five values of r_t (row 5) from 1975 to 1979 is 11.81%. However, inspection of the annual values reveal that although r was abnormally depressed in 1975, its values for the next three years exhibited a definite upward trend, and then only declined slightly in 1979. Investors now might well believe that the material rise in the cost of capital between 1975 and 1979 justifies the rates of return the company realized in the more recent years, in which case they would rely primarily on the 1978 and 1979 figures in forecasting the company's future rate of return. A simple average of these figures is 13.05% and it seems reasonable that investors might conclude that 13% represents the best estimate of the long-term return AT&T is expected to earn on common equity.

For the retention rate that investors expect, we first note that a simple average of the five values of b_t (row 9) from 1975 to 1979 is 37.23%. However, this average is affected by the low retention rate in 1975, and in recent years, 1977-1979, the retention rate has averaged 38.93%. It seems reasonable that on the basis of this data, investors might use these recent years, and arrive at 39% as the best estimate of AT&T's retention ratio.

Combining the above values (obtained by using historical values in Equation (8) for P_0, D_1, b, and r) provides an estimate of AT&T's cost of equity capital as of March 28,

1980, of:

$$k = \frac{D_1}{P_0} + br$$

$$= \frac{\$\ 5.00}{\$48.50} + (.39)(.13)$$

$$= .1031 + .0507 = 15.38\% \ .$$

However, before accepting this result it may be
instructive to pose the following question: What would
have been the estimate for k as of June 30, 1979?

3.   <u>Growth Rate - Recent Developments</u>

On June 30, 1979, Value Line estimated that AT&T's
1979 earnings would be $8.00 per share. The actual
value of earnings per share for 1979 was $8.04.
Since we would have been reluctant to estimate k at that
time without 1979 data, we would have relied on the Value
Line forecast to complete the 1979 annual data, a proce-
dure we have used in the past. Since the Value Line
estimates were extremely close to the actual 1979 results,
using these estimates and the historical data would have
produced the same estimates of b and r obtained previously.
It is obvious that if the data and analysis do not change
materially, we would obtain the same measurement of the growth
rate at any point between June 30, 1979, and March 28, 1980.
The estimates which would have been obtained on two
previous dates are provided below:

Date	D_1/P_0	+	br	=	k
June 30, 1979	8.99%		5.07%		14.06%
November 19, 1979	9.39%		5.07%		14.46%

An estimate is provided for November 19, 1979, for comparative purposes, since an estimate of k was obtained for Rochester Telephone Co. on that date of 14.85%.[1] The difference in k between Rochester Telephone and AT&T may be attributed to AT&T's slightly lower business risk due to its greater diversification.

The problem can now be easily seen. The estimate of 15.38% obtained for AT&T is correct only if we assume that the large increase in the expected rate of inflation (which raised the dividend yield on AT&T from 8.99% on June 30, 1979, to 10.31% on March 28, 1980) had no effect on the anticipated growth in the dividend.

It is extremely unlikely that investors believe that to be true. The rise in the expected rate of inflation has not only increased interest rates, but also the expected rate at which AT&T's other costs of production, such as materials and labor, will grow. A continued expectation that the company will earn a return on common of 13% and retain 39% of earnings would require the belief that the rate of growth in its revenues will rise to match

---

[1] Myron J. Gordon, Direct Testimony, Before the State of New York Public Utility Commission, In the Matter of Rochester Telephone Co., November 20, 1979.

the rise in the rate of growth of its costs. However, if
investors fear that the regulatory process will not be
fully responsive to the increase in the rate at which the
company's costs are rising, they will revise their growth
estimate downward. That is, with any regulatory lag in
the pass through of higher costs, a rise in the expected
inflation rate would reduce investor estimates of long-run
return on common equity, and would, therefore, result in a
downward revision of expected growth. In that event, simply
raising the estimate of AT&T's cost of equity capital by the
increase in the dividend yield would result in an overstate-
ment of the required return.

It is our judgment that the response of investors
to the rise in the expected rate of inflation has been a
downward revision in expectations regarding AT&T's rate of
return on common equity, implying a downward revision in
its retention rate also. In support of this position, we
note that Value Line lowered its prediction of 1980 earn-
ings per share for AT&T to $7.50, and lowered its predicted
1980 dividend per share to $5.00.[1] This implies for 1980
an estimate for r of 11.60% and an estimate for b of
33.33%.

Under the present turbulent economic conditions it is
extremely difficult to estimate with precision the extent

---

[1] Value Line, February 1, 1980.

to which these rates have been revised downward. If the
revised figures are a 12.50% return on common equity and
a 37% retention rate, then the estimated growth rate must
be reduced from 5.07% to 4.63%.[1] Adding the latter figure
to the current dividend yield of 10.31% results in a cost
of equity capital of 14.94%. On the other hand, the rise
in interest rates over the past six months may be taken as
evidence that the cost of equity capital has gone up over
the same time period. Hence, in some measure, this rise in
interest rates will lead to an upward revision in the
rate of return allowed by the numerous regulatory commis-
sions that set rates for AT&T. A generous allowance for
a favorable impact of increases in the allowed rates of
return on investment forecasts of the AT&T growth rate is a
rise in its value from the above 4.63% to 5.25%. This latter
growth rate combined with the 10.31% dividend yield re-
presents cost of equity capital of 15.56%. In our judg-
ment, the AT&T cost of equity capital may well be as low as
14.0% but is not likely to be above 15.5%, and 15.25%
represents our base estimate as of March 28, 1980.

---

[1] Using this reasoning, the growth rate was adjusted down-
ward by 69 basis points for Rochester Telephone. Ibid.,
Supplemental Prepared Direct Testimony, March 24, 1980.

A Comprehensive Look at the Empirical Performance of Equity Premium Prediction II

Amit Goyal
Ivo Welch
Athanasse Zafirov

September 23, 2021

Abstract

Our paper reexamines whether 29 variables from 26 papers published after Goyal and Welch (2008), as well as the original 17 variables, were useful in predicting the equity premium in-sample and out-of-sample. Our samples include the original periods in which these variables were identified, but ends later (in 2020). Most variables have already lost their empirical support, but a handful still perform reasonably well. Overall, the predictive performance remains disappointing.

Reader Please Note: Our paper examines many papers with many different variables in many different ways — over a thousand numbers altogether. Therefore, this draft contains myriads of formatting (color, background color, and font-sizing) that are intended to draw the reader's eye towards important results and away from unimportant ones. This is *not* standard journal formatting and will change before the paper is to be submitted to a journal. We also thank the authors of many papers reexamined here for corrections and feedback. Please bring any other errors to our attention. Amit Goyal's website at https://sites.google.com/view/agoyal145 contains a long and detailed appendix of variable definitions and results when the dependent variable is not logged.

Since Goyal and Welch (2008), henceforth GW, a large number of papers predicting the equity premium have been published in top finance journals. It thus seems that academic finance has conquered the problem of investors' predicting time-varying future stock market rates of returns. Many of these papers have further offered strong theoretical foundations for their proposed variables, presumably increasing faith in their forward-looking stability.

Our own paper here reexamines 29 variables proposed in 26 prominent recently published papers (Table 1), for which we could relatively easily reconstruct or obtain the predictive variable. The data in these papers ended between 2000 and 2017. We can replicate and confirm the principal in-sample findings for all but two of the papers, using a simple but consistent predictive framework based on uncontrolled OLS forecasting regressions. (Two papers had data issues.)

We then extend the samples by a few years, ending with stock market returns in December 2020—typically about ten extra years of data.. Because our paper reuses the data that the authors themselves had originally used to discover and validate their variables and theories, all that the predictors had to do in the few added years was not to "screw up" badly. The original results should still hold.

Yet, we find that most variables have already lost their predictive ability. Of 29 variables, 25 variables show lower in-sample significance when we use our extended sample period instead of the authors' original sample period. Only four variables predicted about equally well or better. The widespread deterioration in predictive performance partly reflects the fact that the added years offered great variety. There were three recessions, one in the early 2000s (with 9/11 and the dot-com end), one in 2008 (the Great Recession) and one in 2020 (the Covid-19 recession); and there were two major bear markets from 2000–2002 and in 2008 (plus a minor one in 2018). These (perhaps in-retrospect unusually remarkable) episodes could influence either the independent predictor variable or the dependent predicted equity premium enough to make a difference in the apparent forecasting ability, even though we also included and thus recycled the authors' discovery samples.[1]

Our paper investigates not only whether variables had good and statistically significant in-sample and out-of-sample performance, but also the investment timing performance in some simple investment strategies.

Our first investment strategy was long $1 in the equity premium when the predictive variable was bullish (relative to its prevailing median) and short otherwise. Not a single variable meaningfully (much less statistically) beat the buy-and-hold equity investment (*all-equity-all-the-time*). One can object that being invested in the stock market over the last 20-50 years was a remarkable positive experience that was tough to beat—though doing so is of course the point of market-timing. But good *all-equity-all-the-time* was not the only reason. Half of the predictors performed so poorly that they not only failed to beat *all-equity-all-the-time*, they even *lost* money in absolute terms. Our second investment strategy tilted far more towards equity. It was long the equity premium *unless* the predictor signal was extremely bearish (worse than the

---

[1]The recycling obviously gives a large advantage to the considered predictors and does not make our reexaminations true independent out-of-sample confirmations. However, rather than reflecting type-I errors in the original sample, the in-retrospect perhaps unusual economic and market performance could also have led to more type-II errors. Both are limitations of empiricism that are not possible to overcome within our expected lifetimes.

prevailing 25th percentile). Our third and fourth strategies also scaled the investment based on the Z-magnitude of the signal relative to preceding signals.

Of all variables, only one performed as well the *all-equity-all-the-time* investment strategy on the first strategy untilted unscaled timing strategy. On the equity-tilted strategy, 9 variables had higher returns than *all-equity-all-the-time* (and of these 9, only half were still significant in-sample). With both tilt and scale, the tally improved to 13 variables. Even then, not a single variable outperformed *all-equity-all-the-time* in a statistically significant manner—in fact, none could muster an absolute T-statistic above 1.3. Again, this is despite the fact that all timing strategies could invest during the sample periods in which the variables were identified to begin with.

As already hinted, not all variables performed poorly. The empirical analysis suggests some good candidates. The best and most consistent variable was:

Fourth-Quarter Growth Rate in Personal Consumption Expenditures (gpce) was introduced in Møller and Rangvid (2015). High personal consumption growth rates this year predict poor stock-market returns next year.

Empirically, since the 1970s, gpce has only made one modest misstep in its predictive ability (which was missing the Great Recession bear market). Otherwise, gpce has been a steady performer. (Nevertheless, a risk-averse investor, as defined by Campbell and Thompson (2008), would still not have been better off using gpce.)

A number of other variables have good performance on some but not all criteria. Thus, one could put them on a "watchlist" to monitor whether their performance will improve or deteriorate in the future. In no particular order:

(Aggressive) Accruals (accrul) was introduced by Hirshleifer, Hou, and Teoh (2009). Aggressive corporate accruals predict low future stock returns.

Tempering our enthusiasm, accrul's performance was episodic. In fact, it had only one—though singularly stellar—prediction. In 1999-2001, it strongly and correctly forecast the post-Tech stock market decline of 2000-2002. Since then, accrul has not moved much. Thus, its single outlier performance was enough to at first obtain and subsequently avoid losing its performance in our extended sample. (Incidentally, a risk-averse investor would not have been better off using accrul.)

Credit Standards (crdstd) was introduced by Chava, Gallmeyer, and Park (2015). Optimistic credit standards predict poor stock market returns. It had good OOS performance and usually was the best performer on our investment strategies. Tempering our enthusiasm, crdstd's in-sample T-statistic as of 2020 is only −1.65.

The Investment-Capital Ratio (i/k) was introduced in Cochrane (1991) and included in Goyal and Welch (2008). High investment this quarter predicts poor stock-market returns next quarter. Interestingly, like gpce, i/k associates more outlays today with lower market performance in the future—almost as if the alternative had been stockpiling funds today to allow for more market investment later.

2

For the 13 years from 1975 to 1998, i/k was a poor predictor. In the 22 years since then, i/k has consistently performed well. Thus, it performs better today than it did in Goyal and Welch (2008). Tempering our enthusiasm, its estimated IS coefficient in our sample has declined from –2.17 in the first half to –0.93 in the second half; and i/k could not outperform *all-equity-all-the-time* in three of our four timing strategies.

Again, finding that one handful of variables among nine handfuls have good predictive ability is somewhat disappointing, given not only that these variables were not randomly selected but also that they were already validated in much of the same sample that we are merely recycling. In a sense, within a decade or so, most variables have already become dubious or obsolete.

Our paper now proceeds as follows. Section I lays out our performance criteria. Section II briefly describes the variables and lays out the tables that our analysis refers to. Section III runs through the performance of each of the variables, in alphabetical order of authors. Section IV briefly looks at the performance of the most promising variables from the perspective of a risk-averse investor. Section V takes some liberty in offering some more subjective thoughts on the overall performance tally.

I   Performance Criteria

We first needed a set of variables for which we could confirm the basic predictive results from papers that published them. This means that we had to be able first confirm the authors' results within their sample periods and then be able to extend the variable to 2020. This means we had to exclude variables that are proprietary and not available to us.[2]

In our opinion, to be considered a reliable and useful predictor of the equity premium on a forward-looking basis, a variable should satisfy at least the following set of criteria:

1. The variable should be able to predict the equity premium at a conventional statistical significance level using OLS in an in-sample regression in our extended sample period. The absolute T-statistic should at least be 1.65. If this fails, there is little reason to proceed.

2. The model should be reasonably stable, i.e., a variable should *not* have statistically significantly different IS coefficients and/or a sign change in predicting the equity premium in our sample's first half and second half—for us, at least at the ≈5% level. If this fails, there is little reason to proceed.

3. The variable should have positive rolling OOS R^2, suggesting basic improvement of the conditional residuals over the unconditional residuals (the latter from a simple prevailing equity-premium average model).[3] If this fails, there is little reason to proceed.

---

[2]We admit to giving the original paper the benefit of the doubt in trying to follow its methods somewhat more closely in Table 2 than we do in subsequent tables. Thus, we may use the preferred data frequency and overlap handling of the authors.

[3]We note that Campbell and Thompson (2008) discuss reasons when a researcher may want to focus on IS prediction rather than both IS and OOS prediction, as we do. These reasons usually apply when testing theories in which the researcher is sure that the model is stable and known by market participants in time.

4. On our four simple timing investment strategies (untilted and equity-tilted; unscaled and scaled), the variable should earn higher rates of return than the *all-equity-all-the-time* unconditional strategy.

Furthermore, we take into account two heuristic concerns, though they are not "make-or-break":

7. The variable's performance should not be driven almost entirely by its performance in one or two unusual years only. It should also show reasonably good performance over the last 20-30 years.

8. The variable should offer positive ex-post utility improvement for a quadratic-utility investor with parameter 5, as suggested by Campbell and Thompson (2008).

We are however tolerant of two problems:

1. We ignore the fact that a variable that has the choice to be statistically significant in one of three frequencies (say, monthly, quarterly, annually) should be viewed more critically. Simulations suggest that one should use not the 10% significance level of 1.65 when allowing consideration of monthly, quarterly, and annual frequencies, but more appropriately a 10% level of 2.1 on the best of the three. Our failure to impose this more stringent criterion is partly counterbalanced by the fact that we expect variables to offer performance not just on IS prediction, but also on other criteria.

2. We ignore the fact that, collectively, the profession has examined many more variables and that the variables we observe are themselves already highly selected (Lo and MacKinlay (1990), Harvey, Liu, and Zhu (2016)).

A sufficiently skeptical researcher may therefore want to impose even more stringent criteria. Of course, a researcher with sufficiently strong positive priors on the model may want to discount our empirical evidence altogether.

4

II Variables and Tables

A. Variables

Table 1 contains the glossary of recently published papers and variables that we investigate. It explains their meaning and sample availability briefly. This will be followed, in more detail, in our paper by a paper discussion below; and in most detail in the appendix.

The variables can broadly be grouped into six categories:

Macroeconomic: sbdlev, pce, govik, crdstd, ogap, ndrbl, gpce (and gip), house.

Sentiment: accrul (and cfacc), sntm, ygap, shtint.

Variance-Related: impvar, vrp.

Stock Cross-Section: lzrt, skew, skvw, tail, fbm, rdsp, avgcor.

Other Stock Market: tchi, dtoy (and dtoat), disag;

Commodities: wtexas.

Most stock-market related variables are monthly, most macroeconomic variables are quarterly or annual.

In addition, our paper also looks again at the performance of 17 variables already investigated in Goyal and Welch (2008): the dividend-price ratio (d/p), the dividend-yield (d/y), the earnings-price ratio (e/p), the dividend-payout ratio (d/e), as in Campbell and Shiller (1988); stock volatility (svar), as in Guo (2006); book-market (b/m), as in Kothari and Shanken (1997) and Pontiff and Schall (1998); net issuing activity (ntis), as in Boudoukh, Michaely, Richardson, and Roberts (2007); equity issuing activity (eqis), as in Baker and Wurgler (2000); the T-Bill rate (tbl), as in Campbell (1987); the long-term yield (lty), the long-term bond rate of return (ltr), the term-spread (tms), the default yield (dfy), the default rate of return (dfr), as in Fama and French (1989), the inflation rate (infl), as in Fama and Schwert (1977); private investment (i/k), as in Cochrane (1991), and "cay, " as in Lettau and Ludvigson (2001). For precise definitions, please refer to Goyal and Welch (2008).

5

B. Tables

To examine the predictive performance of 46 variables while fitting into the space of a standard article, we have to be frugal in our descriptions. This is best accomplished by following a standard format discussing each variable, while referring to a set of common tables. We will do so as follows.

Our first task is to confirm that we can create variables that match the performance proposed by the original papers. In most cases, the authors have posted or shared their data series, allowing us to confirm their key results using our own calculations.[4]

Table 2 shows our ability to replicate the basic results of the original paper using the original sample period, and (where possible) the same controls.

Once we have confirmed that we can obtain similar results, we can extend the sample to 2020. Our key results examining IS and OOS performance appear in four tables:

Table 3 shows the prediction performance of log equity premia for variables available on monthly frequency.[5]

Table 4 does the same, but for variables available only on a quarterly frequency;

Tables 5 and 6 do the same, but for variables available only on annual frequency—Table 5 for the calendar year, Table 6 for July-to-June performance with a 6-month recording lag (i.e., the predictor being measured as of the previous December).

For the IS performance, we predict the equity-premium based on each variable using a standard OLS regression. We also look at the stability of the model by dividing the sample into two halves and estimating the coefficients separately. This gives equal billing to the first and the second half, thereby not disadvantaging the first few predictions as in our OOS prediction. For the OOS performance, we focus on the in-time OOS R^2,

$$R^2_{\text{oos}} = 1 - \frac{\sum_t (r_t - \hat{r}_{t-1})^2}{\sum_t (r_t - \bar{r}_{t-1})^2},$$

where $\hat{r}_{t-1}$ is the conditional forecast at time $t-1$ and $\bar{r}_{t-1}$ is the prevailing mean at time $t-1$. We star this "OOS R^2" using the MSE-F statistic of McCracken (2007).[6] The variables are *always* constructed on a real-time basis—for example, when variables require filters or regression coefficients for construction (such as pce), these coefficients are always based only on prevailing historical values.

---

[4]The exceptions were Kelly and Jiang (2014), Piazzesi, Schneider, and Tuzel (2007), and Pollet and Wilson (2010).

[5]We are not predicting lower-frequency stock returns with higher-frequency predictors. Thus we need not worry about overlapping observations. In a previous draft, we found that higher frequency variables generally did not do better predicting lower-frequency equity premia, either with or without overlap.

[6]We use MSE-F statistic because we are interested in population-level predictive ability (whether a variable has any predictive content). One can test finite-sample predictive ability (whether a variable has useful predictive content given that parameters are estimated). Giacomini and White (2006) study such a question in the context of rolling regressions (where the null hypothesis then, necessarily, depends on window length).

6

The results are almost the same if we predict simple rather than log equity premia (available upon request). We experimented with more sophisticated forecasting, but the inference was similar enough to recommend the brevity and simplicity of an exposition based on plain OLS forecasting techniques. This includes our consideration of forecasting and techniques from Kostakis, Magdalinos, and Stamatogiannis (2015) and Cederburg, Johnson, and O'Doherty (2019).[7]

Our OOS period always starts 20 years after the IS period, but never earlier than 1946. Authors can (perhaps legitimately) complain that there are good reasons why they started their own analyses earlier or later. Obviously, different starting periods can lead to different results, just like different ending periods. Our choice was the same as that in Goyal and Welch (2008), and dictated by the desire to keep the same scheme across our 29 variables. Importantly, our figures make it easy to assess how different starting period would affect the results.

Next, we show the performance of a risk-neutral investor who seeks to time her investments. Performance is always based on zero-investment strategies (i.e., either the value-weighted stock market financed with bills, or bills financed by shorting value-weight stocks).[8] The unconditional investment strategy is earning the equity premium itself. We name this *all-equity-all-the-time*. The other investment strategy is timed, i.e., conditioned on the variable. When the timing investor is bullish (i.e., in the market), the unconditional and conditional strategies invest and earn the same. When the timing investor is bearish, the conditional strategy earns the opposite of the unconditional strategy.

We consider four variants based on *scaled* and *unscaled* timing strategies, and *equity-tilted* and *untilted* timing strategies in Tables 7-10.

The *untilted, unscaled* timed investment strategy (**Table 7**) invests either +\$1 in the market (financed by bills) when it is bullish or –\$1 in the stock market (saved in bills) when it is bearish. This conditional strategy decides based on whether the variable is bullish or bearish by looking above or below its historical median in time, according to the sign of the prevailing coefficient. The *equity-tilted* strategy (**Tables 9 and 10**) switches from long stocks to long bills only if the signal is very bearish, i.e., at the 25th percentile rather than the median. The *scaled* strategy (**Tables 8 and 10**) first calculates a Z-score in time, i.e., it subtracts the prevailing median (untilted) or first-quartile (tilted) from the x variable and then divides by the prevailing standard deviation. It then scales the investment by this Z-score. For example, when the prevailing forecasting coefficient is positive (so being above the x cutoff [median or first quartile] means bullish), if the Z-score calculates –0.5, the conditional strategy would short \$0.50 in the market and purchase \$0.50 of T-bills. The comparative unconditional strategy would long \$0.50 in the market and purchase \$0.50 in T-bills.

[Table 7 here: 'Untilted \$1-Unscaled Investment Strategy']

[Table 8 here: 'Untilted Z-Scaled Investment Strategy']

[Table 9 here: 'Equity-Tilted \$1-Unscaled Investment Strategy']

[Table 10 here: 'Equity-Tilted Z-scaled Investment Strategy']

---

[7]We do however highly recommend both. The latter further looks at a good number of recent prediction variables. Recent finance literature investigates the pitfalls associated with multiple hypothesis testing. The common approaches are to control family-wise error rate (Romano and Wolf (2005) and White (2001)) or false discovery rate (Benjamini and Hochberg (1995) and Benjamini and Yekutieli (2001)). **However, these approaches are not suitable for the nested models that we study here.** We thank Todd Clark and Michael McCracken for clarifying these issues for us.

[8]Zero-investment strategies can always be viewed as "overlays." Thus, they are comparable but never mutually exclusive.

C. Results Preview

There are only five variables that have both a statistically significant in-sample coefficient and a positive OOS R^2 (all of which happen to be statistically significant at least at the 10% level). On a monthly frequency, this is only the T-bill rate (tbl), as in Campbell (1987). On a quarterly frequency, these are credit standards (crdstd), as in Chava, Gallmeyer, and Park (2015); and the investment-capital ratio (i/k), as in Cochrane (1991). On an annual frequency, these are corporate accruals (accrul), as in Hirshleifer, Hou, and Teoh (2009), and the fourth-quarter growth-rate of personal consumption (gpce), as in Møller and Rangvid (2015).

Of these five variables, the T-bill rate does not help much in our investment strategies. The other five are somewhat inconsistent in how much they help—it depends on their exact deployment. Credit standards and accruals are usually the best performers. However, none yields returns that are statistically significant

Of these five variables, only three would have made a risk-averse investor no better off: the T-bill rate, credit standards, and the investment-capital ratio. Only one would have left the risk-averse investor statistically significantly better off: credit standards.

Of these five variables, accruals was a "one-trick pony." It helped greatly in the dot-com aftermath bear market. Sentiment was somewhat similar. gpce was most consistent.

III   Empirical Performance

We are now ready to describe the performance of the variables proposed in each recent paper, in alphabetical order of authors. Our standard discussion template for papers presents each variable as follows:

1. A modified version of the original abstract that focuses on relevant aspects. For the complete version, please refer to the original paper.

2. A basic intuitive explanation of the variable and sample period. This explanation is almost always insufficient to replicate our version of the variable. The fully detailed discussion appears in our appendix.

3. A discussion of the performance in four parts: [A] IS performance, including stability statistics (first half vs second half); [B] OOS R^2; [C] OOS investment performance; and [D] graphical performance.

4. Our somewhat subjective assessment.

1. AMP: Atanasov, Møller, and Priestley (2020)

We are now prepared to begin our discussion of AMP.

Abstract: *[AMP] introduce a novel consumption-based variable, cyclical consumption, and examine its predictive properties for stock returns. Future expected stock returns are high (low) when aggregate consumption falls (rises) relative to its trend and marginal utility from current consumption is high (low). [They] show that the empirical evidence ties consumption decisions of agents to time-variation in returns in a manner consistent with asset pricing models based on external habit formation.*

Variable: The key variable, pce, measures NIPA seasonally adjusted consumption on nondurables and services, provided by the Bureau of Economic Analysis, relative to a trend that is identified by using a filter. pce is available quarterly.

Performance: The performance of pce is as follows:

[A (IS)] We can confirm the strong negative and statistically significant IS coefficient of pce prior to 2017 in their sample also in our own data (Table 2). We then investigate our extended sample, which ends in December 2020. Being of quarterly frequency, our key results appear in Table 4. The two left-most columns show IS performance. We can confirm that pce also has negative IS significance in our extended sample. The three middle columns show that the AMP model is reasonably stable across its two halves. (The IS coefficient is modestly weaker in the second half but not statistically significantly so.) Given good IS performance, it makes sense to continue and consider OOS performance.

[B (OOS)] pce performed poorly on OOS prediction, as shown in the two right-most column in Table 4. The OOS R^2 is a negative −3.44% in our sample.[9]

[C (Investment)] Table 7 shows that the unbiased untilted OOS annual timing strategy underperformed the *all-equity-all-the-time* non-timing equivalent by about 2.5% per year. The three other investment strategies do not show performance better than *all-equity-all-the-time*, either. The scaled strategies in Tables 8 and 10 suggest slightly negative (–0.2%/year) performance, while the equity-tilted but unscaled strategy suggests slightly positive performance (0.1%/year).

[D (Graphical)] Our performance figures (Goyal and Welch (2008)) show when a variable performed well and when it did not. Intuitively, in these figures, when the prediction based on the conditioning variable (here pce) does well, the line increases; when the variable underperforms (the prevailing mean for the OOS lines), the line decreases. The solid lines use simple returns, the dashed lines use log returns. The black lines are IS predictions, the blue lines are OOS predictions (which means the conditional prediction in time is compared to the unconditional prediction at time t, the prevailing mean). A variable that

---

[9]In the original paper, the authors began OOS prediction in 1980. This avoided the first 7 years of poor OOS performance in our sample. It was enough to keep pce out of the red zone, though not enough to show meaningfully positive OOS performance (much less with statistical significance). Further unreported investigation shows that our OOS starting forecasting quarter was particularly unfortunate for pce. The OOS turns positive with later starting points, though not statistically significantly so.

9

is statistically significant should lie solidly above the 0 line.[10] The authors' original end of sample is shown with a vertical dotted red line.

Figure 1 shows that the predictive performance of pce was quite good in-sample (IS), although much of its good IS performance appeared in the first 20 years. Since about 1975, the IS performance has been more modest. In contrast, the OOS performance was poor for the first 10 years, reaching its lowest cumulative point when (mis-)predicting the equity-premium in Q2-1982 with pce of Q1-1982. It was largely unremarkable thereafter. The red line shows that the variable did perform well OOS in 2020, which was after the original sample had ended in 2017.

Evaluation: We dismiss pce as a useful predictor of equity premia, based on poor and insignificant OOS performance. Presumably, if the evidence in Atanasov, Møller, and Priestley (2020) was consistent with asset pricing models based on habit formation, the extended evidence should now be viewed as unsupportive.

[Figure 1 here: 'IS and OOS Predictive Performance of AMP pce (quarterly)']

2. AMS: Adrian, Mönch, and Shin (2010)

Abstract: *[AMS] document that financial intermediary balance sheet aggregates contain strong predictive power for excess returns on a broad set of equity ... portfolios. [These] results provide support to the hypothesis that financial intermediary balance sheet quantities matter in the determination of risk premia...Our findings point to the importance of financing frictions in macroeconomic dynamics and asset pricing.*

Variable: AMS entertain a number of potential variables and use Lasso to select, as their strongest candidate, the quarterly variable 'ySBRDLRlevg'. Unfortunately, their definition of ySBRDLRlevg can and does cause negative denominators in their ratio, raising doubts about its definitional validity. We modify their definition to a variant, sbdlev.[11] sbdlev is available quarterly.

---

[10]The blue range is the ± 2 standard deviation range for OOS prediction, based on an MSE-T statistic Diebold and Mariano (1995), which is related to but not identical to the MSE-F statistic used to star the OOS R^2 in the tables.

[11]AMS do not want to measure the ratio of world assets over world equity, but (presumably a proxy for) the ratio of domestic assets over domestic equity. They thus calculate

$$\text{ySBRDLRlevg} \equiv \log \left(\frac{\overbrace{\text{World Domestic \& Foreign Financial Assets}}^{\text{FDI Assets}}}{\underbrace{\text{World SBD Equity} - \underbrace{(\text{FDI Equity} + \text{FDI Non-Equity})}_{\text{FDI Assets}}}_{\text{Domestic Equity Proxy If FDI Non-Equity is small}}} \right),$$

where SBD is "security-broker-dealer" and FDI is "foreign direct investment." FDI equity alone is unfortunately not available, making it impossible to accurately calculate Domestic Equity. They thus subtract FDI total assets (not just equity) in the denominator, which is reasonable if FDI non-equity investment is very small. (It is also not clear to us why they use world assets in the numerator.) We can modestly improve on their variable and avoid zero or negative denominators by using "World SBD equity - RoW FDI Equity * (SBD FDI/RoW FDI)" where RoW is the result of the world. We dub our variable sbdlev. sbdlev has good correlation with a version of ySBRDLRlevg emailed to us by the authors.

10

Performance: **[A]** Table 2 shows that our coefficient of –0.03 (T of –1.04) cannot replicate AMS' significant negative coefficient of –0.09 (T of –3.01) in the same sample (–2009). The table also shows that the sbdlev model was unstable. The IS coefficient switches sign from positive in the first half ("H1") to negative in the second half ("H2") in our sample. This is also the case in our extended sample ending in 2020. Table 4 shows that the IS coefficient switched from positive to negative, with the overall coefficient having a T-statistic of 0.87. Thus, with poor IS performance, further OOS investigation seems unwarranted. (**[B-D]** The OOS and investment performance is always poor, too. Thus, we also do not graph pce's performance.)

Evaluation: We dismiss sbdlev as a useful predictor of equity premia, due to lack of replicability and both poor IS and OOS performance. Presumably, if the evidence in Adrian, Mönch, and Shin (2010) was consistent with a role for financial intermediary frictions, the extended evidence should now be viewed as inconsistent.

3. BPS: Bakshi, Panayotov, and Skoulakis (2011)

Abstract: *[BPS] present an option positioning that allows [them] to infer forward variances from option portfolios. The forward variances [they] construct from equity index options help to predict ... (iii) stock market returns... .*

Variable: BPS synthesize the exponential of integrated variance using a strip of European calls and puts, written on the market index. impvar is available monthly.

Performance: **[A]** We can confirm the strong positive and statistically significant IS coefficient of impvar in the original sample period (–2008). In our extended sample (–2020), the IS coefficient is no longer statistically significant (Table 3). Moreover, the model was always unstable: The IS coefficient switches sign from the first half (H1) to the second half, both in the original and in our own sample. Thus, with poor IS performance, further OOS investigation seems unwarranted. (**[B]** The OOS R^2 of impvar is negative. **[C]** The investment performance of impvar was poor. When not tilted towards equity, impvar not only does not beat *all-equity-all-the-time*, it even loses money in absolute terms. When tilted towards equity and unscaled, it barely manages to avoid such exceptionally bad performance.) **[D]** Figure 2 shows why our results are so different from the authors': impvar collapsed completely in the Great Recession, just after the BPS sample had ended in Sep 2008. Specifically, impvar's Sep and Oct 2008 values failed to predict the –18% and –8.5% drops in the value-weighted market rate of return in Oct and Nov 2008.

Evaluation: We dismiss impvar as a useful predictor of equity premia, based on poor IS and OOS performance. Presumably, if the evidence in Bakshi, Panayotov, and Skoulakis (2011) was consistent with a role for implied volatility, the extended evidence should now be viewed as inconsistent.

[Figure 2 here: 'IS and OOS Predictive Performance of BPS **impvar** (monthly)']

11

4. BTZ: Bollerslev, Tauchen, and Zhou (2009)

Abstract: *Motivated by the implications from a stylized self-contained general equilibrium model incorporating the effects of time-varying economic uncertainty, [BTZ] show that the difference between implied and realized variation, or the variance risk premium, is able to explain a nontrivial fraction of the time-series variation in post-1990 aggregate stock market returns, with high (low) premia predicting high (low) future returns. [The] empirical results depend crucially on the use of "model-free," as opposed to Black-Scholes, options implied volatilities, along with accurate realized variation measures constructed from high-frequency intraday as opposed to daily data. The magnitude of the predictability is particularly strong at the intermediate quarterly return horizon...* BTZ is the most-cited paper in our set, with about 1,500 Google scholar citations..

Variable: Unlike other variables, we did not compute vrp ourselves. Instead, it is updated regularly by the authors themselves and posted on their website. vrp is available monthly.

Performance: **[A]** We can confirm the strong positive and statistically significant IS coefficient of vrp in the original sample period (–2007). In our extended sample (–2020), the IS coefficient is no longer statistically significant. The IS T-statistic is now 0.12. Moreover, the model has become unstable. The coefficient is now negative in the second half of the extended sample. Thus, with poor IS performance, further OOS investigation seems unwarranted. (**[B]** The OOS R^2 of vrp is negative. **[C]** The investment performance of vrp was poor. In fact, it is between –0.7%/year and 4.3%/year, always greatly underperforming *all-equity-all-the-time* (6.4%/year and 7.7%/year).)

   [D] Figure 3 shows that vrp did well following the Great Recession. However, it collapsed badly in early 2020. In Feb 2020, vrp predicted +3.52%, much above the prevailing mean of +0.66%. Because the actual Mar 2020 equity premium was –12.32%, the relative errors were –15.84% vs. –12.98%, with a squared difference of about –0.8%. In Mar 2020, vrp reversed itself, predicting –14.57% for Apr 2020 (vs. 0.62%). Because the actual Apr 2020 equity premium was +12.89%, the relative errors were 27.46% vs. 12.26%. This increased the cumulative squared difference by a further dramatic 6%, thereby falling off our common (monthly return) chart scale of –3% to +3%. Obviously, this poor performance after their sample had ended explains why our inference is different.

Evaluation: We dismiss vrp as a useful predictor of equity premia, based on poor IS and OOS performance.

   Presumably, if the evidence in Bollerslev, Tauchen, and Zhou (2009) was consistent with their stylized self-contained general-equilibrium model with time-varying economic uncertainty, the extended evidence should now be viewed as inconsistent.

[Figure 3 here: 'IS and OOS Predictive Performance of BTZ vrp (monthly)']

12

5. BY: Belo and Yu (2013)

Abstract: *[BY find that] high rates of government investment in public sector capital forecast high risk premiums.... This result is in sharp contrast with the well-documented negative relationship between the private sector investment rate and risk premiums. To explain the empirical findings, [BY] extend the neoclassical q-theory model of investment and specify public sector capital as an additional input in the firm's technology. [They] show that the model can quantitatively replicate the empirical facts with reasonable parameter values if public sector capital increases the marginal productivity of private inputs.* Naturally, their finding has a strong policy implication, in that it suggests that governments may want to tax and invest more in infrastructure on the margin.

Variable: Their key variable, govik, measures government investment (in contrast to i/k described later, which measures corporate investment). Their original paper's Figure 1 also shows that govik peaked in 1950, then declined until 1982, increased sharply during the Reagan years, then stayed constant, and finally declined again from 2002 to 2010. govik is available quarterly.

Performance: **[A]** We can confirm the (small) positive and statistically significant IS coefficient of govik in the original sample period (–2010). In our extended sample, the T-statistic drops to 1.67 (Table 4). The model was always unstable. Both in the original and our own sample period, the IS coefficent turned negative in the second half. Thus, with poor IS performance, further OOS investigation seems unwarranted. (**[B]** The OOS R^2 of govik is negative. **[C]** The investment performance of govik was poor—indeed exceptionally poor. Except for the unscaled equity-tilted strategy, not only did govik not beat *all-equity-all-the-time*, it even lost money.)

[D] Figure 4 shows that all of the good IS performance was due to early performance. Since about 1960, govik has not had any good IS power. The OOS performance had some good predictions, specifically in 1970 and again during the oil-crisis from 1973 to 1974, but has underperformed ever since.

Evaluation: We dismiss govik as a useful predictor of equity premia, based on "ancient-only" IS performance and poor OOS performance. Presumably, if the evidence in Belo and Yu (2013) was consistent with a role for useful government infrastructure investment, the extended evidence should now be viewed as inconsistent.

[Figure 4 here: 'IS and OOS Predictive Performance of BY **govik** (quarterly)']

6. CEP: Chen, Eaton, and Paye (2018)

Abstract: *[CEP] constructs and analyzes various measures of trading costs in US equity markets covering the period 1926-2015. These measures contain statistically and economically significant predictive signals for stock market returns and real economic activity. [They]...find strong evidence that the component of illiquidity uncorrelated with volatility forecasts stock market returns...*

Variable: lzrt is the log of the number of zero returns. The series has structural break adjustments for tick-size reductions in 1997 and 2001 (these are included by regressing the series on binary variables that take the value of 1 after the tick-size reductions, and 0 otherwise, then taking the residuals as the final series). lzrt is available monthly.

Performance: **[A]** We can confirm the strong positive and statistically significant IS coefficient of lzrt in the original sample period (–2015). In our extended sample (–2020), the IS coefficient

13

is no longer statistically significant. The T-statistic falls to 0.96. Thus, with poor IS performance, further OOS investigation seems unwarranted. **[B]** The OOS R^2 was positive (Table 3). **[C]** The investment performance of lzrt was poor. Without the heavy equity tilt, lzrt even loses money in absolute terms. With equity tilt, lzrt still greatly underperforms *all-equity-all-the-time*.) **[D]** Figures 5 illuminates the performance. On a monthly basis, Chen, Eaton, and Paye (2018) caught the variable nearly at its best. It had outperformed in the Great Recession. However, lzrt collapsed in the Covid year of 2020. Otherwise, lzrt was fairly unremarkable.

Evaluation: We dismiss lzrt as a useful predictor of equity premia, due to poor IS performance, poor investment performance, and only-episodic superior OOS R^2 performance (in the Great Recession). Presumably, if the evidence in Chen, Eaton, and Paye (2018) was consistent with a role for illiquidity, the extended evidence should now be viewed as unsupportive.

[Figure 5 here: 'IS and OOS Predictive Performance of CEP **lzrt** (monthly)']

7. CGMS: Colacito, Ghysels, Meng, and Siwasarit (2016)

Abstract: *[CGMS] document that the first and third cross-sectional moments of the distribution of GDP growth rates made by professional forecasters can predict equity excess returns, a finding that is robust to controlling for a large set of well-established predictive factors...time-varying skewness in the distribution of expected growth prospects in an otherwise standard endowment economy can substantially increase the model-implied equity Sharpe ratios, and produce a large amount of fluctuation in equity risk premiums.*

Variable: CGMS kindly worked with us to isolate the cause for the difference between their data series and our own recalculation. The principal reason is that the data provided by the vendor are different than the data used by CGMS.

Performance: **[A]** We cannot confirm the significant IS coefficient of skew with the correct vendor data. Our own skew calculation shows no useful predictive ability.

Evaluation: We dismiss skew as a useful predictor of equity premia, due to irreproducibility.

8. CGP: Chava, Gallmeyer, and Park (2015)

Abstract: *[CGP analyze] U.S. stock return predictability using a measure of credit standards ('Standard') derived from the Federal Reserve Board's Senior Loan Officer Opinion Survey on Bank Lending Practices. Standards is a strong predictor of stock returns at a business cycle frequency, especially in the post-1990 data period. Empirically, a tightening of Standards predicts lower future stock returns. Standards perform well both in-sample and out-of-sample and is robust to a host of consistency checks. Standards captures stock return predictability at a business cycle frequency and is driven primarily by the ability of Standards to predict cash flow news.*

Variable: crdstd is as obtained from the survey data by the Fed. crdstd is available quarterly.

Performance: **[A]** We can confirm the positive and statistically significant IS coefficient of crdstd in the original sample period (–2013). However, in our extended sample (–2020), the IS T-statistic drops to 1.65. Moreover, the coefficient is also not climbing but falling, having declined from the first to the second half of the sample (albeit not statistically significantly so). **[B]** The OOS R^2 of crdstd is positive. **[C]** The investment performance of crdstd was mostly good. With

14

either scaling or equity-tilting, crdstd performed well, earning between 2%/year and 6%/year more than *all-equity-all-the-time*. Only in a no-scaling no-equity-tilt strategy did it modestly underperform *all-equity-all-the-time*. **[D]** Figure 6 shows that crdstd had great performance early on—predicting well from 2000 to mid-2002. Since 2003, crdstd performance has been unremarkable, with a short temporary spike around the time of the Great Recession (predicting Q1-Q2 2009).

Evaluation: We are concerned that crdstd has an IS T-statistic this low, and that practically all its good performance originates from its first four years in the sample. However, we consider crdstd worth watching. It is one of the variables mentioned in our introduction.

[Figure 6 here: 'IS and OOS Predictive Performance of CGP **crdstd** (quarterly)']

9. CP: Cooper and Priestley (2009)

Abstract: *[CP show that] the output gap, a production-based macroeconomic variable, is a strong predictor of U.S. stock returns. It is a prime business cycle indicator that does not include the level of market prices, thus removing any suspicion that returns are forecastable due to a "fad" in prices being washed away. The output gap forecasts returns both in-sample and out-of-sample, and it is robust to a host of checks...*

Variable: The output gap (ogap) is the deviation of the log of industrial production from a trend that incorporates both a linear and a quadratic term. ogap is available monthly.

Performance: **[A]** We can confirm the strong positive and statistically significant IS coefficient of ogap in the original sample period (–2005). In our extended sample (–2020), the IS coefficient is no longer statistically significant. The IS T-statistic is now –0.62. Thus, with poor IS performance, further OOS investigation seems unwarranted. (**[B]** The OOS R^2 of ogap is negative. **[C]** The investment performance of ogap was poor. It always underperforms *all-equity-all-the-time*.[12]) **[D]** Figure 7 shows that the IS performance was steady. However, the OOS performance early on was very poor, so the (unremarkable) improvements from 1950 to 2020 are insufficient to make much difference one way or another. The variable simply did not move much.

Evaluation: We dismiss ogap as a useful predictor of equity premia, based on its insignificant IS coefficient (and poor OOS performance). Presumably, if the evidence in Cooper and Priestley (2009) was consistent with a role for the output gap, the extended evidence should now be viewed as unsupportive

[Figure 7 here: 'IS and OOS Predictive Performance of CP **ogap** (monthly)']

10. DJM: Driesprong, Jacobsen, and Maat (2008)

Abstract: *[DJM show that] changes in oil prices predict stock market returns worldwide...These results cannot be explained by time-varying risk premia as oil price changes also significantly predict negative excess returns. Investors seem to underreact to information in the price of oil. A rise in oil prices drastically lowers future stock returns. Consistent with the hypothesis of a delayed reaction by investors, the relation between monthly stock returns and lagged monthly oil price changes strengthens once we introduce lags of several trading days between monthly stock returns and lagged monthly oil price changes.*

---

[12]The authors showed positive OOS significance, because they started predicting in 1948 rather than 1926.

15

Variable: wtexas is the price of West-Texas Intermediate crude oil, as obtained from *Global Financial Data* services. We also extend the sample backward from 1973, when Driesprong, Jacobsen, and Maat (2008) begin. wtexas is available monthly.

Performance: **[A]** We can confirm the strong positive and statistically significant IS coefficient of wtexas in the original sample period (–2004). In our extended sample (–2020), the IS coefficient is no longer statistically significant, with a T-statistic of –1.47. Thus, with poor IS performance, further OOS investigation seems unwarranted. (**[B]** The OOS R^2 of wtexas is negative (–0.12). **[C]** The investment performance of wtexas was inconsistent. When unscaled, it performed terribly, even losing money in absolute terms. When scaled, it performed about as well as *all-equity-all-the-time*, even beating it by a tiny 0.3% per year.) **[D]** Figure 8 shows that wtexas had good annual OOS R^2 performance in the 1973 oil crisis, specifically in Oct and Nov 1973, when the oil price went from \$3.51/b to \$13.37/b. It collapsed in June 2008, when the oil price dropped from \$139/b to \$39/b. The latter occurred just after Driesprong, Jacobsen, and Maat (2008) was published, which explains the difference between their results and our own.

Evaluation: We dismiss wtexas as a useful predictor of equity premia, based on its insignificant IS coefficient and poor OOS R^2. Presumably, if the evidence in Driesprong, Jacobsen, and Maat (2008) was consistent with models of delayed reaction by investors (offering simple high trading profits), the extended evidence should now be viewed as inconsistent

[Figure 8 here: 'IS and OOS Predictive Performance of DJM **wtexas** (monthly)']

11. HHT: Hirshleifer, Hou, and Teoh (2009)

Abstract: *[HHT] examine whether the firm-level accrual and cash flow effects extend to the aggregate stock market. In sharp contrast to previous firm-level findings, aggregate accruals is a strong positive time series predictor of aggregate stock returns, and cash flows is a negative predictor...These findings suggest that innovations in accruals and cash flows contain information about changes in discount rates, or that firms manage earnings in response to marketwide undervaluation.*

Variable: HHT introduce two variables: cfacc and accrul. The latter is the difference between earnings and cash flows. HHT use these variables only on annual frequency. For our purposes, it is important to recognize that the two variables are reported by corporations only a few months *after* the closing of their fiscal years. (Our Jan-to-Dec numbers assume no reasonable reporting lag.) Ergo, our focus are on the Jul-to-Jun numbers reported below, which are the only investable ones.

➤ The Accruals Component (accrul)

[Figure 9 here: 'Time-Series of Accruals (ac-crul) and Equity Premia']

Performance: **[A]** We can confirm the strong positive and statistically significant IS coefficient of accrul in the original sample period (–2005). Tables 5 and 6 shows that this also holds in our extended sample (–2020) and especially in our Jul-Jun mid-years. **[B]** Remarkably, accrul offers good OOS R^2, too. Somewhat unexpectedly, the OOS R^2 is higher than the IS R^2. **[C]** Only the untilted and unscaled timing strategy underperformed *all-equity-all-the-time* (Table 7). (Because of its stability (low standard deviation), with its negative investment performance, accrul also had the single-worst Sharpe ratio in our \$1 investment table.) However, as soon as accrul is scaled (Table 8) or tilted towards equity (Table 9), accrul timing outperforms *all-equity-all-the-time*. Intuitively, Both tilting and scaling place more emphasize on accrul's strong and decisive

16

calls from 1999–2001, with good prediction of the poor market performance in 2000–2002. **[D]** Figures 9 and 10 explain why accrul performed so well. Figure 9 shows that aggregate accruals were perennially quite flat, with two stark exceptions: 1973–1974 (conservative) and 1999–2001 (aggressive). The former occurred before our OOS analysis begins. Figure 10 shows that the latter was a great call. The market declined greatly in 2000–2002, following the dot-com years. In "ordinary years," aggregate accruals were unremarkable. They barely budged.

Evaluation: accrul is a difficult variable to assess due to its episodic performance.

One can share the view of HHT that managers' over-optimism or over-pessimism anticipated the (opposite) reversal of investors' sentiment in a particular kind of market exuberance followed by its predictable collapse. (Of course, corporate managers would have had to have the appropriate prescience, ignored by funds and other market participants.)

Or one can take the view that the 1999–2001 event was too singular a period to make it likely that accrul will help again predict equity premia in the future. (We will also briefly discuss below that a risk-averse investor would not want to use accrul for timing.)

[Figure 10 here: 'IS and OOS Predictive Performance of HHT **accrul** (annual/jun)']

➤ **The Cash Flow Component (cfacc)**

Performance: **[A]** We can confirm the strong positive and statistically significant IS coefficient of cfacc in the original sample period (–2005). The key problem for cfacc is that it performs well only if there is no reporting lag (Jan-Dec but not Jul-Jun). With a reporting lag, the IS T-statistic falls from –3.08 to –1.42. **[B]** The OOS R^2 of cfacc is negative in the investible Jul-Jun version.

Evaluation: We dismiss cfacc as a useful predictor of equity premia, based on poor IS and OOS performance in the investable Jul-to-Jun data set.

[Figure 11 here: 'IS and OOS Predictive Performance of HHT **cfacc** (annual/jun)']

12. HJTZ: Huang, Jiang, Tu, and Zhou (2015)

Abstract: *[HJTZ] propose a new investor sentiment index that is aligned with the purpose of predicting the aggregate stock market. By eliminating a common noise component in sentiment proxies, the new index has much greater predictive power than existing sentiment indices have both in and out of sample, and the predictability becomes both statistically and economically significant. In addition, it outperforms well-recognized macroeconomic variables and can also predict cross-sectional stock returns sorted by industry, size, value, and momentum. The driving force of the predictive power appears to stem from investors' biased beliefs about future cash flows.*

HJTZ can be viewed as combining the sentiment measure of Baker and Wurgler (2007), which was designed for the cross-section and not for market timing, with the in-sample optimization method of Kelly and Pruitt (2013).

Variable: sntm uses the Baker and Wurgler (2007) six sentiment variables, but weights them to optimize the predictive performance in sample using the technique pioneered in Kelly and Pruitt (2013).

[Figure 12 here: 'Time-Series of Sentiment (**sntm**) and Equity Premia']

Figure 12 plots the time-series of sntm. Sentiment was very pessimistic in 1968–1969, 1982, and 2000–2001; and very optimistic in 1974–1976. Oddly, sentiment does not have intuitive time-series behavior. Figure 12 shows that sntm was not particularly optimistic in 1998–1999

17

(though it did collapse later in 2001–2002), and that it has remained fairly steady throughout the Great Moderation, the Great Recession and Covid.

Performance: **[A]** We can confirm the strong positive and statistically significant IS coefficient of sntm in the original sample period (–2010). The T-statistic is 2.6 and 2.7 in our basic and extended sample. We note that this seems low, given the use of the Kelly and Pruitt (2013) in-sample optimization. (**[B]** The OOS R^2 of sntm is negative (–1.24). **[C]** sntm always underperforms *all-equity-all-the-time*. **[D]** Figure 13 explains when sntm performed well: Like accrul, sntm moves relatively little most of the time. However, the times when sntm did move are different from the times when accrul moved. sntm was very low, spiking down in 1969-70, 1982, and 2001-2. Only the latter is in our OOS R^2 sample. This was a good call.

Evaluation: Like accrul, sntm is a difficult variable to assess. Since 2002, sntm has not moved much. However, just before then, it appropriately predicted ongoing good performance from 1995 to 1998 and called the poor performance in 2001 and 2002.

13. JT: Jones and Tuzel (2013)

Abstract: *[JT] investigate the asset pricing and macroeconomic implications of the ratio of new orders (NO) to shipments (S) of durable goods. NO/S measures investment commitments by firms, and high values of NO/S are associated with a business cycle peak. We find that NO/S proxies for a short-horizon component of risk premia not identified in prior work. Higher levels of NO/S forecast lower excess returns on equities...at horizons from one month to one year. These effects are generally robust to the inclusion of common return predictors and are significant on an out-of-sample basis as well...*

Variable: ndrbl is the ratio of new orders to shipments of durable goods, obtained from the Census Bureau. Jones and Tuzel (2013) interpret their variable as a forecast of future investment growth. ndrbl is available monthly.

Performance: **[A]** We can confirm the strong negative and statistically significant IS coefficient of ndrbl in the original sample period (–2009). However, the IS coefficient has weakened, with declining H2 performance. In our extended sample (–2020), the IS coefficient is just barely statistically significant, with a T-statistic of –1.73. [13] **[B]** The OOS R^2 of ndrbl is negative. [13] **[C]** The investment performance of ndrbl was poor. **[D]** Figure 14 shows that much of the good **in-sample** performance was due to the performance in 1974–1975, i.e., predicting the oil-shock bear market. This was too early to be included in our OOS prediction. Since then, ndrbl has also been unremarkable in-sample. In the aftermath of the Tech collapse, ndrbl performed poorly, mispredicting 2001 and 2002, which ruined its OOS performance. Otherwise, with modest spikes in the Great Recession and Covid, ndrbl was mostly unremarkable.

---

[13]Our OOS periods always begin 20 years after a variable is available. In contrast, Jones and Tuzel (2013, Table 8) start after 5 or 10 years. Thus, with a data start of 1958, they still include the stellar oil-crisis 1974–1975 performance of ndrbl (see Figure 14), whereas we do not. Looking back to Table 2, the model coefficients also declined over their sample, though they did not do so in a statistically significant manner. JT also use quarterly frequency data for their OOS prediction, whereas we remain with the frequency of the main results, monthly. The reader can thus consider the OOS performance of JT to be sensitive rather than negative.

18

Evaluation: We dismiss ndrbl as a useful predictor of equity premia, based on its marginal IS significance and poor OOS performance.

14. JZZ: Jondeau, Zhang, and Zhu (2019)

Abstract: *[JZZ find that] average skewness, which is defined as the average of monthly skewness values across firms, performs well at predicting future market returns. This result still holds after controlling for the size or liquidity of the firms or for current business cycle conditions. [They] also find that average skewness compares favorably with other economic and financial predictors of subsequent market returns. The asset allocation exercise based on predictive regressions also shows that average skewness generates superior performance.*

Variable: skvw is as described in the authors' abstract. Note that this is *not* the average time-series skewness of the market index itself, but a cross-sectional skewness. skvw is available monthly.

Performance: **[A]** We can confirm the strong positive and statistically significant IS coefficient of skvw in the original sample period (–2016). In our extended sample (–2020), the IS coefficient is no longer statistically significant, with a T-statistic of 0.04. The model is unstable, as its IS coefficient switches from positive to negative in the second half. Thus, with poor IS performance, further OOS investigation seems unwarranted. **[B]** The OOS R^2 of skvw is negative. The decline in significance OOS in our extended sample relative to their original sample mirrors the decline in significance IS. **[C]** The investment performance of skvw was poor. The strategies not tilted towards equity even lost money in absolute terms. **[D]** Figure 15 shows that skvw always performed poorly.

Evaluation: We dismiss skvw as a useful predictor of equity premia, based on its poor IS and OOS performance. Presumably, if the evidence in Jondeau, Zhang, and Zhu (2019) was consistent with a role for average of individual skewnesses, the extended evidence should now be viewed as unsupportive

15. KJ: Kelly and Jiang (2014)

Abstract: *[KJ] propose a new measure of time-varying tail risk that is directly estimable from the cross-section of returns. [They] exploit firm-level price crashes every month to identify common fluctuations in tail risk among individual stocks. [The] tail measure is significantly correlated with tail risk measures extracted from S&P 500 index options and negatively predicts real economic activity. We show that tail risk has strong predictive power for aggregate market returns.*

Variable: tail is as described in the authors' abstract. Note that tail is *not* the tail risk of the market index itself, but a cross-sectional statistic. tail is available monthly.

Performance: **[A]** We can confirm the positive and statistically significant IS coefficient of tail in the original sample period (–2010). In our extended sample (–2020), the IS coefficient is no longer statistically significant, with a T-statistic of 0.21. The model is also unstable, with the IS coefficient turning from negative in the first half to positive in the second half. Thus, with poor IS performance, further OOS investigation seems unwarranted. **[B]** The OOS R^2 of tail is negative. The decline in significance OOS in our extended sample relative to their original

19

sample mirrors the decline in significance IS. **[C]** The investment performance of tail was poor. **[D]** tail offered no remarkable performance or episodes. The variable barely budged and the predictive coefficient was not large, which is why the performance relative to the equity premium remains rather flat and unremarkable.

Evaluation: We dismiss tail as a useful predictor of equity premia, based on its poor IS and OOS performance. Presumably, if the evidence in Kelly and Jiang (2014) was consistent with a role for a new measure of risk, the extended evidence should now be viewed as inconsistent

[Figure 16 here: 'IS and OOS Predictive Performance of KZ **tail** (monthly)']

16.   KP: Kelly and Pruitt (2013)

Abstract: *[KP find that] returns and cash flow growth for the aggregate U.S. stock market are highly and robustly predictable. Using a single factor extracted from the cross-section of book-to-market ratios, [they] find an out-of-sample return forecasting R^2 of 13% at the annual frequency (0.9% monthly)... We present a model linking aggregate market expectations to disaggregated valuation ratios in a latent factor system. Spreads in value portfolios' exposures to economic shocks are key to identifying predictability and are consistent with duration-based theories of the value premium.* Oliveira (2022 exp) show that Kelly and Pruitt (2013) is sensitive to a number of implementation choices, although this is in turn disputed by Kelly and Pruitt (2022 exp).

Variable: fbm constructs its variable based on a partial least squares (PLS) technique that extracts a latent factor most relevant for predicting returns by exploiting the relationship between the predictors and the returns being forecast. fbm is an outlier in terms of its predictive IS performance, because it was optimized (fitted in PLS) based on ex-post data to do so. fbm is available monthly.

Performance: **[A]** We can confirm the strong positive and statistically significant IS coefficient of fbm in the original sample period (–2010). Note that fbm is optimized to maximize the IS performance. Thus, it also performs very well in our extended sample. **[B]** The OOS R^2 of fbm is negative. Further unreported analysis shows that the discrepancy in reported OOS between KP and us can be traced to three issues: [1] their OOS prediction started in 1980, our's in 1946 (for consistency across all papers); [2] their OOS prediction ended in 2010, our's in 2020; and [3] they predict (log) market returns, we predict (log) equity premia. Each of these matters and together they explain why they have a positive OOS R^2 and we have a negative one. **[C]** The investment performance of fbm was poor. **[D]** Figure 17 shows consistently inferior OOS performance over the entire sample period. (It makes no sense to plot fbm IS, because it is highly optimized in this respect.) This suggests that fbm is simply overfitted.

Evaluation: We dismiss fbm as a useful predictor of equity premia, based on its poor OOS performance. Presumably, if the evidence in Kelly and Pruitt (2013) was consistent with high and robust predictive ability and a model linking aggregate market expectations to disaggregated valuation ratios (with spreads in value portfolios being key together with duration-based theories of the value premium), the extended evidence should now be viewed as inconsistent

[Figure 17 here: 'IS and OOS Predictive Performance of KP **fbm** (monthly)']

20

17. LY: Li and Yu (2012)

Abstract: *Motivated by psychological evidence on limited investor attention and anchoring, [LY] propose two proxies for the degree to which traders under- and overreact to news, namely, the nearness to the Dow 52-week high and the nearness to the Dow historical high, respectively. [LY] find that nearness to the 52-week high positively predicts future aggregate market returns, while nearness to the historical high negatively predicts future market returns....*

Variable: LY introduce two variables: dtoy and dtoat. The former is the scaled current difference to the 52-week high of the Dow Jones index, the latter is the current difference to the lifetime high. Because the Dow-Jones was mostly moving up, the distance was often near its maximum of 1. The variables are available monthly.

➤ Distance to Historical Maximum Price of the Dow-Jones index (dtoy)

Performance: **[A]** We can confirm the positive and statistically significant IS coefficient of dtoy in the original sample period (–2009). In our extended sample (–2020), the IS coefficient is no longer statistically significant, with a T-statistic of 0.40. Moreover, the model seems unstable. The coefficient turns from positive to negative in the second half. Thus, with poor IS performance, further OOS investigation seems unwarranted. **[B]** The OOS R^2 of dtoy is negative. **[C]** The investment performance of dtoy was poor. Unless equity-tilted, dtoy not only underperforms *all-equity-all-the-time*, it even loses money in absolute terms. When equity-tilted, it "merely" underperforms *all-equity-all-the-time*.)

Evaluation: We dismiss dtoy as a useful predictor of equity premia, based on poor IS and OOS performance. Presumably, if the evidence in Li and Yu (2012) was consistent with models of psychological evidence on limited investor attention and anchoring, the extended evidence should now be viewed as inconsistent.

➤ Distance to Maximum Price Lifetime (dtoat)

Performance: **[A]** We can confirm the strong negative and statistically significant IS coefficient of dtoat in the original sample period (–2009). In our extended sample (–2020), the IS coefficient is no longer statistically significant, with a T-statistic of –0.32. Thus, with poor IS performance, further OOS investigation seems unwarranted. **[B]** The OOS R^2 of dtoat is negative. **[C]** The investment performance of dtoat was poor. dtoat not only underperforms *all-equity-all-the-time*, it even loses money in Tables 7, 8, and 9. It "merely" underperforms *all-equity-all-the-time* in Table 9, the equity-tilted unscaled strategy.)

Evaluation: We dismiss dtoat as a useful predictor of equity premia, based primarily on poor IS and OOS performance. Presumably, if the evidence in Li and Yu (2012) was consistent with models of psychological evidence on limited investor attention and anchoring, the extended evidence should now be viewed as inconsistent.

18. Maio(13): Maio (2013)

Abstract: *The focus of this article is on the predictive role of the stock-bond yield gap—the difference between the stock market earnings (dividend) yield and the 10-year Treasury bond yield—also known as the "Fed model". The results show that the yield gap forecasts positive excess market returns...the yield gap has reasonable out-of-sample predictability for the equity premium when the comparison is made against a simple historical average, especially when one imposes a restriction of positive equity premia....An investment strategy based on the forecasting ability of the yield gap produces significant gains in Sharpe ratios.*

Variable: Maio (2013) calculates the Fed model as the dividend-price ratio net of the 10-year government bond yield, the latter multiplied by 10. We use a corrected definition, which removes the multiplication factor. ygap is available monthly.

Performance: [A] We can confirm the strong positive and statistically significant IS coefficient of ygap in the original sample period (–2008), though with lower statistical significance. In our extended sample (–2020), the IS coefficient is no longer statistically significant, with a T-statistic of 0.67. Thus, with poor IS performance, further OOS investigation seems unwarranted. ([B] The OOS R^2 of ygap is negative. [C] The investment performance of ygap was poor, with three out four strategies not just not beating *all-equity-all-the-time*, but losing money in absolute terms.)

Evaluation: We dismiss ygap as a useful predictor of equity premia, based on poor IS and OOS performance. Presumably, if the evidence in Maio (2013) was consistent with the Fed Model as a predictor, the extended evidence should now be viewed as inconsistent.

19. Maio(16): Maio (2016)

Abstract: *[Maio] examines whether stock return dispersion (RD) provides useful information about future stock returns. RD consistently forecasts a decline in the excess market return at multiple horizons, and compares favorably with alternative predictors used in the literature. The out-of-sample performance of RD tends to beat the alternative predictors, and is economically significant as indicated by the certainty equivalent gain associated with a trading investment strategy.*

Variable: rdsp is the cross-sectional standard deviation on the set of 100 size and book-to-market portfolios. rdsp is available monthly.

Performance: **[A]** We can confirm the strong positive and statistically significant IS coefficient of rdsp in the original sample period (–2013). In our extended sample (–2020), the IS coefficient is no longer statistically significant, with a T-statistic of 0.74. The predictive sign has changed between the first and second halves of our sample, too. Thus, with poor IS performance, further OOS investigation seems unwarranted. (**[B]** The OOS R^2 of rdsp is negative. The decline to significance OOS in our extended sample relative to their original sample mirrors the decline in significance IS. **[C]** The investment performance of rdsp was poor. In fact, it was so poor that all four strategies not only did not beat *all-equity-all-the-time*, three all lost money in absolute terms, the fourth broke even.)

Evaluation: We dismiss rdsp as a useful predictor of equity premia, based on its poor IS and OOS performance. Presumably, if the evidence in Maio (2016) was consistent with a role for useful information in stock dispersion, the extended evidence should now be viewed as inconsistent.

20. Mrtn: Martin (2017)

Abstract: *[Mrtn] uses the SVIX index as a proxy for the equity premium and argues that the high equity premia available at times of stress largely reflect high expected returns over the very short run.*

The relationship between squared volatility and returns makes sense, e.g., in a CRRA framework. We note that equity options had been used to establish bounds on the equity premium in Martin (2011), Backus, Chernov, and Zin (2014), Welch (2016) and Seo and Wachter (2019). Remarkably, Martin (2017) was first to test whether the constraint could be binding, i.e., whether the implied volatility bounds could be related directly to the equity premium.

Variable: Mrtn measures his predictor rsvix as the risk-neutral variance index, SVIX^2. The 1-month rsvix has 99.5% correlation with the more common squared CBOE volatility index (VIX), which is also based mostly on a 1-month horizon. Mechanically, predicting equity premia with the 1-month rsvix index is therefore functionally equivalent to predicting it with the squared VIX. As a volatility index, rsvix is available on a daily basis. Mrtn uses different horizon rsvix to predict different horizon equity premia: 1-month, 2-month, 3-month, 6-month, and 12-month ahead equity premia. (The longer horizon rsvix series still has 94% correlation with the squared 1-month VIX squared and performs roughly as well in predicting future equity premia as the same-horizon rsvix numbers.)

Performance: Martin (2017) did not cherry pick its presentation: Table II shows that rsvix has predictive in-sample power *only* on the 6-month horizon, but not on monthly, bi-monthly, quarterly, and annual horizons. We have to deviate from our standard presentation scheme, because we cannot focus on monthly, quarterly, or annual regressions as we could for other papers. (They merely confirm the insignificant performance in Martin (2017).)

[Table 12 here: 'Sensitivity of Martin (2017)']

Panel A of Table 12 walks through how our specification differs from Martin's. It matters little whether one uses daily or monthly frequency observations, and/or log or simple equity premium (though results may drop in or out of statistical significance), and/or a sample ending in 2012 or 2020. The real differences are elsewhere.

First, the coefficients are sensitive whether one uses overlapping or non-overlapping observations. With non-overlapping observations, all in-sample coefficients including those on the 6-month horizon turn insignificant in the original sample. In the extended sample, the quarterly but not the semi-annual turn significant.

Second, there is an important difference in how we calculate OOS R^2 for all our variables: Our predictions always take variables, run predictive regressions on the prevailing history of the equity premium, and then predict the future equity-premium with the resulting regression. In contrast, under his theory, Martin's variable is already a meaningful expected return. Thus he uses rsvix directly as the forecast (equivalent if we set our predictive regression intercept to 0 and the predictive slope to 1). His direct prediction is *always* better than our regression-intermediated prediction. However, even his OOS performance is never statistically significant.

Panel B further investigates the performance on his preferred semi-annual horizon. The rsvix variable has no predictive power when used to predict semi-annual equity premia in December

24

and June for first and second halves of calendar years, respectively. It only has good performance when used to predict in March (i.e., at the end of Q1 for equity premia in Q2 and Q3) and September (at the end of Q3 for Q4 and Q1).

The test intent of Martin (2017) is intrinsically different from our own. Our premise is that a variable has to have a coefficient different from 0. In a loose sense, we ask variables to prove that they are better than nothing. The null hypothesis receives 95%, the variable has to beat the 5%. Martin's premise is that rsvix has a coefficient of 1. In a loose sense, he asks skeptics to prove that it is not 1. The variable receives the 95%, the alternative gets the 5%.

This makes sense from the perspective that the theory is correct, but it is empirically weak. That is, most of the regression coefficients can similarly not reject (atheoretical) hypotheses that the coefficient on rsvix is -1, 0, or +3, just as they cannot reject that it is 1. With a good prior in favor of the theory, most of the regression evidence is not contradictory. There is one ironical exception not shown in the table. The theory *can* be rejected when rsvix had the best performance against the null hypothesis that it is 0. Reverting back to simple rates of return, and using Mar/Sep semi-annual prediction, the hypothesis that the coefficient is 1.0 can be rejected with a T-statistic of 2.3 ending in 2012 and 2.4 ending in 2020. Again, with a strong prior, this evidence can be ascribed to sampling.

Figure 18 plots the performance both IS and OOS based on the only specification with good performance, i.e., the Mar/Sep semiannual predictions. The figure shows that rsvix' positive OOS performance is based on three months: 2009/03 (predicting to 2009/09), 2011/09 (to 2012/03) and 2020/03 (to 2020/09).[14] The 2009/03 performance merely undoes the preceding poor 2008/09 prediction. However, the two other observations are principally responsible. Removing either the 2011/09 or 2020/03 prediction both drops the OOS performance into the red and removes the statistical significance of the IS regression. As noted in Section 11., it should be left to the reader how to think about outliers.

Evaluation: With respect to the null hypothesis that the coefficient is zero, we dismiss rsvix as useful general predictors of equity premia, primarily based on the sensitivity and specificity in performance. Stock-market volatility fails to predict equity premia on monthly, bimonthly, quarterly, annual, and standard semi-annual frequencies. Even on its only good predictive specification (the semi-annual frequency with offset), its good (but not statistically significant) performance was due to two specific outliers.

---

[14]The IS performance further benefited from its prediction of the 1999/09 (to 2001/03).

25

21. MR: Møller and Rangvid (2015)

Abstract: *[MR] show that macroeconomic growth at the end of the year (fourth quarter or December) strongly influences expected returns on risky financial assets, whereas economic growth during the rest of the year does not. We find this pattern for many different asset classes, across different time periods, and for US and international data.*

It is worth noting that the paper's perspective that the fourth quarter data is special and was motivated by Jagannathan and Wang (2007).

Variable: MR introduce two variables: gpce and gip. The former is the growth rate in personal consumption expenditures, the latter is the growth rate in industrial production. The variables are available on a quarterly basis, but Møller and Rangvid (2015) use them only on an annual basis, presumably due to the special fourth-quarter perspective.

➤ The Growth Rate in Personal Consumption Expenditures (gpce)

Performance: **[A]** We can confirm the strong negative and statistically significant IS coefficient of gpce in the original sample period (–2009). In our extended sample (–2020), the IS coefficient still has a T-statistic of –3.40. **[B]** The OOS R^2 of gpce is positive, with an OOS R^2 of 4.86% (not as good as the IS R^2 of 11.91%, but quite respectable). **[C]** Except for the untilted, unbiased timing strategy (Table 7), the gpce-based timing strategies outperformed *all-equity-all-the-time* (Tables 8-10). **[D]** Figure 19 plots the time-series of gpce. There are no obvious patterns. In recessions, gpce declines. Figure 20 shows that gpce had good performance beginning with its 1974 prediction for 1975, with the only misprediction being its bad 2007 call for the 2008 Great Recession.

Evaluation: **The gpce (fourth-quarter) variable was the best equity-premium predictor in our sample:** Years in which consumers spent more are followed by bear stock markets. gpce also performed well after the original authors' sample period had ended. The only caveat is that the reader must assess whether the ex-post choice of the fourth quarter raises data-snooping as a concern.

[Figure 19 here: 'Time-Series of Personal Expenditures Growth (**gpce**) and Equity Premia']

[Figure 20 here: 'IS and OOS Predictive Performance of MR **gpce** (annual/jun)']

➤ The Growth Rate in Industrial Productions (gip)

Performance: **[A]** We can confirm the strong negative and statistically significant IS coefficient of gip in the original sample period (–2009). In our extended sample, the IS coefficient is no longer statistically significance, with its T-statistic of –0.10. Thus, with poor IS performance, further OOS investigation seems unwarranted. (**[B]** The OOS R^2 of gip is negative. **[C]** The investment performance of gip was poor.)

Evaluation: We dismiss gip as a useful predictor of equity premia, based on poor IS and OOS performance.

However, the poor performance of gip is less driven by the fact that our sample extends forward to 2020 and more by the fact that gip (unlike gpce) has been available since 1926. Remaining consistent with our treatment of other variables, our paper uses the entire data, while the authors focus on the shared sample beginning in 1947. After 1947, gip had better performance.

22. NRTZ: Neely, Rapach, Tu, and Zhou (2014)

Abstract: *Technical indicators display statistically and economically significant in-sample and out-of-sample predictive power, matching or exceeding that of macroeconomic variables.*

Variable: tchi is the first principal component of 14 technical indicators, themselves principally versions of moving price averages, momentum, and ("on-balance") dollar-trading volume. tchi is available monthly.

Performance: **[A]** We can confirm the modest positive and statistically significant IS coefficient of tchi in the original sample period (–2011). In our extended sample (–2020), the IS coefficient is no longer statistically significant, with a T-statistic of 1.61. Thus, it is unclear whether further investigation is warranted. (**[B]** The OOS R^2 of tchi is positive. **[C]** The timing investment performance was good when a strong equity-tilt is maintained and poor otherwise.) **[D]** Figure 21 shows that tchi predicted well in late 2008 and early 2009. It reached its very brief high point in the Great Recession, i.e., in Feb 2009, predicting Mar 2009. Since then, tchi has consistently underperformed. This also explains why our findings differ from those in NRTZ, which ended in 2011.

Evaluation: We dismiss tchi as a useful predictor of equity premia, based on its marginal IS performance and consistently poor performance since the Great Recession.

[Figure 21 here: 'IS and OOS Predictive Performance of NRTZ tchi (monthly)']

23. PST: Piazzesi, Schneider, and Tuzel (2007)

Abstract: *[PST] consider a consumption-based asset pricing model where housing is explicitly modeled both as an asset and as a consumption good...the model predicts that the housing share can be used to forecast excess returns on stocks. We document that this indeed true in the data. The presence of composition risk also implies that the riskless rate is low which further helps the model improve on the standard CCAPM.*

Variable: house is a measure of the dollar amount spent on rent or estimates of how much owners would rent their houses for. (The paper shows that effective rents as a fraction of income declined from 1930s to the 1980s and then stabilized.) house is available annually.

Performance: **[A]** We can confirm the strong positive and statistically significant IS coefficient of house in the original sample period (–2001). In our extended sample (–2020), the IS coefficient is no longer statistically significant, with a T-statistic of 0.99 without reporting delay and 0.63 with. The model is also unstable now, with a positive coefficient in the first half and a negative coefficient in the second half. Thus, with poor IS performance, further OOS investigation seems unwarranted. (**[B]** The OOS R^2 of house is positive. **[C]** The investment performance of house was poor. The non-equity-tilted strategies not only did not beat *all-equity-all-the-time*, they lost money in absolute terms.)

Evaluation: We dismiss house as a useful predictor of equity premia, based primarily on its poor IS performance. Presumably, if the evidence in Piazzesi, Schneider, and Tuzel (2007) was consistent with a CCAPM with housing, the extended evidence should now be viewed as inconsistent

27

24. PW: Pollet and Wilson (2010)

Abstract: ...[PW show that] higher aggregate risk can be revealed by higher correlation between stocks. [PW] show that the average correlation between daily stock returns predicts subsequent quarterly stock market excess returns.

Variable: avgcor is the average correlation among the 500 largest stocks (by capitalization). The daily pairwise correlations of stock returns are multiplied by the product of both stock's weights relative to total sample market capitalization, then summed to create the measure. avgcor is available monthly. Because the authors remark that avgcor performs better on a quarterly rather than monthly frequency, we look at both.

Performance: **[A]** We can confirm the positive and statistically significant IS coefficient of avgcor in the original sample period (–2007). In our extended sample (–2020), the IS coefficient is no longer statistically significant on a monthly frequency, with a T-statistic of 0.89 (Table 3). The quarterly performance is indeed better, but the T-statistic is still only 1.43 (Table 4). Thus, with poor IS performance, further OOS investigation seems unwarranted. (**[B]** The OOS R^2 of avgcor is positive on a monthly frequency. On a quarterly frequency, it further increases, but it is still not statistically significant. **[C]** The investment performance of avgcor was poor.)

Evaluation: We dismiss avgcor as a useful predictor of equity premia, based primarily on its poor IS performance.

25. RRZ: Rapach, Ringgenberg, and Zhou (2016)

Abstract: [RRZ] show that short interest is arguably the strongest known predictor of aggregate stock returns. It outperforms a host of popular return predictors both in and out of sample, with annual R2 statistics of 12.89% and 13.24%, respectively. In addition, short interest can generate utility gains of over 300 basis points per annum for a mean-variance investor... Overall, our evidence indicates that short sellers are informed traders who are able to anticipate future aggregate cash flows and associated market returns. (This hypothesis further requires that other intelligent investors ignore publicly available short interest information.)

Variable: shtint is the aggregate short interest in the stock market, calculated as the log of the equal-weighted mean of short interest (as a percentage of shares outstanding) across publicly listed US stocks. shtint is available monthly.

Performance: **[A]** We can confirm the negative and statistically significant IS coefficient of shtint in the original sample period (–2014). In our extended sample (–2020), the IS coefficient is no longer statistically significant, with a T-statistic of –1.51. Thus, with poor IS performance, further OOS investigation seems unwarranted. (**[B]** The OOS R^2 of shtint is positive. **[C]** The investment performance of shtint was just about the same as *all-equity-all-the-time*, with scaled versions beating it by about 1.5% per year, non-scaled versions losing to it by about the same amount.) **[D]** Figure 22 shows the performance pattern of shtint: it did well from about mid-2008 to mid-2011. Otherwise, shtint was mostly unremarkable.

Evaluation: We dismiss shtint as a useful predictor of equity premia, based on very mediocre IS performance, as well as single-episode good performance during the Great Recession, only.

28

Presumably, if the evidence in Rapach, Ringgenberg, and Zhou (2016) was consistent with short investors being better informed and able to anticipate the market (*and* other traders ignoring the publicly available information in shtint), the extended evidence should now be viewed as unsupportive

26. Y: Yu (2011)

Abstract: *[Y] provides evidence that portfolio disagreement measured bottom-up from individual-stock analyst forecast dispersion...is negatively related to ex post expected market return. Contemporaneously, an increase in market disagreement manifests as a drop in discount rate. These findings are consistent with asset pricing theory incorporating belief dispersion.*

Variable: disag is the dispersion of earnings-per-share long-term growth rate forecasts by analysts from the I/B/E/S data base, value-weighted across stocks. disag is available monthly.

Performance: **[A]** We can confirm the negative and statistically significant IS coefficient of disag in the original sample period (–2005) In our extended sample (–2020), the IS coefficient is no longer statistically significant, with a T-statistic of 0.06. The model is also unstable, with a coefficient sign change between the first and second half of the sample. Thus, with poor IS performance, further OOS investigation seems unwarranted. (**[B]** The OOS R^2 of disag is negative. **[C]** The investment performance of disag was poor. However, the equity-tilted strategies outperform *all-equity-all-the-time*.) **[D]** Figure 23 shows that after good performance from mid-2008 to early 2009 (the early Great Recession), disag has performed poorly.

Evaluation: We dismiss disag as a useful predictor of equity premia, based on its poor IS and OOS performance. Presumably, if the evidence in Yu (2011) was consistent with a role for market disagreement, the extended evidence should now be viewed as inconsistent

27. Variables in Goyal and Welch (2008)

Our paper is a good opportunity to revisit the 17 variables from Goyal and Welch (2008). Since publication, 15 years have passed, allowing to investigate not just "pretend"-OOS performance, but true OOS performance. Tables 3–8 thus include the relevant IS and OOS performance and investment statistics, too.

Most variables in Goyal and Welch (2008) were of monthly frequency.

[A] On a monthly frequency, of the fourteen variables, only three variables (e/p, tbl, and ltr) show IS statistical significance—with the largest T-statistic being a scant 1.72. Of these three variable, the e/p IS coefficient drops from 0.97 to 0.04 across the two sample halves; tbl drops from –0.50 to –0.26; and only ltr increases (from 0.16 to 0.27). However, our enthusiasm is tempered by the fact that ltr's overall IS coefficient is only 1.53. **[B]** Moreover, among the three, only tbl has positive OOS R^2.[15] **[C]** All predictors underperformed *all-equity-all-the-time* on the untilted unscaled investment strategy (Table 7). The only meaningful investment improvement came from ltr in the scaled strategies (Table 8 with 1.4%/year and 10 with

---

[15]Indeed, only tbl and infl have statistically significant OOS statistics among all monthly variables, ignoring whether IS statistics are significant or not.

2.0%/year). **[D]** Figure 24 shows that even the best variable, ltr, was largely unremarkable and due to a good run long ago (from 1970 to 1987) that neutralized most of its bad run from 1950 to 1970. Figure 25 shows that most of the good performance of the second variable (tbl) dates back to the 1974–1975 oil-shock episode.

<u>Evaluation:</u> None of the variables show good predictive power on monthly horizons.

Goyal and Welch (2008) examined only two quarterly variables, i/k and cay. cay performs poorly both IS and OOS, as well as for investment purposes. i/k however has good IS and OOS performance. Figure 26 shows that it has improved since Goyal and Welch (2008). Nevertheless, it did not help an investor outperform *all-equity-all-the-time* on any of our four timing strategies.

Goyal and Welch (2008) examined only one annual variable, eqis. It still performs well IS, but has negative OOS R^2.

Not shown, since Goyal and Welch (2008), the OOS performance of 13 out of the 17 variables has further deteriorated, the performance of three variables (ltr, lty, and tbl) have increased very slightly, and only i/k improved visibly (Figure 26). The overall tally is worse than chance, which would have had 8 out 17 perform better.

IV Risk-Averse Investors' Certainty Equivalence

We now take a further look at the predictors from the perspective of a risk-averse investor. We focus only on the most promising variables from the previous tables.[16]

Our principal metric is

$$\Delta\text{UTIL} = \left(\hat{\mu}_{\text{Mdl}} - \frac{\gamma}{2}\hat{\sigma}^2_{\text{Mdl}}\right) - \left(\hat{\mu}_{\text{Unc}} - \frac{\gamma}{2}\hat{\sigma}^2_{\text{Unc}}\right),$$

where $\hat{\mu}_{\text{Mdl}}$ and $\hat{\sigma}^2_{\text{Mdl}}$ are the realized mean and variance of the portfolio of conditional strategy, and $\hat{\mu}_{\text{Unc}}$ and $\hat{\sigma}^2_{\text{Unc}}$ are the corresponding statistics for the unconditional strategy (or all-equity-all-the-time strategy). We set the risk-aversion coefficient γ to be 5 and impose investment limits of 0 and 1.5 times equity, as suggested by Campbell and Thompson (2008) and enhanced by Löffler (2022 (exp)). We calculate the statistical significance of the utility difference following the procedure in footnote 16 of DeMiguel, Garlappi, and Uppal (2009).

Table 11 shows that the standout performers for such an investor would have been crdstd on a quarterly basis and i/k on a quarterly or annual basis, followed closely by lzrt (annual), shtint (monthly) and tchi (monthly). The latter can beat unconditional investment strategies, though not always statistically significantly so. Remarkably, despite its good performance for a risk-neutral investor, accrul (and gpce) would not have helped a risk-averse investor. The additional realized variance would have been bad enough to negate the advantage of the improved mean prediction.

---

[16]Two variables, crdstd and i/k, were useful to a risk-averse investor, not because they predicted future equity premia, but because they reduced the variance in realized returns.)

30

V Conclusion

As noted in the introduction, only one variable (crdstd) outperformed the *all-equity-all-the-time* investment strategy with the simplest naive long-short market timing strategy ("$1") and it did so by only 0.2% per year. A few did better on equity-tilted and/or scaled investment strategies. Yet, none could beat *all-equity-all-the-time* in a statistically significant way.

Nevertheless, as already noted in the introduction, some variables (especially gpce) showed good performance on other dimensions. Even though we required candidates to pass more than just one criterion in predicting the equity premium, it still seems underwhelming that only a handful of variables—even with reuse of the original identifying data—succeeded. These were, after all, variables from high-quality papers important enough to have been published in the top academic journals—with manuscript rejection rates has high as 18-19 in 20 papers.

We want to end our paper by taking the liberty to voice some more subjective concerns.

We absolutely do not want to imply that the authors of the papers we examined here made inappropriate choices. Instead, we are inclined to agree with Lo and MacKinlay (1990) and Harvey, Liu, and Zhu (2016), who worry primarily about our collective academic research enterprise. For every predictive variable stumbled upon and published by a lucky researcher, there are probably hundreds that failed and were never published.

Our published results have conveyed a distorted picture of reality, perhaps more obvious to participating producers than to outside consumers. Readers, referees, editors, and journals like papers with impressive results. Moreover, academic finance audiences like impressive results even more so if the results can be justified based on a "strong theoretical basis." Not surprisingly, motivated authors often obligingly offer them. Many of our papers seemed remarkably confident in expressing *strong support* of "theory,"[17] regardless of whether these theories are neoclassical or behavioral. However, the presence of these theories seems not to have offered the desired solid and stable forward-looking performance that theory is intended to provide.

In addition to the problem that academic research gives the wrong impression, i.e., that it is possible or even easy to predict the stock market, there is the secondary problem of crowding out. Authors that write more mundane papers, which fail to show remarkable powers, are likely not to be published and thus disappear from the academic rat-race. The incentives imposed by our collective on its members are clear.

Many of the reexamined papers' claims seem (to us) to defy common sense. The more aggressive models were promising unusually high timing rates of returns to their readers, based on exploiting the ignorance or strange risk preferences of ordinary investors. Yet, easy profits forward-looking seems absurd to us. *It has not been easy for a very long time (Fama (1970)) to predict the equity premium or the performance of large publicly-traded stocks.* Large U.S. stocks are traded in highly competitive markets, with even the smartest funds struggling to perform well.

The somewhat less aggressive adherents of risk-factor based models—though the nature and the measure of the "risks" in the factors are typically themselves a mystery—have also

---

[17]For example, papers "point to the importance of theory x", "delayed reaction by investors", "agents recognize market-wide undervaluation," "investors' biased beliefs," "key to identifying predictability," "psychological evidence," "arguably the strongest predictor."

31

underperformed more often than not. A long procession of academics who have been involved in market-timing and/or stock-selection based funds can attest to it.

We remain comfortable with the original claims in Goyal and Welch (2008). Standing here today in 2021, even as risk-neutral investors willing to take on more risk, we do not believe that we know what variables should help us today to predict the equity premium forward-looking for 2022.

References

Adrian, Tobias, Emanuel Mönch, and Hyun Song Shin (2010). *Financial intermediation, asset prices, and macroeconomic dynamics*. Tech. rep. Staff Report 422.

Atanasov, Victoria, Stig Vinther Møller, and Richard Priestley (2020). "Consumption fluctuations and expected returns". In: *The Journal of Finance* 75.3, pp. 1677–1713.

Backus, David, Mikhail Chernov, and Stanley Zin (2014). "Sources of entropy in representative agent models". In: *The Journal of Finance* 69.1, pp. 51–99.

Baker, Malcolm and Jeffrey Wurgler (2000). "The equity share in new issues and aggregate stock returns". In: *The Journal of Finance* 55.5, pp. 2219–2257.

— (2007). "Investor sentiment in the stock market". In: *Journal of Economic Perspectives* 21.2, pp. 129–152.

Bakshi, Gurdip, George Panayotov, and Georgios Skoulakis (2011). "Improving the predictability of real economic activity and asset returns with forward variances inferred from option portfolios". In: *Journal of Financial Economics* 100.3, pp. 475–495.

Belo, Frederico and Jianfeng Yu (2013). "Government investment and the stock market". In: *Journal of Monetary Economics* 60.3, pp. 325–339.

Benjamini, Yoav and Yosef Hochberg (1995). "Controlling the false discovery rate: A practical and powerful approach to multiple testing". In: *Journal of the Royal Statistical Society. Series B (Methodological)* 57.1, pp. 289–300.

Benjamini, Yoav and Daniel Yekutieli (2001). "The control of the false discovery rate in multiple testing under dependency". In: *Annals of Statistics* 29.4, pp. 1165–1188.

Bollerslev, Tim, George Tauchen, and Hao Zhou (2009). "Expected stock returns and variance risk premia". In: *Review of Financial Studies* 22.11, pp. 4463–4492.

Boudoukh, Jacob, Roni Michaely, Matthew Richardson, and Michael R. Roberts (2007). "On the importance of measuring payout yield: Implications for empirical asset pricing". In: *The Journal of Finance* 62.2, pp. 877–915.

Campbell, John Y. (1987). "Stock returns and the term structure". In: *Journal of Financial Economics* 18.2, pp. 373–399.

Campbell, John Y. and Robert J. Shiller (1988). "Stock prices, earnings, and expected dividends". In: *The Journal of Finance* 43.3, pp. 661–676.

Campbell, John Y. and Samuel B. Thompson (2008). "Predicting excess stock returns out of sample: Can anything beat the historical average?" In: *Review of Financial Studies* 21.4, pp. 1509–1531.

Cederburg, Scott, Travis L. Johnson, and Michael S. O'Doherty (2019). *On the Economic Significance of Stock Return Predictability*. Tech. rep. SSRN 3372629.

Chava, Sudheer, Michael Gallmeyer, and Heungju Park (2015). "Credit conditions and stock return predictability". In: *Journal of Monetary Economics* 74, pp. 117–132.

Chen, Yong, Gregory W. Eaton, and Bradley S. Paye (2018). "Micro (structure) before macro? The predictive power of aggregate illiquidity for stock returns and economic activity". In: *Journal of Financial Economics* 130.1, pp. 48–73.

Cochrane, John H. (1991). "Production-based asset pricing and the link between stock returns and economic fluctuations". In: *The Journal of Finance* 46.1, pp. 209–237.

Colacito, Riccardo, Eric Ghysels, Jinghan Meng, and Wasin Siwasarit (2016). "Skewness in expected macro fundamentals and the predictability of equity returns: Evidence and theory". In: *Review of Financial Studies* 29.8, pp. 2069–2109.

Cooper, Ilan and Richard Priestley (2009). "Time-varying risk premiums and the output gap". In: *Review of Financial Studies* 22.7, pp. 2801–2833.

DeMiguel, Victor, Lorenzo Garlappi, and Raman Uppal (2009). "Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy?" In: *Review of Financial Studies* 22.5, pp. 1915–1953.

Diebold, Francis and Roberto Mariano (1995). "Comparing predictive accuracy". In: *Journal of Business and Economic Statistics* 13.3, pp. 253–263.

Driesprong, Gerben, Ben Jacobsen, and Benjamin Maat (2008). "Striking oil: another puzzle?" In: *Journal of Financial Economics* 89.2, pp. 307–327.

Fama, Eugene F. (May 1970). "Efficient Capital Markets: A Review of Theory and Empirical Work". In: *The Journal of Finance* 25.2, pp. 383–417.

Fama, Eugene F. and Kenneth R. French (1989). "Business conditions and expected returns on stocks and bonds". In: *Journal of Financial Economics* 25.1, pp. 23–49.

Fama, Eugene F. and G. William Schwert (1977). "Asset returns and inflation". In: *Journal of Financial Economics* 5.2, pp. 115–146.

Goyal, Amit and Ivo Welch (2008). "A comprehensive look at the empirical performance of equity premium prediction". In: *Review of Financial Studies* 21.4, pp. 1455–1508.

Guo, Hui (2006). "On the out-of-sample predictability of stock market returns". In: *Journal of Business* 79.2, pp. 645–670.

Harvey, Campbell R., Yan Liu, and Heqing Zhu (2016). "... and the cross-section of expected returns". In: *Review of Financial Studies* 29.1, pp. 5–68.

Hirshleifer, David, Kewei Hou, and Siew Hong Teoh (2009). "Accruals, cash flows, and aggregate stock returns". In: *Journal of Financial Economics* 91.3, pp. 389–406.

Huang, Dashan, Fuwei Jiang, Jun Tu, and Guofu Zhou (2015). "Investor sentiment aligned: A powerful predictor of stock returns". In: *Review of Financial Studies* 28.3, pp. 791–837.

Jagannathan, Ravi and Yong Wang (2007). "Lazy investors, discretionary consumption, and the cross-section of stock returns". In: *The Journal of Finance* 62.4, pp. 1623–1661.

Jondeau, Eric, Qunzi Zhang, and Xiaoneng Zhu (2019). "Average skewness matters". In: *Journal of Financial Economics* 134.1, pp. 29–47.

Jones, Christopher S. and Selale Tuzel (2013). "New orders and asset prices". In: *Review of Financial Studies* 26.1, pp. 115–157.

Kelly, Bryan and Hao Jiang (2014). "Tail risk and asset prices". In: *Review of Financial Studies* 27.10, pp. 2841–2871.

Kelly, Bryan and Seth Pruitt (2013). "Market expectations in the cross-section of present values". In: *The Journal of Finance* 68.5, pp. 1721–1756.

— (2022 exp). "Dissecting market expectations in the cross-section of book-to-market ratios: A Comment". In: *Critical Finance Review*.

Kostakis, Alexandros, Tassos Magdalinos, and Michalis P. Stamatogiannis (2015). "Robust econometric inference for stock return predictability". In: *Review of Financial Studies* 28.5, pp. 1506–1553.

Kothari, Smitu P. and Jay Shanken (1997). "Book-to-market, dividend yield, and expected market returns: A time-series analysis". In: *Journal of Financial Economics* 44.2, pp. 169–203.

Lettau, Martin and Sydney Ludvigson (2001). "Consumption, aggregate wealth, and expected stock returns". In: *The Journal of Finance* 56.3, pp. 815–849.

Li, Jun and Jianfeng Yu (2012). "Investor attention, psychological anchors, and stock return predictability". In: *Journal of Financial Economics* 104.2, pp. 401–419.

Lo, Andrew W. (2002). "The statistics of Sharpe ratios". In: *Financial Analysts Journal* 58.4, pp. 36–52.

Lo, Andrew W. and A. Craig MacKinlay (1990). "Data-snooping biases in tests of financial asset pricing models". In: *Review of Financial Studies* 3.3, pp. 431–467.

Löffler, Gunter (2022 (exp)). "Equity Premium Forecasts Tend to Perform Worse Against a Buy-and-Hold Benchmark". In: *Critical Finance Review*, forthcoming.

Maio, Paulo (2013). "The "Fed model" and the predictability of stock returns". In: *Review of Finance* 17.4, pp. 1489–1533.

— (2016). "Cross-sectional return dispersion and the equity premium". In: *Journal of Financial Markets* 29, pp. 87–109.

Martin, Ian (2011). *Simple variance swaps*. Tech. rep. National Bureau of Economic Research.

— (2017). "What is the Expected Return on the Market?" In: *The Quarterly Journal of Economics* 132.1, pp. 367–433.

McCracken, Michael W. (2007). "Asymptotics for out of sample tests of Granger causality". In: *Journal of Econometrics* 140.2, pp. 719–752.

Møller, Stig Vinther and Jesper Rangvid (2015). "End-of-the-year economic growth and time-varying expected returns". In: *Journal of Financial Economics* 115.1, pp. 136–154.

Neely, Christopher J., David E. Rapach, Jun Tu, and Guofu Zhou (2014). "Forecasting the equity risk premium: the role of technical indicators". In: *Management Science* 60.7, pp. 1772–1791.

Oliveira, Thiago de (2022 exp). In: *Critical Finance Review*.

Piazzesi, Monika, Martin Schneider, and Selale Tuzel (2007). "Housing, consumption and asset pricing". In: *Journal of Financial Economics* 83.3, pp. 531–569.

Pollet, Joshua M. and Mungo Wilson (2010). "Average correlation and stock market returns". In: *Journal of Financial Economics* 96.3, pp. 364–380.

Pontiff, Jeffrey and Lawrence D. Schall (1998). "Book-to-market ratios as predictors of market returns". In: *Journal of Financial Economics* 49.2, pp. 141–160.

Rapach, David E., Matthew C. Ringgenberg, and Guofu Zhou (2016). "Short interest and aggregate stock returns". In: *Journal of Financial Economics* 121.1, pp. 46–65.

Romano, Joseph P. and Michael Wolf (2005). "Stepwise multiple testing as formalized data snooping". In: *Econometrica* 73.4, pp. 1237–1282.

Seo, Sang Byung and Jessica A Wachter (2019). "Option prices in a model with stochastic disaster risk". In: *Management Science* 65.8, pp. 3449–3469.

Welch, Ivo (2016). "The (time-varying) importance of disaster risk". In: *Financial Analysts Journal* 72.5, pp. 14–30.

White, Halbert (2001). "A reality check for data snooping". In: *Econometrica* 68.5, pp. 1097–1126.

Yu, Jialin (2011). "Disagreement and return predictability of stock portfolios". In: *Journal of Financial Economics* 99.1, pp. 162–183.

VI   TABLES

Table 1: Glossary of Recently Published Papers and Variables

Code	Reference and Variable
AMP	Atanasov, Møller, Priestley (JF 2020), » *Consumption Fluctuations and Expected Returns*
pce	aggregate consumption to its trend (1953:1-2020:4)
AMS	Adrian, Mönch, Shin (FRBNY 2010), » *Financial intermediation, asset prices, and macroeconomic dynamics*
sbdlev	growth rate of security broker-dealer leverage (1951:4-2020:4)
BPS	Bakshi, Panayotov, Skoulakis (JFE 2011), » *Improving the predictability of real economic activity and asset returns with for.*
impvar	forward implied variances (1996:01-2020:12)
BY	Belo and Yu (JME 2013), » *Household & government investment and the stock market*
govik	public-sector investment (1947:1-2020:4)
BTZ	Bollerslev, Tauchen, Zhou (RFS 2009), » *Expected Stock Returns and Variance Risk Premia*
vrp	variance risk premium (1990:01-2020:12)
CEP	Chen, Eaton, Paye (JFE 2018), » *Micro(structure) before macro? The predictive power of aggregate illiquidity for.*
lzrt	9 illiquidity measures (1926:01-2020:12)
CGMS	Colacito, Ghysels, Meng, Siwasarit (RFS 2016), » *Skewness in Expected Macro Fundamentals and the Predictability of Equity Ret.*
skew	skewness of GDP growth forecasts (1951:2-2019:2)
CGP	Chava, Gallmeyer, Park (JME 2015), » *Credit conditions and stock return predictability*
crdstd	loan officer credit standards (1990:2-2020:4)
CP	Cooper and Priestley (RFS 2009), » *Time-Varying Risk Premiums and the Output Gap*
ogap	output gap of industrial production (1926:01-2020:12)
DJM	Driesprong, Jacobsen, Maat (JFE 2008), » *Striking oil: Another puzzle?*
wtexas	oil price changes (1926:01-2020:12)
HHT	Hirshleifer, Hou, Teoh (JFE 2008), » *Accruals, cash flows, and aggregate stock returns*
accrul, cfacc	aggregate accruals and cash flows (1965:2020/1965:2020)
HJTZ	Huang, Jiang, Tu, Zhou (RFS 2015), » *Investor Sentiment Aligned: A Powerful Predictor of Stock Returns*
sntm	optimized investor sentiment index (1965:07-2018:12)
JT	Jones and Tuzel (RFS 2013), » *New Orders and Asset Prices*
ndrbl	new orders to shipments of durable goods (1958:02-2020:12)

38

JZZ   Jondeau, Zhang, Zhu (JFE 2019), » *Average Skewness Matters*
skvw   average stock skewness (1926:07-2020:12)

KJ   Kelly and Jiang (RFS 2014), » *Tail Risk and Asset Prices*
tail   tail risk from cross-section (1926:07-2020:12)

KP   Kelly and Pruitt (JF 2013), » *Market Expectations in the Cross-Section of Present Values*
fbm   single factor from B/M cross-section (1926:06-2020:12)

LY   Li and Yu (JFE 2012), » *Investor attention, psychological anchors, and stock return predictability*
dtoy,dtoat   nearness to Dow 52-week high (1926:01-2020:12/1926:01-2020:12)

Maio$_{(13)}$ Maio (RF 2013), » *The Fed Model and the Predictability of Stock Returns*
ygap   stock-bond yield gap (1953:04-2020:12)

Maio$_{(16)}$ Maio (JFM 2016), » *Cross-sectional return dispersion and the equity premium*
rdsp   stock-return dispersion (1926:09-2020:12)

Mrtn   MA (QJE 2017), » *Expected Return on the market*
rsvix   scaled risk-neutral vix (1996:01-2020:12)

MR   Møller and Rangvid (JFE 2015), » *End-of-the-year economic growth and time-varying expected returns*
gpce, gip   year-end economic growth (1947-2020/1926-2020)

NRTZ   Neely, Rapach, Tu, Zhou (MS 2014), » *Forecasting the Equity Risk Premium: The Role of Technical Indicators*
tchi   14 technical indicators (1951:02-2020:12)

PST   Piazzesi, Schneider, Tuzel (JFE 2007), » *Housing, consumption, and asset pricing.*
house   share of housing in consumption (1929-2020)

PW   Pollett and Wilson (JFE 2010), » *Average correlation and stock market returns*
avgcor   average correlation of daily stock returns (1926:03-2020:12)

RRZ   Rapach, Ringgenberg, Zhou (JFE 2016), » *Short interest and aggregate stock returns*
shtint   short stock interest (1973:01-2020:12)

Y   Yu (JFE 2011), » *Disagreement and return predictability of stock portfolios*
disag   analyst forecast disagreements (1981:12-2020:12)

39

Table 2: Basic-Replication IS Sample Results

		Paper	Vrbl Name	Reported Coef b	Reported (T)	Author Sample	Similar Spec Coef b	Similar Spec (T)		IS Halves b_{H1}	b_{H2}	T_Δ	oos
1	2020	AMP	pce	−0.43	−3.28	Q 1953:Q3 2017:Q4	−0.44	−3.30	✓	−0.54**	−0.38**	✓	*
2	2010	AMS	sbdlev	−0.09	−3.01	Q 1986:Q1 2009:Q4	−0.03	−0.94	✓?	0.03	−0.19**	Δ	
3	2011	BPS	impvar†	15.56	3.30	M 1998:09 2008:09	13.58	3.55	✓	17.88***	−1.37	Δ	
4	2009	BTZ	vrp	0.47	2.86	Q 1990:01 2007:12	0.12	4.35	✓	0.13***	0.08**	✓	✓
5	2013	BY	govik	1.02	2.11	Q 1947:Q2 2010:Q4	1.07	2.06	✓	1.66**	−1.88	Δ	
6	2018	CEP	lzrt	2.59	2.84	M 1948:01 2015:12	2.30	2.51	✓	1.88	2.84**	✓	✓
7	2016	CGMS	skew	−0.02	−2.66	SA 1951H1 2010H2	−0.02	−1.28	✓?	−0.04**	0.02	Δ	
8	2015	CGP	crdstd	−0.10	−2.45	Q 1990:Q2 2013:Q4	−0.10	−2.20	✓	−0.09	−0.11	✓	✓
9	2009	CP	ogap	−0.11	−4.08	M 1948:01 2005:12	−0.09	−3.68	✓	−0.13***	−0.05	✗	*
10	2008	DJM	wtexas	−0.09	−3.57	M 1973:10 2004:04	−0.10	−3.19	✓	−0.08	−0.11***	✓	✓
11	2008	HHT_acc	accrul	0.07	3.33	A 1965 2005	0.06	2.79	✓	0.03	0.11***	✗	✓
	2008	HHT_cf	cfacc	−0.05	−2.42	A 1965 2005	−0.06	−2.86	✓	−0.05**	−0.11	✓	✓
12	2014	HJTZ	sntm	0.58	3.04	M 1965:07 2010:12	0.48	2.59	✓	0.53	0.46	✓	✓
13	2013	JT	ndrbl	−0.46	−3.21	M 1958:02 2009:12	−0.31	−2.47	✓	−0.47**	−0.20	✓	*
14	2019	JZZ	skvw	−0.13	−3.10	M 1963:08 2016:12	−0.11	−2.58	✓	−0.10	−0.11	✓	*
15	2014	KJ	tail	4.54	2.08	M 1963:01 2010:12	5.09	2.38	✓	4.79	8.51	✓	*
16	2013	KP	fbm	n/a	2.85	M 1930:01 2010:12	0.16	3.30	✓	0.14***	0.51***	✗	*
17	2012	LY_dtoy	dtoy†	0.32	2.09	M 1958:01 2009:12	0.22	1.92	✓	0.28	0.18	✓	
	2012	LY_dtoat	dtoat†	−0.48	−3.79	M 1958:01 2009:12	−0.29	−3.54	✓	−0.36***	−0.26***	✓	*
18	2013	Maio_(13)	ygap	0.20	2.94	M 1953:04 2008:12	0.01	1.84	✓	0.01	0.01	✓	*
19	2016	Maio_(16)	rdsp	−3.45	−2.55	M 1963:07 2013:12	−2.32	−2.24	✓	−1.39	−2.81*	✓	*
20	2017	Mrtn	rsvix	2.10	2.46	SA 1996H1 2012H2	2.12	2.19	✓	2.10	2.12**	✓	
21	2015	MR_gpc	gpce	−14.61	−4.88	A 1948 2009	−14.07	−4.07	✓	−16.17***	−10.24	✓	✓
		MR_gip	gip	−3.78	−5.74	A 1948 2009	−3.81	−5.17	✓	−4.40***	−2.55	✓	*
22	2014	NRTZ	tchi	0.12	2.12	M 1951:01 2011:12	0.26	1.88	✓	0.21	0.31	✓	✓
23	2007	PST	house	8.44	3.65	A 1936 2001	4.74	2.64	✓	5.05***	−19.88	Δ	
24	2010	PW	avgcor	0.06	2.66	M 1963:02 2007:01	0.05	2.57	✓	0.07***	0.04	✓	*
25	2016	RRZ	shtint	−0.50	−2.50	M 1973:01 2014:12	−0.43	−2.15	✓	−0.47	−0.37	✓	✓
26	2011	Y	disag	−0.17	−2.59	M 1981:12 2005:12	−0.16	−3.98	✓	−0.09**	−0.22***	✗	

†: Regression includes multivariate variables also included in the original paper.

Explanations: Papers and variables are defined in Table 1. The "reported" statistics appeared in the original paper. The "author sample" shows the original frequency and sample period. The "similar spec" is our replication of the original paper's main finding in the same sample period with the same frequency. Our own similar specification is usually based only on the key independent variable itself, except for regressions marked with daggers (BPS impvar, JMV metal, LY dtoy and dtoat), in which we also included control variables similar to those used by the authors. The "IS Halves" columns on the right are from a different regression, in which the intercept and slope are interacted with a dummy for the first vs. second half of the authors' period. In a stable specification, this coefficient should be zero. It is crossed if the two halves are statistically significant different (a high hurdle) or the sign changes. The OOS column is checkmarked if we have positive OOS performance in the sample used by the authors, left blank if we could not locate OOS statistics in the original paper, and * if discrepancies in our and their analysis will be explained in the more detailed discussion in the text.

Table 3: Predicting Monthly Log Equity Premia

Variable		IS OLS b	IS NW T	IS Halves b_{H1}	IS Halves b_{H2}	T_Δ	IS/OOS IS R^2	IS/OOS OOS R^2
BPS	impvar	0.14	0.35	0.83	−0.07	Δ	0.10	−3.44
BTZ	vrp	0.07	0.12	1.03	−0.08	Δ	0.03	−8.88
CEP	lzrt	0.26	0.96	0.30	0.23		0.23	0.11[*]
CP	ogap	−0.21	−0.62	−0.08	−1.33		0.15	−0.24
DJM	wtexas	−0.29	−1.47	−0.60	−0.25		0.29	−0.12
HJTZ	sntm	0.45[***]	2.66[***]	0.48	0.30		1.04	−1.24
JT	ndrbl	−0.33[*]	−1.73[*]	−0.50	−0.22		0.60	−0.48
JZZ	skvw	0.01	0.04	0.18	−0.33	Δ	0.00	−0.57
KJ	tail	0.03	0.21	−0.08	0.80	Δ	0.00	−0.37
KP	fbm	1.05[***]	3.44[***]	1.08	0.48		3.79	−2.21
LY_{dtoy}	dtoy	0.17	0.40	0.34	−0.23	Δ	0.10	−0.57
LY_{dtoat}	dtoat	−0.09	−0.32	−0.02	−0.69		0.03	−0.07
$Maio_{(13)}$	ygap	0.13	0.67	0.18	0.29		0.10	−1.24
$Maio_{(16)}$	rdsp	0.28	0.74	0.39	−0.22	Δ	0.27	−1.39
NRTZ	tchi	0.29	1.61	0.33	0.22		0.47	0.20[*]
PW	avgcor	0.20	0.84	0.00	0.44		0.13	−0.17
RRZ	shtint	−0.29	−1.28	0.06	−0.39	Δ	0.41	0.74[**]
Y	disag	0.01	0.06	−0.31	0.36	Δ	0.00	−1.07
CS_{dp}	d/p	0.20	0.84	0.54	0.14		0.14	−0.12
CS_{dy}	d/y	0.26	1.03	0.74	0.15		0.22	−0.32
CS_{ep}	e/p	0.30[*]	1.72[*]	0.97	0.04		0.31	−1.24
CS_{de}	d/e	−0.10	−0.34	−0.64	0.11	Δ	0.03	−1.19
G	svar	−0.08	−0.18	0.00	−0.18		0.02	−0.32
KS	b/m	0.31	0.95	1.02	−0.02	Δ	0.33	−1.35
BMRR	ntis	−0.36	−1.44	−0.64	−0.09		0.44	−0.48
Cmpl	tbl	−0.28[*]	−1.71[*]	−0.50	−0.26		0.26	0.14[*]
FF_{lty}	lty	−0.22	−1.43	−0.77	−0.21		0.16	−0.63
FF_{ltr}	ltr	0.24	1.53	0.16	0.27		0.20	−0.94
FF_{tms}	tms	0.18	1.11	0.19	0.21		0.11	0.02
FF_{dfy}	dfy	0.07	0.15	0.02	0.24		0.02	−0.11
FF_{dfr}	dfr	0.23	0.90	−0.02	0.39	Δ	0.18	−0.34
FS	infl	−0.23	−1.01	−0.15	−0.43		0.17	0.11[*]

(continues on next page)

(Table 3 continued)

Explanations:

Papers, variables and sample periods are defined in Table 1. The prediction frequencies in this table are monthly. Only monthly variables are considered. The dependent variable is always the log equity premium. The independent variable is always the (lagged) variable as proposed by the author, but Z-normalized (which matters for the coefficient magnitude but not the T-statistic) and without further controls. (For coefficient estimates more comparable with those reported by the authors, consult Table 2.)

The IS sample period always begins when data is available, but no earlier than 1926. The OOS sample period usually begins 20 years after the IS period.

The IS statistics report the OLS beta, casually starred based on the Newey-West T-statistic (one lag)—at absolute levels of 1.65 (*, about 90%), 2.0 (**, about 95%), and 2.5 (***, about 99%).

The next three statistics (IS Halves) are first-half and second-half independent OLS IS regressions, plus the significance of the difference. Sign changes are noted as $_\Delta$.

The next two statistics (IS/OOS) are the IS and OOS R^2, in percent. In a stable model with a large number of observations, the two would be the same. A negative OOS R^2 means that the prediction error of the variable is worse than that of the prevailing unconditional mean rate of return. The OOS R^2 is starred based on the McCracken (2007) (one-sided) MSE-F statistic.

The table highlights variables with both good IS performance ($|T| > 1.5$) and positive OOS performance in yellow.

Interpretation: The most promising variables are sntm and tbl. Other variables have either insignificant IS coefficients or negative OOS R^2.

42

Table 4: Predicting Quarterly Log Equity Premia

Variable	IS			IS Halves			IS/OOS	
	OLS b	NW T		b_{H1}	b_{H2}	T_Δ	IS R^2	OOS R^2
AMP pce	-1.68^{***}	-3.57^{***}		-1.84	-1.57		4.30	-1.17
AMS sbdlev	0.52	0.87		1.05	-1.80	Δ	0.43	-1.91
BY govik	0.67^{*}	1.67^{*}		1.36	-2.86	Δ	0.72	-1.25
CGP crdstd	-1.73	-1.65		-2.20	-1.24		4.52	2.61^{**}
Mrtn rsvix	2.00^{*}	1.69^{*}		2.51	1.74		5.66	-2.90
PW avgcor	1.00	1.37		0.21	1.96		0.90	1.09^{**}
Crn i/k	-1.56^{**}	-3.42^{***}		-2.17	-0.93		3.90	2.26^{**}
LL cay	0.34	0.51		2.64	-0.54	Δ	0.17	-11.57

Explanations: See Table 3. However, this table is for variables that are available only on a quarterly basis.

Interpretation: crdstd and i/k have significant IS coefficients and OOS R^2. crdstd is based on the Fed's Senior Loan Officer Opinion Survey which is conducted once at the end of each quarter and released a month into the following quarter.

Table 5: Predicting Annual Calendar-Year (Jan-Dec) Log Equity Premia

Variable		IS		IS Halves			IS/OOS	
		OLS b	NW T	b_{H1}	b_{H2}	T_Δ	IS R^2	OOS R^2
CGMS	skew	0.65	0.28	–2.66	7.85	Δ	0.16	–3.73
HHT_{ac}	accrul	5.15**	2.73***	2.69	8.54	↑	9.61	12.49**
HHT_{cf}	cfacc	–5.35**	–3.08***	–6.15	–4.58		10.53	5.16**
MR	gpce	–5.64**	–3.40***	–6.02	–4.48		11.91	4.86**
MR	gip	–0.29	–0.10	0.42	–10.26	Δ	0.02	–1.98
PST	house	1.94	0.99	2.81	–1.48	Δ	1.00	0.59*
BW	eqis	–5.48***	–2.67***	–8.10	–0.15	↓	7.96	–0.32

Explanations: See Table 3. However, this table is for variables that are available only on an annual basis. Statistically significant differences with a T-statistic greater than 2.0 between the "IS Halves" are indicated with an uparrow (↑) or downarrow (↓).

Interpretation: The three outstanding variables are accrul, cfacc, and gpce.

Table 6: Predicting Annual Mid-Year (Jul-Jun) Log Equity Premia, With Reporting Delay

Variable		IS			IS Halves			IS/OOS	
		OLS b	NW T		b_{H1}	b_{H2}	T_Δ	IS R^2	OOS R^2
CGMS	skew	1.76	0.83		1.20	2.93		1.34	−0.26
HHT$_{ac}$	accrul	5.46**	2.98***		4.70	7.16	↑	12.82	16.68**
HHT$_{cf}$	cfacc	−3.32	−1.42		−4.67	−2.22	↓	4.86	−7.34
MR	gpce	−3.86*	−1.77*		−3.46	−4.78		6.37	6.85**
MR	gip	−0.42	−0.13		−0.10	−4.92	↑	0.03	−0.54
PST	house	1.34	0.63		2.06	−4.95	Δ	0.34	0.04
BW	eqis	−4.48	−1.65		−8.05	2.03	Δ	3.70	−15.68

Explanations: See Table 5. However, this table is based on Jul-to-Jun predictive regressions, based on independent variable data available the prior December (i.e., with a 6 months implementation lag).

Interpretation: cfacc loses relevance, leaving accrul and gpce as good candidates.

45

Table 7: Untilted $1-Unscaled Investment Strategy

Fq	Ppr	Var	Conditional R (V) Long	Short	L–S	#Obs Bull	Bear	Unconditional R (U) L (Eq)	S (TB)	L–S	Δ V – U Mean	SR
M	BPS	impvar	2.6	9.2	−6.6	40	140	10.7	1.1	9.6	−16.2	−0.22
M	BTZ	vrp	6.7	2.6	4.1	86	166	7.7	1.6	6.2	−2.1	−0.03
M	CEP	lzrt	6.5	9.0	−2.5	263	637	11.7	3.8	7.9	−10.4	−0.12
M	CP	ogap	10.4	5.1	5.3	685	215	11.7	3.8	7.9	−2.6	−0.05
M	DJM	wtexas	10.6	4.9	5.8	597	303	11.7	3.8	7.9	−2.1	−0.04
M	HJTZ	sntm	9.5	5.6	4.0	309	117	12.0	3.1	8.9	−4.9	−0.08
M	JT	ndrbl	9.8	7.2	2.6	358	157	12.8	4.3	8.5	−5.9	−0.12
M	JZZ	skvw	5.6	9.9	−4.3	399	495	11.7	3.9	7.8	−12.2	−0.16
M	KJ	tail	8.9	6.6	2.3	531	363	11.7	3.9	7.8	−5.5	−0.09
M	KP	fbm	8.6	6.9	1.7	478	417	11.6	3.9	7.8	−6.1	−0.08
M	LY$_{dtoat}$	dtoat	5.8	9.7	−3.9	121	779	11.7	3.8	7.9	−11.8	−0.14
M	LY$_{dtoy}$	dtoy	7.5	8.0	−0.5	364	536	11.7	3.8	7.9	−8.4	−0.12
M	Maio$_{(16)}$	rdsp	4.1	11.6	−7.5	179	713	11.8	3.9	8.0	−15.4	−0.18
M	Maio$_{(13)}$	ygap	6.7	9.4	−2.7	203	370	11.6	4.4	7.2	−9.9	−0.12
M	NRTZ	tchi	9.3	6.7	2.6	333	267	11.6	4.4	7.2	−4.6	−0.06
M	PW	avgcor	9.4	6.1	3.3	412	486	11.7	3.8	7.9	−4.6	−0.07
M	RRZ	shtint	10.2	3.1	7.2	219	106	11.0	2.3	8.7	−1.6	−0.03
M	Y	disag	9.0	1.8	7.2	215	14	9.5	1.2	8.3	−1.1	−0.03
M	BMRR	ntis	10.2	5.7	4.6	560	329	12.0	3.9	8.1	−3.6	−0.06
M	CS$_{de}$	d/e	4.8	10.7	−5.9	125	775	11.7	3.8	7.9	−13.7	−0.15
M	CS$_{dp}$	d/p	6.2	9.3	−3.0	179	721	11.7	3.8	7.9	−10.9	−0.12
M	CS$_{dy}$	d/y	6.4	9.1	−2.7	177	722	11.6	3.8	7.8	−10.5	−0.12
M	CS$_{ep}$	e/p	7.6	8.0	−0.4	298	602	11.7	3.8	7.9	−8.2	−0.10
M	Cmpl	tbl	6.0	9.6	−3.6	245	655	11.7	3.8	7.9	−11.5	−0.13
M	FF$_{dfr}$	dfr	9.0	6.6	2.4	469	431	11.7	3.8	7.9	−5.4	−0.07
M	FF$_{dfy}$	dfy	7.6	7.9	−0.4	382	518	11.7	3.8	7.9	−8.2	−0.12
M	FF$_{ltr}$	ltr	9.7	5.8	3.9	441	459	11.7	3.8	7.9	−4.0	−0.06
M	FF$_{lty}$	lty	6.7	8.8	−2.1	224	676	11.7	3.8	7.9	−10.0	−0.12
M	FF$_{tms}$	tms	8.4	7.2	1.2	403	497	11.7	3.8	7.9	−6.7	−0.09
M	FS	infl	10.4	5.1	5.3	411	489	11.7	3.8	7.9	−2.6	−0.04
M	G	svar	7.4	8.1	−0.7	373	527	11.7	3.8	7.9	−8.6	−0.14
M	KS	b/m	5.9	9.6	−3.6	265	635	11.7	3.8	7.9	−11.5	−0.14
Q	AMP	pce	10.7	5.8	5.0	134	55	12.1	4.4	7.6	−2.6	−0.06
Q	AMS	sbdlev	9.1	7.3	1.8	111	86	11.9	4.5	7.4	−5.6	−0.12
Q	BY	govik	4.5	11.6	−7.1	12	204	11.6	4.5	7.1	−14.2	−0.23
Q	CGP	crdstd	8.1	1.4	6.7	46	37	8.0	1.5	6.5	0.2	0.00
Q	Mrtn	rsvix	7.8	4.3	3.4	19	41	11.0	1.1	9.9	−6.5	−0.15
Q	Crn	i/k	9.6	6.6	3.0	88	128	11.6	4.5	7.1	−4.1	−0.08
Q	LL	cay	7.4	8.9	−1.6	60	136	11.9	4.5	7.4	−9.0	−0.16
D	MR	gip	11.8	4.7	7.1	52	23	12.5	3.9	8.6	−1.4	−0.08
D	MR	gpce	10.8	5.8	5.0	31	23	12.0	4.7	7.4	−2.4	−0.12
D	PST	house	7.0	10.2	−3.2	23	49	13.0	4.1	8.9	−12.2	−0.40
D	BW	eqis	9.8	7.0	2.8	43	31	12.8	4.0	8.8	−6.0	−0.23
J	HHT$_{acc}$	accrul	7.5	7.6	−0.1	15	20	11.9	3.2	8.7	−8.8	−0.39
J	HHT$_{cf}$	cfacc	10.9	4.2	6.7	23	12	11.9	3.2	8.7	−2.0	−0.10

Explanations: Variables and sample periods are defined in Table 1. Investment begins 20 years after the sample has started. 'D' means calendar year, 'J' is mid-year investing with reporting delay. The conditional timing strategy invests $1 in the equity premium financed by the T-bill if the predictive coefficient is positive and V is above its median, or the coefficient is negative and V is below its median (both "bullish"); and the opposite otherwise ("bearish"). The unconditional strategy is always bullish and earns the equity premium. All measures are annualized, incl. the Sharpe Ratio (SR), and (none are strong enough to be) starred for statistical significance as in Lo (2002).

Table 8: Untilted Z-Scaled Investment Strategy

	Variable (V)		Conditional R (V)			#Obs		Unconditional R (U)			Δ V – U	
Fq	Ppr	Var	Long	Short	L–S	Bull	Bear	L (Eq)	S (TB)	L–S	Mean	SR
M	BPS	impvar	1.5	7.0	−5.5	34	146	7.9	0.5	7.4	−12.9	−0.22
M	BTZ	vrp	4.9	4.9	0.1	63	189	8.6	1.2	7.4	−7.4	−0.02
M	CEP	lzrt	4.6	5.8	−1.2	266	634	7.6	2.8	4.8	−5.9	−0.05
M	CP	ogap	7.1	2.1	5.1	678	222	7.5	1.7	5.8	−0.7	−0.04
M	DJM	wtexas	8.5	3.0	5.5	523	377	8.3	3.1	5.2	0.3	0.00
M	HJTZ	sntm	9.6	4.6	4.9	358	68	10.7	3.5	7.2	−2.2	−0.05
M	JT	ndrbl	10.7	5.3	5.4	356	159	12.8	3.2	9.6	−4.2	−0.10
M	JZZ	skvw	2.8	5.4	−2.7	423	471	6.2	2.0	4.2	−6.9	−0.08
M	KJ	tail	6.5	4.8	1.7	543	351	8.8	2.5	6.2	−4.5	−0.07
M	KP	fbm	4.0	1.8	2.2	736	159	4.7	1.1	3.5	−1.3	−0.07
M	LY$_{dtoat}$	dtoat	3.5	8.4	−4.8	69	831	8.9	3.0	5.9	−10.8	−0.14
M	LY$_{dtoy}$	dtoy	4.0	4.9	−0.9	209	691	6.9	2.1	4.8	−5.7	−0.14
M	Maio$_{(16)}$	rdsp	2.3	6.6	−4.3	81	811	6.8	2.1	4.7	−9.0	−0.17
M	Maio$_{(13)}$	ygap	7.3	9.4	−2.1	194	379	11.4	5.3	6.1	−8.2	−0.06
M	NRTZ	tchi	7.8	3.8	4.0	407	193	7.8	3.9	3.9	0.1	0.00
M	PW	avgcor	7.2	4.0	3.2	349	549	8.7	2.5	6.2	−3.0	−0.05
M	RRZ	shtint	15.4	1.5	13.8	231	94	14.5	2.3	12.2	1.6	0.02
M	Y	disag	11.2	1.0	10.2	200	29	11.4	0.8	10.6	−0.4	−0.02
M	BMRR	ntis	6.6	2.4	4.2	682	207	6.7	2.3	4.4	−0.3	−0.02
M	CS$_{de}$	d/e	5.6	10.5	−4.9	91	809	12.2	3.8	8.4	−13.3	−0.12
M	CS$_{dp}$	d/p	6.1	10.4	−4.4	165	735	12.3	4.2	8.1	−12.5	−0.09
M	CS$_{dy}$	d/y	6.1	10.2	−4.0	164	735	12.0	4.2	7.8	−11.8	−0.08
M	CS$_{ep}$	e/p	7.2	7.6	−0.4	272	628	11.0	3.8	7.1	−7.6	−0.06
M	Cmpl	tbl	8.3	7.7	0.7	284	616	10.1	5.9	4.1	−3.4	−0.02
M	FF$_{dfr}$	dfr	5.4	3.6	1.8	466	434	6.3	2.7	3.6	−1.8	−0.01
M	FF$_{dfy}$	dfy	3.3	5.3	−2.1	193	707	6.6	2.0	4.6	−6.7	−0.13
M	FF$_{ltr}$	ltr	11.9	4.0	7.8	435	465	11.2	4.7	6.5	1.4	0.01
M	FF$_{lty}$	lty	10.4	11.3	−0.9	307	593	14.4	7.3	7.1	−8.0	−0.04
M	FF$_{tms}$	tms	8.7	4.8	3.9	435	465	9.3	4.2	5.1	−1.2	−0.02
M	FS	infl	3.7	1.5	2.1	403	497	3.3	1.9	1.4	0.8	0.01
M	G	svar	0.9	3.5	−2.6	119	781	3.2	1.3	1.9	−4.5	−0.15
M	KS	b/m	5.9	9.0	−3.1	249	651	11.0	3.9	7.1	−10.2	−0.10
Q	AMP	pce	13.2	4.4	8.7	135	54	13.3	4.3	8.9	−0.2	−0.01
Q	AMS	sbdlev	5.1	5.3	−0.2	107	90	6.8	3.6	3.2	−3.4	−0.05
Q	BY	govik	4.7	10.9	−6.2	0	216	10.9	4.7	6.2	−12.4	−0.17
Q	CGP	crdstd	6.4	−1.8	8.2	53	30	3.3	1.4	1.9	6.3	0.08
Q	Mrtn	rsvix	9.6	3.6	6.0	17	43	12.3	0.9	11.4	−5.4	−0.18
Q	Crn	i/k	8.2	5.3	2.9	102	114	9.4	4.0	5.4	−2.6	−0.04
Q	LL	cay	10.9	13.5	−2.6	70	126	18.4	6.0	12.4	−15.0	−0.15
D	MR	gip	3.1	1.2	1.9	33	42	3.4	1.0	2.4	−0.5	−0.09
D	MR	gpce	5.3	1.8	3.5	27	27	5.0	2.2	2.8	0.7	0.06
D	PST	house	3.3	5.4	−2.1	9	63	6.3	2.3	4.0	−6.1	−0.38
D	BW	eqis	6.4	2.3	4.1	50	24	6.2	2.5	3.7	0.4	0.04
J	HHT$_{acc}$	accrul	3.5	−0.9	4.5	19	16	0.9	1.7	−0.7	5.2	0.19
J	HHT$_{cf}$	cfacc	8.4	3.1	5.3	27	8	9.2	2.2	7.0	−1.7	−0.11

Explanations: See Table 7. Here, the decision cutoff is not the median but the 25th percentile (bearish). Thus, the strategy typically invests more in equities.

47

Table 9: Equity-Tilted $1-Unscaled Investment Strategy

Fq	Ppr	Var	Long	Short	L–S	Bull	Bear	L (Eq)	S (TB)	L–S	Mean	SR
	Variable (V)		Conditional R (V)			#Obs		Unconditional R (U)			Δ V – U	
M	BPS	impvar	6.3	5.5	0.8	96	84	10.7	1.1	9.6	−8.9	−0.18
M	BTZ	vrp	5.2	4.1	1.1	144	108	7.7	1.6	6.2	−5.0	−0.08
M	CEP	lzrt	9.3	6.3	3.0	520	380	11.7	3.8	7.9	−4.9	−0.07
M	CP	ogap	11.5	4.0	7.5	856	44	11.7	3.8	7.9	−0.4	−0.01
M	DJM	wtexas	10.5	5.0	5.5	603	297	11.7	3.8	7.9	−2.3	−0.04
M	HJTZ	sntm	11.6	3.5	8.1	408	18	12.0	3.1	8.9	−0.8	−0.03
M	JT	ndrbl	11.5	5.6	5.9	464	51	12.8	4.3	8.5	−2.7	−0.10
M	JZZ	skvw	8.3	7.2	1.0	674	220	11.7	3.9	7.8	−6.8	−0.12
M	KJ	tail	9.7	5.8	3.9	706	188	11.7	3.9	7.8	−3.9	−0.09
M	KP	fbm	10.5	5.0	5.5	796	99	11.6	3.9	7.8	−2.2	−0.06
M	LY$_{dtoat}$	dtoat	8.0	7.5	0.4	471	429	11.7	3.8	7.9	−7.4	−0.13
M	LY$_{dtoy}$	dtoy	10.2	5.3	4.8	639	261	11.7	3.8	7.9	−3.0	−0.08
M	Maio$_{(16)}$	rdsp	7.0	8.7	−1.8	429	463	11.8	3.9	8.0	−9.7	−0.14
M	Maio$_{(13)}$	ygap	9.7	6.4	3.3	321	252	11.6	4.4	7.2	−3.8	−0.05
M	NRTZ	tchi	11.8	4.3	7.4	469	131	11.6	4.4	7.2	0.3	0.00
M	PW	avgcor	10.6	5.0	5.7	631	267	11.7	3.8	7.9	−2.2	−0.05
M	RRZ	shtint	10.5	2.8	7.7	263	62	11.0	2.3	8.7	−1.0	−0.03
M	Y	disag	9.9	0.8	9.1	228	1	9.5	1.2	8.3	0.9	0.07
M	BMRR	ntis	11.6	4.3	7.3	775	114	12.0	3.9	8.1	−0.8	−0.02
M	CS$_{de}$	d/e	7.1	8.4	−1.3	340	560	11.7	3.8	7.9	−9.2	−0.12
M	CS$_{dp}$	d/p	7.3	8.2	−1.0	286	614	11.7	3.8	7.9	−8.8	−0.11
M	CS$_{dy}$	d/y	7.3	8.2	−0.9	286	613	11.6	3.8	7.8	−8.7	−0.11
M	CS$_{ep}$	e/p	9.6	6.0	3.6	497	403	11.7	3.8	7.9	−4.3	−0.06
M	Cmpl	tbl	8.9	6.6	2.4	434	466	11.7	3.8	7.9	−5.5	−0.07
M	FF$_{dfr}$	dfr	10.6	4.9	5.7	647	253	11.7	3.8	7.9	−2.2	−0.03
M	FF$_{dfy}$	dfy	8.6	7.0	1.6	651	249	11.7	3.8	7.9	−6.2	−0.13
M	FF$_{ltr}$	ltr	11.1	4.4	6.7	587	313	11.7	3.8	7.9	−1.2	−0.02
M	FF$_{lty}$	lty	9.2	6.3	2.9	468	432	11.7	3.8	7.9	−5.0	−0.07
M	FF$_{tms}$	tms	10.6	5.0	5.6	600	300	11.7	3.8	7.9	−2.2	−0.04
M	FS	infl	11.5	4.1	7.4	702	198	11.7	3.8	7.9	−0.5	−0.01
M	G	svar	9.3	6.2	3.0	618	282	11.7	3.8	7.9	−4.9	−0.11
M	KS	b/m	7.9	7.6	0.3	408	492	11.7	3.8	7.9	−7.6	−0.10
Q	AMP	pce	12.2	4.3	8.0	173	16	12.1	4.4	7.6	0.4	0.02
Q	AMS	sbdlev	10.9	5.5	5.5	148	49	11.9	4.5	7.4	−2.0	−0.06
Q	BY	govik	8.3	7.9	0.4	90	126	11.6	4.5	7.1	−6.7	−0.13
Q	CGP	crdstd	9.4	0.1	9.3	65	18	8.0	1.5	6.5	2.8	0.07
Q	Mrtn	rsvix	8.7	3.4	5.3	37	23	11.0	1.1	9.9	−4.7	−0.14
Q	Crn	i/k	11.4	4.8	6.5	155	61	11.6	4.5	7.1	−0.6	−0.01
Q	LL	cay	8.3	8.0	0.3	97	99	11.9	4.5	7.4	−7.1	−0.15
D	MR	gip	12.8	3.6	9.2	73	2	12.5	3.9	8.6	0.6	0.13
D	MR	gpce	12.4	4.2	8.2	46	8	12.0	4.7	7.4	0.8	0.08
D	PST	house	10.9	6.2	4.7	52	20	13.0	4.1	8.9	−4.3	−0.25
D	BW	eqis	13.4	3.4	10.0	63	11	12.8	4.0	8.8	1.2	0.10
J	HHT$_{acc}$	accrul	13.0	2.1	10.9	31	4	11.9	3.2	8.7	2.2	0.23
J	HHT$_{cf}$	cfacc	12.4	2.7	9.7	30	5	11.9	3.2	8.7	1.0	0.08

Explanations: See Table 7. Here, the investment is not $1, but $Z (V realization minus prevailing mean, divided by prevailing standard deviation), i.e., the signal strength influences not just the direction but also the size of the investment. The unconditional strategy is always bullish and invests |$Z|.

48

Table 10: Equity-Tilted Z-scaled Investment Strategy

	Variable (V)		Conditional R (V)			#Obs		Unconditional R (U)			Δ V − U	
Fq	Ppr	Var	Long	Short	L–S	Bull	Bear	L (Eq)	S (TB)	L–S	Mean	SR
M	BPS	impvar	3.7	2.1	1.5	96	84	5.3	0.5	4.8	−3.2	−0.17
M	BTZ	vrp	6.0	4.2	1.8	144	108	9.4	0.8	8.6	−6.8	−0.02
M	CEP	lzrt	7.3	3.4	3.9	520	380	8.0	2.7	5.3	−1.5	−0.02
M	CP	ogap	12.5	3.1	9.4	856	44	12.6	3.0	9.6	−0.2	−0.04
M	DJM	wtexas	8.4	2.9	5.6	603	297	8.2	3.1	5.0	0.5	0.00
M	HJTZ	sntm	18.1	6.2	11.9	408	18	18.4	5.8	12.6	−0.7	−0.05
M	JT	ndrbl	16.8	5.4	11.3	464	51	17.5	4.7	12.9	−1.5	−0.06
M	JZZ	skvw	5.1	3.9	1.2	674	220	6.3	2.7	3.6	−2.4	−0.04
M	KJ	tail	13.4	4.7	8.6	706	188	13.6	4.5	9.1	−0.5	−0.02
M	KP	fbm	6.5	2.1	4.4	796	99	6.9	1.7	5.2	−0.8	−0.06
M	LY$_{dtoat}$	dtoat	3.8	2.6	1.2	471	429	5.2	1.1	4.0	−2.9	−0.11
M	LY$_{dtoy}$	dtoy	6.4	2.0	4.4	639	261	6.5	1.9	4.7	−0.2	−0.06
M	Maio$_{(16)}$	rdsp	1.6	1.6	0.0	429	463	2.3	0.9	1.5	−1.5	−0.11
M	Maio$_{(13)}$	ygap	11.0	8.3	2.6	321	252	12.9	6.3	6.6	−4.0	−0.04
M	NRTZ	tchi	15.7	5.3	10.4	469	131	15.3	5.6	9.7	0.7	0.01
M	PW	avgcor	12.1	2.9	9.1	631	267	12.3	2.7	9.6	−0.5	−0.02
M	RRZ	shtint	20.5	1.8	18.7	263	62	19.5	2.8	16.6	2.1	0.04
M	Y	disag	18.5	1.6	17.0	228	1	18.5	1.6	16.9	0.1	0.07
M	BMRR	ntis	8.8	2.8	6.0	775	114	8.8	2.8	6.0	−0.1	−0.01
M	CS$_{de}$	d/e	5.0	5.2	−0.1	340	560	7.9	2.3	5.6	−5.7	−0.09
M	CS$_{dp}$	d/p	7.3	6.1	1.2	286	614	10.0	3.4	6.6	−5.4	−0.06
M	CS$_{dy}$	d/y	7.5	6.0	1.5	286	613	10.0	3.5	6.5	−5.1	−0.05
M	CS$_{ep}$	e/p	11.0	5.6	5.4	497	403	12.4	4.1	8.3	−2.9	−0.03
M	Cmpl	tbl	8.5	3.9	4.5	434	466	8.1	4.3	3.8	0.7	0.01
M	FF$_{dfr}$	dfr	8.1	3.2	4.9	647	253	8.3	3.0	5.3	−0.4	−0.00
M	FF$_{dfy}$	dfy	5.8	2.4	3.4	651	249	6.2	2.1	4.1	−0.8	−0.06
M	FF$_{ltr}$	ltr	14.9	3.7	11.2	587	313	13.6	5.0	8.6	2.6	0.03
M	FF$_{lty}$	lty	12.7	8.0	4.8	468	432	14.6	6.1	8.5	−3.7	−0.03
M	FF$_{tms}$	tms	12.9	4.0	8.9	600	300	12.5	4.3	8.2	0.7	0.01
M	FS	infl	7.4	1.1	6.2	702	198	6.5	2.0	4.6	1.7	0.03
M	G	svar	1.5	0.8	0.7	618	282	1.6	0.8	0.9	−0.2	−0.09
M	KS	b/m	7.7	5.3	2.4	408	492	9.2	3.8	5.4	−3.0	−0.06
Q	AMP	pce	20.8	6.4	14.3	173	16	20.8	6.3	14.5	−0.1	−0.03
Q	AMS	sbdlev	8.4	3.6	4.8	148	49	8.6	3.3	5.3	−0.5	−0.01
Q	BY	govik	2.6	3.1	−0.5	90	126	3.7	2.0	1.7	−2.2	−0.05
Q	CGP	crdstd	10.9	−0.4	11.3	65	18	9.0	1.6	7.4	4.0	0.06
Q	Mrtn	rsvix	13.4	1.0	12.4	37	23	13.9	0.4	13.5	−1.1	−0.21
Q	Crn	i/k	12.2	4.0	8.2	155	61	12.6	3.6	9.0	−0.8	−0.02
Q	LL	cay	13.8	12.5	1.3	97	99	19.0	7.3	11.8	−10.5	−0.13
D	MR	gip	8.4	1.9	6.5	73	2	8.4	1.9	6.5	0.0	0.16
D	MR	gpce	9.5	2.2	7.3	46	8	9.0	2.6	6.4	0.9	0.14
D	PST	house	4.2	1.4	2.8	52	20	4.7	0.8	3.9	−1.1	−0.31
D	BW	eqis	11.0	3.2	7.8	63	11	10.8	3.3	7.5	0.3	0.05
J	HHT$_{acc}$	accrul	7.9	−0.5	8.4	31	4	5.1	2.3	2.8	5.6	0.24
J	HHT$_{cf}$	cfacc	11.6	4.0	7.6	30	5	12.6	3.0	9.5	−1.9	−0.14

Explanations: See Tables 8 and 9. Here, the decision is both equity-tilted and scaled (relative to the 25th percentile, not the mean).

49

Table 11: Performance for a Risk-Averse Investor

Freq	Variable		Δu(Mdl–Unc) CEV	T	Δu(Mdl –100% Eq) CEV	T
M	Ca	tbl	0.58	0.90	0.24	0.21
Q	CGP	crdstd	6.35	3.13***	5.03	1.69*
Q	CO	i/k	2.09	1.43	1.78	1.10
A	HHT	accrul	–1.24	–1.35	–3.83	–1.78*
A	HJTZ	sntm	–2.48	–1.63	–4.90	–2.15**
A	MR	gpce	–0.10	–0.08	–2.26	–1.39

Explanations: Papers, variables, and sample periods are defined in Table 1. The table describes the performance of selected variables for a quadratic risk-averse investor with parameter 5 and investment limits of 0 and 1.5 times equity, just as in Campbell and Thompson (2008). The 'Unc' investor ignores the signal and always holds equity according to the predicting performance and uncertainty; the '100% Eq' holds only equity.

Interpretation: The standout performer was crdstd. It would have been statistically significantly preferred by a risk-averse investor over both benchmark alternatives. accrul, sntm, and gpce would have been of no use to a risk-averse investor.

Panel A: Various Prediction Horizons, Log vs Simple Premia, and Sample Ends

	3 month				6 month				12 month			
			OOS R^2				OOS R^2				OOS R^2	
F O E	Coef	NW-T	Reg	Con	Coef	NW-T	Reg	Con	Coef	NW-T	Reg	Con
					Reported in Martin (2017)							
D Y S	1.01	0.62	—	1.5	2.10	2.46	—	4.9	1.67	1.32	—	4.7
					Sample ending 2012							
D Y S	1.06	0.74	−14.9	1.9	2.12	2.19	4.8	5.4	1.71	1.59	−10.1	6.1
M Y S	1.24	0.66	−21.2	2.3	2.34	2.19	5.5	6.0	1.71	1.59	−8.4	6.7
Ntv S	2.62	1.44	−10.9	4.1	1.23	0.96	9.4	4.0	1.68	1.46	−21.2	6.8
Ntv L	2.29	1.25	−11.6	3.3	1.05	0.87	−11.2	2.6	1.50	1.28	−22.9	5.2
					Sample ending 2020							
D Y S	1.39	1.07	−8.4	0.7	1.84	2.13	2.5	1.4	1.10	1.17	−11.9	−2.6
M Y S	1.33	0.81	−13.8	0.3	1.89	1.98	1.8	1.1	1.06	1.15	−10.9	−2.8
Ntv S	2.95	1.91	−1.2	2.8	1.15	0.92	−6.7	−2.2	0.85	0.82	−17.6	−4.2
Ntv L	2.61	1.69	−2.9	3.2	0.90	0.77	−8.5	−0.5	0.63	0.60	−19.7	−0.4

Panel B: Semi-Annual Predictions: Alignments and Sample Ends

	Dec/Jun				Mar/Sep			
			OOS R^2				OOS R^2	
	Coef	NW-T	Reg	Con	Coef	NW-T	Reg	Con
−2012	1.05	0.87	−11.2	2.6	3.47	2.64	4.5	6.8
−2020	0.90	0.77	−8.5	−0.5	3.04	2.80	3.8	5.9

Explanations: This table shows the effects of various specification choices. The predictor rsvix is always measured on the last day of the preceding time-period. We use Newey-West instead of Hansen-Hodrick T-statistics. F means data frequency, either daily (D) or monthly (M). O means overlapping observation. Ntv means use of native frequency non-overlapping natural intervals. E means type of equity premium, either simple (S) or Log (L). The "reg" out-of-sample performance uses a prevailing regression with rsvix to predict equity premia. The "con" uses rsvix unconstrained. (This is possibly only because the natural units of rsvix are equity premia.) In Panel B, the predictor is measured either on Dec 31 and Jun 30 on the left or on Mar 31 and Sep 30 on the right (or earlier if these were holidays). Panel B predicts log equity premia and shows Newey-West T-statistics.

Interpretation: The 3-month and 12-month results are all insignificant. The only good rsvix predictions occurred semi-annually on March ends and September ends.

51

Figure 1: IS and OOS Predictive Performance of AMP pce (quarterly)
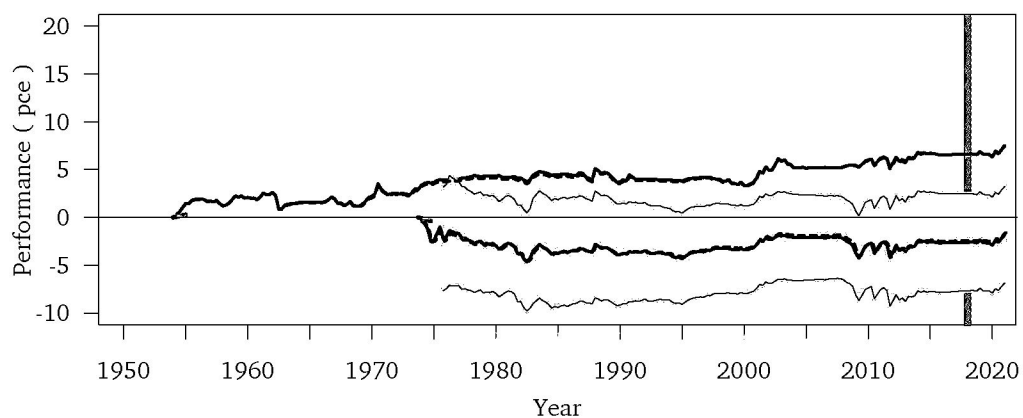
Figure 2: IS and OOS Predictive Performance of BPS impvar (monthly)

52
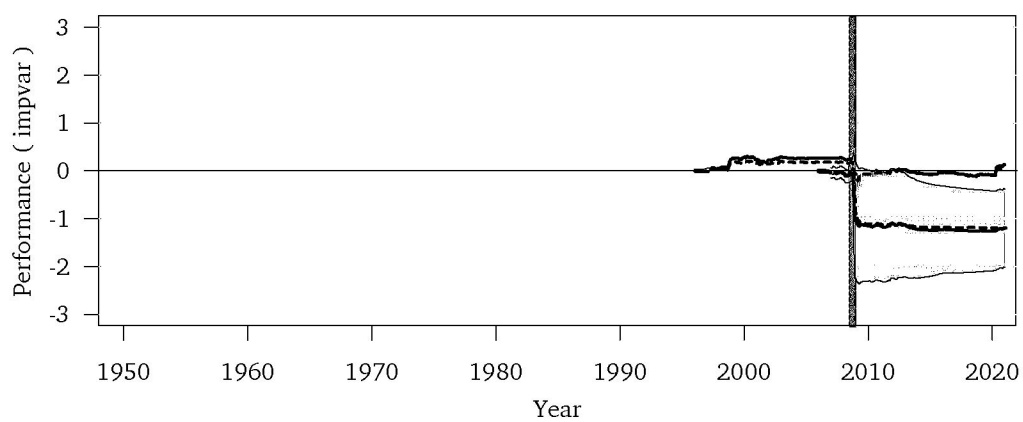
Figure 3: IS and OOS Predictive Performance of BTZ vrp (monthly)
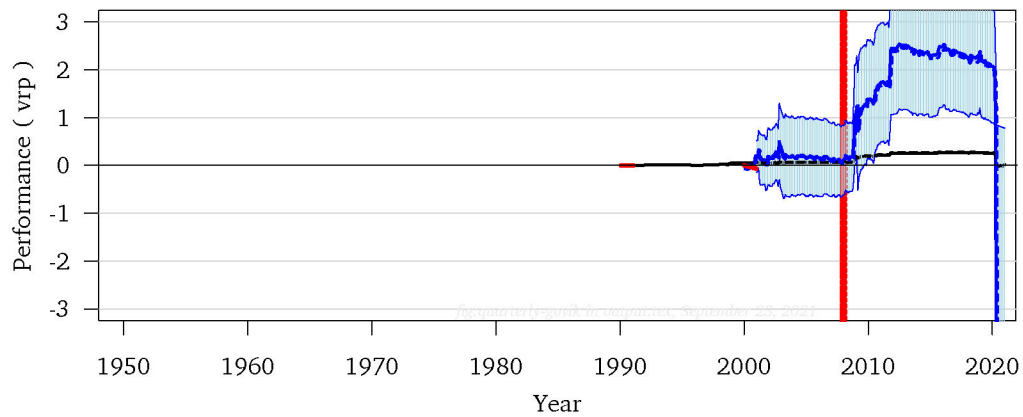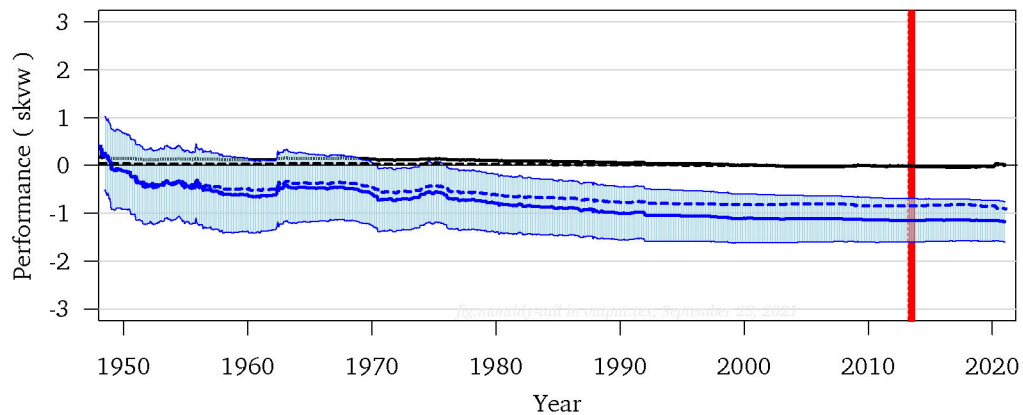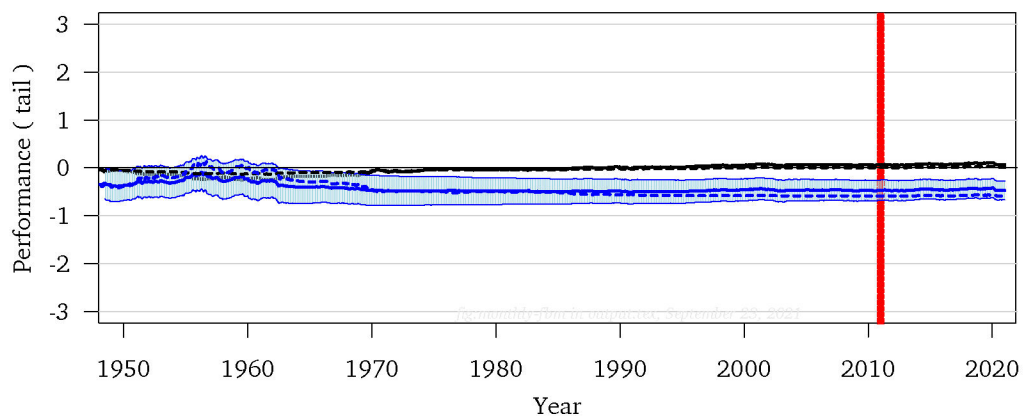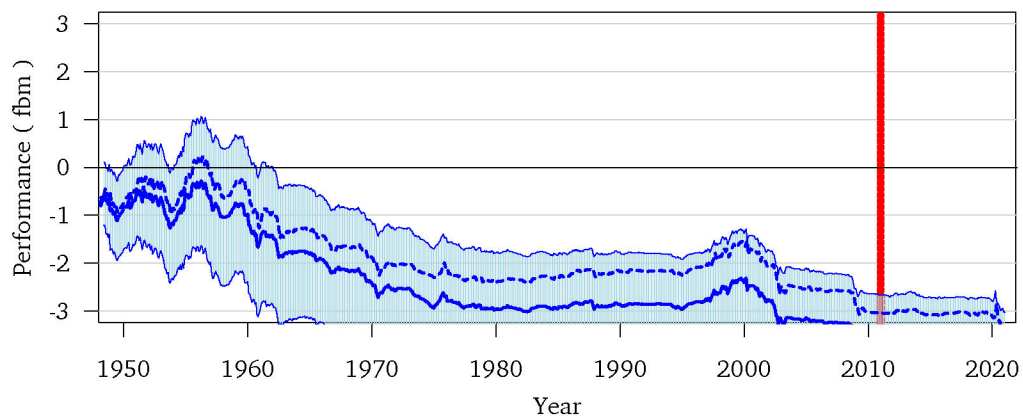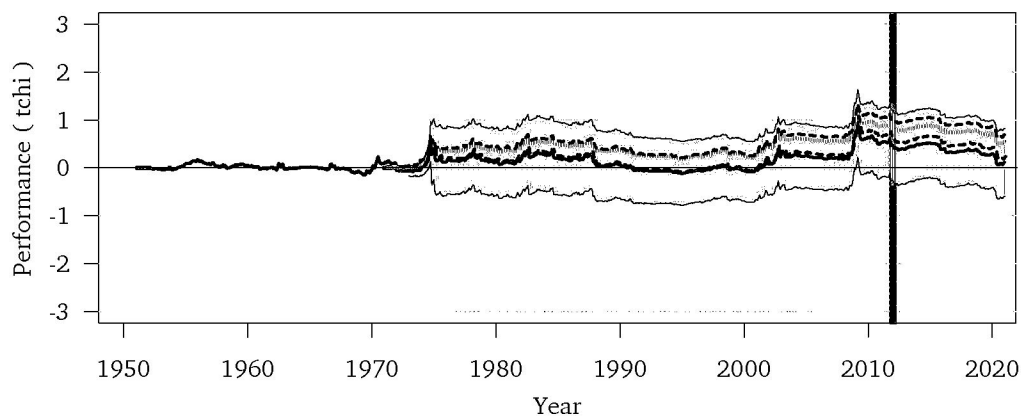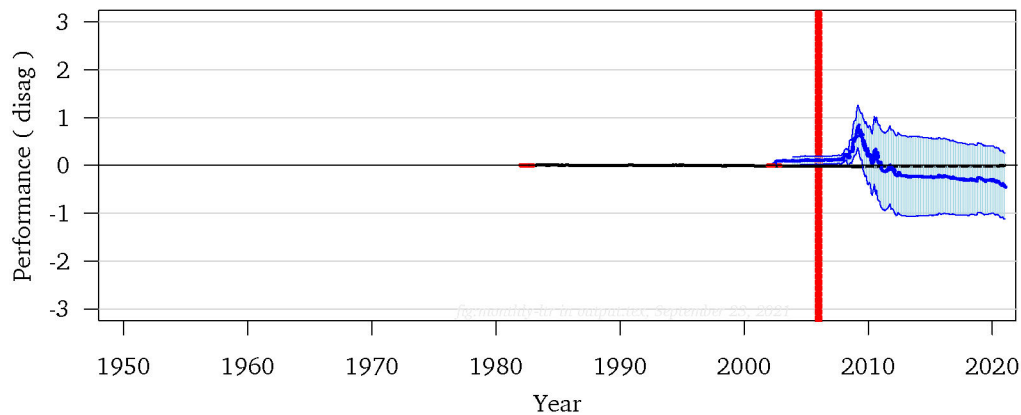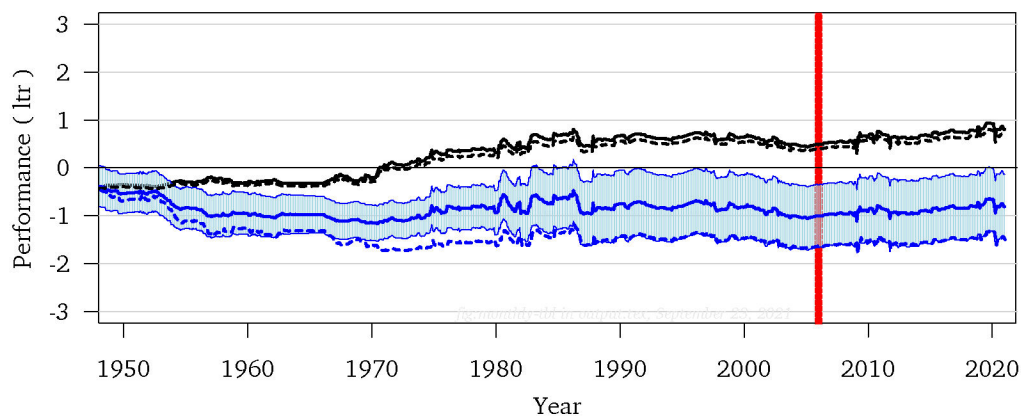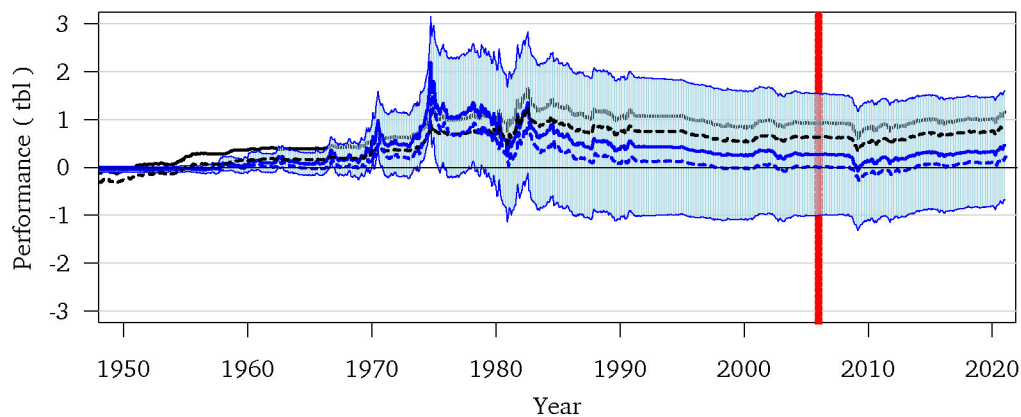

Figure 4: IS and OOS Predictive Performance of BY govik (quarterly)
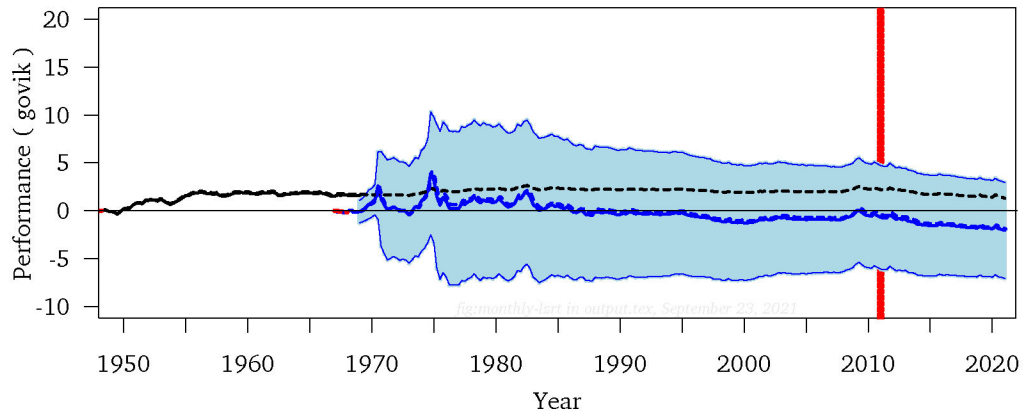
Figure 5: IS and OOS Predictive Performance of CEP lzrt (monthly)

53
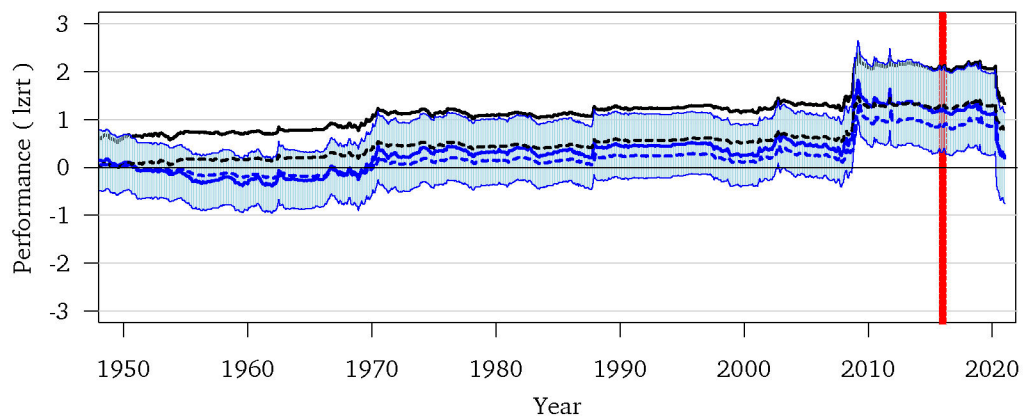
Figure 6: IS and OOS Predictive Performance of CGP crdstd (quarterly)
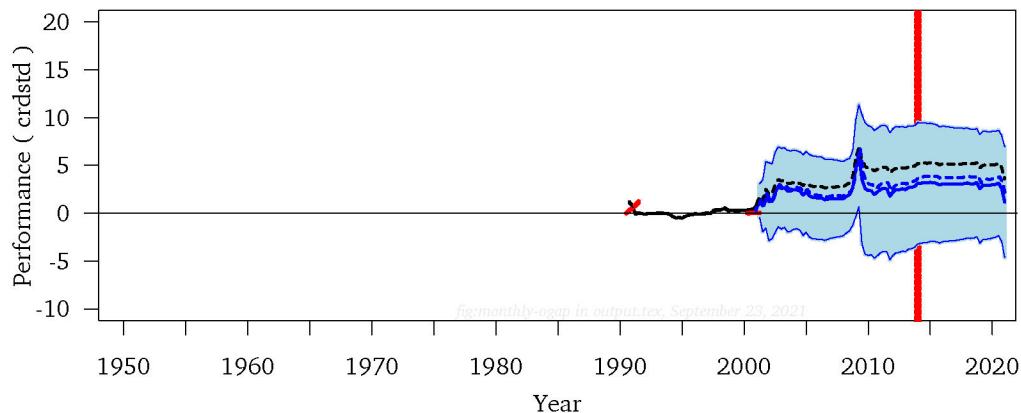
Figure 7: IS and OOS Predictive Performance of CP ogap (monthly)
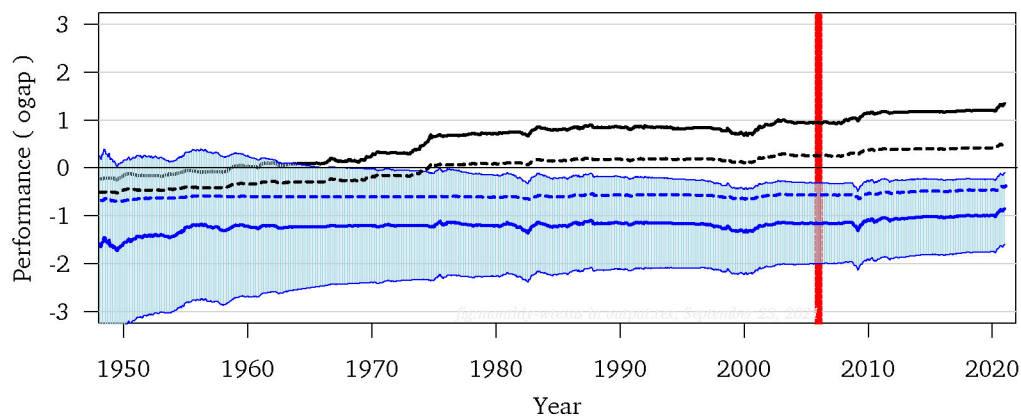
Figure 8: IS and OOS Predictive Performance of DJM wtexas (monthly)

54
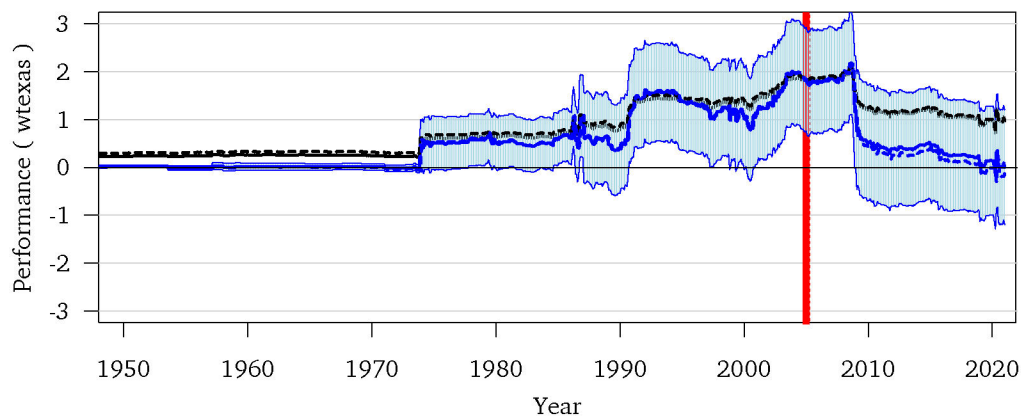
Figure 9: Time-Series of Accruals (accrul) and Equity Premia
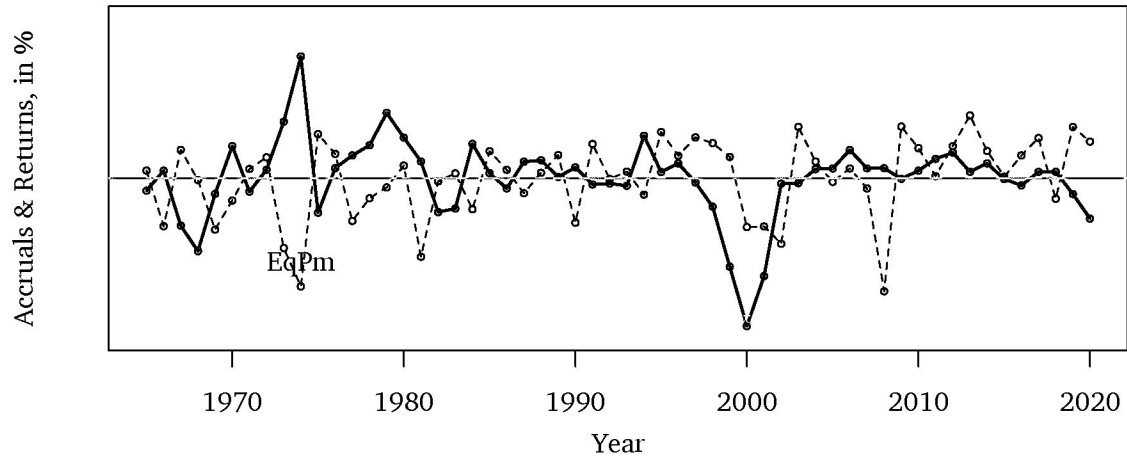
Figure 10: IS and OOS Predictive Performance of HHT accrul (annual/jun)
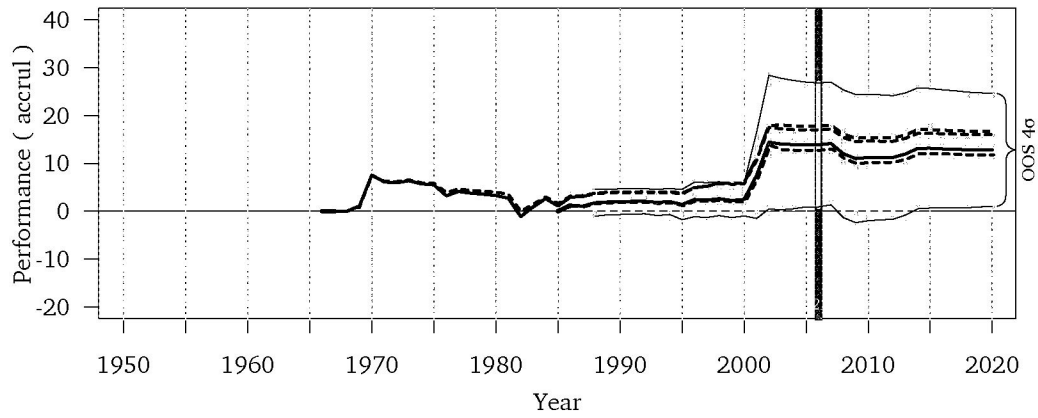
Figure 11: IS and OOS Predictive Performance of HHT cfacc (annual/jun)

55
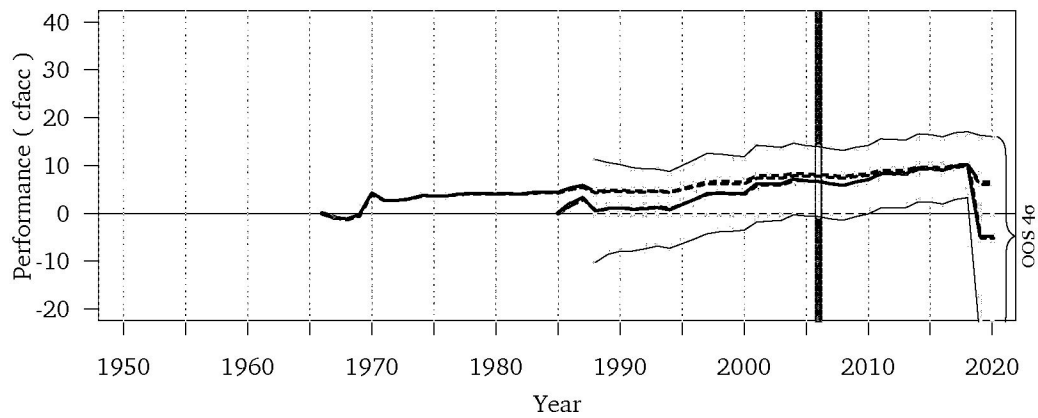
Figure 12: Time-Series of Sentiment (sntm) and Equity Premia
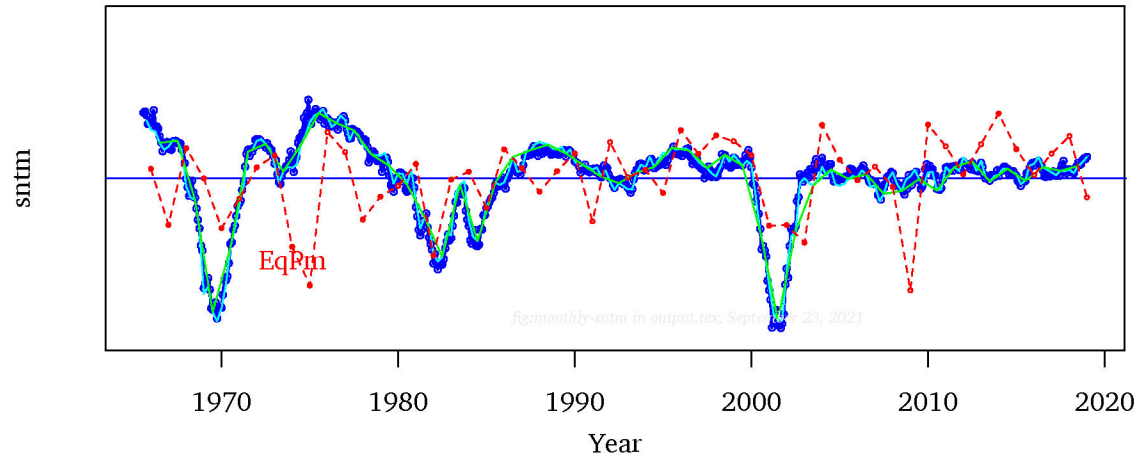
Figure 13: IS and OOS Predictive Performance of HJTZ sntm (monthly)
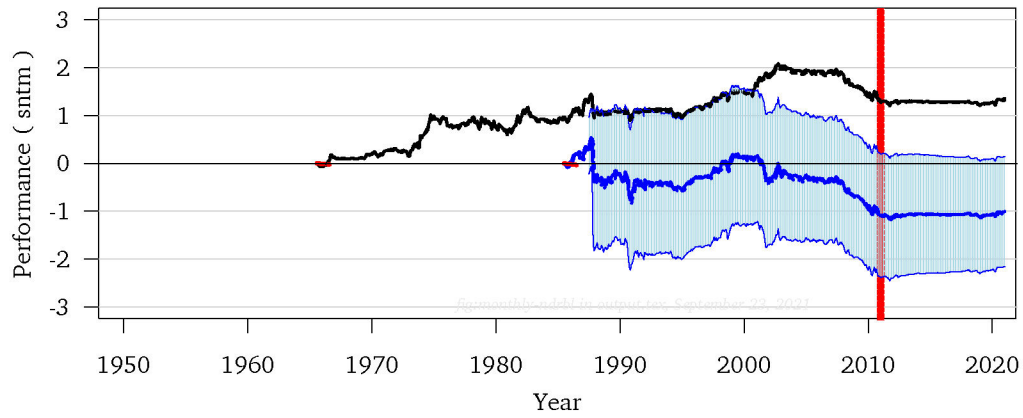
Figure 14: IS and OOS Predictive Performance of JT ndrbl (monthly)

56
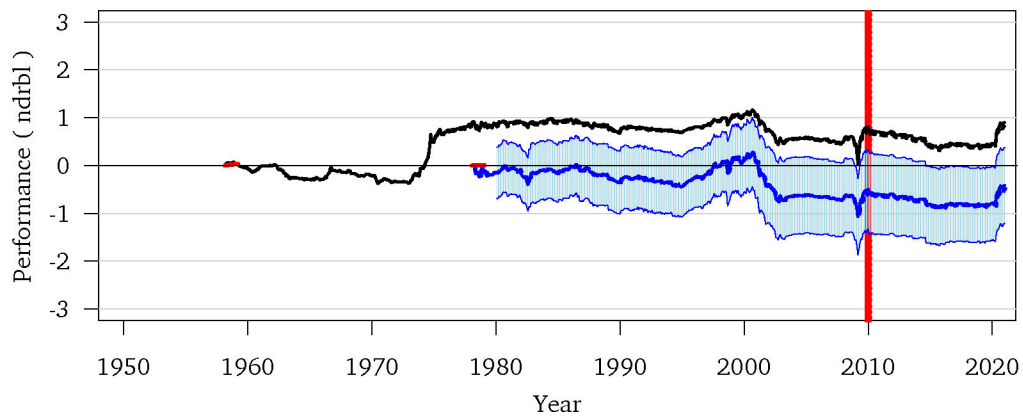
Figure 15: IS and OOS Predictive Performance of JZZ skvw (monthly)

Figure 16: IS and OOS Predictive Performance of KZ tail (monthly)

Figure 17: IS and OOS Predictive Performance of KP fbm (monthly)

57

Figure 18: Time-Series of Implied Volatility (rsvix) and Equity Premia
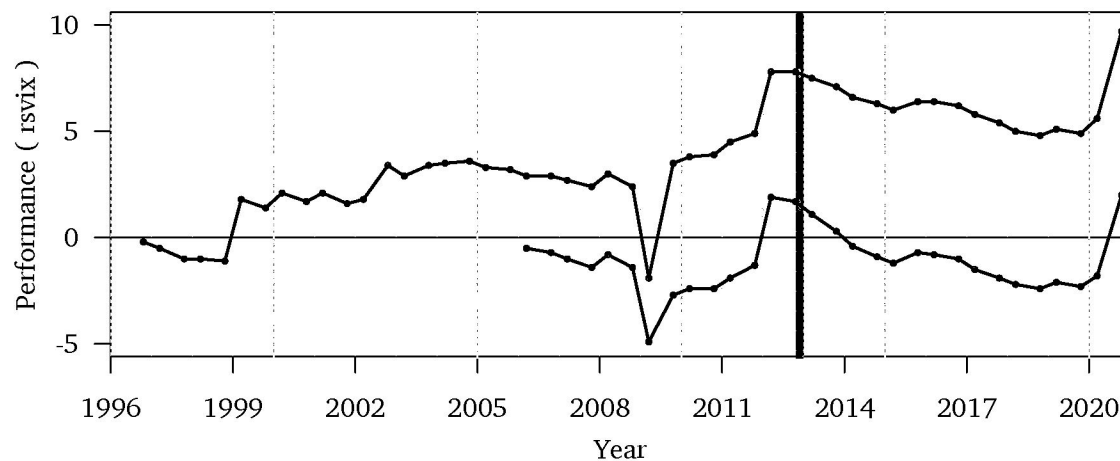
Figure 19: Time-Series of Personal Expenditures Growth (gpce) and Equity Premia

58
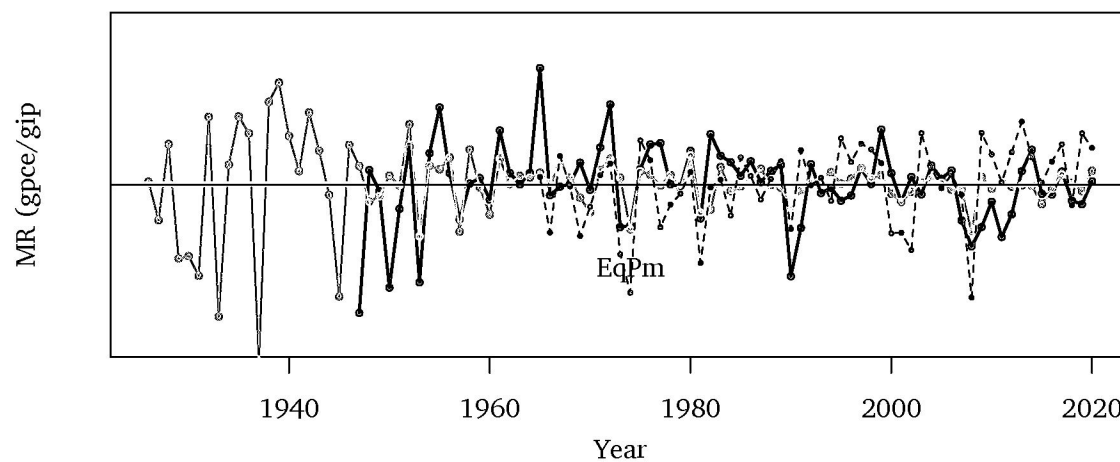
Figure 20: IS and OOS Predictive Performance of MR gpce (annual/jun)
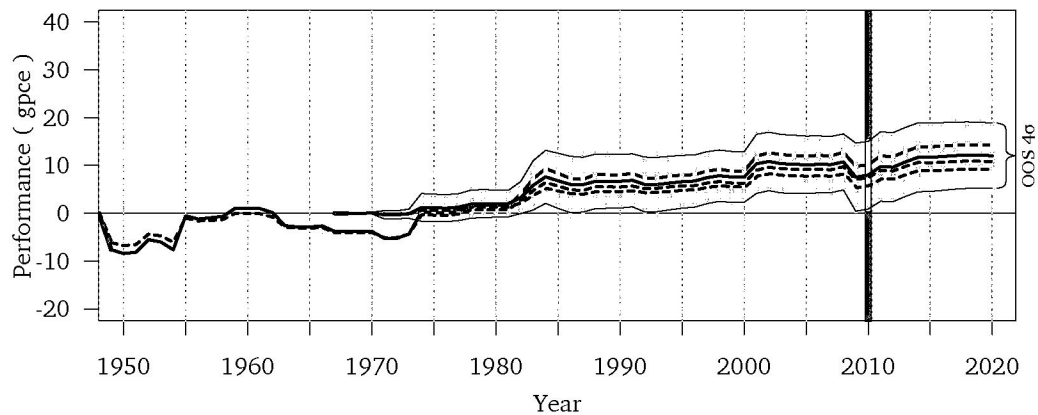
Figure 21: IS and OOS Predictive Performance of NRTZ tchi (monthly)

Figure 22: IS and OOS Predictive Performance of RRZ shtint (monthly)

59
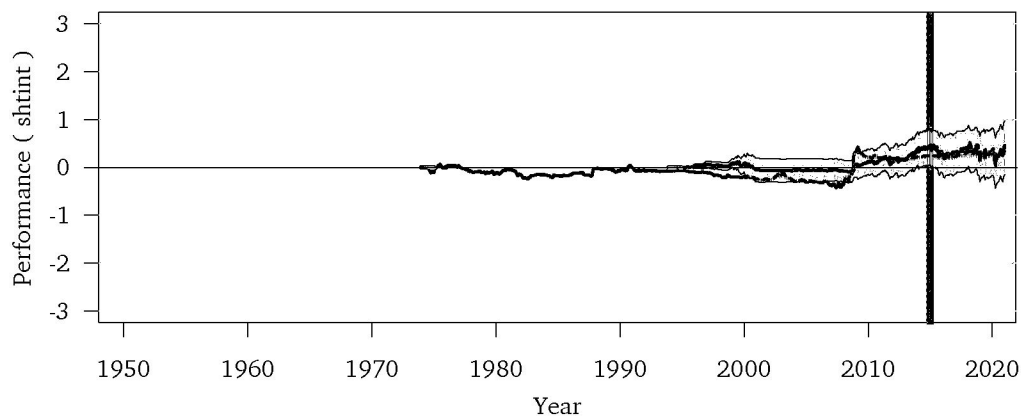
Figure 23: IS and OOS Predictive Performance of Y disag (monthly)

Figure 24: IS and OOS Predictive Performance of FF ltr (monthly)

Figure 25: IS and OOS Predictive Performance of Ca tbl (monthly)

60

Figure 26: IS and OOS Predictive Performance of Co i/k (quarterly)

61
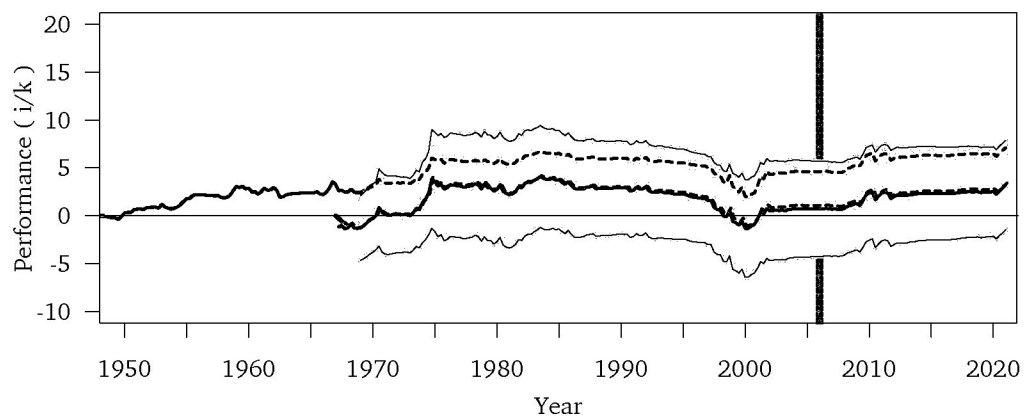
The Equity Risk Premium in 2018

John R. Graham

Fuqua School of Business, Duke University, Durham, NC 27708, USA
National Bureau of Economic Research, Cambridge, MA 02912, USA

Campbell R. Harvey[*]

Fuqua School of Business, Duke University, Durham, NC 27708, USA
National Bureau of Economic Research, Cambridge, MA 02912, USA

ABSTRACT

We analyze the history of the equity risk premium from surveys of U.S. Chief Financial Officers (CFOs) conducted every quarter from June 2000 to December 2017. The risk premium is the expected 10-year S&P 500 return relative to a 10-year U.S. Treasury bond yield. The average risk premium is 4.42% and is somewhat higher than the average observed over the past 18 years. We also provide results on the risk premium disagreement among respondents as well as asymmetry or skewness of risk premium estimates. We also link our risk premium results to survey-based measures of the weighted average cost of capital and investment hurdle rates. The hurdle rates are significantly higher than the cost of capital implied by the market risk premium estimates.

---

*Corresponding author, Telephone: +1 919.660.7768, E-mail: cam.harvey@duke.edu. We appreciate the research assistance of Joy Tong. Version: March 28, 2018

Introduction

We analyze the results of the most recent survey of Chief Financial Officers (CFOs) conducted by Duke University and *CFO* Magazine. The survey closed on December 7, 2017 and measures expectations beginning in the first quarter of 2018. In particular, we poll CFOs about their long-term expected return on the S&P 500. Given the current U.S. 10-year Treasury bond yield, we provide estimates of the equity risk premium and show how the premium changes through time. We also provide information on the disagreement over the risk premium as well as average confidence intervals. Finally, we link the equity risk premium to measures used to evaluate firm's investments: the weighted average cost of capital (WACC) and the investment hurdle rate.

1. Method

2.1 Design

The quarterly survey of CFOs was initiated in the third quarter of 1996.[1] Every quarter, Duke University polls financial officers with a short survey on important topical issues (Graham and Harvey, 2009). The usual response rate for the quarterly survey is 5%-8%. Starting in June of 2000, a question on expected stock market returns was added to the survey. Fig. 1 summarizes the results from the risk premium question. While the survey asks for both the one-year and ten-year expected returns, we focus on the ten-year expected returns herein, as a proxy for the market risk premium.

  The executives have the job title of CFO, Chief Accounting Officer, Treasurer, Assistant Treasurer, Controller, Assistant Controller, or Vice President (VP), Senior VP or Executive VP of Finance. Given that the majority of survey respondents hold the CFO title, for simplicity we refer to the entire group as CFOs.

---

[1] The surveys from 1996Q3-2004Q2 were partnered with a national organization of financial executives. The 2004Q3 and 2004Q4 surveys were solely Duke University surveys, which used Duke mailing lists (previous survey respondents who volunteered their email addresses) and purchased email lists. The surveys from 2005Q1 to present are partnered with *CFO magazine*. The sample includes both the Duke mailing lists and the *CFO* subscribers that meet the criteria for policy-making positions.

2.2 Delivery and response

In the early years of the survey, the surveys were faxed to executives. The delivery mechanism was changed to the Internet starting with the December 4, 2001 survey. Respondents are given four business days to fill out the survey, and then a reminder is sent allowing another four days. Usually, two-thirds of the surveys are returned within two business days.

The response rate of 5-8% could potentially lead to a non-response bias. There are six reasons why we are not overly concerned with the response rate. First, we do not manage our email list. If we deleted the email addresses that had not responded to the survey in the past 12 quarters, our response rate would be in the 15-20% range – which is a good response rate. Second, Graham and Harvey (2001) conduct a standard test for non-response biases (which involves comparing the results of those that fill out the survey early to the ones that fill it out late) and find no evidence of bias. Third, Brav, Graham, Harvey and Michaely (2005) conduct a captured sample survey at a national conference in addition to an Internet survey. The captured survey responses (to which over two-thirds participated) are qualitatively identical to those for the Internet survey (to which 8% responded), indicating that non-response bias does not significantly affect their results. Fourth, Brav et al. contrast survey responses to archival data from Compustat and find archival evidence for the universe of Compustat firms that is consistent with the responses from the survey sample. Fifth, Campello, Graham, and Harvey (2011) show that the December 2008 response sample is fairly representative of the firms included in the commonly used Compustat database. Sixth, Graham, Harvey, Popadak and Rajgopal (2017) update the non-response bias test in a survey of 1,900 CFOs and find no evidence of non-response bias.

2.3 Data integrity

In each quarter, implement a series of rules to ensure the integrity of the data. We have, on average, 351 responses each quarter. There are a total of 24,812 survey observations. There are six key pieces of data: 1) the 10-year forecast (LT); 2) lower 10% of 10-year forecast (LLT); and 3) upper 10% of the 10-year forecast (ULT). We collect the analogous information for the one-year

2

S&P 500 forecasts too (ST). This paper focuses on the 10-year forecasts but the short-term forecasts factor into our data filters.

Our exclusion rules are the following:

1. Delete all missing forecasts, LT, ST
2. Delete all negative LT forecasts (not ST forecasts)
3. Delete all observations that failed to use percentages (forecasts<1.0 for both ST and LT)
4. Delete observations where they failed to annualize, i.e. delete if LT>30% (does not apply to ST)
5. Delete is ST>100%.
6. Delete if lower intervals inconsistent, i.e. LST>=ST or LLT>=LT.
7. Delete if upper intervals inconsistent, i.e. UST<=ST or ULT<=LT.
8. Delete if ST-LST and UST-ST both equal 1 (we call this a lazy answer)
9. Delete if LT-LLT and ULT-LT both equal 1 (again, a lazy answer)

2.4 The 2018 results

The expected market return questions are a subset of a larger set of questions in the quarterly survey of CFOs. The survey usually contains between eight and ten questions. Some of the questions are repeated every quarter and some change through time depending on economic conditions. The historical surveys can be accessed at http://www.cfosurvey.org. Appendix 1 shows the risk premium question in the most recent survey.

While the survey is anonymous, we collect demographic information on seven firm characteristics, including industry, sales revenue, number of employees, headquarters location, ownership (public or private), and proportion of foreign sales.

During the past 18 years, we have collected almost 25,000 responses to the survey. Panel A of Table 1 presents the date that the survey window opened, the number of responses for each survey, the 10-year Treasury bond rate, as well as the average and median expected excess returns. There is relatively little time variation in the risk premium. This is confirmed in Fig. 1a, which displays the historical risk premiums contained in Table 1. The current premium, 4.42%, is above the historical average of 3.64%. The December 2017 survey shows that the expected

annual S&P 500 return is 6.79%% (=4.42%+2.37%) which is slightly below the overall average of 7.11%. The total return forecasts are presented in Fig. 1b.[2]

Panel B of Table 1 presents some summary statistics that pool all responses through the 18 year history of the survey. The overall average ten-year risk premium return is 3.64%.[3] The standard deviation of the individual responses is 2.93% (see Panel B). The standard deviation of the quarterly risk premium estimates is 0.58%.

Figure 1a

10-year forecasted S&P 500 (mean) annual returns over and above the 10-year Treasury bond yield

---
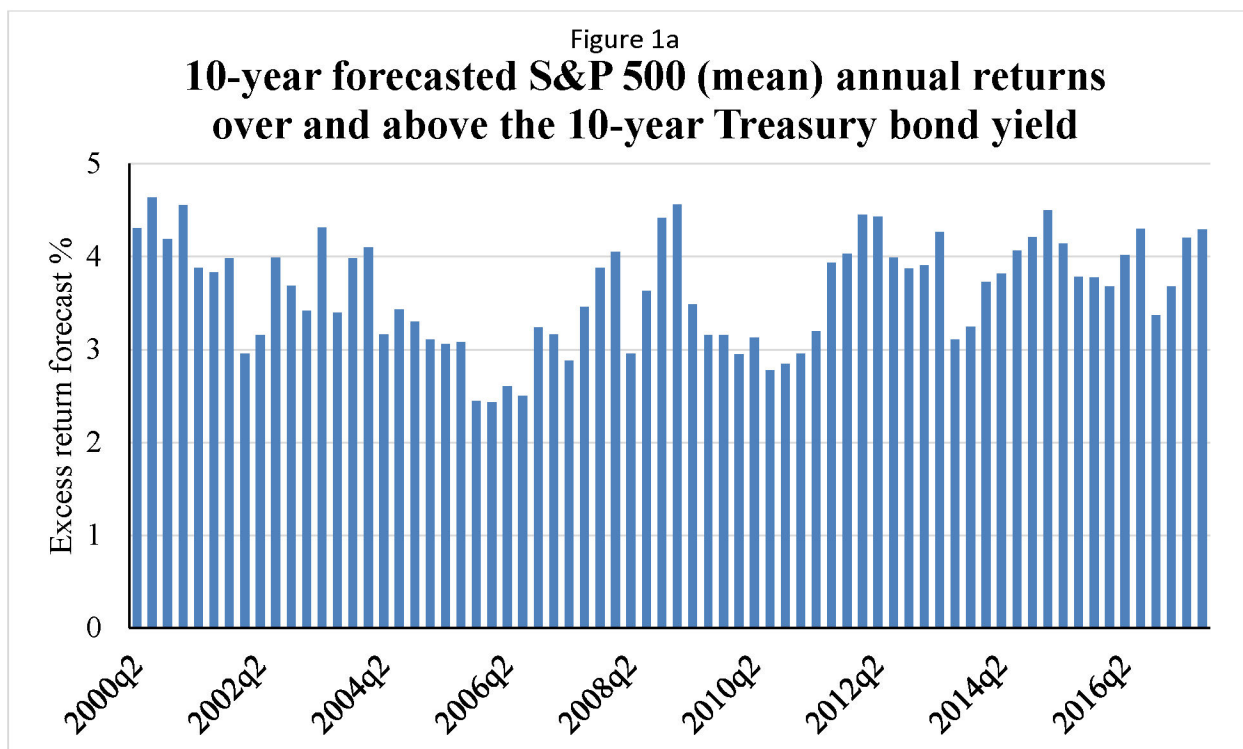
[2] See, for example, Ghysels (1998), Welch (2000, 2001, 2009), Ghysels (1998), Fraser (2001), Harris and Marston (2001), Pástor and Stambaugh (2001), Fama and French (2002), Goyal and Welch (2003), Graham and Harvey (2003), Ang and Bekaert (2005), Fernandez (2004, 2006, 2009) for studies of the risk premium.

[3] Using the Ibbotson Associates data from January 1926 through July 2010, the arithmetic (geometric) average return on the S&P 500 over and above the 30-day U.S. Treasury bill is 7.75% (5.80%). Using data from April 1953-July 2010, the arithmetic (geometric) risk premium is 6.27% (5.12%). The risk premium over the 10 year bond should be reduced by 212 basis points for the arithmetic premium and 174 basis points for the geometric premium. Fama and French (2002) study the risk premium on the S&P 500 from 1872-2000 using fundamental data. They argue that the ex ante risk premia is between 2.55% and 4.32% for 1951-2000 period. Ibbotson and Chen (2001) estimate a long-term risk premium between 4 and 6%. Also see Siegel (1999), Asness (2000), Heaton and Lucas (2000) and Jagannathan, McGratten and Scherbina (2001). A recent treatment is Sharpe and Suarez (2013).

Figure 1b

10-year forecasted S&P 500 total (mean) annualized returns
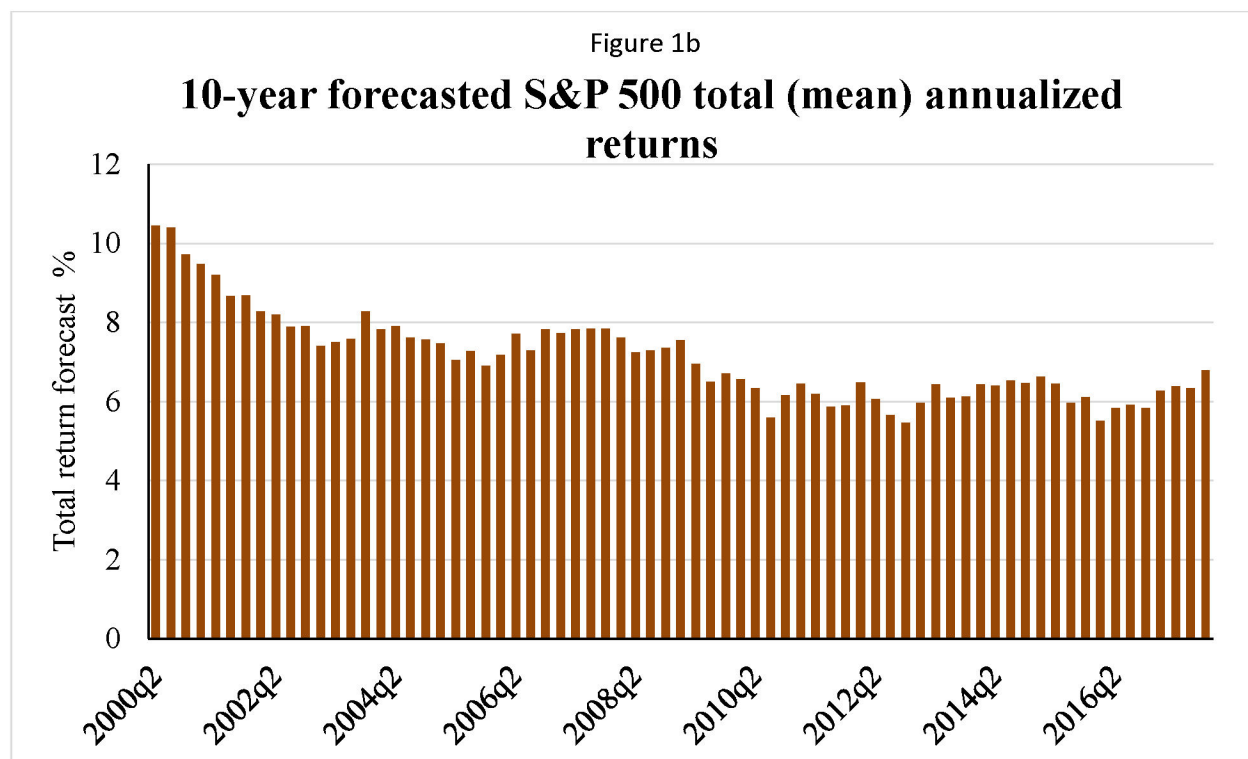
The cross-sectional standard deviation across the individual CFO forecasts in a quarter is a measure of the disagreement or dispersion of the participants in each survey. Dispersion sharply increased during the global financial crisis. The average disagreement in 2005 was 2.39%. Disagreement increased in 2006 to 2.64%. As the crisis began in 2007, disagreement increased to 2.98 by March 2008. The peak disagreement was recorded in February 2009 (4.13%). The most recent observation is 3.49%.

We also report information on the average of the CFOs' assessments of the one in ten chance that the market will exceed or fall below a certain level. In the most recent survey, the worst case total return is +0.85% which is lower than the historic average of 1.45%. The best-case return is 10.97% which is very close to the historical average of 10.90%.

With information on the 10% tails, we construct a probability distribution for each respondent. We use Davidson and Cooper's (1976) method to recover each respondent's probability distribution:

$$\text{Variance} = ([x(0.90)-x(0.10)]/2.65)^2$$

where $x(0.90)$ and $x(0.10)$ represent the 90th and 10th percentiles of the respondent's distribution,

ULT and LLT. Keefer and Bodily (1983) show that this simple approximation is the preferred method of estimating the variance of a probability distribution of random variables, given information about the 10th and 90th percentiles. Like disagreement, the average of individual volatilities peaked in February 2009 at 4.29%. The current level, 3.80%, is very close to the overall average, 3.52%.

There is also a natural measure of asymmetry in each respondent's response. We look at the difference between each individual's 90% tail and the mean forecast and the mean minus the 10% tail. Hence, if the respondent's forecast of the excess return is 6% and the tails are -8% and +11%, then the distribution is negatively skewed with a value of -9% (=5%-14%). As with the usual measure of skewness, we cube this quantity and standardize by dividing by the cube of the individual standard deviation. In every quarter's survey, there is on average negative skewness in the individual forecasts. The average asymmetry -0.66 which is slightly lower than the average of -0.48.

Table 1
Summary statistics based on the responses from the
71 CFO Outlook Surveys from June 2000 to Sept 2017 (Maximums in red, minimums in green)

A. By quarter

#	Survey date	Survey quarter	Number of survey responses	10-year bond yield	Total market return forecast	Average risk premium	Median risk premium	Disagreement (standard deviation of risk premium estimates)	Average of individual standard deviations	Average of individuals' worst 10% market return scenario	Average of individuals' best 10% market return scenario	Skewness of risk premium estimates	Average of individuals' asymmetry	% who forecast negative excess return
1	6/6/2000	2000Q2	209	6.14	10.45	4.31	3.86	3.22				0.95		9.09
2	9/7/2000	2000Q3	188	5.76	10.40	4.64	4.24	3.03				0.83		4.79
3	12/4/2000	2000Q4	243	5.53	9.72	4.19	4.47	2.52				0.53		4.12
4	3/12/2001	2001Q1	140	4.92	9.47	4.55	4.58	2.91				0.78		3.57
5	6/7/2001	2001Q2	208	5.33	9.21	3.88	3.67	2.64				0.58		5.77
6	9/10/2001	2001Q3	199	4.84	8.67	3.83	3.16	2.53				0.13		3.52
7	12/4/2001	2001Q4	279	4.70	8.68	3.98	3.30	2.43				0.61		2.15
8	3/11/2002	2002Q1	233	5.33	8.29	2.96	2.67	2.43	3.28	3.68	12.42	1.06	-0.28	11.16
9	6/4/2002	2002Q2	316	5.04	8.20	3.16	2.96	2.61	3.50	3.00	12.28	1.86	-0.39	10.44
10	9/16/2002	2002Q3	361	3.90	7.89	3.99	4.10	2.31	3.39	3.05	12.03	0.86	-0.25	2.77
11	12/2/2002	2002Q4	285	4.22	7.91	3.69	3.78	2.56	3.23	3.32	11.87	1.24	-0.28	4.91
12	3/19/2003	2003Q1	184	3.98	7.40	3.42	3.02	2.37	3.59	1.95	11.47	0.83	-0.62	4.35
13	6/16/2003	2003Q2	366	3.18	7.50	4.32	4.82	2.34	3.74	2.16	12.07	0.90	-0.33	3.28
14	9/18/2003	2003Q3	167	4.19	7.58	3.39	3.81	2.07	2.83	3.31	10.83	0.35	-0.43	6.59
15	12/10/2003	2003Q4	220	4.30	8.29	3.98	3.70	2.66	3.29	3.40	12.10	1.74	-0.45	2.27
16	3/24/2004	2004Q1	206	3.73	7.83	4.10	4.27	2.37	3.46	2.85	12.02	0.50	-0.29	3.88
17	6/16/2004	2004Q2	177	4.74	7.90	3.16	3.26	2.61	3.10	3.14	11.34	2.14	-0.40	6.21
18	9/10/2004	2004Q3	179	4.19	7.62	3.43	3.31	2.92	3.27	2.61	11.29	2.02	-0.52	8.94
19	12/3/2004	2004Q4	287	4.27	7.57	3.30	3.23	2.66	3.05	3.10	11.17	1.89	-0.37	5.92
20	2/28/2005	2005Q1	272	4.36	7.46	3.10	3.39	2.52	3.06	3.13	11.23	1.29	-0.33	6.62
21	5/31/2005	2005Q2	316	4.00	7.06	3.06	3.00	2.22	3.22	2.39	10.93	0.46	-0.26	6.65
22	8/29/2005	2005Q3	321	4.20	7.28	3.08	2.80	2.61	3.36	2.15	11.06	2.42	-0.52	7.48
23	11/21/2005	2005Q4	338	4.46	6.91	2.45	2.54	2.20	3.48	2.23	11.44	0.41	-0.23	9.76
24	3/6/2006	2006Q1	276	4.74	7.17	2.43	2.26	2.40	3.44	2.07	11.18	1.02	-0.37	8.70
25	6/1/2006	2006Q2	494	5.11	7.72	2.61	2.89	2.74	3.29	3.00	11.70	1.84	-0.24	18.02
26	9/11/2006	2006Q3	460	4.80	7.30	2.50	2.20	2.49	3.32	2.53	11.33	1.32	-0.33	7.83
27	11/21/2006	2006Q4	386	4.58	7.82	3.24	3.44	2.93	3.36	2.94	11.82	1.91	-0.30	6.99
28	3/1/2007	2007Q1	380	4.56	7.72	3.16	3.44	2.39	3.38	2.73	11.67	1.80	-0.39	5.53
29	6/1/2007	2007Q2	419	4.95	7.83	2.88	3.05	2.14	3.21	3.08	11.58	0.56	-0.37	3.58
30	9/7/2007	2007Q3	479	4.38	7.84	3.46	3.62	2.82	3.12	3.33	11.59	1.80	-0.34	5.22
31	11/30/2007	2007Q4	458	3.97	7.85	3.88	4.03	2.75	3.31	2.93	11.70	1.38	-0.32	3.28
32	3/7/2008	2008Q1	381	3.56	7.61	4.05	4.44	2.99	3.21	3.08	11.58	2.23	-0.30	3.94
33	6/13/2008	2008Q2	384	4.27	7.23	2.96	2.73	2.60	3.32	2.44	11.24	1.50	-0.41	9.38
34	9/5/2008	2008Q3	432	3.66	7.29	3.63	3.34	2.79	3.31	2.30	11.06	1.71	-0.42	4.63
35	11/28/2008	2008Q4	534	2.93	7.35	4.42	4.07	3.19	3.73	1.77	11.64	1.94	-0.37	2.81
36	2/26/2009	2009Q1	443	2.98	7.54	4.56	4.02	4.13	4.29	1.18	12.54	1.80	-0.47	5.87
37	5/29/2009	2009Q2	427	3.47	6.96	3.49	3.53	3.12	3.73	1.37	11.26	1.79	-0.42	6.56
38	9/11/2009	2009Q3	536	3.34	6.50	3.16	2.66	2.88	3.87	0.62	10.86	1.82	-0.46	10.82
39	12/11/2009	2009Q4	457	3.55	6.71	3.16	2.45	3.56	3.86	0.64	10.88	2.38	-0.52	9.85
40	2/26/2010	2010Q1	478	3.61	6.56	2.95	2.95	3.28	3.96	0.39	10.86	2.31	-0.68	9.41
41	6/4/2010	2010Q2	444	3.20	6.33	3.13	2.80	3.08	3.90	0.33	10.64	2.61	-0.64	9.91
42	9/10/2010	2010Q3	451	2.81	5.59	2.78	2.19	2.53	4.21	-1.16	9.99	0.77	-0.67	8.65
43	12/10/2010	2010Q4	402	3.32	6.17	2.85	2.68	2.62	3.91	0.26	10.63	1.89	-0.55	10.70
44	3/4/2011	2011Q1	429	3.49	6.45	2.96	2.51	2.92	4.16	-0.27	10.76	2.44	-0.70	8.16
45	6/3/2011	2011Q2	406	2.99	6.18	3.19	3.01	2.90	3.90	0.12	10.45	2.09	-0.68	5.17
46	9/9/2011	2011Q3	397	1.93	5.86	3.93	3.07	3.11	3.79	0.04	10.09	2.41	-0.54	2.02
47	12/16/2011	2011Q4	439	1.86	5.89	4.03	3.14	2.98	4.07	-0.11	10.68	1.91	-0.36	3.42
48	3/1/2012	2012Q1	406	2.03	6.48	4.45	3.97	2.97	4.07	0.30	11.08	2.25	-0.59	2.71
49	5/30/2012	2012Q2	338	1.63	6.06	4.43	4.37	2.96	3.94	0.00	10.42	1.96	-0.59	2.37
50	9/7/2012	2012Q3	675	1.67	5.66	3.99	3.33	3.00	3.66	-0.01	9.67	2.04	-0.58	2.37
51	12/6/2012	2012Q4	325	1.59	5.46	3.87	3.41	2.59	3.69	-0.49	9.25	1.42	-0.62	3.08
52	3/8/2013	2013Q1	418	2.06	5.97	3.91	3.94	2.73	3.84	-0.14	10.02	2.01	-0.64	4.55
53	5/31/2013	2013Q2	300	2.16	6.43	4.27	3.84	2.91	4.02	0.10	10.76	1.63	-0.67	2.67
54	9/5/2013	2013Q3	404	2.98	6.09	3.11	3.02	2.73	3.41	0.75	9.77	1.71	-0.53	6.68
55	12/5/2013	2013Q4	320	2.88	6.13	3.25	3.12	2.95	3.81	0.18	10.26	1.69	-0.50	7.19
56	3/4/2014	2014Q1	291	2.70	6.43	3.73	3.30	2.63	3.32	1.35	10.13	0.64	-0.69	5.15
57	6/5/2014	2014Q2	325	2.59	6.41	3.82	3.41	3.23	3.76	0.50	10.46	1.89	-0.64	7.08
58	9/4/2014	2014Q3	316	2.45	6.52	4.07	3.55	3.33	3.69	0.90	10.68	2.56	-0.60	3.16
59	12/4/2014	2014Q4	398	2.25	6.46	4.21	4.50	2.51	3.79	0.46	10.51	1.22	-0.59	2.26
60	3/13/2015	2015Q1	414	2.13	6.63	4.50	3.87	3.50	3.72	0.81	10.68	1.92	-0.55	5.80
61	6/4/2015	2015Q2	399	2.31	6.45	4.14	3.69	3.03	3.96	0.20	10.68	1.93	-0.72	4.26
62	9/3/2015	2015Q3	376	2.18	5.96	3.78	2.82	3.17	3.48	0.28	9.49	2.72	-0.72	3.99
63	12/3/2015	2015Q4	347	2.33	6.11	3.78	2.67	3.58	3.55	0.54	9.94	1.92	-0.52	9.22
64	3/3/2016	2016Q1	476	1.83	5.51	3.68	3.17	2.55	3.12	1.04	9.29	0.99	-0.34	3.15
65	6/2/2016	2016Q2	472	1.81	5.83	4.02	3.19	3.24	3.52	0.39	9.71	2.14	-0.63	2.54
66	9/8/2016	2016Q3	372	1.61	5.91	4.30	3.64	2.93	3.45	0.64	9.77	1.90	-0.55	1.61
67	12/1/2016	2016Q4	263	2.45	5.82	3.37	3.55	2.69	3.34	0.56	9.39	2.24	-0.68	5.32
68	3/9/2017	2017Q1	278	2.60	6.28	3.68	3.40	3.31	3.27	1.64	10.31	2.29	-0.42	5.76
69	6/8/2017	2017Q2	300	2.19	6.39	4.20	3.81	3.06	3.60	0.90	10.43	2.60	-0.43	2.67
70	9/7/2017	2017Q3	301	2.05	6.34	4.29	3.95	2.77	3.88	0.20	10.48	1.64	-0.52	4.99
71	12/7/2017	2017Q4	212	2.37	6.79	4.42	3.63	3.49	3.80	0.85	10.97	2.06	-0.66	4.25
	Average of quarters		351	3.48	7.11	3.63	3.39	2.81	3.56	1.46	10.90	1.58	-0.48	5.73
	Standard deviation			1.20	1.13	0.58	0.62	0.38	0.33	1.29	0.81	0.66	0.14	3.03

B. By individual responses

	Survey for													
	All dates		24,812	3.32	6.94	3.62	3.33	2.93	3.60	1.33	10.86	1.67	-0.48	5.79

2.5 Risk premia, weighted average cost of capital and hurdle rates

The risk premia that we measure can be used in the calculation of the cost of capital. In a simple capital asset pricing model, the cost of equity capital would be the product of the company's beta times the risk premium along with the risk free rate. The average firm's cost of equity capital would be 6.79% (assuming a beta=1). Assuming the Baa bond yield (4.22%) is the borrowing rate and a 25% marginal tax rate as well as a 75%-25% debt equity split, the weighted average cost of capital would be about 5.89%.

In previous surveys, we have asked CFOs about their weighted average cost of capital. For example, in March of 2011, companies told us that their internally calculated weighted average cost of capital was 10% (averaged across respondents). At the time, the cost of equity capital was similar to today, 6.45%. The bond yields were higher, with the Baa yielding 6.09%. The average firm (assuming average beta is 1.0) without any debt would have a WACC of 6.45%. When debt is introduced, the WACC would be less than 6.45% -- which is sharply lower than the reported 10%.

Why is there such a divergence? One possible reason is that companies consider other factors in calculating the WACC – perhaps a multifactor model.[4] However, there is no evidence supporting this hypothesis. For example, consultants often add a premium for smaller firms based on the results in many research papers of a size premium. However, in our survey the average WACC for firms with less than $25 million in revenue is 10.6% and the WACC for the largest firms with annual revenue greater than $10 billion is 10.5%.

This analysis was replicated in June of 2012 with similar results. Given the same assumptions, the WACC is 5.37%. However, the average self-reported WACC is 9.3%. Again, there is no evidence of a size premium. The smallest firms reported a WACC of 9.3% and the largest firms 9.7%.

The WACC should not be confused with the investment hurdle rate. The WACC is an analytical calculation that combines a model-based cost of equity (such as the CAPM) and the after-tax cost

---

[4] Graham and Harvey (2001) find that most companies use a one-factor CAPM model for cost of capital calculations.

of debt (reflected in current borrowing rates). Given capital constraints, firms often impose a higher hurdle rate on their investments. For example, to allocate capital to an investment that promises a projected return exactly at the firm's WACC is equivalent to accepting a zero net present value project.

The June 2012 survey also asked for the investment hurdle rates. They are much higher than the WACCs. The average rate was 13.5% (compared to the survey-reported WACC of 9.3% and the implied WACC from the survey based risk premium of 5.7%. Similar to the WACC results, there is no evidence that the hurdle rates are higher for small firms. Our evidence shows that the reported average hurdle rate for the smallest firms is 13.1% and for the largest firms the rate is 14.2%.

Even though we know from Graham and Harvey (2001) that three quarters of companies use the capital asset pricing model, there is a large gap between an imputed WACC and the WACC that people use. One way to reconcile this is that companies use very long term averages of equity and bond premia in their calculations. For example, suppose the cost of capital is being calculated with averages from 1926. Ibbotson (2013) reports an arithmetic average return of 11.8% over the 1926-2012 period. The average return on corporate bonds is 6.4%. Using the same parameters, we get an imputed WACC of 9.7%. This is very close to the average reported WACC and, indeed, identical to the WACC reported by the largest firms in our survey.

We learn the following: 1) the equity risk premium is much lower today than averages used over long-periods (e.g. from 1926) such as reported in Morningstar (2013) and Duff and Phelps (2015); 2) the survey questions asking directly about a company's WACC is consistent with companies routinely using long-horizon averages for inputs; and 3) WACCs should be thought as lower bounds – the Hurdle Rates used for actual investment decisions are 400bp higher than the stated WACCs.[5]

---

[5] Also see Sharpe and Suarez (2013) and Jagannathan et al. (2016) who analyze our CFO survey data.

2.6 Recessions, the financial crisis and risk premia

Our survey spans two recessions: March 2001-September 2001 as well as the recession that begins in December 2007 and ends in June 2009. Financial theory would suggest that risk premia should vary with the business cycle. Premiums should be highest during recessions and lowest during recoveries. Previous research has used a variety of methods including looking at ex post realized returns to investigate whether there is business-cycle like variation in risk premia.

While we only have 60 observations and this limits our statistical analysis, we almost no differences. During recessions, the risk premium is 3.52% and during non-recessions, the premium is 3.68%.

2.7 Explaining variation in the risk premium

While we document the level and a limited time-series of the long-run risk premium, statistical inference is complicated by the fact that the forecasting horizons are overlapping. First, we have no way of measuring the accuracy of the risk premiums as forecasts of equity returns. Second, any inference based on regression analysis is confounded by the fact that from one quarter to the next, there are 36 common quarters being forecasted. This naturally induces a moving-average process.

We do, however, try to characterize the time-variation in the risk premium without formal statistical tests. Figure 2 examines the relation between the mean premium and previous one-year returns on the S&P 500.

Figure 2

The mean one year premium and past one-year returns on the S&P 500 index

Past 1-year S&P 500 return %
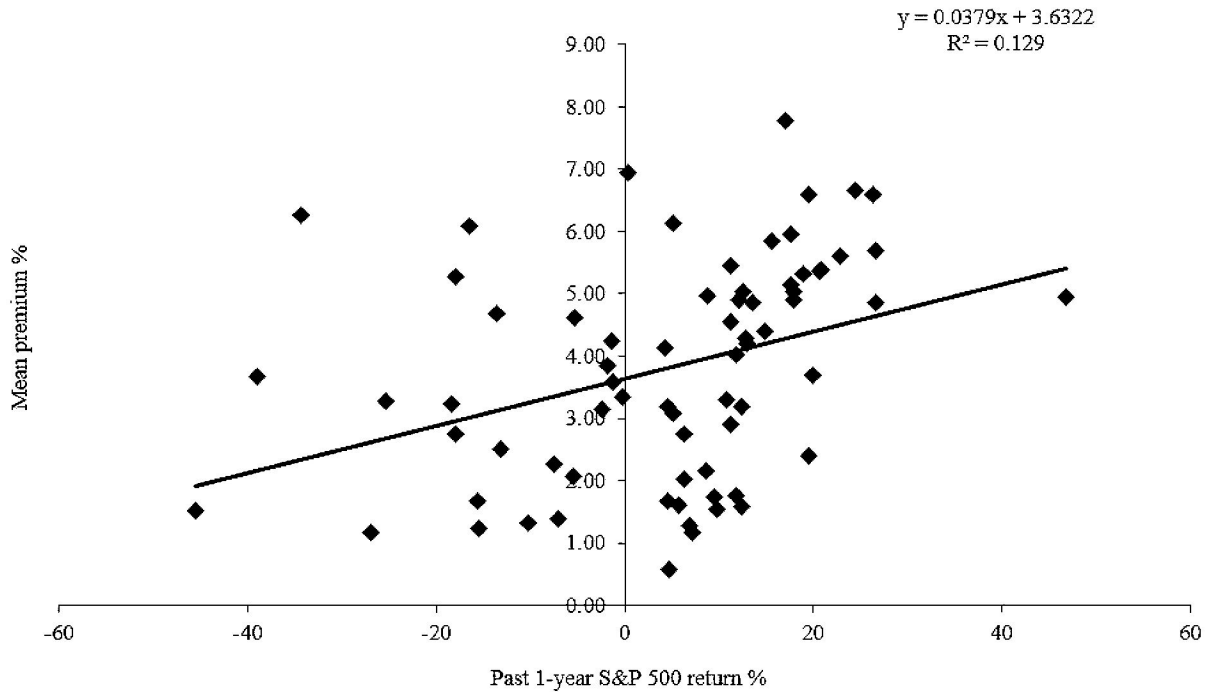
The evidence suggests that there is a weak negative correlation between past returns and the level of the long-run risk premium. This makes economic sense. When prices are low (after negative returns), expected return increase.

An alternative to using past-returns is to examine a measure of valuation. Figure 3 examines a scatter of the mean premium versus the forward price-to-earnings ratio of the S&P 500.

Figure 3

The equity risk premium and the S&P 500 forward price-to-earnings ratio

$y = 0.0386x + 2.9184$
$R^2 = 0.0607$
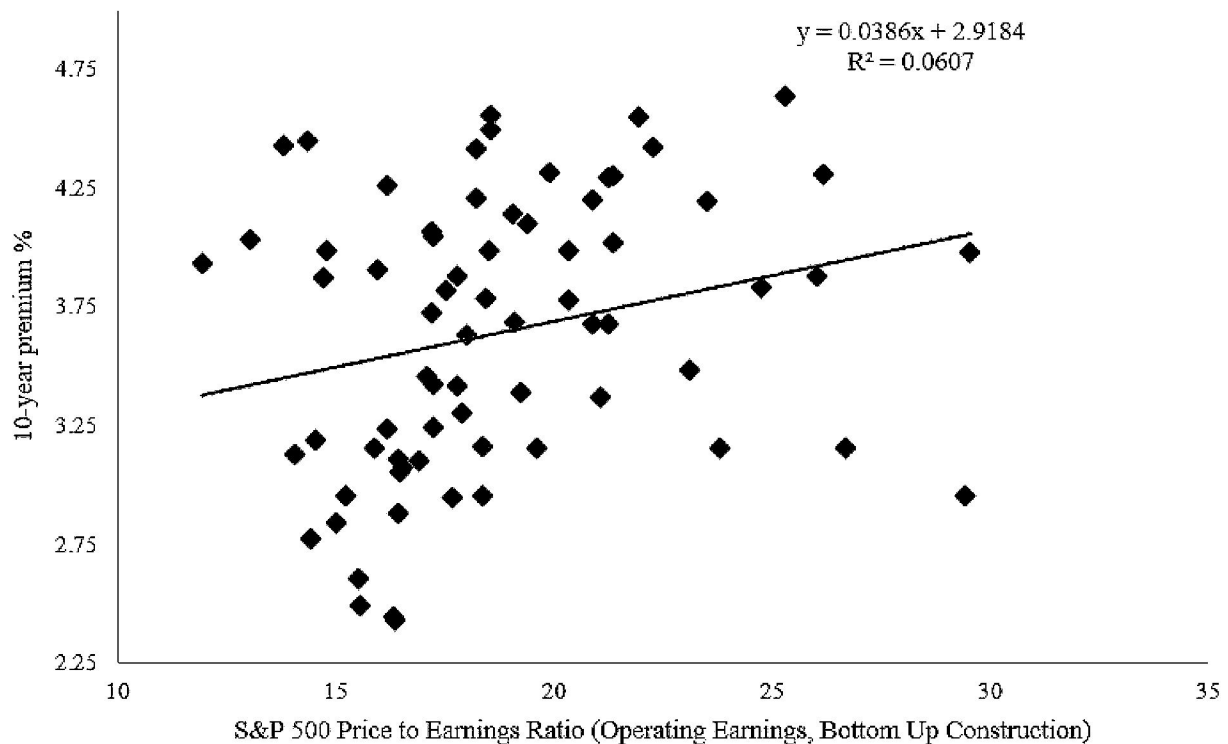
Looking at the data in Figure 3, it appears that the inference may be complicated by a non-linear relation. At very high levels of valuation, the expected return (the risk premium) was low.

We also examine the real yield on Treasury Inflation Indexed Notes. The risk premium is like an expected real return on the equity market. It seems reasonable that there could be a correlation between expected real rates of return stocks and bonds. Figure 4 examines the 10-year on the run yield on the Treasury Inflation Indexed Notes.

Overall, there is a negative correlation of -0.517. However, this correlation is driven by the negative TIPS yields. This is consistent with the idea that in periods of heightened uncertainty, investors engage in a flight to safety and accept low or negative TIPS yields – and at the same time demand a high risk premium for investing in the equity market.

Figure 4

The equity risk premium and the real yield on Treasury Inflation Indexed Notes
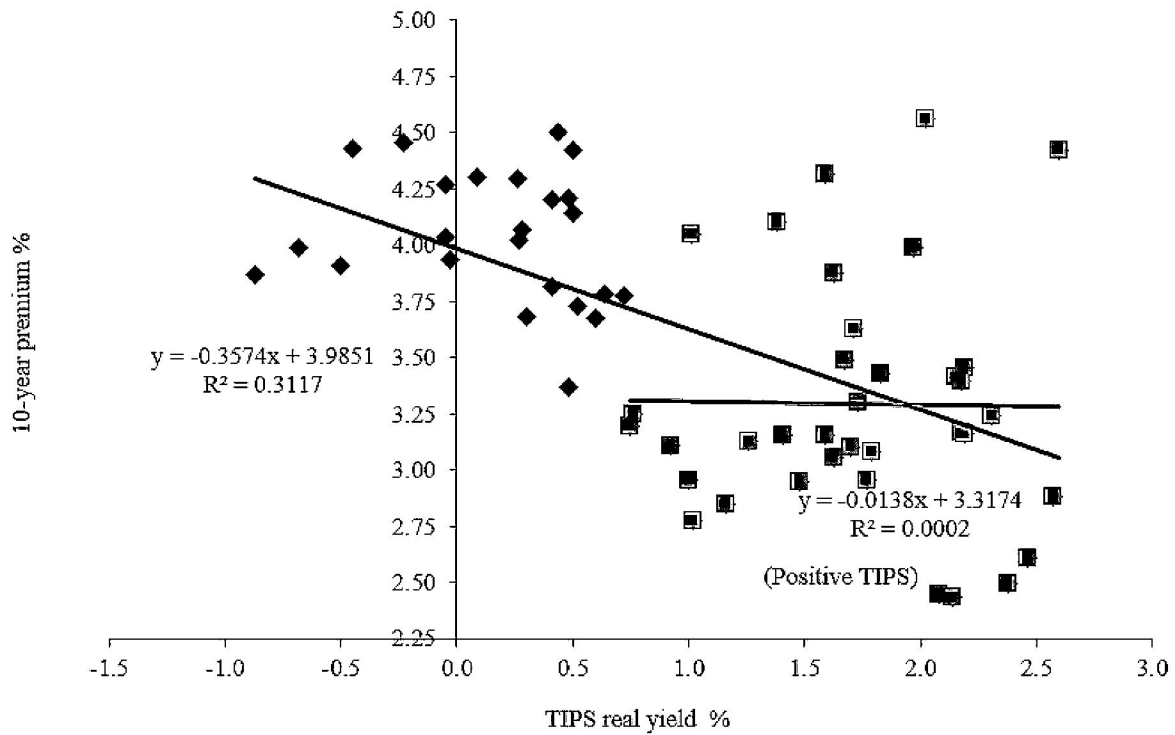
Finally, we consider two alternative measures of risk and the risk premium. Figure 5 shows that over our sample there is evidence of a strong positive correlation between market volatility and the long-term risk premium. We use a five-day moving average of the implied volatility on the S&P index option (VIX) as our volatility proxy. The correlation between the risk premium and volatility is 0.26. If the closing day of the survey is used, the correlation is roughly the same. Asset pricing theory suggests that there is a positive relation between risk and expected return. While our volatility proxy doesn't match the horizon of the risk premium, the evidence, nevertheless, is suggestive of a positive relation. Figure 5 also highlights a strong recent

divergence between the risk premium and the VIX.

Figure 5

The equity risk premium and the implied volatility on the S&P 500 index option (VIX)

Quarter surveyed
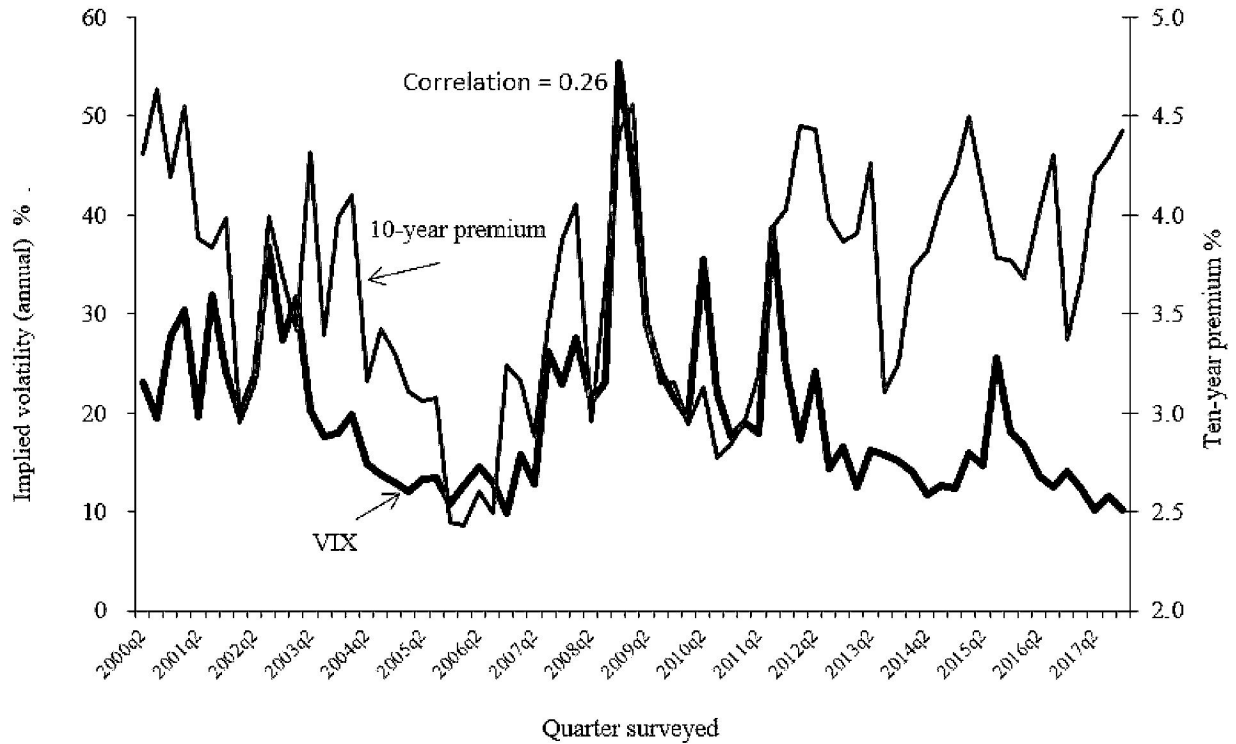
We also consider an alternative risk measure, the credit spread. We look at the correlation between Moody's Baa rated bond yields less the 10-year Treasury bond yield and the risk premium. Figure 6 shows a highly significant relation between the time-series with a correlation of 0.42. Similar to Figure 5, there is a strong recent divergence.

Figure 6

The equity risk premium and credit spreads
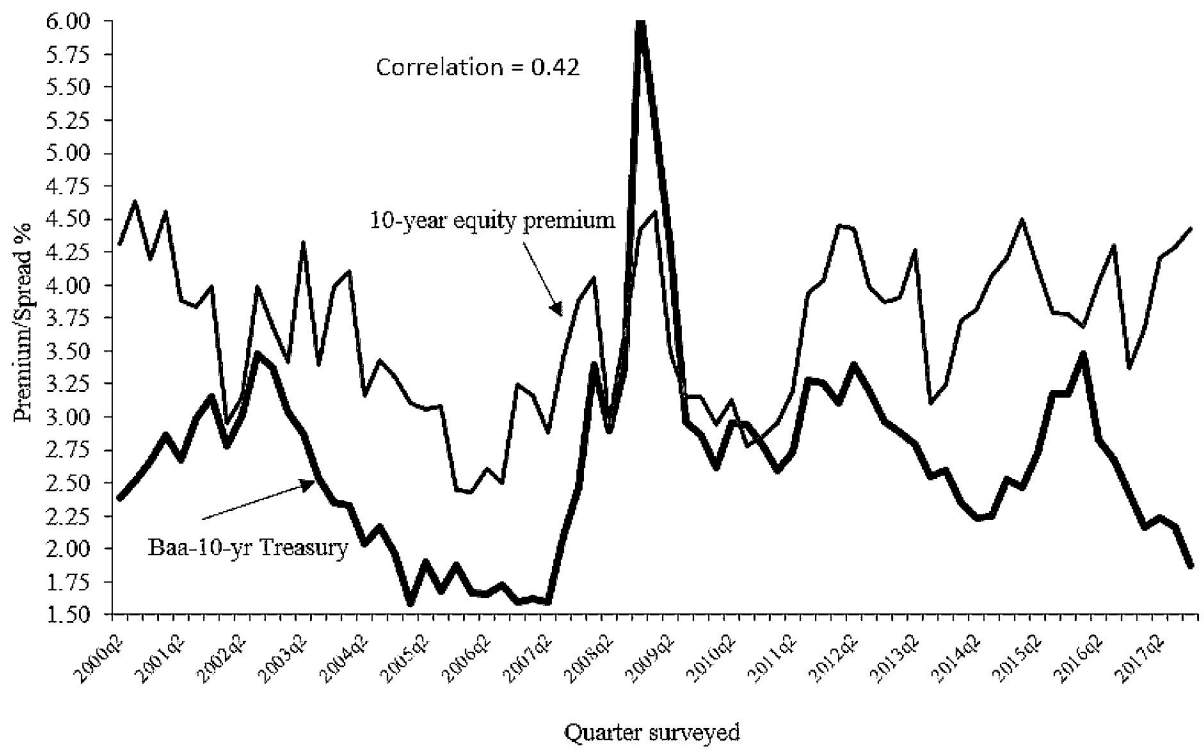
2.8 *Other survey questions*

The June 2016 survey contains a number of other questions. http://www.cfosurvey.org presents the full results of these questions. The site also presents results conditional on demographic firm characteristics. For example, one can examine the CFOs views of the risk premium conditional on the industry in which the CFO works.

2.9 *Risk premium data and corporate policies*

Research by Ben-David, Graham and Harvey (2013) uses the one-year risk premium forecasts as a measure of optimism and the 80% confidence intervals as a direct measure of overconfidence. By linking email addresses that respondents provide to archival corporate data, Ben-David et al. find that the tightness of the confidence intervals is correlated with corporate investment. Overconfident managers invest more.

Campello, Graham and Harvey (2010) use the survey during the financial crisis and the higher risk premiums to examine the implications of financial constraints on the real activities of the

firm. They provide new evidence on the negative impact of financial constraints on firms' investment plans.

Campello, Giambona, Graham and Harvey (2011) use the survey during the financial crisis to study how firms managed liquidity during the financial crisis.

Graham, Harvey and Puri (2013) administer a psychometric test using the survey instrument and link CEO optimism and risk aversion to corporate financial policies.

Graham, Harvey and Puri (2015) use survey data to study how capital is allocated within the firm and the degree to which CEOs delegate decision making to CFOs.

Graham, Harvey and Rajgopal (2005) use survey data to study how managers manipulate earnings. Dichev, Graham, Harvey, and Rajgopal (2013) study earnings quality.

Graham, Harvey, Popadak and Rajgopal (2017) use a similar survey sample to study corporate culture.

2.10 CFO Survey compared to other surveys

Table 2 compares the predictive ability of the Duke-CFO survey with other popular surveys. The table reports the correlations between the current quarter Duke-CFO survey of either optimism about the economy or optimism about the firm's prospects with the subsequent quarter's realization for five surveys: UBS-Gallup, CEO Survey, Conference Board Consumer Confidence, University of Michigan Consumer Confidence and ISM Purchasing Manager's Index. Both of the Duke-CFO optimism measures significantly predict all five of these popular barometers of economic confidence. Related analysis shows that our CFO survey anticipates economic activity sooner (usually one quarter sooner) than do the other surveys.

Table 2
The ability of the Duke CFO survey to predict other surveys

| | Predictive correlations | |
| | Optimism about | Optimism about |
Survey	economy	firm's prospects
UBS-Gallup	0.289	0.380
CEO Survey	0.814	0.824
Conference Board Consumer Confidence	0.513	0.767
University of Michigan Consumer Confidence	0.341	0.253
ISM Purchasing Managers Index	0.694	0.497

3. Conclusions

We provide a direct measure of ten-year market returns based on a multi-year survey of Chief Financial Officers. Importantly, we have a 'measure' of expectations. We do not claim it is the true market expectation. Nevertheless, the CFO measure has not been studied before.

While there is relatively little time-variation in the risk premium, premia are higher during recessions and higher during periods of uncertainty. We also link our analysis to the actual investment decisions of financial managers. We are able to impute the weighted average cost of capital given the CFO estimates of equity risk premia, current corporate bond yields and marginal tax rates. This imputed measure is significantly less than the WACCs that CFOs report using in project evaluation. One way to reconcile this is that CFOs use very long-term averages of equity premia and bond rates when calculating WACCs. We provide evidence on the actual hurdle rates used by companies. These hurdle rates are, on average, 400bp higher than the reported WACCs.

While we have nearly 25,000 survey responses in 18 years, much of our analysis uses summary statistics for each survey. As such, with only 71 unique quarters of predictions and a variable of interest that has a 10-year horizon, it is impossible to evaluate the accuracy of the market excess return forecasts. For example, the November 30, 2007 10-year annual forecast was 7.85% and the realized annual S&P 500 return through December 7, 2017 is approximately 6%. Our analysis shows some weak correlation between past returns, real interest rates and the risk premium. In contrast, there is significant evidence on the relation between two common measures of economic risk and the risk premium. We find that both the implied volatility on the S&P index as

well as a commonly used measure of credit spreads are correlated with our measured equity risk premium.

References

Asness, C. S., 2000, Stocks vs. bonds: Explaining the equity risk premium, *Financial Analysts Journal*, May/June.

Ben-David, I., J. R. Graham, and C. R. Harvey, 2013, Managerial miscalibration. *Quarterly Journal of Economics*, 128:4, 1547-1584. Available at SSRN: http://ssrn.com/abstract=1640552

Brav, A., J. R. Graham, C. R. Harvey, and R. Michaely, 2005, Payout policy in the 21st century, *Journal of Financial Economics* 77:3, 483-529. Available at SSRN: http://ssrn.com/abstract=571046

Campello, M., J. R. Graham and C. R. Harvey, 2010. The real effects of financial constraints: Evidence from a financial crisis, *Journal of Financial Economics* 97, 470-487. Available at SSRN: http://ssrn.com/abstract=1318355

Campello, M., E. Giambona, J. R. Graham and C. R. Harvey, 2011. Liquidity management and corporate investment during a financial crisis, *Review of Financial Studies* 24:6, 1944-1979. Available at SSRN: http://ssrn.com/abstract=144400

Claus, J. and J. Thomas, 2001, Equity premia as low as three percent: Evidence from analysts' earnings forecasts for domestic and international stock markets, *Journal of Finance* 56, 1629-1666.

Davidson, L. B., and D. O. Cooper, 1976, A simple way of developing a probability distribution of present value, *Journal of Petroleum Technology*, September, 1069-1078.

Dichev, I., Graham, J. R., C. R. Harvey, and S. Rajgopal, 2013, Earnings quality: Evidence from the field, *Journal of Accounting and Economics* 56, 1-33. Available at SSRN: http://ssrn.com/abstract=2078535

Duff and Phelps, 2015. *2015 Valuation Handbook*. John Wiley and Sons, NJ,

Fama, E. F. and French, K. R., 2002, The equity premium, *Journal of Finance* 57, 637-659.

Fernandez, P., 2004 Market risk premium: Required, historical and expected, Unpublished working paper, University of Navarra. Available at SSRN: http://ssrn.com/abstract=601761

Fernandez, P., 2006 Market risk premium in 100 textbooks, Unpublished working paper, University of Navarra. Available at SSRN: http://ssrn.com/abstract=1148373

Fernandez, P., 2009 Market risk premium used in 2008 by professors: A survey with 1,400 answers, Unpublished working paper, University of Navarra. Available at SSRN: http://ssrn.com/abstract=1344209

Gebhardt, W. R., C. M. C. Lee, and B. Swaminathan, 2001, Toward an implied cost of capital, *Journal of Accounting Research* 39, 135-176.

Goyal, A. and I. Welch, 2003, Predicting the risk premium, *Management Science* 49, 639-654.

Graham, J. R. and C. R. Harvey, 2001, Theory and practice of corporate finance: Evidence from the field, *Journal of Financial Economics* 60, 187-243. Available at SSRN: http://ssrn.com/abstract=220251

Graham, J. R. and C. R. Harvey, 2003, Expectations of equity risk premia, volatility and asymmetry from a corporate finance perspective, Working paper, Available at SSRN: http://ssrn.com/abstract=292623

Graham, J. R., and C. R. Harvey, 2009, The CFO Global Business Outlook: 1996-2009. http://www.cfosurvey.org.

Graham, J. R., C. R. Harvey and M. Puri, 2013, Managerial attitudes and corporate actions, *Journal of Financial Economics,* 109:1, 103-121. Available at SSRN: http://ssrn.com/abstract=1432641

Graham, J. R., C. R. Harvey and M. Puri, 2015, Capital allocation and delegation of decision-making authority within firms, *Journal of Financial Economics, 115:3,* 449-470. Available at SSRN: http://ssrn.com/abstract=1527098

Graham, J. R., C. R. Harvey, and S. Rajgopal, 2005, The economic implications of corporate financial reporting, *Journal of Accounting and Economics* 40, 3-70. Available at SSRN: http://ssrn.com/abstract=491627

Graham, J. R., C. R. Harvey, J. Popadak, and S. Rajgopal, 2017. Corporate culture: Evidence from the field, Unpublished working paper, Available on SSRN: http://ssrn.com/abstract=2805602

Harris, R. S. and F. C. Marston, 2001, The market risk premium: Expectational estimates using analysts' forecasts, *Journal of Applied Finance* 11, 6-16.

Harvey, C. R., 2001, The specification of conditional expectations, *Journal of Empirical Finance* 8, 573-638.

Jagannathan, R., E. R. McGrattan and A. Scherbina, 2001, The declining U.S. equity premium, *Quarterly Review*, Federal Reserve Bank of Mineapolis.

Jagannathan, R., D. A. Matsa, V. Tarhan and I. Meier. 2016. Why Do Firms Use High Discount Rates?. *Journal of Financial Economics*. 120(3): 445-463.

Keefer, D. L. and S. E. Bodily, 1983, Three-point approximations for continuous random variables, *Management Science* 29, 5 595-609.

Morningstar, 2013. *Ibbotson SBBI 2013 Valuation Yearbook*. Morningstar, Chicago, IL.

Pástor, L. And R. Stambaugh, 2001, The equity premium and structural breaks, *Journal of Finance* 56, 1207-1239.

Poterba, J. M. and Summers, L. H., 1995, A CEO survey of U.S. companies' time horizons and hurdle rates, *Sloan Management Review*, Fall, 43-53.

Sharpe, S. A. and G. A. Suarez, 2013, The insensitivity of investment to interest rates: Evidence from a survey of CFOs, Working paper, Federal Reserve Bank, Washington, DC.

Siegel, J. J., 1999, The shrinking equity premium, *Journal of Portfolio Management*, 10-17.

Welch, I., 2000, Views of financial economists on the equity premium and other issues, *Journal of Business* 73 (October): 501-37.

Welch, I., 2001, The equity premium consensus forecast revisited, Unpublished working paper, Cowles Foundation for Research in Economics, Yale University, New Haven, CT. Available at SSRN: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=285169.

Welch, I., 2009, Views of economists about the equity premium and policy, Unpublished working paper, Brown University. Available at http://welch.econ.brown.edu/academics/equpdate-results2009.html

Appendix A

Excerpt from the Survey Instrument

10. On November 14, 2017 the annual yield on 10-yr treasury bonds was 2.37%. Please complete the following:

a. Over the next 10 years, I expect the average annual S&P 500 return will be:

Worst Case: There is a 1-in-10 chance the actual average return will be less than:

Best Guess: I expect the return to be:

Best Case: There is a 1-in-10 chance the actual average return will be greater than:

[____] %   [____] %   [____] %

b. During the next year, I expect the S&P 500 return will be:

Worst Case: There is a 1-in-10 chance the actual return will be less than:

Best Guess: I expect the return to be:

Best Case: There is a 1-in-10 chance the actual return will be greater than:

[____] %   [____] %   [____] %

Please check one from each category that best describes your company:

a. Industry (choose best option)

- ○ Retail/Wholesale
- ○ Banking/Finance/Insurance/Real Estate
- ○ Mining/Construction
- ○ Transportation & Public Utilities
- ○ Energy
- ○ Services, Consulting
- ○ Agriculture, Forestry, & Fishing

- ○ Public Administration
- ○ Communication/Media
- ○ Tech [software/biotech/hardware]
- ○ Manufacturing
- ○ Healthcare/Pharmaceutical
- ○ Other: [_____]

b. Sales Revenue	**c. Number of Employees**
○ Less than $25 million	○ Fewer than 100
○ $25-$99 million	○ 100-499
○ $100-$499 million	○ 500-999
○ $500-$999 million	○ 1,000-2,499
○ $1-$4.9 billion	○ 2,500-4,999
○ $5-$9.9 billion	○ 5,000-9,999
○ More than $10 billion	○ More than 10,000

d. Where are you personally located?

- ○ Northeast U.S.
- ○ Mountain U.S.
- ○ Midwest U.S.
- ○ South U.S.
- ○ South Atlantic U.S.
- ○ Pacific U.S.

- ○ Canada
- ○ Latin America
- ○ Europe
- ○ Asia
- ○ Africa
- ○ Other

e. Ownership

- ○ Public, NYSE
- ○ Public, NASDAQ/AMEX
- ○ Private
- ○ Government
- ○ Nonprofit

f. Foreign Sales

- ○ 0%
- ○ 1-24%
- ○ 25-50%
- ○ More than 50%

g. Return on assets (ROA=operating earnings/assets) (e.g., -5%, 6.2%)

[____] % Approximate ROA in 2017
[____] % Expected ROA in 2018

The Accuracy, Bias and Efficiency of Analysts' Long Run Earnings Growth Forecasts

RICHARD D.F. HARRIS*

1. INTRODUCTION

Considerable research has now been undertaken into professional analysts' forecasts of companies' earnings in respect of both their accuracy relative to the predictions of time series models of earnings, and their rationality. The evaluation of the reliability of analysts' earnings growth forecasts is an important aspect of research in accounting and finance for a number of reasons. Firstly, many empirical studies employ analysts' consensus forecasts as a proxy for the market's expectation of future earnings in order to identify the unanticipated component of earnings. The use of consensus forecasts in this way is predicated on the assumption that they are unbiased and efficient forecasts of future earnings growth. Secondly, institutional investors make considerable use of analysts' forecasts when evaluating and selecting individual shares. The quality of the forecasts that they employ therefore has important practical consequences for portfolio performance. Finally, from an academic point of view, the performance of analysts' forecasts is interesting because it sheds light on the process by which agents form expectations about key economic and financial variables.

Address for correspondence: Richard D.F. Harris, School of Business and Economics, University of Exeter, Exeter EX4 4PU, UK.
e-mail: R.D.F.Harris@exeter.ac.uk

Nearly all of the research to date, however, has been concerned with analysts' forecasts of quarterly and annual earnings per share.[1] While the properties of analysts' short run forecasts are undoubtedly important in their own right, it is long run expectations of earnings growth that are more relevant for security pricing (see, for instance, Brown et al., 1985). A number of papers have suggested that there is substantial mis-pricing in the stock market as a consequence of irrational long run earnings growth forecasts being incorporated into the market expectation of earnings growth (DeBondt, 1992; La Porta, 1996; Bulkley and Harris, 1997; and Dechow and Sloan, 1997). The evaluation of the performance of analysts' long run forecasts is clearly important as corroborating evidence.

This paper provides a detailed study of the accuracy, bias and efficiency of analysts' long run earnings growth forecasts for US companies. It identifies a number of characteristics of forecast earnings growth. Firstly, the accuracy of analysts' long run earnings growth forecasts is shown to be extremely low. So low, in fact, that they are inferior to the forecasts of a naïve model in which earnings are assumed to follow a martingale. Secondly, analysts' long run earnings growth forecasts are found to be significantly biased, with forecast earnings growth exceeding actual earnings growth by an average of about seven percent per annum. Thirdly, analysts' forecasts are shown to be weakly inefficient in the sense that forecast errors are correlated with the forecasts themselves. In particular, low forecasts are associated with low forecast errors, while high forecasts are associated with high forecast errors. The bias and inefficiency in analysts' long run forecasts are considerably more pronounced than in their short run and interim forecasts.

It is investigated whether analysts incorporate information about future earnings that is contained in current share prices. It is demonstrated that consistent with their short run and interim forecasts, analysts' long run earnings growth forecasts can be enhanced by assuming that each individual firm's earnings will evolve in such a way that its price-earnings ratio will converge to the current market average price-earnings ratio. Analysts therefore neglect valuable information about future earnings that is readily available at the time that their forecasts are made.

The source of analyst inaccuracy is explored by decomposing the mean square error of analysts' forecasts into two systematic components, representing the error that arises as a result of forecast bias and forecast inefficiency, and a random, unpredictable component. In principle, the systematic components of analysts' forecast errors can be eliminated by taking into account the bias and inefficiency in their forecasts. However, it is shown that the bias and inefficiency of analysts' forecasts contribute very little to their inaccuracy. Over eighty-eight percent of the mean square forecast error is random, while less than twelve percent is due to the systematic components. This is an important result for the users of analysts' forecasts since it means that correcting forecasts for their systematic errors can potentially yield only a small improvement in their accuracy.

A second decomposition is used to examine the level of aggregation at which forecast errors are made. The mean square forecast error is decomposed into the error in forecasting average earnings growth in the economy, the error in forecasting the deviation of average growth in each industry from average growth in the economy, and the error in forecasting the deviation of earnings growth for individual firms from average industry growth. It is demonstrated that the error in forecasting average earnings growth in the economy contributes relatively little to analysts' inaccuracy. Over half of total forecast error arises from the error in forecasting deviations of individual firm growth from average industry growth. The error in forecasting deviations of average industry growth from average growth in the economy is smaller, but also significant. However, there is evidence that this pattern is changing over time, with increasing accuracy at the industry level, and diminishing accuracy at the individual firm level.

Finally, it is shown that the performance of analysts' long run earnings growth forecasts varies substantially both with the characteristics of the company whose earnings are being forecast and of the forecast itself. The accuracy, bias and efficiency of analysts' forecasts is examined for sub-samples of firms partitioned by market capitalisation, price-earnings ratio, market-to-book ratio and the level of the forecast itself. The most reliable earnings growth forecasts are low forecasts issued for large companies with low price-earnings ratios and high

market-to-book ratios. Again, this is of considerable practical importance since it offers users of analysts' forecasts some opportunity to discriminate between good and bad forecasts.

The organisation of this paper is as follows. The following section gives a detailed description of the data sources and the sample selection criteria. Section 3 describes the methodology used to evaluate forecast accuracy, bias and efficiency. Section 4 reports the results, while Section 5 concludes.

2. DATA

The sample is drawn from all companies listed on the New York, American and NASDAQ stock exchanges. Data on long run earnings growth expectations are taken from the Institutional Brokers Estimate System (IBES). The data item used in this paper is the 'expected EPS long run growth rate' (item 0), which has been reported by IBES since December 1981, and is defined as:

> the anticipated growth rate in earnings per share over the longer term. IBES Inc. requests that contributing firms focus on the five-year interval that begins on the first day of the current fiscal year and make their calculations based on projections of EPS before extraordinary items.

The expected long term growth rate is therefore taken to be the forecast average annual growth in earnings per share before extraordinary items, over the five year period that starts at the beginning of the current fiscal year.[2] The measure used in this paper is the median forecast calculated and reported in April of each year, t. The analysis was also conducted using the mean forecast, but the quantitative results are virtually identical, and the qualitative conclusions unchanged.[3]

Only December fiscal year end companies are included in the sample and so the use of the consensus forecast reported in April should ensure that the previous fiscal year's earnings are public information at the time that the individual forecasts that make up the consensus forecast are made (see Alford, Jones and Zmijewski, 1994). Restricting the sample to December fiscal year-end companies ensures that observations for a particular fiscal year span the same calendar period, thus allowing the identification of macroeconomic shocks that contemporaneously affect the earnings of all firms.

Actual growth in earnings is calculated using data on earnings per share, excluding extraordinary items, taken from the Standard and Poor's Compustat database (item EPSFX). Average annual earnings growth is computed as the average change in earnings over each five year period, from December of year $t-1$ to December of year $t+5$, scaled by earnings in December of year $t-1$. The need for five years' subsequent earnings growth data limits the sample period to the eleven years 1982–92. Data on a number of other variables are also used in the analysis. The share price and market capitalisation are both taken at the end of April of year t (Compustat items PRCCM and MKVALM). The market price-earnings ratio, used to test whether information contained in the share price is incorporated in analysts' forecasts, is computed as the price at the end of April in year t (item PRCCM) divided by earnings per share in the fiscal year ending December $t-1$ (item EPSFX). The market-to-book ratio is computed as the market value of the company in April of year t (item MKVALM) divided by the book value of the company in the fiscal year ending December of year $t-1$ (item CEQ).

There are a total of 7,660 firm-year observations that satisfy the data requirements for all the variables used in the analysis, and that have a December fiscal year-end. However, for 658 of these, earnings reported at the end of the preceding fiscal year are zero or negative. These are omitted from the sample since forecast growth has no natural interpretation when earnings in the base year are non-positive.[4] When initial earnings are close to zero, actual growth in earnings may take extreme values, resulting in outliers that have a disproportionately high degree of influence on the least squares regression results. There is no immediately obvious way to circumvent this problem without dropping some observations from the sample. The approach most commonly adopted is to omit observations for which the calculated growth rate, the forecast growth rate or the forecast error is above a certain threshold in absolute value, or for which calculated initial earnings are below a certain level. For instance, Fried and Givoly (1982) truncate observations for which forecast error exceeds 100%. Elton et al. (1984) include in their sample only those companies for which initial earnings are above 0.20 dollars per share. O'Brien (1988), in order to test the robustness of her results to outliers, also uses 0.20 dollars as a threshold value.

Capstaff et al. (1995) omit observations for which forecast earnings growth or forecast error exceeds 100%, while Capstaff et al. (1998) exclude companies for which forecast earnings growth or actual earnings growth exceeds 100%. In this paper, all observations for which actual earnings growth or forecast earnings growth exceeds 100% in absolute value are omitted from the analysis, reducing the sample by a further 336 firm-year observations. The final pooled sample comprises 6,666 firm-year observations.[5]

3. METHODOLOGY

(i) Forecast Accuracy

The metric used to evaluate forecast performance is the forecast error, defined as the difference between actual and forecast earnings growth:

$$fe_{it} = g_{it} - g_{it}^f \tag{1}$$

where fe_{it} is the forecast error for firm i corresponding to the forecast made at date t, g_{it} is actual earnings growth over the five year forecast period and g_{it}^f is forecast five year earnings growth. Forecast accuracy is evaluated using the mean square forecast error, which is computed in each year t as:

$$MSFE_t = \frac{1}{N}\sum_{i=1}^{N}(g_{it} - g_{it}^f)^2. \tag{2}$$

The mean square forecast error for the pooled sample is computed over all firms and years. The mean square forecast error was chosen in preference to the mean absolute forecast error to maintain consistency with the subsequent analysis which uses the former measure rather than the latter. However, it should be noted that the use of the mean square forecast error is consistent with a quadratic loss function of risk averse economic agents (see Theil, 1964; and Mincer and Zarnovitz, 1969). It can be reported that the conclusions drawn about forecast accuracy are not sensitive to the choice of measure.

As a benchmark against which to compare the accuracy of analysts' long run forecasts, the performance of two 'naïve'

forecasts is also considered. The first is the forecast generated by a martingale model of earnings, in which expected earnings growth is zero. The second is the forecast generated by a sub-martingale model, in which expected earnings is equal to a drift parameter that is identical for all firms. In each forecast year, the common drift parameter is set equal to the average growth rate in earnings over all firms, over the previous five year period.[6] This choice of naïve forecasts is motivated by the early evidence on the time series properties of earnings, which suggests that annual earnings follow a random walk, or a random walk with drift (see, for instance, Brooks and Buckmaster, 1976; or Foster, 1977). Although more recent evidence finds that annual earnings may have a mean reverting component (see Ramakrishnan and Thomas, 1992), the martingale and sub-martingale models of earnings nevertheless provide simple alternative models that are approximately consistent with the reported evidence.

(ii) Forecast Bias

In order for a forecast to be unbiased, the unconditional expectation of the forecast error must be zero. If the average forecast error is greater than zero then analysts are systematically over-pessimistic (since their forecasts are on average exceeded) while if the average forecast error is less than zero analysts are systematically over-optimistic (since their forecasts are on average unfulfilled). Unbiasedness is tested using the mean forecast error, which is computed in each year t as:

$$\text{MFE}_t = \frac{1}{N} \sum_{i=1}^{N} (g_{it} - g_{it}^f). \tag{3}$$

The mean forecast error for the pooled sample is computed over all firms and years. The hypothesis that the mean forecast error is zero is tested using the standard error of the mean forecast error across all firms and years for the pooled sample, and across all firms for each of the annual samples.

(iii) Forecast Efficiency

A forecast is efficient if it optimally reflects currently available information, and is therefore associated with a forecast error that

is unpredictable. If a forecast is strongly efficient, the forecast error is uncorrelated with the entire information set at time t. Strong efficiency is a stringent condition, and so more usually forecasts are instead tested for weak efficiency, which requires that the forecast error is uncorrelated with the forecast itself (see Nordhaus, 1987). Weak efficiency is tested by estimating the following regression:

$$g_{it} = \alpha + \beta g_{it}^f + v_{it}. \tag{4}$$

Under the null hypothesis that analysts' forecasts are weakly efficient, the intercept, α, should be zero, while the slope coefficient, β, should be unity. If β is significantly different from one then conditioning on the forecast itself, the forecast error is predictable.[7] If β is significantly less than one then analysts' forecasts are too extreme, in the sense that high forecasts are associated with high forecast errors, while low forecasts are associated with low forecast errors. If β is significantly greater than one then forecasts are too compressed.

(iv) The Incremental Information Content of Price-Earnings Based Forecasts

A stronger form of forecast efficiency can be tested by examining whether analysts' forecasts incorporate particular sources of publicly available information. One such source of information is the current share price. In an efficient market, the share price is the present discounted value of all rationally expected future economic earnings of the company, and hence it should reflect, *inter alia*, the market's expectation of long run earnings growth. To extract the information about future earnings embodied in the share price, some assumption must be made about the company's cost of equity, or risk. The simplest assumption is that all companies face the same constant cost of equity in the long run, so that the earnings of each company evolve in such a way that its price-earnings ratio converges to the current market average price-earnings ratio. The earnings growth forecast that is implicit in this assumption can then be used to supplement the analysts' earnings growth forecast in the following regression:

$$g_{it} = \alpha + \beta g_{it}^f + \gamma g_{it}^p + v_{it}, \tag{5}$$

where

$$g_{it}^p = \frac{p_{it}/pe_{mt} - e_{it}}{e_{it}}, pe_{mt} = \frac{1}{N}\sum_{i=1}^{N}\frac{p_{it}}{e_{it}}$$

and p_{it} is the share price of firm i at time t. If analysts incorporate all information contained in the current share price, the coefficient, γ, should be zero (see Capstaff et al., 1995 and 1998). Naturally, the assumption that all firms have the same long run price-earnings ratio is a strong simplification, and a superior forecast would almost certainly be obtained by assuming that price-earnings ratios differ between industries. Nevertheless, the assumption of a single market-wide long run price-earnings ratio has been shown to forecast earnings growth over shorter horizons (see, for instance, Ou and Penman, 1989).

(v) Forecast Error Decomposition

In order to analyse the source of analysts' forecast errors, two decompositions of the mean square forecast error are used. The first decomposes the mean square forecast error into systematic and unsystematic components. The systematic component is further divided into a component due to forecast bias and a component due to forecast inefficiency. In each year t, the decomposition of the MSFE is given by:

$$\text{MSFE}_t = \frac{1}{N_t}\sum_{i=1}^{N_t}(g_{it} - g_{it}^f)^2 = (\overline{g}_t - \overline{g}_t^f)^2 + (1 - \beta_t)^2\sigma_{g^f t}^2 + (1 - \rho_t^2)\sigma_{gt}^2$$

$$(6)$$

where N_t is the sample size in year t, $\overline{g}_t$ and $\overline{g}_t^f$ are the average values of g_{it} and g_{it}^f, β_t is the slope coefficient from regression (4), above, ρ_t is the correlation coefficient between g_{it} and g_{it}^f, and $\sigma_{g^f t}^2$ and σ_{gt}^2 are the variances of g_{it} and g_{it}^f. The first term in the decomposition gives the error that is due to the inability of analysts to forecast earnings growth for the whole sample. When computed over all years, it is therefore a measure of the error that is due to forecast bias. The second term captures the error that is due to forecast inefficiency. Together, these two terms capture the systematic error in analysts' forecasts. In contrast, the third term captures the component of the error that is purely random. This decomposition is particularly useful since it reveals

to what extent forecasts can be improved through 'optimal linear correction' procedures (see Mincer and Zarnovitz, 1969; and Theil, 1966). For instance, if the main component of mean square error is systematic, rather than random, then assuming that the data generating process for both the actual data and the forecast data remains constant, the accuracy of analysts' forecasts can be substantially improved by using the predicted values from regression (4), above, rather than the forecasts themselves. The extent to which this reduces the inaccuracy of the forecasts depends upon the fraction of the mean square forecast error that is due to the systematic component.

The second decomposition breaks the mean square forecast error into economy, industry and firm components. The decomposition of the MSFE is given each year t by:

$$\text{MSFE}_t = \frac{1}{N} \sum_{i=1}^{N_t} (g_{it} - g_{it}^f)^2$$

$$= (\overline{g}_t - \overline{g}_t^f)^2 + \frac{1}{N_t} \sum_{j=1}^{J_t} N_{jt} [(\overline{g}_{jt} - \overline{g}_t) - (\overline{g}_{jt}^f - \overline{g}_t^f)]^2 \qquad (7)$$

$$+ \frac{1}{N_t} \sum_{i=1}^{N_t} [(g_{it} - \overline{g}_{jt}) - (g_{it}^f - \overline{g}_{jt}^f)]^2,$$

where J_t is the number of industries in the sample, N_{jt} is the number of firms in industry j, $\overline{g}_{jt}$ and $\overline{g}_{jt}^f$ are the average values of g_{it} and g_{it}^f in industry j. The decomposition has the following interpretation. As before, the first term measures the error that is due to analysts' inability to forecast the average growth for the whole sample, which in this context may be interpreted as their inability to forecast earnings growth for the economy. The second term measures the error that is due to an inability to forecast the deviation of average growth in an industry from average growth in the economy. The third term measures the error that is due to an inability to forecast deviation of individual firm growth from average growth in its industry. The decomposition for the pooled sample is computed by taking the weighted average of the decomposition for the annual samples, with weights proportional to the sample size each year. Such a decomposition is useful because it reveals the level of aggregation at which

forecast errors are made, and may reflect the particular approach used to generate earnings growth forecasts (see Elton, Gruber and Gultekin, 1984). In the present study, each industry is defined by a two digit SIC code. This yields a total of 56 industries, with an average of about twelve firms in each industry. The use of three digit SIC codes yields a large number of industries that comprise only a single firm. In these cases, the firm-specific error and industry specific error are not separately identifiable, and are reflected in the third component of the decomposition. The effect of using two digit, rather than three digit SIC codes is therefore to increase the firm specific error and reduce the industry specific error.

For both decompositions, it is convenient to express each term as a percentage of the total mean square forecast error. For the pooled samples, the mean square forecast error components are averaged over the individual years, with weights proportional to the sample size each year.

(vi) The Performance of Analysts' Forecasts Conditional on Firm and Forecast Characteristics

In order to explore possible heterogeneity in the performance of analysts' long run earnings growth forecasts, the sample is partitioned by various characteristics of the firm whose earnings are being forecast and of the forecast itself. Specifically, the sample is split into equally sized quintiles on the basis of market capitalisation, market-to-book ratio, price-earnings ratio and the level of the forecast itself. Forecast accuracy, bias and efficiency is then examined for each sub-sample. Forecast accuracy is measured by the mean square forecast error given by (2), forecast bias is measured by the mean forecast error given by (3), while forecast efficiency is measured by the estimated slope parameter in regression (4).

In order to identify the marginal effects of each of the firm and forecast characteristics on forecast accuracy, bias and weak form efficiency, the following regressions are estimated:

$$(g_{it} - g_{it}^f)^2 = \alpha + \beta_1 \ln m_{it} + \beta_2 mb_{it} + \beta_3 pe_{it} + \beta_4 g_{it}^f + v_{it}, \quad (10)$$

$$g_{it} - g_{it}^f = \alpha + \beta_1 \ln m_{it} + \beta_2 mb_{it} + \beta_3 pe_{it} + \beta_4 g_{it}^f + v_{it} \quad (11)$$

and

$$(g_{it}^f - \overline{g}_t^f)[(g_{it} - \overline{g}_t) - (g_{it}^f - \overline{g}_t^f)] = \alpha + \beta_1 \ln m_{it} + \beta_2 mb_{it}$$
$$+\beta_3 pe_{it} + \beta_4 g_{it}^f + v_{it}, \tag{12}$$

where $\ln m_{it}$ is the natural logarithm of the market capitalisation of firm i at the beginning of the forecast period, mb_{it} is the market-to-book ratio and pe_{it} is the price-earnings ratio. The dependent variables in the three regressions are the summands in (a) the mean square forecast error, (b) the mean forecast error and (c) the estimated covariance between $(g_{it} - g_{it}^f)$ and g_{it}^f.[8]

(vii) Estimation Procedure

In order to allow for time specific market wide shocks, each of the regression equations (4), (5), (9), (10), (11) and (12) is estimated by OLS, including fixed time effects. However, inference based on OLS estimates of the variance-covariance matrix of the disturbance term may be misleading since both heteroscedasticity and cross-sectional correlation are likely to be present in the data. One potential solution is to use GLS, in which the heteroscedasticity and cross-section correlation are parameterised and estimated. However, in the present case, GLS is infeasible since the number of cross-section observations is large relative to the number of time series observations. This paper employs instead the non-parametric approach of Froot (1989), which is robust to both contemporaneous correlation and heteroscedasticity. This involves partitioning the data by a two digit SIC code and assuming that the intra-industry correlation is zero. This then allows the consistent estimation of the parameter covariance matrix. The Froot estimator is modified using the Newey-West (1987) procedure in order to allow for the serial correlation in the regression error term that is induced by the use of overlapping data.

4. RESULTS

(i) Forecast Accuracy

Panel A of Table 1 reports the mean square forecast error, given by (2), for the pooled sample and for each individual year. It also