Predicting Relative Returns*

Valentin Haddad
UCLA & NBER

Serhiy Kozak
University of Michigan

Shrihari Santosh
University of Maryland

October 24, 2017

Abstract

Across a variety of asset classes, we show that relative returns are highly predictable in the time-series in and out of sample, much more so than aggregate returns. Dominant principal components of equity anomalies, a portfolio of Treasuries sorted by maturity, and a currency carry portfolio are more predictable than the index return in their respective asset classes. We show that the common practice of predicting each individual asset separately obscures predictability of relative returns and is often equivalent to predicting only the index. Our approach to predictability uncovers multiple statistically robust and economically relevant sources of discount-rate variation.

*Haddad: valentin.haddad@anderson.ucla.edu; Kozak: sekozak@umich.edu; Santosh: shrihari@umd.edu.
We thank Mikhail Chernov, John Cochrane, Julien Cujean, Robert Dittmar, Bryan Kelly, Ralph Koijen, Mark Loewenstein, Tyler Muir, Stefan Nagel, Nikolai Roussanov, Avanidhar Subrahmanyan, Michael Weber and seminar participants at Maryland and Michigan for helpful comments and suggestions.

1 Introduction

A central observation in asset-pricing research is that aggregate asset returns are predictable over time, in particular by current aggregate valuations (e.g., Shiller 1981, Fama and French 1988, Fama and Bliss 1987). For instance, when the price-dividend ratio of the market portfolio is large, subsequent stock returns are lower than when it is small. Going further, the cross-section of valuations contains additional information about aggregate returns (e.g., Cochrane and Piazzesi 2005, Polk et al. 2006, Kelly and Pruitt 2013). In this paper we study the time series properties of cross-sections of returns. Across multiple asset classes, we find that relative returns exhibit substantial predictability, more so than aggregate returns. Furthermore, we find that expected aggregate and expected relative returns are only weakly correlated. These results highlight both the presence and the structure of multiple important components of risk premia, thereby challenging popular models of time-varying expected returns which feature a single source of predictability.[1]

For each asset class, we reduce the cross-section of assets to a few portfolios along dimensions suggested by economic analysis, statistics, or both. We then study the predictability of these portfolios following standard methods for aggregate returns. We apply this approach to Treasury bonds, stocks and exchange rates. For bonds, a number of common predictors put forward in previous literature forecast the average level of returns with R^2's around 20%. We show that a portfolio trading long against short maturities exhibits a similar degree of predictability, predictability that would be missed by looking at the standard individual asset-level regressions. For stocks we also find strong predictability beyond the market return. Long-short portfolios of so-called anomalies exhibit common variation that we extract using principal components (PCs). These components are also predictable, more so than the market. Finally, a currency "carry" return strategy long high- and short low-interest-rate countries is more predictable than a strategy of trading all currencies against the dollar. Across all of these asset classes, predictability of relative returns is also more robust than of aggregate returns, with stark differences in out-of-sample R^2's. Table 1 summarizes these results.

The starting point of our analysis is to reduce the dimensionality of the cross-section. To understand the importance of doing so, consider first running individual predictive regressions,

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

[1] See for instance Campbell and Kyle (1993), Campbell and Cochrane (1999), Bansal and Yaron (2004), Barberis et al. (2015).

Table 1: In-sample and out-of-sample predictive R^2 across asset classes

We summarize IS and OOS R^2 of predicting aggregate and relative returns for bonds, equities, and FX rates. Specification details are in Table 7 (bonds), Table 11 (stocks), and Table 15 (FX).

	Bonds		Stocks		FX	
	IS	OOS	IS	OOS	IS	OOS
Aggregate returns (index)	23%	-198%	9%	-5%	4%	-6%
Relative returns	37%	18%	42%	42%	19%	5%

The broad question we wish to answer is whether X_t predicts returns; do risk premia vary with X_t? If we impose no structure on the problem, we can perform a standard Wald test of joint significance of the b_i's. A significant Wald test identifies that there exists a linear combination of returns which is predictable; that is, X_t predicts "something". The problem is this test might capture predictability arising from smaller components of the data — which we illustrate by simulation analysis. These components are likely unstable and economically uninteresting. The other extreme is to impose the restriction that X_t predicts "everything", or that all coefficients b_i have a similar pattern and are statistically distinct from 0. Such a test is too stringent since it is likely to uncover aggregate patterns, but tends to ignore predictability of relative returns. We show this in two steps. First, we derive a bound on how aggregate predictability translates into individual asset predictability. Second, we study analytically and by simulation the power to detect relative predictability.

Our approach fits between the extremes of "something" and "everything" by judging whether X_t predicts "something important". We operationalize the notion of importance by choosing specific linear combinations of returns, implemented as portfolios. When possible, these portfolios are simply economically motivated, for instance sorted on maturity for bonds. In the absence of clear guidance, we favor the use of principal components. Beyond preventing overfitting, our methodology allows us to naturally adapt the large body of work developed to study predictability of aggregate returns to relative returns.

Given their well-known strong and stable factor structure, Treasury bond returns constitute a natural empirical setting for our approach. Cochrane and Piazzesi (2005) find that the same combination of yields predicts bond returns at all maturities; that is, b_i is approximately proportional across maturities. Given the strong factor structure of returns, this joint predictive success is actually equivalent to univariate success in predicting the level factor of returns. To explore predictability of relative returns, we consider a maturity sorted portfolio, which is approximately the mimicking portfolio for changes in the slope of the yield curve. We find that yields forecast returns on this slope portfolio with similar explanatory power to level, but this predictability is only evident upon direct examination of slope. We further illustrate how individual tests can fail to capture patterns of predictability by considering a number of business cycle indicators, including the Chicago Fed National Activity Index (CFNAI).[2] Whereas CFNAI doesn't significantly predict any individual bond return, it predicts the second principal component of returns with a similar magnitude as bond yields. Further, the out-of-sample R^2 is negative for predicting the level return, whereas for the slope return it is only moderately lower than in full sample. This difference highlights the greater stability of predicting relative returns. Also, even though the unconditional Sharpe ratio on the slope portfolio is not significantly different from zero, a managed portfolio with weights proportional to expected returns earns an annualized Sharpe ratio of about 0.7, both in and out-of-sample. This result shows that focusing first on unconditional differences in expected returns can hide some important sources of variation in expected returns in the cross-section.

We next apply our method to stocks, an asset class with weaker factor structure and where "important" relative returns are less obvious. Rather than forecasting individual stock returns, we consider 50 well-known anomaly long-short portfolios, which we orthogonalize with respect to the aggregate market return. The residuals exhibit a moderate factor structure, with two components explaining close to 40% of the remaining variation. Again, we find evidence that these additional dimensions —relative returns— are also highly and robustly predictable. Using the cross-section of book-to-market ratios as predictors, we find an out-of-sample R^2 of 42% for relative returns when predicting the first component of long-short anomalies. In contrast this statistic is negative for the aggregate market. This result mimics the unreliability of aggregate predictability in bonds. In contrast, relative returns in both asset classes show robust out-of-sample forecastability. Strong predictability of large common components of stock returns also imply substantial levels of predictability of individual equity anomalies, stemming from their loading on these highly predictable PCs. We show that most anomalies are indeed robustly predictable in the time-series when using our restricted estimation method. On the contrary, naive methods which ignore the factor structure typically suffer from overwhelming spurious in-sample predictability.

Finally, we consider the returns of currency strategies. Following the literature on the carry trade, we use interest rate differentials as predictors (Lustig et al., 2014). We find that

[2] Joslin et al. (2014) include CFNAI as an "unspanned" factor in an affine term structure model.

an index of all currencies against the dollar is less predictable than a long-short portfolio of currencies sorted by interest rate differential. Specifically, whereas the out-of-sample R^2 is negative when predicting the aggregate dollar carry portfolio, it is reliably around 5% for the relative carry returns of high against low interest rate differential currencies.

Taken together, these results illustrate that our approach provides a simple organizing framework to establish stylized facts on predictability for multiple assets. In particular we highlight that relative returns are systematically more predictable than aggregate returns. Further the dynamics of expected relative returns often appear only weakly related to those of aggregate expected returns, suggesting that there are multiple sources of time-varying expected returns within each asset class we study. Using a test inspired by Lewellen and Nagel (2006) we formally reject the hypothesis that predictability of relative returns is due to time-varying exposure to the aggregate index factor, implying there are multiple fundamental sources of variation in risk premia. This result poses a challenge to models such as Campbell and Kyle (1993), Campbell and Cochrane (1999), Bansal and Yaron (2004), Barberis et al. (2015) which attribute all time-series predictability to a single source, i.e. habits, stochastic volatility, or sentiment. If one takes the view that fluctuations in expected returns reflect market inefficiencies, the patterns we document are consistent with Samuelson's dictum separating "macro" and "micro" efficiency.[3] Focusing on stocks, the significant and common predictability we find for anomaly portfolios suggests many of them are not spurious artifacts of data-mining. The commonality we uncover in expected returns has practical implications for tactical asset allocation, suggesting that investment strategies which focus on timing broad components are more robust than timing each anomaly individually.

The remainder of the paper proceeds as follows. We next situate our findings in the existing literature. We present our methodology and contrasts it with the main existing approaches in Section 2. In Section 3, we use our framework to study the expected returns of Treasury bonds. Section 4 and Section 5 study stock anomalies and currency portfolios respectively. Section 6 derives ways to quantify the relation between individual predictability and predictability of important components. Section 7 concludes by outlining the implications of our findings for finance theory and practice.

[3]Samuelson argues that inefficiencies in relative prices are less persistent than for aggregate prices; see for instance Samuelson et al. (1998), Jung and Shiller (2005). The faster process of arriving at efficient pricing leads to *greater* short-run predictability of relative returns.

Related literature

This paper builds on the long literature which studies time series predictability of returns, starting from Shiller (1981) and Fama and French (1988) for stocks, or Fama and Bliss (1987) for bonds.[4] While the early evidence is mostly about aggregate returns, our main focus is on understanding predictability of cross-sections of returns. Vuolteenaho (2002) studies the properties of firm-level returns using a VAR and present-value restrictions. While he notices common variation in estimated expected returns, his analysis does not provide a way to characterize these underlying common forces. Cochrane (2011) summarizes this challenge, arguing that one should ask "what is the linear combination of forecasting variables that captures common movement in expected returns across assets". Cochrane and Piazzesi (2005) propose one way to do so, by estimating a single combination of yields which predicts bonds at all maturities. Our evidence suggests that there is interesting common variation in expected returns beyond only one common predictor. Lochstoer and Tetlock (2016) use a bottom-up approach of aggregating firm-level estimates to anomaly portfolios in order to decompose variation in returns into discount rate and cash-flow news. To focus directly on common variation in risk premia, we use a more top-down methodology. We directly measure the predictability of aggregate components and then project it back onto individual assets.

Interestingly, predictability of relative returns sometimes arises in the context of different exercises. For instance, Stambaugh et al. (2012), Akbas et al. (2015) study whether investor sentiment or fund flows, respectively, drive common variation in stock anomaly expected returns. Brooks and Moskowitz (2017) study the predictability of bond returns across countries, summarizing them by factors similar to our important components, rather than considering individual bonds. Finally, predictability arising from a few factors is also sometimes a part of completely specified models of stochastic discount factors. Dynamic term structure models—e.g., Joslin et al. (2014)—or dynamic asset pricing models—e.g., Adrian et al. (2015)—can include and estimate this predictability. Because the focus of our paper is on predictability alone, we favor a more reduced-form approach.

2 Methodology

We start by reviewing two standard approaches to predictability with multiple returns, highlighting their limitations, then move on to our proposed approach. The problem we are

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

interested in is the time-series predictability of a family of N returns, $R_t = \{R_{i,t}\}$, where i is an asset and t is time. That is, we are interested in estimating time-varying risk premia. We take as given a set of candidate predictors, captured by the vector X_t. We focus on the case of linear predictability, summarized by the equation

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

for each return. This linear representation is preserved by rotation into the space of principal components of returns, which offers an alternative revealing point of view. The principal component portfolios are $F_t = Q' R_t$ where Q is the matrix of eigenvectors of the covariance matrix of returns Σ, $\Sigma = Q \Lambda Q'$. Rewriting Equation 1 in terms of PC portfolios gives:

$$F_{i,t+1} = \alpha_i + \beta_i' X_t + e_{i,t+1}. \tag{2}$$

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, which we discuss now. Given the equivalence of Equation 1 and Equation 2, any method of aggregating the b_i implies an aggregation of the β_i.[5]

2.1 Standard Approaches and Limitations

Predict "something" — spurious predictability. We could test whether there exists a linear combination of the coefficients b_i — or equivalently β_i — that is statistically distinct from 0. This corresponds exactly to a standard Wald test. This notion of predictability, just asking if X_t predicts "something", is intuitively too lax. For instance, our conclusion about the predictive value of X_t could be driven by its ability to predict the lowest variance PC portfolios, or only a few assets. We show in Section 6.3 that a small amount of noise in measured returns can lead to significant spurious predictability of the smallest PC portfolios in population. This issue is exacerbated in small samples. Finding this type of predictability is at odds with our goal of finding economically important variation in risk premia.

Predict "everything" — lack of power beyond the first component. The other extreme is to impose the restriction that X_t predicts "everything", or that all coefficients b_i are statistically distinct from zero. For instance Cochrane and Piazzesi (2005) obtain such

[5] The mapping between the two representation is given by: $\alpha_i = q_i' a$, $\beta_i = q_i' b$ and $e_{i,t+1} = q_i' \varepsilon_i$.

a pattern predicting Treasury bond returns of various maturities using the cross-section of yields, concluding to the presence of one common factor in expected returns. While this approach can uncover interesting patterns, it is likely to be too stringent and ignores predictability of relative returns. We show in Section 6.1 that such a test is often equivalent to testing whether X_t predicts the first component of return, that is $\beta_1 = 0$. In other words, finding uniform predictability across all assets simply finds predictability of the "level" factor in returns. In contrast, we show in Section 6.2 that if a predictor is useful for forecasting relative returns, captured by a long-short portfolio, but not for aggregate returns, individual asset predictive regressions are unlikely to uncover such predictability.

2.2 Our Approach: Predict "Something Important"

To strike a balance between those two notions, we focus on particular linear combinations of the coefficients b_i. We form a few portfolios — in practice two or three — of the returns, $R_{p,t} = \sum_i \omega_i R_{i,t}$. We then estimate individually whether each portfolio is predictable by X_t:

$$R_{p,t+1} = \eta_p + \delta_p' X_t + \nu_{p,t+1}.$$

Testing for the significance of δ_p is exactly equivalent to testing significance of $\sum_i \omega_i b_i$. Therefore we restrict our attention to whether X_t predicts "something important". Restricting to a low-dimensional set of portfolios avoids the issue of the Wald test by focusing on the main dimensions of the data. But it also avoids the other extreme of only focusing on the first component of returns, and allows us to study patterns of relative returns.

Of course the choice of portfolio weights ω_i is somewhat arbitrary. A natural choice for these portfolios are long-short strategies formed by sorting returns based on characteristics, such as maturity for bonds, where characteristic selection is driven by economic motivation. Alternatively one can use guidance from statistical analysis, focusing on the largest principal components of the family of returns. In our applications, it turns out that these two approaches are closely related. For instance, for bonds, we consider the average return across maturity and a long-short strategy across maturities, the most natural dimension of heterogeneity in bonds. These two portfolios correspond closely to the first two principal components of bond returns.

Our approach of treating small principal components of returns as unpredictable can also be motivated by recent papers addressing issues related to large cross-sections of returns. Kozak et al. (2017a,b) argue that for typical families of returns, a stochastic discount factor that prices these portfolios should be well approximated by the first few PCs if there are no near-arbitrage opportunities.[6] We show in Appendix A that imposing this restriction on the SDF is equivalent to imposing that small PCs of returns have zero conditional mean; that is, they are unpredictable.

In the remainder of the paper, we implement this simple approach successively to Treasury bond returns, stocks, and exchange rate strategies. We document the predictability of important components of the cross-section. In addition we revisit the discussion of this section in the context of these applications. Finally, in Section 6 we provide formal statistical arguments and show ways to quantify the statistical properties that lead us to favor this methodology.

3 Predicting Bond Returns

In this section we study the predictability of Treasury bond returns. We find substantial predictability of relative bond returns across maturities in addition to the existing evidence on aggregate predictability. We show this relative predictability is masked when forecasting individual bond returns. In contrast, it is substantial and clearly visible when directly forecasting relative returns.

3.1 Data

We obtain yields on zero-coupon Treasury bonds with maturities from 1 to 15 years from Gürkaynak et al. (2006).[7] Following Joslin et al. (2014), our sample is 1985-2015. We calculate log excess returns from these log yields.[8] We then rescale the excess returns of bonds of maturity n, $rx_{t+1}^{(n)}$, by dividing them by $n-1$:

$$rx_{t+1}^{(n)} \equiv -y_{t+1}^{(n-1)} + \frac{n}{n-1}y_t^{(n)} - \frac{1}{n-1}y_t^{(1)}.$$

Table 2 shows summary statistics for rescaled returns. The table shows that our rescaling largely eliminates scale effects across bond returns. Scaled returns have approximately equal

[6] Giglio and Xiu (2017), Kelly et al. (2017) also argue in favor of low-dimensional stochastic discount factors.

[7] We assess the robustness of our main results to using yields constructed as in Fama and Bliss (1987) in Table 19 in the Appendix.

[8] As standard for one-year holding period returns, we use the one-year zero-coupon yield as the risk-free rate. See Cochrane and Piazzesi (2008) for a comprehensive exposition of bond yields, forward rates, and returns.

Table 2: Bond Return Summary Statistics

The table shows mean, standard deviation, skewness and Sharpe ratio for annual zero-coupon Treasury bonds excess returns with the indicated maturities. All bond returns are first normalized to have unit duration.

	3Y	5Y	7Y	9Y	11Y	13Y	15Y
Mean (%)	0.79	0.72	0.66	0.60	0.55	0.51	0.47
Std. Dev. (%)	1.26	1.15	1.07	1.01	0.97	0.94	0.92
Skewness	0.12	0.12	0.17	0.25	0.35	0.47	0.60
Sharpe Ratio	0.62	0.63	0.62	0.59	0.57	0.54	0.51

standard deviation whereas this statistic varies by a factor of 10 for unscaled returns. The transformed excess returns all have modified duration equal to unity. Additionally, our sample exhibits the well-known phenomenon of Sharpe ratios which slightly decline with maturity.

Following Cochrane and Piazzesi (2008), we define log forward rates as

$$f_t^{(n)} \equiv p_t^{(n-1)} - p_t^{(n)},$$

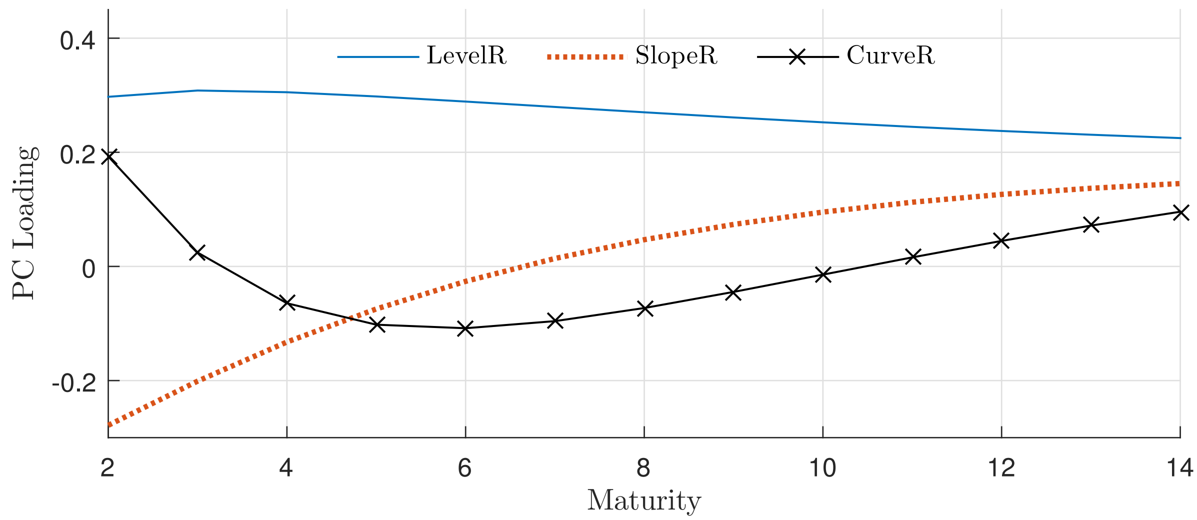
where $p_t^{(n)}$ denotes the time t log price of an n-year bond. Likewise, log forward spreads are

$$s_t^{(n)} \equiv f_t^{(n)} - y_t^{(1)},$$

where $y_t^{(1)}$ is the time t yield on a one-year zero-coupon bond. Given the strong factor structure in forward spreads, we use only the first three principal components of spreads as predictive variables. We denote these by FS1, FS2, and FS3. For ease of comparison we rescale each predictive variable to have 1% standard deviation.

3.2 Principal Components of Returns

Figure 1 plots the first three eigenvectors of the covariance matrix of returns. We obtain "level", "slope", and "curve" factors for returns. The inset table shows that the first two factors, LevelR and SlopeR, capture more than 99% of the variation in realized bond returns. There are essentially only two uncorrelated portfolios that can be formed from bond returns. This suggests we should take advantage of this strong factor structure when forecasting returns.



	LevelR	SlopeR	CurveR
% of variance	90.1	9.3	0.5

Figure 1: Factor Structure in Realized Returns. The top panel plots the first three eigenvectors of realized zero-coupon bond excess returns, termed LevelR, SlopeR, and CurveR. The bottom panel shows the percent of total variance contributed by each factor.

We use these first two PCs of returns as important components. Though these are statistically derived, they are in fact economically interesting. LevelR has 100% correlation with the average return portfolio, formed by equally weighting all maturities. Further, it has -97% correlation with changes in the level of the yield curve and nearly zero correlation with changes in the slope of the yield curve.[9] Hence it is approximately the level-mimicking portfolio, which captures duration risk. Similarly, SlopeR has 94% correlation with a long-short portfolio which has weights that are linear in maturity and that sum to zero.[10] SlopeR has 95% correlation with changes in the slope of yields and nearly zero correlation with changes in the level. Therefore it is the mimicking portfolio for changes in the slope of the yield curve and captures convexity risk.[11]

3.3 Predicting Aggregate and Relative Bond Returns

[9]The level of yields is their first principal component. Slope is the second PC.
[10]The weight of a bond with maturity n is $n - \frac{17}{2}$.
[11]The slope of yields is their second principal component.

Table 3: Predicting Bond Returns with Forward Spreads

We report predictive coefficients and absolute t-statistics (in parentheses) from predictive regressions of zero-coupon bond excess returns on the the first three principal components of lagged forward spreads. Each column presents results for the indicated maturity (in years).

	3Y	5Y	7Y	9Y	11Y	13Y	15Y
FS1	0.06	0.19	0.25	0.29	0.30	0.30	0.29
	(0.34)	(1.21)	(1.87)	(2.32)	(2.57)	(2.67)	(2.66)
FS2	-0.43	-0.40	-0.37	-0.34	-0.32	-0.30	-0.29
	(2.59)	(2.79)	(2.95)	(3.01)	(3.02)	(2.99)	(2.96)
FS3	-0.38	-0.32	-0.26	-0.20	-0.15	-0.10	-0.06
	(2.23)	(2.20)	(2.03)	(1.73)	(1.36)	(0.96)	(0.59)
R^2	0.20	0.21	0.23	0.23	0.23	0.23	0.22

Individual bonds. We start by revisiting the predictability of individual bond returns by the current yield curve, along the lines of Fama and Bliss (1987), Cochrane and Piazzesi (2005), Cieslak and Povala (2015). We forecast scaled excess bond returns with maturities 2 to 15 years using three PCs of forward spreads. Table 3 shows the estimated coefficients, circular block bootstrapped t-statistics (Politis and Romano, 1992), and R^2.[12] We find R^2s around 20% at all maturities, which is substantial and similar to results in other studies.[13] Coefficients on the second PC of forward spreads, FS2, are all significant and are similar across maturities. Coefficients on FS1 and FS3 show a nearly monotone pattern across maturities, but have mixed significance. These patterns suggest there are not fourteen independent left-hand-side variables. There is clearly cross-maturity structure, which can be exploited to improve power and reduce noise by focusing on important components.

PC portfolios. We repeat the above forecasting exercise, except we now directly forecast the first two PCs of returns, LevelR and SlopeR. For ease of comparison, we normalize each portfolio return to have 5% standard deviation in the full sample. Table 4 shows summary statistics for these PC portfolios. The first two columns are for the base, static portfolios. The last two characterize dynamic managed portfolio returns, which we discuss further below. The important takeaway from the table is that LevelR has a large unconditional

[12]We use thirty-six month block length. Results are robust to choice of block size. HAC (Newey-West) t-statistics are generally slightly larger.

[13]See for instance Cochrane and Piazzesi (2005), Cooper and Priestley (2008), Ludvigson and Ng (2009), Cieslak and Povala (2015), Cieslak (2016), Haddad and Sraer (2017).

Table 4: Bond PC Portfolios Summary Statistics

The table shows mean, standard deviation, skewness and Sharpe ratio of returns on Base and Dynamic managed portfolio strategies for the first two principal component portfolios of zero-coupon Treasury bond returns, LevelR and SlopeR.

	Base Portfolios		Dynamic Strategies	
	LevelR	SlopeR	LevelR	SlopeR
Mean (%)	3.11	-0.73	0.06	0.09
Std. Dev. (%)	5.00	5.00	0.16	0.15
Skewness	0.16	-0.00	2.07	1.10
Sharpe Ratio	0.62	-0.15	0.35	0.60

Sharpe ratio, while the Sharpe ratio of SlopeR is nearly zero economically and statistically. Focusing only on unconditional pricing, it would seem that SlopeR is unimportant. Finally, remember that LevelR and SlopeR are uncorrelated by construction.

Table 5 reports the estimated coefficients and R^2 from predictive regressions of these first two principal components on the first three PCs of lagged forward spreads. Measured by R^2, the long-short portfolio, SlopeR, is as predictable as the aggregate portfolio, LevelR. Interestingly, the correlation of their estimated expected returns, which we denote $\mathbb{E}_t\,[\text{LevelR}]$ and $\mathbb{E}_t\,[\text{SlopeR}]$, is only 8%.[14] More than half of the total R^2 for LevelR is generated by FS2, which is irrelevant for predicting SlopeR. FS1 positively predicts both portfolio returns, but with nearly twice the magnitude for SlopeR. Finally, LevelR and SlopeR load on FS3 with equal magnitude, but opposite sign. We conclude that relative returns are as predictable as aggregate returns, and there are at least two independent drivers of time-varying risk premia.

Detecting predictability using individual bonds. In Section 2.1 we argue that inference regarding a variable which predicts SlopeR but not LevelR is difficult using individual bond regressions. As a first empirical analysis of this point, consider forecasting individual bond returns with the estimated forecasts $\mathbb{E}_t\,[\text{LevelR}]$ and $\mathbb{E}_t\,[\text{SlopeR}]$. We know from Table 5 these predictive variables forecast LevelR and SlopeR with equal R^2. Do they forecast individual bond returns? Table 6 shows the estimation results. The first row is the R^2 from

[14]This low correlation could obtain due to a particular pattern of time variation in risk premia combined with time variation in factor loadings, even with a conditional single-factor model. In Appendix D.1 we formally reject such a model with LevelR_t as the single factor.

Table 5: Predicting PC Returns with Forward Spreads

We report predictive coefficients and absolute t-statistics (in parentheses) from predictive regressions of the first two principal components of zero-coupon bond excess returns on the the first three principal components of lagged forward spreads.

	LevelR	SlopeR
FS1	1.08	1.88
	(1.68)	(2.63)
FS2	-1.79	0.22
	(2.97)	(0.34)
FS3	-1.16	1.28
	(1.94)	(1.98)
R^2	0.22	0.22

the unrestricted regressions reported in Table 4. The next two rows show that $\mathbb{E}_t$ [LevelR] predicts all individual returns nearly as well as the unrestricted regressions. All estimated coefficients are statistically significant; this is essentially the finding in Cochrane and Piazzesi (2005). The last block shows the result of including $\mathbb{E}_t$ [SlopeR] as an additional predictor. Since the correlation between $\mathbb{E}_t$ [LevelR] and $\mathbb{E}_t$ [SlopeR] is only 8%, the coefficients on $\mathbb{E}_t$ [LevelR] are nearly unchanged. R^2 values barely improve. Importantly, *none* of the estimated coefficients with respect to $\mathbb{E}_t$ [SlopeR] are significant.[15] Therefore, a researcher who forecasts bond returns equation-by-equation would conclude $\mathbb{E}_t$ [SlopeR] is not an important predictor of excess returns though it significantly predicts the return on an interesting maturity-sorted portfolio.

Revisiting the role of economic activity in bond risk premia. There is a long literature arguing for a relationship between macroeconomic variables and bond returns, with a number of other papers highlighting issues related to these findings.[16] We revisit this issue with a focus on predicting relative, rather than aggregate, returns. Our analysis

[15]This is result could be affected by the fact that we estimate $\mathbb{E}_t$ [SlopeR] in sample. However, the absolute t-stats are biased upward, increasing the probability of rejecting the null. Furthermore, simulation analysis in Section 6.2 shows the same result holds when the researcher has an economically motivated predictor, which happens to be $\mathbb{E}_t$ [SlopeR].

[16]See Estrella and Mishkin (1997), Evans and Marshall (2001, 1998), Ang and Piazzesi (2003), Cooper and Priestley (2008), Ludvigson and Ng (2009), Bikbov and Chernov (2010), Joslin et al. (2014), Cieslak and Povala (2015) among others. Issues are pointed out for instance in Duffee (2013), Bauer and Rudebusch (2016).

Table 6: Predicting Bond Returns with Expected PC Return

We report predictive coefficients and absolute t-statistics (in parentheses) from predictive regressions of bond PC returns on various predictors. The first row reports the unrestricted R^2 from Table 3 using three principal components of forward spreads as forecasting variables. The next block shows the results using $\mathbb{E}_t$ [LevelR], the expected return on LevelR based on the coefficients reported in Table 5 to forecast individual bonds. The last block gives the results from including $\mathbb{E}_t$ [SlopeR] as an additional predictor.

	3Y	5Y	7Y	9Y	11Y	13Y	15Y
	Unrestricted — Predictors: FS1, FS2, FS3						
R^2	0.20	0.21	0.23	0.23	0.23	0.23	0.22
	Predictors: $\mathbb{E}_t$ [LevelR]						
$\mathbb{E}_t$ [LevelR]	-0.29	-0.29	-0.29	-0.27	-0.25	-0.24	-0.22
	(-2.82)	(-3.43)	(-3.78)	(-3.87)	(-3.79)	(-3.63)	(-3.43)
R^2	0.17	0.21	0.23	0.23	0.22	0.20	0.18
	Predictors: $\mathbb{E}_t$ [LevelR], $\mathbb{E}_t$ [SlopeR]						
$\mathbb{E}_t$ [LevelR]	-0.30	-0.30	-0.29	-0.28	-0.26	-0.24	-0.22
	(2.85)	(3.34)	(3.71)	(3.86)	(3.83)	(3.69)	(3.50)
$\mathbb{E}_t$ [SlopeR]	-0.58	-0.25	-0.05	0.09	0.18	0.25	0.29
	(1.65)	(0.82)	(0.18)	(0.38)	(0.85)	(1.21)	(1.48)
R^2	0.20	0.21	0.23	0.23	0.23	0.23	0.22

reveals that variables capturing economic growth are important predictors of SlopeR, but not of LevelR, providing an explanation for the fragility of the relationship when using standard reduced-form methods.

Following Joslin et al. (2014) we start by proxying for macroeconomic conditions with GRO, the Chicago Fed National Activity Index. Table 7 shows the estimates from adding GRO as an additional predictive variable.[17] GRO is statistically insignificant and does not improve R^2 when forecasting LevelR. In contrast, GRO dramatically increases the predictive R^2 for SlopeR. Lack of predictive power for LevelR but significant incremental predictability of SlopeR suggests GRO will not be statistically significant in individual bond regressions. Figure 2 shows the estimated coefficient on GRO by bond maturity with ± 2 standard errors. The coefficients show an upward sloping pattern across maturity, as expected based on the

[17]For ease of comparison we rescale GRO to have 1% standard deviation.

Table 7: Predicting PC Returns with Forward Spreads and GRO

We report predictive coefficients and absolute t-statistics (in parentheses) from predictive regressions of the first two principal components of zero-coupon bond excess returns on the the first three principal components of lagged forward spreads and GRO (Chicago Fed National Activity Index).

	LevelR			SlopeR		
FS1	1.08	0.98	-	1.88	2.36	-
	(1.69)	(1.52)		(2.60)	(4.01)	
FS2	-1.79	-1.92	-	0.22	0.81	-
	(2.99)	(3.09)		(0.35)	(1.47)	
FS3	-1.16	-0.94	-	1.28	0.27	-
	(1.96)	(1.48)		(1.99)	(0.45)	
GRO	-	-0.50	-0.55	-	2.33	1.78
		(0.69)	(0.81)		(3.60)	(2.54)
R^2	0.22	0.23	0.01	0.22	0.37	0.13
Wald test p-value	0.00	0.01	0.72	0.03	0.00	0.04

results in Table 7, but none of the coefficients are statistically significant at conventional levels, except for the 2-year bond. This demonstrates again that individual bond forecasting regressions often miss predictability across maturities which is statistically and economically significant.

The predictability of SlopeR can have multiple economic reasons. SlopeR experiences poor performance following a steepening in the yield curve. In periods where such a steepening is likely to be good economic news, the risk premium of SlopeR should be lower as it acts as a hedge. This may be more likely during recessions when growth is low and a steepening of the yield curve indicates increased expectations of future growth and accompanying inflation. Changes in reach-for-yield behavior (see e.g., Hanson and Stein 2015) could also create predictable changes in the relative returns of long- and short-maturity bonds.

To further explore the link between macroeconomic conditions and bond risk premia, we consider a variety of business cycle variables in addition to GRO. Specifically, we use: (i) PCE, annual real change in per capital personal consumption expenditures; (ii) GDP, annual real change in gross domestic product; (iii) IND, annual change in real industrial production; (iv) LNF1, the first PC of 132 measures of economic activity constructed as in Ludvigson and Ng (2009); and (v) CAY, the consumption/wealth ratio constructed as in

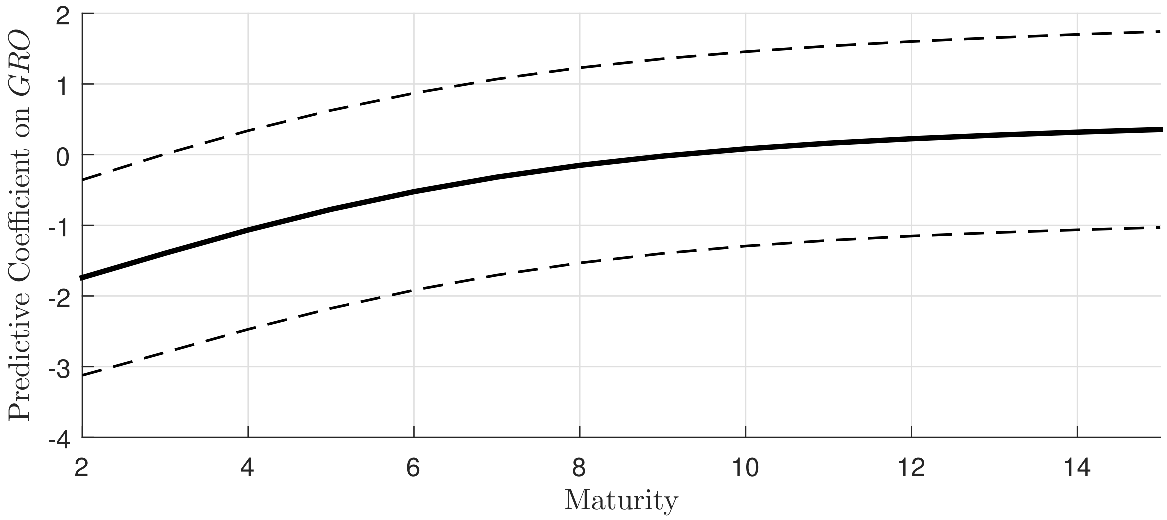


Figure 2: Predicting Bond Returns with Economic Activity. We plot predictive coefficients on the Chicago Fed National Activity Index GRO (and ±2 standard error bands) by maturity from predictive regressions of zero-coupon bond excess returns on GRO and the first three principal components of lagged forward spreads.

Lettau and Ludvigson (2001).[18] Since PCE, GDP, and CAY are only available quarterly, we limit all analysis to that frequency. As above for GRO, we forecast LevelR and SlopeR using the three PCs of forward spreads and each of these measures, one at a time. For ease of comparison, all predictive variables are normalized to have 1% standard deviation. Table 8 shows the estimated coefficients on the business cycle variables as well as in and out-of-sample R^2 statistics. As before, none of the business cycle variables is statistically or economically significant in predicting LevelR. In contrast, all variables besides CAY substantially predict SlopeR, with similar coefficients. Note that LNF1 is essentially a "real" factor that "loads heavily on measure of employment and production ... [but] displays little correlation with prices or financial variables" (Ludvigson and Ng, 2009). Hence, any of the real business cycle measures gives a similar result that SlopeR has low expected returns during downturns. CAY, which is a well-known predictor of equity risk premia, seems to have little power in forecasting either aggregate or relative bond excess returns. In summary, we find that stepping away from individual bond-level predictions and directly analyzing level and slope portfolios disentangles the role of various predictors of the yield curve. This clarification challenges the view that real variables are not useful to forecast returns across the yield curve,

[18]PCE and GDP are from the Bureau of Economic Analysis. IND is from the Federal Reserve Board of Governors. LNF1 and CAY are available at Sydney Ludvigson's website, https://www.sydneyludvigson.com/data-and-appendixes.

Table 8: Predicting PC Returns with Business Cycle Indicators

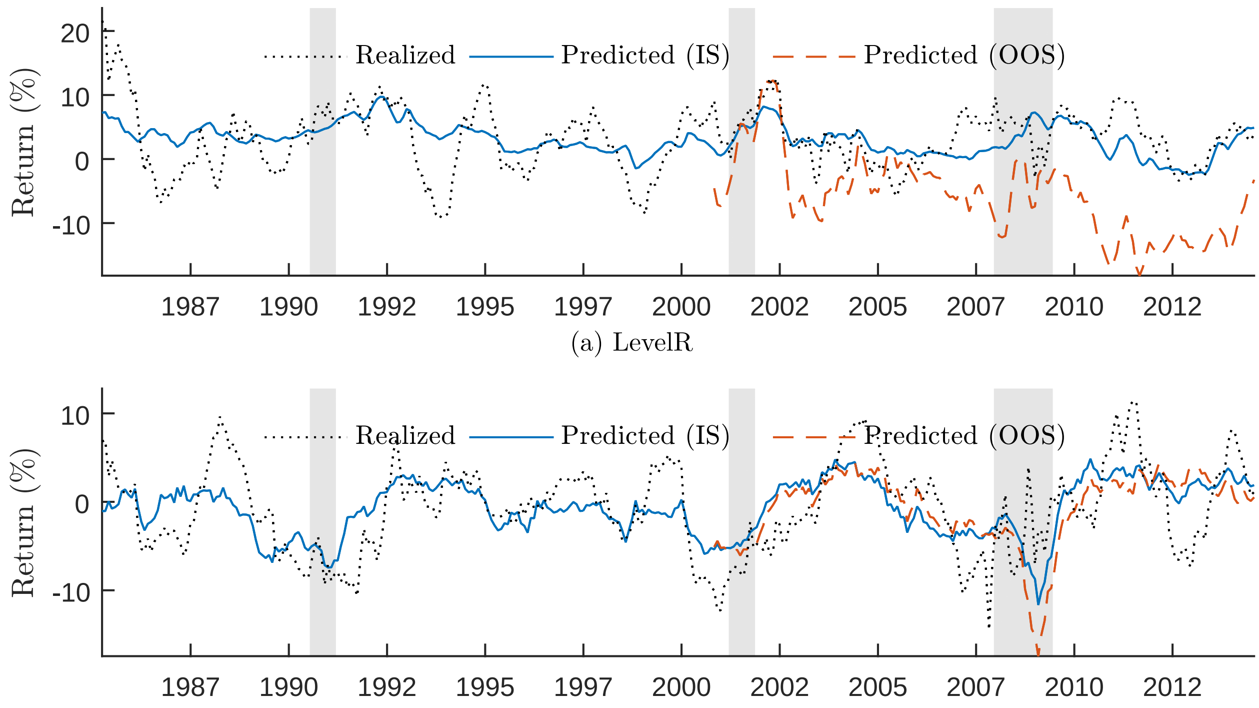
We report predictive coefficients and absolute t-statistics (in parentheses) from predictive regressions of the first two principal components of zero-coupon bond excess returns on the the first three principal components of lagged forward spreads and various business cycle indicators. The data is quarterly from 1985-2015. The indicators are the Chicago Fed National Activity Index (GRO), real per-capita annual growth in personal consumption expenditures (PCE), real per-capita annual growth in gross domestic product (GDP), real annual growth in industrial production (IND), the first PC of a broad set of macro variables (LNF1), and the consumption/wealth ratio (CAY). Reported coefficients are β in the forecasting regression:

$$rx_{i,t+1} = a_i + b_i' \text{FS}_t + \beta_i x_t + \varepsilon_{i,t+1},$$

where x_t is the business cycle indicator and FS_t are the first three PCs of forward spreads. Absolute circular block bootstrap t-statistics are in parentheses. R-square values are computed as before.

	LevelR			SlopeR		
	β	IS R^2	OOS R^2	β	IS R^2	OOS R^2
GRO	-0.45	0.22	-2.43	2.39	0.40	0.15
	(0.59)			(3.56)		
PCE	-0.62	0.22	-2.15	2.78	0.37	0.35
	(0.59)			(2.98)		
GDP	-0.51	0.22	-2.24	3.90	0.47	0.18
	(0.46)			(4.54)		
IND	-0.22	0.22	-2.19	2.19	0.33	-0.06
	(0.25)			(2.68)		
LNF1	0.02	0.21	-2.40	2.64	0.37	0.11
	(0.02)			(3.17)		
CAY	-0.83	0.23	-2.06	-0.30	0.23	0.01
	(1.00)			(0.34)		

by demonstrating these variables explain a particular important component of this cross-section of returns. Interestingly, our explanation for lack of success in predicting individual bond returns does not interact much with those centered around difficulties to draw statistical inference with persistent predictors — e.g. Bauer and Hamilton (2017) — because for SlopeR, both yield-based and macroeconomic predictors appear to evolve mostly at the business cycle frequency.



(a) LevelR

(b) SlopeR

Figure 3: Predicted and Realized Returns. The top and bottom panels shows predicted and realized returns of LevelR and SlopeR, the 1st and 2nd PCs of bond excess returns, respectively. The solid line represents forecasts using parameters estimated with the full sample. The dashed line give forecasts using parameters estimated in the first half (pre-2000) and the dotted line gives realized returns.

3.4 Expected Returns Dynamics

We now turn to the dynamic behavior of our return forecasts and their robustness. Figure 3 plots predicted and realized returns for the first two PCs of bond returns, LevelR and SlopeR, with predictions formed using the estimated coefficients from Table 7. We perform an out-of-sample (OOS) exercise to assess the robustness and usefulness of return predictability, as suggested for example by Welch and Goyal (2008). Specifically, we split the sample into two equal halves, estimate regression coefficients in the first period and use these to forecast returns in the second period. The red-dashed lines indicate the OOS forecasts. For LevelR, the OOS forecasts are quite different from the full-sample predicted returns, suggestive of spurious predictability or substantial parameter uncertainty. R^2 during the OOS period drops from 24% using full-sample estimated parameters to -198% OOS.[19] For

[19] We define R^2 as $1 - \frac{\mathrm{var}(\varepsilon)}{\mathrm{var}(r)}$, both in and out-of-sample.

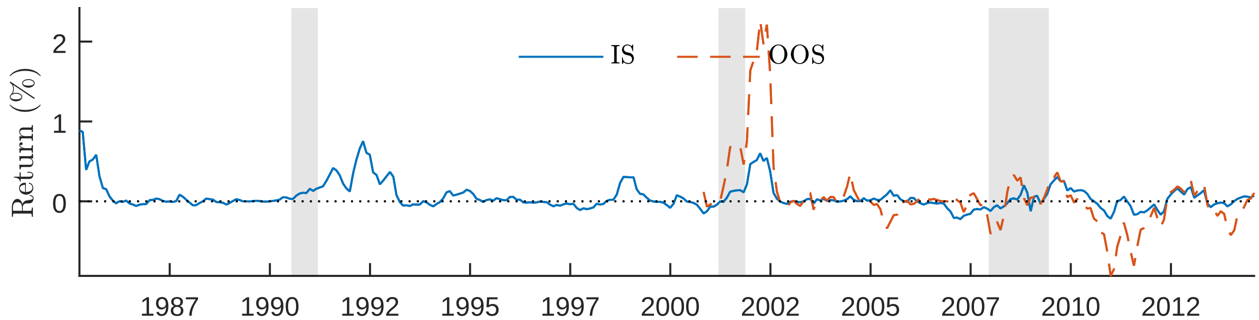
SlopeR, in contrast, the OOS forecasts are nearly unchanged relative to the full-sample. Further, the R^2 declines from 38% to 18% with the reduction mostly due to an extreme outlier during the financial crisis when GRO declined dramatically. Indeed, the PCE variable from Table 8—which is highly correlated with GRO—experienced a milder decline and thus shows similar levels of predictive R^2 in and out of sample. These findings suggest that predictability of SlopeR is exploitable by investors.

We assess these benefits by comparing two strategies: (1) a static strategy which invests a constant fraction of wealth proportional to an asset's unconditional mean and (2) a dynamic strategy which invests proportional to the conditional mean. For full sample results, we construct conditional means using parameter estimates from Table 7. For OOS, we use parameters estimated using the first half of the data. For ease of comparison, we plot the difference between the returns on the static and dynamic strategies. This is equivalent to the return on a strategy with zero weight on average, but which is sometimes long and other times short the asset. Hence, it measures the relative gains and losses from using conditioning information.

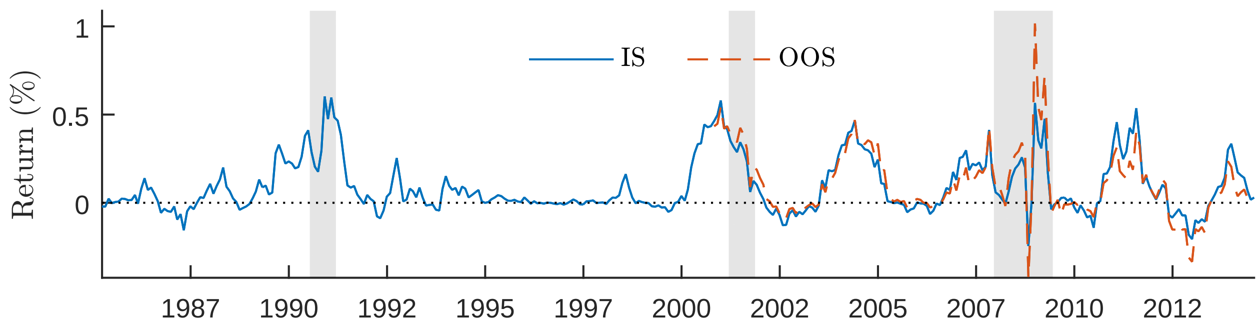
Figure 4 shows realized returns for the managed portfolios with NBER recessions indicated by gray bars. The full-sample Sharpe ratio for LevelR is 0.35 compared to 0.60 for SlopeR. Recall from Table 4 that SlopeR has nearly zero *unconditional* Sharpe ratio, whereas it is 0.6 for LevelR. This result is reversed for *conditional* Sharpe ratios. For LevelR it is quite small whereas the Sharpe ratio of the SlopeR managed portfolio is substantial. Furthermore, dynamic SlopeR tends to have high returns just before and during recessions, whereas dynamic LevelR profits mainly during economic post-recession recoveries. The cyclicality of SlopeR returns echoes our potential theoretical justifications.

Out-of-sample Sharpe ratios mimic the findings for R^2; during the OOS period, the Sharpe ratio falls from 0.20 to 0.12 for LevelR.[20] For SlopeR, in contrast, it barely declines from 0.67 to 0.59 indicating predictability which is valuable in real-time. Beyond Sharpe ratios, we can examine the distribution of managed portfolio returns. Figure 5 shows kernel density estimates for LevelR and SlopeR, both for full- and out-of-sample. In full-sample, both exhibit significant right-skewness, which is to be expected. Out of sample, however, LevelR is nearly symmetric whereas SlopeR remains highly positively skewed, which is desirable to investors.

[20]For OOS evaluation, we normalize the static and dynamic strategies so the difference between their weights has zero mean.



(a) LevelR

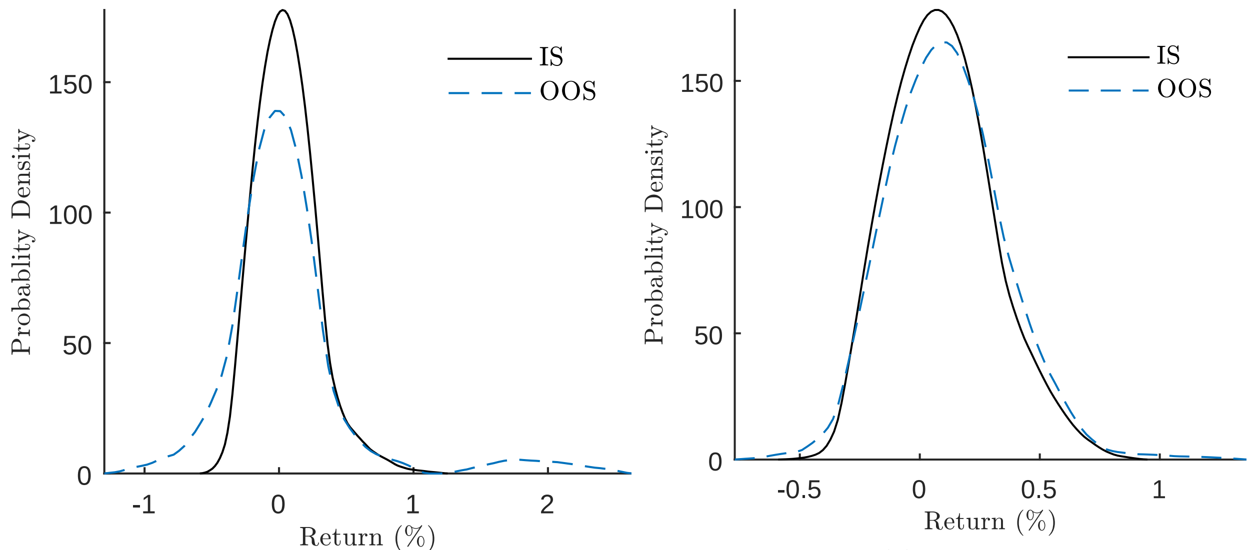


(b) SlopeR

Figure 4: Dynamic Strategy Returns. The figure reports realized returns on dynamic relative to static strategies of LevelR and SlopeR, the 1st and 2nd PCs of bond excess returns.

4 Predicting Stock Returns

We now turn our attention to equity returns. Stock returns have a weaker factor structure than bonds, with more PCs having meaningful variance and therefore potentially implying a richer structure of the time-series of expected returns. We proceed with predicting two PCs of long-short equity anomaly strategies and show that their predictability is stronger and more robust than the predictability of the aggregate market return. Strong predictability of these important components of stock returns also leads to substantial implied predictability of individual equity anomalies, which stems from their loading on these components. We show that most anomalies are indeed highly predictable in the time-series when using our restricted estimation method, even though unrestricted naïve methods typically suffer from overwhelming spurious in-sample predictability.



(a) LevelR (b) SlopeR

Figure 5: Dynamic Strategy Return Distribution. The figure shows kernel densities of realized returns on dynamic (relative to static) strategies of LevelR and SlopeR, 1st and 2nd PCs of bond excess returns.

4.1 Data

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. (2017b) and listed in Table 20 in the Appendix. Their dataset is primarily based on anomaly definitions in Novy-Marx and Velikov (2014), Kogan and Tian (2015), and McLean and Pontiff (2016). Portfolios include all NYSE, AMEX, and NASDAQ firms; however, the breakpoints use only NYSE firms as in Fama and French (2016). The sample is monthly from November 1973 to December 2015.

We construct 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 as the difference in log book-to-market ratios of portfolio 10 and portfolio 1.[21]

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

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

Table 9: Percentage of variance explained by anomaly PCs

Percentage of variance explained by each PC of pooled anomaly portfolio returns. The top panel shows PC1-PC10 of both long and short ends of anomalies (100 portfolios). The bottom panel focuses on 50 long-short market-neutral strategies.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
	PCs of pooled P1, P10 portfolios									
% var. explained	80.8	3.3	3.0	2.1	0.9	0.7	0.6	0.5	0.5	0.4
Cumulative	80.8	84.2	87.2	89.3	90.1	90.8	91.4	91.9	92.4	92.8
	PCs of Long-Short strategies									
% var. explained	19.6	17.5	10.9	5.6	4.2	3.7	3.3	3.2	2.1	2.1
Cumulative	19.6	37.1	48.0	53.6	57.8	61.4	64.8	68.0	70.1	72.1

and can be verified in Table 20 in the Appendix. In this time period, 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 the literature.

4.2 Portfolios and Predictors

PC portfolios. We are interested in the joint predictability of anomaly portfolio returns. Based on the logic of Section 2 we construct PCs from the 50 anomaly long-short (decile 10 - decile 1) portfolios and study their predictability. Table 9 shows that anomaly portfolio returns exhibit a relatively weaker factor structure than Treasury bonds. In the top panel we pool all long and short ends of each strategy (portfolios 1 and 10, that is, 100 portfolios in total). The first PC in the top panel, thus, roughly corresponds to the aggregate market. We see that it accounts for about 80% of the total return variation. The second and third principal components account for a much smaller but similar percentage of total variance.

The bottom panel focuses on our primary dataset — 50 long-short anomaly strategies. To focus purely on relative returns, we first orthogonalize each anomaly strategy with respect to the market portfolio by subtracting its market beta times the return on the market each period.[22] The factor structure of the market-neutral anomalies is weaker than for bond returns: the first three PCs account for about 50% of the total variation. There is, therefore,

[22]We estimate the market beta using the full sample.

more potential for predicting higher-order PCs than there is with bond returns. In our analysis we focus on predicting the market and the first two PCs of long-short strategies.

In Figure 11 in the Appendix, we explore the eigenvector loadings of these PCs. The loadings have a natural interpretation. We can broadly view the PC1 portfolio as long half of the anomalies and short the other half. PC2 is essentially long most of the anomalies besides momentum and profitability-like strategies. Notably, the two PCs span 82% of the return variation of the average anomaly strategy that equal-weights all anomalies. PC2 alone is responsible for 77% of that variation, while PC1 explains an additional 5%.

Predictors. To predict these important components we also need to focus on a low-dimensional set of predictors. We follow the large literature using valuation ratios to predict stock returns.[23] We construct an aggregate BE/ME ratio as well as two restricted linear combinations of log BE/ME ratios of anomaly returns to construct our two cross-sectional predictors. We combine the individual anomaly BE/ME ratios with weights matching their loading on the two portfolios we predict. For example, if PC1 of long-short anomalies is long all strategies with equal weights, our predictor becomes an average of all anomalies' log BE/ME ratios. Recall that an anomaly's log BE/ME ratio is the difference of log BE/ME of its portfolios 10 and 1. The predictor is then simply the sum of all log BE/ME of portfolios 10 — long ends of each strategy — minus the sum of all BE/ME ratios of portfolios 1 — short ends of each strategy.

Formally, consider the eigenvalue decomposition of anomaly excess returns, $\operatorname{cov}(R) = Q\Lambda Q^{'}$, where Q is a matrix of eigenvectors and Λ is a diagonal matrix of eigenvalues. Let B_t be a vector of log BE/ME values of long-short anomaly strategies. We form the two forecasting variables as $bm_{1,t} \equiv q_1^{'} B_t$ and $bm_{2,t} \equiv q_2^{'} B_t$, where the eigenvectors q_1 and q_2 are the first and second columns of Q, respectively. Finally, our last candidate predictor is the log of the aggregate book-to-market ratio, which we denote by $\overline{bm}_t$.

The advantage of such an approach is that we significantly reduce the dimensionality of the space of predictors while imposing intuitive restrictions on how to combine many valuation ratios. Using eigenvectors of strategy returns rather than eigenvectors of BE/ME ratios in constructing our predictors is more robust due to the high persistence of BE/ME ratios in our sample. There is no such distinction for bonds, for instance, because their dominant yield-based and return-based eigenvectors effectively coincide.

[23] See for instance Campbell and Shiller (1988), Fama and French (1988), Cohen et al. (2003), Binsbergen et al. (2010), Kelly and Pruitt (2013).

Table 10: Summary Statistics of Stock PC portfolios

The table shows mean, standard deviation, skewness and Sharpe ratio of returns on Base and Dynamic managed portfolio strategies for the market and first two PCs of anomalies.

	Base Portfolios			Dynamic Strategies		
	MKT	PC1	PC2	MKT	PC1	PC2
Mean (%)	1.89	1.52	2.94	0.02	0.11	0.06
Std. Dev. (%)	5.00	5.00	5.00	0.10	0.21	0.20
Skewness	-0.95	-0.14	0.30	2.98	3.17	3.72
Sharpe Ratio	0.38	0.30	0.59	0.24	0.51	0.31

4.3 Predicting Aggregate and Relative Stock Returns

PC portfolios. We now analyze the predictability of anomaly PC portfolios and the aggregate market. Table 10 shows summary statistics for these variables. For ease of comparison, we normalize each portfolio return to have 5% standard deviation in the full sample. Unlike for bonds, both PC1 and PC2 of anomaly strategies exhibit high unconditional Sharpe ratios in our sample, comparable to the Sharpe ratio of the aggregate market return.

Table 11 shows the results of the three predictive regressions. Column 1 shows that the market is slightly predictable in sample, with an R^2 of 9%, consistent with prior evidence (e.g., see Cochrane, 2008, 2011). The two restricted linear combinations of BE/ME ratios are insignificant in predicting the aggregate market. These variables, however, are highly significant in forecasting PC1 and PC2 of anomalies, respectively. The PC1-restricted linear combination of log BE/ME ratios, bm_1, forecasts PC1 with a t-statistic of 6.18 and contributes much of the total R^2 of 42%. Similarly, the PC2-restricted linear combination of BE/ME ratios, bm_2, has a t-statistic of 3.48 in forecasting PC2 of long-short anomaly returns and obtains a R^2 of 25%. The Wald test of joint significance of the three predictors (last row of the table) rejects the null of no predictability when predicting returns on PC1 and PC2, but not the aggregate market. Based on this evidence we conclude that PC1 and PC2 of long-short anomaly portfolio returns are highly forecastable, much more so than the aggregate market.

The correlation of the estimated expected returns on PC1 and PC2 with the aggregate market are 0.48 and 0.39, respectively. This suggests that more than 75% of variance of expected returns of either of the two PCs is unexplained by market expected returns. Expected returns on two PCs are nearly uncorrelated with each other; their correlation is

Table 11: Predicting PC returns of all anomalies with BE/ME ratios

We report predictive coefficients and absolute t-statistics (in parentheses) from predictive regressions of excess market returns and two PCs of long-short anomaly returns on three predictors: (i) log of the aggregate BE/ME ($\overline{bm}$), (ii) a restricted linear combination of anomalies' log BE/ME ratios with weights given by the first eigenvector of pooled long-short strategy returns (bm_1); and (iii) a restricted linear combination of anomalies' log BE/ME ratios with weights given by the second eigenvector of pooled long-short strategy returns (bm_2). The last two rows show regression R^2 and p-value of the Wald test of joint significance of all regression coefficients. Circular block bootstrapped standard errors in parentheses.

	MKT	PC1	PC2
$\overline{bm}$	0.024	-0.026	0.032
	(1.36)	(1.92)	(1.99)
bm_1	-0.006	0.036	0.002
	(0.80)	(6.18)	(0.23)
bm_2	-0.012	0.010	0.026
	(1.65)	(1.51)	(3.48)
R^2	0.092	0.423	0.246
Wald test p-value	0.246	0.000	0.004

0.04. Overall, this evidence indicates that there are multiple sources of time-varying expected returns in equities.

Individual anomalies. We now study the implications of these results for the predictability of individual anomalies. A naïve approach would be to forecast each anomaly individually using our three predictors from Table 11. Throughout this paper we argued, however, that such an approach is sub-optimal in light of the issues of spurious predictability of small economically irrelevant components of returns and low power in detecting predictability of important but not dominant sources of return variation.

Columns 1 and 3 of Table 12 illustrate these issues. We show the R^2 of predicting each anomaly individually with the three predictors. Column 1 focuses on the in-sample predictability. Not surprisingly, the in-sample R^2 reported in this column are high, but deteriorate drastically in the out-of-sample test shown in column 3, consistent with our argument above. The results demonstrate clear evidence of spurious predictability which does not generalize well to other samples.

In columns 2 and 4 we show the IS and OOS R^2 of predicting individual anomalies by

Table 12: Part I: Predicting individual anomaly returns: R^2 (%)

Predictive R^2 of individual anomalies returns using three forecasting variables from Table 11 (columns 1 and 3) and implied fitted values based on PC forecasts (columns 2 and 4). Columns 1 and 2 provide estimates in full sample. Columns 3 and 4 show out-of-sample R^2.

	Unrestricted IS	Restricted IS	Unrestricted OOS	Restricted OOS
1. Size	30.7	27.3	-68.4	35.3
2. Value (A)	18.2	17.8	4.8	8.8
3. Gross Profitability	5.3	-1.5	-119.5	-40.1
4. Value-Profitablity	17.6	10.7	13.5	19.7
5. F-score	13.5	5.0	-58.2	5.5
6. Debt Issuance	17.1	15.1	10.5	13.6
7. Share Repurchases	29.9	28.0	-1.0	20.1
8. Net Issuance (A)	34.7	16.0	3.9	14.5
9. Accruals	2.2	0.5	-2.0	-0.2
10. Asset Growth	15.6	12.9	7.3	17.7
11. Asset Turnover	7.2	2.2	-18.0	0.2
12. Gross Margins	9.1	7.3	-69.2	-8.2
13. Dividend/Price	14.9	13.2	-16.1	11.9
14. Earnings/Price	12.7	12.5	-2.7	19.3
15. Cash Flows/Price	13.8	9.9	9.3	-1.5
16. Net Operating Assets	20.6	17.5	-27.9	25.8
17. Investment/Assets	21.8	15.6	-18.2	14.3
18. Investment/Capital	10.5	1.0	7.2	-1.5
19. Investment Growth	20.0	14.0	12.3	13.5
20. Sales Growth	10.7	6.1	10.5	5.7
21. Leverage	25.6	21.3	19.7	25.8
22. Return on Assets (A)	32.4	29.2	39.0	24.2
23. Return on Book Equity (A)	29.4	27.7	29.5	23.1
24. Sales/Price	16.9	15.5	12.7	17.3
25. Growth in LTNOA	7.1	-7.7	-91.6	-2.5
26. Momentum (6m)	21.0	13.6	-26.4	10.6
27. Value-Momentum	3.8	3.0	-27.4	2.9
28. Value-Momentum-Prof.	11.1	-0.7	-42.3	-0.4
29. Short Interest	16.4	8.4	-3.9	15.5
30. Momentum (12m)	16.3	10.2	-40.1	15.1
31. Industry Momentum	11.5	7.3	-75.0	0.6
32. Momentum-Reversals	33.5	6.9	29.3	16.2
33. Long Run Reversals	34.4	29.0	26.0	30.3
34. Value (M)	30.1	29.6	7.9	25.0
35. Net Issuance (M)	15.4	12.1	-42.4	12.7

Table 12: Part II: Predicting individual returns: R^2 (%)

	Unrestricted IS	Restricted IS	Unrestricted OOS	Restricted OOS
36. Earnings Surprises	6.8	-9.6	-180.6	-2.4
37. Return on Book Equity (Q)	23.3	23.0	-13.1	16.7
38. Return on Market Equity	18.0	15.0	-45.1	18.4
39. Return on Assets (Q)	27.9	25.4	9.7	18.0
40. Short-Term Reversals	10.1	6.6	5.1	11.1
41. Idiosyncratic Volatility	40.7	38.2	29.0	38.8
42. Beta Arbitrage	16.3	10.2	15.8	9.9
43. Seasonality	0.9	-2.2	-112.4	-2.1
44. Industry Rel. Reversals	31.1	9.0	8.4	14.3
45. Industry Rel. Rev. (L.V.)	42.5	13.0	19.3	19.9
46. Ind. Mom-Reversals	11.8	-1.8	-37.1	-5.0
47. Composite Issuance	10.5	8.2	-52.2	7.7
48. Price	42.7	38.3	5.0	41.2
49. Age	32.2	25.8	32.8	28.5
50. Share Volume	15.3	11.5	9.7	11.0
Mean R^2	19.2	12.9	-16.3	12.3
Median R^2	16.7	12.3	1.5	14.0
Std. Dev. of R^2	10.7	11.0	44.3	13.7

imposing restrictions that any predictability comes only from forecasting PC1 and PC2. Concretely, we start with the estimated expected returns on PC1 and PC2 from Table 11. Next, we combine these forecasts using the loadings of each anomaly on the principal components. This individual forecast, therefore, is fully restricted; the only additional parameter we estimate for each anomaly is an intercept, to allow for time-invariant heterogeneity in expected returns not captured by the first two PCs. Column 2 confirms that imposing this restriction leads to a mechanically lower in-sample R^2 relative to unrestricted regression forecasts, but nonetheless a large fraction of the predictability for many of the anomalies is preserved.

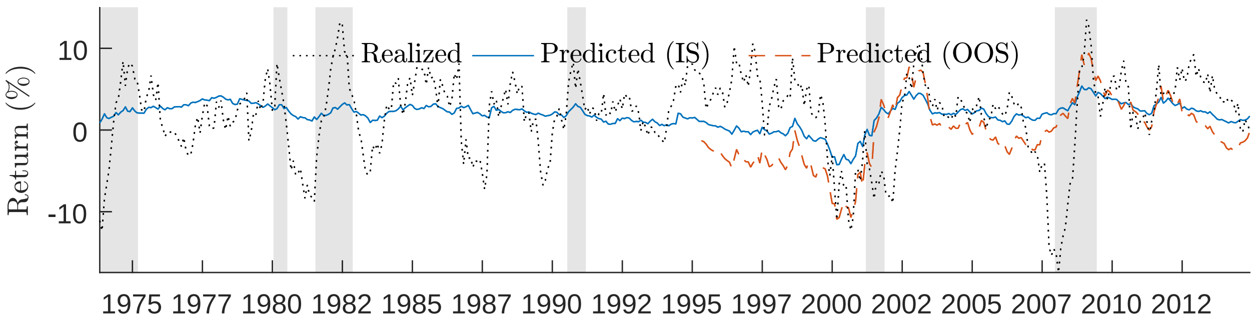
The pattern is vastly different out of sample, shown in the last column. By comparing columns 3 and 4 we see that imposing the restrictions leads to a drastic improvement in OOS predictability. In fact, we are able to predict most of the anomalies, with an average R^2 across all anomalies of 12%, compared to -16% for the unrestricted regressions. We conclude that many individual anomalies are highly predictable out of sample and that our method of restricting predictability to only large PCs helps detect such predictability robustly. On the contrary, if one forecasts each anomaly separately without imposing cross-sectional restrictions, the results are dominated by spurious predictability.

Table 21 in the Appendix decomposes the IS and OOS R^2 of the restricted predictability relation into components that come separately from forecasting PC1 or PC2. Overall, the table illustrates that predictability of PC1 is slightly more important than that of PC2. The average R^2 of predicting individual anomalies by imposing restrictions that any predictability comes only from forecasting PC1 is 7.7%; for PC2 it is 4.5%. PC1 helps explain most of the time-variation in expected returns of momentum, value, reversals, size, ROE/ROA, investment, short interest, and idiosyncratic volatility anomalies. PC2 is responsible for much of the R^2 in forecasting many other anomalies, including beta arbitrage, debt and net issuance, share repurchases, net operating assets, share volume, and some measures of value — dividends/price, earnings/price, sales/price, momentum-reversals.

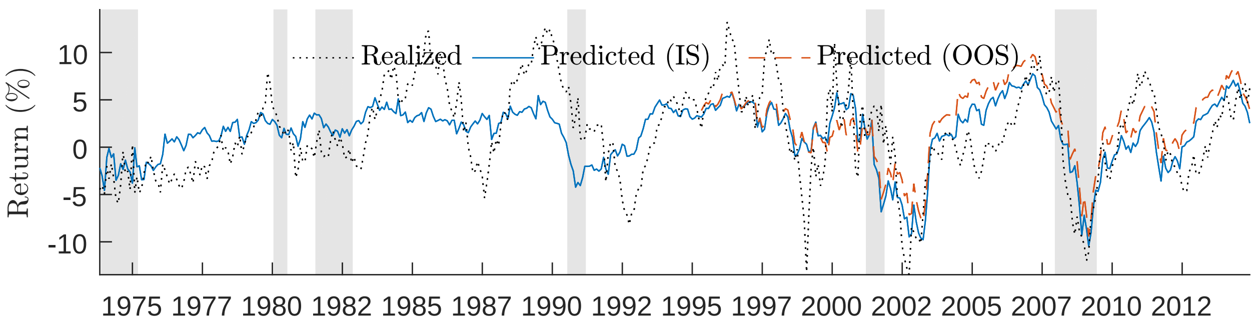
4.4 Expected Returns Dynamics

Figure 6 shows the estimated expected returns on the market and the two PCs of long-short strategies, in and out of sample. These forecasts for the two principal components are remarkably similar in and out of sample. The R^2 confirms this impression with OOS values of 42% and 23%for PC1 and PC2, respectively. The OOS prediction for the market, however, substantially deviates from the IS prediction and results in a poor OOS R^2 of -5%. Based on this evidence we conclude that the aggregate market is not reliably predictable out of sample, while relative returns exhibit substantial and robust time-series predictability.

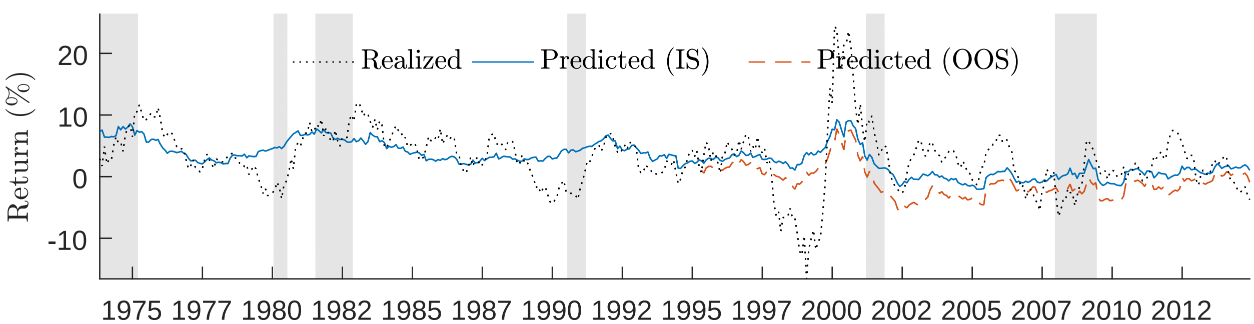
Finally, Figure 7 shows realized returns on dynamic strategies for the market, PC1, and PC2. The full-sample Sharpe ratio for the market is 0.24 compared to Sharpe ratios of 0.51 and 0.31 for PC1 and PC2 respectively. Out-of-sample Sharpe ratio are similar to their in-sample counterparts. The OOS Sharpe ratio of the aggregate market is 0.26. OOS Sharpe ratios of PC1 and PC2 are 0.54 and 0.23, respectively. These numbers suggest there is a substantial amount of time-series variation in expected relative returns, independently of any unconditional premium. An investor can therefore substantially enhance her Sharpe ratio by varying exposures to anomaly strategies over time compared to simply passively holding the anomalies and collecting their unconditional risk premia.



(a) Market



(b) PC1



(c) PC2

Figure 6: Equities Predicted and Realized Returns. The plots show predicted and realized returns of the market (Panel a), PC1 (Panel b), and PC2 (Panel c). The solid line represents forecasts using parameters estimated with the full sample. The dashed line give forecasts using parameters estimated in the first half (pre-2000) and the dotted line gives realized returns.

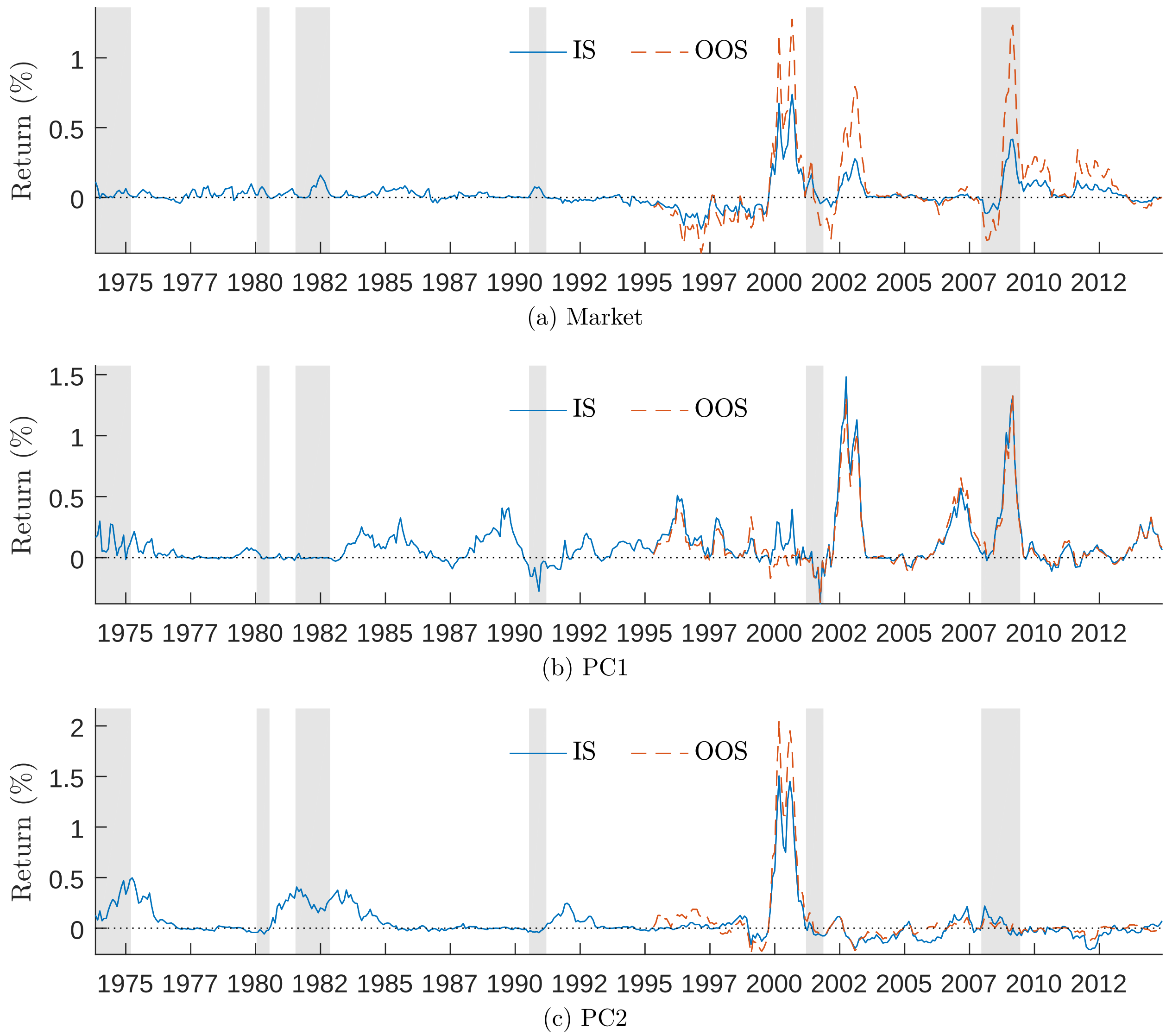


Figure 7: Equities Dynamic Strategy Returns. The figure shows realized returns on dynamic (relative to static) strategies of the market (Panel a), PC1 (Panel b), and PC2 (Panel c).

5 Predicting Foreign Exchange Returns

In this section we study the predictability of foreign exchange (FX) returns. Similarly to previous sections, we find that a relative carry portfolio is more predictable than a basket of all currencies against the dollar.

5.1 Data

We construct a panel of daily foreign exchange returns from Datastream. We compute the annual holding period excess return to investing in each currency j as follows:

$$r_{j,t+1} = f_{j,t} - s_{j,t} - \Delta s_{j,t+1},$$

where $s_{j,t}$ is the spot exchange rate between currency j and the dollar and $f_{j,t}$ is the corresponding forward exchange rate. We also use Eurocurrency Financial Times interest rates and substitute them in place of forward rates according to covered interest rate parity when longer data series are available.[24] Spots, forwards and interest rates come from various datasets: WM/Reuters dataset for spots; WM/Reuters dataset for forwards; BBI/Reuters dataset for developed countries (spots and forwards); Financial Times Eurocurrency for interest rates. We combine all series country by country in order to obtain the longest time series possible. Each country is added to the sample as its data become available. Euro area countries are removed at the dates when the Euro is adopted in each individual country.

We sort the individual currencies into five portfolios based on their forward spread with the US — or equivalently their interest rate differential by covered interest parity — following Lustig et al. (2011). Portfolios are rebalanced daily based on the average of the forward spread in the recent month. We add and drop countries to portfolios as new data becomes available. The sample is from January 1985 until January 2017. In the beginning of the sample we have about three countries per portfolio and about ten per portfolio in the second half of sample.[25] Table 13 shows summary statistics for the sorted portfolios. Similarly to Lustig et al. (2011), the portfolios with higher forward spreads experience higher average

[24] Covered interest parity gives: $f_{j,t} - s_{j,t} = i_{j,t} - i_{\$,t}$, the difference between forward and spot is equal to the difference between foreign and domestic rates.

[25] Our sample contains the following countries: Australia, Austria, Belgium, Bulgaria, Canada, Chile, Colombia, Croatia, Czech Rep., Denmark, EU, Egypt, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, India, Ireland, Israel, Italy, Japan, Jordan, Latvia, Lithuania, Malta, Mexico, Morocco, Netherlands, New Zealand, Norway, Pakistan, Philippines, Poland, Portugal, Romania, Russia, Singapore, Slovakia, Slovenia, South Korea, Spain, Sweden, Switzerland, Taiwan, Thailand, Tunisia, UK.

Table 13: FX portfolios summary statistics

Average forward discounts (row 1), average changes in spot rates (row 2), and excess returns (row 3) for five portfolios (P1–P5) are shown. Hansen-Hodrick standard errors in paratheses. The sample is daily from December 1975 till December 2016.

	P1	P2	P3	P4	P5
Forward spread, $f_t - s_t$	-1.69	-0.12	1.25	2.92	6.42
	(0.35)	(0.32)	(0.30)	(0.31)	(0.53)
Change in spot rates, Δs_{t+1}	-2.10	-1.53	-1.22	-0.23	2.12
	(1.74)	(1.74)	(1.54)	(1.64)	(1.63)
Excess return, $f_t - s_{t+1}$	0.40	1.42	2.47	3.16	4.30
	(1.83)	(1.78)	(1.66)	(1.69)	(1.61)

returns.

5.2 Portfolios and Predictors

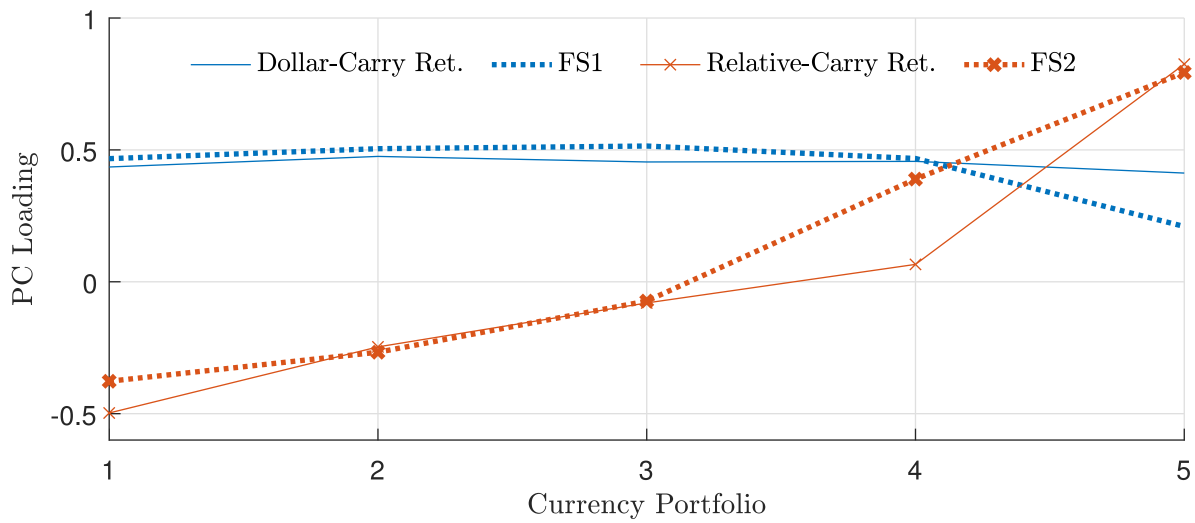
PC portfolios. Figure 8 plots the first two eigenvectors of the covariance matrix of returns.[26] The inset table shows that these first two factors capture more than 90% of the variation in realized currency portfolio returns. Given this strong factor structure, we choose these two PCs as important components, which we denote as Dollar-Carry and Relative-Carry. Dollar-Carry is effectively the return of investing in a basket of all currencies against the U.S. Dollar. Relative-Carry is the return of a portfolio long currencies with high forward spread and short currencies with low forward spread.[27]

Predictors. We use log forward-spreads as forecasting variables. First we average the forward spread across portfolio components at each point in time. This generates five time-series predictors. To reduce the risk of overfitting with relatively short time-series, we use only the first two principal components of forward spreads, which we denote by FS1 and FS2.[28] Figure 8 also shows the first two eigenvectors of forward spreads; they are remarkably similar to the eigenvectors of returns. Lustig et al. (2014) predict Dollar-Carry returns using

[26] As usual, we obtain a "level", "slope", and "curve" type factors from characteristic sorted portfolios.

[27] Lustig et al. (2014) study time-series predictability of the Dollar-Carry factor. Lustig et al. (2011), Lettau et al. (2014) study the Relative-Carry ("Slope") factor in the cross-sectional setting, and Bakshi and Panayotov (2013), Ready et al. (2017) present some evidence of predictability of Relative-Carry.

[28] As with the other asset classes, we first equalize variance across the predictors before performing PCA.



	Dollar-Carry	Relative-Carry	Curve
% of variance	82.5	10.0	3.4

Figure 8: Factor Structure in Realized Returns and Forward Spreads. The top panel plots the first two eigenvectors of realized currency portfolio returns, termed Dollar-Carry, Relative-Carry, and Curve as well as the first two eigenvectors of forward differentials (FS1 and FS2). The bottom panel shows the percent of total variance contributed by each factor.

the average forward differential. As seen from Figure 8, this is essentially equivalent to predicting the first PC of currency returns with FS1.

5.3 Predicting Aggregate and Relative FX Returns

We now analyze the predictability of Dollar-Carry and Relative-Carry portfolios. Table 14 reports their summary statistics. For ease of comparison, we normalize each portfolio return to have 5% standard deviation in the full sample. Both Dollar-Carry and Relative-Carry exhibit high unconditional Sharpe Ratios in our sample, Notably, the Sharpe ratio of the Relative-Carry strategy is more than 50% larger than that of the Dollar-Carry strategy.

Estimation results are shown in Table 15. For Dollar-Carry, the R^2 is small, neither predictive variable is individually significant, and a joint Wald test fails to reject the null of no predictability.[29] For Relative-Carry, in contrast, the R^2 is economically large, both predictors are individually significant, and a joint test convincingly rejects the null. This

[29]Standard errors are estimated using a circular block bootstrap with 36 month block length. Results are robust to varying block length or using HAC (Newey-West) estimation.

Table 14: Summary Statistics of FX PC portfolios

The table shows mean, standard deviation, skewness and Sharpe ratio of returns on Base and Dynamic managed portfolio strategies for the Dollar-Carry and Relative-Carry.

	Base Portfolios		Dynamic Strategies	
	Dollar-Carry	Relative-Carry	Dollar-Carry	Relative-Carry
Mean (%)	1.32	2.21	0.01	0.05
Std. Dev. (%)	5.00	5.00	0.05	0.13
Skewness	-0.05	-0.44	0.37	1.25
Sharpe Ratio	0.26	0.44	0.20	0.38

Table 15: Predicting PC returns of Foreign Exchange rates

Forecasting regression coefficient estimates. We forecast the first two PC portfolios of currencies sorted on forward spreads (Dollar-Carry and Relative-Carry) using the first two PCs of forward spreads (FS1 and FS2). Circular block bootstrapped standard errors in paratheses. The sample is daily from January 1985 until January 2017.

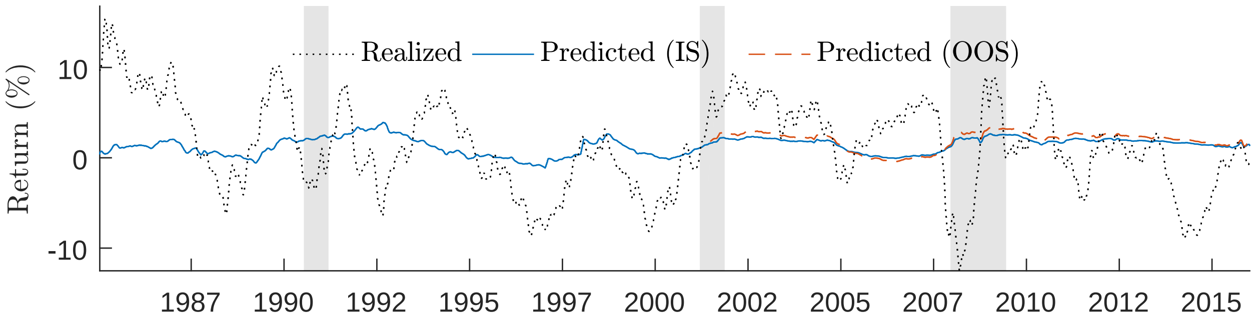
	Dollar-Carry	Relative-Carry
FS1	0.52	-0.89
	(1.23)	(2.76)
FS2	0.07	1.30
	(0.11)	(2.50)
R^2	0.04	0.19
Wald test p-value	0.67	0.00

evidence is in line with our previous findings in other asset classes: relative returns are more predictable in the time series than aggregate returns.

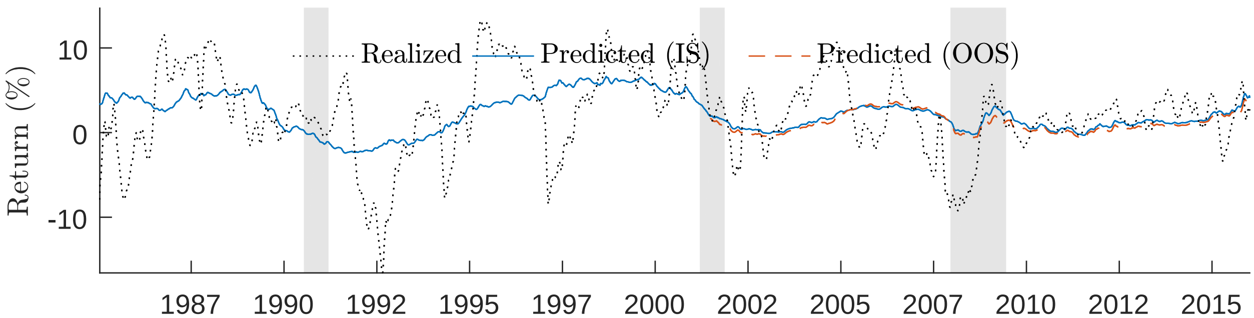
The correlation of expected returns on Dollar-Carry and Relative-Carry is 0.69, suggesting that more than 50% of variance of expected returns of each of the two PCs is unexplained by the other PC. Again, we conclude that there are multiple sources of time-varying expected returns.

5.4 Expected Returns Dynamics

Figure 9 plots predicted and realized returns for both PC portfolios. As before, OOS represents out-of-sample forecasts based on coefficients estimated using the first half of the



(a) Dollar-carry



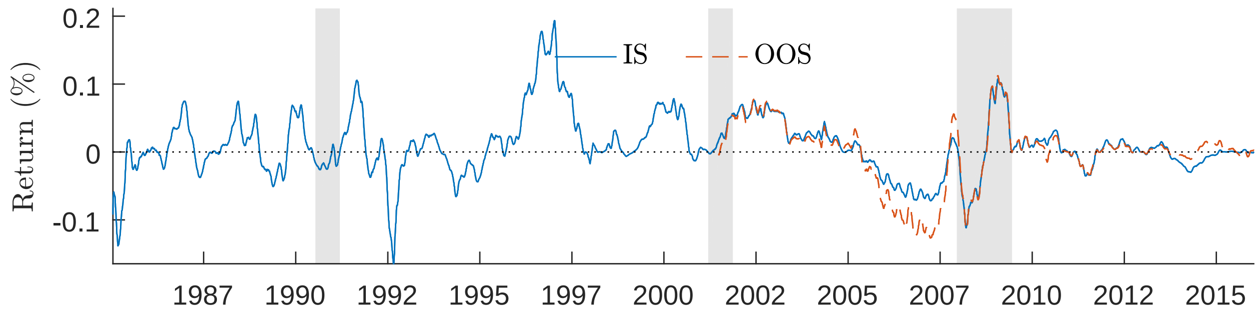
(b) Relative-Carry

Figure 9: FX Expected Returns. The figure shows predictability of the Dollar-Carry and Relative-Carry strategies using forward differentials.

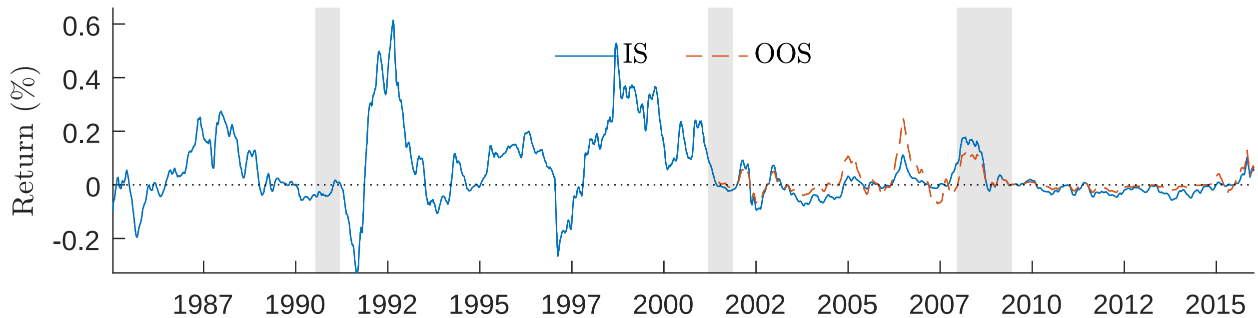
data (1975 - 1999). For both portfolios, there is little evidence of estimation uncertainty, as the OOS forecasts are highly correlated with full sample forecasts. R^2 during the OOS period drops to -6% for Dollar-Carry and to 5% for Relative-Carry. The Relative-Carry premium tends to be low just after recessions, gradually rising to peak one or two years prior to the next recession. Expected returns are significantly less volatile in the OOS period for both portfolios, due to reduced volatility of forward spread differentials.[30] This result is one caution against overemphasis on OOS R^2, since the amount of predictability can be time varying.

Figure 10 shows realized returns on the corresponding dynamic strategies with NBER recessions indicated by gray bars. The full-sample Sharpe ratio for Dollar-Carry is 0.20 compared to 0.38 for Relative-Carry. Relative-Carry managed portfolio are significantly more volatile due to the higher predictive R^2, which leads to more volatile portfolio weights. Lower predictability in the latter half leads to lower volatility of expected returns, and hence less volatility for the managed portfolio returns.

[30] Variances of FS1 and FS2 are lower by a factor of three in the second half of the full sample.



(a) Dollar-Carry



(b) Relative-Carry

Figure 10: FX Dynamic Strategy Returns. The figure shows realized returns on dynamic (relative to static) strategies of the Dollar-Carry and Relative-Carry (1st and 2nd PCs of FX excess returns).

6 Statistical Properties

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. All derivations are in Appendix B.

6.1 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,

$$R_{i,t+1} = \lambda_i F_{t+1} + \varepsilon_{i,t+1}, \tag{3}$$

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

$$F_{t+1} = \beta_1 X_t + u_{t+1}. \tag{4}$$

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

$$R_{i,t+1} = b_i X_t + v_{t+1}. \tag{5}$$

The following proposition characterizes a lower bound on this quantity.

Proposition 1. *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:*

$$R^2_{X,i} \geq \max\left(\sqrt{R^2_{1,i} R^2_X} - \sqrt{\left(1 - R^2_{1,i}\right)\left(1 - R^2_X\right)}, 0\right)^2. \tag{6}$$

Intuitively, if X_t strongly predicts the 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 sizeable 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 $\mathrm{R}^2_{\max}$ on the R-squared of the predictive regression of $\varepsilon_{i,t+1}$ by X_{t+1}.[31] In this case, our bound becomes:

$$\mathrm{R}^2_{X,i} \geq \max\left(\sqrt{\mathrm{R}^2_{1,i}\mathrm{R}^2_X} - \sqrt{\mathrm{R}^2_{\max}\left(1-\mathrm{R}^2_X\right)}, 0\right)^2. \quad (7)$$

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 also assume that the factor receives the same loading for all assets and has the same explanatory power. This corresponds to constant λ_i, and R^2_{1i} across assets. In this case, we obtain that:

$$\mathbb{E}_i\left[\mathrm{R}^2_{X,i}\right] = \mathrm{R}^2_1\mathrm{R}^2_X + \mathrm{var}_i\left(\mathrm{R}^2_{X,i}\right), \quad (8)$$

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

where $\mathbb{E}_i(\cdot)$ and $\text{var}_i(\cdot)$ are the mean and variance in the cross section of individual returns. 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.

In Appendix B.1, we show how to reverse this interpretation, going from the predictability of "everything" to predictability of the common factor. Using the Bhatia-Davis inequality, we show that

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

where $\mathrm{R}_{\max}^2$, as before, is the maximum $\mathrm{R}_{X,i}^2$ from any individual asset forecasting regression. For reasonable values of $\mathrm{R}_{\max}^2$, such as 0.5 or less, the bound implies that ~22% average R^2 we obtain for individual bonds in Table 3 implies at least 18% R_X^2, the R-squared when predicting the aggregate portfolio return.

6.2 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.

I.i.d. predictor. Consider first the case where the forecasting variable X_{t+1} has i.i.d. draws.[32] Suppose that X_t forecasts only one particular principal component j with population R-squared R_X^2 and the remaining principal component returns are i.i.d. Gaussian with known mean.[33] For power analysis, we consider repeated samples of length T.[34]

When directly forecasting the principal component return, $F_{j,t+1}$, the power to correctly

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

[33]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.

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

reject the null with test of nominal size α is

$$\text{power}\left(F_2\right) = G\left(-t_{\alpha/2,T} - z\right) + \left[1 - G\left(t_{\alpha/2,T} - z\right)\right], \tag{9}$$

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

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

$$\text{power}\left(R_i\right) = G\left(-t_{\alpha/2,T} - \zeta\right) + \left(1 - G\left(t_{\alpha/2,T} - \zeta\right)\right), \tag{10}$$

where $\zeta = \sqrt{\mathrm{R}_X^2}\sqrt{T}\left(\left(1 - \mathrm{R}_X^2\right) + \frac{1-\mathrm{R}_{j,i}^2}{\mathrm{R}_{ji}^2}\right)^{-\frac{1}{2}}$.

Because $\zeta \leq z$, we immediately obtain that $\text{power}\left(F_2\right)$ is larger than $\text{power}\left(R_i\right)$ for all assets. Therefore, there is always more information about predictability of the important component by studying it directly.

Persistent predictor. To understand whether these results are a useful approximation for the more general case of a persistent predictor, we turn to simulation. We model X as an $AR\left(1\right)$ process,

$$X_{t+1} = \phi X_t + \nu_{t+1},$$

with the normalization $\text{var}_t\left[\nu_{t+1}\right] = \frac{1}{1-\phi^2}$, so that X has unit unconditional variance. Simulated returns have the same unconditional covariance as in the bond data, but we assume the second principal component is forecastable by X. We simulate 30-year histories and forecast returns of PCs and individual bonds in each simulated sample. We compute the sampling variance of an estimator from the simulated distribution and construct relevant t-statistics.[35]

Table 16 shows the probability of rejecting the null of no predictability under the true distribution. Panels A, B, and C correspond to a persistence of the predictor, ϕ, equal to 0, 0.3 and 0.6, respectively. Each row corresponds to a simulations with the indicated value of R_X^2, the population R-squared obtained when forecasting the second PC of returns with X_t. The first column shows the probability of rejecting the null when forecasting the first principal component of returns. Since it is not predictable by construction, the rejection probability should be 5% for a test with that nominal size. The simulated rejection probabilities are all close to 5%, indicating the t-test has approximately correct rejection probability even

[35]We assume the researcher knows the sampling variance. This circumvents known small-sample issues with HAC variance estimators.

when the predictor is persistent. In fact, the values are very similar across panels: the t-distribution provides a close approximation even when not exact. For individual bonds, we see that power is much lower than for the second principal component for all values of the persistence parameters.[36] These results suggest that the closed form formulas for the i.i.d. case (panel A) constitute a good approximation for settings with persistent predictors.

6.3 Spurious Predictability in Joint Tests

Finally, we consider the issue of spurious predictability likely to be picked up by the Wald test. The Wald test assesses joint significance of the predictive coefficients for individual returns b_i, or equivalently joint significance of the predictive coefficients for all the principal components of the family β_i. Very small components are likely to pick up spurious patterns of predictability, or predictability originating from minor measurement error in prices. This would lead to a rejection of the Wald test, even though there is no economically interesting predictability.

To get a sense of the importance of this issue, we study a simple simulation. We present here an application to the case of Treasury bond returns, but this approach can be adapted to other settings. In the simulation, we introduce a tiny amount of noise in prices, i.e. yields, then predict bond returns. We assume there is no true predictability but observed yields, $\widetilde{y}$, have i.i.d noise with standard deviation of σ_ε:

$$\widetilde{y}_{n,\tau} = y_{n,\tau} + \varepsilon_{n,\tau}.$$

We construct observed returns from observed yields,

$$\begin{aligned} \widetilde{r}_{n,\tau+1} &= -(n-1)\,\widetilde{y}_{n-1,\tau} + n\widetilde{y}_{n,\tau} - \widetilde{y}_{1,\tau} \\ &= r_{n,\tau+1} - (n-1)\,\varepsilon_{n-1,\tau+1} + n\varepsilon_{n,\tau} - \varepsilon_{1,\tau}, \end{aligned} \tag{11}$$

and simulate 30-year histories of "observed" returns based on the sample covariance matrix of realized bond returns and set $\sigma_\varepsilon = 5$bp. For context, annual yield changes have approximately 1% standard deviation. We then consider forecasting returns in the presence of these errors. For each PC portfolio we compute the true — in a simulated sample — size of a nominal 5% Wald test of the null hypothesis of no predictability. Table 17 shows the population R^2 of the predictive regression and probability of rejecting the null for each principal

[36] Across simulations, the power is U-shaped in maturity. This is to be expected for our particular setting since squared loadings on the second principal component are also u-shaped.

Table 16: Power of Predictive Regressions when Only Slope is Predictable

The table gives probability (in %) of rejecting the null hypothesis of no predictability given the alternative hypothesis. ϕ is the annual auto-correlation of the predictive variable. Each row gives results for the indicated theoretical R^2 when forecasting Slope. Level and Slope are the first and second principal components, respectively, of bond returns. The remaining columns are for zero-coupon bonds with the indicated maturity (in years). We compute the probabilities from 100,000 simulations, each of thirty years.

(a) No Persistence, $\phi = 0$

R^2	Level	Slope	3Y	6Y	9Y	12Y	15Y
10%	5	39	9	5	6	8	10
20%	5	74	14	5	7	12	16
30%	5	93	19	5	8	15	21
40%	5	99	25	6	9	18	26

(b) $\phi = 0.3$

R^2	Level	Slope	3Y	6Y	9Y	12Y	15Y
10%	6	40	10	6	7	9	11
20%	6	73	15	6	8	12	16
30%	6	92	20	6	8	15	21
40%	6	99	25	6	9	19	27

(c) $\phi = 0.6$

R^2	Level	Slope	3Y	6Y	9Y	12Y	15Y
10%	7	41	12	7	8	10	11
20%	7	74	16	7	9	14	17
30%	7	91	21	8	10	17	23
40%	7	98	27	8	11	20	29

Table 17: Predicting PC Returns with Noise

The table gives the population R^2 and probability (in %) of rejecting the null of no predictability when forecasting principal components of bond returns with lagged bond yields. The last row reports the size of a Wald test over the first few principal components. We compute these values from 100,000 simulations of 30 years, with yields contaminated by i.i.d. noise with 5bp standard deviation.

	$PC1$	$PC2$	$PC3$	$PC4$
population R^2 (%)	0	4	40	46
5% Wald test size (%)	6	11	70	83
Joint 5% Wald test size (%)	6	10	51	75

component. The first two PC portfolios have R-squared close to zero, and the test size is somewhat higher than the nominal 5%. The second two PCs, on the other hand, have large R-squared and the size of the test is of an order of magnitude larger than 5%. The last row shows the rejection probability for a test that "large" PCs are not predictable, that is a Wald test over the first few principal components. For example, the second column gives the rejection probability for a joint test that PC1 and PC2 are not predictable. The high individual test size for the smaller PCs contaminates this joint Wald test. Indeed, while the size of the test is 10% for the first couple of PCs, it then jumps up and is as high as 75% for the first four PCs. This result exemplifies well the issue with the Wald test: very small, economically meaningless variation in prices tends to get captured in small principal components and generates uninteresting or spurious predictability. Because the Wald test puts them on the same footing as the larger, more interesting sources of variation, it tends to reject the null too often.

7 Concluding Remarks

We have proposed and implemented a systematic approach to study the time-series predictability for cross-sections of assets. Our method relies on estimating the predictability of important components of each family of assets. Across Treasury bonds, stocks, and currencies, a common set of facts emerges, facts that are obscured by standard approaches. First, relative returns—the components of returns beyond the index—are highly predictable, typically more so than the index. Second, this variation in expected relative returns is more robust out of sample than that of expected index returns, making it exploitable by investors in real time. Third, the risk premia of relative returns appear to be only weakly related to aggregate risk premia.

Our findings constitute a novel robust set of facts against which to evaluate theories of asset pricing. In particular, these results highlight the importance of going beyond aggregate predictability to understand common movements in risk premia over time. We outline directions for further study of the implications of our findings. Using a test inspired by Lewellen and Nagel (2006) we confirm the existence of multiple time-varying prices of risk for each asset class we study.[37] This finding presents a challenge to many equilibrium theories of predictability. For instance, the representative agent models in Campbell and Cochrane (1999) and Bansal and Yaron (2004) feature one mechanism driving time-varying in risk premia: habits and stochastic volatility, respectively. Our empirical result highlights the necessity to develop models with allow for multiple mechanisms and hence richer SDF dynamics. Models such as Campbell and Kyle (1993), Barberis et al. (2015) which attribute predictability to some investors' time-varying sentiment are also unable to generate the patterns we find since they do not address the distinction between aggregate sentiment and relative sentiment; a distinction emphasized for instance by Samuelson et al. (1998). Focusing specifically on the cross-section of stock anomaly portfolios, our finding of significant common predictability suggests that at least some anomalies are economically relevant and are not merely statistical artifacts. Practically, our results also imply that factor timing can be improved by jointly predicting anomaly portfolio returns rather than considering each anomaly in isolation.

[37] See Appendix D.1.

References

Adrian, T., R. K. Crump, and E. Moench (2015). Regression-based estimation of dynamic asset pricing models. *Journal of Financial Economics 118*(2), 211–244.

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.

Ang, A. and M. Piazzesi (2003). A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. *Journal of Monetary Economics 50*, 745–787.

Bakshi, G. and G. Panayotov (2013). Predictability of currency carry trades and asset pricing implications. *Journal of Financial Economics 110*(1), 139–163.

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.

Bauer, M. D. and J. D. Hamilton (2017). Robust bond risk premia. Technical report, National Bureau of Economic Research.

Bauer, M. D. and G. D. Rudebusch (2016). Resolving the spanning puzzle in macro-finance term structure models. *Review of Finance*, rfw044.

Bikbov, R. and M. Chernov (2010). No-arbitrage macroeconomic determinants of the yield curve. *Journal of Econometrics 159*(1), 166–182.

Binsbergen, V., H. Jules, and R. S. Koijen (2010). Predictive regressions: A present-value approach. *The Journal of Finance 65*(4), 1439–1471.

Brooks, J. and T. J. Moskowitz (2017). The cross-section of government bond returns.

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.

Cieslak, A. (2016). Short-rate expectations and unexpected returns in treasury bonds.

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

Cochrane, J. H. (2008). The dog that did not bark: A defense of return predictability. *Review of Financial Studies 21*(4), 1533–1575.

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*.

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

Cooper, I. and R. Priestley (2008). Time-varying risk premiums and the output gap. *The Review of Financial Studies 22*(7), 2801–2833.

Duffee, G. R. (2013). *Bond Pricing and the Macroeconomy*, Volume 2, Chapter Vol. 2, Part B, pp. 907–967. Elsevier.

Estrella, A. and F. S. Mishkin (1997). The predictive power of the term structure of interest rates in europe and the united states: Implications for the european central bank. *European economic review 41*(7), 1375–1401.

Evans, C. L. and D. A. Marshall (1998). Monetary policy and the term structure of nominal interest rates: evidence and theory. In *Carnegie-Rochester Conference Series on Public Policy*, Volume 49, pp. 53–111. Elsevier.

Evans, C. L. and D. A. Marshall (2001). Economic determinants of the term structure of nominal interest rates. *manuscript, Federal Reserve Bank of Chicago*.

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 (2016). Dissecting anomalies with a five-factor model. *The Review of Financial Studies 29*(1), 69–103.

Giglio, S. and D. Xiu (2017). Inference on risk premia in the presence of omitted factors. Technical report, National Bureau of Economic Research.

Gürkaynak, R. S., B. Sack, and J. H. Wright (2006, June). The U.S. Treasury Yield Curve: 1961 to the Present.

Haddad, V. and D. Sraer (2017). The banking view of bond risk premia. Technical report.

Hanson, S. G. and J. C. Stein (2015). Monetary policy and long-term real rates. *Journal of Financial Economics 115*(3), 429–448.

Hastie, T. J., R. J. Tibshirani, and J. H. Friedman (2011). *The elements of statistical learning: data mining, inference, and prediction.* Springer.

Joslin, S., M. Priebsch, and K. J. Singleton (2014). Risk Premiums in Dynamic Term Structure Models with Unspanned Macro Risks. *The Journal of Finance 69*(3), 1197–1233.

Jung, J. and R. J. Shiller (2005). Samuelson's dictum and the stock market. *Economic Inquiry 43*(2), 221–228.

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 (2017). Some characteristics are risk exposures, and the rest are irrelevant.

Kogan, L. and M. Tian (2015). Firm characteristics and empirical factor models: a model-mining experiment. Technical report, MIT.

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

Kozak, S., S. Nagel, and S. Santosh (2017a). Interpreting factor models. *The Journal of Finance, forthcoming*.

Kozak, S., S. Nagel, and S. Santosh (2017b). Shrinking the cross-section.

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

Lettau, M., M. Maggiori, and M. Weber (2014). Conditional risk premia in currency markets and other asset classes. *Journal of Financial Economics 114*(2), 197–225.

Lewellen, J. and S. Nagel (2006). The conditional CAPM does not explain asset-pricing anomalies. *Journal of financial economics 82*(2), 289–314.

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

Ludvigson, S. C. and S. Ng (2009). Macro factors in bond risk premia. *The Review of Financial Studies 22*(12), 5027–5067.

Lustig, H., N. Roussanov, and A. Verdelhan (2011). Common risk factors in currency markets. *Review of Financial Studies 24*(11), 3731–3777.

Lustig, H., N. Roussanov, and A. Verdelhan (2014). Countercyclical currency risk premia. *Journal of Financial Economics 111*(3), 527–553.

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

Novy-Marx, R. and M. Velikov (2014). A taxonomy of anomalies and their trading costs. Technical report, National Bureau of Economic Research.

Politis, D. N. and J. P. Romano (1992). A circular block-resampling procedure for stationary sata. *Exploring the Limits of Bootstrap*, 263–270.

Polk, C., S. Thompson, and T. Vuolteenaho (2006). Cross-sectional forecasts of the equity premium. *Journal of Financial Economics 81*(1), 101–141.

Ready, R., N. Roussanov, and C. Ward (2017). Commodity trade and the carry trade: A tale of two countries. *The Journal of Finance*.

Samuelson, P. A. et al. (1998). Summing up on business cycles: opening address. In *Conference Series-Federal Reserve Bank of Boston*, Volume 42, pp. 33–36. Federal Reserve Bank of Boston.

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

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

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

Appendix

A An SDF-based argument for predicting PCs

We provide a formal argument linking time-series predictability regressions of dominant PCs to an SDF-based estimation approach with time-varying risk prices, which achieves regularization by limiting the number of factors in the SDF representation.

Let X_t be a $(P+1)\times 1$ vector of common predictors. The first element of X_t is unity.

Naive predictability approach of all assets. Consider a simple time-series predictability regression for each asset i:

$$R^e_{i,t+1} = a_i' X_t + \varepsilon_{t+1},$$

where a_i is $(P+1)\times 1$. Stack all regressions together to get the following multivariate regression (with no restrictions across a_i's):

$$R^e_{t+1} = a' X_t + \varepsilon_{t+1}, \tag{12}$$

where a' is $N\times(P+1)$ and R^e_{t+1} is $N\times 1$.

$N\times(P+1)$ OLS moment conditions for an estimate of a are:

$$g_T^{\text{TS reg.}}(a) = \mathbb{E}_T\left[\left(R^e_{t+1} - a'X_t\right)\otimes X_t\right] = 0, \tag{13}$$

where $\otimes$ denotes the Kronecker product.

SDF approach. Consider now an SDF that prices all excess returns,

$$M_{t+1} = 1 - b_t'\left(F_{t+1} - \mathbb{E}_t F_{t+1}\right),$$

where F_{t+1} are pricing factors. Kozak et al. (2017a,b) show that for typical sets of portfolios, an SDF that prices those portfolios should be well approximated by the first few PCs. We will therefore assume that F_{t+1} is a $K\times 1$ vector of largest K PCs of returns formed using the eigenvalue decomposition of the conditional covariance matrix of returns, $\Omega = \mathbb{E}_t\left[\left(R^e_{t+1} - \mathbb{E}_t R^e_{t+1}\right)\left(R^e_{t+1} - \mathbb{E}_t R^e_{t+1}\right)'\right] = Q\Lambda Q'$, which we assume is constant. Λ is an $N\times N$ diagonal matrix of eigenvalues of Ω, and Q is a $N\times N$ column matrix of eigenvectors.

Assume Equation 12 holds and the set of predictors X_t fully exhausts all return predictability. Applying the pricing equation $\mathbb{E}_t(M_{t+1}R^e_{t+1}) = 0$ for factors F_t gives,

$$a_F' X_t - b_t'\Sigma = 0,$$

where $\Sigma \equiv \operatorname{cov}_t(F_{t+1}) = \Lambda_K$ since $F_{t+1} = Q_K' R^e_{t+1}$ are orthogonal, where Q_K is a matrix formed from first K columns of Q and Λ_K is $K\times K$ principal submatrix of Λ.

We therefore obtain that b_t' is linear in X_t,

$$b_t' = b' X_t,$$

where $b' = \Sigma^{-1} a_F'$ is an $K\times(P+1)$ matrix.

Plugging b_t' into the SDF yields:

$$M_{t+1} = 1 - \left(b' X_t\right)' \left(F_{t+1} - \mathbb{E}_t F_{t+1}\right).$$

Apply the pricing equation, use all X_t as instruments, and replace expectations with sample moments $\mathbb{E}_T(\cdot)$ to get $K \times (P+1)$ GLS moment conditions (GLS puts all weights on factors themselves to estimate all parameters) for an estimate of b:

$$\mathbb{E}_T\left(F_{t+1} \otimes X_t\right) - \mathbb{E}_T\left[\Lambda_K b' X_t \otimes X_t\right] = 0,$$

We can rewrite the moment conditions as follows,

$$g_T^{\text{SDF}}(b) = \mathbb{E}_T\left[\left(F_{t+1} - \Lambda_K b' X_t\right) \otimes X_t\right] = 0. \tag{14}$$

PC-based predictability approach. Consider now our PC-based approach which prescribes estimating Equation 12 only for factors (dominant PCs), F_{t+1}. Pre-multiply equation (12) by Q_K':

$$\begin{aligned} F_{t+1} \equiv Q_K' R_{t+1}^e &= Q_K' a' X_t + Q_K' \varepsilon_{t+1} \\ &= \tilde{b}' X_t + \tilde{\varepsilon}_{t+1}, \end{aligned} \tag{15}$$

where F_{t+1} denotes a vector of excess returns of K largest principal components and $\tilde{b}'$ is an $K \times (P+1)$ matrix of coefficients. In the case of this PC-based regression, we use the following $K \times (P+1)$ OLS moments:

$$g_T^{\text{PC TS reg.}}(b) = \mathbb{E}_T\left[\left(F_{t+1} - \tilde{b}' X_t\right) \otimes X_t\right] = 0. \tag{16}$$

Equivalency. Comparing equations (16) and (14), we get that two sets of moment conditions corresponding to pricing factors (PCs) are equivalent iff

$$\tilde{b}' \equiv Q_K' a' = \Lambda_K b'. \tag{17}$$

Note that a' is of size $N \times (P+1)$ while b' is $K \times (P+1)$, where K is the number of factors in the SDF and $K \ll N$. The SDF approach (Equation 14) therefore imposes restrictions on the coefficients relative to the naive equation-by-equation predictive regression approach (Equation 13). Predicting only K large PCs and ignoring the rest, as in PC-based predictability approach in Equation 16, respects those restrictions and provides an alternative regression-based method of estimating the SDF. On the contratry, in the naive predictive regression setting these restrictions end up being approximately satisfied for large PCs but might be drastically violated for small PCs.

Therefore, by regressing K PCs of returns on predictive variables X_t, we are recovering re-scaled SDF coefficients $\Lambda_K b'$ in a "robust" fashion (by ignoring predictability of smallest PCs).[38] Our PC-based regression approach therefore effectively imposes SDF-implied restrictions on predictive coefficients in a simple time-series predictability regression setting (applies coefficient shrinkage).

Our method can be also viewed as an economically motivated alternative to the *reduced rank*

[38]If, instead, we standardized all PC returns by dividing them by their respective variances (eigenvalues), we would directly obtain SDF coefficients b.

regression (RRR)[39] — a regularization method in statistics and machine learning designed for dimensionality reduction in the multivariate regression setting. Similar to ridge regression, LASSO, principal-component regression, etc., RRR introduces a shrinkage penalty on the OLS estimate of a in (12), but derives its strength by combining signals from multiple responses (LHS variables). The classical RRR (Izenman 1975) uses eigenvectors derived from the eigenvalue decomposition of the covariance matrix of *fitted* responses (expected returns). Our approach is to use eigenvectors of *realized* returns, which makes the method more robust to the choice of predictors and to spurious predictability in low-order PCs when signals are weak (expected returns account for a tiny amount of variation of outcome variables). In light of this, our method is particularly suitable in the time-series predictability regression setting studied in this paper.

B Statistical Properties: Derivations

B.1 First Principal Component and Individual Regressions

A bound on individual R-squared. Define $\mathrm{R}^2_{1,i}$ the R-squared of the regression of an individual asset on the factor

$$R_{i,t+1} = \lambda_i F_{t+1} + u_{i,t+1},$$

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

$$F_{t+1} = \beta_1 X_t + e_{t+1}.$$

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

We have:

$$\mathrm{R}^2_{1,i} = \frac{\lambda_i^2 \mathrm{var}\left(F_{t+1}\right)}{\mathrm{var}\left(R_{i,t+1}\right)}$$

$$\mathrm{R}^2_X = \frac{\beta_1^2}{\mathrm{var}\left(F_{t+1}\right)}$$

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

$$R_{i,t+1} = b_i X_t + \varepsilon_{i,t+1},$$

which is equal to:

$$\mathrm{R}^2_{X,i} = \frac{b_i^2}{\mathrm{var}\left(R_{i,t+1}\right)}.$$

By linearity of the regression, we immediately have:

$$b_i = \lambda_i \beta_1 + \mathrm{cov}\left(X_t, u_{i,t+1}\right).$$

We can bound the second term:

[39] See Hastie et al. (2011).

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

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

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

Putting the two cases together, we have:

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

In the case where F_{t+1} is the returns of a tradable portfolio, $u_{i,t+1}$ is too. It is natural to consider upper bounds on the R-squared of predicting $u_{i,t+1}$ with X_t, $\mathrm{R}_{X,ui}^2$. If for instance, one believes that predictive regression R-squared are bounded above by a value R_{max}^2, then our original bound becomes:

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

Bound for a family of returns. To understand how these relations affect a whole family of returns, consider the following simple symmetric case. Assume the factor F_{t+1} is the equal-weighted mean of these returns and all returns have the same exposure and R-squared with this factor, that is λ_i and $\mathrm{R}_{1,i}^2$ do not depend of i. We then immediately have:

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

Let us write $\gamma_i = \text{cov}(X_t, u_{i,t+1})$. We then have:

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

Then

$$\begin{aligned} \mathbb{E}_i\left[\mathrm{R}^2_{X,i}\right] &= \mathrm{R}^2_1 \mathrm{R}^2_X + \frac{\mathbb{E}_i\left[\gamma_i^2\right]}{\text{var}(R_{i,t+1})} \\ \mathbb{E}_i\left[\mathrm{R}^2_{X,i}\right] &= \mathrm{R}^2_1 \mathrm{R}^2_X + \text{var}\left(\mathrm{R}^2_{X,i}\right) \end{aligned}$$

From predicting "everything" to aggregate returns. Maintaining the same assumptions, we can rearrange Equation 8 to see what the predictability of "everything" implies for predictability of the common factor. We have:

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

At first this may not seem very powerful since $\text{var}_i\left(\mathrm{R}^2_{X,i}\right)$ 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

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

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 $\mathrm{R}^2_{\max}$ is the maximum R^2 from any individual asset forecasting regression.

B.2 Low Power of Individual Tests

A further reason for directly forecasting PC portfolios rather than individual returns is that tests of predictability of the former have significantly higher power than test of the latter. For simplicity, consider a scalar forecasting variable, x. The following results easily generalize to multivariate prediction. Suppose x forecasts only the second PC portfolio and the remaining PC returns are i.i.d Gaussian with zero mean:

$$\begin{aligned} \operatorname{cov}\left[F_i,\, x\right] &= 0\ \forall i \neq 2 \\ F_{2,\tau} &= \beta_2 x_{\tau-1} + \varepsilon_{2,t} \\ \beta_2 &= \theta\sqrt{d_2} \\ \Rightarrow \sigma_\varepsilon^2 &= d_2\left(1-\theta^2\right) \\ \phi &= 0 \\ &\Rightarrow x\ is\ i.i.d\ \mathcal{N}\left(0,\, 1\right) \end{aligned} \tag{18}$$

This implies that x forecasts F_2 with $R^2 = \theta^2$. Consider a sample forecast regression. predicting F_2 with x (no constant). By usual OLS formulas, $\left(\widehat{\beta}_2 - \beta_2\middle| x\right) \sim \mathcal{N}\left(0,\, \frac{\sigma_\varepsilon^2}{X'X}\right)$ where $X' = \left[\begin{array}{ccc} x_1 & \cdots & x_T \end{array}\right]$. For power analysis, consider repeated samples of length T. Since x is stochastic (in this thought experiment), we must compute the unconditional distribution of $\widehat{b}$. As shown below, $\frac{\sqrt{T}}{\sigma_\varepsilon}\left(\widehat{\beta}_2 - \beta_2\right) \sim t\left(T\right)$; the unconditional distribution of $\widehat{\beta}_2$ (after normalization) is t with T degrees of freedom.[40] Suppose our researcher knows σ_ε and uses a typical two-sided t-test of null hypothesis $\beta_2 = 0$ with size α. We compute the power of such as test as $\text{Prob}\left[\left\|\frac{\sqrt{\mathrm{T}}}{\sigma_\varepsilon}\left(\widehat{\beta}_2 - \beta_2\right)\right\| > \mathrm{t}_{\alpha/2,\mathrm{T}}\right]$:

$$\begin{aligned} \text{power}\left(F_2\right) &= G\left(-t_{\alpha/2,T} - \frac{\beta_2\sqrt{T}}{\sigma_\varepsilon}\right) + \left(1 - G\left(t_{\alpha/2,T} - \frac{\beta_2\sqrt{T}}{\sigma_\varepsilon}\right)\right) \\ &= G\left(-t_{\alpha/2,T} - \frac{\theta\sqrt{d_2}\sqrt{T}}{\sqrt{d_2\left(1-\theta^2\right)}}\right) + \left(1 - G\left(t_{\alpha/2,T} - \frac{\theta\sqrt{d_2}\sqrt{T}}{\sqrt{d_2\left(1-\theta^2\right)}}\right)\right) \\ &= G\left(-t_{\alpha/2,T} - \frac{\theta\sqrt{T}}{\sqrt{\left(1-\theta^2\right)}}\right) + \left(1 - G\left(t_{\alpha/2,T} - \frac{\theta\sqrt{T}}{\sqrt{\left(1-\theta^2\right)}}\right)\right) \end{aligned}$$

where G is the CDF of a t-distribution with T degrees of freedom (with associated PDF g).

Now consider forecasting an individual return, $r_{i,\tau} = q_i^{'}F_\tau$, with x. We can write

$$\begin{aligned} r_{i,\tau} &= q_{i,2}P_{2,\tau} + \left(q_{i,1}P_{1,\tau} + q_{i,3}P_{3,\tau} + \cdots q_{i,N}P_{N,\tau}\right) \\ &= q_{i,2}bx_{\tau-1} + \left(q_{i,2}\varepsilon_{2,\tau} + q_{i,1}P_{1,\tau} + q_{i,3}P_{3,\tau} + \cdots q_{i,N}P_{N,\tau}\right) \\ &= b_i x_{\tau-1} + \nu_{i,\tau} \end{aligned}$$

where $\beta_i = q_{i,2}b$. We know $\operatorname{cov}\left[\nu_{i,t},\, x_{t-1}\right] = 0$ by inspection. Similarly to above, $\frac{\sqrt{T}}{\sigma_\nu}\left(\widehat{b}_i - b_i\right) \sim$

[40] When including a constant in the regression, $\frac{\sqrt{T-1}}{\sigma_\varepsilon}\left(\widehat{b}_2 - b_2\right) \sim t\left(T-1\right)$

$t(T)$ and we again perform a t-test. What is the power?

$$\begin{aligned}
\text{power}(r_i) &= G\left(-t_{\alpha/2,T} - \frac{b_i\sqrt{T}}{\sigma_\nu}\right) + \left(1 - G\left(t_{\alpha/2,T} - \frac{b_i\sqrt{T}}{\sigma_\nu}\right)\right) \\
&= G\left(-t_{\alpha/2,T} - \frac{q_{i,2}b_2\sqrt{T}}{\sqrt{q_{i,2}^2\sigma_\varepsilon^2 + \sum_{j\neq 2} q_{i,j}^2 d_j}}\right) + \left(1 - G\left(t_{\alpha/2,T} - \langle\cdot\rangle\right)\right) \\
&= G\left(-t_{\alpha/2,T} - \frac{q_{i,2}\theta\sqrt{d_2}\sqrt{T}}{\sqrt{q_{i,2}^2 d_2\left(1-\theta^2\right) + \sum_{j\neq 2} q_{i,j}^2 d_j}}\right) + \left(1 - G\left(t_{\alpha/2,T} - \langle\cdot\rangle\right)\right) \\
&= G\left(-t_{\alpha/2,T} - \frac{\theta\sqrt{T}}{\sqrt{\left(1-\theta^2\right) + \frac{1}{d_2 q_{i,2}^2}\sum_{j\neq 2} q_{i,j}^2 d_j}}\right) + \left(1 - G\left(t_{\alpha/2,T} - \langle\cdot\rangle\right)\right).
\end{aligned}$$

Since $d_j \geq 0$, $\sqrt{\left(1-\theta^2\right) + \frac{1}{d_2 q_{i,2}^2}\sum_{j\neq 2} q_{i,j}^2 d_j} > \sqrt{\left(1-\theta^2\right)}$.

Consider the function $h(\xi)$ for $\theta > 0$:

$$\begin{aligned}
h(\xi) &= G\left(-t_{\alpha/2,T} - \frac{\theta\sqrt{T}}{\xi}\right) + \left(1 - G\left(t_{\alpha/2,T} - \frac{\theta\sqrt{T}}{\xi}\right)\right) \\
\frac{\partial h}{\partial \xi} &= \left(\frac{\theta\sqrt{T}}{\xi^2}\right) g\left(-t_{\alpha/2,T} - \frac{\theta\sqrt{T}}{\xi}\right) - \left(\frac{\theta\sqrt{T}}{\xi^2}\right) g\left(t_{\alpha/2,T} - \frac{\theta\sqrt{T}}{\xi}\right) \\
&= \left(\frac{\theta\sqrt{T}}{\xi^2}\right)\left[g\left(-t_{\alpha/2,T} - \frac{\theta\sqrt{T}}{\xi}\right) - g\left(t_{\alpha/2,T} - \frac{\theta\sqrt{T}}{\xi}\right)\right] \\
&< 0 \\
&\text{since} \\
&0 < \left(\frac{\theta\sqrt{T}}{\xi^2}\right) \text{ and} \\
&0 < g\left(-t_{\alpha/2,T} - \frac{\theta\sqrt{T}}{\xi}\right) < g\left(t_{\alpha/2,T} - \frac{\theta\sqrt{T}}{\xi}\right)
\end{aligned}$$

where the last inequality comes from symmetry of g around 0. A similar argument shows $\frac{\partial h}{\partial \xi} < 0$ when $\theta < 0$. Then by inspection, $\forall\theta$, $\text{power}(f_2) > \text{power}(r_i)$. In fact the same argument shows that if a predictor x forecasts only one PC portfolio, directly forecasting that PC has higher power than forecasting any individual return.

In the simple case where the persistence of x is zero the OLS sampling distribution has a known distribution with analytic density (t). Otherwise, however, the exact distribution is unknown. We study this issue further through simulation of the system in Equation 18 except we allow x to follow an AR(1) process: $x_\tau = \phi x_{\tau-1} + \nu$, $\text{var}_{\tau-1}[\nu_\tau] = \frac{1}{1-\phi^2}$. We calibrate the simulation as follows:

1. Obtain annual returns of bonds with maturities $2 \ldots 15$ years from the yields provided in GSW, 2007. Estimate the covariance matrix $\Omega = Q\Lambda Q'$. Let $\widetilde{\Lambda} = \Lambda$ except $\widetilde{\Lambda}_{2,2} = \left(1-\theta^2\right)\Lambda_{2,2}$.

2. We set $T = 30$ years to match the post-1985 sample commonly used (JPS)

3. We vary θ^2 from 10% to 40%, consistent with R^2 values obtained in recent research (Cochrane and Piazzesi, 2008, Joslin et al., 2014, Cieslak and Povala, 2015)

4. We use three values of ϕ: 0, 0.3, and 0.6

5. The stationary mean and variance of x are 0 and 1, respectively

For each simulation:

1. Draw x_1 from $\mathcal{N}(0, 1)$. Simulate $x_2 \cdots x_{30}$ according to the specified AR(1)

2. Draw PC returns as $F_\tau = \begin{bmatrix} 0 & b & 0 & \cdots & 0 \end{bmatrix}' x_{\tau-1} + \varepsilon_\tau$, $\varepsilon_\tau \sim \mathcal{N}\left(0, \widetilde{\Lambda}\right)$. Construct "primitive" bond returns as $R_\tau = QF_\tau$

3. Forecast all returns with x

As before, we compute the sampling variance of an estimator from the simulated distribution and construct relevant t-statistics.

B.3 Noisy Yields

We study spurious predictability in the setting of= bond anomaly returns. We do so by introducing a tiny amount of proportionate noise in yields. The simulated data comes from a no-arbitrage calibration with constant expected returns. We proceed as follows:

1. Estimate $\Omega = \text{cov}(R)$ from realized annual returns (assuming no predictability). This slightly overstates the true conditional variances (by law of total covariance) and hence biases against finding predictability

2. We set $\sigma_\varepsilon = 5bp$. Context: annual yield changes have ~1% standard deviation.

Given the calibrated parameters, we simulate 100,000 realizations of 30-year histories:

1. For each simulation

   (a) Draw R_τ from $\mathcal{N}(0, \Omega)$

   (b) Draw measurement error $\varepsilon_{n,\tau}$ from $\mathcal{N}(0, \sigma_\varepsilon^2)$ with $n = 1 \ldots 5$

   (c) Construct observed returns from Equation 11

   (d) Form PC portfolios from standardized returns

   (e) Forecast each PC returns using five measurement errors $f_{i,\tau} = a + b' x_{\tau-1} + \nu_{i,\tau}$ where $x_\tau = [\varepsilon_{1,\tau} \cdots \varepsilon_{N,\tau}]'$

2. From the distribution of simulated predictive coefficients, compute $\text{cov}[b]$

3. For each simulation, compute Wald statistics for relevant tests using $\text{cov}[b]$ from previous step

C Time-varying Betas

We illustrate the idea using the cross section of bonds as an example. Similar results obtain for equities.

Though LevelR and SlopeR have zero unconditional covariance by construction, it is possible that time-varying loading of SlopeR with respect to LevelR combined with time-varying expected returns on LevelR generate the predictability of SlopeR. We address this possiblity using the methodology of Lewellen and Nagel (2006). Without loss of generality, we can express the conditional expected returns on SlopeR as

$$\mathbb{E}_t\left(R_{s,t+1}\right) = \alpha_{s,t} + \underbrace{\frac{\text{cov}_t\left(R_{s,t+1}, R_{l,t+1}\right)}{\text{var}_t\left(R_{l,t+1}\right)}}_{\beta_{s,t}} \lambda_t, \tag{19}$$

where $R_{l,t+1}$ and $R_{s,t+1}$ are the returns on LevelR and SlopeR, respectively and $\lambda_t = \mathbb{E}_t\left(R_{l,t+1}\right)$ is the conditional expected return on LevelR. Using short window time-series regressions to estimate $\alpha_{s,t}$, Lewellen and Nagel (2006) use a similar setup to test the conditional CAPM. That is, replacing LevelR with the aggregate stock market return and SlopeR with an anomaly portfolio such as the momentum or value factors, their goal is to test whether $\mathbb{E}\left(\alpha_{s,t}\right) = 0$, the average conditional pricing error is zero. Given an unbiased estimator $\hat{\alpha}_{s,t}$ from the short window regressions, the time-series average $\mathbb{E}\left(\hat{\alpha}_{s,t}\right)$ is an unbiased and consistent estimator of $\mathbb{E}\left(\alpha_{s,t}\right)$.

Instead of testing whether $\mathbb{E}\left(\alpha_{s,t}\right) = 0$, we want to test the hypothesis that all time variation in the risk-premium on SlopeR is driven by time varying risk premia on LevelR. That is, we test the null hypothesis that $\alpha_{s,t}$ is a constant:

$$\mathbb{E}_t\left(R_{s,t+1}\right) = \alpha_s + \frac{\text{cov}_t\left(R_{s,t+1}, R_{l,t+1}\right)}{\text{var}_t\left(R_{l,t+1}\right)} \lambda_t. \tag{20}$$

For this test, the approach Lewellen and Nagel (2006) requires modification. Consider an instrument with zero unconditional mean, z_t, which in our example will be the expected return on SlopeR given our predictive regression. Then multiplying Equation 20 through by z_t and using iterated expectations we obtain

$$\mathbb{E}_t\left(z_t R_{s,t+1}\right) = z_t \alpha_s + \frac{\text{cov}_t\left(z_t R_{s,t+1}, R_{l,t+1}\right)}{\text{var}_t\left(R_{l,t+1}\right)} \lambda_t. \tag{21}$$

Hence, a SlopeR managed portfolio with time-varying weight z_t should have conditional α with respect to LevelR equal to $\alpha_{mp,t} \equiv z_t \alpha_s$. Unconditionally we have for the managed portfolio $\mathbb{E}\left(\alpha_{mp,t}\right) = \alpha_s \mathbb{E}\left(z_t\right) = 0$. Hence, the single factor model Equation 20 implies any such managed portfolio should have average conditional α equal to zero.

We construct z_t exactly as above for the managed portfolio, as the difference between the conditional and unconditional expected return on SlopeR implied by the predictive regression in Table 7; by construction $\mathbb{E}\left(z_t\right) = 0$. Following Lewellen and Nagel (2006) we use quarterly and semiannual windows to estimate conditional αs, which we then average over time. Because short-window regressions require high frequency returns, we compute daily zero-coupon bond returns using the parametric yield curve given in Gürkaynak et al. (2006). We construct daily LevelR and SlopeR using the eigenvectors computed from annual holding period returns given in Figure 1. Table 18 gives estimated $\mathbb{E}(\alpha_{mp,t})$ with t-statistics and one-sided p-values for quarterly and semiannual window size. Standard errors are computed using a circular block bootstrap with two-year block size which fully accounts for parameter uncertainty in constructing z_t. For comparison we also show α_{mp} from an unconditional regression of $R_{mp,t+t} \equiv z_t R_{s,t+1}$ with respect to LevelR. The estimates reject the hypothesis that α_s is constant. Further, the average conditional α estimates are quite close to the unconditional α, indicating that time-variation in the loading of SlopeR on LevelR is of negligible importance in generating predictability of SlopeR.

Table 18: Testing Conditional Single Factor Pricing

The table shows estimated $\mathbb{E}(\alpha_{mp,t})$ with t-statistics and one-sided p-values for quarterly and semiannual window size. Standard errors are computed using a circular block bootstrap with two-year block size which fully accounts for parameter uncertainty in constructing z_t. For comparison we also show α_{mp} from an unconditional regression of $R_{mp,t+t} \equiv z_t R_{s,t+1}$ with respect to LevelR (first column).

	Unconditional	Quarterly	Semiannual
α (b.p. monthly)	1.51	1.31	1.38
	(2.35)	(1.83)	(1.97)
p-value (%)	0.95	3.34	2.44

D Additional Results

D.1 Bonds

Table 19: Predicting PC Returns with Forward Spreads and *GRO*

We report predictive coefficients, absolute t-statistics (in parentheses), and R^2s from predictive regressions of the first two principal components of zero-coupon bond excess returns on the the first three principal components of lagged forward spreads and *GRO* (Chicago Fed National Activity Index). Individual bond returns are computed from the Fama-Bliss zero-coupon yields (Fama and Bliss, 1987). Out-of-sample R^2s are calculated using the first half of the sample to estimate coefficients and using these estimates to forecast returns in the second half.

	FS1	FS2	FS3	GRO	R^2
LevelR	0.06	-2.14	-0.92	-1.29	0.26
	(0.10)	(3.37)	(1.45)	(1.80)	
SlopeR	2.68	0.84	-0.71	2.33	0.36
	(4.62)	(1.45)	(1.25)	(3.73)	

D.2 Stocks

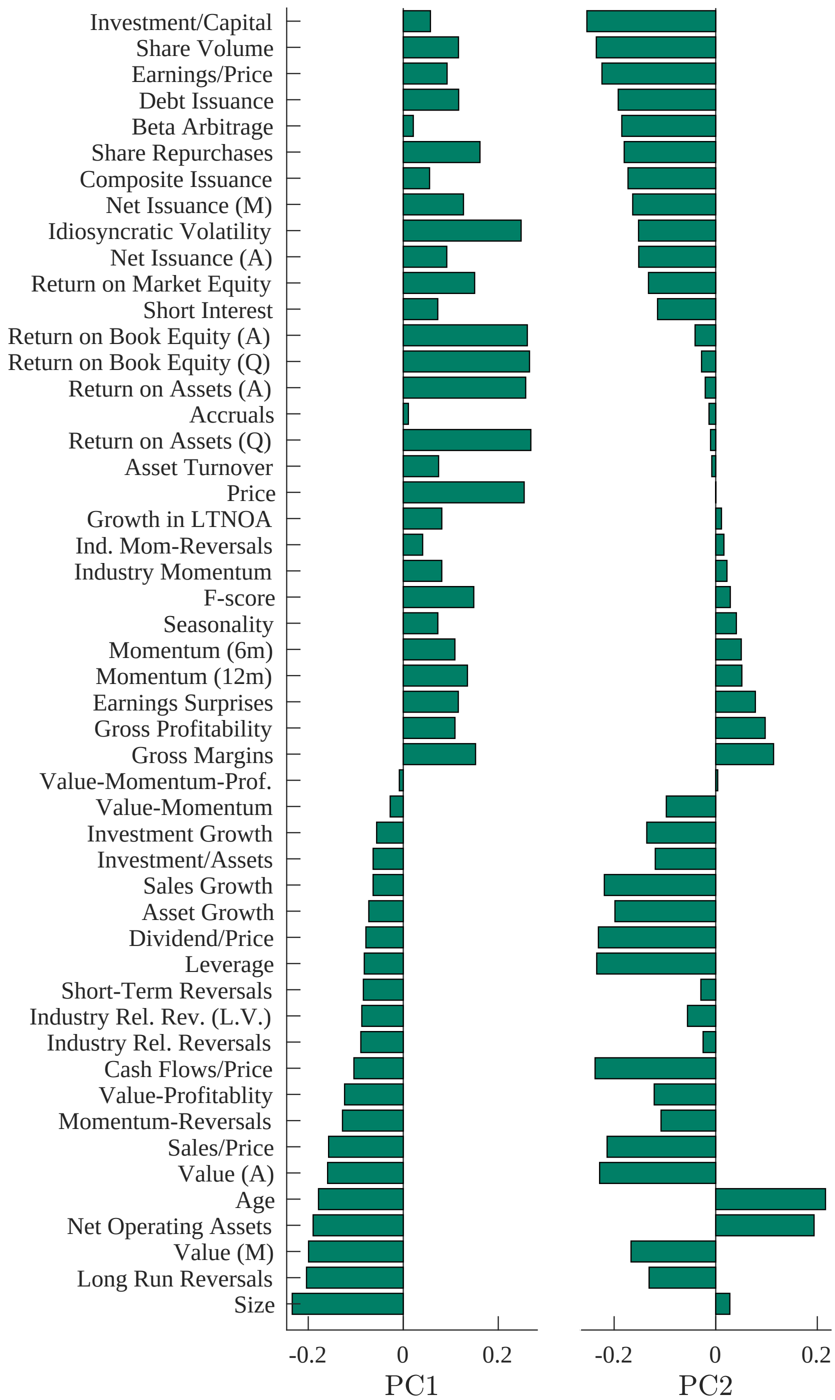


Figure 11: Anomalies Eigenvector Loadings. The figure plots eigenvector loadings of 50 long-short anomalies.

Table 20: Part I: Anomaly porfolios 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 November 1973 to December 2015.

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P10-P1
1. Size	5.8	7.7	8.2	8.8	8.6	9.2	8.9	9.5	9.0	8.9	3.2
2. Value (A)	5.0	7.0	7.9	7.2	8.3	8.3	8.8	8.5	8.5	11.9	6.9
3. Gross Profitability	5.4	5.8	6.3	5.8	7.7	6.7	7.1	6.4	7.0	9.2	3.8
4. Value-Profitablity	4.2	5.5	4.3	6.2	8.5	8.1	10.3	11.3	11.3	13.2	9.0
5. F-score	6.1	-	-	-	-	-	-	-	-	7.0	0.9
6. Debt Issuance	7.9	-	-	-	-	-	-	-	-	6.2	-1.6
7. Share Repurchases	6.2	-	-	-	-	-	-	-	-	7.7	1.5
8. Net Issuance (A)	2.6	5.0	8.4	7.8	7.8	7.2	6.4	8.7	8.1	11.3	8.7
9. Accruals	3.8	6.0	5.0	7.0	6.9	7.2	8.1	7.3	9.4	8.1	4.4
10. Asset Growth	4.9	6.6	7.2	7.2	6.9	7.5	7.3	8.8	10.0	9.8	5.0
11. Asset Turnover	4.5	6.9	6.0	6.3	7.5	7.9	8.4	6.7	9.4	9.2	4.7
12. Gross Margins	6.6	6.8	7.9	7.0	8.4	6.2	7.3	6.7	5.9	6.8	0.2
13. Dividend/Price	5.5	4.8	6.7	6.9	7.1	9.0	9.5	7.9	7.8	8.6	3.1
14. Earnings/Price	4.0	4.8	6.5	7.3	7.1	7.7	9.4	9.0	8.7	11.6	7.6
15. Cash Flows/Price	4.5	7.3	5.9	7.9	8.4	8.2	7.7	9.0	10.8	10.5	6.0
16. Net Operating Assets	6.1	7.1	7.5	7.8	8.3	8.7	8.3	8.5	7.9	7.3	1.1
17. Investment/Assets	4.2	5.1	7.4	6.2	8.1	6.1	8.0	8.4	8.6	10.3	6.0
18. Investment/Capital	5.8	6.7	6.1	7.2	6.9	8.4	7.8	7.7	8.6	9.1	3.3
19. Investment Growth	4.5	7.9	6.5	6.6	6.1	7.3	7.8	8.0	9.8	8.2	3.6
20. Sales Growth	6.8	6.8	7.1	6.5	7.6	8.7	6.6	7.7	8.9	6.8	-0.0
21. Leverage	5.3	6.4	6.7	10.2	7.4	8.3	8.6	8.6	8.8	8.1	2.8
22. Return on Assets (A)	4.1	8.5	7.2	7.1	7.3	7.1	7.1	7.7	6.1	6.8	2.8

Table 20: Part II: Anomaly porfolios mean excess returns, %, annualized

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P10-P1
23. Return on Book Equity (A)	5.9	6.9	6.7	7.6	6.4	7.0	6.4	7.0	6.3	7.5	1.6
24. Sales/Price	4.6	6.1	6.6	8.3	8.9	8.6	9.1	10.7	10.7	12.9	8.2
25. Growth in LTNOA	6.5	6.3	6.7	8.2	5.9	7.0	6.5	7.6	7.8	7.7	1.1
26. Momentum (6m)	8.9	8.8	8.4	8.5	7.5	8.1	6.6	5.3	7.2	10.5	1.6
27. Value-Momentum	6.1	7.6	6.9	6.9	8.3	9.4	9.8	8.4	8.1	11.0	4.8
28. Value-Momentum-Prof.	5.9	7.6	7.5	8.0	7.1	5.5	7.5	8.5	11.6	14.4	8.5
29. Short Interest	6.3	5.7	8.5	8.5	8.0	6.2	6.7	5.9	4.3	5.2	-1.1
30. Momentum (12m)	-2.2	4.2	5.6	6.8	5.5	6.9	6.9	9.1	9.0	12.4	14.6
31. Industry Momentum	5.9	4.6	6.6	7.1	6.0	9.4	9.3	6.3	8.4	9.6	3.8
32. Momentum-Reversals	5.1	6.8	7.2	6.7	7.5	8.6	7.2	8.9	8.7	11.4	6.3
33. Long Run Reversals	6.6	6.9	7.3	8.1	8.1	8.5	8.3	9.5	9.8	11.1	4.5
34. Value (M)	5.5	6.3	6.4	7.0	7.9	7.5	9.0	7.2	11.6	11.7	6.1
35. Net Issuance (M)	3.5	5.2	10.0	8.1	8.7	7.3	7.4	8.0	9.9	10.5	6.9
36. Earnings Surprises	4.0	4.0	4.9	7.2	6.6	7.8	7.1	7.4	8.1	10.8	6.8
37. Return on Book Equity (Q)	1.1	3.8	4.7	6.5	7.1	6.3	7.4	6.9	7.9	8.4	7.3
38. Return on Market Equity	0.1	1.5	6.3	5.9	7.2	7.0	8.0	10.6	11.4	15.6	15.5
39. Return on Assets (Q)	1.8	4.5	6.8	7.1	7.4	6.6	8.0	7.4	6.7	7.6	5.8
40. Short-Term Reversals	3.0	4.3	6.8	6.5	6.9	7.7	8.5	9.1	9.6	7.2	4.2
41. Idiosyncratic Volatility	-0.6	7.8	10.7	7.6	9.8	8.5	7.5	7.3	7.4	6.9	7.5
42. Beta Arbitrage	3.1	3.1	4.1	6.6	8.1	9.3	10.7	11.4	14.3	16.7	13.6
43. Seasonality	3.0	3.4	5.7	5.2	7.6	6.7	7.7	7.1	9.3	12.8	9.8
44. Industry Rel. Reversals	1.5	3.2	4.2	6.0	6.2	7.5	8.6	11.1	12.4	12.3	10.8
45. Industry Rel. Rev. (L.V.)	0.5	4.0	4.6	7.1	5.4	6.5	9.4	10.6	12.7	15.2	14.6
46. Ind. Mom-Reversals	2.6	4.8	5.5	6.1	7.1	7.6	7.9	8.5	10.3	14.1	11.5
47. Composite Issuance	4.0	5.6	5.9	6.1	7.3	7.2	6.9	7.6	9.9	10.2	6.2
48. Price	5.5	8.1	8.4	9.7	8.3	8.2	7.1	7.1	7.3	5.8	0.3
49. Age	6.5	7.9	5.6	9.6	5.6	7.8	9.0	7.4	7.0	6.6	0.1
50. Share Volume	6.2	7.8	6.6	6.4	7.6	6.0	7.6	6.6	6.2	6.2	0.0

Table 21: Part I: Implied anomaly returns by PC: R^2 (%)

Predictive R^2 of individual anomalies returns implied by PC forecasts. Columns labeled "Full" combine forecasts of both PCs; "PC1/PC2 only" focus only on predictability stemming from PC1/PC2.

	IS (Full)	IS (PC1 only)	IS (PC2 only)	OOS (Full)	OOS (PC1 only)	OOS (PC2 only)
1. Size	27.3	28.4	-1.0	35.3	35.6	0.6
2. Value (A)	17.8	8.3	8.6	8.8	12.4	-9.0
3. Gross Profitability	-1.5	1.0	-2.8	-40.1	-22.3	-20.0
4. Value-Profitablity	10.7	-0.4	10.8	19.7	7.2	10.7
5. F-score	5.0	4.7	0.0	5.5	9.4	-5.5
6. Debt Issuance	15.1	7.0	8.7	13.6	5.7	11.2
7. Share Repurchases	28.0	9.8	19.0	20.1	-5.4	29.4
8. Net Issuance (A)	16.0	1.4	14.8	14.5	-5.9	21.5
9. Accruals	0.5	-0.1	0.7	-0.2	-0.1	-0.1
10. Asset Growth	12.9	5.7	6.9	17.7	10.5	5.5
11. Asset Turnover	2.2	1.7	0.5	0.2	-0.8	1.0
12. Gross Margins	7.3	3.7	3.1	-8.2	-8.1	-2.9
13. Dividend/Price	13.2	-0.6	13.3	11.9	-0.3	9.0
14. Earnings/Price	12.5	2.8	10.2	19.3	3.9	17.8
15. Cash Flows/Price	9.9	4.9	4.3	-1.5	7.4	-13.8
16. Net Operating Assets	17.5	16.8	1.7	25.8	10.5	20.3
17. Investment/Assets	15.6	6.6	8.8	14.3	18.2	-5.7
18. Investment/Capital	1.0	0.1	1.2	-1.5	0.0	-0.3
19. Investment Growth	14.0	5.9	7.9	13.5	9.2	3.2
20. Sales Growth	6.1	2.3	3.5	5.7	4.3	-0.8
21. Leverage	21.3	2.4	18.5	25.8	5.0	18.9
22. Return on Assets (A)	29.2	28.1	1.2	24.2	21.1	3.9
23. Return on Book Equity (A)	27.7	26.5	1.5	23.1	16.8	7.9
24. Sales/Price	15.5	1.7	13.1	17.3	2.0	11.6
25. Growth in LTNOA	-7.7	-7.9	0.2	-2.5	-3.4	0.7
26. Momentum (6m)	13.6	14.7	-1.2	10.6	13.3	-3.2
27. Value-Momentum	3.0	-0.6	3.6	2.9	0.2	2.3
28. Value-Momentum-Prof.	-0.7	-0.6	-0.2	-0.4	0.2	-0.6
29. Short Interest	8.4	9.2	-0.5	15.5	14.8	2.1
30. Momentum (12m)	10.2	13.3	-3.3	15.1	17.4	-3.2
31. Industry Momentum	7.3	8.5	-1.3	0.6	5.0	-5.0
32. Momentum-Reversals	6.9	-7.2	13.7	16.2	-0.8	15.3
33. Long Run Reversals	29.0	20.0	8.4	30.3	27.0	-0.4
34. Value (M)	29.6	19.3	9.4	25.0	27.8	-7.8
35. Net Issuance (M)	12.1	5.0	7.7	12.7	-7.9	23.4

Table 21: Part II: Implied anomaly returns by PC: R^2 (%)

	IS (Full)	IS (PC1 only)	IS (PC2 only)	OOS (Full)	OOS (PC1 only)	OOS (PC2 only)
36. Earnings Surprises	-9.6	-0.3	-9.6	-2.4	10.0	-14.9
37. Return on Book Equity (Q)	23.0	22.5	0.7	16.7	14.0	3.6
38. Return on Market Equity	15.0	4.5	11.0	18.4	6.8	13.6
39. Return on Assets (Q)	25.4	24.8	0.6	18.0	16.8	1.6
40. Short-Term Reversals	6.6	4.5	2.0	11.1	10.8	-0.3
41. Idiosyncratic Volatility	38.2	30.2	8.9	38.8	27.3	16.1
42. Beta Arbitrage	10.2	1.1	9.1	9.9	1.9	8.5
43. Seasonality	-2.2	-1.1	-1.2	-2.1	-0.9	-1.7
44. Industry Rel. Reversals	9.0	5.2	3.7	14.3	12.3	1.5
45. Industry Rel. Rev. (L.V.)	13.0	4.7	8.2	19.9	14.2	5.1
46. Ind. Mom-Reversals	-1.8	0.1	-2.0	-5.0	-4.1	-1.1
47. Composite Issuance	8.2	2.6	5.8	7.7	-0.1	9.8
48. Price	38.3	38.3	0.0	41.2	41.2	-0.0
49. Age	25.8	11.7	15.2	28.5	1.7	31.2
50. Share Volume	11.5	5.0	7.1	11.0	1.4	12.1
Mean R^2	12.9	7.9	5.0	12.3	7.7	4.5
Median R^2	12.3	4.8	3.6	14.0	6.2	1.8
Std. Dev. of R^2	11.0	10.2	6.1	13.7	11.6	10.7