NBER WORKING PAPER SERIES

FACTOR TIMING

Valentin Haddad
Serhiy Kozak
Shrihari Santosh

Working Paper 26708
http://www.nber.org/papers/w26708

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2020

We thank John Campbell, Mikhail Chernov, John Cochrane, Julien Cujean, Robert Dittmar, Kenneth French, Stefano Giglio, Bryan Kelly, Ralph Koijen, Hugues Langlois, Lars Lochstoer, Mark Loewenstein, Tyler Muir, Stefan Nagel, Nikolai Roussanov, Avanidhar Subrahmanyan, Michael Weber and seminar participants at AFA, Chicago, FIRS, LSE, Maryland, Michigan, University of Washington, and NBER for helpful comments and suggestions. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Factor Timing
Valentin Haddad, Serhiy Kozak, and Shrihari Santosh
NBER Working Paper No. 26708
January 2020
JEL No. G0,G11,G12

ABSTRACT

The optimal factor timing portfolio is equivalent to the stochastic discount factor. We propose and implement a method to characterize both empirically. Our approach imposes restrictions on the dynamics of expected returns which lead to an economically plausible SDF. Market-neutral equity factors are strongly and robustly predictable. Exploiting this predictability leads to substantial improvement in portfolio performance relative to static factor investing. The variance of the corresponding SDF is larger, more variable over time, and exhibits different cyclical behavior than estimates ignoring this fact. These results pose new challenges for theories that aim to match the cross-section of stock returns.

Valentin Haddad
University of California, Los Angeles
Anderson School of Management
Office C4.19
110 Westwood Plaza
Los Angeles, CA 90024
and NBER
vhaddad@ad.ucla.edu

Serhiy Kozak
University of Maryland
7621 Mowatt Lane
VMH 4453
College Park, MD 20742
https://www.serhiykozak.com
sekozak@umd.edu

Shrihari Santosh
University of Maryland
Van Munching Hall, #4416
7621 Mowatt Lane
College Park, MD 20742
shrisantosh@gmail.com

1 Introduction

Aggregate stock returns are predictable over time (e.g., Shiller 1981, Fama and French 1988), creating the scope for investors to engage in market timing. Factors beyond the aggregate market are sources of risk premia in the cross-section of assets (e.g., Fama and French 1993), creating the basis for factor investing. How valuable is it to combine these two ideas and construct the optimal *factor timing* portfolio, which unifies cross-sectional and time-series predictability of returns? Answering this question has economic importance: the optimal portfolio is equivalent to the stochastic discount factor (SDF). Therefore, if factor timing is relevant for the optimal portfolio, we should account for this fact when estimating the SDF.

Empirically determining the value of factor timing appears difficult because it requires measuring the predictability of many returns, which opens the door for spurious findings. We propose a new approach to overcome this challenge. Imposing that the implied SDF is not too volatile leads us to focus only on estimation of predictability of the largest principal components of the factors. We find that these statistical restrictions are crucial to construct robust forecasts of factor returns.

Taking into account the predictability of the factors leads to an estimated SDF which exhibits drastically different properties than estimates which assume constant factor premia, the standard approach in previous work. Our estimated SDF is more volatile: its variance increases from 1.66 to 2.96. Moreover the benefits to factor timing are strongly time-varying, which results in much more heteroskedasticity of the SDF. These fluctuations in SDF variance exhibit a very different pattern than estimates which only account for the predictability of market returns. They occur mostly at shorter business-cycle frequencies, and are correlated with different macroeconomic variables.

Our empirical analysis focuses on fifty standard stock "anomaly" portfolios that have been put forward in previous work as capturing cross-sectional variation in expected returns. To characterize an optimal portfolio and the SDF, we rely on two restrictions. First, we assume that the SDF which prices stocks implies that a *conditional* factor model holds with respect to these portfolios. Our setting enriches the previous literature by allowing for time-varying loadings of the SDF on the factors. With this assumption, we entertain the possibility that factor timing strategies are profitable. Second, we assume that the pricing kernel does not generate excessively high Sharpe ratios.[1] Such near-arbitrage opportunities would be eliminated by arbitrageurs in equilibrium (Kozak et al., 2018). When the cross-section of returns has a strong factor structure, this assumption implies that the time-series variation in risk premia is mostly driven by time-varying exposure of the SDF to the largest sources of variation in realized returns. If this were not the case, small components would be highly predictable, generating implausibly large Sharpe ratios. The fifty portfolios we consider exhibit such a factor structure, with a stable covariance matrix over time, which allows us to exploit this idea empirically. We focus on the largest sources of variation by restricting our attention to the first five principal components (PCs) of anomaly returns, which explain 60% of the variation in realized returns. This dimension reduction allows for robust estimation of their predictability, and therefore the SDF. As such, our approach is a regularization of the left-hand-side of the predictability problem — "which factors are predictable?" — rather than the right-hand side — "which variables are useful predictors?" We take a simple stance on this second issue by using only the book-to-market ratio of each portfolio as a measure to predict its returns.

We find that the PCs of anomalies are strongly predictable. For the two most predictable components, the first and fourth PCs, their own book-to-market ratios predict future monthly returns with an out-of-sample R^2 around 4%, about four times larger than that of predicting the aggregate market return. We confirm these strong relations are not driven by statistical issues arising in small samples. The predictability of the dominant PC portfolios captures

[1]Using simple economically-motivated restrictions on asset prices to stabilize statistical inference has notable antecedents, for example Cochrane and Saa-Requejo (2000) and Campbell and Thompson (2007).

common variation in risk premia which allows us to form forecasts of individual anomaly returns. These forecasts yield a sizable total out-of-sample monthly R^2 around 1%. The observation that factor returns are robustly predictable lends support to the enterprise of factor timing. We confirm that this conclusion does not rely on the details of our implementation by varying, for example, the number of principal components, the construction of anomaly portfolios, and the horizon of predictability.

The key ingredient for our results is the dimension reduction of the set of portfolios to predict. One could instead separately estimate expected returns for each anomaly portfolio without recognizing the factor structure in returns and imposing the absence of near-arbitrage. We find this approach to be less fruitful: predicting each anomaly return with its own book-to-market ratio generates only half of the predictability our restrictions uncover. The out-of-sample robustness of our approach supports the validity of these restrictions. Interestingly, our approach complements the set of methods developed to choose among many predictors given a portfolio to predict. For example, we find that the 3-pass regression filter of Kelly and Pruitt (2013) provides more robust results when applied to the dominant components of anomalies rather than to each portfolio separately.

We use our results to construct an optimal factor timing portfolio. This allows us to quantify the investment benefits to factor timing. And, more importantly, we use it to characterize the properties of an SDF consistent with the evidence of these factor timing benefits. First, timing expected returns provides substantial investment gains; a pure factor timing portfolio achieves a Sharpe ratio of 0.71. This means that the conditional variance of the SDF is substantially larger than that inferred from static strategies alone. The benefits from timing market-neutral factors largely outweigh those from timing the aggregate market return and are comparable to those obtained by static factor investing. Second, these benefits vary over time: the SDF is strongly heteroskedastic. Variation in the maximum compensation for risk is driven by changes in the means of the factors and to a lesser extent changes in their variances. Again, these fluctuations are much more pronounced than fluctuations in the Sharpe ratio of the market portfolio. Third, the dynamics of the variance of the SDF differ from those of the market risk premium. The SDF variance evolves mostly at business cycle frequency rather than at longer horizons. However, it is not always related to recessions. More broadly, macroeconomic variables capturing variations in the price of market risk often have different relations with the SDF variance. Fourth, the contribution of various anomalies to the SDF exhibit interesting dynamics. For example, the loadings of size and value are procyclical while the loading of momentum is countercyclical.

To summarize, factor timing is very valuable, above and beyond market timing and factor investing taken separately. The changing conditional properties of the pricing kernel are mostly driven by market-neutral factors. The methods and facts we study in this paper are only the beginning of the economic enterprise of understanding the evolution of drivers of risk premia. Our results suggest that theories developed to understand cyclical variation in the price of market risk (e.g. Campbell and Kyle (1993); Campbell and Cochrane (1999); Bansal and Yaron (2004); Barberis et al. (2015)) are unlikely to capture the *dynamic* properties of the cross-section of returns. Indeed, these models generate SDFs that are much less volatile and heteroskedastic than our estimated SDF. Further, they typically focus on a single common force driving variation in risk premia, at odds with the multiple dimensions we uncover. Finally, the properties of our estimated SDF provide a useful set of moments summarizing the properties of the rich cross-section of stocks, moments that future theories should target.

Related literature

This paper builds on the long literature which studies the time series predictability of returns, starting from Shiller (1981) and Fama and French (1988) for stocks, or Fama and Bliss (1987) for bonds.[2] While this early evidence is mostly about aggregate returns, our main focus is on understanding predictability of cross-sections of returns. Early work has extended the ideas of market predictability to specific anomalies: Cohen et al. (2003) for value or Cooper et al. (2004) and Daniel and Moskowitz (2016) for momentum. We aim to tackle the entire cross-section. With a similar goal, various papers such as Stambaugh et al. (2012) and Akbas et al. (2015) examine the ability of a single variable to forecast all anomaly returns, and thus their common component. These papers implicitly assume a single source of time-varying risk premia; our approach entertains multiple. This multiplicity is complementary to that of Kelly and Pruitt (2013): while we study how to predict many returns with a factor-specific explanatory variable (its own valuation ratio), they predict a single return with a wide cross-section of valuation ratios. That is, we ask a question of what we should predict, while they study how to select variables which are most useful in making such a prediction.

Another strand of the literature studies the predictability of returns anomaly by anomaly (or stock by stock) without imposing any structure on the implied pricing kernel. Recent prominent examples are Campbell et al. (2009) and Lochstoer and Tetlock (2016), who use panel VAR techniques to forecast firm-level expected returns, then aggregate the estimates into portfolios. Asness et al. (2000); Cohen et al. (2003); Arnott et al. (2016a,b); Baba Yara et al. (2018) and others use valuation ratios to forecast anomaly returns. Greenwood and Hanson (2012) forecast characteristics based anomalies using their "issuer-purchaser" spread—the difference in the average characteristic for net equity issuers vs repurchasers. Conversely, some papers such as Ilmanen and Nielsen (2015), Asness (2016), and Asness et al. (2017) find that cross-sectional long-short factors are not very predictable by valuation ratios. Irrespective of their conclusion, all of these papers forecast a single return at a time, ignoring the correlation across assets. Implicitly, they assume there are potentially as many independent sources of time-varying risk premia as there are assets. We, instead, study

[2]See Koijen and Van Nieuwerburgh (2011) for a survey of recent work on the topic.

common sources of predictability across all anomalies in a restricted setup and then infer the implied predictability of each anomaly. Such an approach brings important statistical advantages in terms of dimensionality reduction. In Section 3 we show our method yields greater out-of-sample predictability than various alternatives.

Another literature develops methods for dealing with the large dimensionality of the cross-section. Freyberger et al. (2018) use an adaptive group lasso method to test which characteristics provide independent information for the cross section of expected returns on individual stocks. Kozak et al. (2019) model SDF risk prices as linear functions of characteristics, while Kozak (2019) extends this approach to capture arbitrary non-linearities. Kelly et al. (2018) model stock betas as linear functions of characteristics. Light et al. (2017), Kelly et al. (2018), and Giglio and Xiu (2018) employ latent factor analysis. Kozak et al. (2018) and Kozak et al. (2019) use no near-arbitrage to argue for the use of principal components analysis (PCA). All these dimension-reduction techniques are somewhat related. For example, Kelly et al. (2018) show that if the cross-sectional correlation matrix of stock characteristics is constant their latent factors exactly correspond to largest PCs of characteristics-managed anomaly portfolios. We also use PCA to handle a high-dimensional factor space. However, we differ from this previous work in an important dimension. In all these papers, expected returns on anomaly portfolios are approximately constant or vary only due to time-varying volatility. In contrast, we entertain the possibility and find evidence for significant time-variation in prices of risk on these factors.

Finally, some papers highlight the quantitative importance of conditioning information for the SDF. Gallant et al. (1990) use conditioning information and asset prices to derive lower bounds on the unconditional variance of the SDF. We use upper bounds on the unconditional variance to derive restrictions on the impact of conditioning information. Chernov et al. (2018) propose testing asset-pricing models using multi-horizon returns and find that many standard empirical models of the SDF are rejected, exactly because they lack time-variation in how the SDF loads on the factors. Moreira and Muir (2017) document benefits to volatility timing, implying changes in volatility play a meaningful role in the heteroskedasticity of the SDF.

2 Methodology

We are interested in assessing the benefits of timing strategies for a cross-section of excess returns $\{R_{i,t}\}_{i\in I}$; our main empirical setting is the cross-section of stock returns. Studying these timing benefits is important for the purpose of optimal portfolio choice, but also to understand the economic forces shaping equilibrium prices. To measure these benefits requires measuring the dynamics of risk premia. The connection between factor timing benefits, the stochastic discount factor, and predictability is best illustrated by the following decomposition. In Appendix A, we show that if asset returns are uncorrelated, the average maximum conditional Sharpe ratio can be expressed as

$$\mathbb{E}\left(\mathrm{SR}_t^2\right) = \mathbb{E}\left[\mathrm{var}_t\left(m_{t+1}\right)\right] = \sum_i \frac{\mathbb{E}\left[R_{i,t+1}\right]^2}{\sigma_i^2} + \sum_i \left(\frac{R_i^2}{1-R_i^2}\right). \tag{1}$$

The first equality shows that the average maximum squared Sharpe ratio coincides with the expected variance of m_{t+1}, where m_{t+1} is the minimum variance stochastic discount factor which prices the set of returns. The second equality shows that these quantities combine two elements. The first term is an unconditional part reminiscent of static Sharpe ratios: the sum of ratios of the squared average return to σ_i^2, the conditional variance of the asset return.[3] Of interest to us is the second term which encodes predictability. This term is increasing in the R_i^2s, the maximum predictive R-squared when forecasting asset i, $R_i^2 = 1 - \sigma_i^2/\mathrm{var}\left(r_{i,t+1}\right)$.

Without any structure, it is challenging to create robust forecasts for all returns. Spurious results are likely, especially with many assets to predict. In this section, we show how two

[3]For simplicity of exposition, we focus on a homoskedastic setting.

simple restrictions help address these issues. First, we follow the literature and assume that a relatively small number of stock characteristics capture pricing-relevant information. Equivalently, the assets are conditionally priced by a factor model, the main motivation behind factor timing portfolio strategies. Second, we assume that prices feature no near-arbitrage opportunities. These assumptions imply that measuring the predictability of the largest principal components of the set of factors is enough to characterize expected returns.[4] This strong dimension reduction allows us to use the standard tools for forecasting single return series to measure this predictability.

2.1 Factor Model, Factor Investing, and Factor Timing

First, we impose some structure on the pricing kernel. Start with the minimum variance SDF in the span of N individual stock (asset) excess returns R_{t+1} (Hansen and Jagannathan, 1991):

$$m_{t+1} = 1 - b_t' \left(R_{t+1} - \mathbb{E}_t \left[R_{t+1} \right] \right), \tag{2}$$

which satisfies the fundamental relation $0 = \mathbb{E}_t \left[m_{t+1} R_{t+1} \right]$. We restrict the behavior of the loading b_t. As in Kelly et al. (2018), Kozak et al. (2019), Freyberger et al. (2018), and Kozak (2019), we use stock characteristics to reduce the dimensionality of the return space and the SDF. In particular, we assume that cross-sectional heterogeneity in risk prices b_t can be largely captured by K observable stock characteristics, C_t, with $K \ll N$. Time-series variation in the importance of each characteristic is summarized by the vector δ_t of size $K \times 1$.

[4]Specifically, the expected return on any asset is the product of the asset's conditional loading on, and the expected return of these few components.

Assumption 1. *Stock-level SDF loadings can be represented as*

$$\underbrace{b_t}_{N\times 1} = \underbrace{C_t}_{N\times K} \underbrace{\delta_t}_{K\times 1}, \tag{3}$$

where C_t is an $N \times K$ matrix of stock characteristics and δ_t is a $K \times 1$ vector of (possibly) time-varying coefficients, and $K \ll N$.

Substituting Equation 3 into Equation 2, we obtain an alternative SDF representation

$$m_{t+1} = 1 - \delta_t' \left(F_{t+1} - \mathbb{E}_t \left[F_{t+1}\right]\right), \tag{4}$$

where $F_t = C_{t-1}' R_t$ are "characteristics-managed" factor portfolios. For example if one element of C_t is the market capitalization of a firm, the corresponding factor is the market return. If the characteristic is an indicator taking values -1, 0, or 1 depending on whether a stock is in the upper or lower quantiles of a characteristic, the corresponding factor is a standard sort-based portfolio in the style of Fama and French (1992). We can now interpret δ_t as time-varying prices of risk on these factor portfolios.

Assumption 1, therefore, allows us to use characteristics to parsimoniously describe the cross-section but also permits time-variation in a relatively small number of factor risk prices. Consequently, it results in a large dimensionality reduction in the number of variables determining the SDF. However, it is rich enough to consider meaningful variation in factor expected returns, the fundamental idea behind factor timing. For example, this is a richer setting than in Kozak et al. (2019) which assumes the mapping from stock characteristics to SDF coefficients is constant, $b_t = C_t \delta$.[5] In such a specification, there is no scope for factor timing since SDF coefficients are equal to weights in the maximum Sharpe ratio portfolio. Another way to see that our model entertains factor timing is to notice that our SDF can

[5] While they focus on slightly different functional forms, Kelly et al. (2018) and Freyberger et al. (2018) also assume a constant mapping between characteristics and factor exposures.

alternatively be represented as a factor model, with arbitrarily changing expected factor returns.

Lemma 1. *(Conditional Factor Model) A conditional factor model holds:*

$$\mathbb{E}_t\left[R_{j,t+1}\right] = \beta_{jt}'\Sigma_{F,t}\delta_t = \beta_{jt}'\mathbb{E}_t\left[F_{t+1}\right]. \tag{5}$$

The equivalence between the factor model and Equation 4 is given by $\delta_t' = \Sigma_{F,t}^{-1}\mathbb{E}_t\left(F_{t+1}\right)$, where $\Sigma_{F,t}$ is the conditional covariance matrix of the factors. This relation also highlights that our model can generate interesting risk premium variation even in a homoskedastic setting for returns, where the variance-covariance matrix of the factors $\Sigma_{F,t}$ and the betas β_{jt}' are constant over time. This result arises because δ_t controls how the SDF loads on the factors, and therefore changes their price of risk.[6]

In our framework, because the factors completely capture the sources of risks of concern to investors, optimal portfolios can be constructed from only these few factors—the so-called mutual fund theorem. Factor timing strategies are the dynamic counterpart of this observation; as the properties of the factors change, an investor should adjust her portfolio weights accordingly. For example, the maximum conditional Sharpe ratio return is obtained by:

$$R_{t+1}^{\text{opt}} = \mathbb{E}_t\left[F_{t+1}\right]'\Sigma_{F,t}^{-1}F_{t+1}. \tag{6}$$

Knowledge of the conditional risk premia of the factors is crucial to form this and other dynamic strategies.

While going from individual assets to factors provides some useful dimension reduction and stabilization of the covariance structure, we are still left with many factor returns to

[6] Variations in δ_t could occur for multiple reasons. For example, investors' aversion to the various sources of factor risk could change over time. Or, their exposure to these risks, for example through consumption, could change over time.

forecast.[7] A multitude of empirical results and theoretical motivations have put forward a large number of potential factors, leading to the emergence of what Cochrane (2011) calls the "factor zoo." Including potentially irrelevant factors does not affect the theoretical performance of a factor model; the SDF would just have zero loading on these factors. However, including too many factors leads to greater probability of estimating spurious return predictability in finite samples. We now turn to a second assumption which helps discipline our empirical analysis.

2.2 Absence of Near-Arbitrage

Various authors have used the idea of the absence of "good deals" or "no near-arbitrage" opportunities to add economic discipline to statistical exercises. For example, Cochrane and Saa-Requejo (2000) impose an upper bound on the conditional variance of the SDF to derive bounds on option prices. Ross (1976) originally proposed a bound on the squared Sharpe ratio for an unconditional empirical implementation of his APT in a finite-asset economy. Such a bound on the maximum squared Sharpe ratio is immediately equivalent to an upper bound on the variance of the SDF, m_{t+1} (Hansen and Jagannathan, 1991). Kozak et al. (2018) use a similar argument to show that unconditionally, the large principal components of anomaly returns must explain most of the cross-sectional variation in average returns. In our setting with time-varying risk premia, there is no such thing as *the* maximum Sharpe ratio, rather a conditional maximum Sharpe ratio at each point in time. What is an appropriate metric for a "good deal in this setting?" We argue that it is the average maximum conditional squared Sharpe ratio, $\mathbb{E}\left[\mathrm{SR}_t^2\right]$.

Assumption 2. *(Absence of near-arbitrage) There are usually no near-arbitrage opportunities: average conditional squared Sharpe ratios are bounded above by a constant.*

[7] This stabilization role is discussed for example in Brandt et al. (2009); Cochrane (2011); Kozak et al. (2018).

There are two interpretations of this restriction on Sharpe ratios which help clarify its economic content. First, this quantity corresponds to the average certainty equivalent for a mean-variance investor who optimally uses conditioning information. For such an investor with risk-aversion parameter γ, her certainty equivalent at time t is $\frac{\mathrm{SR}_t^2}{2\gamma}$ where SR_t^2 is the maximum conditional squared Sharpe ratio (see Section A.2). Taking the unconditional expectation of SR_t^2 measures, on average, the welfare gain from investing in risky assets. Second, the average conditional variance of the SDF in Equation 4 is also equal to its unconditional variance since it has constant conditional mean. Therefore, our bound is also a bound on the unconditional variance of the SDF. Third, it equivalently provides an upper bound on the maximum unconditional squared Sharpe ratio when considering all possible dynamic factor strategies. This unconditional value measures the welfare gain for a mean-variance investor who does not have direct access to conditioning information but follows a buy and hold strategy from a menu of managed portfolios (Ferson and Siegel, 2001).[8] Both of these interpretations highlight that our bound forbids "good deals" on average, but not always.

We now show that Assumption 2 leads to further dimensionality reduction. First, notice that because the maximum conditional Sharpe ratio is invariant to rotations of the asset space, we can apply Equation 1 with the PC decomposition of returns. Letting $PC_{i,t+1}$ be the ith principal component portfolios of the factors F, and λ_i the corresponding eigenvalue, we have

$$\mathbb{E}\left(\mathrm{SR}_t^2\right) = \sum_{i=1}^{K} \frac{\mathbb{E}\left[PC_{i,t+1}\right]^2}{\lambda_i} + \sum_{i=1}^{K} \left(\frac{R_i^2}{1 - R_i^2}\right), \tag{7}$$

where the summation is across all K PC portfolios. Again, the first term represents the benefits of static factor investing. It is the squared Sharpe ratio of an optimal static factor portfolio. The second term is our focus in this paper and captures the amount of pre-

[8]Gallant et al. (1990) use this estimated unconditional variance to empirically test asset pricing models in the presence of conditioning information.

dictability for each principal component. The more a principal component can be predicted, the better portfolio performance an investor can obtain. This second term represents the incremental benefit of optimally timing the factors.

Second, we can ask how much each PC portfolio contributes to the total predictability of returns. As a measure of the total amount of predictability we define the total R^2 as

$$\begin{aligned} R^2_{\text{total}} &\equiv \frac{\operatorname{tr}\left[\operatorname{cov}\left(\mathbb{E}_t\left[F_{t+1}\right]\right)\right]}{\operatorname{tr}\left[\operatorname{cov}\left(F_{t+1}\right)\right]} = \frac{\operatorname{tr}\left[\operatorname{cov}\left(\mathbb{E}_t\left[PC_{t+1}\right]\right)\right]}{\operatorname{tr}\left[\operatorname{cov}\left(PC_{t+1}\right)\right]} \\ &= \sum_{i=1}^{K}\left(\frac{R_i^2}{1-R_i^2}\right)\frac{\lambda_i}{\lambda}, \end{aligned} \tag{8}$$

where $\lambda = \sum \frac{\lambda_i}{1-R_i^2} \approx \sum \lambda_i$ is the total unconditional variance of returns (Appendix A.3 shows the derivations). This quantity measures the total amount of predictability for the cross-section of returns, and has the useful feature to be invariant by rotation of the asset space. The second line shows that the total R^2 comes from the predictability of each of the PCs, weighted by their importance in explaining the factors. In the case of a single asset, the formula reduces to the standard predictive R^2.

What happens when the set of portfolios exhibits a factor structure, that is some λ_i are large while others are smaller? Combining the total R^2 and maximum squared Sharpe ratio relations, one can see that small principal components cannot contribute meaningfully to predictability, or they would yield too high a Sharpe ratio. Intuitively, while it is entirely possible that each of the many proposed factors are predictable, it is unlikely they all capture independent sources of risk. Otherwise, investors would be able to diversify across them and obtain implausibly large Sharpe ratios. This is the dynamic counterpart to the static reasoning of Kozak et al. (2018), who use a similar argument to conclude that small principal components cannot contribute meaningfully to the cross-sectional dispersion in *average* returns. Hence, the large few PC portfolios must capture both cross-sectional and time-series variation in expected factor returns. We define Z_{t+1} as the vector of the largest principal component portfolios of F_{t+1}. The exact number of PCs to include is an empirical question and depends on the strength of the factor structure. The following proposition summarizes the implications of this result for the SDF and the optimal factor timing portfolio.

Proposition 1. *Under Assumption 1 and Assumption 2, the SDF can be approximated by a combination of the dominant factors:*

$$m_{t+1} \approx 1 - \mathbb{E}_t\left[Z_{t+1}\right]' \Sigma_{Z,t}^{-1}\left(Z_{t+1} - \mathbb{E}_t\left[Z_{t+1}\right]\right). \tag{9}$$

Equivalently, the maximum Sharpe ratio factor timing portfolio can be approximated by

$$R_{t+1}^{opt} \approx \mathbb{E}_t\left[Z_{t+1}\right]' \Sigma_{Z,t}^{-1} Z_{t+1}. \tag{10}$$

Our two assumptions are complementary in the following sense. Assumption 1 delivers the conclusion that factor timing is sufficient; one need not time individual stocks. Even without this assumption, Assumption 2 provides a useful way to time a given set of factors. However, the two together allow us to measure properties of the SDF, shedding light on a fundamental economic quantity.

As we will show in the context of our empirical application, no more than a few principal components explain a sizable fraction of the variation in the factor returns. In addition, because factors are often chosen to offer a stable correlation structure, the extraction of dominant components is readily implementable using the standard unconditional method.[9] We are left with estimating the conditional means and variances of these large principal components. In this paper, we concentrate on estimating the mean forecasts $\mathbb{E}_t\left[Z_{t+1}\right]$, that is produce return forecasts for a low-dimensional set of portfolio returns. The estimation of volatility is typically more straightforward, and we come back to it in Section 5.4.[10] Because

[9] Stock and Watson (2002) provide conditions under which unconditional principal components analysis identifies important components even in the presence of time-varying parameters.

[10] Moreira and Muir (2017) and Engle and Kelly (2012) are examples of work that point to methods for

we are only focusing on few components, we use standard forecasting methods for individual returns.

To summarize, our assumptions lead to the following approach to measure conditional expected returns and engage in factor timing:

1. Start from a set of pricing factors F_{t+1}.

2. Reduce this set of factors to a few dominant components, Z_{t+1}, using principal components analysis.

3. Produce separate individual forecasts of each of the Z_{t+1}, that is measures of $\mathbb{E}_t\left[Z_{t+1}\right]$.

4. To measure the conditional expected factor returns, apply these forecasts to factors using their loadings on the dominant components.

5. To engage in factor timing or estimate the SDF, use these forecasts to construct the portfolio given in Equation 10.

In the remainder of this paper, we implement this approach in the context of the cross-section of stock returns. We show how our method allows one to obtain robust measures of expected returns, which is useful for factor timing. And, by studying the corresponding SDF, we explain how it generates novel empirical facts to discipline economic models. In Appendix C, we provide an alternative, more statistical motivation for our methodology.

3 Factor Return Predictability

3.1 Data

Step 1 of our approach is to start with a set of pricing factors. For equities, we focus on a broad set of fifty "anomaly" portfolios which effectively summarize the heterogeneity in

timing volatility and the benefits it provides.

expected returns, following the logic in Kozak et al. (2019). We present here the construction for our main estimates, and confirm the robustness of our conclusions around these choices in Table 4. We construct these portfolios as follows. We use the universe of CRSP and COMPUSTAT stocks and sort them into 10 value-weighted portfolios for each of the 50 characteristics studied in Kozak et al. (2019) and listed in Appendix Table A.4. Portfolios include all NYSE, AMEX, and NASDAQ firms; however, the breakpoints use only NYSE firms as in Fama and French (2016). Our sample consists of monthly returns from January 1974 to December 2017.

We construct the long-short anomalies as differences between each anomaly's return on portfolio 10 minus the return on portfolio 1. For each anomaly strategy we also construct its corresponding measure of relative valuation based on book-to-market ratios of the underlying stocks. We define this measure, bm, as the difference in log book-to-market ratios of portfolio 10 and portfolio 1.[11]

Most of these portfolio sorts exhibit a significant spread in average returns and CAPM alphas. This finding has been documented in the vast literature on the cross-section of returns and can be verified in Appendix Table A.4. In our sample, most anomalies show a large, nearly monotonic pattern in average returns across decile portfolios, consistent with prior research. Rather than studying unconditional mean returns, our primary focus in this paper is on time variation in conditional expected returns, which has received considerably less attention in prior work.

Finally, we also market-adjust and rescale the data. Specifically, for each anomaly we compute its regression β with respect to aggregate market returns. We then market-adjust our returns and predictors by subtracting $\beta \times r_{mkt}$ for returns and $\beta \times bm_{mkt}$ for the bm ratios. Next, we rescale the market-adjusted returns and bm ratios so that they have equal

[11]The book-to-market ratio, bm, of a portfolio is defined as the sum of book equity relative to the total market capitalization of all firms in that portfolio. Equivalently, it is the market-capitalization weighted average of individual stocks' bm ratios.

Table 1: Percentage of variance explained by anomaly PCs

Percentage of variance explained by each PC of the 50 anomaly strategies.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
% var. explained	25.8	12.4	10.3	6.7	4.8	4.0	3.6	2.8	2.2	2.1
Cumulative	25.8	38.3	48.5	55.2	60.0	64.0	67.6	70.4	72.6	74.7

variance across anomalies. Importantly, the βs and variances used for these transformations are estimated using only the first half of the sample so that out-of-sample (OOS) statistics contain no look-ahead bias.

3.2 Dominant Components of the Factors

Step 2 of our approach is to reduce this set of factors to a few dominant components, its largest PCs. We are interested in the joint predictability of anomaly portfolio returns. We construct PCs from the 50 anomaly portfolios and study their predictability. Formally, consider the eigenvalue decomposition of anomaly excess returns, $\text{cov}\left(F_{t+1}\right) = Q\Lambda Q^{'}$, where Q is the matrix of eigenvectors and Λ is the diagonal matrix of eigenvalues. The i-th PC portfolio is formed as $PC_{i,t+1} = q_i^{\prime} F_{t+1}$ where q_i is the i-th column of Q. To ensure our later OOS results do not have any look-ahead bias, we estimate Q and Λ using only the first half of the data.

Table 1 shows that anomaly portfolio returns exhibit a moderately strong factor structure. For example the first PC accounts for one fourth of the total variation. This is sizable but however, much weaker than what is typically found in other asset classes such as Treasury bonds or foreign exchange. How many components should we use? The relations of Equation 7 and Equation 8 provide some guidance for this choice if we use some plausible priors. First, Campbell and Thompson (2007) show that the monthly R^2 when predicting the market is around 75bp when using various price ratios, so we can use this as a reasonable magnitude for the predictability (total R^2) of the anomalies. Second, a relatively loose upper bound on the maximum Sharpe ratio $\mathbb{E}\left(S_t^2\right)$ is 1 at the annual frequency — about twice that of the market — or 8.3% monthly. Under the view that all included PCs contribute equally to the total R^2, the harmonic mean of their contribution to the total variance of returns must be higher than the ratio of these two numbers, 75bp/8.3% = 9%.[12] Using the eigenvalues in Table 1, this value yields that we should include at most five PCs. Based on this simple calculation, we focus on these for our main analysis, but also explore robustness to other choices. These five components jointly explain nearly two-thirds of the total variation in returns. Since our portfolios are market-neutral, we also include the aggregate market portfolio as a potentially important pricing factor. In other words, we study $Z_{t+1} = (R_{mkt,t+1}, PC_{1,t+1} \;\cdots\; PC_{5,t+1})$.

3.3 Predicting the Large PCs of Anomaly Returns

Step 3 of our approach is to produce individual forecasts of the dominant components of factor returns.

Predictors. We obtain these forecasts using standard predictive regressions on valuation ratios. Valuation ratios are the most commonly used forecasters for the market return, going back to Shiller (1981), Fama and French (1988), and Campbell and Shiller (1988). They have also been used at the individual stock level, by Vuolteenaho (2002); Lochstoer and Tetlock (2016). Cohen et al. (2003) show that value-minus-growth strategies are predictable by their own book-to-market ratios. Kelly and Pruitt (2013) use a large cross-section of book-to-market ratios to predict both the aggregate market return as well as portfolios sorted on various characteristics. This broad use comes from the fact that one should expect them to be informative about expected returns. For example, log-linearizing the clean surplus accounting relation of Ohlson (1995), Vuolteenaho (2002) shows that the log book-to-market

[12] Appendix Section A.4 provides a derivation of this formula.

ratio of a long-only strategy is a discounted sum of all future expected returns for this strategy minus future earning growth. By using valuation ratios as well, our conclusions are readily comparable to this seminal work. However, we are not arguing that other predictors, perhaps motivated by specific economic theories, could not find additional sources of predictability.

Following our broad goal of dimension reduction, we construct a single predictor for each portfolio: we use its net book-to-market ratio. For predicting $PC_{i,t+1}$, we construct its own log book-to-market ratio $bm_{i,t}$ by combining the anomaly log book-to-market ratios according to portfolio weights: $bm_{i,t} = q_i' bm_t^F$. We use the difference between this quantity for the long and short leg of our PCs, thereby capturing potentially useful information about future expected returns. Intuitively, $bm_{i,t}$ keeps track of the relative valuation of stocks of a certain *fixed* combination of "types" (e.g. value vs. growth, large vs. small, etc.), where these "types" are varying across stocks. When the valuation spread is large, it is natural to expect that stocks of the corresponding combination of "types" will experience low future returns. These portfolio-level book-to-market ratios are likely to be stable (even though portfolio composition is changing) because expected returns depend on this combination of "types"; empirically we find they are stable.

In addition, note that this choice of predictors dramatically reduces the dimensionality of the set of predictive variables. In Section 3.6 we explore alternative methods for analyzing high-dimensional data.

Predictability results. We analyze the predictability of anomaly PC portfolios and the aggregate market using a monthly holding period, as in Campbell and Thompson (2007). Table 2 shows the results of these 6 predictive regressions. The first two rows report the predictive coefficient estimate and Newey and West (1987) t-statistic.[13] The third and fourth rows show the bias in coefficient estimate and p-value obtained from a parametric bootstrap.

[13] We use a two-year window for the Bartlett kernel.

Precisely, we first estimate a restricted VAR(1) for a PC's return and *bm* ratio under the null of no predictability. We then simulate 10,000 histories from the VAR system with errors drawn with replacement from the empirical distribution. From these simulations, we obtain the distribution of coefficients and t-statistics. We construct p-values by using the simulated distribution of the t-statistics. Both the asymptotic standard errors and these p-values are useful for inference: the Newey and West (1987) standard errors is consistent under mild assumptions on the data-generating process, while the p-value corrects for potential finite-sample biases and non-normality. The fifth row gives the full sample predictive R^2 and the sixth row reports the OOS R^2.[14] To compute this statistic we divide the sample into two equal halves. We estimate predictive coefficients in the first half and apply these to *bm* ratios in the second half to form OOS forecasts. Since all data construction choices use only the first half data, OOS results are not subject to look-ahead biases.

[14]We define the OOS R^2 as $1 - \frac{\mathrm{var}(r-\hat{r})}{\mathrm{var}(r)}$ where $\hat{r}$ is the forecast formed using parameters estimated in-sample.

Table 2: Predicting dominant equity components with BE/ME ratios

We report results from predictive regressions of excess market returns and five PCs of long-short anomaly returns. The market is forecasted using the log of the aggregate book-to-market ratio. The anomaly PCs are forecasted using a restricted linear combination of anomalies' log book-to-market ratios with weights given by the corresponding eigenvector of pooled long-short strategy returns. The first row shows the coefficient estimate. The second row shows asymptotic t-statistics estimated using the method of Newey and West (1987). The third and fourth rows show the bias and p-value from a parametric bootstrap. The fifth and sixth rows shows the in-sample and out-of-sample monthly R^2. The last three rows give critical values of the OOS R^2 based on the placebo test in Kelly and Pruitt (2013).

	MKT	PC1	PC2	PC3	PC4	PC5
Own bm	0.76	4.32	1.62	1.80	4.86	1.56
	(1.24)	(4.31)	(1.81)	(2.01)	(3.74)	(0.78)
bias	0.68	0.36	0.16	0.18	0.10	0.08
p-value	0.35	0.00	0.10	0.07	0.00	0.48
R^2	0.29	3.96	0.74	0.56	3.59	0.50
OOS R^2	1.00	4.82	0.97	0.47	3.52	0.55
			OOS R^2 Critical Values			
$90th$	0.44	0.49	0.29	0.21	0.71	0.59
$95th$	0.68	0.97	0.48	0.37	1.19	0.96
$99th$	1.35	1.73	0.87	0.84	1.95	1.71

Consistent with previous studies, the estimate is not statistically significant for the market. While the OOS R^2 of 1% is encouraging, the bootstrap exercise reveals that there is substantial Stambaugh bias, almost as large as the estimated coefficient. In contrast, PCs 1 and 4 are unambiguously predictable by their own *bm* ratio. The estimated coefficients are large and significant, with t-statistics around 4 and bootstraped p-values close to 0. Both the in-sample and OOS R^2s are large: around 4% for PC1 and around 3.5% for PC4. The estimated relation for PC2 and PC3 exhibit weaker strength, but still appears statistically significant. T-statistics are 1.81 and 2,01, while bootstraped p-values are 11% and 7%. The OOS R^2s take values around 1% and 0.5%. Finally, PC5 does not appear predictable, with an insignificant coefficient estimate. For these PC portfolios, coefficients are slightly biased upward. In contrast to the market estimates, the biases for the PC portfolios are small relative to the estimated coefficients. In Appendix B.2 we show that this substantially lower bias for the PC portfolios obtains for two reasons. First, their *bm* ratios are less persistent than that of the aggregate market. Second, the correlation of innovations to *bm* ratios and returns is lower for the PCs than for the market.

To further demonstrate that the predictability we uncover does not arise due to mechanical biases, we run placebo tests following Kelly and Pruitt (2013). This allows us to assess whether the OOS R^2s we obtain are statistically significant. Specifically, we generate six AR(1) processes that are specified to have the same mean, autocorrelation, and covariance as the *bm* ratios we use as predictors.[15] Since these are simulated values, they are independent of true return data. We then construct OOS forecasts for actual returns using the simulated data and record the OOS R^2 values. We repeat this procedure 1,000 times to obtain simulated distributions for OOS R^2 statistics. From these distributions, we compute 90%, 95%, and 99% critical values, reported in the last three rows of Table 2. Comparing the estimated OOS R^2 with the critical values reinforces the conclusions we drew above. The estimated OOS R^2 for the PCs are unlikely to be due to mechanical finite-sample biases: they are all significant at the 5% level except for PC5.

In Appendix Figure A.1 we report the time series of realized returns along with full sample and out-of-sample forecasts of returns on the aggregate market and the first five PC portfolios of anomalies. In particular for PC1 and PC4, the out-of-sample forecasts are remarkably close to the full sample values, further confirming that our coefficient estimates are precisely measured.

Importance of restrictions. Our emphasis on the largest principal components is predicated on the idea that they should capture most of the variation in expected returns. In

[15]We match these moments using the first half of the data to avoid any potential look-ahead bias.

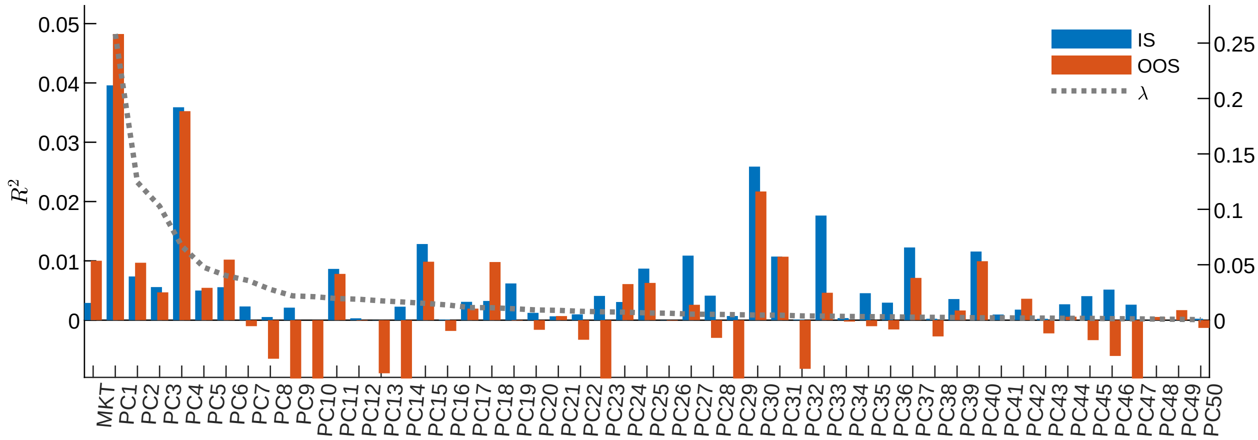


Figure 1: Predictability of equity PCs with own *bm* ratios. The plot shows the in-sample (IS) and out-of-sample monthly R^2 of predicting each PC of anomaly portfolio returns with its own *bm* ratio. The dotted gray line (using the right axis) shows the fraction of total variance explained by each PC.

Figure 1 we report predictability for *all* the principal components of our factors. Only large PCs are strongly predictable in- and out-of-sample by their own *bm* ratios. In this sense, our focus on predicting only the first few PCs is not only motivated by theory, but finds strong support in the data. In addition, by focusing on predicting only these dominant components we uncover robust predictability in the data and largely ignore spurious predictability that stems from small PCs. This latter result echoes the more statistical concerns we put forward as well.

We now turn to forecasting individual anomalies to consider how the restrictions we impose help in this exercise.

3.4 Predicting Individual Factors

Step 4 of our approach is to infer expected return forecasts for the individual factors from the forecasts of our dominant components. Since factors are known linear combinations of the PC portfolios, we can use the estimates in Table 2 to generate forecasts for each anomaly. Notably, each anomaly return is implicitly predicted by the whole cross-section of *bm* ratios.

Table 3 shows the in- and out-of-sample R^2 for each anomaly return using our method. Many of the anomalies are highly predictable; roughly half have OOS R^2 greater than 1% and only two have OOS R^2 below -1%. The total R^2 is 1.03% in-sample and 0.93% OOS.[16] Our approach allows us to uncover these patterns in a robust way.

The substantial anomaly predictability we document in Table 3 also contributes to the recent debate on whether these strategies represent actual investment opportunities or are statistical artifacts which are largely data-mined. For example, Hou et al. (2017) claim that most anomalies are not robust to small methodological or sample variations and conclude there is "widespread *p*-hacking in the anomalies literature." Using a different methodology, Harvey and Liu (2016) argue that the three-factor model of Fama and French (1993) provides a good description of expected returns, and hence, most CAPM anomalies are spurious. Interestingly, we find that some anomalies, such as size and sales growth, which have low unconditional Sharpe ratios are nonetheless highly predictable. This indicates that these strategies at least sometimes represent "important" deviations from the CAPM. This echoes the importance of conditioning information, as emphasized by Nagel and Singleton (2011) and others. More generally these results highlight that a lack of risk premium on average does not necessarily imply a lack of risk premium always: expected returns can be sometimes positive and sometimes negative for the same strategy.

[16] We measure the OOS total R^2 as $1 - \frac{tr[\text{cov}(\varepsilon_{t+1})]}{tr[\text{cov}(F_{i,t+1})]}$ where ε_{t+1} are forecast errors. This quantity can be negative, which typically obtains if forecasts and realizations are negatively correlated out-of-sample.

Table 3: Predicting individual anomaly returns: R^2 (%)

Monthly predictive R^2 of individual anomalies returns using implied fitted values based on PC forecasts. Column 1 (IS) provides estimates in full sample. Column 2 (OOS) shows out-of-sample R^2.

	IS	OOS
1. Size	3.8	4.5
2. Value (A)	1.8	1.9
3. Gross Profitability	-2.2	-4.7
4. Value-Profitablity	3.7	3.8
5. F-score	0.4	-0.2
6. Debt Issuance	0.8	0.5
7. Share Repurchases	0.9	0.7
8. Net Issuance (A)	2.4	3.8
9. Accruals	-0.2	-0.1
10. Asset Growth	1.8	2.6
11. Asset Turnover	0.6	0.8
12. Gross Margins	0.6	-1.0
13. Earnings/Price	0.7	-0.0
14. Cash Flows/Price	0.7	0.4
15. Net Operating Assets	0.6	-0.2
16. Investment/Assets	1.8	1.3
17. Investment/Capital	-0.1	-0.5
18. Investment Growth	1.8	1.9
19. Sales Growth	1.2	2.2
20. Leverage	0.6	0.7
21. Return on Assets (A)	0.9	1.2
22. Return on Book Equity (A)	1.2	-0.1
23. Sales/Price	2.0	1.2
24. Growth in LTNOA	0.5	0.7
25. Momentum (6m)	1.7	1.7
26. Value-Momentum	-0.0	1.2
27. Value-Momentum-Prof.	1.7	2.5
28. Short Interest	-0.5	-0.4
29. Momentum (12m)	1.3	1.4
30. Industry Momentum	0.1	-0.2
31. Momentum-Reversals	2.8	3.3
32. Long Run Reversals	5.7	5.5
33. Value (M)	3.6	3.0
34. Net Issuance (M)	0.8	1.3
35. Earnings Surprises	-0.9	-0.7
36. Return on Book Equity (Q)	0.2	0.0
37. Return on Market Equity	0.7	0.1
38. Return on Assets (Q)	0.3	0.5
39. Short-Term Reversals	0.3	0.5
40. Idiosyncratic Volatility	1.5	0.6
41. Beta Arbitrage	-0.6	-0.6
42. Seasonality	-0.4	0.1
43. Industry Rel. Reversals	1.2	1.0
44. Industry Rel. Rev. (L.V.)	1.9	1.6
45. Ind. Mom-Reversals	0.6	-0.2
46. Composite Issuance	-0.3	0.1
47. Price	4.3	3.5
48. Share Volume	-0.4	-0.6
49. Duration	2.1	2.9
50. Firm age	0.3	0.5
Mean	1.1	1.0
Median	0.8	0.7
Std. Dev	1.4	1.7

3.5 Robustness

Our main estimation includes many choices such as how to construct the raw anomaly returns, the length of the holding period, how many PCs to include, whether to market-adjust, and whether to rescale the data. In Table 4 we explore the robustness of our results to changes in these specifications. For each specification, we report the OOS total R^2 as well as the number of PCs with statistically significant OOS R^2 at the 5% level. The first row repeats the results for our main specification. The next two rows show that our results are robust to how the anomaly portfolios are constructed. Instead of first sorting stocks into deciles for each characteristics, we use quintiles and terciles and obtain similar results. In the next block we consider varying the number of included PCs from one to seven and again obtain similar OOS findings. Adding more components does not meaningfully enhance performance. Reducing the number of PCs below four leads to some reduction in performance. Still, with only one PC we obtain more than half the OOS R^2 of our baseline model. Next we consider not market-adjusting or rescaling returns and *bm* ratios. In fact, the OOS total R^2 improves without these transformations. Finally, we consider different holding periods. With quarterly, half-year, and annual holding periods, the OOS total R^2 increases almost proportionately with horizon. Importantly, this only obtains when estimating the PC eigenvectors using monthly returns. That is, even with a twelve month holding period, we can construct the PCs using the covariance matrix of monthly holding period returns. If instead we construct PCs using only returns at the same frequency, meaningful information about the covariance structure of returns is lost. Since we use non-overlapping returns, increasing the holding period proportionately reduces the sample size. For an annual holding period, the eigenvectors are estimated using only twenty-two observations so the sample covariance matrix is not even of full rank. It is evidently beneficial to use higher frequency data to estimate the covariance matrix and resulting eigenvectors. In Section B.1 we show that our results are robust to using expanding- and rolling-window OOS methods, as well as estimating the regressions in the second half of the sample and measuring OOS performance in the first half.

Table 4: Various data choices

The table reports summary statistics of predictive regressions in Table 2 for various data construction choices. Specifically, we report the OOS total R^2 and the number of PC portfolios for which the OOS R^2 is statistically significant using the placebo test of Kelly and Pruitt (2013). The first column reports the number of portfolios used for the underlying characteristic sorts. The second column reports the holding period in months. For holding periods longer than one month, the third column reports whether principal components are estimated using monthly or holding period returns. The fourth column reports whether the anomaly returns are orthogonalized relative to the aggregate market. The fifth column reports whether the anomaly returns and book-to-market values are normalized to have equal variance.

Anomaly portfolio sort	Holding period (months)	# of PCs	Monthly PCs	Market adjusted returns	Scaled Returns and bm	OOS Total R^2	# PCs signif. OOS R^2
Deciles	1	5	X	X	X	0.93	4
Quintiles	1	5	X	X	X	1.01	4
Terciles	1	5	X	X	X	0.81	3
Deciles	1	1	X	X	X	0.57	1
Deciles	1	2	X	X	X	0.72	2
Deciles	1	3	X	X	X	0.78	3
Deciles	1	4	X	X	X	0.90	4
Deciles	1	6	X	X	X	0.97	5
Deciles	1	7	X	X	X	0.96	5
Deciles	1	5	X	X		1.18	3
Deciles	1	5	X		X	1.27	4
Deciles	1	5	X			1.24	2
Deciles	3	5	X	X	X	2.69	4
Deciles	3	5		X	X	0.99	2
Deciles	6	5	X	X	X	5.42	3
Deciles	6	5		X	X	3.49	3
Deciles	12	5	X	X	X	10.05	3
Deciles	12	5		X	X	4.54	2

3.6 Comparison to Alternative Approaches

Forecasting only large PC portfolios by their own *bm* ratio generates robust OOS total R^2, but there are many other ways one could forecast anomaly returns. Some papers advocate methods which aim to deal with the high-dimensionality of either the forecast targets (returns) or the predictors (*bm* ratios). Others use different predictive variables altogether. In Table 5 we report the OOS total R^2 as well as mean, median, and standard deviation of individual anomaly R^2s for a wide variety of these approaches. All statistics are out-of-sample. The first row shows results from a completely unregularized estimation in which each anomaly return is forecast using all anomaly 50 *bm* ratios. As expected, the OOS total R^2 of -134% is terrible. This highlights the need for dimensionality reduction. The second row reports results from our method.

Instead of predicting each PC return with only its own *bm*, we could expand the information set and allow each PC return to be forecast by any or all of the five PCs' *bm* ratios. Since OLS with even five time-series predictors is likely substantially overfit, we consider various regularization schemes. We first consider ridge regression and Lasso estimation with penalty parameters chosen to allow for exactly one degree of freedom, as in our main estimation. Row 3 shows that ridge regression, even with five predictors, does not deliver robust predictability. Row 4 shows that Lasso does somewhat better, but still substantially worse than using only each portfolio's own *bm*. Belloni et al. (2013) show theoretically that the "OLS post-Lasso estimator performs at least as well as Lasso in terms of the rate of convergence, and has the advantage of a smaller bias." Row 5 confirms this result empirically. Using Lasso to select one predictor, then estimating the coefficient by OLS produces a 0.76% total R^2, nearly as high as our method. Instead of using the *bm* ratios of the PC portfolios, we could have used PCA directly on the anomaly *bm* ratios to reduce the dimensionality of the predictors. Row 6 shows that OLS post-Lasso on the five principal components of anomaly *bm* ratios does reasonably well, but worse than using the *bm* ratios of the PCs themselves. Since price

Table 5: Out-of-sample R^2 of various forecasting methods

The table reports the monthly OOS total R^2 as well as mean, median, and standard deviation of OOS R^2 for individual anomaly portfolios for various forecasting methods. The first column gives the set of assets which are directly forecast, the predictive variables used, and the forecasting method. When omitted, the method is ordinary least squares.

Method	OOS Total R^2	Mean	Median	Std.
1. 50 Anom, BM of Anom, OLS	-133.73	-161.91	-123.12	129.75
2. 5 PCs, Own BM	0.93	1.00	0.69	1.69
3. 5 PCs, BM of PCs, Ridge 1DoF	0.01	0.02	0.02	0.09
4. 5 PCs, BM of PCs, Lasso 1DoF	0.26	0.27	0.19	0.56
5. 5 PCs, BM of PCs, Lasso-OLS 1DoF	0.76	0.83	0.61	1.75
6. 5 PCs, PCs of BM, Lasso-OLS 1DoF	0.52	0.59	0.35	1.17
7. 5 PCs, BM of PCs, 3PRF	0.32	0.36	0.19	0.96
8. 50 Anom, BM of Anom, Lasso-OLS 1DoF	-2.79	-3.27	-1.04	5.10
9. 50 Anom, BM of PCs, Lasso-OLS 1DoF	0.03	-0.06	-0.18	2.33
10. 50 Anom, Own BM	0.50	0.49	0.11	1.42
11. 50 Anom, Own BM, Pooled	0.48	0.51	0.42	1.13
12. 50 Anom, BM of Anom, 3PRF	0.16	0.17	0.12	0.81
13. 50 Anom, Own Characteristic	-2.94	-3.21	0.03	20.67
14. 50 Anom, Sentiment	0.17	0.06	0.01	1.19
15. 5 PCs, Sentiment	0.19	0.06	0.01	1.19
16. 50 Anom, Factor Momentum	-0.49	-0.48	-0.32	1.12
17. 5 PCs, Factor Momentum	-0.08	-0.05	-0.23	1.19

ratios are much more persistent than returns, the sample covariance matrix and resulting eigenvectors are measured with substantially more error, which may contribute to the worse performance. Row 7 uses the five PCs' *bm* ratios, but uses the three-pass regression filter (3PRF) from Kelly and Pruitt (2013) instead of ridge or Lasso. As in that paper, the filter is run separately for each dependent variable (PC return). For this set of portfolios, 3PRF does produce moderate OOS predictability, but less than OLS post-Lasso.

Rows 8 to 12 consider various methods for reducing the dimensionality of the predictors but without any left-hand-side (LHS) dimensionality reduction. These methods predict each of the 50 anomaly returns directly. Row 8 shows that OLS post-Lasso using 50 *bm* fails completely OOS. Instead of using the 50 anomaly *bm* ratios, we could use the five PCs' *bm* ratios as in row 5. This approach does substantially better (row 9), improving to near zero OOS total R^2. However, row 5 shows that with the same information set and same estimation technique, restricting to PCs of returns produces a large OOS total R^2, highlighting the importance of reducing the dimension of the LHS.

In row 10 we predict each anomaly with its own *bm* ratio, as our method does for PC portfolios. Perhaps not surprisingly, this produces a substantial total R^2 of 0.5%. Price ratios seem to be robust predictors of returns. Still, the total R^2 is only slightly greater than half of what we find when directly predicting large PC portfolios. This highlights that while an asset's price ratio may be informative about its expected return, there is valuable information in the whole cross-section of valuation ratios, a point first made by Kelly and Pruitt (2013). There are a number of previous papers using a similar methodology. Asness et al. (2000) and Cohen et al. (2003) find that the value anomaly itself is forecastable by its own *bm* ratio.[17] Arnott et al. (2016b) use various valuation measures (price ratios) to forecast eight anomaly returns; they forecast each anomaly return with its own valuation measures and find statistically significant results. Asness et al. (2017) use each anomaly's *bm* ratio to construct timing strategies for the value, momentum, and low β anomalies. Using their methodology, however, they conclude that the strength of predictability is lacking. Implicitly, these methods allow for as many independent sources of time-varying expected returns as there are assets or factors to predict.[18] This framework is optimal only if one thinks the anomalies and their expected returns are independent, at odds with our Assumption 2.

In row 11, we show results from repeating the exercise of forecasting each anomaly with its own *bm* ratio, but now estimate the predictive coefficients from a panel regression which

[17]Asness et al. (2000) use the ratio of the book-to-market of value stocks to that of growth stocks. Cohen et al. (2003) uses the log difference, as we do.

[18]Formally, the covariance matrix of expected returns is allowed to have full rank.

imposes that the coefficient on bm is the same for all anomalies. Formally, we estimate $\mathbb{E}_t\left(F_{i,t+1}\right) = a_{0,i} + a_1 bm_{i,t}$, allowing each anomaly factor to have a different intercept, thereby allowing different unconditional means, but imposing uniform predictability of returns using bm. The OOS performance is not meaningfully different from unrestricted estimation. In related work, Campbell et al. (2009) and Lochstoer and Tetlock (2016) use a bottom-up approach of aggregating firm-level estimates to portfolios in order to decompose variation in returns into discount rate and cash-flow news. They estimate a panel VAR in which they forecast each stock's return using its own bm ratio, additional stock characteristics, and some aggregate variables.[19] Unlike the previous studies, they impose that the coefficients in the predictive regression are the same for all stocks. While imposing this equality is a form of statistical regularization, it still allows for as many independent sources of time-varying expected returns as there are stocks. Therefore, these restrictions do not put discipline on the Sharpe ratios implied by the predictability estimates.

In row 12 we apply the 3PRF directly to the anomaly returns, using their 50 bms as predictors. This is unlike row 7 where we first reduce by the LHS dimension before estimating with 3PRF. As with PCs, 3PRF generates a moderate OOS total R^2, but substantially less than just using an asset's own bm ratio. These lower R^2s for the 3PRF might seem at odds with the numbers reported in Kelly and Pruitt (2013). One reason for this difference is that we work with market-neutral long-short portfolios whereas that paper focuses on long-only characteristic sorted portfolios which all have market β of around unity. The PLS implementation in Kelly and Pruitt (2013) assumes that each target variable (asset return) loads on only one of the (possibly) few latent factors that drive the cross-section of expected returns, cash flow growth rates, and valuation ratios. Since 3PRF is successful at predicting the aggregate market, it should perform nearly as well in predicting these other

[19] Campbell et al. (2009) do not include any aggregate variables and further cross-sectionally demean stock returns each period. Since most stocks have market β close to unity, their VAR approximately forecasts market-neutral stock returns using each stock's own bm ratio and other characteristics.

portfolios since they all approximately equal the market return plus some smaller other source of variation — their β is close to 1 and they are highly correlated to the market. This conclusion holds even if these other variations are unpredictable, as shown in Appendix C.2. This explanation can rationalize why their cross-sectional portfolios are well-predicted by 3PRF but not our long-short factor portfolios. In fact, their findings for size-sorted portfolios are consistent with this view. The OOS R^2s increase monotonically with firm size, increasing from -18% for the smallest quintile — which is less correlated with the value-weighted market — to +12% for the largest quintile — which is strongly correlated with the market.

Light et al. (2017) develop a related approach to construct forecasts for each stock each month using a large set of characteristics. However, their method does not provide a forecast $\mu_{i,t}$ but rather a scaled forecast $F_t \mu_{i,t}$ where the scaling factor F_t is unobserved and has arbitrary time-varying scale. Hence, their method, while confirming the existence of timing benefits, cannot be directly compared to other forecasting approaches.

In rows 13-17 we consider predictors besides *bm* ratios. In row 13, we present a natural alternative: predicting each anomaly with its own characteristic spread. For example, when sorting firms into deciles based on log market capitalization, the characteristic spread at time t is $\log(ME_{1,t}) - \log(ME_{10,t})$ where $ME_{i,t}$ is the weighted average market capitalization of firms in decile i at time t. As above with *bm* ratios, each anomaly is forecast using its own anomaly specific variable. Somewhat surprisingly, this approach does quite poorly. The performance could possibly be improved through judicious transformation of the sorting variable, though without theoretical guidance on the functional form such an exercise is prone to data mining concerns. In a related paper, Greenwood and Hanson (2012) forecast characteristic-based anomalies using their "issuer-purchaser" spread, or the difference in characteristic (say ME or *bm*) for net equity issuers versus repurchasers. However, as above, a concern for approaches using an anomaly-specific forecasting variable implicitly is that it allows for as many independent sources of time-varying expected returns as there are assets or factors to predict.

There is another literature quite unlike the above alternatives, focusing on one or a few return predictors for all anomalies together. Stambaugh et al. (2012) forecast twelve anomaly strategy returns using the aggregate sentiment index of Baker and Wurgler (2006) and find statistically significant predictability for most of the anomalies they consider. Further, the predictive coefficients are of similar magnitude across anomalies. This is reminiscent of the tradition in bond return forecasting which seeks a single variable that is a significant predictor of excess bond returns of all maturities Cochrane and Piazzesi (2005, 2008); Cieslak and Povala (2015). Using a single predictor for all assets implicitly or explicitly assumes there is only a single source of time-varying expected returns.[20] Stambaugh et al. (2012) do not report R^2 statistics, but a back-of-the-envelope calculation gives values ranging from 0.5%-4% across the anomalies. Rows 14 and 15 show that sentiment does produce moderate OOS predictability, but much less than in-sample. This is true even if we restrict to predicting the five PC portfolios. This suggests that while sentiment is an important variable in estimating variation in expected returns across anomalies, it captures only a small fraction of this total variation. Akbas et al. (2015) start similarly, forecasting eleven anomaly returns individually using aggregate mutual fund and hedge fund flows. Based on the pattern of coefficients, they divide the anomalies into two groups: "real investment factors" and "others". They then form an aggregate portfolio return for each group and forecast these two returns and find substantial R^2 for the "other group" and nearly zero for the investment group.Finally, Ehsani and Linnainmaa (2019) show that for fifteen anomalies, the anomaly's own prior performance significantly positively predicts its return in month t. Rows 16 and 17 show that for our broader set of anomalies, "factor momentum" does not predict returns OOS, with R^2s of -0.49% and -0.08%.[21]

[20]Formally, the covariance matrix of expected returns has at most rank one.

[21]We follow Ehsani and Linnainmaa (2019) and forecast each portfolio with an indicator variable which

4 The Optimal Factor Timing Portfolio

We turn to step 5 of our approach: using our forecasts to form an optimal factor timing portfolio. We find large benefits to factor timing, benefits that appear feasible to collect in practice. This portfolio is also of interest economically, because it informs the properties of the SDF, which we discuss in Section 5.

4.1 Performance

The strong evidence of predictability we document yields substantial investment benefits. By scaling up positions when expected returns are larger, an investor can increase the performance of her portfolio. For example, consider the case of one asset with time-varying expected excess return μ_t and constant variance σ^2. In this situation, the optimal portfolio of a mean-variance investor invests proportionally to μ_t/σ^2 at each point in time. This position generates a certainty equivalent proportional to $\left(\mathbb{E}\left[\mu_t\right]^2 + \text{var}\left[\mu_t\right]\right)/\sigma^2$.[22] The second term, $\text{var}\left[\mu_t\right]/\sigma^2$, is the gain from taking advantage of variation in expected returns. If the same investor invests in a static portfolio with constant weight $\mathbb{E}\left[\mu_t\right]/\sigma^2$, she would only obtain the first term. With multiple assets, the optimal portfolio is given by Equation 10, which yields gains given by Equation 7. Here again, one can see that both the mean and variation in average returns increase the Sharpe ratio. To quantify the gains to factor timing in our setting, we implement this portfolio. Specifically, we use our method to predict component means, $\mathbb{E}_t\left[Z_{t+1}\right]$. We then use these forecasts to construct forecast errors and compute an estimate of the conditional covariance matrix of the market and PC returns, $\Sigma_{Z,t}$, which we for now assume is homoskedastic in order to estimate the role of forecasting means; we revisit this assumption in Section 5.4. We combine these estimates into portfolio

equals one if the portfolio's average monthly return over the past year is positive and zero otherwise.

[22]For an investor with risk aversion γ, the proportionality constant is $1/\gamma$ for the portfolio weight and $1/(2\gamma)$ for the certainty equivalent. Campbell and Thompson (2007) and Moreira and Muir (2017) discuss similar utility calculations.

Table 6: Performance of various portfolio strategies

The table reports the unconditional Sharpe ratio, information ratio, and average mean-variance utility of five strategies: (i) static factor investing strategy, based on unconditional estimates of $\mathbb{E}[Z_t]$; (ii) market timing strategy which uses forecasts of the market return based on Table 2 but sets expected returns on the PC equal to unconditional values; (iii) full factor timing strategy including predictability of the PCs and the market; (iv) anomaly timing strategy which uses forecasts of the PCs based on Table 2 but sets expected returns on the market to unconditional values; and (v) pure anomaly timing strategy sets the weight on the market to zero and invests in anomalies proportional to the deviation of their forecast to its unconditional average, $\mathbb{E}_t[Z_{t+1}] - \mathbb{E}[Z_t]$. All strategies assume a homoskedastic conditional covariance matrix, estimated as the covariance of forecast residuals. Information ratios are calculated relative to the static strategy. Out-of-sample (OOS) values are based on split-sample analysis with all parameters estimated using the first half of the data.

	Factor investing	Market timing	Factor timing	Anomaly timing	Pure anom. timing
IS Sharpe ratio	1.27	1.23	1.19	1.19	0.71
OOS Sharpe ratio	0.76	0.63	0.87	0.96	0.77
IS Inf. ratio	-	-0.17	0.36	0.37	0.35
OOS Inf. ratio	-	-0.64	0.42	0.60	0.59
Expected utility	1.66	1.69	2.96	2.92	1.26

weights $\omega_t = \Sigma_{Z,t}^{-1}\mathbb{E}_t[Z_{t+1}] = \Sigma_Z^{-1}\mathbb{E}_t[Z_{t+1}]$. Importantly, remember that the components Z_{t+1} are fixed linear combinations of the factors F_{t+1}. So, while we express here the factor timing portfolio in terms of the components, it is implicitly trading the underlying factors.

Table 6 reports measures of the performance for versions of this portfolio under various assumptions. We consider five variations of the optimal timing portfolio. "Factor timing" is the portfolio described above. "Factor investing" sets all return forecasts to their unconditional mean, while "market timing" does the same except for the market return. "Anomaly timing" does the opposite: the market is forecast by its unconditional mean, while anomalies receive dynamic forecasts. Finally, the "pure anomaly timing" portfolio sets the weight on the market to zero and invests in anomalies proportional to the deviation of their forecast to its unconditional average. In other words, this portfolio has zero average loading on all factors, and lets us zoom in on the new information of this paper: variation in anomaly expected returns. The following equations summarize these strategies:

$$\text{Factor timing: } \omega_{\text{F.T.},t} = \Sigma_Z^{-1} \left[\mathbb{E}_t\left(R_{mkt,t+1}\right), \mathbb{E}_t\left(PC_{1,t+1}\right), \ldots, \mathbb{E}_t\left(PC_{5,t+1}\right)\right]', \tag{11}$$

$$\text{Factor investing: } \omega_{\text{F.I.}} = \Sigma_Z^{-1} \left[\mathbb{E}\left(R_{mkt,t+1}\right), \mathbb{E}\left(PC_{1,t+1}\right), \ldots, \mathbb{E}\left(PC_{5,t+1}\right)\right]', \tag{12}$$

$$\text{Market timing: } \omega_{\text{M.T.},t} = \Sigma_Z^{-1} \left[\mathbb{E}_t\left(R_{mkt,t+1}\right), \mathbb{E}\left(PC_{1,t+1}\right), \ldots, \mathbb{E}\left(PC_{5,t+1}\right)\right]', \tag{13}$$

$$\text{Anomaly timing: } \omega_{\text{A.T.},t} = \Sigma_Z^{-1} \left[\mathbb{E}\left(R_{mkt,t+1}\right), \mathbb{E}_t\left(PC_{1,t+1}\right), \ldots, \mathbb{E}_t\left(PC_{5,t+1}\right)\right]', \tag{14}$$

$$\text{Pure anomaly timing: } \omega_{\text{P.A.T.},t} = \Sigma_Z^{-1} \left[0, \left[\mathbb{E}_t - \mathbb{E}\right]\left(PC_{1,t+1}\right), \ldots, \left[\mathbb{E}_t - \mathbb{E}\right]\left(PC_{5,t+1}\right)\right]'. \tag{15}$$

The first performance metric we consider is the *unconditional* Sharpe ratio: the ratio of the sample mean and the standard deviation of returns. The factor investing, market timing, anomaly timing, and factor timing portfolio all produce meaningful performance, with Sharpe ratios around 1.2 in sample and between 0.63 and 0.87 out of sample.One might be tempted to conclude from these numbers that factor timing does not improve performance relative to static factor investing. However, it is important to remember that the factor timing portfolio is not designed to maximize the unconditional Sharpe ratio. In a world with predictability, this is not an accurate measure of performance improvement. Ferson and Siegel (2001) show that maximizing the unconditonal Sharpe ratio requires portfolio weights which are highly nonlinear and *nonmonotone* in conditional expected returns. Still, the sizable Sharpe ratio of the pure anomaly timing portfolio is a first piece of evidence that factor timing is valuable. This portfolio does not engage in any static bets, but obtains Sharpe ratios of 0.71 and 0.77 in and out of sample.

A second way to evaluate the value of factor timing is to assess whether the timing portfolios expand the unconditional investment opportunity set captured by the static factor investing portfolio. To do so, we compute the information ratio for the various timing strategy. This number corresponds to the Sharpe ratio that these strategies produce once orthogonalized to the static factor investing portfolio. An advantage of this statistic is that it can be measured without relying on assumptions about return dynamics. However, it only constitutes a *lower bound* on the benefits to factor timing, because it takes the perspective of an uninformed investor. Judged by the information ratio, factor timing, anomaly timing, and pure anomaly timing substantially extend the investment opportunity set. Their information ratios are almost equal at 0.36 in sample, and actually increase to 0.42, 0.60, and 0.59 out of sample. These stable information ratios, in contrast to the more strongly decaying Sharpe ratio of the factor investing portfolio suggests that our conclusions about predictability are actually more robust than the measurement of unconditional return premia.

A third take on the value of factor timing is to ask, given our completely estimated model, what utility would a mean-variance investor expect to obtain on average.[23] Absent statistical issues this is the most accurate characterization of the value of factor timing, because it takes into account the information behind these strategies in the portfolio evaluation. However, this approach does lean more heavily on our estimated model: because expected utility is an ex-ante concept, one needs to take a stand on the distribution to evaluate it. For this reason we can only report full-sample estimates of this quantity. For ease of comparison we report expected utility scaled as in the formulas from the beginning of this section. The gains to factor timing are large: they yield an almost two-fold increase in expected utility from 1.66 to 2.96. Most of that increase, 1.26, comes from pure anomaly timing alone. In contrast, adding market timing to factor investing only slightly increases expected utility by 0.03; removing it from factor timing decreases expected utility by 0.04.

As we have discussed in Section 2.2, another interpretation of this utility calculation is in terms of volatility of the implied SDF. Therefore these numbers suggest large differences in SDF behavior relative to estimates which ignore the evidence of factor predictability. We examine these changes and their economic implications in the next section. Before doing so, we briefly discuss the role of the portfolio rebalancing frequency..

[23] The numbers we report correspond to $\gamma = 1/2$.

4.2 How fast does timing need to be?

Our factor timing portfolios change weights on the anomalies each month. What happens if we trade more slowly? In Table A.5, we report the performance measures for versions of the "pure anomaly timing" strategy which rebalance at lower frequencies. Performance does not deteriorate meaningfully. The unconditional Sharpe ratio actually increases from 0.71 to 0.79 with annual rebalancing. Expected utility declines from 1.26 to 0.81, still a substantial value. We discuss the other results and their construction at more length in Appendix B.6. These results have two implications. First, the potentially less economically meaningful high-frequency variations in our predictors are not the main force behind our predictability results. Second, factor timing strategies might be implementable by actual investors; direct measures of transaction costs would be necessary to make this a firm conclusion.

5 Properties of the SDF

We now use the equivalence between optimal portfolio and SDF to study how including the possibility for factor timing affects the properties of an estimated SDF. This is of particular interest for the construction of economic models: because the SDF encodes asset prices, it can often directly be related to fundamentals of the economy in specific models. For example, the SDF is equal to the marginal utility of consumption of the representative agent in unconstrained economies.

5.1 Volatility

A first property of the SDF which has been the subject of attention by economists is its variance. Hansen and Jagannathan (1991) first show that the variance of the SDF is the largest squared Sharpe ratio attainable in the economy under complete markets; an upper bound in incomplete markets. Therefore the variance of an SDF backed out from asset prices tells us how volatile the SDF has to be in an economic model to possibly account for this evidence. Hansen and Jagannathan (1991) document that the market return alone implies that the SDF is much more volatile than implied by reasonable calibrations of the CRRA model. Subsequent research has proposed models resolving this puzzle.

Naturally, using more conditioning information increases the investment opportunity set, and therefore the variance of the SDF. The results of Table 7 tell us by how much. We report the average conditional variance of the SDF constructed under various sets of assumptions. Remember that the SDF is given by: $m_{t+1} = 1 - \omega_t' (Z_{t+1} - \mathbb{E}_t [Z_{t+1}])$, where ω_t are the weights in the optimal portfolio. The conditional variance of the SDF is therefore:

$$\mathrm{var}_t (m_{t+1}) = \omega_t' \Sigma_Z \omega_t, \tag{16}$$

which we then average over time. “Factor timing” is our full estimate of the SDF, which takes into account variation in the means of the factors, with ω_t given by Equation 11. “Factor investing” imposes the assumption of no factor timing: conditional means are replaced by their unconditional counterpart, Equation 12. Finally, “market timing” only allows for variation in the mean of the market return, Equation 13. These three names coincide with the portfolios of Section 4 because portfolio weights and SDF exposures are equal. However there is a fundamental difference in interpretation. Within the same economy, one can evaluate separately the performance of various strategies. However, there is only one SDF which determines prices, so these various versions are just different estimators of this quantity. They cannot be all correct, and our “factor timing” specification is designed to be the best estimator, while the others force some misspecification in the construction. We find that accounting for the possibility to time factors substantially increases the estimated variance of the SDF. Adding this possibility in addition to timing the market and engaging in static factor investing yields estimates almost twice as large. The magnitude of the SDF variance,

Table 7: Variance of the SDF

We report the average conditional variance of the SDF and its standard deviation constructed under various sets of assumptions. "Factor timing" is our full estimate, which takes into account variation in the means of the PCs and the market. "Factor investing" imposes the assumption of no factor timing: conditional means are replaced by their unconditional counterpart. "Market timing" only allows for variation in the mean of the market return.

	Factor Investing	Market Timing	Factor Timing
$E\left[\mathrm{var}_t\left(m_{t+1}\right)\right]$	1.67	1.71	2.96
$\mathrm{std}\left[\mathrm{var}_t\left(m_{t+1}\right)\right]$	-	0.29	2.17

2.96, is sizable compared to estimates ignoring factor timing, but also substantially larger than in standard models. Bansal and Yaron (2004) report an annualized variance of the SDF of 0.85, while Campbell and Cochrane (1999) obtain a variance that fluctuates roughly between 0 and 1.2.

5.2 Heteroskedasticity

Next, we ask how much does the variance of the SDF change over time. Changes in variance of the SDF can capture how investors' attitude towards risk changes over time: when marginal utility is more volatile, investors ask for larger compensations for bearing risk. The observation that the SDF is heteroskedastic has long been known. Closely related to finding of predictability of market returns in Shiller (1981), many studies document that the Sharpe ratio of the market portfolio fluctuates. Motivated by this observation, a number of models have been developed to capture these changes. In the habit model of Campbell and Cochrane (1999), low realizations of consumption growth make households more risk averse, and renders the SDF more volatile. In the long-run risk model of Bansal and Yaron (2004) periods of high uncertainty about long-run consumption growth yield larger fluctuations in marginal utility.

Because the loadings of our estimated SDF on the components change over time, the SDF

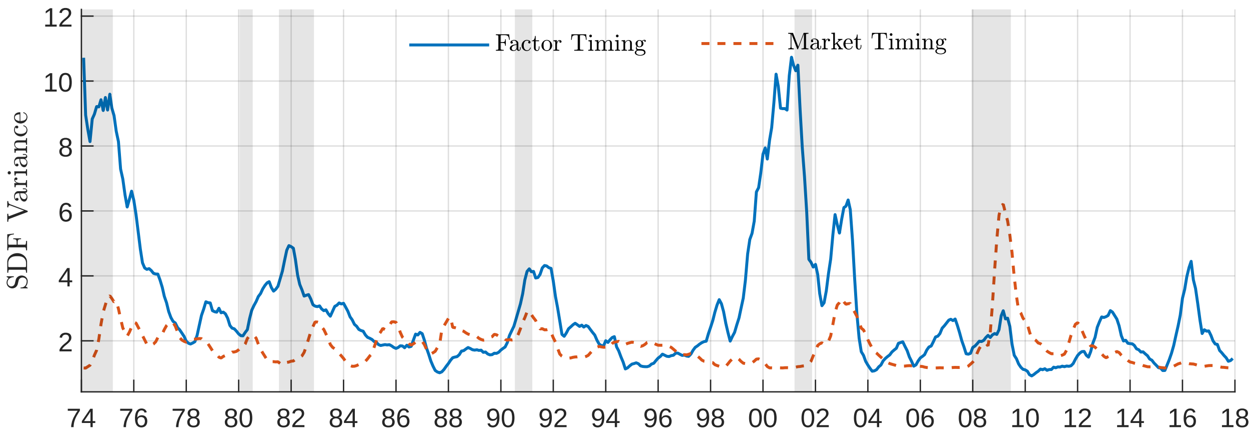


Figure 2: Conditional variance of SDFs. This figure plots the conditional variance of the SDF, constructed in two ways. The solid blue line uses the "factor timing" construction, which allows for variation in the means of the PCs and the market. The red dashed line is a "market timing" estimate that ignores predictability of the anomaly factors. The aggregate market is forecast using aggregate dividend/price, 10-year smoothed earnings/price, realized volatility, term premium, corporate spread, *cay* (Lettau and Ludvigson (2001)), GDP growth, and sentiment (Baker and Wurgler (2006)).

is heteroskedastic even though we assume the component returns themselves are not. Said otherwise, the maximum Sharpe ratio changes over time because prices of risk are changing. With our model estimates, we can compute not only the average variance of the SDF, but also its conditional variance at each point in time using Equation 16. We report this result on the solid blue line in Figure 2. The variance of our estimated SDF varies substantially over time: it fluctuates between low levels close to 0.8 and values as high as 12. The evidence of factor timing is the main driving force behind this result. As a comparison, we report estimates for an SDF estimated under the assumption of constant factor expected returns, but time-variation in market risk premium. We depart from our baseline "market timing" specification and allow many variables other than the book-to-market ratio to forecast market returns.[24] Doing so yields an upper bound on what market return predictions using any combination of these variables imply in terms of SDF heteroskedasticity.[25] Despite this

[24]We include aggregate dividend/price, 10-year smoothed earnings/price, realized volatility, term premium, corporate spread, *cay* (Lettau and Ludvigson (2001)), GDP growth, and sentiment (Baker and Wurgler (2006)).

[25]Appendix Figure A.4 reports the variance of our baseline "market timing" SDF estimate, which is

flexibility, the corresponding SDF variance is much less volatile than our estimate, with a volatility of only 0.29 relative to 2.17. In Appendix Figure A.4 we also report the volatility of our estimated SDF if one imposes constant expected returns on the market and we find no meaningful difference relative to our estimate including factor timing. This observation further corroborates the importance of factor timing as a driver of the time-variation in SDF variance.

The large variations over time we find in SDF variance are at odds with standard theories behind these changes. The variance of the SDF in Campbell and Cochrane (1999) has a standard deviation lower than 0.5. Models focusing on intermediary leverage such as He and Krishnamurthy (2013) generate large spikes in this variance when financial institutions are constrained because they become unwilling to absorb extra risks. However, these theories have a limited appeal in explaining the evidence from factor timing: the large spikes we observe in our estimates do not coincide with periods of intense stress in the financial sector.

5.3 Relation with economic conditions

Economic theories typically focus on specific drivers of variations in the maximum Sharpe ratio. To understand which properties these drivers must exhibit in order to rationalize our findings, we study how the variance of the SDF we estimate relates to measures of economic conditions. We standardized these variables to make estimates comparable.

A first observation is that the variance of the SDF exhibits a moderate degree of persistence. Its yearly autocorrelation is of 0.51. As we noticed when studying performance measures, this implies that following our signals even on a yearly basis provides similar results. Such a result is encouraging for macroeconomic models to explain this variation. However, 0.51 is also a much lower value than what is implied by theories focusing on slow, long-term changes in economic conditions. In line with these short-run patterns, the variance

indeed much less heteroskedastic than this specification.

Table 8: SDF variance and macroeconomics variables

We report univariate regression coefficients and absolute t-statistics from regressions of estimated SDF variance on various macroeconomic variables. Factor timing uses the SDF variance shown in Figure 2. Market timing uses SDF variance assuming the anomaly returns are not predictable and sets the conditional expected market return to the fitted value from a multivariate regression of market returns on the aggregate dividend/price ratio, realized market volatility, term spread, corporate bond yield spread, *cay* (Lettau and Ludvigson, 2001), GDP growth and aggregate sentiment (Baker and Wurgler, 2006). Common idiosyncratic volatility is orthogonalized to market volatility (Herskovic et al., 2016).

	Factor timing		Market timing	
D/P	−0.01	(0.03)	0.22	(2.86)
GDP growth	−0.37	(1.60)	−0.34	(4.93)
Market volatility	0.44	(2.48)	0.30	(5.37)
Sentiment	−0.15	(0.60)	−0.25	(3.35)
Common idio. volatility	0.49	(2.19)	−0.07	(0.97)
cay	−0.42	(1.79)	0.23	(3.00)
Term premium	−0.53	(2.36)	0.24	(3.35)
Inflation	0.75	(3.26)	0.01	(0.08)

of the SDF is, *on average,* related to the state of the business cycle. It averages 4.9 during recessions while only being 2.7 during expansions. However, as one can notice by examining Figure 2, this relation is not systematic. In particular, the depth of the recession does not appear strongly related to the size of the spike in expected returns. This is in part due to the fact that the market component in our "factor timing" SDF does not pick up large variations, unlike the one we use to construct the "market timing" SDF of Figure 2. But this also comes from the weak relation between the intensity of increases in anomaly expected returns with the overall intensity of broad macroeconomic events. For example, these two effects play a role in the low SDF volatility during the 2008 financial crisis relative to the tech boom and bust.

Table 8 relates the variance of the SDF to a variety of measures of economic conditions. For each measure, we report the coefficient and t-statistics of a univariate regression of the SDF variance on the variable. The first column uses our estimated SDF which accounts for the factor timing evidence. The second column uses the estimate of SDF which assumes constant expected returns on the factors, and forecasts market return using a variety of predictors, as in Figure 2. For this second column, because the market forecast is constructed using these specific conditions, it is somewhat mechanical to obtain significant relations. However, the magnitude of the various coefficients still provides a useful benchmark. In contrast, for the first column, we did not use any of these variables are predictors.

We find that the dividend-price ratio of the market, a slow-moving measure of market conditions, does not predict the SDF variance. This is consistent with our discussion on the relatively low persistence of our estimated SDF variance: while D/P contains useful information for the market premium, it does not reveal much about overall fluctuations in Sharpe ratios, which move faster. More transitory measures of business cycle conditions do better: the variance of the SDF is larger following years of lower GDP growth and in periods of high market volatility or low economic sentiment, with loadings similar to those of the pure market timing SDF. Importantly, this similarity in loadings is not mechanically driven by the fact that our estimated SDF times the market: we obtain similar coefficients, -0.33 and 0.45, when setting the expected market return to a constant. Relatedly, the quantity of idiosyncratic risk also becomes important for the variance of the SDF once one accounts for the evidence of factor timing. We find that the common idiosyncratic volatility measure of Herskovic et al. (2016), orthogonalized to market volatility, is strongly positively related to the variance of the factor timing SDF, but not of the the market timing SDF.[26]
Interestingly, some measures of economic conditions that predict the market positively, *cay* and the term premium, forecast the SDF variance negatively. Along these dimensions, risk compensations tend to be low for anomalies when they are large for the market. Finally, last year's inflation rate appears to strongly positively correlate with the SDF variance, while it holds no relation to the overall market return. Upon inspection of the time series (see

[26]We thank Bernard Herskovic who graciously shared his data with us.

Appendix Figure A.5), this relation appears to be driven by the bouts of high inflation in the early part of the sample, and the high inflation during the Internet boom.

Naturally, one should be wary of drawing any causal conclusion from these relations. That being said, economic theories associating variations in the variance of the SDF, or the overall appetite for risk, to these macroeconomic quantities, should confront this evidence to assess if they are consistent with the cross-section of expected returns.

5.4 The role of volatility timing

So far, we have focused our efforts on using information from the predictability of factor returns to discipline the properties of the SDF. Movements in the volatility of factor returns could also play a potentially important role, as suggested for example by Moreira and Muir (2017).[27] In Appendix Section B.8, we measure $\Sigma_{Z,t}$ and revisit the properties of the mean and standard deviation of the SDF variance. Including a time-varying $\Sigma_{Z.t}$ pushes the average SDF variance up from 2.97 from 3.54, about a third of the increase from 1.66 provided by variation in risk premia. In short, both changes in expected returns and variances contribute to the variance of the SDF, but the largest part comes from expected returns. The standard deviation of the squared Sharpe ratio goes from 2.17 to 2.06 when adding volatility timing to factor timing. This small change indicates that including information from changing volatility of returns only has a small effect on the cyclical behavior of the variance of the SDF. Overall, these results imply that the conclusions we have reached so far on our estimates using means only are robust to including volatility timing.

[27] If one focuses on the market return, Moreira and Muir (2017) document that variations in volatility dominate variations in expected returns in creating changes in the squared Sharpe ratio. They also show that volatility timing is useful for factor strategies.

5.5 What are the priced risks?

In addition to specifying how the size of risk compensations evolve over time, economic models typically also specify which sources of risk receive this compensation. For example, the basic tenant of the consumption-CAPM is that the SDF is proportional to aggregate consumption growth. Richer theories include other economic shocks which are priced risks such as long-run shocks to consumption growth, changes in disaster probability, or changes in the health of the financial sector. Going from our estimates to answering which sources of risks are priced is more challenging. When the SDF combines multiple, potentially correlated, sources of risk with time-varying loadings, it is generally not possible to characterize it without a complete structural specification. However, without focusing on a specific model, we can produce a number of statistics that can guide the design of future models.

A first question we answer is whether using a model which targets an SDF which only combines static factor strategies is at least focusing on the right sources of risks. So far, we have showed that such an estimates grossly underestimates the variance and heteroskedasticity of the SDF. However, it might be feasible to easily "patch" univariate dynamics onto the estimate that fits the unconditional properties of returns. Concretely, start from the misspecified static specification of the SDF: $m_{t+1} = 1 - b'\varepsilon_{t+1}$. Is it enough to enrich the model with a single state variable x_t to obtain a SDF $m_{t+1} = 1 - x_t b'\varepsilon_{t+1}$ in order to capture the evidence of factor timing benefits? An economic motivation behind such x_t could for example be a shifter of the risk appetite of the economy, keeping constant the nature of economic risks investors worry about. For example the habit model behaves conditionally a lot like the regular consumption-CAPM, but the level of habit affects risk aversion. In the intermediary asset pricing literature such as He and Krishnamurthy (2013) changes in intermediary wealth (or leverage) drive expected returns in all assets in which they are the marginal investor. Less formally, Stambaugh et al. (2012) articulate a view where investor sentiment coordinates expected returns across anomalies.

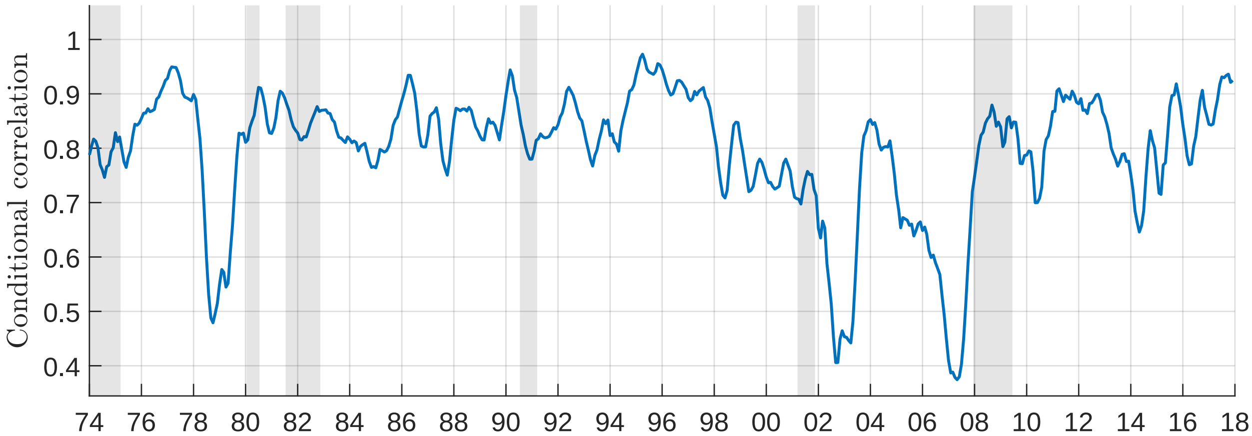


Figure 3: Conditional correlation of the estimated SDF and the misspecified SDF under the assumption of no factor timing benefits. This figure plots the conditional correlation our estimated SDF and a misspecified version which sets conditional factor means to their sample averages. Reported values are six-month averages.

This scaled SDF would be perfectly conditionally correlated with the baseline SDF, but would just add a time-varying loading on it. To assess whether this approach can work, we compute the conditional correlation of the misspecified SDF which is estimated under the assumption of constant expected returns ("factor investing") with our complete estimate of the SDF ("factor timing"). This correlation is given by

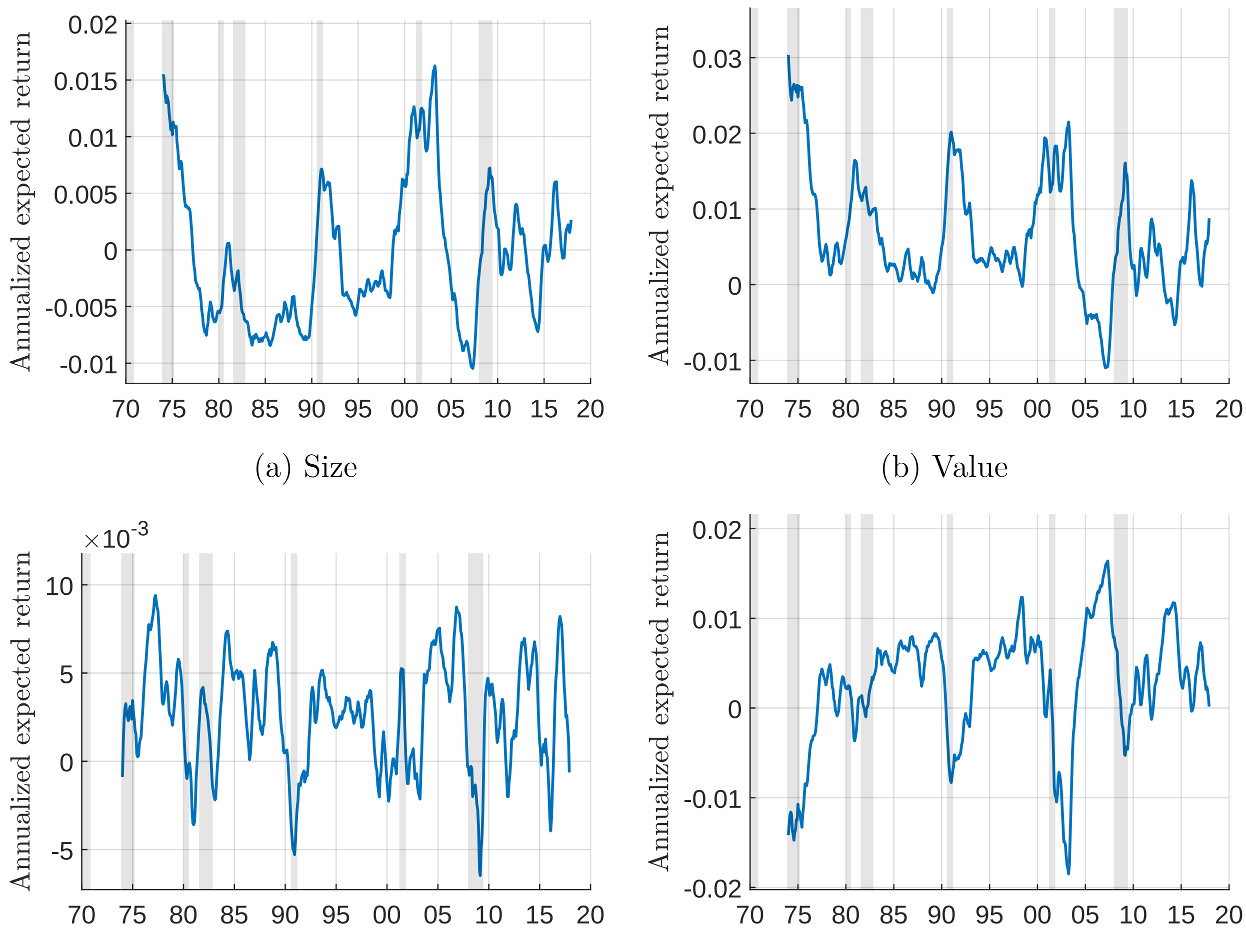
$$\mathrm{corr}_t\left(m_{\mathrm{F.T.},t+1}, m_{\mathrm{F.I.},t+1}\right) = \frac{\omega'_{\mathrm{F.T.},t}\Sigma_Z\omega_{\mathrm{F.I.}}}{\sqrt{\left(\omega'_{\mathrm{F.T.},t}\Sigma_Z\omega_{\mathrm{F.T.},t}\right)\left(\omega'_{\mathrm{F.I.}}\Sigma_Z\omega_{\mathrm{F.I.}}\right)}}, \tag{17}$$

where the weights are defined in equations 11 and 12. Figure 3 reports this conditional correlation. On average the two SDF estimates are quite correlated, with values that fluctuate around 0.8. However, the correlation exhibits strong time series variation, with dips to values as low as 0.4. Both the observations of a mean meaningfully below 1 and these changes indicate that our SDF is not just a rescaling of the naive estimate ignoring factor timing evidence.

This first result suggests that one needs to include multiple sources of time-varying loading on shocks to fit the factor timing evidence. What should these loadings look like? We can shed light on this question by considering how the covariance of our estimated SDF covaries with some specific anomalies. While there is no a priori reason that anomaly portfolio returns coincide with economic shocks, many theories offer a clear mapping between structural shocks and characteristic-sorted portfolio returns. For example, Hong and Stein (1999) develop a model of featuring both underreaction and overreaction to news, generating value and momentum type effects. In Papanikolaou (2011), the relative returns of value and growth stocks reveal investment-specific technological shocks. Alti and Titman (2019) study how investor overconfidence and aggregate disruption shocks lead to time-varying expected returns on value, profitability, and asset growth anomaly strategies. Berk et al. (1999) study how the dynamics of firm investment lead to the value and momentum anomalies. With our estimates, it is straightforward to compute the conditional covariance of a specific factor return with the SDF: it is their conditional expected returns. Figure 4 reports the conditional expected return on four standardized factor returns: size, value, momentum and ROA.[28] Several interesting patterns emerge. All four strategies exhibit substantial variation in their correlation to the SDF, with frequent sign changes. Second, the pattern of these correlation differs across assets: size and value often have large correlations with the SDF during recessions, while momentum tends to relate negatively with the SDF during these episodes.[29] These patterns are not necessarily only driven by recessions: a lot of the cyclical movement in the correlation of the ROA factor with the SDF occurs outside of business cycle. Of course, these are just a few of the many potential portfolio sorts one can focus on. Appendix Figure A.3 reports the average return and the standard deviation of conditional expected returns for each of the 50 portfolios. The figure shows a lot of cross-sectional variation in degree of predictability across anomalies. Interestingly, predictability and average

[28]Because the factor strategies are correlated, the covariance of a specific factor with the SDF differs from the loading (or multivariate regression coefficient) of the SDF on this factor. We report these loadings in Appendix Figure A.2.

[29]Interestingly, this last observation is consistent with the findings of Daniel and Moskowitz (2016) who document forecastable momentum crashes, even though we use completely different variables as predictors.



(a) Size

(b) Value

(c) Momentum

(d) ROA

Figure 4: Anomaly expected returns. The plot shows the time-series of conditional expected returns on four anomaly strategies: size, value, momentum, and return on assets.

performance are not related: the cross-sectional correlation between the mean and standard deviation numbers is close to zero. This result further highlights that there is not a single variable scaling all risk premia up and down.

The broader point is that one can, with our estimates in hand, ask if their theory generates a time series pattern of loadings of the SDF on a specific shock or factor which is consistent with the evidence from the entire cross-section.

6 Concluding Remarks

In this paper we study factor timing, which combines the ideas of long-short factor investing and market timing. Measuring the benefits to such strategies is relevant given their recent popularity, but more importantly, because they affect inference about the stochastic discount factor. However, taking a holistic approach to this question is challenging: there are many potential factors one could time, which increases the scope for finding spurious results. We use the idea of no near-arbitrage to overcome this issue. We show this principle disciplines the estimation of the dynamics of expected returns. Thus guided, we obtain robust estimates of factor predictability. We find that factor timing is very valuable, generating superior performance relative to market timing and factor investing alone.

This conclusion has important consequences for the SDF. When we include our findings in the estimation of the SDF, we uncover a behavior which differs strongly from misspecified SDF estimates ignoring factor predictability. The variance of the implied SDF is much larger and more variable over time. These observations pose strong challenges for existing economic models, which understate these quantities, if they aim to match the cross-section of returns. But the difficulty is not only about explaining a more volatile and heteroskedastic SDF. We find that the variance of the SDF exhibits different cyclical pattern than standard estimates, suggesting that some of the previously established drivers of variation in risk premia are less important when looking at the cross-section. In addition, the dynamics of risk premia are heterogenous across different factors.

In short, our results suggest that it is at least as important for economic theories to understand the dynamic properties of the cross-section as its average properties. We anticipate these facts will be helpful in guiding future research.

References

Akbas, F., W. J. Armstrong, S. Sorescu, and A. Subrahmanyam (2015). Smart money, dumb money, and capital market anomalies. *Journal of Financial Economics 118*(2), 355–382.

Alti, A. and S. Titman (2019). A dynamic model of characteristic-based return predictability. Technical report, National Bureau of Economic Research.

Arnott, R. D., N. Beck, and V. Kalesnik (2016a). Timing “smart beta" strategies? of course! buy low, sell high!

Arnott, R. D., N. Beck, and V. Kalesnik (2016b). To win with “smart beta" ask if the price is right.

Asness, C., S. Chandra, A. Ilmanen, and R. Israel (2017). Contrarian factor timing is deceptively difficult. *The Journal of Portfolio Management 43*(5), 72–87.

Asness, C. S. (2016). The siren song of factor timing aka “smart beta timing" aka “style timing". *The Journal of Portfolio Management 42*(5), 1–6.

Asness, C. S., J. A. Friedman, R. J. Krail, and J. M. Liew (2000). Style timing: Value versus growth. *Journal of Portfolio Management 26*(3), 50–60.

Baba Yara, F., M. Boons, and A. Tamoni (2018). Value return predictability across asset classes and commonalities in risk premia.

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

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.

Barberis, N., R. Greenwood, L. Jin, and A. Shleifer (2015). X-CAPM: An extrapolative capital asset pricing model. *Journal of Financial Economics 115*(1), 1–24.

Belloni, A., V. Chernozhukov, et al. (2013). Least squares after model selection in high-dimensional sparse models. *Bernoulli 19*(2), 521–547.

Berk, J. B., R. C. Green, and V. Naik (1999). Optimal investment, growth options, and security returns. *The Journal of Finance 54*(5), 1553–1607.

Brandt, M. W., P. Santa-Clara, and R. Valkanov (2009). Parametric portfolio policies: Exploiting characteristics in the cross-section of equity returns. *Review of Financial Studies 22*(9), 3411–3447.

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

Campbell, J. Y. and J. H. Cochrane (1999). By Force of Habit: A Consumption Based Explanation of Aggregate Stock Market Behavior. *Journal of Political Economy 107*(2), pp. 205–251.

Campbell, J. Y. and A. S. Kyle (1993). Smart money, Noise Trading and Stock Price Behaviour. *The Review of Economic Studies 60*(1), 1–34.

Campbell, J. Y., C. Polk, and T. Vuolteenaho (2009). Growth or glamour? fundamentals and systematic risk in stock returns. *The Review of Financial Studies 23*(1), 305–344.

Campbell, J. Y. and S. B. Thompson (2007). Predicting excess stock returns out of sample: Can anything beat the historical average? *The Review of Financial Studies 21*(4), 1509–1531.

Chernov, M., L. A. Lochstoer, and S. R. Lundeby (2018). Conditional dynamics and the multi-horizon risk-return trade-off. Technical report, National Bureau of Economic Research.

Cieslak, A. and P. Povala (2015). Expected returns in Treasury bonds. *Review of Financial Studies*, hhv032.

Cochrane, J. H. (2011). Presidential Address: Discount Rates. *The Journal of Finance 66*(4), 1047–1108.

Cochrane, J. H. and M. Piazzesi (2005). Bond Risk Premia. *American Economic Review*, 138–160.

Cochrane, J. H. and M. Piazzesi (2008). Decomposing the yield curve. *Graduate School of Business, University of Chicago, Working Paper*.

Cochrane, J. H. and J. Saa-Requejo (2000). Beyond arbitrage: Good-deal asset price bounds in incomplete markets. *Journal of Political Economy 108*(1), 79–119.

Cohen, R. B., C. Polk, and T. Vuolteenaho (2003). The value spread. *The Journal of Finance 58*(2), 609–641.

Cooper, M. J., R. C. Gutierrez Jr, and A. Hameed (2004). Market states and momentum. *Journal of Finance 59*(3), 1345–1365.

Daniel, K. and T. J. Moskowitz (2016). Momentum crashes. *Journal of Financial Economics 122*(2), 221–247.

Ehsani, S. and J. T. Linnainmaa (2019). Factor momentum and the momentum factor. Technical report, National Bureau of Economic Research.

Engle, R. and B. Kelly (2012). Dynamic equicorrelation. *Journal of Business & Economic Statistics 30*(2), 212–228.

Fama, E. and R. Bliss (1987). The information in long-maturity forward rates. *The American Economic Review*, 680–692.

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

Fama, E. F. and K. R. French (1992). The Cross-Section of Expected Stock Returns. *The Journal of Finance 47*(2), pp. 427–465.

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

Fama, E. F. and K. R. French (2016). Dissecting anomalies with a five-factor model. *The Review of Financial Studies 29*(1), 69–103.

Ferson, W. E. and A. F. Siegel (2001). The efficient use of conditioning information in portfolios. *Journal of Finance 56*(3), 967–982.

Freyberger, J., A. Neuhierl, and M. Weber (2018). Dissecting characteristics nonparametrically.

Gallant, A. R., L. P. Hansen, and G. Tauchen (1990). Using conditional moments of asset payoffs to infer the volatility of intertemporal marginal rates of substitution. *Journal of Econometrics 45*(1-2), 141–179.

Giglio, S. and D. Xiu (2018). Asset pricing with omitted factors. *Chicago Booth Research Paper* (16-21).

Greenwood, R. and S. G. Hanson (2012). Share issuance and factor timing. *The Journal of Finance 67*(2), 761–798.

Griffin, J. M. and J. Xu (2009). How smart are the smart guys? a unique view from hedge fund stock holdings. *The Review of Financial Studies 22*(7), 2531–2570.

Hansen, L. P. and R. Jagannathan (1991). Implications of Security Market Data for Models of Dynamic Economies. *Journal of political economy 99*(2), 225–262.

Harvey, C. R. and Y. Liu (2016). Lucky factors. Technical report.

He, Z. and A. Krishnamurthy (2013). Intermediary asset pricing. *American Economic Review 103*(2), 732–70.

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

Hong, H. and J. Stein (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets. *Journal of Finance 54*(6), 2143–2184.

Hou, K., C. Xue, and L. Zhang (2017). Replicating anomalies. Technical report, National Bureau of Economic Research.

Ilmanen, A. and L. N. Nielsen (2015). Are defensive stocks expensive? a closer look at value spreads.

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

Kelly, B. and S. Pruitt (2013). Market Expectations in the Cross-Section of Present Values. *The Journal of Finance 68*(5), 1721–1756.

Kelly, B. T., S. Pruitt, and Y. Su (2018). Characteristics are covariances: A unified model of risk and return. Technical report, National Bureau of Economic Research.

Koijen, R. S. and S. Van Nieuwerburgh (2011). Predictability of returns and cash flows. *Annu. Rev. Financ. Econ. 3*(1), 467–491.

Kozak, S. (2019). Kernel trick for the cross section. *Available at SSRN 3307895*.

Kozak, S., S. Nagel, and S. Santosh (2018). Interpreting factor models. *The Journal of Finance 73*(3), 1183–1223.

Kozak, S., S. Nagel, and S. Santosh (2019). Shrinking the cross-section. *Journal of Financial Economics*, Forthcoming.

Lettau, M. and S. Ludvigson (2001). Consumption, Aggregate Wealth, and Expected Stock Returns. *Journal of Finance 56*, 815–849.

Light, N., D. Maslov, and O. Rytchkov (2017). Aggregation of information about the cross section of stock returns: A latent variable approach. *The Review of Financial Studies 30*(4), 1339–1381.

Lochstoer, L. A. and P. C. Tetlock (2016). What drives anomaly returns?

Marriott, F. and J. Pope (1954). Bias in the estimation of autocorrelations. *Biometrika 41*(3/4), 390–402.

Moreira, A. and T. Muir (2017). Volatility-managed portfolios. *The Journal of Finance 72*(4), 1611–1644.

Nagel, S. and K. J. Singleton (2011). Estimation and Evaluation of Conditional Asset Pricing Models. *The Journal of Finance 66*(3), 873–909.

Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica: Journal of the Econometric Society*, 703–708.

Ohlson, J. A. (1995). Earnings, book values, and dividends in equity valuation. *Contemporary accounting research 11*(2), 661–687.

Papanikolaou, D. (2011). Investment shocks and asset prices. *Journal of Political Economy 119*(4), 639–685.

Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory 13*(3), 341–360.

Shiller, R. (1981). Do stock prices move too much to be justified by subsequent changes in dividends? *American Economic Review 71*, 321 – 436.

Stambaugh, R. F. (1999). Predictive regressions. *Journal of Financial Economics 54*(3), 375–421.

Stambaugh, R. F., J. Yu, and Y. Yuan (2012). The short of it: Investor sentiment and anomalies. *Journal of Financial Economics 104*(2), 288–302.

Stock, J. H. and M. W. Watson (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association 97*(460), 1167–1179.

Vuolteenaho, T. (2002). What drives firm-level stock returns? *The Journal of Finance 57*(1), 233–264.

Internet Appendix

A Key formulas

We derive key results. First, start with a set of returns $R_{t+1} = [r_{1,t+1} \cdots r_{N,t+1}]'$. The maximum conditional squared Sharpe ratio is

$$\mathrm{SR}_t^2 = \mathbb{E}_t \left[R_{t+1}\right]' \Sigma_{R,t}^{-1} \mathbb{E}_t \left[R_{t+1}\right]. \tag{A1}$$

A.1 Expected conditional squared Sharpe ratio, $\mathbb{E}\left[\mathrm{SR}_t^2\right]$

If the assets are conditionally uncorrelated, then $\Sigma_{R,t}$ is diagonal and the formula becomes

$$\mathrm{SR}_t^2 = \sum_{i=1}^{N} \frac{\mathbb{E}_t \left[r_{i,t+1}\right]^2}{\sigma_{i,t}^2}, \tag{A2}$$

where σ_i^2 is the conditional variance of the return on asset i. Assuming returns are homoskedastic and taking unconditional expectations we obtain

$$\mathbb{E}\left[\mathrm{SR}_t^2\right] = \sum_{i=1}^{N} \frac{\mathbb{E}\left(\mathbb{E}_t \left[r_{i,t+1}\right]^2\right)}{\sigma_i^2}. \tag{A3}$$

Substituting in the identity $\mathbb{E}\left(\mathbb{E}_t \left[r_{i,t+1}\right]^2\right) = \mathbb{E}\left[r_{i,t+1}\right]^2 + \mathrm{var}\left(\mathbb{E}_t \left[r_{i,t+1}\right]\right)$ we have

$$\mathbb{E}\left[\mathrm{SR}_t^2\right] = \sum_{i=1}^{N} \frac{\mathbb{E}\left[r_{i,t+1}\right]^2}{\sigma_i^2} + \sum_{i=1}^{N} \frac{\mathrm{var}\left(\mathbb{E}_t \left[r_{i,t+1}\right]\right)}{\sigma_i^2}. \tag{A4}$$

Using the definition of R-squared, $R_i^2 = 1 - \frac{\sigma_i^2}{\mathrm{var}(\mathbb{E}_t[r_{i,t+1}])+\sigma_i^2}$, we obtain $\frac{R_i^2}{1-R_i^2} = \frac{\mathrm{var}(\mathbb{E}_t[r_{i,t+1}])}{\sigma_i^2}$, which can be substituted in to get the formula in the paper:

$$\mathbb{E}\left[\mathrm{var}_t\left(m_{t+1}\right)\right] = \mathbb{E}\left(\mathrm{SR}_t^2\right) = \sum_i \frac{\mathbb{E}\left[r_{i,t+1}\right]^2}{\sigma_i^2} + \sum_i \left(\frac{R_i^2}{1-R_i^2}\right). \tag{A5}$$

A.2 Expected utility

We consider the perspective of an investor with mean-variance utility risk aversion parameter equal to γ. We maintain the assumption that $\Sigma_{R,t}$ is diagonal, constant, and known. Consider two portfolio strategies: (1) the static (s) strategy which does not use conditioning information and (2) the dynamic (d) strategy which can condition on $\mu_t = \mathbb{E}_t\left(R_{t+1}\right)$. The optimal portfolio weights are given by

$$w_{t,s} = \frac{1}{\gamma}\Sigma_{R,t}^{-1}\mu \tag{A6}$$

$$w_{t,d} = \frac{1}{\gamma}\Sigma_{R,t}^{-1}\mu_t \tag{A7}$$

where $\mu = \mathbb{E}\left(R_{t+1}\right)$ is the unconditional mean return. The date t expected utility for the agent under the two strategies are given by

$$U_{t,s} = \frac{1}{\gamma}\mu'\Sigma^{-1}\mu_t - \frac{1}{2\gamma}\mu'\Sigma^{-1}\mu \tag{A8}$$

$$U_{t,d} = \frac{1}{2\gamma}\mu_t'\Sigma^{-1}\mu_t, \tag{A9}$$

where expectations are taken under the same measure. Computing unconditional expectations we obtain

$$\mathbb{E}\left[U_{t,s}\right] = \frac{1}{2\gamma}\mu'\Sigma^{-1}\mu = \frac{1}{2\gamma}\sum_i \frac{\mathbb{E}\left[r_{i,t+1}\right]^2}{\sigma_i^2} \tag{A10}$$

$$\mathbb{E}\left[U_{t,d}\right] = \frac{1}{2\gamma}\mathbb{E}\left[\mu_t'\Sigma^{-1}\mu_t\right] = \frac{1}{2\gamma}\sum_i \frac{\mathbb{E}\left[r_{i,t+1}\right]^2}{\sigma_i^2} + \frac{1}{2\gamma}\sum_i\left(\frac{R_i^2}{1-R_i^2}\right). \tag{A11}$$

Therefore the "timing" term $\sum_i\left(\frac{R_i^2}{1-R_i^2}\right)$ exactly captures the increase in average utility obtained by using conditioning information.

A.3 Total $\mathbf{R^2}$

Again start with a set of returns $R_{t+1} = \left[r_{1,t+1} \cdots r_{N,t+1}\right]'$ with arbitrary cross-correlations. Define the total R^2 as

$$R^2_{\text{total}} \equiv \frac{\mathrm{tr}\left[\mathrm{cov}\left(\mathbb{E}_t\left[R_{i,t+1}\right]\right)\right]}{\mathrm{tr}\left[\mathrm{cov}\left(R_{i,t+1}\right)\right]}, \tag{A12}$$

where tr is the trace function. By similarity invariance of trace, this is equal to

$$R^2_{\text{total}} \equiv \frac{\mathrm{tr}\left[Q'\mathrm{cov}\left(\mathbb{E}_t\left[R_{i,t+1}\right]\right)Q\right]}{\mathrm{tr}\left[Q'\mathrm{cov}\left(R_{i,t+1}\right)Q\right]}, \tag{A13}$$

where and Q is *any* orthogonal matrix $\left(Q' = Q^{-1}\right)$. Next assume returns are homoskedastic, that is, $\Sigma_{R,t}$ is constant. This leads to the eigendecomposition $\Sigma_{R,t} = Q\Lambda Q'$. Denoting $PC_{t+1} = Q'R_{t+1}$ and substituting in we have

$$R^2_{\text{total}} = \frac{\mathrm{tr}\left[\mathrm{cov}\left(\mathbb{E}_t\left[PC_{t+1}\right]\right)\right]}{\mathrm{tr}\left[\Lambda\right] + \mathrm{tr}\left[\mathrm{cov}\left(\mathbb{E}_t\left[PC_{t+1}\right]\right)\right]}, \tag{A14}$$

where we use $\mathrm{cov}\left(PC_{t+1}\right) = \Lambda + \mathrm{cov}\left(\mathbb{E}_t\left[PC_{t+1}\right]\right)$ and additivity of trace. Next use $\frac{R_i^2}{1-R_i^2} = \frac{\mathrm{var}(\mathbb{E}_t[PC_{i,t+1}])}{\lambda_i}$ to obtain

$$R^2_{\text{total}} = \sum_{i=1}^{K}\left(\frac{R_i^2}{1-R_i^2}\right)\frac{\lambda_i}{\lambda}, \tag{A15}$$

where

$$\lambda = \text{tr}\left[\Lambda\right] + \text{tr}\left[\text{cov}\left(\mathbb{E}_t\left[PC_{t+1}\right]\right)\right] \tag{A16}$$

$$= \sum_{i=1}^{K} \frac{\lambda_i}{1-R_i^2} \approx \sum \lambda_i. \tag{A17}$$

A.4 Number of PCs

Start with prior beliefs on the maximum squared Sharpe ratio $\mathbb{E}\left[\text{SR}_t^2\right] \leq s^\star$ and the total R^2 $R^2_{\text{total}} \geq r^\star$.[30] Given these beliefs how many PCs should we include? Under the view that all included k PCs contribute equally to the total R^2, using Equation 7 and Equation 8, we can equivalently write

$$r^\star \leq k\left(\frac{R_i^2}{1-R_i^2}\right)\frac{\lambda_i}{\lambda} \tag{A18}$$

$$s^\star \geq \sum_{i=1}^{k}\left(\frac{R_i^2}{1-R_i^2}\right). \tag{A19}$$

Note this is analogous to the setup in Kozak et al. (2018) who assume that all included PCs contribute equally to cross-sectional heterogeneity in expected returns when determining the number of PCs to include. Combining these expressions we obtain the final formula:

$$\frac{r^\star}{s^\star} \leq \left[\frac{1}{k}\sum_{i=1}^{k}\frac{\lambda}{\lambda_i}\right]^{-1}. \tag{A20}$$

By inspection, the weaker the factor for a given set of assets the fewer PCs one may include given prior beliefs.

B Additional Results

We report supplemental empirical results.

B.1 Out-of-sample

Our main out-of-sample analysis uses a sample split where all parameters are estimated using the first half and used to construct OOS forecasts in the second half of the data. We consider two alternatives, expanding and rolling window analysis. For both, the OOS analysis begins on the same date as the main estimation, but predictive regression coefficients are reestimated each month. For rolling window, we use a twenty year (240 month) sample. Table A.1 presents results from the alternative OOS methods. The first row shows the coefficient estimate. The second row shows asymptotic t-statistics. The third and fourth rows show coefficients estimated from the first and second half data, respectively. The fifth shows the in-sample R^2. The next three rows give

[30] Here we ignore the static component of the Sharpe ratio.

Table A.1: Predicting Dominant Equity Components with BE/ME ratios

We report results from predictive regressions of excess market returns and five PCs of long-short anomaly returns. The market is forecast using the log of the aggregate book-to-market ratio. The anomaly PCs are forecast using a restricted linear combination of anomalies' log book-to-market ratios with weights given by the corresponding eigenvector of pooled long-short strategy returns. The first row shows the coefficient estimate. The second row shows asymptotic t-statistics estimated using the method of Newey and West (1987). The third and fourth rows show coefficients estimated from the first and second half data, respectively. The fifth shows the in-sample R^2. The next three rows give OOS R^2 based on split sample, expanding window, and 240 month rolling window analysis. The last row reports a reverse OOS R^2 where estimation is conducted in the second half of the sample and performance is measured in the first half.

	MKT	PC1	PC2	PC3	PC4	PC5
Own *bm* Full	0.76	4.32	1.62	1.80	4.86	1.56
	(1.24)	(4.31)	(1.81)	(2.01)	(3.74)	(0.78)
Own *bm* 1st	1.46	3.77	1.37	2.62	5.66	2.74
Own *bm* 2nd	2.79	4.91	7.68	2.83	4.31	2.14
R^2 Full	0.29	3.96	0.74	0.56	3.59	0.50
Split R^2	1.00	4.82	0.97	0.47	3.52	0.55
Expanding R^2	-0.53	4.43	-0.32	-0.30	2.59	-1.31
Rolling R^2	-0.28	3.04	-0.50	-1.10	2.47	-1.61
Reverse R^2	0.09	2.40	-7.30	1.17	2.65	1.62

OOS R^2 based on split sample, expanding window, and rolling window analysis. PCs 1 and 4 show remarkable stability of estimated coeffcients and substantial OOS R^2 using all three methods. This stability reflects the precision of the coefficient estimates documented by the t-statistics. Finally, the last row reports a reverse OOS R^2 where estimation is conducted in the second half of the sample and we evaluate the performance of the prediction in the first half of the sample. Here again, PC1 and PC4 have sizable R^2. Interestingly, PC3 and PC5 have larger R^2 than in the baseline, while PC2 does poorly, consistent with the instability of the predictive coefficient across periods. Overall, these various OOS approaches lead to similar conclusions to our baseline in terms of predictability of the dominant components.

B.2 Finite Sample Bias

The relative lack of Stambaugh-type bias for the PCs may be surprising given that bias for the aggregate market is large. However, this difference arises for two reasons. Assuming an AR(1) process for the predictor, Stambaugh (1999) shows that

$$E\left(\hat{\beta}-\beta\right)=c\rho_{xy}E\left(\hat{\rho}_x-\rho_x\right),$$

where $\hat{\beta}$ is the estimated predictive coefficient, ρ_{xy} is the contemporaneous correlation of innovations to x and y, and ρ_x is the autocorrelation of x. Marriott and Pope (1954) show that the bias in $\hat{\rho}_x$ is approximately proportional to ρ_x itself. Hence, the overall bias in $\hat{\beta}$ is proportional to $\rho_x \rho_{xy}$. We empircally estimate these quantities for the market and the anomaly PC portfolios to help decompose the lower simulated bias for PCs shown in Table 2.

In the first row of Table A.2 we report annualized AR(1) coefficients estimated by OLS from monthly data, $\rho_{annual} = \rho_{monthly}^{12}$. Unsurprisingly, they are much smaller for PCs relative to the aggregate market. By estimating the restricted VAR(1) assumed in Stambaugh (1999), we obtain estimates of the error correlation, shown in the second row of Table A.2. The error correlation is also substantially smaller for the PCs, further reducing the bias in estimated predictive coefficients.

Table A.2: Stambaugh Bias

The first row reports annualized AR(1) coefficients of *bm* ratios, estimated from monthly data $\left(\rho_{annual} = \rho_{monthly}^{12}\right)$. The second row reports the contemporaneous correlation of innovations to returns and *bm* ratios assuming a VAR(1) data-generating process.

	MKT	PC1	PC2	PC3	PC4	PC5
Persistence	0.87	0.52	0.61	0.53	0.27	0.44
Error correlation	-0.84	-0.67	-0.33	-0.34	-0.30	-0.15

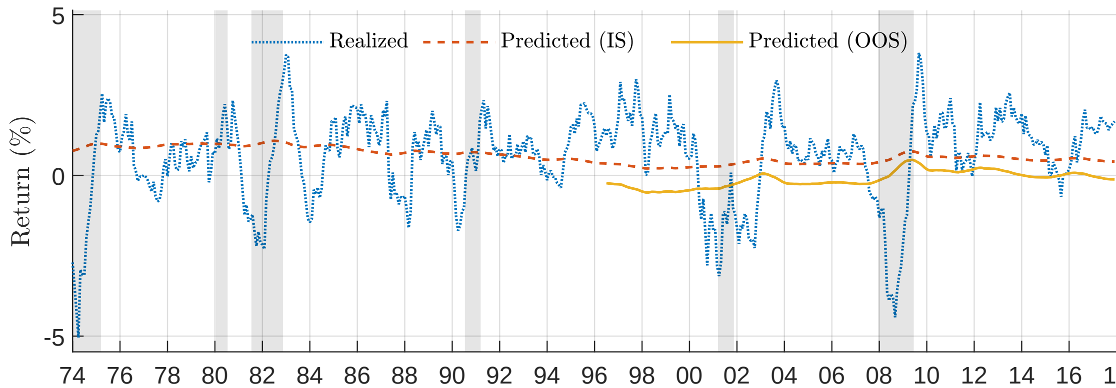
B.3 Macro Predictors

It is possible that price ratios are useful return forecasters of anomaly returns, but their predictive ability is subsumed by standard aggregate return predictors. We explore this by including the aggregate dividend-to-price ratio (D/P), cyclically-adjusted earnings-to-price (CAPE), lagged realized volatility, the term premium, corporate bond yield spread, consumption-to-wealth ratio from Lettau and Ludvigson (2001) (CAY), GDP growth, and aggregate sentiment from Baker and Wurgler (2006). We include each of these additional predictors one at a time to the regressions of the market and PC returns on their own *bm*. In Table A.3 we report the multivariate coefficients, t-statistics on the *bm* ratios and full sample R^2 values. The first row repeats the baseline estimates from Table 2. The remaining rows show that macro variables do not even partially drive out the price ratios when predicting returns. This is not surprising. Even if we knew the "true" macro variables that drive time-variation in expected returns, the empirically measured values are likely extremely noisy since quantities like consumption, wealth, and gdp are not directly observable. Price ratios, by contrast, are likely much better measured expected return proxies. Daniel and Moskowitz (2016) find that aggregate market volatility predicts returns on the momentum portfolio. Among the largest five PCs, we find that market volatility predicts only PC2, which has a large loading on momentum.

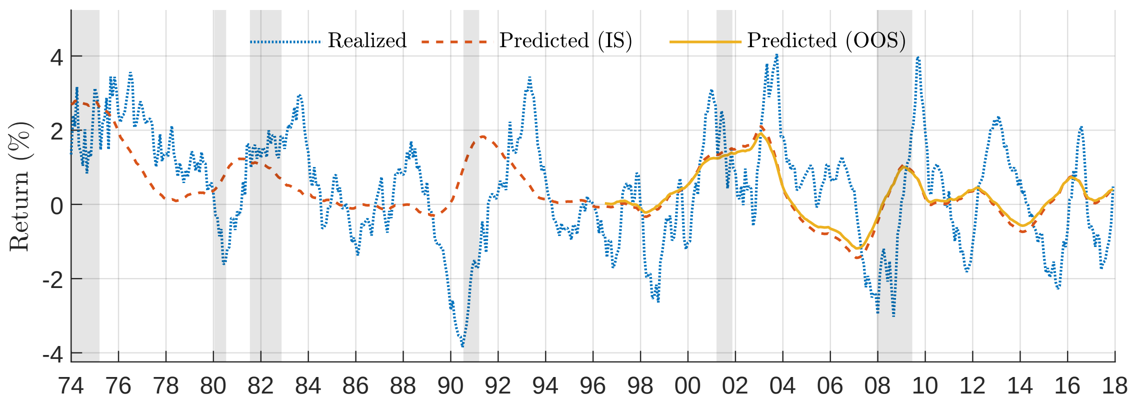
Table A.3: Including Macro Predictors

We we report the multivariate coefficients, t-statistics on the bm ratios and full sample R^2 values. The first row repeats the baseline estimates from Table 2.

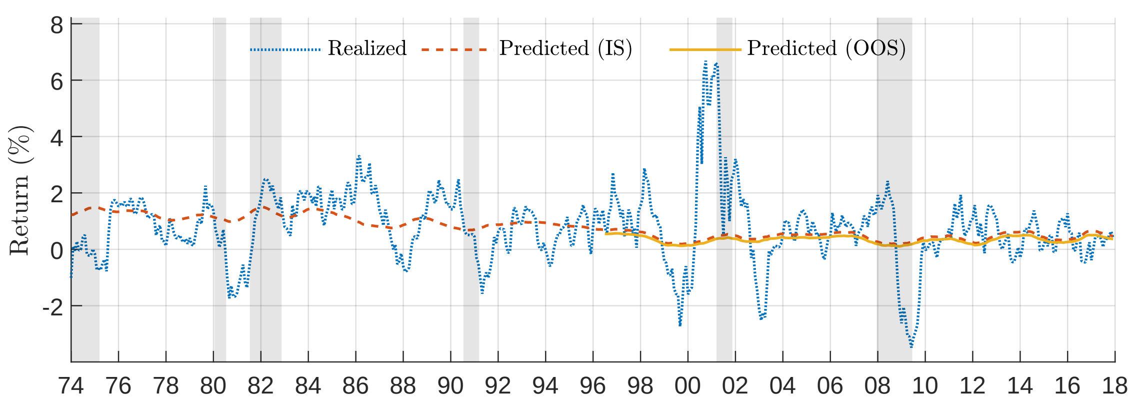
	MKT	PC1	PC2	PC3	PC4	PC5
Baseline	0.76	4.32	1.62	1.80	4.86	1.56
	(1.24)	(4.31)	(1.81)	(2.01)	(3.74)	(0.78)
	0.29	3.96	0.74	0.56	3.59	0.50
D/P	1.09	4.17	6.82	3.02	4.48	2.45
	(0.50)	(4.36)	(3.40)	(2.49)	(3.29)	(1.47)
	0.30	4.02	4.55	0.83	3.79	0.87
CAPE	1.21	4.31	1.97	2.64	4.44	1.89
	(1.42)	(4.29)	(1.90)	(2.96)	(3.65)	(1.06)
	0.39	3.97	0.77	0.87	4.14	0.77
Volatility	0.74	4.57	1.03	1.82	4.88	1.47
	(1.17)	(4.09)	(1.40)	(2.05)	(3.89)	(0.74)
	0.48	4.09	2.04	0.84	3.66	0.59
Term Premium	0.82	4.32	1.61	1.79	5.02	1.36
	(1.33)	(4.24)	(1.82)	(1.85)	(3.60)	(0.66)
	0.53	4.06	0.74	0.56	3.83	0.58
Corp. Spread	0.68	4.02	1.35	1.77	4.64	1.55
	(0.99)	(3.38)	(1.62)	(1.96)	(3.57)	(0.77)
	0.36	4.13	1.42	0.58	4.07	0.52
CAY	0.75	4.30	1.61	1.87	5.19	1.68
	(1.25)	(4.17)	(1.74)	(1.78)	(4.32)	(0.79)
	0.57	3.97	0.74	0.57	4.55	0.57
GDP growth	0.69	4.62	1.39	1.81	4.85	1.51
	(1.09)	(4.36)	(1.75)	(2.06)	(3.78)	(0.75)
	0.44	4.09	1.55	0.59	3.59	0.70
Sentiment	0.65	4.15	1.88	1.49	3.85	1.80
	(1.15)	(4.12)	(2.40)	(1.75)	(2.71)	(0.89)
	0.37	4.06	2.28	1.17	4.62	1.29



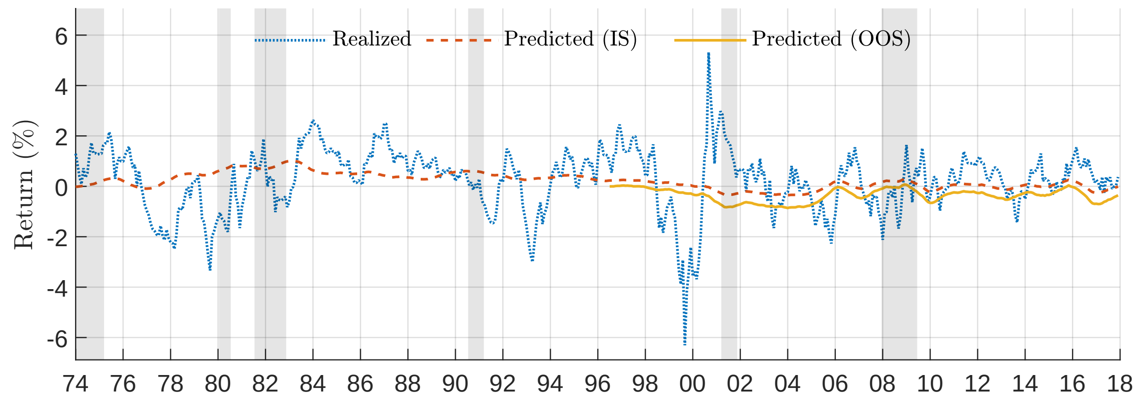
(a) Market



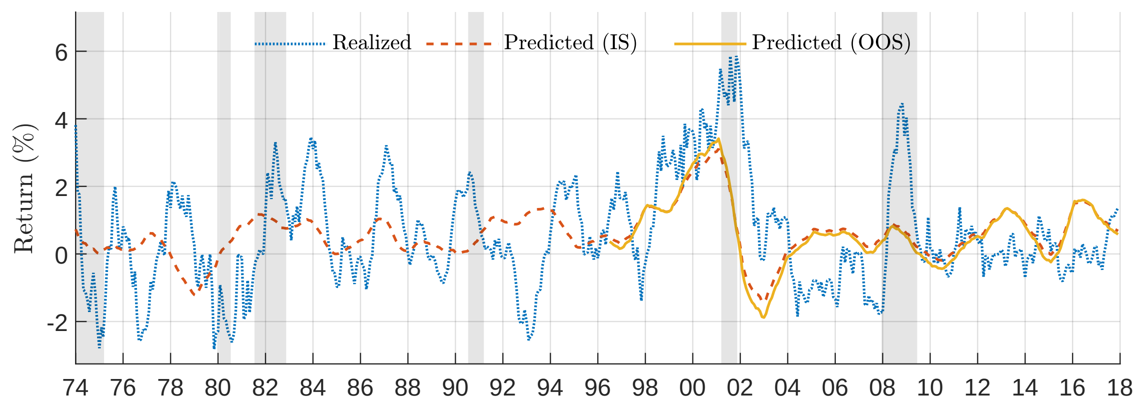
(b) PC1



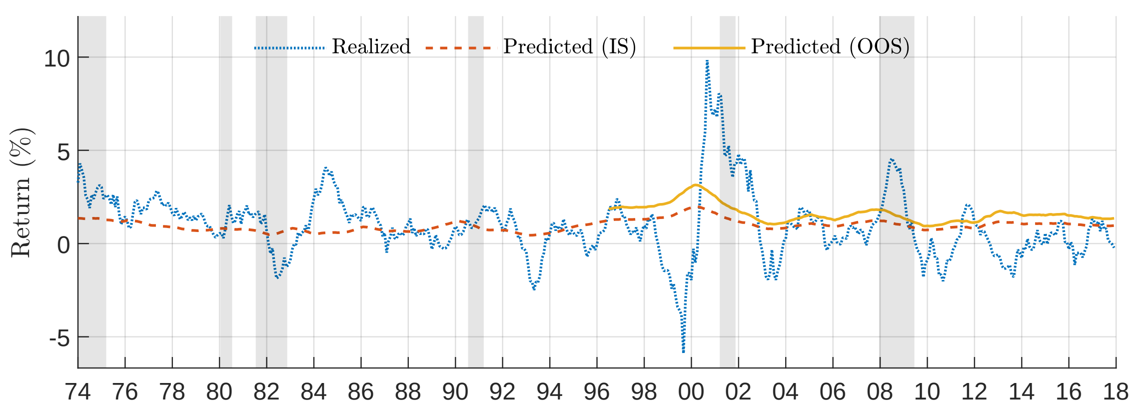
(c) PC2



(d) PC3



(e) PC4



(f) PC5

Figure A.1: Realized and Predicted Return. The plot shows realized returns along with full sample and out-of-sample forecasts of returns on the aggregate market and first five PC portfolios of the fifty anomalies.

B.4 Forecast and Realized Returns

Figure A.1 shows realized returns along with full sample and out-of-sample forecasts of returns on the aggregate market and first five PC portfolios of anomalies.

B.5 Anomaly Return Properties

Table A.4 shows annualized mean excess returns on the fifty anomaly long-short portfolios as well as the underlying characteristic-sorted decile portfolios.

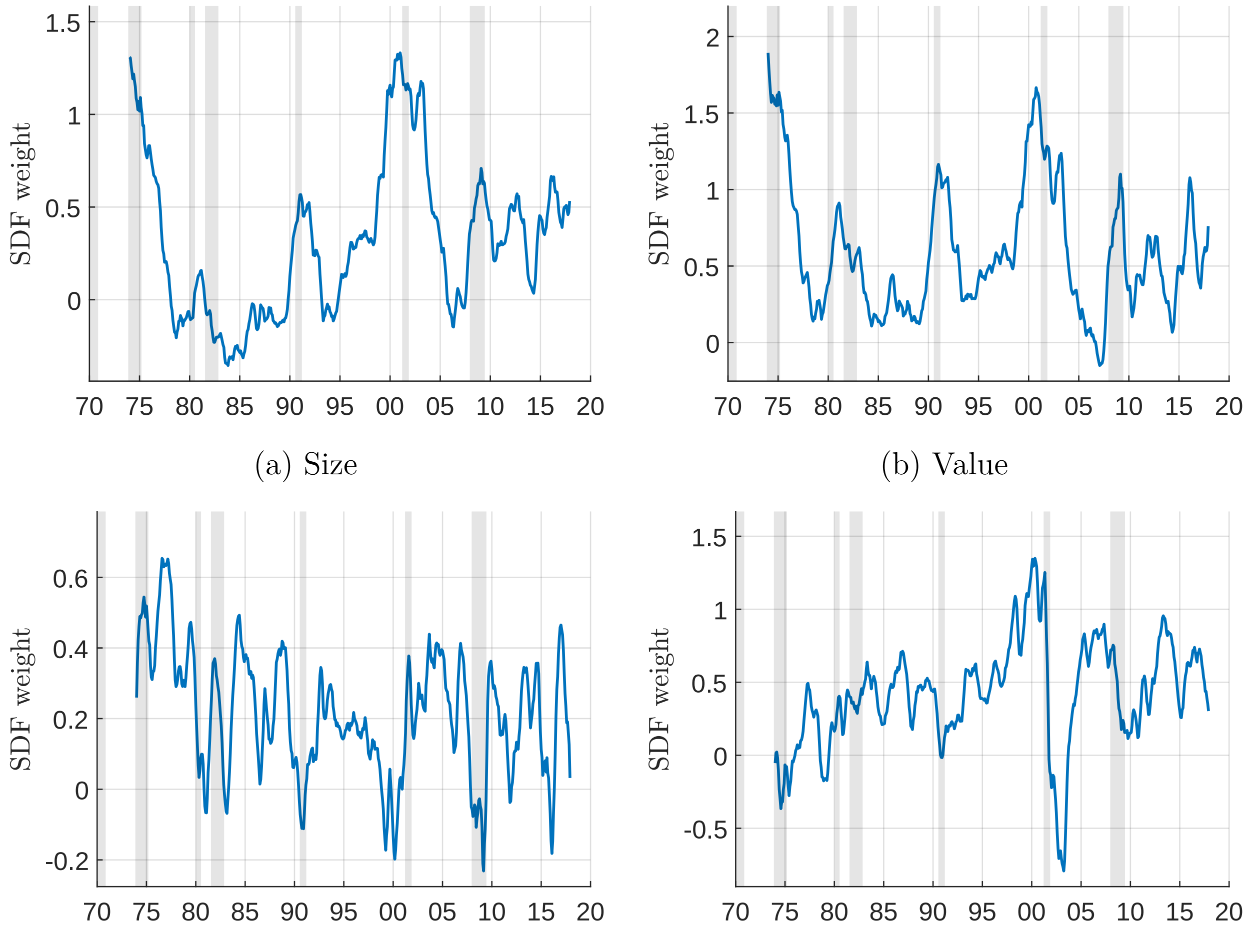
Table A.4: Part I: Anomaly portfolios mean excess returns, %, annualized

Columns P1 through P10 show mean annualized returns (in %) on each anomaly portfolio net of risk-free rate. The column P10-P1 lists mean returns on the strategy which is long portfolio 10 and short portfolio 1. Excess returns on beta arbitrage portfolios are scaled by their respective betas. F-score, Debt Issuance, and Share Repurchases are binary sorts; therefore only returns on P1 and P10 are reported for these characteristics. Portfolios include all NYSE, AMEX, and NASDAQ firms; however, the breakpoints use only NYSE firms. Monthly data from January 1974 to December 2017.

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P10-P1
1. Size	6.5	8.5	9.0	9.7	9.4	10.1	9.7	10.5	9.9	9.6	3.1
2. Value (A)	5.8	8.0	9.0	7.6	8.8	9.1	9.0	9.2	9.0	12.2	6.4
3. Gross Profitability	6.0	6.0	6.9	6.4	8.5	7.6	8.2	7.1	7.9	9.8	3.8
4. Value-Profitablity	4.7	6.4	4.9	7.1	9.1	8.6	11.1	11.9	12.1	13.7	9.0
5. F-score	6.9	-	-	-	-	-	-	-	-	7.9	1.0
6. Debt Issuance	7.0	-	-	-	-	-	-	-	-	8.7	1.7
7. Share Repurchases	7.0	-	-	-	-	-	-	-	-	8.4	1.4
8. Net Issuance (A)	3.5	5.8	9.4	8.8	7.8	7.8	7.1	9.1	8.9	11.8	8.3
9. Accruals	5.0	6.7	6.1	7.5	7.8	7.8	8.7	7.7	10.3	9.0	4.0
10. Asset Growth	5.8	7.4	7.9	7.8	8.1	7.7	7.9	9.3	10.6	10.0	4.2
11. Asset Turnover	4.8	7.3	6.8	7.0	8.2	9.1	9.6	7.6	10.2	9.8	5.0
12. Gross Margins	6.9	7.5	8.7	7.7	8.7	7.2	8.1	7.5	6.5	7.5	0.6
13. Earnings/Price	4.6	5.8	7.2	7.9	7.7	7.9	10.4	9.3	9.7	12.3	7.6
14. Cash Flows/Price	5.3	8.1	6.7	8.6	8.7	9.1	8.4	9.7	11.4	11.2	5.9
15. Net Operating Assets	3.8	7.0	7.5	4.5	8.3	8.1	8.5	8.3	9.4	9.1	5.2
16. Investment/Assets	5.2	5.7	8.3	7.0	8.9	7.2	8.1	9.2	9.0	11.0	5.8
17. Investment/Capital	7.0	7.3	6.9	8.0	7.6	9.0	7.8	8.2	9.0	9.9	2.9
18. Investment Growth	5.4	8.7	7.4	7.1	7.0	7.9	8.6	8.4	10.3	9.1	3.7
19. Sales Growth	7.9	7.6	7.9	7.0	8.1	9.1	7.3	8.4	9.3	7.3	-0.5
20. Leverage	6.2	7.3	7.4	10.8	7.9	8.6	9.2	9.2	9.4	8.9	2.7
21. Return on Assets (A)	4.5	8.8	7.9	8.1	7.7	7.6	7.9	8.3	7.1	7.7	3.2
22. Return on Book Equity (A)	6.4	7.3	7.0	8.2	7.0	8.1	7.1	8.0	6.9	8.4	2.0

Table A.4: Part II: Anomaly portfolios mean excess returns, %, annualized

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P10-P1
23. Sales/Price	5.5	6.7	7.4	8.9	9.5	9.4	9.9	11.4	11.5	13.1	7.7
24. Growth in LTNOA	7.4	6.9	7.1	9.1	6.5	7.8	7.3	8.6	8.6	8.4	1.0
25. Momentum (6m)	9.9	9.6	9.1	9.1	8.3	8.6	7.5	6.1	7.9	11.2	1.3
26. Value-Momentum	6.8	8.6	7.4	8.1	8.9	9.7	10.1	9.1	8.4	11.5	4.7
27. Value-Momentum-Prof.	6.4	8.3	8.2	8.8	7.6	5.9	8.2	9.3	11.9	14.8	8.4
28. Short Interest	7.1	6.6	9.1	9.4	8.7	7.1	7.7	6.5	5.1	5.8	-1.4
29. Momentum (12m)	-0.3	5.6	7.0	8.1	6.6	7.4	7.6	9.6	9.4	12.7	13.0
30. Industry Momentum	6.6	6.2	8.5	5.8	8.2	10.4	8.2	7.4	9.6	9.5	2.9
31. Momentum-Reversals	5.6	7.5	8.0	7.6	8.0	9.7	7.8	9.8	9.5	12.5	6.9
32. Long Run Reversals	7.4	7.4	8.2	9.0	8.6	9.0	8.8	9.9	10.6	11.8	4.4
33. Value (M)	6.4	7.0	7.3	7.4	8.7	7.9	9.7	7.8	12.8	12.3	5.9
34. Net Issuance (M)	4.6	6.1	10.8	8.8	9.2	7.7	7.9	8.7	10.6	11.2	6.6
35. Earnings Surprises	5.0	5.2	5.8	7.7	7.4	8.3	7.6	8.0	8.9	11.5	6.4
36. Return on Book Equity (Q)	2.4	6.3	7.4	5.4	6.4	7.1	8.2	8.2	7.8	9.8	7.5
37. Return on Market Equity	1.3	2.2	7.1	6.4	7.8	7.7	8.7	11.1	12.1	15.8	14.4
38. Return on Assets (Q)	2.8	5.3	8.1	7.9	7.9	7.6	8.8	8.1	7.6	8.6	5.8
39. Short-Term Reversals	4.0	5.0	7.2	7.3	7.4	8.5	9.5	9.9	10.3	8.4	4.4
40. Idiosyncratic Volatility	0.9	8.9	11.4	8.5	10.6	9.1	8.3	8.2	7.9	7.5	6.7
41. Beta Arbitrage	3.9	4.0	5.1	7.3	8.6	10.2	11.2	11.8	14.6	17.2	13.3
42. Seasonality	4.0	4.4	6.7	6.3	8.5	7.4	7.9	7.6	9.8	13.2	9.2
43. Industry Rel. Reversals	2.6	4.2	4.9	6.3	6.8	8.2	9.6	11.6	13.3	13.1	10.6
44. Industry Rel. Rev. (L.V.)	1.7	5.2	5.2	6.8	6.6	7.4	9.8	10.8	13.8	15.6	13.9
45. Ind. Mom-Reversals	4.1	5.3	6.2	6.3	8.4	7.9	8.4	9.5	10.4	14.7	10.6
46. Composite Issuance	4.7	6.5	6.6	7.1	8.0	8.0	7.5	8.1	10.3	10.8	6.1
47. Price	6.1	9.5	9.2	10.8	9.2	8.9	7.9	7.9	7.9	6.5	0.5
48. Share Volume	7.2	8.7	7.3	7.6	8.1	6.8	8.4	7.3	6.9	6.8	-0.4
49. Duration	5.4	7.5	9.0	8.3	9.3	10.1	9.8	9.5	11.0	11.8	6.5
50. Firm age	7.0	9.1	6.0	9.8	6.4	8.8	10.0	8.5	7.3	7.7	0.7



(a) Size

(b) Value

(c) Momentum

(d) ROA

Figure A.2: Anomaly SDF Weights. The plot shows implied SDF coefficients on the size, value, momentum and ROA anomaly portfolios..

Figure A.2 gives the time-series of implied SDF coefficients on the size, value, momentum and ROA anomaly portfolios.

Figure A.3 shows the annualized unconditional mean return and standard deviation of conditional mean return on the fifty anomaly portfolios. Unconditional mean returns are computed as sample average returns. Standard deviation of conditional mean return are model implied based on the expected returns of five PC portfolios. The cross-sectional correlation of these two quantities is -20%.

B.6 Role of the Rebalancing Frequency

Table A.5 studies the role of the rebalancing frequency for factor timing. To focus on anomaly timing, we report statistics for the "pure anomaly timing" strategy which always has zero weight

Table A.5: Rebalancing frequency

We report the average unconditional Sharpe ratio, expected utility for a mean-variance investor, monthly portfolio turnover, and correlation with the monthly strategy with various rebalancing frequencies. Turnover is measured as the sum of absolute changes in portfolio weights divided by the sum of absolute initial portfolio weights.

	Monthly	Quarterly	Semi-annual	Annual
Sharpe ratio	0.71	0.72	0.82	0.79
Expected utility	1.26	1.08	0.88	0.81
Turnover	0.41	0.20	0.13	0.08
Correlation	1.00	0.98	0.92	0.88

in the market and has zero weight on average in each of the fifty anomaly portfolios.[31] We change portfolio weights on the anomalies monthly, quarterly, semi-annually, or annually. The first two rows report the Sharpe ratio and expected utility performance measures. The third row reports portfolio turnover, which we construct as follows. Given portfolio weights $w_{i,t}$ on anomaly i at date t, we construct period t turnover as $\sum_i |w_{i,t} - w_{i,t-1}| \,/\, \sum_i |w_{i,t-1}|$ which measures absolute trading scaled by gross exposure.[32] We report the average of this monthly measure over our sample. The last row of Table A.5 reports the correlation of each of the portfolios with the baseline monthly rebalanced return. Interestingly, the performance of the portfolios does not deteriorate meaningfully. The unconditional Sharpe ratio actually increases from 0.71 to 0.79 with annual rebalancing. Expected utility declines from 1.26 to 0.81, still a substantial value. For comparison, the static factor investing strategy yields an expected utility of 1.66 so even with annual rebalancing, timing benefits are economically meaningful. The correlation of the slower strategies with our baseline further confirms that the strategies are not that different. Even with annual rebalancing the correlation drops to only 0.88, showing that lowering the rebalancing frequency does not generate substantial tracking error. The signaling value of the predictors we use is sufficiently persistent to be used without continuous tracking.

Our anomaly timing strategy has a monthly turnover of 41%. Changing nearly half of positions each month might seem large, but it is important to remember that nothing in the construction of our strategy imposes a smooth trading path. The lower rebalancing frequencies drastically lower the turnover rate down to 8% with annual rebalancing. It is tempting to conclude that that these strategies are implementable in practice. Indeed, these numbers are in line with usual trading activity of investment funds. Griffin and Xu (2009) show that the median hedge fund has 8.5% monthly turnover and even the median mutual fund has 5% turnover. However, to reach a firm conclusion in terms of implementability, one would need a clear model of transaction costs. In addition, the transaction costs would likely depend of the scale at which the strategies are implemented.

[31] The factor timing strategy has about 25-30% lower turnover at all rebalancing frequencies.

[32] Since our portfolios are all zero cost excess returns, the standard definition which divides by portfolio equity makes little sense in this context.

Table A.6: Volatility timing

We report the mean and standard deviation of the conditional variance of the SDF based on three estimates. The first column uses the SDF variance shown in Figure 2 based on return forecasts in Table 2 and assumes returns are homoskedastic. The second column assumes returns are not predictable and uses estimates of conditional return variances constructed from a regression of squared forecast errors on lagged realized variance. The final column combines both mean and variance forecasting.

	Means	Variances	Both
$E\left[\mathrm{var}_t\left(m_{t+1}\right)\right]$	2.96	2.19	3.54
$\mathrm{std}\left[\mathrm{var}_t\left(m_{t+1}\right)\right]$	2.17	0.74	2.06

B.7 Conditional Variance of the SDF

Figures A.4 and A.5 show conditional variance of SDFs, as well as the relationship between SDF variance and inflation.

B.8 Volatility timing

As discussed in Moreira and Muir (2017), optimal timing strategies rely not only on estimates of conditional expected returns, but also conditional volatilities. Going back to our one-asset example at the beginning of Section 4.1, consider the situation where volatility changes independently from expected returns. Then the average squared Sharpe ratio becomes

$$\left(\mathbb{E}\left[\mu_t\right]^2+\mathrm{var}\left[\mu_t\right]\right)\left(\mathbb{E}\left[\frac{1}{\sigma_t}\right]^2+\mathrm{var}\left[\frac{1}{\sigma_t}\right]\right),$$

the gains from timing returns and volatility are multiplicative.

In our multivariate setting, we need to construct estimates of $\Sigma_{Z,t}$. We proceed as follows. For each of our five principal components and the market returns, we compute the realized volatility of daily returns during the previous month. We use these realized variances to create a forecast of the squared monthly prediction errors in the following month using a simple regression for each return series. These forecasts constitute the diagonal elements of $\Sigma_{Z,t}$. We confirmed that using GARCH(1,1) volatility forecasts leads to similar conclusions. We further assume that the five principal components and the market are *conditionally* orthogonal, and set the off-diagonal elements to 0.[33]

In Section 5.4 we report statistical properties of estimated stochastic discount factors which incorporate time-varying means, variances, or both. We compute the mean and standard deviation of the corresponding SDF variances. Figure A.6 shows the time-series of the conditional SDF variance implied by each of these three estimates. Examining the two time series, we note that the largest volatility spikes tend to mitigate the effect of high expected returns on SDF variance. For

[33]The five components and the market are *unconditionally* orthogonal by construction.

example this coincidence occurs during the Internet boom and bust, and also during the financial crisis of 2008.

C Statistical Approach

We first discuss an alternative statistical motivation behind our methodology, then derive some useful statistical properties.

C.1 An Alternative Statistical Motivation

Another way to approach our empirical exercise is to look for common sources of variation in risk premia across base assets or factors. For example, starting from a vector of candidate predictors X_t, we want to assess their usefulness to forecast the returns. In a linear setting, this corresponds to studying the vector of coefficients b_i' in the panel regression:

$$R_{i,t+1} = a_i + b_i' X_t + \varepsilon_{i,t+1}, \tag{A21}$$

where one can replace $R_{i,t+1}$ by $F_{j,t+1}$ if focusing on factors. There are multiple ways to aggregate the information in the estimated coefficients of interest, b_i, to judge the success of X_t as a predictor.

One can ask if X_t predicts "something": is there a linear combination of the coefficients $b = [b_1 \cdots b_n]$ that is statistically distinct from zero? This corresponds exactly to a standard Wald test. This notion of predictability, is intuitively too lax. For instance, our conclusion about the predictive value of X_t could be driven by its ability to predict only a few assets or the lowest variance PC portfolios. A small amount of noise in measured returns can lead to significant spurious predictability of the smallest PC portfolios, even in population. This issue is exacerbated in small samples.

The other extreme is to ask whether elements of X_t predicts "everything", or that all coefficients in a row of b are statistically distinct from zero. For instance Cochrane and Piazzesi (2005) obtain such a pattern predicting Treasury bond returns of various maturities using the cross-section of yields, concluding to the presence of a single common factor in expected returns. While this approach can uncover interesting patterns, it is likely to be too stringent. We show in Section C.2 that such a test is often equivalent to testing whether X_t predicts the first principal component of realized returns. In other words, finding uniform predictability across all assets simply finds predictability of the "level" factor in returns. In contrast, we show in Section C.3 that if a predictor is useful for forecasting index-neutral factor returns, captured by a long-short portfolio, but not for aggregate returns, individual asset predictive regressions are unlikely to uncover such predictability.

Our approach strikes a balance between these two extremes by asking whether X_t predicts the largest principal components of returns. In other words, we focus on common predictability along the few dimensions explaining a large fraction of realized returns. Focusing on components with a large explanatory power avoids the issue of the Wald test. Entertaining multiple dimensions avoids the other extreme of only focusing on the first component of returns, and allows us to study time-series predictability of cross-sectional strategies.

Our approach to the predictability of cross-sections of returns is focused on predicting important dimensions of the data rather than considering regressions at the individual asset level. In this section, we study more systematically the relation between predicting important components of returns and predicting individual returns.

We consider three features that were relevant in our empirical applications and provide ways to quantify them more generally. First, there is a strong link between predicting the first principal component of returns and predicting each individual return. Second, it is difficult to detect predictability of the second or higher components of returns in individual regressions when the first component is large. Third, joint tests of significance in individual regressions are susceptible to picking up small unimportant patterns of predictability.

C.2 First Principal Component and Individual Regressions

A common empirical situation is that a family of returns $\{R_{i,t+1}\}_{i\in I}$ has a strong common component F_{t+1}. When this component is predictable by a variable X_t, does this imply that the individual returns are predictable by X_t? We answer this question quantitatively by deriving a series of bounds linking the predictability of F_{t+1} with the individual predictability of asset returns. We first zoom in on one particular return before considering properties for an entire family of returns.

One individual return: a purely statistical bound. Define $\mathrm{R}^2_{1,i}$ as the population R-squared of the contemporaneous regression of an individual asset on the common component,

$$R_{i,t+1} = \lambda_i F_{t+1} + \varepsilon_{i,t+1}, \tag{A22}$$

and R^2_X as the R-squared of the predictive regression of the factor,

$$F_{t+1} = \beta_1 X_t + u_{t+1}. \tag{A23}$$

We are interested in $\mathrm{R}^2_{X,i}$, the R-squared of the predictive regression

$$R_{i,t+1} = b_i X_t + v_{t+1}. \tag{A24}$$

The following proposition characterizes a lower bound on this quantity.[34]

Proposition 2. *If a variable X_t predicts a factor F_{t+1} with R-squared R^2_X and an individual return is explained by this factor with R-squared $R^2_{1,i}$, then a lower bound for the R-squared $R^2_{X,i}$ of predicting this return using X_t is given by:*

$$R^2_{X,i} \geq \max\left(\sqrt{R^2_{1,i} R^2_X} - \sqrt{\left(1 - R^2_{1,i}\right)\left(1 - R^2_X\right)}, 0\right)^2. \tag{A25}$$

Proof. By the definition of a regression R^2 we have $\mathrm{R}^2_{1,i} = \frac{\lambda_i^2 \mathrm{var}(F_{t+1})}{\mathrm{var}(R_{i,t+1})}$, $\mathrm{R}^2_X = \frac{\beta_1^2}{\mathrm{var}(F_{t+1})}$, and $\mathrm{R}^2_{X,i} = \frac{b_i^2}{\mathrm{var}(R_{i,t+1})}$. The the linearity of regression we have $b_i = \lambda_i \beta_1 + \mathrm{cov}\left(X_t, u_{i,t+1}\right)$. We can bound the

[34]Without loss of generality, we assume that the predictor X_t has unit variance.

second term in this expression:

$$\begin{aligned}\left|\operatorname{cov}\left(X_t, u_{i,t+1}\right)\right| &= \left|\operatorname{corr}\left(X_t, u_{i,t+1}\right)\right| \sqrt{\operatorname{var}\left(u_{i,t+1}\right)} \\ &\leq \sqrt{1-\mathrm{R}_{1,i}^2}\sqrt{\operatorname{var}\left(u_{i,t+1}\right)},\end{aligned}$$

where the bound comes from the fact that the correlation matrix of $u_{i,t+1}$, F_{t+1} and X_{t+1} has to be semidefinite positive and therefore have a positive determinant.

If $|\lambda_i \beta_i| \leq \sqrt{1-\mathrm{R}_{1,i}^2}\sqrt{\operatorname{var}\left(u_{i,t+1}\right)}$, then 0 is a lower bound for $\mathrm{R}_{X,i}^2$. In the other case, we obtain the following bound:

$$\begin{aligned}\mathrm{R}_{X,i}^2 &\geq \frac{\left(\lambda_i \beta_1 - \sqrt{1-\mathrm{R}_{1,i}^2}\operatorname{var}\left(u_{i,t+1}\right)\right)^2}{\operatorname{var}\left(R_{i,t+1}\right)} \\ &\geq \left(\sqrt{\frac{\lambda_i^2 \beta_1^2}{\operatorname{var}\left(R_{i,t+1}\right)}} - \sqrt{1-\mathrm{R}_{1,i}^2}\sqrt{\frac{\operatorname{var}\left(u_{i,t+1}\right)}{\operatorname{var}\left(R_{i,t+1}\right)}}\right)^2 \\ &\geq \left(\sqrt{\mathrm{R}_{1,i}^2 \mathrm{R}_X^2} - \sqrt{\left(1-\mathrm{R}_{1,i}^2\right)\left(1-\mathrm{R}_X^2\right)}\right)^2\end{aligned}$$

Putting the two cases together gives Equation A25. □

Intuitively, if X_t strongly predicts the common factor, and the factor has high explanatory power for individual returns, then X_t should predict the individual returns as well. The bound is indeed increasing in the R-squared of these two steps. However, it is lower than the product of the two R-squared — a naive guess that assumes "transitivity" of predictability. This is because the predictor X_t might also predict the residual $\varepsilon_{i,t+1}$ in a way that offsets the predictability coming from the factor. The orthogonality of F_{t+1} and $\varepsilon_{i,t+1}$ limits this force, but does not eliminate it.

To get a quantitative sense of the tightness of this bound, consider the case of bond returns. The level factor explains about 90% of the variation in individual returns, and it can be predicted with an R-squared around 25%. Plugging into our bound, this implies a predictive R-squared of at least 4% for a typical individual bond return. This is a sizable number, but also much less than the 22.5% implied by a naive approach.

One individual return: a bound with an economic restriction. One reason this bound is relatively lax is that it does not take into account the nature of the variable $\varepsilon_{i,t+1}$. Indeed, if, as is the case in our setting, the component F_{t+1} is itself an excess return, the residual $\varepsilon_{i,t+1}$ is one too. It is therefore natural to make the economic assumption that it cannot be "too"' predictable by the variable X_t. This corresponds to imposing an upper bound $\mathrm{R}_{\max}^2$ on the R-squared of the predictive regression of $\varepsilon_{i,t+1}$ by X_{t+1}.[35] In this case, our bound becomes:

$$\mathrm{R}_{X,i}^2 \geq \max\left(\sqrt{\mathrm{R}_{1,i}^2 \mathrm{R}_X^2} - \sqrt{\mathrm{R}_{\max}^2\left(1-\mathrm{R}_X^2\right)}, 0\right)^2. \tag{A26}$$

[35] One way to determine a reasonable bound on $\mathrm{R}_{\max}^2$ is to note that the standard deviation of an asset's conditional Sharpe ratio equals $\sqrt{\frac{\mathrm{R}_{x,i}^2}{1-\mathrm{R}_{x,i}^2}}$.

Such an approach can considerably tighten the bound. For instance, in our example for treasuries, one could impose an upper bound of 25% for predicting the residual. This yields a lower bound on predicting the return $R_{i,t+1}$ of 10%, a much larger number, statistically and economically.

Family of returns: the symmetric case. Another reason that predictability of the common factor must transmit to predictability of individual returns is that by design it absorbs common variation across all those returns. To highlight this point, we consider the following simple symmetric case. We assume that the factor is the average of all the individual returns, $F_{t+1} = \frac{1}{N}\sum_i R_{i,t+1}$. We further assume that all assets have the same loading on the factor and the factor has the same explanatory power for each return. This corresponds to constant λ_i, and R^2_{1i} across assets. We then immediately have:

$$\sum_i u_{i,t+1} = 0$$
$$\sum_i \mathrm{cov}\left(X_t, u_{i,t+1}\right) = 0.$$

Letting $\gamma_i = \mathrm{cov}\left(X_t, u_{i,t+1}\right)$ we then obtain an expression for an individual asset:

$$\begin{aligned}\mathrm{R}^2_{X,i} &= \frac{\left(\lambda_i \beta_1 + \gamma_i\right)^2}{\mathrm{var}\left(R_{i,t+1}\right)} \\ &= \mathrm{R}^2_1 \mathrm{R}^2_X + \frac{\gamma_i^2}{\mathrm{var}\left(R_{i,t+1}\right)} + 2\gamma_i \frac{\lambda_i \beta_1}{\mathrm{var}\left(R_{i,t+1}\right)}\end{aligned}$$

Finally, taking averages across assets we have:

$$\mathbb{E}_i\left[\mathrm{R}^2_{X,i}\right] = \mathrm{R}^2_1 \mathrm{R}^2_X + \mathrm{var}_i\left(\mathrm{R}^2_{X,i}\right), \tag{A27}$$

where $\mathbb{E}_i(\cdot)$ and $\mathrm{var}_i(\cdot)$ are the mean and variance in the cross section of individual returns and we use the fact that we use the fact that $\mathbb{E}_i\left[\gamma_i\right] = 0$. This formula implies that the average explanatory power is now at least as large as given by the transitive formula. This would correspond to 22.5% in our example, almost the same value as the predictive R-squared for the common factor. Furthermore, the more unequal this predictive power is across assets, the stronger it must be on average. That is, if the variable X_t does less well than the transitive R-squared for some particular returns, it must compensate more than one-to-one for the other assets.

From predicting "everything" to aggregate returns. Maintaining the same assumptions, we can rearrange Equation A27 to see what the predictability of "everything" implies for predictability of the common factor. We have:

$$\mathrm{R}^2_X = \frac{\mathbb{E}_i\left[\mathrm{R}^2_{X,i}\right] - \mathrm{var}_i\left(\mathrm{R}^2_{X,i}\right)}{\mathrm{R}^2_1}.$$

At first this may not seem very powerful since $\mathrm{var}_i\left(\mathrm{R}^2_{X,i}\right)$ could be large. This maximal variance, however, is related to the average $\mathbb{E}_i\left[\mathrm{R}^2_{X,i}\right]$. Consider the simple example of only two assets. Then, if the average $\mathbb{E}_i\left[\mathrm{R}^2_{X,i}\right]$ is 10%, the maximal variance is only 1%, which obtains when $\mathrm{R}^2_{X,1} = 0\%$ and $\mathrm{R}^2_{X,2} = 20\%$. In general with two assets we have

$$\mathrm{var}_i\left(\mathrm{R}^2_{X,i}\right) \leq \left(0.5 - \left|\mathbb{E}_i\left[\mathrm{R}^2_{X,i}\right] - 0.5\right|\right)^2$$

which gives the bound

$$\mathrm{R}^2_X \geq \frac{\mathbb{E}_i\left[\mathrm{R}^2_{X,i}\right] - \left(0.5 - \left|\mathbb{E}_i\left[\mathrm{R}^2_{X,i}\right] - 0.5\right|\right)^2}{\mathrm{R}^2_1}.$$

For large N, the Bhatia-Davis inequality gives:

$$\mathrm{R}^2_X \geq \frac{\left(1 - \mathrm{R}^2_{\max}\right)\mathbb{E}_i\left[\mathrm{R}^2_{X,i}\right] + \mathbb{E}_i\left[\mathrm{R}^2_{X,i}\right]^2}{\mathrm{R}^2_1},$$

where $\mathrm{R}^2_{\max}$, as before, is the maximum $\mathrm{R}^2_{X,i}$ from any individual asset forecasting regression. For reasonable values of $\mathrm{R}^2_{\max}$, such as 0.5 or less, the bound implies that ~22% average R^2 we obtain for individual bonds implies at least 18% R^2_X, the R-squared when predicting the aggregate portfolio return.

C.3 Low Power of Individual Tests

While individual regressions are strongly related to predicting the first common component of returns, they can face challenges in detecting predictability of other factors. We provide a way to quantify this issue by characterizing the statistical power of a test of significance for a predictor that only predicts one particular component of returns. We illustrate this idea in the simple case of an i.i.d. predictor. Simulations confirm these ideas extend to a situation with persistent predictors.

I.i.d. predictor. Consider first the case where the forecasting variable X_{t+1} has i.i.d. draws.[36] Suppose that X_t forecasts only one particular principal component j with population R-squared R^2_X and the remaining principal component returns are i.i.d. Gaussian with known mean.[37] For power analysis, we consider repeated samples of length T.[38]

When directly forecasting the principal component return, $F_{j,t+1}$, the power to correctly reject the null with test of nominal size α is

$$\mathrm{power}\left(F_2\right) = G\left(-t_{\alpha/2,T} - z\right) + \left[1 - G\left(t_{\alpha/2,T} - z\right)\right], \tag{A28}$$

where G is the CDF of a t-distribution with T degrees of freedom, $z = \sqrt{\mathrm{R}^2_X}\sqrt{T}\left(1 - \mathrm{R}^2_X\right)^{-\frac{1}{2}}$, and $t_{\alpha/2,T}$ is the $\frac{\alpha}{2}$ critical value from the t-distribution.

[36]The formulas hereafter admit simple generalizations to multivariate prediction.

[37]More generally, the components need not be principal components. They must be uncorrelated and only one particular component must be forecastable by our predictor. If the mean is unknown, the results below are unchanged except that the degrees of freedom are $T - 1$ instead of T.

[38]The analysis treats X as stochastic. With fixed X the distribution is normal instead of a Student t.

In contrast, when directly forecasting an individual return, $R_{i,t+1}$, the power is

$$\text{power}\left(R_i\right) = G\left(-t_{\alpha/2,T} - \zeta\right) + \left(1 - G\left(t_{\alpha/2,T} - \zeta\right)\right), \tag{A29}$$

where $\zeta = \sqrt{\mathrm{R}_X^2}\sqrt{T}\left(\left(1-\mathrm{R}_X^2\right) + \frac{1-\mathrm{R}_{j,i}^2}{\mathrm{R}_{ji}^2}\right)^{-\frac{1}{2}}$. By symmetry of the t-distribution and because $\zeta \leq z$, we immediately obtain that power (F_2) is larger than power (R_i) for all assets. Therefore, there is always more information about predictability of the important component by studying it directly.

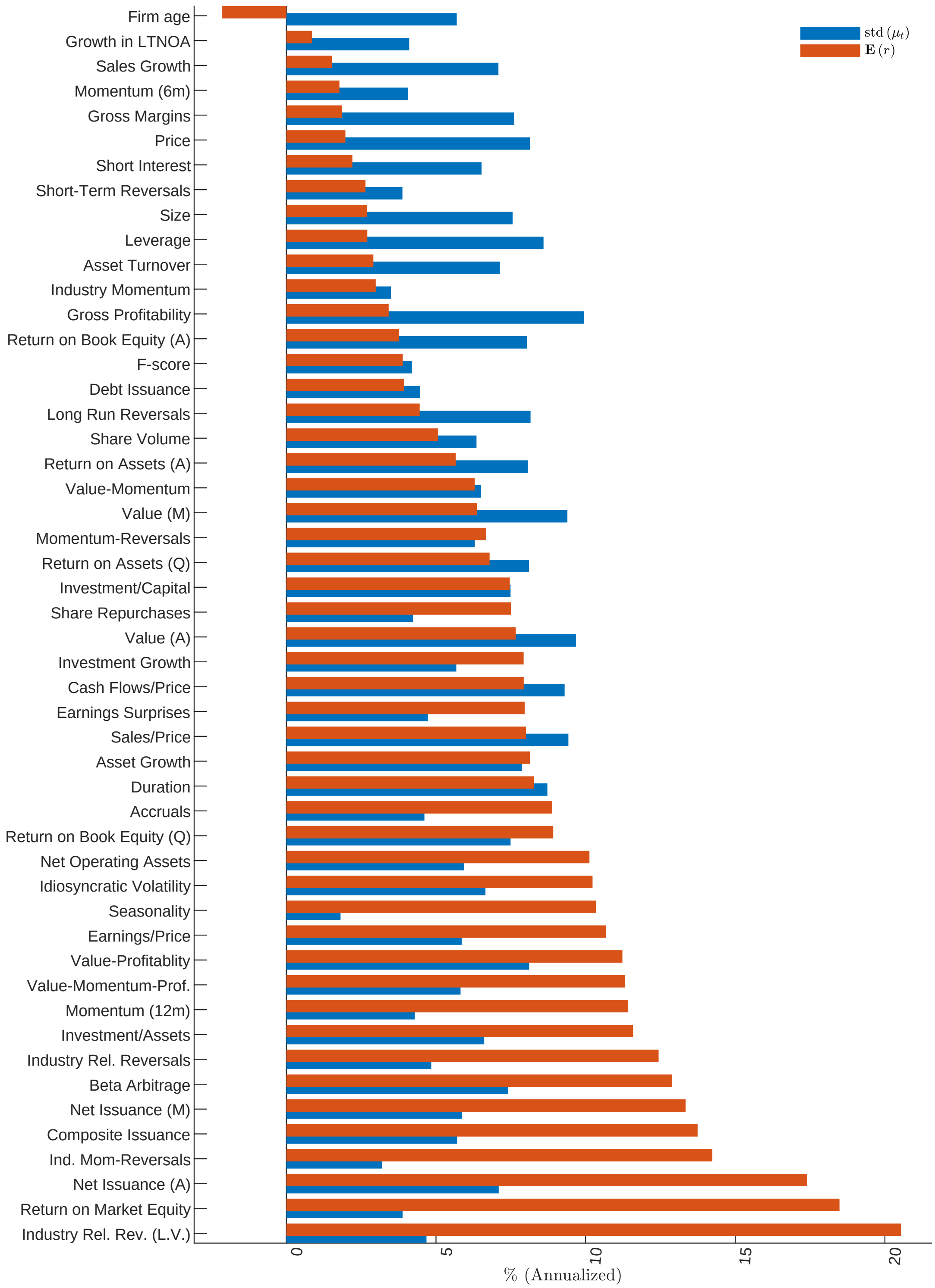


Figure A.3: Anomaly Expected Returns. The plot shows the annualized unconditional mean return and standard deviation of conditional mean return on the fifty anomaly portfolios.

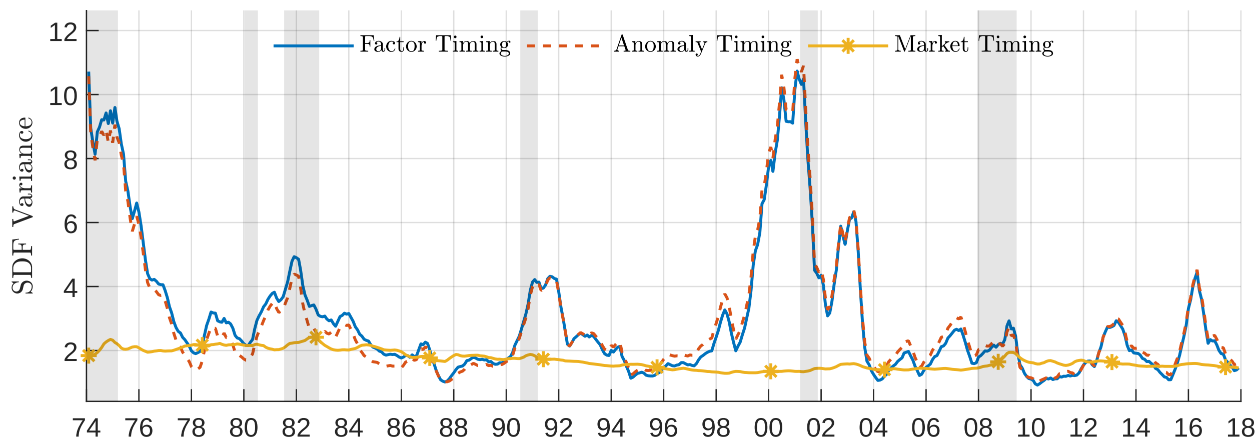


Figure A.4: Conditional Variance of SDFs. This figure plots the conditional variance of the SDF constructed under various sets of assumptions. "Factor timing" (solid blue line) is our full estimate, which takes into account variation in the means of the PCs and the market. "Anomaly timing" (dashed red line) imposes the assumption of no market timing: the conditional expectation of the market return is replaced by its unconditional counterpart. Conversely, "Market timing" (starred yellow line) allows for variation in the mean of the market return, but not the means of the factors.

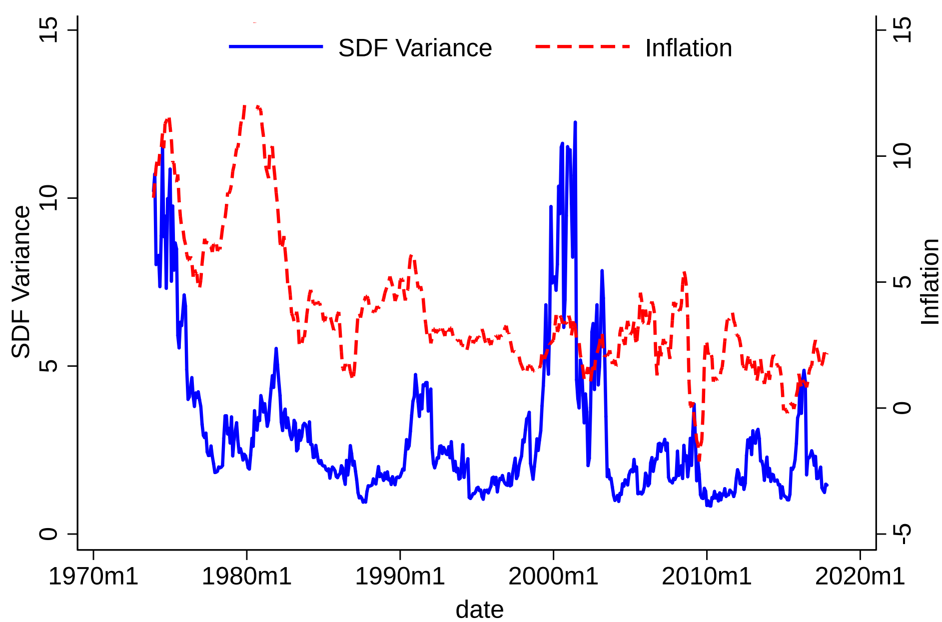


Figure A.5: Variance of the SDF and inflation. This figure plots the conditional variance of the SDF (solid blue line), and inflation rate over the previous year (dashed rate line). The SDF variance is constructed using the predictive regressions reported in Table 2. The inflation rate is the annual log change in the CPI.

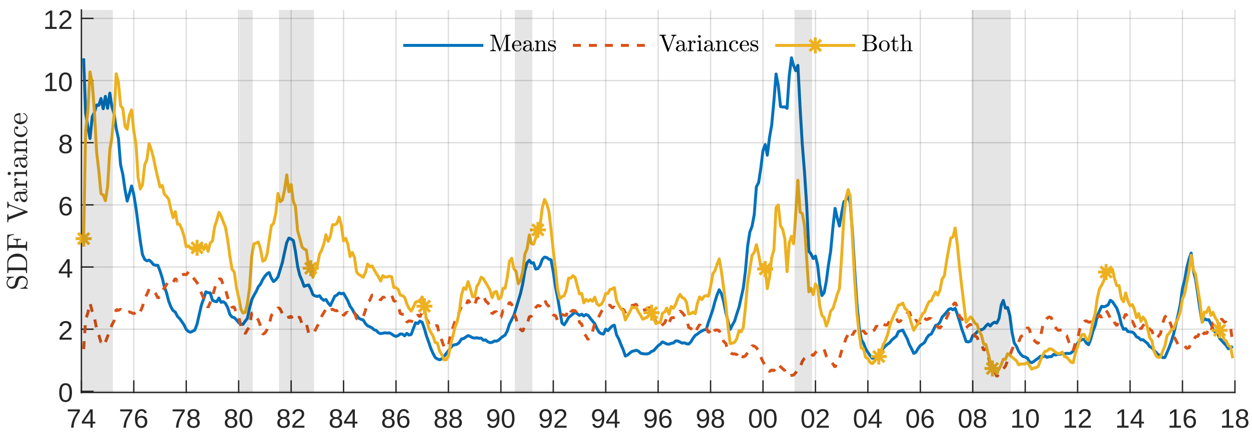


Figure A.6: Conditional Variance of SDFs. This figure plots the model-implied conditional variance of the SDF constructed in three ways. The solid blue line uses only timing of conditional means but constant variance. The red dashed line ignores predictability of returns but times variances. The yellow starred line times both means and variances.