NBER WORKING PAPER SERIES

# CONDITIONAL DYNAMICS AND THE MULTI-HORIZON RISK-RETURN TRADE-OFF

Mikhail Chernov
Lars A. Lochstoer
Stig R. H. Lundeby

Working Paper 25361
http://www.nber.org/papers/w25361

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2018

We thank Hank Bessembinder, Ian Dew-Becker, Greg Duffee, Valentin Haddad, Francis Longstaff, Tyler Muir, Christopher Polk, Seth Pruitt, Sri Santosh, Tommy Stamland, and Stan Zin for comments on earlier drafts, as well as participants in the seminars and conferences sponsored by the ASU Winter Finance Conference, BI, David Backus Memorial Conference at Ojai, LAEF, NBER LTAM conference, NBIM, Norwegian School of Economics, the SFS Finance Cavalcade, Stanford GSB, the UBC Winter Finance Conference, and UCLA. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Conditional Dynamics and the Multi-Horizon Risk-Return Trade-Off
Mikhail Chernov, Lars A. Lochstoer, and Stig R. H. Lundeby
NBER Working Paper No. 25361
December 2018, Revised August 2020
JEL No. C51,G12

**ABSTRACT**

We propose testing asset-pricing models using multi-horizon returns (MHR). MHR effectively generate a new set of test assets that are endogenous to the model and that identify a broad set of possible conditional misspecifications. We apply MHR-based testing to prominent linear factor models and show that these models typically do a poor job of pricing longer-horizon returns, with pricing errors of similar magnitude as the risk premiums they were designed to explain. We trace the errors to the conditional factor dynamics. Explicitly incorporating factor timing in the models often makes mispricing worse, posing a challenge for future research.

Mikhail Chernov
Anderson School of Management
University of California, Los Angeles
110 Westwood Plaza, Suite C-417
Los Angeles, CA 90095
and CEPR
and also NBER
mikhail.chernov@anderson.ucla.edu

Lars A. Lochstoer
University of California at Los Angeles
lars.lochstoer@anderson.ucla.edu

Stig R. H. Lundeby
Norwegian School of Economics (NHH)
Helleveien 30
Bergen
Norway
stig.lundeby@nhh.no

The latest version is available at https://sites.google.com/site/mbchernov/CLL_MHR_latest.pdf

# 1 Introduction

In this paper we propose a new asset-pricing test and apply it to a set of leading linear factor models. Do we need another test, you might ask. Current tests take the set of test assets as given, although it is well-understood that this decision on the part of the researcher is critical for test performance. Over time novel expected return patterns observed in the historical data prompt modifications in the existing models to account for these patterns. As this process unfolds, viewing the test assets as "outside" the model, as the tests assume, becomes tenuous.[1]

We argue that using multi-horizon returns (MHR) offers a useful way to address these issues. Specifically, we show formally that MHR effectively generate a set of test assets that are endogenous to the model at hand and that allow for testing most, if not all, aspects of conditional model misspecification. No matter how much conditioning information a model already accounts for in the construction of the model's factors, the MHR-based test generates new test assets that leverage this information in an endogenous fashion.

Our test is derived using the standard no-arbitrage condition and by formulating models in terms of their implications for the stochastic discount factor (SDF). No-arbitrage implies that the $h$-period SDF equals the product of the $h$ corresponding single-period SDFs. It is therefore straightforward to derive a model's implication for returns at any horizon. Thus, with MHR we are testing overidentifying restrictions of the model.

[1]A classic test evaluates whether test assets have zero "alpha" (Gibbons, Ross, and Shanken, 1989). Lo and MacKinlay (1990) discuss the effects of data snooping. Lewellen, Nagel, and Shanken (2010) and Daniel and Titman (2012) cover the effects of test asset factor structure.

In our empirical contribution we show that misspecification of the implicit temporal dynamics in state-of-the-art models of the SDF, as uncovered by MHR, indeed are quantitatively large. Specifically, we consider eight linear factor models: the CAPM, a two-factor model related to Black, Jensen, and Scholes (1972) (the market factor plus a betting-against-beta factor), the Carhart (1997) four-factor model (the three Fama and French (1993) factors plus momentum), the Fama and French (2015) five-factor model, the Daniel, Mota, Rottke, and Santos (2020) five-factor model, the Stambaugh and Yu (2017) four-factor model, the Hou, Xue, and Zhang (2015) four-factor model, and the Haddad, Kozak, and Santosh (2020) six-factor model. These models which are either workhorse or recent cutting-edge models for empirical risk-return modeling. We test the minimal requirement that a model prices its own factors at multiple return horizons.

As an example of the test results, consider the market factor in the Fama-French model. The $h$-period gross return to this factor is simply the product of the one-period gross returns from $t$ to $t + h$. The model trivially prices the one-period return to this factor, but quickly generates pricing errors when we consider the model's implications for longer-period returns. At the four-year horizon, the model's annualized pricing error for the market factor is 7% – about the same as the market risk premium itself.

This example is not unique. The average annualized pricing error across all factors and models is 4.5% when tested jointly on horizons of 1, 3, 6, 12, 24, and 48 months. This is about the same magnitude as the average annualized factor risk premiums the models where designed to explain in the first place. With the exception of the CAPM, all models are rejected at the 5% level.

These MHR-test rejections imply that the models fail to price the factors conditionally. In order to obtain further intuition, we, as a baseline, note that if factor returns are i.i.d., a model that prices single-horizon returns will also price MHR. In the data, the factors in many cases turn out to be surprisingly far from i.i.d. As a simple metric, we compute variance ratios of the log gross return to the mean-variance efficient (MVE) combination of the factors in each model. Variance ratios measure the cumulative autocorrelations of log returns, where an i.i.d. process has a variance ratio of one at all horizons. For the market portfolio, there is a well-known slight increase above one and then a subsequent decline after the 15-month horizon.

Many of the other models, however, have much stronger patterns. For instance, the Daniel, Mota, Rottke, and Santos (2020) model has a variance ratio that increases to about three at the four-year horizon. Thus, there are strong persistent components in the returns to these factor portfolios.

In fact, i.i.d. returns is not the only scenario that delivers unconditional spanning by the candidate factors. Consider the SDF $M_{t,t+1} = 1 - b^{\top}(F_{t,t+1} - E(F_{t,t+1}))$, where $F$ is a vector excess returns on traded portfolios. Under the null of this model, $F$ is unconditionally mean-variance efficient. This implies that the conditional expected return to $F$ is proportional to its conditional second moment. Thus, while it is not true that the variance ratio of an unconditionally MVE portfolio (UMVE) needs to be one at all horizons, the degree of persistence in returns that we document is much higher than that typically estimated for conditional second moments of returns.

Since the test assets in our empirical applications are MHR to the factors themselves, there does exist an SDF using timed versions of the factors that will price these MHR. The timing strategy is a function of the conditional mean and covariance matrix of the factors. We therefore estimate these for each model in-sample. This leads to models with higher maximal Sharpe ratios, as one would expect. Nevertheless, we reject the timed versions of the models using our MHR-based test (exceptions are the CAPM, as well as the the market factor plus a betting-against-beta factor). In fact, the pricing errors and Information ratios remain economically large, often larger than in the original versions of the models. This may seem surprising, but conditional dynamics are notoriously hard to estimate and persistent errors in these estimates show up as pricing errors in MHR.

As a final step in our empirical analysis, we move outside our own modeling of the UMVE portfolio and apply state-of-the-art approaches from the literature. Specifically, we consider out-of-sample volatility timing of Moreira and Muir (2017) and expected return timing using book-to-market ratio of Haddad, Kozak, and Santosh (2020). We apply their approaches to the Carhart (1997), Fama and French (2015), and the static version of Haddad, Kozak, and Santosh (2020) models.[2] We again reject the models using our test, and the MHR pricing errors are of a similar magnitude or larger than in the versions of the models without factor timing.

Our rejection of the most advanced models in the literature suggests the extreme challenge of correctly estimating the conditional factor dynamics. In fact, as the number of factors in a model increases, so does the complexity of their conditional dynamics. Simultaneously, those same rejections indicate that our test has good statistical power properties. Taken together our evidence poses a challenge for future research in terms of understanding and estimation of the conditional risk-return trade-off. The MHR-based test that we propose should serve as a useful guidance

[2] Nagel, Kozak, and Santosh (2018) is the antecedent to static Haddad, Kozak, and Santosh (2020). Throughout, we refer to the model as Haddad, Kozak, and Santosh (2020), or HKS for convinience and because we use the same dataset.

for this endeavor.

*Related literature.* There are many papers that test conditional versions of factor models. For instance, Boguth, Carlson, Fisher, and Simutin (2011), Ferson and Harvey (1999), Farnsworth, Ferson, Jackson, and Todd (2002), Jagannathan and Wang (1996), Kelly, Pruitt, and Su (2019), Lettau and Ludvigson (2001), Lewellen and Nagel (2006), and Moreira and Muir (2017). Our contribution relative to this literature is to show that MHR in asset pricing tests effectively serve as conditioning variables endogenous to the model and that, empirically, multi-horizon factor returns indeed are informative in terms of uncovering novel conditional dynamics of prominent factor models.

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 Backus, Boyarchenko, and Chernov (2018), 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), Koijen, Lustig, and Nieuwerburgh (2017), Lustig, Stathopoulos, and Verdelhan (2013), and Zviadadze (2017).

A related strand of the literature considers multiple frequencies of observations when testing models (e.g., Brennan and Zhang, 2018, Daniel and Marshall, 1997, Jagannathan and Wang, 2007, Kamara, Korajczyk, Lou, and Sadka, 2016, Parker and Julliard, 2005), though none of these consider the implications of a joint test across horizons.

Baba Yara, Boons, and Tamoni (2020) consider the predictive power of characteristics lagged at different horizons. Favero and Melone (2020) analyze a factor model that also incorporates long-run relationships through cointegration. Bessembinder, Cooper, and Zhang (2020) model and document changes in measures of mutual fund performance at long horizons.

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

# 2 Linear factor models and multi-horizon returns

This section has three main objectives. First, we offer a unified view of model construction in cross-sectional asset pricing. Second, we highlight difficulties in testing and evaluating progress in improving such models. Third, we introduce a testing approach that is essentially a dual to the dominant testing paradigm in the literature. This new approach is attractive because it allows sidestepping many issues that we describe.

## 2.1 Conditional MVE portfolio

A long-standing paradigm in asset pricing is that of the construction of the mean-variance frontier (MVF). Applications include linear beta-pricing models of the cross-section of expected returns, as well as a more general understanding of the properties of the minimum variance stochastic discount factor (see, e.g., Cochrane, 2004). While the literal construction of MVF is infeasible, it is instructive to start with that unattainable reference point.

Let $R^{ei}_{t,t+1}$ represent a one-period excess return on an asset $i$. Stack excess returns on all assets into a $N_t \times 1$ vector $R^e_{t,t+1}$. Then

$$R^C_{t,t+1} = (w^C_t)^\top R^e_{t,t+1}, \quad (1)$$

$$w^C_t = k_t^{-1} V_t \left(R^e_{t,t+1}\right)^{-1} E_t \left(R^e_{t,t+1}\right) \quad (2)$$

is a conditional MVE (CMVE) portfolio. Here $k_t$ is any positive constant known at time $t$, governing the leverage of the portfolio. The CMVE "prices" any combination of these assets. That is, there exists an SDF, $M^*_{t,t+1}$, derived from the CMVE that prices $R^e_{t,t+1}$ conditionally and unconditionally, $E(M^*_{t,t+1}R^e_{t,t+1}) = E_t(M^*_{t,t+1}R^e_{t,t+1}) = 0$. Specifically,

$$M^*_{t,t+1} = 1 - k_t(R^C_{t,t+1} - E_t(R^C_{t,t+1})), \quad (3)$$

$$k_t = V_t^{-1} \left(R^C_{t,t+1}\right) E_t \left(R^C_{t,t+1}\right). \quad (4)$$

## 2.2 Unconditional MVE portfolio

In practice, researchers are searching for the unconditional MVE (UMVE) portfolio, which is the portfolio with the highest maximal unconditional Sharpe ratio. This portfolio is perfectly negatively correlated with the projection of the SDF on the set of excess returns, and it is also a CMVE (Hansen and Richard, 1987). The reason for the focus on finding the UMVE is that is easier to test unconditional model implications.

UMVE is a timed version of the CMVE in Equation (1),

$$R^U_{t,t+1} \quad = \quad \delta_t R^C_{t,t+1} = \delta_t (w^C_t)^\top R^e_{t,t+1} = (w^U_t)^\top R^e_{t,t+1}, \tag{5}$$

where the timing variable $\delta_t$ is

$$\delta_t = (1+\theta_t)^{-1} k_t,$$

and where $\theta_t$ is the maximal conditional squared Sharpe ratio:

$$\theta_t = E_t\left(R^e_{t,t+1}\right)^\top V_t\left(R^e_{t,t+1}\right)^{-1} E_t\left(R^e_{t,t+1}\right).$$

As a by-product, we obtain a particularly transparent expression for the UMVE portfolio weights

$$w^U_t = \delta_t w^C_t = (1+\theta_t)^{-1} \cdot V_t\left(R^e_{t,t+1}\right)^{-1} E_t\left(R^e_{t,t+1}\right). \tag{6}$$

The corresponding SDF is

$$M_{t,t+1} \quad \equiv \quad 1 - (1 - E(R^U_{t,t+1}))^{-1} \cdot (R^U_{t,t+1} - E(R^U_{t,t+1})). \tag{7}$$

See Appendix A.1.

$M_{t,t+1}$ prices all excess returns conditionally by construction (linear transformation of $M^*_{t,t+1}$). It also prices excess returns unconditionally (by the law of iterated expectations). This SDF is perfectly negatively correlated with the UMVE portfolio, which implies that the UMVE portfolio indeed has the maximal attainable unconditional

Sharpe ratio.

## 2.3 Construction of the unconditional MVE portfolio

Implementation of the expression in Equation (5) faces three major hurdles: handling all assets (stocks) is an intractable problem for a variety of reasons, the full information set implicit in the subscript $t$ is not known, and computation of correct conditional mean and variance of returns is not possible without knowing their true distribution at each point in time $t$. In response to these challenges, the literature evaluates portfolios of stocks (Black, Jensen, and Scholes, 1972) and considers various conditioning variables as explicit proxies for the information set, such as the cross-section of market-to-book ratios (Fama and French, 1993) or the aggregate dividend-price ratio (Fama and French, 1988). All of these approaches translate to the following form of the UMVE weights:

$$(w_t^U)^\top = b_t^\top C_t,$$

where $C_t$ is a $K \times N_t$ matrix of stock-level characteristics, and $b_t$ is a $K \times 1$ timing vector.

The characteristics define a set of $K$ factors,

$$F_{t,t+1} \equiv C_t R^e_{t,t+1},$$

that, if the model is true, conditionally span a CMVE portfolio for the set of assets in $R^e_{t,t+1}$. For instance, a row of $C_t$ could be firms' cross-sectionally demeaned book-to-market ratios to create a value factor, or the cross-section of value-weights to create the market factor. The $K \times 1$ factor timing vector, $b_t$, optimally combines these factors to get to the UMVE portfolio. If $b$ is constant, the factors defined by the characteristics $C_t$ unconditionally span the UMVE portfolio.

Factor timing, as studied in the literature, can be generically represented as

$$b_t = D_0 + D_1 z_t, \tag{8}$$

where $D_0$ and $D_1$ are a $K \times 1$ vector and a $K \times L$ matrix of parameters, respectively, while $z_t$ is a $L \times 1$ vector with observable conditioning variables. Such variables include the aggregate dividend-price ratio (Fama and French, 1988), the *cay*-variable (Lettau and Ludvigson, 2001), the default spread (Keim and Stambaugh, 1986), and the term spread (Fama and French, 1989). As one example, if the market dividend-price ratio ($dp_t$) is used as a conditioning variable for each factor, $z_t = dp_t$, while $D_1$ is a $K \times 1$ vector. See Ferson and Harvey (1999), Jagannathan and Wang (1996), and Lettau and Ludvigson (2001). As another example, conditioning variables may be factor-specific as well. For instance, Moreira and Muir (2017) advocate using the inverse conditional variance as the timing variable for each factor. In this case, $z_t$ is a $K \times 1$ vector with factor $k$'s inverse conditional variance in row $k$, while $D_1$ is a diagonal $K \times K$ matrix. Similarly, Haddad, Kozak, and Santosh (2020) use the book-to-market ratio. Kelly, Pruitt, and Su (2019) offer a systematic approach towards selecting significant $z_t$ out of a large set of candidates.

Given the optimal factor weights, $b_t$, the SDF can now be written as:

$$M_{t+1} = 1 - \left(b_t^\top F_{t,t+1} - E(b_t^\top F_{t,t+1})\right). \tag{9}$$

This matches Equation (7) with $R^U_{t,t+1} = b_t^\top F_{t,t+1}(1 - E(b_t^\top F_{t,t+1})/(1 + E(b_t^\top F_{t,t+1}))$.

In summary, a linear factor model consists of several important ingredients: the initial set of assets, the set of conditioning variables $C_t$ and $z_t$, and how they are incorporated into the model.

## 2.4 Testing linear factor models

Consider a version of the SDF in Equation (9) where $b_t = b$ is constant, which implies that the factors span the UMVE portfolio

$$M_{t,t+1} \;\; = \;\; 1 - b^\top (F_{t,t+1} - E(F_{t,t+1})). \tag{10}$$

As a result, any asset's risk premium is linear in the factor risk premiums, both conditionally and unconditionally. Thus, with $N$ test assets we have

$$E(R^e_{t,t+1}) \;\; = \;\; \beta E(F_{t,t+1}), \tag{11}$$

$$E_t(R^e_{t,t+1}) \;\; = \;\; \beta_t E_t(F_{t,t+1}), \tag{12}$$

where $\beta$ and $\beta_t$ are $N \times K$ matrices.[3] The model is commonly tested via a regression

$$R^e_{t,t+1} = \alpha + \beta F_{t,t+1} + \varepsilon_{t+1}, \tag{13}$$

where, $cov(\varepsilon_{t+1}) \equiv \Sigma_\varepsilon$. If Equation (11) holds, $\alpha = 0$. Ferson and Harvey (1999) and Kelly, Pruitt, and Su (2019) are prominent examples of tests of Equation (12)

[3] Here we modify our notation from $N_t$ to $N$ in recognition of the fact that test assets are typically represented by a fixed number of portfolios. This distinction is not material for our analysis.

with $\beta_t$ posited to be an affine function of conditioning variables $z_t$.

The standard methods in the literature for testing factor models is either a time-series test such as that above (traded factors; see Gibbons, Ross, and Shanken, 1989) or a cross-sectional test (necessary for non-traded factors; see Fama and MacBeth, 1973 and Shanken, 1992). In both cases, the tests are asking if it is possible to improve upon the maximal Sharpe ratio combination of the factors (or factor-mimicking portfolios for non-traded factors) by forming a portfolio consisting of both the factors and the test assets. Obviously, this is always possible in-sample and so the tests account for sampling uncertainty.

The testing challenge is to find a set of test portfolios that would sufficiently informative about a given model. The ideal test asset is the UMVE portfolio, which is unattainable (e.g., Barillas and Shanken, 2017). Thus, in practice, researchers instead search for test portfolios that have two properties: (i) large spread in average returns and (ii) returns that not spanned by the model factors (e.g., Daniel and Titman, 2012, Lewellen, Nagel, and Shanken, 2010).

If a model is rejected based on a set of test assets, we know how to modify the model so that it prices the test assets in sample. As explained in MacKinlay (1995), we need to add a factor with portfolio weights proportional to $\Sigma_\varepsilon^{-1}\alpha$ from regression (13). That is, we tend to use information from the construction of the test assets for the construction of new factors. This insight informs the search for additional characteristics and timing variables, which refine the conditioning information implicit in $z_t$ and $C_t$.

This iterative process has led to a number of recent prominent factor models that improve upon the original Fama and French (1993) model. Examples include Daniel, Mota, Rottke, and Santos (2020), Fama and French (2015), Stambaugh and Yu (2017), among others. A logical conclusion of this process is the explicit data-mining approach of Nagel, Kozak, and Santosh (2020), which considers all functions of all characteristics used in prior research in the model test and construction.

Thus, "the model" is in practice the union of the factors and the test assets. In the next section, we show that MHR provide a way to test this broader notion of a model that, unlike existing tests, does not rely on specifying further conditioning information beyond $C_t$ and $z_t$.

## 2.5 The role of multiple horizons

In this paper we propose using MHR in model tests by making use of the models' no-arbitrage implications for returns across different horizons. It turns out there is a tight connection between correct conditional pricing and unconditional pricing across multiple horizons. In order to describe the connection, we have to extend factor models to multiple horizons.

As is known from extant literature, see e.g., Grossman, Melino, and Shiller (1987), Levhari and Levy (1977), and Longstaff (1989), a factor model does not apply across all horizons. To see this, consider the two-period SDF implied by Equation (10):

$$\begin{aligned} M_{t,t+2} &= M_{t,t+1}M_{t+1,t+2} = (a - b^\top F_{t,t+1})(a - b^\top F_{t+1,t+2}) \\ &= a^2 - ab^\top F_{t,t+1} - ab^\top F_{t+1,t+2} + b^\top F_{t,t+1}F_{t+1,t+2}^\top b, \end{aligned}$$

where $a = 1 + b^\top E(F_{t,t+1})$. This implies that the corresponding regression for the two-period return $R_{t,t+2}^{ei}$ will essentially feature a new set of factors even if the original single-horizon model is correctly specified.

The SDF-based approach is a natural way to translate the model in Equation (11) into its counterpart at any longer horizon $h$. Denote one-period risk-free rate by $R^f_{t,t+1}$, and the asset's return by $R^i_{t,t+1} = R^{ei}_{t,t+1} + R^f_{t,t+1}$. The multi-horizon SDF and returns are simple products of their single-horizon counterparts:

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

Thus, we cast analysis in this paper in terms of SDFs. Switching over to the SDF language means that the focus on the magnitude of $\alpha$ changes to the focus on whether

$$E(M_{t,t+h}R^i_{t,t+h}) = 1. \tag{14}$$

Simply put, in a correctly specified model the present value of any \$1 investment is indeed \$1.

**Proposition 1.** *Consider the Euler equation (14) at the single and $h-$period horizons.*

1. *Suppose $E_t(M_{t,t+1}R^i_{t,t+1}) = 1$ for any $i$, then $E(M_{t,t+h}R^i_{t,t+h}) = 1$ for any $i$, $h$.*
2. *Suppose $E(M_{t,t+h}R^i_{t,t+h}) = 1$ for any $i$ and $h$, then for any $i$, the conditional pricing error, $E_t(M_{t,t+1}R^i_{t,t+1})-1$, is zero-mean and uncorrelated with the lagged "Euler equation errors", $M_{t-h,t}R^i_{t-h,t} - 1$, for any $h$.*

We present the proof here because it is helpful in developing intuition about the meaning of the Proposition.

**Proof.**

1. By recursively iterating on the following equation for $h = 1, 2, \ldots,$ we have:

$$\begin{aligned} & E[M_{t-h,t+1}R^i_{t-h,t+1}] = E[M_{t-h,t}R^i_{t-h,t}M_{t,t+1}R^i_{t,t+1}] \\ = \; & E[M_{t-h,t}R^i_{t-h,t}E_t[M_{t,t+1}R^i_{t,t+1}]] = E[M_{t-h,t}R^i_{t-h,t}] = 1. \end{aligned}$$

2. First note that conditional pricing errors, $E_t(M_{t,t+1}R^i_{t,t+1}) - 1$, have zero mean because $E(M_{t,t+1}R^i_{t,t+1}) - 1 = 0$. Next, consider two-period returns:

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

If (14) holds at both one- and two-period horizons, then

$$Cov(M_{t-1,t}R^i_{t-1,t} - 1, E_t(M_{t,t+1}R^i_{t,t+1}) - 1) = 0.$$

A generalization to $(h+1)$-period returns is

$$\sum_{j=1}^{h} Cov(M_{t-j,t}R^i_{t-j,t} - 1, E_t(M_{t,t+1}R^i_{t,t+1}) - 1) = 0. \tag{15}$$

These equations tell us that conditional pricing errors are uncorrelated with Euler equation errors for any $h$.

Second part of the Proposition forms the basis for the test of asset-pricing models that we propose in this paper. Specifically, we advocate testing if $E(M_{t,t+h}R^i_{t,t+h}) = 1$ jointly for a set of different $h$ and $i$. In order to appreciate what the rejection of this null tells us, we revisit some elements of the proof.

Express $M_{t,t+h}R_{t,t+h} - 1$ as $\eta_t^{(h)} + \nu_{t,t+h}$, where $\eta_t^{(h)}$ is the pricing error, $\eta_t^{(h)} = E_t(M_{t,t+h}R_{t,t+h}) - 1$, and $\nu_{t,t+h}$ is the innovation, which is conditionally uncorrelated with $\eta_t^{(h)}$ by the properties of the conditional expectation. Then, Equation (15) can be re-written as:

$$\sum_{j=1}^{h} Cov(\eta_{t-j}^{(j)} + \nu_{t-j,t}, \eta_t^{(1)}) = 0. \tag{16}$$

Thus, rejection of the null implies that either pricing errors are persistent, $Cov(\eta_{t-j}^{(j)}, \eta_t^{(1)}) \neq 0$, or errors are contemporaneously correlated with innovations, $Cov(\nu_{t-j,t}, \eta_t^{(1)}) \neq 0$, or both.

Thus, exploring a model's pricing implications over multiple horizons appears to be a promising avenue. It allows to test for conditional pricing, $E_t(M_{t,t+1}R^i_{t,t+1}) = 1$, using model-implied conditioning information, without the need to specify auxiliary conditioning variables from outside the model. Given that the informational variables $z_t$ and $C_t$ are already implicit in the candidate SDF, $M_{t,t+1}$, MHR-based testing is automatically incorporating whatever conditioning is advocated in the literature.

The next section develops our testing methodology which is applicable to any model that respects the Law of One Price (LOOP) and to any set of test assets. Before we do so, we offer two examples that illustrate the Proposition further.

## 2.6 Examples

Consider two examples to get a sense of what kind of models can be rejected by the proposed test.

### Persistent pricing errors

Suppose that a misspecified one-factor model is

$$\widetilde{M}_{t+1} = 1 - \left(bF_{t,t+1} - E\left[bF_{t,t+1}\right]\right), \quad b = V^{-1}\left(F_{t,t+1}\right) \cdot E\left(F_{t,t+1}\right), \tag{17}$$

where the factor is the excess return to a traded portfolio. The correct model, however, is:

$$M_{t,t+1} = 1 - \left(b_t F_{t,t+1} - E\left[b_t F_{t,t+1}\right]\right)$$

with $b_t = B_0 + B_1 b_{t-1} + u_t$ where $u_t$ is an error term. That is, $F_{t,t+1}$ is only CMVE, not UMVE. The candidate model prices factor returns unconditionally:

$$E\left[\widetilde{M}_{t,t+1} F_{t,t+1}\right] = 0.$$

However, due to the misspecification, we have:

$$E_t\left[\widetilde{M}_{t,t+1} F_{t,t+1}\right] = \left(\frac{1 + bE\left[F_{t,t+1}\right]}{1 + E\left[b_t F_{t,t+1}\right]} b_t - b\right) E_t\left[F_{t,t+1}^2\right].$$

That is, the model does not correctly price the factor conditionally. See Appendix A.2. The pricing error, $\eta_t^{(1)}$, is persistent since the true $b_t$ is persistent. Thus,

$Cov(\eta_{t-j}^{(j)}, \eta_t^{(1)}) \neq 0$. This type of misspecification arises when the factors span the CMVE, but not the UMVE.

### Correlated pricing errors and innovations

Another example is short-term dependence in returns, as seen in short-term reversal. Again, let the proposed model be as the one in Equation (17). What is different is that now the factor returns are i.i.d. Thus, the model prices the factors both conditionally and unconditionally. However, a test asset's returns are not. In particular, consider:

$$R_{t,t+1}^{ie} = \beta_i F_{t,t+1} + \varepsilon_{i,t+1} + \theta \varepsilon_{i,t},$$

where $\varepsilon_{i,t+1}$ is an i.i.d. error term uncorrelated with $F_{t,t+1}$ at all leads and lags. This model represents reversal if $\theta < 0$. The SDF prices $R_{t,t+1}^{ie}$ unconditionally. See Appendix A.3. Thus, a Gibbons, Ross, and Shanken (1989) test with $R_{t,t+1}^{e}$ as the test assets would fail to reject this model because all alphas are equal to zero.

However, the model does not correctly price excess returns conditionally:

$$E_t \left[ \widetilde{M}_{t,t+1} R_{t,t+1}^{ie} \right] = \theta \varepsilon_{i,t}.$$

See Appendix A.3.

Using the notation in Equation (16), the pricing error $\theta \varepsilon_{i,t} = \eta_t^{(1)}$. Thus, in this example pricing errors are not persistent, so $Cov(\eta_{t-j}^{(j)}, \eta_t^{(1)}) = 0$. However, the one-period pricing error is correlated with the innovation, $Cov(\nu_{t-1,t}, \eta_t^{(1)}) \neq 0$. See Appendix A.3.

To summarize, failure to reject a model using MHR does not imply that the model prices assets conditionally. The model could still have errors with two properties: (i) the errors are not persistent, and (ii) the errors have no contemporaneous correlation with innovations. While, mathematically, it is possible to have a model that is misspecified along these lines, we could not think of any model contemplated in the literature that would match this description. Thus, Proposition 1.2 justifies the use of MHR to test for many important, although formally not all, conditional pricing implications.

# 3 Testing linear factor models using MHR

## 3.1 Testing general asset pricing models using MHR

In this section, we develop a GMM-based test using MHR that is applicable to any asset pricing model that satisfies LOOP. As shown in Proposition 1.1, such a model implies Equation (14) for any asset $i$, which we repeat here for convenience:

$$E(M_{t,t+h}R^i_{t,t+h} - 1) = 0$$

for any asset $i$ and horzion $h$. This condition can be easily tested jointly for multiple horizons $h$ in a GMM framework.

The proof of the Proposition in section 2.5 demonstrates that these MHR-based moments are equivalent to Equation (15). The Equation implies moment conditions that we ultimately use in our testing. Specifically, for test assets $i = 1, ..., I$

$$f^i_{t+1} = \begin{pmatrix} M_{t,t+1}R^i_{t,t+1} - 1 \\ z^{(h_2)}_{i,t}(M_{t,t+1}R^i_{t,t+1} - 1) \\ \vdots \\ z^{(h_n)}_{i,t}(M_{t,t+1}R^i_{t,t+1} - 1) \end{pmatrix}, \tag{18}$$

where the conditioning variable is

$$z^{(h)}_{i,t} = \sum_{j=0}^{h-1} M_{t-h+j,t}R^i_{t-h+j,t}, \tag{19}$$

$n$ is the number of horizons used in the test, and $\{h_j\}^n_{j=2}$ are the set of horizons used in addition to the single-period horizon. The null hypothesis is $E(f^i_{t+1}) = 0$ for all $i$, and the test is thus an unconditional test of the conditional properties of the asset pricing model as explained in the second part of the Proposition.

The virtue of the moments in Equation (18) is that the associated residuals are not serially correlated under the null hypothesis. See Appendix A.4. Imposing this additional restriction when estimating the covariance matrix of the moment conditions improves the small-sample properties of the standard errors and test statistics. See Hodrick (1992) for a similar argument in the context of overlapping observations in regressions.

The test falls into the standard GMM framework, where:

$$g_T(\theta) = \frac{1}{T}\sum_{t=1}^{T}\begin{pmatrix} f_t^1(\theta) \\ f_t^2(\theta) \\ \vdots \\ f_t^I(\theta) \end{pmatrix},$$

where $\theta$ are the parameters in the SDF to be estimated. The objective function is as usual:

$$\underset{\theta}{\operatorname{argmin}}\, g_T(\theta)^\top W g_T(\theta),$$

where $W$ is an $(I \times n) \times (I \times n)$ positive definite weighting matrix (e.g., Hansen and Singleton, 1982). Relevant test statistics and parameter standard errors can be found using the usual GMM toolkit.

## 3.2 Adopting the general test to linear models

We slightly re-write the $K$-factor model in Equation (10) as

$$M_{t,t+1} = 1 - b^\top(F_{t,t+1} - \mu),$$

to emphasize the need to estimate $\mu = E(F_{t,t+1})$. Guaranteeing that this SDF prices the risk-free rate conditionally requires adding auxiliary assumptions that are not explicit in the settings that are traditionally used for testing linear factor models. Because our goal is to assess the original models' performance, we make a slight adjustment to the moment conditions to ensure we do not reject the models based on mispricing of the multi-period risk-free rates, something that they were not designed to match.

Specifically, we note that predicting discounted gross returns, $MR^i$, as in the covariance condition in Equation (15), is equivalent to predicting discounted excess returns, $M(R^i - R^f)$, if the model prices the risk-free asset. We therefore replace the moment conditions in Equation (18) with the following ones:

$$f^i_{t+1} = \begin{pmatrix} M_{t,t+1}(R^i_{t,t+1} - R^f_{t,t+1}) \\ z^{(h_2)}_{i,t} M_{t,t+1}(R^i_{t,t+1} - R^f_{t,t+1}) \\ \vdots \\ z^{(h_n)}_{i,t} M_{t,t+1}(R^i_{t,t+1} - R^f_{t,t+1}) \end{pmatrix}.$$

The resulting $K + I \times n$ GMM moments are:

$$g_T(b, \mu) = \frac{1}{T} \sum_{t=1}^{T} \begin{pmatrix} F_{t,t+1} - \mu \\ f^1_t(b, \mu) \\ f^2_t(b, \mu) \\ \vdots \\ f^I_t(b, \mu) \end{pmatrix}.$$

Note that the managed portfolio weights $z^{(h)}_{i,t}$ in each $f^i$ are exactly the same as in Equation (19), that is, they still depend on gross returns rather then excess ones.

## 3.3 Data

We select our models based on their historical importance, recent advancements, and data availability. Specifically, we include the CAPM, CAPM combined with the BAB factor (Frazzini and Pedersen, 2014, Black, Jensen, and Scholes, 1972, Novy-Marx and Velikov, 2016), Fama and French five-factor model, FF5, (Fama and French, 2015), a version of the FF5 model with hedged unpriced risks (Daniel, Mota, Rottke, and Santos, 2020), Fama and French three-factor model with momentum (Carhart, 1997, Fama and French, 1993), the four-factor models of Hou, Xue, and Zhang (2015) and Stambaugh and Yu (2017), and the six-factor model of Haddad, Kozak, and Santosh (2020).

The Fama-French five-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). These data and the momentum factor MOM (Jegadeesh and Titman, 1993) are provided on Kenneth French's webpage. The returns are monthly and the sample is from July 1963 to June 2017.

The hedged versions of these factors studied by Daniel, Mota, Rottke, and Santos (2020) (DMRS) are available on Kent Daniel's webpage. The sample period is July 1963 to June 2017. The factors studied by Hou, Xue, and Zhang (2015) (HXZ) are MKT, SMB, I/A (investment-to-assets) and ROE (return on equity), and are available on Lu Zhang's website. The sample is from January 1967 to December 2017. Stambaugh and Yu (2017) propose two factors intended to capture stock mispricing, in addition to the existing MKT and SMB factors: PERF and MGMT. We denote this four-factor model as SY. These data are available on Robert Stambaugh's webpage. The sample period for these factors starts January 1963 and ends December 2016. Haddad, Kozak, and Santosh (2020) (HKS) propose, in addition to MKT, factors that are the first five principal components (PC1-5) of fifty anomaly portfolios that are entertained in the literature. Their sample period is January 1974 to December 2017.

Given the recent critique by Novy-Marx and Velikov (2016), we depart from the BAB factor construction of Frazzini and Pedersen (2014). We use the value-weighted beta- and size-sorted portfolios on Kenneth French's webpage as the building blocks for constructing this factor, following Fama and French (2015) and Novy-Marx and Velikov (2016). See Appendix A.5.

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

### 3.4 Test assets

We consider the factors themselves as the set of test assets. This choice limits us to testing only for persistence in pricing errors as in the first example of section 2.6. The reason for this choice of test assets is three-fold.

First, the factors in these models are created from mechanical trading strategies to price documented empirical spreads in the cross-section of expected returns. Thus, a natural and minimal requirement for a well-specified model is that the model can price these strategies at any horizon. As an example, in a factor model that uses the Fama and French (1993) HML factor, the present value of a $1 investment in the risk-free rate and a position in the HML (value) factor should be \$1 regardless of the holding period.

Second, it is clear that a model can price the single-horizon excess returns associated with its factors unconditionally. We will in fact estimate $b$ such that the single-horizon returns to the factors themselves are priced without error and set $\mu$ equal to the factor sample means. We choose the weighting matrix accordingly to ensure these are the only moments used to identify the parameters. That is in line with the standard Black, Jensen, and Scholes (1972) regressions in Equation (13), as the regression imposes the sample mean of the factors in the estimation of $\alpha_i$. Thus, any rejection is due to the joint test of the models' pricing of longer-horizon returns.

Third, this choice of test assets implies that there exists an SDF with time-varying loadings $b_t$ as in Equation (9) that does price the factors conditionally and, therefore, prices these factors unconditionally at any horizon per the first part of the Proposition in section 2.5. We discuss this alternative hypothesis in more detail in a later section.

### 3.5 MHR pricing errors and model tests

In the tests we use the horizons 3, 6, 12, 24, and 48 months in addition to the one-period (monthly) horizon. Because the evaluated factors are designed as zero-investment long-short portfolios, we construct $R^i = R^f + F^i$ for each factor $i$ when evaluating $z_{i,t}^{(h)}$ in Equation (19).

We start by computing pricing errors for each factor in each model across horizons. The pricing errors should be understood as the net present value of an $h$-period \$1 buy-and-hold investment in the gross factor return. Since the models are estimated to match one-period returns unconditionally, non-zero net present values are due to mispricing of the conditional factor return. To facilitate comparison we annualize errors so that each reported number reflects the same period irrespective of the horizon. Thus, the pricing error for a factor $F^i$ at horizon $h$ is $12/h \times E_T(z_{i,t}^{(h)} M_{t,t+1} F_{t,t+1}^i)$. The horizons for reported errors range from 1 to 48 months.

Figure 1 displays the pricing errors for the first four factor models. The top left panel shows that the pricing errors of the CAPM are small across horizons, always less than 1% annualized. Thus, for the market model, a constant $b$ coefficient in the SDF works well for pricing market returns at any horizon.

The top right panel shows the MKT+BAB model. In this case, pricing errors are much larger for both factors. For the BAB factor, the annualized pricing error increases with horizon (in absolute value) to almost 10% per year at the 48-month horizon. That is about twice the average annualized monthly returns on this factor.

The bottom left plot shows the Carhart model (FF3+MOM), where the pricing errors get very large, exceeding 50% p.a. for the 4-year MOM return. The bottom right panel shows the corresponding pricing errors for the FF5 model. Again pricing errors increase in absolute value with horizon. Three of the five factors (MKT, RMW, and CMA) have absolute pricing errors in excess of 5% p.a. at the 4-year horizon.

Panel A of Table 1 gives the $p$-values of the $J$-test of each model. With the exception of the CAPM, the models are rejected (MKT+BAB is marginally rejected with a $p$-value of 0.048).[4] We calculate the mean absolute pricing error (MAPE) for each model as the mean of the absolute value of the annualized pricing errors across the

[4] We would like to emphasize the meaning of failure to reject in this context. The MKT factor is capable of pricing itself at multiple horizons. That does not imply, however, that the MKT model is well-specified. As we know, it is easily rejected by a cross-section of equity returns.

factors and horizons. For the CAPM, the MAPE is only 0.7%, for the CAPM+BAB it is 3%, for the Carhart model it is 7.6%, and for the FF5 it is 1.9%.

Figure 2 shows the pricing errors for the remaining four models. The top left panel shows pricing errors for the $\text{FF5}_{DMRS}$ model. Its pricing errors exceed 10% p.a. for two factors (their versions of the MKT and SMB factors) and 5% for their version of the CMA factor. The remaining plots show the pricing errors for the SY, HXZ, and HKS models. For these models, pricing errors are even larger, with the largest pricing error exceeding 100% p.a. (PERF in the SY model). As $p$-values in Panel B of Table 1 indicate, all these models are rejected with high levels of confidence.

Overall, Table 1 shows the average MAPE across all eight models is 4.4%, which is about the same as the annualized factor risk premiums that these models were originally designed to match. We conclude from this that the current benchmark models for risk-adjustment do a poor job accounting for MHRs.

Table 1 also displays the annualized maximal Sharpe ratios, $[E(F)^{\top}V(F)^{-1}E(F)]^{1/2}$, implied by each factor model. A higher Sharpe ratio suggests that the factors are closer to spanning the unconditional MVE portfolio. As is well-known, the Sharpe ratio of the MKT factor is much lower than the maximal Sharpe ratios in more recent multi-factor models. For instance, the SY model has an annualized Sharpe ratio of 1.7 compared to 0.4 for the CAPM.

Figure 3 shows each model's MAPE plotted against the respective maximal Sharpe ratios. Interestingly, there is a positive relation. The higher a model's Sharpe ratio the closer it should be to spanning the unconditionally mean-variance efficient portfolio and thus the lower the pricing errors should be. The opposite being the case indicates that the search for high Sharpe ratio models has increased the complexity of the conditional dynamics, consistent with the findings in Haddad, Kozak, and Santosh (2020). Thus, there is a need for understanding the economic effects and drivers of these dynamics.

Finally, Table 1 reports the annualized maximal Information ratio for each model as an alternative economic measure of the mispricing implied by MHR. In particular, we run regressions as in Equation (13), where the left hand side test assets returns are $z_{i,t}^{(h)} \times F_{t,t+1}^{i}$ for factor $i$ and horizon $h$. The reported Information ratio is then $\sqrt{12\alpha^{\top}\Sigma_{\varepsilon}^{-1}\alpha}$. The models' information ratios are economically large, ranging from 0.61 to 1.14 for the rejected models, similar in magnitude to the model's maximal Sharpe ratios.

In sum, the MHR-based tests indicate that the constant $b$ versions of the models are not able to price important dynamics in the factor returns.

### 3.6 Factor dynamics and long-horizon investment

In order to gain more intuition about the rejection results, we evaluate both statistical and economic metrics that are relevant for long-horizon investors. Because we have implemented formal inference in the previous section, we no longer test the models. We rather highlight their properties that are responsible for the reported rejections.

We focus on the in-sample mean-variance efficient combination of each model's factors in order to facilitate cross-model comparison and to reduce the overall dimensionality of the presentation. In particular, we calculate for each model the unconditional MVE combination of the factors and scale the positions such that its volatility is the same as the volatility of market returns. We denote (excess) returns on the MVE portfolio by $MVE_{t,t+1}$. The returns $R_{t,t+1}$ are computed by adding the gross-risk free rate.

**Statistical assessment**

The null hypothesis that the SDF in Equation (10) is correctly specified tells us something about dynamics of factors. Specifically, it implies that conditional mean of factor returns is proportional to their conditional second moment. See Appendix A.6. That our test rejects the models implies that this requirement does not hold in the data.

One way to illustrate dynamics of returns is to compare variance ratios of MVE returns across multiple horizons. We take logs of returns ($r = \log R$). The $h$-horizon variance ratio is then calculated as:

$$VR(h) = \frac{V(r_{t,t+1} + r_{t+1,t+2} + ... + r_{t+h-1,t+h})}{h \times V(r_{t,t+1})}.$$

For further intuition, note that the variance ratio can be written as a weighted average of autocorrelations of log returns at lags up to horizon $h$. One benchmark for return dynamics is that of i.i.d. In this case, the variance ratio is equal to 1 at any horizon.

Figure 4 displays the variance ratios for each model. The market factor in the CAPM displays familiar dynamics where the variance ratio increases slightly from 1 to almost 1.2 at the annual horizon and subsequently decreases towards 1 at the 4-year horizon. The latter decrease is consistent with a long-run mean-reverting component in market returns.

All the other models display markedly stronger departures from the i.i.d. baseline. To start with the most extreme cases, the $FF5_{DMRS}$ and the MKT+BAB models have strongly increasing variance ratios exceeding 2 at 12 to 24 months. For $FF5_{DMRS}$, the 4-year variance ratio is about 3. That is, a 4-year investor holding this portfolio is subject to, per unit of time, triple the variance of an investor with a monthly holding period.

The variance ratio of the FF5 model peaks at about 1.8 at the 2-year horizon, while the SY and HKS models have a variance ratio in excess of 2 at the 4-year horizon. The Carhart and HXZ models have slowly increasing variance ratios that end up at about 1.6 at the 4-year horizon.

The finding is suggestive of the potential difficulty of spanning UMVE statically. Clearly, there are strong persistent components in the factor return dynamics. While, in theory, conditional volatility of the factors could adjust to compensate for their persistence to make the constant $b$ assumption valid, it is hard to imagine that the degree of persistence in returns that we document is matched by that of second moments of returns. The fact that we reject all models except the CAPM in Section 3.4 implies that indeed it is not the case.

**Economic assessment**

We offer another perspective on this conclusion by assessing the economic impact of these dynamics for long-horizon investors. We calculate the annualized Sharpe ratio by horizon for each of the MVE portfolios. Excess returns at horizon $h$ is the average $h$-period gross return minus the $h$-period gross risk-free rate, where both of these are calculated by multiplying together $h$ one-period gross returns. The Sharpe ratio is then the mean excess return divided by the standard deviation of excess returns. We annualize by multiplying with $\sqrt{12/h}$.

To obtain a benchmark Sharpe ratio that corresponds to the null hypothesis, we draw from the original MVE and risk-free returns with replacement $10,000$ artificial histories of the same length as our sample. This procedure imposes i.i.d. dynamics on the bootstrapped returns. The procedure retains the same unconditional distribution of returns and does not impose normality.

Figures 5 and 6 display the Sharpe ratios for each model and their benchmarks under the null. Across all models Sharpe ratios are declining in the horizon. As the benchmark Sharpe ratios illustrate, the pattern in of itself is not surprising. What is different for some models is much steeper decline than in the benchmark. The larger steepness coincides with instances of strong increases in the variance ratio.

The persistent returns lead to higher long-run variance which depresses long-run Sharpe ratios. The economic magnitude is particularly large for FF5, FF5$_{DMRS}$, SY, and HKS. For these models, the 4-year Sharpe ratio is less than a half of that in the benchmark.

# 4 Pricing factor MHR

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

## 4.1 Estimating factor timing

Per part one of Proposition 1, an SDF that prices the factors conditionally also prices MHR to the factors unconditionally. Following the discussion in Sections 2.1 and 2.2, the UMVE portfolio formed by dynamic trading in the factors prices them both conditionally and unconditionally. The corresponding constant $b$ SDF is

$$M_{t,t+1} = 1 - (1 - E(F^U_{t,t+1}))^{-1} \cdot (F^U_{t,t+1} - E(F^U_{t,t+1})), \tag{20}$$

where $F^U_{t,t+1}$ is that UMVE portfolio. We follow the logic of Equations (1) – (6) to construct $F^U_{t,t+1}$. We simply replace excess returns on all assets $R^e_{t,t+1}$ by factors $F_{t,t+1}$. Specifically,

$$\begin{aligned} F^U_{t,t+1} \equiv w_t^\top \cdot F_{t,t+1} &= (1+\theta_t)^{-1} \left[V_t\left(F_{t,t+1}\right)^{-1} E_t\left(F_{t,t+1}\right)\right]^\top \cdot F_{t,t+1}, \\ \theta_t &= E_t\left(F_{t,t+1}\right)^\top V_t\left(F_{t,t+1}\right)^{-1} E_t\left(F_{t,t+1}\right). \end{aligned} \tag{21}$$

Under the null of correct conditional pricing of the factors $F_{t,t+1}$, the timed combination of the factors that is the UMVE portfolio $F^U_{t,t+1}$ should be priced as well. Thus, we include this model-implied timed portfolio into the set of test assets in our MHR-based test, in addition to the original factors. This is an important step because in a model with a time-varying $b_t$, the UMVE portfolio embodies conditional information not captured by the factors themselves.

In order to obtain $w_t$, we explicitly estimate $E_t(F_{t,t+1})$ and $V_t(F_{t,t+1})$ for each model. We emphasize that, because of the illustrative nature of our exercise, the estimation is in-sample. We estimate the conditional monthly variance-covariance matrix of the factor returns using the multivariate CCC-GARCH method of Bollerslev (1990). We estimate conditional mean of each element $k$ of the vector of factors $F$ using a simple regression model that is motivated by the uncovered strong dependencies in factor returns documented in Section 3.5:

$$F^i_{t,t+1} = \beta_{i,0} + \sum_{j=1}^{n} \beta_{i,h_j} x^{(h_j)}_{i,t} + \varepsilon^i_{t+1}, \tag{22}$$

where $x^{(h)}_{i,t} = \sum_{j=1}^{h} F^i_{t-j,t-j+1}$. We use the same horizons $h$ as in our GMM tests.

We follow the post LASSO approach of Belloni and Chernozhukov (2013) to estimate the regression (22) for each factor. That is, we use the LASSO to select strong predictive variables and, because the LASSO yields biased return estimates, we next use OLS with these selected regressors to get conditional expectation $E_t(F^i_{t,t+1})$.

Lastly, because dividing by estimated variance introduces a bias, we rescale each element $i$ of the estimated version of portfolio weights in (21) by a constant, $b^i$. We mitigate the bias by ensuring, via $b^i$, for $i = 1, ..., K$ that the unconditional factor SHRs are priced correctly. That is also consistent with our testing strategy in the constant $b$ case of the preceding section. Finally, $b^i$ absorbs the theoretical scaling constant in Equation (20). Thereby, this approach connects with factor-timing, $b_t$, as described in Equation (8), with $D_0 = 0$, diagonal $D_1$ with element $i$ equal to $b^i$, and $z_t = w_t$, where the latter are UMVE portfolio weights as defined in Equation (21).

## 4.2 Estimation results

Table 2 presents test results along with pricing errors and Sharpe ratios. The results suggest that, overall, we reject the same models as in the case of constant $b$. There are some differences. MKT+BAB is marginally rejected when $b$ is constant, while we fail to reject when it is time-varying. HXZ's $p-$value is 0.039, a marginal rejection, when $b_t$ is time-varying.

HKS gives us an opportunity to compare our estimation of $b_t$ with an alternative because they provide their own in-sample estimates of $E_t$ and $V_t$. In particular, estimation of $E_t$ relies on a conceptually different approach by using book-to-market ratio instead of past returns for timing as the basis for expected returns. The resulting $p-$value of 0.030 (not reported in the Table) is lower than 0.049 that is obtained with our estimates. Our version of in-sample timing of the HKS factors is barely rejected. To be clear, HKS correctly advocate out-of-sample timing, so this result could be a manifestation of in-sample overfitting using our method.

Table 2 shows the Sharpe ratio of the additional test asset, which is constructed as the UMVE portfolios $F^U_{t,t+1}$ implied from each model's estimated conditional factor means and covariance matrices as described in Equation (21). This is the maximal Sharpe ratio, comparable to that given in Table 1 for the constant $b$ models. As one would expect, the Sharpe ratio reported for the models with time-varying $b_t$ are generally higher than those from the constant $b$ versions of the models. SY's maximal Sharpe ratio decreases, indicating poor estimates of the conditional factor dynamics for this model.

Table 2 reports MAPE across the original factors and horizons and is, thereby, comparable to the ones in Table 1. As an example, the constant $b$ MAPE for the

$FF5_{DMRS}$ model is 3.4%, while it is 19.9% for the time-varying $b_t$ case. This occurs despite the maximal Sharpe ratio increasing from 1.59 to 1.82 with factor timing. That is the case more generally. Despite the increase in the maximal Sharpe ratios, the MAPEs across models are in many cases higher than those reported for the constant $b$ case.

While it might seem surprising that pricing errors increase when the model's maximal Sharpe ratio increases, thus presumably getting closer to the true UMVE portfolio, this is a manifestation of the Herculean task of estimating the correct conditional dynamics. Misspecification in the conditional mean and variance processes likely results in persistent errors, which in turn show up in MHR per Equation (15).

As in Section 3.5, we report the maximal annualized Information ratio for each model by running regressions (13). The left hand side test assets returns are $z_{i,t}^{(h)} \times F_{t,t+1}^{i}$ for factor $i$ and horizon $h$, where we again add the timing UMVE portfolio as a test asset. In this case, the right-hand side factor is the implied UMVE portfolio. The Information ratios are similar to those of the constant $b$ models, although these are not directly comparable due to the addition of the MHR-based return of UMVE portfolio in each model as a test asset.

## 4.3 Out-of-sample tests

There are two concerns about our implementation of factor timing. For one, it is clear that it is misspecified as there does exist portfolio weights that result in correct factor pricing. Also, an in-sample exercise does not reflect the real-time nature of dynamic portfolio allocation. While we cannot address these concerns in all generality, we turn to most recent research on this topic to evaluate alternative methods. Specifically, Haddad, Kozak, and Santosh (2020) and Moreira and Muir (2017) propose factor timing approaches, which are both out-of-sample and rely on estimates of $E_t(F_{t,t+1})$ and $V_t(F_{t,t+1})$ as prescribed by theory.

As discussed in Section 2.3, both their approaches can be represented in the form of Equation (8). Moreira and Muir (2017) time volatility via $z_t^i = V_t^{-1}(F_{t,t+1}^i)$, which is estimated using squared realized daily factor returns. HKS use $z_t^i = E_t(F_{t,t+1}^i)$, where out-of-sample conditional expectations are constructed for their factors using each factor's value spread. HKS also contemplate a version with $z_t^i = E_t(F_{t,t+1}^i) \cdot V_t^{-1}(F_{t,t+1}^i)$ with $V_t$ estimated in-sample.

Table 3 reports testing results. Because of the specifics of the methodologies employed in these papers, we cannot test all the models that we have considered heretofore. In the case of Volatility Timing, we report FF3+MOM, and FF5. In the case of HKS, we report two versions of the model: out-of-sample timing with $E_t$ only, and hybrid timing with the same $E_t$ and in-sample $V_t$. All models are rejected with economically large MAPEs for the MHR to the model factors.

For the Volatility Timing case, the Sharpe ratios of the timing portfolio are lower and MAPE higher than those in the constant $b$ case. That suggests that volatility timing is insufficient for spanning the UMVE correctly. For the HKS factor timing results, we note that the sample is only the last half of the original sample, due to the estimation of out-of-sample means in HKS. Thus, a direct comparison to the corresponding model in Table 1 cannot be made. That said, the MAPEs are economically large, more than 5% annualized on average per factor. For all of these timing strategies the Information ratios, computed the same way as for Table 2, are still large, all above 1.

Overall, the MAPEs and the Information ratios are still large even when considering models that attempt to account for factor timing. While factor timing does tend to increase the maximal Sharpe ratio of the model, this is a one-period metric. Testing the model with MHR, however, brings in moments that are particularly sensitive to persistent misspecification in the conditional factor dynamics. This new perspective indicates that there is substantial work to do in improving estimates of these dynamics.

# 5 Conclusion

We propose a new unconditional test of conditional asset-pricing models. It is GMM-based and uses multi-horizon returns (MHR) to evaluate the ability of any model to price risky cash flows that accrue at different horizons. We argue that MHR are appealing because they effectively provide a set of test assets that are endogenous to the model being tested and because these assets identify a broad set of conditional model misspecification. Thus, the test does not require any conditioning variables beyond those used in the construction of a model.

Our empirical exercise, involving a number of prominent linear factor models, suggests that our test has statistical power. We reject most of these models under three variations of the null hypothesis: (i) the model factors span the unconditional mean variance portfolio (UMVE); (ii) the timed in-sample model factors span UMVE; and (iii) the timed out-of-sample model factors span UMVE. The associated pricing errors and Information ratios are economically large.

The reason the models do a poor job pricing longer-horizon returns is that the model-implied conditional properties of risk pricing are strongly at odds with dynamic properties of the factors associated with these models. Because long-run investment entails exposure to conditional return dynamics even in the absence of factor timing, these dynamics show up as large mispricing in longer-run returns.

Many economic applications, such as capital budgeting and consumption-savings decisions, require discounting cash flows accruing at multiple horizons. Our evidence suggests that there is still much to be done to arrive at models that can successfully be applied to such problems. In particular, correct specification of the joint conditional dynamics of pricing factors appears even more quantitatively important than previously emphasized in the literature.

# References


Baba Yara, Fahiz, Martijn Boons, and Andrea Tamoni, 2020, New and old characteristic-sorted portfolios: Implications for asset pricing, working paper.

Backus, David K., Nina Boyarchenko, and Mikhail Chernov, 2018, Term structures of asset prices and returns, *Journal of Financial Economics* 129, 1–23.

Barillas, Fancisco, and Jay Shanken, 2017, Which alpha?, *The Review of Financial Studies* 30, 1316–1338.

Belloni, Alexandre, and Victor Chernozhukov, 2013, Least squares after model selection in high-dimensional sparse models, *Bernoulli* 19, 521–547.

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

Bessembinder, Hendrik, Michael J. Cooper, and Feng Zhang, 2020, Mutual fund performance at long horizons, working paper.

Black, Fischer, Michael Jensen, 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.).

Boguth, Oliver, Murray Carlson, Adlai Fisher, and Mikhail Simutin, 2011, Conditional risk and performance evaluation: Volatility timing, overconditioning, and new estimates of momentum alphas, *Journal of Financial Economics* 102, 363–389.

Bollerslev, Tim, 1990, Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model, *Review of Economics and statistics* 72, 498–505.

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

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, 2020, The cross section of risk and return, *Review of Financial Studies* 33, 1927–1979.

Daniel, Kent, and Sheridan Titman, 2012, Testing factor-model explanations of market anomalies, *Critical Finance Review* 1, 103–139.

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

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, A five-factor asset pricing model, *Journal of Financial Economics* 116, 1–22.

Fama, Eugene, and James D. MacBeth, 1973, Risk, return, and equilibrium: Empirical tests, *Journal of Political Economy* 81, 607–636.

Fama, Eugene F., and Kenneth R. French, 1989, Business conditions and expected returns on stocks and bonds, *Journal of Financial Economics* 25, 23–49.

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

Favero, Carlo, and Alessandro Melone, 2020, Asset pricing vs asset expected returning in factor-portfolio models, working paper.

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

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.

Grossman, Sanford, Angelo Melino, and Robert Shiller, 1987, Estimating the continuous-time consumption-based asset pricing model, *Journal of Business and Economic Statistics* 5, 315–327.

Haddad, Valentin, Serhiy Kozak, and Shri Santosh, 2020, Factor timing, *Review of Financial Studies* 33, 1980–2018.

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

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

Hansen, Lars Peter, and Kenneth J. Singleton, 1982, Generalized instrumental variables estimation of nonlinear rational expectations models, *Econometrica* 50, 1269–1286.

Hodrick, Robert J., 1992, Dividend yields and expected stock returns: Alternative procedures for inference and measurement, *The Review of Financial Studies* 5, 357?386.

Hou, Kewei, Chen Xue, and Lu Zhang, 2015, Digesting anomalies: An investment approach, *The Review of Financial Studies* 28, 650–705.

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.

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

Keim, Donald, and Robert Stambaugh, 1986, Predicting returns in teh bind and stock markets, *Journal of Financial Economics* 17, 357–390.

Kelly, Bryan, Seth Pruitt, and Yinan Su, 2019, Characteristics are covariances: A unified model of risk and returns, *Journal of Financial Economics* 134, 501–524.

Koijen, Ralph, Hanno Lustig, and Stijn Van Nieuwerburgh, 2017, The cross-section and time series of stock and bond returns, *Journal of Monetary Economics* 88, 50–69.

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.

Levhari, David, and Haim Levy, 1977, The capital asset pricing model and the investment horizon, *Review of Economic Studies* 59, 92–104.

Lewellen, Jonathan, and Stefan Nagel, 2006, The conditional capm does not explain asset pricing anomalies, *Journal of Financial Economics* 79, 289–314.

——— , and Jay Shanken, 2010, A skeptical appraisal of asset-pricing tests, *Journal of Financial Economics* 96, 175–194.

Lo, Andrew, and Craig A. MacKinlay, 1990, Data-snooping biases in tests of financial asset pricing models, *The Review of Financial Studies* 3, 431–467.

Longstaff, Francis, 1989, Temporal aggregation and the continuous-time capital asset pricing model, *Journal of Finance* 44, 871–887.

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, Serhiy Kozak, and Shrihari Santosh, 2018, Interpreting factor models, *Journal of Finance* 73, 1183–1223.

——— , 2020, Shrinking the cross-section, *Journal of Financial Economics* 135, 271–292.

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

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

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

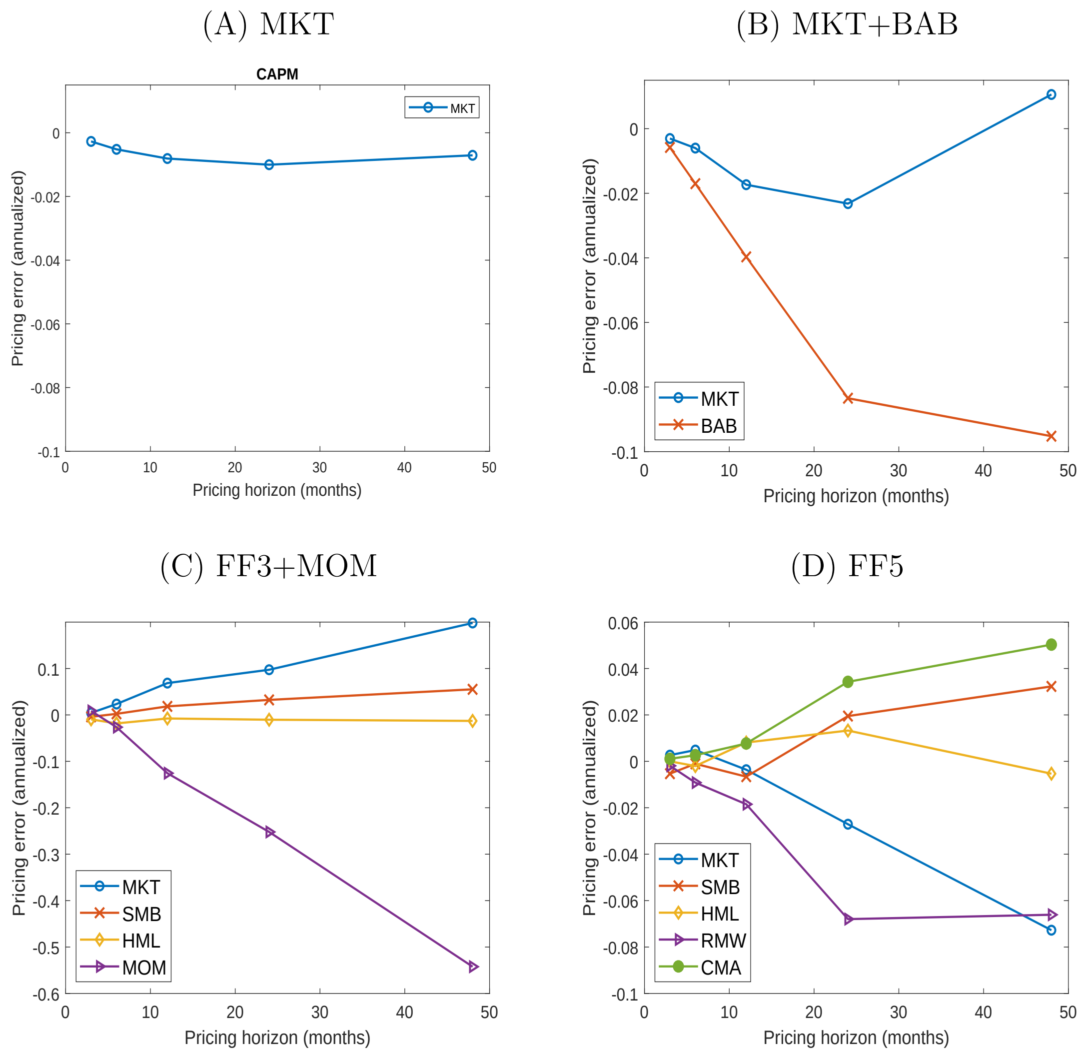
Shanken, Jay, 1992, On the estimation of beta-pricing models, *The Review of Financial Studies* 5, 1–33.

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

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

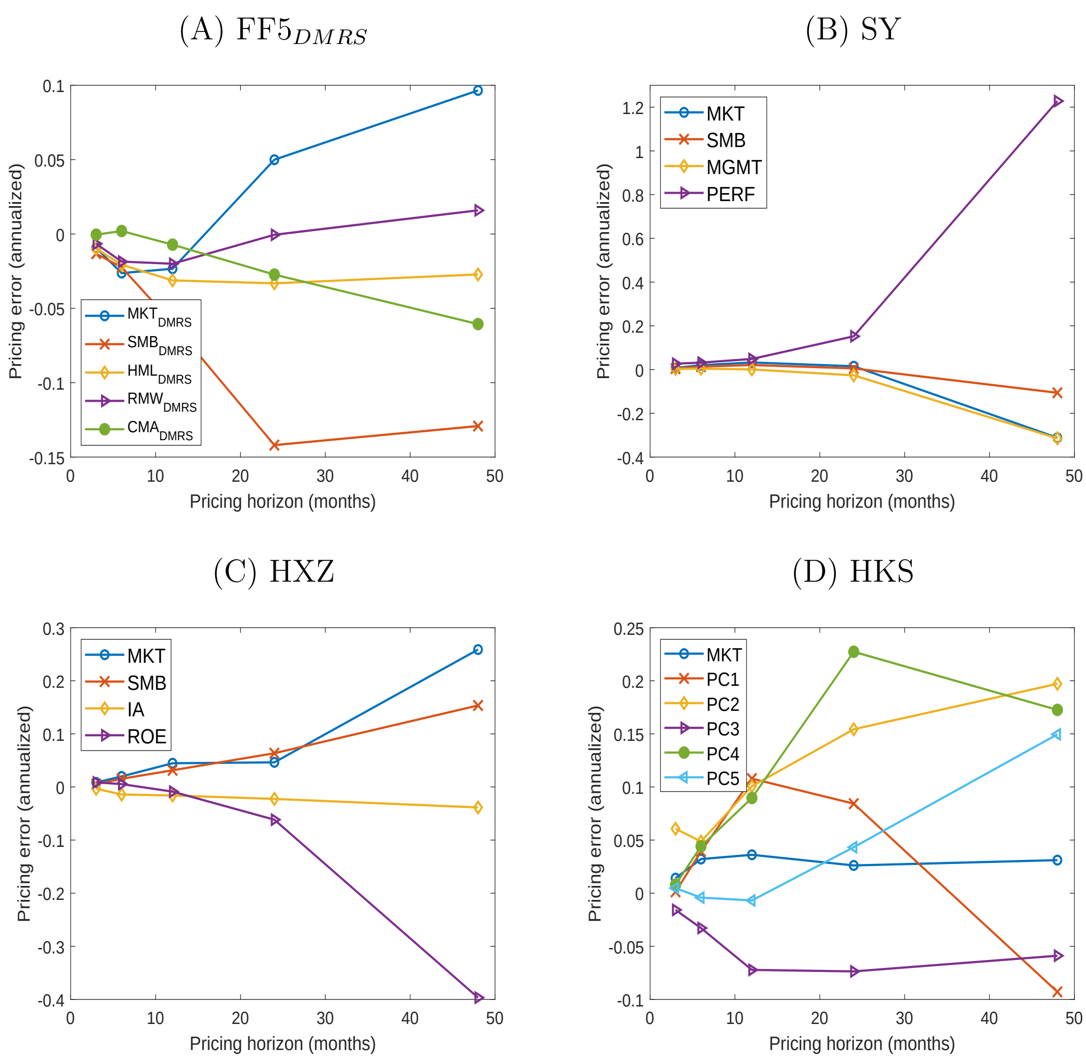
Zviadadze, Irina, 2017, Term-structure of consumption risk premia in the cross-section of currency returns, *Journal of Finance* 72, 1529–1566.

## Figure 1
## Term structure of annualized factor pricing errors I



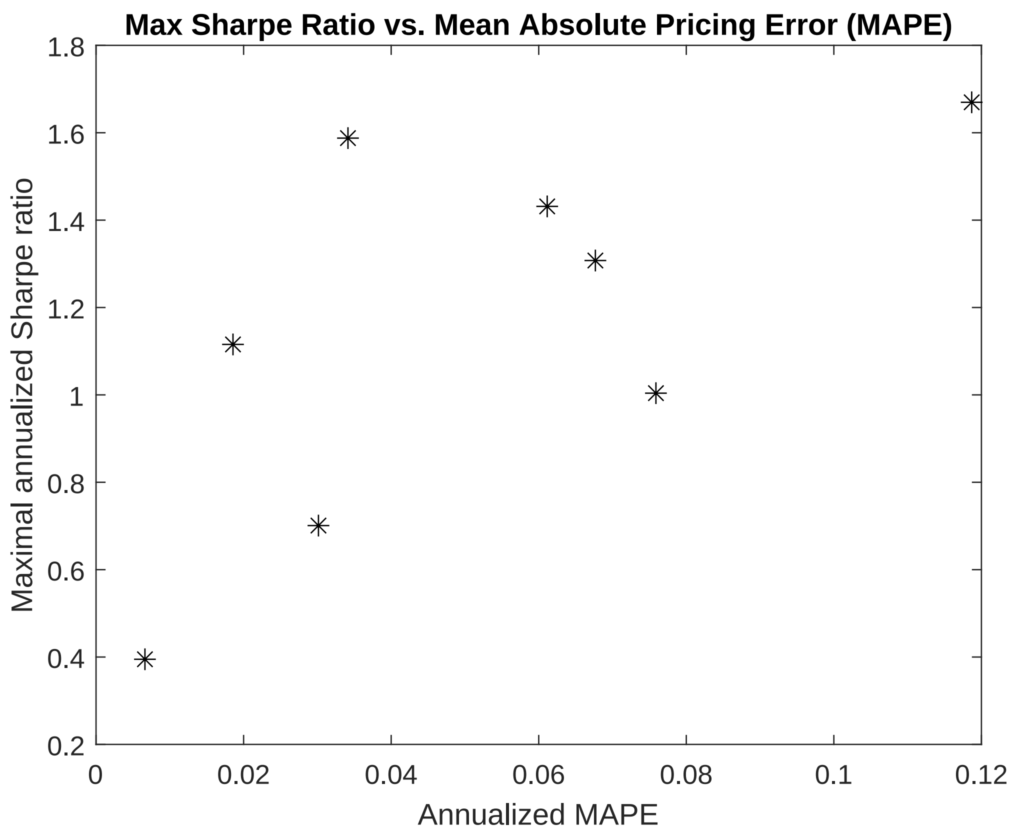

The panels show factor pricing errors for various models at horizons 3, 6, 12, 24, and 48 months. Annualized pricing errors at horizon $h$ are $12/h \times E_T(z_{i,t}^{(h)} M_{t,t+1} F_{t,t+1}^i)$, where $E_T$ denotes the sample average, $z_{i,t}^{(h)}$ is the endogenous conditioning variable for factor $i$ at horizon $h$ described in the main text, and $F_{t,t+1}^i$ is the return to factor $i$. The population average of a correctly specified model is zero. The sample is monthly, from 1963 to 2017.

## Figure 2
## Term structure of annualized factor pricing errors II



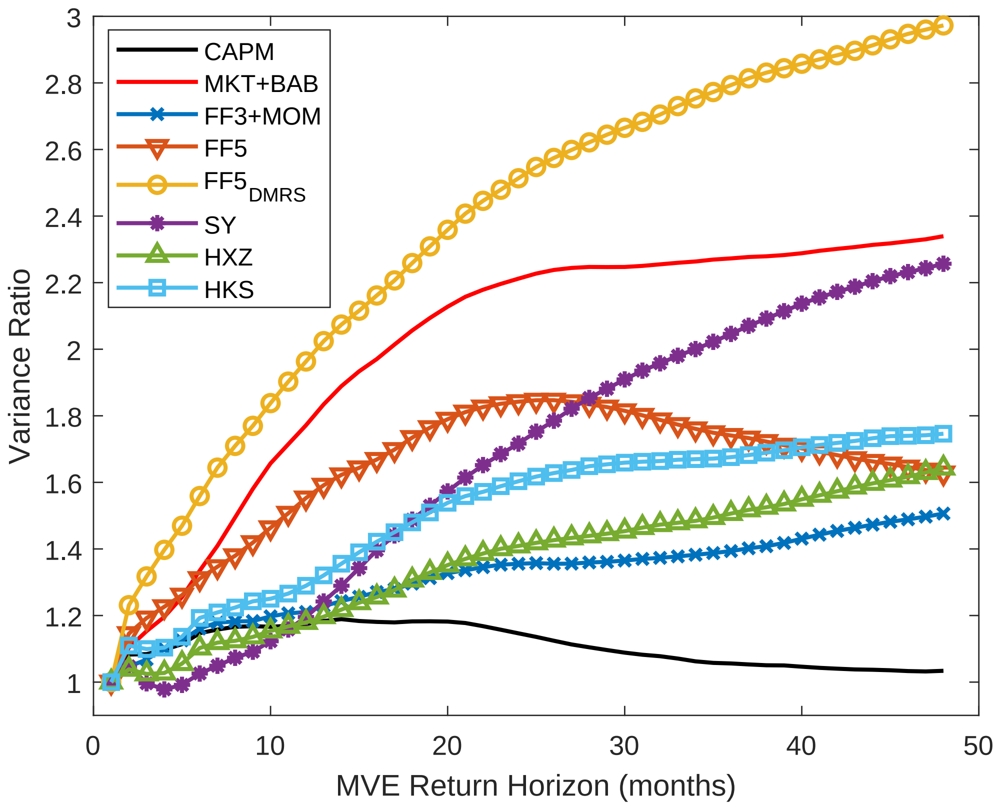

The panels show factor pricing errors for various models at horizons 3, 6, 12, 24, and 48 months. Annualized pricing errors at horizon $h$ are $12/h \times E_T(z_{i,t}^{(h)} M_{t,t+1} F_{t,t+1}^i)$, where $E_T$ denotes the sample average, $z_{i,t}^{(h)}$ is the endogenous conditioning variable for factor $i$ at horizon $h$ described in the main text, and $F_{t,t+1}^i$ is the return to factor $i$. The population average of a correctly specified model is zero. The sample is monthly, from 1963 to 2017 for $FF5_{DMRS}$, 1963 to 2016 for SY, 1967 to 2017 for HXZ, and 1974 to 2017 for HKS.

## Figure 3
## Max Sharpe ratio of single-horizon factor model vs. multi-horizon pricing errors



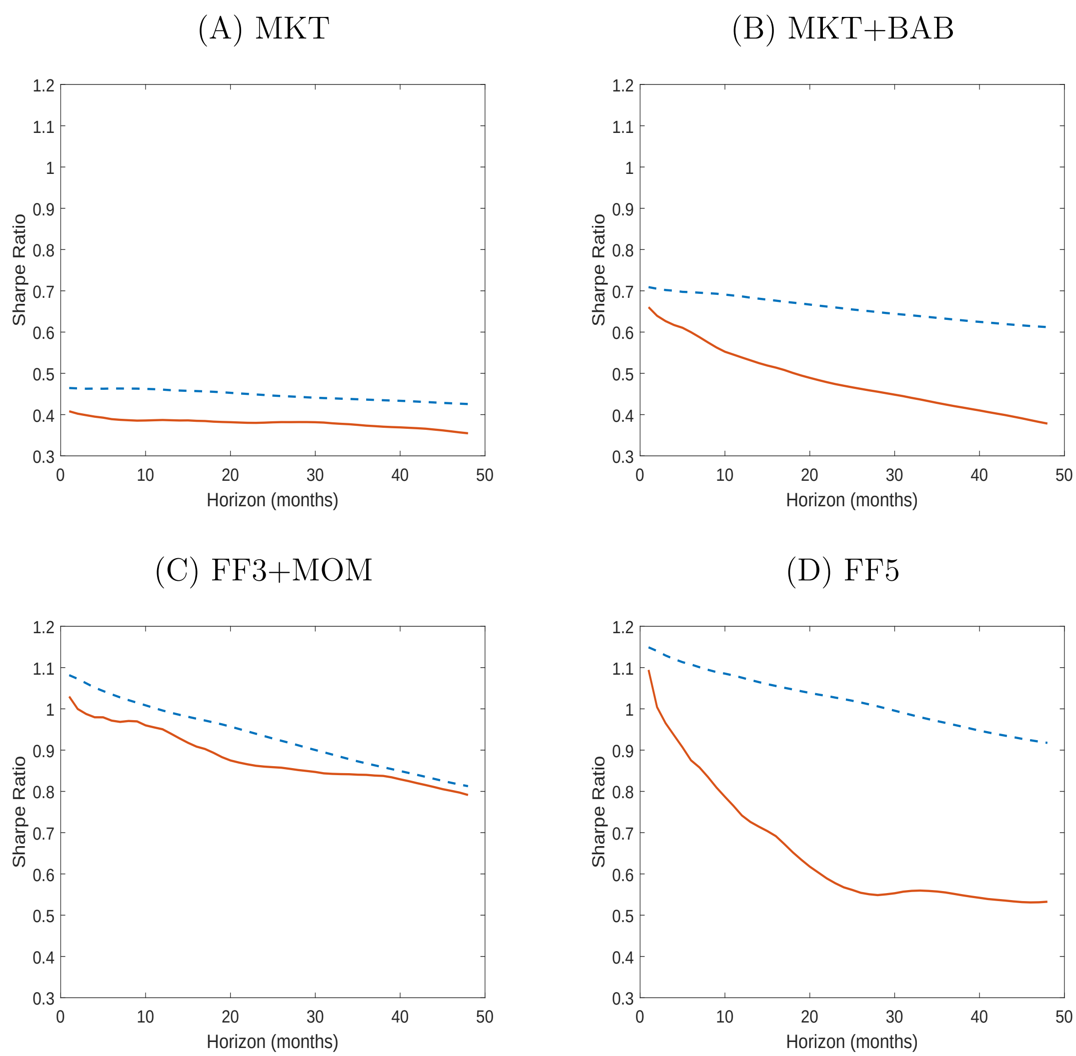

The figure plots the annualized maximal in-sample Sharpe ratio combination of the factors in each model against the annualized mean absolute pricing error (MAPE) of the corresponding model, when the model is estimated using one-period returns and tested on excess factor returns with horizons 1, 3, 6, 12, 24, and 48 months. The sample is monthly, from 1963 to 2017 for all models except SY, which is 1963 to 2016, HXZ, which is 1967 to 2017, and HKS, which is 1974 to 2017.

## Figure 4
## Variance Ratios



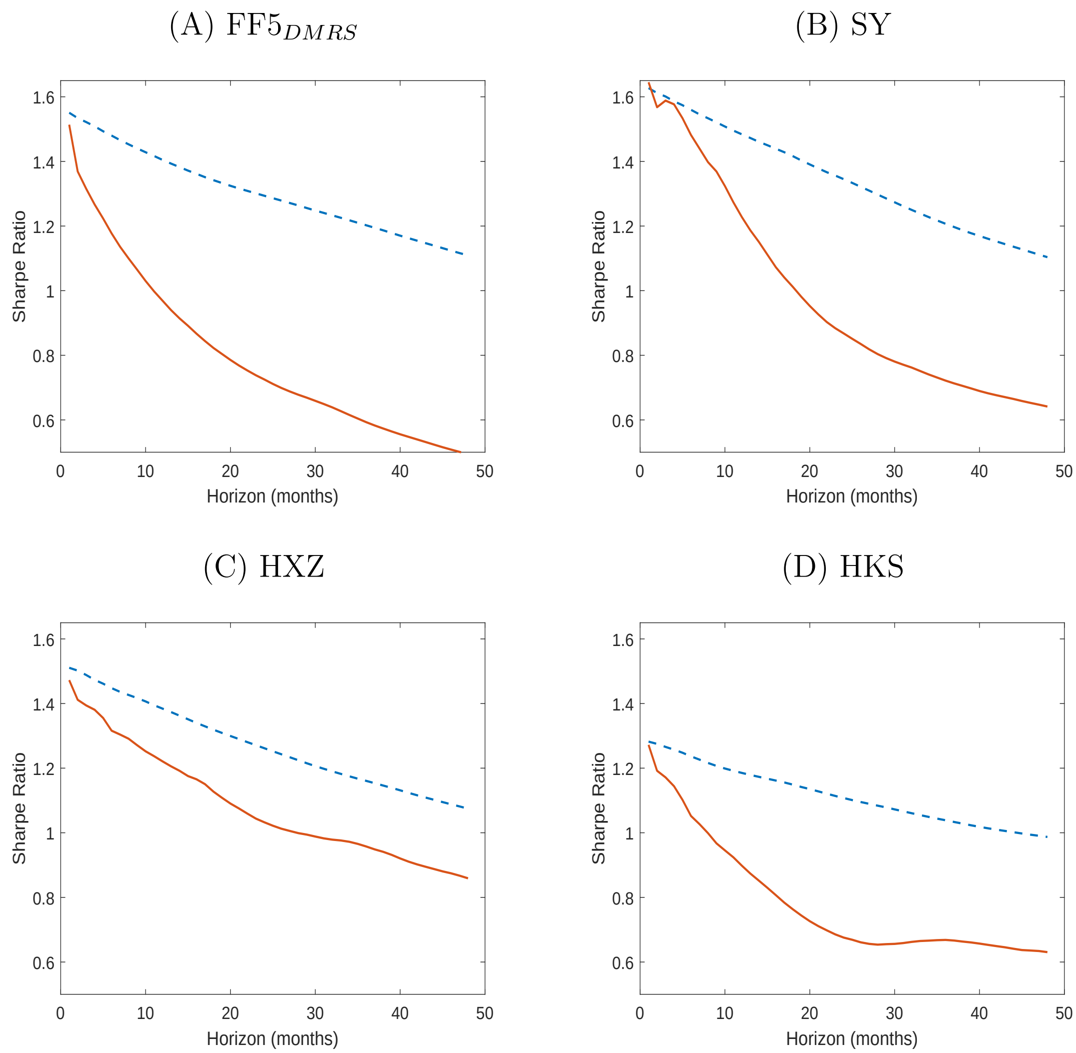

The figure plots variance ratios for the MVE portfolio from each factor model. In particular, for each model we consider the in-sample MVE combination of the factors normalized to have the same return volatility as the market factor. We then add the gross real risk-free rate to this factor return, take logs and compute the variance ratio for each model from horizons 1 to 48 months. If the factor returns are i.i.d., the variance ratio is 1 at all horizons. The variance ratio at horizon $h$ is related to the cumulative autocorrelations of the return series from horizons 1 through $h$. The sample is monthly, from 1963 to 2017 for all models except SY, which is 1963 to 2016, HXZ, which is 1967 to 2017, and HKS, which is from 1974 to 2017.

# Figure 5
# Term structure of Sharpe ratios I




The figure plots the annualized Sharpe ratio of each model's MVE portfolio for different holding periods. For each model we consider the in-sample MVE combination of the factors normalized to have the same return volatility as the market factor. We then add the gross real risk-free rate to this factor return and get $h$-period returns to this portfolio as $R_{t,t+h} = R_{t,t+1} \times R_{t+1,t+2} \times ... \times R_{t+h-1,t+h}$. The $h$-period risk-free rate is found in the same way. We then calculate the $h$-period annualized Sharpe ratio as $\sqrt{12/h} \times E(R_{t,t+h} - R^f_{t,t+h})/V^{1/2}(R_{t,t+h} - R^f_{t,t+h})$ for horizons 1 to 48 months (solid, red line). The dashed, blue line gives the corresponding Sharpe ratios using a bootstrap approach that creates i.i.d. factor returns. The sample is monthly, from 1963 to 2017 for all models.

## Figure 6
## Term structure of Sharpe ratios II




The figure plots the annualized Sharpe ratio of each model's MVE portfolio for different holding periods. For each model we consider the in-sample MVE combination of the factors normalized to have the same return volatility as the market factor. We then add the gross real risk-free rate to this factor return and get $h$-period returns to this portfolio as $R_{t,t+h} = R_{t,t+1} \times R_{t+1,t+2} \times ... \times R_{t+h-1,t+h}$. The $h$-period risk-free rate is found in the same way. We then calculate the $h$-period annualized Sharpe ratio as $\sqrt{12/h} \times E(R_{t,t+h} - R^f_{t,t+h})/V^{1/2}(R_{t,t+h} - R^f_{t,t+h})$ for horizons 1 to 48 months (solid, red line). The dashed, blue line gives the corresponding Sharpe ratios using a bootstrap approach that creates i.i.d. factor returns. The sample is monthly, from 1963 to 2017 for $FF5_{DMRS}$, from 1963 to 2016 for SY, 1967 to 2017 for HXZ, from 1974 to 2017 for HKS.

Table 1: MHR tests of linear factor models

| **Panel A:** | **CAPM** | **MKT+BAB** | **FF3+MOM** | **FF5** |
|---|---|---|---|---|
| p-value | 0.191 | 0.048 | 0.032 | 0.013 |
| MAPE | 0.007 | 0.030 | 0.076 | 0.019 |
| Sharpe Ratio | 0.395 | 0.701 | 1.004 | 1.116 |
| Information Ratio | 0.380 | 0.613 | 0.909 | 1.025 |

| **Panel B:** | **FF5**$_{DMRS}$ | **SY** | **HXZ** | **HKS** |
|---|---|---|---|---|
| p-value | 0.006 | 0.017 | 0.007 | 0.009 |
| MAPE | 0.034 | 0.119 | 0.061 | 0.068 |
| Sharpe Ratio | 1.588 | 1.670 | 1.431 | 1.308 |
| Information Ratio | 1.017 | 0.939 | 0.907 | 1.135 |

The first row of each panel gives the $p$-value from the GMM $J$-test, where the linear factor models are estimated on the one-period factor returns and tested on multi-horizon factor returns. The second row displays the annualized mean absolute price error ($MAPE$) across the test assets. The returns horizons used are 1, 3, 6, 12, 24, and 48 months. The table also reports the sample Sharpe ratio of the in-sample MVE combination of each model's factors, as well as maximal annualized Information ratio implied by the MHR returns. The sample is monthly, from 1963 to 2017 for all models except SY, which is 1963 to 2016, HXZ, which is 1967 to 2017, and HKS, which 1974 to 2017.

Table 2: MHR tests of conditional linear factor models

| **Panel A:** | **CAPM** | **MKT+BAB** | **FF3+MOM** | **FF5** |
|---|---|---|---|---|
| p-value | 0.923 | 0.263 | 0.021 | 0.009 |
| MAPE | 0.001 | 0.006 | 0.021 | 0.109 |
| Sharpe Ratio | 0.396 | 0.912 | 1.206 | 1.372 |
| Information Ratio | 0.326 | 0.633 | 0.941 | 1.001 |

| **Panel B:** | **FF5**$_{DMRS}$ | **SY** | **HXZ** | **HKS** |
|---|---|---|---|---|
| p-value | 0.018 | 0.000 | 0.039 | 0.049 |
| MAPE | 0.199 | 0.126 | 0.038 | 0.416 |
| Sharpe Ratio | 1.816 | 1.582 | 1.581 | 1.472 |
| Information Ratio | 0.985 | 1.244 | 0.817 | 1.205 |

This table reports test statistics from the factor models with time-varying SDF loadings $b_t$, as opposed to the constant $b$ model tests given in Table 1. The first row of each panel gives the $p$-value from the GMM $J$-test. The returns horizons used in the test are 1, 3, 6, 12, 24, and 48 months. The second row gives the mean absolute pricing errors (MAPE) of the model factors across horizons, excluding the pricing errors of the timing portfolio so as to be comparable to the MAPEs in Table 1. The third row gives the sample annualized Sharpe ratio of the unconditional MVE portfolio as implied by the model with time-varying $b_t$, as well as maximal annualized Information ratio implied by all MHR returns. The sample is monthly, from 1963 to 2017 for all models except SY, which is 1963 to 2016, HXZ, which is 1967 to 2017, and HKS, which 1974 to 2017.

Table 3: MHR tests of conditional linear factor models: out-of-sample conditioning

| **Panel A:**<br>**Volatility Timing** | **FF3+MOM** | **FF5** |
|---|---|---|
| p-value | 0.002 | 0.000 |
| MAPE | 0.052 | 0.517 |
| Sharpe Ratio | 0.917 | 0.967 |
| Information Ratio | 1.224 | 1.302 |

| **Panel B:**<br>**HKS Factor Timing** | $E_t$ **only** | $E_t$ **and** $V_t$ |
|---|---|---|
| p-value | 0.002 | 0.000 |
| MAPE | 0.126 | 0.055 |
| Sharpe Ratio | 1.093 | 1.016 |
| Information Ratio | 1.564 | 1.667 |

This table reports test statistics from the factor models with time-varying SDF loadings $b_t$, as opposed to the constant $b$ model tests given in Table 1. In contrast to Table 2, $b_t$ are computed out-of-sample. The exception is HKS, where $E_t$ is computed out-of-sample but the conditional covariance $V_t$ is computed in-sample. The first row of each panel gives the $p$-value from the GMM $J$-test. The returns horizons used in the test are 1, 3, 6, 12, 24, and 48 months. The second row gives the mean absolute pricing errors (MAPE) of the model factors across horizons, excluding the pricing errors of the timing portfolio so as to be comparable to the MAPEs in Table 1. The third row gives the sample annualized Sharpe ratio of the unconditional MVE portfolio as implied by the model with time-varying $b_t$, while the fourth row gives the maximal Information ratio implied by all MHR returns. The sample is monthly, from 1963 to 2017 for Volatility Timing, and from 1996 to 2017 for HKS.

# A Appendix

## A.1 UMVE from CMVE

Start with the SDF derived from CMVE:

$$M^*_{t,t+1} = 1 + k_t E_t\left(R^C_{t,t+1}\right) - k_t R^C_{t,t+1}.$$

Note that if $E_t(M^*_{t,t+1}R^e_{t,t+1}) = 0$, then $E_t(a_t \cdot M^*_{t,t+1}R^e_{t,t+1}) = 0$, if $a_t$ is known at time $t$. Thus, dividing $M^*_{t,t+1}$ by $1 + k_t E_t(R^C_{t,t+1})$ we obtain another SDF that prices the same set of assets:

$$\begin{aligned}
\widetilde{M}_{t+1} &= 1 - \frac{k_t}{1 + k_t E_t\left(R^C_{t,t+1}\right)} R^C_{t,t+1} \\
&= 1 - \frac{k_t}{1 + E_t\left(R^e_{t,t+1}\right)^\top V_t\left(R^e_{t,t+1}\right)^{-1} E_t\left(R^e_{t,t+1}\right)} R^C_{t,t+1} \\
&= 1 - \delta_t R^C_{t,t+1} \\
&= 1 - E(R^U_{t,t+1}) - \left(R^U_{t,t+1} - E(R^U_{t,t+1})\right),
\end{aligned}$$

where $R^U_{t,t+1} \equiv \delta_t R^C_{t,t+1}$ .

Lastly, divide by the constant $1 - E(R^U_{t,t+1})$ to get the final version of the SDF that still prices the same set of assets:

$$M_{t,t+1} = 1 - (1 - E(R^U_{t,t+1}))^{-1}\left(R^U_{t,t+1} - E(R^U_{t,t+1})\right).$$

## A.2 Misspecified model with persistent errors

We have:

$$\begin{aligned}
E_t\left[\widetilde{M}_{t+1}F_{t+1}\right] &= E_t\left[(1 - b(F_{t+1} - E\left[F_{t+1}\right])F_{t+1}\right] \\
&= E_t\left[F_{t+1}\right]\left(1 + bE\left[F_{t+1}\right]\right) - bE_t\left[F^2_{t+1}\right].
\end{aligned}$$

The correctly specified model implies that

$$E_t\left[F_{t+1}\right]\left(1 + E\left[b_t F_{t+1}\right]\right) - b_t E_t\left[F^2_{t+1}\right] = 0.$$

Thus,

$$E_t\left[F_{t+1}\right] = \frac{b_t}{1+E\left[b_t F_{t+1}\right]} E_t\left[F_{t+1}^2\right],$$

and

$$\begin{aligned} E_t\left[\widetilde{M}_{t+1} F_{t+1}\right] &= E_t\left[F_{t+1}\right]\left(1+bE\left[F_{t+1}\right]\right) - bE_t\left[F_{t+1}^2\right] \\ &= \left(\frac{1+bE[F_{t+1}]}{1+E\left[b_t F_{t+1}\right]} b_t - b\right) E_t\left[F_{t+1}^2\right]. \end{aligned}$$

## A.3 Misspecified model with i.i.d. errors

The SDF prices $R_{t,t+1}^{ie}$ unconditionally:

$$\begin{aligned} E\left[\widetilde{M}_{t,t+1} R_{t,t+1}^{ie}\right] &= E\left[\left(1-b\left(F_{t,t+1}-E\left[F_{t,t+1}\right]\right)\right)\left(\beta_i F_{t,t+1}+\varepsilon_{i,t+1}+\theta\varepsilon_{i,t}\right)\right] \\ &= \beta_i E\left[F_{t,t+1} - \frac{E\left[F_{t,t+1}\right]}{Var\left(F_{t,t+1}\right)}\left(F_{t,t+1}-E\left[F_{t,t+1}\right]\right) F_{t,t+1}\right] = 0. \end{aligned}$$

However, that is not the case conditionally:

$$\begin{aligned} E_t\left[\widetilde{M}_{t,t+1} R_{t,t+1}^{ie}\right] &= E_t\left[\left(1-b\left(F_{t,t+1}-E\left[F_{t,t+1}\right]\right)\right)\left(\beta_i F_{t,t+1}+\varepsilon_{i,t+1}+\theta\varepsilon_{i,t}\right)\right] \\ &= \beta_i E_t\left[F_{t,t+1} - \frac{E\left[F_{t,t+1}\right]}{Var\left(F_{t,t+1}\right)}\left(F_{t,t+1}-E\left[F_{t,t+1}\right]\right) F_{t,t+1}\right] \\ &\quad + E_t\left[\left(1-b\left(F_{t,t+1}-E\left[F_{t,t+1}\right]\right)\right)\theta\varepsilon_{i,t}\right] \\ &= \theta\varepsilon_{i,t}. \end{aligned}$$

Next, we show that $Cov\left(\nu_{t-1,t}, \eta_t^{(1)}\right) \neq 0$. Because

$$\eta_{t-1}^{(1)} + \nu_{t-1,t} = \left(1-b\left(F_{t-1,t}-E\left[F_{t-1,t}\right]\right)\right)\left(\beta_i F_{t-1,t}+\varepsilon_{i,t}+\theta\varepsilon_{i,t-1}\right),$$

we have that:

$$Cov\left(\eta_{t-1}^{(1)} + \nu_{t-1,t}, \theta\varepsilon_{i,t}\right) = \theta Var\left(\varepsilon_{i,t}\right).$$

Because $Cov(\eta_{t-1}^{(1)}, \theta\varepsilon_{t,t}) = Cov(\eta_{t-1}^{(1)}, \eta_t^{(1)}) = 0$, $Cov\left(\nu_{t-1,t}, \eta_t^{(1)}\right) = \theta Var\left(\varepsilon_{i,t}\right).$

## A.4 No serial correlation in residuals

That residuals are not autocorrelated follows from Equation (14). For simplicity, consider one horizon, $h$. We have that

$$E(f_t^i \cdot f_{t+1}^i) = E(f_t^i \cdot E_t(f_{t+1}^i)) = E(f_t^i \cdot z_{i,t}^{(h)} E_t(M_{t,t+1} R_{t,t+1}^i - 1)) = 0,$$

because, under the null, $E_t(M_{t,t+1} R_{t,t+1}^i - 1) = 0$ for all $t$.

## A.5 Construction of the BAB factor

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

The factor return is then constructed as follows:

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

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

## A.6 Factor dynamics implied by constant SDF loadings

First, apply the SDF in Equation (10) to unconditional pricing of the factors themselves

$$
\begin{aligned}
E(M_{t,t+1}F_{t,t+1}) &= E((1 - b^\top(F_{t,t+1} - \mu))F_{t,t+1}^\top) \\
&= \mu^\top - b^\top\Sigma \\
&= 0,
\end{aligned}
$$

where we denote $E(F_{t,t+1}) = \mu$ and $V(F_{t,t+1}) = \Sigma$. This implies that $b = \Sigma^{-1}\mu$. Second, apply the same SDF to conditional pricing of the factors, $E_t(M_{t,t+1}F_{t,t+1}) = 0$, using this value for $b$. This yields:

$$
0 = E_t((1 + \mu^\top\Sigma^{-1}\mu - \mu^\top\Sigma^{-1} \cdot F_{t,t+1})F_{t,t+1}^\top).
$$

Therefore,

$$
E_t(F_{t,t+1}^\top) = (1 + \mu^\top\Sigma^{-1}\mu)^{-1}\mu^\top\Sigma^{-1} \cdot E_t(F_{t,t+1}F_{t,t+1}^\top).
$$

In words, the conditional mean of the factors that unconditionally span the UMVE portfolio is proportional to the conditional second moment.