Comment on Cakici, Fieberg, Neumaier, Poddig, and Zaremba: Pockets of predictability: A replication

Leland E. Farmer
University of Virginia

Lawrence D. W. Schmidt
Massachusetts Institute of Technology

Allan Timmermann
*University of California, San Diego**

March 26, 2024

ABSTRACT

A coding error in Farmer et al. (2023) (FST) meant that future information leaked into the classification scheme used to identify pockets with elevated return predictability. While the underlying return forecasts are unaffected by the error, their usage to identify predictability pockets *ex ante* becomes noisier under the corrected classification scheme. A simple modification that smooths the underlying forecasts prior to pocket classification retains all the main conclusions of FST including stronger out-of-sample return predictability inside pockets, economic gains from exploiting this return predictability, and the greater alignment of return predictability pockets with a sticky expectations model than with existing asset pricing models.

Key words: Pockets of predictability; Bayesian estimation, kernel regressions, investment performance, sticky expectations.

*Allan Timmermann (corresponding author, atimmermann@ucsd.edu) is at University of California, San Diego. Leland E. Farmer (lef2u@virginia.edu) is at University of Virginia and Lawrence D. W. Schmidt (ldws@mit.edu) is at Massachusetts Institute of Technology.

We would like to thank Shaozhe Zhao for helpful feedback and comments. The replication code for this comment can be downloaded from Dropbox here. We have read the *Journal of Finance* disclosure policy and have no conflicts of interest to disclose.

I. Introduction

Farmer et al. (2023) (FST) argue that predictability of U.S. stock market returns is concentrated in blocks in time ("pockets"). Their empirical analysis uses local kernel regressions to capture time variation in return predictability and an auxiliary classification scheme to identify pockets. Next, they examine whether existing asset pricing models can generate pockets of predictability. Finding that conventional models struggle to generate sufficient time variation in return predictability concentrated around pockets, they propose an alternative asset pricing model with sticky expectations which seems more consistent with the empirical evidence Finally, they document empirical correlations between analysts' forecast errors and their return forecasts consistent with the sticky expectations mechanism.

In a replication study of FST, Cakici et al. (2024) (CFNPZ) identify a coding error which implies that the original approach for detecting pockets of predictability in FST would not have been implementable in real time. They also present evidence that some of the results in FST become substantially weaker after addressing the error.

In this note, we discuss the error, then propose a modest change to the empirical specification in FST which preserves the key results in the original paper by making the out-of-sample forecasting approach more robust and, finally, present updated results for the main specifications in the FST paper.

Our analysis begins by discussing the coding error detected in the replication files by CFNPZ, and its implications for the results from the exact method described in the paper. The error does not affect the underlying time-varying (kernel) return forecasts which are truly one-sided and hence do not benefit from any leakage of future information. Only that part of the algorithm that classifies periods into predictability pockets gets affected by the leakage of future information through the usage of a two-sided window. Hence, only the question of whether the pockets were discoverable in real time is affected. The coding error also does not affect the methodological contributions in the paper such as the introduction of the integrated R^2 (IR^2) measure to capture local return predictability or the analytical results demonstrating differences in the ability of conventional asset pricing models and a sticky expectations model to generate local return predictability.

Next, motivated by some of the issues which are apparent from looking at the forecasts of expected returns within pockets identified after fixing the coding error, we introduce a simple and intuitive Bayes regularization into the estimation. As we explain below, the effect of correcting the code and using a one-sided classification scheme is that pocket identification becomes noisier which generates many short-lived pockets which are associated with large and volatile return forecasts. (Under the previous pocket classification, these forecasts were excluded from the earlier analyses of predictability within pockets.) These undesirable features, however, are easily handled by smoothing the original time-varying forecasts using one of the many regularization techniques in widespread use. Specifically, we use a Bayesian approach which shrinks the original (one-sided kernel) return forecast towards zero. This has the effect of reducing estimation error in the original return forecast and, in turn, the real-time pocket identification procedure. We then use these smoothed forecasts for out-of-sample exercises and pocket classification, as in the original paper.

Once this slight adjustment to our original methodology is implemented, we find that all the conclusions from the original study carry over: (i) there are clear pockets with elevated return predictability and these can be identified ex-ante. As a corollary, we find many periods of time with little evidence of return predictability; (ii) return predictability inside such pockets can also be used to improve risk-adjusted return performance through the same trading rule adopted by FST; (iii) the asset pricing models studied by FST continue to struggle to match the values of statistical and economic return performance measures observed in the data; (iv) in contrast, a sticky expectations model is closely aligned with both the statistical and economic measures of return predictability observed in the data; (v) The new return forecasts continue to be positively correlated with the errors in professional forecasters' predictions of pro-cyclical measures such as GDP and industrial production growth and negatively correlated with survey forecast errors of the (countercyclical) unemployment rate.

Hence, while correcting the coding error affects the numerical results in FST, a simple modification to the real-time prediction scheme re-establishes the main economic findings in FST. The remainder of this note revisits the main analyses conducted in the published FST paper. Section II explains the coding error, Section III proposes a simple solution for smoothing the forecast, while Section IV reports empirical results under this modified forecasting rule. Section V reports simulations from workhorse asset pricing models, Section VI discusses model simulation and empirical results related to the sticky expectations model, while Section VII concludes.

II. Key Issue in Prior Analysis

CFNPZ replicate the results in the study by FST but discover an error in one of the functions used to implement the replication code. FST use a two-step approach to generate market excess return forecasts inside and outside of pockets. First, a one-sided kernel is used to estimate the parameters of return prediction models of the form:

$$r_t = x'_{t-1}\beta_{t-1} + \varepsilon_t, \tag{1}$$

where r_t is the market excess return at time t, x_{t-1} is the predictor observed at time $t-1$ and β_{t-1} is the estimated coefficient(s) based on this regression. The one-sided kernel only uses backward-looking data known at time $t-1$ and so the resulting estimate $\widehat{\beta}_{t-1}$ can be used to generate forecasts $\widehat{r}_{t|t-1} = x'_{t-1}\widehat{\beta}_{t-1}$ in real-time. Importantly, this part of the code is correctly implemented throughout the FST paper, so these first-step market excess return forecasts can be considered out-of-sample.

The second step in the FST method is used to identify pockets with return predictability. To determine if the market is currently in a pocket, FST define the squared error difference:

$$SED_t = (r_t - \overline{r}_{t|t-1})^2 - (r_t - \widehat{r}_{t|t-1})^2, \tag{2}$$

where $\overline{r}_{t|t-1}$ is the return forecast from a simple prevailing mean and $\widehat{r}_{t|t-1}$ is the forecast from a kernel-weighted linear return prediction model. When $SED_t > 0$, the local kernel-based forecast is more accurate, in a squared error sense, than the prevailing mean. Periods where this is expected to hold *ex ante* are labeled "pockets of predictability" by FST. A pocket classification regression is used to determine if this condition holds:

$$SED_t = \gamma_{0t} + \gamma_{1t}t + v_t. \tag{3}$$

The code implemented by FST to execute this step erroneously uses a two-sided kernel rather than a one-sided kernel to estimate the parameters, $\widehat{\gamma}_{0t}, \widehat{\gamma}_{1t}$ and generate forecasts $\widehat{SED}_{t|t-1} = \widehat{\gamma}_{0t} + \widehat{\gamma}_{1t}t$. This has the effect of leaking future information into the $\widehat{\gamma}_{0t}, \widehat{\gamma}_{1t}$ estimates and means that they were not, as implemented, relying purely on historically available information.

The key issue, therefore, is not whether or not the degree of stock market return predictability fluctuates over time and, thus, whether pockets exist. Rather, it is related to whether such pockets could have been identified–and exploited–in real time.

We acknowledge this error, and we are grateful to CFNPZ for bringing it to our attention. Errors are humbling to discover. This subtle bug in the code emerged as a result of having, in one instance, the kernel weighting function embedded inside the regression routine and in the other case calling the kernel scheme as an external function. To speed up the execution of the code, the main subroutines for estimating predictive kernel regressions were compiled into mex functions in MATLAB. One version of this function, the oldest version which was a leftover from an earlier version of the analysis in which we were using two-sided kernels, was mistakenly compiled and given a name suggesting it was a one-sided kernel regression function. Separate versions of the function were created for handling different sign restrictions on the regression coefficients which we imposed in the paper, where everything is properly one-sided. In our main file, we used the old function to estimate the regression for the SED measure in (3) but the newer versions to estimate the regression coefficients for the predictive regressions of stock returns themselves.

How does this error affect the main conclusion in FST? As pointed out by CFNPZ, when pockets are identified using the one-sided (backward-looking) rule in FST, many additional short-lived pockets appear. Some of these pockets only last a single period and so can be considered spurious in the sense that they reflect sampling variation, i.e., noise in the $\widehat{\gamma}_{0t}, \widehat{\gamma}_{1t}$ estimates used to identify pockets. As a consequence, the statistical and economic evidence of return predictability for the market portfolio based on the pockets identified using the one-sided (real-time) estimates of $\widehat{\gamma}_{0t}, \widehat{\gamma}_{1t}$ is notably weaker and often disappears when compared to the evidence established for the pockets identified using a two-sided kernel estimate of the parameters in equation (3).

Because the market excess return forecasts, $\widehat{r}_{t|t-1}$, use the correct one-sided kernel and the two-sided kernel is only used to identify the pockets, the coding error does not necessarily affect the conclusion that pockets of return predictability exist. It only affects the conclusion that these pockets were discoverable in real time using the one-sided regression estimates of the $\widehat{\gamma}_{0t}, \widehat{\gamma}_{1t}$ estimates in equation (3).

An analogy is the common finding that predictability of stock market returns is higher during recessions than in expansions, see Dangl and Halling (2012), Henkel et al. (2011), and Rapach et al. (2010). Recession and expansion periods are known only ex-post with a delay of several months, and many other macroeconomic series are only available with some delay. In real time, an investor cannot use this information ex-ante to condition her return forecasts on being in a recession. Nonetheless, it is still of interest to understand whether predictability differs across macroeconomic environments. Similarly, the leakage of future information into the kernel estimates used to determine whether the market is in a pocket means these pockets were not discovered in real time. However, this does not imply that there does not exist pockets in time with much higher return predictability than at other times.[1]

[1]That being said, we do not wish to push this analogy too far due to possible biases in the estimates of the pocket identification scheme arising from future return information contaminating the parameter estimates.

The main effect of the coding error resulting in the use of a two-sided kernel to estimate the parameters of the pocket classification scheme in (3) is to produce notably smoother forecasts of SED_t compared to those based on a one-sided window.

The key finding in CFNPZ is that the "in pocket" one-sided market excess return forecasts from equation (1), $\widehat{r}_{t|t-1} = x'_{t-1}\widehat{\beta}_{t-1}$, are considerably noisier when the pocket classification scheme is properly implemented using a one-sided kernel. A symptom of this is the sharp increase in the number of pockets generated under the one-sided identification scheme – with most of them being extremely short-lived – and more volatile in-pocket forecasts. This higher volatility helps explain why so many short-lived pockets are generated under the one-sided window and is an unattractive feature that makes the resulting periods classified as pockets more prone to be dominated by noise. Results presented in CFNPZ illustrate that this increase in noise substantially weakens the evidence on out-of-sample predictability, especially when using the Clark-West and Diebold-Mariano tests, and partially weakens the evidence on economic significance of our return forecasts.

We next propose a simple solution that handles the noise in the one-sided pocket classification scheme by regularizing the underlying market excess return forecasts in a way that makes them smoother prior to the pocket classification step.

III. A Simple Solution

Above, we argued that correcting the SED regression in (3) used for pocket classification led to considerably noisier in-pocket forecasts relative to what was considered in the original FST paper. In light of this issue, we consider a simple and effective solution to reduce noise in our return forecasts. We use a Bayesian shrinkage procedure, which is just one of many available regularization methods, for smoothing the original market excess return forecasts, $\widehat{r}_{t|t-1}$ prior to implementing the one-sided pocket classification regression and out-of-sample forecast evaluation/economic significance exercises. Reducing the volatility in the original one-sided return forecasts brings down the estimation error in the parameters of the pocket identification scheme, $\widehat{\gamma}_{0t}, \widehat{\gamma}_{1t}$, and also results in better-behaved forecasts overall.

To this end we apply a simple Bayesian approach that shrinks the one-sided kernel-based market excess return forecasts, $\widehat{r}_{t|t-1}$, towards zero. To see how this works, consider a regression of market excess returns, r_t, on $\widehat{r}_{t|t-1}$:

$$r_t = \lambda_{1,t-1}\widehat{r}_{t|t-1} + \varepsilon_{rt}. \tag{4}$$

To produce a forecast of r_t we estimate $\lambda_{1,t-1}$ over the rolling window from $t-s$ to $t-1$, where s is the length of the window. Let $\widehat{\boldsymbol{r}}_{t-s:t-1} = (\widehat{r}'_{t-s|t-s-1}, ..., \widehat{r}'_{t-1|t-2})'$ be an $s \times 1$ vector of return forecasts from the time-varying coefficient model and $\boldsymbol{r}_{t-1} = (r_{t-s}, ..., r_{t-1})'$ an $s \times 1$ vector of returns. Both are known at time $t-1$. To dampen the effect of estimation error on our forecasts, we use a simple and intuitive G-prior Bayesian shrinkage scheme along the lines proposed by Zellner (1986). Let $\widehat{\lambda}_{1,t-1} = \left[\widehat{\boldsymbol{r}}'_{t-s:t-1}\widehat{\boldsymbol{r}}_{t-s:t-1}\right]^{-1}\widehat{\boldsymbol{r}}'_{t-s:t-1}\boldsymbol{r}_{t-1}$ and λ_0 be the rolling window estimate of $\lambda_{1,t-1}$ and the shrinkage target, respectively. Then the G-prior estimator for λ_1 that uses returns up to time $t-1$, $\widehat{\lambda}^G_{1,t-1}$, can be written as (see, e.g., Diebold and Pauly (1990))

$$\widehat{\lambda}^G_{1,t-1} = \lambda_0 + \frac{1}{1+g}\left(\widehat{\lambda}_{1,t-1} - \lambda_0\right). \tag{5}$$

We set $\lambda_0 = 0$, shrinking return forecasts towards zero for maximal stability, consistent with the benchmark excess return forecast used by Gu et al. (2020).[2] In effect, we form a portfolio of the original forecast $\widehat{r}_{t|t-1}$ and a zero forecast, with the latter playing a role similar to a risk-free asset. Our main analysis sets $g = 2$ corresponding to putting two-thirds weight on the shrinkage target and one-third weight on the kernel forecasts. We also conduct a sensitivity analysis with regards to the choice of g and show that the results are robust to other choices of this parameter.

Before estimating $\hat{\lambda}_{1,t-1}$, we winsorize 2.5% of the return forecasts obtained from our time-varying parameter model in both tails, with the winsorization performed recursively and in real-time (i.e., using an expanding window) so as to avoid look-ahead bias. Furthermore, following standard practice from forecast combinations, we restrict the weights to lie between zero and one and we apply this restriction to the initial rolling-window estimate $\widehat{\lambda}_{1,t-1}$ (note that the G-prior shrinkage estimate, $\widehat{\lambda}^G_{1,t-1}$, will also lie between 0 and 1 by construction).

To determine if the market is in a pocket with significant return predictability, as in (2), we next compute the SED measure for comparing the squared errors of the prevailing mean and shrinkage forecasts:

$$SED^G_t = (r_t - \overline{r}_{t|t-1})^2 - (r_t - \widehat{r}^G_{t|t-1})^2. \tag{6}$$

Finally, in the interest of simplicity, we simplify the pocket classification regression in (3) to only include a constant

$$SED^G_t = \gamma_{0t} + v_t. \tag{7}$$

[2]Other shrinkage targets such as the prevailing mean could of course be used. However, these will introduce additional variation in the return forecasts and so are likely to require larger values of g to achieve a similar level of robustness of the resulting forecasts.

Leaving out the time-trend from the regression in (3) is not important to our results, so we decided to adopt the simpler specification used here. Further, to avoid ultra-short pockets with no economic meaning, we define a pocket as a period when the predicted SED_t^G is greater than 0 for a minimum of one month (21 trading days).[3]

To test for differences between return predictability inside versus outside of pockets, we also report the outcome of a two-sample CW t-test. Here the null hypothesis is that return predictability is the same inside and outside the pockets identified by our classification scheme, while positive values indicate that return predictability is higher inside pockets and negative values suggest the opposite.

IV. Empirical Results

We next report results based on the modified return forecasts described above, following closely the tables and figures in FST. The only key changes are that we use the regularized one-sided kernel forecasts in place of the original one-sided kernel forecasts, and we use the simplified pocket classification dummies discussed above.

A. Full-sample Return Regressions

Table I shows the slope coefficients, t-statistics, and $\overline{R}^2$ values for univariate regressions estimated on the full sample, in-pocket, and out-of-pocket periods. While only one of the t-statistics (T-bill rate) is significant in the out-of-pocket regressions, two of the four (dividend-price ratio and term spread) are significant in-pocket and in the full sample (T-bill rate and term spread).[4] This is particularly noteworthy since the number of in-pocket observations is so much smaller than the number of observations in the full sample. $\overline{R}^2$ values are also notably higher in-pocket for the dividend-price ratio and term spread: 0.099 and 0.083 in-pocket versus 0.005 and 0.041 in the full sample and -0.006 and 0.013 out-of-pocket (all in percent). Conversely, return predictability from the T-bill rate seems stronger out-of-pocket than in-pocket.

[3]We also tried a 10-day and 15-day minimum duration and our results remain virtually unchanged. The higher level of smoothness of the earlier two-sided SED measure ended up meaning that pockets this short were not encountered.

[4]These conclusions are based on one-sided tests with the expected signs of the coefficients or two-sided tests with a 10% significance level.

B. *Pocket Characteristics*

Table II reports pocket statistics for our simple pocket identification scheme applied to the daily frequency. The number of pockets identified by our Bayesian shrinkage approach ranges from 20 for the return forecasts based on the term spread predictor to 54 for the return forecasts based on the dividend-price ratio. Sample lengths vary across the individual variables, so the fraction of the sample identified as pockets is a more telling statistic. This measure ranges from 27% for the term spread and dividend-price models to 32% for the forecasts based on the T-bill rate. Pocket lengths range from 22 days (our imposed minimum) to more than three years.[5]

The last three rows in II report the integral R^2 measure, IR^2. The mean IR^2 ranges from 1.00 for the model based on the dividend-price ratio to 2.03 for the model that uses the T-bill rate as the predictor.

C. *Spurious Pockets*

The pocket identification scheme implemented by FST relies on repeated testing which introduces a risk of detecting spurious pockets. To deal with this issue, FST propose a bootstrap based on individual pockets' IR^2 values with only the pockets that generate the highest IR^2 values deemed to be non-spurious. FST find that short-lived pockets with low IR^2 values tend to be insignificant while longer-lived pockets have a higher chance of being significant.

To see if this continues to hold, in Figure 1 we plot the individual pockets' IR^2 values against their duration. We see a clear upward sloping relation with the longest-lived pockets generating the highest IR^2 values while short-lived pockets frequently produce very small positive and, in a few cases, negative IR^2 values.

Building on this insight, we apply the FST bootstrap to classify pockets into significant ones (pockets with an IR^2 value matched in less than 5% of the simulations) and spurious pockets (pockets with IR^2 values matched in at least 5% of the simulations). These are marked in red and blue colors, respectively, in Figure 2. We see clear evidence that the significant pockets tend to be longer-lived than the pockets deemed to be spurious. There is also a notably stronger correlation across predictions in the pockets identified through the Bayesian regularization scheme used here as compared to the pockets identified in the original FST analysis. In our new analysis, the correlation between pocket indices range

[5]Note that the first 21 days of a pocket do not count towards any of the statistical or economic significance statistics that we compute.

between 0.6 and 0.9 across all variables whereas in the original FST analysis these ranged between -0.02 and 0.59.

The right section of Table II reports summary statistics on the pockets deemed to be significant. Requiring pockets to be non-spurious significantly reduces their numbers. For the dividend-price ratio model the number of pockets is reduced from 54 to only 11, while for the term spread model we go from 20 pockets to only seven significant ones. Furthermore, the fraction of the sample spent inside pockets now ranges from 13% for the dividend-price ratio to 23% for the T-bill rate. Minimum pocket duration is also drastically increased to range between 70 and 173 days. Finally, the minimum IR^2 values for the non-spurious pockets are never negative, ranging from 1.58 to 2.68 with mean values ranging between 3.50 for the dividend-price ratio and 6.68 for the T-bill rate.

D. Statistical and Economic Performance

Panel A of Table III reports Clark and West (2007) (CW) test statistics for the unrestricted daily market return forecasts (left columns), sign-restricted forecasts (middle columns) and sign-and-slope restricted forecasts (right columns).

For the unrestricted return forecasts, the in-pocket CW test statistics are statistically significant at the 5% level for two of the predictors (T-bill rate and term spread) and at the 10% level for the realized variance. The multivariate model, principal components forecast and two of the three combination forecasts also generate highly significant CW test statistics inside pockets. In contrast, there is no evidence of return predictability outside the pockets where all CW test statistics are negative and insignificant. The strong in-pocket return predictability carries over to the full sample for two of the forecast combinations but not for the individual predictors and potentially reflects the additional stability generated via regularization of the local kernel forecasts, especially outside of pockets. Moreover, the two-sample t-test for equal return predictability in- and out-of-pocket is statistically significant at the 1% level in all but one case (dividend-price ratio) which is significant at the 5% level. Hence, we find considerable statistical evidence consistent with out-of-sample return predictability being significantly stronger inside the pockets identified by our approach than outside these pockets.

Imposing the positivity constraint on the return forecasts or further imposing constraints on the sign of the slope of the regression coefficients improves the predictive accuracy of most of the return forecasts although results vary across the predictors. For example, the CW t-statistics increase by a large margin for the T-bill rate, term spread and realized variance and also for the principal component and the three combination schemes. Improvements are positive but smaller for the dividend-price ratio. Again, differences in predictive accuracy in-pocket versus out-of-pocket is highly significant for all individual predictors and forecasting schemes.

Turning to the financial performance measures listed in Panel B of Table III, we find that the alpha estimates are statistically significant at the 5% level or higher for all of our approaches (using one-sided critical values as in the FST paper) with annualized alpha estimates of 2.74, 5.48, 4.42, and 3.41 for the four individual predictor variables and associated t-statistics of 1.94, 4.01, 3.17, and 2.43. Sharpe ratios associated with the unrestricted return forecasts exceed that of the prevailing mean (0.46), ranging from 0.48 for the dividend-price ratio to 0.71 for the T-bill ratio. Imposing non-negativity of the return forecasts and adding a slope restriction increases alpha estimates which remain highly statistically significant, although the margin of improvement over the unconstrained excess return forecasts is smaller than that seen for the statistical (CW) tests and Sharpe ratios also do not improve from imposing these constraints.

E. Robustness to Window Length and Shrinkage Parameter g

Table IV reports CW test statistics for different combinations of the two look-back windows used to estimate the parameters in the return regression and the regression used for pocket classification. We consider window lengths between two and three years to estimate the parameters in the return regression in (1) and window lengths ranging from six to 18 months for the pocket classification regression in (7). In Table IV, we use our baseline value of shrinkage $g = 2$, but we will demonstrate robustness to other values of g in Table V below.

In-pocket values of the test statistic are significantly positive for most of these scenarios for the individual predictor variables as well as for the multivariate and principal components models and the three combinations. The main exception to this finding is when a short (six month) window is used in the classification scheme which leads to somewhat lower test statistics, though the degradation of performance in market-timing alphas (bottom right quadrant) is less severe in some cases. Thus the main take-away from these exercises is to use a classification window of one year or longer.

In contrast, during the out-of-pocket part of our sample the CW test statistics are almost always negative. This suggests that the shrinkage return forecasts are less accurate than the prevailing mean forecasts out-of-pocket.

The test statistics for the null of equal in-pocket and out-of-pocket return predictability are reported in the bottom left quadrant. These results continue to be strong, again with the exception of the shortest six-month window used in the pocket classification step. This shows that the ability of our forecasting and classification scheme to identify periods with elevated return predictability is robust across a wide range of settings. Measures of economic significance such as the alpha estimate (bottom right quadrant) also remain strong for the vast majority of parameter settings considered here.

Next, consider the sensitivity of our results with regards to the degree of shrinkage, g. Table V shows that the statistical significance of the in-pocket CW test statistics holds up in most cases both when we reduce the amount of shrinkage towards the zero target ($g = 1$) and when we increase it ($g = 3$). As in the benchmark case with $g = 2$, the vast majority of out-of-pocket test statistics are negative and the test statistics for the null of equal in-pocket and out-of-pocket return predictability is strongly rejected against the alternative of higher in-pocket predictability. We also report measures of economic significance, which illustrates that the alpha and t-statistic estimates from our market timing strategies are insensitive to our choice of the degree of shrinkage g.

F. Monthly Return Forecasts

Table VI shows pocket characteristics for the monthly data. The number of pockets identified at the monthly frequency varies from 19 for the term spread model to 41 for the realized variance model. Pocket duration ranges from 21 to 651 days.[6] Between 28% (term spread) and 32% (T-bill rate) of the monthly samples are identified as being pockets.

The mean IR^2 values fall in a somewhat narrower band for the monthly than for the daily data, ranging between 1.26 for the forecasts based on the realized variance to 1.98 for the forecasts based on the T-bill rate.

Table VII shows statistical and economic performance results for the monthly data. First consider the statistical accuracy of the forecasts as reported in Panel A. During the in-pocket sample periods, the CW test statistics generated by the unrestricted forecasts are significantly positive at the 10% level for three of the individual predictors (the T-bill rate, term spread and realized variance) and positive but insignificant for the dividend-price ratio. The multivariate, principal component and three forecast combination schemes also generate CW test statistics that are significant at the 10% level of higher inside pockets. In contrast, the out-of-pocket CW test statistics are all negative except for one case with a small positive value (0.09 for the dividend-price ratio). Moreover, the tests for differences in predictive

[6] We report the duration in trading days using the convention of 21 days per month.

accuracy in-pocket versus out-of-pocket are all positive and significant at the 5% level or higher for all models except for the dividend-price ratio model which generates a test statistic that is significant at the 10% level.

Imposing the non-negativity constraint on the sign of the monthly return forecasts mostly improves forecast accuracy with CW test statistics that in many cases go from being significant at the 10% level to being significant at the 5% level. Further adding the sign restriction on the slope does not make much of a difference to the results.

Next, consider the economic performance measures reported in Panel B in Table VII. In the unrestricted case all four individual predictors generate highly significant and positive alpha estimates between 2% and 4% per year. The multivariate, principal components and two of the three forecast combination schemes also generate alphas exceeding 3.5% per year. Sharpe ratios range from 0.46 to 0.59 for the individual predictors compared to a value of 0.49 for the prevailing mean. Imposing economic restrictions on our forecasts, we find a notable improvement in the alpha estimate for the dividend-price ratio model but smaller effects for the other prediction models. Sharpe ratios do not change much as a result of imposing these restrictions and are in fact a bit smaller in some cases.

G. *Size and Value factor returns*

CFNPZ find that many of the strong return prediction results reported by FST continue to hold for the Fama-French size and value portfolios even after correcting the coding error. Their findings use the original pocket classification scheme in FST which, as we argue above, is quite noisy and tends to generate too many short-lived pockets.

To address this point, we also examine predictions generated for the Fama-French size and value portfolio return series based on our Bayesian forecasting and pocket classification scheme. Table VIII presents our results for the size (small minus big, or SMB) and value factor (high minus low, or HML) portfolios. For the SMB return series, the in-pocket CW t-statistics associated with the individual predictors fall in a range from 1.69 to 2.76 and exceed three for the three forecast combination schemes and the multivariate model. However, we now find that the out-of-pocket CW test statistics are even higher than their in-pocket counterparts for two of the predictors (T-bill rate and term spread). Even so, the test for differences in predictive accuracy reported in the fourth column continues to be positive in all comparisons of in-pocket and out-of-pocket return predictability and are statistically significant at the 10% level of better for all but one comparison (T-bill rate).[7]

[7]This can happen even in cases where the CW t-statistic is higher in the out-of-pocket than in-pocket

For the HML portfolio, in-pocket CW test statistics are all significant at the 5% or 1% level while only one of these statistics is significant and positive out-of-pocket (for the multivariate model). The far stronger in-pocket return predictability manifests itself in the form of significantly positive test statistics for the difference in predictive accuracy.

The strong statistical tests for return predictability carry over to the economic performance measures (Panel B). For the size factor (SMB) returns, the alpha estimates fall in a range between 1.31% and 4.15% with all but one estimate being statistically significant at the 1% level. Sharpe ratios for the SMB returns fall in a range between 0.41 and 0.95 versus 0.17 for the prevailing mean.

Alpha estimates for the value factor (HML) fall in a range between 1.46% and 2.73% and are again significant at the 1% level for all but one of the forecasting models. Sharpe ratios range from 0.64 to 0.79 for the four individual predictor variables versus 0.62 for the prevailing mean which is a smaller improvement than that seen for the SMB portfolio.

V. Asset Pricing Models

In the context of a range of canonical asset pricing models, FST simulate excess returns data and values for three of the four predictors (the dividend-price ratio, risk-free rate, and the realized variance) which are generated endogenously by these models. Next, they test whether the resulting data is consistent with the statistical tests for return predictability observed in the empirical data as well as the evidence that this could have been exploited to generate improved economic performance. They find that all asset pricing models under consideration fall short of matching the in-pocket CW test statistics computed for the actual data as well as the risk-adjusted return performance measures (alpha).

We undertake the same analysis based instead on the Bayesian forecasting scheme. Results from these simulations are shown in Table IX. The first column in the table shows the empirically observed sample estimates which the simulations from the asset pricing models are compared against.[8] For the regression that uses the dividend-price ratio as a predictor, all asset pricing models succeed in matching the CW test statistics both in-pocket and out-of-pocket, as well as for the full sample. However, they fail (with p-values around 0.02) to match the CW test of equality of in-pocket and out-of-pocket return predictability.

sample because mean squared forecast errors are lower (relative to the benchmark) for the in-pocket than out-of-pocket forecasts. These are cases where the higher out-of-pocket CW t-statistic simply reflects the larger out-of-pocket sample size as opposed to a lower mean squared forecast error.

[8]The reported p-values measure the success of the asset pricing models in achieving this with a low p-value suggesting that a model is struggling to match the corresponding moment observed in the actual data.

All asset pricing models also fail to match the empirically observed in-pocket CW test statistics for the specification that uses the risk-free rate as a predictor and, with p-values typically around 0.05-0.10, also struggle to match the data for the realized variance specification. For these two predictors, again none of the models matches the test statistic for equal in-pocket and out-of-pocket return predictability.

Turning to the economic performance measures, none of the models come close to matching the alpha estimates and t-tests for alpha computed for the data. They do succeed in matching some of the Sharpe ratio estimates, however, which is unsurprising since Sharpe ratios have substantial variability across simulated samples and improvements to this measure from applying our prediction scheme to the data tend to be smaller by comparison to the alphas.

VI. Sticky Expectations

In the last part of the paper, FST argue that a model with sticky expectations, with a degree of information rigidity disciplined by other studies, can reproduce the evidence of pockets observed in the data. Further, they find that the expected return forecasts generated by such a model are correlated with forecast errors with signs consistent with the model. We demonstrate that both conclusions still hold using the revised estimation procedure.

A. Simulations from the sticky expectations model

Table X shows simulation results for the sticky expectations model proposed in FST based on a one-sided 2.5-year estimation window mirroring the baseline empirical results. As for the regular asset pricing models, we only present results for the three predictor variables generated endogenously by this model. As in Table IX, the first column shows the sample moments that the models should be compared against. First consider the dividend-price ratio variable. The baseline standard calibration model with sticky expectations, the baseline calibration with rational expectations ($\lambda = 0$), and the rational expectations model recalibrated to match our original moment targets can all match the CW test statistics across the full sample, in-pocket and out-of-pocket periods. However, only the baseline sticky expectations model can match the in-pocket versus out-of-pocket difference in squared error performance test computed for the data. The baseline sticky expectations model is also the only model to succeed in matching the alpha and t_α estimates in the data, whereas the baseline calibration with rational expectations ($\lambda = 0$) and the rational expectations models fall well short of this.

Next consider the risk-free rate predictor. Our sticky expectations model succeeds in matching both the statistical (CW) measures of return predictability along with the three economic measures. The baseline rational expectations model with $\lambda = 0$ and the rational expectations model fail to match the CW difference statistic as well as all three economic performance measures computed for the data.

Finally, the sticky expectations model easily matches all the statistical and economic measures of return predictability when the realized variance is used as the predictor variable. In contrast, the baseline model with $\lambda = 0$ and rational expectations model fail to match the CW difference statistic as well as the alpha and t-statistic of alpha computed for the data.

B. Direct evidence from forecast errors

FST also provide more direct evidence on the mechanism that might generate return predictability pockets. They do so by linking expected return forecast and biases in the beliefs of participants in the Survey of Professional Forecasters (SPF). Specifically, they regress average quarterly SPF forecast errors for three macroeconomic variables (GDP growth (gy), the unemployment rate (ue) and real industrial production growth (ie)) on return forecasts. The sticky expectations model suggests that there should be a positive relation between the time-varying return forecasts and the forecast errors of gy and ie and a negative relation between the return forecasts and the forecast errors of ue.

Implementing the same steps and regressions as in FST, we obtain the results depicted in Figure 3. Consistent with the original results in FST, we find positive and highly significant correlations between our return forecast errors and survey forecast errors for gy and ip and negative and highly significant correlations between our return forecast errors and survey errors in forecasts of the unemployment rate.

VII. Conclusions

CFNPZ identify a coding error in the replication files accompanying the paper by FST. While this error does not affect how the underlying market excess return forecasts are generated, it does affect the extent to which pockets can be discovered in real time by letting future information leak into the parameter estimates of the pocket classification scheme. CFNPZ show that, once corrected, pocket classification becomes noisier with too many short-lived pockets being triggered.

In this note, we propose a simple way to address this additional noise in the forecasts through a Bayes regularization scheme with G-priors that smooths the underlying (one-sided) market excess return forecasts before applying a simplified pocket determination rule. This scheme dampens the noise in the return forecasts and produces significantly more accurate return forecasts inside pockets for most of the predictor variables and prediction schemes considered in the analysis.

In conclusion, while we acknowledge the coding error discovered by CFNPZ, we find that all the main conclusions and economic insights in FST continue to hold once the underlying market excess return forecasts are smoothed so as to prevent too many short-lived pockets from being detected. Therefore, modulo the additional need to regularize the forecasts to reduce estimation error, we find that the FST conclusions are preserved.

REFERENCES

Cakici, Nusret, Christian Fiedberg, Tobias Neumaier, Thorsten Poddig, and Adam Zaremba, 2024, Pockets of predictability: A replication, *Journal of Finance* Forthcoming.

Clark, Todd E., and Kenneth D. West, 2007, Approximately normal tests for equal predictive accuracy in nested models, *Journal of Econometrics* 138, 291–311.

Dangl, Thomas, and Michael Halling, 2012, Predictive regressions with time-varying coefficients, *Journal of Financial Economics* 106, 157–181.

Diebold, Francis X, and Peter Pauly, 1990, The use of prior information in forecast combination, *International Journal of Forecasting* 6, 503–508.

Farmer, Leland E, Lawrence Schmidt, and Allan Timmermann, 2023, Pockets of predictability, *The Journal of Finance* 78, 1279–1341.

Gu, Shihao, Bryan Kelly, and Dacheng Xiu, 2020, Empirical asset pricing via machine learning, *The Review of Financial Studies* 33, 2223–2273.

Henkel, Sam, J. Martin, and Federico Nardari, 2011, Time-varying short-horizon return predictability, *Journal of Financial Economics* 99, 560–580.

Rapach, David E., Jack K. Strauss, and Guofu Zhou, 2010, Out-of-sample equity premium prediction: Combination forecasts and links to the real economy, *The Review of Financial Studies* 23, 821–862.

Zellner, Arnold, 1986, On assessing prior distributions and bayesian regression analysis with g-prior distributions, *Bayesian inference and decision techniques* .

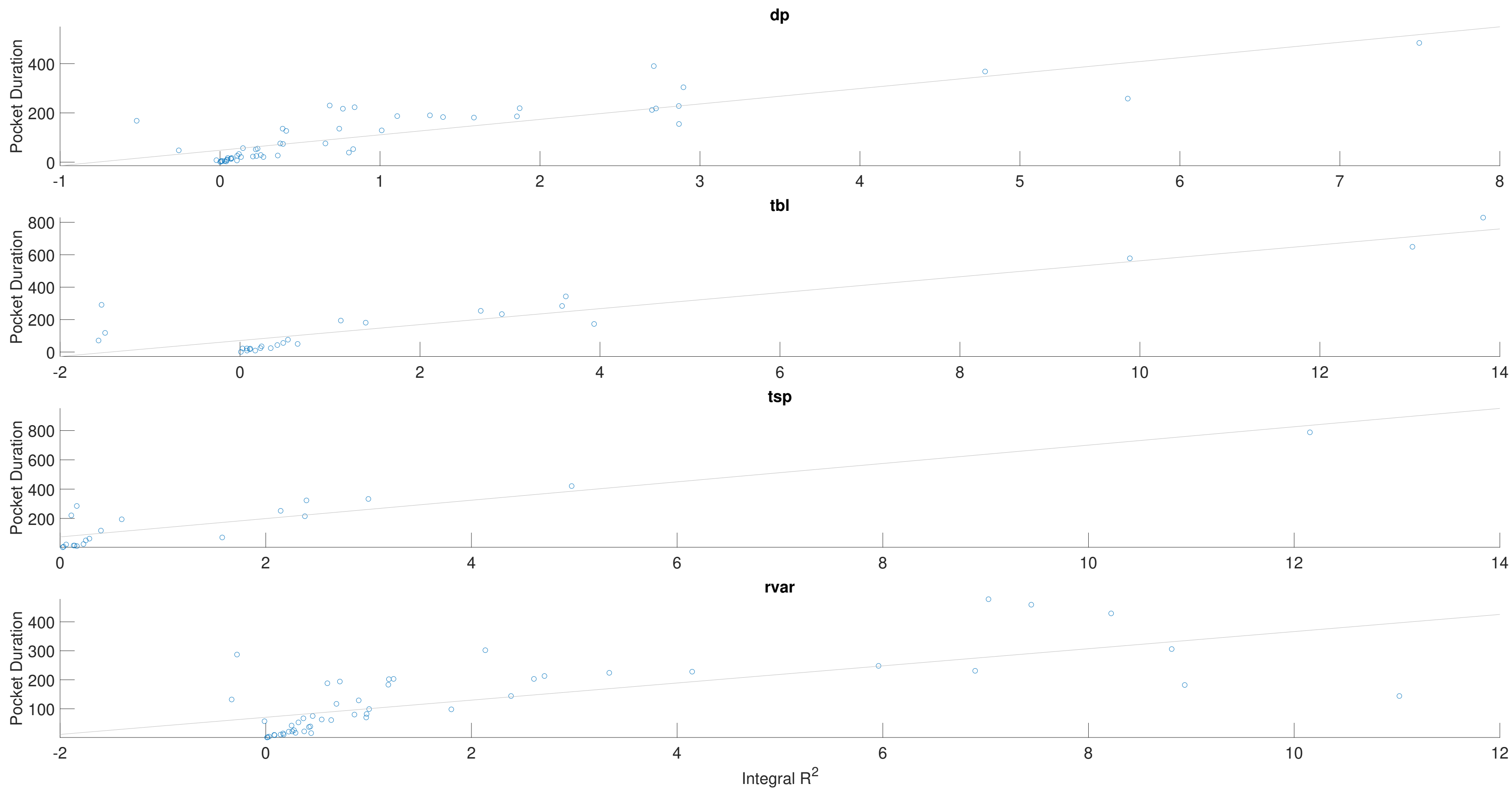

Figure 1: Scatter Plot of Integral R^2 of pockets (x-axis) vs. pocket duration in days (y-axis) This figure shows scatter plots documenting the relationship between the integral R^2 and duration (in days) of pockets identified in our baseline daily empirical specification for each univariate prediction model. Each panel also plots a best fit line. Coefficients are estimated using a one-sided kernel with a 2.5-year effective sample size, with return forecasts regularized through a Bayesian shrinkage scheme, and pockets are identified as periods in which a fitted squared forecast error differential (relative to a prevailing mean forecast and estimated using a one-sided kernel with a one-year effective sample size) is above zero for at least 21 trading days (1 month).

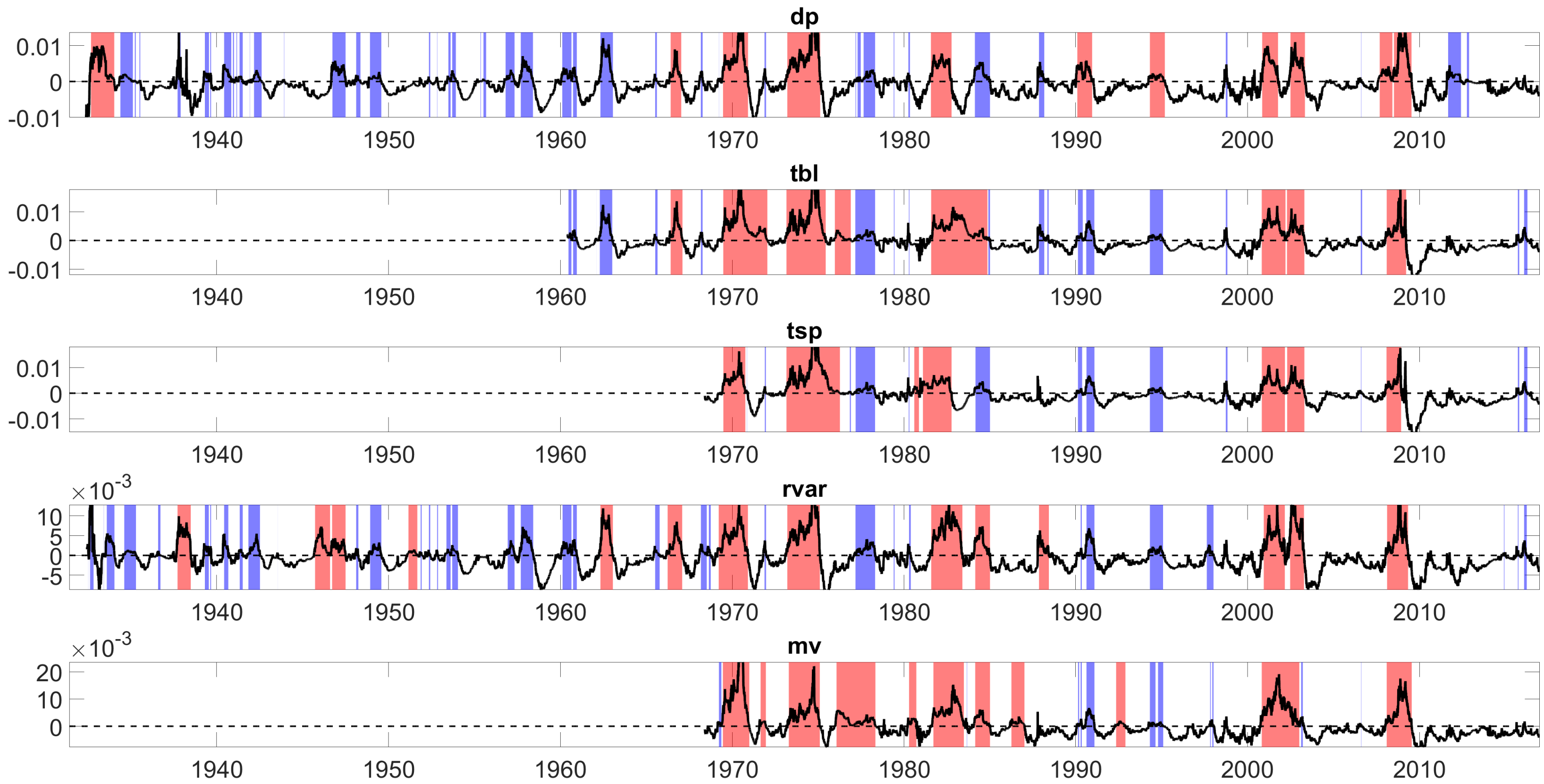

Figure 2: Local return predictability. The first four panels plot one-sided nonparametric kernel estimates of the fitted squared forecast error differential $\widehat{SED}_t$ (estimated using a one-sided kernel with a one-year effective sample size) from a regression of daily excess stock returns on each of the four predictor variables using an effective sample size of 2.5 years. The final panel plots the local $\widehat{SED}_t$ from a four-variable regression specification with coefficients estimated using a product kernel. The shaded areas represent periods when $\widehat{SED}_t > 0$, with areas in red representing pockets that have less than a 5% chance of being spurious and areas in blue representing pockets that have more than a 5% chance of being spurious. The sampling distributions used to determine spuriousness come from an EGARCH(1,1) residual bootstrap design.

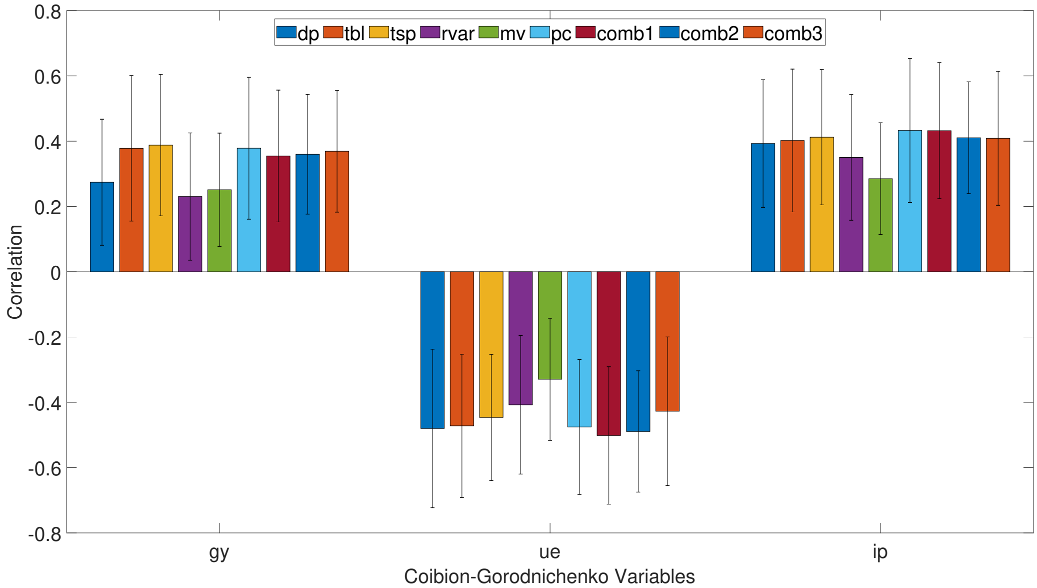

Figure 3: Correlation of Coibion-Gorodnichenko forecast errors with excess return forecasts. This figure shows correlations between forecast errors of three macroeconomic variables from the Survey of Professional Forecasters (SPF) and excess return forecasts from our time-varying coefficient models. The three sets of bar graphs correspond to forecast errors for real GDP growth (gy), the unemployment rate (ue), and real industrial production growth (ip). The height of the nine colored bars represents correlations of those forecast errors with the excess return forecasts from our time-varying predictor models. Each bar is bracketed by 95% confidence intervals computed using HAC standard errors. Since the SPF respondents send in their forecasts in the middle of each quarter, we only use excess return forecasts from the first month of each quarter to make the information sets consistent.

Table I
Constant-Coefficient Regression Results

This table reports slope coefficient estimates, t-statistics (computed using Newey-West standard errors), and $\overline{R}^2$ values for univariate regressions of daily excess stock returns on the lagged predictor variables listed in the rows. The three panels report results for three different sub periods. Panel A reports results for the full sample, Panel B reports results for the concatenation of periods determined to be pockets, and Panel C reports results for the concatenation of all periods not classified as pockets. The start dates for each series are: November 5, 1926 for the dividend price ratio (dp), January 4, 1954 for the three-month Treasury bill (tbl), January 2, 1962 for the term spread (tsp), and January 15, 1927 for the realized variance (rvar). All series run through the end of 2016.

Variables	**Slope coefficient**	***t*-statistic**	$\overline{R}^2$ **(in %)**	**No. of obs.**
		Panel A: Full sample		
dp	0.025	1.142	0.005	23,786
tbl	−0.007	−2.783	0.053	15,860
tsp	0.017	2.311	0.041	13,846
rvar	0.000	0.536	0.000	23,727
		Panel B: In-pocket		
dp	0.121	2.625	0.099	5,980
tbl	−0.004	−0.658	−0.009	4,608
tsp	0.029	1.881	0.083	3,429
rvar	−0.000	−0.349	−0.011	6,536
		Panel C: Out-of-pocket		
dp	0.002	0.082	−0.006	16,446
tbl	−0.006	−1.913	0.037	9,892
tsp	0.012	1.334	0.013	9,057
rvar	0.000	1.134	0.057	15,831

Table II
Pocket Statistics (Daily)

This table reports statistics on the duration of pockets (in days) and their integral R^2 for pockets estimated with daily data. Coefficients are estimated using a one-sided kernel with a 2.5-year effective sample size, with return forecasts regularized through a Bayesian shrinkage scheme, and pockets are identified as periods in which a fitted squared forecast error differential (relative to a prevailing mean forecast and estimated using a one-sided kernel with a one-year effective sample size) is for a minimum of 21 trading days (1 month). The right half of the panel reports the statistics only for pockets which are significant at the 5% level according to our bootstrap integral R^2 analysis.

Statistics	**dp**	**tbl**	**tsp**	**rvar**	**dp**	**tbl**	**tsp**	**rvar**
Num pockets	54.00	27.00	20.00	51.00	11.00	8.00	7.00	14.00
Fraction of sample	0.27	0.32	0.27	0.29	0.13	0.23	0.19	0.17
Duration								
Min	22.00	22.00	24.00	22.00	155.00	173.00	70.00	144.00
Mean	131.74	191.67	192.45	149.16	274.73	418.00	343.14	270.79
Max	505.00	850.00	809.00	499.00	484.00	829.00	788.00	478.00
Integral R^2								
Min	−0.52	−1.57	0.03	−0.33	1.86	2.68	1.58	2.13
Mean	1.00	2.03	1.56	1.95	3.50	6.68	4.09	5.83
Max	7.50	13.81	12.15	11.02	7.50	13.81	12.15	11.02

Table III

Out-of-Sample Measures of Forecasting Performance (Daily Benchmark Specification)

Panel A reports the Clark and West (2007) test statistics for out-of-sample return predictability measured relative to a prevailing mean forecast. Panel B reports three measures of economic significance associated with returns on a portfolio that uses the time-varying coefficient model forecast in-pocket and the prevailing mean forecast out-of-pocket to allocate between the risk-free asset and the market (portfolio weights are limited to be between zero and two): the annualized estimated alpha in percentage points, the heteroskedasticity- and autocorrelation-consistent t-statistic for the estimated alpha, and the annualized Sharpe ratio of the portfolio. We use a backward-looking kernel with an effective sample size of 2.5 years to compute forecasts and regularize the forecasts through a Bayesian shrinkage scheme with shrinkage paramter $g = 2$. "pc" is a recursively computed first principal component of the four predictor variables. "mv" is a four-variable multivariate forecast estimated using a product kernel. "comb1," "comb2," and "comb3" refer to using a simple average of the univariate forecasts. "comb1" sets an individual predictor's forecast to the time-varying coefficient model forecast during a pocket and to the prevailing mean otherwise. "comb2" is the same as "comb1" except that it ignores individual predictor forecasts when that variable is not in a pocket but at least one other variable is in a pocket. "comb3" makes no distinction between in-pocket and out-of-pocket periods and always uses the simple equal-weighted average of all four univariate models. The CW test statistics approximately follow a normal distribution, with positive values indicating more accurate out-of-sample return forecasts than the prevailing mean benchmark and negative values indicating the opposite. A pocket is classified as a period in which a fitted squared forecast error differential (estimated using a one-sided kernel with a one-year effective sample size) is above zero for a minimum of 21 trading days (1 month). Consider a particular statistic of interest, β. $*, **$, and $***$ represent statistical significance at the 10%, 5%, and 1% level from a hypothesis test of $\beta = 0$ against an alternative hypothesis of $\beta > 0$. $\dagger$, $\dagger\dagger$, and $\dagger\dagger\dagger$ represent statistical significance at the 10%, 5%, and 1% level from a hypothesis test of $\beta = 0$ against an alternative hypothesis of $\beta < 0$.

Panel A: Clark-West statistics

	Unrestricted				+ excess return forecasts				All sign restrictions			
Variables	**Full sample**	**In-pocket**	**Out-of-pocket**	**Diff.**	**Full sample**	**In-pocket**	**Out-of-pocket**	**Diff.**	**Full sample**	**In-pocket**	**Out-of-pocket**	**Diff.**
dp	0.32	1.25	−1.03	2.00^{**}	0.19	1.32^{*}	−0.98	1.94^{**}	0.36	1.49^{*}	−0.90	2.08^{**}
tbl	1.34^{*}	2.29^{**}	−1.05	3.14^{***}	1.76^{**}	2.98^{***}	−0.74	3.41^{***}	1.74^{**}	3.14^{***}	−0.74	3.35^{***}
tsp	0.24	1.85^{**}	$-1.62^{\dagger}$	2.87^{***}	0.88	2.29^{**}	−1.17	3.03^{***}	1.05	2.58^{***}	$-1.37^{\dagger}$	3.53^{***}
rvar	0.57	1.39^{*}	−0.83	2.07^{**}	0.58	1.76^{**}	−1.02	2.45^{***}	0.69	1.77^{**}	−0.90	2.47^{***}
mv	1.55^{*}	2.38^{***}	−0.38	2.88^{***}	1.81^{**}	2.10^{**}	0.51	1.71^{**}	1.81^{**}	2.10^{**}	0.51	1.71^{**}
pc	0.40	1.76^{**}	$-1.38^{\dagger}$	2.77^{***}	1.06	2.27^{**}	−0.92	3.00^{***}	1.06	2.27^{**}	−0.92	3.00^{***}
comb1	1.81^{**}	1.84^{**}	–	–	2.21^{**}	2.23^{**}	–	–	2.39^{***}	2.41^{***}	–	–
comb2	1.94^{**}	1.97^{**}	–	–	2.37^{***}	2.40^{***}	–	–	2.52^{***}	2.54^{***}	–	–
comb3	0.67	1.52^{*}	−1.27	2.57^{***}	0.90	2.07^{**}	−1.23	3.10^{***}	1.04	2.09^{**}	−1.14	3.05^{***}

Panel B: Economic significance

	Unrestricted				+ excess return forecasts				All sign restrictions			
Variables	$\hat{\alpha}$	$t_{\hat{\alpha}}$	**Sharpe Ratio**		$\hat{\alpha}$	$t_{\hat{\alpha}}$	**Sharpe Ratio**		$\hat{\alpha}$	$t_{\hat{\alpha}}$	**Sharpe Ratio**	
dp	2.74^{**}	1.94	0.48	–	2.99^{**}	1.93	0.48	–	3.26^{**}	2.10	0.50	–
tbl	5.48^{***}	4.01	0.71	–	5.64^{***}	3.64	0.66	–	5.95^{***}	3.80	0.68	–
tsp	4.42^{***}	3.17	0.62	–	4.84^{***}	2.99	0.59	–	5.16^{***}	3.26	0.63	–
rvar	3.41^{***}	2.43	0.52	–	3.54^{***}	2.33	0.51	–	3.37^{**}	2.26	0.51	–
mv	3.59^{***}	2.72	0.56	–	3.56^{***}	2.43	0.53	–	3.56^{***}	2.43	0.53	–
pc	4.57^{***}	3.45	0.65	–	4.97^{***}	3.10	0.60	–	4.97^{***}	3.10	0.60	–
comb1	4.51^{***}	3.25	0.62	–	4.65^{***}	3.08	0.60	–	4.85^{***}	3.25	0.62	–
comb2	4.47^{***}	3.05	0.59	–	4.72^{***}	3.01	0.59	–	4.98^{***}	3.19	0.61	–
comb3	2.26^{**}	1.90	0.46	–	2.56^{**}	1.72	0.45	–	2.57^{*}	1.59	0.45	–
pm	-0.26^{*}	−1.62	0.46	–	-0.26^{*}	−1.62	0.46	–	-0.26^{*}	−1.62	0.46	–

Table IV

Robustness of Out-of-Sample Measures of Forecasting Performance (Window Length)

This table reports the Clark and West (2007) test statistics for out-of-sample return predictability measured relative to a prevailing mean forecast for different combinations of bandwidths for both the coefficient from the predictive regression and fitted squared forecast error differential estimation. In each column header, the first duration corresponds to the effective sample size for estimating the coefficient and the second duration corresponds to the effective sample size for estimating the fitted squared forecast error differential. The shrinkage parameter is set to $g = 2$ across specifications. "pc" is a recursively computed first principal component of the four predictor variables. "mv" is a four-variable multivariate forecast estimated using a product kernel. "comb1," "comb2," and "comb3" refer to using a simple average of the univariate forecasts. "comb1" sets an individual predictor's forecast to the time-varying coefficient model forecast during a pocket and to the prevailing mean otherwise. "comb2" is the same as "comb1" except it ignores individual predictor forecasts when that variable is not in a pocket but at least one other variable is in a pocket. "comb3" makes no distinction between pocket and nonpocket periods and always uses the simple equal-weighted average of all four univariate models. The CW test statistics approximately follow a normal distribution with positive values indicating more accurate out-of-sample return forecasts than the prevailing mean benchmark and negative values indicating the opposite. A pocket is classified as a period in which a fitted squared forecast error differential is above zero for a minimum of 21 trading days (1 month). Consider a particular statistic of interest, β. $*, **$, and $***$ represent statistical significance at the 10%, 5%, and 1% level from a hypothesis test of $\beta = 0$ against an alternative hypothesis of $\beta > 0$. †, ††, and † † † represent statistical significance at the 10%, 5%, and 1% level from a hypothesis test of $\beta = 0$ against an alternative hypothesis of $\beta < 0$.

Variable	**In-pocket**					**Out-of-pocket**				
	2.5yCoef, 1ySED	**2yCoef, 1ySED**	**3yCoef, 1ySED**	**2.5yCoef, 6mSED**	**2.5yCoef, 1.5ySED**	**2.5yCoef, 1ySED**	**2yCoef, 1ySED**	**3yCoef, 1ySED**	**2.5yCoef, 6mSED**	**2.5yCoef 1.5ySED**
dp	1.25	1.18	1.28	0.74	0.67	−1.03	−0.91	−1.11	−0.49	−0.42
tbl	2.29**	2.15**	2.41***	1.44*	2.34***	−1.05	−0.20	−1.02	0.29	−1.02
tsp	1.85**	1.97**	1.82**	0.53	1.66**	−1.62†	−1.38†	−1.50†	−0.16	−1.50†
rvar	1.39*	1.62*	1.42*	1.00	1.05	−0.83	−0.43	−0.80	−0.35	−0.53
mv	2.38***	1.66**	2.52***	1.91**	2.12**	−0.38	−0.19	−0.28	−0.07	−0.35
pc	1.76**	1.79**	1.66**	0.70	1.60*	−1.38†	−1.15	−1.22	−0.13	−1.21
comb1	1.84**	1.89**	1.87**	1.01	1.58*	–	–	–	–	–
comb2	1.97**	2.03**	1.94**	0.76	1.64*	–	–	–	–	–
comb3	1.52*	1.53*	1.50*	0.44	1.36*	−1.27	−0.68	−1.26	0.46	−1.03
Variable	**Difference**					**Economic significance - $\hat{\alpha}$**				
	2.5yCoef, 1ySED	**2yCoef, 1ySED**	**3yCoef, 1ySED**	**2.5yCoef, 6mSED**	**2.5yCoef, 1.5ySED**	**2.5yCoef, 1ySED**	**2yCoef, 1ySED**	**3yCoef, 1ySED**	**2.5yCoef, 6mSED**	**2.5yCoef 1.5ySED**
dp	2.00**	1.83**	2.07**	1.13	1.05	2.74**	2.60**	2.79**	1.62	1.65
tbl	3.14***	2.50***	3.17***	1.73**	3.13***	5.48***	4.99***	5.05***	3.56***	5.45***
tsp	2.87***	2.94***	2.85***	0.71	2.63***	4.42***	4.19***	4.33***	1.74*	4.09***
rvar	2.07**	2.15**	2.10**	1.39*	1.56*	3.41***	3.06**	3.42***	2.19*	2.42**
mv	2.88***	2.01**	3.01***	2.29**	2.59***	3.59***	3.56***	3.74***	3.02**	3.48***
pc	2.77***	2.71***	2.57***	0.93	2.51***	4.57***	4.28***	4.29***	2.36**	4.16***
comb1	–	–	–	–	–	4.51***	4.23***	4.33***	2.55**	3.96***
comb2	–	–	–	–	–	4.47***	4.18***	4.48***	2.04*	3.64***
comb3	2.57***	2.31**	2.53***	0.77	2.32**	2.26**	2.61**	2.11**	2.18**	2.18**

Table V

Robustness of Out-of-Sample Measures of Forecasting Performance (Shrinkage Parameter g)

This table reports the Clark and West (2007) test statistics for out-of-sample return predictability measured relative to a prevailing mean forecast for different choices of the Bayesian shrinkage parameter g. In each column header, we indicate the choice of g associated with each set of results. Window lengths for estimation and pocket identification are as in Table III. "pc" is a recursively computed first principal component of the four predictor variables. "mv" is a four-variable multivariate forecast estimated using a product kernel. "comb1," "comb2," and "comb3" refer to using a simple average of the univariate forecasts. "comb1" sets an individual predictor's forecast to the time-varying coefficient model forecast during a pocket and to the prevailing mean otherwise. "comb2" is the same as "comb1" except it ignores individual predictor forecasts when that variable is not in a pocket but at least one other variable is in a pocket. "comb3" makes no distinction between pocket and nonpocket periods and always uses the simple equal-weighted average of all four univariate models. The CW test statistics approximately follow a normal distribution with positive values indicating more accurate out-of-sample return forecasts than the prevailing mean benchmark and negative values indicating the opposite. A pocket is classified as a period in which a fitted squared forecast error differential is above zero for a minimum of 21 trading days (1 month). Consider a particular statistic of interest, β. $*, **$, and $***$ represent statistical significance at the 10%, 5%, and 1% level from a hypothesis test of $\beta = 0$ against an alternative hypothesis of $\beta > 0$. †, ††, and † † † represent statistical significance at the 10%, 5%, and 1% level from a hypothesis test of $\beta = 0$ against an alternative hypothesis of $\beta < 0$.

Variable	In-pocket			Out-of-pocket			Difference		
	$g=1$	$g=2$	$g=3$	$g=1$	$g=2$	$g=3$	$g=1$	$g=2$	$g=3$
dp	0.80	1.25	1.44*	−0.79	−1.03	−1.06	1.38*	2.00**	2.19**
tbl	2.60***	2.29**	2.23**	−1.02	−1.05	−1.00	3.58***	3.14***	3.02***
tsp	2.00**	1.85**	1.54*	−2.05††	−1.62†	−1.21	3.25***	2.87***	2.36***
rvar	0.89	1.39*	1.50*	−0.34	−0.83	−0.82	1.33*	2.07**	2.18**
mv	2.23**	2.38***	2.59***	−0.05	−0.38	−0.74	2.80***	2.88***	3.24***
pc	1.71**	1.76**	1.56*	−1.33†	−1.38†	−1.05	2.71***	2.77***	2.38***
comb1	1.78**	1.84**	1.78**	–	–	–	–	–	–
comb2	1.96**	1.97**	1.92**	–	–	–	–	–	–
comb3	1.35*	1.52*	1.55*	−1.35†	−1.27	−1.13	2.29**	2.57***	2.63***

Variable	$\hat{\alpha}$			$t_{\hat{\alpha}}$			Sharpe Ratio		
	$g=1$	$g=2$	$g=3$	$g=1$	$g=2$	$g=3$	$g=1$	$g=2$	$g=3$
dp	1.72*	2.74**	3.35**	1.36	1.94	2.25	0.43	0.48	0.51
tbl	4.80***	5.48***	5.64***	4.03	4.01	3.86	0.72	0.71	0.69
tsp	4.28***	4.42***	4.28***	3.40	3.17	2.93	0.66	0.62	0.59
rvar	2.17**	3.41***	3.57***	1.77	2.43	2.40	0.46	0.52	0.52
mv	2.72***	3.59***	4.30***	2.37	2.72	3.05	0.52	0.56	0.60
pc	3.75***	4.57***	4.25***	3.15	3.45	2.99	0.63	0.65	0.60
comb1	3.97***	4.51***	4.51***	3.27	3.25	3.09	0.62	0.62	0.61
comb2	3.77***	4.47***	4.96***	2.84	3.05	3.22	0.56	0.59	0.61
comb3	2.26**	2.26**	2.26**	1.90	1.90	1.90	0.46	0.46	0.46

Table VI
Pocket Statistics (Monthly)

This table reports statistics on the duration of pockets (in days) and the integral R^2 of pockets for pockets estimated with monthly data. Coefficients are estimated using a one-sided kernel with a 2.5-year effective sample size and we regularize the resulting forecasts through a Bayesian shrinkage scheme with parameter $g = 2$. Pockets are identified as periods in which a fitted squared forecast error differential (relative to a prevailing mean forecast and estimated using a one-sided kernel with a one-year effective sample size) is above zero for a minimum of one month.

Statistics	**dp**	**tbl**	**tsp**	**rvar**
Num pockets	34.00	23.00	19.00	41.00
Fraction of sample	0.31	0.32	0.28	0.31
Duration				
Min	42.00	42.00	42.00	42.00
Mean	214.94	224.61	204.47	183.37
Max	546.00	672.00	651.00	609.00
Integral R^2				
Min	-0.11	-1.21	-0.28	-0.31
Mean	1.52	1.98	1.43	1.26
Max	6.12	8.50	8.54	7.11

Table VII

Out-of-Sample Measures of Forecasting performance (Monthly Benchmark Specification)

Panel A reports Clark and West (2007) test statistics for out-of-sample return predictability measured relative to a prevailing mean forecast. Panel B reports 3 measures of economic significance associated with returns on a portfolio that uses the time-varying coefficient model forecast in-pocket and the prevailing mean forecast out-of-pocket to allocate between the risk-free asset and the market (portfolio weights are limited to be between zero and two): the annualized estimated alpha in percentage points, the HAC t-statistic for the estimated alpha, and the annualized Sharpe ratio of the portfolio. We use a backward-looking kernel with an effective sample size of 2.5 years to compute forecasts and regularize the forecasts through a Bayesian shrinkage scheme. "pc" is a recursively computed first principal component of the four predictor variables. "mv" is a four-variable multivariate forecast estimated using a product kernel. "comb1," "comb2," and "comb3" refer to a simple average of the univariate forecasts. "comb1" sets an individual predictor's forecast to the time-varying coefficient model forecast during a pocket and to the prevailing mean otherwise. "comb2" is the same as "comb1" except it ignores individual predictor forecasts when that variable is not in a pocket but at least one other variable is in a pocket. "comb3" makes no distinction between in-pocket and out-of-pocket periods and always uses the simple equal-weighted average of all four univariate models. The CW test statistics approximately follow a normal distribution with positive values indicating more accurate out-of-sample return forecasts than the prevailing mean benchmark and negative values indicating the opposite. A pocket is classified as a period in which a fitted squared forecast error differential (estimated using a one-sided kernel with a one-year effective sample size) is above zero in the preceding period, for a minimum of 21 days. Consider a particular statistic of interest, β. $*, **$, and $***$ represent statistical significance at the 10%, 5%, and 1% level from a hypothesis test of $\beta = 0$ against an alternative hypothesis of $\beta > 0$. $\dagger$, $\dagger\dagger$, and $\dagger\dagger\dagger$ represent statistical significance at the 10%, 5%, and 1% level from a hypothesis test of $\beta = 0$ against an alternative hypothesis of $\beta < 0$.

Panel A: Clark-West statistics

	Unrestricted				**+ excess return forecasts**				**All sign restrictions**			
Variables	**Full sample**	**In-pocket**	**Out-of-pocket**	**Diff.**	**Full sample**	**In-pocket**	**Out-of-pocket**	**Diff.**	**Full sample**	**In-pocket**	**Out-of-pocket**	**Diff.**
dp	1.06	1.23	0.09	1.62*	0.93	1.42*	−0.08	1.67**	1.07	1.27	0.29	1.35*
tbl	1.33*	1.59*	−0.22	2.16**	1.48*	1.99**	−0.10	2.41***	1.73**	1.81**	0.56	1.71**
tsp	0.88	1.52*	−0.96	2.47***	0.92	1.58*	−0.49	2.21**	1.07	1.65**	−0.34	2.28**
rvar	0.93	1.51*	−0.78	2.28**	1.02	1.71**	−0.47	2.26**	1.39*	1.57*	0.39	1.74**
mv	0.83	1.70**	−0.54	2.19**	1.56*	1.85**	0.35	1.90**	1.56*	1.85**	0.35	1.90**
pc	0.79	1.58*	$-1.32^{\dagger}$	2.66***	0.85	1.58*	−0.68	2.26**	0.85	1.58*	−0.68	2.26**
comb1	1.51*	1.50*	–	–	1.59*	1.68**	–	–	1.52*	1.59*	–	–
comb2	1.57*	1.51*	–	–	1.49*	1.55*	–	–	1.48*	1.55*	–	–
comb3	1.10	1.33*	−0.40	2.15**	1.12	1.27	0.08	1.95**	1.38*	1.13	0.87	1.42*

Panel B: Economic significance

	Unrestricted				**+ excess return forecasts**				**All sign restrictions**			
Variables	$\hat{\alpha}$	$t_{\hat{\alpha}}$	**Sharpe Ratio**		$\hat{\alpha}$	$t_{\hat{\alpha}}$	**Sharpe Ratio**		$\hat{\alpha}$	$t_{\hat{\alpha}}$	**Sharpe Ratio**	
dp	2.11**	1.65	0.46	–	3.03**	1.84	0.48	–	2.80**	1.68	0.47	–
tbl	3.99***	3.04	0.59	–	4.34***	2.73	0.56	–	4.03***	2.38	0.53	–
tsp	3.15***	2.34	0.53	–	3.18**	1.99	0.49	–	3.27**	2.05	0.50	–
rvar	3.34***	2.60	0.55	–	3.64**	2.30	0.52	–	3.49**	2.19	0.51	–
mv	3.79***	2.93	0.59	–	3.85***	2.39	0.53	–	3.85***	2.39	0.53	–
pc	3.37***	2.46	0.54	–	3.23**	1.95	0.50	–	3.23**	1.95	0.50	–
comb1	3.51***	2.71	0.56	–	3.72***	2.40	0.53	–	3.54**	2.25	0.52	–
comb2	3.57***	2.62	0.56	–	3.60**	2.11	0.51	–	3.58**	2.11	0.51	–
comb3	0.81	0.69	0.40	–	1.17	0.63	0.40	–	2.41*	1.39	0.45	–
pm	−0.36**	−1.88	0.49	–	−0.36**	−1.88	0.49	–	−0.36**	−1.88	0.49	–

Table VIII

Out-of-Sample Measures of Forecasting Performance
(Fama-French Factor Portfolio Excess Returns, Daily)

Panel A reports Clark and West (2007) test statistics for out-of-sample return predictability measured relative to a prevailing mean forecast. Panel B reports three measures of economic significance associated with returns on a portfolio that uses the time-varying coefficient model forecast in-pocket and the prevailing mean forecast out-of-pocket to allocate between small and big or high and low (portfolio weights are limited to be between zero and two): the annualized estimated alpha in percentage points, the t-statistic on the estimated alpha, and the annualized Sharpe ratio of the portfolio. Significance of the estimated alpha is assessed using a t-statistic estimated using HAC standard errors. We use a backward-looking kernel to compute forecasts and regularize the forecasts through a Bayesian shrinkage scheme. "pc" is a recursively computed first principal component of the four predictor variables. "comb1," "comb2," and "comb3" refer to a simple average of the univariate forecasts. "comb1" sets an individual predictor's forecast to the time-varying coefficient model forecast during a pocket and to the prevailing mean otherwise. "comb2" is the same as "comb1" except it ignores individual predictor forecasts when that variable is not in a pocket but at least one other variable is in a pocket. "comb3" makes no distinction between pocket and nonpocket periods and always uses the simple equal-weighted average of all four univariate models. The CW test statistics approximately follow a normal distribution with positive values indicating more accurate out-of-sample return forecasts than the prevailing mean benchmark and negative values indicating the opposite. A pocket is classified as a period in which a fitted squared forecast error differential (estimated using a one-sided kernel with a one-year effective sample size) is above zero in the preceding period, for a minimum of 21 days. Consider a particular statistic of interest, β. $*, **$, and $***$ represent statistical significance at the 10%, 5%, and 1% level from a hypothesis test of $\beta = 0$ against an alternative hypothesis of $\beta > 0$. $\dagger$, $\dagger\dagger$, and $\dagger\dagger\dagger$ represent statistical significance at the 10%, 5%, and 1% level from a hypothesis test of $\beta = 0$ against an alternative hypothesis of $\beta < 0$.

Panel A: Clark-West statistics

	SMB				HML			
Variable	**Full sample**	**In-pocket**	**Out-of-pocket**	**Difference**	**Full sample**	**In-pocket**	**Out-of-pocket**	**Difference**
dp	2.72***	2.36***	1.86**	1.90**	1.68**	2.04**	0.49	1.62*
tbl	2.78***	1.69**	2.33***	0.82	1.40*	1.86**	0.33	1.58*
tsp	2.80***	1.99**	2.18**	1.28*	2.27**	3.14***	0.30	2.75***
rvar	2.60***	2.76***	0.85	2.40***	0.48	2.78***	$-2.03^{\dagger\dagger}$	3.73***
mv	4.81***	3.11***	4.60***	1.32*	4.19***	4.09***	1.97**	2.80***
pc	3.37***	2.51***	2.61***	1.81**	2.09**	2.45***	0.73	1.84**
comb1	3.06***	3.12***	–	–	2.57***	2.69***	–	–
comb2	3.64***	3.81***	–	–	2.63***	2.76***	–	–
comb3	3.80***	3.41***	1.81**	1.73**	1.61*	2.12**	0.04	2.30**

Panel B: Economic significance

	SMB				HML			
Variable	$\hat{\alpha}$	$t_{\hat{\alpha}}$	**Sharpe ratio**		$\hat{\alpha}$	$t_{\hat{\alpha}}$	**Sharpe ratio**	
dp	2.75***	2.51	0.65	–	1.65***	2.50	0.68	–
tbl	1.31**	2.08	0.41	–	1.46**	2.18	0.64	–
tsp	1.33***	2.60	0.50	–	2.21***	3.17	0.74	–
rvar	2.05***	3.04	0.59	–	2.34***	3.46	0.79	–
mv	2.73***	2.78	0.63	–	2.51***	3.82	0.80	–
pc	1.44***	2.64	0.48	–	1.88***	2.66	0.69	–
comb1	3.04***	3.64	0.79	–	2.08***	3.12	0.75	–
comb2	4.15***	4.25	0.95	–	2.43***	3.17	0.75	–
comb3	3.89***	3.93	0.83	–	2.73***	2.77	0.76	–
pm	-0.38	-0.75	0.17	–	$-0.17^{\dagger}$	-1.55	0.62	–

Table IX

OOS Asset Pricing Model Simulations (Unrestricted - new baseline identification of pockets)

This table reports Monte Carlo simulation results of our one-sided kernel estimation applied to data simulated from four different asset pricing models (this includes two specifications of Wachter's rare disasters model, one of which omits data from disaster episodes). We report seven statistics. The first four are Clark and West (2007) t-statistics relative to a prevailing mean benchmark in the full sample, in-pocket, out-of-pocket, and the difference between in and out-of-pocket. The final three are economic statistics associated with returns on a portfolio that uses the time-varying coefficient model forecast in-pocket and the prevailing mean forecast out-of-pocket to allocate between the risk-free asset and the market (portfolio weights are limited to be between zero and two): the annualized estimated alpha in percentage points, the HAC t-statistic associated with that alpha, and the annualized Sharpe ratio of the portfolio. Column (2) presents the corresponding statistics from the data for reference.

		Bansal-Yaron			**Campbell-Cochrane**			**Garleanu-Panageas**			**Wachter**			**Wachter (no disasters)**		
Stats	**Sample**	**Avg.**	**Std. err.**	**p-val**	**Avg.**	**Std. err.**	**p-val**	**Avg.**	**Std. err.**	**p-val**	**Avg.**	**Std. err.**	**p-val**	**Avg.**	**Std. err.**	**p-val**
Panel A: dp																
CW_{FS}	0.32	0.09	1.00	0.40	0.16	0.64	0.41	0.23	0.91	0.46	0.32	0.99	0.50	0.40	1.20	0.47
CW_{IP}	1.25	0.07	0.97	0.11	−0.04	0.89	0.08	0.10	0.98	0.13	0.10	1.02	0.13	0.22	1.10	0.15
CS_{OOP}	−1.03	0.04	1.00	0.85	0.19	0.79	0.93	0.19	0.94	0.91	0.26	1.02	0.90	0.33	1.15	0.89
CW_{DIFF}	2.00	0.02	0.97	0.02	−0.14	0.97	0.02	−0.04	0.97	0.02	−0.04	1.02	0.03	−0.02	1.01	0.02
$\hat{\alpha}$	2.74	−0.36	1.72	0.03	0.01	0.97	0.00	−0.04	0.87	0.00	0.16	2.04	0.08	0.22	1.44	0.04
$t_{\hat{\alpha}}$	1.94	−0.21	0.97	0.01	0.02	0.99	0.02	−0.04	1.00	0.02	0.22	1.08	0.06	0.16	1.02	0.04
SR	0.48	0.44	0.13	0.37	0.47	0.07	0.46	0.33	0.11	0.07	0.46	0.12	0.42	0.58	0.10	0.87
Panel B: risk-free																
CW_{FS}	1.34	0.10	1.00	0.12	0.16	0.65	0.04	0.22	0.90	0.11	0.33	0.99	0.14	0.41	1.20	0.18
CW_{IP}	2.29	0.08	0.96	0.01	−0.04	0.89	0.00	0.10	0.98	0.01	0.10	1.03	0.01	0.23	1.10	0.04
CS_{OOP}	−1.05	0.05	1.00	0.86	0.20	0.80	0.94	0.18	0.94	0.91	0.26	1.01	0.91	0.32	1.16	0.89
CW_{DIFF}	3.14	0.03	0.96	0.00	−0.14	0.96	0.00	−0.03	1.00	0.00	−0.04	1.03	0.00	−0.01	1.02	0.00
$\hat{\alpha}$	5.48	−0.34	1.72	0.00	0.01	0.97	0.00	−0.04	0.87	0.00	0.16	2.03	0.01	0.22	1.43	0.00
$t_{\hat{\alpha}}$	4.01	−0.20	0.98	0.00	0.03	0.99	0.00	−0.05	0.99	0.00	0.22	1.08	0.00	0.16	1.01	0.00
SR	0.71	0.44	0.13	0.02	0.47	0.07	0.00	0.33	0.11	0.00	0.46	0.12	0.02	0.58	0.10	0.08
Panel C: rvar																
CW_{FS}	0.57	0.10	1.02	0.34	0.07	0.67	0.22	0.19	0.91	0.33	0.22	1.00	0.35	0.32	1.21	0.37
CW_{IP}	1.39	0.08	0.99	0.09	−0.17	0.92	0.04	0.04	0.98	0.08	−0.01	1.02	0.07	0.12	1.09	0.10
CS_{OOP}	−0.83	0.04	1.01	0.80	0.17	0.80	0.90	0.17	0.94	0.86	0.21	1.01	0.85	0.28	1.15	0.84
CW_{DIFF}	2.07	0.04	1.05	0.02	−0.25	1.07	0.01	−0.05	1.09	0.03	−0.10	1.07	0.03	−0.06	1.07	0.03
$\hat{\alpha}$	3.41	−0.40	1.85	0.02	−0.07	1.02	0.00	−0.05	0.89	0.00	0.04	2.04	0.04	0.11	1.47	0.01
$t_{\hat{\alpha}}$	2.43	−0.24	1.01	0.00	−0.07	1.01	0.01	−0.08	0.98	0.01	0.11	1.04	0.01	0.04	1.01	0.01
SR	0.52	0.44	0.13	0.26	0.47	0.07	0.23	0.33	0.11	0.04	0.45	0.12	0.28	0.58	0.09	0.71

Table X

Sticky Expectations Model Simulation Results

This table reports Monte Carlo results for the one-sided kernel empirical results using simulated data from the sticky expectations model. We generate 500 bootstrap samples of the same sample size as is available for each predictor in the data for three separate calibrations. “Baseline” refers to the standard calibration with sticky expectations, “Baseline ($\lambda = 0$)” refers to the “Baseline” calibration but with rational expectations (i.e., $\lambda = 0$), and “RE Recalibrated” refers to a recalibration of the rational expectations model to match the target moments. “dp” refers to the log dividend-price ratio, “rf” refers to the log risk-free rate, and “rvar” refers to realized variance on a 60-day trailing window. A pocket is classified as a period in which a fitted (using a one-sided kernel with a one-year effective sample size) squared forecast error differential is above zero in the preceding period. For each predictor and each calibration, we report six statistics. The first four are Clark and West (2007) t-statistics relative to a prevailing mean benchmark in the full sample, in-pocket, out-of-pocket, and the difference between in and out-of-pocket. The final three are economic statistics associated with returns on a portfolio that uses the time-varying coefficient model forecast in-pocket and the prevailing mean forecast out-of-pocket to allocate between the risk-free asset and the market (portfolio weights are limited to be between zero and two): the annualized estimated alpha in percentage points, the HAC t-statistic associated with that alpha, and the annualized Sharpe ratio of the portfolio. The column “Data” reports the corresponding statistics from the data for reference.

		Baseline			**Baseline ($\lambda = 0$)**			**RE Recalibrated**		
Stats	**Data**	**Avg**	**Std. err**	**p-val**	**Avg**	**Std. err**	**p-val**	**Avg**	**Std. err**	**p-val**
					dp					
CW_{fs}	0.32	1.31	1.13	0.83	0.30	0.97	0.50	0.29	0.91	0.50
CW_{ip}	1.25	2.85	1.22	0.90	0.11	1.00	0.14	0.03	0.94	0.11
CW_{oop}	−1.03	−0.61	0.97	0.69	0.27	0.99	0.89	0.30	0.93	0.92
CW_{diff}	2.00	2.90	1.25	0.77	−0.08	1.00	0.01	−0.16	0.92	0.01
α	2.74	1.98	1.59	0.31	0.01	1.82	0.08	0.16	1.36	0.03
t_α	1.94	1.39	1.08	0.30	0.01	0.98	0.03	0.11	0.98	0.03
SR	0.48	0.52	0.18	0.58	0.32	0.11	0.07	0.43	0.10	0.28
					rf					
CW_{fs}	1.34	2.38	1.13	0.82	0.22	0.95	0.12	0.08	0.87	0.07
CW_{ip}	2.29	3.15	1.17	0.76	−0.01	0.99	0.01	−0.13	0.95	0.01
CW_{oop}	−1.05	0.42	1.01	0.93	0.26	0.93	0.92	0.17	0.89	0.91
CW_{diff}	3.14	2.94	1.30	0.43	−0.12	1.10	0.00	−0.21	1.06	0.00
α	5.48	3.78	1.68	0.16	−0.18	1.82	0.00	−0.37	1.35	0.00
t_α	4.01	2.55	1.07	0.09	−0.13	0.98	0.00	−0.28	0.96	0.00
SR	0.71	0.61	0.17	0.30	0.32	0.11	0.00	0.43	0.10	0.01
					rvar					
CW_{fs}	0.57	1.84	1.13	0.87	0.19	0.95	0.34	0.05	0.86	0.29
CW_{ip}	1.39	2.98	1.25	0.90	0.08	0.99	0.09	−0.11	0.98	0.05
CW_{oop}	−0.83	−0.04	0.98	0.80	0.16	0.93	0.87	0.11	0.89	0.85
CW_{diff}	2.07	2.79	1.28	0.74	−0.01	1.04	0.02	−0.16	1.06	0.01
α	3.41	3.18	1.65	0.44	−0.36	1.90	0.02	−0.54	1.39	0.01
t_α	2.43	2.09	1.05	0.39	−0.22	0.99	0.01	−0.39	0.96	0.00
SR	0.52	0.56	0.18	0.60	0.32	0.11	0.05	0.43	0.10	0.18