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: December 4, 2018

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 all state-of-the-art models that imply high maximum Sharpe ratios in a single-horizon setting. Thus, the models do a poor job in accounting for the risk-return trade-off at longer horizons. Across the different models, the mean absolute pricing errors associated with MHR are positively related to the magnitude of maximal Sharpe ratio in the single-horizon setting. Model misspecification manifests itself in strong intertemporal dynamics of the factor loadings in the SDF representation. We suggest that misspecification of the dynamics of loadings arises from the common approach towards factor construction via portfolio sorts.

JEL Classification Codes: G12, C51.

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

* We thank Valentin Haddad, Francis Longstaff, Tyler Muir, and Tommy Stamland for comments on earlier drafts and participants in the seminars and conferences sponsored by LAEF, NHH, 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

Specifying an asset-pricing model via the stochastic discount factor (SDF) is a powerful analytical tool that encodes the risk-return trade-off across multiple assets and across multiple horizons. The former set of implications has enjoyed virtually endless amount of theoretical and empirical research. The focus is on finding, at a single horizon, the maximum Sharpe ratio (SR) portfolio whose return is inversely related to the SDF.

In this paper, we argue that there are important dynamics of the SDF that substantially affects its multi-horizon empirical performance. Hence, evaluating SDFs using multi-horizon returns (MHR) should feature more prominently in model selection. We show that an MHR-based test amounts to a stringent evaluation of conditional implications of a model.

Our approach is motivated by the fact that most economic decisions, such as, corporate investments, private equity, or consumption-savings plans, involve longer horizons. For instance, Campbell and Viceira (2002), among many others, emphasize the importance of return predictability for optimal long-term portfolios. By implication, the multi-horizon SDF is an important framework for evaluating economic models.

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 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 MHR-based tests could be interpreted as adding a set of instruments to standard single-horizon tests in the context of GMM-based estimation. As a result, MHR serve as a natural guard against persistent conditional misspecification.

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 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. Further, we show that the dynamics of the factors, as manifested by multi-horizon variance ratios, are grossly inconsistent with the dynamics implied under the null of the factor model. In other words, current state-of-the-art 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.

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 b_t as a latent process and estimate it and the corresponding a_t by including MHR in the set of test assets, thus asking the model to jointly price returns at multiple horizons. The estimated factor loading has large variation and little relation to the standard conditioning variables used in asset-pricing.

The evidence prompts us to investigate the economic origins of the time variation in b. We hypothesize that the typical factor construction methodology, which uses characteristics-based rank portfolios, is likely responsible for the persistent misspecification. When forming long-short portfolios, the decision to use decile extremes, or the characteristic itself, or the specific type of a double-sort, are ad hoc. We show that these decisions easily yield persistent misspecification in the scale of the portfolio. If one can identify a time-varying b_t reflecting the scale-misspecification, then one could form a new factor $\widetilde{F}_{t+1} = b_t^\top F_{t+1}$ that would feature a constant (unitary) loading in the SDF.

In practice, b is sometimes close to a constant and the model is not rejected. That happens when F is comprised of the market excess returns and other factors from the FF3 model (Fama and French, 1993), the value (HML) and size (SMB) factors. But, in most cases, including betting against beta (BAB), profitability, investment, momentum, and versions of the five Fama-French factors hedged for unpriced risks

(see Daniel, Mota, Rottke, and Santos, 2018), the constant b hypothesis is strongly rejected.

It may seem intuitive that as one adds more factors, there is more scope for complicated dynamics that require time-varying b. Economically, however, this is not so clear. As one considers models with increasingly higher short-run maximal SR, and, presumably, gets closer to the MVE portfolio, it would be natural to expect less evidence of time-variation in b. In this sense, our results are surprising.

In a robustness section, we run standard "alpha-regressions" and use Gibbons, Ross, and Shanken, 1989 (GRS) to test whether instruments (managed portfolios) based on MHR can deliver alpha in the standard, short-horizon setting. The models are rejected with high maximal information ratios. That is also the case when considering factors scaled by volatility and dividend-price ratio on the right-hand side. In sum, standard observable instruments do not capture the effects that we find in this paper.

There are many papers that test conditional versions of factor models. 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 multi-factor models with constant factor loadings in the SDF. Another example is Moreira and Muir (2017) who find that factors normalized by their conditional variance outperform the original factors. That is equivalent to a b that varies with the inverse of the conditional factor variance. Indeed, we find that the estimated b_t is significantly negatively related to the conditional factor variance and significantly positively related to the dividend-price ratio. However, the R^2 of a regression of b_t on these variables is low, indicating that most of the variation is due to other sources.

Our paper 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).

Notation. We use E for expectations and V for variances. 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 (e.g., Fama and French, 1993, 2015).

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:[1]

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

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., $E(M_{s,s+1}R_{s,s+1}) = 1$ for any s. 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,

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

[1] 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.

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 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

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

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

In general, lower-frequency returns can be thought of as applying a particular conditioning instrument in the sense of Hansen and Richard (1987). In particular, consider an instrument in investors' information set, z_t, and the returns to the managed portfolio $z_t R_{t,t+1}$. LOOP implies:

$$E(M_{t,t+1}z_t R_{t,t+1}) = E(M_{t,t+1}R_{t,t+1}) \cdot E(z_t) + Cov(z_t, M_{t,t+1}R_{t,t+1}).$$

Instruments allow for testing of the model's conditional properties as they test, through the term $Cov(z_t, M_{t,t+1}R_{t,t+1})$, whether pricing errors are predictable.

In our case, instruments vary with horizon and we can define them as $z_t^{(h)} = M_{t-h,t}R_{t-h,t}$. The variable $z_t^{(h)}$ is not a traditional instrument used in GMM tests because it has to satisfy $E(z_t^{(h)}) = 1$. This condition is tested implicitly when implementing unconditional tests of (1) or (2).

The instrument-based interpretation of the tests connects to the optimal instruments of Hansen (1985), see also Nagel and Singleton (2011). Optimal instruments are selected to minimize the covariance matrix of parameter estimation error. In general, one cannot implement the true optimal instrument because it entails computation of the conditional expectation of the derivative of the moment with respect to parameters. That makes it hard to directly compare them to $z_t^{(h)}$.

We view our approach as attractive and pragmatic complement to the optimal perspective. Our instruments are economically motivated, not data-mined, and straightforward to implement. The remaining issue is whether they are informative about the candidate models. In contrast to optimal instruments, we cannot offer a mathematical result. Thus, instead, we will proceed with evaluating informativeness of these instruments in the context of a specific application: linear factor models.

2.4 Linear factor models

While our approach is applicable to any asset-pricing model featuring an SDF, we consider cases where SDFs can be expressed as

$$M_{t,t+1} = a - b^\top F_{t,t+1}, \tag{3}$$

where $F_{t,t+1}$ is a vector of factor excess returns.

Our test assets are the factors themselves, as well as the risk-free rate. The SDF coefficients, a and b, can be estimated using only one-period returns in an exactly-identified unconditional GMM estimation of the model. We will be able to also test the model by adding multi-horizon excess factor returns. Thus, we will be testing if a one-period factor model can price the MHR of its factors when compounded to the corresponding horizons. This is the minimal requirement one would want to impose on a model in terms of MHR pricing. In order to test the condition (2), we construct factor MHR via:

$$R_{t,t+h} = \prod_{j=1}^{h} (F_{t+j-1,t+j} + R_{f,t+j-1,t+j}).$$

Because the test assets are the SDF factors themselves, there exists an SDF with time-varying coefficients,

$$M_{t,t+1} = a_t - b_t^\top F_{t,t+1}, \tag{4}$$

that prices all test assets correctly.[2] Equations (1) and (2) imply that

$$a_t = R_{f,t,t+1}^{-1} + b_t^\top E_t(F_{t,t+1}), \tag{5}$$

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

Our null hypothesis is $a_t = a$, and $b_t = b$.

This null hypothesis is equivalent to saying that there exists a linear combination of the factors that is perfectly negatively correlated with an SDF. Such a linear combination is the MVE portfolio (Cochrane, 2004). The large literature on unconditional beta-pricing factor models (e.g., Carhart, 1997; Fama and French, 1993) is exactly

[2] If the SDF conditionally prices the factors and the risk-free rate correctly, it also prices any trading strategy in these base assets correctly. As shown earlier, MHR can be viewed as trading strategies in one-period returns. If LOOP holds, it is immediate that there exists some a_t and b_t such that this alternative SDF prices the base assets conditionally, as we have as many degrees of freedom as we have moments at each time t.

concerned with finding factors that unconditionally span the unconditional MVE portfolio.

Note that the constant b does not imply constant factor risk premiums—it implies that conditional factor risk premiums must be proportional to the conditional factor variance multiplied by the conditional gross risk-free rate. A number of authors, e.g., Campbell and Shiller (1988), Fama and French (1988), Haddad, Kozak, and Santosh (2018), find strong return predictability in common pricing factors. These predictability findings do not by themselves contradict our null hypothesis. Our tests are, however, a test of whether the conditional factor risk premiums line up with the conditional variance in the way Equation (6) prescribes.

Yet another perspective on our analysis arises from representing the SDF in terms of prices of risk, λ_t, and innovations. Decompose the SDF into its conditional mean, $E_t(M_{t,t+1}) = R^{-1}_{f,t,t+1}$, and a shock:

$$M_{t,t+1} = R^{-1}_{f,t,t+1} - \lambda_t^\top \varepsilon_{t+1}, \tag{7}$$

where ε_{t+1} is i.i.d., zero mean, and unit variance. Further, assume that the shocks to the SDF also drive factor returns in the following manner:

$$F_{t,t+1} = R_{f,t,t+1} V_t^{1/2}(F_{t,t+1})\lambda_t + V_t^{1/2}(F_{t,t+1})\varepsilon_{t+1}.$$

Then, we can represent the SDF in terms of F

$$M_{t,t+1} = R^{-1}_{f,t,t+1} + R_{f,t,t+1}\lambda_t^\top \lambda_t - \lambda_t^\top V_t^{-1/2}(F_{t,t+1})F_{t,t+1}. \tag{8}$$

Comparing Equations (4) and (8), we see that

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

The null hypothesis $b_t = b$ implies that prices of risk must line up with the amounts of risk.

2.5 Results: a case of MKT + BAB

Our initial discussion is based on a two-factor case: excess return on the market (MKT) and the betting against beta (BAB) strategy. The first factor corresponds to the oldest factor model, the CAPM. The second factor corresponds to the oldest anomaly: low (high) beta stocks have positive (negative) alphas (Jensen, Black, and Scholes, 1972). Frazzini and Pedersen (2014) propose the BAB factor that exploits

this anomaly while hedging the market risk. We later show that the issues highlighted in this case are pervasive and hold for many factor models.

The model (3) specializes to

$$M_{t,t+1} = a - b_m MKT_{t,t+1} - b_{bab} BAB_{t,t+1}, \tag{10}$$

where $MKT_{t,t+1} = R_{m,t,t+1} - R_{f,t,t+1}$ and $BAB_{t,t+1}$ is the (excess) return associated with the BAB long-short strategy. The test assets are the 1-, 12-, and 24-month excess returns to the factors in the model at hand, as well as the 1-month risk-free rate. The sample is from July 1963 to June 2017. We implement the two-stage efficient GMM. We also consider the unconditional CAPM by setting $b_{bab} = 0$.

In addition to the GMM estimation results, we report mean absolute pricing errors (MAPE). MAPE should be understood as the present value of a zero-investment strategy with a notional of \$1. Results are reported in Table 1.

The leftmost columns in the table show, as a reference, a "one-horizon test", which is simply the two models, MKT or MKT+BAB, estimated using one-month returns to the factor(s) and the risk-free rate. The "multi-horizon test" reports the results of estimation including the 12- and 24-month excess factor returns as test assets.

The SDF coefficient estimates for the unconditional CAPM, or the MKT factor, are basically the same across the two estimations. The row labeled MAPE shows the annualized mean absolute pricing error of the factors across the 1-, 12-, and 24-month horizons for both single- and multi-horizon estimations. The annualized error is less than 0.5% in both estimations and the J-test in the multi-horizon test does not reject. The maximal annualized prices of risk of this model (the maximal SR), reported as $\sqrt{12}V_T^{1/2}(M_{t,t+1})/E_T(M_{t,t+1})$, are equal to 0.383 and 0.365 in the single-horizon and multi-horizon tests, respectively. In other words, for the MKT-model the assumption of constant a and b is not rejected and the addition of multi-horizon returns in the test therefore does not affect the results much.

For the MKT+BAB model, however, multi-horizon excess factor returns reveal substantial model misspecification. In particular, in the single-horizon case, the annualized MAPE is 31%. Considering that the risk premium on these factors are in the order of 6-8% p.a. the mispricing is economically very large. In the multi-horizon test, the MAPE is reduced considerably to 4.4%, which is still economically a large number. This reduction is achieved by strongly altering the coefficient estimates. Most notably, the risk loadings b_m and b_{bab} are reduced from 3.0 and 8.2 in the single-horizon case to 0.4 and 1.7 in the multi-horizon case, respectively. The estimated maximal SR is reduced from 0.979 to 0.198. Put differently the SR is too low even to account for the market itself.

That the estimated parameters have changed substantially upon introduction of MHR is a prima facie evidence of the model misspecification. The J-test confirms that, rejecting the model at the 1% level. Hence, this factor model does not successfully price its own factors both at short and long horizons. This is a startling fact. One might have thought that it is innocuous to focus on short horizon returns. But that is not the case. A model fit that does well at short horizons may do poorly at long horizons, even when only pricing its own components. Clearly, it is important to consider factor model fit across multiple horizons.

Figure 1 graphically conveys the same message. The panels report pricing errors (raw, not annualized) for each factor at all horizons from 1 to 24 months. Panels A and C show that MKT is pricing itself across horizons quite well in the MKT model, whether the model is estimated using single- or multi-horizon returns. For the MKT+BAB model, however, pricing errors increase strongly with horizon for both factors, and especially for the BAB factor. Panel B shows that when estimating the model using only 1-month returns, the 24-month pricing error is more than \$1.6 in absolute value – economically, an astonishing failure relative to the \$1 scale of the zero-investment strategy being tested. While estimating this model on multi-horizon returns reduces the pricing errors substantially, the model is still rejected.

Note that the failure to reject the MKT-model should not be interpreted as this being the "best" model. The non-rejection is a function of the test assets, which in this case is only the risk-free rate and excess MKT returns at different horizons. Indeed, the SDF loading on the BAB-factor b_{bab} is statistically different from zero in both estimations of the MKT+BAB model, thus rejecting the MKT model.

3 Dynamics implicit in a model specification

The MKT+BAB model rejection is a consequence of factor dynamics unaccounted for by our model. In this section, we consider these dynamics.

3.1 Constant factor loadings

In order to frame the evidence, recall that our null hypothesis, $a_t = a$ and $b_t = b$, implies that the portfolio return defined by

$$\widetilde{F}_{t,t+1} = b^\top F_{t,t+1} \tag{11}$$

is the unconditionally MVE portfolio (Hansen and Richard, 1987). Equation (6) implies for this MVE portfolio:

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

Therefore, under the null $b_t = b$, the portfolio dynamics can be written as

$$\widetilde{F}_{t,t+1} = R_{f,t,t+1} V_t(\widetilde{F}_{t,t+1}) + \widetilde{\varepsilon}_{t+1}, \tag{12}$$

where $\widetilde{\varepsilon}_{t+1} \equiv b^\top V_t^{1/2}(F_{t,t+1})\varepsilon_{t+1}$. Rejection of the null $b_t = b$ implies that true dynamics of the portfolio are different.

3.2 Variance ratios

Equation (12) implies that if $\widetilde{F}_{t,t+1}$ is the MVE portfolio, then the residuals $\widetilde{\varepsilon}$ should be uncorrelated over time. That is not the case in the data. To illustrate this point, we use variance ratios of these residuals across multiple horizons. Variance ratios have been used earlier to assess multi-horizon Euler equation errors (see, e.g., Flood, Hodrick, and Kaplan, 1994).

Recall that the variance ratio at horizon h is:

$$VR(h) = \frac{V\left(\sum_{j=1}^{h} \widetilde{\varepsilon}_{t+j}\right)}{h \times V(\widetilde{\varepsilon}_{t+1})}. \tag{13}$$

As is well known, if $\widetilde{\varepsilon}$ is uncorrelated over time, then $VR(h) = 1$ for any h. If $\widetilde{\varepsilon}$ is positively autocorrelated, then $VR(h) > 1$, whereas if $\widetilde{\varepsilon}$ is negatively autocorrelated, then $VR(h) < 1$.

We use b estimated by GMM to construct $\widetilde{F}_{t,t+1}$ and estimate an EGARCH(1,1) to obtain $V_t(\widetilde{F}_{t,t+1})$ under the maintained assumption that the two factors (MKT and BAB) are uncorrelated. Continuing with the examples of section 2.5, we report variance ratios for the shocks in the case of MKT and MKT+BAB in Figure 2.

Not surprisingly, there is not much of a departure from 1 in the case of the MKT model. The variance ratios for MKT+BAB are much bigger than the null, however, implying a strong degree of intertemporal dependence that is not accounted for by the model with $b_t = b$. The variance ratio in this case increases rapidly with horizon and exceeds 3(!) at the 24-month horizon.

Thus, the MVE combination of the MKT and BAB portfolios is too highly autocorrelated relative to what is implied by the model and the estimated conditional

return variance.[3] Time-variation in b_t that prices MHR correctly would "undo" this autocorrelation and create a residual with variance ratios equal to 1 at all horizons.

3.3 Time-varying factor loadings

The rejection of the model (10) prompts us to consider an alternative SDF

$$M_{t,t+1} = a_t - b_m MKT_{t,t+1} - b_{bab,t} BAB_{t,t+1}.$$

To illustrate how much variation in b_t we are missing, we use MHR to estimate the SDF. We assume that $b_{bab,t}$ follows a latent AR(1) process:

$$b_{bab,t} = (1-\phi)b + \phi b_{bab,t-1} + \sigma\epsilon_t.$$

While b_m could be time-varying, we set it to a constant for simplicity as most of the action is clearly emanating from the BAB factor. Further, we assume that the MKT and BAB factors are conditionally uncorrelated. That is consistent with the BAB factor construction, which weights the underlying stocks by the inverse of estimated market betas to make the BAB factor's forward-looking market beta approximately zero.

The model is estimated via GMM using the same moment conditions as in the multi-horizon test case. We use the Kalman filter to estimate $b_{bab,t}$ via the observation equation:

$$[R_{f,t,t+1} V_t(BAB_{t,t+1})]^{-1} BAB_{t,t+1} = b_{bab,t} + e_{t+1}$$

that is inferred from Equation (6). We infer the corresponding a_t using Equation (5).

Table 2 presents estimation results. The $J-$test fails to reject the model, and the annualized MAPE is 0.013, which is more than three times smaller than that of the model with a constant b, as estimated using MHR, and more than 20 times smaller than the case with constant b estimated using only 1-month returns. Figure 3 displays the pricing errors at each horizon. Consistent with MAPE, they are smaller than those of the constant-b model. Also, they do not exhibit the horizon-related trend evident in Figure 1D.

The estimated $b_{bab,t}$ is quite persistent with $\phi = 0.91$ and volatile with unconditional volatility of 6.04, which is large relative to its mean of 9.22. To get a more intuitive

[3] In unreported results, we find that it is the BAB factor itself that has a strongly increasing variance ratio over time.

interpretation for these numbers we can use Equation (9), which implies that the price of risk (SR) associated with exposure to the BAB-shock, $\lambda_{bab,t}$, is equal to $b_{bab,t}V_t^{1/2}(BAB_{t,t+1})$. The annualized volatility of this price of risk is 0.66. This is large compared to its mean of 0.91. Figure 4 displays the time-series of $b_{bab,t}$ and $\lambda_{bab,t}$. Both variables appear to be related to business cycles, but it is clear that there is much additional variation. This prompts the question if we could relate the dynamics to standard observable instruments used in the literature.

Table 3 displays the results of regressing $b_{bab,t}$ and $\lambda_{bab,t}$ on a recession indicator, a beta decile spread, the conditional variances of the market and BAB returns as estimated earlier, the log market dividend to price ratio, and the 10 year over 1 year term spread. The beta decile spread is the spread between the high and low decile market-beta sorted portfolios on Kenneth French's webpage. The beta of the high and low decile portfolio are the historical 5-year betas of the constituent firms. We then shrink the betas towards 1, with a weight of 0.6 on the historical estimate and 0.4 on 1. We use the Kenneth French data in lieu of the beta spread in the actual BAB portfolio as this is not available to us. The spread we use is likely to be highly correlated with the beta-spread of the underlying stocks in the BAB portfolio.

The overall R^2 is modest, in the 20-30% range with sporadic significance of some of the variables. Notably, $b_{bab,t}$ is significantly *negatively* correlated with the NBER recession indicator and the conditional variances of BAB and MKT. One may have thought that in these periods, the SDF would load more heavily on the factor, but that is not the case. Both $b_{bab,t}$ and $\lambda_{bab,t}$ load positively and significantly on the beta spread.

This evidence leaves us with a question of the driving forces behind the variability in b_t. Equation (9) says that there is going to be variation whenever factor prices of risk and their variances are not proportional. In the case of MKT, the two objects appear to line up close enough to fail to reject the null of constant b_m. That is not the case for BAB.

4 The impact of sorting on the factor dynamics

Our thesis is that the common characteristic-sort-based approach to construction of factors often leads to persistent misspecification of the factors. In particular, in order to construct factors that correspond to an SDF with a constant b, the sort needs to deliver factors that have a conditional covariance matrix that lines up with the conditional prices of risk of the underlying shocks, as in equation (9). It is unlikely

that a characteristic-sort will deliver such a factor as there is no mechanism in the sorting procedure that guarantees it.

4.1 An example: betting against beta

We specialize the SDF in (7) to

$$M_{t,t+1} = R_{f,t,t+1}^{-1} - \lambda_m \varepsilon_{m,t+1} - \lambda_{bab,t} \varepsilon_{bab,t+1}, \tag{14}$$

where the shocks are uncorrelated with each other and over time. Since the market factor is not our focus, we assume the market shock's price of risk is constant and let excess market returns follow:

$$MKT_{t,t+1} = \sigma_m \lambda_m + \sigma_m \varepsilon_{m,t+1}. \tag{15}$$

Further, consider a simplified version of the BAB factor of Frazzini and Pedersen (2014) where there are only two underlying assets – a high market beta portfolio and a low market beta portfolio with returns:

$$\begin{aligned} R_{h,t,t+1} - R_{f,t,t+1} &= \beta_{h,t} MKT_{t,t+1} + [(1-\beta_{h,t})\lambda_{bab,t} + (1-\beta_{h,t})\varepsilon_{bab,t+1}], \\ R_{\ell,t,t+1} - R_{f,t,t+1} &= \beta_{\ell,t} MKT_{t,t+1} + [(1-\beta_{\ell,t})\lambda_{bab,t} + (1-\beta_{\ell,t})\varepsilon_{bab,t+1}]. \end{aligned}$$

Then, construct the BAB factor as in Frazzini and Pedersen (2014):

$$\begin{aligned} BAB_{t,t+1} &\equiv \beta_{\ell,t}^{-1}(R_{\ell,t,t+1} - R_{f,t,t+1}) - \beta_{h,t}^{-1}(R_{h,t,t+1} - R_{f,t,t+1}) \\ &= \Delta_t \lambda_{bab,t} + \Delta_t \varepsilon_{bab,t+1}, \end{aligned} \tag{16}$$

where

$$\Delta_t \equiv (\beta_{h,t}\beta_{\ell,t})^{-1}(\beta_{h,t} - \beta_{\ell,t})$$

is the "beta spread," which measures the difference in the betas between the high and low beta portfolios at each time t.

This beta spread is highly time-varying and persistent in the data. Figure 5 displays the time-series of Δ_t. The annual volatility of the beta spread is 0.57 and the annual persistence is 0.79. Thus, the choice to weight the portfolios with the inverse of the portfolio beta is potentially not innocuous – it induces persistent dynamics in the conditional mean and volatility of BAB returns, per equation (16) and the observed beta spread.

Substitute Equations (15) and (16) into Equation (14) to express the SDF as a function of the MKT and BAB factors:

$$M_{t,t+1} = R_{f,t,t+1}^{-1} + \lambda_m^2 + \lambda_{bab,t}^2 - \lambda_m \sigma_m^{-1} MKT_{t,t+1} - \lambda_{bab,t} \Delta_t^{-1} BAB_{t,t+1}. \tag{17}$$

That is a version of the general specification in (8). Note that the price of risk $\lambda_{bab,t}$ has to be proportional to Δ_t for b_{bab} to be constant. Our evidence is that this not the case. Further, while the sorting procedure chooses Δ_t, there is nothing explicit in the sorting procedure that ensures that it lines up correctly with $\lambda_{bab,t}$.

Empirically, as shown in table 3A, the beta spread is positively and significantly related to $b_{bab,t}$. Thus, the price of risk for the BAB shock, $\lambda_{bab,t}$, must be varying strongly with the beta spread so as to undo the effect of the inverse of the beta spread present in the term multiplying BAB, $\lambda_{bab,t}\Delta_t^{-1}$.

One might think of constructing a new factor $\widetilde{BAB}_{t,t+1} = \lambda_{bab,t}\Delta_t^{-1} BAB_{t,t+1}$ that would lead to a constant b. However, one needs to know ex-ante both the conditional price of risk $\lambda_{bab,t}$ and the time-varying scale Δ_t in order to accomplish that.

Our example is stylized and there could be other dynamics at play. For instance, the BAB shock itself may have time-varying volatility that is not driven by the beta spread. The main takeaway, however, is that there is no direct solution to this problem by adjusting the time t sort based on the time t cross-section of the characteristic itself. The argument we put forth in this paper is that MHR factor returns provide informative moments to detect this type of misspecification.

4.2 General observations

Although we are using BAB as a specific example, the phenomenon that we are describing is pervasive. Cross-sectional portfolio sorts may easily have drifting loading on the true underlying risk factor. The two standard methods used in the literature for constructing factors – Fama-MacBeth regressions of returns on lagged characteristics and rank characteristic-sorts – both normalize the cross-section of the characteristics at each t in particular ways that are likely to induce persistent misspecification. To illustrate the mechanism we highlight in this paper, we consider Fama-MacBeth regressions as the method for factor construction.

Assume the true SDF follows (3), where F is a vector of orthogonal factor excess returns, and excess factor returns and the risk-free rate are priced by this SDF. Let excess returns to assets follow a factor model

$$R_{i,t,t+1} - R_{f,t,t+1} = \beta_{i,t}^{\top} F_{t,t+1} + \epsilon_{i,t+1},$$

where residuals are uncorrelated across stocks i and time t. Assume there is a unit mass of firms such that the idiosyncratic shocks are completely diversifiable, that the cross-sectional distribution of factor betas is non-degenerate, and that $R^{-1}_{f,t,t+1} = E_t(M_{t,t+1})$. If we knew the $\beta_{i,t}$ we could run a cross-sectional regression at each time $t+1$ to retrieve the factors F:

$$R_{i,t,t+1} - R_{f,t,t+1} = \gamma_{0,t+1} + \gamma_{1,t+1}^{\top}\beta_{i,t} + \varepsilon_{i,t+1}, \tag{18}$$

In this case, $\gamma_{1,t+1}$ gives the vector of long-short factor excess returns, $F_{t,t+1}$. Typically, we don't have these conditional betas and researchers instead use observable characteristics, such as the book-to-market ratio.

Consider characteristics that have the following relation to betas:

$$X_{i,t} = \delta_{0,t} + \delta_{1,t}\beta_{i,t}.$$

Thus, the characteristic is not unconditionally proportional to the true factor betas, even if we allow for a time fixed effect.[4] For ease of exposition, but without loss of generality, assume that there is only one factor. The Fama-MacBeth cross-sectional regression

$$R_{i,t,t+1} - R_{f,t,t+1} = \widehat{\gamma}_{0,t+1} + \widehat{\gamma}_{1,t+1}X_{i,t} + \varepsilon_{i,t+1},$$

then delivers the factor

$$\widehat{\gamma}_{1,t+1} = \frac{Cov\left(\delta_{1,t}\beta_{i,t}, \beta_{i,t}F_{t,t+1}\right)}{V\left(\delta_{1,t}\beta_{i,t}\right)} = \frac{V\left(\beta_{i,t}\right)\delta_{1,t}F_{t,t+1}}{\delta_{1,t}^2 V\left(\beta_{i,t}\right)} = \delta_{1,t}^{-1}F_{t,t+1},$$

where the variance and covariance are taken cross-sectionally across assets i at each point in time. Thus, the factor used by researchers is

$$\hat{F}_{t,t+1} = \delta_{1t}^{-1}F_{t,t+1}.$$

The intercept $\delta_{0,t}$ cancels out in the long-short portfolio, but the proportional component $\delta_{1,t}$ does not. There is no control variable that naturally "cleans up" the characteristic in the Fama-MacBeth regression.

[4] One way to motivate this relation is to consider the effect of time variation in prices of risk. For simplicity, consider a one-factor model version of (18), which implies $E_t(R_{i,t,t+1} - R_{f,t,t+1}) = E_t(\gamma_{1,t+1})\beta_{i,t}$. Next, consider a characteristics-based model of excess returns: $E_t(R_{i,t,t+1} - R_{f,t,t+1}) = \delta_0 + \delta_1 X_{i,t}$ (e.g., Daniel and Titman, 1997). Setting the conditional expected returns equal implies that $X_{i,t} = \delta_{0,t} + \delta_{1,t}\beta_{i,t}$ where $\delta_{0,t} = -\delta_0/\delta_1$ and $\delta_{1,t} = E_t(\gamma_{1,t+1})/\delta_1$. In this case, $\delta_{1,t}$ will vary persistently if the conditional factor risk premium varies persistently.

As a result, the standard approach identifies a conditionally correctly specified model, but an unconditionally misspecified model. That is, researchers specify:

$$M_{t+1} = \hat{a} - \hat{b}\hat{F}_{t+1}.$$

A conditional version of this model is true:

$$\begin{aligned} M_{t+1} &= a - b_t\hat{F}_{t+1}, \\ b_t &= b\delta_{1t}. \end{aligned}$$

This is exactly the same issue that we have highlighted generally in equation (8) and, specifically in the context of BAB, in equation (17). In sum, MHR, as test assets, are likely to be useful for most factor models. The next section confirms this intuition.

5 Testing factor models

In order to demonstrate the breadth of our conclusions we implement the MHR-based testing of a broad set of linear factor models. These are state-of-the-art specifications that boast high single-horizon SR. In this section we describe what MHR reveal about these models.

5.1 Additional factors

Our roster of models includes the Fama and French 3- and 5-factor models, FF3 and FF5, respectively (Fama and French, 1993, Fama and French, 2015), and FF5+MOM+BAB, where MOM refers to the Momentum factor (Carhart, 1997, Jegadeesh and Titman, 1993) as given on Kenneth French's webpage. 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 versions of these models with unpriced risks hedged out from the five Fama-French factors as proposed by Daniel, Mota, Rottke, and Santos (2018) (DMRS). All these models feature very high short-run SR. Thus, the hope is that these models get us closer to the portfolio that is perfectly negatively correlated with an SDF. Given Equation (3), such a portfolio has constant a and b in the SDF and lies on the unconditional MVE frontier.

Table 4 reports the same GMM tests we have implemented for MKT, and MKT+BAB (Table 1) for this roster of models. To help with comparison, we repeat the MAPE

and $J-$tests for these two models. We see that all models, with the exception of FF3 (and MKT) are rejected. Out of the rejected models, the annualized MAPE in the multi-horizon estimation ranges from 0.004 from the original MKT model to 0.047 for the DMRS versions of the MKT+BAB and the FF5+MOM+BAB models. Again, given that risk premiums on these factors is of the order of 4-8%, the MAPEs are also economically large, with the exception of the non-rejected original MKT and FF3 models.

The table also reports maximum SR implied by the model (max. price of risk). The one-period maximum SR – a dimension along which the candidate models are optimized – appears to be positively related to MAPE corresponding to the MHR-based tests. We demonstrate this observation by constructing a scatter plot of (log) MAPE versus (log) maximum SR in Figure 6. 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 (3) span the unconditional MVE portfolio with a constant b in the SDF specification.

Recall that the multi-horizon test assets are the model factors themselves at longer horizons. So, a failure to reject MKT or FF3 does not necessarily imply that these models would be able to price the full cross-sections of assets at multiple horizons. Nevertheless, the results are suggestive of the implicit dynamics of b_t. It appears that the time-varying volatility of MKT and FF3 factors do not have a dramatic impact on b_t for these two models as the prices of risk in these models line up reasonably well with the conditional volatilities, as per equation (9).

Related, the price of risk for the remaining models, estimated using multi-horizon returns, typically vary strongly from the prices of risk estimated using one-month returns. Thus, when viewed through this lens, the rejection of the constant b models is again economically large.

We investigate this further via variance ratios that follow Figure 2. Figure 7 reports the patterns of variance ratios for all models. We observe variance ratios that deviate from 1 (the null hypothesis) for all models. The extreme pattern that we've documented in the MKT+BAB case is also observed for the FF5+MOM+BAB, the DMRS3, and the DMRS5 models. In these cases, variance ratios are well in excess of 2 from the 12-month horizon and on.

The FF3+MOM+BAB and FF5 models exhibit slowly increasing variance ratios, ending at about 1.6 and 1.4, respectively, at the 60-month horizon, indicating small but highly persistent misspecification. The MKT and FF3 models all have variance ratios that increase relatively quickly, with maximal values around 1.3-1.4 at the 12- to 24-month horizons. At longer horizons, these variance ratios revert towards 1.

Thus, all the models exhibit too high autocorrelation of factor returns relative to the volatility dynamics, and there is considerable variation in the strength of these patterns. Interestingly, the early factor models (MKT and FF3) have substantially smaller deviations from the null hypothesis of a variance ratio of 1 relative to the DMRS versions of these models. It appears that the refinement of the original factors, which yields higher short-run maximal SR, has come at a cost of introducing persistent misspecification that shows up in MHR and the models' variance ratios.

To summarize, we document pervasive, persistent misspecification in the current, most successful factor models, as judged by their one-month maximal SR. This misspecification leads to large pricing errors over longer horizons. Thus, while these models may do well in terms of characterizing the local, one-month, risk-return trade-off, they do surprisingly poorly in terms of longer-horizon risk-return trade-offs. The nature of our tests allow us to trace this misspecification to the current dominant procedure for factor construction that easily produces factors with the wrong conditional scale.

5.2 Robustness checks

In this section we show that our results are robust to using shorter horizon excess factor returns as test assets in the GMM. We also consider the standard monthly return "alpha-regressions" in the literature with short-term factor returns managed over time using instruments motivated by the MHR test. The results are robust also to this restricted, but more familiar test. Furthermore, allowing for time-varying b's through standard instruments such as dividend-price ratios or realized variance in the GRS tests does not fix the problem – the models are still rejected when using the horizon-managed factor returns as test assets.

5.2.1 Alternative horizons

Table 5 reports the two-stage efficient GMM-results when using 1-, 2-, 6- and 12-month excess factor returns in addition to the 1-month risk-free rate as test assets. The shorter maximal horizon means there is less overlap in the data and give more statistical power. The downside is that it is less likely to capture the most persistent sources of misspecification. The models considered in Panel A are MKT, MKT+BAB, FF3, FF5 and FF5+BAB+MOM. The p-values of the J-test shows that all models except MKT are rejected at the 1 percent level, whereas the market has a p-value of 0.8. From Panel A we see that the MAPEs range from 0.007 for MKT to 0.115 for

MKT+BAB. In Panel B we consider the same models with DMRS in place of MKT and the FF factors. All models in Panel B are rejected at the 1 percent level and the MAPE are generally larger, ranging from 0.014 for DMRS5 to 0.092 for the DMRS version of MKT+BAB.

5.2.2 GRS test

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. For our purposes, the test assets will be 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.,

$$R^e_{i,t,t+1} \equiv z^{(h)}_{k,t} F_{k,t,t+1},$$

where $z^{(h)}_{k,t} = M_{t-h,t} R_{k,t-h,t}$ 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 $z^{(h)}_{k,t}$ 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.

In sum, we run the time-series regression

$$R^e_{i,t,t+1} = \alpha_i + \beta_i^\top F_{t,t+1} + \varepsilon_{i,t+1}$$

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

Because the instruments $z^{(h)}_{k,t}$ 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.

Table 6 reports the results of GRS tests on a variety of factor models using lagged discounted returns for horizons 2, 6, 12, and 24 months for each of the factors, i.e., four managed portfolios for each factor. In Panel A we consider MKT, MKT+BAB, FF3, FF5 and FF5+BAB+MOM. 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

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

model and FF5, which both have p-values using in-sample instruments of about 0.05 but when using out-of-sample instruments these p-values goes up to 0.206 and 0.234 respectively.

Also reported are the annualized maximum information ratios (IR) from MacKinlay (1995), which is the component of the maximum SR unexplained by the factors.[6] To get a sense of magnitudes, it is natural to compare the maximum IR to the SR of the factors. In FF5, the maximum IR is 0.49 compared to a SR of the factors of 1.09. For all the other models, the maximum IR is even larger relative to the SR of the corresponding factors. In other words, there seems to be a lot left on the table. Panel B reports even stronger rejections for the DMRS models. Except for the out-of-sample p-value of 0.05 for DMRS5+BAB+MOM, all p-values are below 1 percent. Similarly, all maximum IRs are greater than 40 percent of the corresponding factor SR.

The GRS test also serves an additional purpose as a robustness exercise. Throughout the paper we have assumed there exists a set of excess return factors with a linear combination that is perfectly correlated with the SDF, i.e., Equation (3). This may not be the case as the linear space of asset returns is bigger than that of excess returns. That is because the former additionally includes linear functions of the risk-free rate.

If the unconditional MVE portfolio based on excess returns, $\widetilde{F}_{t,t+1} = b^\top F_{t,t+1}$ (see Equation (11)), is not perfectly correlated with the projection of the SDF in the asset payoff space, then the SDF can be represented $M_{t,t+1} = a_t(1-b^\top F_{t,t+1})$. If the risk-free rate is not used for estimation, a_t is not identified, and one can consider, equivalently, $M_{t,t+1} = 1 - b^\top F_{t,t+1}$. By construction, the loading on $F_{t,t+1}$ is a constant.

Thus, if the GRS test rejects, it rejects that b is constant in this more general SDF specification. That has the familiar interpretation of the proposed factors being unable to price all excess returns. In sum, our results in this section show that the failure of the models to account for MHR is not driven by conditional mispricing in the risk-free rate as we reject the models also with one-period, excess return GRS tests.

5.2.3 Conditioning information

Models with time-varying SDF coefficients have been investigated in earlier literature (e.g., Jagannathan and Wang, 1996, Farnsworth, Ferson, Jackson, and Todd, 2002,

[6] 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.

Lettau and Ludvigson, 2001). The time-variation is typically assumed to be linear function of observable variables, motivated by the evidence of return and variance predictability (e.g., Campbell and Shiller, 1988, Engle, 1982, Fama and French, 1988). In this subsection, we in the same spirit add observable variables to allow for a conditional model in the GRS test.

Assume there are K original factors. Then, let $x_t^{(k)}$ be an $L \times 1$ vector of observable variables that predict returns or variance for factor k, as well as the number 1. We then run:

$$R^e_{i,t,t+1} = \alpha_i + \sum_{k=1}^{K} \widetilde{\beta}_i^{(k)} (F^{(k)}_{t,t+1} \otimes x_t^{(k)}) + \varepsilon_{i,t+1},$$

and test whether $\alpha_i = 0$ for all i for this expanded linear factor model. In this case, $\widetilde{\beta}^{(k)}$ is an $1 \times L$ vector of factor loadings for each original factor k. The test assets are the same as in the previous GRS tests – that is, the original factor excess returns at horizons 1, 2, 6, 12, and 24.

We consider two common conditioning variables for each factor: the dividend-price ratio (when sufficient data are available) and realized variance of each of the factors. The dividend-price (dp) ratio of a factor is calculated as the difference between the dp ratio of the long- and short leg of each factor. The realized variance (RV) is the variance of daily factor returns over the preceding month.

Table 7 reports the p-values of GRS tests of extended factor models using either dp or RV as conditioning information for each factor, that is, $x_t^{(k)} = [1 \; RV_t^{(k)}]^\top$ or $x_t^{(k)} = [1 \; dp_t^{(k)}]^\top$. The models considered are MKT, FF3 and FF5 and their DMRS counterparts. For the latter, only RV is used as conditioning information as DMRS do not report both ex- and cum-dividend returns. For comparison, we also report the results of the static models tested in Table 6.

Panel A shows results for the MKT factor. The baseline test, where simply $x_t = 1$, has a GRS test p-value of 4.9 percent. When conditioning on RV and dp the GRS p-values are 8.2 and 4.0 percent, respectively. The maximum IRs are similar to that of the baseline model, although, as a fraction of factor SR, they decline somewhat. For the DMRS MKT model, we reject at the 1 percent level both in the baseline case and when adding the RV managed factor. The IR increases relative to the baseline case.

Panel B shows results for the FF3 model. The FF3 is rejected both when conditioning on dp or RV, as well as in the benchmark case. The IRs are large, about the same or larger than in the baseline case. The DMRS version of this model rejects at the

1 precent level both in the baseline case and when we add the RV-managed factors, with the IRs about the same as in the baseline case.

Panel C shows results for the FF5 model. In this case, the model is rejected at the 1 percent level both in the baseline case and when conditioning on dp_t. When conditioning on RV, the p-value increases to 0.136 and we do not reject. In the DMRS version of this model, we reject in both cases at the 1 percent level. The IRs are all about the same magnitude.

In summary, the conditioning variables we consider here do not fix the problem. The IRs remain high for the same set of test assets, despite adding factors managed using RV or dp, and the significance levels are largely unchanged. Note that adding these managed factors does help increase one-period return maximal SR. The test assets reveal that the models still exhibit persistent misspecification. Thus, MHR can provide valuable information to a test that employs observable conditioning information to the factors as well.

6 Conclusion

The literature typically characterizes the risk-return trade-off, or, equivalently, the SDF, at a single horizon. This focus is somewhat surprising given that the multi-horizon risk-return trade-off is an important input to a range of economic decisions. We propose testing SDFs jointly across multiple horizons and analyze the implications of these additional moments for a wide roster of the state-of-the-art empirical linear factor models.

We show that tests employing multi-horizon returns (MHR) are highly informative about asset-pricing models. MHR relate to familiar conditioning instruments that are common in GMM-based tests. This new type of conditioning information allows us to uncover large mispricing for a wide range of benchmark factor models when faced with MHR. We identify the current practice of ad hoc portfolio sorts for factor construction as a likely culprit for the persistent misspecification and argue that MHR provides valuable additional moments that can help in the construction and testing of asset pricing models.

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

Campbell, John, and Robert Shiller, 1988, Stock prices, earnings, and expected dividends, *Journal of Finance* 43, 661–676.

Campbell, John, and Luis Viceira, 2002, *Strategic Asset Allocation: Portfolio Choice for Long-Term Investors* (Oxford University Press: Oxford, UK) first edn.

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.

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.

Daniel, Kent, and Sheridan Titman, 1997, Evidence on the characteristics of cross sectional variation in stock returns, *Journal of Finance* 52, 1–33.

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

Engle, Robert, 1982, Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation, *Econometrica* 50, 987–1007.

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

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

——— , 2015, Incremental variables and the investment opportunity set, *Journal of Financial Economics* 117, 470–488.

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

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.

Haddad, Valentin, Serhiy Kozak, and Shri Santosh, 2018, The economics of factor timing, working paper.

Hansen, Lars, 1985, A method for calculating bounds on the asymptotic covariance matrices of generalized method of moments estimators, *Journal of Econometrics* 30, 203–238.

——— , 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.

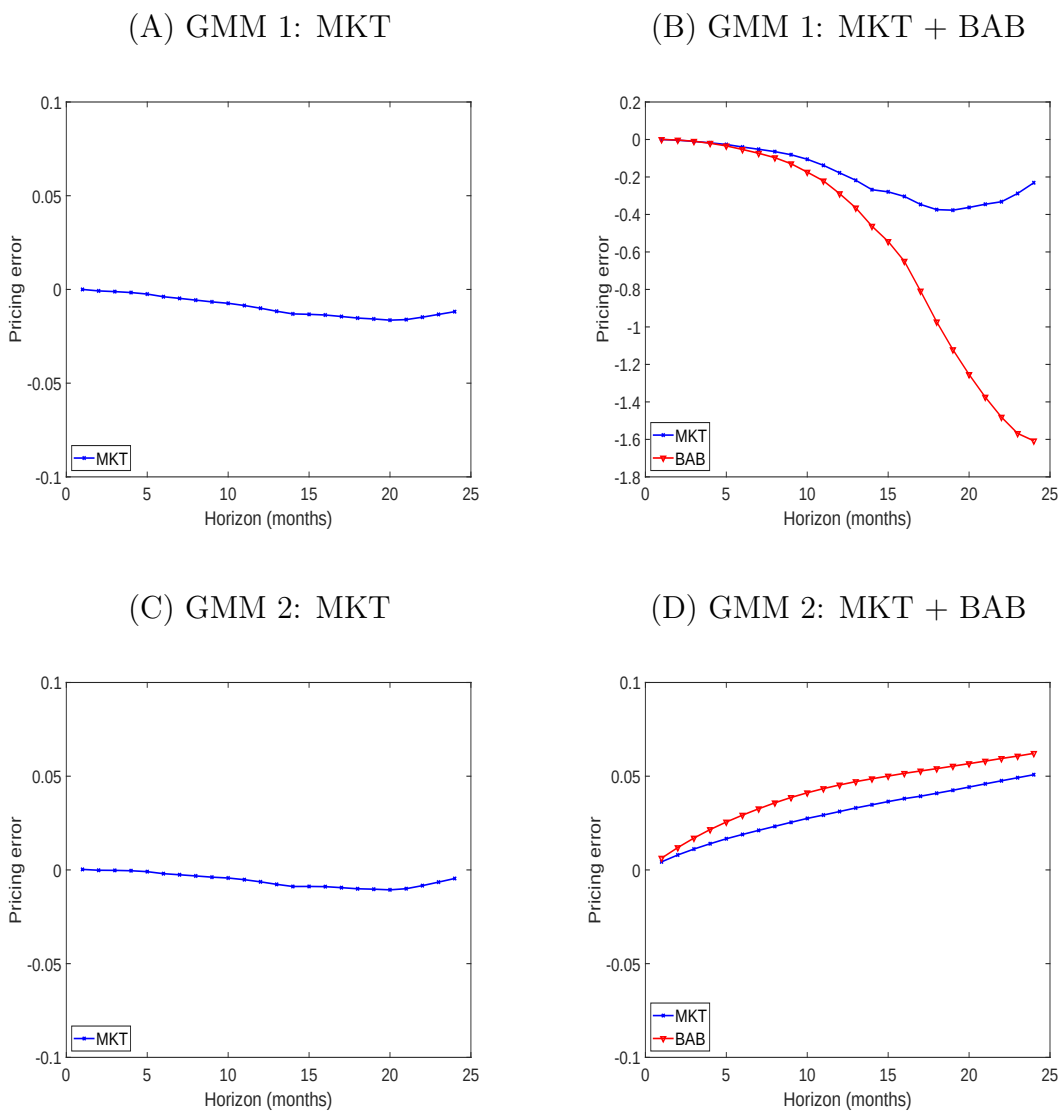
Nagel, Stefan, and Kenneth Singleton, 2011, Estimation and evaluation of conditional asset pricing models, *Journal of Finance* 66, 873–909.

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

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

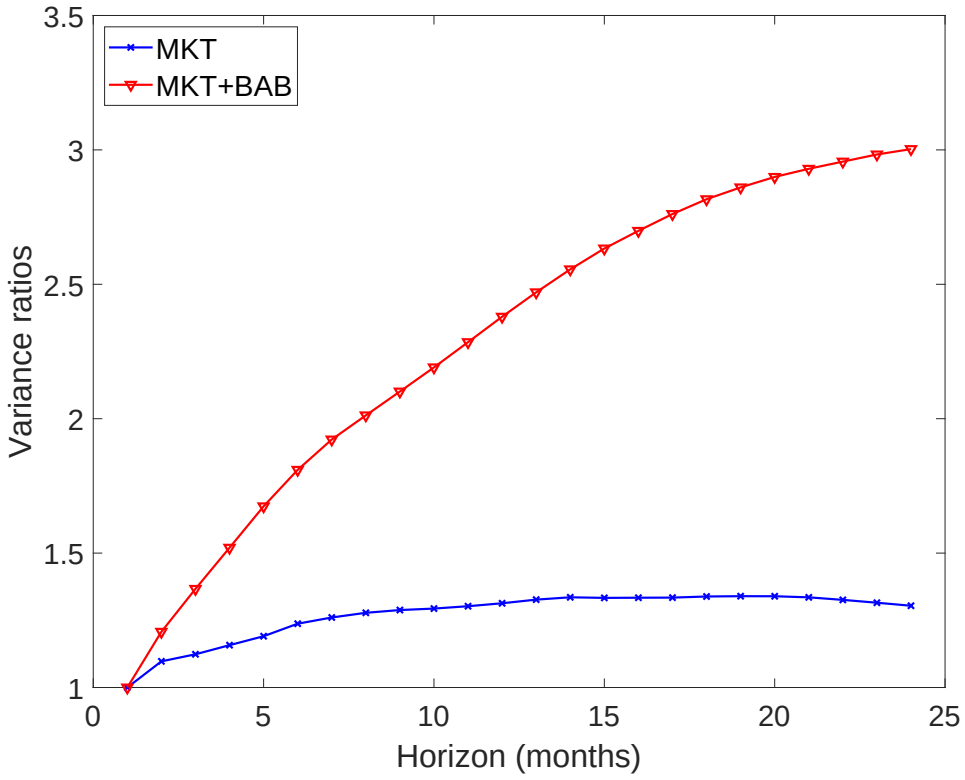
Zviadadze, Irina, 2017, Term-structure of consumption risk premia in the cross-section of currency returns, *Journal of Finance* 72, 1529–1566.

Figure 1
Term structure of pricing errors, $E(M \cdot (R - R_f))$



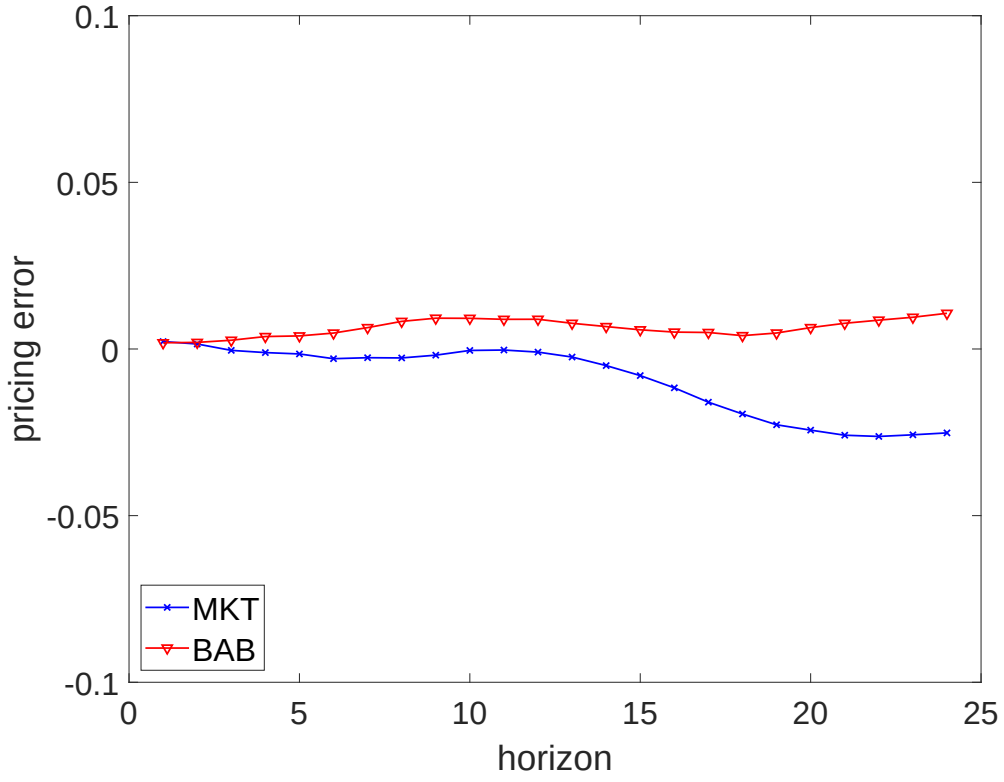
The panels show multi-horizon factor pricing errors for two models: $M_t = a - b_m \cdot MKT_t$ (left) and $M_t = a - b_m \cdot MKT_t - b_{bab} \cdot BAB_t$ (right). Pricing errors at horizon h are $E_T(M_{t,t+h}(R_{t,t+h} - R_{f,t,t+h}))$, where E_T denotes the sample average, $R_{t,t+h}$ is the gross 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 0. Panels A and B give pricing errors when the SDF is estimated using 1-month returns (GMM 1), whereas Panels C and D give pricing errors when the SDF is estimated also using the 12 and 24 month horizon excess factor returns (GMM 2). Note the scale of the y-axis for Panel B relative to the other panels. Sample: July 1963 to June 2017.

Figure 2
Variance ratios of factor return model-implied residuals



The figure shows variance ratios for model-implied residuals to the estimated linear combination of factors given by $b^\top F$ across different horizons, where b is taken from the SDF $M_t = a - b^\top F_t$ estimated using multi-horizon returns, as in the multi-horizon estimation in Table 1. The residuals equal $b^\top F_{t,t+1} - E_t(b^\top F_{t,t+1})$, where the model gives the conditional expectation. Under the null hypothesis that the model correctly captures the conditional expected returns, the residuals should not be autocorrelated and the variance ratios should therefore equal 1 at all horizons. See the main text for the precise calculation behind the factor residuals. The sample is monthly from July 1963 to June 2017.

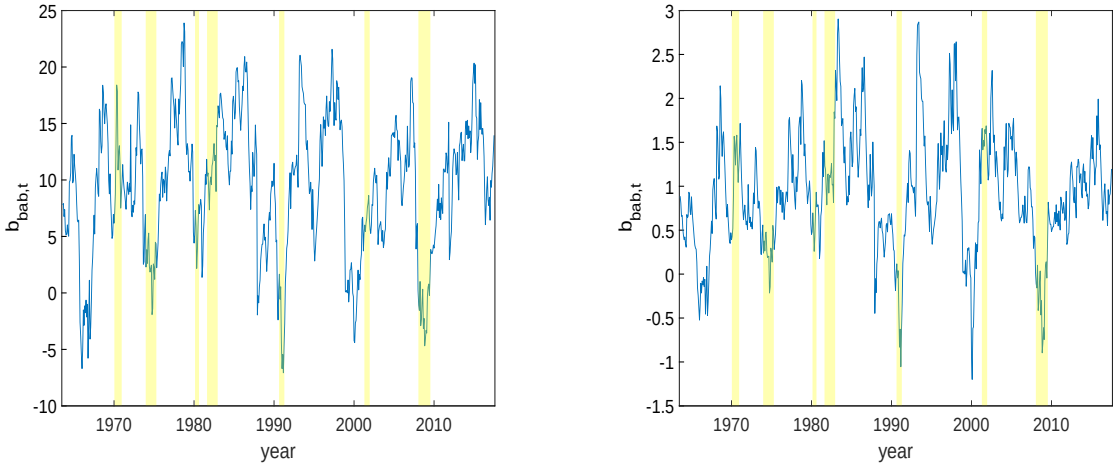
Figure 3
Pricing errors for model with latent b_t



The figure shows the pricing errors for the MKT and BAB factors vs. return horizon from the model $M_t = a_t - b_m \cdot MKT_t - b_{bab,t} \cdot BAB_t$ (see main text for details). Holding period h pricing errors are $E_T(M_{t,t+h}(R_{t,t+h} - R_{f,t,t+h}))$, 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 the h-period holding return for the (1-period) riskless bond. The population average of a correctly specified model is 0. The SDF is estimated using 1-month returns, as well as the 12 and 24 month horizon excess factor returns. The sample is from July 1963 to June 2017.

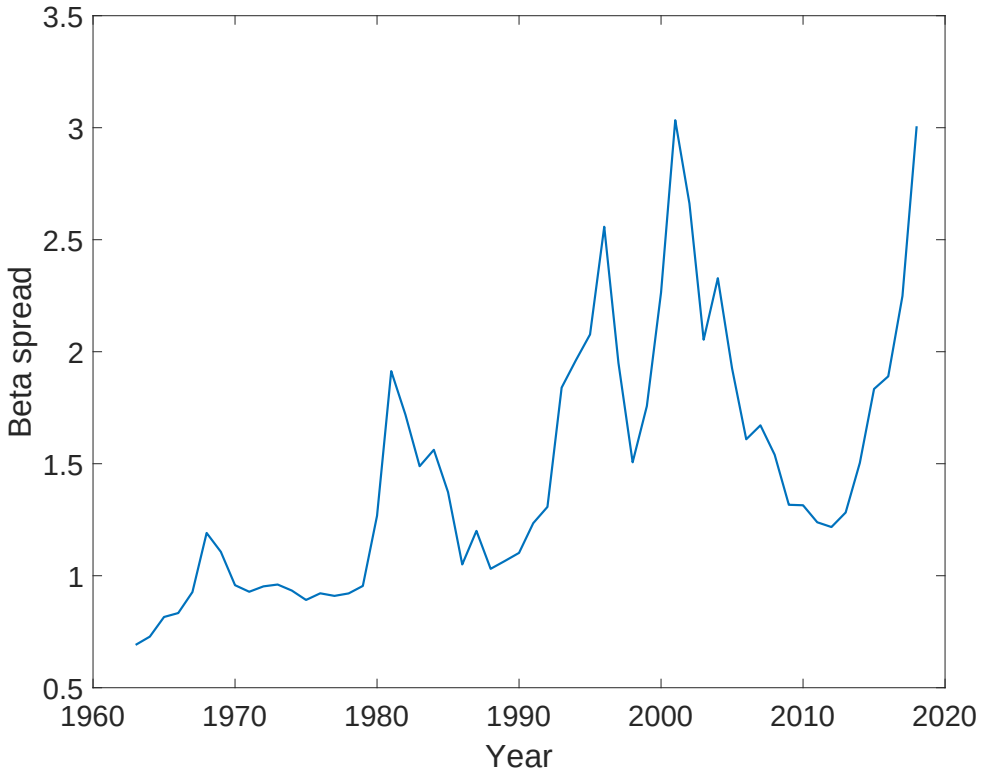
Figure 4
Time-series of $b_{bab,t}$ and the price of risk, $\lambda_{bab,t}$

(A) Time-series of $b_{bab,t}$ (B) Time-series of $\lambda_{bab,t}$



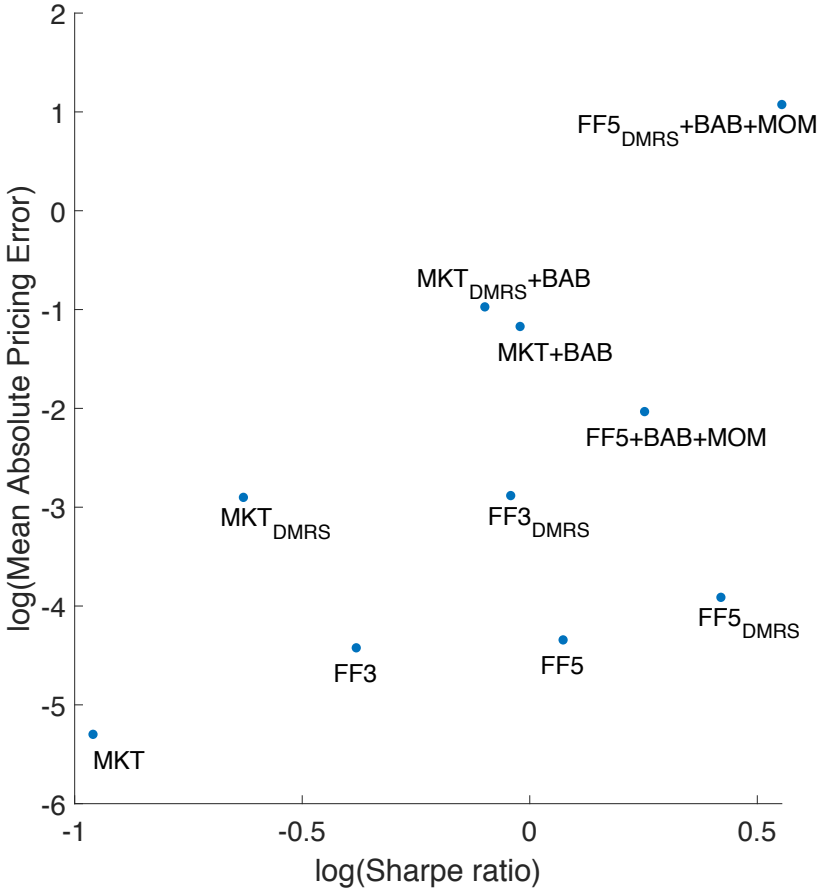
Panel A shows the filtered time series of $b_{bab,t}$ (blue line) along with NBER recession indicators (yellow bars) for the estimated model $M_{t+1} = a_t - b_m \cdot MKT_{t+1} - b_{bab,t} \cdot BAB_{t+1}$ (see main text for details), using long-run excess factor returns in the model estimation. Panel B shows the annualized conditional price of risk (Sharpe ratio) of the BAB shock, denoted $\lambda_{bab,t}$, where $\lambda_{bab,t} = b_{bab,t} \cdot \sqrt{12 \cdot V_t(BAB_{t+1})}$). The sample is from July 1963 to June 2017.

Figure 5
Beta spread between the beta-sorted decile portfolios



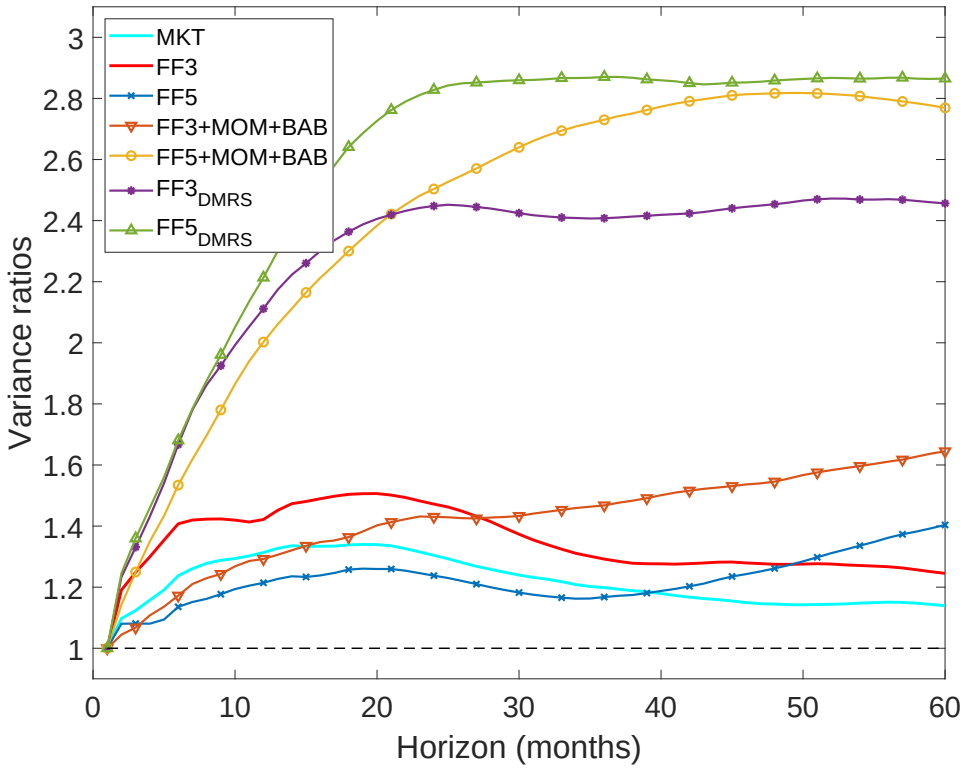
The figure shows the difference between the conditional market betas of the top minus the bottom beta-sorted decile portfolios as defined in the main text ($\Delta_t = (\beta_{h,t}\beta_{\ell,t})^{-1}(\beta_{h,t} - \beta_{\ell,t})$). The betas of the beta sorted portfolios are obtained from Ken French's website. For each portfolio, the conditional beta is the market beta of the constituent firms estimated using the past 60-month monthly return window, shrunk towards 1 (the value-weighted average beta) using a weight of 0.6 on the historical estimate and 0.4 on 1.

Figure 6
Max Sharpe ratio of single-horizon factor model vs. multi-horizon pricing errors



The figure plots the log (annualized) maximal Sharpe ratio of commonly used models factor models against the log of the (annualized) mean absolute pricing error (MAPE) of the corresponding model. The max Sharpe ratios are calculated as the sample Sharpe ratio (annualized) of the monthly returns to the mean-variance efficient combination of the factors in a given model. The MAPEs are computed by evaluating the fit of the models estimated using only 1-month returns on the excess factor returns at the 1-, 12-, and 24-month horizons (see the main text for details). The factor abbreviations are: MKT (the market factor), BAB (the betting-against-beta factor), FF3 (the Fama-French 3-factor model), FF5 (the Fama-French 5-factor model), MOM (the momentum factor). The DMRS subscripts refer to the Daniel et al (2017) unpriced-risks-hedged versions of the original factors. The sample is from July 1963 to June 2017.

Figure 7
Variance ratios of excess returns, actual vs. theoretical. All models



The figure shows the variance ratio $VR(h) = \frac{V\left(\sum_{j=1}^{h} \widetilde{\varepsilon}_{t+j}\right)}{h \times V(\widetilde{\varepsilon}_{t+1})}$ across horizons for each models' implied MVE portfolio return residuals, $\widetilde{\varepsilon}$. The null hypothesis from each model implies that variance ratios should equal 1 for all horizons. The factor abbreviations are: MKT (the market factor), BAB (the betting-against-beta factor), FF3 (the Fama-French 3-factor model), FF5 (the Fama-French 5-factor model), MOM (the momentum factor). The DMRS subscripts refer to the Daniel et al (2017) unpriced-risks-hedged versions of the original factors. The sample is from July 1963 to June 2017.

Table 1: Multi-horizon GMM tests
Models with the Market and Betting-Against-Beta factors

	Single-horizon estimation		Multi-horizon estimation	
	$F = MKT$	$F = [MKT, BAB]^\top$	$F = MKT$	$F = [MKT, BAB]^\top$
a	1.012***	1.082***	1.019***	0.995***
	(0.010)	(0.041)	(0.007)	(0.007)
b_m	2.506**	3.000***	2.386***	0.400
	(1.080)	(1.131)	(0.842)	(0.538)
b_{bab}		8.226***		1.722***
		(2.474)		(0.627)
p-value GMM J-test	n/a	n/a	0.663	0.000***
Factor MAPE	0.005	0.310	0.004	0.044
Max. price of risk	0.383	0.979	0.365	0.198

The table gives results from two-stage GMM estimation of two factor models: the market model, with excess returns to the market as the factor (MKT), and the two-factor model with the Market and Betting-Against-Beta (BAB) factors. The SDF is of the form $M_t = a - b^\top F_t$. "Single-horizon estimation" means the model is estimated using only monthly returns to the factor(s) and the risk-free rate. The "Multi-horizon estimation" additionally uses excess returns to the factor(s) at the 12- and 24-month horizons. The estimated model parameters are given along with GMM standard errors in parentheses. The maximal price of risk is the annualized standard deviation of the estimated monthly stochastic discount factor divided by its mean. Factor MAPE is the annualized mean absolute pricing error for the excess factor returns across horizons 1, 12, and 24 months. The p-values are for the J-tests from the GMM estimation (n/a for the exactly identified "single-horizon test" cases). The GMM procedure corrects for autocorrelation, cross-correlation, and heteroscedasticity in the moments using the Newey-West method with lag length equal to 1.5 times the maximum horizon or 12 months, whichever is larger. The sample is from July 1963 to June 2017. Three asterisks denote statistical significance at the 1% level, two at the 5% level, and one at the 10% level.

Table 2: Estimated Market and BAB factor model with time-varying SDF coefficient

Market SDF loading: b_m	1.283
BAB SDF loading: $b_{bab,t}$	
- $\hat{b}$	9.222
- $\hat{\phi}$	0.909
- $\hat{\sigma}$	2.520
J-test p-value	0.247
Factor MAPE	0.013
St.Dev. of $b_{bab,t}$	6.042
Annualized mean of $\lambda_{bab,t}$	0.911
St.Dev. of annualized $\lambda_{bab,t}$	0.661

The table gives results from two-stage GMM estimation of the two-factor model with the market (MKT) and Betting-Against-Beta (BAB) factors. The loading on the BAB factor, $b_{bab,t}$, follows a latent AR(1) process ($b_{bab,t} = (1-\phi)b + \phi \cdot b_{bab,t-1} + \eta_t$). The test assets are the 1-, 12-, and 24-month excess returns to the factors, as well as the short-term real risk-free rate. The annualized conditional price of risk (Sharpe ratio) of the BAB shock is denoted $\lambda_{bab,t}$ and equals $b_{bab,t} \cdot \sqrt{12 \cdot V_t(BAB_{t+1})}$). The table reports its mean and standard deviation. Factor MAPE is the annualized mean absolute pricing error for the excess factor returns across horizons 1, 12, and 24 months. The p-values are for the J-tests from the GMM estimation. The GMM procedure corrects for autocorrelation, cross-correlation, and heteroscedasticity in the moments using the Newey-West method with lag length equal to 1.5 times the maximum horizon or 12 months, whichever is larger. The sample is from July 1963 to June 2017.

Table 3: Dynamics of time-varying BAB loading: Relation to standard instruments

Panel A: $b_{bab,t}$ regressed on	Coefficient	(std. error)
NBER recession dummy	-3.483^{*}	(2.032)
Beta decile spread	1.462^{**}	(0.678)
Conditional variance of market return	-10.777^{***}	(3.129)
Conditional variance of BAB return	-10.071^{*}	(5.427)
Log market dividend-price ratio	5.007^{**}	(2.457)
Term Spread (10yr - 1yr)	0.708	(0.459)
R^2	0.273	
Panel B: $\lambda_{bab,t}$ regressed on:		
NBER recession dummy	-0.091	(0.068)
Beta decile spread	0.050^{**}	(0.022)
Conditional variance of market return	-0.417^{***}	(0.157)
Conditional variance of BAB return	0.332	(0.302)
Log market dividend-price ratio	0.187^{**}	(0.090)
Term Spread (10yr - 1yr)	0.026^{*}	(0.015)
R^2	0.200	

The table shows regressions of the time-varying SDF coefficient $b_{bab,t}$ and the price of risk for the BAB-shock ($\lambda_{bab,t} = b_{bab,t} \times \sqrt{12V_t(BAB_{t+1})}$) regressed on common conditioning variables: an NBER recession dummy, the spread in market beta between the top minus bottom decile portfolios sorted on market beta, the conditional variance of market and BAB 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 from July 1963 to June 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 4: GMM tests of other benchmark factor models

	Standard Factors				
	MKT	MKT+ BAB	FF3	FF5	FF5+ BAB+MOM
Panel A:					
Single-horizon estimation:					
Max. price of risk (annualized)	0.383	0.979	0.683	1.076	1.287
Factor MAPE (annualized)	0.005	0.310	0.012	0.013	0.131
Multi-horizon estimation:					
Max. price of risk (annualized)	0.365	0.198	0.528	1.165	2.998
Factor MAPE (annualized)	0.004	0.044	0.006	0.038	0.021
p-value GMM J-test	0.663	0.000^{***}	0.824	0.000^{***}	0.000^{***}

	DMRS Factors				
	MKT	MKT + BAB	FF3	FF5	FF5+ BAB+MOM
Panel B:					
Single-horizon estimation:					
Max. price of risk (annualized)	0.533	0.906	0.959	1.522	1.740
Factor MAPE (annualized)	0.055	0.378	0.056	0.020	2.929
Multi-horizon estimation:					
Max. price of risk (annualized)	0.119	0.183	0.510	1.184	0.196
Factor MAPE (annualized)	0.036	0.047	0.012	0.023	0.047
p-value GMM J-test	0.000^{***}	0.000^{***}	0.063^{*}	0.000^{***}	0.000^{***}

The table gives the p-values from the J-tests from two-stage GMM estimation of the stochastic discount factor implied by each factor model. The stochastic discount factor is assumed to be affine in the excess factor returns. The test assets are excess returns to each factor in the model at hand at horizons 1-, 12-, and 24-month horizons, in addition to the short-term real risk-free rate. Factor MAPE is the annualized mean absolute pricing error for the factors in the model. The p-values are from the J-tests from the GMM estimation. The GMM procedure corrects for autocorrelation, cross-correlation, and heteroskedasticity in the moments using the Newey-West method with lag length equal to 1.5 times the maximum horizon or 12 months, whichever is larger. The sample is from July 1963 to June 2017. The statistical significance of the parameters and the J-tests are given as follows: Three asterisks denote significance at the 1% level, two at the 5% level, and one at the 10% level.

Table 5: GMM tests of factor models using alternative horizons

	Standard Factors				
	MKT	MKT+ BAB	FF3	FF5	FF5+ BAB+MOM
Panel A:					
Factor MAPE (annualized)	0.007	0.115	0.012	0.010	0.033
p-value GMM J-test	0.800	0.000^{***}	0.000^{***}	0.000^{***}	0.000^{***}
	DMRS Factors				
	MKT	MKT + BAB	FF3	FF5	FF5+ BAB+MOM
Panel B:					
Factor MAPE (annualized)	0.028	0.092	0.026	0.014	0.024
p-value GMM J-test	0.000^{***}	0.000^{***}	0.003^{*}	0.000^{***}	0.001^{***}

The table gives the p-values from the J-tests from two-stage GMM estimation of the stochastic discount factor implied by each factor model using shorter horizons relative to earlier results as additional moments. The stochastic discount factor is assumed to be affine in the excess factor returns. The test assets are excess returns to each factor in the model at hand at horizons 1-, 2-, 6-, and 12-month horizons, in addition to the short-term real risk-free rate. Factor MAPE is the annualized mean absolute pricing error for the factors in the model. The p-values are from the J-tests from the GMM estimation. The GMM procedure corrects for autocorrelation, cross-correlation, and heteroscedasticity in the moments using the Newey-West method with lag length equal to 1.5 times the maximum horizon or 12 months, whichever is larger. The sample is from July 1963 to June 2017. The statistical significance of the parameters and the J-tests are given as follows: Three asterisks denote significance at the 1% level, two at the 5% level, and one at the 10% level.

Table 6: Long-run factor mispricing as instruments for monthly factor returns: GRS tests

	Standard Factors				
	MKT	MKT + BAB	FF3	FF5	FF5+ BAB+MOM
Panel A:					
Max Sharpe ratio of factors	0.388	0.995	0.690	1.089	1.308
Max Information ratio	0.329	0.722	0.576	0.491	0.781
p-value GRS test	0.049**	0.000***	0.005***	0.052*	0.008***
p-value, out-of-sample instruments	0.206	0.001***	0.001***	0.234	0.009***
	DMRS Factors				
	MKT	MKT + BAB	FF3	FF5	FF5+ BAB+MOM
Panel B:					
Max Sharpe ratio of factors	0.540	0.920	0.971	1.544	1.768
Max Information ratio	0.433	0.759	0.793	0.823	0.811
p-value GRS test	0.009***	0.000***	0.000***	0.000***	0.002***
p-value, out-of-sample instruments	0.008***	0.000***	0.000***	0.001***	0.050**

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 horizons considered are 2, 6, 12, and 24 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 from July 1963 to June 2017. Three asterisks denote significance at the 1% level, two at the 5% level, and one at the 10% level.

Table 7: Long-run factor mispricing using observable conditioning information: GRS tests

	MKT			DMRS MKT	
Panel A:	Baseline	RV managed	d/p managed	Baseline	RV managed
Max SR of model factors	0.388	0.438	0.410	0.540	0.656
Max Information Ratio	0.329	0.290	0.344	0.433	0.594
p-value GRS test	0.049**	0.082*	0.040**	0.009***	0.000***
	Fama-French 3-factor model			DMRS 3-factor	
Panel B:	Baseline	RV managed	d/p managed	Baseline	RV managed
Max SR of model factors	0.690	0.752	0.780	0.971	1.040
Max Information Ratio	0.576	0.521	0.738	0.793	0.760
p-value GRS test	0.005***	0.014***	0.000***	0.000***	0.000***
	Fama-French 5-factor model			DMRS 5-factor	
Panel C:	Baseline	RV managed	d/p managed	Baseline	RV managed
Max SR of model factors	1.089	1.356	1.198	1.544	1.642
Max Information Ratio	0.491	0.395	0.525	0.823	0.863
p-value GRS test	0.052*	0.136	0.040**	0.000***	0.000***

The table gives the results from GRS regressions of managed factor returns, where the time-varying factor weights are based on lagged multihorizon pricing errors, on the given model's factors. The horizons considered are 2, 6, 12, and 24 months. For each factor model, the test assets are managed positions in the factors of the factor model at hand. So, for the CAPM, the test assets are four managed portfolios (corresponding to the four pricing error horizons) in the MKT factor. The baseline case repeats the quantities found in Table 1, whereas the "RV managed" and "d/p managed" cases refer to the use of additional managed factors in the factor model, as described in the main text, to account for relevant conditioning information. "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 from July 1963 to June 2017. Three asterisks denote significance at the 1% level, two at the 5% level, and one at the 10% level.