Pockets of Predictability

Leland E. Farmer
University of Virginia

Lawrence Schmidt
University of Chicago

Allan Timmermann
University of California, San Diego

March 28, 2018

Abstract

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 little or no evidence of return predictability. We document this empirically using a flexible non-parametric approach and explore possible explanations of this finding, including time-varying risk-premia. We find that short-lived predictability pockets are inconsistent with a broad class of affine asset pricing models. Conversely, pockets of return predictability are more in line with a model with investors' incomplete learning about a highly persistent growth component in the underlying cash flow process which undergoes occasional regime shifts.

Key words: Predictability of stock returns; incomplete learning; Markov switching predictive systems; cash flows; affine asset pricing models.

1 Introduction

Are stock market returns predictable and, if so, how often, for how long and by how much? Even answering the first of these questions has proven surprisingly elusive as illustrated by the overarching conclusion from many empirical studies that return predictability tends to be highly unstable, varying greatly across time and across different markets.[1]

Existing evidence on return predictability has mostly been established using linear, constant-coefficient regressions which pool information across long historical spans of time and thus are designed to establish whether stock returns are predictable "on average," i.e., 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 such concerns, this paper adopts a new estimation strategy capable of identifying patterns in return predictability that is "local" in time. Unlike conventional methods that impose tight restrictions on how return predictability evolves over time, we do not need to take a stand on the return generating process. Instead, our approach lets the data determine both how large any predictability is at a given point in time and how long it lasts.[2]

Using this approach, we present new empirical evidence that return predictability is far more concentrated or "local" in time and tends to fall in certain (contiguous) "pockets."For example, using the T-bill rate as a predictor variable over a sixty three year period, our approach identifies eight pockets whose duration lasts between one month and one and a half years. In total, eleven percent of the sample–or roughly twice as much as would be expected by random chance for a test with a size of 5%–is spent inside pockets with return predictability.

To quantify the amount of "local" return predictability and to calibrate what amount of predictability to expect under conventional asset pricing models, we introduce an integrated R^2 (IR^2) measure which is the sum of local R^2 estimates within a particular pocket. This measure allows us to explore if the evidence on return predictability identified by our non-parametric approach is consistent with random variations generated under the null of no return predictability (constant expected returns) or under a time-varying risk premium model with constant coefficients. In particular, we bootstrap stock returns from these types of models and compare the estimated values of the IR^2 measure in the simulations to the values observed in the actual data. We find that both the constant expected return and the time-varying risk premium models fail to match the amount of

---

[1]For early studies, see, e.g., Campbell (1987), Fama and French (1988, 1989), Keim and Stambaugh (1986), and Pesaran and Timmermann (1995). Lettau and Ludvigsson (2010) and Rapach and Zhou (2013) review the extensive literature on return predictability. Paye and Timmermann (2006), Rapach and Wohar (2006), and Chen and Hong (2012) find evidence of model instability for stock market return prediction models, while 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 model predictability in stock returns.

[2]Studies such as Henkel et al. (2011), Dangl and Halling (2012), and Johannes et al. (2014) propose models with time-varying coefficients. However, these studies introduce strong parametric assumptions about changes in the return generating model using regime switching or time-varying parameter models.

return predictability observed in the longest pockets identified in the actual return data, although they can match predictability for the shortest pockets. We conclude from this evidence that the conventional constant-coefficient linear return predictability model fails to generate time-variation in expected returns that is consistent with the empirical evidence we observe.

Having quantified return predictability, we next address "why" stock returns appear to be locally predictable.[3] We make two contributions to the debate on what generates predictability in stock returns. First, we provide a new theoretical result which shows that linear constant-coefficient return predictability models are consistent with a broad class of affine asset pricing specifications in common use, including models that allow for time-varying volatility and compound Poisson jumps. Since our simulations show that this type of model fails to match the empirical evidence, this lends less support to a time-varying risk premium explanation of local return predictability.

Second, we propose an alternative explanation of return predictability. Stock prices depend on expected cash flows that occur in the distant future and so are surrounded by considerable uncertainty. The high sensitivity of aggregate stock prices to even minor variations in investor beliefs about future cash flow growth rates means that incomplete learning about cash flows could be an important source of return movements.[4] Consistent with this intuition, we show that a new type of cash flow learning dynamics can generate return predictability patterns that look like time-varying risk premia in a setting where, by construction, the risk premium is constant.

Building on the predictive systems model of Pastor and Stambaugh (2010), we assume that the cash flow process can be decomposed into a persistent, unobserved component that tracks expected cash flows and a temporary shock that is not predictable.[5] Although the true expected cash flow process is unobserved, investors observe a state variable that is correlated with variation in the persistent component in expected cash flows and thus can be used to predict future cash flows.

Generalizing the predictive systems approach, we allow both the drift in the expected cash flow process as well as its correlation with the observed state (predictor) variable to undergo discrete changes captured through a regime switching process. For commonly used predictor variables such as the T-bill rate and the term spread, it is plausible to expect that the extent to which these variables are informative over future cash flows will vary over time and depend on the underlying

---

[3]Answers to this question face the challenge of the joint hypothesis testing problem, i.e., without positing a fully specified asset pricing model, it is not possible to determine if return predictability is due to a time-varying risk premium or due to market inefficiencies. Balvers, Cosimano, and McDonald (1990), Bansal and Yaron (2004), Campbell and Cochrane (1999), and Cecchetti, Lam, and Mark (1990) present models in which return predictability is consistent with market efficiency.

[4]In a model with paradigm shifts, Hong, Stein and Yu (2007) find that investors learning about the underlying model that generates dividends can give rise to predictable variation in returns and help to match volatility and skewness patterns in returns. In their analysis, investors switch between models that are under-dimensioned representations of the true dividend generating process.

[5]A key difference to Pastor and Stambaugh (2010) is that we model the unobserved component in expected cash flows and use an asset pricing model to study its implications for prices and returns. Instead, Pastor and Stambaugh directly model the dynamics in expected returns and use economic arguments to constrain the sign of the correlation between innovations in the predictive system. As these constraints do not apply to the cash flow process, they are not imposed in our analysis.

monetary policy regime.

We use our regime switching predictive systems model to compare two scenarios. In the first "no-learning" scenario investors observe the regime process underlying the cash flow process. In the second "learning" scenario investors do not observe the underlying regime and so have to recursively update their estimates of the state probabilities using information on returns and the predictor variable to track the state of the economy.

Next, we simulate asset prices under the no-learning and learning scenarios. By construction, the ex-ante risk premium is constant in these simulations. We find that the no-learning model cannot match the empirical evidence on return predictability pockets in the historical data, particularly the presence of long-lived pockets with considerable amounts of return predictability. In contrast, the model with incomplete learning about cash flow growth is capable of generating pockets with similar return predictability characteristics as those we observe in the actual returns data. These simulations suggest that investors' learning about the underlying cash flow process can induce patterns that look, ex-post, like local return predictability even in a model in which ex-ante expected returns are constant.

To focus on the effect on return predictability of incomplete learning about cash flow growth, our simple learning model assumes that risk premia are constant. However, as suggested by authors such as Veronesi (2000), in practice it can be difficult to distinguish between pure learning and risk premium stories as investors' learning may itself command a risk (uncertainty) premium. It is also likely that such risk premia could compound the learning effects we document here. Hence, our results should not be interpreted to exclude time-varying risk premia as an explanation of return predictability that is local in time. Rather, they illustrate the extent to which investors' learning about cash flow growth can produce predictability patterns consistent with what we find in the data.

Authors such as Schwert (2003), Green, Hand, and Soliman (2011), and McLean and Pontiff (2016) also find evidence that return predictability patterns can be learned away over time. These papers suggest that the strength of the evidence of return predictability obtained from time series or cross-sectional regressions tends to weaken as the knowledge of such patterns becomes more widespread. A plausible mechanism is that investors' attempts at exploiting predictive patterns leads to their self-destruction as new money flows into undervalued assets or out of overvalued assets. Our model offers a mechanism for explaining how these effects unfold. We assume that investors learn about the cash flow process and the asset price is derived endogenously as a function of investors' expectations about discounted cash flows. We use our model to quantify how long it takes for this cash flow learning mechanism to be completed to the point where no additional return predictability is detectable and we characterize the amount of return predictability that is present in the interim.[6]

---

[6]We distinguish between learning about a fixed number of parameters–which eventually (asymptotically) will reveal the true value of the parameters–and incomplete learning for which agents will never learn the true value. The

Studies such as Pesaran and Timmermann (1995) and Welch and Goyal (2008) emphasize the distinction between in-sample and out-of-sample return predictability. In-sample return predictability uses full-sample information to estimate model parameters and so could not have been exploited by investors in "real time", while out-of-sample return predictability impose the constraint that only information that was available at a given point in time could be used to generate return predictions. Along with much of the literature, our main analysis is concerned with in-sample return predictability but we also analyze out-of-sample return predictability using a one-sided kernel estimator and a simple scheme for identifying pockets of return predictability in "real time". We find that our daily out-of-sample return forecasts with time-varying predictors are marginally more accurate than the simple "prevailing mean" benchmark of Goyal and Welch (2008) inside the pockets, whereas they perform significantly worse than this benchmark outside the pockets.

It is worth highlighting some key differences between our analysis and earlier studies. Our analysis uses daily stock market returns. This differs from existing studies of return predictability which generally use monthly, quarterly, or annual returns. Using daily stock market returns enables us to more accurately date the timing of pockets with local return predictability which is likely to be missed by returns sampled at monthly or longer horizons. This also introduces interesting econometric issues which we further discuss in the paper.

There are also key differences between our findings of local return predictability pockets and earlier empirical evidence. For example, Henkel et al. (2011), Dangl and Halling (2012), and Rapach, Strauss, and Zhou (2010) argue that return predictability is closely linked to the economic cycle. Although there exists a link between economic recessions and return predictability pockets, we find that this link is weak and the stage of the economic cycle only explains a very small part of the time-variation in expected returns that we document.

The rest of the paper proceeds as follows. Section 2 discusses conventional approaches to modeling return predictability, establishes the class of affine asset pricing models consistent with conventional constant-coefficient return predictability regressions, and introduces our nonparametric methodology for identifying pockets with return predictability. Section 3 introduces our daily data and presents empirical evidence on return predictability pockets using a variety of predictor variables from the literature on return predictability. This section also uses simulations to address whether the pockets could be generated spuriously as a result of the repeated use of correlated tests for local return predictability and conducts a variety of robustness tests. Section 4 introduces our Markov switching predictive systems model for cash flows and presents evidence on the extent to which incomplete learning about cash flows can generate return predictability pockets that are similar to those found in the data. Section 5 discusses possible alternative explanations and sources of return predictability pockets and explores out-of-sample return predictability inside and outside

---

latter situation arises in settings with a latent state whose dimension increases with the time period. Learning about an unobserved state variable is an example of incomplete learning since the dimension of the state vector increases with the sample size and so the current state cannot be consistently estimated.

of ex-ante identified pockets. Section 6 concludes. Two appendices contain additional technical material.

2 Prediction Models and Estimation Methodology

In this section we derive a theoretical result that establishes the class of asset pricing models that is consistent with the benchmark linear regression specification commonly used in empirical studies of return predictability. Next, we introduce the alternative non-parametric regression methodology that we use to measure and quantify time variation in return predictability.

2.1 Return Prediction Model with Constant Coefficients

We start by establishing a set of conditions under which the conventional constant-coefficient return prediction model holds almost exactly within a fairly general endowment economy which nests many canonical specifications considered in the literature. We parameterize cash flow risks and investor preferences in the economy, allowing for time variation in either the quantity or the price of risk. To this end, let z_t be an $L \times 1$ vector of state variables capturing the aggregate state of the economy. We assume that this evolves according to the following law of motion:

Assumption 1 *The aggregate state of the economy follows a stationary VAR process:*

$$z_{t+1} = \mu + F z_t + \epsilon_{t+1}, \tag{1}$$

with z_0 given, where the $L \times L$ matrix F has all of its eigenvalues inside the unit circle and $E[\epsilon_{t+1}] = 0$. Moreover, the log of aggregate dividend growth, Δd_{t+1}, equals $S_d' z_{t+1}$ for some $L \times 1$ vector S_d.

Assumption 1, which is quite standard, states that aggregate dividend growth can be captured by a linear combination of the elements of a finite-dimensional, stationary vector autoregressive process, z_t. We will place further restrictions on the vector of innovations below.

In addition to the restrictions on the cash flow process in Assumption 1, we put restrictions on investor preferences. In particular, Assumption 2 will impose that the log risk-free rate and pricing kernel are "essentially" affine functions of the z_t vector that summarizes the aggregate state of the economy, possibly with time-varying prices of risk.

Assumption 2 *The continuously compounded risk-free rate, $r_{f,t+1}$, satisfies*

$$r_{f,t+1} = A_{0,f} + A_f' z_t, \tag{2}$$

and the continuously compounded return on any financial asset, $r_{a,t+1}$, satisfies the Euler equation

$$1 = E_t[\exp(-\Lambda_t'\epsilon_{t+1} - \log E_t \exp[\Lambda_t'\epsilon_{t+1}] + r_{a,t+1} - r_{f,t+1})] \tag{3}$$

where Λ_t is an $L \times 1$ vector of risk prices.

A large class of models have risk-free rates and pricing kernels which fit into this class. For example, Assumption 2 holds approximately in a representative agent model where agents have Epstein and Zin (1989) preferences when aggregate consumption growth is also an affine function of the state vector.[7] Thus, our results will apply to many of the specifications considered in the literature on consumption-based asset pricing models with long-run risks and rare disasters. This property also holds in an incomplete markets setting with state-dependent higher moments of uninsurable idiosyncratic shocks.[8] We also allow, with some restrictions discussed below, for time-variation in the price of risk, Λ_t, which enables our results to nest many models which have been used to characterize the term structure of interest rates as well as the log-linearized stochastic discount factor of the Campbell-Cochrane habit formation model.

Finally, we provide two alternative sets of restrictions on risk prices and quantities which ensure that, up to a log-linear approximation, price-dividend ratios and market returns are exponential affine functions of z_t.[9] We also define a partition of the set of state variables z_t in a way which will be useful later.

Assumption 3 *Partition the state vector $z_t = [z_{1t}', z_{2t}']'$, where $dim(z_{1t}) = L_1 \le L$. One of the following sets of conditions is satisfied:*

1. *Risk prices are constant: $\Lambda_t = \Lambda$. In addition, for any $\gamma \in \mathbb{R}^L$, the conditional Laplace transform of ϵ_{t+1} satisfies*

$$\log E_t[\exp(\gamma'\epsilon_{t+1})|z_t] = f(\gamma) + g(\gamma)'z_{1t}, \tag{4}$$

   where $f(\gamma)\colon \mathbb{R}^L \to \mathbb{R}$ and $g(\gamma)\colon \mathbb{R}^L \to \mathbb{R}^{L_1}$

2. *Risk prices satisfy $\Lambda_t = \Lambda_0 + \Lambda_1 z_{1t}$, where Λ_1 is an $L \times L_1$ matrix, and $\epsilon_{t+1} \overset{iid}{\sim} MVN(0, \Sigma)$, where Σ is a positive semi-definite matrix.*

---

[7]See, e.g., Bansal and Yaron (2004), Hansen, Heaton, and Li (2008), Eraker and Shaliastovich (2008) and Drechsler and Yaron (2011).

[8]See, e.g., Constantinides and Duffie (1996), Constantinides and Ghosh (2017), Schmidt (2016), and Herskovic et al (2016).

[9]Note that we can get exact exponential affine expressions for the price-dividend ratio and returns of dividend strips–i.e., the value as of time t of a dividend paid at time $t + k$ for any k–and returns. The linearization is only necessary because the market return is a weighted average of these individual dividend strip returns which is not exactly affine in the state vector. Some authors, such as Lettau and Wachter (2011), have elected to work with the exact dividend strip formulas.

Assumption 3 characterizes two sets of assumptions which are commonly made to get affine valuation ratios. In the first case, we assume that risk prices are constant but risk quantities are time-varying. z_{1t} is the subset of variables (e.g., stochastic volatility and/or Poisson jump intensities) that are useful for predicting the quantity of risk, while z_{2t} contain additional variables useful for predicting cash flows or the risk-free rate. We have summarized our main restriction on the distribution of ϵ_{t+1} in terms of its cumulant generating function, which is the logarithm of its moment generating function. The affine structure greatly facilitates analytical tractability and is satisfied for a wide class of distributions used in the theoretical asset pricing literature.[10] For instance, suppose that $\epsilon_{t+1} \sim MVN(0, \sigma_t^2 \Sigma)$ for some positive semi-definite matrix Σ. Then $f(\gamma) = 0$ and $g(\gamma)' z_{1t} = \frac{1}{2} \gamma' \Sigma \gamma$ with $z_{1t} = \sigma_t^2$.

In the second case, we allow for risk prices to be affine in a subset of the state variables, z_{1t}, but restrict the innovations ϵ_{t+1} to be homoskedastic and multivariate normally-distributed.[11] In this case, z_{1t} indicates the subset of variables which characterize time-variation in the price of risk Λ_t. These assumptions are quite common in the bond pricing literature as well as for models featuring time-varying risk aversion and are identical to those in Lustig, van Nieuwerburgh, and Verdelhan (2013), among others.

To solve for asset prices in this economy, we apply the Campbell and Shiller (1988) log-linearization of the stock market return, $r_{s,t+1}$, in excess of the risk-free rate, $r_{f,t+1}$, as a function of the log-dividend growth rate, Δd_{t+1}, and the log price-dividend ratios at time $t+1$ and t, pd_{t+1} and pd_t:

$$r_{s,t+1} \approx c + \Delta d_{t+1} + \rho \cdot pd_{t+1} - pd_t. \tag{5}$$

Here c and $\rho < 1$ are linearization constants. Using this linearization and assumptions 1-3, we can show the following result:

Proposition 1 *Suppose Assumptions 1, 2, and 3 are valid and that a solution exists to the log-linearized asset pricing model. Then, the following properties are satisfied*

(i) The market price-dividend ratio is $pd_t = A_{0,m} + A_m' z_t$;

(ii) The expected excess return is $E_t[r_{s,t+1}] - r_{f,t+1} = \beta_0 + \beta' z_{2t}$,

where $A_{0,m}, A_{0,f}, \beta_0$ are scalars and $A_m \in \mathbb{R}^L$ and $\beta \in \mathbb{R}^d$.

Part (i) of Proposition 1 shows that the log price-dividend ratio is an affine function of the aggregate state vector, which immediately implies that the log-linearized market return is also

---

[10]For example, the property holds for affine jump-diffusion models, e.g., Eraker and Shaliastovich (2008) and Drechsler and Yaron (2011). In these models, ϵ_{t+1} is the sum of Gaussian and jump components and the variance-covariance matrix for the Gaussian shocks and the arrival intensities for the jump shocks are affine functions of z_t. See also Bekaert and Engstroem (2017) and Creal and Wu (2016) for alternative stochastic processes with affine cumulant generating functions.

[11]Creal and Wu (2016) provide some restrictions which permit both risk prices and quantities to vary while keeping valuation ratios in the affine class. We do not detail these assumptions here, but note that the constant coefficient result should obtain for this more general case as well.

an affine function of z_{1t} and ϵ_{t+1}. Part (ii) of the proposition characterizes the extent of return predictability. In particular, it shows that risk premia–expected log excess returns–are an affine function of z_{2t}, variables used to forecast cash flows and the risk free rate. The expressions for β may be found in Appendix A. For a set of predictors x_t chosen to be elements of the underlying state variables (z_{2t}), Proposition 1 justifies using linear return prediction models of the form

$$r_{s,t+1} - r_{f,t+1} = \beta_0 + x_t'\beta + \varepsilon_{t+1}, \tag{6}$$

where x_t is a $(d \times 1)$ vector of covariates (predictors), and ε_{t+1} is an unobservable disturbance with $\mathbb{E}\left[\varepsilon_{t+1}|x_t\right] = 0$. Thus, our result can be used to motivate why a large empirical literature summarized in Goyal and Welch (2008) and Rapach and Zhou (2013) studies predictability of stock returns using the constant-coefficient model in (6).

Part (ii) of Proposition 1 also indicates the extent to which the theory allows for some degree of dimension reduction. In principle, one could allow for a very large number of state variables to predict cash flow growth, each of which could have innovations which may even be priced. Nonetheless, if these variables do not predict time-variation in the quantity of risk (under the conditions of Assumption 3, part 1) or the price of risk (under the conditions of Assumption 3, part 2), they may safely be omitted from the predictive regression. On the other hand, if the true state variables z_{1t} are not spanned by the choice of predictors, x_t, included in the return regression, as could be the case if there are additional drivers of risk prices or quantities omitted from the regression, it need not necessarily be the case that the projection of $r_{s,t+1} - r_{f,t+1}$ on the empirical proxies would have constant coefficients.

Below, we examine the extent to which β is constant in various return prediction models. As is the case for many asset pricing tests, we can only test the joint hypothesis that the model is correctly specified (i.e., we have the correct predictors) and the theoretical restrictions (constant coefficients). Thus, an important caveat on our results is that any evidence we provide which is inconsistent with the constant coefficient null could potentially be explained by omitted factors, as opposed to the learning story we provide below.

2.2 Nonparametric Identification of Pockets

The assumption of constant regression coefficients in the linear return regression (6) has been challenged in numerous studies such as Paye and Timmermann (2006), Rapach and Wohar (2006), Chen and Hong (2012), Dangl and Halling (2012), and Johannes, Korteweg, and Polson (2014), all of which find strong statistical evidence that this assumption is empirically rejected for U.S. stock returns using standard predictor variables.

Define the excess return of the stock market relative to the risk-free rate as $r_{t+1} \equiv r_{s,t+1} - r_{f,t+1}$. Following insights from these studies, we generalize (6) to allow for time-varying return

predictability of the form:

$$r_{t+1} = x_t'\beta_t + \varepsilon_{t+1}, \tag{7}$$

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 $\sigma_t^2 \equiv \mathbb{E}\left[\varepsilon_t^2|x_t\right] = \sigma^2\left(x_t\right)$. The constant coefficient model in (6) is obtained as a special case of (7) when $\beta_t = \beta$ for all t. To economize on notation, here and in the remainder of the paper 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). We want to use an approach that is valid regardless of whether the linear return prediction model in (6) 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, e.g., 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 "anatomy" of the pockets, e.g., the duration and frequency of pockets and the amount of return predictability inside the pockets. Such characteristics can provide important clues about the economic sources of return predictability.

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 $\beta_t = \beta\left(\frac{t}{T}\right)$.[12]

Appendix B provides details on how we implement the nonparametric analysis. Specifically, we use a local constant model to compute the estimator of β_t as

$$\hat{\beta}_t = \arg\min_{\beta_0 \in \mathbb{R}^d} \sum_{s=1}^{T} K_{hT}\left(s-t\right)\left[r_{t+1} - x_s'\beta_0\right]^2. \tag{8}$$

The weights on the local observations get controlled through the kernel $K_{hT}\left(u\right) \equiv K\left(u/hT\right)/\left(hT\right)$, where h is the bandwidth. The estimator in (8) can be viewed as a series of weighted least squares regressions with Taylor expansions of α around each point t/T. The weighting of observations in (8) can be contrasted with the familiar rolling window estimator which uses a flat kernel that puts equal weights 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 $K_{hT}(s-t) = 0$. A weakness of the conventional rolling window approach is that it assigns the same weight to local observations, making it less suited for picking up time variation in α if the build-up and disappearance of such patterns is more gradual, as we might expect a priori.

To identify periods with return predictability ("pockets"), we need a decision rule for determin-

---

[12]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 $T \to \infty$.

ing what constitutes significant return predictability. To this end we compute asymptotic standard errors for the local slope coefficients, $\hat{\beta}_t$, and evaluate their statistical significance. For the estimator of a particular ordinate $\hat{\beta}_t$, the estimated asymptotic variance-covariance matrix is given by

$$\hat{\Sigma}_{\beta,t} = \frac{\kappa_2}{hT} \left(\sum_{s=1}^{T} K_{hT}\left(s - t\right) \hat{e}_s^2 \right) \left(\sum_{s=1}^{T} K_{hT}\left(s - t\right) x_s x_s' \right)^{-1}, \tag{9}$$

where $\hat{e}_s = r_s - x_{s-1}' \hat{\beta}_{s-1}$ is the residual at time s obtained from the nonparametric regression and $\kappa_2 \equiv \int_0^1 K^2\left(u\right) du$ is an integration constant. The limiting distribution of $\hat{\beta}_t$ is normal and thus a valid $100\gamma\%$ pointwise confidence interval for the i_{th} element of $\hat{\beta}_t$, $\hat{\beta}_{i,t}$, is given by

$$\left[\hat{\beta}_{i,t} - q_{(1-\gamma)/2} \hat{\Sigma}_{\beta,t}^{1/2}\left(i,i\right), \hat{\beta}_{i,t} + q_{(1-\gamma)/2} \hat{\Sigma}_{\beta,t}^{1/2}\left(i,i\right) \right], \tag{10}$$

where $q_{(1-\gamma)/2}$ is the $(1-\gamma)/2$ quantile of the standard normal distribution.

We quantify the degree of local return predictability through the local R^2 measured at time t, R_t^2:

$$R_t^2 = 1 - \frac{\sum_{s=1}^{T} K_{hT}\left(s - t\right) \hat{e}_s^2}{\sum_{s=1}^{T} K_{hT}\left(s - t\right) y_s^2}, \tag{11}$$

To identify local variations in the regression coefficients of our model (7), we use a two-sided Epanechnikov Kernel and an effective sample size of one year, i.e., six months of data before and six months after each observation. The Epanechnikov Kernel function has an inverted parabola shape and takes the form

$$K(u) = \frac{3}{4}\left(1 - u^2\right) 1\left\{|u| \leq 1\right\}. \tag{12}$$

Thus, for each day in the sample, we nonparametrically estimate the return prediction model in (7) after trimming the first and last six months of the data. At each point we test if the local slope coefficient is significantly different from zero (using a two-sided test), assigning a value of unity to the pocket indicator $\mathcal{I}_t = 1\{|\hat{\beta}_t / se(\hat{\beta}_t)| > c\}$, where c is a cutoff value that determines the size of the test.

The overlap in adjacent kernel weighting schemes for nearby dates t, t' yields a sequence of highly correlated test statistics. Moreover, repeatedly calculating the pocket test statistic multiple times can be expected to generate false rejections that might identify spurious evidence of return predictability. We address this concern in Section 3.3 by simulating from different data generating processes for returns and calculating to what extent different models can match the characteristics of the pockets of predictability identified by our methodology.

11

2.3   Measuring Pocket Characteristics

Pocket characteristics are measured in a variety of ways. At the most basic level, we want to know how many contiguous pockets our procedure detects. We refer to this as N_p and define it as the number of times we observe a shift in the pocket indicator from zero to one so that, in a sample with T observations, $N_p = \sum_{t=1}^{T-1}(1 - \mathcal{I}_t)\mathcal{I}_{t+1}$. Second, we want to know how long the pockets last. To this end, let $\mathcal{I}_{jt} = 1$ for time-series observations inside the jth pocket, while $\mathcal{I}_{jt} = 0$ outside pockets. Letting t_{0j} and t_{1j} be the start and end date of the jth pocket, the duration of pocket j, Dur_j, is given by

$$Dur_j = \sum_{\tau=1}^{T} \mathcal{I}_{j\tau} = t_{1j} - t_{0j} + 1, \quad j = 1, ..., N_p. \tag{13}$$

We characterize the distribution of pocket durations by reporting the mean, minimum, and maximum durations and also report the fraction of observations inside a pocket, i.e., $\sum_{j=1}^{N_p} Dur_j / T$. We would expect it to be easier for investors to detect and exploit long-lived pockets as the power of tests for the presence of pockets grows with the length of the pocket.

Pocket durations do not quantify the total amount of predictability which accounts for both the duration and the magnitude of the local predictability. This matters a great deal because investors are more likely to identify local return predictability if the R^2 is high. We capture the total amount of return predictability inside a pocket by means of the integral R^2 measure which, for the jth pocket, is defined as

$$IR_j^2 = \sum_{\tau=t_{0j}}^{t_{1j}} R_\tau^2 = \sum_{\tau=1}^{T} \mathcal{I}_{j\tau} R_\tau^2. \tag{14}$$

Visually, this measure captures the area marked under a time-series plot of the local R_τ^2 values in (11), summed across each of the pocket indicators. We report the mean and maximum values of IR_j computed across the pockets $j = 1, ..., N_p$.

Pockets are more detectable either when the degree of predictability within a pocket is very high, possibly for a brief period of time, or when a pocket lasts long, even with low average predictability, or both. By combining the duration of a pocket with the magnitude of the predictability inside this pocket, the integral R^2 measure provides both economic insights into how much predictability is present as well as the possibility that investors can detect and exploit this predictability.[13]

---

[13]Note the analogy to the integral R^2 measure from the literature on breakpoint testing which finds that tests for breaks cannot easily distinguish between frequent, but small breaks to parameters versus rare, but large breaks that move the parameters by the same distance over a particular sample, see, e.g., Elliott and Muller (2006).

3 Empirical Results

This section introduces our data on stock returns and predictor variables, presents empirical evidence from applying the non-parametric approach to identifying local return predictability pockets and, finally, tests whether this evidence is consistent with the affine class of asset pricing models described in Section 2.

3.1 Data

Most studies on predictability of stock returns use monthly, quarterly, or annual returns data. However, since we are concerned with local return predictability which may be of a relatively short-lived nature, we use daily data on both stock returns and the predictor variables. Data observed at the standard frequencies are likely to miss episodes with return predictability at times when the slope coefficients (β_t) change relatively quickly and will not allow us to accurately date the timing of such episodes.

Following studies such as Goyal and Welch (2008), Dangl and Halling (2012), Johannes et al. (2014), and Pettenuzzo, Timmermann, and Valkanov (2014), our main empirical analysis considers univariate prediction models that include one time-varying predictor at a time, i.e., $r_{t+1} = x_t \beta_t + \varepsilon_{t+1}$. 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 also discuss multivariate extensions at the end of the section.

In all our return regressions, the dependent variable is the value-weighted CRSP US stock market return minus the one-day return on a short T-bill rate. Turning to the predictor variables, we consider four variables that have been used in numerous studies on return predictability and are included in the list of predictors considered by Goyal and Welch (2008). First, we use the lagged dividend yield, 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 3-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.[14] Finally, we also 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 12/31/2016 for all series. However, the beginning of the data samples varies across the four predictor variables. Specifically, it begins in 11/4/1926 for the dividend yield (23,786 observations), 1/4/1954 for the 3-month T-bill rate (15,860 obs.), 1/2/1962 (13,846 obs.) for the term spread, and 1/15/1927 (23,727 obs.) for the realized variance.

---

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

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 6-month moving average which is a common procedure for variables such as the nominal interest rate even at longer horizons such as monthly data, see, e.g., Ang and Bekaert (2007). However, we found that the results do not change very much due to this type of detrending and so go with the simpler approach of using the raw data. In practice, 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 1 confirms that this is indeed the case. The table shows full-sample coefficient estimates obtained from the linear regression model in (6) along with t-statistics and R^2 values. Only the regressions that use the T-bill rate (t-statistic of -2.77) and the term spread (t-statistic of 2.32) 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 $4.29 \times 10^{-4}\%$ for the realized variance measure to a maximum of 0.053% (i.e., 0.00053) for the regression that uses the T-bill rate as a predictor.

Campbell and Thompson (2008) suggest comparing the R^2 of return regressions such as (6) to the squared Sharpe ratio of returns to get a measure of the economic value of return predictability. For our daily data, the Sharpe ratio is 0.0255 and so the squared Sharpe ratio is $S^2 = 0.00065$. Using equations (13) and (14) in Campbell and Thompson (2008), the in-sample R^2 value for the dividend yield regression translates into a gain of 0.27% in the return of a mean-variance investor with a coefficient of risk aversion of three or, equivalently, a 12% proportional increase in the investor's utility.[15] Even ignoring the fact that these are in-sample estimates and omit any transaction costs (and trading limits) associated with exploiting the prediction signals, this shows that there would not have been great economic benefits to investors from exploiting daily return predictability from the dividend yield. Notably bigger values are seen for the regression based on the T-bill rate for which the R^2 value of 0.053 translates into an increase in the expected return of 1.8% per annum, assuming again a coefficient of risk aversion of three. We emphasize again that these are not feasible gains and instead should be viewed as an upper bound on the economic value of the daily return predictability signals from the constant coefficient regression model.

For each of the predictor variables, Figures 1-4 provide graphical illustrations of the pockets identified by our nonparametric procedure. When estimating the time-varying coefficient models, we standardize the excess returns and the predictor variables by subtracting their respective means and dividing by their standard deviations. All coefficient magnitudes should therefore be interpreted in standard deviation units. However, since the standard deviation of daily returns (in percentages) is very close to 1 already, the y-axis can roughly be interpreted in daily return percentages. The

---

[15]These numbers are computed by comparing the expected return of an investor with access to the (in-sample) predictions relative to the return of the same investor who assumes a constant expected return.

top panel in each figure plots time series of non-parametric kernel estimates of the local slope coefficient $(\hat{\beta}_t)$ from regressions of daily excess stock returns on the lagged predictors. Dashed lines surrounding the solid line represent plus or minus two standard error bands, calculated using equation (9). The bottom panel in each figure plots the local R^2 measure against time. Shaded areas underneath the local R^2 curve represent the integral R^2 measure (14) for periods identified as pockets of predictability at the 5% significance level. Using a bootstrap simulation methodology described below, areas colored in red represent pockets that have less than a 5% chance of being spurious, areas colored in orange represent pockets that have between a 5% and a 10% chance of being spurious, and areas colored in yellow represent pockets with more than a 10% chance of being spurious. We comment more on this below.

First consider the predictability plots for the dividend yield predictor, shown in Figure 1. The plots for this variable indicate the existence of 13 separate pockets with significant return predictability. The two longest pockets occur during the Second World War and around the Korean War. Moreover, both the frequency and average duration of the pockets has come down over time with only four pockets appearing after 1970 and no pocket showing up in the last 30 years of our sample. For all but two short-lived pockets, the coefficient on the dividend yield is positive inside the pocket. Inside pockets, the R^2 goes as high as 0.035 in the pocket in 1954, but mostly hovers substantially below this level at around 0.01.

For the T-bill rate predictor (Figure 2), we identify eight pockets, only one of which occurs after 1990. Unlike the plots for the dividend yield–and consistent with existing studies such as Ang and Bekaert (2007)–the local coefficient estimates for the T-bill rate are mostly negative, the only exception being the pocket in 1994. The local R^2-values exceed 0.02 during two of these episodes, but are very low during most of the remaining sample, including the period after 2000 which saw low and downward trending interest rates.

The plots for the term spread (Figure 3) identify three pockets–all with positive coefficients– in 1969, 1973-74, and in 1981-82. Interestingly, the last pocket coincides with changes to the Federal Reserve's operating procedures during the monetarist experiment in 1979-1982 which led to significantly higher and more volatile interest rates. The local R^2 is notably higher during these three episodes, ranging between 0.015 and 0.025. Finally, the plots for the realized variance (Figure 4) identify eight pockets. Interestingly, whereas the estimated coefficients on this variable are negative during the four pockets identified in the first half of the sample up to 1960, they switch sign and become positive in the three longest pockets identified in the second half of the sample. The instability in the sign of the coefficient of this predictor is consistent with the difficulty the finance literature has experienced in establishing a consistently positive risk-return trade-off.

3.2   Anatomy of Pockets

Having illustrated the presence of pockets with return predictability, we move on to study the properties of such pockets in more detail for the different predictor variables. To this end, the first five columns in Table 2 show statistics on the number of pockets identified by our methodology. This includes the minimum, maximum, and average pocket lengths and the fraction of the total sample for which a pocket is identified. Results in Panel A use a 5% significance level to identify pockets, while results in Panel B use a 1% significance level.

The length of the pockets varies significantly, even for a given predictor variable. For example, using a 5% significance level to identify pockets (Panel A), the model based on the dividend yield finds a pocket that lasts only 41 days (a little less than two months) while the longest pocket lasts 448 days, or a little less than two years. Similar, if less extreme, variations in pocket length are observed for the other predictor variables. The average pocket duration varies from 177 days (eight months) for the dividend yield variable to 385 days (18 months) for the term spread.

Figures 1-4 show that the number of pockets identified by our approach also varies substantially across predictors–from 13 for the dividend yield model to only three for the term spread. This translates into differences in the proportion of the sample spent inside pockets. For the dividend yield and T-bill rate predictors, 10-11 percent of the sample is spent inside pockets while the pocket frequencies for the term spread (8.5%) and the realized variance (7.7%) predictors are a little lower. In all cases, these numbers are higher than what we would expect by random chance: Since we use a 5% test size and repeat the test multiple times, we should expect to find pockets 5% of the time even under the null of no return predictability.[16]

Comparing the periods spent inside pockets (columns 1-5) to periods spent outside pockets (columns 6-10), we find that the average duration of spells outside pockets is far greater than that spent inside pockets. This is, of course, a reflection of the fact that most of the time (at least 89% of the sample) is spent outside pockets, but the duration measures for the "out-of-pocket" episodes show that there are decade-long periods with no significant return predictability.

Panel B repeats the analysis in Panel A, now using a significance level of 1%. The advantage of using this more stringent level of significance is that it is likely to trigger fewer cases of "spurious" pockets due to the repeated use of the pocket test statistic. Although the number of pockets, as well as their average and maximum length decline from the case with a 5% significance level (Panel A), we see continued strong evidence of pockets even for this more stringent threshold. For the dividend yield, T-bill rate, term spread and realized variance predictors, pockets occupy 3.0%, 3.9%, 4.9%, and 4.0% of the sample, respectively. This is between three and five times higher than the frequency (1%) expected due to the repeated use of the pocket test statistic.

---

[16]In unreported results that use the default spread, i.e., the difference in the daily yield on BAA and AAA-rated bonds, we find a lower-than-expected pocket frequency of only 1.1%. This result is established using a substantially shorter sample than that used for the other predictors and so we do not further pursue the results for this predictor variable here.

Panels C and D in Table 2 report sample statistics on the mean, standard deviation, skewness, kurtosis and persistence of returns inside the predictability pockets identified by our methodology (left columns) as well as outside the pockets (right columns). Focusing on the results based on the 5% cutoff (Panel C), distributions of stock returns inside versus outside the pockets can differ by large amounts. For example, the daily mean return inside the pockets identified by the dividend yield predictor is 6.4 basis points (bps) per day which is more than twice as high as outside the pockets (2.4 bps). Even larger differences are observed for the pockets identified by the T-bill rate and the term spread predictors, for which we observe negative mean returns (-3.0 and -2.6 bps, respectively) in the pockets, but positive means (3.3 and 3.0 bps, respectively) outside the pockets.

Returns inside the pockets also tend to be less volatile (with the exception of pockets identified by the term spread) with positive skews for three of the four predictors (the exception being the realized variance). The positive skews inside pockets contrast with the large negative skews observed outside pockets. Kurtosis is also markedly smaller inside the pockets than outside for three of four variables. This suggests that returns inside the pockets overall have lower risk than during non-pocket periods.[17]

We conclude from these results that return predictability varies significantly over time. Our nonparametric regression approach detects local pockets of return predictability and the return distribution appears to be quite different inside versus outside such pockets. Of course, we have not yet conducted any formal inference on these findings–a topic we turn to next.

3.3 Separating Spurious from Non-spurious Pockets

Because we use a new approach for identifying local return predictability, it is worth further exploring its statistical properties. For example, we are interested in knowing to what extent our approach spuriously identifies pockets of return predictability. Since our approach repeatedly computes 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 issue is whether shorter pockets are more likely to be spurious than the longer ones and whether the degree of return predictability (as measured by the local R^2) inside pockets is consistent with standard models for return dynamics. A third issue is the effect of using highly persistent predictor variables.

To address these questions, we consider two different models for return dynamics. Our simplest model assumes a random walk with a drift for stock prices and so takes the form

$$r_{t+1} = \mu + \varepsilon_{t+1}. \tag{15}$$

---

[17]This finding may in part be mechanical because periods with higher return volatility are less likely to be identified as pockets as the test statistic underlying the pocket indicator may have less power to identify return predictability during such times.

To allow returns to follow a non-Gaussian distribution, we draw the zero-mean innovations, $\hat{\varepsilon}_{t+1} = r_{t+1} - \hat{\mu}$, by means of an i.i.d. bootstrap, where $\hat{\mu}$ is the sample mean of returns. This is clearly not a very good model for daily stock returns, but it serves as a benchmark that allows us to gauge the importance of adding more realistic features of return dynamics.

To account for the pronounced time-varying volatility in daily returns, we estimate a GARCH(1,1) model which has been used extensively to characterize stock market volatility. Moreover, to account for persistence in the regressors, in addition to allowing for volatility dynamics in returns, we incorporate (constant) return predictability from a time-varying state variable, x_t, whose volatility is also time-varying, so that the second model we simulate from takes the form

$$
\begin{aligned}
r_{t+1} &= \gamma x_t + \varepsilon_{rt+1} \equiv \gamma x_t + \sqrt{h_{rt}} u_{rt+1}, \quad u_{rt+1} \sim (0,1), \quad (16) \\
h_{rt+1} &= \omega + \alpha_1 \varepsilon_{rt}^2 + \beta_1 h_{rt}, \\
x_t &= \rho x_{t-1} + \varepsilon_{xt} \equiv \rho x_{t-1} + \sqrt{h_{xt}} u_{xt}, \quad u_{xt+1} \sim (0,1), \\
h_{xt+1} &= \omega_x + \alpha_x \varepsilon_{xt}^2 + \beta_x h_{xt},
\end{aligned}
$$

where u_{rt+1} and u_{xt+1} are mutually independent. The specification in (16) is very flexible: We allow for time-varying volatility both in the return shocks and in the predictor variable and shocks to returns and the predictor variable can be correlated. This constant-coefficient specification nests as a special case the conventional return prediction model used in the empirical literature. Moreover, the GARCH(1,1) model in equation (16) allows for the possibility that local pockets of return predictability could arise due to periods with large variations in the predictor variable, provided that $\gamma \neq 0$.

To simulate from the model in (16), we first estimate the parameters $\gamma, \omega, \alpha_1, \beta_1, \rho, \omega_x, \alpha_x$ and β_x by fitting GARCH(1,1) models to daily values of excess returns and the predictors. Using these estimates, we next construct values of x_t as $\hat{\rho} x_{t-1} + \sqrt{\hat{h}_{xt}} \hat{u}_{xt}$, where $\hat{h}_{xt}$ is the fitted variance of x_t from a GARCH(1,1) model and $\hat{u}_{xt}$ is obtained by bootstrapping (with replacement) from the normalized residuals of the x process. Finally, we construct a series of conditional variances $\left\{ \hat{h}_{rt+1} \right\}_{t=0}^{T-1}$ and obtain normalized residuals $\{\hat{u}_{rt+1}\}_{t=0}^{T-1}$, where $\hat{u}_{rt+1} = (r_{t+1} - \hat{\gamma} x_t)/\sqrt{\hat{h}_{rt+1}}$. Specifically, we construct 1,000 bootstrap samples by first drawing $T+1$ bootstrap residuals $\left\{ \hat{u}_{rt}^b \right\}_{t=0}^{T}$ at random from $\{\hat{u}_{rt+1}\}_{t=0}^{T-1}$ with replacement, then construct a bootstrap sample of excess returns $\left\{ r_{t+1}^b \right\}_{t=0}^{T-1}$ from (16), with $\hat{h}_{r0}^b = \hat{\omega}/(1 - \hat{\alpha}_1 - \hat{\beta}_1)$.

Our simulations follow the empirical analysis and define pockets as periods where the estimated coefficient on the lagged predictor variable is found to be significant at the 5% level. For each bootstrap sample, we record the number of such pockets, along with the minimum, maximum and average values for the pocket duration (measured in days), the R^2 and the integral R^2, described in equations (11) and (14), along with the fraction of time spent inside pockets, measured as a proportion of the full sample.

18

Table 3 shows results for the actual data (first column) and the bootstrapped average, standard errors and p-values–the latter computed as the proportion of simulations that, for each measure listed in a given row, generates a value as large as or bigger than that found in the actual data. Columns two through four assume the simple random walk return generating model in (15), while columns five through seven present results for the GARCH(1,1) model in (16).

First consider the results for the model that uses the dividend yield as a predictor variable (Panel A). On average there are 6.5 pockets in the simulations as compared with 13 in the actual data and this difference is statistically significant: Only 1.5% of the random walk simulations generate at least 13 pockets. The simulations can match the minimum integral R^2 value but, with p-values of 0.045 and 0.002, fail to match the mean and maximum integral R^2 measures. Finally, the fraction of the sample spent inside pockets is 5% in the simulations (as we would expect) which is significantly smaller than the 10% observed for the actual data.

For many of the measures of local return predictability, similar patterns are found for the other predictor variables: Although simulations based on the benchmark specifications in (15) and (16) can generate the same number of pockets as in the original sample and also match the minimum IR^2, they have a much harder time matching the mean or maximum IR^2 values. The evidence is a bit more mixed for the fraction of time spent inside pockets. For this measure, we get p-values of 0.013, 0.113, 0.092 for the T-bill rate, term spread and realized variance predictor variables.

Looking across the different benchmark specifications, it makes very little difference to the results if the random walk with a constant expected return or the GARCH model with a constant slope coefficient is used in the simulations.[18]

Two conclusions emerge from these simulations. First, the overall patterns of return predictability identified by our nonparametric return regressions cannot be explained by either of the return generating models considered here. In particular, since the model in equation (16) 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. Second, the shortest predictability pockets can be due to "chance" as they are matched in many of our simulations. Conversely, neither the model with zero coefficients and constant expected returns (15) or the model with a constant slope coefficient and time-varying volatility (16) comes close to matching the amount of predictability observed in some of the longer-lived pockets.

---

[18]To test if there is evidence of significant time variation in the slope coefficient of the predictors, we also conducted an analysis that defines pockets relative to a constant-coefficient benchmark. Pockets defined in this manner can be thought of as contiguous periods with evidence of significant time variation in the slope coefficient. We detect a similar number of pockets and continue to find that the mean and maximum values of the integral R^2 as well as the fraction of days with a significant pocket indicator, cannot, in most cases, be matched in the simulations. This is evidence of significant time variation in the regression coefficients of the univariate return prediction models and, as shown in Proposition 1, evidence against the class of affine asset pricing models.

19

3.3.1 Analysis of Individual Pockets

We previously discussed the concern that our local, non-parametric approach may detect spurious pockets due to the repeated use of pocket detection tests based on overlapping data. This naturally raises the question whether we can tell if some of the pockets identified by our approach are more or less likely to be spurious. Table 3 shows that simulations with zero or constant return predictability can match some properties of the pocket distribution but fail to match others. This suggests that we can discriminate between spurious and non-spurious pockets by looking at each individual pocket's IR^2 value–a measure found to be hard to match in the simulations–and computing the percentage of simulations with at least one pocket matching this value. This produces an odds ratio with small values indicating how difficult it is to match the total amount of predictability observed for the individual pockets.

Following this idea, for each of the pockets shown in Figures 1-4, Table 4 reports the associated IR^2 measure and the proportion of simulated pockets that generate a value at least as large. First consider the 13 pockets identified by the return prediction model that uses the dividend yield as a predictor (first column). Some of the pockets are highly unlikely to be due to chance–for example, the fourth and seventh pockets generate very high IR^2 values of seven and 11, respectively–with less than 1% of the simulations being able to match these values. Other pockets, notably the first three and the last three pockets, are more likely to be spurious as their integral R^2 values are matched in at least ten percent of the simulations. In total, five of the 13 pockets generate p-values below 5% and so are unlikely to be spurious.

Similarly, six of the eight pockets identified by the T-bill rate regressions generate IR^2 values with p-values less than 10%. Similarly, all three pockets identified by the term spread regressions and five of eight pockets identified using the realized variance appear to be non-spurious at the 5% critical level.

Using the analysis in Table 4, Figures 1-4 mark in red the pockets with less than a 5% chance of being spurious, pockets colored in orange have between a 5% and a 10% chance of being spurious, while pockets colored in yellow have more than a 10% chance of being spurious. As expected, pockets that are more short-lived and have lower peaks in the IR^2 measure are more likely to be deemed spurious.

These results suggest that roughly half of the identified pockets are non-spurious in the sense that the amount of predictability in these pockets cannot be matched by the return prediction models from which we simulate and so we are more confident that these pockets represent periods where returns were genuinely predictable.

3.4 Multivariate Predictions

The results reported so far all use univariate prediction models, but it is of economic interest to see to what extent the pockets identified in this manner are correlated across the different predictors.

A strong positive correlation might suggest that the pockets have a common economic source and represent periods during which the identity of the particular predictor is not too critical. Conversely, weaker correlations are suggestive that the predictable pockets are variable-specific. Note that there might be good reasons for the pocket indicators to be only weakly correlated as the pocket indicators will depend on how informative the individual predictors are for a particular episode with return predictability.

For each of the predictor variables, Table 5 reports estimates of the pairwise correlations between pocket indicators (above the main diagonal) and estimates of the pairwise correlation between local R^2 measures. The correlations are all positive with values ranging from 0.10 to 0.54 for the pocket indicator correlations and values ranging between 0.34 and 0.80 for the local R^2 value.

These findings suggest the presence of a sizeable common component in the return predictability pockets identified by our approach. In the presence of such a common component, using a multivariate regression model could help improve the power to identify episodes with local return predictability. Moreover, our theoretical analysis in Section 2 suggests that the inclusion of more state variables as predictors can bring benefits such as making our results more robust to omitted variable biases.

For these reasons we next extend our approach to a multivariate setting. Multivariate kernel regressions suffer from the curse of dimensionality, so instead of including all four of our predictor variables we consider a bivariate model that includes the T-bill rate and the term spread as predictors. Figure 5 reports results from the multivariate estimation. In order to identify a pocket in the multivariate model, we conduct an F-test of the joint significance of the coefficient vector β_t at every time t. The top panel plots the p-value from this F-test over time with the horizontal black line representing a cutoff of 5%. The shaded gray regions represent periods in which the p-value is less than the 5% cutoff. The bottom panel plots the local R^2 from these nonparametric regressions over time.

In total we find evidence of five pockets, four of which are deemed to be significant at the 5% level using the simulation methodology described earlier. Moreover, the pockets identified by the multivariate model appear to capture the same three periods in the late 60s through the early 80s identified by both the T-bill rate and term spread models. Additionally, the pocket in the mid-90s identified by the T-bill rate model is captured.

3.5   Robustness of Results

We next explore the robustness of our results with regards to our choice of bandwidth, the effect of persistence in the predictor variables, and autocorrelation in returns.

3.5.1 Choice of Bandwidth

Our decision to use a one-year kernel bandwidth to identify local pockets of predictability is not based on any considerations for econometric optimality but reflects our priors for a reasonable length of a predictive pocket. Nevertheless, one might reasonably have argued for the usage of a different bandwidth and so it is important to explore how sensitive our results are to this choice. The trade-offs in choice of the bandwidth are clear: a smaller bandwidth is likely to lead to a more "noisy" determination of pockets, and hence to an increase in the number of pockets, while a larger bandwidth will have the reverse effect. It is less clear how the proportion of the sample identified to be pockets of predictability changes with the bandwidth as the power of our test procedure also depends on the bandwidth.

To explore how the bandwidth changes our results, we compute results that alternatively use bandwidths of 6, 18, and 24 months. Table 6 reports the results with each panel capturing a different predictor variable. For the dividend yield, the number of pockets varies from seven (for a kernel bandwidth of 24 months) to 21 (for a bandwidth of 6 months) and so this measure is quite sensitive to the choice of bandwidth. However, the average length of the pockets is shorter, the smaller the bandwidth. As a consequence, the fraction of the sample spent inside pockets is a far more robust measure that only fluctuates between 7.9% versus 10.4%, compared to the baseline case with a 9.8% significance rate. The mean IR^2 measure is also quite robust, fluctuating between 2.36 and 2.99 for the longest and shortest bandwidth values, respectively.

For the model that uses the T-bill rate as a predictor, using a six-month bandwidth leads to the identification of 15 pockets and a fraction of the sample spent inside pockets that equals 10.8%. This compares with eight pockets and 11% of the sample spent inside pockets for the baseline scenario with a one-year bandwidth. The chief effect of decreasing the bandwidth is again to break up the longer pockets identified by the one-year bandwidth into shorter ones. Increasing the bandwidth to 18 and 24 months has the effect of reducing the number of pockets to seven and five, while the proportion of the sample spent inside pockets increases to 13.8% and 17.5%, respectively.

Figure 6 plots the local R^2 measure along with the pockets identified for the T-bill rate model using bandwidth values of 12 (top panel), 6, 18, and 24 (bottom panel) months. Across these very different choices of bandwidth, pockets in 1957, 1970, 1973, and 1995 are identified. Moreover, the shortest bandwidth (6 months) has a stronger tendency to identify what appears to be spurious, short-lived pockets compared with the other choice of bandwidth. This clearly illustrates how using too small a bandwidth can lead to results that are too noisy.

Similar findings emerge for the term spread and realized variance predictors, although the number of pockets is particularly sensitive to using a short (six month) bandwidth for the term spread variable.

We conclude from these findings that the shorter the kernel bandwidth, the larger the number of pockets identified, but the shorter the average pocket length. Conversely, the longer bandwidth

values tend to identify fewer pockets with a longer duration. While there is a tendency for the mean IR^2 and the fraction of the sample spent inside pockets of predictability to be higher for the largest bandwidth, these measures are more robust than the noisier "number of pockets" measure. Reassuringly, our results on the existence of pockets of return predictability appear to be robust across a wide range of values of the kernel bandwidth and across predictor variables.

3.5.2 Persistence of Predictor Variables

All four of our predictors are highly persistent. This is by no means unique to our setup–our predictors are in common use in the finance literature–but arguably persistence could be more of a concern when dealing with daily return regressions as the predictors become even more persistent at the daily frequency compared with the more common monthly or quarterly frequencies.

To deal with this issue, we explore the robustness of our results by using the first-differenced value, $\Delta x_t = x_t - x_{t-1}$, of the predictors. The bottom row of each panel in Table 6 show the results from using each of the four predictors detrended in this way. If anything, we tend to identify more pockets and classify a higher fraction of the sample as spent inside the pockets for the first-differenced data.[19]

These results again illustrate that the presence of pockets with predictable stock returns does not reflect the high persistence of the predictors although, of course, the transformation of the predictor variables and the econometric test used to measure the pockets may affect the exact location of the pockets.

3.5.3 Autocorrelation in returns

We also explore the sensitivity of our results with regards to the presence of mild autocorrelation in daily stock market returns which may be induced by nonsynchroneous trading and/or market microstructure effects. We emphasize that the serial correlation is very mild–the first-order autocorrelation in daily stock returns is 0.068 for the full sample. Indeed, when we apply our non-parametric approach to the stock return series from which we have filtered out the first-order serial correlation, we find 11, 7, 3, and 7 pockets for the four predictors (dividend-price ratio, T-bill rate, term spread, and realized variance), as compared to the 13, 8, 3, and 8 pockets we found for the original returns data. Moreover, the percentage of the sample taken up by pockets of predictability is very similar to what we find in the original returns.

---

[19]We also considered an alternative way of detrending the predictors, namely by subtracting an exponentially weighted average of past daily values of each predictor, i.e., $\tilde{x}_t = x_t - [\lambda(1-\lambda)/(1-\lambda^p)] \sum_{j=1}^{p}$, where λ ranges between 0.97 and 0.99 and the cutoff, p, is set at one year. This way of detrending the data is appropriate if the regressor follows an integrated moving average process. Again, we found that the presence of pockets is robust to detrending the regressors in this manner.

4 Learning About Cash Flow Growth

This section explores whether the evidence of local pockets with return predictability is consistent with learning dynamics induced by an asset pricing model where expected returns are constant but the cash flow process is partially predictable. We propose a new specification for cash flow dynamics that builds on, and generalizes, the predictive systems approach pioneered by Pastor and Stambaugh (2009). We assume that cash flows consist of an unobserved expected growth component that is highly persistent and a temporary "unexpected growth" shock. This unobserved process is correlated with a set of observable state variables which, through their correlation with expected growth, gain predictive power over future cash flows.

A novel feature of our approach is that it allows the correlation between expected cash flows and predictors to be state dependent. This feature is likely to more accurately reflect the time-varying predictive power of economic state variables over future cash flows. For example, shifts in correlations between term structure variables and cash flows could result from changes in the monetary policy regime. Indeed, the predictive content of interest rates or the term spread over future cash flows is unlikely to be the same under quantitative easing or "zero lower bound" regimes compared to under a more conventional monetary policy regime. We capture this idea by assuming that the underlying predictor is correlated with cash flow growth only in one regime while the correlation is zero in the other regime. These particular assumptions can of course be relaxed, but make our results easier to interpret.

The following subsection introduces our predictive systems model with regime switching.

4.1 A Predictive Systems Model With Regime Switching

We develop a model for the dividend process that captures a small predictable component in cash flows. This is consistent with recent empirical findings such as van Binsbergen and Koijen (2010), Kelly and Pruitt (2013), and Pettenuzzo, Sabbatucci, and Timmermann (2018). Specifically, let $\Delta d_{t+1} = \log(D_{t+1}/D_t)$ be the growth rate in (log-) dividends and assume that this can be decomposed into an expected cash flow component, μ_t, and a purely temporary shock, u_{t+1} :

$$\Delta d_{t+1} = \mu_t + u_{t+1}. \tag{17}$$

We capture persistence in daily cash flow growth by means of an autoregressive component in μ_t. In addition, the mean of the expected cash flow process, μ_t, is affected by a state variable, s_t, which captures discrete shifts to the process and, thus, also can induce persistence in cash flow growth.[20] Finally, expected cash flows are affected by a transitory shock, w_{t+1} :

$$\mu_{t+1} = \mu_{\mu,s_{t+1}} + \rho_\mu \mu_t + w_{t+1}. \tag{18}$$

---

[20]See also David and Veronesi (2013) for a related asset pricing model with regime switching.

While investors do not observe the expected cash flow process, they are assumed to observe a predictor variable, x_{t+1}, that is driven by the same state variable, s_{t+1}, and follows a similar dynamic process:

$$x_{t+1} = \mu_{x,s_{t+1}} + \rho_x x_t + v_{t+1}. \tag{19}$$

We assume that the innovations to the processes in (17) - (19) are normally distributed with mean zero, $(u_{t+1}, w_{t+1}, v_{t+1})' \sim N(0, \Sigma_{s_{t+1}})$, where $\Sigma_{s_{t+1}}$ is a state-dependent variance-covariance matrix:

$$\Sigma_{s_{t+1}} = \begin{bmatrix} \sigma_u^2 & \sigma_{uv} & \sigma_{uw} \\ \sigma_{uv} & \sigma_{v,s_{t+1}}^2 & \sigma_{vw,s_{t+1}} \\ \sigma_{uw} & \sigma_{vw,s_{t+1}} & \sigma_{w,s_{t+1}}^2 \end{bmatrix}. \tag{20}$$

Note that we constrain this covariance matrix to have a particular form since only the variance of the expected cash flow (σ_{w,s_t}^2) and predictor variable (σ_{v,s_t}^2), in addition to their covariance (σ_{vw,s_t}), are state dependent. In contrast, the variance of the purely temporary shocks to dividend growth (σ_u^2), or their correlation with the other shocks in the model $(\sigma_{uv}, \sigma_{uw})$, do not depend on the underlying state variable, s_t. We impose these constraints to ensure that the identified states capture changes to how informative the predictor variable, x_{t+1}, is with respect to the expected value of the cash flow process, μ_{t+1}.

We focus on the case with two states so that $s_t \in \{1, 2\}$ and assume that s_t follows a first-order Markov chain with transitions

$$\pi_{ii} = \mathbb{P}(s_{t+1} = i | s_t = i), \quad i = 1, 2. \tag{21}$$

Moreover, s_t is assumed to be independent of all past, current and future values of $(u_t, w_t, v_t)'$. We collect the state transitions in a 2×2 transition probability matrix Π_s and define the unconditional mean, volatility and ergodic state probabilities

$$\boldsymbol{\mu_{\mu,s}} = \begin{bmatrix} \frac{\mu_{\mu,1}}{1-\rho_\mu} \\ \frac{\mu_{\mu,2}}{1-\rho_\mu} \end{bmatrix}, \quad \boldsymbol{\sigma_{w,s}} = \begin{bmatrix} \sigma_{w,1} \\ \sigma_{w,2} \end{bmatrix}, \quad \bar{\boldsymbol{\pi}} = \begin{bmatrix} \mathbb{P}(s_t = 1) \\ \mathbb{P}(s_t = 2) \end{bmatrix}.$$

Using results from Timmermann (2000), the unconditional mean and variance of the μ_t process are given by

$$E[\mu_t] = \bar{\boldsymbol{\pi}}' \boldsymbol{\mu_{\mu,s}}, \tag{22}$$

$$Var(\mu_t) = \bar{\boldsymbol{\pi}}' \left((\boldsymbol{\mu_{\mu,s}} - \mu_\mu \boldsymbol{\iota}) \odot (\boldsymbol{\mu_{\mu,s}} - \mu_\mu \boldsymbol{\iota}) + \frac{\boldsymbol{\sigma_{w,s}^2}}{1 - \rho_\mu^2} \right). \tag{23}$$

We use these expressions in the simulations and to calibrate the parameters governing dividend growth. In particular, we ensure that the unconditional mean and variance of μ_t are matched. For

any choice of state transition probabilities and the regime 1 mean and conditional variance of μ_t, these expressions imply choices for the regime 2 mean and conditional variance. This is explained in greater detail in section 4.4.

4.2   Filtering of the State Variables

The predictive systems model in (17) - (19) is a nonlinear state space model as it contains a combination of linear and regime-switching dynamics and thus standard Kalman filtering methods cannot be used to filter the states and evaluate the likelihood. To address this issue, we approximate the likelihood function of the model using a discretization of μ_t as proposed in Farmer (2017). We briefly explain how this is done.

In the first step, we construct a discrete approximation to the stochastic process governing the dynamics of the state variables, μ_t and s_t. Because the shocks to the measurement and state equations in (17) - (19) are correlated, the distribution of the state in the next period conditional on the state in the current period depends on the values of the observables and so the transition matrix constructed to approximate the dynamics will be time-varying. We handle this issue as follows. Using properties of correlated normal random variables, we have

$$w_t \,|u_t, v_t, s_t \sim N\left(\mu_{w|t}, \sigma^2_{w|t}\right),$$

where $\mu_{w|t}$ and $\sigma_{w|t}$ are given by

$$\mu_{w|t} = \frac{\left(\sigma_{uw}\sigma^2_{v,s_t} - \sigma_{vw,s_t}\sigma_{uv}\right)u_t + \left(\sigma_{vw,s_t}\sigma^2_u - \sigma_{uw}\sigma_{uv}\right)v_t}{\sigma^2_u\sigma^2_{v,s_t} - \sigma^2_{uv}},$$

$$\sigma^2_{w|t} = \sigma^2_{w,s_t} - \frac{\sigma^2_{uw}\sigma^2_{v,s_t} + \sigma_{vw,s_t}\sigma^2_u}{\sigma^2_u\sigma^2_{v,s_t} - \sigma^2_{uv}}.$$

Next, define a new random variable $\mu_{t,M}$ which takes M discrete values $\left(\mu^1, \ldots, \mu^M\right)$. For a given choice of M, define a grid from the set of M equally spaced points between $E\left[\mu_t\right] - \sqrt{(M-1)\,\mathrm{Var}\left(\mu_t\right)}$ and $E\left[\mu_t\right] + \sqrt{(M-1)\,\mathrm{Var}\left(\mu_t\right)}$. Let each point μ^m be associated with the interval $\left[\underline{\mu}^m, \bar{\mu}^m\right]$ where $\underline{\mu}^1 = -\infty$, $\bar{\mu}^M = \infty$, and $\bar{\mu}^m = \underline{\mu}^{m+1} = \frac{\mu^m + \mu^{m+1}}{2}$ for $m = 1, \ldots, M-1$. Further, construct the transition probabilities for $\mu_{t,M}$ at time t as

$$\mathbb{P}\left(\mu_{t+1,M} = \mu^{m'} \,|\mu_{t,M} = \mu^m, u_{t+1}, v_{t+1}, s_{t+1}\right) = \Phi\left(\frac{\bar{\mu}^{m'} - \mu_{\mu,s_{t+1}} - \rho_\mu\mu^m - \mu_{cond,t+1}}{\sigma_{w|t+1}}\right)$$

$$- \Phi\left(\frac{\underline{\mu}^{m'} - \mu_{\mu,s_{t+1}} - \rho_\mu\mu^m - \mu_{cond,t+1}}{\sigma_{w|t+1}}\right), \quad (24)$$

where Φ is the standard normal CDF. This way of computing transition probabilities of discrete approximations to continuous stochastic processes was first proposed by Tauchen (1986).

In the presence of two variance regimes, we have a total of $2 \times M$ states in the discrete chain. The ordering convention we adopt for the states is

$$
\psi = \begin{bmatrix}
\mu_1 & s_1 \\
\mu_2 & s_1 \\
\vdots & \vdots \\
\mu_M & s_1 \\
\mu_1 & s_2 \\
\vdots & \vdots \\
\mu_M & s_2
\end{bmatrix}
\tag{25}
$$

The filtered state probabilities, $\hat{\xi}_{t|t}$, can therefore be represented by a $(2 \times M) \times 1$ vector whose individual entries refer to the probability of being in each of the state pairs listed in the ψ matrix in (25). Because the innovations to the state and observation equations are correlated, the discrete Markov chain is non-homogeneous with transition matrix at time t given by Π_t. Each element of P_t is defined as the probability of transitioning between a particular pair of states. For example, assuming $M \geq 2$, the $(2, 1)$ element of P_t corresponds to the probability of transitioning from state (μ_1, s_1) at time t to state (μ_2, s_1) at time $t + 1$.

The forecast of next period's state is given by

$$
\hat{\xi}_{t+1|t} = P_t \hat{\xi}_{t|t},
\tag{26}
$$

while the state probabilities are updated recursively according to

$$
\hat{\xi}_{t+1|t+1} = \frac{\hat{\xi}_{t+1|t} \odot \eta_{t+1}}{1' \left(\hat{\xi}_{t+1|t} \odot \eta_{t+1} \right)},
\tag{27}
$$

where 1 is a $(2 \times M) \times 1$ vector of ones and η_{t+1} denotes the joint conditional densities

$$
\eta_{t+1} = \begin{bmatrix}
p\left(u_{t+1}, v_{t+1} \,\middle|\, \mu_t = \mu^1, s_{t+1} = 1\right) \\
\vdots \\
p\left(u_{t+1}, v_{t+1} \,\middle|\, \mu_t = \mu^M, s_{t+1} = 1\right) \\
p\left(u_{t+1}, v_{t+1} \,\middle|\, \mu_t = \mu^1, s_{t+1} = 2\right) \\
\vdots \\
p\left(u_{t+1}, v_{t+1} \,\middle|\, \mu_t = \mu^M, s_{t+1} = 2\right)
\end{bmatrix}.
$$

From the time series $\left\{ \hat{\xi}_{t|t} \right\}_{t=1}^{T}$ we can construct filtered estimates of μ_t as

$$\hat{\mu}_{t|t} = (\psi \iota_1)' \, \hat{\xi}_{t|t} \tag{28}$$

where ι_1 is the first column of the (2×2) identity matrix.

4.3 Asset Prices and Returns

We next develop a simple model for pricing stocks under the assumption that dividends follow the Markov switching predictive systems model in (17) - (19). To this end, we use a simple log-linearized present value model. Following Campbell and Shiller (1988), the logarithm of the approximate present value stock price can be written as

$$p_t = d_t + \frac{c}{1-\rho} + E_t \left[\sum_{j=0}^{\infty} \rho^j [\Delta d_{t+1+j} - r_{t+1+j}] \right], \tag{29}$$

where p_t and d_t denote the log of the stock price and dividends, respectively, Δd_{t+j} and r_{t+j} are the dividend growth rate and (log-) returns in period $t+j$, and c, ρ are (linearization) constants as in equation (5). Under the assumption that expected returns are constant, $E_t[r_{t+j}] = \bar{r}$ for all j and so (29) simplifies to

$$p_t = d_t + \frac{c - \bar{r}}{1-\rho} + E_t \left[\sum_{j=0}^{\infty} \rho^j \Delta d_{t+1+j} \right]. \tag{30}$$

Thus, calculating the stock price in (30) only requires us to compute the expected future dividend growth, Δd_{t+1+j} for $j \geq 0$. Recall from (28) that the filtered state estimate of μ_t is given by $\hat{\mu}_{t|t} = (\psi \iota_1)' \, \hat{\xi}_{t|t}$. To compute an expression for the expected value of future dividend growth, we assume that the transition matrix remains as P_t, which amounts to assuming that agents do not account for the effect of their future learning on prices when projecting future cash flows. Under this assumption, we have

$$E_t[\Delta d_{t+1+j}] = (\psi \iota_1)' \left(P_t^j \hat{\xi}_{t|t} \right). \tag{31}$$

Using (30), and assuming that $(I - \rho P_t)$ is invertible for all P_t, we can compute an expression

for the log stock price

$$p_t = d_t + \frac{c - \bar{r}}{1 - \rho} + \sum_{j=0}^{\infty} \rho^j E_t \left[\Delta d_{t+1+j} \right]$$

$$= d_t + \frac{c - \bar{r}}{1 - \rho} + \sum_{j=0}^{\infty} \rho^j \left(\psi \iota_1 \right)' \left(P_t^j \hat{\xi}_{t|t} \right)$$

$$= d_t + \frac{c - \bar{r}}{1 - \rho} + \left(\psi \iota_1 \right)' \left((I - \rho P_t)^{-1} \hat{\xi}_{t|t} \right). \tag{32}$$

From this equation, we obtain an expression for stock returns:

$$r_{t+1} = c + \rho \left(p_{t+1} - d_{t+1} \right) + d_{t+1} - p_t$$

$$= c + \rho \frac{c - \bar{r}}{1 - \rho} + \left(\psi \iota_1 \right)' \left((I - \rho P_t)^{-1} \hat{\xi}_{t+1|t+1} \right)$$

$$+ \Delta d_{t+1} - \frac{c - \bar{r}}{1 - \rho} - \left(\psi \iota_1 \right)' \left((I - \rho P_t)^{-1} \hat{\xi}_{t|t} \right)$$

$$= \bar{r} + \Delta d_{t+1} + \left(\psi \iota_1 \right)' (I - \rho P_t)^{-1} \left[\hat{\xi}_{t+1|t+1} - \hat{\xi}_{t|t} \right] \tag{33}$$

We use equations (32) and (33) along with the dividend process in (17) - (19) to simulate the state and predictor variable, dividends, stock prices, and stock returns.

4.4 Calibration of Model Parameters

In practice, the daily dividend process is not observed, and so we have to use a proxy for Δd_{t+1}. To this end, we use the ADS index proposed by Aruoba, Diebold, and Scotti (2009). This is a daily business cycle index that is constructed using daily updates to "real" economic variables observed at different frequencies such as weekly payroll figures, monthly industrial production, and quarterly GDP growth. The ADS index is updated daily by the Federal Reserve Bank of Philadelphia and closely tracks the business cycle. The daily ADS time series is highly persistent but is constructed to revert to a mean of zero. As observed by Rossi and Timmermann (2015), the ADS index is a good candidate for picking up a slow-moving component in consumption or dividend growth.[21] We rescale the ADS index so that when it is simulated at a daily frequency, the time-aggregated mean and standard deviation at a yearly frequency are 5.44% and 5.71% respectively. These numbers are the unconditional mean and standard deviation of annual dividend growth computed using dividends on the CRSP value-weighted index over the same sample period as the T-bill rate. As our predictor variable, we use the T-bill yield which, as we saw earlier, captures a number of pockets with return predictability.

---

[21] Rossi and Timmermann (2015) find that the correlation between an economic activity index, constructed using a similar methodology to that used for the ADS index, and growth in real personal, nondurable consumption is 15.4% and 39.7% at the quarterly and annual horizons, respectively.

Values of the calibrated parameters are listed in Table 7. We choose the diagonal elements of the state transition matrix ($\pi_{11} = 0.995, \pi_{22} = 0.9986$) so that the expected duration of the first and second regimes are 200 and 700 days, respectively. The mean of the observable predictor variable (the T-bill rate) in the first regime, $\mu_{x,1}$ is set so that its expectation in this state equals 2.15% while its expectation in the second state ($\mu_{x,2}$), at 5.86%, is calibrated such that its unconditional mean across the two states matches the overall sample mean of the 3-month Treasury bill rate.

We choose a value of the autoregressive parameter for the daily expected cash flow growth process, $\rho_\mu = 0.9953$, which implies an annualized persistence in cash flows of about 0.3. The level of persistence chosen in our simulations is quite a bit lower than that assumed in the literature on "long-run risk", see, e.g., Bansal and Yaron (2004).[22]

Given our choice for ρ_μ, the annualized means of the expected growth process, 8.32% in regime 1 and 4.62% in regime 2, are again calibrated so that the unconditional mean matches the sample mean of the rescaled ADS series. The same is true for the standard deviation of the innovations to x. The standard deviation of the innovations to μ are chosen to be equal across regimes. That is, we impose $\sigma_{w,1} = \sigma_{w,2}$. We also impose that $\sigma_{uv} = 0$ across both regimes, and that $\sigma_{vw,1} = 0$. Lastly, we choose a correlation between shocks to expected cash flows and shocks to the observed predictor variable to be -0.2 in regime 1 and 0 in regime 2, so that the predictor is informative over innovations to cash flow growth in the first but not in the second regime. While the sign of the correlation is not important for the model's ability to generate return predictability pockets, the negative correlation can be thought of as reflecting that higher interest rates tend to be associated with lower real growth.

4.5   Simulation Results

Using the calibrated parameters of the regime switching predictive systems model, we generate simulated data and run the local, non-parametric regressions exactly as in the empirical specification to identify pockets and to see if the characteristics of such pockets match the characteristics of the pockets identified in the actual data. Since we are interested in matching the pocket evidence in the actual data (Table 2), we generate samples whose length match that of the predictor variable.

We show results for the two significance levels (5% and 1%) considered in our study. We compare the simulated statistics to the empirical results from the model that uses the T-bill rate as a predictor variable (second row in Table 2). For the case with no learning, investors are assumed to know the state of the underlying Markov chain and μ_t is set equal to the value corresponding to the interval in which the true continuous value lies. Under incomplete learning, investors do not observe the state, s_t, and instead have to form beliefs about the probability that they are in any

---

[22]For example, the coefficient on the expected growth rate in Bansal and Yaron (2004) is 0.979 for monthly data which translates into roughly 0.999 at the daily frequency, assuming 21 days in a month. Similarly, Bansal, Kiku, and Yaron (2012) analyze a model whose implied persistence of the expected growth rate and of volatility is 0.9988 and 0.99995, respectively, at the daily frequency.

particular (μ_t, s_t) regime.

Before studying the ability of the simulated model to generate predictability pockets, first consider the overall area under the local R^2 curves $(\bar{R}^2)$ displayed in Figures 1-4 as well as the areas above $(\bar{R}^2_+)$ and below $(\bar{R}^2_-)$ the zero line, labeled "positive" and "negative", respectively. Defining indicator variables $I^+_\tau = 1$ if $R^2_\tau \geq 0$, and $I^+_\tau = 0$ otherwise, while $I^-_\tau = 1 - I^+_\tau$, we have

$$\overline{R}^2 \;=\; \frac{1}{T}\sum_{\tau=1}^{T} R^2_\tau,$$

$$\overline{R}^2_+ \;=\; \left(\sum_{\tau=1}^{T} I^+_\tau\right)^{-1} \sum_{\tau=1}^{T} I^+_\tau R^2_\tau,$$

$$\overline{R}^2_- \;=\; \left(\sum_{\tau=1}^{T} I^-_\tau\right)^{-1} \sum_{\tau=1}^{T} I^-_\tau R^2_\tau,$$

If a return model does not match the average R^2, this suggests that it does not generate much predictability. Table 8 shows that the average local $(\bar{R}^2)$ equals 0.41%. This value cannot be matched by the simulations with no learning which generate, on average, a local R^2 of 0.26%, whereas the models with learning easily match this measure (average local R^2 of 0.50%). The key reason for the no-learning model's failure to match the amount of return predictability observed in the data is that it generates too small positive values of the local R^2 (0.30% on average)–something that is not matched in the actual sample (0.53%) which in turn closely lines up with the learning model (0.54%).

Turning to the emergence of return predictability pockets, first consider the results based on the 5% significance level (Panel A in Table 9). The model with no learning generates an average of only 3.9 pockets as opposed to the eight pockets observed in the actual sample and only 3.8% of the sample is spent inside pockets compared to 11.0% in the actual data. The no-learning model also does not get close to matching the values observed in the actual data of the mean or maximum integral $\bar{R}^2$ statistics or the number of pockets.

In sharp contrast, the model with learning dynamics is capable of matching all sample statistics based either on the value of the integral R^2 measure or the length of the pockets. For example, the number of pockets is eight in the sample as compared to an average value of 7.1 in the simulations, and the simulations with learning also match the fraction of the sample spent inside pockets (13.2% versus 11.0%) quite closely. Even the mean and maximum values of the IR^2 are matched in the simulations with learning effects.

These findings carry over to the results that use the 1% significance level to identify pockets (Panel B in Table 9). For example, whereas the model with no learning only generates 0.7 pockets on average, the model with learning generates an average of 3.4 pockets, a number that, while slightly below the four pockets observed in the data, is within sampling error of that number.

The fraction of the sample spent inside pockets with predictability in the actual data (3.9%) is also matched more closely in the simulations with learning (5.8%) than in the simulations without learning (0.5%).

To further illustrate these findings, Figure 7 shows the cumulative distribution functions of the integral R^2 measure under no-learning and learning in the cash flow process of the simulated predictive systems model with regime switching. We also show (as vertical red bars) the IR^2 values identified in the actual data. Clearly the largest values of the IR^2 are extremely unlikely under the no-learning model but much more likely to occur under the model with learning.

We conclude the following from these simulations of our Markov switching predictive systems model. First, in the absence of learning, a model with discrete changes in how informative the observed predictor is over the (unobserved) mean of the cash flow process cannot match the local nature (pockets) of the temporal patterns we observe in return predictability. Second, a model that introduces learning about the underlying state process driving cash flows is capable of generating return predictability pockets with similar features as those observed in the actual data. Significantly, both the number of pockets and the average time spent in pockets is matched by this model. Third, since our simulations assumed a constant risk premium, the results suggest that learning about cash flow dynamics could be an alternative explanation to the time-variation in return predictability that we document in the first part of the paper.

4.6   Learning Effects and Pockets

The simulation results in the previous section show that investor learning in a predictive systems model with regime switching dynamics is capable of generating the pockets of return predictability that we find empirically. In particular, contrasting the results in Table 9 under learning and no learning suggests that investors' perceptions of the underlying state can be a key driver of return predictability. To explore if this is indeed the case, we next investigate the ability of an investor's misperception of expected growth to explain the rise and fall of pockets. To this end, define the belief discrepancy measure

$$\Delta \tilde{\mu}_t \equiv \hat{\mu}_{t|t} - \mu_t, \tag{34}$$

which is the difference between an agent's inference about expected growth ($\hat{\mu}_{t|t}$) and the true value of expected growth at time t, μ_t. To link local return predictability to this measure of belief discrepancy, we consider a variety of regression specifications of the following form

$$y_{it}^{pocket} = \alpha + x_{it}'\beta + \gamma \mathbb{1}\left\{edge_{it}\right\} + \mathbb{1}\left\{edge_{it}\right\} x_{it}'\delta + \varepsilon_{it}. \tag{35}$$

Here the i subscript refers to the simulation number and t refers to the time period within a simulated sample, from 1 to 15,860 (the sample size for the 3-month Treasury bill). The dummy variable $edge_{it}$ takes the value 1 for the first and last 126 (half of the kernel regression bandwidth)

periods of the sample, and zero otherwise. The dependent variable y_{it}^{pocket} is chosen to be either an indicator for whether a pocket is identified at period t in sample i or the local R^2 measure from the kernel regression. The vector x_{it} contains different functions of $\Delta\tilde{\mu}_{it}$, namely (i) $x_{it} = \Delta\tilde{\mu}_{it}$, (ii) $x_{it} = \Delta\tilde{\mu}_{it}^2$, (iii) $x_{it} = |\Delta\tilde{\mu}_{it}|$, and (iv) $x_{it} = [\ \Delta\tilde{\mu}_{it}^+ \quad \Delta\tilde{\mu}_{it}^- \ ]$, where $\Delta\tilde{\mu}_{it}^+ = \max(\Delta\tilde{\mu}_{it}, 0)$ and $\Delta\tilde{\mu}_{it}^- = \min(\Delta\tilde{\mu}_{it}, 0)$. In addition to these four discrepancy measures, we also consider a measure that captures the uncertainty about the underlying regime s_{it}, namely (v) $\hat{\pi}_{it|t}(1 - \hat{\pi}_{it|t})$.

To summarize the results across the 1,000 simulations, we report the coefficient estimate as the average coefficient estimate across simulations. Standard errors of the coefficient estimates are computed as the standard deviation of the estimates across simulations scaled by the square root of the number of simulations. These standard errors are then used to compute p-values. We consider both choices for y_{it}^{pocket} and allow for 5% and 1% significance thresholds for identifying pockets. Note that the choice of significance threshold does not affect the local R^2 results, only the pocket indicator variable results.

The results, reported in Table 10, show that the belief discrepancy measure, $\Delta\tilde{\mu}_t$, is strongly negatively correlated with both the pocket indicator and the local R^2 measure so that periods in which investors are overly pessimistic about dividend growth are more likely to coincide with pockets of return predictability. Moreover, $\Delta\tilde{\mu}_t$ explains a fairly high proportion of the variation in the two pocket measures with R^2 values between 13% and 19%. An even stronger result is obtained when we introduce the squared belief discrepancy, $\Delta\tilde{\mu}_t^2$, which explains between 22% and 30% of the variation in the two pocket measures.

Figure 8 uses a single simulation to illustrate the relation between the pocket indicator and the belief discrepancy measure, $\Delta\tilde{\mu}_{it}$. In this particular simulation there are three instances in which agents substantially underestimate the true growth rate of cash flows and the figures shows how each of these episodes is associated with a pocket of predictability.

In unreported results we find that the switching indicator is not significantly correlated with either measure of return predictability. This happens because regime switches are quite rare in our sample. Thus, the mere possibility of a regime switch seems to generate local return predictability even in cases where a regime switch fails to materialize.

This finding is related to the effect of the uncertainty about the current state, $\hat{\pi}_{t|t}(1 - \hat{\pi}_{t|t})$ on local return predictability. We find that uncertainty about the current state is significantly positively correlated with both the pocket indicator and the local R^2, although at 4-5%, the explanatory power of this variable is quite low.

We conclude from these findings that variation in investors' learning about a highly persistent growth rate of the cash flow process can create pockets of return predictability. It is when the belief discrepancy measure and uncertainty about the underlying state are largest that the effect tends to be biggest.

4.6.1 Difficulty of Investors' Learning Problem

Our model effectively captures investors' learning about a persistent risk component in the cash flow process. The presence of return predictability pockets is directly related to the difficulty of the learning problem for the investor. If the learning problem is too difficult, investors never detect the presence of pockets in which the predictor variable is correlated with cash flow growth. Conversely, if investors can estimate the underlying cash flow growth regime very accurately, cash flow growth will be predictable while stock returns will *not* be predictable as the stock price immediately incorporates any news about an underlying regime switch.

The difficulty of investors' learning problem seems to most significantly depend on three parameters: the persistence of expected cash flows, the relative duration of regime 1 to regime 2, and the correlation between the predictor variable and expected cash flows in the first regime. In particular, we find that for a given value of the correlation, the difficulty of the learning problem is similar to the baseline calibration for a more persistent cash flow process and a relatively more infrequent regime 1 (higher persistence of regime 2). Similarly, for a given value of the relative duration of the two regimes, the difficulty of the learning problem is similar to the baseline calibration for a more persistent cash flow process and a higher value of the correlation between expected cash flows and the predictor variable in regime 1.

Intuitively, if regime 1 is more infrequent, then for the same persistence in expected cash flows, the investor will have a more difficult learning problem, because they may more often misattribute large fluctuations in dividend growth to a regime change as opposed to just noise. Similarly, if expected cash flows and the observed predictor are more strongly correlated, then for the same persistence in expected cash flows, the investor will have an easier learning problem because they can be more confident that unusual fluctuations in dividend growth are due to a regime change.

Fixing the values of all parameters except for the persistence of cash flow growth, values comparable to those used by Bansal and Yaron (2004) and Bansal, Kiku, and Yaron (2012) lead to an overidentification of pocket periods. In other words, more persistent cash flow growth all else equal makes the learning problem 'too difficult' and so long periods where there exists genuine predictability of cash flows through x_t go unincorporated into asset prices. This is because it takes longer for agents to detect a switch to the regime with non-zero correlation. Conversely, if cash flow growth is even less persistent than in the calibration, the learning problem is 'too easy.' Agents are able to quickly identify switches between regimes and thus information is incorporated into prices too quickly to see long periods of predictability.

5   Economic Sources of Local Return Predictability

We argued earlier in the paper that the return predictability pockets detected by our analysis can be used as a diagnostic that helps identify the sources of return predictability. We next use

this idea to explore whether the evidence of local return predictability is associated with business cycle movements, changes in variables known to track market sentiment, and shifts in broker-dealer leverage. Moreover, we also study whether the pockets with return predictability could have been detected in real time. This is a question with implications for whether investors could have exploited localized return predictability and hence affects the interpretation of how our findings are related to market efficiency.

5.1 Pockets and Variation in the Business Cycle

Studies such as Rapach, Strauss, and Zhou (2010), Henkel et al. (2011), and Dangl and Halling (2012) find a systematic relationship between return predictability in the stock market and economic recessions. To explore this relationship, we regress the pocket indicator generated by our univariate linear regressions, I_t^{pocket}, on a constant and the NBER recession indicator, $NBER_t$

$$I_t^{pocket} = \mu + \beta NBER_t + \varepsilon_t. \tag{36}$$

A positive coefficient β suggests that return predictability pockets are more likely to occur during economic recessions while a negative value of β suggests the opposite.

To see whether the extent of return predictability depends on the state of the economy, we also regress the local R^2 measure on the NBER indicator

$$R_t^2 = \mu + \beta NBER_t + \varepsilon_t. \tag{37}$$

Here a positive coefficient indicates that return predictability tends to be higher during recessions, while a negative coefficient would indicate the opposite

The results, reported in the top panel of Table 11, show that local predictability of stock returns is indeed related to the business cycle, with seven of eight coefficients being positive. However, recession risk does not appear to be the main driver of local pockets of return predictability as the R^2-values of these regressions are very low, less than five percent for all predictors with exception of the term spread for which the R^2 is 17% in either regression.[23] Moreover, results from the regression in (36) are more mixed with a negative estimate of β for the dividend yield predictor and with only two of the four predictors (T-bill rate and the term spread) generating significantly positive coefficients at the 5% critical level.

5.2 Pockets and Variation in Sentiment

Our second regression uses the sentiment indicators proposed by Baker and Wurgler (2006, 2007) as a means to see whether return predictability is correlated with market sentiment. We first

---

[23]We find similar results when we project the pocket indicator on an early recession indicator (the three months after the peak of the cycle) or a late recession indicator (three months before the trough).

assign to each day within a given month the value of the Baker-Wurgler sentiment indicator, BW, of the same month. Then, analogously to the analysis of a business cycle component in return predictability, we estimate daily regressions

$$
\begin{aligned}
R_t^2 &= \mu + \beta BW_t + \varepsilon_t, \\
I_t^{pocket} &= \mu + \beta BW_t + \varepsilon_t.
\end{aligned}
\tag{38}
$$

The second panel in Table 11 shows evidence that large values of the BW index are associated with a greater degree of local return predictability for the T-bill rate and the realized variance, but not for the dividend yield and the term spread. Moreover, the R^2 from these regressions is quite low with exception of the realized variance for which it is around 10%.

5.3   Pockets and Variation in Leverage

Our third regression uses seasonally adjusted changes in U.S. broker-dealer leverage as constructed by Adrian, Etula, and Muir (2014) to see whether return predictability is correlated with the presence of funding constraints. Their measure is constructed using quarterly *Flow of Funds* data on the assets and liabilities of security brokers and dealers. Again, we assign each day within a given quarter the value of the leverage factor, LF, of that same quarter. We then estimate daily regressions

$$
\begin{aligned}
R_t^2 &= \mu + \beta LF_t + \varepsilon_t, \\
I_t^{pocket} &= \mu + \beta LF_t + \varepsilon_t.
\end{aligned}
\tag{39}
$$

The bottom panel in Table 11 shows that decreases in broker dealer leverage are associated with a greater degree of local return predictability for the dividend yield and the T-bill rate but not for the term spread and the realized variance. Because lower leverage is associated with lower availability of arbitrage capital (tighter funding constraints), the negative sign of the significant slope coefficients is what we would expect if return predictability is likely to be lower when there is more arbitrage capital available to exploit such predictability.

Turning to the magnitude of the relations, a one-standard deviation decrease in leverage, which corresponds to roughly a 14 percentage point drop, is associated with an increase in the local R^2 of up to 0.056 percentage points for the T-bill rate. Despite the economic significance of these results, the R^2 from the regressions is quite low, with the highest being 1.48% for the T-bill rate.

5.4   Out-of-sample Return Predictability

So far our methods for identifying return predictability used two-sided kernels, i.e., windows consisting of data both before and after the point at which local return predictability is being tested. In

real time, investors only have access to data prior to and including the point at which the forecast is being generated and so must use a one-sided window to estimate their model.

Assuming that return predictability is not driven by a time-varying risk premium, a one-sided prediction approach should not be able to generate better return forecasts than a simple model with a constant equity premium. To test if this implication (absence of one-sided return predictability) holds, we estimate the same model as in Section 3 but use a one-sided analog of the Epanechnikov Kernel in (12):

$$K(u) = \frac{3}{2}\left(1 - u^2\right) 1\left\{-1 < u < 0\right\}, \tag{40}$$

so that only past data are used to estimate the time varying relationship between y and x as indicated by $1\left\{-1 < u < 0\right\}$.[24]

We construct two forecasts of excess returns at time $t + 1$. The first uses the prevailing mean benchmark of Goyal and Welch (2008):

$$\bar{r}_{t+1|t} = \frac{1}{t}\sum_{s=1}^{t} r_s. \tag{41}$$

The second forecast is generated by the nonparametric model:[25]

$$\hat{r}_{t+1|t}^{local} = \bar{r}_{t+1|t} + x_t'\hat{\beta}_t. \tag{42}$$

To see if local return predictability could have been exploited in real time, we test the null of equal predictive accuracy (equal squared forecast errors) for the prevailing mean model in (41) and the time varying mean model in (42). To this end, we consider values of the test statistic proposed by Diebold and Mariano (1995) which is based on the difference in squared forecasts errors

$$\Delta SE_{t+1} = \left(r_{t+1} - \bar{r}_{t+1|t}\right)^2 - \left(r_{t+1} - \hat{r}_{t+1|t}^{local}\right)^2. \tag{43}$$

Positive values of ΔSE_{t+1} show that the time-varying mean model produced more accurate return forecasts, while negative values suggest that the constant equity premium (prevailing mean) model produced the most accurate one-sided forecasts.

Next, we compute the sample mean of ΔSE_{t+1} across the out-of-sample period, $T_0,..., T$:

$$\hat{\mu}_{\Delta MSE} = \frac{\sum_{t=t_{pre}}^{T-1} \Delta SE_{t+1}}{T - T_0},$$

---

[24]Note that the multiplicative factor becomes $\frac{3}{2}$ instead of $\frac{3}{4}$ in (12) so that the kernel function in (40) still integrates to one.

[25]To get a forecast of the level of stock returns, $x_t'\hat{\beta}_t$ is rescaled by the unconditional standard deviation of r.

and use this to compute the Diebold-Mariano test statistic as

$$DM = \frac{\hat{\mu}_{\Delta MSE}}{SE(\hat{\mu}_{\Delta MSE})}, \tag{44}$$

where $SE(\hat{\mu}_{\Delta MSE})$ is a Newey-West (HAC) estimate of the standard error of $\hat{\mu}_{\Delta MSE}$.

The first column in Panel A of Table 12 shows that, on average, across all days in the out-of-sample period, the prevailing mean model (41) outperforms the model with a local time-varying mean, (42) for all predictor variables. Additionally, all of these test statistics are statistically significant at conventional levels.

These results show that local return predictability could not have been exploited in real time to produce return forecasts that "on average" were more accurate than a model that assumes a constant equity premium. In fact, the one-sided estimates of the regression coefficients in (7) are notably noisier than their two-sided equivalents. The stark difference between the one-sided and two-sided results can thus be explained by the latter's use of more information, and its improved power to identify local return predictability.[26]

The results reported in the first panel of Table 12 pertain to the average out-of-sample performance of the local return prediction models. However, it is worth exploring whether the time-varying mean model outperformed the prevailing mean in periods that were identified, ex-ante, as pockets using a one-sided window.

To evaluate whether evidence of return predictability pockets could have been used in real time, we use left-sided kernel estimates of the model parameters in (8) as well as a left-sided estimate of the local R^2 value in (11). We then define a real-time return predictability pocket as a period in which the left-sided R^2, estimated using a backward-looking kernel, exceeds 0.01 for at least 10 days. We use this simpler definition rather than relying on one-sided standard errors of the β estimates because there is no result establishing consistency for one-sided standard errors of the form in equation (9). Moreover, an R^2 value of 1% broadly corresponds to the pockets identified in Figures 1-4.[27] The requirement that a pocket has lasted 10 days reduces the chance of picking up very short-lived, spurious pockets, but the results are not sensitive to our choice of the choice of a 10-day threshold.[28]

To compute DM tests inside versus outside pockets, let $\mathbb{1}\{pocket_t\}$ be an indicator variable for whether or not time t was identified as belonging to a pocket period in our one-sided estimation

---

[26]Lettau and van Nieuwerburgh (2008) report a similar finding for a return predictability model with breaks to the dividend yield.

[27]We did not experience with alternative cutoff values for the R^2 value, but it is clear what the trade-offs are: higher cutoff values will reduce the period of time spent inside pockets, which can reduce the power of the CW and DM test statistics as they are computed on a smaller sample. Lower cutoff values will increase the percentage of the sample inside pockets but may also lead to the inclusion of periods with spurious return predictability.

[28]For example, the results are not sensitive to whether a five or a twenty day minimum pocket length is used.

exercise. We then define in- and out-of-pocket versions of the DM test statistic as

$$DM_{in} = \frac{\frac{\sum_{t=T_0}^{T-1} \Delta SE_{t+1} \mathbb{1}\{pocket\}_{t+1}}{T_{pocket}}}{SE \left(\frac{\sum_{t=T_0}^{T-1} \Delta SE_{t+1} \mathbb{1}\{pocket\}_{t+1}}{T_{pocket}} \right)},$$

$$DM_{out} = \frac{\frac{\sum_{t=T_0}^{T-1} \Delta SE_{t+1}(1-\mathbb{1}\{pocket\}_{t+1})}{T-T_0-T_{pocket}}}{SE \left(\frac{\sum_{t=T_0}^{T-1} \Delta SE_{t+1}(1-\mathbb{1}\{pocket\}_{t+1})}{T-T_0-T_{pocket}} \right)},$$

where $T_{pocket} = \sum_{t=T_0}^{T-1} \mathbb{1}\{pocket\}_{t+1}$ is the number of days spent inside ex-ante identified pockets.

The second and third columns of Table 12 show results for the real-time identified pockets. Inside the pockets, the DM test statistics are positive for all four predictor variables, though they fail to be statistically significant. Conversely, outside the return predictability pockets, we see strong evidence for all four predictors of negative and highly significant DM tests. These results show that while the backward-looking kernel forecasts are significantly less accurate than forecasts from the prevailing mean model outside the return pockets, they are actually more accurate inside the ex-ante identified pockets, though not at a statistically significant level.

We can also test if the relative performance of the prevailing mean and kernel prediction model is similar inside versus outside pockets. The final column in Table 12 show the results from this test which compares the MSE performance of the left-sided kernel forecasts to the prevailing mean, reporting the p-value for a one-sided test with small p-values indicating that the kernel forecasts are relatively more accurate inside pockets compared to outside the pockets. For the forecasts based on the dividend yield and the T-bill rate we obtain p-values of 0.06 and 0.04, respectively, while the p-values for the term spread and realized variance predictors are 0.14 and 0.15, respectively. These results lend further credibility to the conclusion that the out-of-sample performance of at least a subset of our prediction models is notably better inside compared to outside pockets that have been identified using only historically available information.

An issue with using one-sided estimation windows for the out-of-sample forecasts is that they tend to generate coefficient estimates that are quite noisy. To deal with this, we use the "reflection" technique proposed by Chen and Hong (2012). Specifically, in order to estimate β_t at a given time t, we set the data $(X_{t+1}, \ldots, X_{t+hT-1}) = (X_{t-1}, \ldots, X_{t-hT+1})$ and similarly set $(y_{t+2}, \ldots, y_{t+hT}) = (y_t, \ldots, y_{t-hT+2})$. That is, we "reflect" the past data around time t to now also be the future data. We then use the same two-sided Epanechnikov kernel as in our main empirical specification to estimate β_t. Thus, even though we are using a two-sided kernel, we are still only using historical data and these estimates can be used to construct valid out-of-sample forecasts. This same method is then used to construct a real-time measure of the local R^2 by reflecting the residuals from the original regression around each time period t. Results from this procedure are shown in panel B of Table 12.

The out-of-sample forecasts are now uniformly better compared to the prevailing mean model. For the full sample, the DM test statistics go from being negative (Panel A) to being positive (Panel B). While the largest improvements are seen during the out-of-pocket periods, we also see sizeable improvements in the in-pocket performance which, measured relative to the benchmark, continues to be statistically better at the 10% level for three of the four predictors, with the fourth (realized variance) obtaining a p-value of 0.18.

Figure 9 provides a graphical illustration of the difference between our return predictability results inside versus outside pockets. The figure plots the cumulative sum of squared forecast errors using the forecasts from the prevailing mean model minus the squared forecast errors from the nonparametric kernel model:

$$CSSED_{\tau_E} = \frac{1}{\sqrt{\tau_E}} \sum_{\tau=1}^{\tau_E} \left[\left(r_{\tau+1} - \bar{r}_{\tau+1|\tau} \right)^2 - \left(r_{\tau+1} - \hat{r}_{\tau+1|\tau}^{local} \right)^2 \right] \mathbb{1}\{pocket\}_{t+1},$$

where $\mathbb{1}\{pocket\}_{t+1}$ is again defined using only ex-ante available information, and τ_E is the length of a pocket episode. We compute a similar measure for observations spent outside pockets.

For each predictor variable we average this measure across the pockets identified ex-ante in our sample and plot this against the number of days since a pocket started. Finally, we scale this measure by the square root of the length of the period over which the difference is being cumulated, τ_E, so as to make the variance of the plot comparable across different values of τ_E. Positive and rising values of the graph indicate that the kernel forecasts produce smaller squared forecast errors than the prevailing mean and so are more accurate. Negative values suggest the opposite. In each case, we find that there is evidence of return predictability inside the pockets (left column) but not outside the pockets (right column).

We conclude from these results that it would have been difficult in real time to identify local return predictability pockets and use such information to generate forecasts that were significantly more accurate than those produced by the prevailing mean model. However, using real-time information on the presence of return predictability pockets could, at the very least, have been used to avoid (out-of-pocket) periods with significantly worse forecasting performance than the prevailing mean.

6 Conclusion

We develop a robust nonparametric approach to test for the presence of pockets with local predictability of stock market returns. Empirically, we find evidence that stock returns are predictable more often than one would expect from a large class of asset pricing models which imply that expected stock returns are an affine function of economic state variables with constant coefficients. Such models fail in matching the longer-lived pockets found in our data which account for the largest amount of return predictability.

We next develop a new predictive systems model with regime switching that explains the presence of local return predictability by means of investors' incomplete learning about a persistent component in the dividend growth process. Predictability of stock market returns turns out to be well suited for studying learning effects due to the dependency of stock prices on cash flows expected to occur in the distant future and the considerable uncertainty surrounding such expectations. The high sensitivity of aggregate stock prices to even minor variations in beliefs about future cash flow growth rates means that cash flow learning effects are likely to be an important source of return movements. Through simulations from our new model, we find that investor learning about cash flow growth can induce patterns in return predictability that closely resemble those found in the data.

Our findings contribute to several areas of the finance literature in which a better understanding of both the patterns of return predictability and the source of such predictability matters. Indeed, the belief that returns are predictable has influenced key areas of finance such as asset allocation (e.g., Ait-Sahalia and Brandt (2001), Barberis (2000), Campbell and Viceira (1999), and Kandel and Stambaugh (1996)), performance evaluation of mutual funds (e.g., Ferson and Schadt (1996), Avramov and Wermers (2006), and Banegas et al. (2013)), and theoretical asset pricing models (e.g., Bansal and Yaron (2004)). Our empirical findings that stock return predictability is more local in time than previously thought and need not represent a time-varying risk premium may lead to revisions in how investment performance is being benchmarked and how asset pricing models are being tested. Some of our new measures of local return predictability could be used as diagnostics for determining whether a particular asset pricing model matches the return predictability patterns observed in the data.

References

[1] Adrian, T., E. Etula, and T. Muir, 2014, Financial Intermediaries and the Cross-Section of Asset Returns. Journal of Finance 69 (6), 2557–2596.

[2] Aït-Sahalia, Y. and M.W. Brandt, 2001, Variable selection for portfolio choice. Journal of Finance 56 (4), 1297–1351.

[3] Ang, A. and G. Bekaert, 2007, Stock Return Predictability: Is it There? Review of Financial Studies 20 (3), 651–707.

[4] Aruoba, S., F. Diebold, and C. Scotti, 2009, Real-Time Measurement of Business Conditions, Journal of Business and Economic Statistics 27 (4), 417–427.

[5] Avramov, D. and R. Wermers, 2006, Investing in mutual funds when returns are predictable. Journal of Financial Economics 81 (2), 339–377.

[6] Baker, M. and J. Wurgler, 2006, Investor sentiment and the cross-section of stock returns. Journal of Finance 61 (4), 1645–1680.

[7] Baker, M. and J. Wurgler, 2007, Investor sentiment in the stock market. Journal of Economic Perspectives 21 (2), 129–152.

[8] Balvers, R.J., T.F. Cosimano, and B. McDonald, 1990, Predicting Stock Returns in an Efficient Market. Journal of Finance 45 (4), 1109–1128.

[9] Banegas, A., B. Gillen, A. Timmermann, and R. Wermers, 2013. The cross section of conditional mutual fund performance in European stock markets. Journal of Financial economics 108 (3), 699–726.

[10] Bansal, R. and A. Yaron, 2004, Risks for the long run: A potential resolution of asset pricing puzzles. The Journal of Finance 59 (4), 1481–1509.

[11] Barberis, N., 2000, Investing for the Long Run when Returns Are Predictable. Journal of Finance 55 (1), 225–264.

[12] Bansal, R., D. Kiku, and A. Yaron, 2012, An Empirical Evaluation of the Long-Run Risks Model for Asset Prices, Critical Finance Review, 1 (1), 183–221

[13] Bekaert, G., and E. Engstrom, 2017, Asset Return Dynamics under Habits and Bad Environment Good Environment Fundamentals. Journal of Political Economy 125 (3), 713–760.

[14] van Binsbergen, J.H., and R.S.J. Koijen, 2010, Predictive regressions: A present value approach. Journal of Finance 65 (4), 1439–1471.

[15] Cai, Z., 2007, Trending time-varying coefficient time series models with serially correlated errors. Journal of Econometrics, 136 (1), 163–188.

[16] Campbell, J.Y., 1987. Stock returns and the term structure. Journal of Financial Economics 18 (2), 373–399.

[17] Campbell, J.Y. and J.H. Cochrane, 1999, By force of habit: A consumption-based explanation of aggregate stock market behavior. Journal of Political Economy 107 (2), 205–251.

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

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

[20] Campbell, J.Y. and L.M. Viceira, 1999, Consumption and portfolio decisions when expected returns are time varying. Quarterly Journal of Economics 114 (2), 433–495.

[21] Cecchetti, S. G., Lam P.S., and Mark N.C., 1990, Mean Reversion in Equilibrium Asset Prices. American Economic Review 80 (3), 398–418.

[22] Chen, B. and Y. Hong, 2012, Testing for smooth structural changes in time series models via nonparametric regression. Econometrica 80 (3), 1157–1183.

[23] Chu, C-S J., M. Stinchcombe, and H. White, 1996, Monitoring Structural Change. Econometrica 64 (5), 1045–1065.

[24] Clark, T.E. and K.D. West, 2007, Approximately normal tests for equal predictive accuracy in nested models. Journal of Econometrics 138 (1), 291–311.

[25] Constantinides, G.M. and D. Duffie, 1996, Asset Pricing with Heterogeneous Consumers. Journal of Political Economy 104 (2), 219–240.

[26] Constantinides, G.M. and A. Ghosh, 2017, Asset Pricing with Countercyclical Household Consumption Risk. Journal of Finance 72 (1), 415–460.

[27] Creal, D., and C. Wu, 2016, Bond Risk Premia in Consumption Based Models. Unpublished manuscript, University of Chicago.

[28] Dangl, T. and M. Halling, 2012, Predictive regressions with time-varying coefficients. Journal of Financial Economics 106 (1), 157–181.

[29] David, A., and P. Veronesi, 2013, What ties return volatilities to price valuations and fundamentals? Journal of Political Economy 121 (4), 682–746.

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

[31] Drechsler, I. and A. Yaron, 2011, What's Vol Got to Do with It. Review of Financial Studies, 24 (1), 1–45.

[32] Elliott, G. and U. Muller, 2006, Efficient tests for general persistent time variation in regression coefficients. Review of Economic Studies 73 (4), 907–940.

[33] Epstein, L.G. and S.E. Zin, 1989, Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework. Econometrica 57 (4), 937–969.

[34] Eraker, B. and I. Shaliastovich, 2008, An Equilibrium Guide to Designing Affine Asset Pricing Models. Mathematical Finance, 18 (4), 519–543.

[35] Fama, E.F., French, K.R., 1988. Dividend yields and expected stock returns. Journal of Financial Economics 22 (1), 3–25.

[36] Fama, E.F., French, K.R., 1989. Business conditions and expected returns on stocks and bonds. Journal of Financial Economics 25 (1), 23–49.

[37] Farmer, L., 2017, The Discretization Filter: A Simple Way to Estimate Nonlinear State Space Models. SSRN Working Paper No. 2780166. Available at SSRN: https://ssrn.com/abstract=2780166

[38] Ferson, W. E. and R. W. Schadt, 1996, Measuring fund strategy and performance in changing economic conditions. Journal of Finance 51 (2), 425–461.

[39] Green, J., R.M. Hand, and M.T. Soliman, 2011, Going, Going, Gone? The Apparent Demise of the Accruals Anomaly. Management Science 57 (5), 797–816.

[40] Hansen, L.P., J.C. Heaton, and N. Li, 2008, Consumption Strikes Back? Measuring Long-Run Risk. Journal of Political Economy 116 (2), 260–302.

[41] Harvey, C.R., Y. Liu, and H. Zhu, 2016, ...and the Cross-Section of Expected Stock Returns. Review of Financial Studies 29 (1), 5–68.

[42] Henkel, S.J., J.S. Martin, and F. Nardari, 2011, Time-varying short-horizon predictability. Journal of Financial Economics 99 (3), 560–580.

[43] Herskovic, B., B. Kelly, H. Lustig, S. van Nieuwerburgh, 2016, The common factor in idiosyncratic volatility: Quantitative asset pricing implications. Journal of Financial Economics 119 (2), 249–283.

[44] Hong, H., J.C. Stein, and J. Yu, 2007, Simple forecasts and paradigm shifts. Journal of Finance 62 (3), 1207–1242.

[45] Johannes, M., A. Korteweg, and N. Polson, 2014, Sequential learning, predictive regressions, and optimal portfolio returns. Journal of Finance 69 (2), 611–644.

[46] Kandel, S. and R.F. Stambaugh, 1996, On the predictability of stock returns: An asset-allocation perspective. Journal of Finance 51 (2), 385–424.

[47] Keim, D.B. and R.F. Stambaugh, 1986. Predicting returns in the stock and bond markets. Journal of Financial Economics 17 (2), 357–390.

[48] Kelly, B. and S. Pruitt, 2013, Market expectations in the cross-section of present values. Journal of Finance 68 (5), 1721–1756.

[49] Lettau, M. and S.C. Ludvigson, 2010, Measuring and modeling variation in the risk-return trade-off. Volume 1 of Handbook of Financial Econometrics, 617–690. Elsevier.

[50] Lettau, M. and S. van Nieuwerburgh, 2008, Reconciling the return predictability evidence. Review of Financial Studies 21 (4), 1607–1652.

[51] Lettau, M. and J.A. Wachter, 2011, The term structure of equity and interest rates. Journal of Financial Economics 101 (1), 90–113.

[52] Lustig, H., S. van Nieuwerburgh, and A. Verdelhan, 2013, The Wealth-Consumption Ratio. Review of Asset Pricing Studies 3 (1), 38–94.

[53] McLean, R.D., and J. Pontiff, 2016, Does Academic Research Destroy Stock Return Predictability? Journal of Finance 71 (1), 5-32.

[54] Pastor, L. and R.F. Stambaugh, 2009. Predictive systems: living with imperfect predictors. Journal of Finance 64 (4), 1583–1628.

[55] Paye, B.S., 2012, 'Déjà vol': Predictive regressions for aggregate stock market volatility using macroeconomic variables. Journal of Financial Economics 106 (3), 527–546.

[56] Paye, B.S. and A. Timmermann, 2006, Instability of return prediction models. Journal of Empirical Finance 13 (3), 274–315.

[57] Pesaran, M. H., and Timmermann, A., 1995, Predictability of Stock Returns: Robustness and Economic Significance. Journal of Finance 50 (4), 1201–1228.

[58] Pettenuzzo, D., R. Sabbatucci, and A. Timmermann, 2018, High Frequency Cash Flow Dynamics. Unpublished manuscript, Brandeis University, Stockholm School of Economics, and UCSD.

[59] Pettenuzzo, D., A. Timmermann, and R. Valkanov, 2014, Forecasting stock returns under economic constraints. Journal of Financial Economics 114 (3), 517–553.

[60] Rapach, D.E. and M.E. Wohar, 2006, Structural breaks and predictive regression models of aggregate U.S. stock returns. Journal of Financial Econometrics 4 (2), 238–274.

[61] Rapach, D.E. and G. Zhou, 2013, Forecasting stock returns. Handbook of economic forecasting Vol. 2, 328–383.

[62] Robinson, P.M., 1989, Nonparametric estimation of time-varying parameters. In Statistical Analysis and Forecasting of Economic Structural Change, 253–264. Springer Berlin Heidelberg.

[63] Rossi, A.G. and A. Timmermann, 2015, Modeling Covariance Risk in Merton's ICAPM. Review of Financial Studies 5 (1), 1428–1461.

[64] Schmidt, L., 2016, Climbing and Falling Off the Ladder: Asset Pricing Implications of Labor Market Event Risk. SSRN Working Paper No. 2471342. Available at SSRN: https://ssrn.com/abstract=2471342

[65] Schwert, G.W., 2003, Anomalies and market efficiency. In George M. Constantinides, Milton Harris, Rene M. Stulz (eds.) Handbook of the Economics of Finance. North Holland: Amsterdam.

[66] Tauchen, G., 1986, Finite state markov-chain approximations to univariate and vector autoregressions. Economic Letters 20 (2), 177-181.

[67] Timmermann, A., 2000, Moments of Markov switching models. Journal of Econometrics 96 (1), 75–111.

[68] Timmermann, A., 2008, Elusive Return Predictability. International Journal of Forecasting 24 (1), 1–18.

[69] Veronesi, P., 2000, How Does Information Quality Affect Stock Returns? Journal of Finance 55 (2), 807–837.

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

Appendix A: Proof of Proposition 1

Proof. To show part (i) of Proposition 1, we conjecture and verify that the price-dividend ratio is $pd_t = A_{0,m} + A'_m z_t$.

By Assumption 1, $\Delta d_t = S'_d z_t$. Suppose that 3.1 holds. Using $r_{s,t+1} \approx k + \rho(p_{t+1} - d_{t+1}) + \Delta d_{t+1} + d_t - p_t$ and plugging the log-linearized return into the Euler equation, we have

$$
\begin{aligned}
1 &= \exp[-A_{0,f} - A'_f z_t - \log E_t \exp[\Lambda'_t \epsilon_{t+1}] + \kappa + (\rho - 1)A_{0,m} - A'_m z_t] \\
&\quad \times E_t \left[\exp\left\{ -\Lambda' \epsilon_{t+1} + [S'_d + \rho A'_m] z_{t+1} \right\} \right] \\
0 &= -A_{0,f} - A'_f z_t + \kappa + (\rho - 1)A_{0,m} - A'_m z_t + [S'_d + \rho A'_m](\mu + F z_t) \\
&\quad + [f(-\Lambda' + S'_d + \rho A'_m) - f(-\Lambda)] + [\tilde{g}(-\Lambda' + S'_d + \rho A'_m)' - \tilde{g}(-\Lambda')'] z_t,
\end{aligned}
$$

where $\tilde{g}(u) \equiv [g(u)', \mathbf{0}']'$ and the second line takes logs and applies assumption 1(ii). Rearranging yields the $(L+1)$-dimensional system of equations in $A_{0,m}$ and A_m

$$
f(-\Lambda + S_d + \rho A_m) - f(-\Lambda) - A_{0,f} + \kappa + (\rho - 1)A_{0,m} + (S'_d + \rho A'_m)\mu = 0, \tag{45}
$$

$$
\tilde{g}(-\Lambda + S_d + \rho A_m) - \tilde{g}(-\Lambda) - A_f - (I - \rho F')A_m + F'S_d = 0. \tag{46}
$$

This system does not have an analytical solution in the general case; however, it is relatively straightforward to solve the system numerically. We note that Assumption 3.(2) for the data generating process is identical to those in Lustig et al. (2013). Therefore, we refer the interested reader to the proof of their Proposition 1 for full derivations of the $A_{0,m}$ and A_m coefficients in that case.

To show part (ii), we follow a very similar argument to Drechsler and Yaron (2011). We can write expected returns as follows:

$$
E_t[\exp(r_{s,t+1})] = \exp[E_t r_{s,t+1}] E_t[\exp([S'_d + \rho A'_m]\epsilon_{t+1})] \equiv \exp[E_t r_{s,t+1}] E_t[\exp(B'_m \epsilon_{t+1})] \tag{47}
$$

$$
\exp(-r_{f,t+1}) \equiv \exp[E_t m_{t+1}] E_t[\exp(-\Lambda'_t \epsilon_{t+1})]. \tag{48}
$$

Next, using the Euler equation in (3) and the law of iterated expectations, we have

$$
1 = \exp[E_t r_{s,t+1}] \exp[E_t m_{t+1}] E_t \exp[(-\Lambda'_t + B'_m)\epsilon_{t+1}] \tag{49}
$$

$$
\frac{E_t \exp[B'_m \epsilon_{t+1}] E_t \exp[-\Lambda_t \epsilon_{t+1}]}{E_t \exp[(-\Lambda'_t + B'_m)\epsilon_{t+1}]} = \exp[E_t r_{s,t+1}] \exp[E_t m_{t+1}] E_t \exp[B'_m \epsilon_{t+1}] E_t \exp[-\Lambda'_t \epsilon_{t+1}] \tag{50}
$$

$$
= E_t[\exp(r_{s,t+1})] \exp(-r_{f,t+1}) \tag{51}
$$

Taking logs and noting that $E_t r_{s,t+1} = \log E_t \exp(r_{s,t+1}) - \log E_t \exp[B'_m \epsilon_{t+1}]$, we get

$$E_t[r_{s,t+1}] - r_{f,t+1} = \log E_t \exp[-\Lambda_t \epsilon_{t+1}] - \log E_t \exp[(-\Lambda'_t + B'_m)\epsilon_{t+1}].$$

Suppose that Assumption 3.1 holds. Then, the expression in (??) simplifies to

$$E_t[r_{s,t+1}] - r_{f,t+1} = f(-\Lambda) - f(-\Lambda + B'_m) + [g(-\Lambda) - g(-\Lambda + B'_m)]'x_t,$$

which establishes the claim. If Assumption 3.2 holds, we can evaluate each one of the expressions in (??) using the cumulant generating function of the normal distribution:

$$E_t[r_{s,t+1}] - r_{f,t+1} = -\tfrac{1}{2}B'_m \Sigma B_m + B'_m \Sigma \Lambda_t = -\tfrac{1}{2}B'_m \Sigma B_m + B'_m \Sigma[\Lambda_0 + \Lambda_1 x_t],$$

which also establishes the claim. The first term is due to Jensen's inequality, while the second captures the covariance between the market return and the priced risk factors. Collecting terms in front of x_t in the two equations above yields the expressions for β under the two sets of assumptions. ∎

Appendix B: Details of Nonparametric Estimation

Robinson (1989) and Cai (2007) consider local constant and local linear approximations of β respectively, but this approach can easily be generalized to accommodate polynomials of arbitrary order. In particular, we can approximate the function β_t as a pth-order Taylor expansion about the point $\frac{t}{T}$ (where $p \geq 0$). To this end, define the quantities:

$$\mathbf{W}_{st} = \left(1, \frac{s-t}{T}, \ldots, \left(\frac{s-t}{T}\right)^p\right)', \tag{52}$$

$$K_{st} = K\left(\frac{s-t}{hT}\right), \tag{53}$$

$$\mathbf{Q}_{st} = \mathbf{W}_{st} \otimes x_s, \tag{54}$$

for $s, t = 1, \ldots, T$, where K is a kernel function and $h \equiv h(T)$ is the bandwidth. More formally, $K : [-1, 1] \to \mathbb{R}^+$ is a function that is symmetric about 0 and integrates to 1, and $h \in [0, 1]$ satisfies $h \to 0$ and $hT \to \infty$ as $T \to \infty$.

The local polynomial estimator $\beta = (\beta'_0, \beta'_1, \ldots, \beta'_p)'$ is obtained by solving

$$\min_{\beta \in \mathbb{R}^{pd}} \sum_{s=t-\lfloor hT \rfloor}^{t+\lfloor hT \rfloor} K_{st}\left[r_{s+1} - \beta'_0 x_s - \beta'_1\left(\frac{s-t}{T}\right)x_s - \ldots - \beta'_p\left(\frac{s-t}{T}\right)^p x_s\right]^2$$

$$= \sum_{s=t-\lfloor hT \rfloor}^{t+\lfloor hT \rfloor} K_{st}\left(r_{s+1} - \beta'\mathbf{Q}_{st}\right)^2. \tag{55}$$

48

Solving this optimization problem for α gives the solution

$$\hat{\beta}_t = \left(\sum_{s=t-\lfloor Th \rfloor}^{t+\lfloor Th \rfloor} K_{st} \mathbf{Q}_{st} \mathbf{Q}'_{st} \right)^{-1} \sum_{s=t-\lfloor Th \rfloor}^{t+\lfloor Th \rfloor} K_{st} \mathbf{Q}_{st} r_{s+1}, \tag{56}$$

where our object of interest, β_{1t}, is the first element of β_t. That is, the estimator of β_{1t} is given by

$$\hat{\beta}_{1t} = \left(\mathbf{e}'_1 \otimes \mathbf{I}_d \right) \hat{\beta}_t, \tag{57}$$

where $\mathbf{e}_1$ is the first standard basis vector of $\mathbb{R}^{p+1}$, $\mathbf{I}_d$ is a $(d \times d)$ identity matrix, and d is the dimension of x_t. This can also be thought of as the OLS estimator of β_0 in the transformed model

$$K_{st}^{1/2} y_{s+1} = K_{st}^{1/2} x'_s \sum_{q=0}^{p} \beta_q + \varepsilon_{s+1}. \tag{58}$$

The asymptotic properties of these estimators are studied in Robinson (1989) and Cai (2007). Under various regularity conditions, it can be shown that the estimator $\hat{\beta}_t$ in (57) is consistent and asymptotically normal.

Our main empirical results adopt a local constant (Nadarya-Watson) estimation procedure and so set $p = 0$. The motivation behind this choice is that the nonparametric procedures require very large amounts of data to perform well in finite samples and every additional degree of approximation requires that we estimate dT additional parameters. However, we also repeated the analysis using local linear models $(p = 1)$ and found very similar results.

Variables	Full sample beta	t-statistic	$\overline{R}^2$ (in %)	Start date	No. of obs.
dy	0.067	1.355	0.008	11/5/1926	23,786
tbl	-0.007	-2.765	0.053	1/4/1954	15,860
tsp	0.018	2.316	0.042	1/2/1962	13,846
rvar	6.43×10^{-5}	0.524	4.29×10^{-4}	1/15/1927	23,727

Table 1: Full sample regression statistics. This table reports full-sample beta estimates, t-statistics (computed using Newey-West standard errors), and $\overline{R}^2$ values for univarate regressions of daily excess stock returns on the lagged predictor variables listed in the rows. All series run through the end of 2016.

Variables	In-pocket					Out-of-pocket				
	Num pockets	Min length	Max length	Avg. length	Frac signif	Num pockets	Min length	Max length	Avg. length	Frac signif
					Panel A: 5% pocket statistics					
dy	13	41	448	177	0.098	13	142	8,448	1,633	0.902
tbl	8	19	415	214	0.110	8	19	5,357	1,737	0.890
tsp	3	313	498	385	0.085	4	774	8,492	3,109	0.915
rvar	8	32	329	225	0.077	9	726	4,423	2,408	0.923
					Panel B: 1% pocket statistics					
dy	3	171	320	236	0.030	4	851	14,487	5,706	0.970
tbl	4	56	221	152	0.039	4	839	10,737	3,749	0.961
tsp	3	122	326	221	0.049	4	1,016	8,536	3,232	0.951
rvar	5	78	246	189	0.040	6	894	8,729	3,755	0.960

Variables	In-pocket					Out-of-pocket				
	Avg. R^2	Mean	Std. dev	Skew	Kurtosis	Avg. R^2	Mean	Std. dev	Skew	Kurtosis
					Panel C: 5% return statistics					
dy	0.016	0.064	0.915	0.409	23.153	0.004	0.024	1.090	-0.140	19.148
tbl	0.015	-0.030	0.850	0.127	5.995	0.003	0.033	0.967	-0.594	19.967
tsp	0.018	-0.026	1.049	0.367	4.592	0.003	0.030	0.979	-0.607	20.437
rvar	0.021	0.053	0.816	-0.507	8.080	0.003	0.026	1.093	-0.093	19.498
					Panel D: 1% return statistics					
dy	0.024	0.115	0.608	-0.950	6.974	0.004	0.025	1.085	-0.100	19.281
tbl	0.020	-0.008	0.828	-0.284	4.020	0.004	0.028	0.960	-0.542	19.369
tsp	0.020	-0.022	1.165	0.365	3.903	0.003	0.027	0.975	-0.581	20.109
rvar	0.026	0.043	0.677	-0.359	4.391	0.004	0.027	1.088	-0.105	19.295

Table 2: Pocket summary statistics. This table reports summary statistics on the number of pockets with significant return predictability from the predictor variable listed in the left column, using a non-parametric kernel regression approach with significance levels of 5% (Panel A) or 1% (Panel B) to identify pockets. We show, for each predictor variable, the number of pockets identified, the minimum, maximum and average pocket length (all measured in days) as well as the fraction of day in the sample identified to have significant local return predictability. Left columns show summary statistics for periods inside pockets while, for comparison, right columns show summary statistics for periods outside pockets. Panels C and D report summary statistics for daily excess returns inside (left panels) and outside (right panels) pockets, including the average value of the local $\overline{R}^2$, the mean, standard deviation, skewness and kurtosis of daily returns. The sample periods vary across the predictor variables and begin in 11/5/1926 for the dividend yield (23,786 observations), 1/2/1954 for the 3-month T-bill rate (15,860 obs.), 1/2/1962 (13,846 obs.) for the term spread, and 1/15/1927 (23,727 obs.) for the realized variance.

Stats	Sample	Random Walk			GARCH		
		Avg.	Std. err.	p-val	Avg.	Std. err.	p-val
dy							
Num pockets	13	6.510	2.471	0.015	6.512	2.464	0.013
Min integral $\overline{R}^2$	0.237	0.269	0.520	0.319	0.229	0.506	0.282
Mean integral $\overline{R}^2$	2.784	0.801	1.028	0.045	0.713	0.964	0.048
Max integral $\overline{R}^2$	11.539	1.391	1.634	0.002	1.254	1.557	0.003
Frac signif	0.098	0.048	0.021	0.019	0.048	0.022	0.019
tbl							
Num pockets	8	4.381	1.882	0.057	4.430	1.906	0.059
Min integral $\overline{R}^2$	0.174	0.295	0.510	0.430	0.287	0.541	0.415
Mean integral $\overline{R}^2$	3.260	0.738	0.876	0.027	0.746	0.920	0.030
Max integral $\overline{R}^2$	6.889	1.272	1.517	0.014	1.299	1.501	0.009
Frac signif	0.110	0.047	0.025	0.013	0.049	0.026	0.020
tsp							
Num pockets	3	3.743	1.851	0.729	3.887	1.848	0.758
Min integral $\overline{R}^2$	4.924	0.256	0.541	0.002	0.279	0.679	0.004
Mean integral $\overline{R}^2$	6.768	0.635	0.938	0.002	0.665	1.024	0.005
Max integral $\overline{R}^2$	8.632	1.093	1.667	0.004	1.103	1.505	0.006
Frac signif	0.085	0.048	0.029	0.113	0.050	0.028	0.116
rvar							
Num pockets	8	6.714	2.421	0.371	6.557	2.423	0.343
Min integral $\overline{R}^2$	0.367	0.236	0.492	0.183	0.218	0.424	0.200
Mean integral $\overline{R}^2$	4.738	0.783	0.942	0.007	0.769	0.912	0.008
Max integral $\overline{R}^2$	8.754	1.451	1.569	0.005	1.418	1.552	0.004
Frac signif	0.077	0.049	0.021	0.092	0.048	0.021	0.090

Table 3: Statistical significance tests for pocket diagnostics (zero coefficient benchmark). This table reports the outcome of Monte Carlo simulations of daily excess returns using either a random walk model with constant mean and volatility (columns 2-4) or a model that allows for a time-varying expected return and time-varying volatility (columns 5-7). Using these respective models, each simulation draws a sample with the same length as the original sample for the respective predictor variables and computes the pocket measures listed in each row, including the number of pockets, the minimum, maximum and average length (in days) of the pockets, the minimum, mean and maximum integral $\overline{R}^2$, the fraction of the sample with a significant pocket indicator, the average and maximum values of the $\overline{R}^2$ inside pockets. The average values, standard errors and p-values for the pocket measures are computed using 1,000 simulations and are based on a zero coefficient benchmark.

Pocket #	dy	tbl	tsp	rvar
1	1.131	2.341	4.924	1.316
	(0.207)	(0.092)	(0.016)	(0.203)
2	0.448	4.746	8.632	7.214
	(0.427)	(0.016)	(0.003)	(0.004)
3	0.623	6.889	6.747	8.754
	(0.348)	(0.003)	(0.007)	(0.001)
4	7.032	0.752	-	3.352
	(0.005)	(0.327)		(0.044)
5	3.464	0.174	-	7.065
	(0.032)	(0.654)		(0.005)
6	1.580	6.438	-	6.958
	(0.138)	(0.004)		(0.005)
7	11.539	2.393	-	0.367
	(0.001)	(0.088)		(0.519)
8	4.103	2.351	-	2.875
	(0.025)	(0.091)		(0.056)
9	0.413	-	-	-
	(0.449)			
10	3.756	-	-	-
	(0.029)			
11	0.237	-	-	-
	(0.581)			
12	0.498	-	-	-
	(0.403)			
13	1.370	-	-	-
	(0.165)			

Table 4: Integral $\overline{R}^2$ measure and p-values for individual pockets. This table reports the integral $\overline{R}^2$ measure for each of the pockets identified by our nonparametric kernel regression approach, assuming a 5% cutoff value to define pockets with p-values in brackets. To compute p-values We use the Monte Carlo simulations in Table 3 to compute the proportion of simulated pockets that have an integral $\overline{R}^2$ measure as high as the value associated with a particular pocket identified in the data.

Variable	dy	tbl	tsp	rvar
dy	1.000	0.346	0.102	0.290
tbl	0.672	1.000	0.543	0.477
tsp	0.338	0.619	1.000	0.380
rvar	0.662	0.799	0.450	1.000

Table 5: Correlation between different pocket measures. The upper right correlations are the pairwise correlations between the pocket indicators for the two variables being compared. The bottom left correlations are the pairwise correlations between the local R^2 measures for the two variables being compared. Each correlation is computed over the longest common subsample available.

Specification	Num pockets	Mean length	Max length	Mean integral R^2	Max integral R^2	Frac signif
			dy			
12-month kernel	13	177	448	2.784	11.539	0.098
6-month kernel	21	88	356	2.991	15.384	0.079
18-month kernel	9	270	710	2.751	9.174	0.104
24-month kernel	7	314	598	2.361	7.076	0.094
First differencing	12	175	363	2.993	9.194	0.089
			tbl			
12-month kernel	8	214	415	3.260	6.889	0.110
6-month kernel	15	113	326	3.173	8.755	0.108
18-month kernel	7	305	737	3.243	8.098	0.138
24-month kernel	5	536	933	4.494	9.268	0.175
First differencing	11	261	930	5.401	32.036	0.184
			tsp			
12-month kernel	3	385	498	6.768	8.632	0.085
6-month kernel	12	86	217	2.185	5.910	0.075
18-month kernel	3	586	779	7.273	10.084	0.131
24-month kernel	3	767	1,020	7.576	11.978	0.173
First differencing	11	261	651	5.068	17.340	0.211
			rvar			
12-month kernel	8	225	329	4.738	8.754	0.077
6-month kernel	19	112	340	3.763	13.235	0.090
18-month kernel	7	302	589	3.797	6.965	0.091
24-month kernel	5	456	704	3.958	6.711	0.098
First differencing	11	248	476	4.860	15.692	0.116

Table 6: Robustness of pocket statistics across specifications (zero coefficient benchmark). This table reports a number of statistics, including the number of pockets, the mean and maximum length of a pocket, the mean and maximum integral R^2 across pockets, and the fraction of the sample that is classified as pockets, across different empirical specifications. The baseline specification uses the predictor variables in levels with an effective sample size of 12 months. Alternative specifications considered use the predictors in levels with effective sample sizes of 6, 18, and 24 months. Additionally, we consider two detrending procedures: subtracting off an exponentially-weighted moving average of the prior 12 months of data with a $\lambda = 0.99$, and first differencing.

Parameter	Value	Description
π_{11}	0.995	Probability of staying in regime 1 (average duration of 200 days)
π_{22}	0.9986	Probability of staying in regime 2 (average duration of 700 days)
ρ_x	0.882	Annualized persistence of observed predictor variable x
$\frac{\mu_{x,1}}{1-\rho_x}$	2.15%	Unconditional mean of observed predictor variable in regime 1
$\frac{\mu_{x,2}}{1-\rho_x}$	5.86%	Unconditional mean of observed predictor variable in regime 2
ρ_μ	0.305	Annualized persistence of expected cash flows
$\frac{\mu_{\mu,1}}{1-\rho_\mu}$	8.32%	Annualized unconditional mean of expected cash flows in regime 1
$\frac{\mu_{\mu,2}}{1-\rho_\mu}$	4.62%	Annualized unconditional mean of expected cash flows in regime 2
$\frac{\sigma_{v,1}}{\sqrt{1-\rho_x^2}}$	2.75%	Unconditional standard deviation of observed predictor variable in regime 1
$\frac{\sigma_{v,2}}{\sqrt{1-\rho_x^2}}$	2.55%	Unconditional standard deviation of observed predictor variable in regime 2
$\frac{\sigma_w}{\sqrt{1-\rho_\mu^2}}$	0.012%	Annualized unconditional standard deviation of expected cash flows
σ_u	0.66%	Annualized standard deviation of realized cash flows
$\frac{\sigma_{vw,1}}{\sigma_{v,1}\sigma_{w,1}}$	-0.2	Correlation between innovations to observed predictor variable and expected cash flows in regime 1

Table 7: **Calibrated parameters of predictive systems model.** This table reports the values and descriptions for the calibrated parameter values in the predictive systems model.

		No learning			Learning		
	Sample	Avg.	Std. err.	p-val	Avg.	Std. err.	p-val
Positive	0.531	0.297	0.070	0.002	0.547	0.217	0.477
Negative	-0.122	-0.032	0.030	0.993	-0.044	0.022	0.990
Net	0.414	0.262	0.067	0.017	0.508	0.209	0.634

Table 8: Average integral $\overline{R}^2$. This table reports the average integral $\overline{R}^2$ conditional on it being positive, negative, and over the whole sample. For the "positive" measure, the average of the local $\overline{R}^2$ is taken over all periods where it is positive and multiplied by 100. The analogous procedure is done for the "negative" measure. For the "net" measure, the average of the local $\overline{R}^2$ is taken over the whole sample and multiplied by 100. These statistics are computed for the actual data under "Sample," and the average, standard error, and one-sided p-values are computed for the predictive systems model simulations under both the no learning and learning specifications.

Stats	Sample	No learning			Learning		
		Avg.	Std. err.	p-val	Avg.	Std. err.	p-val
		5% significance results					
Num pockets	8	3.878	1.930	0.036	7.066	3.382	0.442
Min pocket length	19	78.442	62.319	0.866	88.903	80.111	0.861
Avg. pocket length	214	144.474	58.192	0.090	276.597	98.915	0.768
Max pocket length	415	231.578	91.663	0.019	591.708	281.226	0.744
Min integral $\overline{R}^2$	0.174	0.240	0.477	0.339	0.634	1.125	0.545
Mean integral $\overline{R}^2$	3.260	0.486	0.744	0.016	2.335	3.045	0.225
Max integral $\overline{R}^2$	6.889	0.752	1.092	0.000	5.026	6.989	0.223
Fraction significant	0.110	0.038	0.021	0.004	0.132	0.077	0.556
		1% significance results					
Num pockets	4	0.702	0.844	0.007	3.415	2.408	0.410
Min pocket length	56	46.395	65.894	0.330	122.279	116.429	0.687
Avg. pocket length	152.25	50.907	67.793	0.105	231.588	131.291	0.768
Max pocket length	221	57.866	76.654	0.040	388.279	251.471	0.763
Min integral $\overline{R}^2$	1.057	0.201	0.563	0.056	0.932	1.896	0.240
Mean integral $\overline{R}^2$	3.127	0.233	0.613	0.010	2.071	3.247	0.188
Max integral $\overline{R}^2$	4.771	0.265	0.695	0.003	3.601	5.813	0.214
Fraction significant	0.039	0.005	0.007	0.002	0.058	0.048	0.567

Table 9: Simulations from predictive systems learning model. This table presents simulation results from the predictive systems model with regime switching in the cash flow growth rate. Investors observe a predictor variable that is correlated with the latent process driving the mean dividend growth rate, but whose correlation is also affected by the regime switching. In the scenario with no learning (columns 2-4), investors are assumed to observe the latent state variable while in the scenario with learning (columns 5-7), investors update their estimates of the mean dividend growth rate based on their probability estimates of the underlying state. The reported sample average, standard errors and p-values for the simulated data are based on 1,000 simulations of the same length as the sample for the T-bill rate and assume a mean dividend-price ratio of 0.038 which is the historical sample average. Pockets in both the actual and simulated data sample are computed around a zero coefficient benchmark.

	Pocket Indicator					Local R^2				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
5% significance results										
$\hat{\mu}_{t\|t} - \mu_t$	-0.091*** (<0.01)	-	-	-	-	-0.225*** (<0.01)	-	-	-	-
$(\hat{\mu}_{t\|t} - \mu_t)^2$	-	0.094*** (<0.01)	-	-	-	-	0.218*** (<0.01)	-	-	-
$\|\hat{\mu}_{t\|t} - \mu_t\|$	-	-	0.216*** (<0.01)	-	-	-	-	0.448*** (<0.01)	-	-
$(\hat{\mu}_{t\|t} - \mu_t)^+$	-	-	-	0.204*** (<0.01)	-	-	-	-	0.448*** (<0.01)	-
$(\hat{\mu}_{t\|t} - \mu_t)^-$	-	-	-	-0.251*** (<0.01)	-	-	-	-	-0.601*** (<0.01)	-
$\hat{\pi}_{t\|t}(1 - \hat{\pi}_{t\|t})$	-	-	-	-	0.038*** (<0.01)	-	-	-	-	0.086*** (<0.01)
$\overline{R}^2$	13.47%	22.58%	20.03%	24.55%	4.19%	19.14%	29.93%	27.03%	33.72%	5.58%
1% significance results										
$\hat{\mu}_{t\|t} - \mu_t$	-0.058*** (<0.01)	-	-	-	-	-0.225*** (<0.01)	-	-	-	-
$(\hat{\mu}_{t\|t} - \mu_t)^2$	-	0.060*** (<0.01)	-	-	-	-	0.218*** (<0.01)	-	-	-
$\|\hat{\mu}_{t\|t} - \mu_t\|$	-	-	0.128*** (<0.01)	-	-	-	-	0.448*** (<0.01)	-	-
$(\hat{\mu}_{t\|t} - \mu_t)^+$	-	-	-	0.101*** (<0.01)	-	-	-	-	0.448*** (<0.01)	-
$(\hat{\mu}_{t\|t} - \mu_t)^-$	-	-	-	-0.155*** (<0.01)	-	-	-	-	-0.601*** (<0.01)	-
$\hat{\pi}_{t\|t}(1 - \hat{\pi}_{t\|t})$	-	-	-	-	0.025*** (<0.01)	-	-	-	-	0.086*** (<0.01)
$\overline{R}^2$	14.07%	21.33%	17.27%	22.19%	4.83%	19.14%	29.93%	27.03%	33.72%	5.58%

Table 10: Panel regressions of pocket diagnostics on belief discrepancies. This table reports coefficient estimates and p-values (in brackets) from regressions of the local $\overline{R}^2$ (in percentage points) measure for return predictability or the binary pocket indicator that is one inside pockets with return predictability and is zero otherwise on an intercept, a dummy for being within the first or last 126 (half of the kernel regression bandwidth) periods of the sample, functions of the difference between the true simulated expected cash flows and the agent's filtered beliefs about expected cash flows, and their interactions with the dummy. The last regressor is the variance of the agent's filtered probability of being in regime 1. These measures of belief discrepancy and variance are normalized to have standard deviation 1. All specifications contain the intercept, the dummy, and its interactions with any other regressors that are included. The coefficient estimate is the average coefficient across simulations, and the p-values are computed by dividing the average by the standard deviation of the estimates across samples and multiplying by the square root of the number of simulations. The $\overline{R}^2$ is computed as the average $\overline{R}^2$ across simulations for each regression specification.

	Local $\overline{R}^2$			Pocket Indicator		
Variables	**Slope**	**p-val**	**$\overline{R}^2$ (in %)**	**Slope**	**p-val**	**$\overline{R}^2$ (in %)**
			NBER Recession indicator			
dy	0.195***	(0.00)	1.74	-0.030*	(0.06)	0.15
tbl	0.306***	(0.00)	4.15	0.141***	(0.00)	2.36
tsp	0.634***	(0.00)	17.03	0.350***	(0.00)	17.58
rvar	0.152***	(0.00)	0.82	0.033*	(0.08)	0.22
			BW Index			
dy	0.001	(0.91)	-0.01	-0.022***	(0.00)	1.31
tbl	0.123***	(0.00)	6.29	0.030**	(0.04)	0.98
tsp	0.012	(0.62)	0.04	0.010	(0.51)	0.12
rvar	0.193***	(0.00)	11.60	0.081***	(0.00)	9.67
			AEM Leverage Factor			
dy	-0.034***	(0.00)	0.75	-0.024***	(0.00)	1.35
tbl	-0.053***	(0.00)	1.04	-0.040***	(0.00)	1.48
tsp	-0.012	(0.34)	0.04	-0.011	(0.14)	0.11
rvar	0.003	(0.83)	-0.01	-0.005	(0.41)	0.03

Table 11: Regressions of pocket diagnostics on economic indicators. This table reports coefficient estimates and p-values (in brackets) along with the $\overline{R}^2$ value from regressions of the local $\overline{R}^2$ (in %) measure for return predictability or the binary pocket indicator that is one inside pockets with return predictability and is zero otherwise on an intercept and either the NBER recession indicator (top panel), the Baker-Wurgler sentiment index (middle panel), or the Adrian-Etula-Muir leverage factor (bottom panel). All regressions use daily data with the samples described in the caption to Table 1 intersected with the samples available for the factors. The data for the BW index goes through the end of 2016, while the data for the AEM leverage factor goes through the end of 2009. Both the BW index and AEM leverage factors are normalized to have standard deviation 1 over the regression sample. The p-values are computed using Newey-West standard errors.

Variables	Full sample	In-pocket (real time)	Out-of-pocket (real time)	Difference in MSFE
		Panel A: 1-sided Kernel estimates		
dy	-3.708	1.346	-3.932	0.013* (0.06)
tbl	-2.922	0.455	-4.054	0.022** (0.04)
tsp	-3.594	0.296	-3.962	0.017 (0.14)
rvar	-2.146	0.334	-2.155	0.037 (0.15)
		Panel B: 2-sided Kernel estimates with reflected data		
dy	1.223	3.452	0.675	0.011* (0.06)
tbl	0.868	1.510	-0.651	0.034*** (< 0.01)
tsp	0.860	1.470	0.301	0.022** (0.04)
rvar	0.705	0.854	0.153	0.019 (0.18)

Table 12: Out-of-sample measures of forecasting performance. This table reports the Diebold and Mariano (1995) test statistics for out-of-sample return predictability measured relative to a prevailing mean model that assumes constant expected excess returns. Panel A uses a purely backward-looking kernel to compute forecasts. Panel B uses a two-sided kernel combined with data reflection as described in Chen and Hong (2012) (note that this is still a real-time forecast because only historical data is used). The DM 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 "real-time" pocket is defined as a period where the local R^2 of the time-varying coefficient model estimated using a backward-looking kernel is above 1% for more than 10 days. p-values for the difference in mean squared forecast errors (MSFE) are computed using a permutation test. The differences in forecast errors are randomly permuted 10,000 different ways and the the empirical distribution of resulting differences in MSFE are computed using the same sample split for in vs. out-of-pocket periods. The p-values are computed as the fraction of observations in the simulated distribution of differences that are greater than the actual sample statistic.
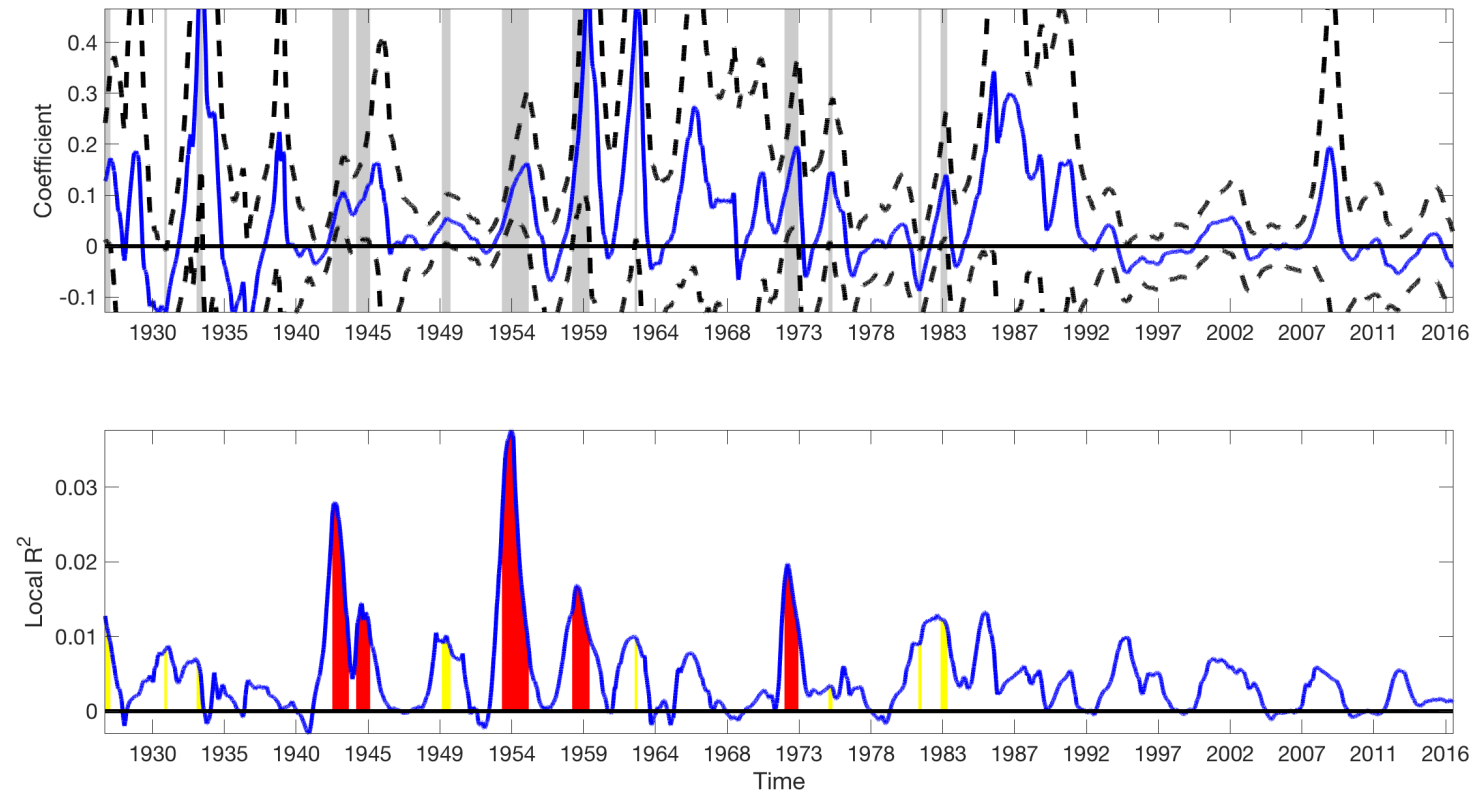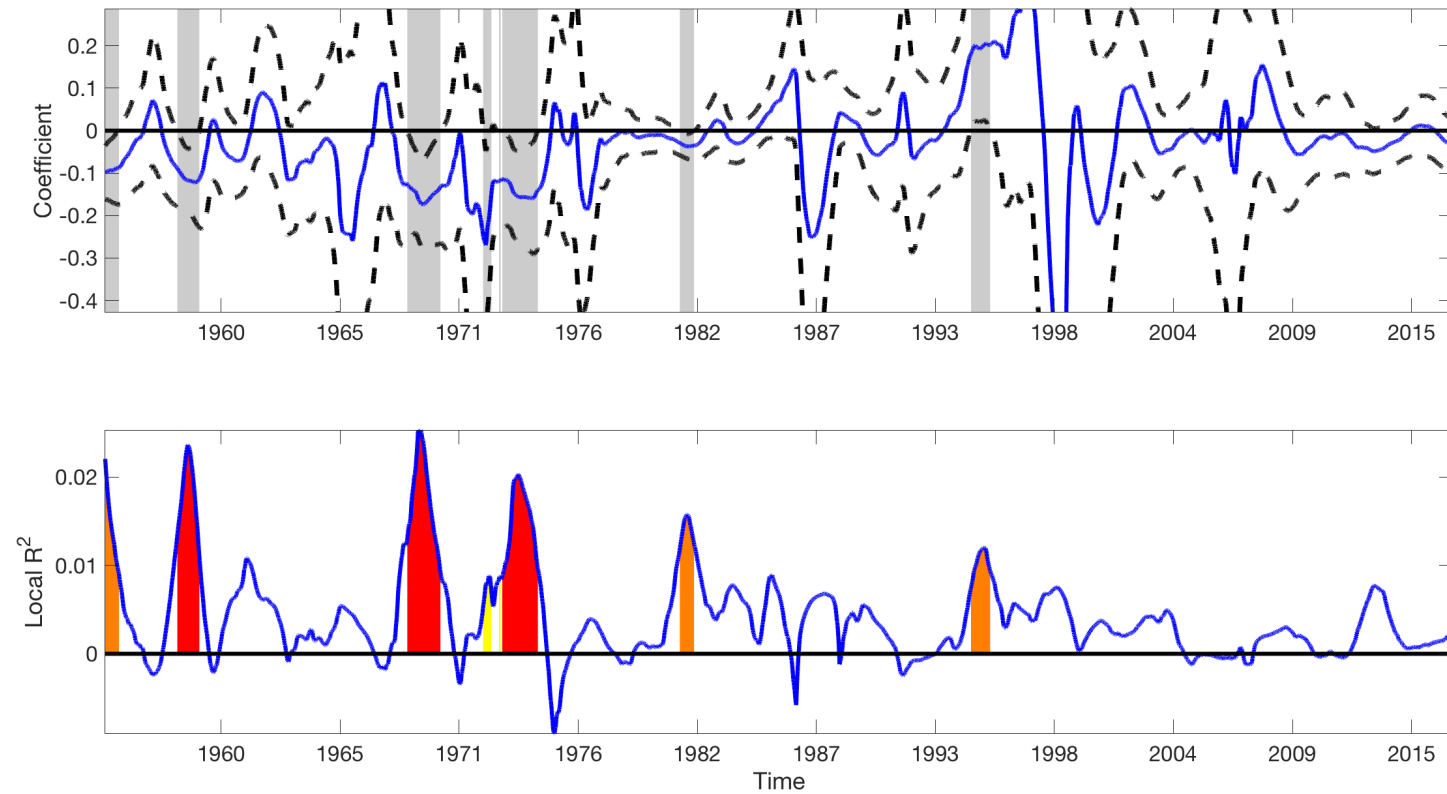
Figure 1: Local return predictability from the dividend yield. The top panel in this figure plots non-parametric kernel estimates of the local slope coefficient from a regression of daily excess stock returns on the lagged dividend yield. Dashed lines represents plus or minus two standard error bands. The bottom panel plots the local $\overline{R}^2$ measure with shaded areas tracking periods identified as pockets of return predictability using a 5% critical value. The shaded areas represent the integrated $\overline{R}^2$ inside pockets with areas colored in red representing pockets that have less than a 5% chance of being spurious, areas colored in orange representing pockets that have between a 5% and a 10% chance of being spurious, and areas colored in yellow representing pockets that have more than 10% chance of being spurious.
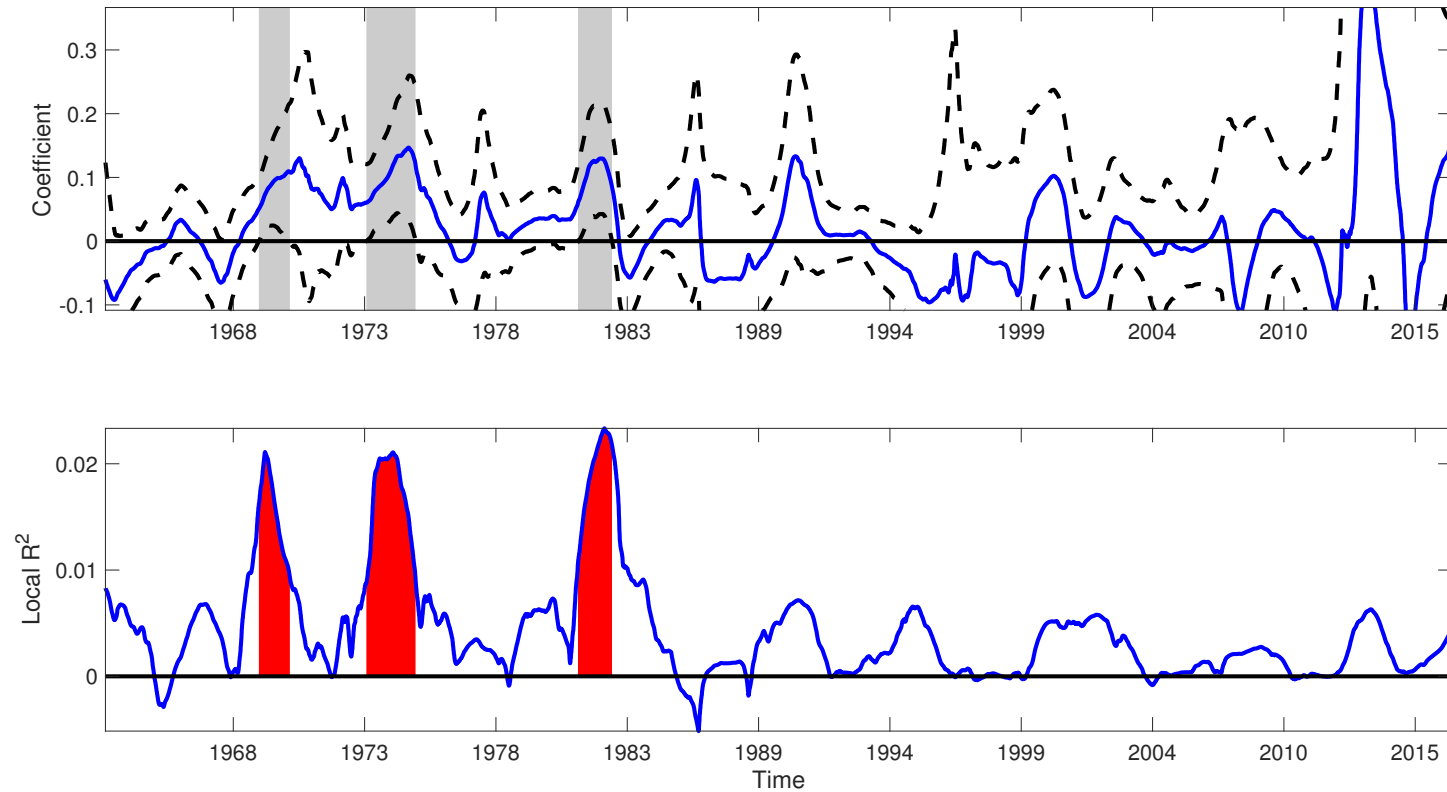
Figure 2: Local return predictability from the T-bill rate. The top panel in this figure plots non-parametric kernel estimates of the local slope coefficient from a regression of daily excess stock returns on the lagged T-bill rate. Dashed lines represents plus or minus two standard error bands. The bottom panel plots the local $\overline{R}^2$ measure with shaded areas tracking periods identified as pockets of return predictability using a 5% critical value. The shaded areas represent the integrated $\overline{R}^2$ inside pockets with areas colored in red representing pockets that have less than a 5% chance of being spurious, areas colored in orange representing pockets that have between a 5% and a 10% chance of being spurious, and areas colored in yellow representing pockets that have more than 10% chance of being spurious.
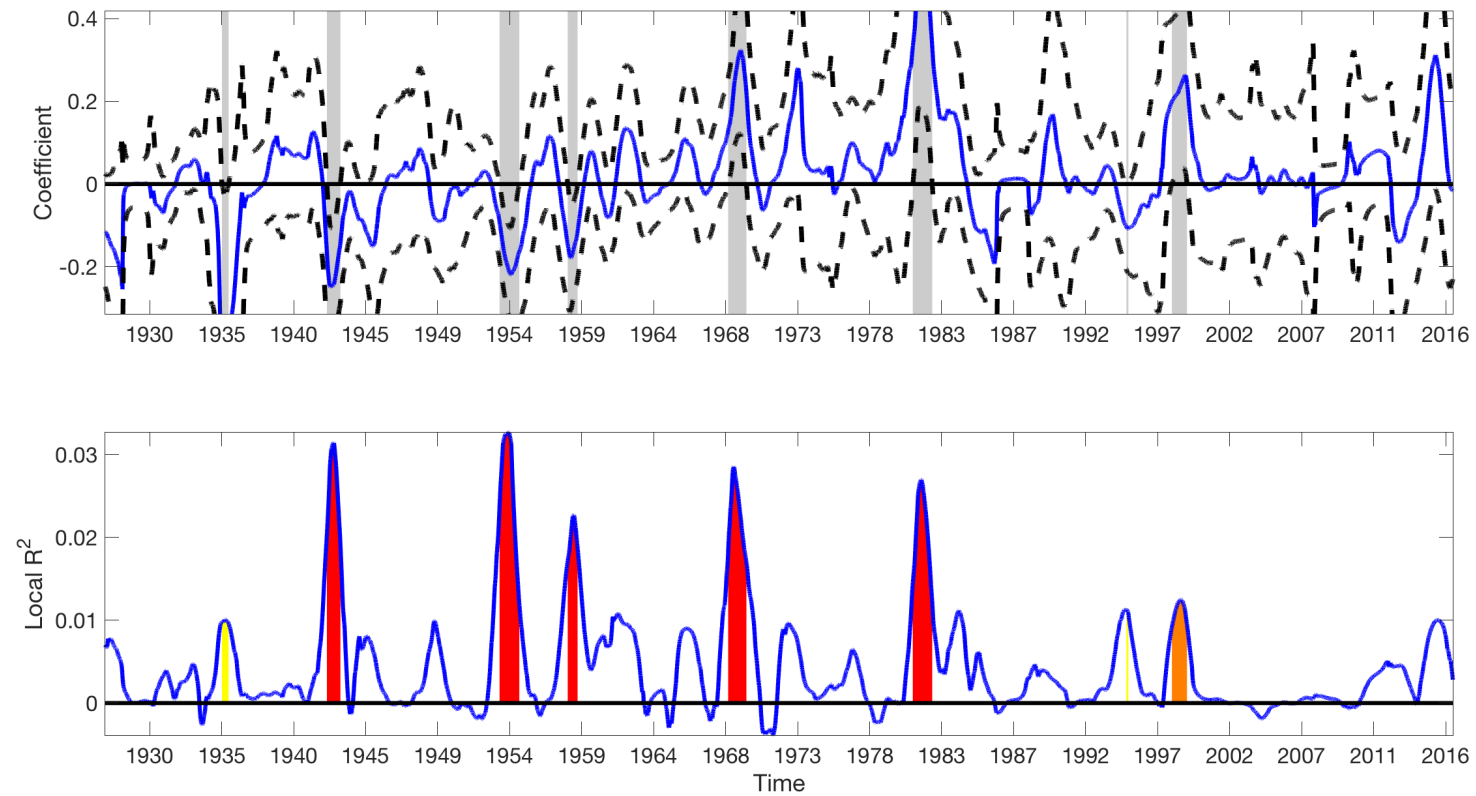
Figure 3: Local return predictability from the term spread. The top panel in this figure plots non-parametric kernel estimates of the local slope coefficient from a regression of daily excess stock returns on the lagged term spread. Dashed lines represents plus or minus two standard error bands. The bottom panel plots the local $\overline{R}^2$ measure with shaded areas tracking periods identified as pockets of return predictability using a 5% critical value. The shaded areas represent the integrated $\overline{R}^2$ inside pockets with areas colored in red representing pockets that have less than a 5% chance of being spurious, areas colored in orange representing pockets that have between a 5% and a 10% chance of being spurious, and areas colored in yellow representing pockets that have more than 10% chance of being spurious.
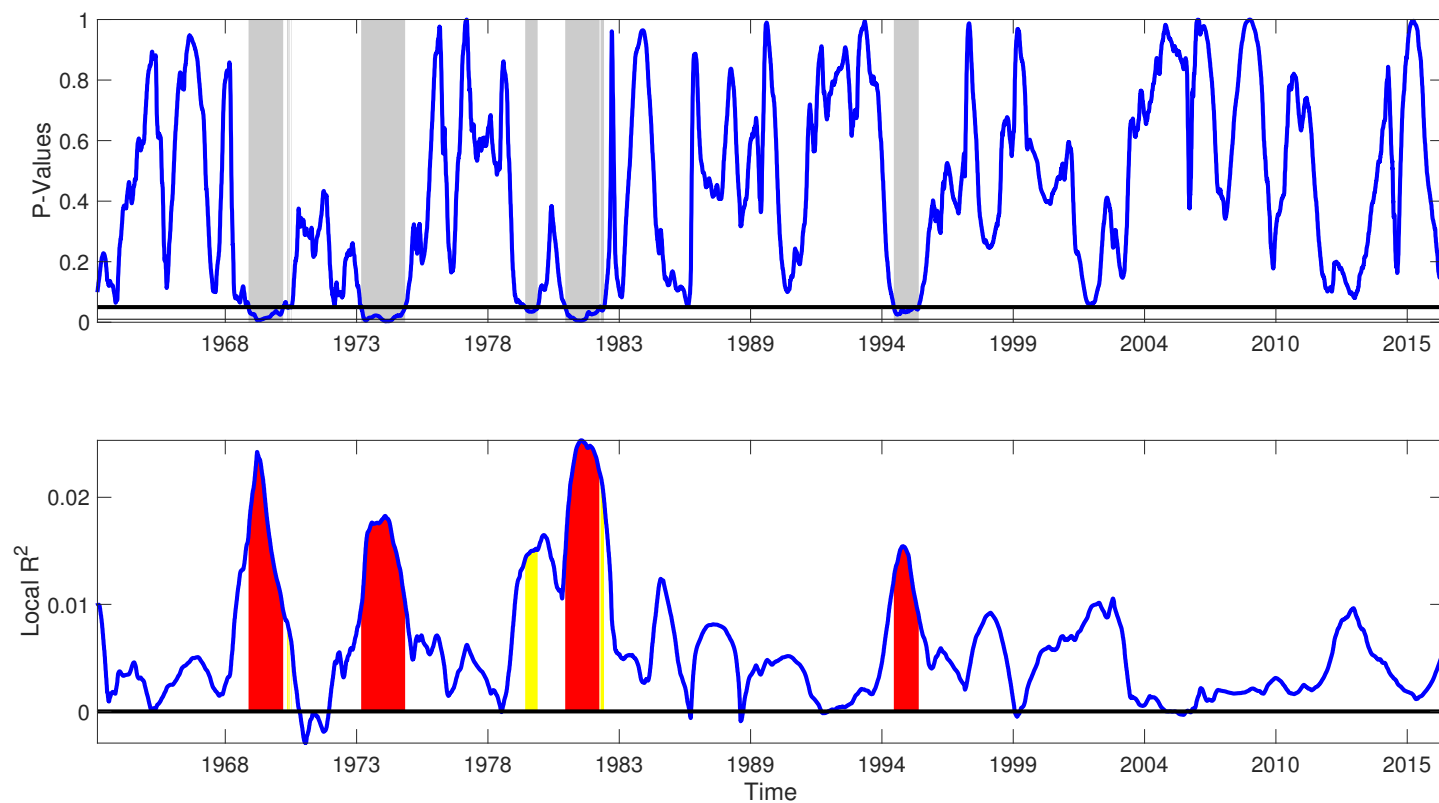
Figure 4: Local return predictability from the realized variance. The top panel in this figure plots non-parametric kernel estimates of the local slope coefficient from a regression of daily excess stock returns on the lagged realized variance. Dashed lines represents plus or minus two standard error bands. The bottom panel plots the local $\overline{R}^2$ measure with shaded areas tracking periods identified as pockets of return predictability using a 5% critical value. The shaded areas represent the integrated $\overline{R}^2$ inside pockets with areas colored in red representing pockets that have less than a 1% chance of being spurious, areas colored in orange representing pockets that have between a 5% and a 10% chance of being spurious, and areas colored in yellow representing pockets that have more than 10% chance of being spurious.

Figure 5: Local return predictability from a bivariate model with the T-bill rate and the term spread. The top panel in this figure plots the p-value from an F-test of the joint significance of non-parametric kernel estimates of the local slope coefficients from a regression of daily excess stock returns on the lagged T-bill rate and lagged term spread. The bottom panel plots the local $\bar{R}^2$ measure with shaded areas tracking periods identified as pockets of return predictability using a 5% critical value. The shaded areas represent the integrated $\bar{R}^2$ inside pockets with areas colored in red representing pockets that have less than a 1% chance of being spurious, areas colored in orange representing pockets that have between a 5% and a 10% chance of being spurious, and areas colored in yellow representing pockets that have more than 10% chance of being spurious.
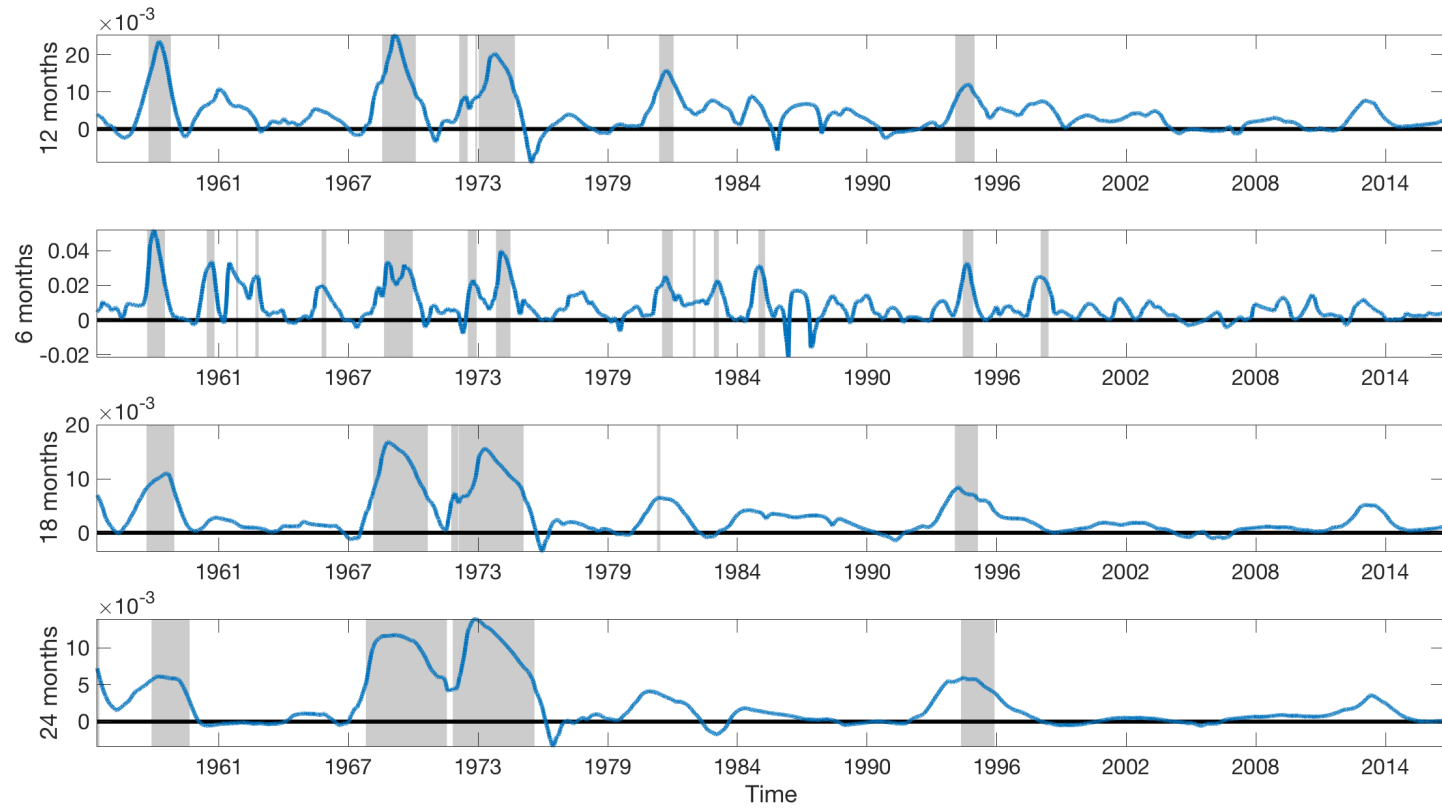
Figure 6: Robustness of local R^2 measures. The labels refer to different empirical specifications for identifying pockets. The number refers to the effective sample size (in months) of the kernel used for estimation. The letters refer to whether the T-bill rate was used in levels (ND), or was detrended using a 12-month trailing average (D). Each line corresponds to the pocket indicator for the corresponding specification.
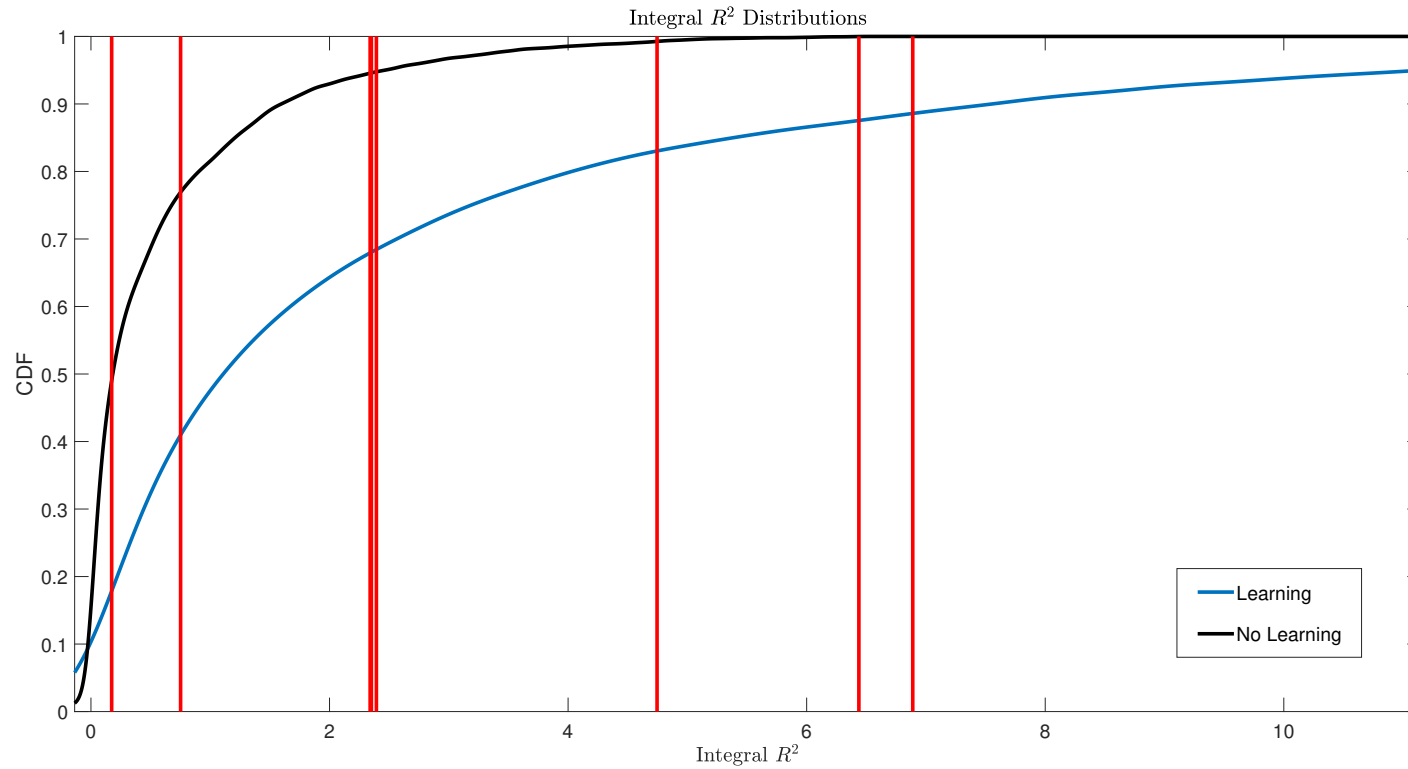
Figure 7: Cumulative distribution functions of integral R^2 under no-learning and learning in the cash flow process (5% significance). The figure shows kernel-density estimated cdfs of integral R^2 values computed using simulations from the predictive systems model with regime switching in the cash flow process assuming either observed states and no learning or unobserved states and learning. Pockets are defined using a 5% critical value.
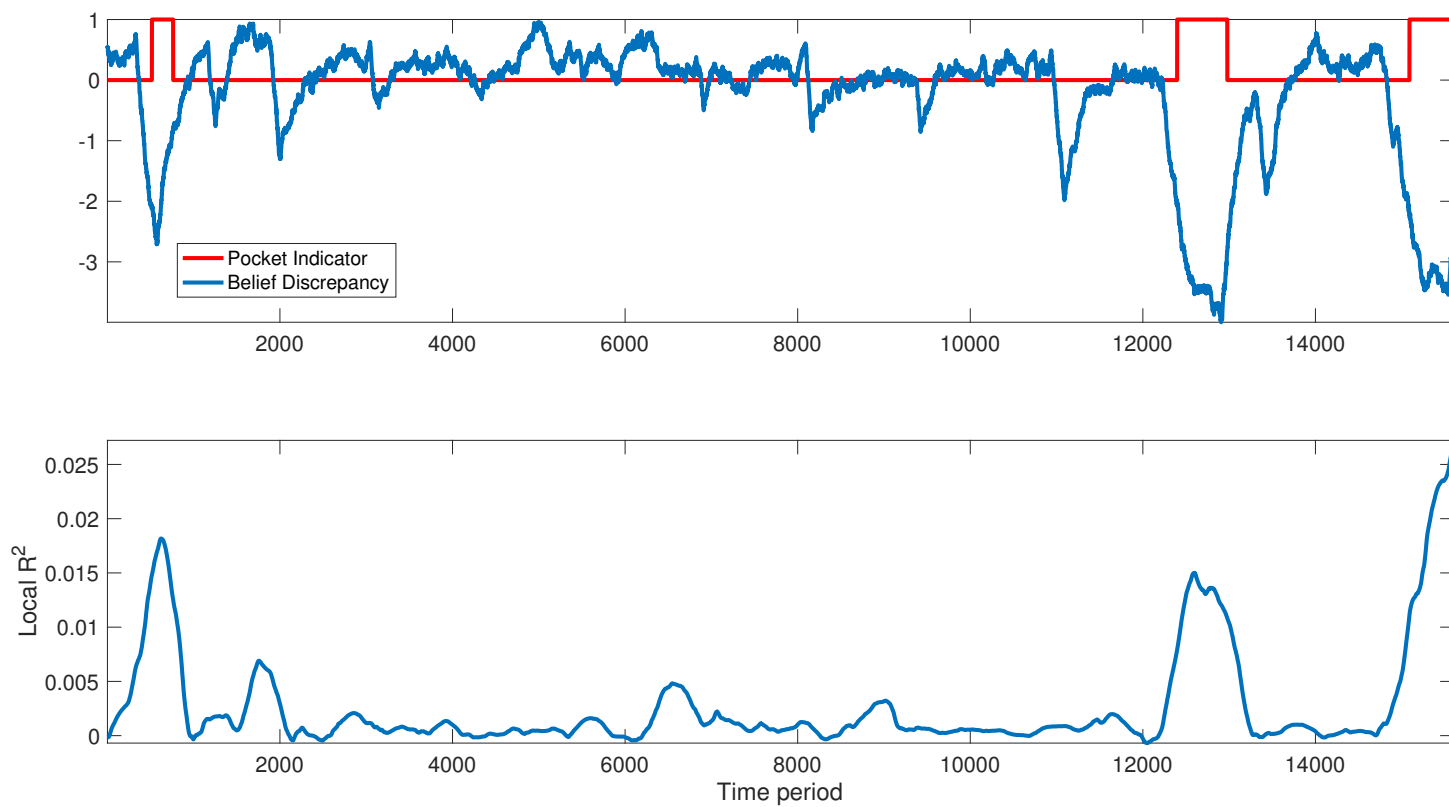
Figure 8: Correlation of belief discrepancy with return predictability pockets. This figure plots a particular realization from the learning model simulations. The top panel plots the belief discrepancy along with the indicator for periods identified as return predictability pockets. The bottom panel plots the local R^2 from the two-sided kernel regression of simulated excess returns on the simulated lagged T-bill rate.
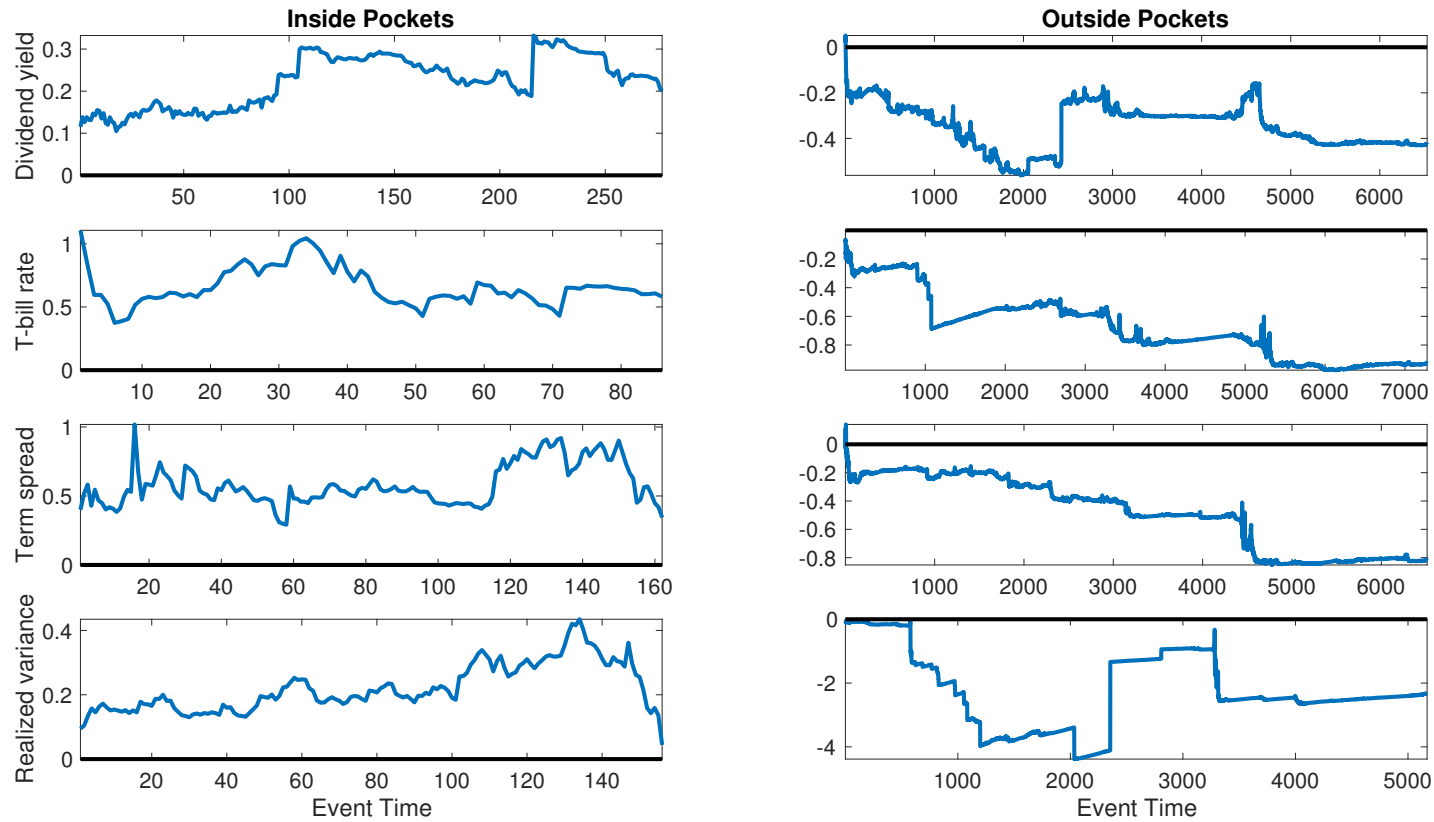
Figure 9: Cumulative sum of squared forecast errors averaged across pockets of predictability (left column) or periods outside pockets of predictability (right column). The cumulative sums are normalized by the square root of the number of periods. Pockets are identified as consecutive periods of more than 10 days with a local R^2 greater than 1% computed using a backward-looking kernel. The predictor variables used in the analysis are listed to the left of each row of panels.