Conditional dynamics and the multi-horizon risk-return trade-off*

Mikhail Chernov,† Lars A. Lochstoer,‡ and Stig R. H. Lundeby§

First Draft: May 12, 2018
This Draft: April 29, 2019

Abstract

We propose testing asset-pricing models using multi-horizon returns (MHR). MHR serve as powerful source of conditional information that is economically important and not data-mined. We apply MHR-based testing to linear factor models. These models seek to construct the unconditionally mean-variance efficient portfolio. We reject many popular models that deliver high maximum Sharpe ratios in a single-horizon setting. Model misspecification manifests itself in strong intertemporal dynamics of the factor loadings in the SDF representation.

JEL Classification Codes: G12, C51.

Keywords: multi-horizon returns, stochastic discount factor, linear factor models.

* We thank Ian Dew-Becker, Greg Duffee, Valentin Haddad, Francis Longstaff, Tyler Muir, Christopher Polk, Seth Pruitt, and Tommy Stamland for comments on earlier drafts, as well as participants in the seminars and conferences sponsored by the ASU Winter Finance Conference, David Backus Memorial Conference at Ojai, LAEF, NBIM, NHH, the UBC Winter Finance Conference and UCLA. We appreciate any comments including inadvertently omitted citations. The latest version is available at https://sites.google.com/site/mbchernov/CLL_MHR_latest.pdf.

† Anderson School of Management, UCLA, NBER, and CEPR; mikhail.chernov@anderson.ucla.edu.

‡ Anderson School of Management, UCLA; lars.lochstoer@anderson.ucla.edu.

§ Norwegian School of Economics; stig.lundeby@nhh.no.

1 Introduction

For the past five decades researchers have been actively thinking about optimal multi-horizon asset allocation. There would be no difference between the single- and multi-horizon allocations only if asset returns were i.i.d. (or investors would have log preferences). Normative analysis that solves for optimal allocation is very difficult as emphasized by many authors (Brennan, Schwartz, and Lagnado, 1997, Campbell and Viceira, 1999, Cochrane, 2014, among many others). That is why researchers and practitioners alike continue solving the myopic problem and then correct the solution in ad-hoc ways to account for persistence in returns.[1]

One can flip the perspective and work to see what observed returns could tell about horizon-based investments. What returns would be helpful, you might ask. Differences between short- and long-horizon returns on a given asset should reflect their dynamic properties. Thus, multi-horizon returns (MHR) should be able to serve as a useful diagnostic of the models that one could potentially use for estimating expected asset returns and for multi-horizon allocation amongst these assets. In this paper, we propose a methodology that explicitly incorporates MHR into testing of asset pricing models.

Our tests are based on a given model's implications for the stochastic discount factor (SDF). Specifying an asset-pricing model via the SDF is a powerful analytical tool that encodes the risk-return trade-off across multiple assets and across multiple horizons, which provides the link between MHR and testable implications from the model.

Surprisingly, there is little research on evaluation of SDF models using MHR. Theoretical work has developed tools allowing researchers to characterize properties of equilibrium models at different horizons (e.g., Hansen and Scheinkman, 2009, Backus, Chernov, and Zin, 2014). Empirical work highlights properties of MHR but does not

[1]For instance, Cochrane (2014) observes: "Even highly sophisticated hedge funds typically form portfolios with one-period mean-variance optimizers – despite the fact that mean-variance optimization for a long-run investor assumes i.i.d. returns, while the funds' strategies are based on complex models of time-varying expected returns, variances, and correlations. Beyond formal portfolio construction, their informal thinking and marketing is almost universally based on one-period mean-variance analysis ... Institutions, endowments, wealthy individuals, and regulators struggle to use even the discipline of mean-variance analysis in place of name-based buckets ..."

use them to explicitly test models of the SDF (e.g., Campbell, 2001, Kamara, Korajczyk, Lou, and Sadka, 2016).

We make two key contributions. First, we show that an MHR-based test amounts to a stringent evaluation of conditional implications of a model. In particular, persistent misspecification shows up in long-run pricing errors. We develop MHR-based GMM estimation and testing of asset pricing models. In particular, we take care to setup moment conditions in such a way that their residuals are not serially correlated under the null of a model. That improves the small-sample performance of the test in a setting with overlapping observations.

Second, we show that misspecification of the temporal dynamics in state-of-the-art models of the SDF, as uncovered by MHR, indeed are quantitatively large. In particular, we consider linear factor models of the form SDF $= a - b^\top F$, which arguably are the workhorse models for empirical risk-return modeling. It is well-known that such an SDF implies a linear beta-pricing model of expected returns (e.g., Cochrane, 2004). Under the null of these SDFs, the traded factors F span the unconditional mean-variance efficient (MVE) portfolio. We impose the minimal requirement on a factor model that it prices its own factors. The new idea in our approach is to impose this requirement on factor returns at multiple horizons.

We consider nine linear model that either viewed as classic in the literature, or that are state-of-the-art recent additions. Specifically, we include CAPM, CAPM and Betting Against Beta (Frazzini and Pedersen, 2014, Jensen, Black, and Scholes, 1972, Novy-Marx and Velikov, 2016), Fama and French 3- and 5-factor models (Fama and French, 1993, Fama and French, 2016), versions of the later two models with unpriced risks hedged (Daniel, Mota, Rottke, and Santos, 2018), a volatility-managed version of the Fama and French 5-factor model (Moreira and Muir, 2017), and the recent 4-factor model of Stambaugh and Yuan (2017). We show that the dynamics of the factors are grossly inconsistent with the dynamics implied under the null of the factor model. In other words, current models of the risk-return trade-off, which are typically estimated at a monthly or quarterly frequency, do a poor job accounting for the risk-return trade-off at longer horizons. Our MHR-based tests reject all these models.

The pricing errors are economically large, ranging from 1% to 3.5% per year. Fur-

ther, we find that models deemed successful on the basis of single-horizon returns (thus having factors that yield a high maximal SR) tend to do worse at long horizons, that is, they have large MHR pricing errors. As one considers models with increasingly higher short-run maximal SR, and, presumably, gets closer to the unconditional MVE portfolio, it would be natural to expect improved pricing at longer horizons. In this sense, our results are surprising. Finally, as a robustness check, we run standard "alpha-regressions" and use Gibbons, Ross, and Shanken (1989) (GRS) to test whether managed portfolios based on MHR can deliver alpha in the standard, short-horizon setting. The models are rejected with high maximal information ratios indicating economic significance via an alternative metric.

The reason the models reject is that the model-implied SDFs do not price the test assets conditionally. If conditional pricing errors are persistent (autocorrelated), they may affect longer-horizon excess factor returns. Specifically, persistent pricing errors that appear "small" at the monthly horizon can become large over multiple horizons as they accumulate over time. This is the kind of misspecification MHR is detecting when used as test assets.

An unconditionally MVE portfolio prices assets also conditionally (Hansen and Richard, 1987). Thus, the rejection suggests that the factors do not unconditionally span the MVE portfolio. However, because we evaluate the models using MHR to the pricing factors themselves, it follows that there must exist time-varying a's and b's in the SDF that do price MHR correctly. To assess the nature of the misspecification in more detail, we model time-varying b_t and estimate it together with the corresponding a_t by including MHR in the set of test assets. The estimated factor loading has large variation and little relation to the standard conditioning variables used in asset-pricing. The evidence is puzzling from the perspective of existing evidence and theory.

There are many papers that test conditional versions of factor models. In our notation, that amounts modeling time-varying b_t as a linear function of various instrumental variables. For instance, Jagannathan and Wang (1996) and Lettau and Ludvigson (2001) consider conditional versions of the CAPM and the consumption CAPM using proxies for the conditioning information related to aggregate discount

rates, such as the dividend-price ratio, to arrive at unconditional linear multi-factor models with constant factor loadings in the SDF. Other examples include Ferson and Harvey (1999) and Farnsworth, Ferson, Jackson, and Todd (2002). Our contribution relative to this literature is to use MHR to infer time-varying b_t and to document that little of variation in b_t is connected to those instrumental variables.

Our paper also makes a connection with a literature that seeks to characterize multi-horizon properties of "zero-coupon" assets, such as bonds, dividends strips, variance swaps, and currencies. Such work includes Belo, Collin-Dufresne, and Goldstein (2015), van Binsbergen, Brandt, and Koijen (2012), Dahlquist and Hasseltoft (2013), Dew-Becker, Giglio, Le, and Rodriguez (2015), Hansen, Heaton, and Li (2008), Lustig, Stathopoulos, and Verdelhan (2013), and Zviadadze (2017). There is also an earlier literature that considers multiple frequencies of observations when testing models (e.g., Brennan and Zhang, 2018, Daniel and Marshall, 1997, Jagannathan and Wang, 2007, Parker and Julliard, 2005).

Notation. We use E for expectations and V for variances (a covariance matrix if applied to a vector). A t-subscript on these denotes an expectation or variance conditional on information available at time t, whereas no subscript denotes an unconditional expectation or variance. We use double subscripts for time-series variables, like returns, to explicitly denote the relevant horizon. Thus, a gross return on an investment from time t to time $t+h$ is denoted $R_{t,t+h}$.

2 Testing asset pricing models using MHR

In this section, we evaluate the implications of jointly testing whether a proposed model of the SDF can explain the empirical risk-return trade-off across horizons. Intuitively, risks that appear very important at a short horizon may be less important at longer-horizons relative to other, more persistent risks, and vice versa. Analysis of such dynamics has received little attention in the previous work, despite the relevance for theoretical models and investment practice.

2.1 Preliminaries

Let the one-period stochastic discount factor (SDF) from time t to $t+1$ be $M_{t,t+1}$. The Law of One Price (LOOP) then states that

$$E_t(M_{t,t+1}R_{t,t+1}) = 1,$$

or

$$E_t(M_{t,t+1}(R_{t,t+1} - R_{f,t,t+1})) = 0.$$

Here $R_{t,t+1}$ and $R_{f,t,t+1}$ are one-period gross return on a risky and the risk-free assets. The overwhelming majority of empirical asset pricing papers are concerned with tests of these relationships, where a period is typically a month or a quarter.

2.2 The MHR Test

The framework offers a natural way to propagate the model implications across multiple horizons. The multi-horizon SDF and returns are simple products of their one-period counterparts:

$$\begin{aligned} M_{t,t+h} &= \prod_{j=1}^{h} M_{t+j-1,t+j}, \\ R_{t,t+h} &= \prod_{j=1}^{h} R_{t+j-1,t+j}, \\ R_{f,t,t+h} &= \prod_{j=1}^{h} R_{f,t+j-1,t+j}. \end{aligned}$$

LOOP still holds:[2]

$$E_t(M_{t,t+h}R_{t,t+h}) = 1, \tag{1}$$

or, for excess returns,

$$E_t(M_{t,t+h}(R_{t,t+h} - R_{f,t,t+h})) = 0. \tag{2}$$

The unconditional version of this condition can be easily tested jointly for multiple horizons h in a GMM framework.

2.3 Interpretation

To see what MHR add to the evaluation of a candidate stochastic discount factor, assume the model unconditionally prices one-period returns correctly, i.e., $E(M_{t,t+1}R_{t,t+1}) = 1$. Applying the model to two-period returns from time $t-1$ to time $t+1$, we can write:

$$\begin{aligned} E(M_{t-1,t+1}R_{t-1,t+1}) &= E(M_{t-1,t}R_{t-1,t}) \cdot E(M_{t,t+1}R_{t,t+1}) \\ &+ Cov(M_{t-1,t}R_{t-1,t}, M_{t,t+1}R_{t,t+1}). \end{aligned}$$

Therefore, if (1) holds at both one- and two-period horizons,

$$Cov(M_{t-1,t}R_{t-1,t}, M_{t,t+1}R_{t,t+1}) = 0.$$

To appreciate the meaning of this restriction of the covariance, define one-period pricing errors as $M_{s,s+1}R_{s,s+1} - E_s(M_{s,s+1}R_{s,s+1})$. Because R is a gross return, the conditional expectation of its product with the SDF is equal to one. Thus, the

[2]It is immediate from the law of iterated expectations, that an SDF that prices a set of one period returns conditionally (i.e., $E_t[M_{t,t+1}R_{t,t+1}] = 1$), also prices multi-horizon returns to the same set of assets (i.e., $E[M_{t-h,t+1}R_{t-h,t+1}] = 1$ for any $h \geq 1$). This can be seen by recursively iterating on the following equation for $h = 1, 2, ...$, etc.: $E[M_{t-h,t+1}R_{t-h,t+1}] = E[M_{t-h,t}R_{t-h,t}M_{t,t+1}R_{t,t+1}] = E[M_{t-h,t}R_{t-h,t}E_t[M_{t,t+1}R_{t,t+1}]] = E[M_{t-h,t}R_{t-h,t}]$, where the last equality follows if the model prices the one-period returns conditionally.

condition above that the covariance term equals zero implies that past pricing errors cannot predict future pricing errors. This is the hallmark of a correctly specified model.

A generalization to $(h+1)$-period returns, under covariance-stationarity, is

$$E(M_{t-h,t+1}R_{t-1,t+1}) = 1 + \sum_{j=0}^{h-1} Cov(M_{t-h+j,t}R_{t-h+j,t}, M_{t,t+1}R_{t,t+1}). \tag{3}$$

Thus, adding information about MHR is a test of whether multi-horizon pricing errors predict one-period pricing errors.

The virtue of this representation is that it allows us to construct a test on the basis of moment conditions whose residuals are not serially correlated. This improves the small-sample performance of the test. Specifically, we consider moment conditions of the form:

$$f_{t+1} = \begin{pmatrix} M_{t,t+1}R_{t,t+1} - 1 \\ (\sum_{j=0}^{h_1-1} M_{t-h_1+j,t}R_{t-h_1+j,t}) \cdot (M_{t,t+1}R_{t,t+1} - 1) \\ \dots \\ (\sum_{j=0}^{h_n-1} M_{t-h_n+j,t}R_{t-h_n+j,t}) \cdot (M_{t,t+1}R_{t,t+1} - 1) \end{pmatrix}, \tag{4}$$

where n is the number of horizons used in the test. The null hypothesis is $E(f_{t+1}) = 0$. See Hodrick (1992) for a similar argument.[3]

2.4 Implementation

We estimate and test models using Equation (4) in two ways. First, we estimate the models using single-horizon returns (SHR), $R_{t,t+1}$, of N assets and test on MHR of the same assets. That is, in the language of GMM, we use a weighting matrix with upper $N \times N$ elements equal to the N-dimensional identity matrix. That corresponds to the first row in Equation (4). With n horizons, we set the remaining elements in

[3]That residuals are not autocorrelated follows from Equation (1). For simplicity, consider one horizon, h. Define $z_t = \sum_{j=0}^{h-1} M_{t-h+j,t}R_{t-h+j,t}$. Then, we have that $E(f_t \cdot f_{t+1}) = E(f_t \cdot E_t(f_{t+1})) = E(f_t \cdot z_t E_t(M_{t,t+1}R_{t,t+1} - 1)) = 0$, because, under the null, $E_t(M_{t,t+1}R_{t,t+1} - 1) = 0$ for all t.

the $(n-1)N \times (n-1)N$ weighting matrix to zero. We test each model by testing if the average of the $(n-1)N$ MHR moment conditions jointly equal zero. This test is standard and described in, e.g., Cochrane (2004). The spectral density matrix is estimated as the sample covariance matrix of the moment vector in (4) because there is no autocorrelation in the moment conditions under the null hypothesis.

Second, we estimate models using all the moments, i.e. including the MHR moments. In that case, we use the inverse of the spectral density matrix estimated in our former approach as weighting matrix. Thus, this second test is a two-stage efficient GMM test, where the information in MHR moments is allowed to affect the estimated parameters in the stochastic discount factor.

The MHR-based moment conditions are model-implied overidentifying moments. The geometric nature of aggregating MR makes the moment conditions highly non-normal. Thus, the asymptotic GMM tests may be badly behaved in the samples of the length that is available to us. To ensure the p-values account for such issues, we compute bootstrapped values of the GMM objective function imposing the null of the model for both tests. We then compare the objective function in the data to the bootstrapped distribution and find the p-value from this bootstrapped version of the so-called "J-test."

The null hypothesis implies that $M_{t,t+1}R_{t,t+1}$ is unpredictable. We impose this in our bootstrap simply by drawing (with replacement) the cross-section of one-period returns to the test assets as well as the variables in the SDF at random time t. This is a valid approach as under the null hypothesis the MR are not autocorrelated. We then create MHR and moment conditions in the bootstrapped samples in the same way as in the data. We create 2,500 artificial samples of the same length as the data and estimate the model for each sample to find 2,500 GMM objective function values. These bootstrapped samples construction impose the property of the null hypothesis that $M_{t,t+1}R_{t,t+1}$ is uncorrelated with any lagged MHR-based moment.

3 Linear factor models

3.1 Preliminaries

We explore linear factor models, the most pervasive set of asset pricing models, as an empirical application of our approach. The focus of linear factor-based modeling is to find a set of excess return factors that span the conditionally mean-variance efficient (MVE) portfolio. The reason researchers are interested in this pursuit is that besides an efficient trading strategy, the MVE portfolio is connected to the SDF. Specifically, let $F_{t,+1}$ be a $K \times 1$ vector of factor excess returns. Then the SDF

$$M_{t,t+1} = a_t - b_t^\top F_{t,t+1}, \tag{5}$$

where a_t is a scalar that ensures the one-period risk-free bond is priced correctly, and b_t is a $K \times 1$ vector proportional to the conditional MVE portfolio weights. This SDF prices all traded assets, that is, Equation (1) holds.

Equations (1) and (5) imply, when applied to the one-period risk-free bond, that:

$$a_t = R_{f,t,t+1}^{-1} + b_t^\top E_t(F_{t,t+1}). \tag{6}$$

Equations (2) and (5) imply, when applied to the factors themselves, that:

$$b_t = [R_{f,t,t+1} V_t(F_{t,t+1})]^{-1} E_t(F_{t,t+1}). \tag{7}$$

Equation (7) tells us that the conditional MVE weights are related to the dynamic properties of the factors, that is, the conditional mean and covariance matrix of the factors. Equations (5) to (7) also combine to give another common representation of the SDF:

$$M_{t,t+1} = R_{f,t,t+1}^{-1} - \lambda_t^\top \varepsilon_{t+1}, \tag{8}$$

where $\varepsilon_{t+1} = V_t(F_{t,t+1})^{-1/2}(F_{t,t+1} - E_t(F_{t,t+1}))$ is i.i.d., zero mean, and unit variance,

and where

$$\lambda_t^\top = b_t \cdot V_t(F_{t,t+1})^{1/2}, \tag{9}$$

are commonly referred to as the conditional prices of risk.

Unfortunately, a_t and b_t are not observed. Thus, unless one is willing to explicitly model a_t and b_t, it is impossible to measure them. For instance, the literature on conditional factor models specifies a_t and b_t as a linear function of instrumental variables (e.g., Ferson and Harvey (1999), Jagannathan and Wang (1996), Lettau and Ludvigson (2001), among many others).

Note that if b_t is proportional to a_t, $b_t = a_t \cdot k$, the factors span both the conditional and the unconditional MVE portfolio. In particular, the portfolio with constant weights $k^\top F_{t,t+1}$ is both conditionally and unconditionally MVE (e.g., Hansen and Richard, 1987). This result prompts most of the literature to search for factors whose excess returns span the *unconditional* MVE portfolio (e.g., Fama and French, 1993). Many of the aforementioned models feature very high maximal short-run Sharpe ratio (SR). Thus, the hope is that these models get us close to the unconditional MVE portfolio.

The remaining issue is the modeling of a_t. The literature circumvents this problem by focusing on pricing excess returns only. In this case, the level of the corresponding SDF is not identified and the scalar is typically set to 1 (see, e.g., Cochrane, 2004). We cannot use this approach when testing the model implications for MHR. The use of MHR prompts the need to model the level of the SDF and inclusion of the one-period risk-free rate as a test asset.

The easiest way to model the SDF level when MVE portfolio weights are constant is to assume that a_t is constant. This will be our leading case in the paper. Subsequently, we will show robustness of our results to the case of time-varying a_t. Thus, we study the SDF:

$$M_{t,t+1} = a - b^\top F_{t,t+1}. \tag{10}$$

3.2 Data

We select our models based on their historical importance, recent advancements, and data availability. Specifically, we include the CAPM, CAPM and BAB (Frazzini and Pedersen, 2014, Jensen, Black, and Scholes, 1972, Novy-Marx and Velikov, 2016), Fama and French 3- and 5-factor models, FF3 and FF5, respectively (Fama and French, 1993, Fama and French, 2016), version of the FF3 and FF5 models with unpriced risks hedged (Daniel, Mota, Rottke, and Santos, 2018), a volatility-managed version of the FF5 model (Moreira and Muir, 2017), and the recent 4-factor model of Stambaugh and Yuan (2017).

The Fama-French 5-factor model includes the market factor (MKT), the value factor (HML), the size factor (SMB), the profitability factor (RMW; see also Novy-Marx, 2013), and the investment factor (CMA; see also Cooper, Gulen, and Schill, 2008). The Fama-French 3-factor model only have the first 3 factors. We also consider FF5+MOM, where MOM refers to the Momentum factor (Carhart, 1997, Jegadeesh and Titman, 1993). These data are provided on Kenneth French's webpage. The returns are monthly and the sample is from July 1963 to June 2017.

We also study versions of these models with unpriced risks hedged out as proposed by Daniel, Mota, Rottke, and Santos (2018) (DMRS; data available on Kent Daniel's webpage). Their alternative factor construction methodology substantially increases the Sharpe ratios of the factors, which indicates that these revised factors are closer to spanning the unconditional mean-variance efficient portfolio. See Daniel, Mota, Rottke, and Santos (2018) for details regarding the factor construction. We consider two models based on these factors: FF3_{DMRS} and FF5_{DMRS}. The sample period is July 1963 to June 2017.

Moreira and Muir (2017) show that volatility-timed factors tend to have 'alpha' relative to the corresponding original factors. In the spirit of their findings, we consider a volatility managed version of the FF5 model: FF5_{VolMan}. In particular, we use the daily return data for these original factors (also available on Kenneth French's webpage) to construct realized monthly factor variance. We then divide each factors return in month $t+1$ by the same factor's realized variance in month t. The factor sample for this model is then August 1963 to June 2017.

Stambaugh and Yuan (2017) propose two factors intended to capture stock mispricing, in addition to the existing MKT and SMB factors: PERF and MGMT. We denote this four-factor model as SY. These data are available on Robert Stambaugh's webpage. The sample period for these factors starts January 1963 and ends December 2016. See Stambaugh and Yuan (2017) for details on the factor construction.

We depart from the BAB factor construction of Frazzini and Pedersen (2014). We use the value-weighted beta- and size-sorted portfolios on Kenneth French's webpage as the building blocks for constructing this factor, following Fama and French (2016) and Novy-Marx and Velikov (2016). Specifically, we construct four value-weighted portfolios: (1) small size, low beta, (2) small size, high beta, (3) big size, low beta, and (4) big size, high beta. The size cutoffs are the 40th and 60th NYSE percentiles. For betas, we use the 20th and 80th NYSE percentiles. Denote these returns as $R_{s\ell}, R_{sh}, R_{b\ell}, R_{bh}$, respectively, where s denotes small size, ℓ denotes low beta, b denotes big size, and h denotes high beta. We also compute the prior beta for each of the four portfolios and shrink towards 1 with a value of 0.5 on the historical estimate. We denote these as $\beta_{s\ell,t}$, $\beta_{b\ell,t}$, $\beta_{sh,t}$, and $\beta_{bh,t}$. We construct these portfolios using the 25 size and market beta sorted portfolio returns, as well as the corresponding market values and 60-month historical betas, given on Kenneth French's webpage.

The factor return is then constructed as follows:

$$\begin{aligned} BAB_{t+1} &= \frac{1}{\beta_{\ell,t}}\left(\frac{1}{2}R_{s\ell,t+1}+\frac{1}{2}R_{b\ell,t+1}-R_{f,t,t+1}\right) \\ &- \frac{1}{\beta_{h,t}}\left(\frac{1}{2}R_{sh,t+1}+\frac{1}{2}R_{bh,t+1}-R_{f,t,t+1}\right), \end{aligned} \tag{11}$$

where $\beta_{\ell,t} = \frac{1}{2}\beta_{s\ell,t} + \frac{1}{2}\beta_{b\ell,t}$, and $\beta_{h,t} = \frac{1}{2}\beta_{sh,t} + \frac{1}{2}\beta_{bh,t}$. As a result, the conditional market beta of BAB is close to zero, as in Frazzini and Pedersen (2014).

Finally, we get the monthly risk-free rate from CRSP and create the real risk-free rate by subtracting realized monthly inflation from the nominal rate. The inflation data are from CRSP as well.

3.3 Initial evidence of factor dynamics

As we emphasized earlier, Equation (7) implies that b is connected to the dynamic properties of the model factors. A benchmark case is when factor returns are i.i.d. and the risk-free rate is constant. In that case, the solution to portfolio allocation and savings problems are unrelated to the investment horizon, the constant coefficient SDF prices the assets, and there is nothing to learn from considering multi-horizon factor returns in addition to the usual one-period returns.

However, as we show here, the factor return dynamics of the anomaly factors are strikingly far from i.i.d. – much more so than what is the case for the market factor. As prima facie evidence of these dynamic properties, we report, in Figure 1, Panels A-C, cumulative autocorrelations for the factors appearing in the considered models.

The market factor exhibits slightly negative autocorrelation at longer horizons, which the literature attributes to a mean-reverting conditional market risk premium. In contrast, most of the anomaly factors exhibit strong positive autocorrelation that is increasing with horizon. Further, the cumulative anomaly autocorrelations in many cases increase quickly the during first 12 months, in contrast to the market autocorrelations which hover around zero and become negative only at the 20-month horizon.

The factor returns are excess returns. Return autocorrelations are therefore driven by persistent variation in the conditional factor risk premiums. A take-away from these plots is that the dynamics of the conditional anomaly factor risk premiums appear to be quite different from that of the conditional market risk premium.

As another illustration of the factor dynamics, Panel D of Figure 1 shows variance ratios of the log returns to the in-sample MVE portfolio of each of the models we consider. Recall that the excess portfolio return defined by

$$\widetilde{F}_{t,t+1} = b^\top F_{t,t+1} \tag{12}$$

is the unconditionally MVE portfolio. The corresponding simple return is

$$R_{t,t+1} = R_{f,t,t+1} + \widetilde{F}_{t,t+1}.$$

We consider horizons from 1 to 60 months. Recall that the variance ratio at horizon h is:[4]

$$VR(h) = \frac{V(\log R_{t,t+h})}{h \times V(\log R_{t,t+1})}.$$

As is well known, if $\log R_{t,t+1}$ are uncorrelated over time, then $VR(h) = 1$ for any h. If log returns are positively autocorrelated, then $VR(h) > 1$, whereas if they are negatively autocorrelated, then $VR(h) < 1$.

The variance ratios measure how the variance of (log) returns scale over time relative to the i.i.d. case – an informative metric for long-horizon investors that is exposed to longer-run return variance. The plots are striking. The market factor again exhibits a familiar moderate pattern of first slightly increasing and then decreasing variance ratios with horizon. For most of the multi-factor models, however, that is not the case. In fact, several models have variance ratios in the 1.7 to 2.5 range at horizons of 2 to 5 years. That is, an investor that holds these "MVE" portfolios are exposed to about twice as high variance at the longer-horizons than at the short (e.g. 1-month) horizons.

These strong non-i.i.d. dynamics implies that the investment horizon becomes important for portfolio choice and that Equation (7) may not hold in the case of constant b, unless expected factor returns line up with conditional variance of factor returns in exactly the right way. The question we answer next is whether the presented evidence is in fact associated with statistically and economically significant pricing errors in the linear SDF. To that effect, we apply the MHR-based test that we have developed in section 2. We will be able to compute pricing errors associated with the models and assess their statistical magnitude.

[4]Variance ratios have been used earlier to assess multi-horizon Euler equation errors (see, e.g., Flood, Hodrick, and Kaplan, 1994).

3.4 Results

3.4.1 Model specification tests

We start by computing pricing errors for the models estimated using only SHR. Thus, these models match the one-period risk-free rate and factor returns perfectly. The horizons for reported errors range from 1 to 24 months. The errors that should be understood as the present value of a zero-investment strategy with a notional of \$1 for each horizon. Figure 2 displays the results.

The main takeaway is that the pricing errors for many factors increase with horizon and become large. For instance, the error is 40% at 24 months for the Stambaugh-Yuan performance (PERF) factor, 20% at the same horizon for the BAB factor and the DMRS version of the SMB factor. For comparison, the factor risk premiums are broadly around 5% per year, so these largest 2-year pricing errors are about 2 to 4 times bigger than the anomaly risk premiums we are trying to explain.

We formally evaluate all the models of the form outlined in Equation (10) in two different ways, using SHR and MHR for estimation as described in section 2.4 with one minor modification. Since the factor models we consider are developed to price excess returns, we formulate the MHR-based moments as excess return moments. In particular, the GMM moment vector is:

$$f_t = \begin{pmatrix} M_{t-1,t} R_{f,t-1,t} - 1 \\ M_{t-h_1,t}(R_{t-h_1,t} - R_{f,t-h_1,t}) \\ \dots \\ M_{t-h_n,t}(R_{t-h_n,t} - R_{f,t-h_n,t}) \end{pmatrix}. \tag{13}$$

In our actual tests, we again apply the somewhat more involved covariance representation from Equation (3) to Equation (13) to obtain MHR-based moments that are not autocorrelated under the null.

We choose horizons $h_j \in \{1, 2, 3, 6, 12, 24\}$ months for $j = 1, 2, ..., 6$. Recall, $R_{f,t-h_j,t}$ is the gross risk-free rate realized over the period from $t - h_j$ to t, while $R_{t-h_j,t}$ is the vector of gross factor returns realized over the period from $t - h_j$ to t. Thus, for

a K-dimensional factor model, we have $K \cdot n + 1$ moments, where the number of horizons in our case is $n = 6$.

Figure 3 displays the testing results. Panel (A) shows $p-$values indicating rejection of all models with the exception of the MKT model (the unconditional CAPM) for both testing approaches. We would like to emphasize the meaning of failure to reject in this context. The MKT factor is capable of pricing itself at multiple horizons. That does not imply, however, that the MKT model is well-specified. As we know, it is easily rejected by a cross-section of equity returns. The failure to reject tells us that the conditional expectation and variance of the MKT factor line up sufficiently closely with each other so that the coefficient b in the SDF is close to a constant, per equation (7).

3.4.2 Economic magnitudes

To gain economic intuition of the rejections of the multi-factor models, we also report annualized mean absolute pricing errors (MAPE) in Panel (B). MAPE should be understood as the present value of a zero-investment strategy with a notional of \$1. MAPE is computed in two steps. First, because our tests assets span different horizons, we annualize each horizon's pricing error via geometric compounding. Second, we report MAPE as the average of the absolute values of the annualized pricing errors for the MHR-based moments.

In the case of SHR-based estimation, the MAPE for rejected models range between 1% and 3.5%. This is of the same order of economic importance as factor risk premiums in these models. The MAPE decline substantially when MHR are used for estimation. That is natural because of the change in the objective function. The models change the intercept and price of risk parameters in the SDF to price long-horizon returns better. This by itself indicates MHR returns are informative for estimation of a given SDF. The better fit across horizons means we have to give up on zero pricing errors at the one-period horizon. MAPE now range from 0.3% to 1.75%. These numbers are still economically large as factor returns across horizons are positively correlated. Thus, a substantial amount of the variation in a multi-horizon return strategy can

be hedged with a static position in the single-horizon returns. A horizon-related mispricing with annualized MAPE of, say, 1% of can be evidence of large unexplained SRs (typically referred to as Information Ratios). We discuss this below in the context of standard GRS "alpha" regressions.

Before that, we note that the one-period maximal SR – a dimension along which the candidate models are optimized – appears to be *positively* related to MAPE of longer-horizon factor excess returns. We demonstrate this observation by constructing a scatter plot of MAPE versus maximum SR in Figure 4. While there is no ex-ante reason for this relationship to hold, it is interesting that the pursuit of the maximum SR, or equivalently, the MVE portfolio, takes us further away from a constant b. One might have thought the opposite would be the case given that the factors in Equation (10) span the unconditional MVE portfolio with a constant b in the SDF specification.

3.4.3 GRS test using horizon-managed portfolios

An alternative to the GMM-tests presented so far is the GRS test of whether a given set of one-period factor returns span the mean-variance frontier of a set of one-period test asset returns. This test is not a test of whether the model can price MHR, but it is useful to also couch our results in the typical single-horizon setting most studies apply.

Furthermore, this test serves an additional purpose as a robustness exercise. Because we use the one-period risk-free rate to construct MHR, one could be concerned that model rejections are due the misspecification of our model of the SDF in (10) with respect to the conditional risk-free rate. Indeed, as we discussed, b_t that is proportional to a_t corresponds to the unconditional MVE portfolio. So far, we have discussed a particular case of that when both a and b are constant. Because the GRS test is implemented using single-horizon excess returns, its results are mechanically unrelated to the properties of the risk-free rate.

For our purposes, the test assets are trading strategies in the factors themselves. In particular, the test asset $i = (k, h)$ is a strategy in factor k, where the time-varying

weights are given by lagged h-period discounted returns in factor k, i.e.,

$$R_{i,t,t+1} - R_{f,t,t+1} \equiv z_{k,t}^{(h)} F_{k,t,t+1},$$

where $z_{k,t}^{(h)} = M_{t-h,t} R_{k,t-h,t}$ and $R_{k,t-h,t} = \prod_{s=1}^{h} (R_{f,t+s-1,t+s} + F_{k,t+s-1,t+s})$ is the gross return on factor k from $t-h$ to t.[5] The M used in construction of the instrument $z_{k,t}^{(h)}$ is $M_{t,t+1} = a - b^\top F_{t,t+1}$, where a and b are estimated by requiring the model to match the sample means of risk-free rate and the factors.

Then we run the time-series regression

$$R_{i,t,t+1} - R_{f,t,t+1} = \alpha_i + \beta_i^\top F_{t,t+1} + \varepsilon_{i,t+1}$$

for each test asset i and report the standard joint (GRS) test that $\alpha_i = 0$ for all i.

Because the instruments $z_{k,t}^{(h)}$ are constructed using the candidate SDF, which in turn depends on the parameter estimates of a and b, we consider both in- and out-of-sample instruments. The former use the entire sample to estimate the SDF, whereas the latter use information only up to time t. We then test whether the factors F span the mean-variance frontier of these managed portfolios.

Panels A and B of Table 1 report the results of GRS tests on the classic and recent factor models using lagged pricing errors, $z_{k,t}^{(h)}$, for horizons 1, 2, 5, 11, and 23 months for each of the factors, i.e., five managed portfolios for each factor. The motivation for these horizons is to, when multiplying with the current factor return the total horizons are the same as before (2, 3, 6, 12, and 24). Similar to the case for the GMM test, it is clear that a conditional version of the factor models would price the horizon-managed portfolios. However, if the factors span the unconditional mean-variance efficient portfolio, they should also account for the managed factor portfolios in an unconditional test.

As the table shows, most models are strongly rejected regardless of whether one uses the in-sample instruments (see p-value GRS test) or out-of-sample instruments (p-value, out-of-sample instruments). The exceptions are the market model and

[5] The weights in a factor are unrestricted since the factors are excess returns.

$FF5_{VolMan}$. The former is rejected at the 5%-level based on the in-sample instruments, but not rejected based on the out of sample instruments. The $FF5_{VolMan}$ model is not rejected based on either in-sample or out-of-sample instrument p-values when the managed portfolios are based on discounted returns to the volatility managed FF5-factors. If the managed portfolios are based on the original FF5-factors, however, the model is rejected based on the in-sample instruments. The p-value is marginally insignificant at 0.11 for the out-of-sample instrument case. Overall, the findings are in line with the GMM tests in Figure 3.

Also reported are unbiased estimates of the annualized maximum information ratios (IR) attainable using the horizon-managed factor returns. Recall, the squared monthly information ratio can be obtained from the regressions as $\alpha^\top \Sigma^{-1} \alpha$, where Σ is the covariance matrix of the regression residuals. MacKinlay (1995) provides a small-sample adjustment to arrive at an unbiased estimate of this quantity. We report the maximum of this quantity and zero. To get a sense of magnitudes, it is natural to compare the maximal IR to the maximal SR combination of the factors. In FF5, the maximum IR is 0.57 compared to a SR of the factors of 1.09. For the other models, with the exception of the $FF5_{VolMan}$, the maximum IR tends to be similarly large relative to the SR of the corresponding factors. In other words, the GRS tests show that there is substantial scope for market timing of the factors, providing an easy-to-interpret measure of the economic significance of the model rejections.

4 Dynamics implicit in a model specification

The rejection of the models is a consequence of factor dynamics unaccounted for in the linear SDF specification. In this section, we consider these dynamics. Full accounting for the uncovered role of dynamics and proposing a convincing alternative to the models is beyond the scope of this paper. We have a more modest objective of providing an illustration of what accounting for these dynamics might entail and to suggest a path for future research.

4.1 Construction of b_t

The rejection of the model (10) prompts us to consider the alternative SDF in Equation (5). Due to our choice of test assets (MHR to the factors themselves), there exists an SDF of the form in Equation (5) that prices all test assets at all horizons. To illustrate how much variation in a_t and b_t we are missing, we use MHR to estimate the SDF.

In theory, Equation (7) offers a direct way to construct b_t. The elements in the denominator are easy to obtain. $R_{f,t,t+1}$ is observed. Various estimators of conditional variance are typically similar to each other. In contrast, obtaining conditional expectations of factors is very hard. A large literature is debating the mere presence of predictability, and predictors are a subject of an active debate. Because the denominator could be viewed as observable, at least for our illustrative purposes, we can measure something close to b_t by replacing conditional expectation in the numerator with the actual factor. We exploit the relation between this "observed" b_t and true b_t to estimate the latter.

We construct an estimate of b_t in four steps. First, given the illustrative nature of the exercise we posit a one-factor structure for b_t regardless of a considered model. Second, we simplify the complicated objective of analyzing multiple factors. In order to do that, we construct in-sample unconditional MVE portfolio as in Equation (12). Third, we construct "observed" b_t :

$$\tilde{b}_{t+1} = [R_{f,t,t+1} V_t(\tilde{F}_{t,t+1})]^{-1} \tilde{F}_{t,t+1},$$

where the conditional variance of $\tilde{F}_{t,t+1}$ is estimated by EGARCH(1,1) over the full sample. Equation (7) implies that $E_t(\tilde{b}_{t+1}) = b_t$. Fourth, we use that relationship to estimate a specific proposed functional form of b_t using the MHR-based moment conditions in (13).

The specific functional form of b_t that we entertain is:

$$b_t = c_0 + \sum_{j=1}^{12} c_j \tilde{b}_{t-j+1} + c_{13} [R_{f,t,t+1} V_t(\tilde{F}_{t,t+1})]^{-1},$$

where, for parsimony, $c_3 = \ldots = c_{12}$. Thus, given that $\tilde{b}_{t+1} = b_t + e_{t+1}$, we effectively assume $\tilde{b}_t$ to be a (restricted) AR(12) process, which is designed to capture its fast and slow moving components. The last term with a loading c_{13} reflects the theoretical relation between b_t, the risk-free rate, and the conditional factor variance.

Further, combine Equation (6) with Equation (7) to get:

$$a_t = R^{-1}_{f,t,t+1} + b_t^2 R_{f,t,t+1} V_t(\tilde{F}_{t,t+1}).$$

This completes the specification of the alternative SDF:

$$M_{t,t+1} = a_t - b_t \tilde{F}_{t,t+1}.$$

4.2 Estimation results

Figure 5 displays GMM estimation results. As a reference point, we juxtapose them with MHR-based estimation for the case of constant b from section 3.4. Panel A shows that we fail to reject the models. FF5_{DMRS} is one exception as it is marginally rejected at the 10% level.

The pricing errors on Panel B drop across the board relative to the case of constant b, sometimes six-fold (MKT+BAB), sometimes just 1.5-fold (FF5_{DMRS}). There are two exceptions: FF5, where the two MAPEs are nearly identical, and FF3, where the constant b MAPE is two times smaller.

There could be at least two reasons for some inconsistencies between the two panels. First, MAPE is different from GMM's objective function, so failure to reject does not guarantee a drop in the pricing error and vice versa. Second, our illustrative specification of b_t could be more misspecified for some models than for others.

We further demonstrate economic magnitude of the improvement by displaying maximal SR's for SHR in Figure 6. Note that we were not targeting SR's during estimation. In particular, we provide the annualized Sharpe ratio for each model's in-sample MVE portfolio with constant b, $b^\top F_{t+1}$, versus the annualized Sharpe ratio of the timed versions of the same MVE portfolio with monthly returns $(b_t/a_t) \cdot b^\top F_{t+1}$, as estimated

using our MHR-based moments. The Sharpe ratios increase as we switch from constant to time-varying SDF coefficients. Thus, we see benefits to factor timing even in the monthly return case. The magnitude of the improvement is reflected in annualized Information Ratios of a regression of the timed MVE portfolio returns on the static that averages 0.64 and ranges between 0.3 and 0.9 per year.

4.3 Sources of variation in b_t

The estimation procedure delivers nine different b_t's, one for each model. That gives us an opportunity to study possible drivers behind the variation in b. Analyzing nine different series is cumbersome, and we are looking for big picture implications rather than the details of a specific model. That objective calls for studying common variation in b_t's. These are not easy objects to study because they reflect conditional variance of factors, which could vary between factors within and across models. Thus, the more appropriate objects to study would be market prices of risk, λ_t.

Equation (9) gives the relation between b_t and λ_t as:

$$b_t = \lambda_t^\top V_t^{-1/2}(\widetilde{F}_{t,t+1}). \tag{14}$$

Thus, we scale the estimated b_t by the respective conditional volatilities of the MVE portfolios.

Next, we study common variation in λ_t's via principal component analysis. The first two principal components explain 45% and 19% of variation in λ_t, respectively. We focus on the properties of these two components for simplicity. The first principal component loads positively on all models' λ_t and has correlation of 0.96 with the average λ_t across models, so we can interpret its sign accordingly. The average loading on PC1 is 0.26. The annualized range of this average price of risk is from about 0 to 1.5.

Figure 7 displays the time series of PC1 and PC2. Both variables appear to be related to business cycles. This prompts the question if we could relate the dynamics to standard observable instruments used in the literature.

Figure 8 displays the results of regressing of the PCs on a recession indicator, the BAB beta spread, the conditional variance of the market (estimated by EGARCH(1,1)), the log market dividend to price ratio, and the 10 year over 1 year term spread. The beta spread is the spread between the high and low market-beta sorted portfolios in the BAB factor, as described earlier. This variable is included both as an example of a variable that captures a degree of cross-sectional characteristic heterogeneity.

For PC1, which sign we can easily interpret, the price of risk is significantly negatively related to the NBER recession indicator and market variance. This is striking and unexpected. We would venture to say most market participants view recessions and spikes in volatility as bad times, but this is not reflected in the prices of risk of these factor models. Either the factor models are missing an important component of risk or investor expectations are not rational and the price of risk dynamics we estimate in the data reflects investor expectational errors in these bad times. For instance, investors could be persistently surprised by the severity of recessions and/or volatility spikes, leading to lower expected returns in bad times as prices initially under-react.

More reassuringly, the dividend-price ratio and the term spread are positively associated with fluctuations in PC1, as theory and intuition suggests should be the case. The BAB beta spread is positively associated with the price of risk, indicating that cross-sectional heterogeneity reflects factor risk premium dynamics. PC2 also load significantly on the NBER recession indicator and conditional market return variance, but is unrelated to the other variables.

The overall R^2's are 11.5% and 17.5% for PC1 and PC2, respectively. Thus, the lion's share of the estimated variation in the prices of risk in these models is unexplained. This evidence leaves us with a question of the driving forces behind the variability in λ_t, and, therefore, b_t.

5 Conclusion

We have developed a GMM-based test that uses multi-horizon returns (MHR) to evaluate asset pricing models. We apply this test to a set of nine linear factor models. All

of the multi-factor models are rejected with high statistical and economic significance. The reason the models are rejected is that the true dynamics of the factors (excess returns on portfolios of assets) are quite different from the ones implicitly prescribed by the null hypothesis.

That MHR are informative about these dynamics goes at the heart of the optimal multi-horizon asset allocation. In that literature differences in horizons matters only if returns are non-i.i.d. We use our framework to estimate how the factors in the candidate models should be timed to be consistent with the observed MHR. We find that the variation has to be substantial and is not strongly connected to existing conditional variables in obvious ways. This sort of variation is puzzling. Future research should focus on understanding the economic forces behind the evidence.

References

Backus, David, Mikhail Chernov, and Stanley Zin, 2014, Sources of entropy in representative agent models, *Journal of Finance* 69, 51–99.

Belo, Frederico, Pierre Collin-Dufresne, and Robert Goldstein, 2015, Dividend dynamics and the term structure of dividend strips, *Journal of Finance* 70, 1115–1160.

Brennan, Michael, Eduardo Schwartz, and Roland Lagnado, 1997, Strategic asset allocation, *Journal of Economic Dynamics and Control* 21, 1377–1403.

Brennan, Michael, and Yuzhao Zhang, 2018, Capital asset pricing with a stochastic horizon, working paper.

Campbell, John, and Luis Viceira, 1999, Consumption and portfolio decisions when expected returns are time varying, *Quarterly Journal of Economics* 130, 433–495.

Campbell, John Y., 2001, Why long horizons? A study of power against persistent alternatives, *Journal of Empirical Finance*.

Carhart, Mark, 1997, On persistence in mutual fund performance, *Journal of Finance* 52, 57–82.

Cochrane, John, 2004, *Asset Pricing* (Princeton University Press: Princeton, NJ) revised edn.

———, 2014, A mean-variance benchmark for intertemporal portfolio theory, *Journal of Finance* 69, 1–49.

Cooper, Michael, Huseyn Gulen, and Michael Schill, 2008, Asset growth and the cross-section of stock returns, *Journal of Finance* 63, 1609–1651.

Dahlquist, Magnus, and Henrik Hasseltoft, 2013, International bond risk premia, *Journal of International Economics* 90, 17–32.

Daniel, Kent, and David Marshall, 1997, Equity-premium and risk-free-rate puzzles at long horizons, *Macroeconomic Dynamics*.

Daniel, Kent, Lira Mota, Simon Rottke, and Tano Santos, 2018, The cross section of risk and return, working paper.

Dew-Becker, Ian, Stefano Giglio, Anh Le, and Marius Rodriguez, 2015, The price of variance risk, working paper.

Fama, Eugene, and Kenneth French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

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

Farnsworth, Heber, Wayne Ferson, David Jackson, and Steven Todd, 2002, Performance evaluation with stochastic discount factors, *Journal of Business* 75, 473–503.

Ferson, Wayne E., and Campbell R. Harvey, 1999, Conditioning variables and the cross-section of stock returns, *Journal of Finance* 54, 1325–1360.

Flood, Robert P., Robert J. Hodrick, and Paul Kaplan, 1994, An evaluation of recent evidence on stock price bubbles, in Robert P. Flood, and Peter M. Garber, ed.: *Speculative Bubbles, Speculative Attacks and Policy Switching* (MIT Press).

Frazzini, Andrea, and Lasse Pedersen, 2014, Betting against beta, *Journal of Financial Economics* 111, 1–25.

Gibbons, Michael, Stephen A. Ross, and Jay Shanken, 1989, A test of the efficiency of a given portfolio, *Econometrica* 59, 1121–1152.

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

Hansen, Lars, and Scott Richard, 1987, The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models, *Econometrica* 55, 587–613.

Hansen, Lars Peter, and Jose Scheinkman, 2009, Long term risk: an operator approach, *Econometrica* 77, 177–234.

Jagannathan, Ravi, and Yong Wang, 2007, Lazy investors, discretionary consumption, and the cross-section of stock returns, *Journal of Finance* 62, 1623–1661.

Jagannathan, Ravi, and Zhenyu Wang, 1996, The conditional capm and the cross-section of expected returns, *Journal of Finance* 51, 3–53.

Jegadeesh, Narasimhan, and Sheridan Titman, 1993, Returns to buying winners and selling losers: Implications for stock market efficiency, *Journal of Finance* 48, 65–91.

Jensen, Michael, Fischer Black, and Myron Scholes, 1972, The capital asset pricing model: Some empirical tests, in Michael Jensen, ed.: *Studies in the Theory of Capital Markets* (Praeger Publishers Inc.).

Kamara, Avraham, Robert A. Korajczyk, Xiaoxia Lou, and Ronnie Sadka, 2016, Horizon pricing, *Journal of Financial and Quantitative Analysis* 51, 1769–1793.

Lettau, Martin, and Sydney Ludvigson, 2001, Resurrecting the (c)capm: A cross-sectional test when risk premia are time-varying, *Journal of Political Economy* 109, 1238–1287.

Lustig, Hanno, Andreas Stathopoulos, and Adrien Verdelhan, 2013, The term structure of currency carry trade risk premia, working paper.

MacKinlay, A. Craig, 1995, Multifactor models do not explain deviations from the capm, *Journal of Financial Economics* 38, 3–28.

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

Novy-Marx, Robert, 2013, The other side of value: The gross profitability premium, *Journal of Financial Economics* 108, 1–28.

——— , and Mihail Velikov, 2016, A taxonomy of anomalies and their trading costs, *The Review of Financial Studies* 29, 104–147.

Parker, Jonathan, and Christian Julliard, 2005, Consumption risk and the cross-section of expected returns, *Journal of Political Economy* 113, 185–222.

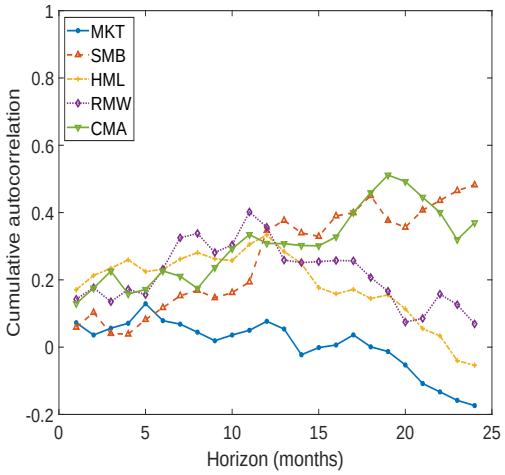
Stambaugh, Robert, and Yu Yuan, 2017, Mispricing factors, *The Review of Financial Studies* 30, 1270–1315.

van Binsbergen, Jules, Michael Brandt, and Ralph Koijen, 2012, On the timing and pricing of dividends, *American Economic Review* 102, 1596–1618.

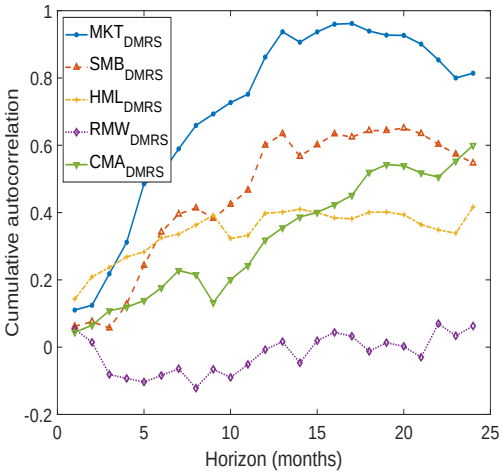
Zviadadze, Irina, 2017, Term-structure of consumption risk premia in the cross-section of currency returns, *Journal of Finance* 72, 1529–1566.

Figure 1
Cumulative autocorrelations and variance ratios of factors

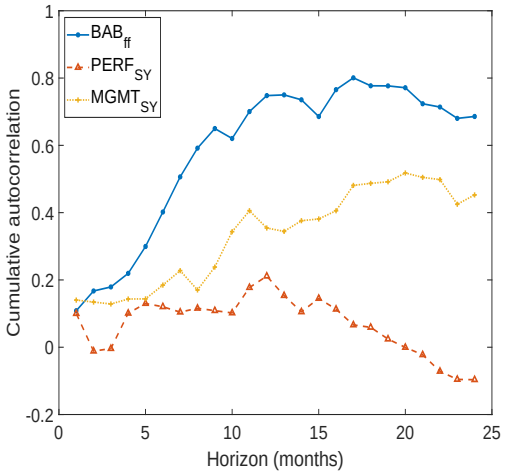
(A) Fama-French factors



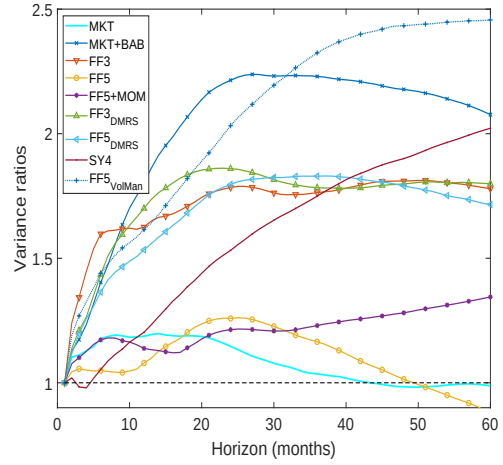
(B) DMRS factors



(C) Stambaugh-Yuan factors

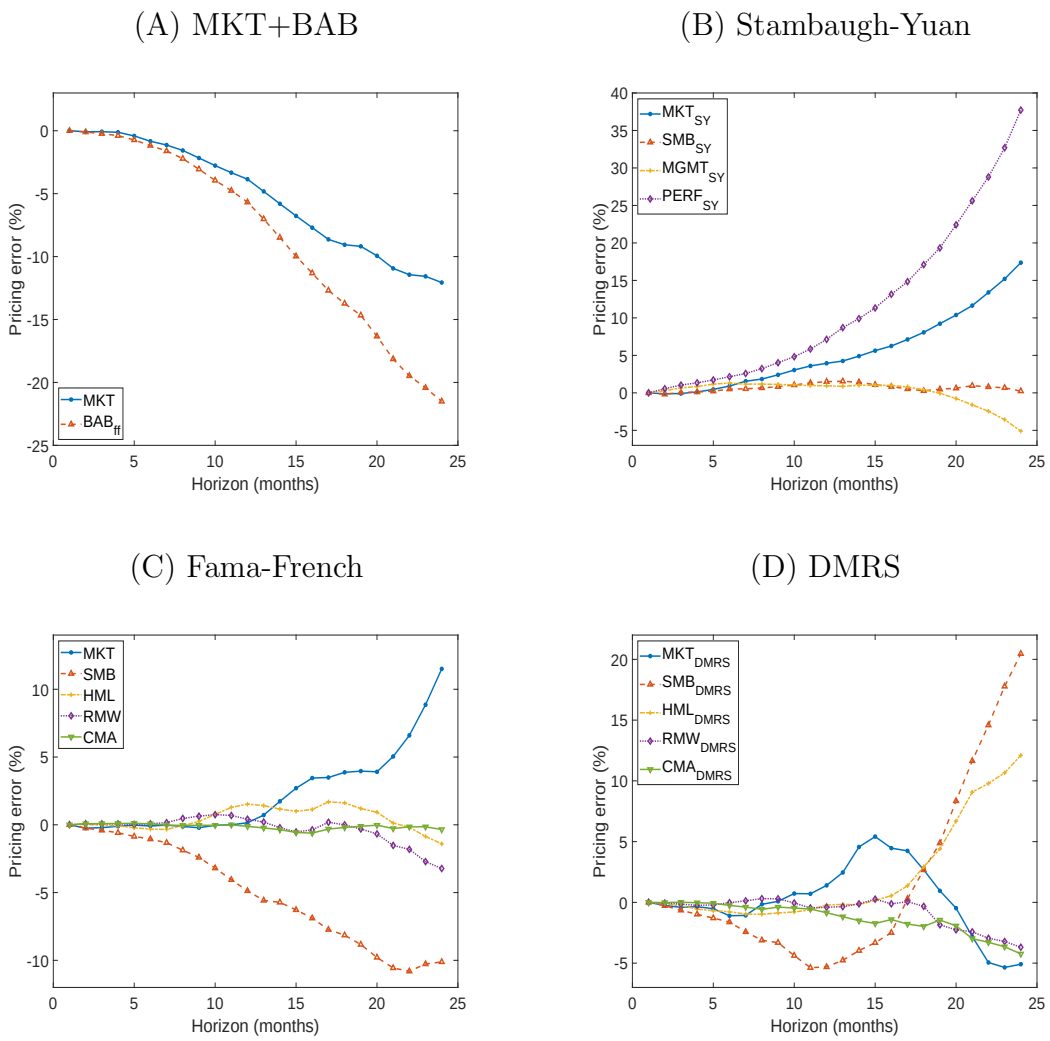


(D) Variance ratios



Panels A, B, and C show the cumulative autocorrelation of the factors in the FF5, FF5$_{DMRS}$, and SY models, respectively, for horizons of 1 to 24 months. See the main text for a description of the individual factors. The sample is monthly, from 1963 to 2017. Panel D shows the variance ratio of the in-sample MVE combination of the factors in each of the models we consider, for horizons from 1 to 60 months. No autocorrelation corresponds to a variance ratio of 1 at all horizons.

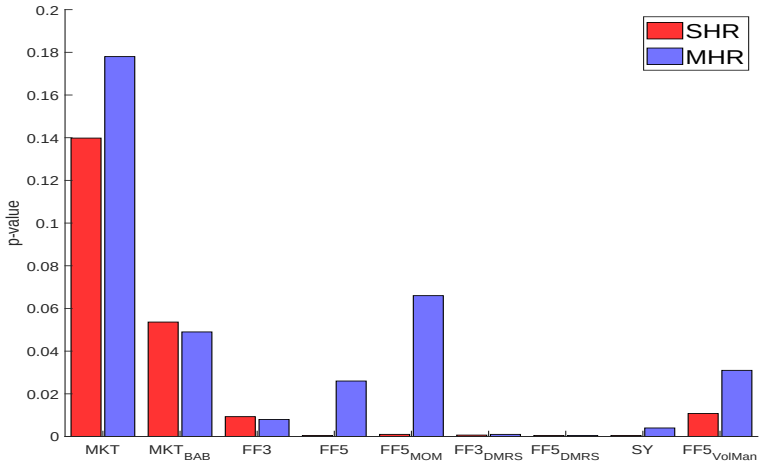
Figure 2
Term structure of pricing errors, $E(M \cdot (R - R_f))$



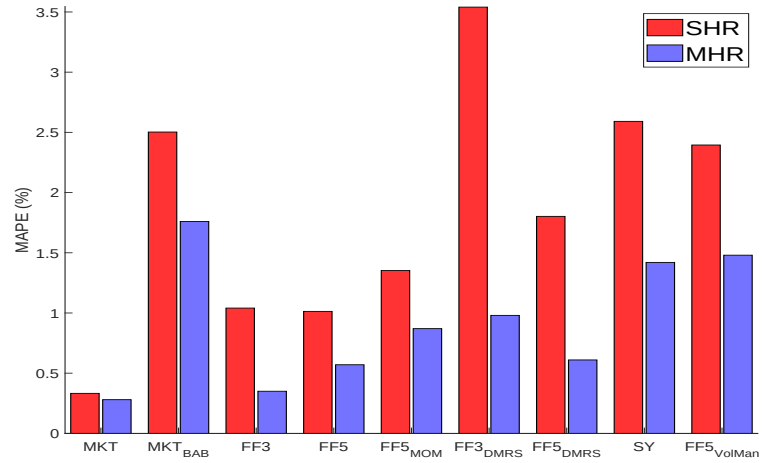
The panels show factor pricing errors for four factor models for horizons from 1 to 24 months. Annualized pricing errors at horizon h are $(1 + E_T(M_{t,t+h}(R_{t,t+h} - R_{f,t,t+h})))^{12/h} - 1$, where E_T denotes the sample average, $R_{t,t+h}$ is the gross factor return from month t to $t + h$, and $R_{f,t,t+h}$ is gross h-period return from rolling over the riskless bond. The population average of a correctly specified model is zero. The sample is monthly, from 1963 to 2017.

Figure 3
GMM tests of linear factor models

(A) $p-$values

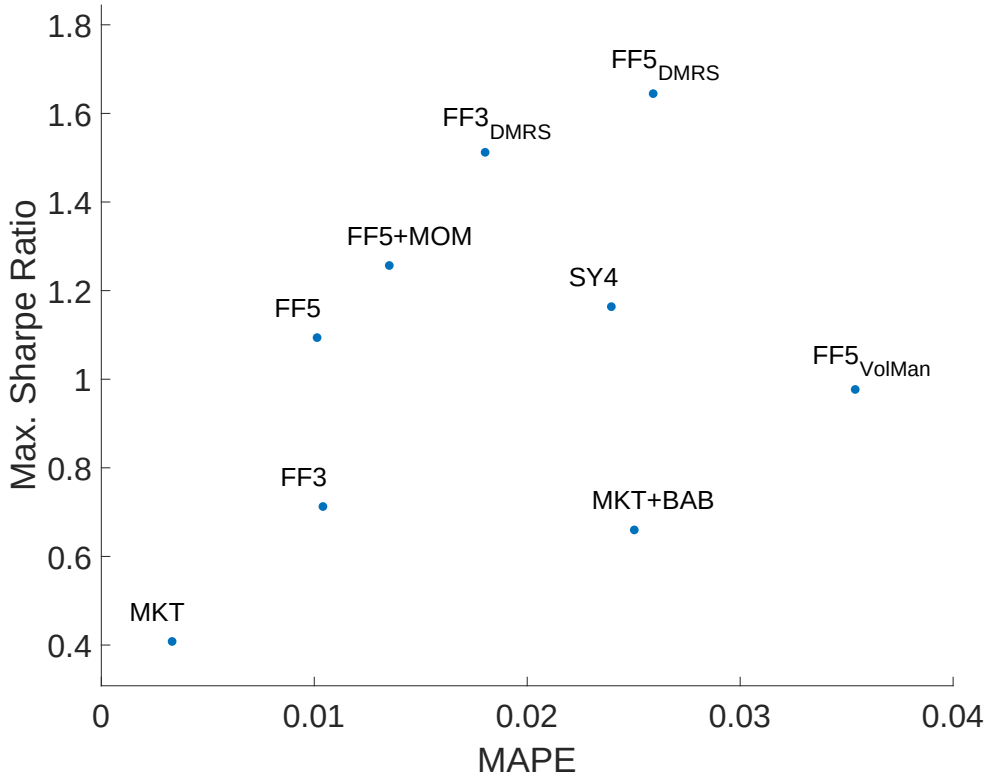


(B) MAPE



Panel A gives bootstrapped model p-values from one- and two-stage GMM estimation of each models' SDF. One-stage estimation obtains the SDF parameters by fitting average one-period factor returns and risk-free rate. Two-stage GMM in addition uses the MHR-based moments to estimate parameters. The test assets are excess factor returns at the 1, 2, 3, 6, 12, and 24-month horizons, as well as the short-term real risk-free rate. Panel B shows MAPE – the annualized mean absolute pricing error for the multi-horizon excess factor returns. See the main text for a description of each of the models. The sample is monthly, from 1963 to 2017.

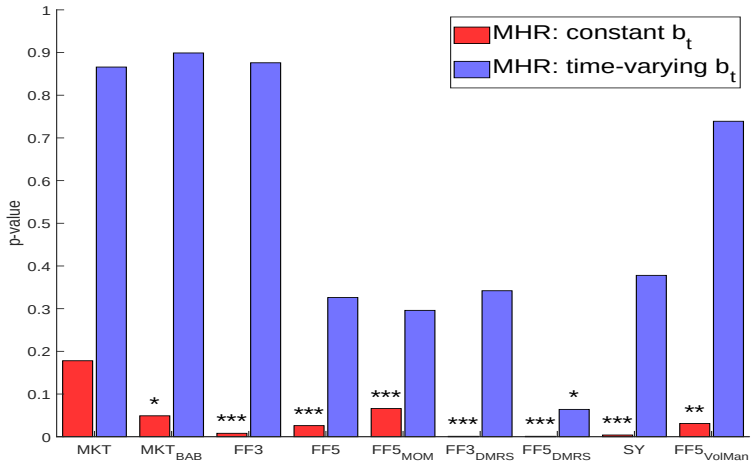
Figure 4
Max Sharpe ratio of single-horizon factor model vs. multi-horizon pricing errors



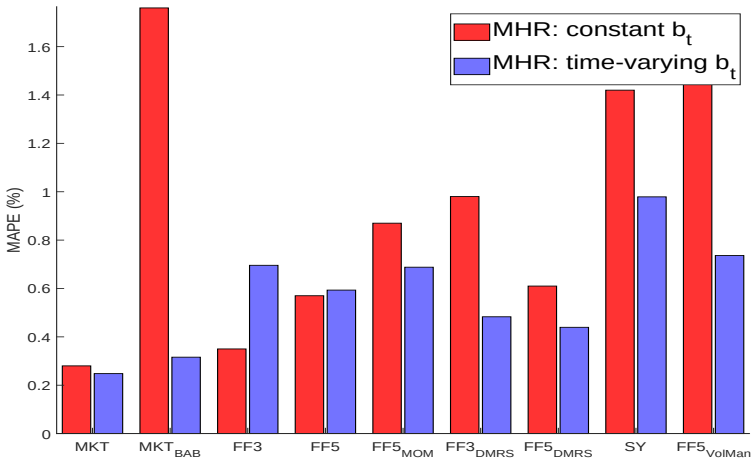
The figure plots the annualized maximal in-sample Sharpe ratio combination of the factors in each model against the annualized mean absolute pricing error (MAPE) of the corresponding model, when the model is estimated using one-period returns and tested on excess factor returns with horizons 1, 2, 3, 6, 12, and 24 months. The sample is monthly, from 1963 to 2017.

Figure 5
GMM tests of linear factor models with time-varying b

(A) $p-$values

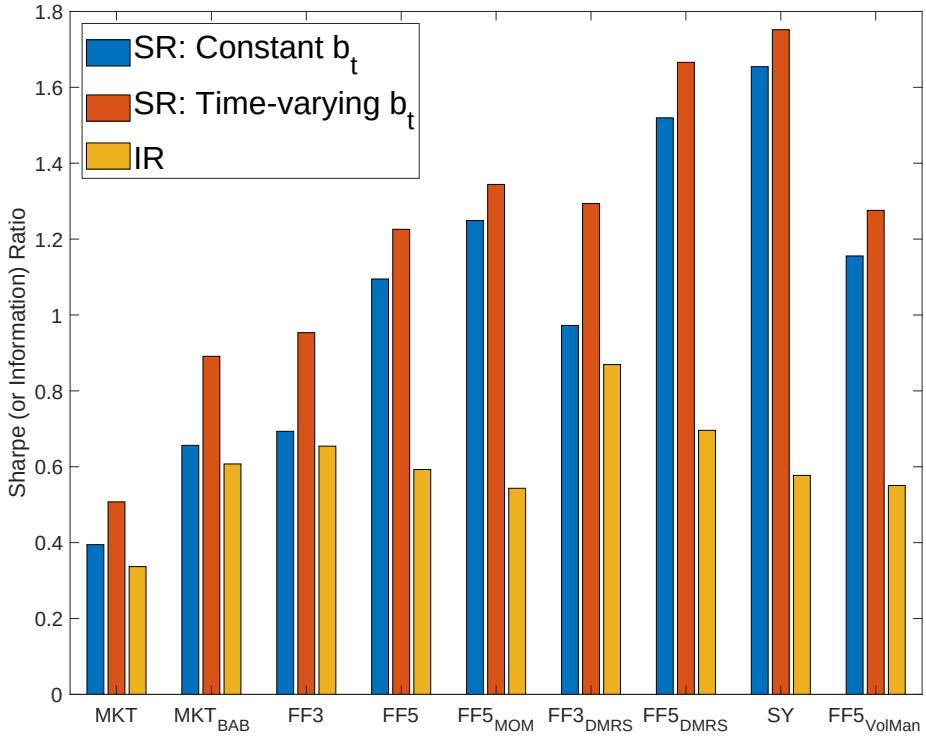


(B) MAPE



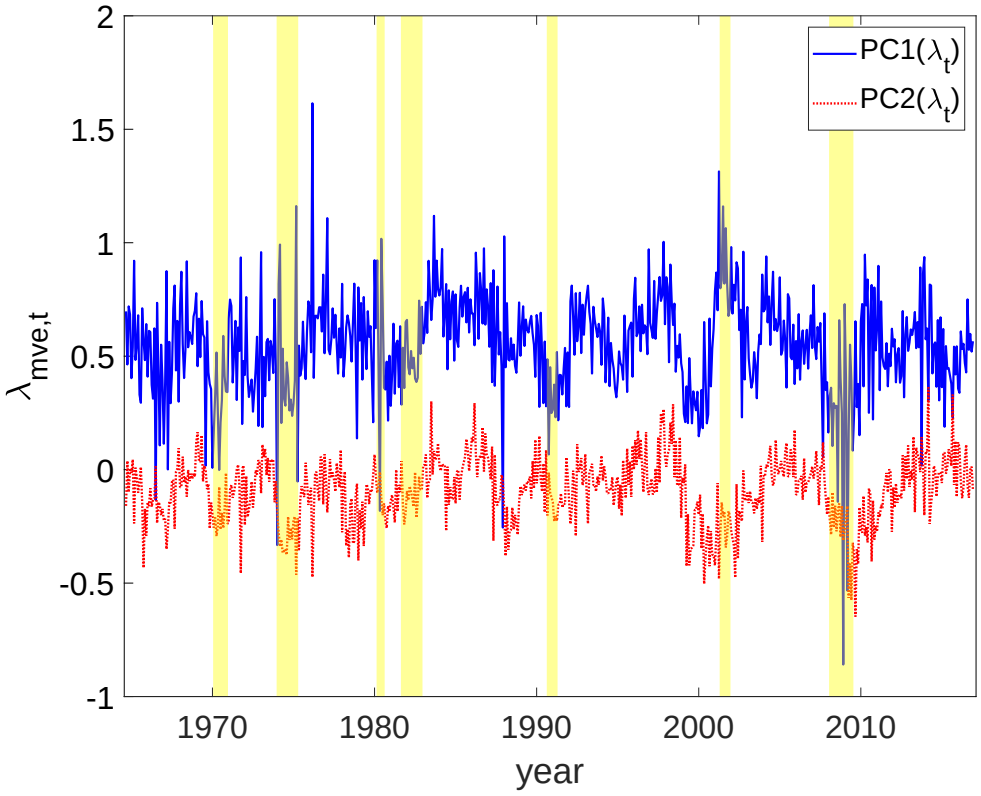
Panel A gives the bootstrapped p-values from the J-tests from two-stage GMM estimation of the SDF implied by each factor model, including multi-horizon excess factor returns in the test assets at the 1, 2, 3, 6,12, and 24-month horizons. The blue bars corresponds to the case where we have time-varying a_t and b_t in the SDF, while the red bars to the case of constant a and b. Panel B shows MAPE – the annualized mean absolute pricing error for the multi-horizon excess factor returns. See the main text for a description of each of the models. The sample is monthly, from 1963 to 2017.

Figure 6
MVE portfolio Sharpe ratios when b is constant and time-varying



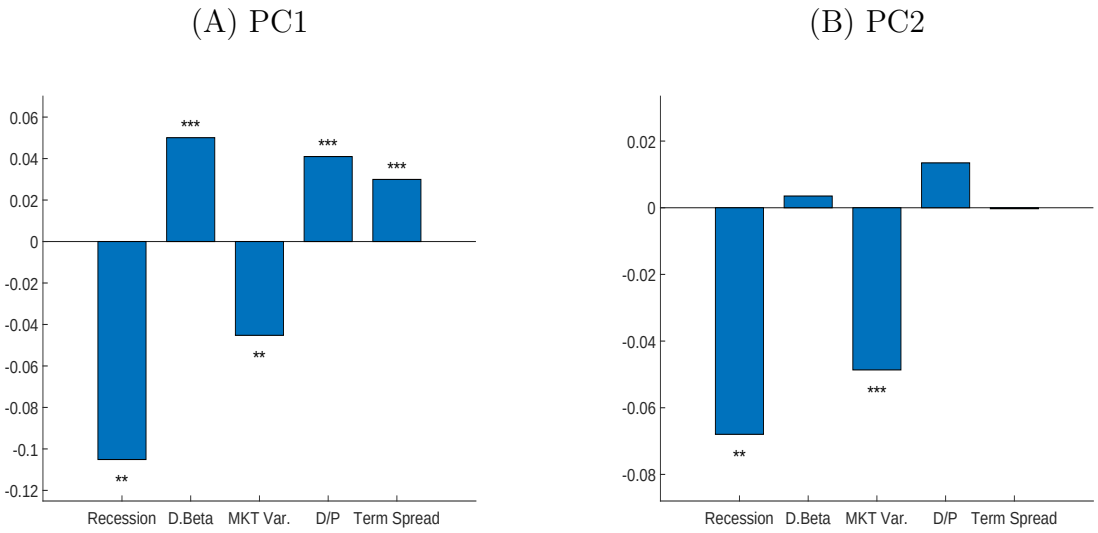
The figure plots the annualized maximal in-sample Sharpe ratio combination of the factors in each model when b is constant (blue bars) and time-varying (red bars). To assess the economic importance of the difference, we report the annualized Information ratio obtained from a regression of the MVE portfolio constructed using the b_t/a_t as the managed portfolio weight on the MVE portfolio constructed from a constant b.

Figure 7
Time-series of principal components λ_t



Panel A shows the time-series of the first and second principal components of each of the nine factor models' monthly price of risk, λ_t, along with NBER recession indicators (yellow bars). The prices of risk were estimated using multi-horizon factor excess returns as moments. Note that for each model, the price of risk is related to b_t as follows: $\lambda_t = b_t * Var_t(MVE_{t,t+1})$. The sample is from July 1963 to June 2017.

Figure 8
Dynamics of time-varying prices of risk: Relation to standard instruments



The table shows regressions of the prices of risk λ_t regressed on common conditioning variables: an NBER recession dummy, the spread in market beta for the BAB factor, the conditional variance of market returns, the log market price-dividend ratio, and the Treasury bond term spread. Standard errors are computed using the Newey-West method with 12 lags. The sample is monthly, from 1963 to 2017. The statistical significance of the parameters are given as follows: Three asterisks denote significance at the 1% level, two at the 5% level, and one at the 10% level.

Table 1: Long-run factor mispricing as monthly factor return instruments: GRS tests

	Standard Factor Models				
Panel A:	MKT	MKT+ BAB	FF3	FF5	FF5+MOM
Max Sharpe ratio of factors	0.388	0.657	0.690	1.089	1.246
Max Information ratio	0.411	0.429	0.486	0.572	0.531
p-value GRS test	0.018**	0.038**	0.032**	0.028**	0.063*
p-value, out-of-sample instruments	0.309	0.047**	0.001***	0.022**	0.035**

	Alternative Factor Models				
Panel B:	$FF3_{DMRS}$	$FF5_{DMRS}$	SY4	$FF5_{VolMan}$	(FF5 assets) $FF5_{VolMan}$
Max Sharpe ratio of factors	0.971	1.544	1.653	1.159	1.159
Max Information ratio	0.870	0.821	0.473	0.000	0.589
p-value GRS test	0.000***	0.001***	0.081*	0.790	0.023**
p-value, out-of-sample instruments	0.000***	0.002***	0.033**	0.300	0.111

The table gives the Gibbons-Ross-Shanken (GRS) test p-values from regressions of managed factor returns, where the time-varying factor weights are based on lagged multi-horizon pricing errors. The lagged horizons considered are 1, 2, 5, 11, and 23 months. For each factor model, the test assets are managed positions in the factors of the factor model at hand. So, for the market model, the test assets are four managed portfolios (corresponding to the four pricing error horizons) in the MKT factor. "Out-of-sample instruments" means the pricing errors used as instrument are estimated using only data up to time t, when considering time t+1 returns. Max. Information Ratio (IR) refers to the square root of the max expected Sharpe ratio one can achieve when combining the test assets and hedge out the unconditional factor exposures. See MacKinlay (1995) and the main text for the formula leading to an unbiased estimate of this quantity. The sample is monthly, from 1963 to 2017. Three asterisks denote significance at the 1% level, two at the 5% level, and one at the 10% level.