ALMA MATER STUDIORUM — UNIVERSITÀ DI BOLOGNA

Dottorato di ricerca in Economics

Ciclo 34

Settore Concorsuale: 13/A1 - ECONOMIA POLITICA

Settore Scientifico Disciplinare: SECS-P/01 - ECONOMIA POLITICA

Essays on Asset Pricing

Presentata da: Fabio Franceschini
f.franceschini@unibo.it

Coordinatore Dottorato
Andrea Mattozzi

Supervisore
Martín Gonzalez-Eiras

Co-supervisore
Mariano Massimiliano Croce

Esame finale anno 2023

I'm not even supposed to be here.

– LeBron James, *2013 NBA Finals Trophy presentation*

Acknowledgements

Special thanks to Martín Gonzalez-Eiras and Max Croce for their valuable supervision and the many insightful discussions. I am also grateful to Howard Kung and Svetlana Bryzgalova for conversations from which I greatly benefited. Finally, I also thank Oliviero Pallanch, Luca Fanelli, Giuseppe Cavaliere and seminar participants at the DSE PhD Forum, WiP series and Macro-Finance reading group for their useful comments.

Abstract

The first chapter provides evidence that aggregate Research and Development (R&D) investment drives a persistent component in productivity growth and that this embodies a risk priced in financial markets. In a semi-endogenous growth model, this component is identified by the R&D in excess of equilibrium levels and can be approximated by the Error Correction Term in the cointegration between R&D and Total Factor Productivity. Empirically, the component results being well defined and it satisfies all key theoretical predictions: it exhibits appropriate persistency, it forecasts productivity growth, and it is associated with a cross-sectional risk premium.

CAPM is the most foundational model in financial economics, but is known to empirically underestimate expected returns of low-risk assets and overestimate those with high risk. The second chapter studies how risks omission and funding tightness jointly contribute to explaining this anomaly, with the former affecting the definition of assets' riskiness and the latter affecting how risk is remunerated. Theoretically, the two effects are shown to counteract each other. Empirically, the spread related to binding leverage constraints is found to be significant at 2% yearly. Nonetheless, average returns of portfolios that exploit this anomaly are found to mostly reflect omitted risks, in contrast to their employment in previous literature.

The third chapter concerns 'sustainability' of assets: when it is valued, its effect on the discount rates does not only depend on the sustainability of the asset priced, but it is intrinsically mediated by the risk profile of the asset itself. This has implications for the assessment of the sustainability-related spread and for hedging changes in the sustainability concern. Specifically, (1) long-short portfolios of assets sorted on sustainability can even average returns with the opposite sign of the spread and (2) the effectiveness of more sustainable assets in hedging changes to the sustainability premium will depend on their 'sustainability intensity' and their risk *jointly*. The main implications are tested on the ESG score dimension for US data and is inconclusive regarding the existence of a ESG-related premium in the first place. Also, the risk profile of the long-short ESG portfolio is not likely to impact the sign of its average returns.

Contents

List of Figures

List of Tables

Acronyms

ADF Augmented Dickey-Fuller

CAPM Capital Asset Pricing Model

CF Cash-flow

DOLS Dynamics Ordinary Least Squares

ESG Environmental, Social, and corporate Governance

LRA Low Risk Anomaly

PCA Principal Component Analysis

R&D Research and Development

SDF Stochastic Discount Factor

TFP Total Factor Productivity

1

The long-run innovation risk component

1 Introduction

To reconcile consumption-based asset pricing theory with the data, Bansal and Yaron (2004) focused on a ‘small’ but persistent component of consumption growth, named the ‘long-run risk’ (LRR) component. This process can add little variance to consumption growth despite heavily impacting the whole consumption path. Therefore, when coupled with preferences that are sensitive to uncertainty in future consumption expectations, as in Epstein and Zin (1989), it becomes a significant source of risk. Risks of this kind have proven useful in studying various macro-financial phenomena.[1] However, detecting LRR components empirically proves challenging, undermining the validation of mechanisms relying on them and drawing significant criticism towards the entire framework.[2] Given the extensive literature that has developed around the LRR concept, it is essential to provide evidence that supports its establishment: in this paper I contribute by directly documenting a LRR component related to innovation efforts, plotted in figure 1.1. More specifically, I empirically show that aggregate Research and Development (R&D) investment intensity is highly persistent, it forecasts Total Factor Productivity (TFP) growth, and it is associated to a positive risk premium in financial markets.

Existence and relevance of long-run risks have been previously corroborated either by directly tackling the statistical difficulties in its detection, as done for example by Ortu et al. (2013), Dew-Becker and Giglio (2016) and Schorfheide et al. (2018), or by framing their origin in richer structural models,

[1] For example, exchange rates dynamics as in Colacito and Croce (2011), climate change pricing as in Bansal, Ochoa, et al. (2016), term structures as in Ai et al. (2018), or oil dynamics as in Ready (2018).

[2] Most notably, Beeler and Campbell (2012) and Epstein, Farhi, et al. (2014).

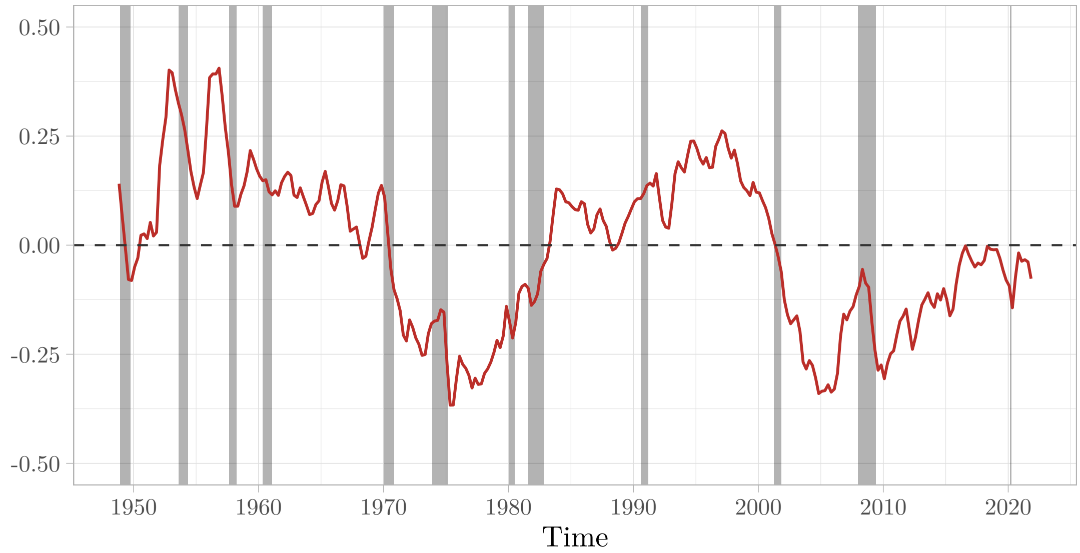

Figure 1.1: R&D excess intensity, see section 3 for details. Shaded areas mark NBER recessions.

which provide additional implications to test. Following the latter approach, Kaltenbrunner and Lochstoer (2010) first showed in general equilibrium how the long-run risk component can arise in consumption growth with standard productivity dynamics. Then, Croce (2014) went a step further, providing both theoretical arguments and empirical evidence for a long-run consumption risk component being originated in the persistence of the productivity growth process. This found additional support in Ortu et al. (2013), which found high correlation between the components with half-life within eight and sixteen years of consumption and TFP growth rates. Kung and Schmid (2015) moved one further step upstream, acknowledging the well-established role of R&D in spurring productivity growth and showing how a long-run risk component in consumption could ultimately be driven by an endogenous and persistent aggregate R&D investment intensity. The empirical evidence they provided to support this claim, however, relied on a measure of R&D intensity that, with updated data, shows undesirable statistical properties, most importantly an apparent non-stationarity. This paper improves on this, providing empirical evidence for a long-run risk originating in R&D efforts that is based on a more reliable R&D intensity measure. The crucial difference from Kung and Schmid (2015) is the definition of R&D intensity, which in this paper stems from a semi-endogenous growth model rather than a fully endogenous one: in semi-endogenous models R&D and TFP level are approximately cointegrated and it is the related Error Correction Term (ECT) to reflect the fluctuations in R&D intensity around the equilibrium level, which drive conditional expectations of productivity growth. As the estimated Error Correction Term proves being stationary, analysis relying on it are less likely to produce spurious results. A further novelty of this paper with respect to Kung and Schmid (2015) consists of the direct employment of the long-run innovation risk component in a cross-sectional test, which returns a positive and significant risk premium associated to assets' exposure to it, as expected.

The definition of R&D intensity is illustrated exploiting a semi-endogenous 'lab-equipment' R&D growth model, specifically.[3] The allocation of resources to R&D is modelled as a stochastic rule of thumb rather than by welfare optimality conditions, in a similar fashion to Jones (2005). This already provides enough structure to interpret the emerging cointegrating relation between R&D and TFP and identify the long-run innovation risk component in R&D fluctuations, while still ensuring a Balanced Growth Path in the economy. From a methodological point of view, as often done in the recent macro-finance literature, the cointegrating relation is estimated with the Dynamic OLS methodology studied in Phillips and Loretan (1991), Saikkonen (1991), and Stock and Watson (1993). In macroeconomics, cointegration methods has already been employed to study the relation between R&D and technological progress in different studies, such as Ha and Howitt (2007), Bottazzi and Peri (2007), and more recently Herzer (2022) and Kruse-Andersen (2023). These papers are mostly concerned with the assessment of foreign spillovers and the comparison of fully- versus semi-endogenous growth models, with more recent evidence leaning towards semi-endogenous ones, as also backed by Bloom et al. (2020). This paper does not contribute directly to the debate on the modelling comparison, rather it leverages the growing empirical evidence in favor of semi-endogenous growth models to perform more effective empirical analysis and test it on the new grounds, namely that of financial economics. More broadly, cointegration in macroeconomic variables has been widely exploited to price assets, with notable examples in Lettau and Ludvigson (2001) and Melone (2021), but to my knowledge this is the first application to relate aggregate R&D and asset pricing.

On the financial side of the analysis, the most important contribution of the paper is providing empirical evidence that persistent swings in R&D activities is a priced risk in the markets. This evidence comes from cross-sectional pricing tests pioneered by Bansal, R. F. Dittmar, et al. (2005), and developed by Bansal, R. Dittmar, et al. (2009) among others. Like Bansal, R. F. Dittmar, et al. (2005), in this paper I focus on the risk premium related to cash-flow growth rates' sensitivities rather than returns' sensitivities, as discount rates may be impacted by many more factors. In this regard, I

[3] This class of models was introduced in Romer (1987), and is characterized by the use of units of the final output good to produce ideas, instead of using labor as in more traditional cases à la Romer (1990).

obviously depart from them by focusing on sensitivities to the estimated R&D intensity, which ends up showing a stronger risk premium than those to consumption growth. I also contribute by considering a wider set of test asset portfolios: in particular, I include portfolios sorted on firm-specific R&D. This is interesting because this sorting leads to the greatest dispersion in cash-flow growth rates across portfolios and it is a dimension likely relevant for heterogeneity in sensitivities to aggregate R&D. Indeed, a clear pattern emerge: cash-flows of more R&D intensive firms prove being much more positively sensitive to aggregate R&D intensity, meaning that cash-flows of more R&D-intensive firms grow more when other firms invest more in R&D too. This is in line with both R&D-intensive firms showing higher excess returns and spillover effects being stronger than the fishing-out effect, as previously shown by Jiang et al. (2016). For the cross-sectional pricing test, a traditional Fama and Macbeth (1973) is employed.

The rest is structured as follows: in section 2 I show the fragilities of the R&D intensity measure studied in Kung and Schmid (2015) and I outline the emergence of a long-run innovation risk component in a semi-endogenous growth models as well as its asset pricing role; in section 3 I show the results from the estimation of this measure and I proceed illustrating its proprieties, forecasting power with respect to TFP and relation with markup and funding conditions, which are known to interact with R&D investment; in section 4 I carry out the cross-sectional pricing test; in section 5 I conclude.

2 The R&D component of long-run productivity risk

2.1 Background

The starting point of this study is the law of motion of the aggregate intangible capital stock N in Kung and Schmid (2015), which, using their notation is

$$N_{t+1} = (1-\phi)N_t + \chi \left(\frac{S_t}{N_t}\right)^{\eta} N_t, \tag{1.1}$$

where N_t can be interpreted as a measure of patented ideas and S_t of R&D expenditure. ϕ controls ideas' obsolescence rate, while η captures both the duplication effects in innovative efforts and spillover effects from past innovations. In their economy, the intangible capital contributes to equilibrium final goods production via technology

$$Y_t = Z_t K_t^{\alpha} L_t^{1-\alpha} \qquad \text{where} \qquad Z_t \equiv \bar{A}(e^{a_t} N_t)^{1-\alpha}, \tag{1.2}$$

with K_t being the physical capital, L_t the amount of labour, Z_t the standard Solow residual, and α the capital share. a_t is assumed to be a stationary process with innovations ε_t. When a_t is highly persistent and the TFP growth rate is small, the latter can be simply approximated as

$$\Delta \ln Z_{t+1} \approx (1-\alpha)\left[\chi\left(\frac{S_t}{N_t}\right)^{\eta} - \phi\right] + (1-\alpha)\varepsilon_{t+1}. \tag{1.3}$$

This formulation highlights the crucial role of the ratio S_t/N_t, there defined as 'R&D intensity': as its dynamics drive conditional expectations of TFP growth, any persistent movement in it translates into a source of long-run productivity risk in the sense of Croce (2014).

Kung and Schmid (2015) show theoretically that innovation efforts endogenously driven by the fluctuations of profitability level set by the exogenous process a_t spontaneously lead to persistence in growth prospects, jointly rationalizing macroeconomic and asset prices dynamics. To support the idea that R&D intensity identifies a long-run risk component, they study its empirical counterpart formed as the raw ratio of US annual private R&D expenditure from the National Science Foundation, measuring S_t, over the R&D stock series estimate by the US Bureau of Labor Statistics, representing intangible capital N_t. This measure of R&D intensity proves indeed being highly persistent and co-moving at low frequencies with the price-dividend ratio as well as forecasting the growth rates of consumption, GDP and TFP.

However, this approach has a few potential shortcomings. Indeed, their R&D intensity measure, as in the natural logarithm of S_t/N_t, is *extremely* persistent, with a point estimate of the yearly first autocorrelation equal to 0.987 and a standard error of 0.005.[4] It should be noted that the R&D stock series has been updated by the Bureau of Labor Statistics with respect to the one used in their paper and now covers a slightly different time period. Anyway, the 95% confidence interval, which spans from 0.977 to 0.998, highlights two potential issues with the use of this measure: the upper bound, being so close to 1, shows that the process could well be non-stationary while the lower bound is still so high that makes it unlikely for this process to identify a long-run risk component in the economy related to productivity. Sample non-stationarity is not critical to the validity of the measure and the theory it is used to support: from a statistical point of view, R&D intensity is expected to be persistent and the more persistent a process is, the harder it becomes to assess its stationarity in finite samples, so the generating process could still be stationary. Then, even if R&D intensity really was non-stationary the key mechanism studied by Kung and Schmid (2015) could still hold, at the price

[4] The standard error is obtained using the Delta method and 1-step GMM estimates of the fundamental moments.

Table 1.1: statistics of R&D intensity measure from Kung and Schmid (2015). In the first column, S is yearly R&D expenditure from the National Science Foundation and N is the R&D stock from Bureau of Labor Statistics, spanning 1963 to 2020; in the second and third column, S is quarterly real R&D expenditure from Bureau of Economic Analysis and Z is the quarterly utilization-adjusted TFP from Fernald (2012), spanning from 1947 Q1 to 2021 Q4. ADF u.r. stat is the statistic of the unit root coefficient in an Augmented Dickey-Fuller test with a time trend. AC(1) is the first autocorrelation, estimated as cross-correlation with the lagged value via 1-step GMM, and in parenthesis there are the HAC standard error recovered via Delta-method.

	$(\ln S_t - \ln N_t)$	$(\ln S_t - \frac{1}{1-\alpha} \ln Z_t)$	
α	–	0.35	0.3
ADF u.r. stat	-2.55	-2.11	-2.09
AC(1)	0.989	0.999	1.000
	(0.006)	(0.000)	(0.000)
Num. obs.	57	299	299

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

of a more complex model. Nonetheless, non stationarity of R&D intensity undermines the regressions in which it is employed, as any results would essentially be spurious. Unfortunately, the Augmented Dickey Fuller test with trend delivers for this series a statistic for the unit-root coefficient of -2.21, which is well above the 10% critical value of -3.15, thereby suggesting that the series is highly likely non-stationary. Furthermore, even with enough evidence backing its stationarity, another concerning issue is that this series' first autocorrelation is likely at least 0.977, which implies a half-life of shocks over 30 years for a yearly AR(1) process. This suits the long-run risk component in consumption calibrated by Bansal and Yaron (2004), but it is way more persistent than the component that Ortu et al. (2013) find consumption and productivity to share more strongly in the data, which has half-life between eight and sixteen years, corresponding to a maximum autocorrelation of 0.957 if modelled as a yearly AR(1). Therefore, all in all, this measure could well identify a long-run risk source in the economy, but the empirical evidence for it effectively identifying the long-run productivity risk originated in the R&D investment is fragile.

There are a few details that may drive this measure away from its aim. First, the measure of intangible capital stock used, which in this case is the stock of R&D. This might open a wedge between the model and the data because intangible capital, as shown in (1.1), is formed in a very different way than simple accumulation and depreciation of R&D expenses – R&D investments unlikely have constant marginal returns, considering duplication and spillover effects. Another related issue is that the production of ideas is likely not to rely solely on domestic R&D expenditure and stock of ideas anyway. This makes it difficult to rely on any measure of intangible capital stock for an empirical analysis because it would require to account for all spillover sources relevant to the formation of new patented ideas and make strong assumptions on the functional form to combine them. A way to bypass this issue could be to utilize directly the variable that the very concept of ideas' stock was born to drive and explain: Total Factor Productivity in the form of Solow residual. This quantity was born in the data and requires little structure to be identified.

Combining (1.2) and (1.3), with the assumption that the first moment of $\ln(S/N)$ exists and the log-ratio does not deviate too much from it, TFP growth can be fairly approximated in terms of S and Z as

$$\mathbb{E}_t \left[\Delta \ln Z_{t+1}\right] \approx \gamma_0 + \gamma_1 a_t + \gamma_1 \left(\ln S_t - \frac{1}{1-\alpha} \ln Z_t\right). \tag{1.4}$$

Here, conditional expectations of ΔZ_{t+1} are impacted by R&D investment via $(\ln S_t - \frac{1}{1-\alpha} \ln Z_t)$, which contains the R&D fluctuations that moves TFP growth expectations around; a_t enters the equation mechanically following the substitution of N_t with $\bar{A}^{\frac{-1}{1-\alpha}} e^{-a_t} Z^{\frac{1}{1-\alpha}}$ and can be thought as merely compensating its implicit presence in the 'modified R&D intensity' term. On one side, this formulation is empirically convenient because it provides a way to measure TFP growth expectations involving variables that have more obvious empirical counterparts than (1.3) – a_t is not directly observable but can be represented by a combination of processes that are able to forecast TFP growth. Note that similar manipulations could be performed to express TFP growth in terms of other observables too, but it would require involving much more theoretical structure than just TFP definition and ideas' law of motion. An effective description of the data would then require more complex models, with greater chances of being misspecified. On the other side, however, this formulation makes it harder to identify the source of fluctuations in growth prospect: even if R&D intensity S/N and expected TFP growth were actually constant, $(\ln S_t - \frac{1}{1-\alpha} \ln Z_t)$ would still fluctuate, even persistently, because of a_t in Z_t. This could be amended recovering a_t from the forecasting regression and filtering it out of $(\ln S_t - \frac{1}{1-\alpha} \ln Z_t)$, but this would need γ_1 to be estimated consistently, which is difficult given the high non stationary behaviour shown by $(\ln S_t - \frac{1}{1-\alpha} \ln Z_t)$: this can be observed looking at table 1.1 where it is reported the unit root coefficient statistic from the ADF test performed on the series of $(\ln S_t - \frac{1}{1-\alpha} \ln Z_t)$ built using quarterly US R&D expenditure from the Bureau of Economic Analysis and US utility-adjusted TFP estimated by Fernald (2012),[5] for two different values of α.

A further point of departure of the theory from reality might be represented

[5]More details on the data are provided in the next section.

by the strong scale effects in the model. As highlighted by Bloom et al. (2020), there is wide evidence for decreasing research productivity in the data, so this is likely a realistic feature that is necessary for a model to be applied empirically. To take this into account, in the next section I formulate a model based on the key dynamics of Kung and Schmid (2015) reframed in a semi-endogenous framework, where I directly model R&D expenditure with a stochastic thumb rule, which allows to address R&D intensity dynamics more explicitly.

2.2 A simple data-driven model

Consider a discrete-time model where goods Y_t are produced using only labour L_t and intangible capital I_t with technology

$$Y_t = e^{a_t} I_t^{\xi} L_t, \tag{1.5}$$

characterized by the degree of increasing returns ξ. a_t is a stationary exogenous process with innovations ε_t^a to keep track of possible external factors affecting the dynamics of output level. As common in the 'lab-equipment' literature, final goods can be employed in consumption C_t and R&D expenditures S_t,

$$Y_t = C_t + S_t. \tag{1.6}$$

R&D investment implicitly employs both capital and labor, and produces new intangible capital with a schedule that embodies the insight of the semi-endogenous growth theory that ideas get harder to find. Specifically, the law of motion of the intangible capital stock is

$$I_{t+1} = (1-\phi)I_t + {S_t}^{\eta} {I_t}^{\psi}. \tag{1.7}$$

Here, duplication effects are controlled via η and are independent from how easier/harder finding ideas gets, set by ψ, which is the net result of spillovers and fishing-out effects. ϕ controls ideas' obsolescence rate. This formulation nests both fully endogenous models, which are described by setting $\psi = 1$, and Kung and Schmid (2015), which requires $\psi = 1 - \eta$.

In this economy, the Solow residual, or TFP, Z_t is equal to $e^{a_t} I_t^{\xi}$ so, performing approximations similar to those that led to (1.4), its growth rate can be written as

$$\Delta \ln Z_{t+1} \approx \gamma_0 + \gamma_1 \left(\ln S_t - \Psi \ln I_t \right) + \varepsilon_{t+1}^a, \tag{1.8}$$

where Ψ is short-hand for $\frac{1-\psi}{\eta}$, which is also the key difference with (1.3). In a way, the semi-endogenous approach makes obtaining an empirical measure of R&D intensity, now intended as S_t/I_t^{Ψ}, more difficult because it demands the calibration of an additional coefficient. On the other hand, the linear relation of $\Delta \ln Z_t$ and $(\ln S_t - \Psi \ln I_t)$ combined with the wide support for a stationary $\Delta \ln Z_t$ implies a cointegration relationship between S_t and I_t, thereby enabling a direct estimation of Ψ.

The cointegration between S_t and I_t suggests a stochastic rule of thumb for the allocation of the R&D expenditures:

$$S_t = e^{\bar{r}+\tilde{r}_t}\ I_t^{\Psi}, \tag{1.9}$$

with $\tilde{r}_t$ being a stationary process. This rule can be interpreted as S_t being a stochastic proportion of I_t^{Ψ} fluctuating around the fixed value $e^{\bar{r}+\mathrm{Var}[\tilde{r}]/2}$. The steady state in this economy is on a balanced growth path where output, consumption and R&D investment grow at rate $g = \frac{\Psi}{\Psi-\xi} g_L$ while TFP grows at rate $g_{\mathrm{TFP}} = \frac{\xi}{\Psi-\xi} g_L$. S_t could be set or described in a different manner, but this way of approximating and representing it allows to directly address the fluctuations in conditional expectations of TFP growth, which, depending only on how S_t and I_t^{Ψ} relate to each other, are completely described in one process, $\tilde{r}_t$. Following this, I will refer to $\tilde{r}_t$ as 'excess R&D intensity', or 'excess innovative efforts', as in excess of the long-run equilibrium level. Finally to close the model, for simplicity of exposition, assume labour being entirely devoted to final goods production and growing at an exogenous rate g_L.

This rule of thumb directly implies cointegration between $\ln S_t$ and $\ln I_t$, with $\tilde{r}_t$ being the error correction term, and implies a neat expression for TFP growth:

$$\Delta \ln Z_{t+1} \approx \mu + \gamma \cdot \tilde{r}_t + \varepsilon^a_{t+1}. \tag{1.10}$$

This formulation traces very closely the productivity process used in Croce (2014), illustrating the mapping between the typical productivity long-run component x_t and the excess R&D intensity $\tilde{r}$, which has the potential to be a 'long-run *innovation* risk component'. The only missing piece for the two specifications to be completely equivalent concerns the substantial persistence in x_t dynamics, which leaves open the issue of whether $\tilde{r}$ is persistent enough to identify it. Kung and Schmid (2015) have provided an exhaustive theoretical answer showing its emergence in the optimization solution, I answer this empirically in the following section.

The thumb rule can naturally be expressed in terms of TFP instead of intangible capital:

$$\ln S_t - \frac{\psi}{\xi} \ln Z_t = \bar{r} + \underbrace{\tilde{r}_t - \frac{\psi}{\xi} a_t}_{\hat{r}_t}, \tag{1.11}$$

meaning that the residual from the regression of $\ln S_t$ on $\ln Z_t$, $\hat{r}_t$, actually includes the level of any process that affect productivity levels other than intangible capital. In terms of $\hat{r}_t$, TFP growth is then:[6]

$$\Delta \ln Z_{t+1} \approx \mu + \gamma \frac{\psi}{\xi} a_t + \gamma \cdot \hat{r}_t + \varepsilon^a_{t+1}. \tag{1.12}$$

Note that the coefficient of $\tilde{r}$ and $\hat{r}$ is the same.

2.3 The pricing of long-run innovation risk

The key object of study in asset pricing is the Stochastic Discount Factor (SDF) m_t. This is a stochastic process that tracks the growth in marginal utility of investors in a market, thus reflecting the shocks to the economy state variables that are relevant to them. In a typical long-run risk models this takes the form

$$m_{t+1} = \bar{m}_t - b_x \varepsilon_{x,t+1} - b_s \varepsilon_{s,t+1} \tag{1.13}$$

with $\bar{m}_t$ being its expectations conditional on previous-period information, and $\varepsilon_{x,t+1}$ and $\varepsilon_{s,t+1}$ being innovations that affect consumption marginal utility persistently and transiently with loadings b_x and b_s, respectively.

The SDF plays a relevant role in studying assets price dynamics because, in perfect markets, assets' expected returns in excess of the risk-free return are determined by the level of return innovations exposure to it, as the following holds:

$$\mathbb{E}_t \left[R^i_{t+1}\right] - R^f_t = -R^f_t \cdot \mathrm{Cov}_t \left[m_{t+1}, R^i_{t+1}\right]. \tag{1.14}$$

I complement these conditions with the assumption of the following factor structure for returns,

$$R^i_{t+1} = \bar{R}^i_t + \beta^i_x \varepsilon_{x,t+1} + \beta^i_s \varepsilon_{s,t+1} + e^i_{t+1}, \tag{1.15}$$

where β^i_j is the sensitivity of the return i to shocks of the variable j. A potential driver of heterogeneity in sensitivities to the long-run innovation risk is represented by the firm-specific R&D intensity: evidence from Jiang et al. (2016) shows that R&D spillovers do get priced in financial markets, so it is

[6] Full derivation in Appendix A.

sensible to hypothesize that fluctuations in the aggregate R&D investment leads to different return dynamics depending on the externality a firm can enjoy. This will be explored in the empirical analysis by looking at the sensitivity distribution over firm-specific-R&D-sorted stocks portfolios, but, as more mechanisms could play a role, there is ground for further investigation. Combining (1.13), (1.14) and (1.15) yields the main reduced-form pricing equation,

$$\mathbb{E}_t\left[R^i_{t+1}\right] - R^f_t = \lambda_x \beta^i_x + \lambda_s \beta^i_s, \tag{1.16}$$

with λ_j being the so called 'risk premium' associated with risk factor j.

The main take-away of the long-run risk models is that the heavy lifter in explaining stocks risk premium is, by orders of magnitude, the risk of exposure to the factor affecting the marginal utility in a persistent manner, i.e. β^i_x. So, for the sake of presentation clarity, I will proceed focusing on the cross-sectional pricing equation

$$\mathbb{E}_t\left[R^i_{t+1}\right] - R^f_t = \lambda_x \beta^i_x \tag{1.17}$$

instead. The long-run risk factor in the seminal paper by Bansal and Yaron (2004) is a persistent component of consumption directly; in Croce (2014) this is a persistent component in productivity growth instead; Kung and Schmid (2015) imputes this last one to persistency in R&D investment, which, following the framework of the previous section, should be embedded in $\tilde{r}_t$. This translates to $\tilde{r}$ being a risk factor priced in the cross-section, which will be investigated later by looking at the significance of $\lambda_{\tilde{r}}$. This obviously represents a joint test of the theory and of the empirical identification of $\tilde{r}$, which I additionally address before performing the financial analysis.

Next, consider the decomposition shown in Campbell (1996), where returns innovations can be approximated as the sum of news to cash-flow growth rates and to discount rates:

$$\ln R^i_{t+1} - \mathbb{E}_t\left[\ln R^i_{t+1}\right] = \delta^i_{D,t+1} - \delta^i_{R,t+1} \qquad \text{where}$$

$$\delta^i_{D,t+1} = \{\mathbb{E}_{t+1} - \mathbb{E}_t\}\left[\sum_{j=0}^{\infty} \kappa^j \Delta \ln D_{i,t+j}\right] \quad \text{and} \quad \delta^i_{R,t+1} = \{\mathbb{E}_{t+1} - \mathbb{E}_t\}\left[\sum_{j=1}^{\infty} \kappa^j \ln R^i_{t+j}\right]. \tag{1.18}$$

Then, as

$$\beta^i_r = \frac{\mathrm{Cov}\left[R^i_t, \tilde{r}_t\right]}{\mathrm{Var}\left[\tilde{r}_t\right]} \approx \frac{\mathrm{Cov}\left[\delta^i_{D,t}, \tilde{r}_t\right]}{\mathrm{Var}\left[\tilde{r}_t\right]} - \frac{\mathrm{Cov}\left[\delta^i_{R,t}, \tilde{r}_t\right]}{\mathrm{Var}\left[\tilde{r}_t\right]} = \beta^i_{r,D} - \beta^i_{r,R}, \tag{1.19}$$

the cross-sectional pricing equation can be expressed as

$$\mathbb{E}_t\left[R^i_{t+1}\right] - R^f_t = \lambda_r \beta^i_{r,D} - \lambda_r \beta^i_{r,R}. \tag{1.20}$$

Following Bansal, R. F. Dittmar, et al. (2005), being the long-run risk premium a phenomenon that mainly concerns assets fundamentals, rather than conditional discount rates, I focus on the cash-flows' exposure $\beta_{r,D}$ to aggregate R&D intensity, as key dimension of risk to explain excess returns.

3 Empirical long-run excess R&D intensity

The long-run innovation risk is embodied in $\tilde{r}_t$, which is the key object of the analysis. From the theoretical assumption of the allocation rule in (1.9) or the simple combination of assuming the ideas production function in (1.7) and the empirical observation of a stationary TFP growth, this could be directly estimated as the error correction term of the cointegration relationship between $\ln S_t$ and $\ln I_t$. I explore this by regressing a measure of ideas stock on a measure of R&D expenditure. The former is built in a spirit similar to Bottazzi and Peri (2007), i.e. recursively adding new patents, from the quarterly series from USPTO, to a depreciated value of past patents stock. The depreciation rate is assumed to be 0.15, a value that is in line with most of the literature, while higher than the one used by Bottazzi and Peri (2007) and lower than the one advocated by W. Li and Hall (2016). Different values leads to similar conclusions, so they are not shown. The empirical measure of R&D that I employ here, and through out the rest of analysis, is the quarterly private R&D expenditures series expressed in chained 2012 US Dollar prices provided by the Bureau of Economic Analysis in the National Income and Product Accounts tables.[7]

The results of a DOLS estimation of the cointegrating relationship are shown in the first two columns of Table 1.2, and they are not encouraging: the cointegration coefficient β_S is statistically insignificant with the addition of a time trend and the error correction terms are never stationary. This was to be expected to some extent, as patents are widely considered to be not a good measure of successful innovation, see for example Reeb and Zhao (2020) and Herzer (2022). Therefore, the analysis will focus instead on the error correction term $\hat{r}_t$ of the empirical counterpart of (1.11),

$$\ln S_t = b_0 + b_1 \ln Z_t + \hat{r}_t. \tag{1.21}$$

[7] The real series is obtained deflating the nominal R&D series Y006RC of table 5.3.5 by the deflator series Y006RG of table 5.3.4.

It should be remarked that $\hat{r}_t$, which equals $\tilde{r}_t - \frac{\psi}{\xi} a_t$, does not directly identify the R&D intensity $\tilde{r}$, which is the persistent component in TFP growth conditional expectations. In fact even with fixed $\tilde{r}$, one could still observe fluctuations in $\hat{r}$, with these being due to external factors acting on the level of TFP but not on the growth rate expectations at all. Nonetheless, assuming that a_t is spanned by some available factors $\mathbf{f}_t$, one could still measure the impact of $\tilde{r}$ on expected TFP growth rates, i.e. 'γ' in (1.10) and (1.12), by estimating k_r with

$$\Delta Z_{t+1} = k_0 + \mathbf{k}_f' \, \mathbf{f}_t + k_r \hat{r}_t + u_{t+1}. \tag{1.22}$$

In principle this also allows to explicitly recover $\tilde{r}_t$ by exploiting its definition, $\tilde{r}_t = \hat{r}_t + \frac{\psi}{\xi} a_t$:[8]

$$\tilde{r}_t = \hat{r}_t + \frac{\mathbf{k}_f' \, \mathbf{f}_t}{k_r}. \tag{1.23}$$

However, as will be shown later, it turns out that $\hat{r}$ is likely to identify $\tilde{r}$ already, for the purposes that are relevant to this project at least. I will now first go through the estimation of $\hat{r}$ and then proceed illustrating its case to be a close approximation of $\tilde{r}$.

3.1 Estimation of $\hat{r}$

I estimate parameters of (1.21) via DOLS to avoid imposing any structure on the short-term dynamics. This implies estimating

$$\ln Z_t = \beta_0 + \beta_S \ln S_t + \beta_{tt} t + \sum_{i=-J}^{K} \beta_{\Delta i} \Delta S_{t-i} + u_t. \tag{1.24}$$

Terms are later re-arranged to form $\hat{r}_t = \ln S_t - \frac{1}{\beta_S} \ln Z_t + \frac{\beta_0}{\beta_S} + \frac{\beta_{tt}}{\beta_S} t$. The TFP series employed to measure Z is the quarterly utilization-adjusted series by Fernald (2012), which, paired with the R&D series, cover from 1947 to 2021. Since R&D is a flow variable that measures expenditures all along the quarter ending at time t, which in principle is continuously chosen by the agents between $t-1$ and t, while TFP level is a stock variable, to match the timing of economy state and economic choices at best, the main specification will refer to Z_t as the interpolated value of TFP between $t-1$ and t. At the same time, ΔZ_{t+1} will simply be the difference between utility-adjusted TFP at time $t+1$ and time t, to make TFP movements completely subsequent to

[8] Another way is to add the same factors to the cointegration estimation and directly obtain $\tilde{r}$. However this option seems less sensible because a large number of potential regressors leads to poor estimation accuracy and the estimation of an impractical number of regressions to perform a formal model selection.

Table 1.2: cointegration results. HAC Standard Errors in parenthesis, computed as advised by Lazarus et al. (2018). BIC values refer to the estimation of the same specifications on a sample where the first 32 observations were trimmed to allow for a fair comparison in model selection. AC(1) is the first autocorrelation estimated as cross-correlation with the lagged value, via 1-step GMM, whose HAC Standard Errors, below in parenthesis, are obtained via Delta-method.

	$\ln I$		$\ln Z$		$\ln Z$ (unadj.)
	(1)	(2)	(1)	(2)	(3)
β_S	0.09***	0.06	0.20***	0.18***	
	(0.01)	(0.11)	(0.01)	(0.03)	
$\beta_{S,T}$					0.17***
					(0.02)
β_{tt}		0.0004		0.0003	0.0005
		(0.0019)		(0.0004)	(0.0003)
J	0	0	0	0	0
K	11	11	14	15	11
BIC	−1087.9	−1088.5	−1095.3	−1092.0	
	$\tilde{r}_t$		$\hat{r}_t$		
Num. obs.	171	171	283	282	283
SD	11.8%	16.72%	18.2%	21.5%	18.8%
ADF u.r. stat	−3.09	−2.85	−3.89**	−3.82**	−3.44**
AC(1)	0.986	0.987	0.979	0.982	0.964
	(0.002)	(0.002)	(0.002)	(0.002)	(0.004)

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

any R&D expenditure between time $t-1$ and time t. This peculiar timing structure is a further motivation to estimate the cointegrating parameters using DOLS instead of estimating a full Vector Error Correction Model. As a robustness check, the results from the estimation using the raw TFP series from Fernald (2012) (no utilization adjustment and no timing-adjustment) are also reported in the last column of Table 1.2, where also a broader measure of R&D is employed - private plus government R&D expenditure. The results are extremely similar, with a cross correlation with the $\hat{r}$ of the specifications marked in Table 1.2 as (1) and (3) of 0.97%.

The formulation in (1.24) nests all the specifications tested. As advised by Choi and Kurozumi (2012), numbers of leads and lags are selected independently, i.e. J needs not be equal to K, and the selection is based on the Bayes Information Criterion (BIC). Specifically, leads of ΔS_t turn out to never be significant, so I focus on specifications with $J = 0$ and compare BIC values of the models estimated on a trimmed sample that allows fair comparisons up to $K = 32$ (8 years). Leads of ΔS_t never being significant also motivates keeping Z on the left-hand side: with this formulation all the first-differences of the regressor are lags, making the estimation based on the most recent observations of Z and S levels; vice versa, leaving S as a dependent variable would otherwise make the estimation performed on a dataset without the most recent observations in levels, lost to the missing leads of ΔZ_t. Table 1.2 shows the estimation results of the best performing specification with and without a time trend.

β_S and unit root ADF statistic of the error correction term are found to be significant, supporting the cointegration of S and Z. The time trend existence, on the other hand, finds little support. Therefore, the preferred specification, which will be employed in the rest of the paper, is the one used in column '$\ln Z(1)$' of Table 1.2. For this series, the first autocorrelation is 0.979, which fits in the range expected from an AR(1) long-run productivity risk component per Ortu et al. (2013) results – between 0.979 and 0.989. Correlation among the ECTs of the different specification can be seen in appendix at Table 1.8.

3.2 Forecast the TFP growth

The key property of $\tilde{r}$ is that it is supposed to drive conditional expectations of TFP growth, therefore it should display a strong forecasting ability. Employing $\hat{r}$, this can be tested estimating the regression of equation (1.22), where factors $\mathbf{f}$ are added as controls to capture the exogenous factor a_t hidden in $\hat{r}$, which can bias the estimates. Specifically, note that in the extreme case in which no controls are considered at all, and one is to estimate univariate regression of future TFP growth on $\hat{r}$, the OLS-estimated slope is obviously expected to be biased.[9] Specifically,

$$\hat{k}_r = k_r \frac{\sigma^2_{\tilde{r}}}{\sigma^2_{\tilde{r}}(1 - \frac{\psi}{\xi}d) + \left(\frac{\psi}{\xi}\right)^2 \sigma^2_a}, \tag{1.25}$$

where d is the slope coefficient of the auxiliary regression of a_t on $\tilde{r}$, expected to be positive from Kung and Schmid (2015) and left unspecified in section 2.2. Correlation of a_t with TFP growth on the other hand is assumed to be close to 0, following theory in assuming a persistent a_t. It can be seen that $\hat{k}_r$ gets inflated for $\frac{\psi}{\xi}\sigma_a < \rho_{a,\tilde{r}}\sigma_{\tilde{r}}$, i.e. depending on the degree to which variations in $-\frac{\psi}{\xi}a_t$ go to 'compensate' variations of $\tilde{r}$, compressing the volatility of $\hat{r}$ without affecting its covariance with TFP growth rates.

To avoid this, I consider the sets of factors already employed for the same purpose in Ai et al. (2018): the main one is composed by price-dividend ratio, 3-month Treasury-bill yield, 3- and 5-year Treasury bond yields, and integrated volatility of the CRSP stock market index; a back-up one is formed by the 9 factors studied in Ludvigson and Ng (2009). The first set is preferred because it is available for a significantly longer timespan, starting in 1941 against the other one starting in 1970.

[9] Full derivation in section C.

Table 1.3: TFP growth forecast regression results. TFP growth is the utilization-adjusted TFP growth from Fernald (2012); controls in (BS) specification are the predictive factors used in Bansal and Shaliastovich (2013) plus market integrated volatility, as in Ai et al. (2018); controls in (LN) specification are the factors computed in Ludvigson and Ng (2009)

	(BS)	(LN)	(uv)
(Intercept)	0.0034***	0.0030***	0.0031***
	(0.0007)	(0.0005)	(0.0005)
$\hat{r}_t$	0.0121***	0.0113***	0.0123***
	(0.0031)	(0.0025)	(0.0026)
p.v. (F_{controls})	21.2%	16.0%	–
R^2	9.18%	8.81%	6.90%
Adj. R^2	7.27%	4.88%	6.58%
Num. obs.	292	243	292

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

The results are in table 1.3. The most relevant facts are, first, that $\widehat{k}_r$ is extremely significant with both sets of controls, and well in the confidence interval of the univariate estimate; second, that the control factors coefficients are jointly insignificant. These results can be interpreted in different ways: (1) both sets of factors are simply poor controls of a_t; (2) a_t is not there; (3) a_t is not really persistent. To see why persistency of a_t is relevant here note that the forecasting coefficient of a_t shown in equation (1.12) is a fair approximation only if ρ_a is close to 1, otherwise it reads $\rho_a - 1 + \frac{\psi}{\xi}k_1$: this could well be 0 even with a_t being very much alive. However, note that this, considering the previous estimate $\frac{\psi}{\xi}k_1 \approx 6\%$, would imply a quarterly $\rho_a = 0.94$ – a process with a half-life shorter than 3 years. In this case, it is true that $\hat{r}_t$ would not be strictly identifying $\tilde{r}_t$ because it would be determined by both $\tilde{r}_t$ and a_t, but most of the low-frequency fluctuations in $\hat{r}$, which is what this paper mostly concerns about, would be generated by $\tilde{r}$. Therefore, even in this case, $\hat{r}$ would correctly identify the persistent component originated in R&D intensity for our purposes. For the possibility of both sets of factors being poor controls there are not trivial solutions, other than testing even more sets.

3.3 Investigating fluctuations determinants

While proving causal relations are beyond the scope of this paper, it is informative to outline the dynamic relation of R&D excess intensity with other macroeconomic variables. Specifically, R&D investment in Kung and Schmid (2015) is driven by markup level, but it is well known that financial constraints play a role too in the R&D investments dynamics, see for example Brown et al. (2012) and D. Li (2011). This might matter for both the macroeconomic 'origin' of the long-run innovation risk itself as well as for

Table 1.4: estimates of the $\hat{r}$ regression from the VAR. In parenthesis, estimates' standard errors; 'max |roots|' is the maximum eigenvalue of the companion matrix estimated. Sample from 1970 Q2 to 2017 Q4.

	$\hat{r}$	ΔMark-Up	ΔI.C.R.
Lag: 1	1.520^{***}	-0.058	0.146
	(0.072)	(0.131)	(0.259)
Lag: 2	-0.756^{***}	0.008	0.341
	(0.121)	(0.137)	(0.258)
Lag: 3	0.213^{***}	-0.183	0.654^{**}
	(0.071)	(0.131)	(0.258)
T	R^2	$p(F)$	max |roots|
188	0.978	0	0.976

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

the determination of assets' sensitivities to this risk. To explore the dynamic relation between R&D intensity, mark-up and funding conditions, I estimate a VAR with endogenous variables being $\hat{r}$, the first principal component of the 5 measures of mark-up from Nekarda and Ramey (2020), which predicts 89% of the series' variance, and the intermediary capital ratio from He et al. (2017). Both the mark-up and the intermediaries funding conditions series result being non-stationary, with the ADF test unit-root statistics of -2.63 and -2.38 respectively; for this reasons I employ their first differences. The number of lags fixed for the VAR is 3, chosen by minimizing the AIC over a sample that allowed for a fair comparison up to 10 quarters. In Table 1.4 I report the results of the $\hat{r}$ regression.

The inverse root value indicates that the VAR is not too far from an explosive behaviour, but it is still stationary. What is most impressive from these results is that while mark-up does not show any predictive power with respect to R&D excess intensity, intermediaries' capital ratio does, with a highly significant coefficient when lagged thrice. While this is not conclusive, it is suggestive of a role for aggregate funding conditions on R&D and the long-run risk, which calls for deeper research.

4 Cross-sectional risk premium

The key asset pricing implication of swings in R&D intensity generating persistent fluctuations in expected growth rates of the economy, is that asset returns covarying more with R&D intensity should be regarded as riskier and be held for a higher compensation, i.e. a higher expected excess return. Following Bansal, R. F. Dittmar, et al. (2005), this hypothesis is tested in the cross-section of US stocks, by forming portfolios based on stocks sorts that give rise to a documented spread in average excess returns and testing whether the differences in sensitivities of these portfolios' cash flows to aggregate R&D intensity are related to the differences in excess returns in a manner consistent with theory.

4.1 Test assets

Following Bansal, R. F. Dittmar, et al. (2005), the set of test assets considered here are all stocks portfolios, 10 based on size sorting, 10 on Book/Market equity sorting, 10 on past-year return sorting and 5 on firm-specific R&D intensity. The R&D-sorted portfolios are less than the other sortings to keep a level of diversification inside the portfolio that is homogeneous with the others, considering the severe under-reporting of R&D expenditures which Koh and Reeb (2015) reports being 42% between 1980 and 2006.

Cash-flows growth rates of each portfolio is computed as in Bansal, R. F. Dittmar, et al. (2005). A measure h_t of capital gain is built for each stock and then summed up with those of the other stocks proportionally to the respective portfolio weight, obtaining a portfolio capital gain series $h_{p,t}$. From this series the current value of a dollar invested at the beginning of the series is computed as $V_{p,t+1} = h_{p,t+1} V_t$, where V_t is naturally initialized setting $V_{p,0} = 1$. The measure of cash-flows obtained with such strategy is then $D_{p,t+1} = y_{p,t+1} V_{p,t}$ where $y_{p,t+1}$ is the portfolio dividend-yield, obtained exploiting $R_{p,t} = h_{p,t} + y_{p,t}$. h_t is computed adjusting CRSP ex-dividend returns `RETX` for share repurchases as follows:

$$h_t = \left(\frac{P_{t+1}}{P_t} \right) \cdot \min \left[\left(\frac{n_{t+1}}{n_t} \right), 1 \right]. \tag{1.26}$$

Essentially, capital gains are less than proportional to price appreciation when there is a reduction in (equivalent) shares outstanding, which is likely related to share repurchases, a form of payout not accounted for in dividends records. Then, quarterly dividends series are obtained by simply summing monthly values up and deflating them by the implicit price deflator of nondurable and services consumption shown in Hansen et al. (2005). As the quarterly series still show strong seasonalities, quarterly values are de-seasoned by applying a 4-quarter rolling mean. The series of cash-flows growth rates are then obtained taking the first difference of the log-series of de-seasoned real quarterly dividends.

Monthly stock data is from CRSP, starting at the beginning of 1926 and stopping and the end of 2021. Yearly accounting data is from Compustat Fundamentals dataset, starting in 1950 and ending in 2021. All the monthly returns are compounded to obtain a quarterly figure and then deflated with

Table 1.5: Test asset portfolios returns and cash-flows growth: quarterly summary statistics. All series are from 1947 Q2 to 2022 Q1, a part from the R&D portfolios, which start from 1975 Q1.

Portfolio	Returns Mean	Returns SD	CF growth Mean	CF growth SD
size.01	0.06569	0.18418	0.02767	0.17561
size.02	0.03768	0.15135	0.01470	0.15258
size.03	0.03366	0.14015	0.01166	0.15642
size.04	0.03014	0.13445	0.00962	0.16473
size.05	0.02812	0.13136	0.00491	0.14426
size.06	0.02720	0.11971	0.01022	0.14199
size.07	0.02575	0.11947	0.01026	0.12372
size.08	0.02432	0.11418	0.00699	0.14256
size.09	0.02213	0.10717	0.00659	0.15094
size.10	0.01758	0.09801	0.00241	0.09691
bm.01	0.02476	0.10114	0.02050	0.28804
bm.02	0.02337	0.09135	0.01872	0.25541
bm.03	0.02499	0.08891	0.01845	0.23877
bm.04	0.02297	0.08454	0.01540	0.26606
bm.05	0.02271	0.10876	0.00682	0.14774
bm.06	0.02234	0.10538	0.00496	0.13942
bm.07	0.02118	0.10769	0.00411	0.14106
bm.08	0.02991	0.09674	0.01757	0.22745
bm.09	0.02741	0.11662	0.00933	0.20893
bm.10	0.03312	0.12231	0.01144	0.19516
mom.01	0.01498	0.21583	-0.01330	0.22345
mom.02	0.01176	0.12973	-0.00812	0.16180
mom.03	0.01468	0.11705	-0.00452	0.15323
mom.04	0.01739	0.10650	-0.00042	0.20205
mom.05	0.01824	0.09819	0.00122	0.15589
mom.06	0.01622	0.09966	0.00043	0.15794
mom.07	0.01882	0.09758	0.00126	0.16452
mom.08	0.02378	0.09693	0.00536	0.17665
mom.09	0.02605	0.10352	0.00245	0.26211
mom.10	0.03639	0.12087	-0.00819	0.29443
rd.01	0.02895	0.10367	0.00954	0.15829
rd.02	0.02464	0.08621	0.00528	0.12731
rd.03	0.02935	0.09370	0.01006	0.17154
rd.04	0.03991	0.11387	0.01552	0.16616
rd.05	0.06591	0.19221	0.03406	0.20777

the same deflator used for dividends. The construction of the portfolios closely follows Bansal, R. F. Dittmar, et al. (2005) for comparison purposes; detailed procedure descriptions follow while the main statistics of the formed portfolios' returns and cash-flow growth rates are in table 1.5.

Size-sorted portfolios All firms covered by CRSP are assigned to deciles based on their market capitalization at the end of June of each year relative to NYSE breakpoints. Weights are assigned based on the market capitalization relative to the total capitalization of the portfolio and are re-assigned at the end of every June. Both returns and cash-flows growth display a remarkable reduction for greater-size portfolios, which is in line the usual Small-minus-Big returns spread and cash-flows patterns observed in Bansal, R. F. Dittmar, et al. (2005).

B/M-sorted portfolios All firms covered by both CRSP and Compustat are assigned to deciles based on their book to market ratio and NYSE breakpoints. Portfolios are value-weighted and formed at the end of every June, where for year t the book-to-market ratio is based on book equity of fiscal year $t-1$ and market capitalization at the end of calendar year $t-1$. Both portfolio returns and cash-flows growth rates show an increasing pattern with the B/M ratio, in line with previous evidence on the value premium and Bansal, R. F. Dittmar, et al. (2005).

Momentum portfolios This set of portfolios employs stocks traded on NYSE or AMEX markets only. The assignment of a stock to a decile portfolio is determined at each end-of-quarter month t and is based the rank of the respective stock compound return from the beginning of month $t-12$ to the end of month $t-1$. These portfolios too are value-weighted. In line with previous evidence both returns and cash-flows increase with momentum, with the exception of the cash-flows growth of the most positive momentum portfolio.

R&D-sorted portfolios Firm-specific R&D intensity has been known to be associated to dispersion in excess returns since Chan et al. (2001). I specifically include these portfolios to provide further evidence that can be relevant in the study of the effects of R&D efforts aggregation. If spillover effects are stronger than fishing-out effects, then one would expect more R&D intensive firms to gain more when the whole economy invests more in R&D and the innovation LRR is higher, which leads to sensitivity heterogeneity along the R&D dimension. To enter these portfolios a stock has to be: of ordinary or common type; traded on either NYSE, AMEX, or NASDAQ; not being of a firm working in the utility or financial sectors; have at least one record of R&D expenditure. Similarly to book-market-ratio sorting, at the end of each June each firm is ranked depending on its own R&D intensity, measured by the ratio of R&D expenditure in the previous fiscal year over market capitalization at the end of the previous calendar year. Then, stocks are value weighted. The data highlights higher returns and higher cash-flows growth for higher firm-specific R&D intensity.

Table 1.6: Test assets cash-flows sensitivity to long-run risk components. From 1975 Q1 to 2022 Q1.

Portfolio	β_C	β_Z	$\beta_{\hat{r}}$	$\beta_{\tilde{r}}$
size.01	1.865	6.873	12.360	9.570
size.02	0.773	7.586	8.042	8.557
size.03	-0.072	1.919	1.013	-7.730
size.04	0.655	3.010	2.808	-3.880
size.05	1.231	2.638	2.087	7.114
size.06	1.104	0.541	0.025	0.411
size.07	1.087	3.675	2.443	-9.770
size.08	1.443	-0.924	1.416	-6.097
size.09	1.011	1.257	-3.608	-16.928
size.10	0.190	-0.688	-0.348	-7.804
bm.01	0.631	7.751	-3.374	-8.638
bm.02	0.902	6.230	-0.130	-4.021
bm.03	1.677	7.184	0.152	-2.351
bm.04	1.105	7.258	-2.307	-3.631
bm.05	0.954	2.592	2.052	-1.085
bm.06	0.144	0.354	-1.179	-4.207
bm.07	-0.166	2.900	-0.417	-12.966
bm.08	0.226	6.503	-2.226	-1.546
bm.09	0.129	0.334	0.536	-11.556
bm.10	0.437	-0.834	1.842	-0.085
mom.01	-0.770	-4.243	-0.953	2.946
mom.02	1.496	-2.634	-0.821	-8.086
mom.03	1.305	-2.806	-3.117	-0.962
mom.04	-0.418	-2.404	0.893	-8.338
mom.05	1.971	-2.639	-2.557	-6.691
mom.06	0.488	-3.783	1.353	-6.335
mom.07	-0.125	0.069	3.677	-8.790
mom.08	-0.204	-4.465	1.471	2.005
mom.09	2.283	-3.785	1.352	2.935
mom.10	1.099	-2.887	-2.676	-20.877
rd.01	0.366	-0.605	-1.988	2.864
rd.02	-1.151	3.810	-1.506	-8.603
rd.03	-1.452	2.862	-2.036	-16.581
rd.04	-0.307	7.867	-0.220	-6.877
rd.05	-0.779	4.454	7.927	11.295

4.2 Time-series sensitivities

As in Bansal, R. F. Dittmar, et al. (2005), $\theta_{p,x}$, the sensitivity of portfolio p to a risk factor – the long-run risk component in variable x, is estimated with the following regression:

$$\Delta \ln D_{p,t} = \theta_{p,x} \left(\frac{1}{L} \sum_{l=1}^{L} x_{t-l} \right) + v_{p,t}. \tag{1.27}$$

Both dependent and independent variables are demeaned before estimation. Estimating the coefficient over the rolling mean of the process x_t has the purpose of filtering persistent components of the regressor that should have a long-lasting impact on cash-flows growth. Indeed, the coefficient is asymptotically equivalent to the one estimated in the regression

$$\frac{1}{L}\sum_{l=1}^{L}\Delta \ln D_{p,t+l} = \theta_{p,x} x_t + v'_{p,t}. \tag{1.28}$$

with an inferential advantage in small samples, as illustrated by Hodrick (1992). The long-run risk components studied here are those contained in consumption growth, productivity growth and R&D intensity, i.e. $x \in \{\Delta \ln C,\ \Delta \ln Z,\ \hat{r},\ \tilde{r}\}$, where I also include $\tilde{r}$, the series based on ideas proxied by patents, for robustness. K is fixed to 12, i.e. 3 years, in the main analysis, but results are not significantly different for reasonable changes. Results over the period where all the portfolios are available, i.e. from 1975 to 2022, are shown in table 1.6.

It can be noted that sensitivities to persistent movements in consumption show a pattern for size and BM portfolios, but not quite as much for momentum and R&D portfolios. Long-run *productivity* risk component produce much starker patterns across all sortings and the long-run *innovation* risk component too. Even more interestingly, the sensitivities to R&D intensity increases with firm-specific R&D intensity, meaning that cash-flows of firms investing more in R&D grow more when the whole economy is investing relatively more too. This could support the thesis empirically studied by Jiang et al. (2016) that firms gain from higher R&D investment of peers, here on a economy-wide scale, but changes in payout policies would have to be controlled for in a more formal setting to validate such claim.

4.3 Cross-sectional risk premium

Following Fama and Macbeth (1973), risk premia are estimated with a second-step where each period the returns are regressed on a constant and the risk measure – the cash-flows sensitivities. Estimates are shown in table 1.7.

The most surprising result is that the premium associated to long-run consumption risk is far from significant. This could be related to known measurement error in consumption series,[10] as well as the predominance of other factors in pricing R&D portfolios. Indeed, in estimations over different time periods not shown here, exploiting the series from the beginning of its availability in 1947 and ignoring R&D portfolios, it becomes stronger. Results concerning other risk factors, on the other hand, strongly support the existence of a premium for long-run productivity risk, both directly and through the innovation channel, i.e. related to sensitivities of cash-flows to

[10] See, for example, Savov (2011).

Table 1.7: cross-sectional risk premia estimated following Fama and Macbeth (1973). t-statistics are HAC, computed as advised by Lazarus et al. (2018), and corrected for error-in-variable following Shanken (1992). From 1947 Q2 to 2022 Q1.

	C	Z	$\hat{r}$	$\tilde{r}$
intcpt (%)	1.920***	1.621***	1.730***	2.329***
t-stats	(3.899)	(3.225)	(3.625)	(4.450)
lambda (%)	0.015	0.196***	0.315***	0.096***
t-stats	(0.083)	(3.100)	(3.619)	(3.619)
R^2 (%)	0.01	29.12	55.71	24.93

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

R&D excess intensity. In both cases the premium is significantly different from 0 and the cross-sectional R^2 is remarkable for a single non-traded factor. This is further supported by the premium associated to sensitivity to the R&D intensity measure based on patents being significant too. These results suggest that persistent innovation originated in R&D is indeed priced, as expected by the long-run risk framework.

5 Conclusion

Persistent fluctuations in consumption are theorized to heavily impact investors welfare and how they price financial assets. These swings have also been shown to be originated in persistent swings in productivity, which has, itself, proven to be strictly related to R&D investments in the economy. This paper defines a relevant and empirically-feasible measure of R&D investment intensity and its estimates adhere to theoretical predictions. Specifically, deviations of R&D investment from an equilibrium proportion of TFP level, labelled 'long-run *innovation* risk component', prove being persistent, predict productivity growth rates and are associated to a significant risk premium in the cross section for assets whose cash-flows are more sensitive to them. This provides further support to the existence of a long-run risk component and the relevance of the long-run risk framework.

References

Ai, Hengjie et al. (2018). 'News Shocks and the Production-Based Term Structure of Equity Returns'. In: *The Review of Financial Studies* 31 (7), pp. 2423–2467.

Bansal, Ravi, Robert Dittmar, and Dana Kiku (2009). 'Cointegration and Consumption Risks in Asset Returns'. In: *Review of Financial Studies* 22.3, pp. 1343–1375.

Bansal, Ravi, Robert F. Dittmar, and Christian T. Lundblad (2005). 'Consumption, Dividends, and the Cross Section of Equity Returns'. In: *The Journal of Finance* 60 (4), pp. 1639–1672.

Bansal, Ravi, Marcelo Ochoa, and Dana Kiku (1, 2016). *Climate Change and Growth Risks*. Rochester, NY.

Bansal, Ravi and Ivan Shaliastovich (2013). 'A Long-Run Risks Explanation of Predictability Puzzles in Bond and Currency Markets'. In: *Review of Financial Studies* 26 (1), pp. 1–33.

Bansal, Ravi and Amir Yaron (2004). 'Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles'. In: *The Journal of Finance* 59 (4), pp. 1481–1509.

Beeler, Jason and John Y. Campbell (1, 2012). 'The Long-Run Risks Model and Aggregate Asset Prices: An Empirical Assessment'. In: *Critical Finance Review* 1.1, pp. 141–182.

Bloom, Nicholas et al. (2020). 'Are Ideas Getting Harder to Find?' In: *American Economic Review* 110 (4), pp. 1104–1144.

Bottazzi, Laura and Giovanni Peri (2007). 'The International Dynamics of R&D and Innovation in the Long Run and in the Short Run'. In: *The Economic Journal* 117 (518), pp. 486–511.

Brown, James R., Gustav Martinsson, and Bruce C. Petersen (2012). 'Do financing constraints matter for R&D?' In: *European Economic Review* 56.8, pp. 1512–1529.

Campbell, John Y. (1996). 'Understanding Risk and Return'. In: *Journal of Political Economy* 104 (2), pp. 298–345.

Chan, Louis K.C., Josef Lakonishok, and Theodore Sougiannis (2001). 'The stock market valuation of research and development expenditures'. In: *Journal of Finance* 56 (6), pp. 2431–2456.

Choi, In and Eiji Kurozumi (2012). 'Model selection criteria for the leads-and-lags cointegrating regression'. In: *Journal of Econometrics* 169 (2), pp. 224–238.

Colacito, Riccardo and Mariano Massimiliano Croce (2011). 'Risks for the Long Run and the Real Exchange Rate'. In: *Journal of Political Economy* 119.1, pp. 153–181.

Croce, Mariano Massimiliano (2014). 'Long-run productivity risk: A new hope for production-based asset pricing?' In: *Journal of Monetary Economics* 66, pp. 13–31.

Dew-Becker, Ian and Stefano Giglio (2016). 'Asset Pricing in the Frequency Domain: Theory and Empirics'. In: *Review of Financial Studies* 29.8, pp. 2029–2068.

Epstein, Larry G., Emmanuel Farhi, and Tomasz Strzalecki (1, 2014). 'How Much Would You Pay to Resolve Long-Run Risk?' In: *American Economic Review* 104.9, pp. 2680–2697.

Epstein, Larry G. and Stanley E. Zin (1989). 'Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework'. In: *Econometrica* 57 (4), p. 937.

Fama, Eugene F. and James D. Macbeth (1973). 'Risk, Return, and Equilibrium: Empirical Tests'. In: *Journal of Political Economy* 81 (3), pp. 607–636.

Fernald, John G. (2012). 'A Quarterly, Utilization-Adjusted Series on Total Factor Productivity'. In: *Federal Reserve Bank of San Francisco, Working Paper Series*, pp. 01–28.

Ha, Joonkyung and Peter Howitt (2007). 'Accounting for Trends in Productivity and R&D: A Schumpeterian Critique of Semi-Endogenous Growth Theory'. In: *Journal of Money, Credit and Banking* 39 (4), pp. 733–774.

Hansen, Lars Peter, John C. Heaton, and Nan Li (2005). *Intangible Risk.*

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.

Herzer, Dierk (1, 2022). 'The impact of domestic and foreign R&D on TFP in developing countries'. In: *World Development* 151, p. 105754.

Hodrick, Robert J. (1992). 'Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement'. In: *Review of Financial Studies* 5 (3), pp. 357–386.

Jiang, Yi, Yiming Qian, and Tong Yao (2016). 'R&D Spillover and Predictable Returns*'. In: *Review of Finance* 20 (5), pp. 1769–1797.

Jones, Charles I. (2005). 'Growth and Ideas'. In: *Handbook of Economic Growth.* Vol. 1, pp. 1063–1111. ISBN: 978-0-444-52043-2.

Kaltenbrunner, Georg and Lars A. Lochstoer (2010). 'Long-Run Risk through Consumption Smoothing'. In: *Review of Financial Studies* 23 (8), pp. 3190–3224.

Koh, Ping-Sheng and David M. Reeb (2015). 'Missing R&D'. In: *Journal of Accounting and Economics* 60 (1), pp. 73–94.

Kruse-Andersen, Peter K. (2023). 'Testing R&D-Based Endogenous Growth Models*'. In: *Oxford Bulletin of Economics and Statistics* n/a (n/a).

Kung, Howard and Lukas Schmid (2015). 'Innovation, Growth, and Asset Prices'. In: *The Journal of Finance* 70 (3), pp. 1001–1037.

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

Lettau, Martin and Sydney C. Ludvigson (2001). 'Consumption, Aggregate Wealth, and Expected Stock Returns'. In: *The Journal of Finance* 56 (3), pp. 815–849.

Li, Dongmei (2011). 'Financial Constraints, R&D Investment, and Stock Returns'. In: *Review of Financial Studies* 24.9, pp. 2974–3007.

Li, Wendy C.Y. and Bronwyn Hall (2016). *Depreciation of Business R&D Capital.* Cambridge, MA.

Ludvigson, Sydney C. and Serena Ng (2009). ‘Macro Factors in Bond Risk Premia’. In: *Review of Financial Studies* 22 (12), pp. 5027–5067.

Melone, Alessandro (2021). ‘Consumption Disconnect Redux’. In: *SSRN Electronic Journal*.

Nekarda, Christopher J. and Valerie A. Ramey (2020). ‘The Cyclical Behavior of the Price-Cost Markup’. In: *Journal of Money, Credit and Banking* 52 (S2), pp. 319–353.

Ortu, Fulvio, Andrea Tamoni, and Claudio Tebaldi (2013). ‘Long-Run Risk and the Persistence of Consumption Shocks’. In: *Review of Financial Studies* 26 (11), pp. 2876–2915.

Phillips, Peter C. B. and Mico Loretan (1991). ‘Estimating Long-Run Economic Equilibria’. In: *The Review of Economic Studies* 58 (3), p. 407.

Ready, Robert C (2018). ‘Oil consumption, economic growth, and oil futures: The impact of long-run oil supply uncertainty on asset prices’. In: *Journal of Monetary Economics* 94, pp. 1–26.

Reeb, David M. and Wanli Zhao (2020). ‘Patents do not measure innovation success’. In: *Critical Finance Review* 9.1, pp. 157–199.

Romer, Paul M. (1987). ‘Growth Based on Increasing Returns Due to Specialization’. In: *The American Economic Review* 77.2, pp. 56–62.

— (1990). ‘Endogenous Technological Change’. In: *Journal of Political Economy* 98.5, S71–S102.

Saikkonen, Pentti (1991). ‘Asymptotically Efficient Estimation of Cointegration Regressions’. In: *Econometric Theory* 7 (1), pp. 1–21.

Savov, Alexi (2011). ‘Asset Pricing with Garbage’. In: *The Journal of Finance* 66.1, pp. 177–201.

Schorfheide, Frank, Dongho Song, and Amir Yaron (2018). ‘Identifying Long-Run Risks: A Bayesian Mixed-Frequency Approach’. In: *Econometrica* 86.2, pp. 617–654.

Shanken, Jay (1, 1992). ‘On the Estimation of Beta-Pricing Models’. In: *Review of Financial Studies* 5.1, pp. 1–33.

Stock, James H. and Mark W Watson (1993). ‘A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems’. In: *Econometrica* 61 (4), p. 783.

A R&D-TFP cointegration

A.1 In Kung and Schmid (2015)

Using their notation, the starting conditions are:

$$Z_t = \bar{A}(e^{a_t} N_t)^{1-\alpha} \tag{1.29}$$

$$\frac{N_{t+1}}{N_t} = 1 - \phi + \chi \left(\frac{S_t}{N_t} \right)^{\eta}. \tag{1.30}$$

Then, the intangible capital growth rate is

$$
\begin{aligned}
\Delta \ln N_{t+1} &\approx \chi \left(\frac{S_t}{N_t}\right)^{\eta} - \phi && (1.31)\\
&= \chi \exp\left\{\eta \left(\ln S_t - \ln N_t\right)\right\} - \phi && (1.32)\\
&= \chi \exp\left\{\eta \left(\ln S_t - \ln N_t\right) - \eta\bar{r}\right\} e^{\eta\bar{r}} - \phi && (1.33)\\
&\approx \chi e^{\eta\bar{r}} \left\{1 + \eta \left(\ln S_t - \ln N_t\right) - \eta\bar{r}\right\} - \phi && (1.34)\\
&= \chi e^{\eta\bar{r}}(1 - \eta\bar{r}) - \phi + \chi e^{\eta\bar{r}}\eta \left(\ln S_t - \ln N_t\right) && (1.35)\\
&= a_N + b_N \left(\ln S_t - \ln N_t\right), && (1.36)
\end{aligned}
$$

and the TFP growth rate, in terms of intangible capital is[11]

$$
\frac{Z_{t+1}}{Z_t} = e^{(1-\alpha)(a_{t+1}-a_t)} \left(\frac{N_{t+1}}{N_t}\right)^{(1-\alpha)} \tag{1.37}
$$

$$
\begin{aligned}
\Delta \ln Z_{t+1} &= (1-\alpha)((\rho-1)a_t + \varepsilon_{t+1}) + (1-\alpha)\ln\left[1 - \phi + \chi\left(\frac{S_t}{N_t}\right)^{\eta}\right] && (1.38)\\
&\approx (1-\alpha)((\rho-1)a_t + \varepsilon_{t+1}) + (1-\alpha)\left[\chi\left(\frac{S_t}{N_t}\right)^{\eta} - \phi\right] && (1.39)\\
&= (1-\alpha)((\rho-1)a_t + \varepsilon_{t+1}) + (1-\alpha)\left[\chi e^{\eta(\ln S_t - \ln N_t) - \eta\bar{r}} e^{\eta\bar{r}} - \phi\right] && (1.40)\\
&\approx (1-\alpha)((\rho-1)a_t + \varepsilon_{t+1}) + (1-\alpha)\left[\chi\left(1 + \eta\left(\ln S_t - \ln N_t\right) - \eta\bar{r}\right)e^{\eta\bar{r}} - \phi\right] && (1.41)\\
&= (1-\alpha)((\rho-1)a_t + \varepsilon_{t+1}) + (1-\alpha)\left[\chi e^{\eta\bar{r}}(1-\eta\bar{r}) - \phi + \chi\eta e^{\eta\bar{r}}\left(\ln S_t - \ln N_t\right)\right]. && (1.42)
\end{aligned}
$$

Expressing this in terms of TFP level, from Equation 1.38,

$$
\begin{aligned}
\Delta \ln Z_{t+1} &= (1-\alpha)((\rho-1)a_t + \varepsilon_{t+1}) + (1-\alpha)\ln\left[1 - \phi + \chi\left(\frac{S_t}{Z_t^{\frac{1}{1-\alpha}}\bar{A}^{\frac{1}{\alpha-1}}e^{-a_t}}\right)^{\eta}\right] && (1.43)\\
&= (1-\alpha)((\rho-1)a_t + \varepsilon_{t+1}) + (1-\alpha)\ln\left[1 - \phi + \chi\left(\frac{S_t}{Z_t^{\frac{1}{1-\alpha}}}\bar{A}^{\frac{1}{1-\alpha}}e^{a_t}\right)^{\eta}\right] && (1.44)\\
&\approx (1-\alpha)((\rho-1)a_t + \varepsilon_{t+1}) + (1-\alpha)\left[\chi\left(\frac{S_t}{Z_t^{\frac{1}{1-\alpha}}}\bar{A}^{\frac{1}{1-\alpha}}e^{a_t}\right)^{\eta} - \phi\right] && (1.45)
\end{aligned}
$$

[11] In this simple formulation the presence of a deterministic trend would surely deteriorate the accuracy of the last approximation but would not necessarily invalidate it, depending on its magnitude. Anyway, as shown in the following analysis, the presence of a time trend is statistically rejected.

$$
\begin{aligned}
&= (1-\alpha)((\rho-1)a_t + \varepsilon_{t+1})+ && (1.46)\\
&\quad + (1-\alpha)\left[\chi \cdot \exp\left(\eta(\ln S_t - \frac{1}{1-\alpha}\ln Z_t + \frac{\ln \bar{A}}{1-\alpha} + a_t) - \eta\bar{r}\right) e^{\eta\bar{r}} - \phi\right]\\
&\approx (1-\alpha)((\rho-1)a_t + \varepsilon_{t+1})+ && (1.47)\\
&\quad + (1-\alpha)\left[\chi \left(1 + \eta(\ln S_t - \frac{1}{1-\alpha}\ln Z_t + \frac{\ln \bar{A}}{1-\alpha} + a_t) - \eta\bar{r}\right) e^{\eta\bar{r}} - \phi\right]\\
&= (1-\alpha)((\rho-1)a_t + \varepsilon_{t+1})+ && (1.48)\\
&\quad + (1-\alpha)\left[\chi e^{\eta\bar{r}}(1-\eta\bar{r}) - \phi + \chi e^{\eta\bar{r}}\eta\left(\ln S_t - \frac{1}{1-\alpha}\ln Z_t + \frac{\ln \bar{A}}{1-\alpha} + a_t\right)\right]\\
&= (1-\alpha)((\rho-1+\chi e^{\eta\bar{r}}\eta)a_t + \varepsilon_{t+1})+ && (1.49)\\
&\quad + (1-\alpha)\left[\chi e^{\eta\bar{r}}\left(1-\eta\bar{r} + \frac{\eta\ \ln \bar{A}}{1-\alpha}\right) - \phi + \chi e^{\eta\bar{r}}\eta\left(\ln S_t - \frac{1}{1-\alpha}\ln Z_t\right)\right]\\
&= (1-\alpha)((\rho-1+\chi e^{\eta\bar{r}}\eta)a_t + \varepsilon_{t+1})+ && (1.50)\\
&\quad + (1-\alpha)\left[\chi e^{\eta\bar{r}}\eta\left(\frac{1}{\eta} - \bar{r} + \frac{\ln \bar{A}}{1-\alpha}\right) - \phi + \chi e^{\eta\bar{r}}\eta\left(\ln S_t - \frac{1}{1-\alpha}\ln Z_t\right)\right]\\
&= a_Z + b_Z a_t + c_Z\varepsilon_{t+1} + d_Z\left(\ln S_t - \frac{1}{1-\alpha}\ln Z_t\right). && (1.51)
\end{aligned}
$$

A.2 In my model

The conditions needed for derivation of (1.8):

$$
\begin{aligned}
Z_T &\equiv e^{a_t} I_t^{\xi} && (1.52)\\
I_{t+1} &= (1-\phi)I_t + S_t^{\eta} {I_t}^{\Psi}. && (1.53)
\end{aligned}
$$

Consider the following basic manipulations,

$$
\begin{aligned}
\frac{I_{t+1}}{I_t} &= 1 - \phi + \left(\frac{S_t}{{I_t}^{\psi}}\right)^{\eta} && (1.54)\\
\Delta \ln I_{t+1} &\approx \left(\frac{S_t}{{I_t}^{\psi}}\right)^{\eta} - \phi && (1.55)\\
\ln Z_{t+1} &= a_t + \xi \ln I_t && (1.56)\\
\Delta \ln Z_{t+1} &= (\rho^a - 1)a_t + \varepsilon^a_{t+1} + \xi\Delta \ln I_{t+1} && (1.57)\\
&\approx (\rho^a - 1)a_t + \varepsilon^a_{t+1} + \xi\left[\left(\frac{S_t}{{I_t}^{\psi}}\right)^{\eta} - \phi\right] && (1.58)\\
&= (\rho^a - 1)a_t + \varepsilon^a_{t+1} + \xi\left[\exp\left\{\eta(\ln S_t - \psi \ln I_t)\right\} - \phi\right] && (1.59)\\
&\approx (\rho^a - 1)a_t + \varepsilon^a_{t+1} + \xi\left[1 + \eta(\ln S_t - \psi \ln I_t) - \phi\right] && (1.60)\\
&= (\rho^a - 1)a_t + \varepsilon^a_{t+1} + \xi\left[1 - \phi\right] + \xi\eta(\ln S_t - \psi \ln I_t). && (1.61)
\end{aligned}
$$

Then, assuming $\rho^a \approx 1$,

$$\Delta \ln Z_{t+1} = \gamma_0 + \gamma_1(\ln S_t - \psi \ln I_t) + \varepsilon^a_{t+1}. \tag{1.62}$$

B Half-lives

The half-life of the AR(1) process of interest is between 8 and 16 years,

$$\rho_Y^{N_Y} = 0.5 \quad \Rightarrow \quad \frac{\ln(0.5)}{\ln \rho_Y} = N_Y \in [8, 16]. \tag{1.63}$$

The coefficient ρ_Y such that this is true can range between

$$0.5^{1/8} = 0.9170 < \rho_Y < 0.9576 = 0.5^{1/16}. \tag{1.64}$$

Quarterly,

$$\rho_Q^{N_Q} = 0.5 \quad \Rightarrow \quad \frac{\ln(0.5)}{\ln \rho_Q} = N_Q \in [32, 64]. \tag{1.65}$$

So, the AR(1) coefficient can take values

$$0.5^{1/32} = 0.9786 < \rho_Q < 0.9892 = 0.5^{1/64}. \tag{1.66}$$

C Forecast regression bias

In case of omitted controls for a_t, the regression reads:

$$\Delta \ln Z_{t+1} = \beta_0 + \beta_{\hat{r}} \hat{r}_t + \hat{w}_{t+1} \tag{1.67}$$

Then, $\beta_{\hat{r}}$ is estimated as

$$\beta_{\hat{r}} = \frac{\mathrm{Cov}\left[\Delta \ln Z_{t+1}, \tilde{r}_t\right] - \frac{\psi}{\xi}\mathrm{Cov}\left[\Delta \ln Z_{t+1}, a_t\right]}{\mathrm{Var}\left[\tilde{r}\right] + \left(\frac{\psi}{\xi}\right)^2 \mathrm{Var}\left[a_t\right] - \frac{\psi}{\xi}\mathrm{Cov}\left[\tilde{r}_t, a_t\right]} \tag{1.68}$$

$$= \frac{\mathrm{Cov}\left[\Delta \ln Z_{t+1}, \tilde{r}_t\right]}{\mathrm{Var}\left[\tilde{r}\right]} \frac{1 - \frac{\psi}{\xi}\mathrm{Cov}\left[\Delta \ln Z_{t+1}, a_t\right] / \mathrm{Cov}\left[\Delta \ln Z_{t+1}, \tilde{r}_t\right]}{1 + \left(\frac{\psi}{\xi}\right)^2 \mathrm{Var}\left[a_t\right] / \mathrm{Var}\left[\tilde{r}\right] - \frac{\psi}{\xi}\mathrm{Cov}\left[\tilde{r}_t, a_t\right] / \mathrm{Var}\left[\tilde{r}\right]} \tag{1.69}$$

$$= \beta_{\tilde{r}} \frac{1 - \frac{\psi}{\xi}\mathrm{Cov}\left[\Delta \ln Z_{t+1}, a_t\right] / \mathrm{Cov}\left[\Delta \ln Z_{t+1}, \tilde{r}_t\right]}{1 + \left(\frac{\psi}{\xi}\right)^2 \mathrm{Var}\left[a_t\right] / \mathrm{Var}\left[\tilde{r}\right] - \frac{\psi}{\xi}\mathrm{Cov}\left[\tilde{r}_t, a_t\right] / \mathrm{Var}\left[\tilde{r}\right]}. \tag{1.70}$$

Assuming a_t is extremely persistent,[12] $\text{Cov}\left[\Delta \ln Z_{t+1}, a_t\right] \approx 0$. Further, if one assumes that the relation between $\tilde{r}$ and a can be specified as $a_t = d \cdot \tilde{r}_t + w_t$, where w_t are shocks uncorrelated to $\tilde{r}$ and d is expected from theory to be positive,

$$\beta_{\hat{r}} = \beta_{\tilde{r}} \frac{1}{1 + \left(\frac{\psi}{\xi}\right)^2 \text{Var}\left[a_t\right] / \text{Var}\left[\tilde{r}\right] - \frac{\psi}{\xi} d}. \tag{1.71}$$

So the proportional bias in $\beta_{\hat{r}}$ with respect to $\beta_{\tilde{r}}$ is

$$\frac{\beta_{\hat{r}} - \beta_{\tilde{r}}}{\beta_{\tilde{r}}} = \frac{d - \frac{\psi}{\xi}\left(\frac{\sigma_a}{\sigma_{\tilde{r}}}\right)^2}{\frac{\xi}{\psi} + \frac{\psi}{\xi}\left(\frac{\sigma_a}{\sigma_{\tilde{r}}}\right)^2 - d}, \tag{1.72}$$

which is positive only in the case

$$0 < d - \frac{\psi}{\xi}\left(\frac{\sigma_a}{\sigma_{\tilde{r}}}\right)^2 < \frac{\xi}{\psi} \tag{1.73}$$

or, considering the OLS estimator of d:

$$0 < \rho_{a,\tilde{r}} - \frac{\psi}{\xi}\left(\frac{\sigma_a}{\sigma_{\tilde{r}}}\right) < \frac{\xi}{\psi}\left(\frac{\sigma_{\tilde{r}}}{\sigma_a}\right). \tag{1.74}$$

D Additional tables and graphs

Correlations among the R&D intensity measures are in Table 1.8.

Table 1.8: correlation among specifications of the ECTs. ‘t.t.’ stands for ‘time trend’.

	$\tilde{r}(1)$	$\tilde{r}(2)$	$\hat{r}(1)$	$\hat{r}(2)$	$\hat{r}(1b)$	$\hat{r}(2b)$
$\tilde{r}(1)$	1.000	0.989	0.798	0.817	0.718	0.663
$\tilde{r}(2)$	0.989	1.000	0.765	0.822	0.719	0.675
$\hat{r}(1)$	0.798	0.765	1.000	0.960	0.763	0.723
$\hat{r}(2)$	0.817	0.822	0.960	1.000	0.784	0.761
$\hat{r}(unadj.)$	0.718	0.719	0.763	0.784	1.000	0.969
$\hat{r}(unadj. + t.t.)$	0.663	0.675	0.723	0.761	0.969	1.000

[12] In case it is not, bias in the forecasting regression coefficient would be more easily positive, but concerns for pricing implications about using $\hat{r}$ instead of $\tilde{r}$ would alleviate significantly, since the source of persistency of $\hat{r}$ would be more likely $\tilde{r}$ then a_t.

2

Does CAPM overestimate more the risk or its price?

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 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 2.1 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

[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.

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

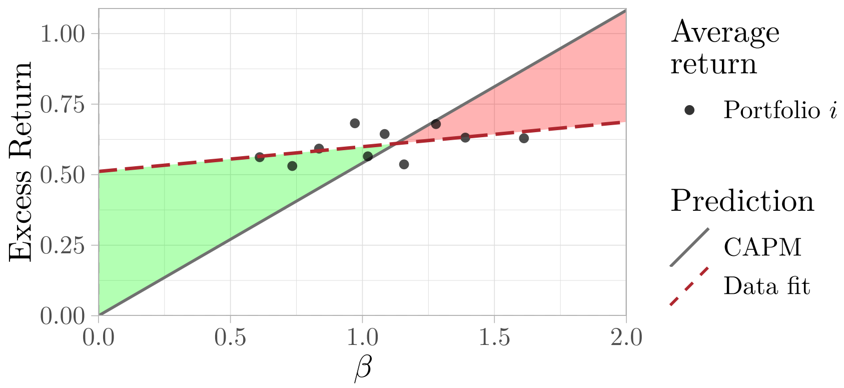

Figure 2.1: 10 US stocks portfolios. Monthly data, 1963-2019, source: K. French data library.

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) 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

[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.

mispricing, the so called 'Betting Against Beta' (BAB) factor, earning a positive return, not explained by other risk factors.[5]

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

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

[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).

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 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

[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.

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 section A 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 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

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

[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.

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 section A, where a more agnostic one is explored); in 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

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

in the amount X_t^s for $s \in \{1, \ldots, S\}$, with budget constraint

$$W_t = X_t^0 + \sum_{s=1}^{S} X_t^s P_t^s. \tag{2.1}$$

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:

$$W_{t+1} = X_t^0 R_t^f + \sum_{s=1}^{S} X_t^s (P_{t+1}^s + D_{t+1}^s). \tag{2.2}$$

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

$$\max_{\{X_t^s\}_0^S} \mathbb{E}_t \left[u\left(W_{t+1}\right)\right]. \tag{2.3}$$

Plugging constraints (2.1) and (2.2) in the objective function (2.3), the problem solved by the agent can be expressed as

$$\max_{\{X_t^s\}_1^S} \mathbb{E}_t \left[u \left(W_t R_t^f + \sum_{s=1}^{S} X_t^s (P_{t+1}^s + D_{t+1}^s - P_t^s R_t^f) \right) \right]. \tag{2.4}$$

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:

$$W_t \geq \left(\sum_{s=1}^{S} X_t^s P_t^s \right) \cdot c_t, \tag{2.5}$$

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

$$\mathbb{E}_t \left[u'(W_{t+1})(R_{t+1}^s - R_t^f) \right] - \psi_t = 0 \tag{2.6}$$

where $R_{t+1}^s = \dfrac{P_{t+1}^s + D_{t+1}^s}{P_t^s}$ is the gross return on asset s and ψ_t is the Lagrange multiplier of (2.5), 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 (2.6) it follows that

$$\mathbb{E}_t\left[r^s_{t+1}\right] = \tilde{\psi}_t - \mathrm{Cov}_t\left[\tilde{u}'_{t+1}, r^s_{t+1}\right], \tag{2.7}$$

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. (2.7) holds similarly for the market, meaning that

$$\mathbb{E}_t\left[r^M_{t+1}\right] = \tilde{\psi}_t - \mathrm{Cov}_t\left[\tilde{u}'_{t+1}, r^M_{t+1}\right]. \tag{2.8}$$

Combining these last two equation one obtains

$$\mathbb{E}_t\left[r^s_{t+1}\right] = \tilde{\psi}_t\left(1 - \frac{\mathrm{Cov}_t\left[\tilde{u}'_{t+1}, r^s_{t+1}\right]}{\mathrm{Cov}_t\left[\tilde{u}'_{t+1}, r^M_{t+1}\right]}\right) + \frac{\mathrm{Cov}_t\left[\tilde{u}'_{t+1}, r^s_{t+1}\right]}{\mathrm{Cov}_t\left[\tilde{u}'_{t+1}, r^M_{t+1}\right]} \cdot \mathbb{E}_t\left[r^M_{t+1}\right]. \tag{2.9}$$

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, (2.9) can be more simply stated as

$$\mathbb{E}_t\left[r^s_{t+1}\right] = \tilde{\psi}_t\left(1 - \gamma^s_t\right) + \gamma^s_t \cdot \mathbb{E}_t\left[r^M_{t+1}\right]. \tag{2.10}$$

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

$${\mathbb{E}_t}^{\mathrm{CAPM}}[r^s_{t+1}] = \beta^s_t \cdot \mathbb{E}_t\left[r^M_{t+1}\right] \qquad \beta^s_t = \frac{\mathrm{Cov}_t\left[R^M_{t+1}, R^s_{t+1}\right]}{\mathrm{Var}_t\left[R^M_{t+1}\right]}, \tag{2.11}$$

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

$$\alpha^s_t = \underbrace{\tilde{\psi}_t(1-\gamma^s_t)}_{\text{Liquidity deviation}} + \underbrace{(\gamma^s_t - \beta^s_t)\mathbb{E}_t\left[r^M_{t+1}\right]}_{\text{Omitted risk deviation}} \tag{2.12}$$

$$= \tilde{\psi}_t\left[(1-\beta^s_t) - (\gamma^s_t - \beta^s_t)\right] + (\gamma^s_t - \beta^s_t)\mathbb{E}_t\left[r^M_{t+1}\right]. \tag{2.13}$$

(2.12) 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. (2.13) 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:

$$\begin{aligned}
\mathbb{E}_t\left[r_{t+1}^{BAB}\right] &= \mathbb{E}_t\left[r_{t+1}^{L}\right] - \mathbb{E}_t\left[r_{t+1}^{H}\right] - (\beta_t^L - \beta_t^H)\mathbb{E}_t\left[r_{t+1}^{M}\right] && (2.14)\\
&= \tilde{\psi}_t\left(\gamma_t^H - \gamma_t^L\right) + \left[(\beta_t^H - \beta_t^L) - (\gamma_t^H - \gamma_t^L)\right]\mathbb{E}_t\left[r_{t+1}^{M}\right]. && (2.15)
\end{aligned}$$

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 vice-versa. It should also be noted that adding coskewness, for example, one expects to lower the

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

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):

$$\begin{aligned} \mathbb{E}_t\left[r_{t+1}^{BAG}\right] &= \mathbb{E}_t\left[r_{t+1}^{LG}\right] - \mathbb{E}_t\left[r_{t+1}^{HG}\right] - (\gamma^{LG} - \gamma^{HG})\mathbb{E}_t\left[r_{t+1}^{M}\right] && (2.16)\\ &= \tilde{\psi}_t\left(\gamma_t^{HG} - \gamma_t^{LG}\right). && (2.17) \end{aligned}$$

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),

$$r_t^{BAG} \approx \mathbb{E}_{t-1}\left[r_t^{BAG}\right] + \sum_{j=0}^{\infty} \rho^j \cdot (\mathbb{E}_t - \mathbb{E}_{t-1})[\Delta \ln D_{t+j}^{BAG}] - \sum_{j=1}^{\infty} \rho^j \cdot (\mathbb{E}_t - \mathbb{E}_{t-1})[r_{t+j}^{BAG}]. \tag{2.18}$$

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

$$\frac{\partial\left(r_t^{BAG} - \mathbb{E}_{t-1}\left[r_t^{BAG}\right]\right)}{\partial\left(\tilde{\psi}_t - \mathbb{E}_{t-1}[\tilde{\psi}_t]\right)} = -\rho\left(\gamma_t^{HG} - \gamma_t^{LG}\right). \tag{2.19}$$

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 Empirical implementation

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 $\boldsymbol{\beta}_{u\mathbf{r}}$,

$$\tilde{u}'^{*}_{t+1} - 1 = (\mathbf{r}_{t+1} - \mathbb{E}\left[\mathbf{r}_{t+1}\right])^{\top} \boldsymbol{\beta}_{u\mathbf{r}}, \tag{2.20}$$

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 (2.6), scaled by $\mathbb{E}\left[u'_{t+1}\right]$, i.e.

$$\mathbb{E}\left[\tilde{u}'^{*}_{t+1}\mathbf{r}_{t+1}\right] = \mathbb{E}\left[\tilde{\psi}_t\right] \mathbf{1}_N. \tag{2.21}$$

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

$$\mathbb{E}\left[(\tilde{u}'^{*}_{t+1} - 1)(\mathbf{r}_{t+1} - \mathbb{E}\left[\mathbf{r}_{t+1}\right])\right] = \mathbb{E}\left[\tilde{\psi}_t\right] \mathbf{1}_N - \mathbb{E}\left[\mathbf{r}_{t+1}\right], \tag{2.22}$$

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

$$\boldsymbol{\beta}_{u\mathbf{r}} = \Sigma_{\mathbf{rr}}^{-1}(\mathbb{E}\left[\psi_t\right] \mathbf{1}_N - \mathbb{E}\left[\mathbf{r}_{t+1}\right]), \tag{2.23}$$

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

$$u'^{*}_{t+1} - 1 = (\mathbf{r}_{t+1} - \mathbb{E}\left[\mathbf{r}_{t+1}\right])^{\top} \Sigma_{\mathbf{rr}}^{-1}(\mathbb{E}\left[\psi_t\right] \mathbf{1}_N - \mathbb{E}\left[\mathbf{r}_{t+1}\right]). \tag{2.24}$$

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

[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).

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

$$\mathbf{r}_{t+1} - \mathbb{E}\left[\mathbf{r}_{t+1}\right] = \mathbf{e}_{t+1} \tag{2.25}$$

$$= \mathrm{B} \cdot \mathbf{f}_{t+1} + \boldsymbol{\epsilon}_{t+1}, \tag{2.26}$$

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

$$\tilde{u}_{t+1}^{\prime *} = 1 + (\mathbf{f}_{t+1}^{\top}\mathrm{B}^{\top} + \boldsymbol{\epsilon}_{t+1}^{\top})\Sigma_{\mathbf{rr}}^{-1}(\mathbb{E}\left[\psi_t\right]\mathbf{1}_N - \mathbb{E}\left[\mathbf{r}_{t+1}\right]), \tag{2.27}$$

or, more succinctly,

$$\tilde{u}_{t+1}^{\prime *} = 1 - \mathbf{f}_{t+1}^{\top}\boldsymbol{\beta}_{u\mathbf{f}} - \boldsymbol{\epsilon}_{t+1}^{\top}\boldsymbol{\beta}_{u\boldsymbol{\epsilon}}, \tag{2.28}$$

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 (2.7) becomes

$$\mathbb{E}_t\left[r_{t+1}^s\right] = \tilde{\psi}_t + \boldsymbol{\beta}_{u\mathbf{f}}^{\top} \cdot \mathrm{Cov}_t\left[\mathbf{f}_{t+1},\ r_{t+1}^s\right] + \boldsymbol{\beta}_{u\boldsymbol{\epsilon}}^{\top} \cdot \mathrm{Cov}_t\left[\boldsymbol{\epsilon}_{t+1},\ r_{t+1}^s\right]. \tag{2.29}$$

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:

$$\mathbb{E}\left[r_{t+1}^s\right] = \mathbb{E}\left[\tilde{\psi}_t\right] + \boldsymbol{\beta}_{u\mathbf{F}}^{\top} \cdot \mathrm{Cov}\left[\mathbf{f}_{t+1}, r_{t+1}^s\right]. \tag{2.30}$$

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 (2.30):

$$\mathbb{E}\left[r^s_t\right] - a - \mathbb{E}\left[\mathbf{b}^\top \hat{\mathbf{f}}_t \cdot \left(r^s_t - \mathbb{E}\left[r^s_t\right]\right)\right] = 0. \tag{2.31}$$

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

$$\gamma^s = \frac{\mathbf{b}^\top \mathrm{Cov}\left[\hat{\mathbf{f}}_t, r^s_t\right]}{\mathbf{b}^\top \mathrm{Cov}\left[\hat{\mathbf{f}}_t, r^M_t\right]}, \tag{2.32}$$

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

[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 2.1: 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)

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 2.6, 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 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 2.1. 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 2.5, while the statistics for the middle third portfolio and the 10-split portfolios over the same time-frame are in Table 2.7, both showing a similar pattern.

4.3 Estimation results

PCs factors

Figure 2.2 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 (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.

Pricing the cross-section

Table 2.2 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

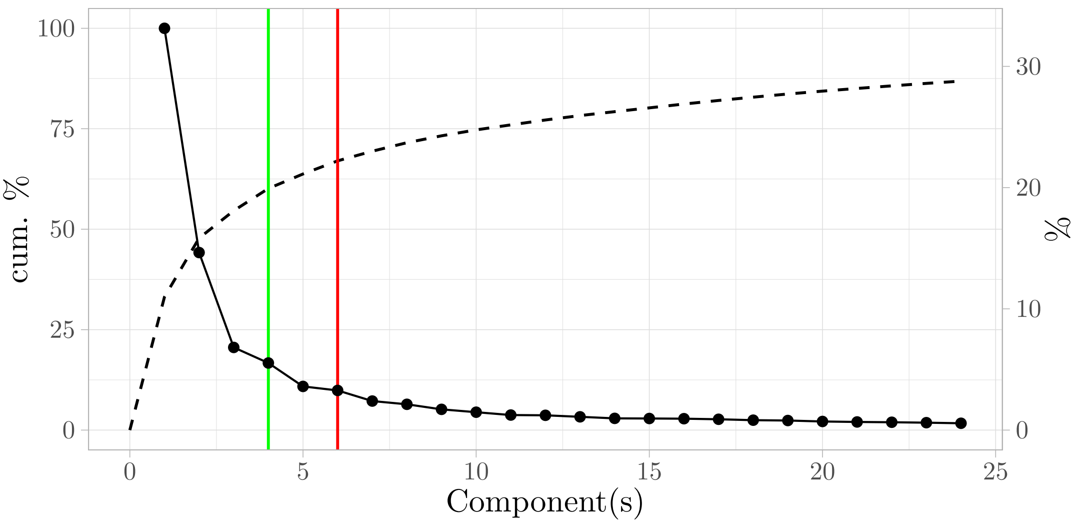

Figure 2.2: Scree plot of principal component analysis. On the horizontal axis the number of component considered, on the vertical axis the share of variance explained by the relative component (solid line) and the cumulative share of variance explained (dotted line). The green vertical line marks the 4^{th} component while the red one the 6^{th}.

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 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

[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.

Table 2.2: 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)

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

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.

Revisiting the low-risk anomaly

Estimates of **b** imply estimates of γs too, which are shown in Table 2.3. The first pattern to emerge is that, when more risks are considered, the spread in the synthetic risk measure 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 (2.15) 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

[17] See Table 2.6 in the appendix.

Table 2.3: 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)

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

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 (2.16) 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 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

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

Table 2.4: 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)

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

significant. Any BAB portfolio is characterized by market-risk neutrality, so a significant γ, interpreted through (2.12), 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 2.4 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

$$\Delta = \frac{\text{Liq. component} - \text{Risk component}}{\text{Tot. prediction}} = \frac{\mathbb{E}\left[\psi\right](1-\gamma^{BAB}) - \gamma^{BAB} \cdot \mathbb{E}\left[r^M\right]}{\mathbb{E}\left[\psi\right](1-\gamma^{BAB}) + \gamma^{BAB} \cdot \mathbb{E}\left[r^M\right]}. \tag{2.33}$$

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

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

generated in the zero-beta asset by the financial frictions. This relation can be further tested with the formation of a Betting-Against-Gamma portfolio, which would also be 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.

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 reexamination 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.

A A non-parametric approach

A.1 Unconditional estimation

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

$$\mathbb{E}\left[m_{t+1} r^s_{t+1} | \mathbf{z}_t\right] = \psi_t \tag{2.34}$$

$$\mathbb{E}\left[m_{t+1} r^s_{t+1}\right] = \mathbb{E}\left[\psi_t\right] \tag{2.35}$$

$$\tag{2.36}$$

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

$$R'\mathbf{m}/T = \mathbf{1}_S \; (\mathbf{1}'_T \, \boldsymbol{\psi})/T \tag{2.37}$$

$$R'\mathbf{m} = \mathbf{1}_S \; (\mathbf{1}'_T \, \boldsymbol{\psi}) \tag{2.38}$$

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

$$RR'\mathbf{m} = R \; \mathbf{1}_S \; (\mathbf{1}'_T \, \boldsymbol{\psi}) \tag{2.39}$$

$$\mathbf{m} = (RR')^{-1} R \; \mathbf{1}_S \; (\mathbf{1}'_T \, \boldsymbol{\psi}), \tag{2.40}$$

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. $\mathbb{E}\left[\psi_t\right]$.

A.2 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 (2.34) by the value z^j, which is contained in the time t information set, to get

$$\mathbb{E}_t\left[(m_{t+1}r^i_{t+1})z^j_t\right] = \psi_t \cdot z^j_t \qquad \forall j \in \{1, \dots, K\}\ , \quad \forall i \in \{1, \dots, N\}. \tag{2.41}$$

Then, assuming stationary variables, the condition

$$\mathbb{E}\left[(m_{t+1}r^i_{t+1})z^j_t\right] = \mathbb{E}\left[\psi_t \cdot z^j_t\right] \qquad \forall j \in \{1, \dots, K\}\ , \quad \forall i \in \{1, \dots, N\} \tag{2.42}$$

is also true. I label $r^i_{t+1}z^j_t$ as d^{ij}_{t+1}, so (2.42) is

$$\mathbb{E}\left[m_{t+1}d^{ij}_{t+1}\right] = \mathbb{E}\left[\psi_t \cdot z^j_t\right] \qquad \forall j \in \{1, \dots, K\}\ , \quad \forall i \in \{1, \dots, N\}. \tag{2.43}$$

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

$$\underbrace{D^\top}_{(NK\times T)} \underbrace{\mathbf{m}}_{(T\times 1)} /T = \underbrace{(Z \otimes \mathbf{1}^\top_N)^\top}_{(NK\times T)} \underbrace{\boldsymbol{\psi}}_{(T\times 1)} /T, \tag{2.44}$$

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

$$DD^\top\mathbf{m} = D(Z \otimes \mathbf{1}^\top_N)^\top\boldsymbol{\psi} \tag{2.45}$$

$$\mathbf{m} = (DD^\top)^{-1}D(Z \otimes \mathbf{1}^\top_N)^\top\boldsymbol{\psi}. \tag{2.46}$$

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

$$\mathbb{E}_t\left[m_{t+1}\right] = 1; \tag{2.47}$$

this implies

$$\mathbb{E}_t\left[m_{t+1}z^j_t\right] = z^j_t \qquad \forall j \in \{1, \dots, K\} \tag{2.48}$$

and

$$\mathbb{E}\left[m_{t+1}z^j_t\right] = \mathbb{E}\left[z^j_t\right] \qquad \forall j \in \{1, \dots, K\}\ . \tag{2.49}$$

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

$$\underbrace{Z^\top}_{(K\times T)} \underbrace{\mathbf{m}}_{(T\times 1)} /T = \underbrace{Z^\top}_{(K\times T)} \underbrace{\mathbf{1}_T}_{(T\times 1)} /T. \tag{2.50}$$

Finally, plugging (2.46), an estimate of the ψ_t time-series is the solution to the system

$$Z^\top (DD^\top)^{-1} D^\top (Z \otimes \mathbf{1}_N^\top)^\top \boldsymbol{\psi} = Z^\top \mathbf{1}_T, \tag{2.51}$$

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

B Additional tables and figures

Table 2.5: 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 2.6: 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 2.7: 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)

3

Are you betting against sustainability?

1 Introduction

To manage climate change , the economy needs to transition into a more sustainable one. This requires a great mobilization of capital, which has in fact started flowing towards investments where Environmental, Social and Governance (ESG) factors are considered in the allocation process. Leaving aside issues concerning other literatures, even critical ones such as what exactly is 'sustainable', the key economic concern about this phenomenon is understanding the *value* of economic activities' 'sustainability'. The financial literature approaches this by studying the ties between this characteristic and the discount rates that are applied to assess projects' value, which directly reflects how investors' welfare is impacted by sustainability. This paper provides further support to the view that investors do value firms' sustainability and highlights how the pricing of sustainability is not independent from assets' riskiness, a phenomenon that can affect the measurement of the sustainability 'premium' and more dynamic considerations, which are relevant towards the hedging of climate concerns changes for example.

More specifically, assets' sustainability can affect investors welfare (1) directly, by providing non-pecuniary benefits, as in Pástor et al. (2021); (2) by affecting consumption with inherently different cash flows from the less sustainable counterparts (e.g. being more profitable or hedging better climate and regulations shocks), such as in Yang (2022); or (3) via both, as in Pedersen et al. (2021). While the second channel can have very rich mechanisms and implications, the first one distinctly predicts more sustainable project to have lower expected return than that of an equivalent non-sustainable counterpart. This, in a way, has been challenged by the empirical observation that in recent years more environmentally-sustainable ('green') assets had higher returns, not lower. However, as clearly shown by Pástor et al. (2022), this is likely due to sustained increases in environmental concerns that intensified the demand for greener assets. This, in turn, while lowering expected returns of greener assets, mechanically lead to greater contemporaneous returns. I build on that, showing that increases in the sustainability premium do not impact returns depending on the assets' sustainability measure only, but on their riskiness too. In more practical terms, in a CAPM world, two assets could be identically sustainable, but, if they had different betas, they would react differently to increases in concerns. The reason is that the spread in returns due to sustainability is not only related to sustainability in the first place: it depends on the difference, in the CAPM example, between beta and the sustainability of the asset relatively to the market. This happens because one can always reach the same level of risk of an asset by levering the market, obtaining, per dollar spent, the market ESG score. It follows that only assets with ESG scores greater than the ESG provided by the equivalently-levered market should be appreciated for the sustainability contribution to the portfolio.

First, I show this mechanism with a simple model, similar to Pástor et al. (2021) in aim and implications, but more flexible in accommodating multiple risk factors and in allowing more interpretations to the sustainability premium's origin. Specifically, a standard investor who maximises expected utility is faced with an inequality constraint on the average ESG score of its portfolio, which, when binding, places a 'sustainability multiplier' in the optimality conditions very similar to that of an agent with linear preferences for sustainability. While the distinction between linear preferences and a binding constraint appears of second order of relevance, I favour this formulation because once sustainability is modelled as affecting welfare, it is not obvious at all why it should do so in a linear way, while a constraint can well be interpreted as a proper requirement set by households to intermediaries, or even as a physical requirement to achieve aggregate environmental targets, e.g. emit less than a certain amount of CO2 to avoid catastrophes. The model boils down to a pricing equation that, with standard CAPM assumptions on either returns distribution or wealth utility functional form, closely adheres to the results in Pástor et al. (2021). Nonetheless, following Franceschini (2023), I also show a way to extend the analysis to a more realistic multi-factor setting, while keeping track of all the risks with a synthetic measure. This theoretical formulation enables empirical analysis that are potentially able to distinguish between the static ESG-'preferences' premium and the premia associated with sustainability-risk motives, via wealth dynamics or sustainability dynamics directly – as displayed in the climate extension of the model of Pástor et al. (2021).

The main result is that, ceteris paribus, more sustainable firms are expected to yield lower returns, which is a known result. However, as the premium for sustainability is proportional to the difference between risk and the asset sustainability relative to the market, determining how assets react to sustainability concern shocks is not a straightforward. Indeed, apart from the reaction being greater the more relatively sustainable the asset is, more interestingly, the reaction sign depends on the riskiness too. Specifically, to be positively covarying with concerns overall, the relative sustainability has to be greater than the risk (again, akin to the beta in the CAPM formulation). This is an important fact to establish because it affects the hedging abilities of most standard 'sustainability' portfolios, such as a market-neutral portfolio that is long on more sustainable assets and short on the least ones. In facts, this portfolio earns the sustainability-spread proportionally to the difference in sustainability of the two components forming it *minus* the difference in the risk levels of the two. Thus, for certain cross-sectional risk distributions, the spread could perfectly be negative while this portfolio's average return is consistently positive. It follows that contemporaneous returns of this sustainability portfolio in reaction to changes in the sustainability spread can make it an effective or a terrible hedge. Also, while this consideration might not have had an impact until now, it might in the future. For example, were the greener assets to be given better funding to invest in innovation, they would become riskier in the cross-section and this would harm their ability of hedging later shocks to environmental concerns.[1]

This theoretically holds for any characteristic that is associated with a constraint on the portfolio weighted value. Empirically, I test whether Refinitiv ESG data is relevant in this sense in US data and how the risk profile of a high-minus-low portfolio impacts its expected returns, which in turn determine its hedging abilities. Proxying the marginal utilities with the 3 factors from Fama and French (1993) or the market and two factors extracted from the test assets returns, the sustainability-related spread is significantly negative. However, it stops being significant in other specifications – when the market is the only risk factor or when many factors extracted from test assets returns are included, none of which result in a worse pricing performance, undermining the view that this particular score is associated to a binding constraint in the market. Despite the relevance of the exercise, at this stage sustainability-related risk factors are not explicitly included among the risk factors, although they are

[1] From an aggregate perspective, green innovation mitigates environmental concerns. So, green firms increasing innovation efforts would work as an hedge to environmental concern, but once the innovation levels are set, and with them the new risk levels, their ability of hedge later shocks to environment concerns would likely worsen. Notice that whether mitigating environmental issues is pro- or counter-cyclical is not obvious, as highlighted by Giglio et al. (2021), so whether greener firms' returns show a positive or negative sensitivity to environment concerns is not so clear either. Consequently, the effect of more aggregate green innovation is not necessarily counter-acting the argument made in the example.

possibly captured by the factors agnostically obtained from test assets returns. Further, the risk of the high-minus-low portfolio is not significantly different from 0 in all specifications beyond the market-only one, meaning that risk is unlikely to have any impact on how the sustainability is reflected in univariate portfolios. Spread, anyway, that is unlikely to be there in the first place, which is why these implications will have to be tested on more sustainability dimensions,[2] such as CO2 emissions, and possibly other countries too. This, for some sustainability measures, will also allow to include sustainable bonds, which reduces the number of relevant parameters to be estimated. Lastly, were these hypothesis to find any grip, a further test could rely on a risk- and sustainability-neutral portfolio, whose contemporaneous returns covariance with concerns is directly indicative of risk impact on the sustainability spread.

This study is clearly related to the recent theoretical literature on ESG investment, most importantly with Pástor et al. (2021) and Pedersen et al. (2021), to which this paper adds the study of the peculiar role that risk, in its generality, has in sustainability premium dynamics. This model resembles the model of Pedersen et al. (2021) if it was only populated by investors who have average ESG score in the utility function. They end up characterizing the security market line in terms of Sharpe Ratio and reach the similar conclusions that returns have alphas relative to the CAPM that depend on individual ESG scores relative to the market's score. As they consider a predictive power of ESG scores on profits, however, a higher relative ESG does not guarantee a lower expected return in their model. Just as with Pástor et al. (2021), this paper departs by considering risk beyond mean-variance optimization set-ups and displays in greater detail its impact on predictions that seem to be independent, such as returns on market-neutral sustainable portfolios.

Further, this paper is related to the recent empirical green finance literature, which has been mainly focused on isolating return dynamics due to concern increases and the 'static' sustainability spread, such as Pástor et al. (2022), Hsu et al. (2020) and Ardia et al. (2022). This paper shows another reason why greener assets have not displayed lower returns besides sustained increases in environmental concerns, i.e. baseline assets' risks. Second, the framework used here allows to include environmental sensitivity as a risk and potentially disentangle the risk-led sustainability premium from the 'static preferences' spread. At the current stage, this is not formally studied in this paper, but it will be, by explicitly adding a sustainability-relevant state variable among the risk factors.[3] The empirical application also makes this paper take

[2]Further motivated by the findings of Berg et al. (2022).

[3]Several indexes that identify aggregate environmental concerns have been proposed, such as Ardia et al. (2022) and Noailly et al. (2021). However, variables that are univocally related with the other pillars of ESG scores are not obvious at all.

part to the literatures employing General Method of Moments (GMM) to estimate marginal utility loadings, and to those modelling marginal utility in an agnostic way that do not employ pre-determined factors.

In section 2 I outline the theoretical set-up and show the impact of risk on sustainability-related pricing implications, in the simple framework of CAPM; in section 3, I consider the existence of a risk-free asset that also provides a sustainability score and the existence of multiple risks priced in the market; in section 4, I explore the main implications on US data employing the Refinitiv ESG scores; in section 5 I conclude.

2 Basic theoretical set-up

Consider a standard two-period economy where there are $I+1$ assets indexed by i, with $i=0$ being a risk-less asset. Then, assume a single agent living in this economy, who is born at time t and simply chooses assets holdings $\{x_t^i\}_{i=1}^I$ to maximize utility from second-period consumption, which corresponds to the entirety of the accumulated wealth W_{t+1}, i.e.

$$\begin{aligned} &\max_{\{x_t^i\}_{i=0}^I} \mathbb{E}_t\left[u(W_{t+1})\right] \\ \text{s.t.} \quad & W_t = x_t^0 + \sum_{i=1}^I x_t^i P_t^i \\ & W_{t+1} = x_t^0 R_t^f + \sum_{i=1}^I x_t^i (P_{t+1}^i + D_{t+1}^i), \end{aligned} \tag{3.1}$$

where P_t^i and D_{t+1}^i are the price and dividend of asset i at time t, respectively, and R_t^f is the gross risk-free return. Growing concerns about sustainability of investments might enforce an exogenous level of greenness of the portfolio held by this agent, either imposed by a second agent in the form of a ruling government or a household whose savings are managed by the agent just described.

A simple way to capture this phenomenon is to require average sustainability of portfolio holdings to be greater than a certain threshold q_t,[4] which can be expressed as the inequality

$$\sum_{i=1}^I (x_t^i P_t^i) \cdot \mathrm{ESG}_t^i > q_t \cdot W_t. \tag{3.2}$$

This is close in spirit to Frazzini and Pedersen (2014), who consider leverage

[4] A key assumption here is that the risk-free asset does not contribute to the greenness of a portfolio, which might can be counterfactual. An extension which allows for this is immediate and will follow in the paper. This also implies more testable conditions.

constraints instead of sustainability requirements. Exploiting $W_{t+1} = W_t R_t^f + \sum_{i=1}^{I} x_t^i (P_{t+1}^i + D_{t+1}^i - P_t^i R_t^f)$, the Lagrangian related to this problem can be expressed as

$$\mathcal{L}_t = \mathbb{E}_t \left[u \left(W_t R_t^f + \sum_{i=1}^{I} x_t^i (P_{t+1}^i + D_{t+1}^i - P_t^i R_t^f) \right) \right] + \lambda_t \left\{ \sum_{i=1}^{I} (x_t^i P_t^i) \cdot \mathrm{ESG}_t^i - q_t \cdot W_t \right\} \tag{3.3}$$

The first-order condition for the optimality of this problem solution when the constraint binds, for each i, is

$$\mathbb{E}_t \left[u'(W_{t+1})(R_{t+1}^i - R_t^f) \right] + \lambda_t \mathrm{ESG}_t^i = 0 \tag{3.4}$$

or

$$\mathbb{E}_t \left[r_{t+1}^i \right] = \frac{-\lambda_t}{\mathbb{E}_t \left[u'(W_{t+1}) \right]} \cdot \mathrm{ESG}_t^i - \mathrm{Cov}_t \left[\frac{u'(W_{t+1})}{\mathbb{E}_t \left[u'(W_{t+1}) \right]}, r_{t+1}^i \right], \tag{3.5}$$

where r_{t+1}^i is the excess return $R_{t+1}^i - R_t^f$. Then, the market, defined by portfolio weights ω_t^i that ensure $\sum_{i=1}^{I} \omega_t^i = 1$, has expected excess return

$$\mathbb{E}_t \left[r_{t+1}^M \right] = \frac{-\lambda_t}{\mathbb{E}_t \left[u'(W_{t+1}) \right]} \mathrm{ESG}_t^M - \mathrm{Cov}_t \left[\frac{u'(W_{t+1})}{\mathbb{E}_t \left[u'(W_{t+1}) \right]}, r_{t+1}^M \right], \tag{3.6}$$

where ESG^M is the ESG score per dollar of market holding. The risk-free, not impacting the sustainability constraint is not determined differently from more standard frameworks.

Taking the approximation $u'(W_{t+1}) \approx u'(W_t) R_t^f + u''(W_t)(R_{t+1}^M - R_t^f)$, $\frac{u'(W_{t+1})}{\mathbb{E}_t[u'(W_{t+1})]}$ can be expressed as $a_t - b_t \cdot r_{t+1}^M$, and from (3.6) the SDF loading b_t can be derived as

$$b_t = \frac{\mathbb{E}_t \left[r_{t+1}^M \right] + \frac{\lambda_t}{\mathbb{E}_t[u'(W_{t+1})]} \mathrm{ESG}_t^M}{\mathrm{Var}_t \left[r_{t+1}^M \right]}. \tag{3.7}$$

Then, any asset abides by the following relation

$$\mathbb{E}_t \left[r_{t+1}^i \right] = \frac{-\lambda_t}{\mathbb{E}_t \left[u'(W_{t+1}) \right]} \cdot \mathrm{ESG}_t^i + \left(\mathbb{E}_t \left[r_{t+1}^M \right] + \frac{\lambda_t}{\mathbb{E}_t \left[u'(W_{t+1}) \right]} \mathrm{ESG}_t^M \right) \frac{\mathrm{Cov}_t \left[r_{t+1}^M, r_{t+1}^i \right]}{\mathrm{Var}_t \left[r_{t+1}^M \right]}, \tag{3.8}$$

or, more simply,

$$\mathbb{E}_t \left[r_{t+1}^i \right] = \tilde{\lambda}_t \left(\frac{\mathrm{ESG}_t^i}{\mathrm{ESG}_t^M} - \beta_t^i \right) + \beta_t^i \cdot \mathbb{E}_t \left[r_{t+1}^M \right]. \tag{3.9}$$

$\tilde{\lambda}_t = \frac{-\lambda_t \cdot \mathrm{ESG}_t^M}{\mathbb{E}_t[u'(W_{t+1})]}$ is a negative constant whose magnitude increases the more constrained the agent is and the higher market ESG score is. The key implication is that an expected excess return is proportional to the respective CAPM beta, as in the standard CAPM, and the difference between the 'ESG-intensity' $\mathrm{ESG}_t^i/\mathrm{ESG}_t^M$ and the beta. This comes from the fact that riskier assets, having greater discounts, provide less ESG-weighted capital per unit of cash-flows claimed by holding that asset. For example, if two assets have identical expected cash-flows and ESG ratings but different betas, the cash-flows of one with the lower beta will be discounted less and have a higher price, thus helping alleviating the constraint more than the other one.

2.1 Responsible portfolios

Let us consider a portfolio that tries to isolate the spread related to sustainability without having any exposure to market risk: arguably, the most intuitive way would be to hold a portfolio S of more sustainable stocks, financing this by shorting a portfolio NS of less sustainable ones and hedge the resulting sensitivity to the market by holding it in an opposite proportion. The expected excess return of this 'sustainability' portfolio, r^{SUS}, would be

$$\mathbb{E}\left[r^{SUS}\right] = \mathbb{E}_t\left[r^S\right] - \mathbb{E}_t\left[r^{NS}\right] - (\beta^S - \beta^{NS})\mathbb{E}_t\left[r^M\right] \tag{3.10}$$

$$= \tilde{\lambda}\left[\left(\frac{\mathrm{ESG}_t^S}{\mathrm{ESG}_t^M} - \frac{\mathrm{ESG}_t^{NS}}{\mathrm{ESG}_t^M}\right) - (\beta^S - \beta^{NS})\right]. \tag{3.11}$$

This portfolio effectively has no exposure to the market, the only source of risk here, but its return will reflect the sustainability spread $\tilde{\lambda}$ only if the risk is distributed evenly across assets with different sustainability. This means that testing the existence of any premium associated to characteristics that can be linked to constraints as those defined in this model on a generic ESG score, cannot rely on plain univariate-sort portfolios, even with controls for risk exposure.[5] If the beta of more sustainable stocks is sufficiently higher than that of less sustainable ones, a portfolio like this would even get to show positive average returns, despite the multiplier on sustainability definitely being negative.

The SUS portfolio effectively has 0 beta, so it might not be obvious from (3.9) why betas' differential appears in (3.11). The reason is that the ESG intensity of this portfolio $\frac{\mathrm{ESG}_t^{SUS}}{\mathrm{ESG}_t^M}$ is not simply $\frac{\mathrm{ESG}_t^S}{\mathrm{ESG}_t^M} - \frac{\mathrm{ESG}_t^{NS}}{\mathrm{ESG}_t^M}$, but is augmented by the ESG intensity of the market portfolio, which is 1, times the position in it, which depends on the S and NS betas' differential.

[5]This has been done for example in Hsu et al. (2020) and Yang (2022).

2.2 Hedging sustainability concerns risk

It follows that risk also impacts how more sustainable assets hedge sustainability concerns shocks too. This can be seen decomposing unexpected returns into changes to expected discount rates and to expected cash-flows growth rates, as in Campbell (1996). Making the simplifying assumption of independence of aggregate sustainability concern shocks with market returns expectations and firm i fundamentals, the contemporaneous relation between an assets return and sustainability concerns is characterized by

$$\frac{\partial(r_t^i - \mathbb{E}_{t-1}\left[r_t^i\right])}{\partial\left(\lambda_t - \mathbb{E}_{t-1}[\lambda_t]\right)} \propto \left(\frac{\mathrm{ESG}_t^i}{\mathrm{ESG}_t^M} - \beta_t^i\right). \tag{3.12}$$

Specifically, for the SUS portfolio, this is

$$\frac{\partial(r_t^{SUS} - \mathbb{E}_{t-1}\left[r_t^i\right])}{\partial\left(\lambda_t - \mathbb{E}_{t-1}[\lambda_t]\right)} \propto \left(\frac{\mathrm{ESG}_t^S}{\mathrm{ESG}_t^M} - \frac{\mathrm{ESG}_t^{NS}}{\mathrm{ESG}_t^M}\right) - (\beta^S - \beta^{NS}), \tag{3.13}$$

meaning that hedging the risk of a high-minus-low sustainable portfolio can be considered an hedge only as long as more sustainable assets are not significantly riskier than lowly or non-sustainable. This is a fact that has relevance for the future as the risk pattern can change over time. For example, being R&D generally considered a risky activity, as more sustainable firm were to become the ones closer to the technology frontier and in need of innovation, as it is happening with the automotive sector, then, the sustainable portfolio can potentially become a terrible hedge for sustainability concerns.[6]

A portfolio that is market- and sustainability-neutral allows to study this effect explicitly. Consider a portfolio investing \$ 1 in the S portfolio, shorting $\mathrm{ESG}^S/\mathrm{ESG}^{NS}$ of the NS portfolio, funding this position shorting $1 - \mathrm{ESG}^S/\mathrm{ESG}^{NS}$ of a risk-free asset and hedging it shorting \$ $\beta^S - \mathrm{ESG}^S/\mathrm{ESG}^{NS}\beta^{NS}$ of the market. The expected returns of this 'Betting Against Sustainability' (BAS) portfolio is

$$\begin{aligned}\mathbb{E}_t\left[r_{t+1}^{BAS}\right] &= \mathbb{E}_t\left[r^S\right] - \frac{\mathrm{ESG}^S}{\mathrm{ESG}^{NS}}\mathbb{E}_t\left[r^{NS}\right] - (\beta^S - \frac{\mathrm{ESG}^S}{\mathrm{ESG}^{NS}}\beta^{NS})\mathbb{E}_t\left[r^M\right] && (3.14)\\ &= \tilde{\lambda}_t\left(\beta_t^{NS} - \beta_t^S\right) && (3.15)\end{aligned}$$

The, the contemporaneous relation between BAS' returns and sustainability concerns shocks is naturally

$$\frac{\partial(r_t^{BAS} - \mathbb{E}_{t-1}\left[r_t^{BAS}\right])}{\partial\left(\tilde{\lambda}_t - \mathbb{E}_{t-1}[\tilde{\lambda}_t]\right)} \propto (\beta^S - \beta^{NS}). \tag{3.16}$$

[6]See footnote 1.

This is helpful in identifying the impact that risk has on sustainability spreads.

3 A more realistic model

3.1 With sustainable bonds

Consider the presence of a bond that contributes to the required sustainability of the portfolio, such as green bonds contribute to environmental sustainability, which is indexed as $i = GB$. Then,

$$W_t = \sum_{i=1}^{I} x_t P_t^i + x_t^0 + x_t^{GB} \tag{3.17}$$

$$W_{t+1} = \sum_{i=1}^{I} x_t (P_{t+1}^i + D_{t+1}^i) + x_t^0 R_t^f + x_t^{GB} R_t^{f,GB}, \tag{3.18}$$

while the sustainability requirement becomes

$$\sum_{i=1}^{I} (x_t^i P_t^i) \cdot \mathrm{ESG}_t^i + x_t^{GB} \cdot \mathrm{ESG}_t^{GB} > q_t \cdot W_t. \tag{3.19}$$

The first-order condition related to the holdings of the sustainable risk-free asset, which can be obtained from a Lagrangian formed in a similar way to the previous section, implies the spread on the sustainable bond being determined as

$$R_t^{f,GB} - R_t^f = -\frac{\lambda_t \mathrm{ESG}_t^{GB}}{\mathbb{E}_t \left[u'(W_{t+1}) \right]}, \tag{3.20}$$

which is non-positive. The optimality condition for the risky asset i, combined with the previous condition, results in a more explicit pricing condition:

$$\mathbb{E}_t \left[r_{t+1}^i \right] = \frac{\mathrm{ESG}_t^i}{\mathrm{ESG}_t^{GB}} \cdot r_t^{f,GB} - \mathrm{Cov}_t \left[\frac{u'(W_{t+1})}{\mathbb{E}_t \left[u'(W_{t+1}) \right]}, r_{t+1}^i \right], \tag{3.21}$$

where r_{t+1}^i is the excess return $R_{t+1}^i - R_t^f$. This essentially deviates from a standard formulation for the term multiplying the sustainable bond rate.

In a CAPM world, this translates to

$$\mathbb{E}_t \left[r_{t+1}^i \right] = r_t^{f,GB} \frac{\mathrm{ESG}_t^M}{\mathrm{ESG}_t^{GB}} \left(\frac{\mathrm{ESG}_t^i}{\mathrm{ESG}_t^M} - \beta^i \right) + \beta^i \mathbb{E}_t \left[r_{t+1}^M \right], \tag{3.22}$$

where $r_t^{f,GB} \frac{\mathrm{ESG}_t^M}{\mathrm{ESG}_t^{GB}}$ essentially provides a proxy of $\widetilde{\lambda}_t$. 'Sustainable' bonds are essentially available for environmental sustainability, but it still useful as it allows an easier empirical test. It follows that the expected returns of a plain high-minus-low sustainability portfolio, previously addressed as 'SUS' and determined by (3.11), is now

$$\mathbb{E}\left[r^{SUS}\right] = r_t^{f,GB} \frac{\text{ESG}_t^M}{\text{ESG}_t^{GB}} \left[\left(\frac{\text{ESG}_t^S}{\text{ESG}_t^M} - \frac{\text{ESG}_t^{NS}}{\text{ESG}_t^M} \right) - (\beta^S - \beta^{NS}) \right] \tag{3.23}$$

and the returns of the sustainability- and market-neutral BAS portfolio, previously defined by (3.14), amounts now to

$$\mathbb{E}\left[r^{BAS}\right] = r_t^{f,GB} \frac{\text{ESG}_t^M}{\text{ESG}_t^{GB}} (\beta^{NS} - \beta^S). \tag{3.24}$$

3.2 Factor-zoo world

Given the remarkable amount of evidence in favour of multi-factor models, it is useful to consider the case in which the marginal utility is affected by risks other than undiversifiable variance. Rearranging (3.5) and (3.6), one can obtain a broader formulation of the pricing condition that any asset has to abide by:

$$\mathbb{E}_t\left[r_{t+1}^i\right] = \tilde{\lambda}_t \left(\frac{\text{ESG}_t^i}{\text{ESG}_t^M} - \frac{\text{Cov}_t\left[u'(W_{t+1}), r_{t+1}^i\right]}{\text{Cov}_t\left[u'(W_{t+1}), r_{t+1}^M\right]} \right) + \frac{\text{Cov}_t\left[u'(W_{t+1}), r_{t+1}^i\right]}{\text{Cov}_t\left[u'(W_{t+1}), r_{t+1}^M\right]} \mathbb{E}_t\left[r_{t+1}^M\right], \tag{3.25}$$

or, more simply,

$$\mathbb{E}_t\left[r_{t+1}^i\right] = \tilde{\lambda}_t \left(\frac{\text{ESG}_t^i}{\text{ESG}_t^M} - \gamma_t^i \right) + \gamma_t^i \cdot \mathbb{E}_t\left[r_{t+1}^M\right]. \tag{3.26}$$

This follows closely what was previously derived, just with a broader measure of risk, $\gamma_t^i = \frac{\text{Cov}_t\left[u'(W_{t+1}), r_{t+1}^i\right]}{\text{Cov}_t\left[u'(W_{t+1}), r_{t+1}^M\right]}$. While $u'(W_{t+1})$ is empirically not easy to identify, all of the asset pricing literature provides alternatives, one of which is its projection on the returns, as in the similar application in Franceschini (2023).

This implies

$$\mathbb{E}\left[r^{SUS}\right] = \tilde{\lambda} \left[\left(\frac{\text{ESG}_t^S}{\text{ESG}_t^M} - \frac{\text{ESG}_t^{NS}}{\text{ESG}_t^M} \right) - (\gamma^S - \gamma^{NS}) \right] \tag{3.27}$$

and

$$\mathbb{E}\left[r^{BAS}\right] = \tilde{\lambda}(\gamma^{NS} - \gamma^S). \tag{3.28}$$

It also naturally extends to the case where sustainable bonds are available, simply plugging $r_t^{f,GB} \frac{\mathrm{ESG}_t^M}{\mathrm{ESG}_t^{GB}}$ for $\tilde{\lambda}_t$.

4 A look at ESG constraints

4.1 Test assets

The analysis focuses on the US stock market; monthly data on stocks is provided by CRSP. The time span of the analysis is ultimately determined by the yearly ESG data, which is from Refinitiv. The series starts in 2002, but it had a remarkable increase in coverage in 2003, which is when the analysis starts from, ending in 2021. Table 3.3 displays the fraction of the US stock market coverage by the ESG measure. By no means the analysis has to be confined to this measure, indeed it will not be, moving on, as many more sustainability measures can be associated to constraints in portfolio formation.

The test assets pool is formed by portfolios of stocks sorted by characteristics that are known to generate dispersion in average returns and are widely used in the literature. Three of these are based on market capitalization (size), three on book-to-market (B/M) equity ratio, three on the cumulated return over the 11 months ending a month before (mom), and three portfolios based on the ESG score. Finally, the SUS portfolio and the market are added, for a total of 14 test assets. The number of portfolios per characteristic is chosen to ensure a minimum of approximately 60 firms per portfolio, given the limited number of firms covered by the ESG measure. Size and B/M portfolios are formed based on NYSE capitalization quantiles. All of the portfolios are value-weighted to obtain the portfolios' returns and portfolios' ESG score at every date. Statistics of the portfolios are in table 3.1.

A clear pattern, obviously, emerges in the esg-sorted portfolio as well as in the size portfolios as well as in the B/M, where ESG scores increase with size and decrease with B/M. Momentum portfolios instead appears to have no meaningful relation to esg scores. Returns mostly conform with the literature, with returns being lower for greater size portfolios and lower B/M ratios. Momentum portfolios show a more bizarre behaviour, with worse performing assets having higher returns, although the unexplained returns by CAPM increase with momentum as expected. More sustainable assets, as expected from theory have lower returns, reflected in the negative average return of the SUS portfolio. To have a first sight of the relation between sustainability and risk, figure 3.1 charts the relative ESG scores and 4-year rolling window of the high- and low-sustainability portfolios. Most of the variation comes from the beta of the least portfolio, 'non-sus', but the difference in risk reached at most half the difference in relative ESG scores, up to 2022.

Table 3.1: Test assets statistics. In parenthesis the HAC standard errors, computed following Lazarus et al. (2018). Monthly returns annualized, sample from Jul 2004 to Dec 2021.

Stat	size.1	size.2	size.3	BM.1	BM.2	BM.3	mom.1	mom.2	mom.3	sus.1	sus.2	sus.3	sus.hml	AVG
r. ESG (%)	43.030	56.621	102.267	103.449	100.233	88.766	99.209	100.736	96.885	36.046	67.859	121.551	85.505	84.781
	(7.204)	(1.460)	(0.827)	(1.785)	(1.814)	(3.451)	(1.422)	(0.567)	(1.440)	(1.881)	(6.531)	(4.708)	(1.693)	
Avg. ret (%)	40.100	17.176	10.366	11.245	8.495	13.126	10.045	10.525	9.745	12.809	11.033	9.952	-4.003	12.355
SD (%)	282.196	89.919	55.454	50.995	55.476	87.238	77.914	50.700	52.918	67.499	62.668	56.340	37.412	78.979
CAPM α (%)	17.820	2.937	0.305	2.151	-1.438	0.521	-2.824	1.438	0.582	1.280	-0.068	0.070	-2.385	1.568
	(17.497)	(4.096)	(1.073)	(1.510)	(0.989)	(3.010)	(2.034)	(1.401)	(1.091)	(1.860)	(1.459)	(1.105)	(1.463)	
CAPM β	2.275	1.454	1.027	0.928	1.014	1.287	1.314	0.928	0.936	1.177	1.134	1.009	-0.165	1.101
	(0.536)	(0.103)	(0.038)	(0.018)	(0.020)	(0.132)	(0.100)	(0.027)	(0.042)	(0.063)	(0.038)	(0.037)	(0.050)	

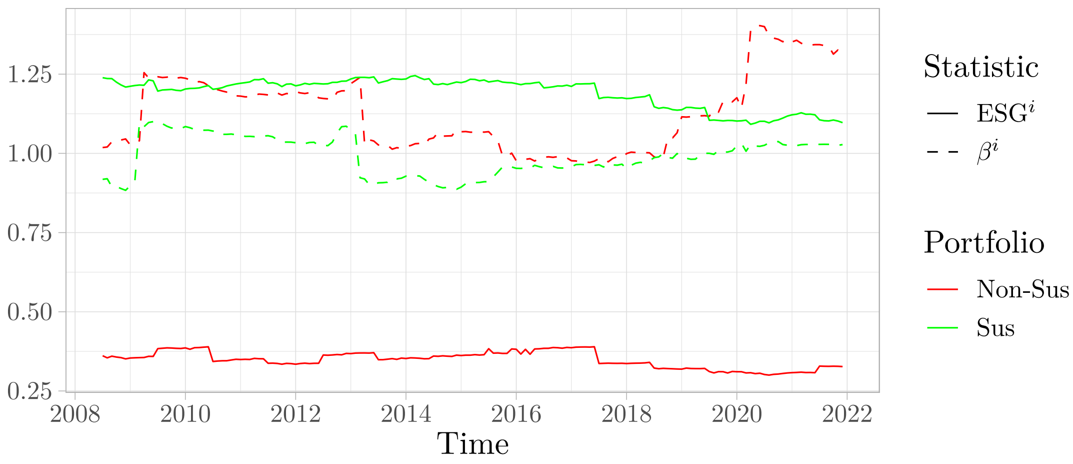

Figure 3.1: ESG intensity and 4-year rolling window of high- and low-sustainability portfolios, both reported on the vertical axis.

4.2 Risk factors

The scaled marginal utility is assumed to be linear in a set of risk factors $\mathbf{f}$, i.e. $\frac{u'(W_{t+1})}{\mathbb{E}[u'(W_{t+1})]} = 1 - \mathbf{f}_t'\mathbf{b}$. Four sets of factors are used: (1) market factor only; (2) the three factors of Fama and French (1993), which include the market, named 'FF3'; (3) the market and the first two principal components of the test assets, orthogonalized with respect to the market to avoid redundant information, named 'F(3)'; (4) the market and six principal components, named 'F(7)'. The number of factors are chosen with two criteria: match the size of FF3 for comparison and not including components that explain less than 5% of returns variance given the cumulative variance explained being around 90% already. The scree plot of the Principal component analysis is in figure 3.2. More sets can be tested, with a special interest in including state variables that capture sustainability-relevant concerns.

4.3 Pricing results

The main condition of interest is (3.5), which traslates into the moment condition

$$\mathbb{E}\left[r_t^i\right] - a \cdot \mathrm{ESG}_{t-1}^i - \mathbb{E}\left[(\mathbf{f}_t'\mathbf{b})(r_t^i - \mathbb{E}\left[r_t^i\right])\right] = 0, \tag{3.29}$$

where a return the unconditional expectations of $\tilde{\lambda}$ and $\mathbf{f}$ are the factors spanning the scaled marginal utility with loadings $\mathbf{b}$, which are to be estimated together with a. The covariances are then estimated contemporaneously exploiting the moment

$$s_i + \mathbb{E}\left[(\mathbf{f}_t'\mathbf{b})(r_t^i - \mathbb{E}\left[r_t^i\right])\right] = 0, \tag{3.30}$$

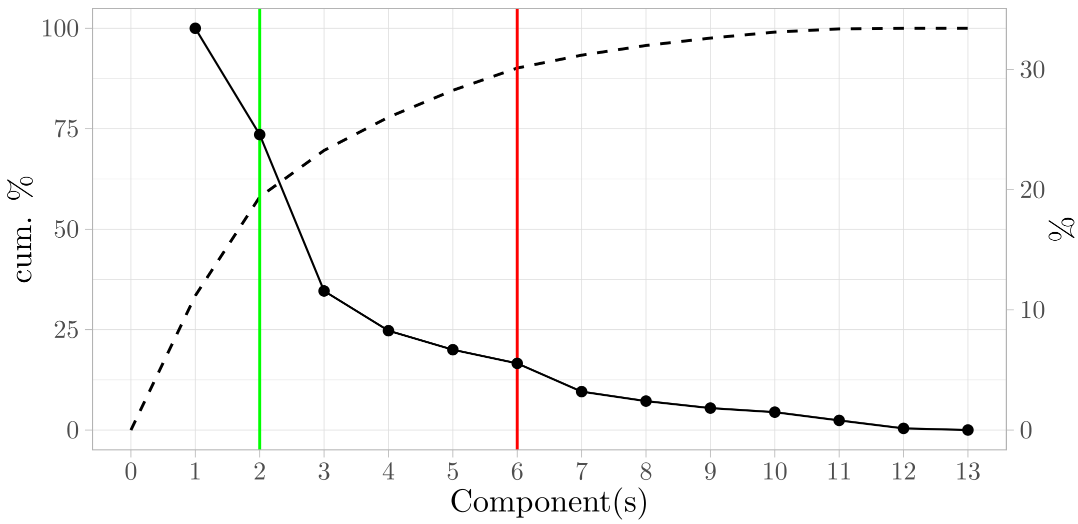

Figure 3.2: Scree plot of principal component analysis. On the horizontal axis the number of component considered, on the vertical axis the share of variance explained by the relative component (solid line) and the cumulative share of variance explained (dotted line). The green vertical line marks the 2^{nd} component while the red one the 6^{th}.

where s_i is the empirical counterpart of $\text{Cov}\left[\frac{u'(W_{t+1})}{\mathbb{E}[u'(W_{t+1})]}, r^i_{t+1}\right]$, employed to obtain estimates of the γs. The estimation results are in table 3.2.

The unconditional sustainability spread $\mathbb{E}\left[\tilde{\lambda}\right] = a \cdot \mathbb{E}\left[\text{ESG}^M_t\right]$ is negative and slightly significant only for 'intermediately' complex models, but not in all of them. The market factor is always positive, and positive in all estimations apart from the biggest model, F(7). Anyway, all the models' validity is rejected by the J-test. Adding factors beyond the market appears to have an effect only for the most complex model of all, F(7), which has remarkably lower Mean Absolute Pricing Errors (MAPE) than the other models. The intermediate models show no material improvement on the single factor model overall, although the SUS portfolio gets priced better, in terms of MAPE. This results overall suggest that this model does not describe reality very accurately; it is likely that the measure of ESG employed here is not one that is related to strongly binding constraints. Nonetheless, estimates provide weak support towards the existence of a negative spread related to sustainability.

To observe the impact of risk on univariate sustainability portfolios such as SUS, the γs implied by the previous estimation are computed and reported in table 3.3. Specifically, the differential in ESG intensity is

$$\Delta\text{ESG}_t = \frac{\text{ESG}^S_t}{\text{ESG}^M_t} - \frac{\text{ESG}^{NS}_t}{\text{ESG}^M_t} \tag{3.31}$$

Table 3.2: cross-sectional pricing estimation, HAC standard errors in parenthesis. Coverage 2004-2021.

	Mkt only	FF3	F(3)	F(7)
$\mathbb{E}\left[\tilde{\lambda}\right]$	-5.753	-8.844^{*}	-5.146^{*}	1.717
	(4.023)	(4.810)	(3.053)	(3.768)
Market factors	0.302^{**}	0.514^{**}	0.274^{**}	0.142
	(0.131)	(0.224)	(0.118)	(0.114)
J-test	94.3^{***}	105.5^{***}	100.2^{***}	97.9^{***}
MAPE	2.431	2.309	2.446	0.831
HML a.p.e.	3.479^{**}	1.864	2.187	-1.392
	(1.575)	(1.701)	(1.494)	(2.217)

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

and the differential in risks is

$$\Delta\gamma = \gamma^{S} - \gamma^{NS}. \tag{3.32}$$

It can be seen that the estimates, read with (3.27), do not suggest risk affecting significantly how the spread $\tilde{\lambda}$ gets reflected into $\mathbb{E}\left[r^{SUS}\right]$: it is unlikely that $\Delta\gamma > \Delta\text{ESG}$, even though it can be seen that considering more risks results in higher $\Delta\gamma$. More accurate analysis regarding the impact of risk through the sustainability channel would stem from exploiting (3.28), but the previous results suggest that other constrained-characteristics, such as carbon emissions, might be better case studies for the mechanisms highlighted in this paper.

5 Conclusion

Assessing how sustainability affects asset prices, and thus investors' welfare, is key to better address current generational challenges. It is shown that a constraint on the average score of the portfolio held by an agent that binds induces a negative spread on the risk premium of an asset that is proportional to its sustainability score. In other words, the sustainability provided by the asset is appreciated by the agent. The spread, however, is not only function of how constrained the agent is or how sustainable the asset is, but also of how risky the asset is. This happens because it is always possible to reach the same risk of an asset by leveraging risk factors portfolios, which provides their own sustainability scores; then, an asset's sustainability should be remunerated only if higher than this equivalently-levered sustainability score. This in turn affects the returns of naive portfolios based on univariate sorting on sustainability, which can be used to capitalize the spread or to hedge shocks to the constraint, i.e. sustainability concerns. Empirical analysis conducted

Table 3.3: ΔESG and $\Delta\gamma$. In parenthesis HAS standard errors. Standard errors of $\Delta\gamma$ are obtained via Delta method from the covariances estimates from the GMM step.

ΔESG	$\Delta\gamma$			
	Market	FF3	F(3)	F(4)
0.855***	−0.165**	0.094	−0.127	−0.637
(0.017)	(0.067)	(0.135)	(0.165)	(0.791)

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$

on an ESG measure covering US data is not conclusive regarding the existence of such a sustainability premium in the first place and that risk actually plays a relevant role either. Anyway, this analysis can be performed on all the other sustainability-related measures and many other countries. Further, was the risk delta not to be found significant at current times, does not mean that it will not be in the future as the economy structure evolves.

References

Ardia, David et al. (16, 2022). 'Climate Change Concerns and the Performance of Green vs. Brown Stocks'. In: *Management Science*.

Berg, Florian, Julian F Kölbel, and Roberto Rigobon (1, 2022). 'Aggregate Confusion: The Divergence of ESG Ratings*'. In: *Review of Finance* 26.6, pp. 1315–1344.

Campbell, John Y. (1996). 'Understanding Risk and Return'. In: *Journal of Political Economy* 104.2, pp. 298–345.

Fama, Eugene F. and Kenneth R. French (1993). 'Common risk factors in the returns on stocks and bonds'. In: *Journal of Financial Economics* 33.1, pp. 3–56.

Franceschini, Fabio (2023). 'Does CAPM overestimate more the risk or its price?'

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 Johannes Stroebel (2021). 'Climate Finance'. In.

Hsu, Po-Hsuan, Kai Li, and Chi-Yang Tsou (2020). 'The Pollution Premium'. In: *SSRN Electronic Journal*.

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

Noailly, Joëlle, Laura Minu Nowzohour, and Matthias Van Den Heuvel (2021). 'Heard the news ? Environmental policy and clean investments'. In: CIES Research Paper (70).

Pástor, Ľuboš, Robert F. Stambaugh, and Lucian A. Taylor (1, 2021). 'Sustainable investing in equilibrium'. In: *Journal of Financial Economics* 142.2, pp. 550–571.

— (1, 2022). 'Dissecting green returns'. In: *Journal of Financial Economics* 146.2, pp. 403–424.

Pedersen, Lasse Heje, Shaun Fitzgibbons, and Lukasz Pomorski (1, 2021). 'Responsible investing: The ESG-efficient frontier'. In: *Journal of Financial Economics* 142.2, pp. 572–597.
Yang, Biao (2022). 'Explaining Greenium in a Macro-Finance Integrated Assessment Model'. In.

A Derivations

The Lagrangian related to the problem with sustainable bonds is

$$\begin{aligned}\mathcal{L}_t = \mathbb{E}_t & \left[u \left(W_t R_t^f + \sum_{i=1}^{I} x_t (P_{t+1}^i + D_{t+1}^i - P_t^i R_t^f) + x_t^{GB} (R_t^{f,GB} - R_t^f) \right) \right] \\ & + \lambda_t \left\{ \sum_{i=1}^{I} (x_t^i P_t^i) \cdot \mathrm{ESG}_t^i + x_t^{GB} \cdot \mathrm{ESG}_t^{GB} - q_t \cdot W_t \right\}\end{aligned}. \tag{3.33}$$

The resulting first order conditions are

$$\mathbb{E}_t \left[u'(W_{t+1}) \right] (R_t^{f,GB} - R_t^f) + \lambda_t \mathrm{ESG}_t^{GB} = 0 \tag{3.34}$$

for the sustainable bond, and

$$\mathbb{E}_t \left[u'(W_{t+1})(R_{t+1}^i - R_t^f) \right] + \lambda_t \mathrm{ESG}_t^i = 0 \tag{3.35}$$

for the risky asset i.

Combining sustainable bond and common asset conditions:

$$\frac{\mathbb{E}_t \left[u'(W_{t+1})(R_{t+1}^i - R_t^f) \right]}{\mathrm{ESG}_t^i} = \frac{\mathbb{E}_t \left[u'(W_{t+1}) \right] (R_t^{f,GB} - R_t^f)}{\mathrm{ESG}_t^{GB}} \tag{3.36}$$

$$\mathbb{E}_t \left[u'(W_{t+1})(R_{t+1}^i - R_t^f) \right] = \frac{\mathrm{ESG}_t^i}{\mathrm{ESG}_t^{GB}} \mathbb{E}_t \left[u'(W_{t+1}) \right] (R_t^{f,GB} - R_t^f). \tag{3.37}$$

So,

$$\mathbb{E}_t \left[r_{t+1}^i \right] = \frac{\mathrm{ESG}_t^i}{\mathrm{ESG}_t^{GB}} \cdot r_t^{f,GB} - \mathrm{Cov}_t \left[\frac{u'(W_{t+1})}{\mathbb{E}_t \left[u'(W_{t+1}) \right]}, r_{t+1}^i \right]. \tag{3.38}$$

Then, the market has expected excess return

$$\mathbb{E}_t \left[r_{t+1}^M \right] = \sum_{i=1}^{I} \omega_t^i \cdot \mathbb{E}_t \left[r_{t+1}^i \right] \tag{3.39}$$

$$= \frac{\mathrm{ESG}_t^M}{\mathrm{ESG}_t^{GB}} \cdot r_t^{f,GB} - \mathrm{Cov}_t \left[\frac{u'(W_{t+1})}{\mathbb{E}_t \left[u'(W_{t+1}) \right]}, r_{t+1}^M \right]. \tag{3.40}$$

Assuming $\frac{u'(W_{t+1})}{\mathbb{E}_t[u'(W_{t+1})]} = a_t - b_t \cdot r^M_{t+1}$,

$$\mathbb{E}_t\left[r^i_{t+1}\right] = \frac{\text{ESG}^i_t}{\text{ESG}^{GB}_t} \cdot r^{f,GB}_t + b_t \text{Cov}_t\left[r^M_{t+1}, r^i_{t+1}\right] \tag{3.41}$$

$$\mathbb{E}_t\left[r^M_{t+1}\right] = \frac{\text{ESG}^M_t}{\text{ESG}^{GB}_t} \cdot r^{f,GB}_t + b_t \text{Var}_t\left[r^M_{t+1}\right]. \tag{3.42}$$

Then, $b_t = \frac{\mathbb{E}_t[r^M_{t+1}] - \frac{\text{ESG}^M_t}{\text{ESG}^{GB}_t} \cdot r^{f,GB}_t}{\text{Var}_t[r^M_{t+1}]}$ and

$$\mathbb{E}_t\left[r^i_{t+1}\right] = \frac{\text{ESG}^i_t}{\text{ESG}^{GB}_t} \cdot r^{f,GB}_t + \left(\frac{\mathbb{E}_t\left[r^M_{t+1}\right] - \frac{\text{ESG}^M_t}{\text{ESG}^{GB}_t} \cdot r^{f,GB}_t}{\text{Var}_t\left[r^M_{t+1}\right]}\right) \text{Cov}_t\left[r^M_{t+1}, r^i_{t+1}\right] \tag{3.43}$$

$$= \frac{\text{ESG}^i_t}{\text{ESG}^{GB}_t} \cdot r^{f,GB}_t + \left(\mathbb{E}_t\left[r^M_{t+1}\right] - \frac{\text{ESG}^M_t}{\text{ESG}^{GB}_t} \cdot r^{f,GB}_t\right) \beta^i \tag{3.44}$$

$$= \frac{r^{f,GB}_t}{\text{ESG}^{GB}_t}\left(\text{ESG}^i_t - \text{ESG}^M_t \beta^i\right) + \beta^i \mathbb{E}_t\left[r^M_{t+1}\right] \tag{3.45}$$

$$= r^{f,GB}_t \frac{\text{ESG}^M_t}{\text{ESG}^{GB}_t}\left(\frac{\text{ESG}^i_t}{\text{ESG}^M_t} - \beta^i\right) + \beta^i \mathbb{E}_t\left[r^M_{t+1}\right] \tag{3.46}$$

For a factor-zoo with a sustainable bond:

$$\frac{\mathbb{E}_t\left[r^i_{t+1}\right] - \frac{\text{ESG}^i_t}{\text{ESG}^{GB}_t} r^{f,GB}_t}{\text{Cov}_t\left[\frac{u'(W_{t+1})}{\mathbb{E}_t[u'(W_{t+1})]}, r^i_{t+1}\right]} = -1 = \frac{\mathbb{E}_t\left[r^M_{t+1}\right] - \frac{\text{ESG}^M_t}{\text{ESG}^{GB}_t} r^{f,GB}_t}{\text{Cov}_t\left[\frac{u'(W_{t+1})}{\mathbb{E}_t[u'(W_{t+1})]}, r^M_{t+1}\right]}, \tag{3.47}$$

so

$$\frac{\mathbb{E}_t\left[r^i_{t+1}\right] - \frac{\text{ESG}^i_t}{\text{ESG}^{GB}_t} r^{f,GB}_t}{\text{Cov}_t\left[\frac{u'(W_{t+1})}{\mathbb{E}_t[u'(W_{t+1})]}, r^i_{t+1}\right]} = \frac{\mathbb{E}_t\left[r^M_{t+1}\right] - \frac{\text{ESG}^M_t}{\text{ESG}^{GB}_t} r^{f,GB}_t}{\text{Cov}_t\left[\frac{u'(W_{t+1})}{\mathbb{E}_t[u'(W_{t+1})]}, r^M_{t+1}\right]} \tag{3.48}$$

$$\mathbb{E}_t\left[r^i_{t+1}\right] - \frac{\text{ESG}^i_t}{\text{ESG}^{GB}_t} r^{f,GB}_t = \frac{\text{Cov}_t\left[\frac{u'(W_{t+1})}{\mathbb{E}_t[u'(W_{t+1})]}, r^i_{t+1}\right]}{\text{Cov}_t\left[\frac{u'(W_{t+1})}{\mathbb{E}_t[u'(W_{t+1})]}, r^M_{t+1}\right]} \left(\mathbb{E}_t\left[r^M_{t+1}\right] - \frac{\text{ESG}^M_t}{\text{ESG}^{GB}_t} r^{f,GB}_t\right) \tag{3.49}$$

$$\mathbb{E}_t\left[r^i_{t+1}\right] = \frac{\text{ESG}^i_t}{\text{ESG}^{GB}_t} r^{f,GB}_t + \frac{\text{Cov}_t\left[u'(W_{t+1}), r^i_{t+1}\right]}{\text{Cov}_t\left[u'(W_{t+1}), r^M_{t+1}\right]} \left(\mathbb{E}_t\left[r^M_{t+1}\right] - \frac{\text{ESG}^M_t}{\text{ESG}^{GB}_t} r^{f,GB}_t\right) \tag{3.50}$$

$$\mathbb{E}_t\left[r^i_{t+1}\right] = \frac{r^{f,GB}_t}{\text{ESG}^{GB}_t}(\text{ESG}^i_t - \text{ESG}^M_t \gamma^i_t) + \gamma^i_t \mathbb{E}_t\left[r^M_{t+1}\right] \tag{3.51}$$

$$\mathbb{E}_t\left[r_{t+1}^i\right] = r_t^{f,GB}\frac{\text{ESG}_t^M}{\text{ESG}_t^{GB}}\left(\frac{\text{ESG}_t^i}{\text{ESG}_t^M} - \gamma_t^i\right) + \gamma_t^i \cdot \mathbb{E}_t\left[r_{t+1}^M\right]. \tag{3.52}$$

B Additional tables and figures

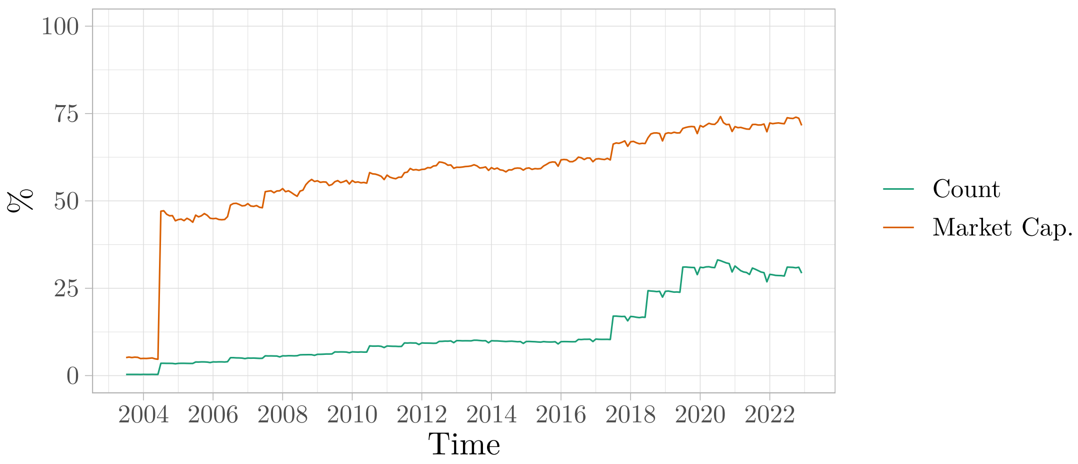

Figure 3.3: Refinitiv ESG scores coverage of the US market stocks in CRSP monthly file.