Granular Betas and Risk Premium Functions

This version: May 5, 2025

Tim Bollerslev[a,*], Andrew J. Patton[b], Rogier Quaedvlieg[c]

[a]*Department of Economics, Duke University, and NBER*
[b]*School of Banking & Finance, UNSW Sydney; Department of Economics, Duke University; School of Economics, Singapore Management University*
[c]*ABM AMRO, Amsterdam*

Abstract

We propose new refined measures of the local covariation between the return on an asset and a risk factor. Our proposed "granular betas" generalize the notion of up- and down-side betas to multi-factor functional measures of covariation. We show how the resulting granular beta functions may be used in the estimation of new "risk premium functions." Implementing the proposed methods with a large cross-section of U.S. equity returns, we find evidence against the traditional (non-granular) CAPM, the Fama-French three and five-factor models, and the Fama-French-Carhart model in favor of the new granular versions of these models. Our empirical results in turn provide new insights into where in the factor-space the compensation for exposures to systematic risks is mostly earned.

Keywords: Cross section of expected returns, factor models, downside risk, granular betas, risk premium functions.
J.E.L. codes: G11, G12, C58.

[⋆]We are grateful to the Editor (Michael Jansson) and two anonymous referees for their constructive comments. We would also like to thank participants at numerous conferences and seminars for their useful comments and suggestions.

[*]Corresponding author: Department of Economics, Duke University, Durham, NC 27708; 919-660-1846; boller@duke.edu.

1. Introduction

Linear factor models, starting with the CAPM (Sharpe, 1964; Lintner, 1965) and its many subsequent generalizations (Fama and French, 1992; Carhart, 1997; Fama and French, 2015, amongst many others), remain ubiquitous in empirical asset pricing. Their ease of implementation and interpretation makes the models an especially useful, if imperfect, tool for understanding systematic risks and explaining cross-sectional variation in returns. The advantages of linear models notwithstanding, the simplicity of such models may mask potentially important non-linear dependencies and pricing. For instance, prospect theory (Kahneman and Tversky, 1979) and disappointment aversion (Gul, 1991) both imply that investors care more deeply about losses than gains, and as a result up- and down-side risks need not be priced the same. Theoretical models involving preferences over higher-order moments (Kimball, 1993) and rare disasters (Wachter, 2013) similarly suggest that investors price tail risks more dearly than "modal" risks near the center of the distribution. Capturing such non-linear dependencies and pricing features requires either the addition of new factors, or new methods.

Rather than adding to the population of the already large factor zoo (Cochrane, 2011), we instead propose to refine the way in which we measure the risk exposures to a given set of existing factors. Drawing on recent developments in high frequency financial econometrics, we increase the information content of a given factor model by measuring the covariation between an asset and a factor in a granular fashion. To do so, we propose new measures of dependence designed to locally capture the strength of the covariance between the return on an asset and a factor across the entire support of the factor. The resulting new "granular betas" generalize the up- and down-side market betas of Ang, Chen and Xing (2006) to allow for multiple factors and a more refined look at the inherent dependencies between an asset and a given set of factors.

These more granular characterizations of systematic risk exposures in turn hold the promise of more accurate asset price predictions compared to the standard linear factor models, which effectively treat the granular beta functions as constants. However, the use of a function or a set of functions, as opposed to a scalar or a vector, to describe the risk exposures also complicates the estimation of the compensation for risks, or risk premiums. To overcome this hurdle, we rely on random field regressions (Cohen and Jones, 1969), now also more commonly referred to as functional linear regressions (see, e.g., Ramsey and Silverman, 2005), in which we regress the returns on a cross-section of assets on the granular beta functions to estimate new "risk premium functions." Our approach naturally extends the traditional Fama-MacBeth approach for estimating scalar risk premiums in linear factor models to the new granular setting. We show how these functional regressions may be meaningfully implemented empirically based on the use of sieve approximations (Andrews, 1991; Newey, 1997; Chen, 2007), together with standard OLS procedures for easily estimating the relevant parameters. This same approach also facilitates the construction of tests for various economic hypotheses of interest concerning the shape of the risk premium functions, including, e.g., that the functions are "flat" as implied by traditional linear factor models.

Implementing the new procedures with U.S. equity return data over the period from 1963 to 2020, we find that the out-of-sample fit of a "granular CAPM" is better than the standard non-granular and downside versions of the CAPM. We further trace the improvements to the premium for covariation with the market factor being especially large in the left most part of its distribution, consistent with the idea that downside tail risk is priced more dearly by investors. The use of granular factor betas similarly improves on the out-of-sample performance of the Fama and French (1992) three-factor model, the Carhart (1997) four-factor model, and the five-factor model of Fama and French (2015). The improvements in fit again accrue because the shape of the best-fitting risk premium function for the market factor is different from being flat, as are the risk premium functions for other factors, most notably the value and momentum factors. These results in turn translate into distinct "expected return functions" that illuminate where in the factor space the returns for different stocks and portfolios are typically earned. Conditioning the estimation of the risk premium functions on various economic indicators further points to non-trivial temporal variation in the compensation for different factor risks and the market's capacity to bear said risks.

The rest of the paper is organized as follows. We begin in Section 2 by placing our work in context of the extant literature. Section 3 presents the key ideas and new estimation and inference procedures. Section 4 discusses our main empirical findings. Section 5 considers various extensions and applications. Section 6 concludes. More detailed explanations of the new procedures, along with supportive empirical analyses and estimation results, are deferred to the Appendix. Additional robustness checks and empirical results are also provided in an online Supplemental Appendix.

2. Related literature

Our work relates to several strands of the asset pricing literature. The most closely related papers are perhaps the aforementioned study by Ang et al. (2006) on the downside CAPM, along with the more recent work by Bollerslev, Patton and Quaedvlieg (2022b), who refine the up- and down-side betas to also condition on the sign of the return on the individual assets. In contrast to both of these studies, however, which rely on fixed thresholds at zero for partitioning the betas, our method allows us to determine where, in the support of the factors, risk exposures earn greater or lesser compensation, as, e.g., in the tails versus the center of the distribution. Motivated by that same idea, the quantile spectral betas proposed by Baruník and Nevrla (2023) are also explicitly designed to focus on tail dependencies, and the compensation therefor, across different return horizons. The importance of allowing for more general non-linear factor structures has also recently been emphasized by Almeida and Freire (2023), who explicitly caution that the rejection of linear factor pricing models may stem from a rejection of linearity, rather than a rejection of the specific set of factors included in the models.

Our empirical finding that the estimated risk premium function for the market factor does indeed appear especially steep for the most left part of the support, also links our paper to earlier work on tails and the pricing of tail risks by Bollerslev and Todorov (2011), Kelly and Jiang (2014), Farago and Tedongap (2018), Bégin, Dorion and Gauthier (2020), Beason and Schreindorfer (2022), and Stoja, Polanski and Nguyen (2025), among many others. Unlike these studies, however, we consider both single and multi-factor models. Chabi-Yo, Huggenberger and Weigert (2022) and Massacci, Sarno and Trapani (2025) have also previously considered the pricing of tail risks in the context of multi-factor models, with the former paper imposing a specific threshold for where the tails begin and risk prices may differ, and the latter paper estimating a specific threshold. Instead, our approach deliberately relies on linear measures of risk, in the from of the proposed granular betas.

By measuring the covariation locally the granular betas can also capture higher-order moment dependencies not accounted for by regular linear factor models and conventional betas. The importance of incorporating preferences over higher-order moments when pricing financial assets, perhaps most notably skewness, have previously been highlighted by Harvey and Siddique (2000), Dittmar (2002), Conrad, Dittmar and Ghysels (2013), Colacito, Ghysels, Meng and Siwasarit (2016), Ghysels, Plazzi and Valkanov (2016), Langlois (2020), Schneider, Wagner and Zechner (2020), and Driessen, Ebert and Koëter (2022), among others.

The way in which we operationalize the granular betas draws directly on developments in high-frequency financial econometrics, and the interpreted of the measures as generalizations of the realized betas proposed Barndorff-Nielsen and Shephard (2004), previously used in empirical work by Andersen, Bollerslev, Diebold and Wu (2006) and Patton and Verardo (2012), among many others. The granular betas also formally, for increasingly finer partitions, encompass the jump and continuous betas defined and analyzed by Todorov and Bollerslev (2010) and Bollerslev, Li and Todorov (2016), and the multi-factor extensions thereof more formally developed by Aït-Sahalia, Jacod and Xiu (2025). The granular betas also extend the notion of realized semicovariances in Bollerslev, Li, Patton and Quaedvlieg (2020), which entail separate covariances for differently signed returns. However, instead of partitioning the covariance into four additive components, in our main analyses we instead partition the support of each of the systematic factors into a set of consecutive non-overlapping segments, more closely resembling the realized partial covariances of Bollerslev, Medeiros, Patton and Quaedvlieg (2022a). As such, the resulting multi-dimensional betas, and potentially asymmetric realized measures, are also reminiscent of the realized semivariances of Barndorff-Nielsen, Kinnebrock and Shephard (2010) and Patton and Sheppard (2015).

At a more general level, our use of finely sampled asset and factor returns for estimating the granular betas combined with more coarsely sample returns for estimating the risk premium functions also echoes the use of mixed-frequency data in the MIDAS approach pioneered by Ghysels, Santa-Clara and Valkanov (2006) and Ghysels, Sinko and Valkanov (2007). However, our reliance on functional linear regressions combined with sieve approximation techniques for quantifying the risk premiums distinguishes our mixed-frequency estimation scheme from the regression-based approach typically employed in the MIDAS literature.

Meanwhile, there is already a large literature, tracing back to Rosenberg (1974), which seek to relate factor loadings, and the cross-sectional variation therein, to firm characteristics and other directly observable variables in an effort to obtain more accurate factor pricing models. Notable econometric advances in this direction include the nonparametric techniques developed by Gregory and Linton (2007) and Conner, Hagmann and Linton (2012) designed to more flexibly estimate the characteristic-beta functions, and Gagliardini, Ossola and Scaillet (2016), who use instruments to capture time variation in risk premia. Related, Raponi, Robotti and Zaffaroni (2020) present econometric methods for “short” panels of asset returns, where the number of assets is large but the number of time periods is small, thereby allowing for slowly-varying risk loadings and premia. More recent work by Cai, Fang and Xu (2022) have also explored the use of functional-coefficient panel data methods for estimating the betas as non-parametric functions of macroeconomic variables. In a similar vein, Pelger and Xiong (2022) and Chen, Pelger and Zhu (2024) propose new inference procedures explicitly allowing for state-dependent factor models and time-varying loadings. Related, Freyberger, Neuhierl and Weber (2020) advocate the use of adaptive LASSO to help select the characteristics that matter the most for explaining the cross-sectional variation in expected returns, while Kelly, Pruitt and Su (2019), Kozak, Nagel and Santosh (2020), and Fan, Ke, Liao and Neuhierl (2023) utilize various dimension reduction techniques, including LASSO and PCA, to tractably incorporate the information from potentially hundreds of characteristics. We differ from these studies in that we do not attempt to augment the factor models with additional information stemming from firm characteristics and/or other economic state variables. Instead, we seek to enhance the explanatory power of already well-established linear factor models through the use of more refined measures of an asset's covariation with the existing risk factors.

The idea of partitioning the factor space to obtain additional information is also related to the concepts of random forests and regression trees routinely employed in the machine learning literature. These techniques have also recently been used by a long list of empirical studies, including, e.g., Gu, Kelly and Xiu (2020), Bianchi, Büchner and Tamoni (2021), and Aleti, Bollerslev and Siggaard (2025), to predict financial asset returns. Bryzgalova, Pelger and Zhu (2025) similarly rely on decision trees in their construction of informative portfolios, or test assets, for the estimation of a stochastic discount factor (SDF), while Cong, Feng, He and He use tree-based methods to generate basis portfolios for spanning the SDF. In contrast to all of these studies, which entail "blind" data-driven procedures for determining non-linear dynamic dependencies, or combining assets into "informative" portfolios, our empirical analyses rely on a set of well-established risk factors. However, instead of traditional linear factor models, we partition the factor loadings, or betas, into finer granular betas based on the joint empirical distributions of the assets and the factors, in turn allowing for the estimation of new risk premium functions and improved return predictions.

Our deliberate use of a small number of factors also sets the paper apart from recent studies that report improvements in out-of-sample forecast performance from the use of an extensive set of candidate factors and complex "black box" prediction models estimated by various techniques adopted from machine learning. In addition to the many studies already cited above, Gu, Kelly and Xiu (2021), Kelly, Malamud and Zhou (2024), and Didisheim, Ke, Kelly and Malamud (2024), in particular, all tout the "virtue of complexity" and the advantages of such models. We do not claim that the more structured approach developed here beat these types of models in terms of forecast accuracy. Instead, we aim to highlight how new economic insights can be gleaned from the use of more refined risk measures for a few well-established risk factors vis-à-vis commonly used small-scale linear factor models.

3. Granular betas and risk premium functions

We begin by formally defining and discussing the "granular betas" and some of their theoretical properties. We then describe how these measures allow for the estimation of corresponding "risk premium functions." For simplicity of exposition, we focus our initial discussion on the simple one-factor case. As discussed further below, the same ideas and new measures readily extend to multi-factor settings.

3.1. Main ideas and definitions

We denote the excess return on the factor and asset i by X and Y_i, respectively. The traditional beta of asset i with respect to the factor X is then simply defined by:

$$\beta_i \equiv \frac{\text{Cov}(Y_i, X)}{\sigma_X^2}, \tag{1}$$

where σ_X^2 refers to the unconditional variance of the factor, and

$$\text{Cov}(Y_i, X) \equiv \mathbb{E}[(Y_i - \mu_i)(X - \mu_X)], \tag{2}$$

where μ_i and μ_X denote the unconditional means of the asset and the factor, respectively. Now, consider a set of G partitions of the support of the factor return based on boundaries $Q_1, ..., Q_{G-1}$, with $Q_0 \equiv -\infty$ and $Q_G \equiv \infty$. Let the j^{th} partition be denoted by $\mathcal{G}_j = (Q_{j-1}, Q_j]$, for $j = 1, 2, ..., G$. Conditioning the expectation of the covariance in (2) on the factor return lying in partition $\mathcal{G}_j$, we define the corresponding granular covariance for asset i as:[1]

$$\text{GCov}(Y_i, X; \mathcal{G}_j) \equiv \mathbb{E}[(Y_i - \mu_i)(X - \mu_X)|X \in \mathcal{G}_j]. \tag{3}$$

Correspondingly, the new set of granular betas are obtained by replacing the unconditional covariance in (1) with the granular covariances:[2]

$$\text{G}\beta_i(\mathcal{G}_j) \equiv \frac{\text{GCov}(Y_i, X; \mathcal{G}_j)}{\sigma_X^2}. \tag{4}$$

To help fix ideas, consider a one-factor model such as the CAPM. It follows readily that the probability weighted sum of the granular betas equals the traditional beta defined in (1):

$$\beta_i = \sum_{j=1}^{G} \text{G}\beta_i(\mathcal{G}_j)\text{Pr}[X \in \mathcal{G}_j]. \tag{5}$$

Thus, in the one-factor setting the new measures provide a simple and easy-to-interpret decomposition of the usual CAPM beta, hence the moniker "granular betas." Moreover, if the quantiles of the market factor are used to obtain a set of equally likely partitions, implying $\text{Pr}[X \in \mathcal{G}_j] = 1/G$ for all j, the usual CAPM beta is simply obtained as the equal-weighted average of the granular betas.

To further appreciate what the granular betas estimate, it is instructive to consider

[1]Note, the expression on the right-hand–side is not a conditional covariance, which would be obtained as $\text{Cov}(Y_i, X|X \in \mathcal{G}_j) = \mathbb{E}[(Y_i - \mathbb{E}[Y_i|X \in \mathcal{G}_j])(X - \mathbb{E}[X|X \in \mathcal{G}_j])|X \in \mathcal{G}_j]$. However, as discussed further below, our definition ensures that the granular betas add up to the traditional beta, thereby facilitating comparisons of the new granular models with traditional linear factor pricing models.

[2]The granular covariances could alternatively be normalized by the corresponding granular variances, $\mathbb{E}[(X - \mu_X)^2|X \in \mathcal{G}_j]$. However, this alternative normalization would destroy the additivity of the betas discussed in equation (5), and the granular beta function obtained under the traditional CAPM, discussed below, would depend on further assumptions about the conditional variance of the idiosyncratic returns, which we avoid with this simpler normalization.

their limit for an increasing number of partitions, or for the width of the partitions converging to zero, say $\Delta \to 0$:

$$\begin{aligned} \mathrm{G}\beta_i((x, x+\Delta]) &= \frac{\mathbb{E}[(Y_i - \mu_i)(X - \mu_X)|x < X \leq x + \Delta]}{\sigma_X^2} \\ &\to \frac{(\mathbb{E}[Y_i|X = x] - \mu_i)(x - \mu_X)}{\sigma_X^2} \equiv \mathrm{G}\beta_i^*(x). \end{aligned} \tag{6}$$

We will refer to this limiting function as the "granular beta function." The functional form of $\mathrm{G}\beta_i^*(x)$ reveals that granular betas are related to the conditional mean of the test asset given the factor. Analogous to the case with a finite number of partitions, the law of iterated expectations again implies that the expectation of this function equals the traditional beta in (1), and in the case of X being the market factor the usual CAPM beta we obtain

$$\beta_i = \mathbb{E}(\mathrm{G}\beta_i^*(x)) = \int_{-\infty}^{\infty} \mathrm{G}\beta_i^*(x) f_X(x) dx, \tag{7}$$

where $f_X(\cdot)$ denotes the PDF of X.

Meanwhile, it is possible that the compensation earned for exposure to the factor X differs across the support of the factor. For example, existing empirical evidence suggests that associations with negative factor returns carry a larger risk premium than those with positive factor returns (Ang et al., 2006) and that market left-tail risk carries a higher risk premium than non-left-tail risk (Chabi-Yo, Ruenzi and Weigert, 2018). By more finely decomposing the covariance of an asset with the factor, we can gain greater insights into *where* in the support of the factor the risk premium for a particular asset tends to be earned.

Specifically, let $\lambda(x)$ denote the risk premium function that characterizes how the compensation varies with x across the support of X. The expected return on Y_i may then be expressed as:

$$\mathbb{E}(Y_i) = \int_{-\infty}^{\infty} \lambda(x) \mathrm{G}\beta_i^*(x) f_X(x) dx. \tag{8}$$

To help further intuit the meaning of the granular betas and risk premium functions defined above, suppose that the traditional CAPM holds, so that the conditional mean of the return on Y_i satisfies $E[Y_i|X] = \beta_i X$, where X denotes the excess return on the market. Substituting this expression into equation (6) shows that the granular beta function implied by the CAPM is a parabola:[3]

$$\mathrm{G}\beta_i^*(x) = \beta_i \frac{(x-\mu_X)^2}{\sigma_X^2}. \tag{9}$$

Moreover, as formally shown in Appendix A, for the expected return to be linear in β_i, the $\lambda(x)$ risk premium function cannot depend on x. In other words, for the traditional CAPM to hold, the lambda function must be constant across all values of x. We will refer to this as the lambda function being "flat."

Of course, there is strong empirical evidence indicating that the traditional CAPM does not hold true, and that the granular beta function with respect to the market factor does not resemble equation (9). To determine what the granular betas actually look like empirically, we turn next to their practical estimation.

3.2. Estimating granular betas

Drawing on the theory for realized semicovariances (Bollerslev et al., 2020), we estimate the time t granular beta for asset i with respect to factor X by:

$$\widehat{\mathrm{G}\beta}_{i,t,j} = \frac{\widehat{\mathrm{GCov}}_{i,t,j}}{\hat{\sigma}_{x,t}}, \tag{10}$$

where

$$\widehat{\mathrm{GCov}}_{i,t,j} = \sum_{s=t-S+1}^{t} (Y_s - \bar{Y}_t)(X_s - \bar{X}_t)\mathbf{1}\{X_s \in \mathcal{G}_j\}, \tag{11}$$

$$\hat{\sigma}_{x,t} = \sum_{s=t-S+1}^{t} (X_s - \bar{X}_t)^2, \tag{12}$$

[3]This parabolic form emerges from our formal definition of the granular betas in equation (4). If instead we had used conditional covariances in the numerator, as discussed in footnote 1, or granular variances in the denominator, as discussed in footnote 2, a different functional form would obtain.

and $\bar{X}_t$ and $\bar{Y}_t$ refer to the sample means of X_t and Y_t over the previous S periods, and the G partitions are similarly determined by the sample distribution of X_t over the past S periods so that each of the $\mathcal{G}_j$ intervals contains the same proportion $1/G$ of the observations.[4] This construction directly parallels the approach in the burgeoning high-frequency realized volatility literature, in which the variance or the covariance over fixed time-intervals is estimated by the sums of within interval more finely sampled squared or cross-products of returns, rather than their averages. Accordingly, the realized beta defined by Barndorff-Nielsen and Shephard (2004) is simply obtain by summing the granular beta estimates defined in equation (10) over all the G partitions:

$$\hat{\beta}_{i,t} = \sum_{j=1}^{G} \widehat{\mathrm{G}\beta}_{i,t,j},$$

mirroring the discussion in Section 3.1 and equation (5) pertaining to the true CAPM beta and the corresponding true granular betas.

3.3. Estimating the risk premium function

Our approach for estimating the risk premium function builds directly on the traditional Fama-MacBeth cross-sectional regression approach for estimating the risk premiums in linear factor pricing models. However, instead of regressing the returns on each of the assets on their estimated factor loadings ("betas") to estimate the scalar risk premiums ("lambdas"), we instead regress the asset returns on the granular beta functions formally defined in equation (6). That is, with a slight abuse of notation, we seek to estimate the non-linear regression:

$$Y_{i,t} = \alpha_t + \int_{-\infty}^{\infty} \lambda_t(x) \mathrm{G}\beta^*_{i,t-1}(x) f_{X,t-1}(x) dx + e_{i,t}, \qquad i = 1, 2, ..., N. \tag{13}$$

[4]In our empirical analyses discussed below, we rely on rolling windows of $S = 1,200$ trading days, or approximately five years, to estimate a given month's granular beta.

This may formally be interpreted as a random field regression (Cohen and Jones, 1969), nowadays also more commonly known as a functional linear regression (see, e.g., Ramsey and Silverman, 2005), in which the scalar asset returns, $Y_{i,t}$, are regressed on the random functions, $\mathrm{G}\beta^*_{i,t-1}(\cdot)$. The output from this "regression" will in turn provide an estimate for the intercept parameter α_t, and the functional slope coefficient $\lambda_t(\cdot)$. In contrast to the estimate for lambda obtained from the traditional Fama-MacBeth approach, which characterizes the average compensation for systematic risk, as measured by beta, the lambda function describes the compensation for systematic risk earned across the entire support of the underlying risk factor, as measured by the granular beta function.

The functional regression in (13) formally constitutes an ill-posed inverse problem (see, e.g., the discussion in Carrasco, Florens and Renault, 2007), necessitating the use of more complicated specialized estimation procedures. Indeed, there is a vast theoretically oriented literature in statistics seeking to address this problem through the establishment of different asymptotically rate optimal functional estimators (e.g., Hall and Horowitz, 2007; James, Wang and Zhy, 2009). The work by Babii, Carrasco and Tsafack (2024) provides an intriguing recent contribution to this literature by establishing optimal convergence rates for functional partial least-squares estimators. However, we do not pursue any of these more involved theoretically-justified rate optimal estimation procedures. Instead, we deliberately take a more pragmatic and easier-to-interpret approach vis-à-vis the ubiquitous Fama-MacBeth procedure.

Intuitively, mimicking the theoretical pricing equation in (8), the functional regression in (13) may naturally be seen as the limit that obtains when the width of the partitions used in the construction of the granular betas converge to zero. In practice, of course, with a finite number of return observations and without the imposition of any additional assumptions, we can only meaningfully estimate a finite number of granular betas. Correspondingly, we resort to the feasible estimates for the G granular betas defined in equations (10)-(12), together with the method of sieves (see Chen, 2007, for a review) for estimating the lambda functions. Specifically, we estimate λ_t based on the time-t

cross-sectional return regression:

$$
\begin{aligned}
Y_{i,t} &= \alpha_t + \sum_{j=1}^{G} \widehat{\mathrm{G}\beta}_{i,t-1,j} \lambda_t(j; G, p) + e_{i,t}, \qquad (14) \\
&= \alpha_t + \sum_{j=1}^{G} \widehat{\mathrm{G}\beta}_{i,t-1,j} \mathbf{LP}(2j/G - 1, p) \boldsymbol{\phi}_t + e_{i,t}, \qquad i = 1, 2, ..., N,
\end{aligned}
$$

where $\boldsymbol{\phi}_t$ denotes the vector of coefficients to be estimated, and $\mathbf{LP}(x, p)$ refers to the deterministic $1 \times (p+1)$-vector of Legendre polynomial transformations that we use for spanning the lambda function.[5,6] Consistent with economic intuition, the sieve approach automatically ensures that the estimate for $\lambda_t(\cdot)$ defined by $\mathbf{LP}(2j/G-1, p)\hat{\boldsymbol{\phi}}_t$ is smooth. As discussed further below, the use of a polynomial basis to span $\lambda_t(\cdot)$ also facilitates the imposition and test of specific economic hypotheses, including "flatness" as implied by the CAPM and other traditional asset pricing models.

Fixing the polynomial order p, the functional relationship in (14) may be easily estimated by OLS. Concretely, let $\mathbf{Y}_t$ denote the N_t-vector of returns, $\widehat{\mathrm{G}\beta}_{t-1}$ denote the $N_t \times G$ matrix of stacked granular betas, and $\mathrm{LP}(p)$ denote the $G \times (p+1)$ matrix of polynomial bases. Further define $\mathbf{Q}_t \equiv [\boldsymbol{\iota} \quad \mathrm{LP}(p)\widehat{\mathrm{G}\beta}_{t-1}]$, where $\boldsymbol{\iota}$ is a vector of ones. The parameter estimates may then readily be expressed as:

$$
[\hat{\alpha}_t \quad \hat{\boldsymbol{\phi}}_t]' = (\mathbf{Q}_t'\mathbf{Q}_t)^{-1}\mathbf{Q}_t' Y_t. \qquad (15)
$$

Thus, conditioning on the specific value of p, standard errors robust to cross-sectional heteroskedasticity are easily obtained from the conventional Eicker-Huber-White covariance

[5]Other basis functions, like B-splines, could, of course, be used instead. However, Legendre polynomials are by construction uncorrelated on the $[-1, 1]$ interval under the uniform measure, which makes them a particularly attractive choice when applied to the equidistant sequence $2j/G - 1$, $j = 1, \ldots, G$. As discussed further below, the Legendre polynomials also easily allow for the imposition of "flatness" and/or "symmetry" by fixing specific coefficients to zero; see also Li, Liao and Quaedvlieg (2022).

[6]By construction, $\int_{-1}^{1} LP(x,p)dx = 0$ for $p > 0$, while LP(x,0) = 1 for all values of x. This naturally ensures that the α_t parameter in (14) is identified. It also implies that when the order of the polynomial base is set to zero, we obtain a traditional Fama-MacBeth style risk premium estimate.

matrix:

$$Var([\hat{\alpha}_t \quad \hat{\boldsymbol{\phi}}_t]') = (\mathbf{Q}_t'\mathbf{Q}_t)^{-1}\mathbf{Q}_t' diag(\hat{\mathbf{e}}_t\hat{\mathbf{e}}_t')\mathbf{Q}_t(\mathbf{Q}_t'\mathbf{Q}_t)^{-1}, \tag{16}$$

where $\hat{\mathbf{e}}_t$ denotes the $N_t \times 1$ vector of residuals from equation (14). Moreover, again treating p as given, inference concerning the estimated conditional lambda function readily follows from the use of the delta method, and the fact that $\hat{\lambda}_t(j; G, p) = \mathbf{LP}(2j/G - 1, p)\hat{\boldsymbol{\phi}}_t$.

Of course, the theory of sieves formally requires that the polynomial order p grows to infinity together with the size of the estimation sample (see, e.g., Chen, 2007). In practice, we obviously have to chose a specific value of p for the $\mathbf{LP}(x, p)$ polynomial. As discussed further in Section 4.2 below, in our empirical analyses we follow common practice in the literature and resort to cross-validation techniques for doing so. This will generally result in a choice of p that is MSE-optimal, in the sense that the bias of the estimated lambda-function is of the same order as the standard error. However, the traditional inference procedures outlined above do not account for any biases in the functional estimates, only variance. We conjecture that a theoretically justified approach to formally circumvent this problem, would be to choose a larger-than-MSE-optimal sieve dimension, so that the bias is of asymptotically smaller order than the sampling error (see, e.g., the discussion in Chen and Christensen (2015)). Still, the interaction of the estimated $\widehat{\mathrm{G}\beta}$ with the lambda-function further complicates the situation relative to the more traditional setting, and we do not attempt such a formal proof, or explicit adjustment procedure, here. Instead, we simply caution that as usual the traditional inference procedures and corresponding confidence intervals outlined above that do not explicitly take into account possible biases that might arise in connection with the use of a sieve approximation and data-dependent tuning parameter need to be interpreted accordingly.

Meanwhile, in addition to the unrestricted estimates obtained for a given value of p, to help mitigate possible over-fitting, we also introduce a second hyperparameter, ω. This additional regularization parameter serves to shrink the unrestricted $\hat{\lambda}_t(\cdot)$ estimates towards the estimated flat lambda function obtained for $p = 0$. Accordingly, our final estimate of $\lambda_t(\cdot)$ is constructed as:

$$\hat{\lambda}_t(j; G, p, \omega) = \omega\hat{\lambda}_t(j; G, 0) + (1 - \omega)\hat{\lambda}_t(j, G, p) \tag{17}$$

where $\omega = 0$ corresponds to no shrinkage, while for $\omega = 1$ the estimate is completely shrunk to a flat function. In parallel to our choice of p, in our empirical analyses we rely on standard cross-validation techniques for choosing the value of ω, as further discussed in Section 4.2 below.

Analogous to the second-stage regression in the traditional Fama-MacBeth approach used for estimating the risk premiums in linear factor pricing models, all of the functional $\hat{\lambda}_t(\cdot)$ estimates discussed above pertain to a single time period t. To obtain a meaningful estimate of the risk premium function and the compensation earned across the entire support of the factor, it is natural to average these period-by-period estimates over the full sample, say $t = 1, 2, ..., T$:

$$\hat{\lambda}(j; G, p, \omega) = \frac{1}{T}\sum_{t=1}^{T}\hat{\lambda}_t(j; G, p, \omega). \tag{18}$$

Following standard econometric procedures, robust pointwise standard errors for the resulting $\hat{\lambda}(\cdot)$ function may readily be constructed based on an estimate of the long-run covariance matrix (e.g., Newey and West, 1987) for the time series of the individual period-by-period lambda function estimates.[7]

3.4. Testing economic restrictions

The sieve approach allows for the meaningful estimation of the individual lambda functions by succinctly parameterizing the functions using polynomial basis functions, effectively reducing the dimensionality of the problem. As previously noted, the use of a polynomial bases to span the lambda functions also conveniently permits the imposition

[7] Estimation and inference pertaining to risk premium functions over different sub-samples, including samples conditioned on specific economic indicators, may be constructed in a similar manner.

of various restrictions on the shape of the period-by-period lambda functions.

Specifically, let $\phi_{t,i}$ denote the coefficient associated with the i^{th} polynomial basis function in period t, with $\phi_{t,1}$ denoting the linear coefficient, $\phi_{t,2}$ denoting the quadratic coefficient, etc. The flatness condition for the lambda function discussed in Section 3.1 above in connection with the implications of the traditional CAPM, corresponding to $\lambda(x) = \bar{\lambda}$ for all values of x, may then simply be imposed by fixing $\phi_{t,i} = 0$ for all i, except $i = 0$.

In a frictionless financial market, without any short-sale or leverage constraints, the lambda function would naturally also be symmetric around zero, as a short position in Y_i merely switches the sign of the granular beta function (see also the related discussion in Bollerslev et al., 2022b). Accordingly, we will refer to lambda functions for which $\lambda(x) = \lambda(-x)$ as satisfying a "symmetry" condition. Of course, legal constraints, higher costs association with short selling, and other impediments, may create limits-to-arbitrage and arbitrage risk (see, e.g., Pontiff, 1996; Schleifer and Vishny, 1997; Hong and Sraer, 2016), thereby causing $\lambda(x) \neq \lambda(-x)$.[8] Intuitively, if X is a market factor, assets that covary more strongly with the market when the market is performing poorly tend to exacerbate downside risk, while assets that covary less with the market when the market is performing poorly help mitigate downside risk. As a result, the latter type of assets may on average demand less of a risk premium causing the lambda function to be asymmetric (as in Ang et al., 2006). In parallel to the flatness condition, this symmetry condition can also easily be imposed and tested by restricting the partitions to be symmetric around zero and fixing $\phi_{t,i} = 0$ for all odd integers i (see Appendix B for additional details).

Testing these types of shape restrictions within a fully nonparametric framework in which p, the polynomial order of the sieve used in the estimation of the lambda function, is formally assumed to increase to infinity together with the sample size presents formidable challenges from a theoretical perspective and would in turn necessitate the

[8]See also the recent discussion and extensive literature review in Gârleanu, Panageas and Zheng (2025) pertaining to the performance of shorting strategies and various economic constraints and mechanisms that might impede arbitrage.

use of specialized inference procedures, typically involving the construction of uniform confidence bands. The earlier studies by Mammen (1991a,b) and Fan and Gijbels (1992) provide a more formal account of the pertinent complications and possible theoretical solutions (see also the review in Chen (2007), along with the more recent discussion in Horowitz and Lee (2017)). We do not pursue any of these more complicated theoretically-justified procedures here. Instead, we take a more pragmatic approach and effectively condition our inference about the shapes of the risk premium functions on the specific choices of p made by our practical cross-validation procedure. We caution that formally this might not produce a valid test, as there is no guarantee that normality holds once the specific data-driven p is conditioned on.

Meanwhile, when considering the estimate of the risk premium function obtained by averaging the period-by-period lambda function estimates over multiple periods, as in (18), it is not possible to directly test for flatness or symmetry based on the period-by-period ϕ_t parameters, as different combinations of the polynomial coefficients for the estimated $\hat{\lambda}_t(\cdot)$ functions may result in non-flat and/or asymmetric average lambda functions $\hat{\lambda}(\cdot)$, even when the average of the individually estimated parameters do not.[9] Hence, we therefore test the hypotheses of interest by directly evaluating the average function estimates at B pre-set fixed points. That is, we test symmetry by:

$$H_0^{(Sym)}: \ \lambda^{(Unr)}(j; G, p, \omega) - \lambda^{(Sym)}(j; G, \tilde{p}, \tilde{\omega}) = 0, \quad \forall\, j = 1, ..., B, \tag{19}$$

where the corresponding unrestricted functional estimate, $\hat{\lambda}^{(Unr)}$, allow all polynomials in the period-by-period lambda functions to have non-zero coefficients, while the symmetric functional estimate, $\hat{\lambda}^{(Sym)}$, only allows even powers in the individually estimated lambda functions to have non-zero coefficients. Similarly, we test the stronger flatness condition

[9] For example, when $p = 4$, similar functions can be obtained with $\phi_2 > 0,\ \phi_4 < 0$ and $\phi_2 < 0,\ \phi_4 > 0$, making it possible that the coefficients are zero on average, while in each single period, and therefore on average, the lambda function is non-flat.

by:

$$H_0^{(Flat)}: \lambda^{(Unr)}(j;G,p,\omega) - \lambda^{(Flat)}(j;G,0,0) = 0, \quad \forall\, j = 1,...,B, \tag{20}$$

where the individual lambda function estimates underlying the averaged flat function estimate, $\lambda^{(Flat)}$, only has an intercept.[10] Again, we caution that our inference concerning the hypotheses in (19) and (20) implicitly treats the polynomial order p as known and thus effectively ignores asymptotic biases in the sieve approximation. As such, this might not produce a valid test since the bias could shift confidence sets away from the truth.[11]

3.5. Granular multi-factor models

All of the estimates and hypothesis tests discussed above pertains to a one-factor setting. The same general ideas and intuition readily extends to the estimation and tests of granular K-factor models.

Specifically, in a direct extension of the one-factor case, consider the situation in which the partitions for the k^{th} factor only depends on that same factor. Analogous to equation (11), the j^{th} granular realized covariance for the k^{th} factor may then similarly be defined as:

$$\widehat{\mathrm{GCov}}_{i,k,t,j} = \sum_{s=t-R+1}^{t} (Y_{i,s} - \bar{Y}_t)(X_{k,s} - \bar{X}_{t,k})\mathbf{1}\{X_{k,s} \in \mathcal{G}_j\}. \tag{21}$$

Stacking the K factors and the j^{th} granular covariances into the $(K \times 1)$ vectors $\mathbf{X}_t = [X_{1,t},...,X_{k,t}]'$ and $\widehat{\mathbf{GCov}}_{i,t,j}$, respectively, the multi-factor granular beta estimates are then simply obtained as:[12]

$$\widehat{\mathbf{G}\boldsymbol{\beta}}_{i,t,j} = \hat{\mathbf{V}}_{x,t}^{-1}\widehat{\mathbf{GCov}}_{i,t,j}, \tag{22}$$

[10] Since the symmetric and flat estimates may rely on different hyperparameters from the unrestricted estimates, we further optimize these hyperparameters in a separate validation sample from that of the test sample.

[11] One way to avoid bias asymptotically is by under-smoothing, thereby ensuring that the bias is of smaller order than the sampling uncertainty; see, e.g., the discussion in Chen and Christensen (2018). This, of course, would not solve other problems associated with a stochastic p.

[12] In parallel to the one-factor case, the granular betas also sum to the usual factor betas when weighted by the probability of each partition. Mirroring the discussion pertaining the CAPM, in a strict K-factor model with orthogonal factors, the granular beta functions are also parabolas, with proportionality coefficients equal to the factor loadings.

where

$$\hat{\mathbf{V}}_{x,t} = \sum_{s=t-R+1}^{t} (\mathbf{X}_s - \bar{\mathbf{X}}_t)(\mathbf{X}_s - \bar{\mathbf{X}}_t)'. \tag{23}$$

Replacing $\widehat{\mathrm{G}\beta}_{i,t-1,j}$ in the sieve regression defined in (14) above with $\widehat{\mathbf{G}\boldsymbol{\beta}}_{i,t,j}$, the joint estimation of the $\lambda_{k,t}(\cdot)$ risk premium functions for each of the K factors, together with inference and hypotheses testing, proceeds exactly as in the one-factor case.

With a total of G partitions for each of the K factors, the factor-specific partitions outlined above results in a total of $K \times G$ granular betas. More involved multi-factor specifications in which the factors are jointly partitioned could, of course, be entertained as well. In an extension of our main empirical analysis based on factor-specific partitions, Section 5.2 briefly considers such a multi-dimensional approach. However, this extension comes at the expense of additional notational complexity, and we defer a more detailed discussion to that section. Instead, we turn next to a discussion of the data and estimation setup that we rely on throughout our empirical analyses.

4. Granular factor model estimates

4.1. Data and granular beta estimates

Our empirical analyses rely on daily return data from the Center for Research in Securities Prices (CRSP) database, spanning the period July 1963 to December 2020. Following standard practice, we consider all stocks with CRSP codes 10 and 11, and remove all penny stocks with prices less than five dollars to alleviate biases arising from price discreteness. We further require five years of daily data to estimate the granular betas. All in all, this leads to a total of 181,804 firm-month observations.

To illustrate the added flexibility afforded by the new granular betas, Figure 1 plots the averages of the traditionally estimated CAPM betas for all firms in the "Hi-Tech" and "Utilities" industries relative to the grand average of the betas for all the stocks in our sample.[13] As the estimated ratios for the traditional betas, labeled "Beta" in the figure,

[13]The industry definitions are based on the 10-SIC classifications on Ken French's website.

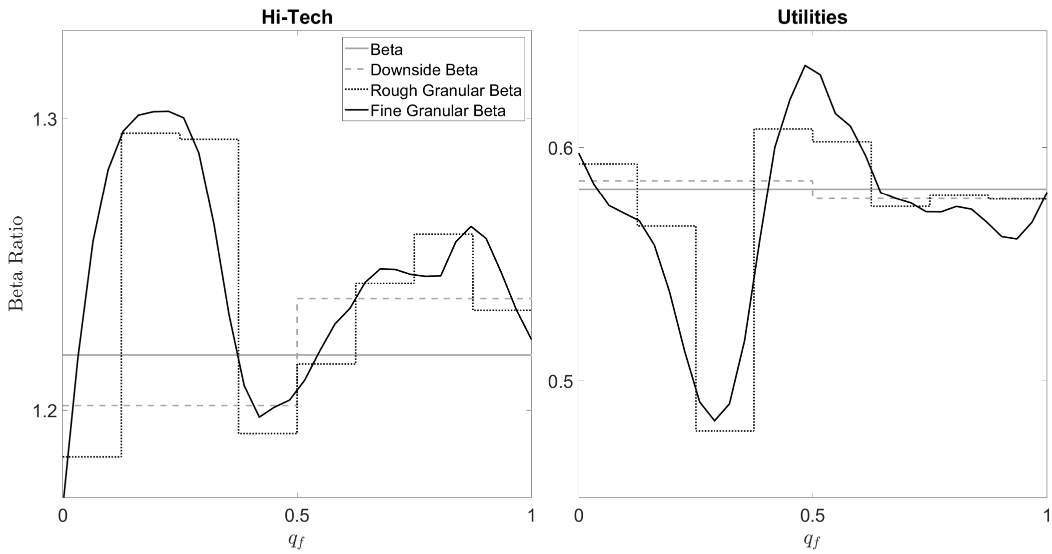

Figure 1: **Relative granular betas for hi-tech and utility stocks**. The figure shows the average CAPM, downside, and granular market betas of "Hi-Tech" and "Utilities" stocks relative to the corresponding average betas for all firms in our sample. The ratios are plotted as a function of the quantiles of the market return, q_f.

show, hi-tech firms tend to have betas in excess of the market average, while the betas of utility stocks tend to be lower. Looking at the dashed lines, labeled "Downside Beta," which report the ratios of the average estimates of the up- and down side betas for the two industries relative to the up- and down-side betas for all stocks, demonstrates that even though the down-side betas for hi-tech stocks tend to be larger than the down-side beta for the average stock, the up-side betas for hi-tech stocks are even larger in a relative sense. By comparison, the up- and down-side betas for utility stocks are both very similar in a relative sense to the average up- and down-side betas for all stocks. Meanwhile, the new "Rough" and "Fine" granular beta estimates, corresponding to $G = 8$ and $G = 64$ in the analysis below, respectively, tell a more nuanced story. In particular, while the betas for hi-tech firms are generally higher than the betas for most other stocks, in extreme down markets, when q_f is close to zero, hi-tech stocks typically move less with the market than their traditional CAPM and downside betas would suggest. Conversely, for utility stocks the more pronounced deviations from the traditional beta estimates occur near the center of the market return distribution.

4.2. Risk premium function estimates

Turning to the risk premium functions, we focus on the one-factor CAPM, the Fama and French (1992) three-factor model (FF3), the Carhart (1997) four-factor model (FF3+Mom), and the Fama and French (2015) five-factor model (FF5). These models have arguably emerged as the leading factor models in the literature. The returns for all of the factors, as well as the risk-free rate required to compute the excess returns, are sourced from Ken French's website. As a reference, Appendix C reports the traditional annualized full-sample Fama-MacBeth risk premium estimates obtained for each of the traditional factor models. For comparison, we also report the estimates obtained by splitting each of the factors into separate up and down components.

As discussed in Section 3.3, our use of a local polynomial approach for estimating the lambda functions introduces two additional hyperparameters, p and ω. As noted in that same section, we rely on standard cross-validation techniques for choosing these parameters. This approach in turn also dictates our choice of estimation and forecasting samples. Specifically, we begin by estimating the requisite granular beta functions using the $1,200$ daily returns up until the start of the month over which the monthly returns that we use in estimating the lambda functions are measured. For a given ordering of the assets, we then use the first 60% of assets to estimate the α_t and $\boldsymbol{\phi}_t$ parameters in (14) on a month-by-month basis for a range of different values of the two hyperparameters. We then use the next 20% of assets as a validation sample to select the polynomial order and degree of shrinkage based on the average R^2 obtained for this collection of assets. Finally, we use the resulting estimated risk premium functions to price the remaining 20% of the assets. Accordingly, the results obtained for these remaining 20% of the stocks effectively constitute an out-of-sample evaluation of the estimated risk premium functions. Since the order of the assets is arbitrary, we repeat this procedure one hundred times for different random permutations of assets, and report the averages obtained across these permutations as our final estimation results.

We commence our empirical analysis by considering a granular version of the standard

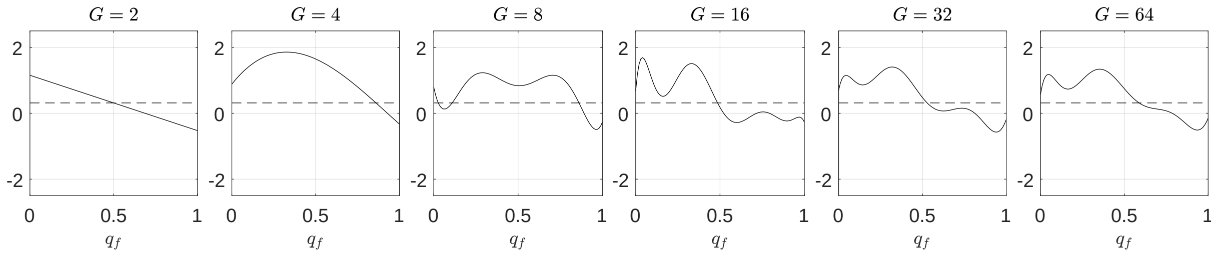

Figure 2: **Market factor risk premium function estimates.** The figure presents the estimated market risk premium function as a function of the market return quantiles for various functional beta granularities (G). The dashed lines report the standard Fama-Macbeth CAPM estimate obtained for $G = 1$.

CAPM. The resulting full-sample risk premium function estimates, $\hat{\lambda}(\cdot)$, are presented in Figure 2 as a function of the quantiles of the market return q_f (ranging from zero to one). To help illustrate the approach, we report the results obtained for different degrees of granularity, increasing from $G = 1$, corresponding to the traditional flat CAPM benchmark reported as the dashed line in all panels, all the way up to $G = 64$.

Looking first at the $G = 2$ case in the left-most panel, the estimates corroborate existing empirical evidence that the market risk premium tend to be larger when the market is down (i.e., for the left half of its support). As G increases, the estimated risk premium function is allowed to take more flexible shapes. However, by $G = 16$ the function has seemingly converged, as there appears to be little systematic change in the general shape when considering larger values of G. Broadly consistent with the up- and down-side CAPM of Ang et al. (2006), the estimated risk premium function is generally positive when the market return is below its median, and close to zero when the market is above its median.

Extending these CAPM-based results, we next consider granular versions of the FF3, FF3+Mom, and FF5 models. To help streamline the presentation, we focus on the most flexible $G = 64$ granular beta models. Figure 3 presents the resulting risk premium functions estimated over the full sample, along with 95% pointwise confidence intervals. The top-left panel in the figure corresponds directly to the right-most panel in Figure 2, with confidence intervals included.

The results for the three-factor FF3 model given in the second row of Figure 3 show that the estimated market risk premium function remains qualitatively the same as in the CAPM model in the top row: the market risk premium is positive when the market return is in its left tail, while the premium is close to zero for points in the support of the market return distribution slightly above its median. The fact that the estimated function appears smoother is attributable to the selected polynomial order for the market factor in the FF3 model being lower than in the CAPM, consistent with a greater need for parsimony in larger models.[14]

Turning to the estimates for the other factors in the FF3 model, the risk premium function for SMB exhibits a U-shape. At the same time, however, the confidence intervals suggest that the function is effectively flat. Related, while the size effect was on average positive and significant up until around 1980 (Banz, 1981; Reinganum, 1981), the size risk premium has seemingly declined over time, and estimates using standard methods has also found it to be insignificant with more recent data (Schwert, 2003; Ahn, Min and Yoon, 2019).

The estimated risk premium function for HML evidence a more pronounced U-shape, with negative premiums when HML is around its median and positive premiums for realizations in either of its tails. Consistent with many other studies documenting that the value premium has substantially weakened over the years, and perhaps even disappeared (e.g., Fama and French), the traditional full-sample Fama-Macbeth risk premium estimate for HML (reported in Appendix C), which effectively averages the lambda function over the support of the HML factor, is also close to zero. However, the more nuanced picture afforded by the functional estimate in Figure 3 tells a more complex story.

The third row of Figure 3 shows the estimation results for the FF3+Mom model. Not surprisingly, the estimated risk premium functions for the first three factors are all broadly in line with the results for the FF3 model. Meanwhile, the premium function for the momentum factor appears mildly non-linear. Looking at the bottom row and

[14] In particular, while the selected polynomial order for the CAPM is eight, in the FF3 it is only three. The cross-validated optimal polynomial and shrinkage parameters for all the models that we consider are reported in Table D.2 in Appendix D.

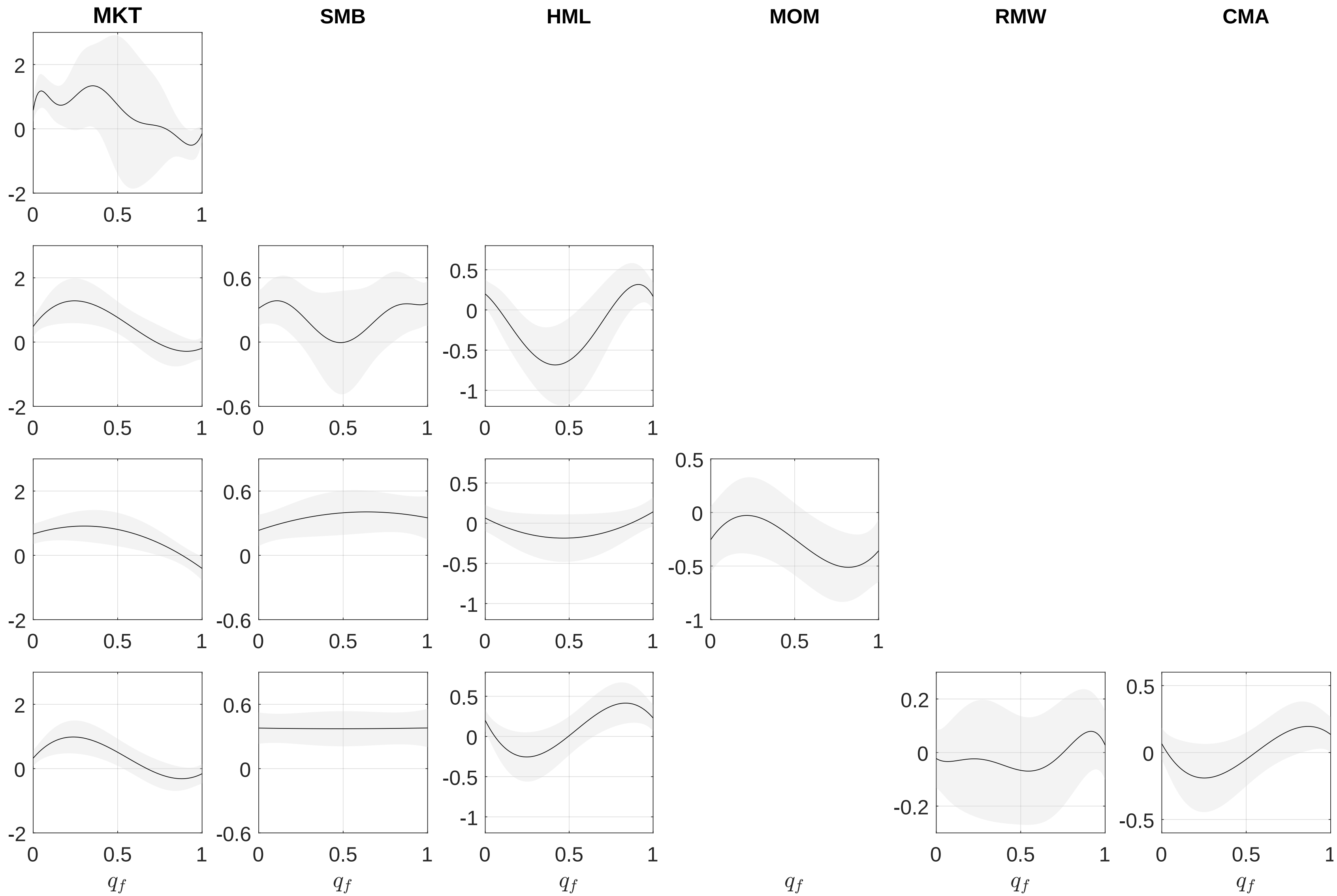

Figure 3: **Multi-factor risk premium function estimates.** The figure presents the estimated risk premium functions based on granular beta functions with $G = 64$, along with pointwise 95% robust confidence bounds. The rows correspond to the CAPM, FF3, FF3+Mom, and FF5 models, respectively.

the estimates for the FF5 model, shows that the risk premium function for profitability (RMW) appears roughly flat, while the estimated function for the investment (CMA) factor is almost a mirror image of the estimate for the market factor: the premium is negative or zero for the left tail of the factor and positive in the right tail.

In order to corroborate these visual impressions, Table 1 presents the results of our more formal tests for flatness and symmetry. We test the restrictions separately for each factor in a given model, as well as jointly across all factors in the model. In accord with the largest value of G used in the estimation of the granular beta functions, we rely on 64 equally-spaced points on the average (across time) estimated risk premium functions in implementing the tests, corresponding to $B = 64$ in equations (19) and (20). Again, echoing the discussion in Section 3.4, we caution that the statistical significance of these tests need to be carefully interpreted, as they condition on the specific choices of p made by our cross-validation procedure. Treating the polynomial order p as known effectively ignore asymptotic biases in the sieve approximation, which could shift confidence sets away from the truth. It also doesn't guarantee that normality holds.

Meanwhile, looking first at Panel A, which considers the null of symmetry as stipulated in (19), we observe that the reported p-values for the joint tests in the left-most column are all zero to three decimal places, indicating a strong rejection of the hypothesis. The tests for each of the factors individually, reported in columns two through seven, reveal that the strongest evidence against symmetry comes, perhaps not surprisingly, from the market factor, followed by the momentum and value factors. Tests of risk premium function symmetry for the remaining three factors, size, profitability, and investment, all seemingly support the null hypothesis.

Panel B in turn tests the stronger restriction that the risk premium functions are flat, as stipulated by the null in (20). This restriction is again seemingly strongly rejected for all of the models, with the reported p-values for the joint tests for the CAPM, FF3 and FF5 models all being less than one percent, and that for the FF3+Mom model being

Table 1: **Tests on risk premium functions.** The table reports p-values from tests of the restrictions on the risk premium function in (19) and (20). The tests are implemented using 64 equally-spaced points on the average (across time) risk premium function estimates. Column 1 reports the results from joint tests for all of the factors in a given model, while columns 2-7 report the results from testing each of the factors individually.

	Joint	MKT	SMB	HML	MOM	RMW	CMA
Panel A. $H_0 : \lambda$ **is symmetric**							
CAPM	0.000	0.000					
FF3	0.000	0.000	1.000	0.089			
FF3+Mom	0.000	0.000	0.969	0.100	0.004		
FF5	0.000	0.000	0.983	0.000		0.231	0.156
Panel B. $H_0 : \lambda$ **is flat**							
CAPM	0.007	0.007					
FF3	0.000	0.000	1.000	0.000			
FF3+Mom	0.029	0.000	1.000	0.091	0.076		
FF5	0.000	0.000	1.000	0.000		1.000	0.114

less than three percent.[15] In line with the previous tests of symmetry, these rejections are again predominantly driven by the estimates for the market, value, and momentum risk premium functions. Indeed, consistent with the visual impression from the estimates depicted in Figure 3, neither size, profitability nor investment exhibit any evidence against flatness.

Having established that the risk premium function estimates for the different factor models apparently differ from flat lines, we next seek to assess the improvement, or lack thereof, in out-of-sample (OOS) fit afforded by the new more flexible granular models. Recall that we always estimate the model parameters and hyperparameters using the first 60% and 20% of stocks respectively, leaving the remaining 20% of stocks for OOS model

[15] Since the hypothesis that the risk premium function is flat formally encompasses the hypothesis that the function is symmetric, one would naturally expect this hypothesis to be more strongly rejected by the data. However, random sampling variation can obviously result in more power to reject the specific null hypotheses stipulated in (19) compared to (20).

Table 2: **Out-of-sample explanatory power and pricing errors.** The table presents the average out-of-sample R^2s and the absolute pricing errors for the CAPM, FF3, FF3+Mom and FF5 factor models, estimated using standard OLS ($G = 1$) and granular betas with an increasing degree of granularity ($G > 1$). The reported values are averaged across the one hundred random cross-validation samples. Diebold and Mariano (1995) t-statistics for testing whether the granular models out-perform their non-granular counterparts in the first row are reported in parentheses. The t-statistics are calculated by a bootstrap procedure that accounts for the cross-sample dependence.

	R^2				Absolute pricing error			
G	CAPM	FF3	FF3+Mom	FF5	CAPM	FF3	FF3+Mom	FF5
1	3.145	4.447	5.204	5.040	0.150	0.053	0.042	0.084
2	3.484	4.966	5.622	5.337	0.169	0.049	0.020	0.068
	(4.834)	(6.410)	(5.951)	(5.367)	(-0.212)	(1.827)	(2.310)	(2.428)
4	3.561	5.019	5.644	5.398	0.145	0.014	0.010	0.042
	(5.848)	(6.543)	(6.385)	(6.030)	(0.462)	(2.443)	(2.468)	(2.710)
8	3.693	5.143	5.692	5.466	0.110	0.018	0.013	0.052
	(7.603)	(6.866)	(6.648)	(6.786)	(1.713)	(2.505)	(2.519)	(3.361)
16	3.805	5.236	5.770	5.536	0.084	0.019	0.026	0.071
	(8.538)	(8.089)	(6.611)	(6.394)	(2.028)	(2.390)	(1.983)	(2.361)
32	3.795	5.282	5.785	5.559	0.077	0.010	0.021	0.066
	(8.317)	(8.766)	(7.094)	(6.910)	(2.811)	(3.516)	(2.230)	(2.436)
64	3.840	5.275	5.787	5.549	0.054	0.002	0.010	0.067
	(8.120)	(8.203)	(7.051)	(7.256)	(3.882)	(3.555)	(3.404)	(2.413)

evaluation. In addition, to alleviate sensitivity to the specific stocks that appear in the OOS sample, we re-do that same analysis one hundred times, randomly permuting the stocks in the different samples.

Table 2 presents the resulting R^2 values and absolute pricing errors averaged across the one hundred OOS permutations. To formally compare the benchmark non-granular ($G = 1$) factor models with their granular counterparts ($G \geq 2$), we also report Diebold and Mariano (1995) (DM) t-tests (in parentheses) to assess the statistical significance of the reported differences, with positive values of the test statistics indicating that the granular models out-perform their traditional non-granular benchmarks. Since the one hundred OOS permutations are correlated (some assets randomly appear in more than one test sample), we rely on a joint bootstrap procedure for calculating a single DM test for all samples that explicitly accounts for the cross-sample dependence.

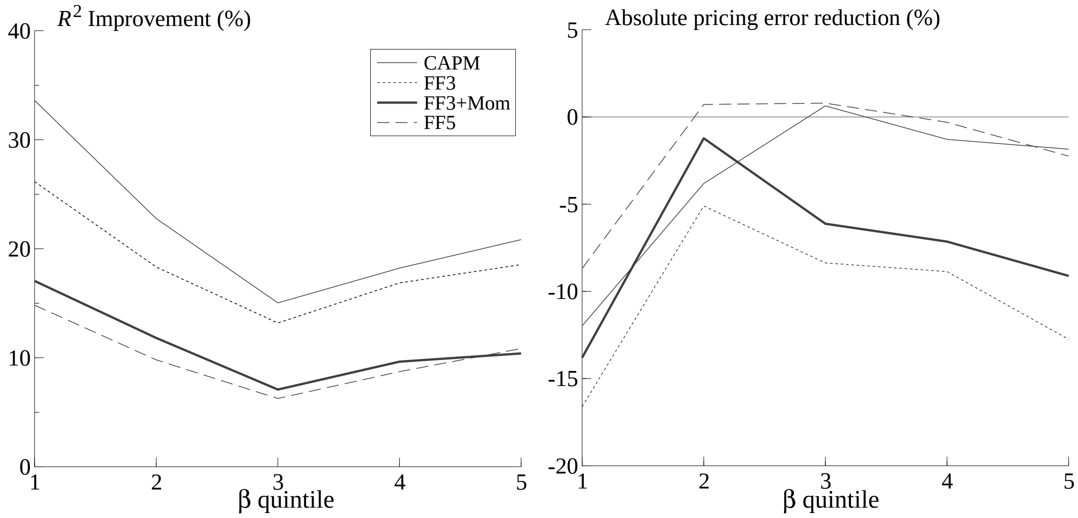

Figure 4: **Out-of-sample R^2s and absolute pricing error improvements across CAPM beta quantiles**. The figure shows the percentage improvements in the out-of-sample R^2s and absolute pricing errors for the granular versions ($G = 64$) of the CAPM, FF3, FF3+Mom and FF5 factor models, relative to the standard versions ($G = 1$) of the same models. The improvements are sorted into quintiles based on the standard ($G = 1$) CAPM betas of the individual stocks averaged across the one hundred random cross-validation samples.

Examining the magnitudes of the R^2 values in the left panel of Table 2 reveals that the "most granular" CAPM has an OOS R^2 that is roughly halfway between the non-granular CAPM and the non-granular FF3 model (3.8% compared with 3.1% and 4.4%, respectively), indicating that substantial gains in explanatory power can be achieved by using granular information, even without exploiting information from additional factors. Further corroborating that thesis, the "most granular" FF3 model out-performs both the non-granular FF3+Mom and FF5 models in terms of their OOS R^2s (5.3% compared with 5.2% and 5.0%, respectively). It is also noteworthy that *all* of the DM t-tests for comparing the granular models to their non-granular counterparts, ranging from $G = 2$ to $G = 64$, are statistically significant at conventional levels. In other words, significant gains in predictive accuracy can be obtained from using the richer information in the new granular betas and associated risk premium functions. The results for the absolute pricing errors, reported in the right panel of Table 2, tell a similar story. All of the granular models based on $G \geq 16$ partitions result in significantly lower average absolute pricing errors.

To further illuminate where these improvements are coming from, Figure 4 depicts the percentage improvements in the out-of-sample R^2s and absolute pricing errors for the granular ($G = 64$) CAPM, FF3, FF3+Mom and FF5 factor models relative to the standard ($G = 1$) non-granular versions of the same models, in which the averages across the one hundred random cross-validation samples are calculated for different quintile sorts defined according to the standard ($G = 1$) CAPM betas of the stocks. As the figure shows, most of the improvements come from improved predictions for low and high beta stocks. For the CAPM and the FF5 models, in particular, it appears that from this perspective almost all of the significant reductions in the magnitudes of the average pricing errors stem from low beta stocks.

To help better understand these cross-sectional results and the origins behind the documented improvements, we next demonstrate how the granular models manifest in differences in expected return functions for a set of specifically defined portfolios.

4.3. Expected return function estimates

Even though the compensation for exposure to the different factor risks are the same across all assets, as explicated in equation (8) the variations in the magnitude and the shape of the granular beta functions, $\mathrm{G}\beta^*_{i,k}(\cdot)$, interact with the risk premium functions, $\lambda_k(\cdot)$, to generate differences in the implied "expected return functions" across assets. To illustrate, Figure 5 plots the expected return functions implied by the CAPM, FF3, FF3+Mom and FF5 granular models as a function of the underlying factor returns averaged across all of the stocks in our sample (All), together with the implied expected return functions for the portfolios of technology stocks (HiTech) and utility stocks (Utils) previously discussed in Figure 1 in the introduction.[16]

Looking first at the left most column, all three portfolios exhibit roughly similar average expected return functions in regards to the market factor, although the Utils

[16]The Online Supplemental Appendix further details the expected returns for other industry portfolios, and how said returns may be attributed to the different factors in the CAPM, FF3, FF3+Mom and FF5 granular models.

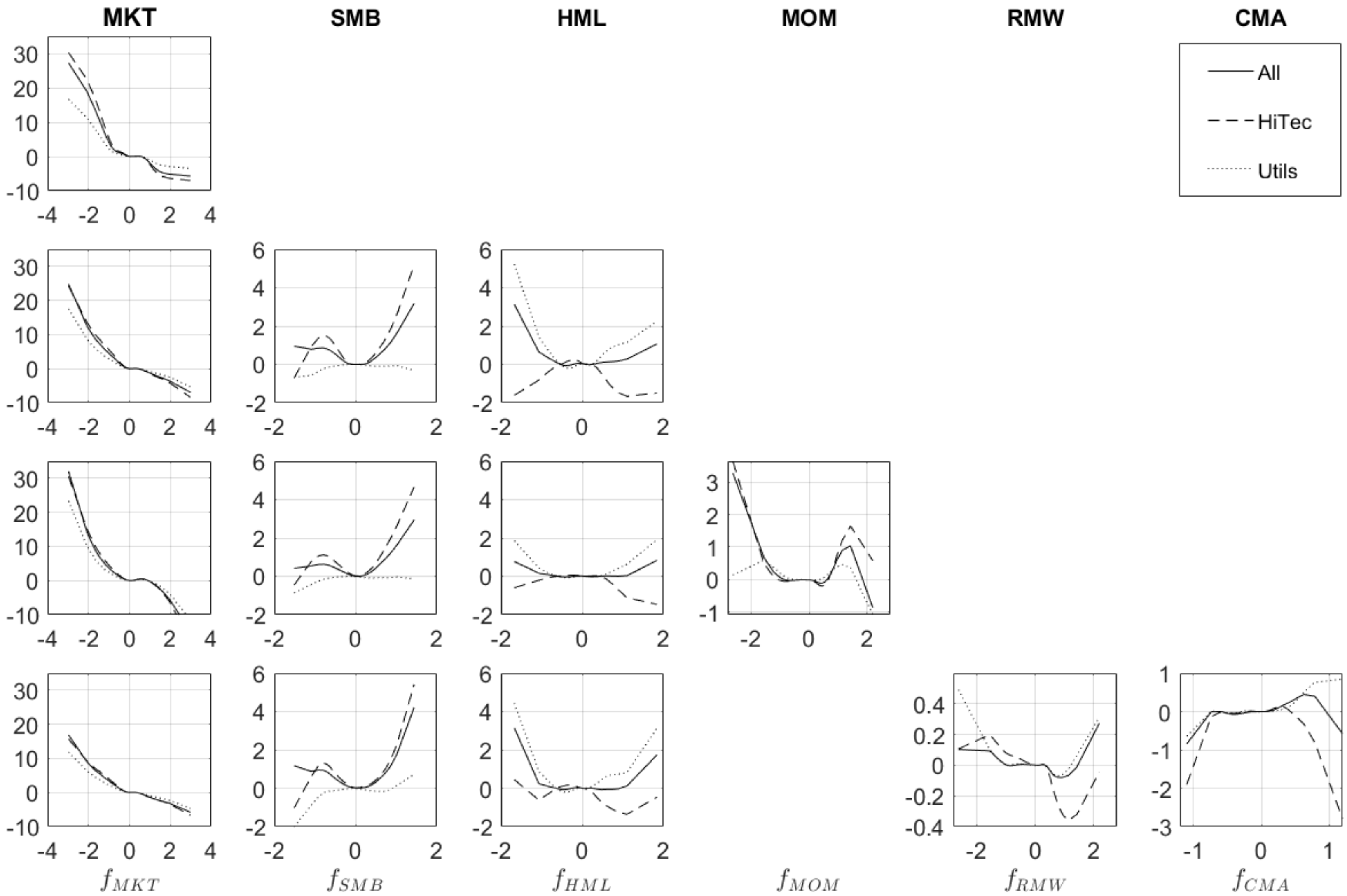

Figure 5: **Expected return functions.** The figure presents the expected return functions implied by the CAPM, FF3, FF3+Mom and FF5 granular models averaged across all of the stocks in the sample (All), together with portfolios comprised of technology stocks (HiTech) and utility stocks (Utils). The expected return functions are plotted as a function of the underlying factor returns.

curves appear slightly flatter than the other two curves, regardless of which other factors are included in the model. The pronounced steepness in the curves for especially low market returns, is consistent with the suggestion by Lu and Murray (2019) that the CAPM needs to be augmented with an additional factor to capture expected returns observed during deep market declines, or bear markets.

Even though all three portfolios earn little or no return when HML is near its median, when the value factor is in its tails the expected returns on utility stocks and most other stocks are high, while the expected returns on high tech stocks are low. Meanwhile, high tech stocks generally earn higher returns when the momentum factor is in its left tail, as do most other stocks, while the expected returns on utility stocks are mostly unaffected by the momentum factor. Similarly, the expected return on utility stocks appear largely unaffected by the size factor, while high tech stocks tend to earn higher returns when the size factor is in the right tail of its distribution, as do most other stocks.

In sum, the expected return functions implied by the granular betas and estimated risk premium functions clearly differ across different types of stocks. Importantly, these differences go beyond traditional factor models that restrict the compensation for factor risk to be the same regardless of the realizations of the factors.

5. Extensions and applications

We begin our additional analyses in this section by investigating whether the estimated risk premium functions vary systematically with a set of financial economic indicators. We then consider the use of joint partitioning of the factor space within the context of a multi-factor model. We conclude the section by considering the practical use of the granular betas in the construction of simple long-short portfolio strategies. For the sake of brevity, we focus on the workhorse FF3 model throughout all of these additional analyses.

5.1. Variation in the risk premium functions

Our main empirical analysis and discussion in Section 4 pertain to the estimates of the risk premium functions obtained by averaging the monthly lambda function estimates over the full sample. This directly mirrors the risk premium estimates traditionally reported in the asset pricing literature, which are similarly averaged over longer historical time periods to help reduce the noise in the estimates. However, risk premiums may be sensitive to economic conditions and the market's ability to bear risk (see, e.g., Cochrane, 2017). Our full-sample risk premium function estimates obviously mask any such dependencies.

To shed further light on this issue, we estimate separate risk premium functions by averaging the monthly lambda estimates conditional on three commonly used financial economic indicators: the Chicago Board of Options Exchange's VIX volatility index, the Financial Uncertainty index of Jurado, Ludvigson and Ng (2015)), and the UP versus DOWN market indicator of Cooper, Gutierrez and Hameed (2004).[17] Specifically, for the first two measures, we estimate separate average risk premium functions for months when the measures are above or below their median values. For the binary UP measure, which equals unity for 88% of the months in the sample, we simply estimate separate functions for its UP and DOWN states.[18]

Using the approach discussed in Section 3.4, we can again assess whether these conditional risk premium functions differ from one another. Keeping in mind the same caveats about the formal interpretation of these test statistics as before, Figure 6 plots the estimated functions for which that null is rejected at the 5% significance level. When the test for identical conditional risk premium functions do not reject, we instead plot the full-sample unconditional estimates previously shown in the second row in Figure 3.

As the first column in the figure shows, the risk premium function for the market

[17]We retrieve the VIX from the CBOE website. We use the Financial Uncertainty (one-month ahead) measure retrieved from `www.sydneyludvigson.com`. The UP measure equals one if the current market index price is higher than three years ago, and zero otherwise. We compute this index based on the market factor from Ken French's website.

[18]For comparison, the Online Supplemental Appendix also presents additional risk premium function estimates obtained by splitting the full sample into shorter non-overlapping 20-year periods.

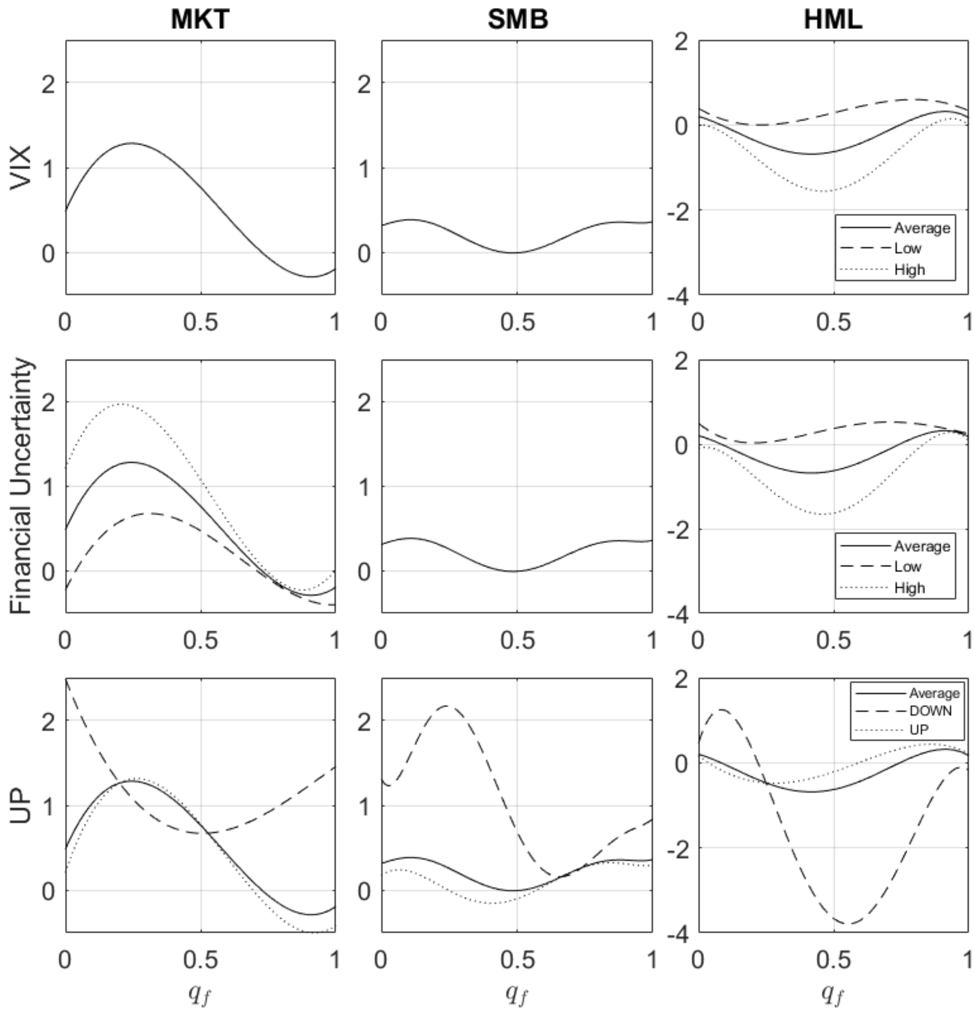

Figure 6: **Risk premium function estimates for different financial market conditions.** The figure presents the estimated risk premium functions for the FF3 model conditional on above and below median values of the VIX and the Financial Uncertainty index, as well as UP and DOWN market states. If our test indicates that the two estimated functions are similar, the figure plots the full-sample estimates.

factor seemingly varies with both the Financial Uncertainty index and the UP/DOWN market indicator. In particular, consistent with the idea that the risk bearing capacity is lower when financial uncertainty is high and during down markets, the corresponding risk premium function estimates are both notably higher in those states compared to their low and up market counterparts. The overall shape of the estimated market risk premium functions seemingly also change between UP and DOWN markets, indicating that the high compensation for left tail risk documented in the extant literature is mostly earned during bear markets.

Looking at the results for the size factor in the second column of the figure, reveals that while the estimated risk premium functions do not vary with the VIX or the Financial Uncertainty index, they do differ between UP and DOWN markets. Indeed, it appears as if the size premium is primarily attributable to higher returns on small firms in DOWN markets. This more nuanced picture is also broadly consistent with Souza (2020), who argues that the size premium is highly non-linear and mostly confined to "bad" economic states.

As previously noted, even though the HML premium as traditionally estimated is not significantly different from zero over the full sample, the estimated risk premium functions for the value factor seemingly differ when conditioning on the state of each of the three financial economic indicators. In particular, while the estimated functions are fairly flat and close to zero when the VIX and Financial Uncertainty are low and the market is in its UP state, the functions become notably more curved in the high and DOWN states, with especially pronounced negative deviations from zero in the center of the distribution of the factor.

These conditional depictions of when and where the factor risk premiums are earned may also potentially help shed new light on predictability of the equity premium as it relates to the phases of the business cycle and general economic conditions (see, e.g., Adrian, Crump and Vogt, 2019; Moench and Stein, 2021). However, we will not pursue that line of questioning any further here. Instead, we turn next to a discussion of the possible use of more refined multi-dimensional partitions.

5.2. Multi-dimensional partitions

The granular betas considered in our empirical analyses so far all rely on partitions determined one factor at a time. In a K-factor model one could entertain more refined partitions based on all K factors jointly. We now consider such an extension of our basic approach.[19]

[19]Building on the notion of partial covariances of Bollerslev et al. (2022a), one could also consider partitions based on other variables in addition to the factor(s), including the assets themselves.

To set out the idea, consider a K-factor model, where the support of each factor is partitioned into G regions. Denote the j^{th} partition of the k^{th} factor by $\mathcal{G}_j^k$. Considering all possible combinations of such partitions across the K factors, $\mathcal{J} \equiv (j_1, j_2, ..., j_K) \in \{1, 2, ..., G\}^K$, results in a total of $K \times G^K$ granular betas. By comparison, the one-dimensional partitions "only" generate $K \times G$ granular betas. Meanwhile, in parallel to the one-dimensional granular covariances and betas defined in (10) and (11), the corresponding multi-dimensional measures are readily obtained as:

$$\widehat{\text{GCov}}_{i,t,k}^{\mathcal{J}} = \sum_{s=t-R+1}^{t} (Y_{i,s} - \bar{Y}_t)(X_{k,s} - \bar{X}_{k,t})\mathbf{1}\{\mathbf{X}_s \in \mathcal{G}_{j_1}^1 \times \mathcal{G}_{j_2}^2 \times \cdots \times \mathcal{G}_{j_K}^K\},$$

and

$$\widehat{\mathbf{G}\boldsymbol{\beta}}_{i,t}^{\mathcal{J}} = \hat{\mathbf{V}}_{x,t}^{-1}\widehat{\mathbf{GCov}}_{i,t}^{\mathcal{J}},$$

where $\widehat{\mathbf{GCov}}_{i,t}^{\mathcal{J}}$ denotes the $K \times 1$ vector of granular covariances across the K factors for the $\mathcal{J}$ multi-dimensional partition. That is, we compute the covariance between the test asset Y_i, and the k^{th} factor X_k, conditioning on the first factor return lying in $\mathcal{G}_{j_1}^1$, the second factor return lying in $\mathcal{G}_{j_2}^2$, and so on up to the K^{th} factor return lying in $\mathcal{G}_{j_K}^K$. Having estimated the multi-dimensional granular betas, mirroring the approach in Section 4.2, we then estimate risk premium functions for each of the granular betas based on the use of sieve approximations. Specifically, we use K-dimensional polynomials of the type $P(\boldsymbol{x}, p) = x_1^a \times x_2^b \times ... \times x_K^k$, where the order $a + b + ... + k$ is again determined by cross-validation.

If the factors are strongly correlated and G^K is "too large," there may be partitions with very few, or no observations. When a given partition has few observations the estimated granular betas will naturally be less precise, and their explanatory power for the cross-section of returns will be diminished.[20] Of course, if the factors are only weakly correlated, as is the case in the FF3 model analyzed below, this is less of a concern.

[20] This is akin to the problem that arise in Bryzgalova et al. (2025) in the context of the construction of informative "test assets," or portfolios.

Meanwhile, by explicitly relying on out-of-sample predictions to asses model performance across a range of values of G, if a given multi-dimensional partition does indeed lead to "too many" poorly estimated granular betas, this should be revealed empirically in that step of our estimation procedure.[21] Related, since the number of polynomials in the basis used for spanning the lambda functions increases much more rapidly as a function of the order in the multi-dimensional case, to help discipline the estimation we consider at most up to fourth-order polynomials. Moreover, to make the results based on the three-way partitions, in which we jointly partition the betas according to the market, size and value factors, directly comparable to the results based on one-way partitions for the same three factors, we fix the orders of G so that the total number of granular betas that appear on the right-hand side of the functional regressions used for estimating the corresponding lambda functions are the same.[22]

Looking first at Panel A in Table 3 and the reference results for the CAPM, we see that the out-of-sample R^2, which as previously reported in Table 2 increases from 3.1% to 3.8% when allowing for one-way partitions, further increases to 4.8% when we allow for three-way partitions based on the FF3 factors. In other words, by more finely measuring the local covariation with the market, we greatly improve on the ability of the CAPM-based model to explain the cross-sectional variation in the returns. Of course, a more direct approach would be to simply include the two additional factors and work directly with the FF3 three-factor model.

However, we already know from our previous results that the FF3 model can also be further improved by considering one-dimensional partitions, with the R^2 increasing from 4.4% to 5.3%. In contrast to the results for the CAPM, however, looking at the new results for the FF3 model in Panel B in Table 3, there appears to be no additional gains available by allowing for three-dimensional partitions. In fact, the average OOS R^2

[21]As a case in point, we find that five-way partitions of the traditional CAPM betas based on the five FF5 factors are simply too "noisy" to be useful for out-of-sample forecasting; for additional details see the Online Supplemental Appendix.

[22]Recall that G one-way partitions of K factors leads to $K \times G$ granular betas, while K-way partitions leads to $K \times G^K$ granular betas.

Table 3: **One-way versus three-way partitions for granular betas.** This table presents a comparison of the out-of-sample R^2s for the CAPM and the FF3 models obtained using one-way (left panel) or three-way (right panel) partitions. The order of the polynomials (G) in each row are selected so that the total number of granular betas (#Gβ) is the same for the one-way and three-way partitions.

	One-way partitions			*Three-way partitions*		
#Gβ	G	Opt. order	R^2	G	Opt. order	R^2
Panel A: CAPM						
1	1	0	3.145	1	0	3.145
8	8	6	3.693	2	2,2,2	4.834
64	64	8	3.840	4	2,2,2	4.789
Panel B: FF3						
3	1	0	4.447	1	0	4.447
24	8	2,2,2	5.236	2	2,2,2	5.213
192	64	3,6,4	5.275	4	2,2,2	5.220

decreases just slightly from 5.3% to 5.2%.

In sum, while the simple one-factor CAPM can be improved by considering multi-dimensional granular betas based on the market, size and value factors, for larger multi-factor models the additional estimation error associated with multi-dimensional partitions seem to outweigh the gains, so that for multi-factor models the simpler granular betas based on one-dimensional partitions generally result in the best performing models.[23]

5.3. Long-short granular beta portfolios

To further underscore the practical value of the granular betas, we now consider a set of simple granular long-short portfolio strategies. Putting the new strategies into perspective, arguably the oldest of all factor related anomalies, dating back to Black, Jensen and Scholes (1972), holds that the empirical relationship between conventional market betas and returns is too flat to be explained by the standard CAPM. Accordingly, portfolios that "bet against beta" by going long in high-beta stocks, and short in low-

[23]These same conclusions are also corroborated by the additional empirical results for the FF3+Mom and FF5 factor models reported in the Online Supplemental Appendix.

beta stocks, tend to earn positive alphas (see, also Frazzini and Pedersen, 2014). In the granular beta setting, however, stock selection cannot simply depend on the risk exposures, but must also take into account the differences in the premiums earned for the different granular risk exposures. Hence, rather than sorting the stocks based on their granular betas, we instead sort the stocks based on their predicted expected returns implied by the granular betas together with the corresponding estimated risk premium functions.

In particular, averaging across the expected return functions for each of the factors, we calculate the expected return for month t and stock i as:

$$\sum_{k=1}^{K}\sum_{j=1}^{G}\widehat{G\beta}_{t-60:t-1,i,k,j}\hat{\lambda}_{t-60:t-1}(j,G),$$

where we rely on a 60-month moving average for both the betas and the risk premium functions. For the sake of brevity, we again focus on the FF3 model, corresponding to $K = 3$. In addition to the expected return predictions from the full model, we also consider portfolios based on the expected return predictions stemming from each of the three individual factors in turn. Specifically, for stock i and factor k, we calculate:

$$\sum_{j=1}^{G}\widehat{G\beta}_{t-60:t-1,i,k,j}\hat{\lambda}_{t-60:t-1}(j,G),$$

where we continue to rely on the lambda functions estimated for the full FF3 model.

Armed with the estimated expected returns for each of the stocks and months in the sample, we then construct long-short portfolios, in which the long leg consists of an equal-weighted portfolio of the 20% of the stocks with the highest expected returns for a given month, and the short-leg consists of an equal-weighted portfolio of the 20% of the stocks with the lowest monthly expected returns.[24] We consider $G = 1, 2$ and 64, representing the traditional FF3 model, the up- and downside version of the FF3 model,

[24]For $G = 1$ and a single factor equal to the market, this naturally reduces to a simple "betting on beta" type portfolio.

as well as the most flexible granular version of the FF3 model considered in Section 4.2 above.

The resulting portfolio performance is summarized in Table 4. Panel A reports the annualized mean and standard deviation of the portfolio returns, along with their Sharpe ratios. We also report the t-statistics for testing whether the Sharpe ratios for the granular-based portfolios are significantly higher than their $G = 1$ counterparts based on the test of Ledoit and Wolf (2008). Panel B in turn provides the corresponding alphas for each of the portfolios with respect to the traditional CAPM and the linear FF3, FF3+Mom and FF5 benchmark models, together with Newey-West robust t-statistics in parentheses.

As the table shows, the granular beta-based portfolios generally result in significantly higher Sharpe ratios. Since the portfolios are explicitly constructed to generate higher returns, the higher Sharpe ratios are almost exclusively driven by higher mean returns. The overall highest average realized returns and Sharpe ratio are, not surprisingly, achieved by the joint FF3 granular specification that utilizes the granular betas and risk premium functions for all three factors to jointly predict the expected returns. Looking at the results for each of the individual factors further shows that stock selection based on the granular betas and expected return function estimates for the MKT factor offer the largest (relative) single-factor improvements, followed by the SMB factor. This is consistent with the visual impression from the earlier Figure 3 and the corresponding tests in Table 1 indicating that the market risk premium function is the furthest from being flat.

These findings are corroborated by the alpha estimates reported in Panel B. For the $G = 1$ portfolios based on the standard linear models and factors, only a single of the sixteen different alphas is significant at the usual 5% level. This, of course, is not surprising as the standard portfolios are formed based on the same factors that define the benchmark models. Meanwhile, consistent with the differential pricing of up- and down-side market risk, the $G = 2$ portfolios based solely on the MKT factor result in significant alphas with respect to both the FF3 and FF3-Mom benchmark models.

Table 4: **Long-short granular beta portfolios** The top panel reports annualized descriptive statistics for long-short granular strategies. The columns marked MKT, HML and SMB, refer to portfolios based on a single factor, while the column marked Joint account for exposures to all three factors jointly. All of the portfolios are self-financing and rebalanced monthly based on a long-position in the 20% of stocks with the highest expected returns, and a short-position in the 20% of stocks with the lowest expected returns. The bottom panel reports the intercept and Newey-West robust t-statistics of time-series regressions for the standard CAPM, FF3, FF3+Mom and FF5 benchmark models. The alphas are reported in annualized percentage terms.

	MKT			SMB			HML			Joint		
G	1	2	64	1	2	64	1	2	64	1	2	64
Panel A: Descriptives												
Mean	0.601	2.831	3.898	1.972	2.725	3.324	0.133	0.103	1.022	2.669	2.862	3.909
StDev	15.81	15.06	14.48	11.41	11.69	12.06	11.32	11.15	11.99	13.95	13.99	13.98
Sharpe	0.038	0.188	0.269	0.173	0.233	0.276	0.012	0.009	0.085	0.191	0.205	0.280
t-stat		(1.927)	(2.923)		(1.927)	(2.110)		(-0.110)	(2.189)		(0.450)	(2.167)
Panel B: Alphas												
CAPM	-0.394	1.671	2.766	1.111	2.073	2.527	0.723	0.224	1.636	1.499	1.718	2.784
t-stat	(-0.346)	(1.549)	(2.674)	(1.426)	(2.574)	(3.016)	(0.874)	(0.274)	(1.856)	(1.472)	(1.659)	(2.408)
FF3	0.651	2.850	3.953	0.046	0.890	1.162	0.518	0.049	1.123	1.245	1.629	2.638
t-stat	(0.576)	(2.636)	(3.830)	(0.060)	(1.141)	(1.435)	(0.580)	(0.056)	(1.222)	(1.157)	(1.505)	(2.281)
FF3+Mom	1.440	3.825	4.668	-0.128	0.710	1.243	1.530	1.272	2.504	3.323	3.571	4.678
t-stat	(1.286)	(3.638)	(4.593)	(-0.163)	(0.891)	(1.525)	(1.722)	(1.450)	(2.728)	(3.222)	(3.405)	(4.331)
FF5	-0.218	1.839	3.229	-0.514	0.742	1.225	0.285	-0.258	1.042	1.356	1.812	2.715
t-stat	(-0.169)	(1.466)	(2.767)	(-0.642)	(0.874)	(1.431)	(0.304)	(-0.278)	(1.063)	(1.172)	(1.491)	(2.119)

However, going one step further and more fully exploiting the granular information and potentially hidden compensation across the support of the different factors not captured by the traditional linear factor models, the $G = 64$ Joint and MKT-based portfolios both result in significant positive alphas with respect to *all* of the benchmark models.

6. Conclusion

Linear factor models remain the workhorse for understanding risk exposures and risk premiums in financial markets. In such models, risk exposures are traditionally measured using simple covariances, or betas, with respect to a given set of factors, with the risk premiums estimated as linear functions of said exposures. Instead, we propose a new approach for boosting the information that can be extracted from existing factors, without the need for any additional data. Drawing on earlier work emphasizing the differential pricing of up- and down-side betas, we propose and estimate new local measures of dependence. These new measures, which we dub granular betas, allow for more refined characterizations of the intrinsic dependencies between an asset and a risk factor. Exploiting the cross-sectional variation in the resulting granular beta functions, we also estimate new risk premium functions. These functional estimates generalize the scalar risk premium estimates from standard factor models and methods, and allow us to precise where, in the support of the factor space, risk premiums are truly earned. Implementing the new procedures with a large cross-section of individual U.S. equity returns, we find that the fit of traditional factor models can be significantly improved by using our new more flexible granular measures.

Our new approach opens up many other empirical and theoretical questions. For instance, the more nuanced depiction of risk exposures provided by the granular betas hints at the opportunity for improved risk management practices by more precisely targeting specific factor risks of concern. Related, new and "smarter beta" investment strategies geared toward specific parts of the factor space may also be possible. Our strong empirical evidence against flatness and symmetry for some of the factor risk premium functions also naturally calls into question what are the economic mechanisms and/or market frictions responsible for these differences in compensation? We leave the answer to all of these tantalizing questions for future research.

Appendix A. The risk premium function under the CAPM

As discussed in Section 3.1, if the CAPM holds and expected returns satisfy $\mathbb{E}[Y_{i,t}|X_t] = \beta_i X_t$, the granular beta function must be a parabola, as defined in equation (9). To appreciate the implications of this for the risk premium function that obtains under the CAPM, it is instructive to decompose the function into a "flat" term, and a term that captures deviations from "flatness," say $\lambda(x) = \mu_X + \ddot{\lambda}(x)$. Equating the expected return on asset i under the CAPM, $\mu_i = \beta_i \mu_X$, with that implied by our functional regression, then yields:

$$\begin{aligned} \mu_i &= \alpha + \mathbb{E}[\lambda(X)\mathrm{G}\beta_i^*(X)] \\ &= \alpha + \mu_X \mathbb{E}[\mathrm{G}\beta_i^*(X)] + \mathbb{E}[\ddot{\lambda}(X)\mathrm{G}\beta_i^*(X)] \\ &= \alpha + \beta_i \mu_X + \beta_i \mathbb{E}\left[\ddot{\lambda}(X)\left(\frac{X-\mu_X}{\sigma_X}\right)^2\right], \end{aligned} \tag{A.1}$$

where the third line uses the fact that $\mathbb{E}[\mathrm{G}\beta_i^*(X)] = \beta_i$ together with the functional form for $\mathrm{G}\beta_i^*(x)$ given in equation (9). Considering a zero-beta asset, it readily follows that $\alpha = 0$. Further, under the CAPM we have $\mu_i = \beta_i \mu_X \ \forall\ i$, and so the third term in equation (A.1) must equal zero for all i. Since β_i may differ from zero, the $\mathbb{E}[\cdot]$ term must therefore be identically equal to zero. Since the squared standardized market return is weakly positive, this can only hold if $\ddot{\lambda}(x) = 0 \ \forall\ x$. In other words, under the CAPM, the intercept α in the regression must be zero, and the lambda function must be flat, and equal to the expected return on the market for all x, that is $\lambda(x) = \mu_X \ \forall\ x$.

Appendix B. Symmetrized quantiles

In order for simple parameter restrictions on the Legendre polynomial terms to be sufficient to ensure symmetry, we need to rely on partition boundaries that are symmetric around zero. If the distributions of the factor returns were known to be symmetric, we could simply use quantiles of these returns. If, however, the factor returns are asymmetrically distributed, as is the case in our empirical analyses, then an alternative set of boundaries are needed. Let $G > 1$ denote the number of desired partitions, and assume that G is even, as is the case in our empirical analyses. We then use the quantiles of the absolute factor returns, defined by $\tilde{q}_\tau = Q^\tau[|X_t|]$, for $\tau = 1/H,\ 2/H,\ ...,\ (H-1)/H$ and $H \equiv G/2$, resulting in the symmetrized $G-1$ partition boundaries, $\left[-\tilde{q}_{(H-1)/H}, ..., -\tilde{q}_{1/H}, 0, \tilde{q}_{1/H}, \tilde{q}_{(H-1)/H}\right]$.

Appendix C. Traditional risk premium estimates

Table C.1: **Annualized Fama-Macbeth risk premium estimates.** The table reports the annualized risk premium estimates obtained for the traditional CAPM, FF3, FF3+Mom and FF5 models over the full sample (denoted G=1 in the table), along with the corresponding estimates for the up and down versions of the models (denoted G=2 in the table). Robust *t*-statistics are reported in parentheses.

	G = 1					G = 2			
	CAPM	FF3	FF3+Mom	FF5		CAPM	FF3	FF3+Mom	FF5
MKT	3.765	3.039	3.068	1.940	MKT^-	26.662	16.583	25.934	15.194
	(1.531)	(1.312)	(1.327)	(0.845)		(2.932)	(1.757)	(2.306)	(1.577)
					MKT^+	-19.167	-9.267	-20.208	-8.613
						(-2.417)	-(0.900)	(-1.703)	(-0.814)
SMB		4.249	3.829	4.621	SMB^-		6.415	3.362	6.575
		(3.227)	(3.119)	(3.568)			(1.288)	(0.672)	(1.195)
					SMB^+		0.821	2.502	1.113
							(0.143)	(0.403)	(0.179)
HML		1.515	0.313	1.865	HML^-		5.285	2.735	3.773
		(1.050)	(0.238)	(1.329)			(1.016)	(0.469)	(0.699)
					HML^+		-2.021	-1.163	0.587
							(-0.389)	(-0.190)	(0.099)
MOM			-3.226		MOM^-			-6.333	
			(-1.248)					(-0.792)	
					MOM^+			-4.601	
								(-0.628)	
RMW				0.148	RMW^-				-2.409
				(0.127)					(-0.669)
					RMW^+				3.380
									(0.863)
CMA				0.701	CMA^-				0.789
				(0.569)					(0.192)
					CMA^+				1.925
									(0.437)
Constant	6.673	6.030	5.878	7.207	Constant	6.375	5.680	5.721	6.020
	(2.889)	(2.873)	(2.782)	(3.235)		(2.798)	(2.722)	(2.704)	(2.659)

Appendix D. Hyperparameters

Table D.2: **Hyperparameter selection.** The table reports the polynomial orders used for estimating the risk premium functions for the different models, together with the shrinkage intensities towards the flat estimates ($G = 1$). The hyperparameters were determined by 60/20/20 cross-validation, as discussed in the main text.

G	Unrestricted	Functional						
		Order						
	ω	MKT	SMB	HML	MOM	RMW	CMA	ω
				CAPM				
1	0.28	0						0.63
2	0.57	1						0.56
4	0.70	3						0.70
8	0.76	6						0.75
16	0.82	8						0.76
32	0.86	8						0.76
64	0.93	8						0.75
				FF3				
1	0.04	0	0	0				0.10
2	0.66	1	1	1				0.66
4	0.78	3	2	3				0.77
8	0.85	2	2	2				0.71
16	0.91	3	4	4				0.76
32	0.97	3	6	4				0.78
64		3	6	4				0.78
				FF3+Mom				
1	0.07	0	0	0	0			0.20
2	0.72	1	1	1	1			0.72
4	0.83	3	2	3	2			0.82
8	0.89	3	2	2	2			0.80
16	0.94	6	4	3	3			0.82
32		3	6	4	3			0.82
64		2	2	2	3			0.78
				FF5				
1	0.02	0	0	0		0	0	0.10
2	0.78	1	1	1		0	0	0.72
4	0.86	3	2	3		2	2	0.84
8		2	4	6		2	4	0.86
16		6	2	4		5	5	0.86
32		3	2	3		5	4	0.84
64		3	2	3		5	4	0.84

References

Adrian, T., Crump, R.K., Vogt, E., 2019. Nonlinearity and flight-to-safety in the risk-return trade-off for stocks and bonds. Journal of Finance 74, 1931–1973.

Ahn, D.H., Min, B.K., Yoon, B., 2019. Why has the size effect disappeared? Journal of Banking and Finance 102, 256–276.

Aït-Sahalia, Y., Jacod, J., Xiu, D., 2025. Continuous-time Fama-MacBeth regressions. Working paper, Princeton University .

Aleti, S., Bollerslev, T., Siggaard, M., 2025. Intraday market return predictability culed from the factor zoo. Management Science forthcoming.

Almeida, C., Freire, G., 2023. Which (nonlinear) factor models? Working Paper, Princeton University .

Andersen, T.G., Bollerslev, T., Diebold, F.X., Wu, G., 2006. Realized beta: Persistence and predictability, in: Fomby, T.B., Terrell, D. (Eds.), Advances in Econometrics: Econometric Analysis of Economic and Financial Time Series. Elsevier Science, Amsterdam. volume 20, pp. 1–40.

Andrews, D.W., 1991. Asymptotic normality of series estimators for nonparametric and semiparametric regression models. Econometrica 59, 307–345.

Ang, A., Chen, J., Xing, Y., 2006. Downside risk. Review of Financial Studies 19, 1191–1239.

Babii, A., Carrasco, M., Tsafack, I., 2024. Functional partial least-squares: Optimal rates and adaptation. Working Paper, UNC-Chapel Hill and University of Montreal .

Banz, R.W., 1981. The relationship between return and market value of common stocks. Journal of Financial Economics 9, 3–18.

Barndorff-Nielsen, O.E., Kinnebrock, S., Shephard, N., 2010. Measuring downside risk: Realised semivariance, in: Bollerslev, T., Russell, J.R., Watson, M.W. (Eds.), Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle. Oxford University Press, Oxford, pp. 117–136.

Barndorff-Nielsen, O.E., Shephard, N., 2004. Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics. Econometrica 72, 885–925.

Baruník, J., Nevrla, M., 2023. Quantile spectral beta: A tale of tail risks, investment horizons, and asset prices. Journal of Financial Econometrics 21, 1590–1646.

Beason, T., Schreindorfer, D., 2022. Dissecting the equity premium. Journal of Political Economy 130, 2203–2222.

Bégin, J.F., Dorion, C., Gauthier, G., 2020. Idiosyncratic jump risk matters: Evidence from equity returns and options. Review of Financial Studies 33, 155–211.

Bianchi, D., Büchner, M., Tamoni, A., 2021. Bond risk premiums with machine learning. Review of Financial Studies 34, 1046–1089.

Black, F., Jensen, M., Scholes, M., 1972. The capital asset pricing model: Some empirical tests, in: Jensen, M. (Ed.), Studies in the Theory of Capital Markets. Praeger, New York, pp. 79–121.

Bollerslev, T., Li, J., Patton, A.J., Quaedvlieg, R., 2020. Realized semicovariances. Econometrica 88, 1515–1551.

Bollerslev, T., Li, S.Z., Todorov, V., 2016. Roughing up beta: Continuous versus discontinuous betas and the cross section of expected stock returns. Journal of Financial Economics 120, 464–490.

Bollerslev, T., Medeiros, M.C., Patton, A.J., Quaedvlieg, R., 2022a. From zero to hero: Realized partial (co)variances. Journal of Econometrics 231, 348–360.

Bollerslev, T., Patton, A.J., Quaedvlieg, R., 2022b. Realized semibetas: Disentangling "good" and "bad" downside risks. Journal of Financial Economics 144, 227–246.

Bollerslev, T., Todorov, V., 2011. Tails, fears, and risk premia. Journal of Finance 66, 2165–2211.

Bryzgalova, S., Pelger, M., Zhu, J., 2025. Forest through the trees: Building cross-sections of stock returns. Journal of Finance forthcoming.

Cai, Z., Fang, Y., Xu, Q., 2022. Testing capital asset pricing models using functional-coefficient panel data models with cross-sectional dependence. Journal of Econometrics 227, 114–133.

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

Carrasco, M., Florens, J.P., Renault, E., 2007. Linear inverse problems in structural econometrics estimation based on spectral decomposition and regularization, in: Heckman, J., Leamer, E. (Eds.), Handbook of Econometrics. Elsevier. volume 6B. chapter 77.

Chabi-Yo, F., Huggenberger, M., Weigert, F., 2022. Multivariate crash risk. Journal of Financial Economics 145, 129–153.

Chabi-Yo, F., Ruenzi, S., Weigert, F., 2018. Crash sensitivity and the cross section of expected stock returns. Journal of Financial and Quantitative Analysis 53, 1059–1100.

Chen, L., Pelger, M., Zhu, J., 2024. Deep-learning in asset pricing. Management Science 70, 714–750.

Chen, X., 2007. Large sample sieve estimation of semi-nonparametric models, in: Heckman, J., Leamer, E. (Eds.), Handbook of Econometrics. 1 ed.. Elsevier. volume 6B. chapter 76.

Chen, X., Christensen, T.M., 2015. Optimal uniform convergence rates and asymptotic normality for series estimators under weak dependence and weak conditions. Journal of Econometrics 188, 447–467.

Chen, X., Christensen, T.M., 2018. Optimal sup-norm rates and uniform inference on nonlinear functionals of nonparametric IV regression. Quantitative Economics 9, 39–84.

Cochrane, J., 2017. Macro-finance. Review of Finance 21, 945–985.

Cochrane, J.H., 2011. Presidential address: Discount rates. Journal of Finance 66, 1047–1108.

Cohen, A., Jones, R.H., 1969. Regression on a random field. Journal of the American Statistical Association 64, 1172–1182.

Colacito, R., Ghysels, E., Meng, J., Siwasarit, W., 2016. Skewness in expected macro fundamentals and the predictability of equity returns: Evidence and theory. Review of Financial Studies 29, 2069–2109.

Cong, L.W., Feng, G., He, J., He, X., . Growing the efficiant frontier on panel trees. Journal of Financial Economics 167, 104024.

Conner, G., Hagmann, M., Linton, O.B., 2012. Efficient semiparametric estimation of the Fama-French model and extensions. Econometrica 80, 713–754.

Conrad, J., Dittmar, R.F., Ghysels, E., 2013. Ex ante skewness and expected stock returns. Journal of Finance 68, 85–124.

Cooper, M.J., Gutierrez, Jr., R.C., Hameed, A., 2004. Market states and momentum. Journal of Finance 59, 1345–1365.

Didisheim, A., Ke, S., Kelly, B., Malamud, S., 2024. Complexity in factor pricing models. Working paper, Yale University .

Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. Journal of Business and Economic Statistics 13, 253–265.

Dittmar, R.F., 2002. Nonlinear pricing kernels, kurtosis preference, and evidence from the cross section of equity returns. Journal of Finance 57, 369–403.

Driessen, J., Ebert, S., Koëter, J., 2022. π-CAPM: The classical CAPM with probability weighting and skewed assets. Working paper, Tilburg University .

Fama, E.F., French, K.R., . The value premium. Review of Asset Pricing Studies 11, 105–121.

Fama, E.F., French, K.R., 1992. The cross-section of expected stock returns. Journal of Finance 47, 427–465.

Fama, E.F., French, K.R., 2015. A five-factor asset pricing model. Journal of Financial Economics 116, 1–22.

Fan, J., Gijbels, I., 1992. Variable bandwidth and local regression smoothers. Annals of Statistics 20, 2008–2036.

Fan, J., Ke, Z., Liao, Y., Neuhierl, A., 2023. Structural deep learning in conditional asset pricing. Working paper, Princeton University .

Farago, A., Tedongap, R., 2018. Downside risks and the cross-section of asset returns. Journal of Financial Economics 129, 69–86.

Frazzini, A., Pedersen, L.H., 2014. Betting against beta. Journal of Financial Economics 111, 1–25.

Freyberger, J., Neuhierl, A., Weber, M., 2020. Dissecting characteristics nonparametrically. Review of Financial Studies 33, 2326–2377.

Gagliardini, P., Ossola, E., Scaillet, O., 2016. Time-varying risk premium in large cross-sectional equity data sets. Econometrica 84, 985–1046.

Gârleanu, N., Panageas, S., Zheng, G., 2025. A long and a short leg make for a wobbly equilibrium. Working paper, Washington University in St.Louis .

Ghysels, E., Plazzi, A., Valkanov, R., 2016. Why invest in emerging markets? the role of conditional return asymmetry. Journal of Finance 71, 2145–2191.

Ghysels, E., Santa-Clara, P., Valkanov, R., 2006. Predicting volatility: How to get most out of returns data sampled at different frequencies. Journal of Econometrics 131, 59–95.

Ghysels, E., Sinko, A., Valkanov, R., 2007. MIDAS regressions: Further results and new directions. Econometric Reviews 26, 53–90.

Gregory, C., Linton, O.B., 2007. Semiparametric estimation of characteristic-based factor model of stock return. Journal of Empirical Finance 14, 694–717.

Gu, S., Kelly, B., Xiu, D., 2020. Empirical asset pricing via machine learning. Review of Financial Studies 33, 2223–2273.

Gu, S., Kelly, B., Xiu, D., 2021. Autoencoder asset pricing models. Journal of Econometrics 222, 429–450.

Gul, F., 1991. A theory of disappointment aversion. Econometrica 59, 667–686.

Hall, P., Horowitz, J.L., 2007. Methodology and convergence rates for functional linear regression. Annals of Statistics 35, 70–91.

Harvey, C.R., Siddique, A., 2000. Conditional skewness in asset pricing tests. Journal of Finance 55, 1263–1295.

Hong, H., Sraer, D.A., 2016. Speculative betas. Journal of Finance 71, 2095–2144.

Horowitz, J.L., Lee, S., 2017. Nonparametric estimation and inference under shape restrictions. Journal of Econometrics 201, 108–126.

James, G.M., Wang, J., Zhy, J., 2009. Functional linear regression that's interpretable. Annals of Statistics 37, 2083–2108.

Jurado, K., Ludvigson, S.C., Ng, S., 2015. Measuring uncertainty. American Economic Review 105, 1177–1216.

Kahneman, D., Tversky, A., 1979. Prospect theory: An analysis of decision under risk.

Econometrica 47, 263–291.

Kelly, B., Jiang, H., 2014. Tail risk and asset prices. Review of Financial Studies 27, 2841–2871.

Kelly, B., Malamud, S., Zhou, K., 2024. The virtue of complexity in return prediction. Journal of Finance 79, 429–503.

Kelly, B., Pruitt, S., Su, Y., 2019. Characteristics are covarinaces: A unified model of risk and return. Journal of Financial Economics 134, 501–524.

Kimball, M., 1993. Standard risk aversion. Econometrica 64, 589–611.

Kozak, S., Nagel, S., Santosh, S., 2020. Shrinking the cross-section. Journal of Financial Economics 135, 271–292.

Langlois, H., 2020. Measuring skewness premia. Journal of Financial Economics 135, 399–424.

Ledoit, O., Wolf, M., 2008. Robust performance hypothesis testing with the sharpe ratio. Journal of Empirical Finance 15, 850–859.

Levi, Y., Welch, I., 2020. Symmetric and asymmetric market betas and downside risk. Review of Financial Studies 33, 2772–2795.

Li, J., Liao, Z., Quaedvlieg, R., 2022. Conditional superior predictive ability. Review of Economic Studies 89, 843–875.

Lintner, J., 1965. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. Review of Economics and Statistics 47, 13–37.

Lu, Z., Murray, S., 2019. Bear beta. Journal of Financial Economics 131, 736–760.

Mammen, E., 1991a. Estimating a smooth monotone regression function. Annals of Statistics 19, 724–740.

Mammen, E., 1991b. Nonparametric regression under qualitative smoothness assumptions. Annals of Statistics 19, 741–759.

Massacci, D., Sarno, L., Trapani, L., 2025. Factor models of asset returns and bear market risk. Working paper, King's College London .

Moench, E., Stein, T., 2021. Equity premium predictability over the business cycle. Working paper, Frankfurt School of Finance and Management .

Newey, W.K., 1997. Convergence rates and asymptotic normality for series estimators. Journal of Econometrics 79, 147–168.

Newey, W.K., West, K.D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.

Patton, A.J., Sheppard, K., 2015. Good volatility, bad volatility: Signed jumps and the persistence of volatility. Review of Economics and Statistics 97, 683–397.

Patton, A.J., Verardo, M., 2012. Does beta move with news? Firm-specific information flows and learning about profitability. Review of Financial Studies 25, 2789–2839.

Pelger, M., Xiong, R., 2022. State-varying factor models of large dimensions. Journal of Business and Economic Statistics 40, 1315–1333.

Pontiff, J., 1996. Costly arbitrage: Evidence from closed-end funds. Quarterly Journal of Economics 111, 1135–1151.

Ramsey, J.O., Silverman, B.W., 2005. Functional Data Analysis, 2nd Ed. Springer Verlag.

Raponi, V., Robotti, C., Zaffaroni, P., 2020. Testing beta-pricing models using large cross-sections. Review of Financial Studies 33, 2796–2842.

Reinganum, M.R., 1981. Misspecification of asset pricing: Empirical anomalies based on earnings yields and market values. Journal of Financial Economics 9, 19–46.

Rosenberg, B., 1974. Extra-market components of covariance in security prices. Journal of Financial and Quantitative Analysis 9, 263–274.

Schleifer, A., Vishny, R.W., 1997. The limits of arbitrage. Journal of Finance 52, 35–55.

Schneider, P., Wagner, C., Zechner, J., 2020. Low-risk anomalies? Journal of Finance 75, 2673–2718.

Schwert, G.W., 2003. Anomalies and market efficiency, in: Constantinides, G.M., Harris, M., Stulz, R.M. (Eds.), Handbook of the Economics of Finance. North Holland.

Sharpe, W.F., 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. Journal of Finance 19, 425–442.

Souza, T.d.O., 2020. Price of risk fluctuation and the size premium. Working paper, University of Southern Denmark .

Stoja, E., Polanski, A., Nguyen, L.H., 2025. The taxonomy of tail risk. Journal of Financial Research forthcoming.

Todorov, V., Bollerslev, T., 2010. Jumps and betas: A new framework for disentangling and estimating systematic risks. Journal of Econometrics 157, 220–235.

Wachter, J.A., 2013. Can time-varying risk of rare disasters explain aggregate stock market volatility? Journal of Finance 68, 987–1035.