

ASSET PRICING ANOMALIES AND TIME-VARYING BETAS:
A NEW SPECIFICATION TEST FOR CONDITIONAL FACTOR MODELS¹

DEVRAJ BASU
ALEXANDER STREMME

Warwick Business School, University of Warwick

JANUARY 2006

address for correspondence:

*Alexander Stremme
Warwick Business School
University of Warwick
Coventry, CV4 7AL
United Kingdom*

e-mail: `alex.stremme@wbs.ac.uk`

¹ We would like to thank Wayne Ferson for helpful discussions and suggestions. We also thank Kenneth French and Doron Avramov for making available the data used in this study. All remaining errors are ours.

ASSET PRICING ANOMALIES AND TIME-VARYING BETAS:
A NEW SPECIFICATION TEST FOR CONDITIONAL FACTOR MODELS

JANUARY 2006

Abstract

In this paper, we develop a new measure of specification error, and thus derive new statistical tests, for conditional factor models, i.e. models in which the factor loadings (and hence risk premia) are allowed to be time-varying. Our test exploits the close links between the stochastic discount factor framework and mean-variance efficiency. We show that a given set of factors is a true conditional asset pricing model if and only if the efficient frontiers spanned by the traded assets and the factor-mimicking portfolios, respectively, intersect. In fact, we show that our test is proportional to the difference in squared Sharpe ratios of these two frontiers.

We draw three main conclusions from our empirical findings. First, optimal scaling clearly improves the performance of asset pricing models, to the point where several of the scaled models are capable of explaining asset pricing anomalies. However, even the optimally scaled models fall short of being true conditional asset pricing models in that they fail to price actively managed portfolios correctly. Second, there is significant time-variation in factor loadings and hence risk premia, which plays a significant role in asset pricing. Moreover, the optimal factor loadings display a high degree of non-linearity in the conditioning variables, suggesting that the linear scaling prevalent in the literature is sub-optimal and does not capture the inter-temporal pattern of risk premia. Third, skewness and kurtosis do matter in the conditional setting, while adding little to unconditional performance.

JEL CLASSIFICATION: C31, G11, G12

1 Introduction

Asset pricing research has documented that the cross-sectional dispersion in expected returns is determined not only by beta, as prescribed by the CAPM, but by firm size, book-to-market ratio, and other factors. Subsequent work has advocated time-varying betas to capture such deviations, often termed financial market anomalies, from the unconditional CAPM. For example, Chan and Chen (1988), and Ball and Kothari (1989) find that beta variation explains the size premium and the profitability of contrarian strategies. Berk, Green, and Naik (1999), and Gomes, Kogan, and Zhang (2003) suggest that beta depends upon the business cycle as well as on firm size and book-to-market ratio. These findings have led to the introduction of *conditional* factor models, in which the factor loadings are allowed to be time-varying, for example as functions of macro-economic or firm-level variables¹.

In this paper, we construct a new measure of specification error which provides necessary and sufficient conditions for a given set of factors to constitute a viable conditional asset pricing model. While traditional asset pricing models are designed to explain unconditional risk premia in terms of factor risk exposure, conditional models capture the cross-sectional *and* intertemporal variation of conditional risk premia. Static models require additional factors to explain the abnormal returns due to asset pricing anomalies. Often, these factors are derived from the very anomalies they are designed to explain². In contrast, the empirical evidence suggests that models with time-varying factor loadings ('betas') can capture much of the cross-section of expected returns without the need for additional factors. Moreover, such models are also able to capture the intertemporal dynamics of these anomalies.

¹Lewellen (1999), Ferson and Harvey (1998, 1999), and Avramov and Chordia (2004), among others, have studied multi-factor models with time-varying factor loadings. Avramov and Chordia (2004) find that time-varying factor loadings could explain size and book-to-market effects.

²For example, the Fama-French factors are constructed by sorting stocks on size and book-to-market ratio.

Our specification test is designed to measure how well a given factor model prices the traded assets, both unconditionally *and* conditionally. It is easy to show that, if a factor model prices the traded assets conditionally correctly, it must necessarily also price all actively managed strategies (unconditionally) correctly. The most difficult active strategies to price are the ones that use conditioning information *optimally* in the sense that they are unconditionally mean-variance efficient. Our test exploits the close links between the stochastic discount factor framework and mean-variance efficiency. In particular, we show that the pricing error of a given factor model is related to the position of the efficient frontier spanned by the factors (or factor-mimicking portfolios) relative to the asset frontier. More specifically, a given set of factors can be a true asset pricing model if and only if these frontiers intersect. We show that our measure is proportional to the difference in squared Sharpe ratios of the asset and the factor frontiers.

In this paper we make several methodological and empirical contributions. We find that our optimal scaling using macro-economic variables improves the performance of asset pricing models quite significantly. We find that a three factor CAPM where the market return is augmented by skewness and kurtosis factors and the factor loadings are time-varying is capable of pricing momentum portfolios and comes quite close to pricing portfolios sorted by size and book-to-market. Our findings suggest that it may be possible to explain size and value premia as well as momentum without having to construct factors that are based on them. However none of the models is capable of pricing the managed portfolios and are thus not true asset pricing models in our setting.

Our theoretical and methodological contributions are fourfold. First, in the case of non-traded factors we explicitly construct the factor-mimicking portfolios in the presence of conditioning information, generalizing the results of Ferson, Siegel, and Xu (2005). Second, we extend the measure of specification error of Hansen and Jagannathan (1997) to the case with conditioning information, and relate our measure to the difference in maximum squared Sharpe ratios. Our test is thus an extension of the Gibbons, Ross, and Shanken (1989) test to the case with conditioning information. Our third contribution is to construct the actively

managed portfolios that attain the maximum Sharpe ratios for both the base assets and the factors. We show that, if the asset and factor frontiers indeed intersect, the corresponding factor portfolio can be lifted to give a valid stochastic discount factor. However, even if our test rejects the model, our methodology allows us to construct the best possible conditional factor model for a given choice of factors in the sense that it minimizes conditional pricing error. In either case, our framework allows us to explicitly construct the optimal factor loadings as functions of the conditioning instruments. Earlier conditional factor models, e.g. Jagannathan and Wang (1996), and Ferson and Harvey (1999) have assumed the factor loadings to be linear in the instruments. However, recent studies have found that this specification can lead to serious pricing errors. For example, Ghysels (1998) shows that a linearly scaled model in fact under-performs an unscaled model with constant factor loadings out-of-sample. Brandt and Chapman (2005) show that if the true model exhibits even mild non-linearities, estimating a linear beta model leads to considerable mis-pricing. In line with these findings, our analysis shows that the optimal use of conditioning information in the construction of factor models leads to highly non-linear factor loadings.

We implement our measure to test the performance of a number of factor models considered in the literature. We use commonly used predictive instruments that measure macro-economic and interest rate risk³.

We first consider the performance of our models on the Fama-French 30 industry portfolios, a benchmark set of base assets that is free of any asset pricing anomalies. Optimal scaling increases the asset Sharpe ratio more than one and a half times while the Sharpe ratio of the CAPM and Fama-French models increase by 48% and 160% respectively. Adding skewness and kurtosis factors to both the CAPM and Fama-French models leads to little improvement in the fixed-weight models but strong effects on the scaled models with the Sharpe ratios

³The instruments are unexpected shocks to inflation, the 1-month Treasury Bill rate, the term spread, convexity which is a measure of curvature of the yield curve and credit spread. These instruments thus capture both information about interest rates, as well as changes in the macro-economic environment.

increasing by 76% for the CAPM and 45% for the Fama-French model. The scaled three factor CAPM prices the base assets. However none of the models are capable of pricing the managed assets with the scaled Fama-French model augmented by skewness, kurtosis and the momentum factors achieving 63% of the scaled asset Sharpe ratio. The outperformance of the optimally scaled version relative to the corresponding fixed-beta specification. is thus in contrast to Ghysels (1998). This shows that our optimal scaling indeed reduces misspecification error relative to fixed-beta models, which is not necessarily the case for the linearly scaled beta models prevalent in most of the literature. Our findings also suggest that, while skewness and kurtosis effects are ‘washed out’ in the unconditional setting by time-aggregation, they have a significant effect in the conditional setting. In other words, the non-normality and in particular asymmetry of returns carries a significant time-varying risk-premium in the short run, which tends to vanish at longer horizons.

We focus next on the size and book-to-market effects and use the 25 portfolios sorted by these characteristics as our base assets. We find that the effect of optimal scaling almost dominates that of sorting assets on these criteria, as while the fixed weight Sharpe ratio of these portfolios is almost double that of the thirty industry portfolios the scaled Sharpe ratios are very close. The unscaled CAPM and Fama-French models are unable to price these portfolios (with pricing errors of 74% and 45%, respectively). In other words, over the time period we consider, even the Fama-French model fails to price the size and value portfolios. Moreover, even with optimal scaling, both models are unable to price the assets (with pricing errors of 68% and 28%, respectively). In contrast, when we augment these models by skewness and kurtosis factors, we find that an optimally scaled three-factor CAPM achieves 90% of the fixed-weight asset Sharpe ratio, while the scaled augmented Fama-French model actually prices the assets correctly. Our results thus show that a scaled three factor CAPM goes a long way towards explaining the size and book-to-market effects.

We also consider the performance of the models in pricing the momentum portfolios. Momentum remains the one CAPM-based anomaly that cannot be explained by the Fama-French model. We use the momentum portfolios constructed by Chordia and Shivakumar (2002)

that span the 1960-1999 period. We find that the fixed-weight Fama-French model achieves 82% of the Sharpe ratio of the assets and, while it performs considerably better than the unscaled CAPM which achieves just over 50% of the asset Sharpe ratio, still does not price the momentum portfolios correctly. We consider a three factor CAPM, that is the market return together with the skewness and kurtosis factors scaled by a number of common business cycle variables. The potential importance of the skewness factor in pricing momentum returns was first observed by Harvey and Siddique (2000) while the ability of business cycle variables to predict momentum profits was analyzed in Chordia and Shivakumar (2002) We find that this scaled three factor CAPM achieves a Sharpe ratio of 1.23 which is higher than the fixed weight asset Sharpe ratio of 1.15 and thus this model prices the momentum portfolios correctly. It is however not a true asset pricing model as it fails to price the scaled momentum portfolios, but still appears to be the first rational asset pricing model that prices the unscaled momentum portfolios.

We draw three main conclusions from our empirical findings. First, optimal scaling clearly improves the performance of asset pricing models, to the point where several of the scaled models are capable explaining some asset pricing anomalies. However, even the optimally scaled models fall short of being true conditional asset pricing models in that they fail to price actively managed portfolios correctly. Second, there is significant time-variation in factor loadings and hence risk premia, which plays a significant role in asset pricing. Moreover, the optimal factor loadings display a high degree of non-linearity in the conditioning variables, suggesting that the linear scaling prevalent in the literature is sub-optimal⁴, and does not capture the inter-temporal pattern of risk premia. Third, skewness and kurtosis do matter in the conditional setting, while adding little to unconditional performance. Our results thus expand upon the findings of Kraus and Litzenberger (1976), Harvey and Siddique (2000), as well as Dittmar (2002).

Finally, we check for the robustness of our results. Our tests are all based on Sharpe

⁴See for example Ghysels (1998).

ratios which can exhibit a high degree of sampling variability particularly in the presence of conditioning information, as observed in Ferson and Siegel (2003). To see to what extent our results may be driven by this sampling variability we simulate a pure noise variable as our predictive instrument and compute the mean and percentiles of the asset and factor Sharpe ratios. We find that although the optimal factor Sharpe ratios are higher than the fixed weight factor Sharpe ratios, they are never as high as the fixed weight asset Sharpe ratios for all our asset sets. Our optimal asset and factor Sharpe ratios using the actual predictive instruments are always well above the simulated 95% confidence levels. This shows that our findings are robust to sampling variability.

The remainder of the paper is organized as follows. In Section 2, we briefly review the theory of conditional factor models in the context of a stochastic discount factor framework. In the following section, we derive explicit expressions for the factor-mimicking portfolios, and develop our test statistic. The results of our empirical analysis are reported in Section 4. All mathematical proofs are given in the appendix.

2 Set-Up and Notation

The aim of this paper is to construct a measure of misspecification for factor models in the presence of conditioning information, based on properties of unconditionally efficient portfolios. To this end, we construct the space of state-contingent pay-offs, and within it the space of traded pay-offs, augmented by the use of conditioning information. For a given set of factors, we then construct the subset of augmented pay-offs that is spanned by the corresponding factor-mimicking portfolios. Our test, developed in the following section, is based on a notion of ‘distance’ which measures whether the factor-mimicking portfolios span the efficient frontier in the augmented base asset space.

2.1 Traded Assets and Managed Pay-Offs

The information flow in the economy is described by a discrete-time filtration $(\mathcal{F}_t)_t$, defined on some probability space (Ω, \mathcal{F}, P) . We fix an arbitrary $t > 0$, and consider the period beginning at time $t - 1$ and ending at t . Denote by L_t^2 the space of all \mathcal{F}_t -measurable random variables that are square-integrable with respect to P . We interpret Ω as the set of ‘states of nature’, and L_t^2 as the space of all (not necessarily attainable) state-contingent pay-off claims, realized at time t .

TRADED ASSETS:

There are n traded risky assets, indexed $k = 1 \dots n$. We denote the gross return (per dollar invested) of the k -th asset by $r_t^k \in L_t^2$, and by $\tilde{R}_t := (r_t^1 \dots r_t^n)'$ the n -vector of risky asset returns. In addition to the risky assets, a risk-free is traded with gross return $r_t^0 = r_f$.

CONDITIONING INFORMATION:

To incorporate conditioning information, we take as given a sub- σ -field $\mathcal{G}_{t-1} \subseteq \mathcal{F}_{t-1}$. We think of \mathcal{G}_{t-1} as summarizing all information on which investors base their portfolio decisions at time $t - 1$. In our empirical applications, \mathcal{G}_{t-1} will be chosen as the σ -field generated by one or more lagged *conditioning instruments*, variables observable at time $t - 1$ that contain information about the distribution of asset returns⁵. To simplify notation, we write $E_{t-1}(\cdot)$ for the conditional expectation operator with respect to \mathcal{G}_{t-1} .

MANAGED PORTFOLIOS:

We allow for the formation of *managed* portfolios of the base assets. To this end, denote by

⁵Examples of conditioning variables considered in the literature include, among others, dividend yield (Fama and French 1988), or interest rate spreads (Campbell 1987),

X_t the space of all elements $x_t \in L_t^2$ that can be written in the form,

$$x_t = \theta_{t-1}^0 r_f + \sum_{k=1}^n (r_t^k - r_f) \theta_{t-1}^k, \quad (1)$$

for \mathcal{G}_{t-1} -measurable functions θ_{t-1}^k . To simplify notation, we write (1) in vector form as $x_t = \theta_{t-1}^0 r_f + (\tilde{R}_t - r_f e)' \theta_{t-1}$, where e is an n -vector of ‘ones’. We interpret X_t as the space of managed pay-offs, obtained by forming combinations of the base assets with time-varying weights θ_{t-1}^k that are functions of the conditioning information.

PRICING FUNCTION:

Because the base assets are defined by their *returns*, we set $\Pi_{t-1}(r_t^k) = 1$ for $k = 0, 1, \dots, n$, and extend Π_{t-1} to all of X_t by conditional linearity. In particular, for an arbitrary pay-off $x_t \in X_t$ of the form (1), it is easy to see that $\Pi_{t-1}(x_t) = \theta_{t-1}^0$. By construction, the pricing rule Π_{t-1} satisfies the ‘law of one price’, a weak form of no-arbitrage condition.

2.2 Stochastic Discount Factors

We use the stochastic discount factor framework to define what it means for a set of factors to give rise to an admissible asset pricing model.

Definition 2.1 *By an admissible stochastic discount factor (SDF) for the model (X_t, Π_{t-1}) , we mean an element $m_t \in L_t^2$ that prices all base assets conditionally correctly, i.e.*

$$E_{t-1}(m_t r_t^k) = \Pi_{t-1}(r_t^k) = 1 \quad \text{for all } k = 0, 1, \dots, n. \quad (2)$$

The existence of at least one SDF is guaranteed by the Riesz representation theorem, but unless markets are complete it will not be unique. Much of modern asset pricing research focuses on deriving plausible SDFs from principles of economic theory, and then empirically testing such candidates against observed asset returns. Note that in our definition, the SDF is required to price the base assets *conditionally*. The vast majority of asset pricing model

tests considered in the literature have used the unconditional version of (2). In our setting, we allow both assets and factors to be dynamically managed, and thus we are testing the conditional version of the pricing equation. Note also that, if m_t is an admissible SDF in the sense of (2), linearity implies $E_{t-1}(m_t x_t) = \Pi_{t-1}(x_t)$ for any arbitrary managed pay-off $x_t \in X_t$. In other words, an SDF that prices all base assets correctly is necessarily compatible with the pricing function Π_{t-1} for managed pay-offs. Taking expectations we obtain,

$$E(m_t x_t) = E(\Pi_{t-1}(x_t)) =: \Pi_0(x_t). \quad (3)$$

In other words, any SDF that prices the base assets (conditionally) correctly must necessarily also be consistent with the unconditional pricing rule Π_0 . In fact, it is easy to show that a candidate m_t is an admissible SDF if and only if (3) holds for all $x_t \in X_t$. We can thus interpret (3) as a set of *moment conditions* that any candidate SDF must satisfy. There are many empirical techniques (e.g. GMM) to estimate and test such restrictions. However, as the space X_t of ‘test assets’ is infinite-dimensional, such tests will typically yield only necessary but not sufficient conditions for the SDF.

This problem can be overcome by exploiting the close link between the SDF framework and mean-variance efficiency. More specifically, one can obtain necessary *and* sufficient conditions by testing how the candidate SDF acts on the unconditionally efficient frontier in the space X_t of managed pay-offs (see Section 3 below). By two-fund separation, this reduces the test to a one-dimensional problem. Motivated by this observation, we set $R_t = \Pi_0^{-1}\{1\}$. In other words, R_t is the set of all managed pay-offs that have unit price and thus represent the *returns* on dynamically managed portfolios.

2.3 Conditional Factor Models

Our focus here is not the *selection* of factors, but rather the construction and testing of models for a given set of factors. Therefore, we take as given m factors, $F_t^i \in L_t^2$, indexed $i = 1 \dots m$. Denote by $\tilde{F}_t = (F_t^1, \dots, F_t^m)'$ the m -vector of factors. In general we do not assume the factors to be traded assets, that is we may have $F_t^i \notin X_t$.

Definition 2.2 We say that the model (X_t, Π_{t-1}) admits a conditional factor structure, if and only if there exist \mathcal{G}_{t-1} -measurable functions a_{t-1} and b_{t-1}^i such that,

$$m_t = \alpha_{t-1} + \sum_{i=1}^m F_t^i b_{t-1}^i \quad (4)$$

is an admissible SDF for the model in the sense of Definition 2.1.

We refer to the coefficients b_{t-1}^i as the *conditional factor loadings* of the model and write (4) in vector notation as $m_t = \alpha_{t-1} + \tilde{F}_t' b_{t-1}$. We emphasize that the above specification defines a *conditional* factor model, in that the coefficients a_{t-1} and b_{t-1}^i are allowed to be functions of the conditioning information. In other words, in this specification the conditional risk premia associated with the factors are allowed to be time-varying. This potentially gives the model the flexibility necessary to price also managed portfolios, since the co-efficients of the model can respond to the same information that is used in the formation of portfolios.

FACTOR-MIMICKING PORTFOLIOS:

Since the factors need not be traded assets, we construct factor-mimicking portfolios within the space R_t of managed returns.

Definition 2.3 An element $f_t^i \in X_t$ is called a factor-mimicking portfolio (FMP) for the factor $F_t^i \in L_t^2$ if and only if $\Pi_{t-1}(f_t^i) = 1$, and

$$\rho^2(f_t^i, F_t^i) \geq \rho^2(r_t, F_t^i) \quad \text{for all } r_t \in X_t \text{ with } \Pi_{t-1}(r_t) = 1. \quad (5)$$

Note that we define an FMP via the concept of maximal correlation with the factor. In the literature, it is also common to characterize factor-mimicking portfolios by means of an orthogonal projection⁶. However, it can be shown that these characterizations are in fact equivalent. To define our test, we now take the factor-mimicking portfolios themselves as

⁶This is for example the approach taken in Ferson, Siegel, and Xu (2005).

base assets, and consider the space of pay-offs attainable by forming managed portfolios of FMPs. Specifically, denote by X_t^F the space of all $x_t \in L_t^2$ that can be written in the form,

$$x_t = \phi_{t-1}^0 r_f + \sum_{i=1}^m (f_t^i - r_f) \phi_{t-1}^i, \quad (6)$$

for \mathcal{G}_{t-1} -measurable functions ϕ_{t-1}^i . By construction, $\Pi_{t-1}(x_t) = \phi_{t-1}^0$ for any $x_t \in X_t^F$ of the form (6). Mimicking the construction in the preceding section, we define the set of *returns* in this space as $R_t^F = R_t \cap X_t^F$.

3 Tests of Conditional Linear Factor Models

In this section, we develop a new measure of model misspecification in the presence of conditioning information. This measure gives rise to a necessary *and* sufficient condition for a given set of factors to constitute a viable asset pricing model. Moreover, we show that our measure is closely related to the shape of the efficient portfolio frontier in the augmented pay-off space.

As a starting point, we take as given an unconditionally efficient benchmark return $r_t^* \in R_t$. Although the results outlined below can be shown to be robust with respect to the choice of benchmark return, we follow Hansen and Jagannathan (1997) and take r_t^* as the return with minimum unconditional second moment in R_t .

Definition 3.1 *For given factors \tilde{F}_t , the model misspecification error is defined as,*

$$\delta_F := \inf_{r_t \in R_t^F} \sigma^2(r_t^* - r_t). \quad (7)$$

In other words, δ_F measures the minimum variance distance between the efficient benchmark return r_t^* and the return space R_t^F spanned by the factor-mimicking portfolios. In the following sections, we prove a series of results that motivate the interpretation of $\delta_F(r_t^*)$ as a measure of model misspecification. Specifically,

- (i) We show (Theorem 3.8) that for given set of factors \tilde{F}_t , the model admits a factor structure in the sense of Definition 2.2 if and only if $\delta_F = 0$. In other words, our measure defines a necessary *and* sufficient condition for a given set of factors to constitute a viable conditional asset pricing model.
- (ii) By construction, r_t^* attains the maximum Sharpe ratio λ_* in the space R_t^G of generalized returns. We show (Proposition 3.6) that any $r_t^F \in R_t^F$ that attains the minimum in (7) also attains the maximum Sharpe ratio λ_F in the space R_t^F spanned by the FMPs. Moreover, we show (Theorem 3.7) that δ_F is proportional to the difference in squared Sharpe ratios, $\lambda_*^2 - \lambda_F^2$. In other words, $\delta_F(r_t^*)$ measures the distance between the efficient frontiers spanned by the base assets and by the FMPs, respectively.

As a consequence of (i) and (ii), it follows that a given factor model is a true asset pricing model if and only if it is possible to construct a dynamic portfolio of the FMPs that is unconditionally mean-variance efficient in the asset return space. Thus, our condition is an extension of the Gibbons, Ross, and Shanken (1989) test to the case with conditioning information. In fact, the resulting test statistic is similar to a standard Wald test. In the following sections, we derive explicit characterizations of the measure δ_F and the return that attains it, in terms of the conditional moments of the base asset returns and the factors. This allows us to implement our test for a variety of factor models considered in the literature.

3.1 Factor-Mimicking Portfolios

We now give an explicit characterization of the factor-mimicking portfolios as ‘managed’ portfolios of the base assets. We define the conditional moments,

$$\mu_{t-1} = E_{t-1}(\tilde{R}_t - r_f e), \quad \text{and} \quad \Lambda_{t-1} = E_{t-1}((\tilde{R}_t - r_f e)(\tilde{R}_t - r_f e)') \quad (8)$$

In other words, excess returns can be written as $\tilde{R}_t - r_f e = \mu_{t-1} + \varepsilon_t$, where ε_t has zero mean and variance-covariance matrix $\Sigma_{t-1} = \Lambda_{t-1} - \mu_{t-1}\mu_{t-1}'$. Similarly, we denote the mixed

conditional moments of the factors by

$$\nu_{t-1} = E_{t-1}(\tilde{F}_t), \quad \text{and} \quad Q_{t-1} = E_{t-1}((\tilde{R}_t - r_f e)\tilde{F}_t') \quad (9)$$

Note that, if an admissible SDF of the form (4) exists, this implies,

$$0 \equiv E_{t-1}((\tilde{R}_t - r_f e)m_t) = a_{t-1}\mu_{t-1} + Q_{t-1}b_{t-1}.$$

Conversely, if a_{t-1} and b_{t-1} exist so that $a_{t-1}\mu_{t-1} + Q_{t-1}b_{t-1} = 0$, then m_t in (4) prices all *excess* returns correctly and can hence be modified to be an admissible SDF. In other words, the model admits a conditional factor structure if and only if the image of the conditional linear operator Q_{t-1} contains μ_{t-1} . This fact will be key to the proof of the equivalence result (Theorem 3.8).

Proposition 3.2 *For a given factor F_t^i , the factor-mimicking portfolio can be written as,*

$$f_t^i = r_f + (\tilde{R}_t - r_f e)' \theta_{t-1}^i \quad \text{with} \quad \theta_{t-1}^i = \Lambda_{t-1}^{-1}(q_{t-1}^i - \kappa_i \mu_{t-1}) \quad (10)$$

where q_{t-1}^i is the column of Q_{t-1} corresponding to factor i , and κ_i is a constant.

Note that the constant κ_i in the above expression is directly related to the unconditional mean of the FMP. In the case where a risk-free asset is present, this constant is not uniquely determined, since the first-order condition arising from maximizing the correlation in (5) is independent of that mean.

PROOF OF PROPOSITION 3.2: Appendix A.1. □

To conclude this section, we derive expressions for the first and second moments of the factor-mimicking portfolios, which we will need for the explicit characterization of the maximum Sharpe ratio spanned by the factors (Corollary 3.5).

Corollary 3.3 *The conditional moments of the factor mimicking portfolios are given by,*

$$E_{t-1}(f_t - r_f e) = Y_{t-1}' \Lambda_{t-1}^{-1} \mu_{t-1}, \quad \text{and} \quad E_{t-1}((f_t - r_f e)(f_t - r_f e)') = Y_{t-1}' \Lambda_{t-1}^{-1} Y_{t-1},$$

where $y_{t-1}^i = q_{t-1}^i - \kappa_i \mu_{t-1}$, and Y_{t-1} is the matrix whose columns are the y_{t-1}^i .

PROOF: Follows directly from Proposition 3.2. □

3.2 Maximum Sharpe Ratios

In this section, we derive explicit expressions for the maximum generalized Sharpe ratios, in the spaces of augmented pay-offs spanned by the base assets and the factors, respectively. Denote by λ_* the maximum Sharpe ratio in the generalized return space R_t^G ,

$$\lambda_* = \sup_{r_t \in R_t^G} \frac{E(r_t) - r_f}{\sigma(r_t)}. \quad (11)$$

Similarly, denote by λ_F the corresponding maximum Sharpe ratio in the space R_t^F of managed returns spanned by the factors.

Proposition 3.4 *The maximum generalized Sharpe ratio in the space R_t^G is given by $\lambda_* = h$,*

$$\text{where } h^2 = E(H_{t-1}^2), \quad \text{and } H_{t-1}^2 = \mu'_{t-1} \Sigma_{t-1}^{-1} \mu_{t-1}, \quad (12)$$

PROOF: Appendix A.2. □

Expression (12) for the maximum Sharpe ratio has many interesting features; first, it extends the expression given in Equation (16) of Jagannathan (1996) to the case with conditioning information. It is well-known (Cochrane 2001) that in the fixed-weight case without conditioning information the maximum (squared) Sharpe ratio is given by an expression of the form (12), with conditional moments replaced by unconditional ones. In other words, H_{t-1} represents the maximum *conditional* Sharpe ratio, once the realization of the conditioning information is known. Hence, the maximum squared *unconditional* Sharpe ratio is simply given by the expectation of the maximum squared conditional Sharpe ratio. For the case of only one risky asset, this result was also shown in Cochrane (1999).

Corollary 3.5 *The maximum generalized Sharpe ratio in the space R_t^F is given by $\lambda_F = h_F$,*

$$\text{where } h_F^2 = E(H_{F,t-1}^2),$$

$$\text{and } H_{F,t-1}^2 = \mu'_{t-1} \Lambda_{t-1}^{-1} Y_{t-1} [Y'_{t-1} \Lambda_{t-1}^{-1} \Sigma_{t-1} \Lambda_{t-1}^{-1} Y_{t-1}]^{-1} Y'_{t-1} \Lambda_{t-1}^{-1} \mu_{t-1}, \quad (13)$$

PROOF: Using the conditional moments from Corollary 3.3, we obtain,

$$\begin{aligned} \Sigma_{t-1}^F &= Y'_{t-1} \Lambda_{t-1}^{-1} Y_{t-1} + Y'_{t-1} \Lambda_{t-1}^{-1} \mu_{t-1} \mu'_{t-1} \Lambda_{t-1}^{-1} Y_{t-1} = Y'_{t-1} \Lambda_{t-1}^{-1} \underbrace{[\Lambda_{t-1} + \mu_{t-1} \mu'_{t-1}]}_{= \Sigma_{t-1}} \Lambda_{t-1}^{-1} Y_{t-1}. \end{aligned} \quad (14)$$

The result then follows from Proposition 3.4, applied to the factor-mimicking portfolios as base assets. \square

Finally, we characterize the weights on the mimicking portfolios of the portfolio that attains the maximum Sharpe ratio in (13). These weights are in fact proportional to the factor loadings in the optimal conditional factor model for given choice of factors.

Proposition 3.6 *The maximum generalized Sharpe ratio in (13) is attained by,*

$$r_t^F = \phi_{t-1}^0 r_f + (f_t - r_f e)' \phi_{t-1} \quad \text{with} \quad \phi_{t-1}^0 = \frac{1 + H_{F,t-1}^2}{1 + h_F^2} \quad (15)$$

$$\text{and } \phi_{t-1} = -\frac{r_f}{1 + h_F^2} [Y'_{t-1} \Lambda_{t-1}^{-1} \Sigma_{t-1} \Lambda_{t-1}^{-1} Y_{t-1}]^{-1} Y'_{t-1} \Lambda_{t-1}^{-1} \mu_{t-1}$$

PROOF: We apply Lemma A.1, using the factor-mimicking portfolios as base assets, and substituting the conditional moments from Corollary 3.3 as in the proof of Corollary 3.5.

\square

3.3 Necessary and Sufficient Condition

We are now in a position to prove the main results of our paper. First, we show that the measure of model misspecification defined in (7) is proportional to the difference in squared Sharpe ratios in the return spaces spanned by the base assets and the factors, respectively. This related our measure to the respective shapes of the efficient portfolio frontiers in the two spaces. Second, we show that for a given set of factors, the model admits a conditional

factor structure if the only if the misspecification error is zero. This establishes a necessary *and* sufficient condition for the factors to constitute a viable conditional asset pricing model.

Theorem 3.7 *The measure δ_F of misspecification error can be written as,*

$$\delta_F(r_t^*) = \left(\frac{r_f}{1 + \lambda_*^2} \right)^2 \cdot (\lambda_*^2 - \lambda_F^2)$$

PROOF: Appendix A.3. □

As a consequence, testing the hypothesis that $\delta_F(r_t^*) = 0$ is similar to a standard Wald test, as shown in Abhyankar, Basu, and Stremme (2005).

Theorem 3.8 *The model (X_t, Π_{t-1}) admits a conditional factor structure if and only if*

$$\delta_F(r_t^*) = 0.$$

PROOF: Appendix A.4. □

As a consequence of the above result, the difference $\lambda_*^2 - \lambda_F^2$ can be interpreted as a test of whether it is possible to construct a conditional linear asset pricing model from a given set of factors. Since by construction $R_t^F \subseteq R_t^G$, we always have $\lambda_F \leq \lambda_*$, with equality if and only if there exists a portfolio $r_t \in R_t^F$ that is unconditionally efficient in the space R_t^G .

Corollary 3.9 *The model (X_t, Π_{t-1}) admits a conditional factor structure if and only if there exist \mathcal{G}_{t-1} -measurable functions ϕ_{t-1}^i with $E(\phi_{t-1}^0) = 1$, such that*

$$\phi_{t-1}^0 r_f + \sum_{i=1}^m \phi_{t-1}^i (f_t^i - r_f) \tag{16}$$

is unconditionally efficient in the space R_t^G of generalized returns.

PROOF: Follows directly from Theorem 3.8. □

4 Empirical Analysis

In this section, we report the results of implementing our test for a variety of asset pricing models, and three sets of base assets (portfolios sorted on industry, size and book-to-market ratio, and momentum).

4.1 Data and Methodology

We specialize the set-up of the preceding sections to the case of a linear predictive model. Let y_{t-1} be a vector of \mathcal{F}_{t-1} -measurable conditioning variables, and set $\mathcal{G}_{t-1} = \sigma(y_{t-1})$. For the estimation, we will use the de-meaned variables $y_{t-1}^0 = y_{t-1} - E(y_{t-1})$. To estimate the conditional moments, we postulate a linear specification of the form,

$$\begin{pmatrix} \tilde{R}_t - r_{fe} \\ \tilde{F}_t \end{pmatrix} = \begin{pmatrix} \mu_0 \\ \bar{\nu} \end{pmatrix} + \begin{pmatrix} \beta \\ \gamma \end{pmatrix} \cdot y_{t-1}^0 + \begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \quad (17)$$

where ε_t and η_t are independent of y_{t-1}^0 with $E_{t-1}(\varepsilon_t) = E_{t-1}(\eta_t) = 0$. Note however that we do *not* assume the ε_t and η_t to be cross-sectionally independent, i.e. the residual variance-covariance matrix need not be diagonal. In the notation of Section 2, we can then calculate the conditional moments as,

$$\mu_{t-1} = \mu_0 + \beta y_{t-1}^0 \quad \text{and} \quad \Lambda_{t-1} = (\mu_0 + \beta y_{t-1}^0)(\mu_0 + \beta y_{t-1}^0)' + E_{t-1}(\varepsilon_t \varepsilon_t')$$

$$\nu_{t-1} = \bar{\nu} + \gamma y_{t-1}^0 \quad \text{and} \quad Q_{t-1} = (\mu_0 + \beta y_{t-1}^0)(\bar{\nu} + \gamma y_{t-1}^0)' + E_{t-1}(\varepsilon_t \eta_t')$$

The maximum generalized Sharpe ratios generated by the base assets and the factor mimicking portfolios, respectively, are then calculated using (12) and (13).

We test the performance of both the unscaled (fixed-beta) as well as the optimally scaled versions of several classic asset pricing models, including the CAPM and the three-factor model of Fama and French (1992). Each of these models is then augmented by additional factors to capture skewness, excess kurtosis, and/or momentum effects. We perform our tests

on three distinct sets of base assets. As a benchmark case, we consider a set of 30 portfolios sorted by industry sector, an asset universe that does not exhibit any of the known asset pricing anomalies. To investigate how well the models succeed in explaining well-known anomalies, we then repeat our tests on Fama and French’s 25 portfolios sorted on size and book-to-market ratio, and the 10 portfolios sorted on momentum (past returns) as used in Chordia and Shivakumar (2002). To optimally scale both assets as well as factor loadings, we employ a set of conditioning instruments designed to capture changes in the overall economic environment⁷. In our empirical analysis, we address the following questions

- (1) Can unscaled (fixed-beta) models explain the size, value and momentum effects?
- (2) Can time-varying betas improve the performance of an asset pricing model?
- (3) Does the inclusion of skewness and kurtosis factors improve model performance?
- (4) Are the Fama-French factors necessary to explain the size and value premia?
- (5) Do we need a momentum factor to explain the momentum effect?
- (6) Can any scaled model price actively managed portfolios?

While we discuss the results of our empirical analysis in detail in the following sections, we summarize our answers to the above questions in Section 4.6.

4.2 Optimal Scaling and Time-Varying Betas

As a benchmark, we first analyze the effect of optimal scaling on both asset as well as model performance, using Fama and French’s 30 industry portfolios as base assets. Table 1 reports

⁷These instruments are the short rate (TB1M), term spread (TSPR), curvature of the yield curve (CONV), credit yield spread (CSPR), and unexpected shocks to inflation (INFL).

the slope of the efficient frontier spanned by assets and factors, respectively, both with and without the optimal use of conditioning information, for a variety of classic asset pricing models.

THE EFFECT OF OPTIMAL SCALING

From row (1) we see that the optimal use of the conditioning instruments dramatically expands the efficient frontier (with Sharpe ratios increasing by a factor of more than 3/2 from 0.88 to 2.36). This considerably raises the bar if we require a model to not only price the base assets but also actively managed portfolios correctly.

On the other hand, optimal scaling improves the performance of the CAPM (row 1 in the table) and Fama-French model (row 6) just as dramatically (with the factor frontiers expanding by 48% and 163%, respectively). In fact, while the unscaled version still has a pricing error of about 40%, the scaled Fama-French model successfully prices static portfolios (with the factor frontier expanding beyond the fixed-weight asset frontier). However both models, even optimally scaled, still fall short by a wide margin of pricing managed portfolios correctly (with pricing errors of 71% and 60%, respectively). Interestingly, the consumption-CAPM shows the most significant boost in performance (with the factor frontier expanding by orders of magnitude from 0.09 to 0.77).

SKEWNESS, KURTOSIS AND MOMENTUM FACTORS

Rows (2) and (7) of Table 1 show that the skewness factor, while having virtually no effect on the performance of the unscaled models, significantly improves the performance of the optimally scaled models. The addition of the skewness factor increases the factor frontier by less than 1% for the unscaled CAPM and Fama-French model, but by 76% and 45% for the scaled versions, respectively. In fact, augmented by the skewness factor, the CAPM now prices static portfolios correctly, and the pricing error for managed portfolios falls below 50%. Similarly, the pricing error for the Fama-French model drops to about 40% for active portfolios. In other words, while skewness has little importance for asset pricing at the level of long-run average returns, it has a dramatic effect on conditional pricing. Intuitively, this

is due to the fact that any skewness in the conditional return distribution is ‘washed out’ by time aggregation in the unconditional distribution (Central Limit Theorem). Conversely, the inclusion of the skewness factor considerably amplifies the effect of optimal scaling. For example, while conditioning information improves the performance of the CAPM by only 48%, once the skewness factor is added, this goes up to 158%.

In contrast, rows (3) and (8) of Table 1 show that kurtosis has a much less dramatic effect, similar in size for both unscaled and scaled models (with an additional expansion of the factor frontier by between 6% and 11%).

TIME-VARYING BETAS

In most of our empirical tests, we find that the optimal factor loadings exhibit a considerable degree of non-linearity in the conditioning instrument, in contrast to the linear scaling prevalent in most of the literature. For example, Figure 2 shows the optimal factor loadings, as functions of the conditioning instruments, for the CAPM augmented by a skewness and kurtosis factor. Interestingly, when we estimated the CAPM without the skewness and kurtosis factors, the non-linearity in the market factor all but disappears. This seems to indicate that it is not only factor risk that matters, but also the *correlations* between the factors.

In order to compare our optimal scaling with the linear specification prevalent in most of the literature, we also estimate the models using linearly scaled factor loadings. We find that the linear models considerably under-perform the optimally scaled versions. For example, for a standard CAPM, the factor Sharpe ratio of the optimally scaled model is 62% higher than that generated by the linearly scaled factor. In fact, the linearly scaled model only marginally out-performs the corresponding model with constant betas. This provides further support for the necessity of non-linear factor risk premia and the sub-optimality of linear scaling, see also Ghysels (1998), and Brandt and Chapman (2005). It should also be noted that, in models with multiple factors and multiple conditioning instruments, linearly scaled models are at risk of ‘over-fitting’ in-sample. For example, a linearly scaled model with m factors

and p instruments has $m \times (p + 1)$ free parameters ('betas') to fit the data.

4.3 The Size and Book-to-Market Effect

Asset pricing research has documented that the cross-sectional dispersion in expected returns is determined not only by beta, as prescribed by the CAPM, but by firm size, book-to-market ratio, and other factors. The CAPM has demonstrated virtually no power to explain the cross section of average returns on assets sorted by size and book-to-market equity ratios (Fama and French 1992).

To capture this effect, we now repeat our tests using Fama and French's 5×5 portfolios sorted on size and book-to-market ratio. The results are reported in Table 2. In our analysis, the size and value effect (Fama and French 1992) manifests itself in the dramatic increase in asset performance (with the fixed-weight asset Sharpe ratio almost doubling from 0.88 to 1.57). Interestingly, the increase in performance for optimally managed portfolios is much less dramatic (increasing from 2.36 to 2.66). This seems to suggest that the superior returns due to the size and value effect are largely washed out by the increase in portfolio performance due to active management. Conversely, because efficient passive portfolios constructed on the basis of size and book-to-market ratio already exhibit much higher returns, the marginal benefit of active management is much smaller in this asset universe (with an increase of the efficient frontier by 69% as compared to 168% for the industry portfolios).

The unscaled CAPM and Fama-French models are unable to price these portfolios (with pricing errors of 74% and 45%, respectively). In other words, over the time period we consider, even the Fama-French model fails to price the size and value portfolios. Moreover, even with optimal scaling, both models are unable to price the assets (with pricing errors of 68% and 28%, respectively). In contrast, when we augment these models by skewness and kurtosis factors, we find that an optimally scaled three-factor CAPM achieves 90% of the fixed-weight asset Sharpe ratio, while the scaled augmented Fama-French model actually prices the assets correctly. However, none of the scaled models comes close to pricing

managed portfolios (with pricing errors still as large as 48% and 40%, respectively). Our results thus show that a scaled three factor CAPM goes a long way towards explaining the size and book-to-market effects.

4.4 The Momentum Effect

We next examine the performance of the scaled models on the momentum portfolios. The momentum portfolios stocks sorted on the basis of prior returns and the pricing of these portfolios has posed a great challenge to standard asset pricing model. Indeed momentum remains the only CAPM-based anomaly that cannot be explained by the static Fama-French model. Several partial risk-based explanations have been proposed. Grundy and Martin (2001) have suggested that the factor loadings should be time-varying to account for the dynamic nature of the momentum strategy, but find that momentum profits persist even when time-varying factor loadings are incorporated. Harvey and Siddique (2000) find that the pricing errors from the Fama-French model are correlated with their skewness factor suggesting that this factor may have significant explanatory power for these portfolios. Recently, Hung (2005) found that stylized momentum effects are partly driven by exposures to co-skewness and co-kurtosis factors. These results indicate that the option like elements of the momentum returns may be captured by nonlinear factors such as skewness and kurtosis. Chordia and Shivakumar (2002) find that profits to momentum strategies can be explained by lagged values of standard macro-economic instruments and that momentum profits disappear once returns are adjusted for their predictability based on their business cycle variables. Their results suggest that time-varying expected returns could be an explanation for momentum profits.

All of the above studies suggest that a three-factor CAPM that is the market return together with skewness and kurtosis factors, optimally scaled by macro-economic variables, might outperform static models in pricing the momentum portfolios. We use the momentum portfolios constructed by Chordia and Shivakumar (2002), covering the period from

January 1960 until January 1999⁸.

We first examine the performance of the static CAPM and Fama-French model on these portfolios. The fixed-weight Sharpe ratio of the momentum portfolios is 1.15 while the static CAPM achieves a Sharpe ratio of 0.64, just over 50% of the momentum portfolios showing that it does a poor job of pricing these. Adding the Fama-French factors improves the performance considerably (with the unscaled factor Sharpe ratio increasing to 0.93), reducing pricing error to about 18%. This shows that although the Fama-French model does much better than the CAPM it still does not completely explain the momentum portfolio returns.

We now examine the performance of the three-factor CAPM. Adding the skewness and kurtosis factors to the unscaled CAPM makes very little difference (with the Sharpe ratio rising only to 0.65). This is in line with our earlier empirical findings where the inclusion of skewness and kurtosis in the unscaled model has very little impact. However, when we incorporate optimal scaling, the Sharpe ratio of the scaled CAPM rises to 0.87, just below that of the static Fama-French model. However the performance of the three-factor CAPM improves dramatically when we scale the factors (with the factor Sharpe ratio rising to 1.23). Thus the scaled three-factor CAPM actually prices the momentum portfolios, which the unscaled Fama-French model is unable to do. Our results thus provide evidence of a rational risk-based explanation for momentum returns, which is in line with earlier studies.

However, the scaled three-factor CAPM still leads to considerable pricing errors for optimally managed portfolios (achieving about 67% of the asset Sharpe ratio). Adding the Fama-French factors, while improving the performance of the model, still does not make it a true asset pricing model (with a minimum achievable pricing error of about 15%).

Nonetheless our model appears to be the first rational asset pricing model that successfully

⁸See Chordia and Shivakumar (2002) for details of the construction of these portfolios. We thank Doron Avramov for making this data available.

prices the momentum portfolios. The time-varying and option-like features of momentum returns appear to be captured mostly by the skewness factor with time-varying factor loadings⁹. We find as in our earlier results that incorporating time-varying factor loadings is absolutely crucial, and that our method of optimal scaling significantly improves the performance of the model.

4.5 Robustness

Our tests are all based on in-sample Sharpe ratios which are known to exhibit high sampling variability particularly in the presence of conditioning information (Ferson and Siegel 2003). This raises the concern that our results may be driven in part by this sampling variability. To this end we conduct the following robustness check. We construct a pure noise variable by simulating from a standard normal random variable and use this as our predictive instrument. We then compute the optimal asset and factor Sharpe ratios for the various sets of assets and factors. The result that is of greatest concern to us is the pricing of the fixed weight assets by the scaled factor models and we want to see if the optimal factor Sharpe ratio using the noise variable is ever as high as the fixed-weight factor Sharpe ratio. We simulate 10,000 values of the noise predictor for all the sets of base assets and in all cases while the optimal factor Sharpe ratio is higher than the fixed weight factor Sharpe ratio, the fixed weight asset Sharpe ratio is above the simulated 99% confidence interval indicating that our results are robust to sampling variability. It is also the case that the optimal factor Sharpe ratio using the true predictive instruments is above the simulated 99% confidence interval. Our conclusion is thus that our findings are robust to sampling variability in the estimation of Sharpe ratio.

⁹Our results are also consistent with Chen, Hong, and Stain (2001), who find that stocks with prior positive returns tend to exhibit high negative skewness.

4.6 Summary

To answer the questions set out in the beginning of this section:

- (1) *Can unscaled (fixed-beta) models explain the size, value and momentum effects?*

No. Although the unscaled CAPM and Fama-French models, augmented by the skewness factor, correctly price industry portfolios, none of the unscaled models tested comes close to pricing size or book-to-market portfolios correctly. While the inclusion of additional factors does improve the performance of the unscaled models, pricing errors are still considerable.

- (1) *Can time-varying betas improve the performance of an asset pricing model?*

The optimal use of conditioning information dramatically improves the performance of all models considered here, although the magnitude of the improvement depends on the choice of base assets. Generally, the marginal benefit of optimal scaling is smaller in the presence of asset pricing anomalies. Scaling affects the Fama-French model more than it does the CAPM. While time-variation in factor loadings reduces pricing errors for all factors, the improvement appears most significant for the skewness factor.

- (3) *Does the inclusion of skewness and kurtosis factors improve model performance?*

For unscaled (fixed-beta) models, the skewness factor has virtually no effect, whatever the base assets. In contrast, the kurtosis factor adds marginally to the performance of the unscaled CAPM and Fama-French model on the size and book-to-market portfolios. The situation is reversed when the factor betas are time-varying. In this case, the skewness factor reduces pricing errors dramatically for both models, while kurtosis has only marginal effect.

- (4) *Are the Fama-French factors necessary to explain the size and value premia?*

No. Surprisingly, an optimally scaled three-factor CAPM, augmented by skewness and kurtosis factors, achieves pricing errors of less than 10% for the size and book-to-

market portfolios, thus in fact clearly out-performing the unscaled Fama-French model (without the skewness and kurtosis factors).

(5) *Do we need a momentum factor to explain the momentum effect?*

No. While the momentum factor adds marginally to the performance of unscaled models, as soon as betas are allowed to be time-varying, skewness and kurtosis clearly dominate the momentum factor. In particular, an optimally scaled three-factor CAPM, augmented by skewness and kurtosis, perfectly prices the momentum portfolios.

(6) *Can any scaled model price actively managed portfolios?*

Although none of the scaled models succeeds in pricing actively managed portfolios correctly, a scaled Fama-French model, augmented by skewness and kurtosis factors, has pricing errors of less than 40%. Interestingly, the model performs much better on those portfolios that exhibit abnormal returns (i.e. the size, book-to-market and momentum portfolios), with pricing errors as low as 15%.

5 Conclusion

In this paper, we develop a new measure of specification error for conditional factor models, leading to new tests for such models. Our test measures how close the frontier spanned by the factors (or factor-mimicking portfolios) is to the unconditionally efficient frontier of the base assets. We can thus show that a conditional factor model is a true asset pricing model if and only if our measure of specification error is zero. Moreover, we show that our test is proportional to the difference in squared Sharpe ratios of the asset and factor frontiers, respectively. Finally, for a given set of factors, our analysis allows us to construct a portfolio of the factor-mimicking portfolios that minimizes pricing error.

We draw three main conclusions from our empirical findings. The first is that there is significant time-variation in factor loadings and that this time-variation plays a significant

role in asset pricing. The second is that our optimal scaling clearly improves the performance of asset pricing models to the point where several of the scaled models are capable of pricing the base assets although they are still not true conditional asset pricing models. The third is that the skewness factor matters for pricing in a conditional setting, that is when the factor loadings are allowed to be time-varying.

We find that a three factor CAPM scaled by business cycle variables is capable of pricing the momentum portfolios. Our results are robust to the high sampling variability observed in Sharpe ratios.

A Mathematical Appendix

A.1 Proof of Proposition 3.2:

Throughout the proof, we omit the time subscript to simplify notation. Denote by f^i the factor-mimicking portfolio for factor F^i .

STEP 1: Let $r \in R$ be an arbitrary return. For $\varepsilon \in \mathbb{R}$, define $r(\varepsilon) = (1 - \varepsilon)f^i + \varepsilon r$. As f^i has maximal correlation with F^i , the first-order condition implies,

$$\begin{aligned} 0 &= \left. \frac{d}{d\varepsilon} \right|_{\varepsilon=0} \rho^2(r(\varepsilon), F^i) \\ &= \text{cov}(r - f^i, F^i) \text{cov}(f^i, f^i) - \text{cov}(f^i, F^i) \text{cov}(f^i, r - f^i) \\ &= \text{cov}(r, \text{var}(f^i)) F^i - \text{cov}(f^i, F^i) f^i. \end{aligned}$$

In other words, $\beta_i f^i$ is the orthogonal projection of F^i onto the space R of managed returns, where $\beta_i = \text{cov}(F^i, f^i) / \text{var}(f^i)$.

STEP 2: Write $f^i = r_f + (\tilde{R} - r_f e)' \theta^i$. For arbitrary ϕ , the first-order condition implies,

$$\begin{aligned} 0 &= \text{cov}\left(F^i - \beta_i(\tilde{R} - r_f e)' \theta^i, (\tilde{R} - r_f e)' \phi\right) \\ &= E\left([F^i - \beta_i(\tilde{R} - r_f e)' \theta^i](\tilde{R} - r_f e)' \phi\right) \\ &\quad - \underbrace{E\left(F^i - \beta_i(\tilde{R} - r_f e)' \theta^i\right)}_{=: \kappa_i} E\left((\tilde{R} - r_f e)' \phi\right) \\ &= E\left([q^i - \beta_i \Lambda \theta^i - \kappa_i \mu]' \phi\right) \end{aligned}$$

As this must hold for all ϕ , we find,

$$\theta^i = \frac{1}{\beta_i} \Lambda^{-1} (q^i - \kappa_i \mu)$$

Finally, as correlation is homogenous of degree 0, we can set $\beta_i = 1$, which determines the value of the constant κ_i .

This completes the proof of Proposition 3.2. □

A.2 Proof of Proposition 3.4:

We use Lemma 3.3 in Hansen and Richard (1987), which shows that the unconditionally efficient generalized return with mean m can be written as $r^* + w z^*$ for some $w \in \mathbb{R}$, where $r^* \in R$ is the unique generalized return with minimal second moment, and z^* is the Riesz representation of the expectation functional on the space Z of excess returns. From this, it follows that the maximum (squared) unconditional Sharpe ratio is given by,

$$\lambda_*^2 = \frac{r_f - E(r^*)}{E(r^*)}. \quad (18)$$

From Lemma A.1 below we obtain the weights of r^* . A straight-forward calculation then yields $E(r^*) = r_f / (1 + h^2)$, which yields the desired result. \square

Lemma A.1 *The generalized return r_t^* with minimum unconditional second moment is,*

$$r_t^* = \theta_{t-1}^0 r_f + (\tilde{R}_t - r_f e)' \theta_{t-1}$$

with $\theta_{t-1} = -\frac{r_f}{1+h^2} \Sigma_{t-1}^{-1} \mu_{t-1}$ *and* $\theta_{t-1}^0 = \frac{1+H_{t-1}^2}{1+h^2}$

PROOF: Throughout the proof, we will omit the time subscript to simplify notation. We use calculus of variation. Let $\theta^0(\varepsilon)$ and $\theta(\varepsilon)$ be differentiable one-parameter families of admissible weights, and define

$$r(\varepsilon) = \theta^0(\varepsilon) r_f + (\tilde{R} - r_f e)' \theta(\varepsilon).$$

We assume $E(\theta^0(\varepsilon)) = 1$ for all ε , so that the $r(\varepsilon)$ are indeed generalized returns for all ε . Suppose that the minimum second moment is attained at $\varepsilon = 0$, then we must have,

$$\left. \frac{d}{d\varepsilon} \right|_{\varepsilon=0} E(r(\varepsilon)^2) = 0 \quad (19)$$

We will use this condition to determine the solutions θ^0 and θ separately.

STEP 1: CHARACTERIZING θ^0

Let ϕ^0 be an arbitrary weight function, and set $a = 1/E(\phi^0)$. Define

$$\theta^0(\varepsilon) = (1 - \varepsilon)\theta^0 + \varepsilon a\phi^0, \quad \theta(\varepsilon) \equiv \theta$$

By construction, $E(\theta^0(\varepsilon)) = 1$ for all ε , so that this specification indeed defines a family of generalized returns. The first-order condition (19) in this case becomes,

$$0 = 2r_f^2 E(\theta^0(a\phi^0 - \theta^0)) + 2r_f E((a\phi^0 - \theta^0)\mu'\theta)$$

Multiplying this equation by $E(\phi^0) = 1/a$, we obtain,

$$E([r_f\theta^0 + \mu'\theta]\phi^0) = \gamma E(\phi^0)$$

for some unconditional constant γ . Since this equation must hold for all ϕ^0 , we obtain

$$r_f\theta^0 + \mu'\theta = \gamma \tag{20}$$

Finally, since $E(\theta^0) = 1$, taking unconditional expectations implies,

$$\gamma = r_f + E(\mu'\theta) \tag{21}$$

STEP 2: CHARACTERIZING θ

Let ϕ be an arbitrary vector of weights for the risky assets, and set

$$\theta^0(\varepsilon) \equiv \theta^0, \quad \theta(\varepsilon) = \theta + \varepsilon\phi$$

Since $E(\theta^0(\varepsilon)) = 1$, this specification generates a family of generalized returns. Hence, we can apply the first-order condition (19) to obtain,

$$0 = 2r_f E(\theta^0\mu'\phi) + 2E(\theta'\Lambda\phi)$$

Since this equation must hold for all ϕ , we obtain,

$$r_f\theta^0\mu = -\Lambda\theta = -[\Sigma + \mu\mu']\theta$$

Substituting from (20) into the left-hand side of this equation, we can write,

$$\theta = -\gamma\Sigma^{-1}\mu,$$

Multiplying this by μ and taking unconditional expectations, we obtain,

$$E(\mu'\theta) = -\gamma E(\mu'\Sigma^{-1}\mu) = -\gamma E(H^2) = -\gamma h^2 \quad (22)$$

Finally, equating (21) and (22),

$$-\gamma h^2 = E(\mu'\theta) = \gamma - r_f, \quad \text{which implies } \gamma = \frac{r_f}{1+h^2}$$

Hence, the weights for the risky assets are given by,

$$\theta = -\gamma \Sigma^{-1}\mu = -\frac{r_f}{1+h^2} \Sigma^{-1}\mu$$

Finally, using (20), we obtain the weight for the risk-free asset,

$$r_f \theta^0 = \gamma - \mu'\theta = \gamma + \frac{r_f H^2}{1+h^2}, \quad \text{which implies } \theta^0 = \frac{1+H^2}{1+h^2}$$

This completes the proof of Lemma A.1. □

A.3 Proof of Theorem 3.7

Throughout the proof, we omit the time subscript to simplify notation. From Proposition 3.2, we know that any portfolio that is constructed from the factor-mimicking portfolios can be written in the form

$$r(\phi) = \phi^0 r_f + (\tilde{R} - r_f e)'\Lambda^{-1}Y\phi$$

for arbitrary ϕ^0 and ϕ . It is then straight-forward to show that

$$\begin{aligned} \text{var}(r^* - r(\phi)) &= \text{const} + r_f^2 E((\phi^0)^2) - 2r_f \frac{h^2}{1+h^2} E(b'\theta) - E(b'\theta)^2 \\ &\quad + E(\phi'[M + bb']\phi) + 2r_f E(\phi^0 b'\phi), \end{aligned}$$

where $b = \mu'\Lambda^{-1}Y$ and $M = Y'\Lambda^{-1}\Sigma\Lambda^{-1}Y$. Now suppose ϕ minimizes the variance in the above expression, and hence attains δ^2 .

STEP 1: CHARACTERIZING ϕ^0

Let θ^0 be arbitrary, set $a = 1/E(\theta^0)$ and define

$$\phi^0(\varepsilon) = (1 - \varepsilon)\phi^0 + \varepsilon a\theta^0$$

As the minimum variance is attained at $\varepsilon = 0$, the first-order condition implies,

$$\begin{aligned} 0 &= \left. \frac{d}{d\varepsilon} \right|_{\varepsilon=0} \text{var}(r^* - r(\phi(\varepsilon))) \\ &= 2r_f^2 E(\phi^0[a\theta^0 - \phi^0]) + 2r_f E([a\theta^0 - \phi^0]b'\phi) \end{aligned}$$

From this follows,

$$E([r_f\phi^0 + b'\phi]\theta^0) = \underbrace{E(r_f(\phi^0)^2 + \phi^0 b'\phi)}_{=: \gamma} E(\theta^0)$$

As this must hold for all θ^0 , we get

$$r_f\phi^0 + b'\theta = \gamma. \tag{23}$$

STEP 2: CHARACTERIZING ϕ

Let θ be arbitrary, and set

$$\phi(\varepsilon) = \phi + \varepsilon\theta$$

As the minimum variance is attained at $\varepsilon = 0$, the first-order condition is,

$$\begin{aligned} 0 &= -2r_f \frac{h^2}{1+h^2} E(b'\theta) - 2E(b'\phi) E(b'\theta) \\ &\quad + 2E(\phi'[M + bb']\theta) + 2r_f E(\theta^0 b'\theta) \\ &\quad - \left[r_f \frac{h^2}{1+h^2} + \underbrace{E(b'\phi)}_{\text{from Step 1}} \right] E(b'\theta) \\ &= \gamma - r_f \text{ from Step 1} \end{aligned}$$

A.4 Proof of Theorem 3.8

We omit the proof of this theorem here due to its length. Details of the proof can be obtained from the authors upon request.

References

- ABHYANKAR, A., D. BASU, and A. STREMME (2005): “The Optimal Use of Asset Return Predictability: An Empirical Analysis,” working paper, Warwick Business School.
- AVRAMOV, D., and T. CHORDIA (2004): “Asset Pricing Models and Financial Market Anomalies,” working paper, University of Maryland.
- BALL, R., and S. KOTHARI (1989): “Nonstationarity in Expected Returns: Implications for Tests of Market Efficiency and Serial Correlations in Returns,” *Journal of Financial Economics*, 25, 51–74.
- BERK, J., R. GREEN, and V. NAIK (1999): “Optimal Investment, Growth Options, and Security Returns,” *Journal of Finance*, 54, 1553–1607.
- BRANDT, M., and D. CHAPMAN (2005): “Linear Approximations and Tests of Conditional Factor Models,” working paper, Boston College.
- CAMPBELL, J. (1987): “Stock Returns and the Term Structure,” *Journal of Financial Economics*, 18, 373–399.
- CHAN, K., and N. CHEN (1988): “An Unconditional Asset Pricing Test and the Role of Firm Size as an Instrumental Variable for Risk,” *Journal of Finance*, 43, 309–325.
- CHEN, J., H. HONG, and J. STAIN (2001): “Trading Volume, Past Returns and Conditional Skewness in Stock Prices,” *Journal of Financial Economics*, 61, 345–381.
- CHORDIA, T., and L. SHIVAKUMAR (2002): “Momentum, Business Cycle and Time-Varying Expected Return,” *Journal of Finance*, 57, 985–1019.
- COCHRANE, J. (1999): “Portfolio Advice for a Multi-Factor World,” *Economic Perspectives, Federal Reserve Bank of Chicago*, 23(3), 59–78.
- COCHRANE, J. (2001): *Asset Pricing*. Princeton University Press, Princeton, New Jersey.

- DITTMAR, R. (2002): “Nonlinear Pricing Kernels, Kurtosis Preference, and Evidence from the Cross Section of Equity Returns,” *Journal of Finance*, 57(1), 369–403.
- FAMA, E., and K. FRENCH (1988): “Dividend Yields and Expected Stock Returns,” *Journal of Financial Economics*, 22, 3–25.
- FAMA, E., and K. FRENCH (1992): “The Cross-Section of Expected Returns,” *Journal of Finance*, 47, 427–465.
- FERSON, W., and C. HARVEY (1998): “Fundamental Determinants of Equity Market Returns: A Perspective on Conditional Asset Pricing,” *Journal of Banking and Finance*, 21, 1625–1665.
- FERSON, W., and C. HARVEY (1999): “Conditioning Variables and the Cross-Section of Stock Returns,” *Journal of Finance*, 54, 1325–1360.
- FERSON, W., and A. SIEGEL (2003): “Stochastic Discount Factor Bounds with Conditioning Information,” *Review of Financial Studies*, 16(2), 567–595.
- FERSON, W., A. SIEGEL, and T. XU (2005): “Mimicking Portfolios with Conditioning Information,” forthcoming, *Journal of Financial and Quantitative Analysis*.
- GHYSELS, E. (1998): “On Stable Factor Structures in the Pricing of Risk: Do Time-Varying Betas Help or Hurt?,” *Journal of Finance*, 53, 549–573.
- GIBBONS, M., S. ROSS, and J. SHANKEN (1989): “A Test of Efficiency of a Given Portfolio,” *Econometrica*, 57(5), 1121–1152.
- GOMES, J., L. KOGAN, and L. ZHANG (2003): “Equilibrium Cross-Section of Returns,” *Journal of Political Economy*, 111, 693–732.
- GRUNDY, B., and S. MARTIN (2001): “Understanding the Nature of the Risks and the Sources to Rewards of Momentum Investing,” *Review of Financial Studies*, 14, 29–78.

- HANSEN, L., and R. JAGANNATHAN (1997): “Assessing Specification Errors in Stochastic Discount Factor Models,” *Journal of Finance*, 52, 557–590.
- HANSEN, L., and S. RICHARD (1987): “The Role of Conditioning Information in Deducing Testable Restrictions Implied by Dynamic Asset Pricing Models,” *Econometrica*, 55(3), 587–613.
- HARVEY, C., and A. SIDDIQUE (2000): “Conditional Skewness in Asset Pricing Tests,” *Journal of Finance*, 55(3), 1263–1295.
- HUNG, D. (2005): “Systematic Risks, Nonlinear Market Models and International Momentum and Size Strategies,” working paper, University of Durham.
- JAGANNATHAN, R. (1996): “Relation between the Slopes of the Conditional and Unconditional Mean-Standard Deviation Frontier of Asset Returns,” in *Modern Portfolio Theory and its Applications: Inquiries into Asset Valuation Problems*, ed. by S. Saito et al. Center for Academic Societies, Osaka, Japan.
- JAGANNATHAN, R., and Z. WANG (1996): “The Conditional CAPM and the Cross-Section of Expected Returns,” *Journal of Finance*, 51(1), 3–53.
- KRAUS, A., and R. LITZENBERGER (1976): “Skewness Preference and the Valuation of Risky Assets,” *Journal of Finance*, 31, 1085–1100.
- LEWELLEN, J. (1999): “The Time-Series Relations among Expected Return, Risk, and Book-to-Market,” *Journal of Financial Economics*, 54, 5–43.

Model	Factors	Sharpe Ratios			
		Asset Frontier		Factor Frontier	
		fw	opt	fw	opt
(1) CAPM	MKT	0.8795	2.3600	0.4700	0.6950
(2) ... + Skewness	MKT			0.4738	1.2225
(3) + Kurtosis	MKT			0.5212	1.3400
(4) ... + Momentum	MKT			0.4778	0.7585
(5) ... + All	MKT			0.5282	1.3886
(6) FF 3-Factor	MKT SMB HML			0.5283	0.9529
(7) ... + Skewness	MKT SMB HML			0.5283	1.3853
(8) + Kurtosis	MKT SMB HML			0.5873	1.4706
(9) ... + Momentum	MKT SMB HML			0.5331	0.9854
(10) ... + All	MKT SMB HML			0.5928	1.4978
(11) C-CAPM	MKT SMB HML			0.0922	0.7716
(12) ... + Skewness	SKEW*			0.0929	0.9039
(13) + Kurtosis	SKEW* KURT*			0.3011	1.0853

Table 1: Test Results (30 Industry Portfolios)

This table reports the maximum (annualized) Sharpe ratios generated by the base assets and the factor-mimicking portfolios, respectively. The factor models tested are the classic CAPM, the three-factor (RMF, HML and SMB) Fama-French model, and the Consumption-CAPM. Also reported are the results when the models are augmented by skewness (SKEW), kurtosis (KURT), and/or momentum (UMD) factors. The base assets are Fama and French's 30 industry portfolios, and the conditioning instruments are the short rate (TB1M), term spread (TSPR), convexity (CONV), credit yield spread (CSPR), and inflation (INFL).

Model	Factors	Sharpe Ratios			
		Asset		Factor	
		fw	opt	fw	opt
(1) CAPM	MKT	1.5712	2.6595	0.4128	0.6620
(2) ... + Skewness	MKT			0.4879	1.0797
(3) ... + Kurtosis	MKT			0.7866	1.3951
(4) ... + Momentum	MKT			0.4155	0.7367
(5) ... + All	MKT			0.7987	1.4310
(6) FF 3-Factor	MKT SMB HML			0.8595	1.1336
(7) ... + Skewness	MKT SMB HML			0.9143	1.3976
(8) ... + Kurtosis	MKT SMB HML			1.0369	1.5958
(9) ... + Momentum	MKT SMB HML			1.0523	1.3353
(10) ... + All	MKT SMB HML			1.2028	1.7184
(11) C-CAPM	MKT SMB HML			0.0561	0.7422
(12) ... + Skewness	SKEW*			0.1676	0.9563
(13) ... + Kurtosis	SKEW* KURT*			0.4855	1.0874

Table 2: Test Results (5×5 Size and Book-to-Market Portfolios)

This table reports the maximum (annualized) Sharpe ratios generated by the base assets and the factor-mimicking portfolios, respectively. The factor models tested are the classic CAPM, the three-factor (RMF, HML and SMB) Fama-French model, and the Consumption-CAPM. Also reported are the results when these models are augmented by skewness (SKEW), kurtosis (KURT), and/or momentum (UMD) factors. The base assets are Fama and French's 5×5 size and book-to-market portfolios, and the conditioning instruments are the short rate (TBIM), term spread (TSPR), convexity (CONV), credit yield spread (CSPR), and inflation (INFL).

	Model	Factors	Sharpe Ratios			
			Asset		Factor	
			fw	opt	fw	opt
(1)	CAPM	MKT	1.1478	1.8565	0.6438	0.8622
(2)	... + Skewness	MKT			0.6456	1.0641
(3) + Kurtosis	MKT			0.6577	1.2369
(4)	... + Momentum	MKT			0.7242	0.9993
(5)	... + All	MKT			0.8604	1.3696
(6)	FF 3-Factor	MKT SMB HML			0.9343	1.3040
(7)	... + Skewness	MKT SMB HML			1.0108	1.4898
(8) + Kurtosis	MKT SMB HML			1.0115	1.5714
(9)	... + Momentum	MKT SMB HML			1.0052	1.4893
(10)	... + All	MKT SMB HML			1.0250	1.6338

Table 3: Test Results (10 Momentum Portfolios)

This table reports the maximum (annualized) Sharpe ratios generated by the base assets and the factor-mimicking portfolios, respectively. The factor models tested are the classic CAPM, and the three-factor (RMF, HML and SMB) Fama-French model. Also reported are the results when these models are augmented by skewness (SKEW), kurtosis (KURT), and/or momentum (UMD) factors. The base assets are 10 portfolios sorted by momentum, and the conditioning instruments are the short rate (TB1M), term spread (TSPR), convexity (CONV), credit yield spread (CSPR), and inflation (INFL).

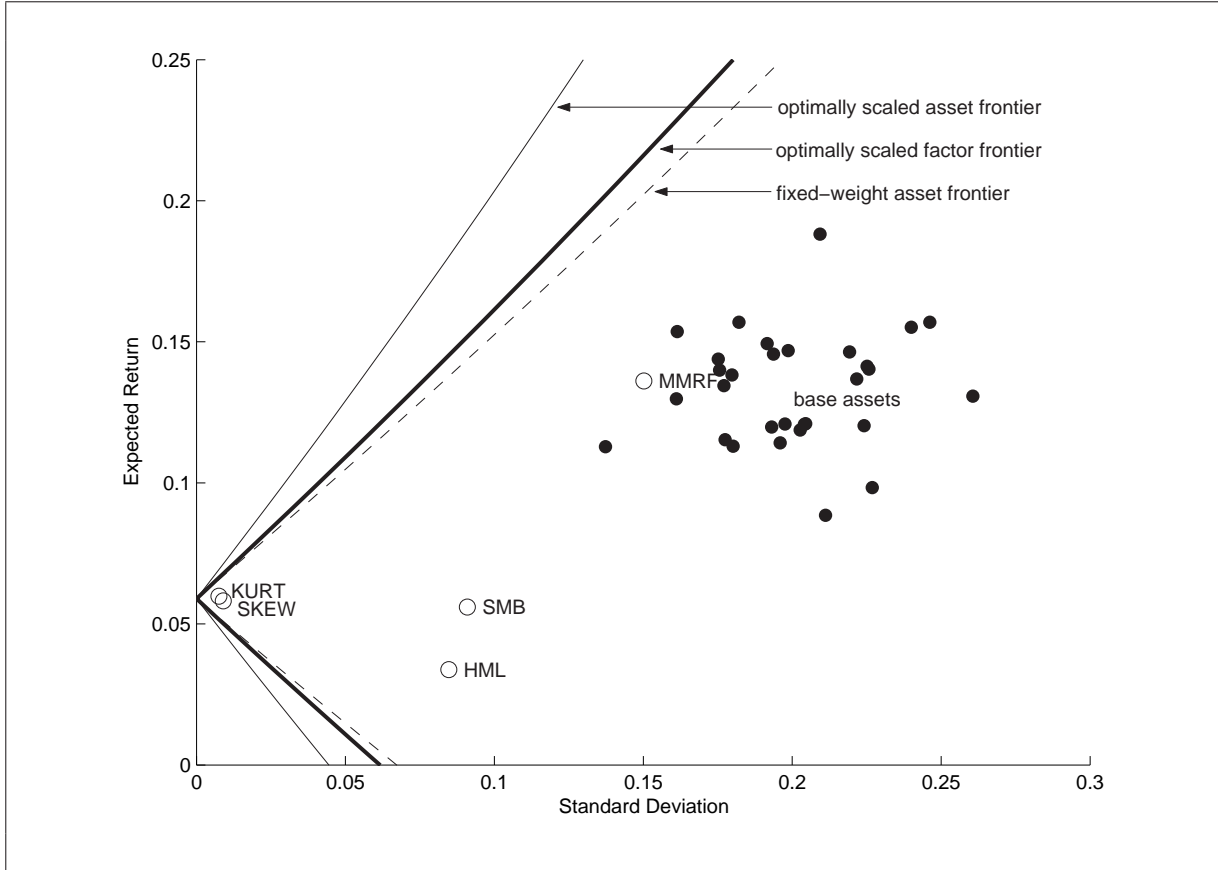


Figure 1: **Efficient Frontiers**

This figure shows the efficient frontiers generated by the base assets, both for fixed-weight (dashed line) and optimally managed (light-weight line) portfolios, and the factor-mimicking portfolios (bold-faced line), respectively. The base assets are the 30 Fama-French industry portfolios, the model is the Fama-French 3-factor model augmented by skewness and kurtosis factors, and the predictive variable is the term spread. The dots indicate the 30 base assets, and the circles depict the factor-mimicking portfolios.

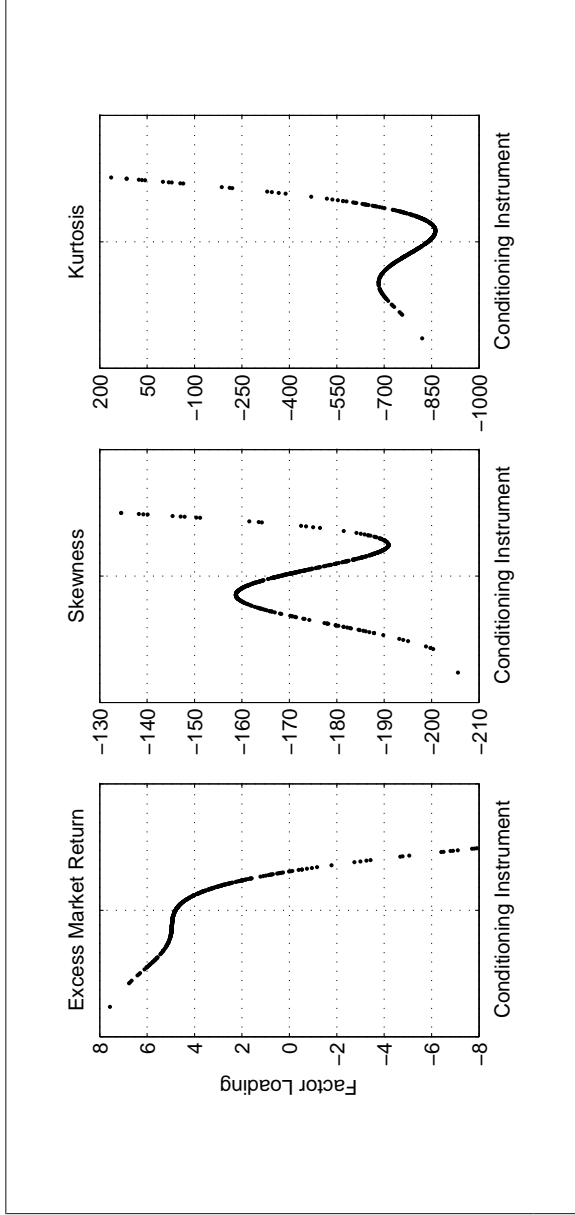


Figure 2: Optimal Factor Loadings

This figure shows the optimal factor loadings as functions of the conditioning instruments (calculated using Proposition 3.6), for the CAPM augmented by skewness and kurtosis factors. The base assets are Fama and French's 5 × 5 size and book-to-market portfolios. The instrument is a linear combination (obtained from the estimation of the predictive regression) of the conditioning variables.