Does CAPM overestimate more the risk or its price?

Fabio Franceschini*

23rd November 2023

Last version

Abstract

CAPM is the most foundational model in finance, but empirically underestimates expected returns of low-risk assets and overestimates those of high risk. This paper first theoretically decomposes this anomaly into the effects of tight financial constraints and of risk factors omission, and then empirically assesses the contribution of each channel in explaining the anomaly. The decomposition highlights a counteracting effect between the two channels, which makes it more relevant to study them jointly. Empirically, it is found that risk factors other than the market and agnostically extracted from test assets end up explaining all of the predicted return of the BAB portfolio as formed by Frazzini and Pedersen (2014) and two thirds of the BAB portfolio formed as in Novy-Marx and Velikov (2022), although the latter is not significantly different from an equal contribution of 50%. Nevertheless, the spread on zero-beta assets originated in leverage constraints proves being significant statistically and economically, at 2% per annum.

Keywords: Asset Pricing, CAPM, Leverage Constraints, Omitted Factors

JEL Codes: G11, G12, G20

1 Introduction

The Capital Asset Pricing Model (CAPM) introduced by Sharpe (1964) has been the first attempt to provide a theoretical framework that relates remuneration and risk in financial markets. It states that the return of every asset with an uncertain payoff is expected to exceed the risk-free rate by a measure of the asset's undiversifiable riskiness times a market-wide unique premium per unit of risk. The only driver of cross-sectional variation in expected returns is then the assets' riskiness, which the CAPM measures with *beta* – the sensitivity of an asset's returns to the market returns, while the common *risk premium* amounts to the expected excess return of the market.[1] Its simplicity contributed to making it the

*Department of Economics, University of Bologna (Italy); f.franceschini@unibo.it.
I thank Martín Gonzalez-Eiras and Svetlana Bryzgalova and for discussions from which I greatly benefited as well as seminar participants at the DSE PhD Forum for their valuable comments.

[1] This follows from assuming perfect markets and variance being the only statistical property of wealth affecting agents' utility besides its first moment; then the market portfolio is the optimal risky holding for any agent because it is the most diversified portfolio possible. From this follows that: first, the risk of every asset is intended as the increase in portfolio variance associated with a greater holding of it, which depends on how much it covaries with the market; second, that such risk addition is compensated competitively just as the variance of the optimal portfolio is, i.e. by the market expected excess return.

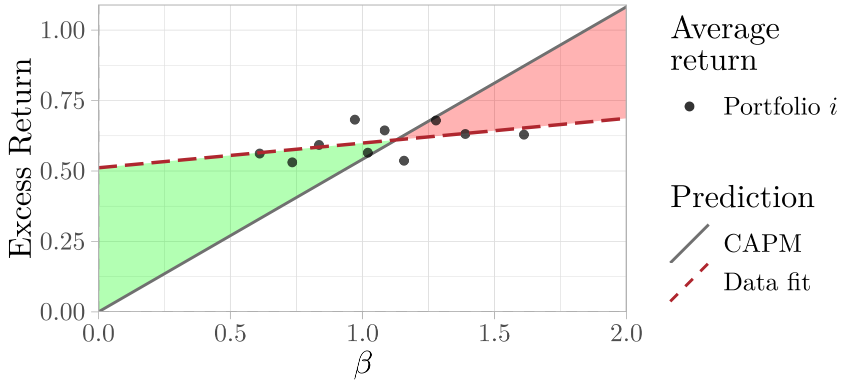

Figure 1: 10 US stocks portfolios. Monthly data, 1963-2019, source: K. French data library.

most seminal model in the field of financial economics and still makes it one of the most common models of use by practitioners to estimate the cost of capital as well as by teachers to introduce the subject.[2] Nonetheless, such simplicity has a cost in terms of empirical performance, making CAPM often a simple benchmark to more advanced models.

One of its most studied fallacies is the empirical regularity of assets' average excess returns growing with the respective beta by less than what implied by the average market excess return in the same period. Picture 1 summarizes this finding graphically for a few US stocks portfolios: it can be seen that the security market line (SML), which describes the relation between risk and remuneration, seems actually flatter than predicted by CAPM. This leads to low-beta portfolios to outperform CAPM predictions and high-beta ones to underperform them, making it in principle possible to build a self-financing portfolio that earns a positive return on average despite having no exposure to the market, i.e. no risk. This phenomenon commonly takes the name of 'low-risk anomaly' and was first shown by Miller and Scholes (1972). Black et al. (1972) immediately noted that from a purely statistical perspective the empirical SML can be flatter than predicted by CAPM because of the error-in-variable problem afflicting the cross-sectional premium estimation stage, which relies on previous estimates of the individual-asset betas as measured regressors. They proposed an easy and effective fix, currently employed by the majority of the studies on the topic: grouping assets in portfolios – but the anomaly survived to these days and has proved being way more pervasive than US stocks, as shown by Baker et al. (2014) and Frazzini and Pedersen (2014). More empirical concerns about estimation of the CAPM still stand as historical averages could be not a good measure of expectations, as highlighted by Elton (1999), and the overall validity of the CAPM when the market portfolio is unobservable is possible but still has no inferential theory behind it to make it viable, as shown in Guermat (2014). Anyway, while pressing, addressing statistical issues with the implementation of CAPM is not a primary concern of this paper, which focuses on the potential economic drivers of the anomaly.

The low-risk anomaly can essentially be rationalized (1) by a lower premium for the beta risk, which would remain the only source of cross-sectional variation in expected returns, or

[2]See Pinto et al. (2019) and Brealey et al. (2022).

(2) by properties of assets other than their CAPM beta being priced besides it. Specifically, in the latter case, assuming the additional property is associated with a positive premium, it needs to be distributed among assets inversely of how the CAPM beta is, to counter-act the increase in risk premium of higher beta assets and produce a flatter SML.[3] The coexistence of these two deviations from CAPM, however, is not trivial: the more the mispricing comes from high-beta assets being more desirable than the beta suggests (e.g. they systematically have low sensitivity to a second factor), the less the mispricing can be originated in a 'cut' on how much undesirability, such as market risk, is remunerated. In other words, the stronger is one argument, the weaker has to be the other. I study this, considering a simple model that features both effects and illustrating a representation that synthesizes the optimal pricing condition into the same two dimensions that entirely describe markets in CAPM: a single risk measure per individual stock, γ, and a premium associated to it that is shared by the whole market, which happens to be the expected market excess return minus a constant ψ. Then, I empirically measure the extent to which CAPM deviations from (a better approximation of) reality can be attributed to mis-measurement of risk, $\gamma - \beta$, and of its price, ψ. Given the non-trivial interaction between the two mechanisms, it is crucial to assess them jointly. Understanding the relevance of one against the other then helps answering the question 'in which way is one more likely to be wrong, when she looks at returns from the simplifying lens of CAPM?' Because of the economics behind the proposed explanations, the answer to this question actually has profound practical implications.

Black (1972) was the first to theoretically address the low-risk anomaly and showed that the addition of restricted borrowing in a standard economy leads to a flatter SML than CAPM prediction because it generates a spread between the zero-beta return with respect to a risk-free rate.[4] Later, Frazzini and Pedersen (2014) built on this idea and suggested to specifically be intermediaries' funding tightness in the form of binding leverage constraints to lower the market risk premium. Then, the low-risk anomaly would then emerge because the risk premium of high-beta assets is reduced to compensate a need of leverage to attain any level of expected payoffs that is lower than with low-beta counterparts, since their higher discount rates make them cheaper. They support this view by testing the implications of a dynamic and more flexible version of Black (1972) and showing a portfolio that embodies the mispricing, the so called 'Betting Against Beta' (BAB) factor, earning a positive return, not explained by other risk factors.[5]

[3]The study of the specific relation between additional assets' features and their beta is what distinguish the low-risk anomaly literature from the multi-factor one, whose successes and current state can be seen in Feng et al. (2020). The sensitivity to a second factor could be significantly priced besides the market-beta, but without evidence of a specific relation with assets' market-beta, nothing can be concluded on how it affects the spread between high market-beta assets and low market-beta assets.

[4]In such a framework the only way to form efficient portfolios that are riskier than the market would be taking a bigger position in the market itself and financing it by shorting a portfolio of risky assets, which is less remunerative per unit of additional portfolio volatility than financing it by borrowing at the risk-free rate. Then, as the market portfolio is a weighted average of constrained and unconstrained agents' portfolios, its expected returns have to be lower than the standard CAPM formulation because it is not maximally diversified.

[5]Further empirical support came in later, most notably from Lu and Qin (2021), Asness et al. (2020) and the studies therein cited.

The zero-beta interest rate has been previously studied, most recently by Di Tella et al. (2023), with strong evidence of a positive spread with respect to the risk-free rate. Notice that a zero-beta spread implies that *any* risk which increases discount rates, i.e. risk premia, and thus makes expected payoffs cheaper to be bought, should get a liquidity-motivated reduction in premium, just as the market risk premium does. This effect weakens a second set of explanations of the anomaly, related to all the assets' features whose impact on discount rates has been found to be inversely related to the assets' betas. For example: assets with high beta, which is associated to *higher* premia, also have higher residual coskewness, which is associated to *lower* premia. Then, if the premium for coskewness is smaller because of funding tightness, it will also explain less of the anomaly.[6] Residual coskewness was shown explaining the anomaly by Schneider et al. (2020), but there are more assets' features that have been shown having explanatory power with respect to the anomaly, such as past highest returns, idiosyncratic volatility and idiosyncratic skewness,[7] all of which Bali, Brown, et al. (2017) shown being strictly related to each other; size, shown in Novy-Marx and Velikov (2022); and extent of overpricing, which leads to higher sensitivity to disagreement, in Hong and Sraer (2016). A point that might be worth making is that also *sensitivity to* funding liquidity, considered for example in Lu and Qin (2021), is likely to enter this category, as high beta assets are expected to be positively related to funding shocks and vice-versa low beta stocks to be negatively related, which would make high beta stocks better hedges for funding shocks, lowering their discount rates, just like residual coskewness does. Nothing precludes for more features have a similar role despite not being yet discovered.

Baker et al. (2014) is a perfect example of why it is useful to understand how much of low-risk anomaly is to be imputed to funding motives versus omitted risk factors: they argue that stricter regulations raises banks' cost of capital building on the assumption that all of the anomaly is to be attributed to a low market risk premium and that the risk of banks is low. However, if omitted risks turned out to be the biggest driver of the anomaly, this conclusion would be challenged, as banks, and the other low-beta equities, would be more likely riskier than measured by CAPM and moving beyond CAPM would be *necessary* to make a proper assessment of how much would changes in the regulation affect their discount rates. Further, this can affect firms' choices regarding capital structure, as firms that are less constrained than the marginal investor can potentially gain from leveraging up. Indeed, if cash obtained at the market risk-free rate is used to pay back shareholders, equity will have a higher beta: in the Frazzini and Pedersen (2014) world the embedded leverage *would* be rewarded with a lower premium that would raise the residual equity valuation, which could be seen as remuneration for liquidity provision by the firm; if it is omitted risks to drive most

[6]The generalization is not trivial: in Frazzini and Pedersen (2014) the multiplier acts as a discount on the price, which means that the cut on a positive premium happens only if the asset that provides such a premium has a positive price. In the coskewness case, one could be buying either 'the market squared', which bring a *negative* premium at a positive price, or derivatives, which bring a *positive* premium at a positive price in the form of opportunity cost of immobilized margin capital. The result is that in the latter case the coskewness premium is reduced just like the market risk premium is, while in the former case the coskenwess premium is even *increased* by funding tighness.

[7]See Ang et al. (2006) and Brunnermeier et al. (2007).

of the anomaly, the residual equity *could* increase in value because of other characteristics gained besides the higher beta, but more elements may prevent this from happening or such characteristics might be undesired by the firm managers.[8] The empirical results show a statistically significant spread averaging above 2% annually even controlling for 8 risk factors, which is 4 times what Lu and Qin (2021) find and a fourth of Di Tella et al. (2023). Nonetheless, most of the BAB factor gets predicted by the risk omission rather than the liquidity spread – two thirds in the BAB factor that is robust to Novy-Marx and Velikov (2022) criticisms and all of it in the original BAB portfolio of Frazzini and Pedersen (2014).

The cornerstone of this analysis is the synthetic measure of risk, γ, which is defined as the ratio of an asset covariance with the marginal utility over the market covariance with the marginal utility. These covariances are the holy grail of the whole asset pricing literature and their estimation is clearly the most delicate step of any empirical analysis. It should be stressed that while it would obviously be ideal to observe the marginal utility directly, the goal of this work is not as presumptuous to effectively study the discrepancies of CAPM from reality, but rather from better representations of it. With the explicit intent to allow for the most agnostic and accurate estimation possible, utility does not have a specific functional form in the theoretical model, although it is pretty straightforward to move in that direction given the simplicity of the set-up. At the current state, the empirical analysis is performed assuming a factor structure in the returns, which justifies the use of a low-dimensionality approximation of the marginal utility projection on all the assets in the market. This, in turn allows for a convenient and contemporaneous estimation of the factors loadings and average premium spread ψ, with GMM. The factors are chosen to be the market, the market squared, since coskewness is a known to be relevant for the anomaly, and the principal components of a large pool of test assets orthogonalized with respect the first two known factors. In Appendix 6 I also outline an extension of Pukthuanthong et al. (2021) to estimate the time series of both marginal utility and ψ non parametrically, for which no formal inferential theory is available at this stage though.[9]

It must also be noted that my approach encompasses only the pricing information coming from the systematic components of assets features, which gets reflected in the pervasive risk factors forming the marginal utility: purely diversifiable behavioural motives are not captured. Nonetheless, the result shown in Schneider et al. (2020) of a coskewness factor capturing much of the idiosyncratic volatility pricing dispersion partly mitigates this concern. It should also be stressed that it is important to include additional and agnostic factors because, sacrificing identification of the risk source, there a lower risk of omitting relevant factors. This is important for two reasons, obtaining a better pricing performance and, more importantly, potentially capture risk factors that are not yet known to affect the relation

[8]For example, with coskewness the residual equity would essentially be valued more for a greater value of the option embedded in equity. It can be observed that whether high leverage actually results in greater value depends on multiple other elements, such as how levered the firm already is; and, more importantly, it is not obvious whether trying to increase equity value by being closer to bankruptcy is desirable for managers.

[9]A time series of the spread ψ would enable direct analysis of the relation between such spread and intermediaries liquidity measures.

beta-return, but they do. Indeed, I find that adding the principal components to market and market squared does impact the risk assessment of the beta-sorted portfolios, with the 'robust' BAB portfolio explanation changing from mostly due to funding motives to omitted risks.[10]

The bottom line is that financial imperfection, in the form of leverage constraints, decisively proved to matter, but from a quantitative perspective the omission of relevant factors is likely to impact more the discrepancies between CAPM and the data.

This study naturally enters the low-risk anomaly literature, providing evidence on how different approaches combine together. Specifically, it strongly supports the relevance of a spread in the zero-beta rate with respect to the risk-free rate in explaining the anomaly when directly facing coskewness contribution. It also provide guidance on how much information in risk factors besides market and coskewness can help in explaining the anomaly. Upon completion, relying on subsection 2.3, further evidence on the relation between the zero-beta rate and funding constraints should be provided. The most closely related papers are naturally Frazzini and Pedersen (2014) and Schneider et al. (2020), of which this is essentially an integrated framework. My work is also related to the work of Di Tella et al. (2023), which, focusing more on the macroeconomic implications, studies the zero-beta rate, but does so employing a wide set of *predetermined* factors, which the BAB factor of Frazzini and Pedersen (2014) proved surviving anyway. Indeed, including coskewness, which is a known factor impacting the anomaly reduces the magnitude of the funding multiplier in an appreciable way.

Further, this paper is related to research that studies imperfect financial markets and asses the impact of intermediaries' funding frictions, such as Adrian et al. (2014) and He et al. (2017), and more specifically quantifies the cost of liquidity tightness, such as Jylhä (2018), Lu and Qin (2021) and Du et al. (2022). In this case, the peculiarity is that the estimates comes directly from the cross-section of stocks, although at this stage no explicit test of the relation between the zero-beta spread and intermediaries' funding tightness is carried out (it will be).

Regarding the method, this work draws from two substantial literatures. One is composed by the studies employing Principal Component Analysis in asset pricing, which was pioneered by Chamberlain and Rothschild (1983) and saw a rise in popularity in recent years, together with other machine learning methods.[11] The second one is composed by the many studies employing GMM to estimate factor loadings of the Stochastic Discount Factor (SDF), such as Croce et al. (2023).

In section section 2 I review the theoretical set-up and how to measure the contributions of deviation in the zero-beta rate from the risk-free rate and the omission of relevant risks in explaining the low-risk anomaly; in section section 3 I show one way to make the analysis empirically feasible (paired with Appendix 6, where a more agnostic one is explored); in

[10] I will have to compare the estimation results with principal components as factors against a set of risk factors with a similar numerosity, such as the Fama-French 6 factors or the Ludvigson-Ng 9 macro-financial factors.

[11] For a review see Giglio et al. (2022).

section section 4 I show the estimates of the coefficients from a cross-sectional pricing exercise and the implied decomposition of the CAPM anomaly; section 5 concludes.

2 Theoretical set-up

2.1 A simple model with leverage constraints

On the lines of Frazzini and Pedersen (2014), consider a simple Overlapping-Generations model where there is a representative agent that is born at time t with wealth W_t, invests it, and finally consumes all of the payouts at time $t+1$, right before dying. Wealth at time t can only be invested in a risk-less asset with price 1, held in the amount X_t^0, and a set of S risky securities, held in the amount X_t^s for $s \in \{1, \dots, S\}$, with budget constraint

$$W_t = X_t^0 + \sum_{s=1}^{S} X_t^s P_t^s. \tag{1}$$

Wealth consumed in the following period will then be determined by the known and exogenously set gross risk-free return R_t^f and risky securities' prices, P_{t+1}^s, and dividends, D_{t+1}^s, by:

$$W_{t+1} = X_t^0 R_t^f + \sum_{s=1}^{S} X_t^s (P_{t+1}^s + D_{t+1}^s). \tag{2}$$

The objective of the agent is then to maximise expected utility from final wealth:

$$\max_{\{X_t^s\}_0^S} \mathbb{E}_t \left[u \left(W_{t+1} \right) \right]. \tag{3}$$

Plugging constraints (1) and (2) in the objective function (3), the problem solved by the agent can be expressed as

$$\max_{\{X_t^s\}_1^S} \mathbb{E}_t \left[u \left(W_t R_t^f + \sum_{s=1}^{S} X_t^s (P_{t+1}^s + D_{t+1}^s - P_t^s R_t^f) \right) \right]. \tag{4}$$

The key addition this standard set-up, to rationalize the low-risk anomaly in the spirit of Frazzini and Pedersen (2014), is a leverage constraint:

$$W_t \geq \left(\sum_{s=1}^{S} X_t^s P_t^s \right) \cdot c_t, \tag{5}$$

which depends on the stylized and exogenously-set margin requirement c_t.

The first order condition with respect to holding X_t^s is then

$$\mathbb{E}_t \left[u'(W_{t+1})(R_{t+1}^s - R_t^f) \right] - \psi_t = 0 \tag{6}$$

where $R^s_{t+1} = \dfrac{P^s_{t+1} + D^s_{t+1}}{P^s_t}$ is the gross return on asset s and ψ_t is the Lagrange multiplier of (5), which is greater than 0 when the leverage constraint binds. As the representative investor in this set-up has to hold all of the assets, a binding constraint happens any time $c_t > 1$. This condition can be interpreted as the assets holder, representing the intermediary sector, being asked to add holdings of the risk-free asset to simply force a safer portfolio onto it, considering that the R^f is not determined by market equilibrium and risky assets are in fixed supply in the short-term.

From (6) it follows that

$$\mathbb{E}_t\left[r^s_{t+1}\right] = \tilde{\psi}_t - \mathrm{Cov}_t\left[\tilde{u}'_{t+1}, r^s_{t+1}\right], \tag{7}$$

where $u'_{t+1} = u'(W_{t+1})$, $\tilde{x} = x/\mathbb{E}_t\left[u'_{t+1}\right]$, and $R^i_{t+1} - R^f_t = r^i_{t+1}$. Remember that R^f_t is exogenously set. (7) holds similarly for the market, meaning that

$$\mathbb{E}_t\left[r^M_{t+1}\right] = \tilde{\psi}_t - \mathrm{Cov}_t\left[\tilde{u}'_{t+1}, r^M_{t+1}\right]. \tag{8}$$

Combining these last two equation one obtains

$$\mathbb{E}_t\left[r^s_{t+1}\right] = \tilde{\psi}_t\left(1 - \frac{\mathrm{Cov}_t\left[\tilde{u}'_{t+1}, r^s_{t+1}\right]}{\mathrm{Cov}_t\left[\tilde{u}'_{t+1}, r^M_{t+1}\right]}\right) + \frac{\mathrm{Cov}_t\left[\tilde{u}'_{t+1}, r^s_{t+1}\right]}{\mathrm{Cov}_t\left[\tilde{u}'_{t+1}, r^M_{t+1}\right]} \cdot \mathbb{E}_t\left[r^M_{t+1}\right]. \tag{9}$$

Labelling $\frac{\mathrm{Cov}_t\left[\tilde{u}_{t+1}, r^s_{t+1}\right]}{\mathrm{Cov}_t\left[\tilde{u}_{t+1}, r^M_{t+1}\right]}$ as γ^s_t, (9) can be more simply stated as

$$\mathbb{E}_t\left[r^s_{t+1}\right] = \tilde{\psi}_t\left(1 - \gamma^s_t\right) + \gamma^s_t \cdot \mathbb{E}_t\left[r^M_{t+1}\right]. \tag{10}$$

Here γ is the only determinant of cross-sectional variation in expected returns and can be interpreted as as assets' comprehensive measure of risk, playing in fact the role that β has in a similar equation in Frazzini and Pedersen (2014). Also in a similar way, ψ measures the zero-beta spread that modifies how risk is rewarded with respect to a perfect market with no financing frictions.

2.2 Deviations from CAPM

Defining the expectations formed following (10) as $\mathbb{E}_t^{\mathrm{FULL}}[r^s_{t+1}]$ and the expectations formed following standard CAPM as

$$\mathbb{E}_t^{\mathrm{CAPM}}[r^s_{t+1}] = \beta^s_t \cdot \mathbb{E}_t\left[r^M_{t+1}\right] \qquad \beta^s_t = \frac{\mathrm{Cov}_t\left[R^M_{t+1}, R^s_{t+1}\right]}{\mathrm{Var}_t\left[R^M_{t+1}\right]}, \tag{11}$$

then deviations from CAPM predictions can be summarized by $\alpha^s_t = \mathbb{E}_t^{\mathrm{FULL}}[r^s_{t+1}] - \mathbb{E}_t^{\mathrm{CAPM}}[r^s_{t+1}]$, which takes value

$$\alpha^s_t = \underbrace{\tilde{\psi}_t(1 - \gamma^s_t)}_{\text{Liquidity deviation}} + \underbrace{(\gamma^s_t - \beta^s_t)\mathbb{E}_t\left[r^M_{t+1}\right]}_{\text{Omitted risk deviation}} \tag{12}$$

$$= \tilde{\psi}_t \left[(1-\beta_t^s) - (\gamma_t^s - \beta_t^s)\right] + (\gamma_t^s - \beta_t^s)\mathbb{E}_t\left[r_{t+1}^M\right]. \tag{13}$$

(12) shows that the deviation of the expected return of a security from CAPM predictions can be split in a part due to a binding leverage constraint, or more generally a spread in the zero-beta rate, and a part due to a different assessment of 'total' risk from the beta measure. (13) makes an even clearer distinction between two stories put forward to explain the CAPM anomaly: Frazzini and Pedersen (2014) finds alpha equating $\tilde{\psi}(1-\beta)$, Schneider et al. (2020) relies on $\gamma - \beta$ where γs explicitly deviate from β because of the market squared in the marginal utility formation. Then, the counteracting effect of the two effects, liquidity motives versus the omitted risk factors, is synthesized by $\gamma - \beta$: the greater is the distance of 'total' risk from β, the more α will be due to risk remuneration of omitted risks, the smaller the term is, the more α will be due to a different market premium which gets captured by portfolios with simply different betas.

To understand in more practical terms the meaning of this, think first of the case in which the market really is the only relevant risk factor: $\gamma - \beta$ will be zeros and one would be exactly back to the pricing equation of Frazzini and Pedersen (2014) where it is the multiplier producing alpha as beta grows. Adding a second priced factor to which assets are sensitive in random way will not change much:[12] high-beta assets will be as riskier than low-beta assets as before, so α in the cross section would still decrease with β with the zero-beta spread driving it. Finally think of a second risk factor to which assets are sensitive to in an opposing way relative to how they are sensitive to the market, such as coskewness. Now high-beta assets are good hedges of market variance shocks and vice-versa low-beta assets are riskier in this second dimension. Bringing the counteracting effect to an extreme, γ could become a constant across assets; then, a liquidity spread would still be present in the market, but the relationship of alphas with beta would now be exclusively driven by the mis-measurement of how risky a security actually is.

Another way to synthesise how flatter the SML is relatively to the theoretical CAPM prediction, and possibly trade on it, is to look at the expected return of a BAB factor. Following Novy-Marx and Velikov (2022) observations, this can be formed holding a set of low CAPM-beta assets, shorting a set of high CAPM-beta ones, and making the portfolio beta-neutral by holding the market proportionally to the beta-tilt of the first two holdings, while financing this with the risk-free asset:

$$\mathbb{E}_t\left[r_{t+1}^{BAB}\right] = \mathbb{E}_t\left[r_{t+1}^L\right] - \mathbb{E}_t\left[r_{t+1}^H\right] - (\beta_t^L - \beta_t^H)\mathbb{E}_t\left[r_{t+1}^M\right] \tag{14}$$

$$= \tilde{\psi}_t\left(\gamma_t^H - \gamma_t^L\right) + \left[(\beta_t^H - \beta_t^L) - (\gamma_t^H - \gamma_t^L)\right]\mathbb{E}_t\left[r_{t+1}^M\right]. \tag{15}$$

From here it is further clear that the origin of the CAPM mispricing critically depends on how different is the assets' 'real' total risk from CAPM assessment. Specifically, the smaller the difference in the γs of the beta-sorted portfolios, the greater the error by CAPM would be due to an omitted risk factor making more market-sensitive assets not so risky after all, and

[12] Random in a cross-sectional sense, not in a time-varying fashion.

vice-versa. It should also be noted that adding coskewness, for example, one expects to lower the gamma-differential as high-beta assets are safer than low-beta ones in the coskewness dimension, but there could even be risk factors with the opposite link that could counteract the counteracting effect of coskewness. On the other side, instead, if actually riskier assets, i.e. those with higher γ, also have high β, either because the market is the only risk priced or because the other risks are neutrally distributed along β dimension, funding-motives will more likely drive the anomaly and show up as a reduction in the premium.

2.3 Betting Against Gamma

With the synthetic risk measure γ capturing risk as β does in CAPM, to isolate the intercept-related spread, one can build a portfolio similar to BAB: long on low-γ (LG), short on high-γ (HG) and hedging the exposure with position $\gamma^{LG} - \gamma^{HG}$ in the market. I call this portfolio 'Betting Against Gamma' (BAG):

$$\mathbb{E}_t\left[r_{t+1}^{BAG}\right] = \mathbb{E}_t\left[r_{t+1}^{LG}\right] - \mathbb{E}_t\left[r_{t+1}^{HG}\right] - (\gamma^{LG} - \gamma^{HG})\mathbb{E}_t\left[r_{t+1}^{M}\right] \tag{16}$$

$$= \tilde{\psi}_t\left(\gamma_t^{HG} - \gamma_t^{LG}\right). \tag{17}$$

This portfolio allows to investigate the sources of variation in the intercept, hypothesized to be related to funding motives, as can be seen through the lens of the decomposition from Campbell (1991),

$$r_t^{BAG} \approx \mathbb{E}_{t-1}\left[r_t^{BAG}\right] + \sum_{j=0}^{\infty}\rho^j \cdot (\mathbb{E}_t - \mathbb{E}_{t-1})[\Delta \ln D_{t+j}^{BAG}] - \sum_{j=1}^{\infty}\rho^j \cdot (\mathbb{E}_t - \mathbb{E}_{t-1})[r_{t+j}^{BAG}]. \tag{18}$$

Assuming, without great loss of generality, constant γs of the high and low portfolio; that shocks to the multiplier are i.i.d; and that shocks to cash flows expectations are independent from those to the leverage constraint, which is reasonable at high frequencies given the interpretation of the model; the contemporaneous relation between changes in ψ and BAG returns is

$$\frac{\partial\left(r_t^{BAG} - \mathbb{E}_{t-1}\left[r_t^{BAG}\right]\right)}{\partial\left(\tilde{\psi}_t - \mathbb{E}_{t-1}[\tilde{\psi}_t]\right)} = -\rho\left(\gamma_t^{HG} - \gamma_t^{LG}\right). \tag{19}$$

Therefore, a further testable implication is that a variable that captures funding tightness should covary negatively with the BAG returns, controlling for other factors.

3 Bringing the theory to the data

The key object of this study, as in most asset pricing models, is the marginal utility, whose covariance with assets' returns determines everything shown so far. It is important to recover its process in the most agnostic way to avoid the risk of omitting variables, which is high in mapping it on arbitrary factors. To do so, I look at its minimum-variance projection on all the assets returns, in the style of Hansen and Jagannathan (1991), and assume a latent factor structure in returns. I then proceed with the contemporaneous estimation of the average zero-beta spread and the marginal utility loadings of the first few principal components only, in a standard GMM procedure.[13]

3.1 A sparse and agnostic approach

As an empirical counterpart of the standardized marginal utility $\tilde{u}'_{t+1}$, consider its linear projection $\tilde{u}'^{*}_{t+1}$ on N returns, with coefficients $\boldsymbol{\beta}_{u\mathbf{r}}$,

$$\tilde{u}'^{*}_{t+1} - 1 = (\mathbf{r}_{t+1} - \mathbb{E}[\mathbf{r}_{t+1}])^{\top}\boldsymbol{\beta}_{u\mathbf{r}}, \tag{20}$$

where $\mathbf{r}_t$ is the vector of N returns at time t. Then, the coefficients can then be retrieved by substituting $\tilde{u}'^{*}_{t+1}$ in the stacked unconditional expectations of all excess returns, as in (6), scaled by $\mathbb{E}[u'_{t+1}]$, i.e.

$$\mathbb{E}\left[\tilde{u}'^{*}_{t+1}\mathbf{r}_{t+1}\right] = \mathbb{E}\left[\tilde{\psi}_t\right]\mathbf{1}_N. \tag{21}$$

Subtracting $\mathbb{E}[\mathbf{r}_{t+1}]$ from both sides, one obtains

$$\mathbb{E}\left[(\tilde{u}'^{*}_{t+1} - 1)(\mathbf{r}_{t+1} - \mathbb{E}[\mathbf{r}_{t+1}])\right] = \mathbb{E}\left[\tilde{\psi}_t\right]\mathbf{1}_N - \mathbb{E}[\mathbf{r}_{t+1}], \tag{22}$$

from which $\boldsymbol{\beta}_{u\mathbf{r}}$ can be obtained by plugging the definition of $\tilde{u}'^{*}_{t+1} - 1$ in,

$$\boldsymbol{\beta}_{u\mathbf{r}} = \Sigma^{-1}_{\mathbf{rr}}(\mathbb{E}[\psi_t]\mathbf{1}_N - \mathbb{E}[\mathbf{r}_{t+1}]), \tag{23}$$

where $\Sigma_{\mathbf{rr}}$ is the covariance matrix of the N returns. The resulting process of the marginal utility projection process is

$$u'^{*}_{t+1} - 1 = (\mathbf{r}_{t+1} - \mathbb{E}[\mathbf{r}_{t+1}])^{\top}\Sigma^{-1}_{\mathbf{rr}}(\mathbb{E}[\psi_t]\mathbf{1}_N - \mathbb{E}[\mathbf{r}_{t+1}]). \tag{24}$$

This involves estimating a great number of parameters and inverting a huge matrix, both concerning $\Sigma_{\mathbf{rr}}$, so this formulation is highly impractical.

Looking for a formulation of the problem that maintains as much information as possible while keeping the problem empirically feasible, I make the further assumption of a perfect factor structure for returns innovations such as

$$\mathbf{r}_{t+1} - \mathbb{E}[\mathbf{r}_{t+1}] = \mathbf{e}_{t+1} \tag{25}$$

$$= \mathrm{B}\cdot\mathbf{f}_{t+1} + \boldsymbol{\epsilon}_{t+1}, \tag{26}$$

where B is the $N \times K$ matrix of factor loadings. Factors $\mathbf{f}$ and residuals $\boldsymbol{\epsilon}$ are independent

[13] To see how realistic is for the first few principal components to capture most of the relevant information *pricing-wise*, an extension of Kozak et al. (2018) is needed, which will likely result in some bound for ψ on the lines of Jiang and Richmond (2022) and Cochrane and Saa-Requejo (2000).

and both zero mean, with $\mathbb{E}\left[\mathbf{f}_{t+1}\mathbf{f}_{t+1}^{\top}\right] = I_K$, and $\mathbb{E}\left[\boldsymbol{\epsilon}_{t+1}\boldsymbol{\epsilon}_{t+1}^{\top}\right]$ being a diagonal matrix filled with the vector $\boldsymbol{\sigma}_{\boldsymbol{\epsilon}}^2$, as most standard applications of factor structures.

The marginal utility projection process can then be expressed as

$$\tilde{u}_{t+1}^{\prime *} = 1 + (\mathbf{f}_{t+1}^{\top}\mathrm{B}^{\top} + \boldsymbol{\epsilon}_{t+1}^{\top})\Sigma_{\mathbf{rr}}^{-1}(\mathbb{E}\left[\psi_t\right]\mathbf{1}_N - \mathbb{E}\left[\mathbf{r}_{t+1}\right]), \tag{27}$$

or, more succinctly,

$$\tilde{u}_{t+1}^{\prime *} = 1 - \mathbf{f}_{t+1}^{\top}\boldsymbol{\beta}_{u\mathbf{f}} - \boldsymbol{\epsilon}_{t+1}^{\top}\boldsymbol{\beta}_{u\boldsymbol{\epsilon}}, \tag{28}$$

where $\boldsymbol{\beta}_{u\mathbf{f}}$ are the loadings of the pervasive factors in marginal utility and $\boldsymbol{\beta}_{u\boldsymbol{\epsilon}}$ is the linear mapping of the marginal utility on the individual innovations residuals.

Using the projected marginal utility, the pricing equation (7) becomes

$$\mathbb{E}_t\left[r_{t+1}^s\right] = \tilde{\psi}_t + \boldsymbol{\beta}_{u\mathbf{f}}^{\top} \cdot \mathrm{Cov}_t\left[\mathbf{f}_{t+1},\ r_{t+1}^s\right] + \boldsymbol{\beta}_{u\boldsymbol{\epsilon}}^{\top} \cdot \mathrm{Cov}_t\left[\boldsymbol{\epsilon}_{t+1},\ r_{t+1}^s\right]. \tag{29}$$

If the factor structure approximates well the returns distribution, the variance of idiosyncratic residuals $\mathrm{Cov}_t\left[\boldsymbol{\epsilon}_{t+1},\ r_{t+1}^s\right] = \sigma_{\epsilon,s}^2$ will be minimal, although not zero. While I have no formal guarantee that $\boldsymbol{\beta}_{u\boldsymbol{\epsilon}}$ tends to 0 in any way at this stage, assuming $\boldsymbol{\beta}_{u\boldsymbol{\epsilon}}^{\top} \cdot \mathrm{Cov}_t\left[\boldsymbol{\epsilon}_{t+1},\ r_{t+1}^s\right] \approx 0$ has the advantage that the following condition involves only a few unknown constants, $\mathbb{E}\left[\tilde{\psi}_t\right]$ and $\boldsymbol{\beta}_{u\mathbf{f}}$, to make meaningful asset pricing predictions and obtain estimates of γs:

$$\mathbb{E}\left[r_{t+1}^s\right] = \mathbb{E}\left[\tilde{\psi}_t\right] + \boldsymbol{\beta}_{u\mathbf{F}}^{\top} \cdot \mathrm{Cov}\left[\mathbf{f}_{t+1}, r_{t+1}^s\right]. \tag{30}$$

A sense of how much information that is relevant to pricing has gotten lost is offered by ex-post performance measures. Again, a model that tends to have no pricing errors would be ideal, but a 'wrong' model can still be good enough to leave no room for flipping the results with a better, or even perfectly, performing model. Therefore, the exercise is likely meaningful even with errors as long as these are reasonably small.

4 Empirical Analysis

4.1 Empirical strategy and test assets

The analysis verges around the estimates of $\mathbb{E}\left[\tilde{\psi}_t\right]$ and $\boldsymbol{\beta}_{u\mathbf{F}}$, to obtain an empirical counterpart of $\tilde{u}_t'$. Assuming a constant ψ_t, this implies estimating the coefficients $\{a, \mathbf{b}\}$ of the following moment condition, given by the pricing equation (30):

$$\mathbb{E}\left[r_t^s\right] - a - \mathbb{E}\left[\mathbf{b}^{\top}\hat{\mathbf{f}}_t \cdot (r_t^s - \mathbb{E}\left[r_t^s\right])\right] = 0. \tag{31}$$

This can be easily used in a GMM estimation, after which the estimates of unconditional γs amount to

$$\gamma^s = \frac{\mathbf{b}^\top \mathrm{Cov}\left[\hat{\mathbf{f}}_t, r_t^s\right]}{\mathbf{b}^\top \mathrm{Cov}\left[\hat{\mathbf{f}}_t, r_t^M\right]}, \tag{32}$$

with standard errors obtainable using the Delta method and GMM estimates of the covariances.

Before moving to GMM the procedure needs an earlier step to form the factors. Indeed, these are composed by (1) risk factors previously known to be relevant for the anomaly and (2) pervasive latent factors of a wide set of test assets. More precisely, the market factor and the market squared excess returns, for coskewness, are the first two. The rest of the factors are the first few Principal Components (PCs) the test assets, whose computation procedure exactly aims at minimizing the assets' residuals' variance. Before performing PCA, the test assets are orthogonalized with respect to the market and the market squared to avoid redundant information.[14] .

The market factor is the monthly CRSP value-weighted market index in excess of the 1-month risk-free, again from CRSP. The test assets pool from which PCs are extracted is mostly populated by the 153 monthly stock portfolios used in Jensen et al. (2021),[15] all of which are available from November 1971 to December 2022. The original BAB portfolio from Frazzini and Pedersen (2014) is among these portfolios. Then, I add 10 beta-sorted stocks portfolios and the key asset of the analysis: the BAB factor, based on a 3-fold split of beta-sorted US stocks, robust to the criticisms of Novy-Marx and Velikov (2022).

To keep the dimensionality of the GMM problem reasonable and the estimation well behaved, the test assets actually used in the GMM step of the analysis are a 'condensed' version of the ones used to obtain $\hat{\mathbf{f}}$. Precisely, this second set of test assets is formed by the market excess returns, the BAB factor, 3 beta-sorted stocks portfolios, and 13 'themed' portfolios build by clustering the 153 portfolios of Jensen et al. (2021), provided by the same authors. Statistics for these portfolios are in Table 6, in the appendix. The clustering technique used to create these 13 theme-portfolios has the precise intent of keeping a great dispersion across-theme and a high degree of within-theme correlation and economic concept similarity. The original BAB portfolios is among those clustered in the 'low risk'-themed portfolio, but I also explicitly add it to the test assets pool of the second step for consistency with the other BAB factor.

4.2 Beta-sorted portfolios

I construct the beta-sorted portfolios using monthly data from CRSP, correcting for delisting returns depending on nature of the delisting, as suggested by Bali, Engle, et al. (2016). Following common practice in the literature, I only consider stocks of share types 10 and

[14] To do so, I simply regress every test asset on the market and the market squared, keeping the residuals.

[15] Available at https://jkpfactors.com.

Table 1: Beta-sorted portfolios statistics. In parenthesis, HAC standard errors obtained as suggested in Lazarus et al. (2018). Monthly annualized returns from Nov 1971 to Nov 2022.

	Bottom	Top	BAB	Orig. BAB
Avg. ret (%)	11.935	12.635	3.881	10.083
SD (%)	43.274	80.635	42.549	40.741
CAPM α (%)	7.283	3.179	4.079	10.52
	(1.219)	(1.446)	(1.986)	(2.773)
CAPM β	0.684	1.391	−0.029	−0.064
	(0.048)	(0.041)	(0.065)	(0.099)
Res. coskew. ($\times 10^3$)	0.217	0.429	−0.111	−1.659
	(0.199)	(0.291)	(0.449)	(0.447)

11, traded on NYSE, Nasdaq or AMEX exchanges. To form portfolios, every month all the stocks are ranked based on the rolling betas estimated on a 5-year window ending the month before. They are then split into 10 or 3 portfolios with weights corresponding to the relative capitalization. The BAB portfolio is built following Novy-Marx and Velikov (2022), i.e. subtracting returns of the top-third beta portfolio from bottom-third beta portfolio and subtracting to this the excess returns of the market, proportionally to the estimated beta of the low-minus-high portfolio in the 5 years ending the month before. Summary statistics of the bottom and top thirds, BAB portfolio and a BAB portfolio as originally formed by Frazzini and Pedersen (2014) are shown in Table 1. The timespan of the analysis is from November 1971 to November 2022, dictated by the availability of portfolios from Jensen et al. (2021).

As expected, the 'top' portfolio earns a higher return than the 'bottom' one, but its CAPM alpha is significantly lower. Top portfolio also shows a higher residual coskewness, obtained regressing CAPM residuals on the market squared, following Schneider et al. (2020). Both BAB portfolios earn a significant alpha with respect to CAPM while having no significant exposure to the market. However, they significantly differ in the magnitude of alphas and the residual coskewness, with the 'original' BAB portfolio earning a higher return unexplained by CAPM and having a lower residual coskewness, which suggests the value-weighting mitigating the coskewness risk. Statistics covering the full sample of the portfolios are in Table 5, while the statistics for the middle third portfolio and the 10-split portfolios over the same time-frame are in Table 7, both showing a similar pattern.

4.3 Estimation results

4.3.1 PCs factors

Figure 2 shows the scree plot relative to the Principal Component Analysis of the 164 test assets. At this stage no formal test is conducted to choose the number of components to keep. Rather, I arbitrarily choose to keep 4 because adding them to the market factor and the squared market factor results in a total of 6 factors, which is a number comparable with the other *sparse* models at the frontier at the time of this work, such as Fama and French

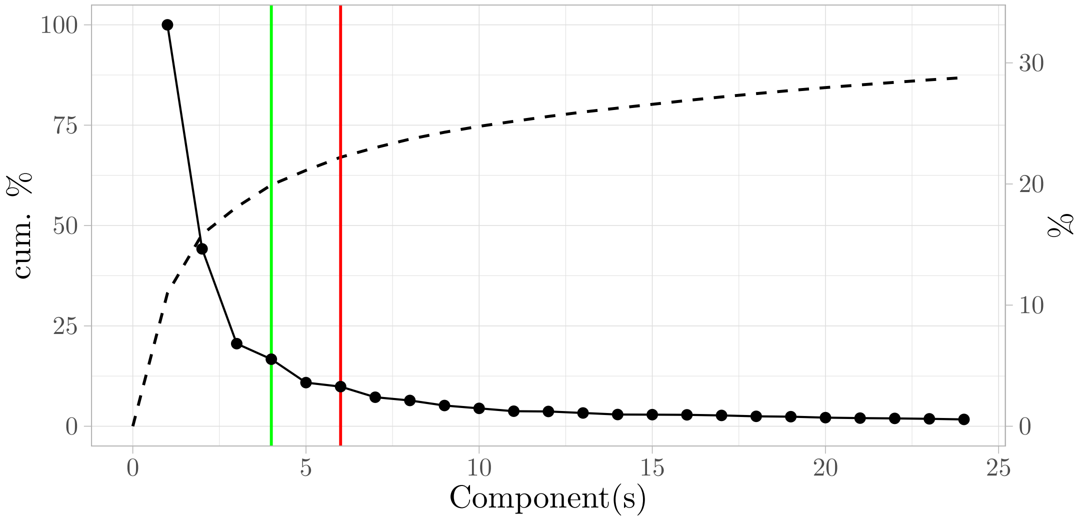

Figure 2: Scree plot of principal component analysis. On the horizontal axis the number of component considered, on the vertical axis the share of variance explained by the relative component (solid line) and the cumulative share of variance explained (dotted line). The green vertical line marks the 4^{th} component while the red one the 6^{th}.

(2018). This also seems a sensible choice because the following component, the 5^{th}, would explain almost half the variance explained by the 4^{th} one, less than 5%. The first 4 PCs end up explaining 60% of test assets' residual variance after orthogonalization to market and market squared; the first 6 ones explain a total of 67%. To control for the gains from adding factors, I also use a specification keeping 6 components, resulting in 8 factors, although it does not change results in a significant way.

4.3.2 Pricing the cross-section

Table 2 shows the estimation results for a few specifications: 'CAPM', where the only risk factor is the market and ψ is fixed at 0; 'F(Market)', where the market is the only factor, but $\mathbb{E}[\psi]$ is freely estimated; 'F(cskw)', which adds the market squared among the factors; 'F(6)', which adds 4 cross-sectional principal components as factors to the previous specification; and finally, 'F(8)', which adds 6 components instead.

Comparing the first two columns, it can be seen that introducing the spread $\mathbb{E}[\psi]$ significantly improves the pricing of the cross-section, cutting the Mean Absolute Pricing Error (MAPE) in half. Indeed, the spread is significantly different from 0 in all specifications and has a substantial magnitude, between 2% and 3.5% annually. The average unexplained return of the 'robust' BAB portfolio immediately becomes statistically and economically insignificant, while the unexplained excess return of the 'original' BAB factor does not. This further remarks the difference in the two and highlights how much do the portfolio formation details matter. The market factor is never significant apart from the CAPM specification.

The market squared enters the marginal utility negatively and quite significantly in all specifications including it. Also, consistently with Schneider et al. (2020), adding the

Table 2: Model estimation results. HAC standard errors in parenthesis. Monthly sample from 1971 to 2022.

	CAPM	F(Market)	F(cskw)	F(6)	F(8)
$\mathbb{E}[\psi]$		3.583^{***}	3.184^{***}	2.050^{***}	2.038^{***}
		(0.423)	(0.532)	(0.256)	(0.258)
Market factors	0.100^{*}	0.060	0.000	0.030	0.000
	(0.046)	(0.046)	(0.085)	(0.080)	(0.103)
Market squared			-0.516°	-0.485°	-0.719°
			(0.281)	(0.288)	(0.385)
J-test	330.140^{***}	351.514^{***}	131.472^{***}	42.699^{***}	30.007^{***}
MAPE	3.431	1.625	1.505	0.709	0.570
BAB a.p.e.	4.149^{*}	0.501	0.591	-0.433	-0.276
	(2.005)	(2.016)	(2.733)	(2.991)	(3.173)
Orig. BAB a.p.e.	10.526^{***}	6.800^{**}	2.230	1.145	1.298
	(2.722)	(2.640)	(3.642)	(3.562)	(4.402)

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{\circ}p < 0.1$

market squared among the factors makes the pricing error in the original BAB portfolio insignificant too. This additional explanatory power, however, appears to be more related to the non-standard practices used to form the original BAB factor,[16] rather than to an intrinsic mechanism of the CAPM anomaly, since the robust BAB portfolio does not experience such a reduction in pricing error. Employing the returns' principal components as factors further reduces MAPEs, decreasing them to less than half of the MAPE in the model with coskewness. Indeed, deviations on both BAB portfolios further decrease too.

All the over-restricting conditions of the models are not valid at the 0.1% of confidence level, although F(8) has a p-value that is only slightly lower than that. This, however, is not too much of a concern: the goal is to see through the lenses of a better reality approximation the sources of CAPM failures and any better-performing model can be used to this aim. Obviously, the more information a model leaves unexplained, the higher the chances of the decomposition being twisted with an even better model. Anyway, despite not achieving a perfect fit, over these test assets with average mean return of around 4%,[17] F(6) and F(8) produce MAPEs that are less than 1%, a fifth of CAPM's starting 3.5%. This certainly leaves room for additional information to improve the analysis, but it seems a rather small space to completely reverse the main takeaways of this exercise. This is further supported in the next sub-section by the uniformity of the decomposition pattern development when increasing the model complexity.

4.3.3 Revisiting the low-risk anomaly

Estimates of **b** imply estimates of γs too, which are shown in Table 3. The first pattern to emerge is that, when more risks are considered, the spread in the synthetic risk measure

[16] As highlighted by Novy-Marx and Velikov (2022), these tilt holdings towards small and illiquid stocks possibly making the portfolio load more on coskewness.

[17] See Table 6 in the appendix.

Table 3: Portolios γ. HAC standard errors in parenthesis, obtained through Delta method from previous GMM estimation. Monthly sample from 1971 to 2022.

	Bottom	Top	HmL	BAB	Orig. BAB
γ CAPM	0.688^{***}	1.395^{***}	0.707^{**}	-0.030	-0.064
	(0.026)	(0.034)	(0.359)	(0.051)	(0.058)
γ Market	0.688^{***}	1.395^{***}	0.707	-0.030	-0.064
	(0.026)	(0.034)	(0.672)	(0.051)	(0.058)
γ Coskew.	0.549^{***}	1.109^{***}	0.560	0.048	1.082
	(0.157)	(0.210)	(0.488)	(0.277)	(0.702)
γ All (4)	0.881^{***}	1.037^{***}	0.156	0.411	1.211^{**}
	(0.177)	(0.223)	(0.429)	(0.328)	(0.551)
γ All (6)	0.786^{***}	1.047^{***}	0.261	0.363	1.116^{**}
	(0.177)	(0.215)	(0.431)	(0.320)	(0.497)

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

decreases, which can be seen by observing the γ of the high-minus-low beta portfolio 'HmL'. This, again, is in line with Schneider et al. (2020), which shows high-beta stocks being safer than what the betas would suggest once coskewness is taken into account, and extends it further to more unidentified risks. On the other hand, the BAB portfolio gets riskier and riskier as the number of risks considered, and arguably the 'realism' of the model, increases, although not significantly. To understand why, the definition in (15) is useful: assuming the extreme case in which all actual risks in the economy makes the high-beta and the low-beta portfolios having the same total risk, then, any holding of the market originally taken to hedge the beta of the low-minus-high beta portfolio will distort the total-risk neutrality. Such risk in this formulation is rewarded with $\mathbb{E}\left[r^M\right]$ per unit of γ, which would be the source of the BAB expected return. In this case, none of the BAB return would be due to liquidity considerations because having the high-beta and the low-beta portfolios the same γ, they would provide cash-flows with the same discount rate, and no embedded leverage would be enjoyed by agents. Note that this would not mean that the liquidity motive is irrelevant: the BAG portfolio defined in (16) would still be entirely determined by the liquidity compensation,[18] just like the BAB return is in Frazzini and Pedersen (2014).

The last example also shows why the CAPM anomaly is not a 'plain' case of omitted factors: to make high-beta assets have the same total risk of low-beta assets, assets have to be riskier in a second dimension inversely proportional to their beta, otherwise the difference in total risk levels of high-beta and low-beta will persist. This means that it does not suffice for the additional factor to have a specific correlation with the market, but that there is a deeper link in the way in which sensitivity to the market relates to sensitivity to this second factor. Notice also that additional risks could also make it *harder* to explain the anomaly, in case the risk pattern relates to beta in the opposite way as coskewness does. However, this does not seem the case, as γs converge towards 1 increasing the risks. All in all, as from F(cskw) and F(6) the difference in risk between high-beta and low-beta decreases and BAB risk increases, the evidence points towards the existence of omitted factors relevant to the

[18] At the current state, the BAG portfolio has not been studied, but it will be included in this work.

Table 4: Deviations' contribution. 'Prd/real' is the ratio of predicted return of the portfolio over the actual average return. HAC standard errors in parenthesis, obtained through Delta method from previous GMM estimation. Monthly sample from 1971 to 2022.

	BAB				Orig. BAB			
	$\psi(1-\gamma)$	$\gamma \cdot \mathbb{E}\left[r^M\right]$	Prd/real%	Δ%	$\psi(1-\gamma)$	$\gamma \cdot \mathbb{E}\left[r^M\right]$	Prd/real%	Δ%
Market	3.69***	−0.21	87.4	111.8***	3.81***	−0.45	33.1	126.4***
	(0.49)	(0.36)		(21.0)	(0.54)	(0.42)		(26.3)
Coskew.	3.03**	0.33	85.2	80.3	−0.26	7.49	78.1	−107.2*
	(1.16)	(1.92)		(109.1)	(2.24)	(5.42)		(59.5)
All (4)	1.21*	2.85	110.9	−40.4	−0.43	8.39*	88.7	−110.9***
	(0.67)	(2.45)		(57.8)	(1.14)	(4.67)		(25.2)
All (6)	1.30*	2.52	106.9	−31.9	−0.24	7.73*	87.2	−106.3***
	(0.67)	(2.36)		(63.8)	(1.02)	(4.25)		(25.2)

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

anomaly, although not yet specifically identified in the literature.

The presence of risks not orthogonally distributed with respect to betas is extremely clear in the original BAB, where the synthetic risk measure is significant. Any BAB portfolio is characterized by market-risk neutrality, so a significant γ, interpreted through (12), suggests the risk component being relevant in explaining the CAPM anomaly, even considering the liquidity spread. It is important however to understand the extent to which the apparent 'mispricing' of CAPM is due to risks remuneration or funding provision, in order to improve assessments such as that in Baker et al. (2014). To do this, I report in Table 4 the measurements of the liquidity component and the risk component, with standard errors obtained via Delta method using previous GMM estimates. I also compute a measure of relative contribution, the share of return prediction that is associated to pure liquidity motives or to pure risk mis-measurement, defined as

$$\Delta = \frac{\text{Liq. component} - \text{Risk component}}{\text{Tot. prediction}} = \frac{\mathbb{E}\left[\psi\right]\left(1-\gamma^{BAB}\right) - \gamma^{BAB} \cdot \mathbb{E}\left[r^M\right]}{\mathbb{E}\left[\psi\right]\left(1-\gamma^{BAB}\right) + \gamma^{BAB} \cdot \mathbb{E}\left[r^M\right]}. \tag{33}$$

This is positive when contribution due to the zero-beta spread is greater than that due to γ-risk, relative to the total return that the model is able to predict, and vice-versa. Once again, remember that BAB portfolios have theoretically and empirically 0 β, so any measure of γ different from 0 is a reflection of mis-measurement of risk in CAPM.

It can be seen that the liquidity spread play a major role when the market risk only is considered: the liquidity component contributes 112% more than the risk component in the prediction of the BAB return and 126% more of the original BAB. Here, the risk component even contributes negatively, with a negative return prediction originated in the residual market risk of the portfolios. However, as more risks are considered, the two components switch roles. Considering coskewness, it can be seen that for the robust BAB portfolio the liquidity component stops significantly contributing more than the risk component, while for the original BAB the relative contribution flips significantly already, again supporting the strong risk mis-measurement motive behind it. Moving to full-fledged models, it can be seen that for both specifications F(6) and F(8) the percentage of average return explained by the omission of factors, which makes a BAB portfolio risky, is higher with no statistical significance for the robust BAB, but with high statistical significance for the original BAB. A Δ not statistically different from zero means that an equal contribution of the liquidity and the risk components cannot be excluded. At the same time, the standard errors of Δ for F(6) and F(8) do not exclude all of the BAB return being due to risk omission, while they essentially rule out the possibility of being all of it imputed to the liquidity component.

Overall, despite possibly suffering of low estimation accuracy, the results support the existence of a funding tightness spread as well as the prominence of omitted risks in explaining the CAPM low-risk anomaly. This could be further studied with a betting-against-gamma portfolio, which can also provide information on the liquidity spread dynamics.[19] The flexibility of the formulation also allows for more complex methods, possibly even those outlined in Didisheim et al. (2023), to be applied.

5 Conclusion

The remuneration of risk, as defined by the CAPM – the most fundamental model in financial economics, is not as high in the data as it is expected to be from theory. This has been hypothesized to be due either because financial frictions reduce such remuneration or because assets do not bring as much risk as they are expected to. In one case acting on financial frictions has an impact on the cost of capital of firms, in the other does not. Also, in one case a firm can expect to gain from exploiting a better funding than investors to leverage up and harvest the zero-beta spread, while in the other case the only effect is that it would become riskier. A formulation of the optimal pricing behaviour of an agent with both a leverage constraint and no specific preferences, in order to accommodate different degrees of realism, illustrates how antagonist to each other the two effects are. In this formulation an inclusive measure of risk can be compared to β, the CAPM risk measure, and inform on how the two differ. It is shown that increasing the risks considered, the return of BAB portfolios are more likely to be compensation for risks omitted by CAPM, despite also supporting an extremely significant role of the spread generated in the zero-beta asset by the financial frictions. This will be further tested with the formation of a Betting-Against-Gamma portfolio, which will be then also instrumental in assessing contemporaneous relations between the zero-beta spread and funding liquidity measures.

References

Adrian, Tobias, Erkko Etula, and Tyler Muir (2014). 'Financial Intermediaries and the Cross-Section of Asset Returns'. In: *The Journal of Finance* 69.6, pp. 2557–2596.

[19] I intend to perform these analysis. Information on the dynamics of the spread would also be obtained with the method outlined in 6.

Ang, Andrew et al. (2006). ‘The Cross-Section of Volatility and Expected Returns’. In: *The Journal of Finance* 61.1, pp. 259–299.

Asness, Cliff et al. (2020). ‘Betting against correlation: Testing theories of the low-risk effect’. In: *Journal of Financial Economics* 135.3, pp. 629–652.

Baker, Malcolm, Brendan Bradley, and Ryan Taliaferro (1, 2014). ‘The Low-Risk Anomaly: A Decomposition into Micro and Macro Effects’. In: *Financial Analysts Journal* 70.2, pp. 43–58.

Bali, Turan G., Stephen J. Brown, et al. (2017). ‘A Lottery-Demand-Based Explanation of the Beta Anomaly’. In: *Journal of Financial and Quantitative Analysis* 52.6, pp. 2369–2397.

Bali, Turan G., R. F. (Robert F.) Engle, and Scott Murray (2016). *Empirical asset pricing: the cross section of stock returns*. Hoboken, New Jersey: Wiley. 494 pp. ISBN: 978-1-118-09504-1.

Black, Fischer (1972). ‘Capital Market Equilibrium with Restricted Borrowing’. In: *The Journal of Business* 45.3, pp. 444–455.

Black, Fischer, Michael C. Jensen, and Myron Scholes (1972). ‘The Capital Asset Pricing Model: Some Empirical Tests’. In: *Studies in the Theory of Capital Markets*. Vol. 81, pp. 79–121.

Brealey, Richard et al. (25, 2022). *Principles of Corporate Finance*. 14th ed. McGraw Hill. ISBN: 978-1-264-08094-6.

Brunnermeier, Markus K., Christian Gollier, and Jonathan A. Parker (2007). ‘Optimal Beliefs, Asset Prices, and the Preference for Skewed Returns’. In: *American Economic Review* 97.2, pp. 159–165.

Campbell, John Y. (1991). ‘A Variance Decomposition for Stock Returns’. In: *The Economic Journal* 101.405, p. 157.

Chamberlain, Gary and Michael Rothschild (1983). ‘Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets’. In: *Econometrica* 51.5, pp. 1281–1304.

Cochrane, John H. and Jesus Saa-Requejo (2000). ‘Beyond Arbitrage: Good-Deal Asset Price Bounds in Incomplete Markets’. In: *Journal of Political Economy* 108.1, pp. 79–119.

Croce, Mariano M, Tatyana Marchuk, and Christian Schlag (1, 2023). ‘The Leading Premium’. In: *The Review of Financial Studies* 36.8, pp. 2997–3033.

Di Tella, Sebastian et al. (2023). *The Zero-Beta Interest Rate*.

Didisheim, Antoine et al. (2023). *Complexity in Factor Pricing Models*.

Du, Wenxin, Benjamin Hébert, and Amy Wang Huber (9, 2022). ‘Are Intermediary Constraints Priced?’ In: *The Review of Financial Studies*.

Elton, Edwin J. (1999). ‘Presidential Address: Expected Return, Realized Return, and Asset Pricing Tests’. In: *The Journal of Finance* 54.4, pp. 1199–1220.

Fama, Eugene F. and Kenneth R. French (1, 2018). ‘Choosing factors’. In: *Journal of Financial Economics* 128.2, pp. 234–252.

Feng, Guanhao, Stefano Giglio, and Dacheng Xiu (2020). ‘Taming the Factor Zoo: A Test of New Factors’. In: *The Journal of Finance* 75.3, pp. 1327–1370.

Frazzini, Andrea and Lasse Heje Pedersen (2014). 'Betting against beta'. In: *Journal of Financial Economics* 111 (1), pp. 1–25.

Giglio, Stefano, Bryan Kelly, and Dacheng Xiu (1, 2022). 'Factor Models, Machine Learning, and Asset Pricing'. In: *Annual Review of Financial Economics* 14.1, pp. 337–368.

Guermat, Cherif (1, 2014). 'Yes, the CAPM is testable'. In: *Journal of Banking & Finance* 46, pp. 31–42.

Hansen, Lars Peter and Ravi Jagannathan (1991). 'Implications of Security Market Data for Models of Dynamic Economies'. In: *Journal of political economy* 99 (2), pp. 225–262.

He, Zhiguo, Bryan Kelly, and Asaf Manela (2017). 'Intermediary asset pricing: New evidence from many asset classes'. In: *Journal of Financial Economics* 126 (1), pp. 1–35.

Hong, Harrison and David A. Sraer (2016). 'Speculative Betas'. In: *The Journal of Finance* 71.5, pp. 2095–2144.

Jensen, Theis Ingerslev, Bryan T. Kelly, and Lasse Heje Pedersen (2021). 'Is There a Replication Crisis in Finance?' In: *SSRN Electronic Journal*.

Jiang, Zhengyang and Robert Richmond (12, 2022). *Convenience Yields and Asset Pricing Models*. Rochester, NY.

Jylhä, Petri (2018). 'Margin Requirements and the Security Market Line'. In: *The Journal of Finance* 73.3, pp. 1281–1321.

Kozak, Serhiy, Stefan Nagel, and Shrihari Santosh (2018). 'Interpreting Factor Models'. In: *The Journal of Finance* 73 (3), pp. 1183–1223.

Lazarus, Eben et al. (2018). 'HAR Inference: Recommendations for Practice'. In: *Journal of Business and Economic Statistics* 36 (4), pp. 541–559.

Lu, Zhongjin and Zhongling Qin (1, 2021). 'Leveraged Funds and the Shadow Cost of Leverage Constraints'. In: *The Journal of Finance* 76.3, pp. 1295–1338.

Miller, Merton H. and Myron Scholes (1972). 'Rates of return in relation to risk: A re-examination of some recent findings'. In: *Studies in the theory of capital markets* 23, pp. 47–48.

Novy-Marx, Robert and Mihail Velikov (2022). 'Betting against betting against beta'. In: *Journal of Financial Economics* 143 (1), pp. 80–106.

Pinto, Jerald E., Thomas R. Robinson, and John D. Stowe (2019). 'Equity valuation: A survey of professional practice'. In: *Review of Financial Economics* 37.2, pp. 219–233.

Pukthuanthong, Kuntara, Richard Roll, and Junbo L. Wang (3, 2021). *An Agnostic and Practically Useful Estimator of the Stochastic Discount Factor*. Rochester, NY.

Schneider, Paul, Christian Wagner, and Josef Zechner (2020). 'Low-Risk Anomalies?' In: *The Journal of Finance* 75 (5), pp. 2673–2718.

Sharpe, William F. (1964). 'Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk*'. In: *The Journal of Finance* 19.3, pp. 425–442.

6 A non-parametric approach

6.A Unconditional estimation

Following the approach used in Pukthuanthong et al. (2021):

$$\mathbb{E}\left[m_{t+1} r^s_{t+1} | \mathbf{z}_t\right] = \psi_t \tag{34}$$

$$\mathbb{E}\left[m_{t+1} r^s_{t+1}\right] = \mathbb{E}\left[\psi_t\right] \tag{35}$$

$$\tag{36}$$

where $m_t = u'_t$. This translates in GMM estimation with sample moments

$$R'\mathbf{m}/T = \mathbf{1}_S \; (\mathbf{1}'_T \, \boldsymbol{\psi})/T \tag{37}$$

$$R'\mathbf{m} = \mathbf{1}_S \; (\mathbf{1}'_T \, \boldsymbol{\psi}) \tag{38}$$

In principle, as long as $S > T$:

$$RR'\mathbf{m} = R \; \mathbf{1}_S \; (\mathbf{1}'_T \, \boldsymbol{\psi}) \tag{39}$$

$$\mathbf{m} = (RR')^{-1} R \; \mathbf{1}_S \; (\mathbf{1}'_T \, \boldsymbol{\psi}), \tag{40}$$

which allows to identify $\mathbf{m}/(\mathbf{1}'_T \, \boldsymbol{\psi})$. With the further assumption that $\mathbb{E}\left[m_t\right] = 1$, the sample counterpart $\mathbf{1}'_T \, \mathbf{m} = T$ allows for an estimate of both the $\mathbf{m}$ time series and $\mathbf{1}'_T \, \boldsymbol{\psi}$, i.e. $\mathbb{E}\left[\psi_t\right]$.

6.B Conditional estimation

This approach can be extended to estimate the whole series of ψ_t, by exploiting conditional expectations. Considering a set of K state variables stored in the $T \times K$ matrix Z, start by multiplying both sides of (34) by the value z^j, which is contained in the time t information set, to get

$$\mathbb{E}_t\left[(m_{t+1} r^i_{t+1}) z^j_t\right] = \psi_t \cdot z^j_t \qquad \forall j \in \{1, \dots, K\} \,, \quad \forall i \in \{1, \dots, N\}. \tag{41}$$

Then, assuming stationary variables, the condition

$$\mathbb{E}\left[(m_{t+1} r^i_{t+1}) z^j_t\right] = \mathbb{E}\left[\psi_t \cdot z^j_t\right] \qquad \forall j \in \{1, \dots, K\} \,, \quad \forall i \in \{1, \dots, N\} \tag{42}$$

is also true. I label $r^i_{t+1} z^j_t$ as d^{ij}_{t+1}, so (42) is

$$\mathbb{E}\left[m_{t+1} d^{ij}_{t+1}\right] = \mathbb{E}\left[\psi_t \cdot z^j_t\right] \qquad \forall j \in \{1, \dots, K\} \,, \quad \forall i \in \{1, \dots, N\}. \tag{43}$$

Under standard assumptions, the sample time-averages of $m_{t+1} d^{ij}_{t+1}$ and $\psi_t z^j_t$ converge to such expectations. Then, labelling D the $T \times NK$ matrix obtained by placing side-by-side all the $\mathbf{d}^{ij}$ vectors, first by i and then by j, these sample averages can be compactly expressed as

$$\underbrace{D^\top}_{(NK\times T)}\underbrace{\mathbf{m}}_{(T\times 1)}/T = \underbrace{(Z\otimes \mathbf{1}_N^\top)^\top}_{(NK\times T)}\underbrace{\boldsymbol{\psi}}_{(T\times 1)}/T, \tag{44}$$

where every entry of the $NK \times 1$ vector on the left-hand side is $\sum_{t=1}^{T} m_{t+1} d_{t+1}^{ij}/T$. It follows that, as long as $NK > T$,

$$DD^\top \mathbf{m} = D(Z\otimes \mathbf{1}_N^\top)^\top \boldsymbol{\psi} \tag{45}$$

$$\mathbf{m} = (DD^\top)^{-1} D(Z\otimes \mathbf{1}_N^\top)^\top \boldsymbol{\psi}. \tag{46}$$

This condition pins down a $\mathbf{m}$ with respect to a $\boldsymbol{\psi}$ and vice-versa, but is not enough to obtain an estimate of the two. To do it, consider the additional condition

$$\mathbb{E}_t\left[m_{t+1}\right] = 1; \tag{47}$$

this implies

$$\mathbb{E}_t\left[m_{t+1} z_t^j\right] = z_t^j \qquad \forall j \in \{1,\dots,K\} \tag{48}$$

and

$$\mathbb{E}\left[m_{t+1} z_t^j\right] = \mathbb{E}\left[z_t^j\right] \qquad \forall j \in \{1,\dots,K\}\ . \tag{49}$$

Exploiting once again the Law of Large Numbers on the time dimension, the sample counterpart is

$$\underbrace{Z^\top}_{(K\times T)}\underbrace{\mathbf{m}}_{(T\times 1)}/T = \underbrace{Z^\top}_{(K\times T)}\underbrace{\mathbf{1}_T}_{(T\times 1)}/T. \tag{50}$$

Finally, plugging (46), an estimate of the ψ_t time-series is the solution to the system

$$Z^\top (DD^\top)^{-1} D^\top (Z\otimes \mathbf{1}_N^\top)^\top \boldsymbol{\psi} = Z^\top \mathbf{1}_T, \tag{51}$$

which is only feasible as long as the K independent columns of Z are greater than the number of time observations T.

7 Additional tables and figures

Table 5: Beta-sorted portfolios statistics. In parenthesis, HAC standard errors obtained as suggested in Lazarus et al. (2018). Monthly annualized returns from Dec 1934 to Nov 2022.

	Bottom	Top	BAB	Orig. BAB
Avg. ret (%)	11.808	13.781	3.837	8.46
SD (%)	43.801	85.02	41.325	36.795
CAPM α (%)	5.801	1.597	3.75	8.943
	(0.983)	(1.159)	(1.598)	(1.808)
CAPM β	0.72	1.461	0.01	-0.058
	(0.034)	(0.056)	(0.05)	(0.064)
Res. coskew.$\times 10^3$	-0.027	0.667	-0.93	-1.335
	(0.155)	(0.228)	(0.487)	(0.332)

Table 6: Test assets statistics. In parenthesis, HAC standard errors obtained as suggested in Lazarus et al. (2018), a part from 'AVG' column: this reports averages of the statistic across portfolios and in parenthesis are standard deviations of such statistics. Monthly annualized returns from Nov 1971 to Nov 2022.

Stat	accruals	debt_iss	invest	low_lev	low_risk	mom	prof_gr	profitab	quality	seasonal	s_t_rev	size	value	AVG
Avg. ret (%)	2.627	2.522	3.544	-0.661	2.168	4.256	1.736	3.235	3.34	1.58	1.366	1.165	4.661	4.112
SD (%)	13.248	8.857	26.877	38.781	47.207	39.366	13.574	27.288	18.644	7.071	15.365	25.89	41.577	32.484
CAPM α (%)	2.799	2.526	5.085	-2.659	6.473	5.476	1.477	4.186	3.522	1.818	1.329	0.4	6.515	3.271
	(0.946)	(0.624)	(1.48)	(1.824)	(1.767)	(1.325)	(0.522)	(1.409)	(0.815)	(0.419)	(0.591)	(1.366)	(2.293)	(1.312)
CAPM β	-0.022	-0.001	-0.198	0.256	-0.552	-0.157	0.033	-0.122	-0.023	-0.031	0.005	0.098	-0.238	0.108
	(0.033)	(0.01)	(0.066)	(0.097)	(0.091)	(0.079)	(0.028)	(0.058)	(0.03)	(0.012)	(0.015)	(0.033)	(0.101)	(0.053)
Res. cskw.$\times 10^3$	0.032	0.116	0.17	0.319	-0.285	-0.868	-0.157	0.215	0.179	-0.053	0.147	-0.263	0.073	-0.09
	(0.176)	(0.066)	(0.271)	(0.343)	(0.412)	(0.381)	(0.088)	(0.228)	(0.115)	(0.073)	(0.169)	(0.183)	(0.376)	(0.242)

Table 7: Additional beta-sorted portfolios statistics.

	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	BAB10
Avg. ret (%)	11.217	11.41	12.561	12.594	12.241	12.268	11.895	12.364	12.381	16.188	3.389
SD (%)	44.274	43.824	48.733	52.167	57.676	61.593	67.967	74.618	88.647	115.511	83.726
CAPM α (%)	7.647	7.078	7.338	6.806	5.628	5.151	3.953	3.652	2.364	4.207	3.702
	(1.55)	(1.486)	(1.218)	(1.163)	(1.101)	(1.189)	(1.016)	(1.027)	(1.885)	(3.159)	(3.877)
CAPM β	0.525	0.637	0.768	0.851	0.972	1.047	1.168	1.281	1.473	1.762	-0.046
	(0.056)	(0.046)	(0.054)	(0.06)	(0.055)	(0.048)	(0.038)	(0.032)	(0.057)	(0.098)	(0.114)
Res. coskew.$\times 10^3$	-0.267	0.302	0.545	0.199	0.438	0.211	0.144	0.107	0.474	1.543	-1.724
	(0.285)	(0.252)	(0.217)	(0.24)	(0.263)	(0.213)	(0.174)	(0.238)	(0.342)	(0.811)	(1.012)