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: November 18, 2018
Comments welcome.

Abstract

We propose testing of 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, that is, an SDF with constant factor loadings. We reject all state-of-the-art models that imply high maximum Sharpe ratios (SR) 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 SR in the single-horizon setting. Model misspecification manifests itself in strong intertemporal dynamics of the factor loadings in the SDF representation. We trace the dynamics of loadings to 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 for comments on earlier drafts and participants in the seminars and conferences sponsored by LAEF, NHH, UCLA. The latest version is available at https://sites.google.com/site/mbchernov/xxx.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. Indeed, under the null of the model, the SDF applies 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.

Most economic decisions, such as, consumption-savings plans, private equity or corporate investments, 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 the implications of an SDF model for intermediate to long horizons. 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 returns at different horizons but does not use them to explicitly test models of the SDF (e.g., Kamara, Korajczyk, Lou, and Sadka, 2016).

In this paper we argue that asset-pricing models should be tested using a set of multi-horizon returns (MHR). That is, models should be tested using, say, 1-, 12-, and 24-month returns of a set of test assets. First, we show that such a test amounts to a stringent evaluation of conditional implications of a model. We show that the MHR test 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 misspecification.

Second, we illustrate the degree of possible misspecification and detect its origins in the context of linear factor models, $\text{SDF} = a - b^\top F$. We ask if a model that is typically evaluated using single-horizon returns is capable of pricing its own factors at longer horizons. We find that models deemed successful on the basis of single-horizon returns (thus having factors that yield a high maximal Sharpe ratio) are more likely to have large MHR errors. Further, we show that dynamics of the factors $F$, as manifested by multi-horizon variance ratios, are grossly inconsistent with the dynamics implied by the constant $b$ specification. In other words, current state-of-the-art models of the risk-return trade-off, typically estimated at a monthly or quarterly frequency, do a poor job accounting for the risk-return trade-off at longer horizons.

This result rejects specifications with constant loadings $b$ in the SDF. Because we evaluate the models using the pricing factors themselves, it follows that there must exist time-varying $a$’s and $b$’s that price MHR correctly. We model $b$ as a latent process and estimate it 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. Thus, testing the SDF jointly across multiple horizons reveal misspecification in the temporal dynamics of the model.

The evidence prompts us to investigate the economic origins of the time variation in $b$. Sometimes, $b$ is close to a constant. In particular, this 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. The goal of much of the factor literature is to find factors that span the unconditional mean-variance efficient portfolio. If this is achieved, $b$ is constant. Thus, as one considers models with increasingly high short-run maximal Sharpe ratios, it would be natural to find less evidence of time-variation in $b$. In this sense, our results are surprising.

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, as well as the type of double-sorts, for instance by size, are ad hoc. We show that these ad hoc 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 $\tilde{F}_{t+1} = b_t^\top F_{t+1}$ that would feature a constant (unitary) loading in the SDF. While a search for the maximum one-period SR would favor such a $\tilde{F}_{t+1}$, there is no mechanism in the current factor construction practice that provides information about how to amend the cross-sectional sorts. We argue that using MHR in tests of the pricing kernel that we develop in this paper provides valuable information about this type of misspecification.

Our paper is complementary to Daniel, Mota, Rottke, and Santos (2018), who argue that the Fama-French method of factor construction is not well-suited to distinguish priced from unpriced risks. Instead, we construct our tests using MHR on the factors themselves and assume the shocks are the right, priced shocks for each model. We analyze, for each model, if there is persistent misspecification over time within the model (i.e., we test the constant $b$ assumption). Thus, we focus on timing effects in the factor construction. In fact, we find large misspecification in the Daniel, Mota, Rottke, and Santos (2018) factors over time, despite the fact that these factors deliver much higher short-run Sharpe ratios than the original factors.

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 – both conditional one-factor models – using proxies for the conditioning information related to aggregate discount rates to arrive at unconditional multi-factor models with constant factor loadings in the SDF. Moreira and
Muir (2017) find that factors normalized by their conditional variance outperform the original factors. This is the same as saying $b$ varies with the inverse of the conditional factor variance. Indeed, when we regress our estimated, latent $b_t$ we find that the conditional factor variance indeed is significantly negatively related to $b_t$ and that the dividend price ratio is positively related to $b_t$. However, as mentioned, the $R^2$ of this regression is low, indicating that most of the variation is due to other sources.

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

Kamara, Korajczyk, Lou, and Sadka (2016) consider factor models in a linear beta-pricing models, like the standard CAPM equation, at different return horizons. However, unlike the SDF, beta-pricing models do not scale with horizon. That is, if the CAPM holds at the monthly horizon, it does not hold at the annual horizon even under the null that the CAPM is true. This is well-known and first pointed out in Levhari and Levy (1977). Further, the SDF framework allows us to understand MHR tests as testing conditional implications of the model.

Hansen and Scheinkman (2009) analyze the infinite-horizon risk-return trade-off, and show how this is valuable relative to what they term “local” risk pricing, which is what typically is considered in the literature. Our empirical work considers the risk-pricing across short and intermediate horizons jointly (up to 24 months). Due to the limited samples available we cannot reliably work with longer horizons without imposing more structure. However, our methodology and the insight that multi-horizon returns tests conditional model implications are general.

The structure of the paper is as follows. Section 2 describes the implications of adding MHR to existing tests and presents an empirical in-depth example. Section 3 discusses what is missing from current models. Section 4 analyzes the impact of cross-sectional characteristics-sort on the factor construction. Sector 5 gives results from a large roster of factor models and shows that the effects we document in Sections 2 and 3 are pervasive. We also provide several robustness tests, using alternate horizons, observable instruments, and GRS tests. Section 6 concludes.

2 Testing asset pricing models using MHR

The stochastic discount factor (SDF) encodes the risk-return trade-off at different horizons. 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 have received little attention in the previous work, despite the relevance for theoretical models and investment practice.

2.1 Preliminaries

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 $R_{t,t+h}$.

Let the one-period stochastic discount factor (SDF) from time $t$ to $t + 1$ spanned by the set of traded assets 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:

$$M_{t,t+h} = \prod_{j=1}^{h} M_{t+j-1,t+j},$$

$$R_{t,t+h} = \prod_{j=1}^{h} R_{t+j-1,t+j},$$

$$R_{f,t,t+h} = \prod_{j=1}^{h} R_{f,t+j-1,t+j}.$$
LOOP still holds:

\[ E_t(M_{t,t+h} R_{t,t+h}) = 1, \quad (1) \]

or, for excess returns,

\[ E_t(M_{t,t+h}(R_{t,t+h} - R_{f,t,t+h})) = 0. \quad (2) \]

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, consider two-period returns from time \( t - 1 \) to time \( t + 1 \). We can write:

\[ 1 = E(M_{t-1,t+1} R_{t-1,t+1}) = E(M_{t-1,t} R_{t-1,t}) \cdot E(M_{t,t+1} R_{t,t+1}) + Cov(M_{t-1,t} R_{t-1,t}, M_{t,t+1} R_{t,t+1}). \]

Therefore \( Cov(M_{t-1,t} R_{t-1,t}, M_{t,t+1} R_{t,t+1}) = 0 \). Define pricing errors as realized \( M \cdot R \) minus its conditional expectation (which is 1 since \( R \) is a gross return). Thus, the condition that the covariance term equals zero implies that past pricing errors cannot predict future pricing errors. A generalization to \((h + 1)\)-period returns, under covariance-stationarity and assuming the model unconditionally prices one-period returns correctly, i.e., \( E(M_{t,t+1} R_{t,t+1}) = 1 \), is

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

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.\(^1\)

\(^1\)In particular, 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 \( h = 1, 2, \ldots \), 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.
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). 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 that features an SDF, in this paper we focus on linear factor models

$$M_{t,t+1} = a - b^T F_{t,t+1}, \quad (3)$$

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 test of the model. We will, however, also add multi-horizon excess factor returns to the test. 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:

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

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

$$M_{t,t+1} = a_t - b_t^T F_{t,t+1}, \quad (4)$$
that prices all test assets correctly.\footnote{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. 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 $t$. Thus, a rejection of the constant coefficient SDF with our set of test assets has a clear implication for the form of the alternative model.} Equation (2) implies that

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

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 the pricing kernel. Such a linear combination is the mean-variance efficient portfolio (Cochrane, 2004). The large literature on unconditional beta-pricing factor models (e.g., Carhart, 1997; Fama and French, 1993) is exactly concerned with finding factors that unconditionally span the unconditional mean-variance efficient portfolio. Thus, under the null of these models (that they price all asset returns), $b$ is constant.

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

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

where $\varepsilon_{t+1}$ is conditionally mean-zero, uncorrelated with $R_{f,t,t+1}^{-1}$, and has unit variance. Further, assume that

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

Then, we can represent the SDF in terms of $F$

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

Comparing equations (4) and (7), we see that

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

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

\section*{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.

In this case, the model (3) specializes to

\[ M_{t,t+1} = a - b_m M_{KT,t+1} - b_{bab} BAB_{t,t+1}, \]

where \( M_{KT,t+1} = R_{m,t,t+1} - R_{f,t,t+1} \) 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 \( b_{bab} = 0 \). 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 (where only the MKT factor is included) 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. For the MKT-model, 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 test, 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. Thus, in this case adding MHR has huge effects on the estimated parameters. The difference between the single- and multi-horizon tests hints at severe model misspecification. The \( J \)-test confirms this, rejecting the model at the 1% level.

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. 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 manifestation of the factor dynamics unaccounted for by our model. The purpose of this section is to elucidate what is missing in the considered specification.

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 \( \tilde{F}_{t,t+1} = b^T F_{t,t+1} \) is an unconditionally mean-variance efficient (MVE) portfolio (Hansen and Richard, 1987). Equation (5) implies for this MVE portfolio:

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

Therefore, the portfolio dynamics can be written as

\[
\tilde{F}_{t,t+1} = R_{f,t,t+1}V_t(\tilde{F}_{t,t+1}) + V_t^{1/2}(\tilde{F}_{t,t+1})\varepsilon_{t+1}.
\] (10)

Rejection of the null \( b_t = b \) implies that true dynamics of the portfolio are different.

3.2 Variance ratios

We use variance ratios to highlight the nature of the misspecification of dynamics of \( \tilde{F} \) in (10) implied by the model (3). Recall that the variance ratio at horizon \( h \) for a random variable \( X_{t,t+1} \) is:

\[
VR(h) = \frac{V\left(\sum_{j=1}^{h} X_{t+j-1,t+j}\right)}{h \times V(X_{t,t+1})}.
\] (11)
As is well known, if \(X\) is iid, then \(VR(h) = 1\) for any \(h\). If \(X\) is positively autocorrelated, then \(VR(h) > 1\), whereas if \(X\) is negatively autocorrelated, then \(VR(h) < 1\).

In our case, the appropriate benchmark for \(VR(h)\) is not 1, but the inter-horizon pattern implied by the dynamics of \(\tilde{F}\) implicit in (10). That pattern would be difficult to construct. Instead, we apply the test to \(X_{t,t+1} = \tilde{F}_{t,t+1} - E_t(\tilde{F}_{t,t+1})\). Under the null, this shock is equal to \(\tilde{F}_{t,t+1} - R_{f,t,t+1} V_t(\tilde{F}_{t,t+1})\) and should be uncorrelated over time. Thus, the null hypothesis implies that \(VR(h) = 1\) for any \(h\) for this model-implied shock. We use \(b\) estimated by GMM to construct \(\tilde{F}_{t,t+1}\) and estimate an EGARCH(1,1) to obtain \(V_t(\tilde{F}_{t,t+1})\) under the maintained assumption that the two factors (MKT and BAB) are uncorrelated. Note from the GARCH-coefficients in Table 2 that the conditional variances are quite persistent.

Continuing with the examples of section 2.5, we report variance ratios for the residuals 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 the autocorrelation implied by the model and the estimated conditional return variance.\(^3\) This spills over into too high positive autocorrelation in the pricing kernel. 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

To illustrate how much variation in \(b_t\) we are missing, we use MHR to estimate the SDF with time-varying \(b\):

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

We assume that \(b_{bab,t}\) follows a latent AR(1) process:

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

It then follows, to make the conditional expectation of the SDF equal to the inverse of the gross risk-free rate, that \(a_t\) can be written:

\[
a_t = R_{f,t,t+1}^{-1} + b_m^2 R_{f,t,t+1} V_t(MKT_{t,t+1}) + b_{bab,t}^2 R_{f,t,t+1} V_t(BAB_{t,t+1})
\]

While \(b_m\) could be time-varying as well, 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

\(^3\)In unreported results, we find that it is the BAB factor itself that has a strongly increasing variance ratio over time.
and BAB factors are conditionally uncorrelated. This is consistent with the BAB factor construction, which weights the underlying stocks by the inverse of estimated market betas so as 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 implied from (5).

Table 3 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 \( \sigma(1 - \phi^2)^{-1/2} = 6.04 \), which appears to be large relative to its mean of 9.22. To get a more intuitive interpretation for these numbers we can use Equation (8), which implies that the price of risk (Sharpe ratio) 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 4 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 variance of BAB. 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 \( b_t \). Equation (8) suggests that time-varying volatility of the factors could be the culprit. Because prices of risk are time-varying as well, it could be the case that the ratio is reasonably close to a constant. Certainly, the evidence about the MKT factor is pointing that way. The question is what could be the driving force behind the strong variation in BAB’s \( b_t \).

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

A related point is made by Daniel, Mota, Rottke, and Santos (2018), who hedge out non-priced risks in the factors. Our point is complementary. We do not consider tests where the issue is that the shock to the factor is contaminated by unpriced risks. Instead, we explicitly construct our tests such that this issue is off the table and focus on how time-dependence (or lack thereof) in factor returns induced by the factor construction procedure leads to long-run pricing errors. We use the BAB strategy as an example to develop this idea.

4.1 An example: betting against beta

We specialize the SDF in (6) to

\[
M_{t,t+1} = R^{-1}_{f,t,t+1} - \lambda_m \varepsilon_{m,t+1} - \lambda_{\text{bab},t} \varepsilon_{\text{bab},t+1},
\]

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:

\[
MKT_{t,t+1} = \sigma_m \lambda_m + \sigma_m \varepsilon_{m,t+1}.
\]

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:

\[
R_{h,t,t+1} - R_{f,t,t+1} = \beta_{h,t} MKT_{t,t+1} + [(1 - \beta_{h,t}) \lambda_{\text{bab},t} + (1 - \beta_{h,t}) \varepsilon_{\text{bab},t+1}],
\]

\[
R_{\ell,t,t+1} - R_{f,t,t+1} = \beta_{\ell,t} MKT_{t,t+1} + [(1 - \beta_{\ell,t}) \lambda_{\text{bab},t} + (1 - \beta_{\ell,t}) \varepsilon_{\text{bab},t+1}].
\]
Then, construct the BAB factor as in Frazzini and Pedersen (2014):

\[ BAB_{t,t+1} \equiv \beta_{h,t}^{-1}(R_{h,t,t+1} - R_{f,t,t+1}) - \beta_{\ell,t}^{-1}(R_{\ell,t,t+1} - R_{f,t,t+1}) = \Delta_t \lambda_{bab,t} + \Delta_t \varepsilon_{bab,t+1}, \]  

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 \( \Delta_t \). In this plot, we use the historical betas of the top and bottom decile beta-sorted portfolios as provided annually on Kenneth French’s webpage, shrunk towards 1 with a weight of 0.6 on the historical beta and a weight of 0.4 on 1. 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 very persistent dynamics in the conditional mean and volatility of BAB returns, per equation (12) and the empirical beta spread shown in figure 5.

If we write the SDF as a function of the MKT and BAB factors, we have:

\[ M_{t,t+1} = R_{f,t,t+1}^{-1} + \lambda_m^2 + \lambda_{bab,t}^2 - \lambda_m \sigma_m^{-1} MKT_{t,t+1} - \lambda_{bab,t} \Delta_t^{-1} BAB_{t,t+1}, \]  

which is a version of the general specification in (7). Note that the price of risk \( \lambda_{bab,t} \) has to be proportional to \( \Delta_t \) for \( b_{bab} \) to be constant. Our evidence is that this not the case. Further, note that while the sorting procedure chooses \( \Delta_t \) there is nothing explicit in the sorting procedure that ensures that it lines up correctly with \( \lambda_{bab,t} \).

A fix is possible, however. Create the new factor \( \widetilde{BAB}_{t,t+1} = \lambda_{bab,t} \Delta_t^{-1} BAB_{t,t+1} \). Then, the true unconditional model is

\[ M_{t,t+1} = \tilde{a} - \tilde{b}_m MKT_{t,t+1} - \widetilde{BAB}_{t,t+1}. \]

However, researchers specify (9) instead, which leads to the large MHR pricing errors we document in Section 2.

Empirically, as shown in table 4A, the beta spread is positively and significantly related to \( b_{bab,t} \), which indicates that the price of risk for the BAB shock, \( \lambda_{bab,t} \), is 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. Our example is stylized and there are more generally 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, consider Fama-MacBeth regressions as the benchmark method for factor construction. It is well-known that when running a cross-sectional regression of excess returns on lagged characteristic and an intercept term, the slope coefficient(s) represent long-short portfolio returns with portfolio weights on the underlying stocks as a function of the characteristics.

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

$$R_{i,t+1} = \beta_{i,t}^T F_{t+1} + \epsilon_{i,t+1},$$

where residuals are uncorrelated across stocks $i$ and time $t$. Note that the proposed SDF prices individual excess returns as well. If we knew $\beta_{i,t}$ we could run a cross-sectional regression at each time $t$ to retrieve the factors $F$:

$$R_{i,t+1} - R_{f,t+1} = \lambda_{0,t+1} + \lambda_{1,t+1}^T \beta_{i,t} + \epsilon_{i,t+1},$$

In this case, $\lambda_{1,t+1}$ gives the vector of long-short factor excess returns, $F_{t+1}$. 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, and this is what causes the issues we highlight in this paper.

For ease of exposition, but without loss of generality, assume that there is only one factor. The Fama-MacBeth cross-sectional regression

$$R_{i,t+1} - R_{f,t+1} = \hat{\lambda}_{0,t+1} + \hat{\lambda}_{1,t+1} X_{i,t} + \epsilon_{i,t+1},$$

then delivers the factor

$$\hat{\lambda}_{1,t+1} = \frac{\text{Cov}(\delta_{1,t} \beta_{i,t}, \beta_{i,t} F_{t+1})}{\text{Var}(\delta_{1,t} \beta_{i,t})} = \frac{\text{Var}(\beta_{i,t}) \delta_{1,t}^2 F_{t+1}}{\delta_{1,t}^2 \text{Var}(\beta_{i,t})} = \delta_{1,t}^{-1} F_{t+1},$$
where the variance and covariance are taken cross-sectionally across assets \( i \) at each time \( t \). Thus, the factor used by researchers is

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

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

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:

\[
M_{t+1} = a - b_{t} \hat{F}_{t+1},
\]

\[
b_{t} = b \delta_{1t}.
\]

This is exactly the same issue that we have highlighted generally in equation (7) and, specifically in the context of BAB, in equation (13). In sum, MHR as test assets is 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 Sharpe ratios (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. 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 the SDF. Such a portfolio has \( a \) and \( b \) in the SDF (4) and lies on the unconditional mean-variance efficient frontier.
Table 5 reports the same GMM tests we have implemented for MKT, and MKT+BAB (Table 1). 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 as a collection of factors that spans the unconditional MVE portfolio should in fact have 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 (8).

Related, the price of risk for the other 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 basically observe the extreme pattern that we’ve documented in the MKT+BAB case for all models, except for the FF3 and MKT models. These models have variance ratios that are relatively close to one, as implied by the null hypothesis of these models, whereas the other models exhibit too high autocorrelation of factor returns relative to the volatility dynamics. Notably, the DMRS versions of the original models, which have even higher short-run Sharpe ratios, have very strongly increasing variance ratios. Thus, it seems the refinement of the original factors has come at a cost of introducing persistent misspecification that shows up in MHR and the models’ variance ratios.

In sum, we document pervasive, persistent misspecification in the current, most successful factor models, as judged by their one-month maximal Sharpe ratios. 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 6 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 Gibbons, Ross and Shanken (1989). GRS is a test of whether a given set of (1-period) factors span the mean-variance frontier of a set of (1-period) test assets. For our purposes, the test assets will be trading strategies in the factors themselves. In particular, the test asset $i = k \cdot h$ is a strategy in factor $k$, where the time-varying weights are given by lagged $h$-period discounted returns in factor $k$, i.e.,

$$R_{i,t,t+1}^e \equiv z_{k,t}^{(h)} F_{k,t+1}$$
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 $t$.\(^4\) Each of the test assets are one-period excess returns as opposed to the multi-period excess returns used in the GMM-test. Loosely speaking, we are testing whether lagged discounted returns predict changes in the one-period risk-return trade-off in the factors.

Using only one-period excess returns is a test of whether

$$E_t[(a_t - \bar{b}_t F_{t,t+1}) R_{e_i,t,t+1}] = 0$$

which is equivalent to testing

$$E_t[(1 - \bar{b}_t F_{t,t+1}) R_{e_i,t,t+1}] = 0$$

where $\bar{b}_t = \frac{b_t}{a_t}$. The GRS test is therefore related to a test of whether $\bar{b}_t$ is constant.

Because the instruments $z_{k,t}^{(h)}$ are constructed using the candidate pricing kernel, 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 pricing kernel, 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 using the GRS test.

Table 7 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 use 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 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, loosely speaking, is the maximum Sharpe ratio unexplained by (orthogonal to) the factors. To get a sense of magnitudes, it is natural to compare the maximum IR to the Sharpe ratio of the factors. In FF5, the maximum IR is 0.49 compared to a Sharpe ratio of the factors of 1.09. For all the other models, the maximum IR is even larger relative to the Sharpe ratios 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 factor based models. Except for the out-of-sample $p$-value for DMRS5+BAB+MOM of 0.05, all $p$-values are below 1 percent. Similarly, all maximum IRs are greater than 40 percent of the corresponding factor Sharpe ratios.

\(^4\)Note that the weights in a factor is unrestricted since the factors are excess returns.
5.2.3 Conditioning information

Given the vast evidence of return and variance predictability of many factors (references), it might not be too surprising that we reject most static models. It will therefore be of some interest to see whether adding conditioning information alleviates the problem. To investigate this, we consider expanding the factor models to allow for time-varying $b$ where the $b$’s are affine functions of conditioning information. In other words, the $b$ of factor $k$ is assumed to of the following form

$$b_{k,t} = b_{k,0} + b_{k,1}^T x_{k,t},$$

where $x_{k,t}$ is some random variable believed to predict the conditional risk-return trade-off known to investors at time $t$.

With this specification of $b_t$ in hand, it is straightforward to formulate a factor model for the pricing kernel with constant $b$

$$M_{t,t+1} = a_t - b_t^T F_{t,t+1} = a_t - \tilde{b}_{t+1}$$

where $\tilde{F}_{t,t+1}$ contains all factors of the type $(1, x_{k,t}^T) F_{k,t+1}$. As an example, consider the case where $F$ is a single factor and $x$ is a single conditioning variable. Then

$$\tilde{F}_{t,t+1} = \begin{bmatrix} F_{t,t+1} \\ x_{t}F_{t,t+1} \end{bmatrix}$$

and

$$M_{t,t+1} = a_t - \begin{bmatrix} b_0 & b_1 \end{bmatrix} \begin{bmatrix} F_{t,t+1} \\ x_{t}F_{t,t+1} \end{bmatrix}$$

We consider two common conditioning variables for each factor: the dividend-price ratio (when sufficient data is 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 8 reports the $p$-values of GRS tests of extended factor models using either $dp$ or $RV$ as conditioning information for each factor. The models considered are MKT, FF3 and FF5 and their DMRS counterparts. For the latter, only $RV$ is used as conditioning information. For comparison, we also report the results of the static models tested in 7.

Panel A shows that conditioning on $RV$ makes the GRS $p$-value go from marginally significant at the 5 percent level to a $p$-value of 8.2 percent, whereas conditioning on $dp$ ratio actually lowers the $p$-value to 4 percent. Also worth noting, is that the maximum IR are similar to that of the baseline model, although, as a fraction of factor Sharpe ratio, they
decline somewhat due to the higher Sharpe ratio afforded by adding a factor. For the DMRS MKT model, we still reject at the 1 percent level and IR increases both in absolute terms and as a fraction of factor Sharpe ratio.

From Panel B we see that the FF3 is rejected both when conditioning on \( dp \) and \( RV \), although the \( p \)-value of the GRS test is now marginally insignificant at the 1 percent level in the \( RV \) case. Again, IRs are still large. The DMRS3 model conditioned on \( RV \) is rejected on the 1 percent level with IR similar to the baseline.

In Panel C, the \( p \)-value of FF5 goes from marginally insignificant at the 5 percent level for the baseline to 13.6 percent in the \( RV \)-conditioned case and the IR also decline from 0.491 to 0.395. On the other hand, for the \( dp \) case, the \( p \)-value falls to 4 percent and IR increases to 0.52. For the DMRS5 model, the \( RV \)-conditioned model performs similar to the baseline and both are rejected at the 1 percent level. In summary, these conditioning variables do not seem to fix the problem.

In most of what we have done so far, we have assumed there exist a linear combination of factor excess returns that are perfectly correlated with an admissible pricing kernel, which is equivalent to assuming constant \( a \) and \( b \) in (4). The case of time-varying \( b \) would imply there exist a trading strategy in the factor excess returns that yields a higher Sharpe ratio and thus takes us closer to an admissible pricing kernel. We therefore have a “recipe” for generating better factors - let factor \( i \) be \( \tilde{F}_{i,t,t+1} = b_{i,t} F_{i,t,t+1} \).

The case with time-varying \( a \) is more subtle and slightly harder to deal with. Note that in (4) the excess return \(-b_t^\top \tilde{F}_{t,t+1}\) is perfectly correlated with an admissible SDF, but if \( a_t \) is time-varying, there does not exist a trading strategy in the factors that is unconditionally perfectly correlated with \( M \). That prompts us to wonder whether the rejections of the models are due only to an assumption of constant \( a \) or if it is also due to time-varying \( b \). The GRS test does not require \( a_t \) to be constant.

6 Conclusion

The literature typically characterizes the risk-return trade-off, or, equivalently, the SDF, at a single horizon. This singular focus is somewhat surprising given that the multi-horizon risk-return trade-off, as encoded in the SDF, is an important input to a range of economic decisions, including long-lived agents’ consumption-savings decisions, private equity investments, and corporate investments. In this paper, 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 models – 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.

We trace the failure of these empirical models to persistent conditional mispricing. The degree of mispricing is, broadly speaking, increasing in the short-run maximal Sharpe ratio (SR) of the models. Economically, as one builds up models that increase one-period SR and approach the mean-variance efficient portfolio, one may think factor loadings in the SDF approach a constant as would be the case for the true mean-variance efficient portfolio. Our evidence shows that this is not the case. The facts we uncover presents a challenge for asset pricing models.
References


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.
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.
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 $\lambda_{bab,t} = b_{bab,t} \cdot \sqrt{12} \cdot \sqrt{V_t(BAB_{t+1})}$. The sample is from July 1963 to June 2017.
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 \( \Delta_t = (\beta_{h,t} - \beta_{l,t}) \). 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.
The figure shows the ratio \( \text{Var}(r_{t,t+h})/(h \cdot \text{Var}(r_{t,t+1})) \), where \( r_{t,t+h} = \log R_{t,t+h} \) denotes \( h \)-period log-returns on an asset (portfolio). Each gross portfolio return \( R_{t,t+1} \) is formed by adding \( R_{f,t,t+1} \) to the monthly in-sample mean-variance efficient combination of a set of excess returns, e.g. \( R_{FF3,t,t+1} = R_{f,t,t+1} + \omega_{FF3}^\top FF3_{t,t+1} \), where \( \omega_{FF3} \) maximizes monthly Sharpe ratio of \( \omega^\top FF3_{t,t+1} \). Because \( \omega_{FF3} \) is only unique up to a choice of scale, we normalize it such that all portfolios have the same variance as the market. The \( h \)-period gross return is \( R_{t,t+h} = \prod_{j=1}^{h} R_{t+j-1,t+j} \), which is the return from holding a portfolio that rebalances to \( \omega_i \) every month from \( t \) to \( t+h \). If factor returns were i.i.d. the ratio should be 1 for all horizons.
### Table 1: Multi-horizon GMM tests

Models with the Market and Betting-Against-Beta factors

<table>
<thead>
<tr>
<th></th>
<th>Single-horizon estimation</th>
<th>Multi-horizon estimation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$F = MKT$</td>
<td>$F = [MKT, BAB]^\top$</td>
</tr>
<tr>
<td>$a$</td>
<td>1.012***</td>
<td>1.019***</td>
</tr>
<tr>
<td></td>
<td>(0.010)</td>
<td>(0.007)</td>
</tr>
<tr>
<td>$b_m$</td>
<td>2.506**</td>
<td>2.386***</td>
</tr>
<tr>
<td></td>
<td>(1.080)</td>
<td>(0.842)</td>
</tr>
<tr>
<td>$b_{bab}$</td>
<td>8.226***</td>
<td>1.722***</td>
</tr>
<tr>
<td></td>
<td>(2.474)</td>
<td>(0.627)</td>
</tr>
<tr>
<td>$p$-value GMM $J$-test</td>
<td>n/a</td>
<td>0.663</td>
</tr>
<tr>
<td></td>
<td>n/a</td>
<td>0.000***</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Multi-horizon estimation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor MAPE</td>
<td>0.004</td>
</tr>
<tr>
<td>Max. price of risk</td>
<td>0.365</td>
</tr>
</tbody>
</table>

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 pricing kernel 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: Conditional monthly variance of market and BAB factors

<table>
<thead>
<tr>
<th></th>
<th>MKT</th>
<th>BAB</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>EGARCH(1,1) parameters</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>$-0.610^{***}$</td>
<td>$-0.503^{***}$</td>
</tr>
<tr>
<td></td>
<td>(0.203)</td>
<td>(0.153)</td>
</tr>
<tr>
<td>Garch(1)</td>
<td>$0.901^{***}$</td>
<td>$0.926^{***}$</td>
</tr>
<tr>
<td></td>
<td>(0.032)</td>
<td>(0.022)</td>
</tr>
<tr>
<td>Arch(1)</td>
<td>$0.203^{***}$</td>
<td>$0.367^{***}$</td>
</tr>
<tr>
<td></td>
<td>(0.068)</td>
<td>(0.068)</td>
</tr>
<tr>
<td>Leverage(1)</td>
<td>$-0.177^{***}$</td>
<td>$-0.059^{***}$</td>
</tr>
<tr>
<td></td>
<td>(0.038)</td>
<td>(0.026)</td>
</tr>
</tbody>
</table>

The table gives the results of estimation of an EGARCH(1,1) model on the MKT and BAB factors. 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 3: Estimated Market and BAB factor model with time-varying SDF coefficient

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Market SDF loading: $b_m$</td>
<td>1.283</td>
</tr>
<tr>
<td>BAB SDF loading: $b_{bab,t}$</td>
<td></td>
</tr>
<tr>
<td>$\hat{b}$</td>
<td>9.222</td>
</tr>
<tr>
<td>$\hat{\phi}$</td>
<td>0.909</td>
</tr>
<tr>
<td>$\hat{\sigma}_b$</td>
<td>2.520</td>
</tr>
<tr>
<td>$J$-test $p$-value</td>
<td>0.247</td>
</tr>
<tr>
<td>Factor MAPE</td>
<td>0.013</td>
</tr>
<tr>
<td>St.Dev. of $b_{bab,t}$</td>
<td>6.042</td>
</tr>
<tr>
<td>Annualized mean of $\lambda_{bab,t}$</td>
<td>0.911</td>
</tr>
<tr>
<td>St.Dev. of annualized $\lambda_{bab,t}$</td>
<td>0.661</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Panel A: ( b_{bab,t} ) regressed on</th>
<th>Coefficient (std. error)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NBER recession dummy</td>
<td>-3.483* (2.032)</td>
</tr>
<tr>
<td>Beta decile spread</td>
<td>1.462** (0.678)</td>
</tr>
<tr>
<td>Conditional variance of market return</td>
<td>-10.777*** (3.129)</td>
</tr>
<tr>
<td>Conditional variance of BAB return</td>
<td>-10.071* (5.427)</td>
</tr>
<tr>
<td>Log market dividend-price ratio</td>
<td>5.007** (2.457)</td>
</tr>
<tr>
<td>Term Spread (10yr - 1yr)</td>
<td>0.708 (0.459)</td>
</tr>
</tbody>
</table>

\( R^2 \) 0.273

<table>
<thead>
<tr>
<th>Panel B: ( \lambda_{bab,t} ) regressed on:</th>
</tr>
</thead>
<tbody>
<tr>
<td>NBER recession dummy</td>
</tr>
<tr>
<td>Beta decile spread</td>
</tr>
<tr>
<td>Conditional variance of market return</td>
</tr>
<tr>
<td>Conditional variance of BAB return</td>
</tr>
<tr>
<td>Log market dividend-price ratio</td>
</tr>
<tr>
<td>Term Spread (10yr - 1yr)</td>
</tr>
</tbody>
</table>

\( R^2 \) 0.200

The table shows regressions of the time-varying SDF coefficient \( b_{bab,t} \) and the price of risk for the BAB-shock (\( \lambda_{bab,t} = b_{bab,t} \times \sqrt{2V_t(BAB_{t+1})} \)) 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 5: GMM tests of other benchmark factor models

<table>
<thead>
<tr>
<th></th>
<th>Standard Factors</th>
<th>DMRS Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MKT</td>
<td>BAB</td>
</tr>
<tr>
<td><strong>Panel A:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single-horizon estimation:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max. price of risk (annualized)</td>
<td>0.383</td>
<td>0.979</td>
</tr>
<tr>
<td>Factor MAPE (annualized)</td>
<td>0.005</td>
<td>0.310</td>
</tr>
<tr>
<td>Multi-horizon estimation:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max. price of risk (annualized)</td>
<td>0.365</td>
<td>0.198</td>
</tr>
<tr>
<td>Factor MAPE (annualized)</td>
<td>0.004</td>
<td>0.044</td>
</tr>
<tr>
<td>p-value GMM J-test</td>
<td>0.663</td>
<td>0.000***</td>
</tr>
</tbody>
</table>

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 6: GMM tests of factor models using alternative horizons

<table>
<thead>
<tr>
<th></th>
<th>Standard Factors</th>
<th></th>
<th>DMRS Factors</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>MKT+</td>
<td>FF3</td>
<td>FF5</td>
</tr>
<tr>
<td>MKT</td>
<td>0.007</td>
<td>0.012</td>
<td>0.010</td>
<td>0.033</td>
</tr>
<tr>
<td>BAB</td>
<td>0.115</td>
<td>0.026</td>
<td>0.014</td>
<td>0.024</td>
</tr>
<tr>
<td>FF3</td>
<td>0.012</td>
<td>0.026</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FF5</td>
<td>0.010</td>
<td>0.014</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BAB+MOM</td>
<td>0.033</td>
<td>0.024</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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***

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

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

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

<table>
<thead>
<tr>
<th>Panel A:</th>
<th>MKT</th>
<th>DMRS MKT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Baseline</td>
<td>RV managed</td>
</tr>
<tr>
<td>Max SR of model factors</td>
<td>0.388</td>
<td>0.438</td>
</tr>
<tr>
<td>Max Information Ratio</td>
<td>0.329</td>
<td>0.290</td>
</tr>
<tr>
<td>p-value GRS test</td>
<td>0.049**</td>
<td>0.082*</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B:</th>
<th>Fama-French 3-factor model</th>
<th>DMRS 3-factor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Baseline</td>
<td>RV managed</td>
</tr>
<tr>
<td>Max SR of model factors</td>
<td>0.690</td>
<td>0.752</td>
</tr>
<tr>
<td>Max Information Ratio</td>
<td>0.576</td>
<td>0.521</td>
</tr>
<tr>
<td>p-value GRS test</td>
<td>0.005***</td>
<td>0.014***</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel C:</th>
<th>Fama-French 5-factor model</th>
<th>DMRS 5-factor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Baseline</td>
<td>RV managed</td>
</tr>
<tr>
<td>Max SR of model factors</td>
<td>1.089</td>
<td>1.356</td>
</tr>
<tr>
<td>Max Information Ratio</td>
<td>0.491</td>
<td>0.395</td>
</tr>
<tr>
<td>p-value GRS test</td>
<td>0.052*</td>
<td>0.136</td>
</tr>
</tbody>
</table>

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.