton School, University of Pennsylvania, and the National Bureau of Economic Research. Email: jwachter@wharton.upenn.edu.

1 Introduction

The search for successful market-timing strategies is likely as old as organized equity markets themselves. As early as 1900, Bachelier posited that stock prices followed an arithmetic random walk, implying that price changes are unpredictable. The notion of unpredictable returns became standard market lore, with Fama's 1970 Efficient Markets Hypothesis and Malkiel's 2003 *Random Walk Down Wall Street*. There is good reason to believe that markets are unpredictable: should investors perceive that stocks have an unusual advantage over bonds, they will quickly shift their portfolios toward stocks, thereby causing stock prices to rise and eliminating such predictability. Nonetheless, Campbell and Shiller (1988) found an economically and statistically significant relation between scaled-price measures and future excess returns. A vast empirical and theoretical literature followed.

Subsequently, Goyal and Welch (2008) evaluated the many variables that the literature had identified as having market-timing ability. They showed that in-sample significance was fleeting, and out-of-sample performance was almost uniformly poor. Campbell and Thompson (2008) showed that performance improved when one imposes the economic requirement that the forecast of the excess return – which equals the conditional equity premium under rational expectations – must be nonnegative. The literature incorporated both the Goyal and Welch out-of-sample metric and the Campbell and Thompson adjustment, making it all the more surprising that newly proposed predictor variables continue to perform poorly (Goyal et al., 2024).

Poor performance in an out-of-sample forecast exercise need not indicate that a predictor variable is useless. We propose an investment-performance-based metric, which we refer to as the *forecast-agnostic* alpha. According to this metric, the majority of the 46 predictor variables studied in Goyal et al. (2024) would have been useful to investors. In fact, the forecast-agnostic alphas are significantly positive overall. Simulation results further show that, under the null of no predictability, alphas are on average zero. Under the alternative of predictability, alphas are on average positive—unlike the out-of-sample R^2, which remains negative. The forecast-agnostic alphas are hence more powerful than out-of-sample R^2 even if used on a standalone basis, responding to Campbell and Thompson (2005)'s critique on out-of-sample statistics.

The construction of the forecast-agnostic portfolios proceeds as follows in 3 steps: de-mean, sign, scale. First, in a given month, we take a predictor variable, de-meaned using its expanding-window mean, and form the portfolio with weight on the market equal to the de-meaned predictor variable. Second, we change the sign of the portfolio if a strategy trading on this de-meaned predictor earned negative average excess returns up to this month. Third, we scale the portfolio weight by the volatility of this portfolio's returns up to that month and multiply by a predetermined target volatility, and truncate the weight on the market at -3 and 3, as leverage beyond these levels is unachievable. The result is a portfolio that assigns a time-varying weight to the market based on a predictor's value, while simultaneously aiming for market neutrality and a constant target volatility.

In contrast to the out-of-sample R^2, this metric of portfolio returns imposes no penalty when the market surprises investors by working too much in their favor. The portfolio is forecast-agnostic, as its construction makes no use of a quantitative forecast of the excess return on the stock market. On this aspect, it is similar to the long-short portfolios used in the literature studying the cross section of stock returns, in particular, the Fama-MacBeth portfolios in Fama and MacBeth (1973). Like the out-of-sample R^2, our forecast-agnostic portfolios also do not use look-ahead information. Even the direction of the trade was informed by past data rather than a economic model-implied sign, which, as Campbell and Thompson (2008), may help trading performance in practice. We do not use those signs in this paper to be conservative about potential look-ahead in formation underneath them.

Although not the focus of their paper, Goyal et al. (2024) also examine the investment performance of predictor variables, specifically by looking at what they term "bull" and "bear" portfolios that remove the dependence on the forecasting coefficient beta.[1] They find

[1]Other papers that considered portfolio returns in the time-series setting include but are not limited to Campbell and Thompson (2008) and Kelly et al. (2024). The logic behind such statistics can further be linked back to earlier work that considered covariance between forecast errors (e.g., Clark and McCracken 2001, Diebold and Mariano 1995). We contribute to the literature by making the points that i) these portfolio

that the portfolios' performance only weakens the use case of the predictor variables. The reason is that Goyal et al. examine the average return and ask whether it is greater or less than what one would have achieved investing in equities. This is too stringent a test for a portfolio that takes on varying degrees of risk, given that the excess return on the market, is positive most of the time. Rather, since a sizable number of predictors are available, the test of whether a long-short portfolio based on a given predictor adds value is whether it has an alpha. If the investor is holding the market, then the relevant question is whether the portfolio has an alpha relative to the CAPM.[2] We drive this point home by adjusting our FA portfolios to have ex-ante zero beta and showing that combining them with the stock market substantially improves utilities of real-time mean-variance investors of various degrees risk aversion.

To better understand the properties of the forecast-agnostic alphas, as well as the out-of-sample R^2 statistics, we start with a simple data-generating process featuring uncorrelated shocks, homoskedasticity, and predictors following AR(1) processes. This simple setting allows us to derive simple, closed-form quantities that provide a useful rule of thumb: in expectation, the out-of-sample R^2 is the population R^2 minus the inverse of the estimation sample size t. Thus, to keep the out-of-sample R^2 positive, an average estimation sample length of 500 months requires a population R^2 of 20 basis points, which quickly appears too high relative to the 14 basis points estimated from the observed t-statistics. In contrast, the penalty in portfolio alphas is much smaller than $1/t$, making them positive in expectation quickly after the inception of data.

We then conduct Monte Carlo simulations in which we seek to match the level of predictability seen in the data along with more complex and realistic features such as heteroskedasticity. Given this level of predictability, a forecast-agnostic approach to out-of-

returns are substantially more powerful statistics when detecting weak predictors and ii) existing return predictors previous believed to have overall failed have in fact been historically successful.

[2]Ultimately, under both the Goyal et al. (2024) interpretation and ours, the investor holds a portfolio that consists of the market, together with the riskfree asset with time-varying weights in the market. However, under our interpretation, the investor tilts around a market portfolio of 100% (or another stable value implied by her risk aversion) depending on the predictor variable, whereas under their interpretation, the tilt is around a weight closer to zero.

sample portfolio performance is superior to a forecast-driven approach. Our simulations do not require any degree of parameter instability; they reflect sampling noise. The in-sample t-statistics indicate a noticeable, but not high, level of statistical significance. Forecast-driven portfolios and out-of-sample R^2s, which rely on expanding-window beta estimates, tend to load on noise especially early in the sample. They thus allocate too much risk budget in earlier periods, which is suboptimal relative to flat risk budget in the forecast agnostic portfolios. Simulations show, however, that if predictability were far stronger and we had ten times the amount of data, a forecast-driven approach would add value to a forecast-agnostic one.

A deeper question is why we do not see sufficient predictability to allow forecast-driven portfolios to outperform. We show that the beta coefficient in a predictive regression equals the excess return on a portfolio – not exactly our forecast-agnostic portfolio – but one that is highly similar, though uses the full sample. If the out-of-sample R^2 were close to the in-sample R^2, then the expanding-window beta coefficients would remain close to the in-sample beta coefficient, namely, the mean excess return on what is approximately the forecast-agnostic portfolio would be near a constant throughout the sample. If it were a constant, it would be an arbitrage opportunity, as it implies an infinite information ratio. Thus, the greater the precision in the beta estimate, the closer the out-of-sample results come to the in-sample results, and the closer the forecast-agnostic portfolio approaches an arbitrage opportunity. The absence of arbitrage is consistent with predictable returns on the market. However, these returns cannot be predicted in a sufficiently predictable way.

In line with the reasoning that arbitrage is at work to limit predictability, we show that the forecast-agnostic alpha associated with a strategy becomes statistically insignificant post-publication of that strategy, echoing Mclean and Pontiff (2016). In contrast, point estimates are similar pre-sample, suggesting that it is indeed arbitrage, and not instability in alpha estimates, that drives variation across samples.

The above reasoning presupposes that return predictability stems from a failure of rational expectations, and is not, say, a result of time-varying risk or risk aversion. Most likely, observed return predictability has more than one underlying cause. However, the degree to which it is a failure of rational expectations, as opposed to time-varying risk, is reflected in the results of several additional tests that we run. We show that while imposing the economic constraint that the risk premium cannot go below zero, as recommended by Campbell and Thompson (2008), does improve performance, nearly the same degree of improvement arises from imposing the symmetric positive constraint. Thus, the improvement appears to arise not so much from imposing an economic constraint, but from reducing the variance in the estimation of the forecasting coefficient. Similarly, for variables where there is a strong economic prior for what the beta should be, imposing this prior also improves out-of-sample performance. The superior performance could arise because the model in question is correct, but it also could be that the model gives a "reasonable" value of the predictive coefficient, and imposing this coefficient leads to reduced noise, even as it introduces bias, similarly to our forecast-agnostic alphas. Finally, post-publication fall-off in results also occurs for predictor variables with risk-based explanations, as does the lack of pre-sample fall-off. Taken together, these patterns cast further doubt on the arguments that these variables capture risk-based fluctuations in the equity premium and that they solely reflect statistical bias, but are consistent with the hypothesis that they got arbitraged away post-publication.

Besides the papers mentioned above, our paper also relates to early work by Kandel and Stambaugh (1996), who emphasize that it is returns, not R^2s, that ultimately matter for investors. Other papers that make use of this bias-variance trade-off include Rapach et al. (2010), who find market-timing capabilities for a portfolio that equally weights component portfolios. Like our forecast-agnostic alphas, equal-weighting throws away what seems to be information but is in fact mostly noise. Our paper also relates to the literature studying the role of estimation risks in the setting of investment, including Kan and Smith (2008), Kan et al. (2024), and Kan and Wang (2024), among others. Wachter and Warusawitharana (2009) show that imposing an economic prior on the R^2 leads to shrinkage of predictive coefficients and superior out-of-sample performance of a forecast-driven portfolio. More recently, Kelly et al. (2024) show that, with a properly chosen shrinkage parameter, the portfolios formed based on their high-complexity forecasts generate positive information ratios even though the high-complexity forecasts also produce negative out-of-sample R^2. Our approach shows that this dichotomy also reflects what might be called a virtue of simplicity – the use of a portfolio that estimates fewer regression coefficients than even a linear projection. Overall, we make 3 contributions to the literature. First, we show that the 46 existing predictors in the literature successfully worked overall. Second, we provide a response to the low power critique of Campbell and Thompson (2005) on OOS R^2. Third, we provide the explanation for the discrepancy between FA portfolios and OOS R^2: different penalties incurred by the estimation of difference parameters.

2 Stock return predictability: Summary of existing evidence

The standard predictive regression assumes

$$r_{t+1} = a + b_x x_t + v_{t+1} \tag{1}$$

where r_{t+1} is the excess return over the Treasury-bill at the monthly or quarterly horizon, x_t is a variable known at time t, and v_{t+1} is mean zero and is uncorrelated with x_t.[3]

For an investor choosing between an aggregate stock portfolio and the Treasury bill, and assuming (unrealistically of course) that parameters are known, $a + b_x x_t$ is an input into portfolio choice. If it is not matched one-for-one with the conditional variance of r_{t+1}, or with changes in risk aversion, then x_t will lead the investor to "time the market," potentially earning far more than holding a static position. It is this possibility that accounts for the enduring interest in this simple regression.

In practice, parameters are not known, and so an investor estimates (1) using ordinary least squares (OLS), and then uses the OLS estimates $\hat{a}$ and $\hat{b}_x$ to form a forecast for stock

[3]Throughout, we use excess returns in levels, so that r_{t+1} is the level return on the market, minus the level return on the riskfree asset.

returns and as an input into portfolio choice.[4]

Over time, the literature has uncovered many such variables. Goyal et al. (2024) organize 46 of these conveniently, expanding the data both backwards and forwards in time (the first two columns of Table 1 give the start and date). The next two columns replicate their in-sample results. Barring researcher error, all variables at one time were statistically significant in at least some specification at standard levels (otherwise they would not have been published). However, only 5 of the 46 achieve t-statistics of above 2 on the extended samples.[5] This tells us that stock return predictability studies usually do not hold up over time.

The next 4 columns display a second problem. Goyal et al. (2024) ask whether the predictors are better forecasts than the expanding-window mean of the stock market excess return itself. Specifically, they form the out-of-sample R^2:

$$\text{out-of-sample } R^2 = 1 - \frac{\sum_{t=0}^{T}(r_t - (\hat{a}_{t-1} + \hat{b}_{x,t-1}x_{t-1}))^2}{\sum_{t=0}^{T}(r_t - \bar{r}_{t-1})^2}$$

Here r_t is the excess return realized at time t and $\hat{a}_t$ and $\hat{b}_{xt}$ are estimated using an expanding window of data , as is the sample mean $\bar{r}_t$.[6] A reduced mean-squared-error (MSE) from using the predictor variables will imply a numerator that is small relative to the denominator, and a positive out-of-sample R^2. The smaller the numerator, the closer the out-of-sample R^2 comes to 1. In contrast, if the numerator is larger than the denominator, the predictor variable adds noise, and the out-of-sample R^2 is negative.

Column 7 of Table 1 reports the out-of-sample R^2. Echoing the main conclusion of Goyal et al. (2024), this table shows that the out-of-sample performance tends to be poor. Slightly more than half of the variables have negative out-of-sample R^2s, implying that the

[4]OLS is consistent for a and b_x but not unbiased (Stambaugh, 1999), with the degree of bias depending on the persistence of x_t and the correlation between shocks to x_t and v_{t+1}. One can incorporate the statistical estimation directly into the decision problem using Bayesian methods, given prior beliefs on a and b_x; see Kandel and Stambaugh (1996).

[5]Assuming these variables are chosen at random and errors are independent, one might expect to see 2 or 3 above 2 by chance. However, neither of these assumptions is satisfied, and this number alone is not decisive regarding rejecting the null of no predictability.

[6]The computation uses all the data available, but never less than twenty years.

predictor variable adds noise in a forecasting exercise.[7] Following Goyal et al. (2024), we also consider out-of-sample R^2 statistics under the adjustment proposed by Campbell and Thompson (2008). In equilibrium, assuming investors are risk averse, the equity premium cannot be negative. Thus any negative values of $\hat{a} + \hat{b}_x x_t$ can only represent noise. Table 1 shows that imposing this constraint leads to out-of-sample R^2 statistics that are less negative on average. The table considers two modifications of the Campbell and Thompson R^2: the first uses a minimum of ten rather than 20 years of data, the second starts in 1946. While individual values of the out-of-sample R^2 change, neither make much of a difference, with the median, in both cases, remaining negative.

3 Forecast-agnostic alphas

In the cross sectional return forecasting literature, it is common practice to form long-short portfolios according to a predictor's values across stocks. For example, Fama and French (1993) create the HML factor, which takes a long position in the 30% of stocks with the highest book-to-market ratios and shorts the 30% of stocks with the lowest book-to-market ratios. Fama and MacBeth (1973) create unit-beta-exposure, market-neutral portfolios that long high-beta stocks and short low-beta stocks with weights linear in betas. Unlike the calculation of an R^2, these portfolios require no explicit forecasts for stock returns. In fact, they only require the predictors underlying those forecasts, and their returns are used to evaluate a predictor's success.

We construct analogous portfolios for time-series predictors. Our construction is not new to the literature—its logic goes back to Fama and MacBeth (1973), but it involves modifications to circumvent look-ahead bias in the portfolio construction process and to target a fixed risk budget. We then ask whether such portfolios created for the same set of predictors in Table 1 can add value in a portfolio. In fact, Goyal et al. (2024) consider similar

[7] We also use data going back to 1871 to form the expanding window mean, whereas Goyal et al. (2024) use the data over the time period for which the predictor is available. Our method leads to lower out-of-sample R^2 statistics on average.

portfolios; however, they arrive at the opposite conclusion to us. This is because they evaluate whether the portfolio outperforms the market, not whether it has a positive risk-adjusted return. However, as Baks et al. (2001) show, α answers the question of whether a strategy should be used in a portfolio. Looking at raw excess returns might mean undercounting performance when it is due to taking on lower levels of risk, overcounting when it takes on higher levels of risk. Furthermore, as Moreira and Muir (2017) emphasize, α helps distinguish between expected returns that are time-varying due to second moments and expected returns that vary separately from these second moments, a crucial difference for how the portfolio would be used by investors.

Given a predictor x, we construct a zero-cost market-timing strategy that rebalances positions in two assets: the aggregate stock market and the riskfree bond. The strategy implements two rules with data available in real time. First, it places proportional weights on the market relative to the predictor value, with a target mean of zero. Second, the portfolio weight is scaled so that the strategy has a target volatility of, say, 5% per month. This volatility target should come from investors' preference, and we use 5% for the purpose of illustration. As shown in Table B1, our main results are not sensitive to this choice.

Let x_t be the predictor value at the end of period t, and let $\overline{x}_{t-1}$ be the expanding-window mean of x computed with data up to $t-1$. Then $x_{t-1}^{dm} = x_{t-1} - \overline{x}_{t-1}$ is a best estimate of x_{t-1} in excess of its long-run mean, using data up to $t-1$. Furthermore, $\overline{r}_{t-1}^{x,unsigned} = \frac{1}{t-1}\sum_{s=1}^{t-1}(x_s^{dm} - \overline{x^{dm}}_{t-1})r_{s+1}$ is the historical return of a trading strategy following x^{dm}. It accounts for variation in demeaning x with real-time data, and at the same time uses all information up to $t-1$ to center the trading strategy. We call it unsigned because it makes no attempt to determine whether the predictor is positively or negatively associated with future returns. We define the raw (signed, unscaled) portfolio weight at the month of $t-1$ as

$$w_{t-1}^{\text{raw}} = \text{sgn}(\overline{r}_{t-1}^{x,unsigned}) \cdot x_{t-1}^{dm}.$$

The unscaled portfolio return is this weight multiplied by the market excess return in period t. The expanding-window unscaled portfolio volatility $\bar{\sigma}_{t-1}^{x}$ is the unscaled portfolio's historical return volatility up to $t-1$. The scaled portfolio weight is $w_{t-1}^{\text{raw}} \frac{0.05}{\bar{\sigma}_{t-1}^{x}}$. We further truncate the weight on the market at -3 and 3, avoiding leverage that is unattainable in practice. The portfolio's construction involves no look-ahead bias, and the scaling can be adjusted according to an investor's risk appetite. We refer to these portfolios as forecast-agnostic (FA) portfolios, whose construction uses no quantitative return forecasts.

The first two columns of the right panel in Table 1 tabulate alphas and their individual t-statistics of the FA portfolios for the 46 predictors. The alphas are appropriate performance statistics for our FA portfolios, because these portfolios target zero average equity weight over time. This is close to, but different from targeting zero average equity beta, because a predictor may correlate with expected stock market volatility. We therefore create zero ex-ante long-term beta FA portfolios by adding $\hat{\beta}_t$—the estimated stock market beta of an FA portfolio estimated in month $t-1$—units of the stock markets. The last two columns report the average returns and the associated t-statistics of these zero ex-ante long-term beta FA portfolios.

Two points are worth noting. First, the alphas and the average returns have a positive mean and median across the predictors. Second, while the out-of-sample R^2 can sometimes lead to catastrophic results (for example, `vrp`), negative alphas are small in magnitude. Table 2 further evaluates the overall statistical significance of these alphas and average returns. In Panel A, the first column conducts a portfolio-month panel regression of these portfolio returns on the contemporaneous market returns and a constant. It shows that the average alpha across the portfolio-month observations, as captured by the coefficient on the constant, is 16.1 basis points per month. The associated t-statistic of 4.57 indicates that the alpha is highly statistically significant at the 1% level. Column (2) starts the evaluation from 1946 and conveys a similar message. Column (3) creates a meta portfolio averaging across all available FA portfolios and evaluates its alpha with a time-series regression. Compared to the previous two rows, it underweights observations in cross sections in which more predictors are available. This average alpha appears higher at 36 basis points per month but comes with a lower t-statistic of 2.87. This is owing to volatile portfolio returns at the beginning of the sample, which, as fewer predictors were available then, are overweighted in this method. The last column excludes this period and behaves similarly to its panel counterpart. Panel B is on the zero ex-ante long-term beta FA portfolios and convey a similar message to Panel A. Even though these portfolios target zero long-term beta ex-ante, their ex-post betas, presented in the first row of Panel B, are not exactly zero. This is realistic, as risk control in real time is imperfect. However, we see do see that these betas are substantially closer to zero than those in Panel A. In this case, both CAPM alphas and average returns are meaningful.

Second, the alphas of the FA portfolios are, perhaps surprisingly, dissimilar to the out-of-sample R^2s. Table 3 shows that, even with the same burn-in period of 20 years, the alphas have correlations in the 0.2-0.3 range with the out-of-sample R^2s in Table 1. This is true both post-1926 and 1946. The point is further illustrated in Figure 2, which shows that sorting the out-of-sample R^2s does not sort the FA alphas. The FA alphas more positively correlate with the in-sample t-statistics, though at 0.46, this correlation is not high. The average returns of the zero ex-ante long-term beta FA portfolios convey a similar message.

Comparing the top and bottom panels in Figure 2, we further see that the start time of evaluation may make a difference on certain metrics. For example, the predictor `de`, the dividend payout ratio, appear to have a high alpha post-1926 but not post-1946. This is because its good investment performance is concentrated in the earlier part of the sample in which the market was more volatile. Our results therefore always feature a version that goes back to 1926, which is based on more data, and a version that goes back to 1946, which is based only on the more modern post-war data. As we have seen in Tables 1 and 2, our main messages remain the same in these two samples.

We further show that our FA portfolios can substantially improve the utility of a mean-variance investor with various degrees of risk aversion in a real-time portfolio choice setting. The practice follows that in Campbell and Thompson (2008). Specifically, we start with a mean-variance investor whose utility is characterized by: $E[r_p] - \frac{\gamma}{2} Var[r_p]$. Here r_p is the return of the investor's portfolio in excess of the risk-free rate, γ is risk aversion, and $E[\cdot]$

and $Var[\cdot]$ compute its realized mean and variance over a given sample period. We consider three types of investors under a range of γs. The first type statically holds 100% of the stock market. The second type optimizes between a risk free bond and the stock market, placing a weight of $\overline{\mu}(mkt)_t/\gamma\overline{\sigma}(mkt)_t^2$ on the stock market. Here, weight is the expanding-window mean of the stock market divided by its expanding-window variance and γ. The third type of investors have an additional position of the aggregate zero ex-ante long-term beta FA portfolio, which is that used by Columns 3 and 4 in Panel B of Table 2. The weight is this aggregate FA portfolio returns' expanding window mean divided by expanding window variance and then γ. This approach uses expanding window quantities to forecast future means and variances, and further assumes that the assets are uncorrelated. The creation of the zero ex-ante long-term beta FA portfolio attempts to make this assumption hold. Optimizing directly on the simple FA portfolios leads to equivalent results.

The results are reported in Table 4. Type 2 investors who optimize between the market and the risk free bond do not always outperform the simple buy-and-hold type 1 investors. This is because the expanding-window mean and variance of the stock market excess returns deviate from the full-sample quantities, which of course contain look-ahead bias. Type 2 investors do much better than type 1 when γ are large or small, because in those cases the static stock market weight of 100% fails to account for these risk aversions and becomes highly suboptimal. The main message in this table is that once the FA portfolios are placed in the opportunity set, the Type 3 investors who take advantage of this additional asset always do substantially better than the Type 2 investors.[8] This shows that the FA portfolios, despite being relatively simple, can be easily incorporated into a portfolio choice problem and substantially improve the welfare of realistic investors in real time.

Lastly, we show that the success of the FA portfolios has little to do with the volatility timing results in Moreira and Muir (2017). Even though the FA portfolios scale with past volatility, the volatility is computed with all available data in real time. This is resulting scaling factor is therefore low-frequency—in fact, it has the lowest frequency possible. Volatility

[8]They also always do better than the Type 1 investors, even though the comparison is not conceptually the most appropriate.

managed portfolios in Moreira and Muir (2017), however, scales by return volatility in the previous month. With monthly data time-series, this is the highest frequency possible. The scaling volatility in FA portfolios and volatility managed portfolios therefore have low correlations on a predictor-month panel both post-1926 and 1946. We follow Moreira and Muir (2017) to create volatility managed portfolios for the market and each FA portfolios. The last row in Panel A and B of Table 2 compute 3-factor alphas that additionally control for the two volatility-managed portfolios. These 3-factor alphas are similar to the CAPM alphas.

The FA alphas paint a starkly different picture from that drawn from the out-of-sample R^2s. The out-of-sample R^2s tell us that the 46 predictors on average fail to forecast stock market returns for investors in real time, even if they employ the predictors at the inception of their availability. The FA alphas tell us that the very same set of predictors add significant value to such investors. This difference is important, as it leads to completely different evaluations of the market return predictability literature as a whole. Below, we provide a thorough explanation for it.

4 Why is it possible for a predictor to have a negative out-of-sample R^2 and a positive alpha?

FA alphas and out-of-sample R^2s evaluate the same set of predictors differently because they ask different things from those predictors. Positive FA alphas require only that the predictor be genuine and, in our implementation, that investors know the sign of the trade based on available data. Positive out-of-sample R^2s further require investors to know how much they will make in the trade. This additional self-awareness requirement turns out to be too high given the amount of available data.

To see where this self-awareness requirement comes from, note first that the R^2 computation involves explicit return forecasts and therefore the estimation of a regression coefficient on the predictor in real time. This coefficient is a representation of the predictor's historical performance: a time-series regression of the market excess return r_t on the predictor x_{t-1}

and a constant yields a coefficient of $b_x = \frac{1}{T}\sum_t \frac{(x_{t-1}-\bar{x})}{\frac{1}{T}\sum_t (x_{t-1}-\bar{x})^2} r_t = \frac{1}{T}\sum_t w_{t-1} r_t$ on x. Here, $\bar{x}$ is the mean of the predictor x in this sample, and T is the total number of periods. Note that b_x is the mean excess return of a portfolio that rebalances between positions in the aggregate stock market (which yields an excess return of r_t) and the riskfree bond (which yields an excess return of 0). The month-t portfolio weight on the stock market $w_{t-1} = \frac{(x_{t-1}-\bar{x})}{\frac{1}{T}\sum_t (x_{t-1}-\bar{x})^2}$ is i) linear in the predictor's value x_{t-1}, ii) on average 0, so that the portfolio is on average market neutral in this sample, and iii) scaled by $\frac{1}{T}\sum_t (x_{t-1}-\bar{x})^2$, so that the portfolio on average has unit exposure to x (i.e., $\frac{1}{T}\sum_t \frac{(x_{t-1}-\bar{x})}{\frac{1}{T}\sum_t (x_{t-1}-\bar{x})^2} x_{t-1} = 1$) in this sample.[9]

Using this coefficient to form a quantitative forecast of the market excess return implicitly assumes that the performance of the predictor's coefficient portfolio is sufficiently stable so that its past performance reliably predicts its future performance at any point in history. This assumption is paradoxical: it implies implausibly high t-statistics, Sharpe ratios, and R^2s. In the most extreme case, if we want the expanding-window coefficient to always equal the in-sample coefficient so that the out-of-sample R^2 equals the in-sample R^2, we need the coefficient portfolio's return volatility to be 0, its Sharpe ratio to be infinite, the predictor's in-sample t-statistic to be infinite, and the in-sample R^2 to be 1. Realistically, given the in-sample R^2 and t-statistics we observe, using (without any adjustment) this expanding-window regression coefficient to compute a quantitative forecast is likely counterproductive, and the out-of-sample R^2s are likely poor even if the predictors are genuine. We formally illustrate this point in the simulation section below.

Our FA portfolio, on the other hand, requires no quantitative prediction of the predictor's performance.[10] The size of the bet on the market is based on a constant risk appetite (5% in our example) and does not change with the past returns of the predictor's coefficient portfolio. This is consistent with the construction of cross sectional factors (e.g., HML, SMB, MOM, etc.), in which the stock weights also do not change with the past factor returns. On the

[9]If x does not have an economically interpretable scale itself, the scaling in step iii) has no economic interpretation either. Our FA portfolio is designed to mimic this coefficient portfolio, except that we replace the in-sample mean with the expanding mean for implementability, adjust the scaling in iii) for economic interpretability, and impose a position constraint that is inevitable in reality.

[10]If one is evaluating predictors *individually*, it is not obvious why such quantitative prediction is beneficial in the first place.

other hand, this constant risk appetite contrasts the calculation of out-of-sample R^2, which is driven by more dispersed forecasts when the betas estimates become more dispersed, even though variations in those estimates are driven in no small part by noise. The improvement is large, as the amount of available data is small (again, more on this in Section 5 and 6). Therefore, the FA portfolio alphas appear better than the out-of-sample R^2s overall. A well-behaved FA portfolio does depend on the stability of the predictor's mean and volatility over time, and, taken as a whole, the predictors exhibit such stability.

The variance risk premium shows the danger of using the predictive coefficients directly. Without the Campbell and Thompson (2008) adjustment, it has an extreme negative out-of-sample R^2 of -19.51%, the lowest of any of the predictor variables. We plot in Figure 1 the univariate expanding-window regression coefficients of the stock market excess return on the predictors and a constant. The predictors are signed so that the full-sample coefficients are positive, and scaled to have a standard deviation of 1 to make the coefficients comparable. The difference between the out-of-sample R^2 and the in-sample R^2 arises because these coefficients are not constant. As the sample size grows over time, the coefficients become increasingly stable towards the right end of the plot, resembling the full-sample coefficients, which are the last observation in each series. Hence, the largest deviations occur at the beginning of the sample. These large deviations drive poor out-of-sample performance that we avoid by using only the sign, and not the magnitude, of the predictive coefficient.

5 Analytical results

The previous section argues that we are unlikely to observe stable, well-estimated predictive coefficients, and hence in-sample performance that translates into high out-of-sample R^2 coefficients. Requiring the out-of-sample R^2 to equal the in-sample R^2 implies an infinite in-sample t-statistic. In contrast, we observe an average in-sample absolute t-statistic of 1.06 and an average in-sample R^2 of 32 basis points. Nonetheless, it is striking that so many out-of-sample R^2 values are negative.

In fact, as we allude to earlier in the introduction, this is a known issue of out-of-sample R^2 (Campbell and Thompson 2005, Goyal and Welch 2008).[11] We first illustrate this lower power issue via the following two questions: What out-of-sample R^2 values should one expect, given the observed amount of in-sample predictability? How large a penalty does the instability in the estimated beta coefficient levy on out-of-sample R^2? We then show that portfolio alphas offer a simple solution to this issue. We start with a simple statistical setting featuring uncorrelated shocks, homoskedasticity, and predictors following AR(1) processes. This simple setting allows us to derive simple, closed-form quantities that provide a useful rule of thumb. We conduct simulation exercises with more complex features in the next section.

Assume the following data-generating process:

$$r_{t+1} = \mu + b_x(x_t - \mu_x) + \epsilon_{t+1}, \quad E[\epsilon_{t+1}] = 0, \quad \mathrm{Var}(\epsilon_{t+1}) = \sigma^2,$$

$$x_{t+1} = \mu_x + \rho(x_t - \mu_x) + u_{t+1}, \quad E[u_{t+1}] = 0, \quad \mathrm{Var}(u_{t+1}) = \sigma_x^2,$$

with u and ϵ uncorrelated. Define

$$\bar{x}_{t-1} = \frac{1}{t}\sum_{s=0}^{t-1} x_s, \quad \bar{r}_t = \frac{1}{t}\sum_{s=1}^{t} r_s, \quad \hat{b}_{xt} = \frac{\sum_{s=0}^{t-1}(x_s - \bar{x}_{t-1})r_{s+1}}{\sum_{s=0}^{t-1}(x_s - \bar{x}_{t-1})^2}, \quad \hat{a}_t = \bar{r}_t - \hat{b}_{xt}\bar{x}_{t-1}, \quad V = \frac{\sigma_x^2}{1-\rho^2}.$$

We derive the following:

Proposition 1. *The mean squared error of $\bar{r}_t$ satisfies:*

$$\mathrm{MSE}(\bar{r}_t) = \mathbb{E}\big[(r_{t+1} - \bar{r}_t)^2\big] = \sigma^2 + b_x^2 V + \frac{1}{t}\left(\sigma^2 + b_x^2 V\right) + O\big(\frac{1}{t^2}\big)$$

The mean squared error of $f_t = \hat{a}_t + \hat{b}_{xt}x_t$ satisfies:

$$\mathrm{MSE}(f_t) = \mathbb{E}\big[(r_{t+1} - f_t)^2\big] = \sigma^2 + \frac{2\sigma^2}{t} + O\big(\frac{1}{t^2}\big)$$

[11] See McCracken (2007) and Zevallos and Rubesam (2026) for comprehensive analyses of a more general class of out-of-sample metrics.

Proof: see Appendix C.

Proposition 1 leads to a simple approximation of the out-of-sample R^2:

$$\text{out-of-sample } R^2 \approx \frac{\text{MSE}(\bar{r}_t) - \text{MSE}(f_t)}{\sigma^2} = \frac{b_x^2 V}{\sigma^2} - \frac{1}{t} \tag{2}$$

Here, $\frac{b_x^2 V}{\sigma^2}$ is approximately the population R^2. This simple rule of thumb discards a term $\frac{b_x^2 V}{t\sigma^2}$, which is proportionally small when the population R^2 is small and the sample size t is reasonably large. The division by σ^2 instead of $\text{MSE}(\bar{r}_t)$ also relies on these assumptions. In our return-forecasting setting, we indeed have low R^2s and reasonably large sample sizes such as 240 months,[12] making the approximation reasonably accurate.

From this rule of thumb, we can see that the out-of-sample R^2 is a penalized version of the population R^2. In our setting, because the population R^2s are low, even when we have large t such as 240, $\frac{1}{t}$ may still exceed the population R^2, rendering the out-of-sample R^2 negative in expectation. For the out-of-sample R^2 to be positive with 20 years of estimation data, the population R^2 needs to exceed 1/240 which is roughly 40 basis points. A quick look at Table 1 reveals that this is clearly too high a threshold, as the average *in-sample* R^2 is only 32 basis points. Furthermore, in-sample R^2 considerably overstates the population R^2. Under a simple estimate of the population R^2 (see Appendix D for the derivation) that is $\max(\frac{(\text{t-stat})^2 - 1}{\text{number of obs}}, 0)$, Table 1 provides an estimate of the average population R^2 of 14 basis points. Assuming an average sample length of 500 months, this average population R^2 is not sufficient to overcome the 20 basis point penalty. It seems that, based on equation (2) and the statistics in Table 1, it is already unsurprising that the out-of-sample R^2s are low on average—a point we will formally show with simulation results.

On the other hand, alphas of portfolios that trade on the predictors suffer from no such penalty. Below, we derive the alphas of the forecast-driven (FD) portfolios, as they are in

[12] On the surface, it may appear that increasing the frequency of the data mechanically improves the out-of-sample R^2 by multiplying t. This is typically not the case. When the predictor is persistent, the population R^2 decreases as the data frequency increases. On the other hand, monthly predictors with low persistence (e.g., those proposed in Guo 2025 and Guo and Wachter 2025a) tend to do well under the out-of-sample R^2 metric.

closed form and provide straightforward intuitions. Expected alphas of the FA portfolios are not in closed form even with the assumption of normality. In Section 6, we compare FA and FD portfolios and explain why we propose the FA portfolios even though FD portfolios are theoretically more convenient.

Proposition 2. *The expected alpha of the following forecast-driven portfolios satisfies:*

$$\mathbb{E}\Big[\hat{b}_{xt}(x_t - \bar{x}_{t-1})(r_{t+1} - \mu)\Big] = b_x^2 V \left(1 - \frac{1}{t} \cdot \frac{\rho}{1-\rho}\right) + O\big(\frac{1}{t^2}\big)$$

The expected alpha of the infeasible portfolio that uses the true b_x value satisfies:

$$\mathbb{E}[b_x(x_t - \bar{x}_{t-1})(r_{t+1} - \mu)] = b_x^2 V \left(1 - \frac{1}{t} \cdot \frac{\rho}{1-\rho}\right) + O\big(\frac{1}{t^2}\big)$$

The expected alpha of the infeasible portfolio that additionally uses the true μ_x value satisfies:

$$\mathbb{E}[b_x(x_t - \mu_x)(r_{t+1} - \mu)] = b_x^2\, V.$$

Proposition 2 tells us that the FD portfolios have the same expected return as the infeasible portfolios that make use of the true value of b_x, up to a second-order approximation. In other words, there is no penalty in mean portfolio return for the fact that $\hat{b}_{xt}$ is estimated. This is because errors in $\hat{b}_{xt}$ depend on $\sum_{s=0}^{t-1}(x_s - \bar{x}_{t-1})\epsilon_{s+1}$, which is orthogonal to both u_t and ϵ_{t+1}. The expected returns equal the expected alphas, as the expected beta loading of the portfolio on the market is zero. A small penalty of $b_x^2 V \frac{1}{t}\frac{\rho}{1-\rho}$ exists relative to the case in which the true values of both μ_x and b_x are known. In our setting, this penalty is substantially smaller than the penalty of $\frac{1}{t}$ in out-of-sample R^2 in relative terms. Even with a high ρ of 0.98, a t size of 50 suffices to make this expectation positive. This helps explain why return-based measures exhibit power to detect predictability where the out-of-sample R^2 does not.

Conceptually, the distinction again traces back to whether a performance metric requires the knowledge of expected profit. An R^2 metric does, and therefore, when the realized returns are in the right direction but too much in scale—say forecasting 3% but receiving only 1%—the metric penalizes the large forecast error and deteriorates relative to the neutral benchmark of 0%. As shown in Appendix C, this leads to a term of $\mathbb{E}\left[(x_t - \bar{x}_{t-1})^2(b_x - \hat{b}_{xt})^2\right]$ in the mean-squared forecast errors that is eventually responsible for the extra penalty of $1/t$. A portfolio, however, requires no such knowledge. Its return would in fact be positive in this situation, as the bullish view drives a long position of the market, on which it receives a smaller-than-expected but positive return. Making less money than expected is thus not penalized but rather viewed as a success. Conceptually, neither approach is definitively better than the other. However, with weak signals—as we likely do when forecasting stock market returns—the portfolio approach has substantially higher power.

There are limits to the result, most importantly the expected return is, of course, measured with noise even if its bias is small. As we will show, using the FA alpha relative to the FD alpha helps reduce this noise.

6 Simulation

6.1 Comparing FA alphas to the out-of-sample R^2

Our simulations consist of 47 monthly time series: stock market excess returns and 46 predictors, with the same start and end months as the historical time series used in Table 1. In Panel A of Table 5, the predictors follow AR(1) processes with a mean of zero and AR(1) coefficients estimated from the real predictor data. Both the estimation and the simulation are done at the same frequency as the historical time series—monthly, quarterly, or annual. The innovations of the predictors are i.i.d. normal shocks with a volatility of 30 basis points multiplied by the square root of one minus the square of the AR(1) coefficients. The value of 30 basis points is the population volatility of all the predictors. It governs the strength of predictability in our simulations and is calibrated to roughly match the historical series. In the first 3 columns with the header "no predictability," the simulated returns equal the full-sample average monthly stock market excess return (58 basis points) plus i.i.d. normal shocks with a mean of zero and a volatility equal to the full-sample volatility of monthly stock market excess returns (4.78%). In the next 3 columns, the simulated returns equal the aforementioned two terms plus 46 additional terms, each being the product of one of the 46 predictors and a beta coefficient randomly drawn from a uniform distribution between -1 and 1.

In the "no predictability" simulations, all predictors are spurious. In the "with predictability" runs, they are all genuine, even though predictors with a beta farther from zero are more important than those close to zero. Panel B presents similar sets of simulations, except that the beta coefficients and the predictors' volatility are estimated from data. The innovation terms in both the predictors and the returns have time-varying conditional variance estimated with a GARCH(1,1) model on the data.

In each simulation, we compute 5 statistics across the 46 simulated predictors: i) the average absolute value of Newey-West t-statistics, ii) the average in-sample R^2, iii) the average out-of-sample R^2, iv) the average out-of-sample R^2 after applying the positive-sign restriction from Campbell and Thompson (2008), and v) the average alphas of the FA portfolios. The quantities are computed using exactly the same approach as in Tables 1 and 2. We report the 5th, 50th, and 95th percentiles across 1000 simulations in the first 6 columns of Table 5, and report the corresponding value computed from historical time series in the last column.

In the with-predictability simulations of Panel A, the medians of the average t-statistics and average in-sample R^2 roughly match those computed from the historical time series. This is by design—the dispersion of the simulated predictors is set to match these two moments, which are computed to capture the amount of in-sample predictability. In the next two rows, we see that the medians of the average out-of-sample R^2 and the average CT out-of-sample R^2 are both negative at -0.22% and -0.10%, even when the predictability is genuine. This shows that, given the amount of in-sample predictability that we observe in real data, we likely won't observe positive out-of-sample R^2s. In Panel B, the message is similar. Here, the average t-statistics and average in-sample R^2 appear somewhat higher than in real data, even though the predictors' dispersion and their regression coefficients are estimated. This is likely because the GARCH(1,1) model does not fully capture the heteroskedasticity structure in the data. We nonetheless observe negative median average out-of-sample R^2s, with or without the positive-sign restriction.

These negative medians reflect a known power issue of the out-of-sample R^2s. Notably, it was pointed out by Campbell and Thompson (2005) in an early draft of their 2008 paper. Goyal and Welch (2008) respond to this critique by arguing that the out-of-sample R^2s should not be used alone but rather as a conditional diagnostic. Specifically, they should be examined only when the in-sample diagnostics are significant. Here, even though the predictability is genuine, the average in-sample t-statistics are only 1.34 and 1.67. Since the in-sample t-statistics cannot reliably detect predictability this weak, the poor out-of-sample R^2s are unsurprising. We further demonstrate this point in Table 6, which shows that if the strength of predictability increases, the average t-statistics rise above the conventional threshold of 2, and the average out-of-sample R^2s become positive.

While we agree with Goyal and Welch (2008)'s response, we provide a simpler response to Campbell and Thompson (2005) in which an out-of-sample statistic is used on a standalone basis. In the last rows of Panels A and B, we compute the median average alphas of the FA portfolios. When the predictors are genuine, these medians are positive and close to the observed values in real data. Furthermore, even the 5th percentiles across the simulations are positive, showing good power of this statistic in detecting the genuine predictability scenario. When the predictors are spurious, we see that the average alphas are roughly centered around 0 as expected, and the 95th percentiles are not much higher than the 5th percentile values among the with predictability simulations. Even at a 5% significance level on the null of no predictability, the average alpha of the FA portfolios achieves power close to 95% against the alternative of real predictability.

Goyal and Welch (2008) make an important point, which is that look-ahead bias in parameter estimation is especially important in a time-series setting and should be carefully removed before a predictor can be fairly evaluated. Out-of-sample diagnostics address this point by removing the look-ahead bias—both out-of-sample R^2s and FA alphas are such diagnostics. Using the average FA alphas as an example, we show that such statistics can and should be used independently of any in-sample statistic.

6.2 How much data and how strong a predictor do we need?

An important economic difference between the FA alphas and the out-of-sample R^2s is whether they try to distinguish between strong and weak predictors. The FA portfolios allocate to each predictor an equal risk budget of 5% per month. This budget does not change with the predictor's past performance. On the other hand, out-of-sample R^2s use forecasts computed from expanding-window regression coefficients, which become more disperse if the predictor historically does well. We show that when return predictability is as weak as the data suggest, giving more dispersion to historically stronger predictors is counterproductive. However, this is not generally true. In fact, with i) more data or ii) stronger predictability, one can more reliably distinguish stronger predictors from weaker ones. These two points can already be seen in the rule of thumb in equation (2). We further demonstrate this point by conducting our simulations with longer samples and stronger predictability.

In these simulations, we form an alternative portfolio for each predictor using a similar approach to that in Campbell and Thompson (2008). Specifically, we first compute the expanding-window coefficients estimated by regressing stock market excess returns on the demeaned predictor (with its expanding-window mean) and the constant. We then take the product of the demeaned predictor and its coefficient, and multiply this product by 200. The multiplier of 200 comes from $\frac{1}{2\times 0.05^2}$. Here, 2 is risk aversion, 0.05 is roughly the monthly volatility of the stock market, and the multiplier converts expected market returns to the position on the market for a risk-averse investor optimizing across the market and the riskfree asset. The risk aversion of 2 is chosen such that the investor would on average choose to place roughly 100% weight on the market. The resulting portfolio weight is then truncated at 3 and -3, as in our FA portfolios. We refer to these portfolios as forecast-driven (FD) portfolios.

As FA and FD portfolios may have different exposures to the predictors, we cannot directly compare their returns or their CAPM alphas. We therefore compute the two-factor alpha the FD portfolios controlling for the market and the FA portfolios, as well as the 2-factor FA alphas controlling for the market and the FD portfolios.

Table 6 reports simulation results with 1-30 times as much data as the original simulation in Panel A of Table 5, although the 20-year estimation periods and the CRSP starting point are fixed. We see that, with 5 times the data, the out-of-sample R^2s are comfortably positive. The average t-statistics at that point are 2.35. With 20 times the data, the FD portfolios start to add alphas on top of the FA portfolios, and the average t-statistics are about 4.41. This point is visually illustrated by the top panel of Figure 3.

In Panel B, we simulate with 10 times the data and stronger predictability. That is, the dispersion of the predictors increases from 30 to 180 basis points across the columns. Here, we see that with 3 times stronger predictability, FD portfolios start to add more alphas on top of the FA portfolios than the other way around. The average t-statistic is around 7.47 at that point. This point is visually illustrated by the lower panel of Figure 3.

Overall, we show that, with a moderate in-sample t-statistic of around 2.3, we can expect to see positive out-of-sample R^2s on average. The practice of chasing past performance (by estimating the expanding-window coefficients) can be productive with more data or stronger predictability, although that requires t-statistics that are rather high.

6.3 Comparing FA and FD portfolios

We saw that the FA portfolios appear to add much higher alphas relative to FD portfolios in simulations with realistic parameter values. The first column of Panel A in Table 6 shows that the 2-factor alpha that FA portfolios add to the market and the FD portfolios is on average 4.57 basis points per month. This is sizable, given that the average CAPM alpha is 11.2 basis points. This can also be observed in terms of the Sharpe ratios of the simulated FA and FD portfolios. Figure 4 is a histogram of one million FA Sharpe ratios and one million FD Sharpe ratios. The FA Sharpe ratios are clearly higher on average. Furthermore, in the real-world monthly panel of the 46 predictors, the FA portfolios add a monthly 2-factor alpha of 4.96 basis points with a t-stat of 3.61 over the market and the FD portfolios, whereas the 2-factor FD alpha is near zero with a t-stat of 0.22. We now pinpoint the source of such improvement from FD to FA portfolios.

One obvious thing to note is that FD portfolios use the values of the estimated betas, whereas the FA portfolios do not explicitly estimate these betas. However, the FA portfolios effectively use the signs of these betas. The second step of the construction of the FA portfolios determines the direction of the trade using $\text{sgn}(\bar{r}_{t-1}^{x,unsigned})$. This is in fact equivalent to using the sign of an expanding window regression coefficient of returns on x^{dm} and a constant, $\text{sgn}(\hat{b}_{xt})$. This does not reduce FA's dependence on estimated coefficients—it merely depends on a different functional form, and this is not the source of FA's advantage.

The source of the improvement lies in how FA portfolios account for the noise in the estimated beta compared to the FD portfolios, and how they allocate risk over time. This is in the third, scaling step of FA portfolio construction. Section 5 shows that the fact that the beta is estimated imposes no penalty on the mean of the FD portfolios. However, there is a penalty in the form of increased portfolio volatility. At the beginning of the sample, less data are used in the estimation of the betas, leading to higher variance in those estimates. As those beta estimates drive the FD portfolios, they are naturally more volatile at the beginning of the sample even though their expected returns are not lower then. An optimal approach to trade on a given predictor across time—whether it is FA, FD, or otherwise—should therefore allocate less risk budget earlier in the sample than later. Earlier in the sample there is simply less data to help construct a trade, however the trade is conducted. The FD portfolio weight is the opposite of that: $\frac{\hat{b}_{xt}(x_t-\bar{x}_{t-1})}{\gamma\sigma_{mkt}^2} = \frac{f_t-\bar{r}_t}{2\times 0.05^2}$ is more volatile at the beginning of the sample.[13] On the other hand, the FA portfolio targets a constant volatility throughout the sample, and this flat risk budget across time is much closer to the optimal trading strategy, albeit not optimal either.

This point is demonstrated in the top panel of Figure 5. The blue solid line and the orange

[13]Campbell and Thompson (2008)'s implementation of the FD portfolio uses a rolling or expanding window to estimate the *stock market* return's volatility σ_{mkt}. The issue remains.

dashed line are the period-by-period Sharpe ratios for the FA and FD portfolios, starting from the 240th month. For a given period, we first compute the average of 1 million FA (FD) portfolio returns, then compute the standard deviation of these 1 million returns, and lastly compute their ratio. This is different from the normal Sharpe ratios in Figure 4, which are computed for each simulated time-series of portfolio returns rather than a given period. The period-by-period Sharpe ratios remove the effect of (potentially time-varying) scaling within those portfolios. They reveal the fundamental risk-return trade-off of the "underlying trades" of the FA and FD portfolios, which are $\mathrm{sgn}(\hat{b}_{xt})(x_t - \bar{x}_{t-1})$ and $\hat{b}_{xt}(x_t - \bar{x}_{t-1})$, respectively. To put it another way, for an arbitrary time series of portfolio weights w_t and scaling factor $c_t > 0$, the period-by-period Sharpe ratios generated by w_t and $c_t w_t$ are exactly identical, as long as c_t is deterministic and does not vary across simulations.

How, then, do the FA portfolios ultimately achieve higher Sharpe ratios, even though the risk-return trade-off of the underlying trades of the FA portfolios is not any better in each individual period? The answer is their risk allocation across those periods, shown in the bottom panel of Figure 5. It plots the standard deviation across the 1 million simulated FA and FD returns, the denominator of the Sharpe ratios in the top panel. As reasoned above, the FD portfolios allocate a higher risk budget to the early part of the sample, and thus the FD portfolios are more volatile there. The FA portfolios target a constant volatility, and the line for the FA portfolio is almost flat, though not perfectly so.

The optimal risk allocation over time would have a similar shape to the period-by-period Sharpe ratio curve. Recall the simple rule of thumb in portfolio optimization: for uncorrelated assets, the optimal portfolio weights are proportional to means over variance. For a given underlying trade (a time series of raw portfolio weights), a time-series of scaling factor is optimal when the resulting scaled portfolio has a flat period-by-period mean over variance. Equivalently, the period-by-period Sharpe ratio should be perfectly correlated with the period-by-period volatility. Since the period-by-period Sharpe ratios are inevitably upward sloping regardless of the underlying trade, the portfolio volatility should also increase over time. The FD portfolios do the opposite, and the FA portfolios do better—but still not optimally—with a flat volatility curve over time.

A few additional points are worth noting. First, our volatility-scaling approach in the FA portfolio is very different from the popular volatility-managed portfolios, à la Moreira and Muir (2017). We scale by the raw portfolio by the expanding-window volatility, whereas the volatility-managed portfolio scales by the unscaled portfolio's variance in the previous *month*. In practice, expanding-window and monthly volatility hardly correlate, as they capture the lowest and a high-frequency component of volatility (e.g., the uncompensated risk in Guo and Wachter 2025b). The critique made by Nagel (2025) therefore does not apply to our FA portfolio. This is clearly demonstrated by the alphas that control for the predictor and market level volatility managed portfolios in Table 2.

Second, it is natural to wonder why this volatility penalty matters at all. We have viewed a trading strategy for a given predictor as an asset among many. From this perspective, the alpha of each asset matters, whereas its volatility does not, assuming that it can be diversified away. This volatility penalty matters precisely because it does not diversify. Early in the sample, data are scarce for all predictors, and the penalty is high across the board. This contrasts with the portfolio returns across predictors in the same period, which do not necessarily correlate with one another if the predictors themselves are dissimilar.

Third, as alluded to above, our FA portfolios are not optimal. Better investment performance can be achieved if i) a properly chosen shrinkage parameter is employed in estimating the forecast model or ii) better risk allocation is done across time. The issue of optimal shrinkage needs its own study, because it needs to address issues like post-publication structural change and parameter uncertainty in general to be practically useful. These issues are outside the scope of our paper, even though in Appendix D, we discuss optimal shrinkage in a simpler setting than that in Section 5. We choose our FA portfolios' construction recipe for three reasons: i) its relative simplicity, ii) its analogy to the cross sectional long-short portfolios that also maintain a constant risk budget over time, and iii) when an economic model does tell us about the sign of the coefficient of the predictor—an opportunity which Campbell and Thompson (2008) argue often presents itself—this piece of information can be directly and fully incorporated in our FA portfolios by replacing the current sign-flipping logic. In that case, our FA portfolio would be fully independent of the estimated coefficient $\hat{b}_{xt}$. This is especially useful for investment purposes. The yellow dotted line in Figure 5a are the period-by-period Sharpe ratios for $\text{sgn}(b_x)(x_t - \bar{x}_{t-1})$, and it is much higher than those of the FA and FD portfolios.[14] We do not use those theory-motivated signs in any of our analyses to be conservative about misspecifications in those theories.

7 The pre- and post-sample alphas

Table 7 shows that post-publication, the alphas are essentially zero. So, taking the world as given, investors should not trade on predictors obtained from academic papers. This, however, can be either because too many investors trade on those papers and "destroy" the return predictability (à la Mclean and Pontiff (2016)) or because the statistical patterns are spurious. Thanks to the data extension of Goyal et al. (2024) and the monthly resolution of the GFD data, we can compute the pre-sample alphas of the FA portfolios. Such pre-sample alphas are valuable because the lack thereof would reflect statistical bias but not post-publication arbitrage activities. Column (1) shows that the average pre-sample alphas are in fact somewhat higher than the in-sample alphas. This speaks against the notion that the predictability patterns are owing to statistical bias, as they work pre-sample (which is out-of-sample). The post-sample alphas are positive but insignificant, suggesting that post-publication arbitraging activities do destroy the alphas. Hence, a marginal investor should not trade on predictors in published papers, but may consider trading on predictors she privately discovers at a similar standard of evidence as in published papers. This contrasts Goyal et al. (2024), according to whom investors should not even trade on those privately discovered predictors.

If a predictor forecasts stock market excess returns via predictable resolution of risks,

[14] A simple way to incorporate the sign information in the FD portfolios is to replace the coefficient with 0 when the estimated value has the wrong sign. This leads to a lower Sharpe ratio curve, as the purple line shows.

then, if investors' attitudes toward such risks do not change, the predictability should not decline post-publication. Column (2) tests this prediction by showing the post-sample alphas for predictors that are published with a rational explanation. It shows that the post-sample alphas are also not statistically distinguishable from zero (although higher than the alphas computed with all predictors). In contrast, they are significantly different from the pre-sample alphas at 1% level. One possible explanation is that investors who do not care about the modeled risk become the marginal investors post-publication. Another explanation is that the proposed risk-based explanation is incorrect in the first place. Regardless of the explanation, the pattern of post-publication fall-off is something one must account for in real-world trading, and is the fourth reason why we choose the constant risk budget in our FA portfolios—an increasing risk budget may not be optimal after all in the presence of the post-publication decline.

8 Obtaining higher out-of-sample R^2s—the value of statistical and economic constraints

One important point made by Campbell and Thompson (2008) is that constraints help improve predictors' out-of-sample performance. In particular, equity premium forecasts below 0 are theoretically hard to justify and should therefore be truncated at zero. Table 1 confirms that implementing this constraint substantially improves the out-of-sample R^2s, although the mean and median of these R^2s remain negative even with this constraint.

A critical assumption behind this constraint is that the predictability arises from predictable resolution of risks. The large post-publication decline shown in the previous section casts this assumption in doubt. It is possible that the non-negative constraint improves forecasting performance for the predictors not because the negative forecasts are economically implausible, but rather because extreme forecasts, whether too high or too low, are statistically implausible. We then implement a symmetric constraint, truncating from above return forecasts exceeding 2 times the expanding-window mean of market excess returns.

Table 8 shows that this positive-end constraint alone improves the median out-of-sample R^2 from -0.25% to -0.09%.[15] This is not far from the improvement to a median of -0.06% produced by the non-negativity constraint. Implementing both constraints further increases this median to -0.01% (and the mean to 0.03%). The improvement from this additional constraint appears appreciable but insufficient.

A less appreciated constraint proposed in Campbell and Thompson (2008) is that imposed by an economic model. Unlike the non-negativity constraint, such a constraint is not uniformly applicable. Even when they apply, the constraints are often not straightforward to obtain. We extend the work in Campbell and Thompson (2008) and apply the economic model-based constraints to 8 predictors.

The most straightforward case in this table is `rsvix`. In the model of Martin (2017), its predictive coefficient is simply 1/12 to convert annual forecasts to monthly units (and the coefficient of the constant term is 0). This constraint can then be applied to implied volatility-based predictors similar to `rsvix`, specifically `vp`, `impvar`, and `vrp`. This is because implied volatility indices such as the VIX differ from Martin (2017)'s SVIX only in weights across options and are similar in their level and variability. Even though these three papers do not have a model that directly provides the coefficient, one can impose the same constraint as in Martin (2017) based on their similarity. As `impvar` appears to be in monthly units, the predictor's coefficient is 1 and that on the constant term is 0. `Vrp` and `vp` are in annualized volatility percentage-point units. We convert them back to fractional monthly variance units and impose a coefficient of 1. We further constrain the coefficient on the constant term to be the expanding-window mean of the squared log market excess returns to be consistent with SVIX.

Another set of predictors on which we can impose economic constraints are the valuation ratios. Campbell and Thompson (2008) use a drifting steady-state valuation model to express the expected stock market return next period as the sum of the level of dividend yield and the

[15]The mean out-of-sample R^2 remains poor at -0.57%, though slightly improved from -0.62% without any constraint. This is entirely owing to vrp, in which an extreme negative (rather than positive) predictor value severely impaired its forecasting performance.

average dividend growth, estimated as the expanding-window average of the aggregate plowback ratio multiplied by the expanding-window average of aggregate return on equity. We combine this relation with equation 5.87 in Campbell (2017), adding half of the expanding-window mean of the squared log market excess returns:

$$\frac{1}{12}(d_t/p_{t-1} + (1 - \overline{de}_t)\overline{roe}_t) + \frac{1}{2}\overline{\sigma^2}_t.$$

This imposes a coefficient of 1/12 on the predictor dy along with a constant of $\frac{1}{12}(1 - \overline{de}_t)\overline{roe}_t + \frac{1}{2}\overline{\sigma^2}_t$. The same constraint can be placed on the predictor `dp`. Following Campbell and Thompson (2008), the equity premium forecasts based on ep and bm are:

$$\frac{1}{12}(ep_t\overline{de}_t + (1 - \overline{de}_t)\overline{roe}_t) + \frac{1}{2}\overline{\sigma^2}_t,$$

and

$$\frac{1}{12}(1 + \overline{de}_t(bm_t - 1))\overline{roe}_t + \frac{1}{2}\overline{\sigma^2}_t,$$

respectively. They impose coefficients of $\frac{\overline{de}_t}{12}$ and $\frac{\overline{de}_t\overline{roe}_t}{12}$ on the predictors, and a coefficient of $\frac{1}{12}(1 - \overline{de}_t)\overline{roe}_t + \frac{1}{2}\overline{\sigma^2}_t$ on the constant term.

Table 9 shows that the economically motivated constraints improve the predictors' forecasting performance overall. In our view, the value of an economic model in this return forecasting exercise is not having a strong prior about the corresponding predictor—that could in fact be counterproductive. Rather, it is to fulfill the self-awareness requirement in the forecasting exercise. That is, to inform investors about the coefficient on the predictor even without any historical data. When the amount of data is insufficient to detect a potentially weak predictor—which is generally the case in our return-predicting setting—this approach is especially valuable.

The point above goes beyond obtaining higher out-of-sample R^2s. It implies that for theories that jointly propose a predictor and a coefficient, out-of-sample R^2 is a much more powerful statistic. Table 10 conducts simulations with true predictability as in Table 5, and additionally computes average true-beta out-of-sample R^2s. The forecasts underlying these true-beta out-of-sample R^2s are the true beta value multiplied by the expanding-window demeaned predictor value plus the expanding window mean of the past returns, and the underlying benchmark remains the expanding window mean of the past returns. The table shows that the 5th percentiles of the average true-beta out-of-sample R^2s are clearly positive, representing substantially improved power. When the theory proposes the value of the coefficient, out-of-sample R^2 can be used on a standalone basis.

The literature has shown considerable interest in statistical methods that generate better out-of-sample R^2s. Examples include a simple average across forecasts (Rapach et al., 2010), shrinking toward a conservative lower bound (Li et al., 2025), isolating the fundamental component (Lin et al., 2025), and varying the strength of the shrinkage according to the predictor (Goyal et al., 2024; Kelly et al., 2024). Even though we push the adoption of FA portfolio alphas, we contribute to the return-forecasting literature by showing that economically motivated coefficient values are an additional set of possible shrinkage targets. We do not attempt to address the issue of optimal shrinkage factors in this paper, as the topic demands its own paper, even though we provide a simple rule in Appendix D.

9 Conclusion

We create forecast-agnostic (FA) portfolios for predictors of aggregate stock market returns. These portfolios target a given volatility level and avoid numerical return forecasts, thus becoming robust to errors in estimated regression coefficients. We show that, given the low strength of predictability we observe, the FA alphas are more powerful statistics than out-of-sample R^2, the classic performance metric used by the literature to evaluate the performance of time-series predictors. Circumventing the estimation of regression coefficients allows the FA portfolios to perform better than the forecast-driven trading strategies proposed in Campbell and Thompson (2008), even for samples longer than those currently available. Making the model less complex than linear regression benefits investors who time the market.

References

Bachelier, L. (1900). *Théorie de la spéculation*. PhD thesis, Faculté des Sciences de Paris.

Baks, K. P., Metrick, A., and Wachter, J. (2001). Should investors avoid all actively managed mutual funds? a study in bayesian performance evaluation. *Journal of Finance*, 56(1):45–86.

Campbell, J. (2017). *Financial decisions and markets: A course in asset pricing*. Princeton University Press, Princeton, NJ.

Campbell, J. and Thompson, S. B. (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? *Review of Financial Studies*, 21(4):1509–1531.

Campbell, J. Y. and Shiller, R. J. (1988). The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies*, 1(3):195–228.

Campbell, J. Y. and Thompson, S. B. (2005). Predicting the Equity Premium Out of Sample: Can Anything Beat the Historical Average?

Clark, T. E. and McCracken, M. W. (2001). Tests of equal forecast accuracy and encompassing for nested models. *Journal of Econometrics*, 105(1):85–110.

Diebold, F. X. and Mariano, R. S. (1995). Comparing Predictive Accuracy. *Journal of Business & Economic Statistics*, 13(3):253–263.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *Journal of Finance*, 25(2):383–417.

Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56.

Fama, E. F. and MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy*, 81(3):607–636.

Goyal, A. and Welch, I. (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, 21(4):1455–1508.

Goyal, A., Welch, I., and Zafirov, A. (2024). A comprehensive 2022 look at the empirical performance of equity premium prediction. *Review of Financial Studies*, 37(11):3490–3557.

Guo, H. (2025). Earnings Extrapolation and Predictable Stock Market Returns. *The Review of Financial Studies*, 38(6):1730–1782.

Guo, H. and Wachter, J. A. (2025a). Correlation neglect in asset prices.

Guo, H. and Wachter, J. A. (2025b). "Superstitious" investors. *Review of Asset Pricing Studies*, 15(1):1–45.

Kan, R. and Smith, D. R. (2008). The Distribution of the Sample Minimum-Variance Frontier. *Management Science*, 54(7):1364–1380.

Kan, R. and Wang, X. (2024). Optimal Portfolio Choice with Unknown Benchmark Efficiency. *Management Science*, 70(9):6117–6138.

Kan, R., Wang, X., and Zheng, X. (2024). In-sample and out-of-sample Sharpe ratios of multi-factor asset pricing models. *Journal of Financial Economics*, 155:103837.

Kandel, S. and Stambaugh, R. F. (1996). On the Predictability of Stock Returns: An Asset-Allocation Perspective. *The Journal of Finance*, 51(2):385–424. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.1996.tb02689.x.

Kelly, B., Malamud, S., and Zhou, K. (2024). The Virtue of Complexity in Return Prediction. *The Journal of Finance*, 79(1):459–503. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jofi.13298.

Li, K., Li, Y., Lyu, C., and Yu, J. (2025). How to Dominate the Historical Average. *The Review of Financial Studies*, 38(10):3086–3116.

Lin, Y., Wu, C., Zhou, G., and Zhu, S. (2025). Market Return Decomposition and Equity Risk Premium Predictability.

Malkiel, B. G. (2003). *A random walk down Wall Street*. W. W. Norton and Company, Inc., New York, NY.

Martin, I. (2017). What is the expected return on the market? *Quarterly Journal of Economics*, 132(1):367–433.

McCracken, M. W. (2007). Asymptotics for out of sample tests of Granger causality. *Journal of Econometrics*, 140(2):719–752.

Mclean, R. D. and Pontiff, J. (2016). Does academic research destroy stock return predictability? *Journal of Finance*, 71(1):5–32.

Moreira, A. and Muir, T. (2017). Volatility-Managed Portfolios. *The Journal of Finance*, 72(4):1611–1644. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jofi.12513.

Nagel, S. (2025). Seemingly Virtuous Complexity in Return Prediction.

Rapach, D. E., Strauss, J. K., and Zhou, G. (2010). Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy. *The Review of Financial Studies*, 23(2):821–862.

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

Wachter, J. A. and Warusawitharana, M. (2009). Predictable returns and asset allocation: Should a skeptical investor time the market? *Journal of Econometrics*, 148:162–178.

Zevallos, M. and Rubesam, A. (2026). Finite-sample properties of the Campbell and Thompson out-of-sample R2<math><msup is=”true”><mrow is=”true”><mi is=”true”>R</mi></mrow><mrow is=”true”><mn is=”true”>2</mn></mrow></msup></math>. *Economics Letters*, 265:113011.

Table 1: Predictor-level estimation

	In-sample				Out-of-Sample							
Signal	Start	End	t	R^2	CT R^2	R^2	CT R^2 (10-yr)	CT R^2 (1946)	Alpha	t	Avg ret (0 ealt β)	t
`vp`	1990-01	2021-12	1.64	1.94	4.72	4.72	-0.46	4.72	0.01	0.06	-0.01	-0.03
`impvar`	1996-01	2023-08	0.63	0.28	0.49	0.49	-2.89	0.49	0.00	-0.01	0.19	0.37
`vrp`	1990-01	2023-12	0.23	0.10	-1.00	-19.51	-0.22	-1.00	-0.11	-0.26	-0.31	-0.71
`lzrt`	1926-01	2024-11	0.96	0.23	0.02	-0.11	0.04	0.02	0.14	1.28	0.20	2.05
`ogap`	1926-01	2024-11	-0.98	0.40	0.72	0.72	0.28	0.72	0.08	3.73	0.25	5.92
`wtexas`	1926-01	2024-11	-1.71	0.34	0.36	0.15	0.32	0.36	0.15	0.95	0.20	1.40
`sntm`	1975-07	2023-11	0.22	0.01	-0.41	-0.41	-0.67	-0.41	-0.10	-0.51	-0.49	-2.01
`ndrbl`	1958-02	2024-11	-1.5	0.40	-0.58	-0.84	-0.08	-0.58	-0.07	-0.35	0.09	0.45
`skvw`	1926-07	2024-11	0	0.00	-0.58	-0.64	-0.46	-0.58	-0.04	-0.56	-0.20	-2.88
`tail`	1926-07	2024-11	0.59	0.03	-0.16	-0.16	-0.29	-0.16	-0.14	-1.16	-0.31	-2.64
`fbm`	1936-06	2024-11	-0.17	0.01	-0.89	-0.89	-0.84	-0.89	0.05	0.40	0.07	0.60
`dtoy`	1926-01	2024-11	-0.06	0.00	-0.19	-0.19	-0.23	-0.19	0.04	0.83	0.21	3.34
`dtoat`	1926-01	2024-11	-0.75	0.15	0.15	0.15	0.11	0.15	0.06	1.05	0.25	3.41
`ygap`	1953-04	2024-11	0.54	0.06	-0.87	-1.35	-0.78	-0.87	0.10	0.34	0.18	0.69
`rdsp`	1926-07	2024-11	1.04	0.63	-1.08	-1.21	-1.62	-1.08	-0.04	-1.22	-0.05	-1.66
`rsvix`	1996-01	2023-08	0.62	0.39	-0.02	-0.02	-3.17	-0.02	-0.63	-0.90	-0.90	-1.25
`tchi`	1951-01	2024-11	1.49	0.37	0.14	-0.02	-0.23	0.14	0.58	2.48	0.62	2.98
`avgcor`	1926-03	2024-11	1.2	0.27	0.37	0.37	-0.02	0.37	0.08	0.73	0.40	3.21
`shtint`	1973-01	2024-11	-1.33	0.32	-0.07	-0.07	-0.79	-0.07	-0.14	-0.49	0.13	0.51
`disag`	1981-12	2024-11	-0.09	0.00	-0.48	-0.90	-0.51	-0.48	-0.35	-0.99	-0.55	-1.50
`ntis`	1926-12	2024-11	-1.54	0.48	-0.54	-0.54	-0.68	-0.54	-0.01	-0.25	0.22	3.40
`tby`	1920-01	2024-11	-2.01	0.34	0.61	0.45	-0.66	0.48	0.33	2.12	0.24	1.73
`dp`	1871-01	2024-11	0.52	0.06	-0.09	-0.59	-0.09	-0.20	0.22	1.12	0.15	0.74
`dy`	1871-02	2024-11	0.87	0.21	0.02	-0.56	0.02	-0.57	0.13	0.69	0.02	0.10
`ep`	1871-01	2024-11	1.6	0.15	0.10	0.07	0.10	0.09	0.26	1.41	0.19	0.98
`de`	1871-01	2024-11	0.08	0.00	0.01	-0.52	0.01	-0.09	0.59	2.60	0.62	2.92
`svar`	1885-02	2024-11	0.21	0.02	-0.06	-1.62	-0.06	-0.52	0.24	1.61	0.36	2.13
`lty`	1919-01	2024-11	-1.73	0.21	0.51	-0.07	-0.59	0.54	0.41	1.91	-0.22	-0.93
`ltr`	1926-01	2024-11	1.48	0.18	-0.47	-0.49	-0.30	-0.47	0.10	0.60	0.28	1.88
`tms`	1920-01	2024-11	1.24	0.13	0.26	0.24	-0.32	0.09	0.10	0.79	0.40	3.32
`dfy`	1919-01	2024-11	0.53	0.20	-0.35	-0.39	-0.29	-0.02	-0.04	-0.96	0.19	3.13
`dfr`	1926-01	2024-11	0.54	0.06	-0.25	-0.32	-0.21	-0.25	0.08	0.60	0.09	0.85
`infl`	1913-02	2024-11	-1.59	0.28	0.44	0.45	0.12	0.56	0.15	1.78	-0.07	-0.69
`bm`	1921-03	2024-11	1.17	0.54	-0.86	-1.15	-0.53	-1.38	0.03	0.59	0.10	1.75
`pce`	1953-12	2024-11	-3.73	1.57	1.09	0.92	1.18	1.09	0.64	2.89	0.82	3.84
`govik`	1947-03	2024-11	1.76	0.24	-0.34	-0.71	-0.35	-0.34	0.12	1.11	0.08	0.72
`crdstd`	1990-06	2024-11	-1.15	0.64	-1.04	-2.41	-0.28	-1.04	-0.25	-0.98	0.54	1.44
`ik`	1947-03	2024-11	-3.76	1.49	0.88	0.77	0.85	0.88	0.55	2.58	0.67	3.22
`cay`	1952-03	2024-11	0.24	0.01	-0.63	-0.87	-0.87	-0.63	0.19	0.80	0.32	1.36
`skew`	1951-12	2020-11	0.28	0.01	-0.29	-0.43	-0.58	-0.29	0.25	1.35	0.33	1.80
`accrul`	1965-12	2024-11	0.05	0.00	-0.78	-0.81	-0.95	-0.78	-0.02	-0.07	-0.22	-0.88
`cfacc`	1965-12	2024-11	-0.84	0.10	-0.70	-0.70	-0.76	-0.70	0.19	0.92	0.12	0.65
`gpce`	1947-12	2024-11	-3.39	1.26	0.17	0.16	0.67	0.17	0.24	1.50	0.32	2.00
`gip`	1926-12	2024-11	-0.34	0.03	0.06	0.06	-1.07	0.06	-0.01	-0.32	-0.07	-2.03
`house`	1929-12	2024-11	0.38	0.03	0.11	0.11	-0.12	0.11	0.08	1.12	0.05	0.74
`eqis`	1927-12	2024-11	-2.32	0.70	0.27	0.15	-0.15	0.27	0.21	1.78	0.26	2.40
Mean				0.32	-0.03	-0.62	-0.40	-0.06	0.10		0.13	
Median				0.21	-0.06	-0.25	-0.29	-0.12	0.08		0.19	

This table shows performance statistics of the 46 predictors in Goyal et al. (2024) forecasting monthly market returns in excess of the riskfree rate. Column CT R^2 reports out-of-sample R^2 imposing the Campbell and Thompson (2008) non-negativity constraint and starting the evaluation sample 20 years after the inception of the predictor or Jan 1926, whichever comes later. The estimation expanding-window regression coefficients and mean market excess returns, however, start at the earliest possible date. Column R^2 removes the non-negativity constraint. Column CT R^2 (10-yr) starts the R^2 evaluation 10 years after the inception of the predictors or Jan 1926, whichever comes later. Column CT R^2 (1946) starts the evaluation sample 20 years after the inception of the predictor or Jan 1946, whichever comes later. Column Alpha reports the alphas of the FA portfolios, and the next column t reports their associated t-statistics. Column Avg ret (0 ealt beta) report the average return of the zero ex-ante long-term beta FA portfolios, and the next column t reports their associated t-statistics. The t-statistics are computed with Newey-West standard errors following Goyal et al. (2024). The R^2s, alphas, and average returns are monthly and in percentage units.

Table 2: Joint estimation of forecast-agnostic portfolios

	(1) Panel	(2) Panel	(3) Time-series	(4) Time-series
	A: FA Portfolios			
Market Beta	-0.139***	-0.099***	-0.285***	-0.097***
	[-5.00]	[-7.92]	[-3.00]	[-7.63]
CAPM Alpha	0.161***	0.112***	0.360***	0.118***
	[4.57]	[4.40]	[2.87]	[4.48]
3-Factor Alpha	0.146***	0.110***	0.277**	0.113***
	[4.36]	[4.40]	[2.50]	[4.41]
	B: Zero ex-ante long-term beta FA Portfolios			
Market Beta	0.028	0.055***	-0.082	0.058***
	[1.15]	[5.12]	[-0.94]	[5.27]
CAPM Alpha	0.145***	0.099***	0.345***	0.104***
	[4.37]	[4.35]	[2.76]	[4.39]
3-Factor Alpha	0.133***	0.097***	0.280**	0.100***
	[4.25]	[4.31]	[2.54]	[4.29]
Avg return	0.165***	0.136***	0.288***	0.143***
	[5.97]	[6.18]	[2.80]	[6.22]
Start date	1926	1946	1926	1946
Sample size	35,395	33,670	1,187	947

In Panel A, the first two rows of Column (1) and (2) conduct the following predictor-monthly level panel regression: $r^{FA}_{i,t} = \alpha + \beta mkt_t + \epsilon_{i,t}$. Here, $r^{FA}_{i,t}$ is the FA portfolio return for predictor i in month t, and mkt_t is the stock market return in excess of the riskfree rate in month t. The 3rd row reports the constant term in $r^{FA}_{i,t} = \alpha + \beta_1 mkt_t + \beta_2 r^{vm}_{mkt,t} + \beta_3 r^{vm}_{i,t} + \epsilon_{i,t}$, where $r^{vm}_{mkt,t}$ is a volatility-managed market portfolio, created following the method in Moreira and Muir (2017), and $r^{vm}_{i,t}$ is the volatility managed FA portfolio predictor i. Column (1) starts the evaluation 20 years after the inception of the predictor or Jan 1926, whichever comes later. Column (2) uses Jan 1946 rather than Jan 1926. The first two rows of columns (3) and (4) conduct the monthly time-series regression: $avg\_r^{FA}_t = \alpha + \beta mkt_t + \epsilon_t$, in which $avg\_r^{FA}_t$ is the average of $r^{FA}_{i,t}$ across i for month t. The third row conducts $avg\_r^{FA}_t = \alpha + \beta_1 mkt_t + \beta_2 r^{vm}_{mkt} + \beta_3 avg\_r^{vm}_t + \epsilon_t$, in which $avg\_r^{vm}_i$ is the average of $r^{vm}_{i,t}$ across i in month t. In columns (3) and (4), the $r^{FA}_{i,t}$ and $r^{vm}_{i,t}$ values included in the calculation of $avg\_r^{FA}_t$ and $avg\_r^{vm}_t$ are the same as in columns (1) and (2), respectively. Panel B is similar to Panel A, except that it uses the zero ex-ante long-term beta FA portfolios. It further adds the fourth row computing the average returns of the portfolios. The t-statistics in columns (1)-(2) and (3)-(4) are computed with Driscoll-Kraay and Newey-West standard errors, respectively, following Goyal et al. (2024).

Table 3: Correlation among evaluation metrics

	abs(t)	in-sample R^2	CT R^2	CT R^2 (1946)	R^2	Alpha	Avg return
abs(t)	1.00						
in-sample R^2	0.79	1.00					
CT R^2	0.40	0.65	1.00				
CT R^2 (1946)	0.40	0.65	0.99	1.00			
R^2	0.25	0.27	0.46	0.46	1.00		
Alpha	0.46	0.24	0.25	0.22	0.18	1.00	
Avg return	0.46	0.32	0.15	0.13	0.21	0.73	1.00

This table shows the correlations among different statistics across the 46 predictors in Table 1.

Table 4: Portfolio choice and welfare gain

Utility (unit: avg return per month)	Risk aversion γ 1	2	3	4	5	6
	A: Sample starts in 1926					
(1) 100% Market	0.55%	0.40%	0.26%	0.11%	-0.03%	-0.18%
(2) Optimize Market + Rf	0.74%	0.37%	0.25%	0.19%	0.15%	0.12%
(3) Optimize Market + Agg FA + Rf	1.12%	0.56%	0.37%	0.28%	0.22%	0.19%
(3)-(2)	0.38%	0.19%	0.13%	0.10%	0.08%	0.06%
Avg equity weight in (3)	196%	98%	65%	49%	39%	33%
Avg FA weight in (3)	289%	144%	96%	72%	58%	48%
	B: Sample starts in 1946					
(1) 100% Market	0.57%	0.49%	0.40%	0.31%	0.22%	0.13%
(2) Optimize Market + Rf	0.94%	0.47%	0.31%	0.24%	0.19%	0.16%
(3) Optimize Market + Agg FA + Rf	1.29%	0.65%	0.43%	0.32%	0.26%	0.22%
(3)-(2)	0.35%	0.17%	0.12%	0.09%	0.07%	0.06%
Avg equity weight in (3)	203%	102%	68%	51%	41%	34%
Avg FA weight in (3)	307%	154%	102%	77%	61%	51%

This table shows the utility of a mean-variance investor, computed as $E[r_p]-\frac{\gamma}{2}Var[r_p]$. Here r_p is the return of her portfolio in excess of the risk-free rate, and $E[\cdot]$ and $Var[\cdot]$ compute its mean and variance over a given sample, which are 1926-2024 and 1946-2024 for Panel A and B, respectively. In row (1), the investor's portfolio is 100% stock market. In row (2), the stock weight equals $\overline{\mu}(mkt)_t/\gamma\overline{\sigma}(mkt)_t^2$, where $\overline{\mu}(mkt)_t$ and $\overline{\sigma}(mkt)_t^2$ are the expanding-window mean and variance of the stock market. Row (3) adds a additional position in the aggregate (equally across predictors) zero ex-ante long-term beta FA portfolios, where the weight is $\overline{\mu}(avg\_r^{zbFA})_t/\gamma\overline{\sigma}(avg\_r^{zbFA})_t^2$, where avg_r^{zbFA} is the time series of the returns of this aggregate zero ex-ante long-term beta FA portfolios. The next two rows report the average weight on the stock market and on the aggregate zero ex-ante long-term beta FA portfolios.

Table 5: Simulation results

Percentiles	No predictability 5	50	95	With predictability 5	50	95	Data —
	A: Random betas						
Avg abs(t)	0.627	0.805	0.988	1.029	1.344	1.796	1.068
Avg in-sample R^2	0.071	0.112	0.172	0.174	0.288	0.509	0.324
Avg R^2	-0.568	-0.382	-0.092	-0.677	-0.222	0.971	-0.620
Avg CT R^2	-0.515	-0.333	-0.046	-0.531	-0.095	0.941	-0.027
Avg alpha	-0.073	-0.004	0.056	0.026	0.112	0.207	0.097
	B: Estimated betas & conditional variance						
Avg abs(t)	0.629	0.813	1.033	1.252	1.665	2.220	1.068
Avg in-sample R^2	0.092	0.156	0.265	0.345	0.617	1.105	0.324
Avg R^2	-0.996	-0.586	-0.266	-1.558	-0.242	1.622	-0.620
Avg CT R^2	-0.758	-0.467	-0.157	-1.026	-0.012	1.550	-0.027
Avg alpha	-0.072	-0.006	0.055	0.053	0.139	0.245	0.097

This table shows the 5th, 50th, and 95th percentiles of in and out-of-sample statistics across 1000 simulations, as well as the corresponding values computed from historical data. The statistics are i) the average absolute value of the t-statistic, ii) the average in-sample R^2, iii) the average out-of-sample R^2, iv) the average Campbell and Thompson R^2, and v) the average forecast-agnostic alphas. Each simulation consists of 47 monthly time series representing stock market excess returns and 46 predictors with the same start and end months as the real data. In Panel A, the predictors follow AR(1) processes with a mean of zero and AR(1) coefficients estimated from the real predictor data. The innovations of the predictors are i.i.d. normal shocks with a volatility of 30 basis points multiplied by the square root of one minus the square of the AR(1) coefficients. In the first 3 columns with the header "no predictability," the simulated returns equal a constant of 58 basis points plus i.i.d. normal shocks with a mean of zero and a volatility of 4.78%. In the next 3 columns, the simulated returns equal the aforementioned two terms plus 46 additional terms, each being the product of one of the 46 predictors and a beta coefficient randomly drawn from a uniform distribution between -1 and 1. Panel B presents two similar sets of simulations, except that we estimate the beta coefficients and the predictors' volatility from the historical time series. The innovation terms in both the predictors and the returns have time-varying conditional variance estimated with a GARCH(1,1) model on the data.

Table 6: Simulation results with longer samples and stronger predictability

	A: Longer sample					
Sample length	1X	2X	5X	10X	20X	30X
Avg abs(t)	1.344	1.670	2.347	3.191	4.405	5.318
Avg in-sample R^2	0.288	0.213	0.162	0.144	0.135	0.131
Avg R^2	-0.222	-0.090	0.009	0.058	0.085	0.096
Avg CT R^2	-0.095	-0.025	0.034	0.072	0.092	0.101
Avg FA 2F alpha	0.0457	0.0325	0.0234	0.0189	0.0148	0.0125
Avg FD 2F alpha	-0.0194	-0.0080	-0.0016	-0.0001	0.0003	0.0004
	B: Stronger predictability with 10X sample length					
Predictability	1X	2X	3X	4X	5X	6X
Avg abs(t)	3.191	5.672	7.466	8.674	9.488	10.035
Avg in-sample R^2	0.144	0.475	0.874	1.245	1.558	1.806
Avg R^2	0.058	0.399	0.802	1.181	1.494	1.738
Avg CT R^2	0.072	0.406	0.738	1.008	1.203	1.339
Avg FA 2F alpha	0.0189	0.0156	0.0166	0.0230	0.0342	0.0475
Avg FD 2F alpha	-0.0001	0.0087	0.0220	0.0387	0.0568	0.0757

This table shows the median percentiles of 6 statistics across 1000 simulations. The 6 statistics are i) the average absolute value of Newey-West t-statistics, ii) the average in-sample R^2, iii) the average out-of-sample R^2, iv) the average out-of-sample R^2 after applying the positive-sign restriction from CT (2008), v) the average 2-factor alphas of the FA portfolios on top of the market and the FD portfolios, and vi) the average 2-factor FD alphas of the portfolios on top of the market and the FA portfolios. Each simulation consists of 47 monthly time series representing stock market excess returns and 46 predictors with the same start and end months as the real data. The predictors follow AR(1) processes with a mean of zero and AR(1) coefficients estimated from the real predictor data. The innovation of the predictors are iid normal shocks with a volatility of 30 basis points multiplied by the square root of one minus the square of the AR(1) coefficients. The simulated returns equal the aforementioned two terms plus 46 additional terms, each being the product of one of the 46 predictors and a beta coefficient randomly drawn from a uniform distribution between -1 and 1. Panel A reports results simulated with 1-30 times the length of the original data. Panel B reports results simulated with 10 times the data and 1-6 times the dispersion of the predictors.

Table 7: Pre- and post-sample behaviors of forecast agnostic alphas

	(1) all	(2) risk
pre-sample	0.315***	0.421***
	[2.97]	[3.05]
original sample	0.201***	0.191***
	[4.86]	[4.19]
post-sample	0.007	0.041
	[0.16]	[0.92]
Sample size	37,325	31,355
p(pre = original)	0.251	0.072
p(post = original)	0.002	0.022
p(pre = post)	0.007	0.009

This table shows results of regressions of all FA portfolio returns on 3 dummy variables indicating whether the observation is before, within, or after the authors' original sample, and their interaction with the contemporaneous market returns. The coefficients on the 3 dummy variables extract the FA alphas pre-, in-, and post-sample. Evaluation begins 20 years after the inception of the predictors, and possibly before 1926. Data are monthly from 1891 to 2024. t-statistics computed with standard errors clustered at the monthly level are reported in square brackets.

Table 8: R^2 in forecasting monthly stock market excess returns–truncating extreme values

	Constraint			
	None	Negative	Positive	Both
vp	4.72	4.72	1.76	1.76
impvar	0.49	0.49	0.49	0.49
vrp	-19.51	-1.00	-18.73	-0.22
lzrt	-0.11	0.02	-0.03	0.10
ogap	0.72	0.72	0.71	0.71
wtexas	0.15	0.36	0.28	0.49
sntm	-0.41	-0.41	-0.41	-0.41
ndrbl	-0.84	-0.58	-0.59	-0.32
skvw	-0.64	-0.58	-0.54	-0.48
tail	-0.16	-0.16	0.01	0.01
fbm	-0.89	-0.89	-0.58	-0.58
dtoy	-0.19	-0.19	-0.13	-0.13
dtoat	0.15	0.15	0.15	0.15
ygap	-1.35	-0.87	-0.96	-0.48
rdsp	-1.21	-1.08	-1.13	-1.00
rsvix	-0.02	-0.02	-0.02	-0.02
tchi	-0.02	0.14	-0.02	0.14
avgcor	0.37	0.37	0.41	0.41
shtint	-0.07	-0.07	-0.05	-0.05
disag	-0.90	-0.48	-0.90	-0.48
ntis	-0.54	-0.54	0.16	0.16
tby	0.45	0.61	0.53	0.69
dp	-0.59	-0.09	-0.59	-0.08
dy	-0.56	0.02	-0.91	-0.32
ep	0.07	0.10	-0.01	0.03
de	-0.52	0.01	-0.52	0.01
svar	-1.62	-0.06	-1.47	0.09
lty	-0.07	0.51	0.15	0.73
ltr	-0.49	-0.47	-0.13	-0.11
tms	0.24	0.26	0.66	0.68
dfy	-0.39	-0.35	-0.37	-0.33
dfr	-0.32	-0.25	-0.16	-0.09
infl	0.45	0.44	0.46	0.45
bm	-1.15	-0.86	-0.64	-0.35
pce	0.92	1.09	0.75	0.92
govik	-0.71	-0.34	-0.71	-0.34
crdstd	-2.41	-1.04	-2.30	-0.93
ik	0.77	0.88	0.76	0.87
cay	-0.87	-0.63	-0.58	-0.33
skew	-0.43	-0.29	-0.46	-0.33
accrul	-0.81	-0.78	-0.78	-0.75
cfacc	-0.70	-0.70	-0.69	-0.69
gpce	0.16	0.17	0.30	0.31
gip	0.06	0.06	0.08	0.08
house	0.11	0.11	0.11	0.11
eqis	0.15	0.27	0.55	0.67
Mean	-0.62	-0.03	-0.57	0.03
Median	-0.25	-0.06	-0.09	-0.01

This table shows performance statistics of the 46 predictors in Goyal et al. (2024) forecasting monthly market returns in excess of the riskfree rate. Column “None” reports out-of-sample R^2 on forecasts computed with freely estimated expanding-window coefficients and is equivalent to out-of-sample R^2 in Table 1. Column “Negative” imposes the Campbell and Thompson (2008) non-negativity constraint and is equivalent to CT R^2 in Table 1. Column “Positive” truncates from above forecasts exceeding 2 times the expanding-window mean of the market excess returns. Column “Both” imposes both constraints from the previous two columns. The R^2s evaluation starts 20 years after the inception of the predictor or Jan 1926, whichever comes later. The estimation expanding-window regression coefficients and mean market excess returns, however, start at the earliest possible date. All R^2s are in percentage units.

Table 9: R^2 in forecasting monthly stock market excess returns–imposing economic constraints

	Constraint				
	None	Negative	Positive	Both	Economic
`vp`	4.72	4.72	1.76	1.76	1.19
`impvar`	0.49	0.49	0.49	0.49	0.94
`vrp`	-19.51	-1.00	-18.73	-0.22	0.47
`rsvix`	-0.02	-0.02	-0.02	-0.02	0.26
`dp`	-0.59	-0.09	-0.59	-0.08	0.44
`dy`	-0.56	0.02	-0.91	-0.32	0.45
`ep`	0.07	0.10	-0.01	0.03	0.32
`bm`	-1.15	-0.86	-0.64	-0.35	0.24
Mean	-2.07	0.42	-2.33	0.16	0.54
Median	-0.29	0.00	-0.30	-0.05	0.45

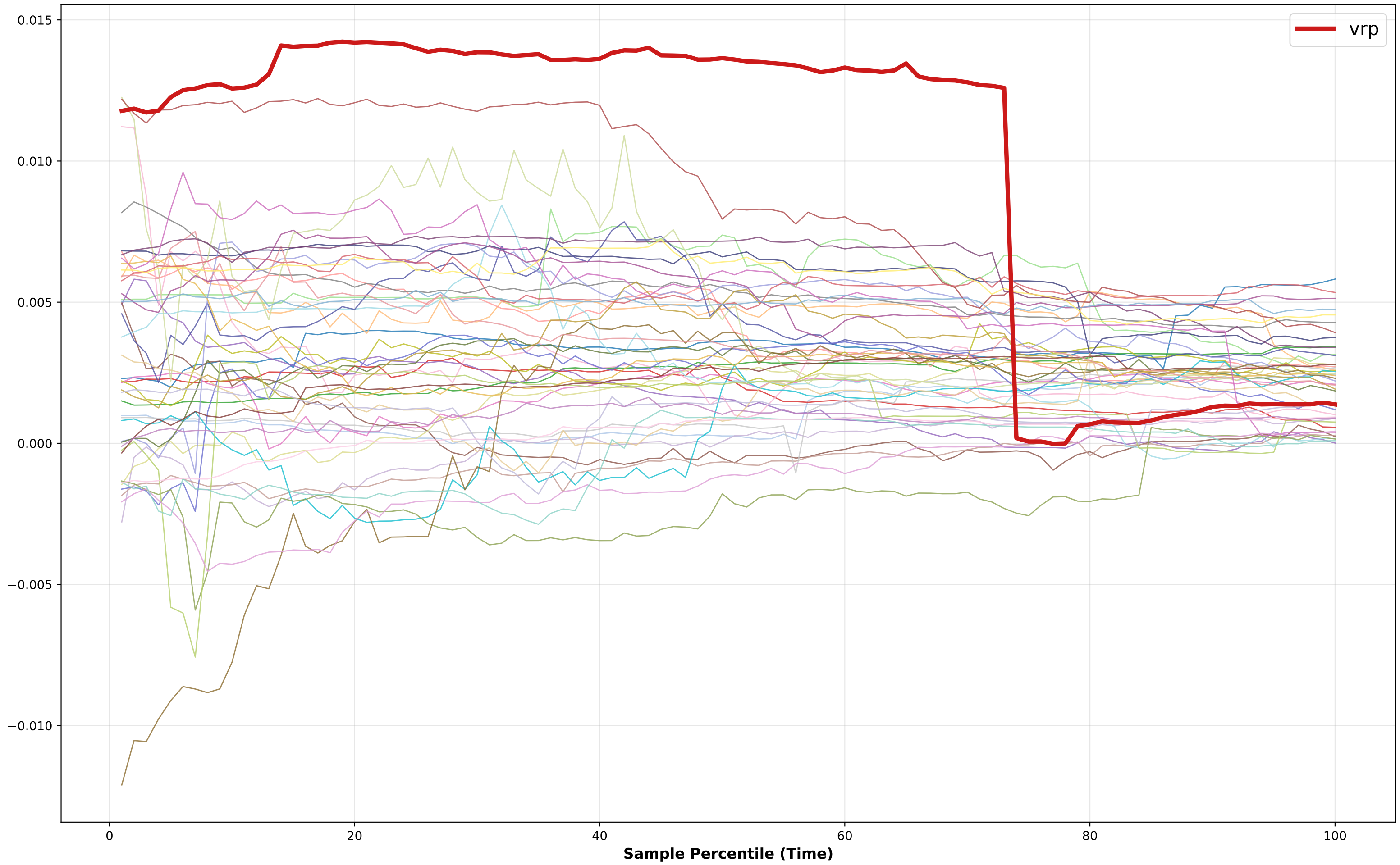
This table shows performance statistics of the 8 predictors in Goyal et al. (2024) forecasting monthly market returns in excess of the riskfree rate. Column "None" reports out-of-sample R^2 on forecasts computed with freely estimated expanding-window coefficients. Column "Negative" imposes the Campbell and Thompson (2008) non-negativity constraint. Column "Positive" truncates from above forecasts exceeding 2 times the expanding-window mean of the market excess returns. Column "Both" imposes both constraints from the previous two columns. Column "Economic" imposes economic constraints specifically applicable to the signal. The R^2s evaluation starts 20 years after the inception of the predictor or Jan 1926, whichever comes later. The estimation expanding-window regression coefficients and mean market excess returns, however, start at the earliest possible date. All R^2s are in percentage units.

Table 10: Out-of-sample R^2s: improved power with true betas

	With predictability		
Percentiles	5	50	95
	A: Random betas		
Avg R^2	-0.677	-0.222	0.971
Avg CT R^2	-0.531	-0.095	0.941
Avg R^2 with true beta	0.013	0.111	0.234
	B: Estimated betas & conditional variance		
Avg R^2	-1.558	-0.242	1.622
Avg CT R^2	-1.026	-0.012	1.550
Avg R^2 with true beta	0.012	0.260	0.569

This table shows the 5th, 50th, and 95th percentiles of out-of-sample R^2 across 1000 simulations. The statistics are i) the average out-of-sample R^2, ii) the average Campbell and Thompson R^2, and iii) the average out-of-sample R^2 with true betas. Each simulation consists of 47 monthly time series representing stock market excess returns and 46 predictors with the same start and end months as the real data. In Panel A, the predictors follow AR(1) processes with a mean of zero and AR(1) coefficients estimated from the real predictor data. The innovations of the predictors are i.i.d. normal shocks with a volatility of 30 basis points multiplied by the square root of one minus the square of the AR(1) coefficients. The simulated returns equal the aforementioned two terms plus 46 additional terms, each being the product of one of the 46 predictors and a beta coefficient randomly drawn from a uniform distribution between -1 and 1. Panel B presents similar simulations, except that we estimate the beta coefficients and the predictors' volatility from the historical time series. The innovation terms in both the predictors and the returns have time-varying conditional variance estimated with a GARCH(1,1) model on the data.

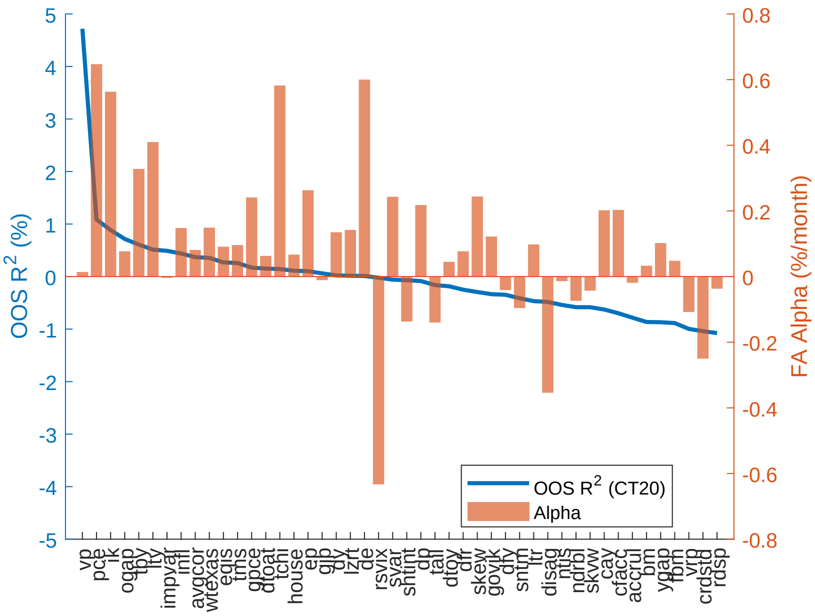
Figure 1: Expanding window coefficients for the 46 predictors



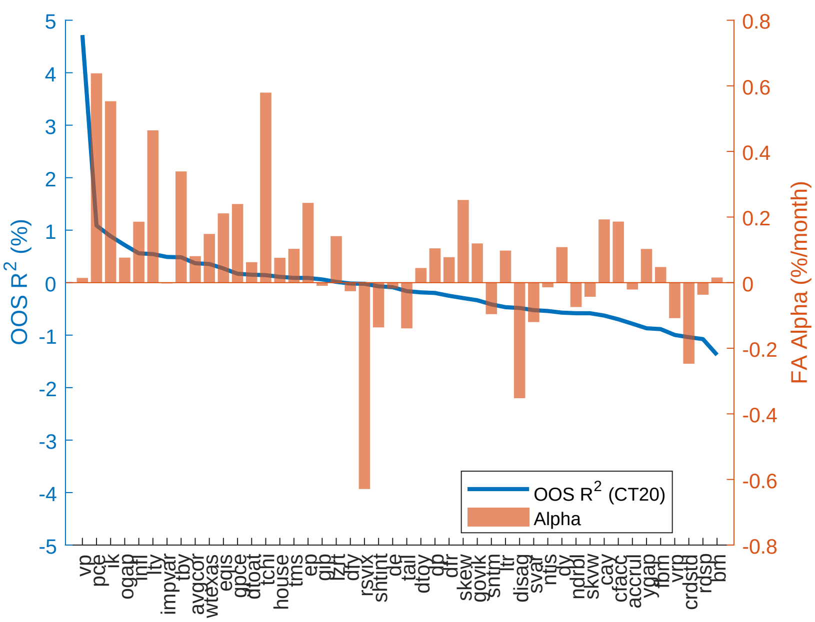
This figure plots the expanding-window regression coefficients of the aggregate stock market excess returns on the 46 predictors in Table 1. The x-axis is the percentile of the sample length of the predictors, with 1 corresponding to the first percentile of the sample, and 100 corresponding to the end of the sample. The coefficients for `vrp` are highlighted.

Figure 2: out-of-sample R^2s and FA alphas across predictors

(a) Earliest evaluation in 1926



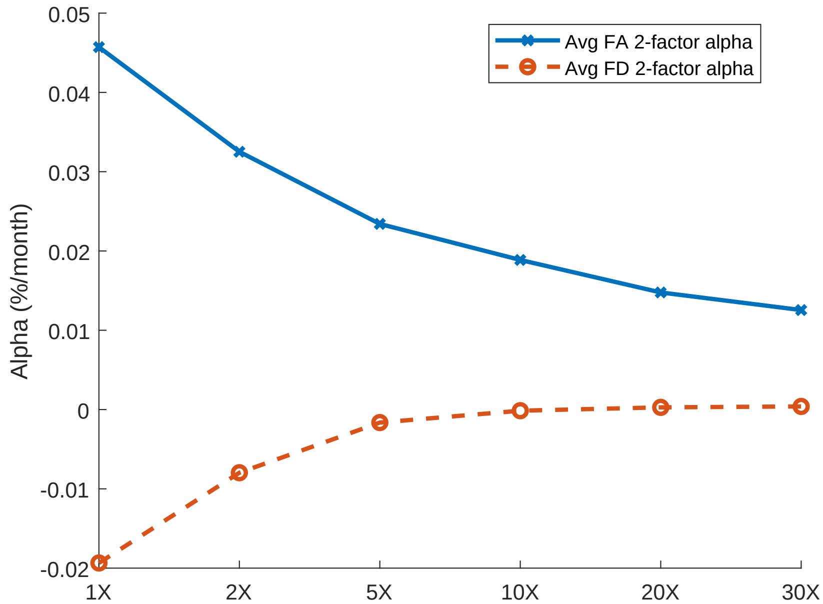
(b) Earliest evaluation in 1946



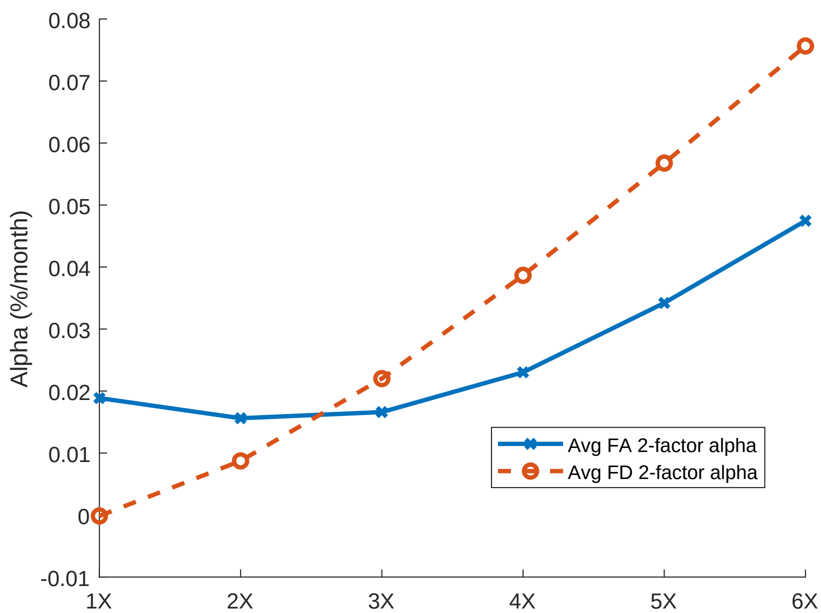
This figure plots out-of-sample R^2s with the Campbell Thompson non-negativity constraint and forecast-agnostic alphas across the 46 predictors in Table 1. The out-of-sample R^2s declines monotonically from left to right. The top panel starts the evaluation of alpha and OOS R^2 in 1926 or 20 years after the inception of the predictor, whichever is later. The bottom panel starts the evaluation of alpha and OOS R^2 in 1946 or 20 years after the inception of the predictor, whichever is later.

Figure 3: Average FA and FD two-factor alphas

(a) Longer samples

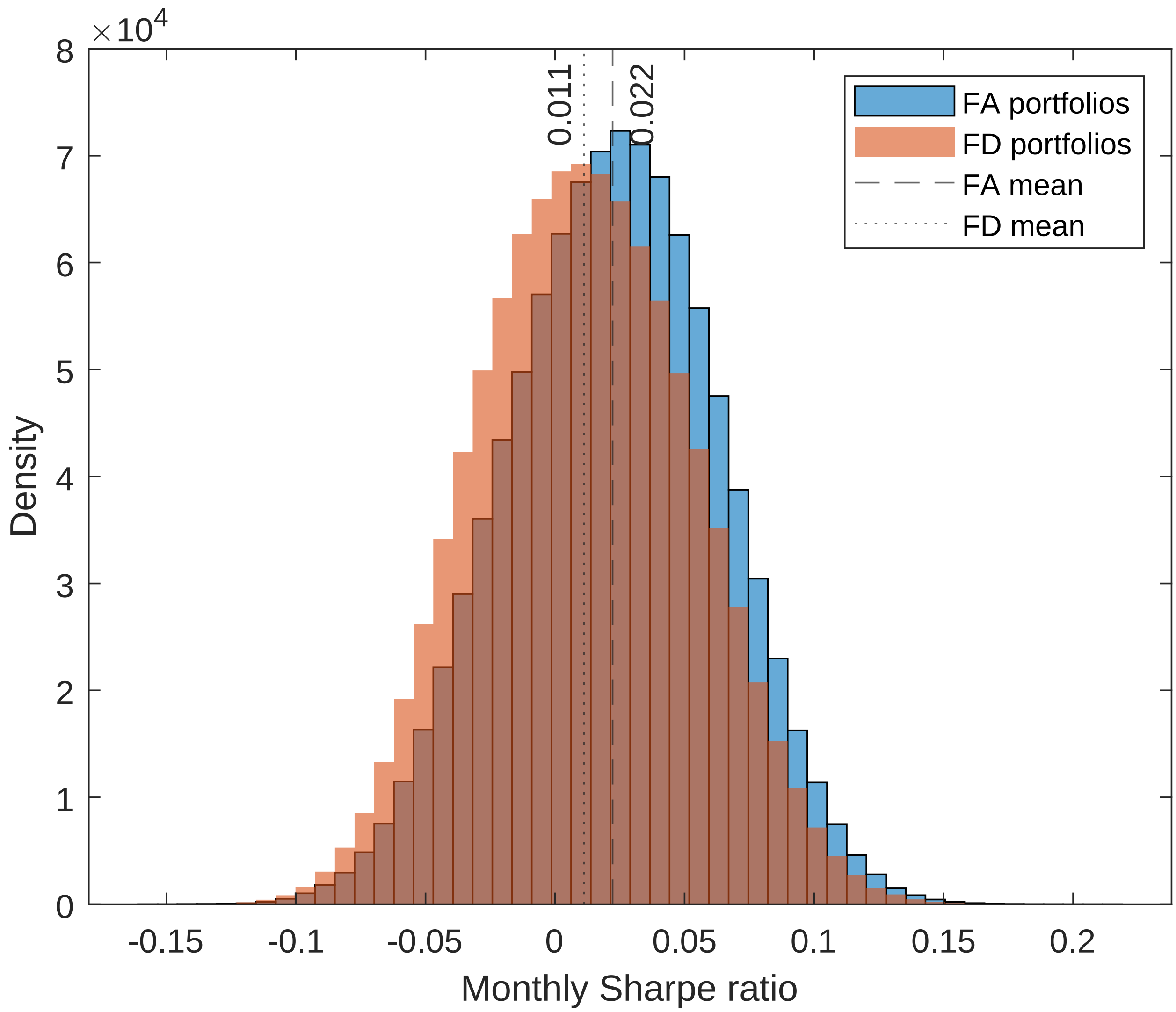


(b) Stronger predictability with 10X sample length



This figure plots the medians of the average (across predictors) FA and FD alphas controlling for each other and the market excess returns across 1000 simulations. Panel (a) conducts the simulations with increasingly longer samples, whereas Panel (b) conducts simulations with 10 times the sample length and increasingly stronger predictability than observed.

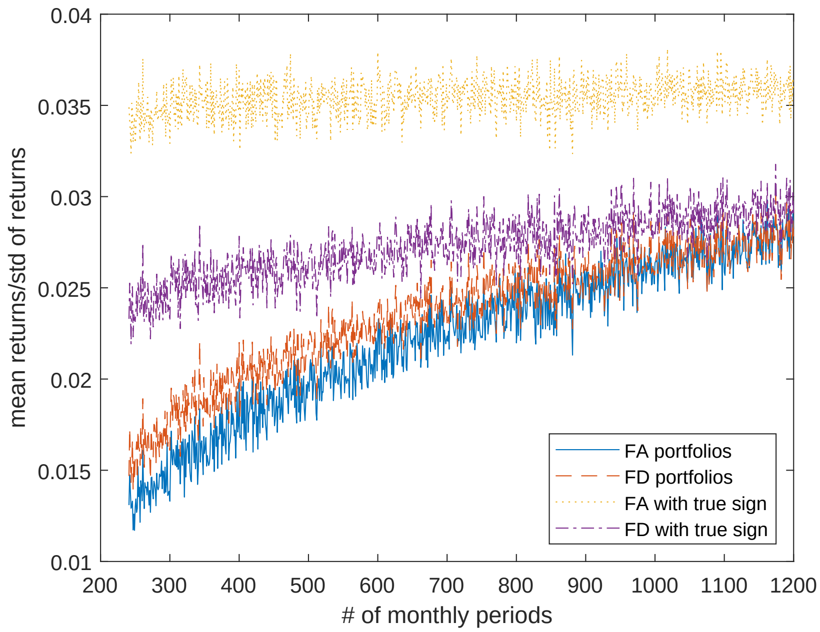
Figure 4: Sharpe ratios of FA and FD portfolios



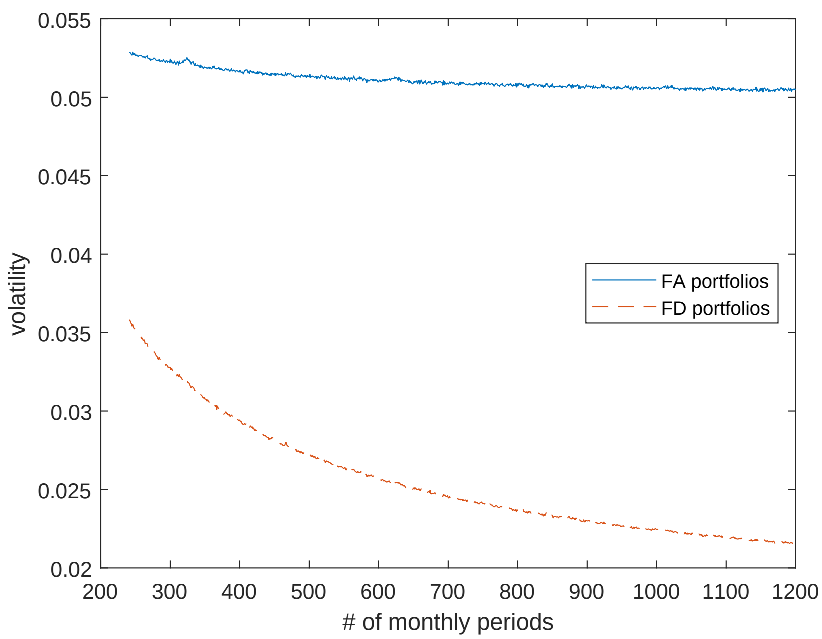
This figure is a histogram of Sharpe ratios of FA and FD portfolios across a million simulations, each of which contains a time-series of 1200 observations. The simulations are of the statistical setup in Section 5 with $\sigma = 0.048$, $\sigma_x = 0.0025$, $\mu = 0.0055$, $\mu_x = 0$, $\rho = 0.9$, and $b_x = 0.7$.

Figure 5: Asset attractiveness and risk allocation over time

(a) Period-by-period Sharpe ratios



(b) Period by period volatility



The top panel of this figure plots the period-by-period Sharpe ratios for four portfolios in 1 million simulations, each of which contains a time-series of 1200 observations. The simulations are of the statistical setup in Section 5 with $\sigma = 0.048$, $\sigma_x = 0.0025$, $\mu = 0.0055$, $\mu_x = 0$, $\rho = 0.9$, and $b_x = 0.7$. For a given period, we first compute the average of 1 million portfolio returns, then compute the standard deviation of these 1 million returns, and lastly compute their ratio. The blue solid line is for the FA portfolios, the orange dashed line the FD portfolios, the yellow dotted line the hypothetical FA portfolio using the true sign of beta, and the purple dotted-dashed line the hypothetical FD portfolio using the true sign. The bottom panel are the standard deviation of the FA and FD portfolio returns across the simulations.

Appendix

A Data

We obtain data on stock returns and on predictor variables from Amit Goyal's website at `https://sites.google.com/view/agoyal145`, and specifically the 2024 vintage. For the purpose of pre-sample testing, we extend the stock market return series back to 1871 using data from Global Financial Data (GFD), which are ultimately based on Standard & Poor's. These return data are available at a monthly frequency. We construct predictor variables using data that would have been available to the investor at that time. This includes `sntm` and `fbm` which embed an estimation in their construction.

B FA alphas with alternative volatility targets

Table B1 calculates panel and time-series alphas of the FA portfolios as in Columns (1) and (3) of Table 2, except they have different volatility targets from 5% per month. The t-statistics remain stable across columns, demonstrating that the FA alphas are highly statistically significant regardless of the specific choice of the volatility target.

Table B1: FA alphas with alternative volatility targets

	2.5%	5%	10%	15%	20%	25%	50%
	A: Panel alphas						
mkt	-0.082***	-0.139***	-0.194***	-0.225***	-0.249***	-0.268***	-0.309***
	[-4.32]	[-5.00]	[-5.44]	[-5.92]	[-6.29]	[-6.45]	[-6.72]
Alpha	0.088***	0.161***	0.232***	0.262***	0.281***	0.295***	0.326***
	[4.51]	[4.57]	[4.79]	[4.78]	[4.73]	[4.63]	[4.32]
Sample size	35,395	35,395	35,395	35,395	35,395	35,395	35,395
	B: Time series alphas						
mkt	-0.181***	-0.285***	-0.360***	-0.374***	-0.369***	-0.371***	-0.400***
	[-2.65]	[-3.00]	[-3.10]	[-3.05]	[-2.88]	[-2.79]	[-2.89]
Alpha	0.203***	0.360***	0.482***	0.527***	0.546***	0.563***	0.622***
	[2.86]	[2.87]	[3.14]	[3.20]	[3.12]	[3.05]	[3.06]
Sample size	1,187	1,187	1,187	1,187	1,187	1,187	1,187

Panel A conducts the following predictor-monthly level panel regression: $r_{i,t}^{FA} = \alpha + \beta r_t + \epsilon_{i,t}$. Here, $r_{i,t}^{FA}$ is the FA portfolio return for predictor i in month t, and r_t is the stock market return in excess of the riskfree rate in month t. Panel B conducts the monthly time-series regression: $avg\_r_t^{FA} = \alpha + \beta r_t + \epsilon_t$, in which $avg\_r_t^{FA}$ is the average of $r_{i,t}^{FA}$ across i for month t. In Panel B, the $r_{i,t}^{FA}$ values included in the calculation of $avg\_r_t^{FA}$ are the same as in Panel A. All regressions start the evaluation 20 years after the inception of the predictor or Jan 1926, whichever comes later. The FA alphas are constructed with different volatility targets, indicated in the column headers. The t-statistics in Panels A and B are computed with Driscoll-Kraay and Newey-West standard errors, respectively, following Goyal et al. (2024).

C Proof of Proposition 1 and 2

Assume the following data-generating process:

$$r_{t+1} = \mu + b_x(x_t - \mu_x) + \epsilon_{t+1}, \quad E[\epsilon_{t+1}] = 0, \quad \mathrm{Var}(\epsilon_{t+1}) = \sigma^2,$$

$$x_{t+1} = \mu_x + \rho(x_t - \mu_x) + u_{t+1}, \quad E[u_{t+1}] = 0, \quad \mathrm{Var}(u_{t+1}) = \sigma_x^2,$$

with u and ϵ uncorrelated. Define

$$\bar{x}_{t-1} = \frac{1}{t}\sum_{s=0}^{t-1} x_s, \quad \bar{r}_t = \frac{1}{t}\sum_{s=1}^{t} r_s, \quad \hat{b}_{xt} = \frac{\sum_{s=0}^{t-1}(x_s - \bar{x}_{t-1})r_{s+1}}{\sum_{s=0}^{t-1}(x_s - \bar{x}_{t-1})^2}, \quad \hat{a}_t = \bar{r}_t - \hat{b}_{xt}\bar{x}_{t-1}, \quad V = \frac{\sigma_x^2}{1-\rho^2}.$$

Proposition 1 Consider the MSE of the expanding-window mean return $\bar{r}_t$. Note that $\mathbb{E}_t[r_{t+1}] = \mu + b_x(x_t - \mu_x)$. First consider the benchmark SSE. Conditionally:

$$\mathbb{E}_t\left[(r_{t+1} - \bar{r}_t)^2\right] = \sigma^2 + \left(\mu + b_x(x_t - \mu_x) - \bar{r}_t\right)^2.$$

So, unconditionally:

$$\mathbb{E}\left[(r_{t+1} - \bar{r}_t)^2\right] = \sigma^2 + \mathbb{E}\left[(\mu + b_x(x_t - \mu_x) - \bar{r}_t)^2\right] = \sigma^2 + b_x^2\,\mathrm{Var}(x_t) + \mathrm{Var}(\bar{r}_t) - 2b_x\,\mathrm{Cov}(x_t, \bar{r}_t).$$

Note that:

$$\mathrm{Var}(x_t) = \frac{\sigma_x^2}{1-\rho^2}, \qquad \mathrm{Cov}(x_s, x_k) = \frac{\sigma_x^2}{1-\rho^2}\rho^{|s-k|}.$$

$$\mathrm{Var}(r_s) = b_x^2\frac{\sigma_x^2}{1-\rho^2} + \sigma^2, \qquad \mathrm{Cov}(r_s, r_k) = b_x^2\frac{\sigma_x^2}{1-\rho^2}\rho^{|s-k|}\ (s \neq k).$$

Therefore:

$$\mathrm{Var}(\bar{r}_t) = \frac{1}{t^2}\left[t\left(b_x^2\frac{\sigma_x^2}{1-\rho^2} + \sigma^2\right) + 2\,b_x^2\frac{\sigma_x^2}{1-\rho^2}\sum_{h=1}^{t-1}(t-h)\rho^h\right] = \frac{1}{t}\left[\sigma^2 + b_x^2\frac{\sigma_x^2}{1-\rho^2}(1 + \frac{2\rho}{1-\rho})\right] + O\big(\frac{1}{t^2}\big),$$

because

$$\sum_{h=1}^{t-1}(t-h)\rho^h = t\frac{\rho(1-\rho^{t-1})}{1-\rho} - \frac{\rho\big(1 - t\rho^{t-1} + (t-1)\rho^t\big)}{(1-\rho)^2} = \frac{t\rho}{1-\rho} + O(1).$$

Furthermore,

$$\mathrm{Cov}(x_t, \bar{r}_t) = \frac{b_x}{t} \sum_{s=1}^{t} \mathrm{Cov}(x_t, x_{s-1}) = \frac{b_x}{t} \cdot \frac{\sigma_x^2}{1-\rho^2} \cdot \frac{\rho(1-\rho^t)}{1-\rho} = \frac{b_x}{t} \cdot \frac{\sigma_x^2}{1-\rho^2} \cdot \frac{\rho}{1-\rho} + O\big(\frac{1}{t^2}\big).$$

Consequently:

$$\boxed{\begin{aligned}\mathbb{E}\big[(r_{t+1} - \bar{r}_t)^2\big] &= \sigma^2 + b_x^2 \,\mathrm{Var}(x_t) + \mathrm{Var}(\bar{r}_t) - 2b_x \,\mathrm{Cov}(x_t, \bar{r}_t) \\ &= \sigma^2 + b_x^2 \frac{\sigma_x^2}{1-\rho^2} + \frac{1}{t}\left(\sigma^2 + b_x^2 \frac{\sigma_x^2}{1-\rho^2}\right) + O\big(\frac{1}{t^2}\big).\end{aligned}}$$

Now consider the forecasting SSE. Unconditionally:

$$\mathbb{E}_t\Big[(r_{t+1} - (x_t - \bar{x}_{t-1})\hat{b}_{xt} - \bar{r}_t)^2\Big] = \sigma^2 + \Big(\mu + b_x(x_t - \mu_x) - (x_t - \bar{x}_{t-1})\hat{b}_{xt} - \bar{r}_t\Big)^2.$$

Conditionally:

$$\mathbb{E}\Big[(r_{t+1} - (x_t - \bar{x}_{t-1})\hat{b}_{xt} - \bar{r}_t)^2\Big] = \sigma^2 + \mathbb{E}\Big[\Big(\mu + b_x(x_t - \mu_x) - (x_t - \bar{x}_{t-1})\hat{b}_{xt} - \bar{r}_t\Big)^2\Big].$$

Define:

$$\Delta_t \equiv \mu + b_x(x_t - \mu_x) - (x_t - \bar{x}_{t-1})\hat{b}_{xt} - \bar{r}_t = b_x(\bar{x}_{t-1} - \mu_x) - (\bar{r}_t - \mu) + (x_t - \bar{x}_{t-1})(b_x - \hat{b}_{xt}).$$

Then:

$$\begin{aligned}\mathbb{E}\big[\Delta_t^2\big] &= b_x^2 \,\mathrm{Var}(\bar{x}_{t-1}) + \mathrm{Var}(\bar{r}_t) - 2b_x \,\mathrm{Cov}(\bar{x}_{t-1}, \bar{r}_t) \\ &\quad + \mathbb{E}\Big[(x_t - \bar{x}_{t-1})^2(b_x - \hat{b}_{xt})^2\Big] + 2b_x \,\mathbb{E}\Big[(\bar{x}_{t-1} - \mu_x)(x_t - \bar{x}_{t-1})(b_x - \hat{b}_{xt})\Big] \\ &\quad - 2\,\mathbb{E}\Big[(\bar{r}_t - \mu)(x_t - \bar{x}_{t-1})(b_x - \hat{b}_{xt})\Big].\end{aligned}$$

Note that:

$$\mathrm{Cov}(\bar{x}_{t-1}, \bar{r}_t) = \frac{b_x}{t^2} \cdot \frac{\sigma_x^2}{1-\rho^2} \sum_{j=0}^{t-1} \sum_{s=1}^{t} \rho^{|j-(s-1)|} = \frac{b_x}{t} \cdot V \cdot (1 + \frac{2\rho}{1-\rho}) + O\big(\frac{1}{t^2}\big) \tag{C1}$$

Because:

$$\sum_{j=0}^{t-1} \sum_{s=1}^{t} \rho^{|j-(s-1)|} = \sum_{j=0}^{t-1} \sum_{k=0}^{t-1} \rho^{|j-k|} = t \;+\; 2\sum_{h=1}^{t-1} (t-h)\,\rho^h,$$

$$\sum_{h=1}^{t-1}(t-h)\,\rho^h = t\sum_{h=1}^{t-1}\rho^h \;-\; \sum_{h=1}^{t-1} h\,\rho^h = t\,\frac{\rho(1-\rho^{t-1})}{1-\rho} \;-\; \frac{\rho\big(1-t\rho^{t-1}+(t-1)\rho^t\big)}{(1-\rho)^2}, \text{ and}$$

$$\sum_{j=0}^{t-1}\sum_{s=1}^{t}\rho^{|j-(s-1)|} = t \,+\, 2\left[t\,\frac{\rho(1-\rho^{t-1})}{1-\rho} \;-\; \frac{\rho\big(1-t\rho^{t-1}+(t-1)\rho^t\big)}{(1-\rho)^2}\right] = t(1+\frac{2\rho}{1-\rho})+O(1).$$

Furthermore,

$$\operatorname{Var}(\bar{x}_{t-1}) = \frac{V}{t^2}\left[t+2\sum_{h=1}^{t-1}(t-h)\rho^h\right] = \frac{V}{t}\cdot(1+\frac{2\rho}{1-\rho})+O\big(\frac{1}{t^2}\big) \tag{C2}$$

because

$$\operatorname{Cov}(x_t,\bar{x}_{t-1}) = \frac{V}{t}\cdot\frac{\rho(1-\rho^t)}{1-\rho}.$$

Note that:

$$\mathbb{E}\big[(x_t-\bar{x}_{t-1})^2\big] = V+\operatorname{Var}(\bar{x}_{t-1})-2\operatorname{Cov}(x_t,\bar{x}_{t-1}) = V+\frac{V}{t^2}\left[t+2\sum_{h=1}^{t-1}(t-h)\rho^h\right]-\frac{2V}{t}\cdot\frac{\rho(1-\rho^t)}{1-\rho}.$$

$$\mathbb{E}[S_{xx}] = \sum_{s=0}^{t-1}\mathbb{E}\big[(x_s-\bar{x}_{t-1})^2\big] = tV - t\operatorname{Var}(\bar{x}_{t-1}) = tV - \frac{V}{t}\left[t+2\sum_{h=1}^{t-1}(t-h)\rho^h\right].$$

Therefore:

$$\mathbb{E}\Big[(x_t-\bar{x}_{t-1})^2(b_x-\hat{b}_{xt})^2\Big] = \sigma^2\cdot\frac{V+\dfrac{V}{t^2}\big[t+2\sum_{h=1}^{t-1}(t-h)\rho^h\big]-\dfrac{2V}{t}\cdot\dfrac{\rho(1-\rho^t)}{1-\rho}}{tV-\dfrac{V}{t}\big[t+2\sum_{h=1}^{t-1}(t-h)\rho^h\big]} = \frac{\sigma^2}{t}+O\big(\frac{1}{t^2}\big). \tag{C3}$$

Note further that:

$$\hat{b}_{xt} = \frac{\sum_{s=0}^{t-1}(x_s-\bar{x}_{t-1})r_{s+1}}{\sum_{s=0}^{t-1}(x_s-\bar{x}_{t-1})^2} = b_x + \frac{\sum_{s=0}^{t-1}(x_s-\bar{x}_{t-1})\,\epsilon_{s+1}}{\sum_{s=0}^{t-1}(x_s-\bar{x}_{t-1})^2},$$

$$b_x-\hat{b}_{xt} = -\frac{\sum_{s=0}^{t-1}(x_s-\bar{x}_{t-1})\,\epsilon_{s+1}}{\sum_{s=0}^{t-1}(x_s-\bar{x}_{t-1})^2}.$$

$$\mathbb{E}\Big[b_x-\hat{b}_{xt}\,\Big|\,\{x_s\}_{s=0}^{t-1}\Big] = 0, \quad \text{since} \quad \mathbb{E}[\epsilon_{s+1}]=0 \text{ and } \epsilon \text{ is independent of } x.$$

It then follows that:

$$\mathbb{E}\Big[(\bar{x}_{t-1}-\mu_x)(x_t-\bar{x}_{t-1})(b_x-\hat{b}_{xt})\Big] = \mathbb{E}\Big[(\bar{x}_{t-1}-\mu_x)(x_t-\bar{x}_{t-1})\,\mathbb{E}\Big[b_x-\hat{b}_{xt}\,\Big|\,\{x_s\}_{s=0}^{t-1}\Big]\Big] = 0. \tag{C4}$$

$$\hat{b}_{xt} = b_x + \frac{\sum_{s=0}^{t-1}(x_s - \bar{x}_{t-1})\,\epsilon_{s+1}}{\sum_{s=0}^{t-1}(x_s - \bar{x}_{t-1})^2}, \quad \Rightarrow \quad b_x - \hat{b}_{xt} = -\sum_{s=0}^{t-1} w_s\,\epsilon_{s+1},$$

$$w_s \equiv \frac{x_s - \bar{x}_{t-1}}{S_{xx}}, \qquad S_{xx} \equiv \sum_{s=0}^{t-1}(x_s - \bar{x}_{t-1})^2.$$

$$\bar{r}_t - \mu = \frac{1}{t}\sum_{j=1}^{t}\big(b_x(x_{j-1} - \mu_x) + \epsilon_j\big) = b_x\,\overline{(x_{j-1} - \mu_x)} \;+\; \bar{\epsilon}_t, \qquad \bar{\epsilon}_t \equiv \frac{1}{t}\sum_{j=1}^{t}\epsilon_j.$$

$$\mathbb{E}\Big[(\bar{r}_t - \mu)(x_t - \bar{x}_{t-1})(b_x - \hat{b}_{xt})\;\Big|\;\{x_s\}_{s=0}^{t-1}\Big] = (x_t - \bar{x}_{t-1})\,\mathbb{E}\left[\bar{\epsilon}_t \cdot \left(-\sum_{s=0}^{t-1} w_s\,\epsilon_{s+1}\right)\;\Big|\;\{x_s\}\right].$$

$$\mathbb{E}\left[\bar{\epsilon}_t \cdot \left(-\sum_{s=0}^{t-1} w_s\,\epsilon_{s+1}\right)\;\Big|\;\{x_s\}\right] = -\frac{1}{t}\sum_{j=1}^{t}\sum_{s=0}^{t-1} w_s\,\mathbb{E}[\epsilon_j\epsilon_{s+1}] = -\frac{\sigma^2}{t}\sum_{s=0}^{t-1} w_s,$$

since $\mathbb{E}[\epsilon_j\epsilon_{s+1}] = \sigma^2$ only when $j = s+1$, and 0 otherwise.

$$\sum_{s=0}^{t-1} w_s = \frac{\sum_{s=0}^{t-1}(x_s - \bar{x}_{t-1})}{S_{xx}} = 0.$$

$$\mathbb{E}\Big[(\bar{r}_t - \mu)(x_t - \bar{x}_{t-1})(b_x - \hat{b}_{xt})\Big] = 0. \tag{C5}$$

Finally, combining equations C1 to C5:

$$\boxed{\mathbb{E}\Big[(r_{t+1} - (x_t - \bar{x}_{t-1})\hat{b}_{xt} - \bar{r}_t)^2\Big] = \sigma^2 + \frac{2\sigma^2}{t} + O\big(\frac{1}{t^2}\big)}$$

The out-of-sample R^2, approximately the difference in SSE divided by σ^2, becomes:

$$\frac{b_x^2 V}{\sigma^2}(1 + \frac{1}{t}) - \frac{1}{t}$$

Proposition 2 Consider the expected alpha of the FD portfolio:

$$\mathbb{E}\Big[\hat{b}_{xt}(x_t - \bar{x}_{t-1})(r_{t+1} - \mu)\Big] = b_x\,\mathbb{E}\Big[\hat{b}_{xt}(x_t - \bar{x}_{t-1})(x_t - \mu_x)\Big] + \mathbb{E}\Big[\hat{b}_{xt}(x_t - \bar{x}_{t-1})\,\epsilon_{t+1}\Big].$$

Immediately, $\mathbb{E}\Big[\hat{b}_{xt}(x_t - \bar{x}_{t-1})\,\epsilon_{t+1}\Big] = 0$. Then:

$$\mathbb{E}\Big[\hat{b}_{xt}(x_t - \bar{x}_{t-1})(x_t - \mu_x)\Big] = b_x\,\mathbb{E}[(x_t - \bar{x}_{t-1})(x_t - \mu_x)].$$

Therefore

$$\mathbb{E}\Big[\hat{b}_{xt}(x_t - \bar{x}_{t-1})(r_{t+1} - \mu)\Big] = b_x^2\,\mathbb{E}[(x_t - \bar{x}_{t-1})(x_t - \mu_x)].$$

Note further that:

$$\mathrm{Var}(x_t) = \frac{\sigma_x^2}{1-\rho^2}, \qquad \mathrm{Cov}(x_t, \bar{x}_{t-1}) = \frac{\sigma_x^2}{1-\rho^2}\cdot\frac{1}{t}\cdot\frac{\rho(1-\rho^t)}{1-\rho}.$$

So:

$$\mathbb{E}[(x_t - \bar{x}_{t-1})(x_t - \mu_x)] = \mathrm{Var}(x_t) - \mathrm{Cov}(x_t, \bar{x}_{t-1}) = \frac{\sigma_x^2}{1-\rho^2} - \frac{\sigma_x^2}{1-\rho^2}\cdot\frac{1}{t}\cdot\frac{\rho(1-\rho^t)}{1-\rho}.$$

Lastly:

$$\boxed{\mathbb{E}\Big[\hat{b}_{xt}(x_t - \bar{x}_{t-1})(r_{t+1} - \mu)\Big] = b_x^2\left[\frac{\sigma_x^2}{1-\rho^2} - \frac{\sigma_x^2}{1-\rho^2}\cdot\frac{1}{t}\cdot\frac{\rho(1-\rho^t)}{1-\rho}\right] = b_x^2 V\left[1 - \frac{1}{t}\cdot\frac{\rho}{1-\rho}\right] + O\big(\frac{1}{t^2}\big)}$$

For the infeasible portfolio that uses the true value of b_x:

$$\mathbb{E}[b_x(x_t - \bar{x}_{t-1})(r_{t+1} - \mu)] = b_x^2\,\mathbb{E}[(x_t - \bar{x}_{t-1})(x_t - \mu_x)] = b_x^2 V\left[1 - \frac{\rho(1-\rho^t)}{t(1-\rho)}\right] = b_x^2 V(1 - \frac{\rho}{t(1-\rho)}) + O\big(\frac{1}{t^2}\big).$$

For the infeasible portfolio that additionally uses the true value of μ_x:

$$\mathbb{E}[b_x(x_t - \mu_x)(r_{t+1} - \mu)] = b_x\,\mathbb{E}\big[(x_t - \mu_x)\big(b_x(x_t - \mu_x) + \epsilon_{t+1}\big)\big] = b_x^2\,\mathbb{E}\big[(x_t - \mu_x)^2\big] = b_x^2\,V.$$

D Optimal shrinkage in a simpler setting

The data-generating process is:

$$r_{t+1} = b_x x_t + \epsilon_{t+1}, \quad E[\epsilon_{t+1}] = 0, \quad \text{Var}(\epsilon_{t+1}) = \sigma^2,$$

$$x_{t+1} = \rho x_t + u_{t+1}, \quad E[u_{t+1}] = 0, \quad \text{Var}(u_{t+1}) = \sigma_x^2,$$

with u and ϵ uncorrelated. Define

$$\hat{b}_{xt} = \frac{\sum_{s=0}^{t-1} x_s r_{s+1}}{\sum_{s=0}^{t-1} x_s^2}, \quad V = \frac{\sigma_x^2}{1-\rho^2}.$$

This is a simplified setting from that in Section 5 as both μ and μ_x are set to 0. Consider the forecast f_t shrunk with the constant c. Specifically, let $f_t(c) = c x_t \hat{b}_{xt}$.

$$\text{MSE}[f_t(c)] = \mathbb{E}\left[(r_{t+1} - c x_t \hat{b}_{xt})^2\right] = \sigma^2 + \mathbb{E}[B_t^2],$$

where

$$B_t = b_x x_t - c\hat{b}_{xt} x_t.$$

It follows that

$$\mathbb{E}[B_t^2] = b_x^2\, V + c^2\, b_x^2\, \mathbb{E}\big[x_t^2\big] - 2c\, b_x^2\, \mathbb{E}\big[x_t^2\big] + c^2\, \sigma^2\, \mathbb{E}\left[\frac{x_t^2}{S_{xx}}\right] = b_x^2\, V + c^2\, b_x^2\, V - 2c\, b_x^2\, V + c^2\, \sigma^2\, \frac{1}{t},$$

where $S_{xx} = \sum_{s=0}^{t-1} x_s^2$. It is clear that the value of c that minimizes $\text{MSE}[f_t(c)]$ is:

$$c^* = \frac{\frac{b_x^2 V}{\sigma^2}}{\frac{b_x^2 V}{\sigma^2} + \frac{1}{t}}.$$

A similar $c*$ for general μ and μ_x is not in closed form unless $\rho = 0$. This is because $\mathbb{E}\left[\frac{(x_t - \bar{x}_{t-1})^2}{\sum_{s=0}^{t-1}(x_s - \bar{x}_{t-1})^2}\right]$ is not in closed form, although a similar formula could be worked out by noting that $\mathbb{E}\left[\frac{(x_t - \bar{x}_{t-1})^2}{\sum_{s=0}^{t-1}(x_s - \bar{x}_{t-1})^2}\right] = \frac{1}{t} + O\big(\frac{1}{t^2}\big)$. The key message in that formula remains identical.

Recall that $\frac{b_x^2 V}{\sigma^2}$ is roughly the population R^2. c^* is determined between the population R^2 and $\frac{1}{t}$. When the population R^2 is positive, c^* increases and approaches 1 as t increases. For a given t, c^* increases with the population R^2. In other words, weaker shrinkage should be applied when one has more data or a stronger predictor.

However, population R^2 is not observed but must be estimated to operationalize the computation of c^*. A simple (yet crude) way to do this is to use the t-statistic: $\frac{\hat{b}_x}{\hat{se}(\hat{b}_x)}$. Specifically, note that:

$$
\begin{aligned}
E[\text{t-stat}^2] &= E[\frac{\hat{b}_{xt}^2}{\hat{se}^2(\hat{b}_{xt})}] \approx E[\frac{\hat{b}_{xt}^2}{se^2(\hat{b}_{xt})}] = E[\frac{\hat{b}_{xt}^2 S_{xx}}{\sigma^2}] = E[E[\frac{\hat{b}_{xt}^2 S_{xx}}{\sigma^2}|\{x_s\}_{s=0}^{t-1}]] = E[\frac{b_x^2 S_{xx} + \sigma^2}{\sigma^2}] \\
&= 1 + t\frac{b_x^2 V}{\sigma^2}
\end{aligned}
$$

This relation is not exact because it replaces the standard error estimates with the standard errors. On the other hand, this crude approximation does lead to a convenient estimate of the population R^2 which is $\frac{\text{t-stat}^2-1}{t}$, and therefore a rule of thumb estimate of c^*:

$$
\frac{\text{t-stat}^2 - 1}{\text{t-stat}^2}
$$

Note that in the event the observed t-stat is less than 1, the above estimation of the population R^2 is clearly incorrect as it is negative. A simple way to account for that in c^* is to use $\max(\frac{\text{t-stat}^2-1}{\text{t-stat}^2}, 0)$ instead. This is the L-multiplier shrinkage commonly used in the machine learning literature (e.g., Lin et al. (2025)) but is different from the forecast combination method used in Goyal et al. (2024).