The Journal of FINANCE
The Journal of THE AMERICAN FINANCE ASSOCIATION

THE JOURNAL OF FINANCE • VOL. LXXVIII, NO. 3 • JUNE 2023

Pockets of Predictability

LELAND E. FARMER, LAWRENCE SCHMIDT, and ALLAN TIMMERMANN[*]

ABSTRACT

For many benchmark predictor variables, short-horizon return predictability in the U.S. stock market is local in time as short periods with significant predictability ("pockets") are interspersed with long periods with no return predictability. We document this result empirically using a flexible time-varying parameter model that estimates predictive coefficients as a nonparametric function of time and explore possible explanations of this finding, including time-varying risk premia for which we find limited support. Conversely, pockets of return predictability are consistent with a sticky expectations model in which investors slowly update their beliefs about a persistent component in the cash flow process.

RESEARCHERS HAVE LONG BEEN INTERESTED in the extent to which stock returns are predictable. Over the last several decades, time-varying risk premia have been widely suggested as a key source of fluctuations in stock prices, and many workhorse macrofinance models seek to exogenously generate large fluctuations in discount rates on the aggregate stock market. Both welfare calculations and normative predictions about optimal investment strategies are often quite different in the presence of return predictability. At the same time, these findings have been met with some skepticism given a number of studies that find empirical evidence that return predictability is highly unstable, varying greatly over time and across markets and being difficult to exploit out-of-sample.[1]

[*]Allan Timmermann is at University of California, San Diego. Leland E. Farmer is at University of Virginia. Lawrence Schmidt is at Massachusetts Institute of Technology. We acknowledge constructive and insightful comments from the Editor, Stefan Nagel, an Associate Editor, and two anonymous referees. We also received helpful comments from Frank Diebold, Xavier Gabaix, Bradley Paye, Hashem Pesaran, and seminar participants at Penn, Boston University, NC State University, UCSD, University of Warwick, University of Virginia, University of British Columbia, Edhec, the 2018 SoFiE meetings in Lugano, the 2018 SITE meetings at Stanford, and the 2018 IAAE conference in Montreal. We thank Yury Olshanskiy and Victor Sellemi for outstanding research assistance. We have read *The Journal of Finance* disclosure policy and have no conflicts of interest to disclose.

Correspondence: Allan Timmermann, Rady School of Management, University of California, San Diego, 9500 Gilman Drive, La Jolla CA 92093-0553, USA; e-mail: atimmermann@ucsd.edu.

[1] For early studies, see, for example, Campbell (1987), Fama and French (1988, 1989), Keim and Stambaugh (1986), and Pesaran and Timmermann (1995). Lettau and Ludvigson (2010) and Rapach and Zhou (2013) review the extensive literature on return predictability. Paye and

DOI: 10.1111/jofi.13229

Existing evidence on return predictability has been established mostly using linear, constant-coefficient regressions that pool information across long historical spans of time and thus are designed to establish whether stock returns are predictable "on average," that is, across potentially very different economic states. Inference on the resulting coefficients may yield misleading and unstable results if, in fact, return predictability shifts over time. To address this possibility, our paper adopts a new estimation strategy capable of identifying patterns in return predictability that are "local" in time. Specifically, we estimate predictive regressions with time-varying parameters based on one-sided kernel regressions that allow the coefficients to follow a smooth, nonparametric function of calendar time. Unlike alternative approaches that impose tight parametric restrictions on how predictive coefficients evolve over time, we do not need to take a stand on the return-generating process. [2] Next, we use a local-trend estimation approach to identify periods in which forecasts from the local kernel regressions were more accurate than those from a prevailing mean benchmark model. Following studies such as Pesaran and Timmermann (1995) and Welch and Goyal (2008), who emphasize the need for out-of-sample return predictability, our approach is fully out-of-sample, avoiding the use of any future data, and we let the data determine both how large predictability is at a given point in time and how long it lasts.

Using this approach, we present new empirical evidence that short-horizon return predictability is quite concentrated, or local in time, and tends to fall in certain (contiguous) "pockets." For example, using the term spread as a predictor variable over a 63-year period, our approach identifies in real time seven pockets whose duration lasts between four months and two years, so that in total 15% of the sample is spent inside pockets with return predictability. As another illustration of the extent to which short-horizon predictability concentrates in time, we estimate univariate regression models with constant coefficients using our predictors over two subsamples — those observations associated with our ex ante identified "pockets" and all other periods. We find strong evidence of in-pocket return predictability and essentially no statistically significant evidence of predictability outside of pockets, despite the fact that the vast majority of our sample falls outside of these pocket periods, where we would have more statistical power to detect predictability.

To quantify the amount of local return predictability, and to calibrate the amount of predictability to expect under conventional asset pricing models, we compute Clark and West (2007) statistics that compare out-of-sample mean-squared prediction errors from the local kernel regressions to those from a

Timmermann (2006), Rapach and Wohar (2006), and Chen and Hong (2012) find evidence of parameter instability for stock market return prediction models.

[2] Several studies adopt parametric assumptions about time variation in the return-generating process. For example, Henkel, Martin, and Nardari (2011) use regime-switching models to capture changes in stock return predictability, while Dangl and Halling (2012) and Johannes, Korteweg, and Polson (2014) use time-varying parameter models to track predictability in stock returns. Like any other nonparametric approach, we do have to pick a bandwidth parameter, but our findings are robust to choices of this parameter across a wide range of values.

prevailing mean benchmark. Next, we conduct a battery of simulation exercises that assess the extent to which we can statistically reject the null hypotheses of no predictability or predictability associated with a constant coefficient model. Mirroring the above analysis, we conduct these tests for the full sample as well as for ex ante identified in-pocket and out-of-pocket subperiods. For the full sample, that is, "on average," we find no statistical evidence that our local kernel regressions outperform the prevailing mean across the univariate or multivariate models that we consider. Results deteriorate substantially outside of pockets; the time-varying coefficient models always underperform prevailing mean forecasts, sometimes by a significant margin. These findings echo a number of empirical results from the literature (Welch and Goyal (2008)) indicating the difficulty of detecting out-of-sample return predictability, a phenomenon that is exacerbated in our context given that our local regressions are subject to larger estimation error relative to standard approaches.

The picture changes substantially, however, inside ex ante identified pockets, where we find strong evidence of return predictability across a range of univariate and multivariate models. Consistent with prior literature, these results generally improve further if we impose economically motivated restrictions on our expected return forecasts or incorporate multivariate information, for example, by combining forecasts from univariate models.[3]

To quantify the economic value of our ability to detect significant out-of-sample return predictability, we construct managed portfolios that use our ex ante expected excess return forecasts to dynamically rebalance a portfolio comprising the market and a risk-free asset. Although such a strategy earns conditional capital asset pricing model (CAPM) alphas of zero by construction, it generates sizable unconditional CAPM alphas; for example, our two best forecast combination-based strategies deliver annualized CAPM alphas (t-statistics) of 6.4% (6.1) and 6.1% (5.7), respectively, while univariate return prediction models generate alphas in the range of 2% to 4% with highly significant t-statistics. These results are robust to controlling for volatility and momentum factors and hold net of proportional transaction costs as high as 10 basis points (bps).

In an additional set of tests, we repeat these analyses with the Fama-French SMB and HML factors. In both cases we find similar, and often even stronger, results. While all predictors underperform outside of pockets, we detect substantial statistical evidence for out-of-sample predictability inside of pockets. Likewise, our market-timing exercises deliver substantial and economically meaningful gains in risk-adjusted performance.

We conduct a battery of additional tests to ensure the robustness of our main results. In particular, we vary the length of the windows used to estimate the parameters of the local kernel regressions and identify pockets, we separately consider null hypotheses with zero or constant slope coefficients

[3] See, for example, Campbell and Thompson (2008), Kelly and Pruitt (2013), Pettenuzzo, Timmermann, and Valkanov (2014), Rapach, Strauss, and Zhou (2010), and Timmermann (2006).

on the state variables, we examine an alternative "local prevailing mean" benchmark that accounts for possible return momentum, and we examine the effect of Stambaugh (1999) bias. In all cases, we show that our empirical findings are not sensitive to the setup of our baseline analysis. Moreover, to make our findings more directly comparable to the extant literature, we apply our real-time, local predictability approach to monthly stock returns. We again find that our local out-of-sample return predictions are significantly more accurate than the prevailing mean benchmark inside ex ante identified pockets while the reverse holds outside of pockets and that our approach can lead to economically important improvements over existing methods from the return predictability literature. [4]

With these new empirical results in hand, we next explore which economic mechanisms are capable of generating pockets of local return predictability. We start by conducting return simulations from four workhorse rational expectations asset pricing models that represent a wide range of mechanisms and are representative of the dynamics of returns and state variables implied by models exhibiting time-varying risk premia. These include the long-run risk model of Bansal and Yaron (2004), the habit formation model of Campbell and Cochrane (1999), the heterogeneous agent model of Gârleanu and Panageas (2015), and the rare disaster model of Wachter (2013). All of these models are calibrated to generate dynamics that are consistent with the data, in the sense that increases in risk premia tend to correspond with slow-moving changes in discount rates. Accordingly, the state variables governing return predictability are highly persistent, signal-to-noise ratios for predictive regressions are extremely low, and innovations to predictors such as the dividend-price ratio have very strong negative correlations with realized returns. As such, positive shocks to the discount rate, especially those large enough to be detectable, will generate large negative realized returns that at least temporarily lead to exactly the wrong inference about the predictive relationship (Stambaugh (1999)). This makes it challenging to detect state-dependent return predictability in such models.

Consistent with this intuition, we find that none of these workhorse models is capable of matching the empirically observed out-of-sample predictive accuracy associated with in-pocket periods.[5] Turning to the economic performance (market-timing) results, the average alpha estimates are usually close to zero and statistically insignificant. Both of these results indicate that the

[4] Henkel, Martin, and Nardari (2011), Dangl and Halling (2012), and Rapach, Strauss, and Zhou (2010) argue that return predictability is largely confined to recession periods. In unreported results, we find that the link between economic recessions and our return predictability pockets is rather weak and that the stage of the economic cycle explains only a small part of the time-variation in expected returns that we document. Movements in an investor sentiment indicator (Baker and Wurgler (2006, 2007)) or changes in broker-dealer leverage (Adrian, Etula, and Muir (2014)) tracking availability of arbitrage capital, also do not correlate strongly with the time-variation in return predictability that we document.

[5] Matching the full-sample or out-of-pocket results is less challenging, in part because the out-of-sample accuracy of our predictive return regressions is fairly weak overall.

benchmark asset pricing models fail to generate short-lived pockets of substantial predictability that is detectable via our local kernel regressions, which suggests that the very features that allow the asset pricing models to replicate a number of stylized facts about equity returns in the data combine to create substantial potential for estimation error to dominate the small amount of true ex ante predictability generated by the time-varying risk premia in the model.

Motivated by a recent and rapidly growing literature at the intersection of macroeconomics and finance (Coibion and Gorodnichenko (2015), Bouchaud et al. (2019)), we finally consider an alternative explanation for our observed results. Specifically, we consider the potential implications for high-frequency return predictability of a model in which agents have sticky expectations, underreacting to news in a manner consistent with both theoretical work and a large body of empirical evidence.[6] We propose a stylized asset pricing model in which agents price cash flows according to a loglinearized dynamic dividend discount model where prices equal the sum of expected cash flows discounted by time-varying subjective discount rates. However, we deviate from the rational expectations benchmark by assuming that agents' beliefs about future cash flows adjust sluggishly to new information relative to the true data-generating process. In other words, whereas agents believe that expected excess returns are governed by a set of slow-moving state variables similar to the workhorse models discussed above, expected returns feature an additional, high-frequency component under the objective probability distribution. This extra term captures the difference between agents' subjective forecasts of expected cash flow growth rates and the true state variable governing expected cash flow growth rates. The presence of this term implies that prices exhibit "local factor momentum": recent changes in valuation ratios signal the likelihood that future valuations will continue to drift upward, a pattern that is counter to the long-run mean reversion in prices that is expected from time-varying discount rates.

We calibrate our model to match a number of observable asset pricing moments. We then perform a number of simulation exercises to assess the extent to which such a model generates pockets of predictability. Importantly, the degree of stickiness of beliefs is disciplined by external estimates based on analysts' forecasts of macroeconomic quantities from Coibion and Gorodnichenko (2015). We next compare simulations from our sticky expectations

[6] Early theoretical papers on sluggish adjustments in expectations include Mankiw and Reis (2002) ,Woodford (2003) and Sims (2003). A number of empirical papers present evidence on underreaction to aggregate news at short horizons. See, for example, Moskowitz and Grinblatt (1999), Hong, Lim, and Stein (2000), Hong, Torous, and Valkanov (2007), Hou (2007), and Bouchaud et al. (2019), who present evidence of slow diffusion of stock- or industry-specific information in stock markets. Katz, Lustig, and Nielsen (2017) also find evidence of underreaction of asset prices to fluctuations in inflation rates across countries. Turning to fixed-income markets, d'Arienzo (2020) and Wang (2020) present evidence indicates that yields underreact to macro news at short horizons but overreact at longer horizons, which relates to a puzzle identified by Giglio and Kelly (2018). See also Bordalo et al. (2020) and Angeletos, Huo, and Sastry (2021) for additional empirical evidence and discussion of the related empirical and theoretical literature on this subject.

benchmark with analogous data simulated from a rational expectations model with the same cash flow and subjective discount rate dynamics. Once again, our local kernel regressions are unable to detect statistically or economically significant out-of-sample return predictability in the specifications that impose rational expectations. However, despite the fact that local predictability is not targeted, we find that the sticky expectations model can replicate the degree of out-of-sample return predictability observed in the data, a pattern that is robust across predictors and econometric specifications.

In our sticky expectations model, one source of return predictability is the "belief discrepancy" between agents' cash flow expectations versus the "correct" forecasts conditional on the true data-generating process. The presence of such a belief distortion acts as an important additional channel through which expected returns are forecastable by the econometrician in such models. [7] We conclude by providing direct evidence linking our expected return forecasts with data on forecast errors of professional forecasters. Consistent with predictions of the theory, above-average forecasts from all of our time-varying coefficient models predict positive forecast errors in the future. In other words, sluggish updating of agents' beliefs implies that returns are predictable because future cash flow "shocks"—deviations between realizations and agents' subjective expectations—are forecastable. Our local return forecasts capture a nontrivial fraction of this variation.[8]

The rest of the paper proceeds as follows. Section I discusses conventional approaches to modeling return predictability and introduces our nonparametric methodology for identifying pockets with local return predictability. Section II introduces our daily data and presents empirical evidence on return predictability pockets. This section also uses simulations to address whether the pockets could be generated spuriously as a result of repeated use of correlated tests for local return predictability. Section III evaluates the statistical and economic performance of our nonparametric return forecasts and conducts a number of robustness checks. Section IV considers whether a suite of workhorse asset pricing models with time-varying risk premia are capable of generating return predictability pockets. Section V presents our framework with sticky expectations, illustrates that a calibrated model can match a number of empirical results, presents empirical evidence linking our ex ante expected return

[7] The effect of such a wedge on local return predictability depends on the sequence of recent shocks to the cash flow, risk premium, and risk-free rate processes in the model. Because the sequence of shocks is never exactly the same as has occurred previously and expectations are sticky, pockets of return predictability will never be "learned away" by agents. This is in contrast to papers such as Green, Hand, and Soliman (2011) and McLean and Pontiff (2016), who imply that patterns of return predictability that can be exploited for economic gains will vanish once discovered by agents. See also Schwert (2003) and Timmermann (2008).

[8] See also Bouchaud et al. (2019) and Gómez-cram (2022) for related evidence using forecast errors aggregated from equity analysts' earnings forecasts. Gómez-cram (2022) introduces a sticky expectations model that relates return predictability to turning points of the business cycle. The mechanism of his model, along with his empirical results, are quite different from ours since we rely on nonparametric methods and find only a weak association between business cycle variation and pockets with local return predictability.

forecasts with future macroeconomic forecast errors. Section VI concludes. An Internet Appendix contains additional technical material and empirical results.[9]

I. Prediction Models and Estimation Methodology

This section briefly discusses the conventional constant-coefficient return prediction model before introducing the nonparametric regression methodology that we use to identify time-variation in return predictability.

A. Conventional Return Predictability Model

A large empirical literature summarized in Welch and Goyal (2008) and Rapach and Zhou (2013) studies predictability of stock returns using linear, constant-coefficient models of the form

$$r_{s,t+1} - r_{f,t+1} = x_t'\beta + \varepsilon_{t+1}, \tag{1}$$

where $r_{s,t+1}$ is the stock market return and $r_{f,t+1}$ is the risk-free rate, both measured in period $t+1$, so that $r_{t+1} \equiv r_{s,t+1} - r_{f,t+1}$ measures the excess return, x_t is a $(d \times 1)$ vector of covariates (predictors) that could include a constant, and ε_{t+1} is an unobservable disturbance with $\mathbb{E}[\varepsilon_{t+1}|x_t] = 0$.

In Section I of the Internet Appendix, we show that the specification in (1) is consistent with a broad class of affine asset pricing models exhibiting time-variation in either the quantity or the price of risk. For example, (1) holds approximately in a representative agent model where agents have Epstein and Zin (1989) preferences when aggregate consumption growth is an affine function of state variables that follow a stationary vector autoregressive process.[10] This setting includes many of the specifications considered in the literature on consumption-based asset pricing models with long-run risks and rare disasters and also holds under incomplete markets with state-dependent higher moments of uninsurable idiosyncratic shocks.[11] As we further demonstrate in this appendix, subject to certain restrictions, (1) can also allow for time-variation in the price of risk and thus nests many models that have been used to characterize the term structure of interest rates as well as the log-linearized stochastic discount factor (SDF) habit formation model of Campbell and Cochrane (1999).

Despite its theoretical appeal, the empirical validity of the assumption of constant regression coefficients in the linear return regression (1) has been challenged in studies such as Paye and Timmermann (2006) , Rapach and Wohar (2006), Chen and Hong (2012), Dangl and Halling (2012), and Johannes,

[9] The Internet Appendix is available in the online version of this article on *The Journal of Finance* website.

[10] See, for example, Bansal and Yaron (2004), Hansen, Heaton, and Li (2008), Eraker and Shaliastovich (2008), and Drechsler and Yaron (2011).

[11] See, for example, Constantinides and Duffie (1996), Constantinides and Ghosh (2017), Schmidt (2020), and Herskovic et al. (2015).

Korteweg, and Polson (2014), all of whom find strong evidence that this assumption is empirically rejected for U.S. stock returns using standard predictor variables. We therefore consider an econometric framework that can accommodate unstable coefficients.

B. A Flexible Time-Varying Parameter Model

We generalize (1) to allow for time-varying return predictability of the form

$$r_{t+1} = x_t'\beta_t + \varepsilon_{t+1}, \tag{2}$$

where the regression coefficients β_t are now subscripted with t to indicate that they are functions of time as a means of allowing for time-varying return predictability. We also allow for general forms of conditional heteroskedasticity $\sigma_t^2 \equiv \mathbb{E}[\varepsilon_t^2|x_t] = \sigma^2(x_t)$. To economize on notation, we let r_{t+1} denote the log excess market return minus its sample mean and assume that the predictor variables x_t are de-meaned prior to running the regression.

To identify periods with return predictability, we follow the nonparametric estimation strategy developed in Robinson (1989) and Cai (2007) that is valid regardless of whether the linear return prediction model in (1) is correctly specified. Using nonparametric methods for pocket identification offers the major advantage that we do not need to take a stand on the dynamics of local return predictability, for example, whether such predictability is short-lived or long-lived and whether it disappears slowly or rapidly. Instead, our nonparametric methods allow us to characterize the "shape" of the pockets, for example, the duration and frequency of pockets and the amount of return predictability inside the pockets that can provide important clues about the economic sources of return predictability.[12]

The nonparametric approach views $\beta : [0, 1] \to \mathbb{R}^d$ as a smooth function of time that can have at most finitely many discontinuities. The problem of estimating β_t for $t = 1, \ldots, T$ can then be thought of as estimating the function β at finitely many points $\beta_t = \beta(\frac{t}{T})$.[13]

Although Section II of the Internet Appendix provides additional details, our basic approach for the nonparametric analysis is as follows. We use a local constant model to compute the estimator of β_t as

$$\hat{\beta}_t = \underset{\beta_0 \in \mathbb{R}^d}{\arg\min} \sum_{s=1}^{T} K_{hT}(s-t)\big[r_{t+1} - x_s'\beta_0\big]^2. \tag{3}$$

The weights on the local observations are controlled through the kernel $K_{hT}(u) \equiv K(u/hT)/(hT)$, where h is the bandwidth. The estimator in (3) can

[12] Although nonparametric kernel regression is not widely used in finance, papers such as Ang and Kristensen (2012) have used this approach to estimate and test conditional CAPM alphas and betas.

[13] Because time t is normalized by the number of observations T, β is a function whose domain is [0,1] as opposed to $[0, T]$. This is useful because we need more and more local information to consistently estimate β_t as $T \to \infty$.

be viewed as a series of weighted least squares regressions with Taylor expansions of β around each point t/T. The weighting of observations in (3) can be contrasted with the familiar rolling-window estimator that uses a flat kernel that puts equal weight on observations in a certain neighborhood. For this estimator $K_{hT}(s-t) = 1$ if $t \in [t - \lfloor hT \rfloor, t + \lfloor hT \rfloor]$, otherwise $K_{hT}(s-t) = 0$. Our preferred estimator differs from the conventional rolling-window approach—which can be a fairly inefficient way to pick up time-variation in β if the build-up and disappearance of such patterns is more gradual (i.e., β_t is smooth), as we might expect a priori — by allowing $K_{hT}(\cdot)$ to be a smooth function that decreases as it moves away further from t.[14]

To test whether local predictability could have been identified in real time, we estimate our model using a one-sided analog of the Epanechnikov kernel,

$$K(u) = \frac{3}{2}(1-u^2)1\{-1 < u < 0\}, \tag{4}$$

ensuring that only past data are used to capture local return predictability. Our baseline results use a 2.5-year one-sided bandwidth, chosen as half the length of a two-sided five-year kernel, which is a standard choice of rolling-window in many finance applications.

As a measure of relative predictive accuracy, define the squared error difference (SED) between some benchmark forecast, $\bar{r}_{t|t-1}$, and the forecast from the local regression model, $\hat{r}_{t|t-1}$:

$$SED_t = (r_t - \bar{r}_{t|t-1})^2 - (r_t - \hat{r}_{t|t-1})^2. \tag{5}$$

Periods in which $SED_t > 0$ indicate that the kernel regression produced a more accurate forecast (in a squared error sense) than the benchmark since it incurred a smaller (squared) forecast error.

To help identify such periods, we project SED_t on a constant and a time trend,

$$SED_t = \gamma_{0,t} + \gamma_{1,t}t + v_t. \tag{6}$$

We estimate $\gamma_{0,t}$ and $\gamma_{1,t}$ again using a one-sided Epanechnikov kernel. We then define predictability pockets as periods for which $\widehat{SED}_t = \hat{\gamma}_{0,t} + \hat{\gamma}_{1,t}t > 0$. At the onset of a pocket, we would expect $\widehat{\gamma}_{1,t} > 0$, indicating that recent values of the benchmark model's squared forecast errors are beginning to exceed those from the local kernel model. Conversely, after the *SED* measure has peaked, we would expect $\widehat{\gamma}_{1,t} < 0$, indicating waning return predictability. [15]

[14] A rolling-window estimator loses some efficiency by not using any information from outside the fitting window and also by assigning the same weight to all observations inside the window. Usually, it is more efficient to give lower weight to observations far away from t relative to observations extremely close to t because the latter are presumably more representative than the former, and thus present a more favorable bias/variance trade-off.

[15] Indeed, this is a consistent pattern that we observe across all predictors in our empirical analysis.

Our estimates of $\gamma_{0,t}$ and $\gamma_{1,t}$ use a shorter one-year bandwidth because the pocket detection regression in equation (6) includes a time trend as a predictor. A priori we would expect such a trend to be very local and not last too long since this would imply an unreasonable buildup in return predictability. Using a shorter window to estimate the γ coefficients will, of course, produce larger estimation errors, but this is not so important here because of our use of a robust pocket identification scheme based on the *sign* of $\widehat{SED}_t$.

Intuitively, combining the time trend in (6) with our local kernel weighting scheme allows us to identify temporary, possibly short-lived, patterns in return predictability, lending our pocket definition a number of advantages. First, a pocket is triggered if the local return prediction model is deemed more accurate than the benchmark in the sense that it produces a lower expected squared forecast error. The definition therefore explicitly accounts for estimation uncertainty: even if the true current value of β_t in (2) is high, this may not produce a pocket if β_t cannot be estimated sufficiently accurately, for example, because returns have been very volatile (heteroskedasticity) or because β_t has not been high for long enough to allow our local estimation scheme in (3) to pick this up.

Second, our definition builds on the practice started by Welch and Goyal (2008) of studying how return predictability evolves over time through sums of squared forecast error differences. Differences in squared forecast errors are also the basis for formal comparisons of economic forecasting performance in the tests of Diebold and Mariano (1995) and Clark and West (2007). However, these tests do not include a local time trend. A novelty of our approach is that it allows us to identify temporary return predictability through a local estimate of the trend in the relative accuracy of the return forecasts.

Third, our pocket definition does not require us to compute standard errors for the estimates $\hat{\gamma}_{0,t}, \hat{\gamma}_{1,t}$ since we do not conduct formal hypothesis tests to identify pockets and hence do not have to decide on a significance level. This is particularly important for out-of-sample estimation since one-sided local kernel estimates of standard errors can be imprecise. Our definition also does not impose any minimum requirements on the length of the pockets. In practice, this means that short-lived pockets will sometimes be triggered ("false alarms"). One could easily impose that a pocket is triggered only after a certain number of periods for which $\widehat{SED}_t > 0$. Such a rule would come at the cost of delaying pocket identification, however, so we do not pursue this idea further.

On a final note, all of our estimates are computed recursively, out-of-sample, using only real-time information available prior to the period for which returns are being predicted. Specifically, we obtain the estimates $\hat{\gamma}_{0,t}$ and $\hat{\gamma}_{1,t}$ in equation (6) from a one-sided kernel using only information known at time t. We then define predictability pockets as periods (days) t for which $\widehat{SED}_t = \hat{\gamma}_{0,t} + \hat{\gamma}_{1,t}t > 0$. If, on day t, $\widehat{SED}_t > 0$, then we use the forecasts of returns in period $t+1$ from the local kernel regression, $\hat{r}_{t+1|t} = x_t'\hat{\beta}_t$, where $\hat{\beta}_t$ again uses only information known at time t.

C. Measures of Pocket Characteristics

To help understand pockets of predictability, we measure their characteristics in a variety of ways. First, we want to know how *many* contiguous pockets, N_p, our procedure detects along with how *long* the pockets last. To this end, let $\mathcal{I}_{jt} = 1$ for time-series observations inside the j^{th} pocket, while $\mathcal{I}_{jt} = 0$ outside this pocket for $t = 1, \ldots, T$. Denoting by t_{0j} and t_{1j} the start and end dates of the j^{th} pocket, the duration of pocket j, Dur_j, is given by

$$Dur_j = \sum_{\tau=1}^{T} \mathcal{I}_{j\tau} = t_{1j} - t_{0j} + 1, \quad j = 1, \ldots, N_p. \tag{7}$$

All else equal, long-lived pockets should be easier for investors to detect and exploit.

Pocket durations do not capture the total amount of predictability, which also depends on the magnitude of the local predictability. We quantify this through the local R^2 at time t, R_t^2:[16]

$$R_t^2 = 1 - \frac{\sum_{s=1}^{T} K_{hT}(s-t)(r_s - \hat{r}_{t|t-1})^2}{\sum_{s=1}^{T} K_{hT}(s-t)(r_s - \bar{r}_{s|s-1})^2}. \tag{8}$$

We measure the total amount of return predictability inside a pocket by means of the integral R^2 measure (IR) which, for the j^{th} pocket, is defined as

$$IR_j^2 = \sum_{\tau=t_{0j}}^{t_{1j}} R_\tau^2 = \sum_{\tau=1}^{T} \mathcal{I}_{j\tau} R_\tau^2. \tag{9}$$

This measure captures the area under a time-series plot of the local R_t^2 values in (8), summed across the pocket indicators. By combining the duration of a pocket with the magnitude of the predictability inside the pocket, the IR^2 measure provides insights into how much predictability is present as well as how feasible it is for investors to detect and exploit such predictability.

II. Empirical Results

This section introduces our data on stock returns and predictor variables, presents empirical evidence from applying the nonparametric approach to identifying local return predictability pockets, and tests whether this evidence is consistent with the conventional constant-coefficient return prediction model in (1).

[16] Note that this measure can be negative in certain periods because our time-varying coefficient model does not nest the prevailing mean model, which is the reference model in the denominator.

A. Data

Empirical studies on predictability of stock returns generally use monthly, quarterly, or annual returns data. Data observed at these frequencies can miss episodes with return predictability at times when the slope coefficients (β_t) change quickly, making it harder to accurately capture and time such episodes. As we are concerned here with local return predictability, which may be relatively short-lived, we therefore start using daily data on both stock returns and the predictor variables.

Following conventional practice in studies such as Welch and Goyal (2008), Dangl and Halling (2012), Johannes, Korteweg, and Polson (2014), and Pettenuzzo, Timmermann, and Valkanov (2014), our main empirical analysis considers univariate prediction models that include one time-varying predictor at a time, that is, $r_{t+1} = x_t\beta_t + \varepsilon_{t+1}$. The univariate approach is well-suited to our nonparametric analysis, which benefits from keeping the dimensionality of the set of predictors low. However, it raises issues related to omitted state variables, so we subsequently also discuss multivariate extensions.

In all of our return regressions, the dependent variable is the value-weighted CRSP U.S. stock market return minus the one-day return on a short T-bill rate. Turning to the predictors, we consider four variables that have been used in numerous studies on return predictability and are included in the list of predictors considered by Welch and Goyal (2008). First, we use the lagged dividend-price (dp) ratio, defined as dividends over the most recent 12-month period divided by the stock price at close of a given day t. This predictor has been used in studies such as Keim and Stambaugh (1986), Campbell (1987), Campbell and Shiller (1988), Fama and French (1988, 1989), and many others to predict stock returns. Second, we consider the yield on a three-month Treasury bill. Campbell (1987) and Ang and Bekaert (2007) use this as a predictor of stock returns. As our third predictor, we use the term spread, defined as the difference in yields on a 10-year Treasury bond and a three-month Treasury bill.[17] Finally, we consider a realized variance measure, defined as the realized variance over the previous 60 days. Again, this variable has been used as a predictor in a number of studies of stock returns.

The final sample date is December 31, 2016 for all series. The beginning of the data samples, however, varies across the four predictor variables. Specifically, it begins on November 4, 1926 for the dp ratio (23,786 observations), January 4, 1954 for the three-month T-bill rate (15,860 observations), January 2, 1962 for the term spread (13,846 observations), and January 15, 1927 for the realized variance (23,727 observations).

The daily predictor variables are highly persistent at the daily frequency, posing challenges for estimation and inference with daily data. We experimented with detrending the predictors by subtracting a six-month moving average which is a common procedure (see, e.g., Ang and Bekaert (2007)), but thus follow we found that the results do not change much and the simpler

[17] See Keim and Stambaugh (1986) and Welch and Goyal (2008) for studies using this predictor.

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	1.14	0.005	23,786
tbl	−0.007	−2.78	0.053	15,860
tsp	0.017	2.31	0.041	13,846
rvar	6.4×10^{-5}	0.54	4.3×10^{-4}	23,727
		Panel B: In-Pocket		
dp	0.084	2.55	0.18	3,483
tbl	−0.014	−3.29	0.37	3,506
tsp	0.073	3.95	1.47	1,810
rvar	4.8×10^{-5}	0.14	−0.02	4,841
		Panel C: Out-of-Pocket		
dp	0.012	0.44	−0.004	18,943
tbl	−0.003	−0.87	−0.002	10,994
tsp	0.006	0.75	−0.005	10,676
rvar	9.5×10^{-5}	0.66	0.004	17,526

approach of using raw data. We address the issue of how persistence affects inference through bootstrap simulations that incorporate the high persistence of our daily predictors along with other features of the daily data such as pronounced heteroskedasticity.

On economic grounds, we would expect return predictability to be very weak at the daily horizon. Table I confirms that this predictor holds. Panel A presents full-sample coefficient estimates obtained from the linear regression model in (1) along with t-statistics and R^2 values. Only the regressions that use the T-bill rate (t-statistic = -2.78) and the term spread (t-statistic = 2.31) generate statistically significant slope coefficients. As expected, the average predictability is extremely low at the daily frequency, with in-sample $\bar{R}^2$ values varying from 0.0004% for the realized variance measure to 0.053% (i.e., 0.00053) for the regression that uses the T-bill rate as a predictor.

Panels B and C of Table I report statistics from full-sample return regressions split into pockets versus nonpockets periods, identified in real time as we explain below. Large differences emerge across these two samples. Specifically,

Table II
Pocket Statistics

This table reports statistics on the duration of pockets (in days) and the integral R^2 of pockets for pockets estimated with both daily and monthly data. Coefficients are estimated using a one-sided kernel with a 2.5-year effective sample size, 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 in the preceding period.

	Daily				Monthly			
Statistics	dp	tbl	tsp	rvar	dp	tbl	tsp	rvar
Num pockets	18	12	7	16	15	15	10	18
Fraction of sample	0.16	0.24	0.14	0.22	0.17	0.21	0.12	0.19
Duration								
Min	16	57	95	25	21	21	42	21
Mean	193.5	292.2	258.6	302.6	243.6	207.2	149.1	226.3
Max	610	672	501	1,302	714	588	378	735
Integral R^2								
Min	−0.24	−0.24	0.28	−0.87	0.04	0.06	0.22	−0.29
Mean	1.51	3.70	2.92	2.77	1.79	2.21	1.59	1.48
Max	4.76	11.69	7.54	16.42	6.27	7.43	5.34	5.84

in-pocket slope coefficients are notably higher for three of the four predictor variables compared to out-of-pocket slope coefficients, the exception being the realized variance. Despite being based on a much shorter sample, the in-pocket regression coefficients are now highly statistically significant for the dp ratio (t-statistic = 2.55), T-bill rate (−3.29), and term spread (3.95). The regression $\overline{R}^2$ values are essentially zero outside pockets but far higher inside pockets for the dp ratio (0.18%), T-bill rate (0.37%), and term spread (1.47%). [18]

In-pocket $\overline{R}^2$ values are thus orders of magnitude higher than the "average" return predictability found in the full sample (Panel A). Although *most of the time* return predictability is extremely low at the daily frequency, some periods appear to exhibit substantially higher predictability. We next provide more details on how we identify those periods and where they are located.

B. Pockets of Local Return Predictability

Table II reports summary statistics for the number of pockets identified by our nonparametric procedure along with minimum, maximum, and mean values for the duration and IR^2 measure. The return regression based on the dp predictor identifies 18 pockets with durations that range from very short (16 trading days) to much longer (610 days), averaging 193 days, or nine months. Overall, pockets are identified for 15% of all days in the sample. Fewer pockets (12) are identified for the model that uses the T-bill rate predictor.

[18] The pockets are identified using ex ante available information, but the $\overline{R}^2$ values are estimated on the full sample.

However, the duration of these pockets is notably longer, ranging from 57 to 672 days and averaging 292 days, or 14 months. These long durations mean that pockets are identified for 24% of the days in the sample.

Seven pockets with a mean duration of 258 days (12 months) are identified for the term spread predictor, while for the realized variance predictor we find 16 pockets whose durations range from 25 to 1,302 days (five years), averaging 302 days (14 months).

We next consider the amount of return predictability computed for the individual pockets. The bottom rows in Table II show that the IR^2 for the dp predictor has a mean of 1.51 and ranges from -0.24 to 4.76. As a reference, note that a one-year (253-trading-day) period with an average daily R_t^2 of 0.004 (or 0.4%) produces an IR^2 of 1. For the T-bill rate predictor, the IR^2 has an average value of 3.70 and a maximum value of 11.69—both far higher than the values found for the dp predictor. The mean IR^2 is 2.92 for the term spread predictor, while the maximum equals 7.54, again higher than for the dp ratio but lower than for the T-bill rate. The very long pockets found for the realized variance predictor generate fairly high IR^2 values averaging 2.77 and peaking at a value of 16.42.[19]

A comparison of returns inside and outside the pockets (available in Internet Appendix Table IA.I) shows that mean returns are marginally higher inside periods identified as pockets. With the exception of the pockets identified by the dp ratio, the first-order autocorrelation of returns is also higher inside the pockets, ranging from 0.12 for realized variance to 0.22 for the T-bill rate. Conversely, returns are less volatile inside pockets and have a larger negative skew for two of the four predictors (dp and realized variance) but only half the kurtosis compared to returns outside the pockets.

We conclude from these results that return predictability varies significantly over time and that our nonparametric regression approach is able to detect local pockets of return predictability in real time. We next conduct more formal tests of these findings.

C. Tests for Spurious Pockets

Because we use a new approach for identifying local return predictability, it is worth exploring its statistical properties. For example, we are interested in knowing to what extent our approach spuriously identifies pockets of return predictability. Since we repeatedly compute local (overlapping) test statistics, we are bound to find evidence of some pockets even in the absence of genuine return predictability. The question is whether we find more pockets than we would expect by random chance, given a reasonable model for the daily return dynamics. Another question is whether shorter pockets or pockets with low IR^2 values are more likely to be spurious than longer ones.

[19] Across our four predictors, pairwise correlations between the local R^2 values range from -0.05 to 0.57.

C.1. Simulation Approach

We consider three different ways of simulating stock returns. To address the effect of using highly persistent predictor variables on pocket detection, all three approaches assume a constant-coefficient null for a predictor variable that follows an AR(1).

The first specification assumes homoskedastic errors and takes the form

$$\begin{aligned} r_{t+1} &= \mu_r + \gamma x_t + \varepsilon_{r,t+1},\ \ \varepsilon_{r,t+1} \sim (0, \sigma_r^2), \\ x_{t+1} &= \mu_x + \rho x_t + \varepsilon_{x,t+1},\ \ \varepsilon_{x,t+1} \sim (0, \sigma_x^2). \end{aligned} \tag{10}$$

We estimate μ_r, γ, μ_x, and ρ using ordinary least squares (OLS). To allow returns to follow a non-Gaussian distribution, we draw the zero-mean innovations $\hat{\varepsilon}_{r,t+1} = r_{t+1} - \hat{\mu}_r - \hat{\gamma} x_t$ and $\hat{\varepsilon}_{x,t+1} = x_{t+1} - \hat{\mu}_x - \hat{\rho} x_t$ by means of an independent and identically distributed bootstrap. Any cross-sectional dependencies are preserved by resampling the residuals in pairs with replacement from $\{\hat{\varepsilon}_{r,t+1}, \hat{\varepsilon}_{x,t+1}\}_{t=0}^{T-1}$. Bootstrap samples of residuals $\{\hat{\varepsilon}_{r,t+1}^b, \hat{\varepsilon}_{x,t+1}^b\}_{t=0}^{T}$ are then used to iteratively construct bootstrap samples for r and x using (10) with $x_0^b = 0$.

To account for the pronounced time-varying volatility in daily returns, we employ two additional variants: a stationary block bootstrap and an EGARCH(1,1) model with t-distributed shocks. The stationary block bootstrap selects the optimal block length using the method proposed by Politis and White (2004) applied to the residuals from the return regression, $\{\hat{\varepsilon}_{r,t+1}\}_{t=0}^{T-1}$ in (10). As in the independent and identically distributed case, blocks of residuals are resampled in pairs with replacement from $\{\hat{\varepsilon}_{r,t+1}, \hat{\varepsilon}_{x,t+1}\}_{t=0}^{T-1}$ to preserve cross-sectional correlation.

The EGARCH(1,1) model is given by

$$\begin{aligned} r_{t+1} &= \mu_r + \gamma x_t + \varepsilon_{r,t+1} \equiv \mu_r + \gamma x_t + \sqrt{h_{r,t}} u_{r,t+1},\ \ u_{r,t+1} \sim t(\nu_r) \\ \ln h_{r,t+1} &= \omega_r + \alpha_r(|u_{r,t+1}| - \mathbb{E}[|u_{r,t+1}|]) + \gamma_r u_{r,t} + \beta_r \ln h_{r,t} \\ x_{t+1} &= \mu_x + \rho x_t + \varepsilon_{x,t+1} \equiv \mu_x + \rho x_t + \sqrt{h_{x,t}} u_{x,t+1},\ \ u_{x,t+1} \sim t(\nu_x) \\ \ln h_{x,t+1} &= \omega_x + \alpha_x(|u_{x,t+1}| - \mathbb{E}[|u_{x,t+1}|]) + \gamma_x u_{x,t} + \beta_x \ln h_{x,t}. \end{aligned} \tag{11}$$

To simulate from this model, we first estimate the parameters and construct normalized residuals $\hat{u}_{r,t+1} = (r_{t+1} - \hat{\mu}_r - \hat{\gamma} x_t)/\sqrt{\hat{h}_{r,t}}$ and $\hat{u}_{x,t+1} = (x_{t+1} - \hat{\mu}_x - \hat{\rho} x_t)/\sqrt{\hat{h}_{x,t}}$. We then sample pairs $\{\hat{u}_{r,t+1}^b, \hat{u}_{x,t+1}^b\}$ independent and identically distributed with replacement from $\{\hat{u}_{r,t+1}, \hat{u}_{x,t+1}\}_{t=0}^{T-1}$. We construct bootstrap samples for r, x, h_r, and h_x using (11), setting $x_0^b = 0$, and setting h_r^b and h_x^b equal to their estimated means. For each of the three specifications, we generate 1,000 bootstrap samples $\{r_{t+1}^b, x_{t+1}^b\}_{t=0}^{T-1}$.

Our simulations follow the empirical analysis and define pockets as periods in which the prevailing mean model is expected to have a larger squared error than the local return predictions. For each bootstrap sample, we record the

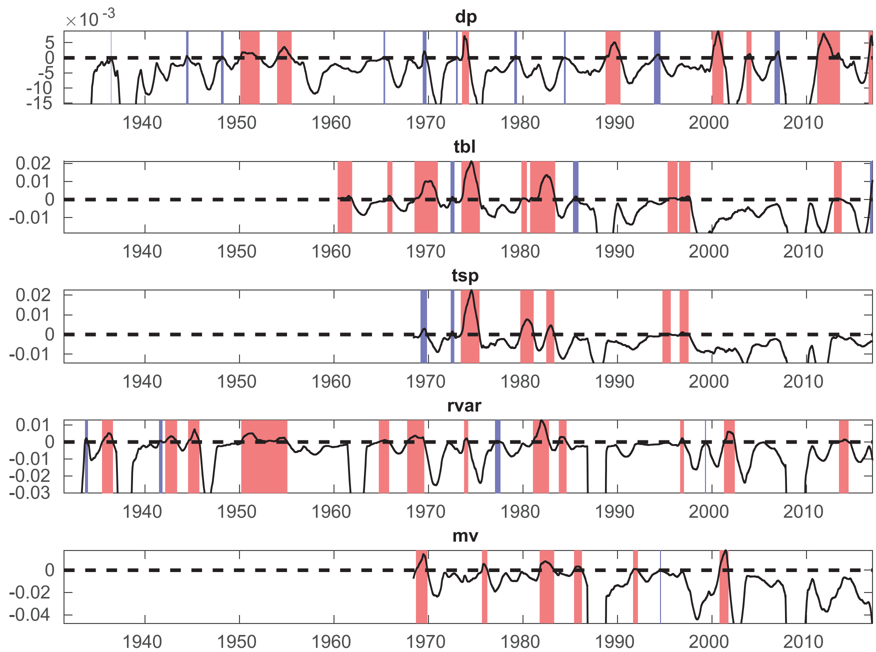

Figure 1. Local return predictability (daily benchmark specification). The first four panels plot one-sided nonparametric kernel estimates of the fitted squared forecast error differential $\widehat{SED}_t$ (estimated using a one-sided kernel with a one-year effective sample size) from a regression of 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. (Color figure can be viewed at wileyonlinelibrary.com)

distribution of IR^2 values from (9) and use this to compute p-values for overall sample statistics for the pocket distribution as well-as for the individual pockets.

C.2. Significance of Individual Pockets

To get a sense of the location and duration of the pockets, Figure 1 plots one-sided nonparametric kernel estimates of $\widehat{SED}_t$ against time for each of the four predictors. Shaded areas represent periods identified as pockets of predictability. We distinguish between spurious and nonspurious pockets by looking at each individual pocket's IR^2 value and computing the percentage of simulations with at least one pocket matching this value. This produces an odds ratio with small values indicating that it is difficult to match the total amount of predictability observed for the individual pockets.[20] We color pocket areas based on whether the pockets have less than (red) or more than (blue) a 5% chance of being randomly generated.[21]

First consider the predictability plot for the dp predictor (top panel). The longest pockets occur during the Korean War, prior to the 1990 recession and in the aftermath of the Great Recession. Conversely, there are relatively long spells without any (long-lasting) pockets prior to 1950 and again between the mid-1970s and late 1980s.[22] Eight of the 18 pockets identified using the dp ratio as our predictor are statistically significant at the 5% level. Conversely, all of the shorter pockets can be attributed to sampling error.

For the T-bill rate predictor (second panel), we locate three long-lived pockets, each lasting at least two years; around 1970, in the aftermath of the early 1970s oil price shocks, and around the Fed's Monetarist Experiment (1979 to 1981). Nine of the 12 pockets identified by the T-bill rate model are significant at the 5% level, leaving only three insignificant pockets.

Most of the pockets identified by the term spread predictor (third panel) occur during the mid-1970s and early 1980s, although we also locate two pockets in the mid-1990s. Five of the seven pockets are significant at the 5% level.

For the realized variance predictor (fourth panel), pocket incidence is fairly evenly spread across the sample, with the longest-lived pocket occurring during the Korean War, just as we find for the dp ratio. Long pockets also occur in the late 1960s and in the aftermath of the Monetarist Experiment. This model identifies 16 pockets, 12 of which are significant at the 5% level.

Pairwise time-series correlations between the four pocket indicators depicted in Figure 1 range from −0.02 to 0.59, indicating some overlap but also a fair amount of independent variation across pockets identified by different predictor variables.

We conclude from these simulations that the majority of return predictability pockets identified by our nonparametric return regressions cannot be explained by any of the return-generating models considered here. This is particularly true for the T-bill rate, term spread, and realized variance predictors. The simulations do not come close to matching the amount of predictability observed in the longer lived pockets. Conversely, the shortest pockets can be due to "chance" and are matched in many of our simulations. This point is particularly relevant for the dp regressions, which are more prone to pick up spurious, short-lived pockets. Reassuringly, since the model in equation (11) allows for highly persistent predictors and time-varying heteroskedasticity, these features of our data do not seem to give rise to the return predictability pockets that we observe.

[20] Returns simulated under the special case of no return predictability yield very similar results to those reported here, as can be seen in Table IA.II of the Internet Appendix.

[21] for the individual pockets identified by our procedure.

[22] Pockets do not necessarily coincide with high values of the estimated slope coefficient, $\hat{\beta}_t$. For example, a sudden spike in $\hat{\beta}_t$ preceded by small values of $\hat{\beta}_t$ will not produce a high value of SED_t and so will not trigger a pocket.

III. Statistical and Economic Performance of Return Forecasts

A large part of the literature on return predictability considers linear, constant-coefficient models based on a single predictor variable. Welch and Goyal (2008) find that such models fail to produce more accurate out-of-sample return forecasts than those from the prevailing mean model.

To address such shortcomings, one approach is to impose economically motivated constraints on the forecasts. Following Campbell and Thompson (2008), we consider three alternative ways of constructing out-of-sample excess return forecasts that, to varying degrees, incorporate economic restrictions: (i) unrestricted forecasts, $\widehat{r}_{t+1|t}$; (ii) nonnegative forecasts that replace negative forecasts with zero, $\max(0, \widehat{r}_{t+1|t})$; and (iii) return forecasts that, in addition to imposing the constraint in (ii), sets $\widehat{\beta}_t = 0$ if the estimated slope coefficient is inconsistent with our prior expectation of its sign (positive for the dp ratio, term spread, and realized variance, and negative for the T-bill rate).

A second approach is to incorporate multivariate information in the return prediction models. We describe alternative ways to do so further below.

A. Performance Measures

We first explain how we evaluate the performance of our local return forecasts using both statistical and economic performance measures. Following Welch and Goyal (2008), we compare our one-sided return forecasts to forecasts from a prevailing mean model, $\bar{r}_{t+1|t} = \frac{1}{t}\sum_{s=1}^{t} r_s$. To test the null of equal predictive accuracy, we use a Clark and West (2007, CW) test, with positive values indicating that the local, one-sided forecasting approach improves on the prevailing mean.

The CW test has three main advantages over conventional test procedures such as those in Diebold and Mariano (1995) and Clark and McCracken (2001). First, unlike the Diebold and Mariano (1995) test, it can be used to compare the accuracy of out-of-sample forecasts from nested prediction models as is frequently encountered in finance. Second, unlike the Clark and McCracken (2001) test, the CW statistic can be compared to critical values from the standard normal distribution and does not rely on simulated critical values. Third, the CW test accounts for the greater finite-sample effect that parameter estimation error can be expected to have on the bigger model (relative to the prevailing mean) and thus better summarizes the true predictive power of the underlying state variable(s) in the bigger model.

To assess the economic significance of our forecasting results, we adopt a strategy similar to that of Gómez-cram (2022) and construct a mean-variance optimized pocket portfolio invested in stocks and T-bills. Each forecasting model is used to compute real-time forecasts of expected excess returns, $E_t[r_{t+1}]$, and form a managed portfolio with excess returns

$$r^p_{t+1} = c \cdot E_t[r_{t+1}] \cdot r_{t+1}, \tag{12}$$

where r_{t+1} is the realized market excess return and the constant c is defined as

$$c \equiv \left[\frac{\mathrm{Var}(r_{t+1})}{\mathrm{Var}(E_t[r_{t+1}]\cdot r_{t+1})}\right]^{1/2}.$$

The weight placed on the market is given by $c \cdot E_t[r_{t+1}]$, which we restrict to be between zero and two, ruling out short sales and capping the leverage ratio at two.

Next, we use the excess returns on the managed pocket portfolio (12) to estimate the risk-adjusted return (α) from the regression

$$r^p_{t+1} = \alpha + \beta r_{t+1} + \epsilon_{t+1}, \quad \epsilon_{t+1} \sim (0, \sigma_\varepsilon).$$

In addition, as is common practice, we compute the Sharpe ratio for the managed portfolio.

B. Univariate Return Forecasts

Table III, Panel A, reports results of the CW tests. Across all days in the out-of-sample period (column (1)), the prevailing mean forecasts and the unrestricted local return forecasts are broadly equally accurate and the null of equal predictive accuracy can not be rejected. Thus, local return predictability could not have been exploited in real time to produce daily return forecasts that "on average" were more accurate than forecasts from a model that assumes a constant equity premium.

Inside the local pockets (column (2)), the CW test statistics are positive and highly statistically significant for all four predictors. Outside the pockets (column (3)), all four predictor models produce very poor forecasting performance with negative CW test statistics that are significant at the 10% level or above. Imposing the economic constraint that forecasts of excess returns cannot be negative (columns (4) to (6)) leads to improvements in all four one-sided kernel forecasts, which, for the T-bill rate, are now significantly more accurate at the 5% level even in the full sample, in addition to being significant at the 1% level for all four predictors inside the pockets. The constraint does not notably improve predictive accuracy out-of-pocket, however. Imposing additional sign restrictions on the slope coefficients (columns (7) to (9)) leads to similar performance as the model that only restricts the sign of the return forecasts.[23]

Table III, Panel B, reports results on economic performance. For the unrestricted univariate prediction models, the risk-adjusted return (α) is economically large and highly statistically significant for the dp ratio (1.69% per annum), T-bill rate (3.57%), term spread (3.14%), and realized variance (2.31%) predictors. The associated Sharpe ratios range from 0.47 for the dp ratio to

[23] Cumulative sum of squared error plots similar to those in Welch and Goyal (2008), described in Section IV of the Internet Appendix and displayed in Figure IA.1, show that the local kernel regressions outperform the prevailing mean model fairly steadily inside pockets while the opposite holds outside pockets.

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 purely backward-looking kernel with an effective sample size of 2.5 years to compute forecasts. "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 in the preceding period. $*$, $**$, and $***$ represent statistical significance at the 10%, 5%, and 1% levels from a hypothesis test of $\beta > 0$. † and †† represent statistical significance at the 10% and 5% levels from a hypothesis test of $\beta < 0$.

Panel A: Clark-West Statistics

	Unrestricted			+ Excess Return Forecasts			All Sign Restrictions		
Variables	Full Sample	In-Pocket	Out-of-Pocket	Full Sample	In-Pocket	Out-of-Pocket	Full Sample	In-Pocket	Out-of-Pocket
dp	−0.74	3.00***	−1.62†	0.40	3.79***	−1.94††	0.68	4.03***	−1.84††
tbl	0.68	3.28***	−1.58†	1.98**	4.75***	−1.33†	2.03**	4.69***	−1.21
tsp	0.15	3.04***	−1.52†	0.95	4.52***	−1.54†	0.79	4.18***	−0.21
rvar	−1.49†	2.88***	−1.77††	−0.79	3.93***	−1.07	−0.44	3.10***	−0.97
mv	−0.99	3.74***	−1.49†	−0.01	4.01***	−1.22	−0.01	4.01***	−1.22
pc	0.99	2.71***	−0.56	1.85**	4.69***	−0.52	1.85**	4.69***	−0.52
comb1	4.48***	4.58***	–	6.35***	6.55***	–	6.20***	6.39***	–
comb2	4.94***	5.05***	–	6.15***	6.33***	–	6.28***	6.48***	–
comb3	−1.03	2.11**	−2.02††	0.23	0.82	−1.23	0.66	2.55***	−1.03

(*Continued*)

Table III—*Continued*

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	1.69**	2.09	0.47	2.50***	2.89	0.54	2.95***	3.26	0.57
tbl	3.57***	4.35	0.79	6.47***	5.56	0.94	6.06***	5.36	0.92
tsp	3.14***	4.26	0.77	5.69***	4.95	0.85	5.03***	4.45	0.84
rvar	2.31***	3.62	0.68	2.88***	3.46	0.71	2.69***	3.03	0.54
mv	2.59***	3.37	0.64	4.79***	4.96	0.85	4.79***	4.96	0.85
pc	3.42***	3.96	0.73	5.86***	5.00	0.86	5.86***	5.00	0.86
comb1	6.38***	6.11	1.00	6.71***	6.70	1.05	6.69***	6.45	0.98
comb2	6.10***	5.66	0.87	8.53***	6.69	0.99	8.35***	6.51	0.95
comb3	0.76*	1.32	0.43	2.35**	1.88	0.46	2.73**	2.08	0.48
pm	−0.26*	−1.62	0.46	−0.26*	−1.62	0.46	−0.26*	−1.62	0.46

0.79 for the T-bill rate.[24] For comparison, the prevailing mean forecasts generate a negative α (-0.25) and a Sharpe ratio of 0.46.

Imposing the restriction that forecasts of mean excess returns should be non-negative leads to improvements in all three performance measures. Alphas now range from 2.51% (dp ratio) to 6.48% (T-bill rate), while Sharpe ratios increase more marginally. Imposing sign restrictions on the slope estimates yields broadly similar risk-adjusted return performance as imposing the sign restriction on the predicted excess return.

C. Incorporating Multivariate Information

We next consider ways in which multivariate information can be incorporated into the forecasts. Based on economic reasoning or more formal model selection methods (Pesaran and Timmermann (1995)), a first approach is to identify a small set of included predictors.[25] In our analysis, we consider a multivariate local kernel regression model (3) that simultaneously uses all four predictors—all of which can be economically motivated—to construct forecasts. The model is estimated on the subsample for which all four predictor variables are available, and we use a product kernel where each variable is assigned the same bandwidth.

Second, dimensionality reduction methods such as principal component analysis (PCA) can be applied directly on the set of predictors to form linear combinations that explain as much of the common variation in the predictors as possible (Pettenuzzo, Timmermann, and Valkanov (2014)). We apply PCA in real time to extract the first principal component (pc) from the four predictors.

Third, forecast combination methods can be used to form averages of the forecasts produced by small (univariate) models; see Rapach, Strauss, and Zhou (2010). We consider three different combination schemes. The first (comb1) sets an individual predictor's forecast to the local kernel forecast ($\widehat{r}^{i}_{t+1|t}$) inside pockets, reverting to the prevailing mean ($\bar{r}_{t+1|t}$) if no pocket is identified by the predictor, before computing an equal-weighted average,

$$\widehat{y}^{\text{comb1}}_{t+1|t} = \frac{1}{4}\sum_{i=1}^{4}\left(1\{\widehat{SED}_{it} \geq 0\}\widehat{r}^{i}_{t+1|t} + 1\{\widehat{SED}_{it} < 0\}\bar{r}_{t+1|t}\right), \tag{13}$$

where the indicator $1\{\widehat{SED}_{it} \geq 0\}$ equals 1 if the expected value of the local squared forecast error differential exceeds zero for predictor i, and 0 otherwise. For example, if the first univariate prediction model identifies a pocket while the remaining models do not, comb1 weights the forecast from the first model by 25% and the prevailing mean by 75%.

[24] The smaller α estimates for the forecasting model that uses the dp ratio are largely a result of this model bumping up against the (upper) constraints on the portfolio weights inside pockets.

[25] Including a large number of predictors ("kitchen sink") generally leads to poor out-of-sample forecasting performance due to estimation error.

The second combination (comb2) ignores forecasts from models that do not currently identify a pocket provided that at least one variable identifies a pocket,

$$\widehat{y}^{\text{comb2}}_{t+1|t} = \begin{cases} \frac{1}{n_t}\sum_{i=1}^{4} 1\{\widehat{SED}_{it} \geq 0\}\widehat{r}^{i}_{t+1|t} & \text{if } n_t \geq 0 \\ \overline{r}_{t+1|t} & \text{if } n_t = 0 \end{cases}, \tag{14}$$

where $n_t = \sum_{i=1}^{4} 1\{\widehat{SED}_{it} \geq 0\}$ is the number of predictors that identify a pocket at time t.

The third combination (comb3) makes no distinction between pocket and nonpocket periods, always using the simple equal-weighted average of all four univariate models:

$$\widehat{y}^{\text{comb3}}_{t+1|t} = \frac{1}{4}\sum_{i=1}^{4}\widehat{r}^{i}_{t+1|t}. \tag{15}$$

C.1. Empirical Results

Rows 5 to 9 in both panels of Table III report results on the multivariate prediction schemes. The fifth row in Panel A shows that the multivariate kernel approach delivers good out-of-sample forecasting performance inside pockets, with CW test statistics of 3.74 and 4.01 for the unrestricted and two sign-restricted forecasts, respectively. Predictive accuracy on out-of-pocket days is comparable to that of the univariate forecasting models.[26]

The table further shows that the PC approach delivers very good out-of-sample forecast performance inside the pockets, with CW test statistics of 2.71 and 4.69 for the unrestricted and two sign-restricted forecasts, respectively. Moreover, while the PC forecasts underperform outside the pockets, they do so to a smaller extent than the univariate forecasts and hence are more accurate in the full sample for 10 of the 12 pairwise comparisons against the univariate models.

Among the combination methods, comb1 and comb2 generate positive and highly significant CW test statistics both for the full sample and for in-pocket periods regardless of whether we combine forecasts from the unrestricted or restricted univariate models. In contrast, the simple equal-weighted average (comb3) performs worse than the underlying univariate forecasts. Since this approach does not distinguish between in-pocket and out-of-pocket periods, this result suggests that such conditioning is important to the benefits from forecast combination.

Examining the economic performance measures, we find that the PC approach performs very well, with alpha estimates and Sharpe ratios close to

[26] The six pockets identified by this approach that includes all four predictors (shown in the bottom panel in Figure 1) overlap to some extent with the pockets identified by the univariate kernel regressions.

those of the best-performing univariate models. The combination methods that condition the underlying forecasts on whether a pocket has been identified (comb1 and comb2) produce the best overall economic performance, while the equal-weighted combination (comb3) delivers poor economic performance.

D. Simulation Evidence

The empirical evidence summarized above demonstrates that the local kernel regressions can generate forecasts that are significantly more accurate than the benchmark inside ex ante identified pockets, though not outside these pockets or in the full sample. As an additional robustness check, we use our Monte Carlo simulation setup from Section II to explore whether similar improvements in predictive accuracy can be achieved by the statistical models introduced earlier.

Table IV summarizes results from simulating the three models and generating forecasts along the unrestricted and restricted schemes described earlier. The simulations are conducted under the null of constant return predictability ($\gamma \neq 0$), but all results are robust to assuming no return predictability ($\gamma = 0$) as shown in Table IA.II of the Internet Appendix.

The results are very clear and easily summarized. For all models, the simulations match both the full-sample and out-of-pocket CW test statistics. Conversely, we find no instance in which the simulations match the in-pocket CW statistic for any predictor or for any of the forecasting schemes. For the economic performance measures, the statistical models match the Sharpe ratio in some cases but fail to match the alphas or alpha t -statistics.[27]

Another possible concern that could affect our results is related to the Stambaugh (1999) bias, which affects the estimated slope coefficient of return prediction models in cases in which the predictor variable follows a highly persistent process and the correlation between innovations to the predictor variable and shocks to the return equation is large. Through a set of simulations described in Section III of the Internet Appendix and displayed in Table IA.IV, we show that this bias does not lead us to spuriously identify pockets, largely because of our use of an out-of-sample pocket identification approach.

E. Local Prevailing Mean Benchmark

So far, we follow studies such as Welch and Goyal (2008) and benchmark our return forecasts against a "global prevailing mean" that uses an expanding estimation window. However, a "local prevailing mean" model provides an interesting alternative benchmark as if this enables us to determine if our kernel regression forecasts are simply picking up local return momentum. To explore this point, let $\tilde{r}^{lpm}_{t|t-1} = \sum_{\tau=1}^{t-1} K(\tau) r(\tau)$ be the prediction from the local prevailing

[27] Alpha t-statistics are added because they have better sampling properties than alpha estimates.

Table IV

OOS Statistical Model Simulations (Daily)

This table reports Monte Carlo simulation results for the empirical one-sided kernel findings. We bootstrap the fitted residuals from a constant-coefficient predictive regression model for excess returns and an AR(1) model for the predictor: (i) an independent and identically distributed heteroskedastic bootstrap, (ii) a stationary block bootstrap where the optimal block length is chosen according to Politis and White (2004), and (iii) an EGARCH(1,1) with t-distributed shocks. All residuals are resampled jointly to preserve the cross-sectional correlation between the innovations to the predictor and excess returns. We generate 1,000 bootstrap samples of the same sample size as is available for each predictor in the data. 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. We report six statistics. The first three are Clark and West (2007) t-statistics relative to a prevailing mean benchmark in the full sample, in-pocket, and out-of-pocket. The second 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 with that estimated alpha, and the annualized Sharpe ratio of the portfolio. Column (2) presents the corresponding statistics from the data for reference.

	Unrestricted							+ Excess Return Forecasts							All Sign Restrictions						
		Independent and Identically Distributed		Block		EGARCH			Independent and Identically Distributed		Block		EGARCH			Independent and Identically Distributed		Block		EGARCH	
Statistic	Actual	Avg.	p-Value	Avg.	p-Value	Avg.	p-value	Actual	Avg.	p-Value	Avg.	p-Value	Avg.	p-Value	Actual	Avg.	p-Value	Avg.	p-Value	Avg.	p-Value
Panel A: dp																					
CW_{FS}	−0.74	−0.14	0.70	−0.22	0.69	−0.25	0.68	0.40	−0.04	0.31	−0.04	0.33	0.03	0.35	0.68	−0.03	0.23	−0.12	0.21	0.11	0.28
CW_{IP}	3.00	−0.16	0.00	0.06	0.00	−0.28	0.00	3.79	−0.10	0.00	0.30	0.00	0.04	0.00	4.03	−0.05	0.00	0.20	0.00	0.01	0.00
CS_{OOP}	−1.62	−0.07	0.94	−0.36	0.90	−0.13	0.92	−1.94	−0.00	0.98	−0.36	0.95	−0.01	0.98	−1.84	−0.01	0.96	−0.32	0.94	0.12	0.98
$\hat{\alpha}$	1.69	−0.23	0.05	0.35	0.11	0.14	0.04	2.50	−0.22	0.01	0.45	0.04	0.48	0.05	2.95	−0.27	0.00	0.24	0.01	0.40	0.03
$t_{\hat{\alpha}}$	2.09	−0.22	0.01	0.31	0.04	0.24	0.03	2.89	−0.19	0.00	0.35	0.00	0.37	0.01	3.26	−0.22	0.00	0.17	0.00	0.30	0.00
SR	0.47	0.44	0.38	0.44	0.41	0.49	0.58	0.54	0.43	0.16	0.44	0.18	0.49	0.36	0.57	0.43	0.07	0.44	0.11	0.49	0.24
Panel B: tbl																					
CW_{FS}	0.68	0.71	0.51	0.67	0.51	0.35	0.38	1.98	0.68	0.13	0.65	0.12	1.05	0.22	2.03	1.21	0.24	1.11	0.23	1.36	0.28
CW_{IP}	3.28	0.53	0.01	0.53	0.01	0.27	0.01	4.75	0.55	0.00	0.50	0.00	0.95	0.02	4.69	0.87	0.00	0.80	0.00	1.03	0.01
CS_{OOP}	−1.58	0.45	0.98	0.39	0.97	0.18	0.95	−1.33	0.39	0.95	0.40	0.94	0.54	0.96	−1.21	0.83	0.97	0.78	0.96	0.91	0.98
$\hat{\alpha}$	3.57	0.36	0.01	0.38	0.02	1.65	0.09	6.47	0.48	0.00	0.39	0.00	2.64	0.04	6.06	0.72	0.00	0.61	0.00	2.66	0.06
$t_{\hat{\alpha}}$	4.35	0.30	0.00	0.30	0.00	1.92	0.04	5.56	0.36	0.00	0.29	0.00	2.06	0.01	5.36	0.57	0.00	0.48	0.00	2.02	0.01
SR	0.79	0.55	0.18	0.53	0.15	0.77	0.45	0.94	0.53	0.07	0.53	0.06	0.77	0.26	0.92	0.53	0.07	0.53	0.06	0.76	0.28

(*Continued*)

Table IV—*Continued*

Panel C: tsp																					
CW_{FS}	0.15	0.61	0.67	0.90	0.75	0.25	0.54	0.95	0.71	0.42	0.97	0.51	0.94	0.49	0.79	0.75	0.47	0.93	0.56	0.94	0.52
CW_{IP}	3.04	0.46	0.01	0.81	0.02	0.08	0.00	4.52	0.55	0.00	0.88	0.00	0.76	0.00	4.18	0.51	0.00	0.85	0.00	0.84	0.00
CS_{OOP}	−1.52	0.38	0.97	0.44	0.97	0.25	0.96	−1.54	0.44	0.97	0.46	0.97	0.57	0.97	−0.21	0.54	0.75	0.54	0.76	0.61	0.77
$\hat{\alpha}$	3.14	0.34	0.02	0.82	0.04	0.84	0.01	5.69	0.46	0.00	0.94	0.00	1.32	0.00	5.03	0.57	0.00	1.07	0.00	1.44	0.01
$t_{\hat{\alpha}}$	4.26	0.27	0.00	0.68	0.00	1.11	0.00	4.95	0.33	0.00	0.70	0.00	1.11	0.00	4.45	0.44	0.00	0.82	0.00	1.23	0.00
SR	0.77	0.42	0.01	0.44	0.03	0.42	0.02	0.85	0.42	0.01	0.43	0.01	0.41	0.01	0.84	0.42	0.01	0.44	0.01	0.42	0.01
Panel D: rvar																					
CW_{FS}	−1.49	−0.03	0.92	−0.21	0.91	−0.37	0.86	−0.79	−0.03	0.77	0.13	0.81	−0.05	0.75	−0.44	0.20	0.73	0.27	0.75	0.19	0.72
CW_{IP}	2.88	−0.06	0.00	−0.01	0.00	−0.35	0.00	3.93	−0.06	0.00	0.28	0.00	−0.13	0.00	3.10	0.07	0.00	0.31	0.00	−0.09	0.00
CS_{OOP}	−1.77	−0.01	0.96	−0.28	0.94	−0.22	0.95	−1.07	−0.02	0.84	−0.12	0.82	0.02	0.85	−0.97	0.17	0.88	0.08	0.85	0.27	0.90
$\hat{\alpha}$	2.31	−0.14	0.02	0.61	0.06	0.34	0.01	2.88	−0.18	0.01	0.95	0.06	0.76	0.05	2.69	0.02	0.02	0.92	0.09	0.55	0.05
$t_{\hat{\alpha}}$	3.62	−0.15	0.00	0.63	0.00	0.51	0.00	3.46	−0.17	0.00	0.76	0.00	0.68	0.00	3.03	−0.01	0.00	0.74	0.01	0.51	0.01
SR	0.68	0.44	0.06	0.46	0.07	0.55	0.21	0.71	0.45	0.04	0.47	0.06	0.55	0.16	0.54	0.46	0.26	0.46	0.31	0.55	0.51

mean (lpm) model and replace equation (5) with

$$SED_t^{lpm} = (r_t - \tilde{r}_{t|t-1}^{lpm})^2 - (r_t - \hat{r}_{t|t-1})^2. \tag{16}$$

We can then apply our kernel regression in (6) to estimate a local trend in SED_t^{lpm} and identify pockets.

The results, reported in Tables IA.V and IA.VI of the Internet Appendix, show that our kernel regression forecasts based on time-varying predictors perform well relative to forecasts from the local prevailing mean model, producing strong economic performance and highly significant CW test statistics in-pocket and small but mostly statistically insignificant test statistics out-of-pocket.

In a second exercise, we revert to using return forecasts from the global prevailing mean model to detect pockets, but instead measure predictive accuracy against the local prevailing mean model so as to explore whether, inside the pockets identified by our predictors, their return forecasts are more accurate than forecasts from the local prevailing mean. This would not hold if our pockets were merely picking up local return momentum.

In results reported in Tables IA.VII and IA.VIII of the Internet Appendix, we continue to find that the time-varying predictors produce strong economic performance and highly significant CW test statistics inside the pockets, though not outside pockets.

As a final exercise, we again use forecasts from the global prevailing mean model to identify local pockets and benchmark our return forecasts. However, we now also consider the pockets identified by the local prevailing mean model by comparing the accuracy of its return forecasts to the return forecasts from the global prevailing mean. Next, to examine whether our time-varying predictors contain additional information that is not present in past returns, we consider the performance of our time-varying predictor models in those periods they identify as pockets that are *not* also identified as pockets by the local prevailing mean model. Pockets identified in this manner can thus be attributed to the additional information in the time-varying predictors that is not contained in the local prevailing mean forecast. In this analysis, only pockets that do not overlap with those identified by the local prevailing mean model are singled out. All other periods are classified as out-of-pocket.

Despite the reduction in the number of in-pocket observations associated with this scheme, for most of the predictors and the first two forecast combination schemes we continue to find significant improvements in predictive accuracy inside the pockets not identified as such by the local prevailing mean. Moreover, these gains in predictive accuracy strengthen notably from imposing economic constraints. We also find significant economic gains for all predictors with the exception of the dp ratio forecasts, whose alpha estimates remain positive, though not significant. Details of these results are reported in Table IA.IX of the Internet Appendix.

F. Choice of Bandwidth

Our pocket identification scheme relies on two windows, namely, the estimation window used by the local kernel regression to generate return forecasts and the performance monitoring window used to capture whether these forecasts are expected to produce a lower squared forecast error than the benchmark model. Our baseline results set these windows to 2.5 years—half of a five-year two-sided window—and one year, respectively. We set the estimation window slightly longer due to the well-known adverse effect of parameter estimation error in an inherently noisy environment, while the shorter monitoring window reflects our prior that local return predictability cannot last too long.

To explore the robustness of our results with regard to these choices, we let the estimation window vary between two and three years—corresponding to two-sided windows of four and six years—while the window used to track SED values varies between six and 18 months.[28]

Table V reports results from the robustness analysis with in-pocket and out-of-pocket results listed in the right and left columns, respectively. In both cases, the first column lists the results from the baseline scenario. The CW test statistics are highly robust to changes in the window lengths—slightly better for the short monitoring window and slightly worse for the longer one—as we continue to find strong evidence that both the univariate and multivariate approaches produce significantly more accurate in-pocket return forecasts than the prevailing mean model, but less accurate forecasts out-of-pocket.

A similar set of robustness tests applied to the economic performance measures yield the same conclusion, namely, that a broad range of choices of the two window sizes leads to highly significant alpha estimates for the managed portfolios that use our pocket methodology.

G. Controlling for Volatility, Momentum, and Transaction Costs

We next examine the effect of controlling for portfolios that manage volatility and momentum. Following Moreira and Muir (2017), we define a volatility factor as

$$f^{\sigma}_{t+1} \equiv \frac{c}{\hat{\sigma}^2_t(r)} r_{t+1},$$

where r_{t+1} is the buy-and-hold excess return on the market, $\hat{\sigma}^2_t(r)$ is a proxy for the portfolio's conditional variance, and c controls the average exposure of

[28] The vast majority of our results continue to hold for additional parameter configurations, including those in which the two bandwidth parameters are identical, for example, both 1.5 years or 2 years. However, the performance of the forecasting models based on the dp ratio and realized variance starts deteriorating when the bandwidth parameters used for pocket detection and parameter estimation are both short. This is what we would expect because these variables have noisier time series, which means that the combined effect of estimation error in the two regression steps starts to dominate for these predictors. We do not observe this effect for the other variables or for the combination approaches. We also find that our results are robust to longer windows such as a kernel estimation window of five years.

Table V

Robustness of Out-of-Sample Measures of Forecasting Performance

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. "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 in the preceding period. Consider a particular statistic of interest, β. $**$ and $***$ represent statistical significance at the 5%, and 1% levels from a hypothesis test of $\beta > 0$. †, ††, and † † † represent statistical significance at the 10%, 5%, and 1% levels from a hypothesis test of $\beta < 0$.

	Panel A: Unrestricted										
		In-Pocket (Real Time)					Out-of-Pocket (Real Time)				
Variable	Full Sample	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	−0.74	3.00***	2.92***	3.43***	3.59***	2.57***	−1.62†	−2.12††	−1.62†	−2.28††	−1.47†
tbl	0.68	3.28***	3.54***	3.32***	3.88***	3.88***	−1.58†	−1.00	−1.73††	−1.61†	−1.98††
tsp	0.15	3.04***	2.83***	3.29***	4.17***	2.77***	−1.52†	−1.08	−1.63†	−1.84††	−1.38†
rvar	−1.49†	2.88***	2.42***	2.88***	4.60***	1.93**	−1.77††	−1.82††	−1.83††	−2.05††	−1.77††
mv	−0.99	3.74***	2.65***	4.08***	4.58***	1.68**	−1.49†	−1.41†	−1.29†	−1.80††	−1.28
pc	0.99	2.71***	3.36***	2.96***	5.05***	3.55***	−0.56	−0.80	−0.56	−2.08††	−1.28
comb1	4.48***	4.58***	4.34***	4.74***	5.27***	4.64***	–	–	–	–	–
comb2	4.94***	5.05***	4.88***	5.25***	5.72***	5.13***	–	–	–	–	–
comb3	−1.03	2.11**	2.68***	2.05**	2.44***	2.31**	−2.02††	−2.34†††	−1.95††	−2.20††	−2.11††

(*Continued*)

Table V—*Continued*

Panel B: + Excess Return Forecasts

Variable	Full Sample	In-Pocket (Real Time) 2.5yCoef, 1ySED	2yCoef, 1ySED	3yCoef, 1ySED	2.5yCoef, 6mSED	2.5yCoef, 1.5ySED	Out-of-Pocket (Real Time) 2.5yCoef, 1ySED	2yCoef, 1ySED	3yCoef, 1ySED	2.5yCoef, 6mSED	2.5yCoef, 1.5ySED
dp	0.40	3.79***	3.95***	4.05***	4.30***	3.11***	−1.94††	−2.09††	−1.90††	−2.47†††	−1.65††
tbl	1.98**	4.75***	3.98***	4.06***	5.15***	4.46***	−1.33†	0.15	−1.23	−1.76††	−0.76
tsp	0.95	4.52***	4.29***	4.24***	6.44***	3.05***	−1.54†	−0.63	−1.28	−2.39†††	−0.48
rvar	−0.79	3.93***	4.50***	3.09***	5.17***	2.11**	−1.07	−1.27	−1.31†	−1.44†	−1.24
mv	−0.01	4.01***	4.13***	4.06***	4.59***	2.98***	−1.22	−1.17	−1.04	−1.52†	−0.92
pc	1.85**	4.69***	3.93***	4.34***	6.62***	4.53***	−0.52	0.46	−0.31	−1.78††	−0.43
comb1	6.35***	6.55***	6.06***	5.50***	7.76***	5.17***	–	–	–	–	–
comb2	6.15***	6.33***	5.86***	6.16***	7.41***	5.29***	–	–	–	–	–
comb3	0.23	0.82	0.64	2.71***	2.90***	3.33***	−1.23	−0.89	−1.43†	−1.69††	−1.55†

Panel C: All Sign Restrictions

Variable	Full Sample	In-Pocket (Real Time) 2.5yCoef, 1ySED	2yCoef, 1ySED	3yCoef, 1ySED	2.5yCoef, 6mSED	2.5yCoef, 1.5ySED	Out-of-Pocket (Real Time) 2.5yCoef, 1ySED	2yCoef, 1ySED	3yCoef, 1ySED	2.5yCoef, 6mSED	2.5yCoef, 1.5ySED
dp	0.68	4.03***	3.82***	4.28***	4.45***	3.15***	−1.84††	−1.94††	−1.75††	−2.46†††	−1.55†
tbl	2.03**	4.69***	4.22***	4.12***	4.98***	4.50***	−1.21	0.52	−1.34†	−1.45†	−0.56
tsp	0.79	4.18***	3.68***	3.99***	5.64***	3.11***	−0.21	1.24	−0.24	−0.19	0.38
rvar	−0.44	3.10***	2.05**	3.24***	4.79***	2.38***	−0.97	−1.05	−1.20	−1.41†	−1.11
mv	−0.01	4.01***	4.13***	4.06***	4.59***	2.98***	−1.22	−1.17	−1.04	−1.52†	−0.92
pc	1.85**	4.69***	3.93***	4.34***	6.62***	4.53***	−0.52	0.46	−0.31	−1.78††	−0.43
comb1	6.20***	6.39***	5.56***	5.64***	7.73***	5.29***	–	–	–	–	–
comb2	6.28***	6.48***	5.93***	6.44***	7.71***	5.43***	–	–	–	–	–
comb3	0.66	2.55***	1.25	2.39***	2.65***	2.78***	−1.03	0.25	−1.02	−1.23	−1.01

the strategy. As in Moreira and Muir (2017), we use the one-month realized variance estimate of excess returns, $\hat{\sigma}_t^2(r)$.

We also define a momentum factor as in Moskowitz, Ooi, and Pedersen (2012),

$$f_{t+1}^{\text{mom}} \equiv \text{sign}(r_{t-252,t})c\frac{r_{t+1}}{\hat{\sigma}_t(r)},$$

where $\text{sign}(r_{t-252,t})$ is the sign of the excess return on the market over the past year (1 if positive, 0 otherwise), and c again controls for the average exposure of the strategy.

Results from regressions extended to include these factors are presented in Table VI. Here, we estimate α as the intercept from regressions of portfolio excess returns, $r_{p,t+1}$, on r_{t+1}, f_{t+1}^{σ}, and f_{t+1}^{mom}. While controlling for these factors reduces performance slightly, all α estimates, except those associated with the equal-weighted combination, remain statistically and economically significant.

As an alternative approach to controlling for volatility, we conduct an additional version of the trading strategy in which we construct portfolio weights by dividing expected returns from each model by our measure of realized variance, rvar, which can be viewed as a proxy for the conditional return variance. If our time-varying mean forecasts are mainly identifying periods with high return volatility (indicating a constant risk-return trade-off), this weighting scheme should result in smoother allocations to the market portfolio. Conversely, if our local kernel return forecasts identify a time-varying risk-return trade-off, we should continue to find strong economic performance for our trading strategy. Compared to our benchmark results, we find that accounting for time-varying variance estimates strengthens our results with regard to estimated alphas and Sharpe ratios (Internet Appendix Table IA.X).

Transaction costs are another concern for the interpretation of our economic performance estimates. To address this issue, we examine the effect on return performance of proportional trading costs of 1 bp, 2 bps, and 10 bps. Due to modest portfolio turnover, we observe only small reductions in alpha estimates as a result of introducing transaction costs. Our alpha estimates for the local kernel prediction models remain strongly statistically and economically significant under all specifications, even for proportional trading costs as high as 10 bps. Results are especially strong for the trading strategies that impose economic restrictions on the forecasts (Table IA.XI).[29]

H. Monthly Return Predictions

Our analysis so far uses daily returns data to account for the possibility that some of the local pockets could be short-lived. However, the majority of studies in the return predictability literature uses monthly or longer data, so it

[29] Equivalently, the proportional trading costs at which the market-timing strategy breaks even are quite high at: 59 bps, 129 bps, 154 bps, and 107 bps for the dp ratio, T-bill rate, term spread, and realized variance predictors, respectively.

Table VI

Robustness of Out-of-Sample Economic Forecasting Performance to the Inclusion of Additional Factors

This table reports annualized estimated alphas in percentage points associated with returns on a daily portfolio that uses utilizes 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). We consider four specifications for estimating alpha: (i) the CAPM model, has excess returns on the market portfolio as the only factor; (ii) a two-factor model that uses the market factor and a volatility factor constructed as in Moreira and Muir (2017); (iii) a two-factor model that uses the market factor and a momentum factor on the market constructed using equation (5) of Moskowitz, Ooi, and Pedersen (2012); (iv) a three-factor model that includes the market, volatility, and momentum factors. Significance of the estimated alpha reported in parentheses below each alpha estimate is assessed using a t-statistic estimated using HAC standard errors. We use a purely backward-looking kernel to compute forecasts. "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 pocket and nonpocket periods and always uses the simple equal-weighted average of all four univariate models. 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. $*$, $**$, and $***$ represent statistical significance at the 10%, 5%, and 1% levels from a hypothesis test of $\beta > 0$. †, ††, and ††† represent statistical significance at the 10%, 5%, and 1% levels from a hypothesis test of $\beta < 0$.

	Unrestricted				+ Excess Return Forecasts				All Sign Restrictions			
Variable	Mkt	Mkt-Vol	Mkt-Mom	Mkt-Vol-Mom	Mkt	Mkt-Vol	Mkt-Mom	Mkt-Vol-Mom	Mkt	Mkt-Vol	Mkt-Mom	Mkt-Vol-Mom
dp	1.69^{**}	1.70^{**}	1.51^{**}	1.48^{**}	2.51^{***}	2.50^{***}	2.14^{***}	2.09^{***}	2.95^{***}	2.94^{***}	2.53^{***}	2.48^{***}
	(2.10)	(2.12)	(1.86)	(1.81)	(2.89)	(2.90)	(2.52)	(2.47)	(3.26)	(3.27)	(2.87)	(2.82)
tbl	3.57^{***}	3.51^{***}	3.23^{***}	3.23^{***}	6.48^{***}	6.21^{***}	5.53^{***}	5.59^{***}	6.07^{***}	5.86^{***}	5.28^{***}	5.32^{***}
	(4.35)	(4.28)	(3.91)	(3.90)	(5.56)	(5.65)	(5.02)	(5.15)	(5.37)	(5.36)	(4.82)	(4.87)
tsp	3.15^{***}	3.09^{***}	2.85^{***}	2.84^{***}	5.70^{***}	5.44^{***}	4.84^{***}	4.91^{***}	5.04^{***}	4.87^{***}	4.27^{***}	4.29^{***}
	(4.26)	(4.21)	(3.85)	(3.84)	(4.95)	(5.08)	(4.40)	(4.53)	(4.45)	(4.49)	(4.02)	(4.04)
rvar	2.31^{***}	2.30^{***}	1.68^{***}	1.64^{***}	2.89^{***}	2.87^{***}	2.21^{***}	2.16^{***}	2.69^{***}	2.67^{***}	2.08^{***}	2.04^{***}
	(3.63)	(3.60)	(2.79)	(2.78)	(3.47)	(3.44)	(2.85)	(2.84)	(3.03)	(3.05)	(2.50)	(2.45)

(*Continued*)

Table VI—*Continued*

	Unrestricted				+ Excess Return Forecasts				All Sign Restrictions			
Variable	Mkt	Mkt-Vol	Mkt-Mom	Mkt-Vol-Mom	Mkt	Mkt-Vol	Mkt-Mom	Mkt-Vol-Mom	Mkt	Mkt-Vol	Mkt-Mom	Mkt-Vol-Mom
mv	2.59***	2.59***	2.51***	2.49***	4.79***	4.80***	4.71***	4.68***	4.79***	4.80***	4.71***	4.68***
	(3.37)	(3.35)	(3.21)	(3.20)	(4.97)	(4.96)	(4.80)	(4.78)	(4.97)	(4.96)	(4.80)	(4.78)
pc	3.43***	3.34***	2.98***	2.99***	5.87***	5.59***	4.82***	4.89***	5.87***	5.59***	4.82***	4.89***
	(3.97)	(3.90)	(3.54)	(3.56)	(5.01)	(5.14)	(4.53)	(4.67)	(5.01)	(5.14)	(4.53)	(4.67)
comb1	6.38***	6.33***	5.71***	5.66***	6.72***	6.55***	5.83***	5.84***	6.69***	6.59***	5.90***	5.88***
	(6.11)	(6.06)	(5.68)	(5.65)	(6.71)	(6.76)	(6.20)	(6.21)	(6.46)	(6.44)	(5.96)	(5.91)
comb2	6.10***	6.08***	5.46***	5.41***	8.53***	8.33***	7.47***	7.50***	8.36***	8.27***	7.44***	7.40***
	(5.66)	(5.64)	(5.09)	(5.03)	(6.69)	(6.78)	(5.92)	(5.95)	(6.51)	(6.50)	(5.87)	(5.80)
comb3	0.76*	0.70	0.27	0.25	2.34**	2.23**	1.27	1.24	2.72**	2.69**	1.82*	1.75*
	(1.32)	(1.22)	(0.52)	(0.50)	(1.87)	(1.80)	(1.14)	(1.11)	(2.07)	(2.07)	(1.52)	(1.47)
pm	−0.25†	−0.29††	−0.39†††	−0.38†††	−0.25†	−0.29††	−0.39†††	−0.38†††	−0.25†	−0.29††	−0.39†††	−0.38†††
	−(1.58)	−(2.00)	−(2.68)	−(2.71)	−(1.58)	−(2.00)	−(2.68)	−(2.71)	−(1.58)	−(2.00)	−(2.68)	−(2.71)

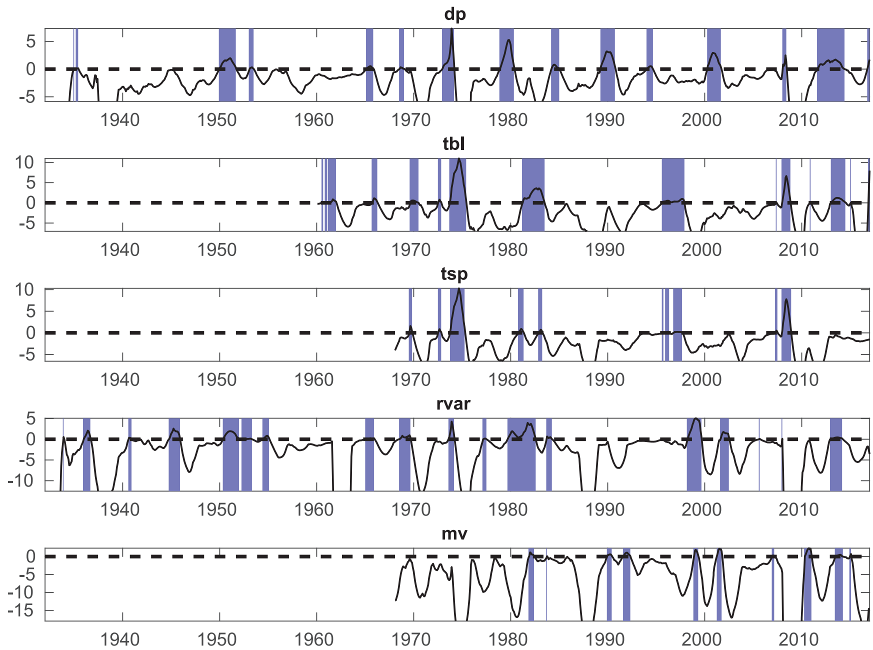

Figure 2. Local return predictability (monthly benchmark specification). The first four panels plot one-sided nonparametric kernel estimates of the fitted squared forecast error differential $\widehat{SED}_t$ (estimated using a one-sided kernel with a one-year effective sample size) from a regression 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. (Color figure can be viewed at wileyonlinelibrary.com)

is important to also conduct our analysis at this frequency to make our results more directly comparable to the literature.

Columns to the right in Table II report pocket statistics for the monthly data. The number of pockets and the proportion of the monthly sample identified as pockets are very similar to those identified for the daily returns data. Pocket durations (converted into days) tend to be a little shorter in the monthly data, and the average IR^2 statistics are substantially lower for three of the four predictor variables, the exception being the dp ratio.

Figure 2 displays the pockets identified at the monthly frequency using the same layout as in Figure 1. As in the daily data, we use a one-sided kernel with a bandwidth of 2.5 years. We find clear similarities between the pockets identified using the daily and monthly data. Indeed, the correlation between the daily pocket indicator (converted into a monthly value) and the monthly pocket indicator is 0.51 for the dp model and 0.65 for the T-bill rate model, which is very high considering we are using a crude scheme for converting monthly values of the pocket indicator to a daily series. For the term spread and realized variance regressions, the corresponding correlations are 0.51 and 0.55. Pockets identified with monthly data are thus very similar to those identified using daily data, which is reassuring from a robustness perspective.

Using a similar simulation setup as that described in Section II, we find that none of the pockets identified with monthly data is statistically significant. This is in marked contrast to the results obtained for the daily data and shows that a notable advantage of using higher frequency data is the associated increase in statistical power.[30]

Table VII reports evidence on the statistical accuracy and economic value of our monthly out-of-sample return forecasts. In the full sample, the statistical accuracy of the return forecasts generated by our local regression approach (Panel A) is indistinguishable from the prevailing mean forecasts. Inside pockets the story is different, however, as the CW test statistics are positive and highly statistically significant for all four predictors. The reason for these findings is again the poor predictive accuracy of the univariate forecasts outside the pockets. Imposing the sign constraint on excess return forecasts does not lead to notably better full-sample performance, as the CW test statistics tend to increase inside pockets but decrease outside pockets relative to the unrestricted forecasts.

As in the daily data, we find that the multivariate PC method performs similarly or a little better than the univariate forecasting methods, depending on whether the unrestricted or restricted forecasts are considered. The first two combination methods again perform very well, generating CW test statistics that are significant both in the full sample and inside pockets, with values that exceed those obtained from the underlying univariate forecasting models. Conversely, the equal-weighted combination (comb3) performs worse than the underlying univariate forecasts.

For the economic performance measures (Panel B), we continue to find strong performance of the univariate monthly forecasting models, with patterns that resemble those found in the daily data. Alphas are positive, economically large, and highly statistically significant, and thus improve notably when we impose either set of economic restrictions. Sharpe ratios start low for the unrestricted forecasts but improve by a sizeable amount once we impose the sign restrictions.

Monte Carlo simulations based on the three statistical models in Section II, but now applied to the monthly returns data, lead to similar conclusions as those reported in Table IV for the daily returns data. Specifically, all three models fail to match the observed in-pocket return predictability, although they easily match out-of-pocket results. The statistical models also fail to get close to matching the alpha estimates observed in the monthly data (Table IA.XII of the Internet Appendix).

[30] The bootstrap procedure has weak power because it only uses information on the IR^2 estimate for each individual pocket and does not pool data across pockets to get a longer evaluation sample.

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 purely backward-looking kernel with an effective sample size of 2.5 years to compute forecasts. "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. $*$, $**$, and $***$ represent statistical significance at the 10%, 5%, and 1% levels from a hypothesis test of $\beta > 0$. †, ††, and ††† represent statistical significance at the 10%, 5%, and 1% levels from a hypothesis test of $\beta < 0$.

Panel A: Clark-West Statistics

	Unrestricted			+ Excess Return Forecasts			All Sign Restrictions		
Variable	Full Sample	In-Pocket	Out-of-Pocket	Full Sample	In-Pocket	Out-of-Pocket	Full Sample	In-Pocket	Out-of-Pocket
dp	0.96	4.05***	−0.09	1.03	4.14***	−3.12†††	1.13	4.14***	−3.01†††
tbl	1.25	3.55***	−0.83	1.23	4.38***	−1.99††	2.40***	4.47***	−1.26
tsp	0.78	2.44***	−1.15	0.28	4.75***	−1.45†	0.46	4.93***	−0.20
rvar	0.64	3.28***	0.00	0.40	3.18***	−2.73†††	1.04	3.58***	−2.10††
mv	1.76**	3.18***	1.28	1.65**	3.70***	−0.69	1.65**	3.70***	−0.69
pc	1.22	3.23***	−1.23	1.04	4.64***	−1.25	1.04	4.64***	−1.25
comb1	4.14***	4.74***	–	4.73***	5.24***	–	4.82***	5.32***	–
comb2	4.48***	5.20***	–	4.81***	5.27***	–	5.48***	6.16***	–
comb3	1.01	2.20**	−2.15††	1.10	2.68***	−2.06††	1.79**	2.18**	−0.83

Panel B: Economic Significance

	Unrestricted			+ Excess Return Forecasts			All Sign Restrictions		
Variable	$\hat{\alpha}$	$t_{\hat{\alpha}}$	Sharpe Ratio	$\hat{\alpha}$	$t_{\hat{\alpha}}$	Sharpe Ratio	$\hat{\alpha}$	$t_{\hat{\alpha}}$	Sharpe Ratio
dp	2.37***	2.53	0.55	4.13***	3.26	0.69	4.13***	3.26	0.69
tbl	3.77***	3.36	0.76	6.08***	4.45	0.86	6.22***	4.43	0.87

(*Continued*)

Table VII—*Continued*

Panel B: Economic Significance									
	Unrestricted			+ Excess Return Forecasts			All Sign Restrictions		
Variable	$\hat{\alpha}$	$t_{\hat{\alpha}}$	Sharpe Ratio	$\hat{\alpha}$	$t_{\hat{\alpha}}$	Sharpe Ratio	$\hat{\alpha}$	$t_{\hat{\alpha}}$	Sharpe Ratio
tsp	2.22***	2.66	0.65	5.14***	3.54	0.76	4.11***	3.24	0.71
rvar	2.21***	2.73	0.55	3.53***	3.27	0.65	3.73***	3.68	0.66
mv	1.29**	2.00	0.47	3.53***	3.30	0.60	3.53***	3.30	0.60
pc	3.76***	3.42	0.77	5.39***	3.84	0.79	5.39***	3.84	0.79
comb1	7.00***	5.20	1.03	6.68***	5.13	1.01	6.61***	5.58	1.09
comb2	6.58***	4.98	0.94	8.02***	6.13	1.01	8.84***	6.18	1.03
comb3	1.38*	1.49	0.46	2.19*	1.32	0.44	4.48***	3.21	0.62
pm	−0.36**	−1.88	0.49	−0.36**	−1.88	0.49	−0.36**	−1.88	0.49

Our combination that averages forecasts from models classified as being in a pocket (comb2) achieves an out-of-sample monthly R^2 of 15.0%. Rapach, Strauss, and Zhou (2010) report quarterly recession R^2 values of 4% to 8% using a forecast combination with 15 underlying predictors.[31]

We conclude from these findings that our local kernel regression approach could also have been also at a frequency similar to that used in the literature (monthly) to identify, in real time, local pockets with a high degree of return predictability.

I. Lumpiness in Return Predictability

The lumpiness that triggers pockets in our empirical exercise comes from our binary decision rule, which classifies pockets according to whether $\widehat{SED}_t > 0$, thus producing a pocket indicator akin to the binary NBER recession indicator used to track fluctuations in economic activity. To explore whether, more broadly, our return forecasts are more accurate when $\widehat{SED}_t$ is large and positive compared to when it is small or negative, we also perform a simple exercise in which we compute the accuracy of our return forecasts, which we sort into four quartiles representing the days with the lowest 25%, second-lowest 25%, second-highest 25%, and highest 25% of days ranked by $\widehat{SED}_t$. For each quartile, we then compute the CW test statistic.

We find that the accuracy of our return forecasts increases monotonically across the $\widehat{SED}_t$-sorted quartiles for three of the four predictor variables, displaying only slight nonmonotonicity for the rvar predictor.[32] Similar patterns emerge with more bins, showing that our pocket identification scheme generates a strong signal about local return predictability.

[31] Note that the two R^2 values are not directly comparable because we choose pockets based on patterns in local return predictability while recession R^2 values are instead based on an (exogenous) economic indicator.

[32] Results are shown in Figure IA.2 of the Internet Appendix.

J. Pockets in Size and Value Factor Returns

The sticky expectations model discussed further in Section V provides a mechanism for generating local predictability pockets not only in aggregate market returns, but also in factor dynamics. We therefore next explore whether local predictability pockets can be identified in the returns on the SMB and HML Fama-French factors. These data, obtained from Ken French's website, are available over the same sample as the excess return and dividend-price ratio data, going back to November 4, 1926.

For the SMB series, the fraction of the sample spent inside pockets ranges between 0.24 (term spread model) and 0.35 (dp ratio). These values are somewhat higher than those found for the market return series, as is reflected in a longer mean duration ranging from 255 days (term spread) to 332 days (T-bill rate). The IR^2 values are also higher for this spread portfolio compared to the market, with mean values ranging from 3.78 (term spread) to 6.31 (realized variance).

Similar findings obtain for the value-growth return series (HML). Pockets take up a fraction of the sample for this series that ranges from 0.25 (realized variance) to 0.34 (term spread), with mean durations ranging from 232 days (realized variance) to 384 days (term spread). Average R^2 values remain high, although a little below those found for the SMB series, ranging from 3.13 for the realized variance predictor to 5.13 for the term spread.[33]

Table VIII reports performance results for local kernel regressions fitted to returns on the SMB and HML portfolios. We focus on the unrestricted model forecasts since it is not clear how to impose sign restrictions on expected return differentials or the slopes of the predictor variables.

First consider the statistical performance measures (Panel A). For both the SMB and the HML return series, and across all four predictors, inside pockets the local kernel regressions generate more accurate out-of-sample return forecasts than the prevailing mean, resulting in highly significant CW statistics. Conversely, the local kernel forecasts tend to be less accurate than the prevailing mean out-of-pocket. In contrast to the results for the market portfolio, the in-pocket results dominate for the full sample, so we now find significantly better full-sample performance for three of four predictors—the exception being the realized variance.

All multivariate approaches—multivariate kernel, PCA, and combinations—generate forecasts that are significantly more accurate than the benchmark both in-pocket and in the full sample, though not during out-of-pocket periods. The first two combinations continue to be better than the simple equal-weighted combination (comb3).

For the economic performance measures (Panel B), the alpha estimates are highly statistically significant, ranging from 2.57% per annum for the term spread predictor to 3.43% for the realized variance predictor applied to the

[33] Figures IA.3 and IA.4 of the Internet Appendix show the pockets identified for the HML and SMB portfolios. Detailed pocket statistics are provided in Table IA.XIII of the Internet Appendix.

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 purely backward-looking kernel to compute forecasts. "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. $*$, $**$, and $***$ represent statistical significance at the 10%, 5%, and 1% levels from a hypothesis test of $\beta > 0$. † represents statistical significance at the 10% level from a hypothesis test of $\beta < 0$.

Panel A: Clark-West Statistics

	SMB			HML		
Variable	Full Sample	In-Pocket	Out-of-Pocket	Full Sample	In-Pocket	Out-of-Pocket
dp	2.03**	4.75***	0.55	1.41*	4.49***	−0.91
tbl	1.77**	4.83***	−0.15	1.90**	3.88***	−0.21
tsp	2.83***	3.58***	0.42	1.93**	3.83***	−1.49†
rvar	0.89	4.98***	0.63	−0.10	4.22***	−1.33†
mv	2.97***	5.94***	1.90**	2.05**	5.18***	−0.09
pc	3.48***	3.60***	1.42*	2.01**	3.35***	−0.74
comb1	5.75***	6.09***	–	5.46***	5.63***	–
comb2	4.67***	4.92***	–	5.33***	5.47***	–
comb3	2.02**	3.52***	0.44	1.49*	4.32***	−1.56†

Panel B: Economic Significance

	SMB			HML		
Variable	$\hat{\alpha}$	$t_{\hat{\alpha}}$	Sharpe Ratio	$\hat{\alpha}$	$t_{\hat{\alpha}}$	Sharpe Ratio
dp	2.95***	3.66	0.81	3.29***	5.13	1.18
tbl	3.35***	4.44	0.90	2.89***	4.40	1.07
tsp	2.57***	4.26	0.96	2.33***	3.74	0.80
rvar	3.43***	4.80	1.00	2.97***	4.28	1.11
mv	3.27***	4.23	0.85	3.74***	4.96	0.99
pc	2.43***	4.07	0.86	1.94***	3.19	0.74
comb1	5.15***	6.22	1.34	4.46***	5.79	1.18
comb2	4.07***	5.56	1.19	3.87***	5.50	1.07
comb3	1.18**	2.12	0.45	1.40**	1.87	0.62
pm	−0.38	−0.75	0.17	−0.17†	−1.55	0.62

SMB portfolio and from 2.33% to 3.29% for the term spread and dp predictors applied to the HML portfolio. The first two forecast combinations boost this performance by anywhere from 0.6% to 1.8% per annum.[34]

IV. Pockets and Asset Pricing Models

Having presented our empirical evidence on the existence of local return predictability, we next use our new measures of pocket characteristics as diagnostics for exploring whether a range of asset pricing models can generate local return predictability patterns similar to those found empirically.

A. Overview of Models Selected

Although it is impossible to explore all possible frameworks, we simulate from four workhorse rational expectations asset pricing models that are representative of the dynamics of returns and state variables implied by models with time-varying risk premia. In all cases, we select versions of these models that are cast in continuous time (making it easy to simulate daily data) and employ global solution algorithms that capture potential nonlinearities inherent in the models. Despite matching a number of common features from the data, the models are quite distinct along a number of dimensions that are representative of different structural explanations of the equity premium puzzle proposed in the literature. We consider the following models:

(i) A continuous-time version of the long-run risk model of Bansal and Yaron (2004), as calibrated by Chen et al. (2009). This model features investors with Epstein-Zin preferences and two state variables, namely, the drift in the consumption growth process and a stochastic volatility process that affects the mean consumption growth process. [35] In the model, time variation in the risk premium is driven almost exclusively by stochastic volatility.

(ii) The habit formation model of Campbell and Cochrane (1999), which features a single state variable capturing investors' "habit level" of consumption that generates time-variation in the effective risk aversion.[36]

[34] As in our main analysis of the market portfolio, we impose limits on portfolio weights between zero and two.

[35] Note that we emphasize a calibration that is more similar to the original Bansal and Yaron (2004) paper. Bansal, Kiku, and Yaron (2012) introduce an alternative calibration in which a larger fraction of variation is explained by fluctuations in a more persistent stochastic volatility variable relative to fluctuations in the persistent expected growth component. Although we have not formally conducted simulation exercises for this specification, our existing results suggest that adding a more persistent risk-premium shifter would strengthen Stambaugh (1999) biases and likely hurt performance relative to the baseline presented here.

Following a sequence of bad shocks, risk aversion and risk premia rise, lowering asset prices.

(iii) The heterogeneous agents model of Gârleanu and Panageas (2015), which features two types of agents with different levels of risk aversion who optimally share claims on the aggregate endowment. The model features a single state variable that captures the share of wealth owned by one of the two types of agents. As the share of wealth owned by risk tolerant agents decreases, risk premia rise, a force that generates excess volatility of asset prices.

(iv) The rare disaster model of Wachter (2013), which features investors with Epstein-Zin preferences and a single state variable capturing the time-varying Poisson arrival rate of a rare disaster, that is, a permanent, large drop in the aggregate endowment.

In Section V of the Internet Appendix, we provide details on how we simulate from these models, while Section VI of the Internet Appendix and Table IA.XIV report a variety of unconditional moment statistics. In addition, Section V below presents (and draws similar conclusions from) a reduced-form present value model in the spirit of Campbell, Lo, and MacKinlay (1997) and van Binsbergen and Koijen (2010).

In each of these models, it is straightforward to construct proxies for three of our state variables, namely, the dp ratio, the risk-free rate, and realized volatility of returns. As such, we can draw initial levels of the state variables, we then simulate daily samples with the same length as our estimation sample. With these simulated times series, we compute our out-of-sample measures of forecasting performance and several associated test statistics. Consistent with the convention of the rare disaster literature, in making comparisons with post-war U.S. data, we also conduct a set of simulations in which we restrict attention to sample paths where no disaster occurs.

B. Pitfalls of Identifying Short-Horizon Predictability

Given that all quantitative asset pricing models seek to rationalize several stylized facts from the data, we first develop some intuition for why precisely these features suggest ex ante that it should be challenging for the canonical asset pricing models discussed above to generate time-varying short-run return predictability consistent with what we find empirically.

Specifically, asset pricing models usually seek to match a fairly similar set of moments observed in the data: (i) dp ratios are stationary but quite persistent and volatile, (ii) discount rates explain a nontrivial fraction of variation in price-dividend ratios, (iii) risk premia, rather than risk-free rates, explain more of the variation in discount rates, and (iv) state variables cap-

[36] We use the continuous-time version of the calibration from Wachter (2005), which also allows habit to affect the risk-free interest rate.

turing both discount rates and risk premia are usually quite persistent. The combination of these features implies that returns are predictable, especially at longer horizons, by the price-dividend ratio, with modest R^2 values over medium-term horizons, consistent with evidence from predictive regressions.

To understand why the canonical asset pricing models with forward-looking rational expectations struggle to generate detectable local return predictability pockets, suppose there is a spike in risk premia. This could happen because a persistent state variable shifts and/or because the sensitivity of the risk premium to the state variable changes in a model with time-varying parameters. Rational, forward-looking agents will then reduce their valuation of the asset, generating an immediate offsetting effect on realized returns. The resulting pattern with a large negative shock to realized returns followed by a sequence of slightly elevated returns is exactly what makes it difficult to detect local return predictability in such models. Further, the more risk premia move, the more volatile realized returns are likely to be, increasing estimation errors in local predictive regression coefficients. A final concern is the Stambaugh bias because shocks to risk premia may be correlated with innovations to the key regressors—an effect that can be particularly strong at the higher (daily) frequency. These effects make local return predictability at high frequencies extremely difficult to detect. Only at longer horizons, as the shock to the persistent risk premium component has had time to build up, do we get more power to detect return predictability.

C. Simulation Results

Building on these observations, Table IX repeats simulation results for the four asset pricing models using the unrestricted return predictions. For each performance measure listed in the rows, the columns show the mean, standard error, and p-value, the latter computed from the proportion of simulations able to match the sample statistic, which for convenience we present in the left-most column. Panels A, B, and C report results for the three predictors generated as part of the asset pricing models, namely, the dp ratio, the risk-free rate, and the realized variance, respectively.

First consider the statistical performance as captured by the CW statistic. Across all three predictors, all asset pricing models can match the full-sample and out-of-pocket accuracy of the local kernel forecasts measured relative to the prevailing mean. None of the asset pricing models gets close to matching the in-pocket accuracy of the kernel regression forecasts, however, regardless of which predictor is used.

Turning to the economic performance measures, the Campbell and Cochrane (1999) and Gârleanu and Panageas (2015) models struggle to match the alphas found in the data. The Bansal and Yaron (2004) and Wachter (2013) models are better able to match alphas for the predictive return regressions that use the dp ratio, but not so much for those that use the risk-free rate or the realized

Table IX

OOS Asset Pricing Model Simulations (Unrestricted)

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 six statistics. The first three are Clark and West (2007) t-statistics relative to a prevailing mean benchmark in the full sample, in-pocket, and out-of-pocket. The second 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.	*SE*	p-Value	Avg.	*SE*	p-Value	Avg.	*SE*	p-Value	Avg.	*SE*	p-Value	Avg.	*SE*	p-Value
								Panel A: dp								
CW_{FS}	−0.74	−0.10	0.98	0.74	0.09	0.98	0.81	−0.03	0.96	0.77	0.31	1.02	0.84	0.65	1.05	0.90
CW_{IP}	3.00	−0.07	1.02	0.00	−0.08	0.99	0.00	−0.09	0.97	0.00	0.15	1.00	0.00	0.34	1.00	0.00
CS_{OOP}	−1.62	−0.10	1.01	0.93	0.15	0.99	0.96	−0.00	0.97	0.94	0.25	1.03	0.96	0.54	1.04	0.97
$\hat{\alpha}$	1.69	−0.38	1.67	0.11	0.05	0.95	0.04	−0.02	0.86	0.02	0.21	1.82	0.19	0.31	1.34	0.15
$t_{\hat{\alpha}}$	2.09	−0.23	1.00	0.01	0.05	1.01	0.02	−0.02	0.99	0.01	0.19	1.08	0.04	0.22	1.02	0.03
SR	0.47	0.44	0.13	0.41	0.47	0.07	0.55	0.33	0.11	0.09	0.46	0.12	0.47	0.58	0.10	0.90

(Continued)

Table IX—*Continued*

Panel B: Risk-Free																
CW_{FS}	0.68	−0.10	0.99	0.22	0.10	0.98	0.28	−0.06	0.96	0.22	0.31	1.02	0.36	0.65	1.06	0.50
CW_{IP}	3.28	−0.09	1.03	0.00	−0.06	0.98	0.00	−0.09	0.97	0.00	0.15	1.00	0.00	0.35	1.00	0.00
CS_{OOP}	−1.58	−0.09	1.01	0.93	0.14	0.99	0.96	−0.04	0.96	0.94	0.25	1.03	0.96	0.54	1.04	0.97
$\hat{\alpha}$	3.57	−0.37	1.67	0.01	0.05	0.95	0.00	−0.03	0.83	0.00	0.21	1.82	0.03	0.31	1.34	0.01
$t_{\hat{\alpha}}$	4.35	−0.22	1.00	0.00	0.05	1.01	0.00	−0.03	0.97	0.00	0.19	1.08	0.00	0.22	1.02	0.00
SR	0.79	0.44	0.13	0.01	0.47	0.07	0.00	0.33	0.11	0.00	0.46	0.12	0.00	0.58	0.10	0.03
Panel C: rvar																
CW_{FS}	−1.49	−0.08	1.03	0.91	−0.26	0.99	0.89	−0.13	0.97	0.92	0.05	1.00	0.95	0.22	1.08	0.95
CW_{IP}	2.88	−0.08	1.02	0.00	−0.24	1.01	0.00	−0.11	0.96	0.00	−0.01	0.97	0.00	0.08	1.00	0.00
CS_{OOP}	−1.77	−0.06	1.01	0.95	−0.15	0.99	0.95	−0.10	0.98	0.95	0.10	1.01	0.97	0.22	1.02	0.98
$\hat{\alpha}$	2.31	−0.38	1.70	0.06	−0.06	0.96	0.01	−0.04	0.85	0.01	0.05	1.29	0.04	0.17	1.33	0.06
$t_{\hat{\alpha}}$	3.62	−0.24	1.02	0.00	−0.07	0.99	0.00	−0.05	0.98	0.00	0.09	1.05	0.00	0.10	1.01	0.00
SR	0.68	0.44	0.13	0.03	0.47	0.07	0.01	0.33	0.11	0.00	0.45	0.12	0.03	0.58	0.10	0.12

variance predictors. None of the asset pricing models is able to match the alpha t-statistic in the empirical data, and they only match the Sharpe ratio for the models that use the dp predictor.[37]

Taking stock, these results suggest that the presence of local return predictability pockets poses a challenge in the sense that such patterns cannot be generated by a range of dynamic asset pricing models spanning a wide spectrum of modeling assumptions. One might suspect that this is due to the omission, by such models, of complicating factors such as time-varying heteroskedasticity or highly persistent predictors whose innovations are correlated with shocks to the return process. However, this is unlikely to be the case here since our earlier simulations of three statistical models incorporates such features and can not produce return patterns that match the local return predictability pockets that we find in the data.

It is important to emphasize that we do *not* preclude the possibility that asset pricing models with rational expectations can generate pockets of return predictability. For instance, one could introduce a moderately persistent variable, s_t, that affects risk premia and risk-free rates by offsetting amounts, thus preserving a signal that is potentially useful and avoids the problem of offsetting noise. Specifically, a predictor such as the risk-free interest rate could be a linear combination of low- and high-frequency components, in which case the projection of returns onto the predictor may be time-varying. Constructing such a model is outside the scope of our current paper, however, and is therefore left for future research.[38]

V. Sticky Expectations and Pockets of Predictability

In the previous section, we argue that our empirical findings of local return predictability pockets pose challenges to a number of workhorse asset pricing models with time-varying risk premia. In this section, motivated by a rapidly growing literature at the intersection of macroeconomics and finance, we propose a model featuring both sluggish adjustment of beliefs, in the spirit of

[37] Tables IA.XVI and IA.XVII of the Internet Appendix show that similar results hold for the return predictions that impose constraints on the sign of the excess return forecasts or restrict the signs of the slope estimates.

[38] Time-variation in intermediaries' net worth is another possible source of local return predictability since it could explain why local return predictability is not arbitraged away in states with only limited access to arbitrage capital. To explore this possibility further, we conducted simulations from the asset pricing model proposed by Di Tella (2017), which emphasizes intermediaries' balance sheets in a model of optimal risk sharing between intermediaries and households and provides a mechanism for generating time-varying risk premia. We found that this model yields results similar to those from the other asset pricing models and does not generate pockets consistent with what we see in the data. Although the key state variables in the model governing risk premia are somewhat less persistent in this framework relative to some of the other models we consider, ultimately we find similar results to Table IX. The key variables fluctuate at business-cycle frequencies, making it difficult to detect pockets of predictability in time to exploit them meaningfully out-of-sample via our local kernel approach. The results are shown in Table IA.XV of the Internet Appendix.

"sticky information" models (Mankiw and Reis (2002), Woodford (2003), Sims (2003), Coibion and Gorodnichenko (2015)) , and departures from market efficiency reflecting tendencies of certain types of information to be incorporated into asset prices slowly.

Our claim is not that a model with sticky expectations is the only, or even the most plausible, way to generate return predictability pockets. However, there are intuitive reasons to expect sticky expectations models to be easier to reconcile with return predictability pockets. Compared to a setup with rational expectations, sticky expectations models reduce the spikiness in asset prices after a large shock to the true growth rate of cash flows, which leads to a predictable drift in realized returns. Instead, the change in price levels is roughly zero on impact and only gradually reflects the change in valuations associated with using the correct cash flow growth rate. Further, sticky expectations can introduce a wedge between agents' expectations and the true conditional mean of the cash flow growth rate process. We show that this wedge is correlated with observable state variables in the sticky expectations model and that, as this wedge cumulates over time, these state variables can be used in simple univariate regression models to identify local return predictability.

A. Present Value Model with Sticky Expectations

Following a modeling approach analogous to Bouchaud et al. (2019) and Gómez-cram (2022), our starting point is a standard log-linearized present value model of asset prices. We first specify the behavior of cash flows; we then turn to agents' beliefs and subjective discount rates. Dividends evolve according to the following law of motion under the objective probability distribution:

$$\Delta d_{t+1} = \mu_d + z_{cf,t} + \epsilon_{d,t+1}, \tag{17}$$

$$z_{cf,t+1} = \rho_{cf} z_{cf,t} + \epsilon_{cf,t+1}. \tag{18}$$

Consistent with the reduced-form representation proposed by Bouchaud et al. (2019) and Coibion and Gorodnichenko (2015), agents have sticky expectations in the spirit of Mankiw and Reis (2002). Letting F_t denote conditional expectations under agents' subjective beliefs at time t, sticky expectations are captured by

$$\begin{aligned} F_t[\Delta d_{t+1+h}] &= \mu_d + (1-\lambda)E_t[z_{cf,t+h}] + \lambda F_{t-1}[\Delta d_{t+1+h} - \mu_d] \\ &= \mu_d + (1-\lambda)\rho_{cf}^h z_{cf,t} + \lambda \rho_{cf}^h F_{t-1}[\Delta d_{t+1} - \mu_d]. \end{aligned} \tag{19}$$

The basic intuition captured by these models is that agents' beliefs about macroeconomic fundamentals are somewhat slow to incorporate new information. Forecasts, even those of professional economists, are therefore subject to predictable biases.

The state variable $z_{cf,t}$ captures a persistent shifter of expected cash flow growth, which is not necessarily observable by agents in the model. Given the substantial debate about the extent to which cash flows are predictable at medium to long horizons (Cochrane (2008)), it seems plausible that a difficult-to-estimate variable like expected cash flow growth for the aggregate stock market might be subject to information rigidities. We allow such a possibility in the model below, and discipline the magnitude of rigidities on estimates from microdata. For parsimony, we assume that agents have rational expectations about all remaining state variables in the model.

Coibion and Gorodnichenko (2015) show that a specification for beliefs like equation (19) obtains from two distinct microfoundations. The first is a sticky expectations model in which a measure $1-\lambda$ of agents update their beliefs about the relevant variable each period. The second is a setting in which $z_{cf,t+1}$ is unobserved but agents individually observe noisy signals about the state variable and update beliefs using the Kalman filter. In such a case, consensus expectations update as a weighted average of the prior and the new signal.[39] The parameter λ captures the degree of sluggishness in the extent to which agents' expectations update to reflect new information about expected macroeconomic fundamentals embedded in the cash flow shock $\epsilon_{cf,t}$. Rational expectations are nested as a special case of (19) when $\lambda = 0$; stickiness increases as λ rises above zero.

To incorporate additional asset pricing dynamics, we introduce exogenous shifters of subjective risk premia and risk-free rates, which (for simplicity) are known, not subject to information rigidities, and follow the laws of motion

$$F_t[r_{t+1} - r_{f,t+1}] = \mu_{rp} + z_{dr,t}, \tag{20}$$

$$z_{dr,t+1} = \rho_{dr}\, z_{dr,t} + \epsilon_{dr,t+1}, \tag{21}$$

$$r_{f,t+1} = \mu_{rf} + \beta_{rf,dr}\, z_{dr,t} + \beta_{rf,cf}\, F_t[\Delta d_{t+1} - \mu_d] + z_{tp,t}, \tag{22}$$

$$z_{tp,t+1} = \rho_{tp}\, z_{tp,t} + \epsilon_{tp,t+1}. \tag{23}$$

Here, $z_{dr,t}$ allows for a "standard" risk premium channel and follows a homoskedastic AR(1) process. The AR(1) state variable, $z_{tp,t}$, allows for additional variables (e.g., time preference shocks) that capture variation in the risk-free rate, which is independent from expected cash flows and discount rates. These

[39] Such a direct interpretation in this context requires that agents do not extract information from common signals such as consensus forecasts and/or prices (as is assumed to be the case for a subset of agents in the model of Hong and Stein (1999)). See also Barberis, Shleifer, and Vishny (1998) and Daniel, Hirshleifer, and Subrahmanyam (1998).

variables generate independent variation in valuations, realized returns, and the risk-free rate. We allow the risk-free interest rate to load on all three state variables, that is, $z_{dr,t}$, $z_{tp,t}$, and subjective expected cash flow growth.

Similar to Katz, Lustig, and Nielsen (2017) and Bouchaud et al. (2019), we assume that asset prices satisfy an approximate present value identity under agents' beliefs.[40] We start with the familiar log-linearized present value model,

$$r_{t+1} \approx k + \rho(p_{t+1} - d_{t+1}) + \Delta d_{t+1} + (d_t - p_t). \tag{24}$$

Iterating on this approximate accounting identity and taking expectations under agents' subjective beliefs yields the present value pricing formula

$$p_t - d_t = \frac{k}{1-\rho} + F_t\left[\sum_{j=0}^{\infty} \rho^j [\Delta d_{t+1+j} - r_{t+1+j}]\right]. \tag{25}$$

As is well-known in this literature, assuming a pricing formula such as (25) is not immediate and involves a departure from full rationality since agents fail to fully incorporate signals—such as information obtainable from local kernel regressions and equilibrium—that could be used to yield more accurate forecasts of expected returns and cash flows.[41]

Under these assumptions, we obtain by direct computation the valuation formula

$$\begin{aligned} p_t - d_t = {} & \frac{k}{1-\rho} + \frac{\mu_d - \mu_{rf} - \mu_{rp}}{1-\rho} + \frac{1-\beta_{rf,cf}}{1-\rho\cdot\rho_{cf}} F_t[\Delta d_{t+1} - \mu_d] \\ & - \frac{1+\beta_{rf,dr}}{1-\rho\cdot\rho_{dr}} z_{dr,t} - \frac{1}{1-\rho\cdot\rho_{tp}} z_{tp,t}, \end{aligned} \tag{26}$$

which we can use to simulate returns under the objective law of motion given the state variables.

B. Subjective and Objective Return Predictability

Next, we consider sources of return predictability in the sticky expectations model. Supposing that all state variables were observed, the expected excess

[40] See also De La O and Myers (2021) and Gómez-cram (2022), who make the same assumption.

[41] Note that we are implicitly assuming that asset prices reflect "consensus" expectations about cash flows of a set of behavioral agents. As noted by Bouchaud et al. (2019), one could potentially introduce a more complicated equilibrium involving interactions between boundedly rational agents with sticky expectations and more sophisticated agents with more accurate beliefs but capital constraints. Consistent with their approach, we do not pursue such an extension here, but we conjecture that it would likely result in similar qualitative dynamics as our simpler specification, albeit attenuated quantitatively toward the rational expectations benchmark. Further, given that our model features a distortion in beliefs about aggregate cash flows, any strategy of the sophisticated agents would be impossible to implement without facing an exposure to substantial nondiversifiable risk. See also Angeletos and Huo (2021) for a more explicit treatment of these issues in a related class of models.

return under the objective measure satisfies

$$E_t[r_{t+1} - r_{f,t+1}] = \mu_r + z_{dr,t} + \left[1 + \frac{(1-\beta_{rf,cf})\rho\rho_{cf}(1-\lambda)}{1-\rho\cdot\rho_{cf}}\right](z_{cf,t} - F_t[\Delta d_{t+1} - \mu_d]). \tag{27}$$

The first term $z_{dr,t}$ captures a standard component associated with agents' subjective risk premium, as in a standard present value model. When agents do not have rational expectations ($\lambda \neq 0$), there is a second term that captures the wedge between the objective forecast an econometrician would make if $z_{cf,t}$ were known and the agent's forecast of the risk premium, which is the sum of two components. First, if $z_{cf,t}$ exceeds agents' subjective expectation of dividend growth, cash flows will tend to surprise in the positive direction. Second, as beliefs about future growth rates gradually mean-revert toward the true expectation (agents become more optimistic), the price-dividend ratio will also continue to drift upward.[42]

By iterative substitution of the state dynamics above, we obtain a Wold decomposition for the difference between subjective and objective expectations of dividend growth:

$$\begin{aligned} z_{cf,t} - F_t[\Delta d_{t+1} - \mu_d] &= \sum_{j=0}^{\infty}\Big[\underbrace{\rho_{cf}^{j}}_{\substack{\text{Rational expectations}\\ MA(\infty)\text{coefficient}}} - \underbrace{\rho_{cf}^{j}(1-\lambda^{j+1})}_{\substack{\text{Sticky expectations}\\ MA(\infty)\text{coefficient}}}\Big]\epsilon_{cf,t-j} \\ &= \sum_{j=0}^{\infty}\rho_{cf}^{j}\lambda^{j+1}\epsilon_{cf,t-j}. \end{aligned} \tag{28}$$

This term is an exponentially weighted moving average of recent shocks to expected cash flow growth, which reflects the sluggish response of beliefs to persistent cash flow information. Estimates of λ from the literature suggest that sluggishness of beliefs is considerably lower than persistence of expected macroeconomic growth rates. This means that the leading term is λ^{j+1} and this term depends mostly on fairly recent shocks, adding a high-frequency component to expected returns.

Return innovations relative to subjective risk premia ($r_{t+1} - \mu_r - z_{dr,t}$) therefore display "local momentum." Each return is a noisy signal of the geometric sum in equation (28), so returns will tend to positively comove (even after netting out subjective risk premia) at short horizons. This point can be made more formally by writing the model in state-space form,

$$\begin{aligned} z_{cf,t+1} - F_{t+1}[\Delta d_{t+2} - \mu_d] \equiv \vartheta_{t+1} &= \rho_{cf}\ \lambda\ \vartheta_t + \lambda\ \epsilon_{cf,t+1}, \\ r_{t+1} - r_{f,t+1} &= \mu_r + z_{dr,t} + \left[1 + \frac{(1-\beta_{rf,cf})\rho\rho_{cf}(1-\lambda)}{1-\rho\cdot\rho_{cf}}\right]\vartheta_t + u_{t+1}, \end{aligned} \tag{29}$$

[42] In our calibration, $1-\lambda$ is larger than $1-\rho\rho_{cf}$ and $(1-\beta_{rf,cf})\rho\cdot\rho_{cf}$ is a bit smaller than 1, so the second term can potentially be quite large (around six in the current calibration).

where the residuals u_{t+1} and $\epsilon_{cf,t+1}$ have a modest positive correlation since pd_{t+1} slightly responds to the current cash flow shock $\epsilon_{cf,t+1}$. Supposing for simplicity that $z_{dr,t}$ was observed, to a fairly close approximation the Kalman filter will imply that an econometrian's forecast of ϑ_t is an exponentially weighted moving average of $r_{t+1} - \mu_r - z_{dr,t}$.[43] High recent past returns signal that future returns are likely to stay high over the near term. The constant term from our local kernel regression involves a weighted moving average of recent data and thus captures similar features. In our local kernel regressions, recent changes in state variables play a dual role: (i) they may be correlated with the subjective risk premium $z_{dr,t}$, and (ii) they provide informative signals on how beliefs about cash flows have changed in the recent past. Both forces combine to allow the econometrician to constructively (though imperfectly) capture an estimate of ex ante expected returns that is detectable in real time, as we demonstrate below.[44]

Low-frequency movements in dp_t reflect persistent variation in risk premia, but also in expected cash flow growth rates and real interest rates, whereas higher frequency movements reflect revisions in agents' beliefs that were unanticipated by agents but lead to predictable movements in valuation ratios. These factors also affect expected excess returns under the objective measure with different signs: recent increases in dp_t signal the likelihood of further upward drift over the near term due to sticky expectations, whereas low-frequency changes in dp_t are expected to gradually mean-revert downward. Further, even though our model has homoskedastic shocks for simplicity, since agents' forecast errors include ϑ_t, realized variance of returns also provides a noisy signal about the absolute value of ϑ_t. Thus, all of the predictive regressions that we consider are misspecified due to omitted variable bias coming from mismeasured predictors. This creates scope for benefits from using multivariate forecasts and/or univariate forecast combinations to further improve performance by controlling for more sources of omitted variable bias and averaging across different sources of misspecification, respectively.

In principle, local return predictability can arise from both $z_{dr,t}$ and ϑ_t. In practice, the latter channel turns out to be far more important than the former in our simulation exercises.[45] The rationale for why a "standard" time-varying risk premium channel does not go very far is quite similar to that discussed earlier when explaining the failure of conventional asset pricing models to match our evidence. Low-frequency movements in state variables that are common over each fitting window are approximately differenced out in the local regressions; in contrast, these effects dominate in constant-coefficient specifi-

[43] If we ignore the fact that u_{t+1} and $\epsilon_{cf,t+1}$ have a slight positive correlation, we have a standard Kalman filtering problem. Given that T is quite large and $\rho_{cf}\lambda$ is fairly far from one, the impact of initial conditions will dissipate rapidly, and the Kalman gain will converge to a constant.

[44] Figure IA.V in Section VII of the Internet Appendix presents impulse responses from large shocks to $z_{dr,t}$ and $z_{cf,t}$.

[45] To see this more formally, we conduct experiments below where we first subtract $z_{dr,t}$ from returns before conducting our out-of-sample experiments. Performance is quite similar, and actually slightly better after doing so, a result that is sensible in light of our findings in the previous section.

cations. Given the high persistence of $z_{dr,t}$ and the substantial negative correlation between return innovations and changes in $z_{dr,t}$, the effects of estimation error more than offset any small potential gains from timing the market using estimates of the subjective risk premium due to a very low signal-to-noise ratio and substantial Stambaugh (1999) bias. In contrast, ϑ_t is considerably less persistent and the correlations that modulate the degree of Stambaugh bias are considerably weaker. Accordingly, there is more scope for our methods to detect pockets of predictability in our sticky expectations framework below.

Specifically, as is clear from (29), expected excess returns under the objective probability measure are a linear combination of the slow-moving subjective risk premium $z_{dr,t}$ and the much less persistent belief discrepancy ϑ_t. In model simulations illustrated in Figure IA.6 of the Internet Appendix, taking the risk-free rate as an example, we find that pockets of predictability are particularly likely to occur shortly (two to three months) after periods in which $|\vartheta_t|$ is large, that is, periods in which a larger fraction of expected excess return variation is explained by the high-frequency belief component. In contrast, the state variable capturing the rational risk premium $z_{dr,t}$ is essentially uncorrelated with the pocket dummy. Related, the time-series correlation between the risk-free rate and ϑ_t as well as the predictive coefficient on the risk-free rate both tend to be larger in absolute value inside pockets. In other words, more of the variation in the state variables reflects changes in the high-frequency component, which makes it easier to capture return predictability using our local regressions.[46]

C. Quantitative Assessment

We next simulate from our calibrated model with sticky expectations and repeat our empirical exercises using model-generated data. In these simulations, we carefully fix the parameters of the sticky expectations model to match moments of the data such as the annualized sample means of dividend growth, the risk-free rate, and expected returns.

Although Section VIII of the Internet Appendix provides further details about how we calibrate our model, it is important to note what is *not* targeted in these calibrations. The central stickiness parameter is fixed ex ante using the empirical estimates of Coibion and Gorodnichenko (2015) so that $\lambda = 0.3^{4/252} \approx 0.981$. In other words, we deliberately fix the degree of information rigidity based on estimates from the literature, and the asset pricing moments selected are fairly standard—as such, they are not explicitly tied to any evidence related to pockets of predictability. We therefore view our examination of the model's ability (or lack thereof) to match evidence related to pockets as a nontargeted validation test of the model.

[46] In addition, the properties of expected returns and the covariance between returns and lagged predictors change in a direction that is favorable for detecting predictability during periods in which true expected cash flow growth rates recently changed substantially (high $|\vartheta_t|$). Our local kernel regression forecasts, by adapting to these changing covariances, are able to detect a meaningful level of out-of-sample predictability.

As additional points of comparison, we consider two alternative models. The first is a rational expectations version of our model with the same true cash flow dynamics but no information rigidities ($\lambda = 0$). The second is a rational expectations model whose parameters are recalibrated with $\lambda = 0$. Since the effects of sticky expectations on unconditional asset pricing moments are fairly modest, these recalibrated parameters are similar to those from our baseline model.

We summarize the results from these experiments in Table X, using a format similar to above with different columns corresponding to different asset pricing models. The first column includes our benchmark sticky expectations model, while the next two columns include the two calibrations of analogous rational expectations models as described above. Each block of results presents CW statistics—computed overall, in-pocket, and out-of-pocket—as well as performance measures from our market-timing regressions. The top three panels present results for the individual predictor variable followed by results from the multivariate kernel specification and the three forecast combinations.

In stark contrast to the asset pricing models considered in Table IX, as well as the rational expectations versions of our model with similar cash flow and subjective discount rate dynamics, the model with sticky expectations is capable of replicating a number of the patterns observed in the data. Local predictive regressions are consistently capable of detecting meaningful out-of-sample predictability, especially in-pocket, whereas they struggle outside of pockets. Across specifications, CW t -statistics are consistently highest (and higher than full-sample coefficients) inside pockets, though full-sample t-statistics are somewhat higher than in the data. The latter feature likely reflects the fact that shocks are Gaussian in the model, so the tendency to overfit large realized return shocks in our simulated samples is more muted relative to the data. [47] The bottom panels of Table X show that the multivariate kernel specification and forecast combinations also work well, in line with the intuition discussed above, with combination forecasts further benefitting from reductions in estimation error due to overfitting.

The middle and right panels of Table X illustrate that the ability to detect pockets of predictability via our local kernel approach does not transfer to the calibrated models with rational expectations. Analogous to the simulation exercises from the asset pricing models from Table IX, estimation error swamps any ability to reliably exploit information from our time-varying forecasts despite the fact that returns are predictable by $z_{rp,t}$. This result obtains in part because our state variables do not perfectly reveal $z_{rp,t}$, but the dominant force is parameter estimation error.

Consistent with our results on the CW statistics, simulated market-timing regressions indicate that an investor could meaningfully improve her Sharpe

[47] In the model, we could easily replicate these features by introducing jumps in $z_{rp,t+1}$, $z_{tp,t+1}$, and/or $\epsilon_{d,t+1}$, but we elected not to introduce these extra parameters in the interest of parsimony. Related, the absence of large jumps likely reduces jumps in realized volatility and likely improves its performance as a predictor relative to the empirical application.

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 three are Clark and West (2007) t-statistics relative to a prevailing mean benchmark in the full sample, in-pocket, and out-of-pocket. The second 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.	*SE*	p-Value	Avg.	*SE*	p-Value	Avg.	*SE*	p-Value
					dp					
CW_{fs}	−0.74	1.43	1.21	0.09	0.13	1.01	0.40	0.26	1.02	0.34
CW_{ip}	3.00	2.39	1.12	0.59	0.00	0.93	0.00	0.03	1.01	0.01
CW_{oop}	−1.62	−0.65	1.03	0.36	0.15	1.01	0.10	0.28	0.98	0.07
α	1.69	1.09	1.62	0.71	0.02	1.92	0.40	0.22	1.44	0.32
t_α	2.10	0.79	1.19	0.28	0.01	0.99	0.05	0.16	1.01	0.07
SR	0.47	0.50	0.18	0.87	0.33	0.12	0.26	0.43	0.11	0.74
					rf					
CW_{fs}	0.68	3.01	1.05	0.04	−0.17	1.04	0.42	−0.31	1.02	0.34
CW_{ip}	3.28	3.54	1.06	0.81	−0.15	0.99	0.00	−0.28	0.99	0.00
CW_{oop}	−1.58	0.42	1.08	0.08	−0.13	1.09	0.20	−0.18	1.06	0.20
α	3.57	3.70	1.58	0.93	−0.49	2.03	0.06	−0.53	1.45	0.01
t_α	4.35	2.62	1.06	0.12	−0.27	1.04	0.00	−0.38	1.02	0.00
SR	0.79	0.62	0.18	0.37	0.33	0.13	0.00	0.43	0.12	0.01
					rvar					
CW_{fs}	−1.49	2.23	1.05	0.00	−0.20	0.98	0.21	−0.45	0.98	0.30
CW_{ip}	2.88	2.99	1.10	0.92	−0.14	1.02	0.01	−0.36	0.98	0.00
CW_{oop}	−1.77	−0.08	1.01	0.11	−0.17	1.00	0.13	−0.30	1.00	0.16
α	2.31	2.65	1.52	0.83	−0.65	1.95	0.15	−0.79	1.40	0.04
t_α	3.63	1.87	1.05	0.11	−0.34	0.99	0.00	−0.56	0.98	0.00
SR	0.68	0.56	0.18	0.51	0.33	0.13	0.01	0.44	0.12	0.05
					comb1					
CW_{fs}	−4.48	3.47	1.00	0.00	−0.13	0.95	0.00	−0.29	0.97	0.00
CW_{ip}	4.57	3.51	1.03	0.32	−0.13	0.95	0.00	−0.29	0.97	0.00
CW_{oop}	–	−0.01	0.96	–	0.02	0.97	0.98	0.01	0.99	0.99

(*Continued*)

Table X—*Continued*

		Baseline			Baseline ($\lambda = 0$)			RE Recalibrated		
Stats	Data	Avg.	*SE*	p-Value	Avg.	*SE*	p-Value	Avg.	*SE*	p-Value
α	6.38	4.06	1.47	0.13	−0.46	1.65	0.00	−0.44	1.16	0.00
t_α	6.11	3.25	1.05	0.01	−0.27	0.94	0.77	−0.37	0.95	0.70
SR	1.00	0.69	0.18	0.11	0.33	0.13	0.00	0.43	0.12	0.00
					comb2					
CW_{fs}	4.94	3.62	1.11	0.25	−0.06	0.95	0.00	−0.18	0.97	0.00
CW_{ip}	5.04	3.62	1.11	0.22	−0.06	0.95	0.00	−0.18	0.97	0.00
CW_{oop}	–	3.62	1.11	–	−0.06	0.95	0.95	−0.18	0.97	0.86
α	6.10	5.67	2.05	0.84	−0.37	2.22	0.01	−0.39	1.65	0.00
t_α	5.66	3.27	1.12	0.05	−0.16	0.94	0.87	−0.23	0.96	0.81
SR	0.87	0.78	0.20	0.65	0.33	0.13	0.00	0.44	0.12	0.00
					comb3					
CW_{fs}	−1.03	2.73	1.07	0.00	−0.09	1.00	0.36	−0.20	0.97	0.41
CW_{ip}	2.32	3.38	1.10	0.35	−0.11	1.02	0.03	−0.22	0.99	0.02
CW_{oop}	−2.12	−0.57	1.09	0.17	0.00	0.99	0.05	−0.05	0.99	0.05
α	0.76	3.21	1.60	0.14	−0.47	1.96	0.54	−0.41	1.39	0.41
t_α	1.32	2.25	1.08	0.40	−0.24	0.99	0.81	−0.30	0.97	0.76
SR	0.43	0.59	0.18	0.39	0.33	0.13	0.43	0.43	0.11	0.99
					mv					
CW_{fs}	−0.99	2.21	1.10	0.01	0.12	1.06	0.31	0.12	1.06	0.31
CW_{ip}	3.74	2.76	1.12	0.39	−0.02	0.98	0.00	−0.02	0.98	0.00
CW_{oop}	−1.49	0.32	1.00	0.09	0.15	1.05	0.13	0.15	1.05	0.13
α	2.59	2.49	1.61	0.95	0.12	2.02	0.24	0.12	2.02	0.24
t_α	3.37	1.78	1.13	0.18	0.07	1.03	0.95	0.07	1.03	0.95
SR	0.58	0.56	0.18	0.89	0.33	0.12	0.06	0.33	0.12	0.06

ratio by adjusting her weights on the market using our local kernel approach. These results obtain across all predictors we consider, and we again find that combination and multivariate approaches work well. However, while our timing strategy generates nontrivial improvements in the Sharpe ratio, such a strategy remains subject to considerable risk. In contrast, market-timing alpha estimates are consistently negative across all specifications in the rational expectations models, despite the fact that the underlying models feature time-varying risk premia.

Further examination shows that the full-sample regression coefficients of excess returns on the log price-dividend ratio (pd)and the risk-free rate are both almost identical for the sticky and rational expectations models, suggesting that the long-run return predictability patterns are similar in these types of models. Accordingly, our results are not incompatible with evidence already established with constant coefficient specifications.

Moreover, the overall degree of "mispricing" is fairly modest in this economy. To see this, we can decompose $\text{Var}[pd_t]$ into the sum of three pieces, namely, the variance of the price that would obtain under the rational expectations beliefs, $\text{Var}[pd_t^*]$, the variance of the difference between the observed price-dividend ratio and this "correct" one, $\text{Var}[pd_t - pd_t^*]$, and two times the covariance between the two terms, $2\,\text{Cov}[pd_t^*, pd_t - pd_t^*]$. We find that

$$1 = \frac{\text{Var}[pd_t]}{\text{Var}[pd_t]} = \underbrace{\frac{\text{Var}[pd_t^*]}{\text{Var}[pd_t]}}_{1.0261 \text{ in model}} + \underbrace{\frac{\text{Var}[pd_t - pd_t^*]}{\text{Var}[pd_t]}}_{0.0098 \text{ in model}} + \underbrace{\frac{2\,\text{Cov}[pd_t^*, pd_t - pd_t^*]}{\text{Var}[pd_t]}}_{-0.0358 \text{ in model}}. \quad (30)$$

The observed price-dividend ratio thus tracks the "true" one fairly closely overall. The variance of the true price-dividend component pd_t^* is more than 100 times larger than the variance of the "pricing error" $pd_t - pd_t^*$ component, which indicates that these two variables are quite similar at low frequencies. However, the two can deviate by nontrivial amounts at higher frequencies.

Intriguingly, one might have thought that there is a tension between the evidence suggesting that return predictability is elusive, almost nonexistent, and/or fragile at high frequencies and the evidence/theoretical work on fundamental drivers of fluctuations in asset prices at lower frequencies. Our model suggests that this is not necessarily the case. Small, high-frequency discrepancies in price levels related to behavioral biases/information frictions can inject considerable noise into short-horizon risk premium estimates without invalidating the insights we glean about predictability at longer horizons from models with rational, low-frequency fluctuations in risk premia.

Finally, while we do not explicitly introduce a cross-section of different assets to be priced here, an extension to different assets with cash flow growth rates that load differentially on our aggregate state variables is straightforward. Although we do not perform a quantitative assessment, such an extension can easily match the market-timing results that we obtain for the size and value portfolios in Table VIII qualitatively. This is because the sticky expectations model quite naturally generates factor momentum, which is an important component of the overall return to momentum strategies (Moskowitz, Ooi, and Pedersen (2012), Ehsani and Linnainmaa (2022)). In our sticky expectations model, due to the sluggish incorporation of news about fundamentals into prices, stocks (and factors) whose prices have recently increased are likely to continue to drift upward. Intriguingly, Ehsani and Linnainmaa (2022) find that factor momentum is particularly concentrated in factors that explain the largest share of variation in the cross-section of realized returns, that is, in portfolios that contain substantial macro information. Thus, sluggish incorporation of macro news into agents' information sets could plausibly be connected to patterns of factor momentum in the data.[48]

[48] Moreover, to the extent that sluggish incorporation of information, especially aggregate information, into beliefs is a general feature of how agents process information about future aggregate payoffs, it is somewhat less surprising to find that momentum appears across a wide variety of

D. Direct Evidence on the Mechanism

Finally, we provide some direct evidence that links the expected return forecasts used in our market-timing strategy and measures of biases in the beliefs of professional forecasters. Specifically, we use the original data considered by Coibion and Gorodnichenko (2015). These data include a measure of the forecast errors made by forecasters in various longitudinal surveys. [49] Consistent with the analysis in section III of their paper, we focus on the quarterly subsample of forecasts from the Survey of Professional Forecasters (SPF). Specifically, Coibion and Gorodnichenko (2015) compute $x_{q+h} - F_{t_q}[x_{q+h}]$ for several macroeconomic variables x_q and at various forecast horizons $h \geq 0$, where $F_{t_q}[x_{q+h}]$ is the consensus forecast for quarter q as of date t_q.[50]

Under rational expectations, forecast errors as defined above should be orthogonal to any information that was available as of time t_q. Given that the information contained in our expected return forecasts from prior to t_q would have been available by the time at which the survey was conducted, our forecasts should be uncorrelated with these forecast errors. In our theoretical model, these forecast errors would map into unexpected cash flow shocks and in turn realized return surprises from the perspective of agents with sticky expectations. Since direct forecasts of dividend growth are not available, we consider three choices for the variable x, all of which capture information about the business cycle: real GDP growth gy, the unemployment rate ue (in percentage points), and real industrial production growth ip. Under sticky expectations, we would expect to see a positive correlation between our return forecasts and forecast errors in procyclical variables like gy and ip and a negative correlation with the countercyclical variable ue.

For each variable and each consensus forecast date, we compute an average of quarterly forecast errors at multiple horizons $h \in \{0, 4\}$ as follows:

$$\hat{\varepsilon}_{x,t_q} \equiv \frac{1}{5} \sum_{h=0}^{4} x_{q+h} - F_{t_q} x_{q+h}, \tag{31}$$

where t_q refers to the time at which the forecast is formed, and $F_{t_q} x_{q+h}$ are h-period-ahead forecasts from the SPF formed at time t_q of the quarterly variable x. We then study the correlation between ex ante return forecasts from our time-varying coefficient models and $\hat{\varepsilon}_{x,t_q}$, using forecasts from both the univariate and multivariate models that switch to the prevailing mean forecast outside of ex ante identified pockets. We convert Coibion-Gorodnichenko fore-

asset classes (Asness, Moskowitz, and Pedersen (2013)) and that momentum strategies might comove. Provided that recent winners include stocks who load disproportionately on macroeconomic factors about which agents were revising beliefs most aggressively, they will also tend to fall the most if these revisions in beliefs turn out to be incorrect. Such a phenomenon could generate momentum crashes around business-cycle turning points (Daniel and Moskowitz (2016)).

[49] Data are available from https://www.aeaweb.org/articles?id=10.1257/aer.20110306.

[50] Note that $h = 0$ corresponds to a "nowcast" of a quarterly variable produced during the middle of the current quarter, the time at which the survey is administered.

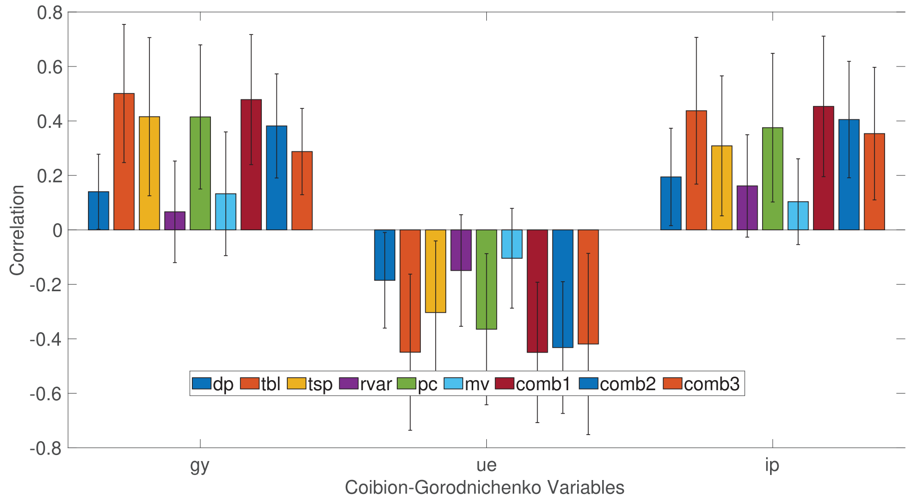

Figure 3. Correlation of Coibion-Gorodnichenko forecast errors with excess return forecasts. This figure shows correlations between forecast errors of three macroeconomic variables from the Survey of Professional Forecasters (SPF) and excess return forecasts from our time-varying coefficient models. 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 a 95% confidence interval 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. (Color figure can be viewed at wileyonlinelibrary.com)

cast errors to a daily frequency by setting $\hat{\varepsilon}_t = \hat{\varepsilon}_{t_q}$ for all days t in quarter q. Because respondents to the SPF send in their forecasts around the middle of each quarter, to avoid possible look-ahead bias we only use return forecasts from the first month of each quarter when estimating these correlations.[51] We then estimate correlations between $\hat{r}_{t|t-1}$ and $\hat{\varepsilon}_t$ and compute Newey-West standard errors using a rule-of-thumb bandwidth.

Our estimated correlations and 95% standard error bands are reported in Figure 3, which contains three groupings of 10 bars. Each grouping corresponds to the forecast errors of one of the variables from Coibion and Gorodnichenko (2015). Each of the nine bars corresponds to a model for forecasting excess returns, and the height of each bar corresponds to the correlation between these forecasts and the forecast errors. Consistent with our proposed sticky expectations mechanism, we find a robust empirical link between our expected return forecasts and future forecast errors. For instance, forecasts based on the T-bill rate have a correlation of around 50% with future forecast errors. Although signs are consistent across all specifications, the correlations

[51] As an additional robustness check, we lead the forecast errors by one additional quarter and repeat the analysis. These correlations, which are reported in Figure IA.7 of the Internet Appendix, are similar in terms of signs and statistical significance but are somewhat attenuated toward zero.

are weaker for the multivariate kernel model and stronger for the three forecast combinations.

As a final observation, pockets tend to be periods of time in which ϑ_t is large in absolute value. Since both r_t and r_{t-1} have components that are linear in ϑ_t and ϑ_{t-1}, respectively, autocorrelation tends to be larger inside versus outside of pockets. We see exactly this pattern in Table IA.I of the Internet Appendix, especially for predictors other than dp.

In conclusion, while the models with rational expectations do not match our evidence related to short-horizon return predictability, a simple model with sticky expectations can account for our evidence both qualitatively and quantitatively. Further, the expectations data in this setting provide direct evidence showing that our ex ante return forecasts explain a nontrivial amount of predictable variation in professional forecasters' expectation errors.

VI. Conclusion

We develop a nonparametric kernel regression approach to detect pockets with local predictability of stock returns. Our out-of-sample approach uses real-time information to monitor for improvements in the accuracy of return forecasts from the local kernel regression model relative to a benchmark no-predictability model. Empirically, we find evidence that while stock returns are unpredictable the vast majority of the time, there are relatively short-lived pockets in which stock returns can be predicted. Moreover, such out-of-sample return predictability is sufficiently large to be exploitable for economic gains, particularly if used in conjunction with economic constraints on the return forecasts or forecast combination methods that incorporate information on which models identify local pockets at a given point in time.

To explore possible sources of return predictability, we simulate returns from a range of statistical models that incorporate features such as highly persistent predictors, time-varying heteroskedasticity, and Stambaugh (1999) bias. We also simulate returns from a set of workhorse asset pricing models representative of the dynamics of returns and state variables consistent with time-varying risk premia. Both types of models fail to match the empirical evidence of in-pocket return predictability and its implications for the investment performance of a simple dynamic trading strategy set up to exploit pockets with return predictability.

Building on recent papers such as Bouchaud et al. (2019), we finally develop a simple asset pricing model in which agents have sticky expectations about future cash flow growth. Our model, which nests rational expectations as a special case, allows for a wedge to form between agents' subjective expectations and forecasts computed under the true cash flow process. For some sequences of shocks to the underlying state variables, this gives rise to local return predictability. We show how this can be captured through familiar state variables such as the dividend-price ratio, the risk-free rate, and realized return volatility, and we also demonstrate why strategies such as forecast combination can be expected to improve forecast accuracy as has been documented in studies such as Rapach, Strauss, and Zhou (2010).

Initial submission: October 18, 2019; Accepted: January 3, 2022
Editors: Stefan Nagel, Philip Bond, Amit Seru, and Wei Xiong

REFERENCES

Adrian, Tobias, Erkko Etula, and Tyler Muir, 2014, Financial intermediaries and the cross-section of asset returns, *Journal of Finance* 69, 2557–2596.

Ang, Andrew, and Geert Bekaert, 2007, Stock return predictability: Is it there?, *Review of Financial Studies* 20, 651–707.

Ang, Andrew, and Dennis Kristensen, 2012, Testing conditional factor models, *Journal of Financial Economics* 106, 132–156.

Angeletos, George-Marios, and Zhen Huo, 2021, Myopia and anchoring, *American Economic Review* 111, 1166–1200.

Angeletos, George-Marios, Zhen Huo, and Karthik A. Sastry, 2021, Imperfect macroeconomic expectations: Evidence and theory, *NBER Macroeconomics Annual* 35, 1–86.

Asness, Clifford S., Tobias J. Moskowitz, and Lasse Heje Pedersen, 2013, Value and momentum everywhere, *Journal of Finance* 68, 929–985.

Baker, Malcolm, and Jeffrey Wurgler, 2006, Investor sentiment and the cross-section of stock returns, *Journal of Finance* 61, 1645–1680.

Baker, Malcolm, and Jeffrey Wurgler, 2007, Investor sentiment in the stock market, *Journal of Economic Perspectives* 21, 129–152.

Bansal, Ravi, Dana Kiku, and Amir Yaron, 2012, An empirical evaluation of the long-run risks model for asset prices, *Critical Finance Review* 1, 183–221.

Bansal, Ravi, and Amir Yaron, 2004, Risks for the long run: A potential resolution of asset pricing puzzles, *Journal of Finance* 59, 1481–1509.

Barberis, Nicholas, Andrei Shleifer, and Robert Vishny, 1998, A model of investor sentiment, *Journal of Financial Economics* 49, 307–343.

Bordalo, Pedro, Nicola Gennaioli, Yueran Ma, and Andrei Shleifer, 2020, Overreaction in macroeconomic expectations, *American Economic Review* 110, 2748–2782.

Bouchaud, Jean-Philippe, Philipp Krueger, Augustin Landier, and David Thesmar, 2019, Sticky expectations and the profitability anomaly, *Journal of Finance* 74, 639–674.

Cai, Zongwu, 2007, Trending time-varying coefficient time series models with serially correlated errors, *Journal of Econometrics* 136, 163–188.

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

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

Campbell, John Y., Andrew W. Lo, and A. Craig MacKinlay, 1997, 7. Present-value relations, in Peter Dougherty, ed., *The Econometrics of Financial Markets* (Princeton University Press, Princeton, NJ).

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

Campbell, John Y., and Samuel B. Thompson, 2008, Predicting excess stock returns out of sample: Can anything beat the historical average?, *Review of Financial Studies* 21, 1509–1531.

Chen, Bin, and Yongmiao Hong, 2012, Testing for smooth structural changes in time series models via nonparametric regression, *Econometrica* 80, 1157–1183.

Chen, Yu, Thomas Cosimano, Alex Himonas, and Peter Kelly, 2009, Asset pricing with long run risk and stochastic differential utility: An analytic approach, *SSRN Electronic Journal*. Available at https://doi.org/10.2139/ssrn.1502968

Clark, Todd E., and Michael W. McCracken, 2001, Tests of equal forecast accuracy and encompassing for nested models, *Journal of Econometrics* 105, 85–110.

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

Cochrane, John H., 2008, The dog that did not bark: A defense of return predictability, *Review of Financial Studies* 21, 1533–1575.

Coibion, Olivier, and Yuriy Gorodnichenko, 2015, Information rigidity and the expectations formation process: A simple framework and new facts, *American Economic Review* 105, 2644–2678.

Constantinides, George M., and Darrell Duffie, 1996, Asset pricing with heterogeneous consumers, *Journal of Political Economy* 104, 219–240.

Constantinides, George. M., and Anisha Ghosh, 2017, Asset pricing with countercyclical household consumption risk, *Journal of Finance* 72, 415–460.

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

Daniel, Kent, David Hirshleifer, and Avanidhar Subrahmanyam, 1998, Investor psychology and security market under and overreactions, *Journal of Finance* 53, 1839–1885.

Daniel, Kent, and Tobias J. Moskowitz, 2016, Momentum crashes, *Journal of Financial Economics* 122, 221–247.

d'Arienzo, Daniele, 2020, Increasing overreaction and excess volatility of long rates, Working paper, Bocconi University.

De La O, Ricardo, and Sean Myers, 2021, Subjective cash flow and discount rate expectations, *Journal of Finance* 76, 1339–1387.

Di Tella, Sebastian, 2017, Uncertainty shocks and balance sheet recessions, *Journal of Political Economy* 125, 2038–2081.

Diebold, Francis X., and Roberto S. Mariano, 1995, Comparing predictive accuracy, *Journal of Business and Economic Statistics* 13, 253–263.

Drechsler, Itamar, and Amir Yaron, 2011, What's vol got to do with it, *Review of Financial Studies* 24, 1–45.

Ehsani, Sina, and Juhani Linnainmaa, 2022, Factor momentum and the momentum factor, *Journal of Finance* 77, 1877–1919.

Epstein, Larry G., and Stanley E. Zin, 1989, Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework, *Econometrica* 57, 937–969.

Eraker, Bjørn, and Ivan Shaliastovich, 2008, An equilibrium guide to designing affine asset pricing models, *Mathematical Finance* 18, 519–543.

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

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

Gârleanu, Nicolae, and Stavros Panageas, 2015, Young, old, conservative, and bold: The implications of heterogeneity and finite lives for asset pricing, *Journal of Political Economy* 123, 670–685.

Giglio, Stefano, and Bryan Kelly, 2018, Excess volatility: Beyond discount rates, *Quarterly Journal of Economics* 133, 71–127.

Gómez-cram, Roberto, 2022, Late to recessions: Stocks and the business cycle, *Journal of Finance* 77, 923–966.

Green, Jeremiah, John R. M. Hand, and Mark T. Soliman, 2011, Going, going, gone? The apparent demise of the accruals anomaly, *Management Science* 57, 797–816.

Hansen, Lars Peter, John C. Heaton, and Nan Li, 2008, Consumption strikes back? Measuring long-run risk, *Journal of Political Economy* 116, 260–302.

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

Herskovic, Bernard, Bryan Kelly, Hanno Lustig, and Stijn Nieuwerburgh, 2015, The common factor in idiosyncratic volatility: Quantitative asset pricing implications, *Journal of Financial Economics* 119, 249–283.

Hong, Harrison, Terence Lim, and Jeremy C. Stein, 2000, Bad news travels slowly: Size, analyst coverage, and the profitability of momentum strategies, *Journal of Finance* 55, 265–295.

Hong, Harrison, and Jeremy C. Stein, 1999, A unified theory of underreaction, momentum trading, and overreaction in asset markets, *Journal of Finance* 54, 2143–2184.

Hong, Harrison, Walter Torous, and Rossen Valkanov, 2007, Do industries lead stock markets?, *Journal of Financial Economics* 83, 367–396.

Hou, Kewei, 2007, Industry information diffusion and the lead-lag effect in stock returns, *Review of Financial Studies* 20, 1113–1138.

Johannes, Michael, Arthur Korteweg, and Nicholas Polson, 2014, Sequential learning, predictability, and optimal portfolio returns, *The Journal of Finance* 69, 611–644.

Katz, Michael, Hanno Lustig, and Lars Nielsen, 2017, Are stocks real assets? Sticky discount rates in stock markets, *Review of Financial Studies* 30, 539–587.

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

Kelly, Bryan, and Seth Pruitt, 2013, Market expectations in the cross-section of present values, *Journal of Finance* 68, 1721–1756.

Lettau, Martin, and Sydney C. Ludvigson, 2010, Measuring and modeling variation in the risk-return trade-off, in Yacine Ait-Sahalia, and Lars Peter Hansen, eds.: *Handbook of Financial Econometrics: Tools and Techniques*, vol. 1, 617–690 (North-Holland, San Diego).

Mankiw, N Gregory, and Ricardo Reis, 2002, Sticky information versus sticky prices: A proposal to replace the New Keynesian Phillips curve, *Quarterly Journal of Economics* 117, 1295–1328.

McLean, R. David, and Jeffrey Pontiff, 2016, Does academic research destroy stock return predictability?, *Journal of Finance* 71, 5–32.

Moreira, Alan, and Tyler Muir, 2017, Volatility-managed portfolios, *Journal of Finance* 72, 1611–1644.

Moskowitz, Tobias J., and Mark Grinblatt, 1999, Do industries explain momentum?, *Journal of Finance* 54, 1249–1290.

Moskowitz, Tobias J., Yao Hua Ooi, and Lasse Heje Pedersen, 2012, Time series momentum, *Journal of Financial Economics* 104, 228–250.

Paye, Bradley S., and Allan Timmermann, 2006, Instability of return prediction models, *Journal of Empirical Finance* 13, 274–315.

Pesaran, M. Hashem, and Allan Timmermann, 1995, Predictability of stock returns: Robustness and economic significance, *Journal of Finance* 50, 1201–1228.

Pettenuzzo, Davide, Allan Timmermann, and Rossen Valkanov, 2014, Forecasting stock returns under economic constraints, *Journal of Financial Economics* 114, 517–553.

Politis, Dimitris N., and Halbert White, 2004, Automatic block-length selection for the dependent bootstrap, *Econometric Review* 23, 53–70.

Rapach, David, and Guofu Zhou, 2013, Forecasting stock returns, in Graham Elliott, and Allan Timmermann, eds.: *Handbook of Economic Forecasting*, vol. 2, 328–383 (Elsevier, Amsterdam).

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.

Rapach, David E., and Mark E. Wohar, 2006, Structural breaks and predictive regression models of aggregate U.S. stock returns, *Journal of Financial Econometrics* 4, 238–274.

Robinson, Peter M., 1989, Nonparametric estimation of time-varying parameters, in Peter Hackl, ed.: *Statistical Analysis and Forecasting of Economic Structural Change*, 253–264 (Springer, Berlin).

Schmidt, Lawrence, 2020, Climbing and falling off the ladder: Asset pricing implications of labor market event risk, SSRN Working Paper No. 2471342.

Schwert, G. William, 2002, Anomalies and market efficiency, Working Paper 9277, National Bureau of Economic Research.

Sims, Christopher A., 2003, Implications of rational inattention, *Journal of Monetary Economics* 50, 665–690.

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

Timmermann, Allan, 2006, Forecast combinations, *Handbook of Economic Forecasting*, vol. 1, 135–196 (Elsevier).

Timmermann, Allan, 2008, Elusive return predictability, *International Journal of Forecasting* 24, 1–18.

van Binsbergen, Jules H., and Ralph S. J. Koijen, 2010, Predictive regressions: A present value approach, *Journal of Finance* 65, 1439–1471.

Wachter, Jessica A., 2005, Solving models with external habit, *Finance Research Letters* 2, 210–226.

Wachter, Jessica A., 2013, Can time-varying risk of rare disasters explain aggregate stock market volatility?, *Journal of Finance* 68, 987–1035.

Wang, Chen, 2020, Under- and overreaction in yield curve expectations, *Working Paper*.

Welch, Ivo, and Amit Goyal, 2008, A comprehensive look at the empirical performance of equity premium prediction, *Review of Financial Studies* 21, 1455–1508.

Woodford, Michael, 2001, Imperfect common knowledge and the effects of monetary policy, Working Paper 8673, National Bureau of Economic Research.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Appendix S1: Internet Appendix.
Replication Code.