The Lost Capital Asset Pricing Model

Daniel Andrei* Julien Cujean† Mungo Wilson‡

November 30, 2020

Abstract

We provide a novel explanation for the empirical failure of the CAPM despite its widespread practical use. In a rational-expectations economy in which information is dispersed, variation in expected returns over time and *across investors* creates an informational gap between investors and the empiricist. The CAPM holds for investors, but appears flat to the empiricist. Variation in expected returns across investors accounts for the larger part of this distortion, which is empirically substantial; it offers a new interpretation of why "Betting Against Beta" works: BAB really bets on true beta. The empiricist retrieves a stronger CAPM on macroeconomic announcement days.

We are grateful for comments received from Hengjie Ai, Laurent Barras, Jonathan Berk, Philip Bond, Michael Brennan, Bernard Dumas, Philip Dybvig, Ken French, Diego Garcia, Lorenzo Garlappi, Mark Grinblatt, Philipp Illeditsch, Christian Julliard, Raymond Kan, Ralph Koijen, Pete Kyle, Hanno Lustig, Christoph Meinerding, Stijn Van Nieuwerburgh, Lasse Pedersen, Marcel Rindisbacher, Richard Roll, Shri Santosh, Avanidhar Subrahmanyam, Laura Veldkamp, Zhenyu Wang, Pierre-Olivier Weill, Brian Weller, and conference and seminar participants at the 2016 OTC Markets and Securities Conference in Santa Barbara, the Macro-Finance lunch at UCLA, the University of Maryland, the University of Colorado Boulder, the University of Washington, the 2017 HEC-McGill Finance Conference, the UBC Winter Finance Conference, the 2017 LA Finance Day, the 2017 SFS Cavalcade, the 2017 WFA, the 2017 Gerzensee Asset Pricing Meeting, HKUST, City University of Hong Kong, University of Hong Kong, the Chinese University of Hong Kong, McGill University, INSEAD, Boston University, Imperial College, the 2018 AFA, and the 2018 NBER 2018 SI Asset Pricing. We thank Michael Tang and Nan Ma for outstanding research assistance.

*McGill University, Desautels Faculty of Management, Office 549, 1001 Sherbrooke Street West, Montréal, Québec, H3A 1G5, Canada, danielandrei.info, daniel.andrei@mcgill.ca.

†University of Berne, Institute for Financial Management, Engehaldenstrasse 4, 3012 Bern, Switzerland, juliencujean.com, julien.cujean@ifm.unibe.ch.

‡Said Business School and Oxford-Man Institute, Oxford University, United Kingdom; +44 (0)1865 288914; mungo.wilson@sbs.ox.ac.uk.

1 Introduction

There is a growing tension between the theory of financial economics and its application. The Capital Asset Pricing Model, a theoretical pillar of modern finance, fails in empirical tests.[1] The consensus among economists is that beta does not explain expected returns, largely shaping the view that the CAPM does not hold. But a flagrant affront to this view is that the CAPM remains to this day the model that investors and firms most widely use.[2] Adding to the controversy, the CAPM does hold on particular occasions, e.g., on announcement days, or at night.[3] Why do economists keep rejecting a theory that practitioners refuse to abandon?

This paper explores the idea that traditional empirical tests reject the CAPM when in fact it is the correct asset pricing model from investors' perspective. Of course, there are many reasons not to believe the CAPM is the correct canonical asset pricing model; and there are as many ways it could fail empirically. However, in this paper, we present a situation in which the CAPM holds from the perspective of investors, but it fails empirically in one specific way: the empiricist perceives a "flat" Securities Market Line, which becomes steeper occasionally, e.g., when public information is released.

We build our argument in a rational-expectations model of informed trading in which a continuum of mean-variance investors trade multiple assets (e.g., Admati, 1985). Investors use their own private information and the information they infer from prices to predict future excess returns. Even though returns are predictable from the perspective of investors, the operation of the law of iterated expectations ensures that they all observe the same unconditional CAPM relation. Yet, return predictability *leaves a mark* on the CAPM relation that the empiricist estimates: the operation of the law of total variance implies that the betas the empiricist measures do depend on the extent to which returns are predictable. Critically, expected returns vary across investors when their information is dispersed. Hence, variation across investors leaves a mark on traditional CAPM tests, the main point of this paper.

We emphasize that our framework is not a standard CAPM environment—investors have information that is dispersed among them (Lintner, 1969) and that the empiricist does not observe (Hansen and Richard, 1987). Hence, the CAPM relation and the sense in which

[1] See Fama and French (2004) for a comprehensive review.

[2] The CAPM is the most widely used model to make investment decisions (Berk and Van Binsbergen, 2016; Barber, Huang, and Odean, 2016) and to compute the cost of capital (Graham and Harvey, 2001).

[3] Savor and Wilson (2014) document a strong relationship between expected returns and betas on days when news about inflation, unemployment, or FOMC interest rate decisions is scheduled to be announced. Hendershott, Livdan, and Rösch (2018) document a strong relationship when the market is closed (at night). Ben-Rephael, Carlin, Da, and Israelsen (2017) document that the CAPM performs better when institutions demand more information.

it holds must be redefined. There is CAPM pricing in the sense that unconditional betas depend on information that investors know, a result that goes back to Admati (1985).[4] Thus, the argument that the CAPM is correctly rejected because it does *not* hold (e.g., Merton, 1987) does not apply. In this framework, the CAPM is incorrectly rejected because the notion of beta that underlies it is not the traditional beta that empiricists commonly compute.[5]

Variation in investors' expectations increases the dispersion in empiricist's betas relative to investors' betas. Since all betas (correct or incorrect) must average one (*market's beta*), the empiricist inflates all betas above market's beta and deflates all others. As a result, the empiricist perceives risky assets as riskier than they really are, and safe assets as safer than they really are. Although the empiricist and investors disagree about betas (by the law of total variance), they agree on unconditional expected returns (by the law of iterated expectations). Since market's beta is the only beta on which they agree, the empiricist's SML rotates clockwise around the market portfolio, which flattens its slope and creates a positive intercept—the SML looks "flat."

The informational distance between investors and the empiricist originates from two sources of variation in investors' expectations. First, there is aggregate (time) variation in consensus expected returns, of the kind studied in Jagannathan and Wang (1996). However, time-series variation alone is often found insufficient to explain asset-pricing anomalies (Lewellen and Nagel, 2006). Our focus is on the second source of variation, which results from dispersion in investors' information. Because investors' information is dispersed there is variation in expected returns across investors, too. In this model time-series and cross-sectional variation in investors' information reinforce each other, leading together to a flat SML. To our knowledge, variation across investors—as opposed to variation over time—has been neglected in empirical tests of the CAPM. We show that the impact of cross-sectional variation on these tests is substantial, stronger than that of time-series variation.

An empirical and methodological contribution of this paper is to measure how expected returns on pairs of stocks covary across investors. We use this measure to correct beta estimates according to our theory, which reveals a much better-performing CAPM. We first obtain proxies for consensus and individual expected returns for a large cross section of stocks. I/B/E/S is one database that offers such a rich panel of individual expectations; it provides one-year price targets made by a large set of analysts for S&P500 firms (among others), starting in 1999. From these price targets we construct expected returns, which we

[4] See Easley and O'Hara (2004), Fama and French (2007), Van Nieuwerburgh and Veldkamp (2010), Banerjee (2010), and Biais, Bossaerts, and Spatt (2010).

[5] See Roll (1978) and Dybvig and Ross (1985).

use to assess quantitatively how variation over time and across investors affect together and separately beta mismeasurement. As the theory predicts, both channels contribute to SML flattening, with variation across investors accounting for the larger part.

Our tests rely on a new form of beta *(dispersion beta)*, which measures how expected returns on a given stock and on the market covary across investors. For instance, consider an investor who systematically under-estimates returns relative to consensus and another investor who does just the opposite. Computed across these two investors, dispersion beta is positive: although investors deviate from consensus in opposite direction, their expectations on the stock and on the market individually deviate from consensus in the same direction. If, further, each investor's deviation from consensus is larger on the stock than it is on the market, dispersion beta will be larger than one. Thus, just as traditional beta compares variation over time of realized returns on a stock to those on the market, dispersion beta compares variation across investors of expected returns on a stock to those on the market.

Our main result that investors' betas are shrunk towards one relative to empiricist's traditional betas is strongly confirmed in the data. Interestingly, this result corresponds to the way practitioners adjust beta estimates (e.g., "ADJ BETA" on Bloomberg terminals). This adjustment in our model and in practice has different origins. Practitioners use it to reduce sampling biases (Vasicek, 1973), which are absent in our model; in this paper it results entirely from the informational distance between investors and the empiricist. We can compare, however, how much shrinkage our theory and empirics imply with how much shrinkage finance textbooks recommend (e.g, Berk and DeMarzo, 2007). The Bloomberg adjustment is: $(2/3)\times$Raw Beta$+(1/3)\times 1$. Our proposed adjustment is: $(1/2)\times$Raw Beta$+$ $(1/2)\times 1$. Thus the adjustment used in practice is likely too conservative.

An alternative explanation is that CAPM flattening is caused by leverage constraints (Black, 1972; Frazzini and Pedersen, 2014). We examine whether there is sufficient variation in expected returns (both over time and across agents) to explain abnormal returns on the "Betting Against Beta" (BAB) anomaly. In our data sample, we cannot reject the theoretical possibility that returns on BAB result from beta mismeasurement, with variation across agents playing a particularly important role in explaining returns on BAB. Although we do not dispute the success of the BAB strategy, our interpretation differs: we claim that a large part of this success is because "betting against *measured* beta is betting on *true* beta."

Variation across investors' information (dispersed information) causes larger CAPM distortion than aggregate variation (public information) does, everything else being equal. Formally, consider two economies that reveal an equivalent amount of information to investors, but one in which this information is dispersed and one in which it is public (Albagli, Hellwig, and Tsyvinski, 2015). In the economy in which all information is public, the empiricist's

SML is steeper. Investors' betas and the unconditional risk premium are, however, identical across the two economies. Thus empiricist's SML steepens through a compression in measured betas when public information dominates (e.g., FOMC meetings). Recent empirical work provides evidence supporting this result (Bodilsen, 2019; Andersen, Thyrsgaard, and Todorov, 2020; Eriksen and Gronborg, 2020).

We examine how our conclusions depend on modeling assumptions. When information is entirely public and the market portfolio is equally weighted, SML flattening is systematic, irrespective of modeling assumptions. When information is dispersed, the structure of payoffs matters; however, absent residual uncertainty in payoffs and under conditions on private signals, dispersed information always amplifies flattening. More generally, the distribution of eigenvalues of investors' covariance matrix dictates whether flattening obtains, which occurs when eigenvalues exhibit little dispersion. We also show that the SML can be downward-sloping although the risk premium is always positive in this model. For instance, this puzzling outcome depends on the composition of the market portfolio: it occurs when assets that have high measured beta simultaneously have little weight in the market portfolio.

There are several established explanations for the finding that the SML is too flat, some of which go back to the 1970's.[6] None result from variation in expected returns across investors. We believe that the fleeting appearance of the CAPM on announcement days (Savor and Wilson, 2014)—let alone its pervasive application in practice—licenses a new look at the finding that the SML is too flat. In addition to aggregate variation (Jagannathan and Wang, 1996; Lewellen and Nagel, 2006), we argue that variation across investors creates CAPM distortion. That CAPM tests fail when the market proxy is not mean-variance efficient is certainly true (Roll, 1977; Stambaugh, 1982; Roll and Ross, 1994) and is not our point. Assuming that the CAPM holds unconditionally, we argue that an empiricist may incorrectly reject it using the correct market proxy.

Section 2 defines the informational distance between an empiricist and investors in the presence of dispersed information. In equilibrium this distance cannot be simply assumed, but arises endogenously in a way we describe in Section 3. Section 4 presents our main result, provides intuition into the distortion in beta estimates, and isolates the role of dispersed information. Section 5 tests our theory, and reinterprets "betting against beta." Finally, Section 6 discusses and relaxes our modeling assumptions, and Section 7 concludes.

[6] These explanations include leverage constraints (Black, 1972; Frazzini and Pedersen, 2014), inflation (Cohen, Polk, and Vuolteenaho, 2005), short-sale constraints and disagreement (Hong and Sraer, 2016), preference for volatile, skewed returns (Kumar, 2009; Bali, Cakici, and Whitelaw, 2011), market sentiment (Antoniou, Doukas, and Subrahmanyam, 2015), stochastic volatility (Campbell, Giglio, Polk, and Turley, 2012), and benchmarking of institutional investors (Baker, Bradley, and Wurgler, 2011; Buffa, Vayanos, and Woolley, 2014).

2 Background

Let $\widetilde{\mathbf{R}}^e$ be the vector of future excess return on an arbitrary cross section of N assets.[7] It does not matter here how these excess returns are computed (e.g., simple returns, log returns, or dollar returns). Suppose a large population of investors observes information about $\widetilde{\mathbf{R}}^e$, or more specifically about "factors" that partially determine $\widetilde{\mathbf{R}}^e$. Importantly, assume that this information is dispersed among investors, meaning that each investor i in the population observes a different information set. Based on her own information set, each investor i forms her own view $\mathbb{E}^i[\widetilde{\mathbf{R}}^e]$ regarding $\widetilde{\mathbf{R}}^e$. We further denote by $\overline{\mathbb{E}}[\widetilde{\mathbf{R}}^e]$ the average expectation across investors, which represents consensus beliefs about excess returns on these assets. To make our point in a simple way, we assume that $\widetilde{\mathbf{R}}^e$ and each investor's information are jointly normally distributed.[8]

Suppose an empiricist observes a dataset containing infinitely many realizations of $\widetilde{\mathbf{R}}^e$. Using the law of total variance, the unconditional variance of $\widetilde{\mathbf{R}}^e$ the empiricist computes can be decomposed into *three* terms:

$$\mathrm{Var}[\widetilde{\mathbf{R}}^e] = \mathrm{Var}^i[\widetilde{\mathbf{R}}^e] + \mathrm{Var}[\overline{\mathbb{E}}[\widetilde{\mathbf{R}}^e]] + \mathrm{Var}[\mathbb{E}^i[\widetilde{\mathbf{R}}^e] - \overline{\mathbb{E}}[\widetilde{\mathbf{R}}^e]]. \tag{1}$$

Our assumption of a Gaussian information structure makes this decomposition particularly simple. First, deviations from consensus beliefs do not covary with consensus beliefs themselves, $\mathrm{Cov}[\mathbb{E}^i[\widetilde{\mathbf{R}}^e] - \overline{\mathbb{E}}[\widetilde{\mathbf{R}}^e], \overline{\mathbb{E}}[\widetilde{\mathbf{R}}^e]] = \mathbf{0}$, and thus this relation only has three terms. This simplification follows from the law of large numbers—because investors are infinitesimally small individual noise washes out in the aggregate. Second, the conditional variance, $\mathrm{Var}^i[\widetilde{\mathbf{R}}^e]$, is nonrandom. We further assume that this variance is homogeneous in the population. A consequence of this assumption is that the last term in Eq. (1) is identical for any investor i, and thus represents the variance of expected returns computed *across the population of investors.*

The first term in Eq. (1) captures investor i's perception of uncertainty about $\widetilde{\mathbf{R}}^e$; the second term measures aggregate variation in *consensus* beliefs; and the third term measures the *dispersion* in beliefs across investors. We know that the law of total variance in Eq. (1) always applies to individual, rational beliefs. When information is homogeneous across investors, individual beliefs are identical to consensus beliefs and the last term in Eq. (1) drops out. But when information is dispersed, the law of total variance must incorporate

[7] Throughout the paper, we will adopt the following notation: we identify random variables with a tilde; we use bold letters to indicate vectors and matrices, and letters in plain font to indicate univariate variables; we use subscripts to indicate individual assets, and superscripts to indicate individual investors (agents). A complete list of symbols is available in Appendix D.

[8] This assumption holds in the main analysis; whenever relevant we explain how it matters for derivations.

variation across individual beliefs, which consensus beliefs average out. The relevance of variation across investors is perhaps unexpected, considering that the empiricist is assumed to observe time variation only. However, what the right-hand side of Eq. (1) shows is that cross-sectional variation *hides* in the variation the empiricist measures.

Without information there can be no variation nor dispersion in beliefs so that the empiricist and investors have identical perceptions of uncertainty, $\mathrm{Var}[\widetilde{\mathbf{R}}^e] = \mathrm{Var}^i[\widetilde{\mathbf{R}}^e]$. Perceptions may only differ if information creates a distance between the empiricist and investors. How do we measure this informational distance—how do we aggregate the matrix relation in Eq. (1) into a single number? In the theoretical framework that we construct in this paper, the market portfolio plays exactly this role. Formally, denote by $\mathbf{M}$ the market portfolio for the cross section of assets. Applying Eq. (1) to this portfolio, dividing it by $\mathrm{Var}[\widetilde{R}^e_{\mathbf{M}}]$ and shifting the first term to the left-hand side yields:

$$\underbrace{1 - \mathrm{Var}^i[\widetilde{R}^e_{\mathbf{M}}]/\,\mathrm{Var}[\widetilde{R}^e_{\mathbf{M}}]}_{\text{Informational distance}} = \underbrace{\mathrm{Var}[\overline{\mathbb{E}}[\widetilde{R}^e_{\mathbf{M}}]]/\,\mathrm{Var}[\widetilde{R}^e_{\mathbf{M}}]}_{\equiv \mathcal{C}^2} + \underbrace{\mathrm{Var}[\mathbb{E}^i[\widetilde{R}^e_{\mathbf{M}}] - \overline{\mathbb{E}}[\widetilde{R}^e_{\mathbf{M}}]]/\,\mathrm{Var}[\widetilde{R}^e_{\mathbf{M}}]}_{\equiv \mathcal{D}^2}, \tag{2}$$

where $\widetilde{R}^e_{\mathbf{M}} \equiv \mathbf{M}'\widetilde{\mathbf{R}}^e$ denotes the excess return on the market portfolio.

Eq. (2) defines the informational distance between investors and the empiricist, and shows that it has two origins. The left-hand side represents the fraction of variation in market excess returns explained by investor i's information—the informational gap between the empiricist and investor i. Aggregate variation in consensus beliefs contributes $\mathcal{C}^2$ to this informational gap. That is, $\mathcal{C}^2$ is the fraction of variation in market excess returns explained by variation in consensus beliefs. Furthermore, because information is dispersed there is variation in beliefs across investors, which accounts for the remaining distance, $\mathcal{D}^2$, i.e., the fraction of variation in market excess returns explained by variation in beliefs across agents. Hence, not only do realized returns vary because of time-varying consensus expected returns, they also vary because of cross-sectional variation in expected returns.

The purpose of this paper is to understand how the two fractions $\mathcal{C}^2$ and $\mathcal{D}^2$ together and separately distort the empiricist's view of the CAPM relation. Distortions resulting from time variation in consensus beliefs, $\mathcal{C}^2$, have been examined extensively in the literature (e.g., Jagannathan and Wang, 1996; Lewellen and Nagel, 2006; Boguth, Carlson, Fisher, and Simutin, 2011). The focus of our paper is mainly on $\mathcal{D}^2$, an estimate of which is missing in the literature. We will show that $\mathcal{D}^2$ is not a typical measure of dispersion in beliefs (e.g., Diether, Malloy, and Scherbina, 2002): it measures how expected returns on each *pair* of stocks *covaries* across investors, for a large cross section of assets. This source of variation in expected returns, to our knowledge, has been neglected in CAPM tests.

3 Model

We build a model of how investors form expectations, imposing an equilibrium structure on excess returns. Consider a one-period economy in which the market consists of one risk-free asset with gross return normalized to 1 and N risky stocks indexed by $n = 1, ..., N$. Suppose the risky stocks have random payoffs, $\widetilde{\mathbf{D}} \equiv [\widetilde{D}_1 \ \dots \ \widetilde{D}_N]'$, realized at the liquidation date (time 1). These payoffs are unobservable at the trading date (time 0) and have a common factor structure:

$$\widetilde{\mathbf{D}} = \begin{bmatrix} D \\ D \\ \vdots \\ D \end{bmatrix} + \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_N \end{bmatrix} \widetilde{F} + \begin{bmatrix} \tilde{\epsilon}_1 \\ \tilde{\epsilon}_2 \\ \vdots \\ \tilde{\epsilon}_N \end{bmatrix} = \mathbf{1}D + \mathbf{\Phi}\widetilde{F} + \tilde{\boldsymbol{\epsilon}}, \tag{3}$$

where D is a strictly positive scalar and $\mathbf{1}$ is a vector of ones of dimension N. The common factor, $\widetilde{F}$, and each stock-specific component, $\tilde{\epsilon}_n$, are independently normally distributed with means zero and precisions, τ_F and τ_ϵ. We assume, without loss of generality, that the vector of loadings of assets' payoffs on the common factor is a unit vector, $\|\mathbf{\Phi}\| = 1$; this normalization is equivalent to scaling τ_F. We further define the mean of this vector, $\bar{\Phi} \equiv \mathbf{1}'\mathbf{\Phi}/N$.

The economy is populated with a continuum of investors indexed by $i \in [0, 1]$, who choose their portfolios at time 0 and derive utility from terminal wealth with constant absolute risk aversion coefficient, γ. Investors know the structure of realized payoffs in Eq. (3), but do not observe the common factor, $\widetilde{F}$. Each investor i forms expectations about $\widetilde{F}$ based on information inferred from prices and information from a private signal, $\widetilde{V}^i$:

$$\widetilde{V}^i = \widetilde{F} + \tilde{v}^i. \tag{4}$$

Signal noises, $\tilde{v}^i$, are unbiased and independently normally distributed with precision τ_v.

In this economy equilibrium prices do not fully reveal investors' private information about the common factor, $\widetilde{F}$. Prices change to reflect new information about final payoffs, but they also change for reasons unrelated to information, e.g., endowments shocks, preference shocks, or private investment opportunities. To model uninformative price changes, we assume that an unmodeled group of agents trade for liquidity needs and/or for non-informational reasons. Liquidity trading prevents prices from revealing $\widetilde{F}$ (Grossman and Stiglitz, 1980), and prevents investors from refusing to trade (Milgrom and Stokey, 1982).[9]

We fix the total number of shares for all assets to $\mathbf{M}$ (hereafter the *market portfolio*), a vector whose elements are all equal to $1/N$. Liquidity traders have inelastic demands of $\widetilde{\mathbf{m}}$ shares, where $\widetilde{\mathbf{m}}$ is normally and independently distributed across stocks with precision τ_m; the remainder, $\mathbf{M} - \widetilde{\mathbf{m}}$, is available for trade to informed investors. This assumption is consistent with the usual noise trading story commonly adopted in the literature (e.g., He and Wang, 1995). Formally, letting $\mathbf{0}$ be a vector of zeros of dimension N and $\mathbf{I}$ denote the identity matrix of dimension N, $\mathbf{M} \equiv [1/N \ \ldots \ 1/N]'$ and $\widetilde{\mathbf{m}} \sim \mathcal{N}(\mathbf{0}, \tau_m^{-1}\mathbf{I})$.

This economy relies on several simplifying assumptions. We have assumed that payoffs in Eq. (3) are driven by a single factor, as opposed to multiple factors; that stocks only differ according to their loading, $\mathbf{\Phi}$, on this common factor; and that stocks have equal weights in the market portfolio, $\mathbf{M}$, and equal precisions across assets for the supply shocks and the idiosyncratic noises. These simplifications serve our purpose of isolating the main result in the clearest and simplest possible terms. We further discuss the generality of our results in Section 6, where we relax some of these assumptions.

We solve for a linear equilibrium of the economy in which prices satisfy:

$$\widetilde{\mathbf{P}} = \mathbf{1}D + \boldsymbol{\xi}_0\mathbf{M} + \boldsymbol{\lambda}\widetilde{F} + \boldsymbol{\xi}\widetilde{\mathbf{m}}, \tag{5}$$

where $\boldsymbol{\lambda}$ is an $N-$dimensional vector and $\boldsymbol{\xi}_0$ and $\boldsymbol{\xi}$ are a $N \times N$ matrices, all of which are determined in equilibrium by imposing market clearing. Because in this framework rates of return are not normally distributed, a convention in the literature is to work with dollar excess returns instead (e.g., Dybvig and Ross, 1985). We follow this convention and refer to $\widetilde{\mathbf{R}}^e \equiv \widetilde{\mathbf{D}} - \widetilde{\mathbf{P}}$, as *excess returns*.[10]

Each investor i forms expectations about excess returns based on her information set:

$$\mathscr{F}^i = \{\widetilde{V}^i, \widetilde{\mathbf{P}}\}. \tag{6}$$

Because private signals, $\widetilde{V}^i$, all have identical precision, and prices, $\widetilde{\mathbf{P}}$, are public, each investor forecasts the common factor, $\widetilde{F}$, with identical precision:

$$\tau \equiv \mathrm{Var}[\widetilde{F}|\mathscr{F}^i]^{-1} = \tau_F + \tau_v + \tau_P. \tag{7}$$

[9]There are different ways to endogenize liquidity trading: private investment opportunities (Wang, 1994), investor specific endowment shocks, or income shocks (Farboodi and Veldkamp, 2017). These alternatives would unnecessarily complicate the analysis, without bringing additional economic insights.

[10]Using rates of returns as opposed to dollar excess returns does not change our main results qualitatively. See Appendix section B.4.1 for a discussion, and also Banerjee (2010) for a similar argument.

The last coefficient in Eq. (7), τ_P, is the sum of squared signal-to-noise ratios over all prices; it is a scalar (to be determined in equilibrium) that measures price informativeness.

Defining $\mathbb{E}^i[\widetilde{\mathbf{R}}^e] \equiv \mathbb{E}[\widetilde{\mathbf{R}}^e|\mathscr{F}^i]$ as investor i's vector of expected returns and $\mathbf{\Sigma} \equiv \text{Var}[\widetilde{\mathbf{R}}^e|\mathscr{F}^i]$ as her conditional covariance matrix of returns (which in this model is constant and identical across investors), investor i's optimal portfolio choice is

$$\mathbf{w}^i = \frac{1}{\gamma}\mathbf{\Sigma}^{-1}\,\mathbb{E}^i[\widetilde{\mathbf{R}}^e], \quad \text{where } \mathbf{\Sigma} = \frac{1}{\tau}\mathbf{\Phi\Phi'} + \frac{1}{\tau_\epsilon}\mathbf{I}. \tag{8}$$

Investors in this model are mean-variance maximizers: *"They've all got their copies of Markowitz and they're doing what he says"* (William F. Sharpe, quoted from Burton, 1998). The key (and single) departure from the standard CAPM framework, however, is that each investor operates under a different information set. In particular, since each investor i forms *conditional views* on future asset returns based her own information $\mathscr{F}^i$, this information determines the parameters of the conditional mean-variance set, $\mathbb{E}^i[\widetilde{\mathbf{R}}^e]$ and $\mathbf{\Sigma}$, that are the inputs of the standard Markowitz (1952) recipe.

Market clearing requires that the demand in Eq. (8) aggregated across informed investors and the demand of liquidity traders sum up to the market portfolio, $\mathbf{M}$:

$$\int_0^1 \mathbf{w}^i \text{d}i + \widetilde{\mathbf{m}} = \mathbf{M}. \tag{9}$$

Let *consensus beliefs* of investors be $\overline{\mathbb{E}}[\widetilde{\mathbf{R}}^e] \equiv \int_0^1 \mathbb{E}^i[\widetilde{\mathbf{R}}^e]di$. Substituting individual portfolios in Eq. (8) into the market-clearing condition implies

$$\overline{\mathbb{E}}[\widetilde{\mathbf{R}}^e] = \gamma\mathbf{\Sigma}(\mathbf{M} - \widetilde{\mathbf{m}}), \tag{10}$$

which represents the expected rate of return that every particular asset must pay for investors to be willing to hold the supplies of the N assets, net of liquidity traders' demand.

The central departure from the traditional CAPM framework is that individual investors do not find it optimal conditionally to hold the market portfolio, $\mathbf{M}$. Each investor instead uses the market portfolio as a starting point, departing from it according to her own views:

$$\mathbf{w}^i = \mathbf{M} + \frac{1}{\gamma}\mathbf{\Sigma}^{-1}\underbrace{\left(\mathbb{E}^i[\widetilde{\mathbf{R}}^e] - \mathbb{E}[\widetilde{\mathbf{R}}^e]\right)}_{\text{Investor } i\text{'s private views}}, \tag{11}$$

a decomposition that follows from Eq. (8) and the unconditional version of Eq. (10). This method of constructing portfolios, which combines a natural starting point (the market portfolio, $\mathbf{M}$) with investors' private views, is reminiscent of the portfolio construction approach advocated by Black and Litterman (1990, 1992). Eq. (11) also implies that the investor who has *average unconditional beliefs*, $\mathbb{E}[\widetilde{\mathbf{R}}^e]$ and $\mathbf{\Sigma}$, holds the market portfolio, $\mathbf{M}$.

Since our framework is not a standard CAPM environment—investors have information that is dispersed among them (Lintner, 1969) and that the empiricist does not observe (Hansen and Richard, 1987)—how do we define a CAPM relation in this context? There is CAPM pricing in the sense that betas are conditional on information that investors know. This result goes back to Admati (1985), and is the sense in which the literature agrees the CAPM holds in this context (e.g., Easley and O'Hara, 2004; Fama and French, 2007; Van Nieuwerburgh and Veldkamp, 2010; Banerjee, 2010; Biais et al., 2010).[11] Formally, conditioning down Eq. (10) expected returns are proportional to the market risk premium and to a new notion of beta, $\boldsymbol{\beta}$, that is based on investors' covariance matrix, $\mathbf{\Sigma}$.

Definition 1. *In this economy, the following linear relation holds:*

$$\mathbb{E}[\widetilde{\mathbf{R}}^e] = \frac{\mathbf{\Sigma M}}{\sigma^2_{\mathbf{M}}}\,\mathbb{E}[\widetilde{R}^e_{\mathbf{M}}] = \boldsymbol{\beta}\,\mathbb{E}[\widetilde{R}^e_{\mathbf{M}}], \tag{12}$$

where $\mathbb{E}[\widetilde{\mathbf{R}}^e] \equiv \mathbb{E}[\overline{\mathbb{E}}[\widetilde{\mathbf{R}}^e]]$ *is a vector of unconditional expected returns on all stocks,* $\mathbb{E}[\widetilde{R}^e_{\mathbf{M}}] \equiv \mathbf{M}'\,\mathbb{E}[\widetilde{\mathbf{R}}^e]$ *is the unconditional expected excess return on the market portfolio, and*

$$\sigma^2_{\mathbf{M}} \equiv \mathbf{M}'\mathbf{\Sigma M} = \frac{\bar{\Phi}^2}{\tau} + \frac{1}{N\tau_\epsilon} \tag{13}$$

is the variance of excess returns on the market portfolio conditional on the information set of any investor $i \in [0,1]$.

The crux of our argument is that the notion of beta, $\boldsymbol{\beta}$, that underlies the CAPM representation in Eq. (12) is not that of an unconditional beta as empiricists commonly compute. Betas in Eq. (12) are computed using investors' covariance matrix, $\mathbf{\Sigma}$, as opposed to the covariance matrix of *realized* returns used in standard empirical tests:

$$\widehat{\beta}_n = \frac{\mathrm{Cov}[\widetilde{R}^e_n, \widetilde{R}^e_{\mathbf{M}}]}{\mathrm{Var}[\widetilde{R}^e_{\mathbf{M}}]}, \tag{14}$$

i.e., the slope of a regression of realized excess returns of asset n on realized excess returns of the market portfolio, $\mathbf{M}$. Therefore, from investors' perspective $\boldsymbol{\beta}$ is the vector of *true betas*

[11] Our definition is identical to that in the literature, except that we focus on the unconditional CAPM, whereas the literature focuses on its conditional version. We can write a conditional CAPM in our model, too, but this relation is not testable, and thus not our focus; furthermore, this conditional CAPM does not hold with respect to the (observable) market portfolio $\mathbf{M}$; it only holds with respect to informed investors' holdings in this portfolio, $\mathbf{M} - \widetilde{\mathbf{m}}$, which are unobservable.

and Eq. (12) is the *true CAPM*. We now elaborate on the origin of the difference between true betas and empiricist's betas.

We take the perspective of an investor i and write how she perceives the cross-section of excess returns conditional on her own information. The following decomposition results directly from Eq. (11) and Definition 1:

$$\mathbb{E}^i[\widetilde{\mathbf{R}}^e] = \boldsymbol{\beta}\,\mathbb{E}[\widetilde{\mathbf{R}}^e_{\mathbf{M}}] + \boldsymbol{\varepsilon}^i, \quad \text{where } \boldsymbol{\varepsilon}^i \equiv \gamma\boldsymbol{\Sigma}(\mathbf{w}^i - \mathbf{M}) \sim \mathcal{N}(\mathbf{0}, \mathrm{Var}[\mathbb{E}^i[\widetilde{\mathbf{R}}^e]]). \tag{15}$$

An investor i views expected returns as a noisy perturbation around the CAPM relation of Definition 1. This perturbation arises because returns are predictable from her perspective (Ferson and Harvey, 1991; Pesaran and Timmermann, 1995; Cochrane, 2007). The more predictable returns are, $\mathrm{Var}[\mathbb{E}^i[\widetilde{\mathbf{R}}^e]] >> \mathbf{0}$, the larger this perturbation is. However, even though there may be substantial predictability at the investor level, the operation of the law of iterated expectations ensures that this perturbation vanishes on average. Thus, after conditioning down Eq. (15) investor i retrieves the CAPM relation of Definition 1. Yet, the perturbation *leaves a mark* on the CAPM relation that the empiricist estimates: the typical betas in Eq. (14)—computed using the covariance matrix of realized returns—do depend on the extent to which returns are predictable (as we will demonstrate formally in Section 4).

We conclude this section by providing an explicit solution to the functional form of equilibrium prices.

Proposition 1. ***(Equilibrium)*** *There exists a unique linear equilibrium in which prices take the linear form in Eq. (5) and are explicitly given by*

$$\widetilde{\mathbf{P}} = \mathbf{1}D - \gamma\left(\frac{1}{\tau}\boldsymbol{\Phi}\boldsymbol{\Phi}' + \frac{1}{\tau_\epsilon}\mathbf{I}\right)\mathbf{M} + \boldsymbol{\Phi}\frac{\tau - \tau_F}{\tau}\widetilde{F} + \left(\frac{\gamma + \sqrt{\tau_m \tau_P}}{\tau}\boldsymbol{\Phi}\boldsymbol{\Phi}' + \frac{\gamma}{\tau_\epsilon}\mathbf{I}\right)\widetilde{\mathbf{m}}, \tag{16}$$

and the scalar τ_P is the unique positive root of the cubic equation:

$$\tau_P(\tau_F + \tau_v + \tau_P + \tau_\epsilon)^2\gamma^2 = \tau_m \tau_\epsilon^2 \tau_v^2. \tag{17}$$

Proposition 1 delivers a simple expression for the true betas, $\boldsymbol{\beta}$, implied by the model.

Corollary 1.1. *The vector of true betas, $\boldsymbol{\beta}$, in this economy is*

$$\boldsymbol{\beta} = \mathbf{1} + \frac{\bar{\Phi}^2}{\tau\sigma^2_{\mathbf{M}}}\left(\frac{\boldsymbol{\Phi}}{\bar{\Phi}} - \mathbf{1}\right). \tag{18}$$

Corollary 1.1 reveals two aspects about true betas in the model. First, cross-sectional heterogeneity in betas results entirely from heterogeneity in exposures, $\boldsymbol{\Phi}$, to the common factor; this implies that, absent this heterogeneity all betas equal unity, and that assets with a stronger exposure to the common factor have larger betas. Second, true betas in Eq. (18) are a weighted average of two vectors, $\mathbf{1}$ and $\mathbf{\Phi}/\bar{\Phi}$. In other words, the whole cross-section of stocks is spanned by just two vectors in equilibrium, an observation that will prove useful for the purpose of analyzing the empiricist's view of this economy.

4 The empiricist's view

We now examine how an empiricist views this economy. We illustrate the difference between the empiricist's view and that of the average investor in the mean-standard deviation diagram; we refer to the *"average investor"* as the investor who holds average unconditional beliefs, $\mathbb{E}[\widetilde{\mathbf{R}}^e]$ and $\mathbf{\Sigma}$. In the analysis that follows, when we say a portfolio is *"minimum-variance,"* we mean that the portfolio is located on the hyperbola of minimum-variance portfolios; when we say a portfolio is *"mean-variance efficient,"* we mean that the portfolio is located on the efficient frontier—the line that goes through 0 and the tangency portfolio. To distinguish the view of the average investor from that of the empiricist, we denote all variables as measured by the empiricist with a hat. Finally, to ease notation, we define $\boldsymbol{\mu} \equiv \mathbb{E}[\widetilde{\mathbf{R}}^e]$ as the vector of unconditional expected excess returns on all assets and $\mu_{\mathbf{M}} \equiv \mathbf{M}' \, \mathbb{E}[\widetilde{\mathbf{R}}^e]$ as the unconditional expected excess return on the market portfolio.

A synonym for Definition 1 is that the market portfolio, $\mathbf{M}$, is mean-variance efficient for the average investor; it commands the highest Sharpe ratio in the economy:

$$\frac{\mu_{\mathbf{M}}}{\sigma_{\mathbf{M}}} = \sqrt{\boldsymbol{\mu}'\mathbf{\Sigma}^{-1}\boldsymbol{\mu}}. \tag{19}$$

This result follows directly from efficient-set mathematics (see, e.g., Roll, 1977).

The main difference between the empiricist and investors is that the empiricist measures the economy *ex post*: the empiricist observes realized returns, but does not observe investors' conditional expected returns. As a result, her information is coarser than that of any investor i in the economy. The law of iterated expectations ensures that the empiricist correctly measures $\boldsymbol{\mu}$ and $\mu_{\mathbf{M}}$. The law of total variance, however, implies that the covariance matrix of excess returns of the empiricist, $\widehat{\mathbf{\Sigma}} \equiv \mathrm{Var}[\widetilde{\mathbf{R}}^e]$, differs from $\mathbf{\Sigma}$:

$$\widehat{\mathbf{\Sigma}} = \mathbf{\Sigma} + \mathrm{Var}[\mathbb{E}^i[\widetilde{\mathbf{R}}^e]], \quad \forall i \in [0,1]. \tag{20}$$

This is precisely where return predictability, $\mathrm{Var}[\mathbb{E}^i[\widetilde{\mathbf{R}}^e]]$, leaves a mark on empiricist's beta.

This relation provides the basis on which we will define the informational distance between the empiricist and any individual investor i.[12] Note that, because modeling assumptions correspond to those of Section 2, $\mathbf{\Sigma} = \mathrm{Var}^i[\widetilde{\mathbf{R}}^e]$ is known and does not vary across investors because in our economy all investors have the same precision of information; hence the same relation holds for all $i \in [0, 1]$. The consequence of Eq. (20), and our first result, is that the empiricist rejects the unconditional CAPM.

Proposition 2. ***(CAPM rejection)*** *For the empiricist, who observes the economy ex post, the market portfolio,* $\mathbf{M}$*, is not mean-variance efficient:*

$$\frac{\mu_{\mathbf{M}}}{\widehat{\sigma}_{\mathbf{M}}} < \sqrt{\boldsymbol{\mu}\widehat{\mathbf{\Sigma}}^{-1}\boldsymbol{\mu}}, \tag{21}$$

where $\widehat{\sigma}_{\mathbf{M}} \equiv \sqrt{\mathbf{M}'\widehat{\mathbf{\Sigma}}\mathbf{M}}$ *is the volatility of realized excess returns on the market.*

Although $\mathbf{M}$ is the correct market portfolio, the empiricist rejects the CAPM. Figure 1 contrasts the view of the empiricist with that of the average investor. The solid line depicts the set of minimum-variance portfolios under average unconditional beliefs. For the average investor, the portfolio $\mathbf{M}$ is the tangency portfolio ($\mathbf{T}$), i.e., the unconditional CAPM holds under average unconditional beliefs, which is the meaning of Definition 1. For the empiricist, all assets have the correct unconditional expected returns, but display systematically larger unconditional variances (due to variation in expected returns, which the empiricist does not observe). Hence, all assets—including the market—move to the right in the mean-standard deviation space, and the minimum-variance set of the empiricist (the dashed line) moves inside that of the average investor. The main result of Proposition 2 is that $\mathbf{M}$ does *not* have the highest Sharpe ratio for the empiricist. That is, for the empiricist $\mathbf{M}$ is not the tangency portfolio, $\widehat{\mathbf{T}}$, nor is it mean-variance efficient and she rejects the CAPM.

[INSERT FIGURE 1 HERE]

Note that the market portfolio for the empiricist, although not mean-variance efficient, remains minimum-variance. Two assumptions jointly lead to this outcome: (i) the market portfolio, $\mathbf{M}$, is equally weighted, and (ii) asset payoffs are driven by a single common factor, $\widetilde{F}$. Relaxing one (or both) of these assumptions would cause $\mathbf{M}$ to move inside the minimum-variance set of the empiricist, and possibly below $\widehat{\mathbf{T}}$. This would introduce further distorting effects on the empiricist's view of the CAPM, which we discuss in Section 6.

[12] In Appendix B.4.2, we reduce the informational distance between investors and the empiricist by allowing the empiricist to control for all publicly available information—i.e., public prices. Because the empiricist cannot possibly control for private information, the results that follow continue to hold and are actually stronger when investors' private information is sufficiently precise.

Based on the market portfolio, $\mathbf{M}$, the vector of betas the empiricist obtains is:

$$\widehat{\boldsymbol{\beta}} = \frac{\widehat{\boldsymbol{\Sigma}}\mathbf{M}}{\widehat{\sigma}^2_{\mathbf{M}}}, \tag{22}$$

which corresponds to the vector formulation of Eq. (14). Roll (1977, Corollary 6) shows that the betas of individual assets with respect to any portfolio are an exact linear function of individual expected excess returns *if and only if* this portfolio is minimum-variance. Since $\mathbf{M}$ is minimum-variance, the empiricist will necessarily observe a linear relation between expected excess returns and betas, $\widehat{\boldsymbol{\beta}}$. To characterize this linear relation, we first define the *zero-beta* portfolio, $\widehat{\mathbf{Z}}$, as the unique minimum-variance portfolio that is uncorrelated with $\mathbf{M}$. In the return-standard deviation space, the expected excess return of $\widehat{\mathbf{Z}}$ represents the intercept of the line that is tangent to the minimum-variance set at $\mathbf{M}$. Figure 1 shows that the empiricist's zero-beta portfolio, $\widehat{\mathbf{Z}}$, lies on the lower limb of the minimum-variance set.

Using the expected excess return of the zero-beta portfolio, Proposition 3 characterizes the linear relation between betas and expected returns that the empiricist measures.

Proposition 3. ***(CAPM tests based on realized returns)*** *The empiricist observes a zero-beta CAPM (Black, 1972):*

$$\boldsymbol{\mu} = \mathbf{1}\mu_{\widehat{\mathbf{Z}}} + \widehat{\boldsymbol{\beta}}(\mu_{\mathbf{M}} - \mu_{\widehat{\mathbf{Z}}}), \tag{23}$$

in which the empiricist's vector of betas, $\widehat{\boldsymbol{\beta}}$, satisfies the proportionality relation:

$$\widehat{\boldsymbol{\beta}} - \mathbf{1} = (1+\delta)(\boldsymbol{\beta} - \mathbf{1}), \tag{24}$$

where the coefficient δ measures the magnitude of the empiricist's distortion of the CAPM:

$$\delta = \frac{\mu_{\widehat{\mathbf{Z}}}}{\mu_{\mathbf{M}} - \mu_{\widehat{\mathbf{Z}}}} \geq 0. \tag{25}$$

The coefficient δ measures the *distortion* of the CAPM relation as estimated by the empiricist relative to the true CAPM in Eq. (12). The zero-beta portfolio has positive expected excess returns when the market portfolio, $\mathbf{M}$, is located above empiricist's tangency portfolio, $\widehat{\mathbf{T}}$; this is always the case in our model, and thus the empiricist perceives a flat Securities Market Line (SML) with a positive intercept.

Further replacing $\mu_{\widehat{\mathbf{Z}}}$ in Eq. (23) produces the following relation:

$$\boldsymbol{\mu} = \mathbf{1}\frac{\delta}{1+\delta}\mu_{\mathbf{M}} + \widehat{\boldsymbol{\beta}}\frac{1}{1+\delta}\mu_{\mathbf{M}}, \tag{26}$$

which describes the biggest failure of the CAPM (e.g., Black, Jensen, and Scholes, 1972, and the literature that followed)—the high returns enjoyed by many apparently low-beta assets and the high intercept of the SML. Figure 2 illustrates this failure. The proportionality relation between betas in Eq. (24) means that the empiricist inflates all betas above market's beta, which is 1, and deflates all others, as denoted with arrows in the plot. Hence, the empiricist perceives risky assets as riskier than they really are, and safe assets as safer than they really are. Even though the empiricist and the average investor disagree about betas (the horizontal axis), they agree on expected returns (the vertical axis) and, since they observe the market portfolio, agree on its location. Thus empiricist's SML rotates clockwise around the market portfolio, which flattens its slope and creates a positive intercept.

[INSERT FIGURE 2 HERE]

The main result is that, in equilibrium, true betas are shrunk towards one relative to empiricist's betas (shrinkage towards $\mathbf{1}$ is a consequence of our assumption that the market portfolio is equally weighted, $\mathbf{M} \equiv \mathbf{1}/N$). The degree of shrinkage is determined by a unique coefficient, δ, which adjusts the empiricist's betas towards true betas. Interestingly, Eq. (24) is identical to the Bayesian estimator proposed by Vasicek (1973), an estimator that is popular in the financial industry ("ADJ BETA" on Bloomberg terminals).[13] We emphasize, however, that the result of Theorem 3 is not due to sampling error. Nor is this result a standard attenuation bias, which commonly plagues the second pass cross-sectional regression in the Fama and MacBeth (1973) method.[14] Rather, in equilibrium shrinkage in betas arises as a consequence of the informational distance between investors and the empiricist.

How much flatter the SML is, and how large its intercept is, depends on the magnitude of the coefficient δ. Eq. (25) offers a way of determining a plausible magnitude for δ. For instance, if the expected excess return on the market is 6% per year and the expected excess return on the zero-beta portfolio is 3%, then δ is 1. We are interested, however, in measuring the coefficient δ in terms of the informational distance between investors and the empiricist, a matter to which we turn next.

[13] This linear adjustment was first proposed by Blume (1971) (due to mean reversion of betas over time) and then by Vasicek (1973) (due to measurement error). See Bodie, Kane, and Marcus (2007), Berk and DeMarzo (2007) among others. Levi and Welch (2017) give best-practice advice for beta-shrinkage.

[14] In Vasicek (1973), the degree of adjustment depends on the sample size and converges to zero as the sample size increases. Similarly, Shanken (1992) shows that the attenuation bias becomes negligible as the length of the sample period grows indefinitely (see also Jagannathan and Wang, 1998; Shanken and Zhou, 2007; Kan, Robotti, and Shanken, 2013). In our case, the adjustment is necessary even in infinite samples.

4.1 Decomposition of the CAPM distortion into aggregate and cross-sectional variation

Following Section 2, we decompose the law of total variance in Eq. (20) into time-series variation in *consensus* beliefs and *dispersion* in investors' beliefs. In our model, dispersion in investors' information implies that their expectations differ from consensus beliefs by the idiosyncratic noise in their private signals, which leads to the following decomposition of return predictability, $\mathrm{Var}[\mathbb{E}^i[\widetilde{\mathbf{R}}^e]]$:

$$\widehat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma} + \underbrace{\mathrm{Var}[\overline{\mathbb{E}}[\widetilde{\mathbf{R}}^e]]}_{\text{Consensus covariance matrix}} + \underbrace{\mathrm{Var}[\mathbb{E}^i[\widetilde{\mathbf{R}}^e] - \overline{\mathbb{E}}[\widetilde{\mathbf{R}}^e]]}_{\text{Matrix of co-beliefs}}. \tag{27}$$

This decomposition corresponds to Eq. (1).

Two sources of variation together lead to the empiricist's rejection of the CAPM. First, there is aggregate (time-series) variation in consensus expected returns, $\overline{\mathbb{E}}[\widetilde{\mathbf{R}}^e]$. The empiricist observes realized returns but does not observe $\overline{\mathbb{E}}[\widetilde{\mathbf{R}}^e]$, and this variation in expected returns creates a *consensus covariance matrix.* Second, dispersed information generates cross-sectional dispersion in expected returns across investors, $\mathbb{E}^i[\widetilde{\mathbf{R}}^e] - \overline{\mathbb{E}}[\widetilde{\mathbf{R}}^e]$. Because the empiricist does not observe investors' individual information, she does not observe cross-sectional dispersion in beliefs. This latter source of variation generates a *matrix of co-beliefs.*

Note that Banerjee (2010) refers to this matrix as *disagreement matrix.* But only the diagonal elements of this matrix, which measure the dispersion of expected returns for a given stock across investors, have the traditional meaning of disagreement. However, its off-diagonal elements reflect the extent to which deviations from consensus beliefs on one stock tend to covary with deviations on another, which is not the same as disagreement. We therefore propose the *"matrix of co-beliefs"* denomination to emphasize this distinction, which will become important in our empirical tests.

The following Lemma further characterizes the *consensus covariance matrix* and the *matrix of co-beliefs* using the equilibrium result of Proposition 1.

Lemma 1. *In equilibrium, the empiricist's covariance matrix of realized excess returns satisfies*

$$\widehat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma} + \frac{\gamma^2}{\tau_m}\left(\frac{1}{\tau_\epsilon}\boldsymbol{\Sigma} + \frac{e_1}{\tau}\boldsymbol{\Phi}\boldsymbol{\Phi}'\right) + \frac{\tau_v}{\tau^2}\boldsymbol{\Phi}\boldsymbol{\Phi}', \tag{28}$$

where e_1 is the unique largest eigenvalue of $\mathbf{\Sigma}$:

$$e_1 = \frac{1}{\tau} + \frac{1}{\tau_\epsilon} > 0. \tag{29}$$

Eq. (28) provides explicit expressions for the consensus covariance matrix (the second term on the right-hand side) and the matrix of co-beliefs (the last term). Remarkably, the informational distance between $\widehat{\mathbf{\Sigma}}$ and $\mathbf{\Sigma}$ is determined in equilibrium by a unique coefficient e_1. This coefficient represents the largest eigenvalue of $\mathbf{\Sigma}$ (the other eigenvalues, of multiplicity $N-1$, being $1/\tau_\epsilon$). This uniqueness result follows from our assumption of a single common factor in payoffs, which implies that $\mathbf{\Sigma}$ has two distinct eigenvalues. Due to this property, the relation between $\widehat{\boldsymbol{\beta}}$ and $\boldsymbol{\mu}$ is affine, and the market portfolio, $\mathbf{M}$, remains minimum-variance for the empiricist (see Figure 1). The same property implies that the CAPM distortion is summarized by a single number, δ.

Although the CAPM distortion δ depends on all elements of the consensus covariance and co-beliefs matrices, Lemma 1 allows us to summarize this dependence with just two numbers. These two numbers, $\mathcal{C}^2$ and $\mathcal{D}^2$, are the ones we introduced conceptually in Section 2. They represent value-weighted averages over each of these two matrices divided by $\widehat{\sigma}^2_{\mathbf{M}}$:

$$\mathcal{C}^2 \equiv \frac{\mathrm{Var}[\bar{\mathbb{E}}[\widetilde{R}^e_{\mathbf{M}}]]}{\widehat{\sigma}^2_{\mathbf{M}}} = \frac{\gamma^2}{\tau_m \tau_\epsilon \tau \widehat{\sigma}^2_{\mathbf{M}}}(\tau \sigma^2_{\mathbf{M}} + \tau_\epsilon e_1 \bar{\Phi}^2), \tag{30}$$

$$\mathcal{D}^2 \equiv \frac{\mathrm{Var}[\mathbb{E}^i[\widetilde{R}^e_{\mathbf{M}}] - \bar{\mathbb{E}}[\widetilde{R}^e_{\mathbf{M}}]]}{\widehat{\sigma}^2_{\mathbf{M}}} = \frac{\tau_v \bar{\Phi}^2}{\tau^2 \widehat{\sigma}^2_{\mathbf{M}}}. \tag{31}$$

These numbers are important for our empirical work: they will help us use analyst forecast data to map the magnitude of the informational gap into that of the CAPM distortion.

The number $\mathcal{C}^2$ represents the fraction of variation in market excess returns explained by variation in consensus beliefs; it can be interpreted as the informational distance between the empiricist and the average investor who holds consensus beliefs. The larger $\mathcal{C}^2$ is, the stronger is the variation in consensus beliefs. The other number $\mathcal{D}^2$ is the fraction of variation in market excess returns explained by variation in beliefs across investors; it measures the extent of dispersion in expectations across investors, and is a purely cross-sectional measure; the larger it is, the more dispersed investors' expectations are. Both numbers belong to the interval $[0, 1]$ and are close to 0.05 in our dataset (see Section 5). Together, $\mathcal{C}^2$ and $\mathcal{D}^2$ account for the informational distance between the empiricist and investors. We now map this distance into the coefficient of CAPM distortion, δ.

Proposition 4. *In equilibrium, the distortion in the CAPM relation to the empiricist is:*

$$\delta = \underbrace{\left(\frac{\tau\sigma_{\mathbf{M}}^2}{\bar{\Phi}^2} - 1\right)}_{\text{Excess volatility}} \left(\mathcal{C}^2 \underbrace{\frac{\tau_\epsilon e_1 \bar{\Phi}^2}{\tau\sigma_{\mathbf{M}}^2 + \tau_\epsilon e_1 \bar{\Phi}^2}}_{\in(0,1)} + \mathcal{D}^2\right) \geq 0. \tag{32}$$

Proposition 4 reveals that both $\mathcal{C}^2$ and $\mathcal{D}^2$ unambiguously increase beta distortion. The impact of aggregate variation in expected returns ($\mathcal{C}^2$ in this model) has been examined extensively in the literature (e.g., Jagannathan and Wang, 1996). Yet, to our knowledge, the impact of dispersion in beliefs has been neglected in CAPM tests; Eq. (32) shows that the impact of the latter is potentially stronger than that of the former (which is given a weight lower than one). We also note that the term $(\tau\sigma_{\mathbf{M}}^2/\bar{\Phi}^2 - 1)$ has a clear economic interpretation: it can be understood as the variation of returns of the market portfolio in excess of the variation of the fundamental.[15] In other words, it measures *excess volatility* in the market (Shiller, 1981). Thus, excess volatility is key in generating distortion in beta.

It is instructive to characterize the trivial cases in which the distortion is zero or in which all betas are equal to 1 (and the distortion has no bearing on betas); this exercise reveals the elements of the model that are necessary ingredients for our results. Table A1 in Appendix B.4 illustrates these cases. We describe here only the case of a vanishing idiosyncratic component in payoffs, $\tau_\epsilon \to \infty$. In this case, $\sigma_{\mathbf{M}}^2 = \bar{\Phi}^2/\tau$, and there can be no distortion in empiricist's SML. There are reasons to believe, however, that the term $(\tau\sigma_{\mathbf{M}}^2/\bar{\Phi}^2 - 1)$ is in fact much larger than zero; as argued above, this term measures excess volatility in the market. We find that a plausible value for $(\tau\sigma_{\mathbf{M}}^2/\bar{\Phi}^2 - 1)$ based on our dataset is as high as 25 (see Section 5.2). This along with other trivial cases from Table A1 reveal that idiosyncratic shocks $\tilde{\boldsymbol{\epsilon}}$, a non-zero risk aversion, imperfect private information, and incomplete revelation of information through public prices are necessary ingredients for our results.

A limiting case that provides particularly transparent intuition into some of our results is that of an infinitely large economy ($N \to \infty$). This case, however, makes little sense in an economy with finitely-many factors driving payoffs. Intuitively, when learning about finitely-many factors, an infinite cross section of stocks reveals these factors perfectly. We relegate the analysis of a large economy to Section 6.3, in which we let both the number of stocks and the number of factors grow to infinity but keeping their ratio finite. This limit is well-defined because the many endogenous stock signals are always accompanied with many factors to learn about. In this context, we characterize conditions under which our flattening result prevails in terms of the distribution of eigenvalues of the matrix of factor loadings; we also

[15] The precision of the fundamental from investors' perspective, $\tau/\bar{\Phi}^2$, is scaled by $\bar{\Phi}^2$ due to our initial assumption of normalizing the vector of loadings on asset payoffs, $\|\mathbf{\Phi}\| = 1$.

take advantage of this limiting case to clarify how the distortion caused by the aggregation of dispersed information magnifies CAPM distortion (see also Section 4.2 below).

In terms of comparative statics, we show in Appendix B.4 (for the case of diffuse priors) that the flattening coefficient, δ, increases with investors' risk aversion, γ, and with the amount of noise in assets' supplies, $1/\tau_m$. This result is perhaps not suprising if one considers that the ratio of the two, γ^2/τ_m, captures aggregate variation in expected returns; Section 6.1 shows that aggregate variation always has a flattening effect on the SML the empiricist observes. The additional effect of variation of dispersion of beliefs, $\mathcal{D}^2$, is central to our paper. In the next section, we cleanly isolate its impact on the CAPM distortion.

4.2 The role of dispersed information

In this section we investigate whether the *type* of information investors receive (public or private) matters for the distortion and if so, how. We compare our model to an otherwise identical economy in which all information is public. We call this economy the *Common Information Economy* (CIE).[16,17] We adopt the following notation: we add "CIE" as a subscript to variables whose value is different in the common information economy. All variables without a "CIE" subscript have identical value to that in the baseline model.

We want to ensure that investors' precisions on the common factor are identical across the two economies. In our baseline economy prices act as endogenous, public signals. Formally, all investors observe N endogenous, public signals with the following structure:

$$\boldsymbol{\xi}^{-1}\widetilde{\mathbf{P}}^a = \frac{\sqrt{\tau_P}}{\sqrt{\tau_m}}\boldsymbol{\Phi}\widetilde{F} + \widetilde{\mathbf{m}}, \tag{33}$$

where $\widetilde{\mathbf{P}}^a \equiv \widetilde{\mathbf{P}} - \mathbf{1}D - \boldsymbol{\xi}_0\mathbf{M}$. We assume that investors in the CIE observe a vector $\widetilde{\mathbf{G}}$ of N

[16] Bacchetta and Wincoop (2006) and Albagli et al. (2015) perform a similar exercise with the same goal of isolating the role of dispersed information. In the former reference, the exercise reveals that dispersed information disconnects the exchange rate from observed macroeconomic fundamentals. In the latter reference, the exercise reveals that dispersed information can result in over-reaction of the market price to public information, i.e., the information that is aggregated through the price.

[17] It is perhaps tempting to create a benchmark economy that features dispersion but no aggregate variation in beliefs ($\mathcal{D}^2 > 0$ and $\mathcal{C}^2 = 0$). However, in our framework this requires eliminating noise in supply (the only source of aggregate variation) and since prices would become fully informative, cross-sectional variation would only persist under additional, behavioral assumptions (e.g., investors "agree to disagree").

exogenous, public signals with same structure:[18]

$$\widetilde{\mathbf{G}} \equiv \frac{\sqrt{\tau_P}}{\sqrt{\tau_m}} \mathbf{\Phi} \widetilde{F} + \tilde{\mathbf{g}}, \quad \text{where } \tilde{\mathbf{g}} \sim \mathcal{N}(\mathbf{0}, \tau_g^{-1} \mathbf{I}). \tag{34}$$

To maintain identical informational content about the common factor, $\widetilde{F}$, in the CIE and in the baseline economy, we choose τ_g such that the precision τ is identical in both economies; the following equation determines τ_g (Appendix B.5 provides analytical details):

$$\tau_F + \tau_v + \tau_P = \tau_F + \frac{\tau_g}{\tau_m} \tau_P. \tag{35}$$

Clearly, for τ to be identical across the two economies, τ_g must be larger than τ_m, meaning that public information must be more informative in the CIE than prices are in the baseline model. Intuitively, investors in the baseline model observe information from private signals, but investors in the CIE do not. Thus for identical posterior precision τ to obtain, the precision on the exogenous, public information in the CIE, $\widetilde{\mathbf{G}}$, must be higher than that on endogenous, public information in the baseline model, $\boldsymbol{\xi}^{-1} \widetilde{\mathbf{P}}^a$.

An important consequence of this construction is that investors' conditional covariance matrix of future expected returns, $\mathbf{\Sigma}$, is identical in the CIE and in the baseline model; this immediately follows from the fact that τ is identical in both economies. As a result, the vector of true betas, $\boldsymbol{\beta}$, is identical to that in the baseline model, and thus investors in the CIE also observe the unconditional CAPM relation in Definition 1. But the view of the empiricist *does* change—the empiricist observes a stronger CAPM relation in the CIE.

Proposition 5. *The distortion of the SML is lower in the common information economy:*

$$\delta_{\text{CIE}} = \left(\frac{\tau \sigma_{\mathbf{M}}^2}{\bar{\Phi}^2} - 1 \right) \mathcal{C}_{\text{CIE}}^2 \frac{\tau_\epsilon e_1 \bar{\Phi}^2}{\tau \sigma_{\mathbf{M}}^2 + \tau_\epsilon e_1 \bar{\Phi}^2} < \delta. \tag{36}$$

Our requirement that precision τ be identical across economies implies that the informational gap resulting from *aggregate* variation, $\mathcal{C}_{\text{CIE}}^2 > \mathcal{C}^2$, is always larger in the CIE. Since there is no dispersion in investors' expectations, $\mathcal{D}_{\text{CIE}}^2 = 0$, it is possible that the gap between $\mathcal{C}^2$ and $\mathcal{C}_{\text{CIE}}^2$ is exactly accounted for by dispersion in beliefs, $\mathcal{D}^2$. This is, however, not the case due to a distortion of the kind studied in Albagli et al. (2015). In particular, comparing

[18] Alternatively, one can consider an economy with a unique public signal on $\widetilde{F}$, but the specification with N public signals keeps the structure as close as possible to the main model. N Gaussian signals can always be aggregated into one single signal, with identical results.

the structure of prices in the CIE and in the baseline model:

$$\widetilde{\mathbf{P}}_{\text{CIE}} = \mathbf{1}D + \frac{\sqrt{\tau_m \tau_P}}{\tau}\mathbf{\Phi}\mathbf{\Phi}'\widetilde{\mathbf{g}} \ \ - \gamma\mathbf{\Sigma}(\mathbf{M} - \widetilde{\mathbf{m}}) + \frac{\tau_P}{\tau}\mathbf{\Phi}\widetilde{F} \tag{37}$$

$$\widetilde{\mathbf{P}} = \mathbf{1}D + \frac{\sqrt{\tau_m \tau_P}}{\tau}\mathbf{\Phi}\mathbf{\Phi}'\widetilde{\mathbf{m}} - \gamma\mathbf{\Sigma}(\mathbf{M} - \widetilde{\mathbf{m}}) + \frac{\tau_P + \tau_v}{\tau}\mathbf{\Phi}\widetilde{F}, \tag{38}$$

shows that the main difference is in the sensitivity of prices to the fundamental (the last term). The precision of private information, τ_v, strengthens this sensitivity in the baseline model, a phenomenon Albagli et al. (2015) interpret as a distortion in how the market aggregates information. Although both economies have identical informational content, investors in the baseline model treat market information as more informative than it truly is, exacerbating price sensitivity to the fundamental. This over-reaction distorts the informational gap in the baseline economy. Due to this distortion, δ is always lower in the CIE.

Another way of understanding the result of Proposition 5 is that empiricist's betas, $\widehat{\boldsymbol{\beta}}$, is a weighted average between true betas $\boldsymbol{\beta}$ and $\mathbf{\Phi}/\bar{\Phi}$, with positive weights. The weight on $\boldsymbol{\beta}$ in the CIE is always larger than that in the baseline model, bringing $\widehat{\boldsymbol{\beta}}_{\text{CIE}}$ closer to $\boldsymbol{\beta}$. In the words of the previous paragraph, over-reaction due to dispersed information increases empiricist's distortion in beta estimates by moving $\widehat{\boldsymbol{\beta}}$ closer to $\mathbf{\Phi}/\bar{\Phi}$ and further away from $\boldsymbol{\beta}$. Thus the empiricist observes a stronger CAPM in the CIE.

Importantly, a direct implication of Proposition 5 is that empiricist's betas are more *compressed* in the CIE. Using Proposition 3 we can link the cross-sectional dispersion in $\widehat{\boldsymbol{\beta}}$ to that in $\boldsymbol{\beta}$ in the baseline model:

$$\sigma_{\widehat{\boldsymbol{\beta}}} = (1 + \delta)\sigma_{\boldsymbol{\beta}}. \tag{39}$$

Since $\sigma_{\boldsymbol{\beta}}$ is identical in the CIE (true betas remain unchanged), for the empiricist a decrease in δ translates into a *beta compression*. In particular, in our model a stronger CAPM in the CIE arises solely from this compression, since the unconditional market risk premium remains the same across economies (investors face, by construction, the same risk). The immediate theoretical prediction is that days on which public information dominates (e.g., important public announcement days, such as FOMC meetings) should be accompanied with a strong beta compression. Recent empirical work provides evidence supporting this result. Bodilsen (2019) and Eriksen and Gronborg (2020) find that betas on low-beta assets rise significantly and betas on high-beta assets drop significantly on FOMC announcement days. Similarly, Andersen et al. (2020) document a sharp compression in betas on these days.

5 Empirical Tests

We have argued theoretically that dispersion in expectations is an important source of variation in returns, in addition to aggregate variation in expectations. Both sources of variation create distortion in beta measurement from the perspective of the empiricist. In this section we want to assess quantitatively how these two sources of variation affect together and separately mismeasurement in beta.

We start by defining two sets of betas ($\boldsymbol{\beta}^{\mathcal{C}}$ and $\boldsymbol{\beta}^{\mathcal{D}}$) that measure aggregate and cross-sectional variation in expected returns. The following proposition shows that these two sets of betas are linearly related to the set of betas, $\boldsymbol{\beta}$, that underlies investor's representation of the CAPM relation and those that the empiricist measures, $\widehat{\boldsymbol{\beta}}$, from realized excess returns.

Proposition 6. *The empiricist's betas are a weighted average of three types of beta:*

$$\widehat{\boldsymbol{\beta}} = (1 - \mathcal{C}^2 - \mathcal{D}^2)\boldsymbol{\beta} + \mathcal{C}^2\boldsymbol{\beta}^{\mathcal{C}} + \mathcal{D}^2\boldsymbol{\beta}^{\mathcal{D}}, \tag{40}$$

where, after defining $\mathbb{E}^{i*}[\widetilde{\mathbf{R}}^e] \equiv \mathbb{E}^i[\widetilde{\mathbf{R}}^e] - \overline{\mathbb{E}}[\widetilde{\mathbf{R}}^e]$ *and* $\mathbb{E}^{i*}[\widetilde{R}^e_{\mathbf{M}}] \equiv \mathbb{E}^i[\widetilde{R}^e_{\mathbf{M}}] - \overline{\mathbb{E}}[\widetilde{R}^e_{\mathbf{M}}]$,

$$\boldsymbol{\beta}^{\mathcal{C}} \equiv \frac{\mathrm{Var}[\overline{\mathbb{E}}[\widetilde{\mathbf{R}}^e]]\mathbf{M}}{\mathrm{Var}[\overline{\mathbb{E}}[\widetilde{R}^e_{\mathbf{M}}]]} \quad \textit{and} \quad \boldsymbol{\beta}^{\mathcal{D}} \equiv \frac{\mathrm{Var}[\mathbb{E}^{i*}[\widetilde{\mathbf{R}}^e]]\mathbf{M}}{\mathrm{Var}[\mathbb{E}^{i*}[\widetilde{R}^e_{\mathbf{M}}]]}. \tag{41}$$

For any individual asset n, the coefficient $\beta^{\mathcal{C}}_n$ is the slope of a *time-series* regression of consensus expected excess returns for asset n on those for the market. Thus, an asset with a high $\beta^{\mathcal{C}}_n$ exhibits greater fluctuations in its consensus expected excess returns relative to the market. For an individual asset n, the coefficient $\beta^{\mathcal{D}}_n$ is the slope of a *cross-sectional* regression (across investors) of individual investors' expected excess returns for asset n on those for the market. Thus, an asset with a high $\beta^{\mathcal{D}}_n$ exhibits greater dispersion in beliefs across investors about its returns relative to the market.

The following simple example illustrates the meaning of $\boldsymbol{\beta}^{\mathcal{D}}$. Two investors, Bull and Bear, hold different views about the future excess returns of asset n and of the market. Bull expects the market (asset n) to *over*-perform by 1% (2%) relative to consensus beliefs. Bear, on the other hand, expects the market (asset n) to *under*-perform by 1% (2%). Thus, in this example $\beta^{\mathcal{D}}_n = 2$. Note first that $\beta^{\mathcal{D}}_n$ is positive, because investors deviate from consensus beliefs in the same direction both on the market and asset n. Second, $\beta^{\mathcal{D}}_n$ is larger than one, because this deviation is larger on asset n than it is on the market. Thus $\beta^{\mathcal{D}}_n$ is a purely cross-sectional measure which captures whether or not investors' expectations on asset n and on the market deviate from consensus in the same direction; and whether deviations from consensus on asset n are inflated or deflated relative to those on the market.

In our model, $\boldsymbol{\beta}^{\mathcal{C}}$ and $\boldsymbol{\beta}^{\mathcal{D}}$ can be computed explicitly.

Corollary 6.1. *In the equilibrium of the model, $\boldsymbol{\beta}^{\mathcal{C}}$ and $\boldsymbol{\beta}^{\mathcal{D}}$ are given by*

$$\boldsymbol{\beta}^{\mathcal{C}} = \boldsymbol{\beta} + \left(\frac{\tau\sigma_{\mathbf{M}}^2}{\bar{\Phi}^2} - 1\right)\frac{\tau_\epsilon e_1 \bar{\Phi}^2}{\tau\sigma_{\mathbf{M}}^2 + \tau_\epsilon e_1 \bar{\Phi}^2}(\boldsymbol{\beta} - \mathbf{1}) \tag{42}$$

$$\boldsymbol{\beta}^{\mathcal{D}} = \boldsymbol{\beta} + \left(\frac{\tau\sigma_{\mathbf{M}}^2}{\bar{\Phi}^2} - 1\right)(\boldsymbol{\beta} - \mathbf{1}). \tag{43}$$

Corollary 6.1 shows that both $\boldsymbol{\beta}^{\mathcal{C}}$ and $\boldsymbol{\beta}^{\mathcal{D}}$ are more dispersed than true betas and thus contribute to the flattening of the SML: when true betas are larger than one, $\boldsymbol{\beta}^{\mathcal{C}}$ and $\boldsymbol{\beta}^{\mathcal{D}}$ are both larger than $\boldsymbol{\beta}$; conversely, when true betas are smaller than one, $\boldsymbol{\beta}^{\mathcal{C}}$ and $\boldsymbol{\beta}^{\mathcal{D}}$ are both smaller than $\boldsymbol{\beta}$. The extent of excess volatility in the market ($\tau\sigma_{\mathbf{M}}^2/\bar{\Phi}^2 - 1$) dictates the magnitude of these effects. We also note that the coefficients of ($\boldsymbol{\beta} - \mathbf{1}$) in Eqs. (42)-(43) are identical to those of $\mathcal{C}^2$ and $\mathcal{D}^2$ in the definition of the distortion δ in Proposition 4.

The empirical challenge in testing our main theoretical argument is to measure the two numbers $\mathcal{C}^2$ and $\mathcal{D}^2$ in Eq. (40), and to obtain the associated sets of betas, $\boldsymbol{\beta}^{\mathcal{C}}$ and $\boldsymbol{\beta}^{\mathcal{D}}$, for a cross section of stocks. This in turn involves finding proxies for consensus beliefs, $\overline{\mathbb{E}}[\widetilde{\mathbf{R}}^e]$, and individual beliefs, $\mathbb{E}^i[\widetilde{\mathbf{R}}^e]$. A database that conveniently offers such a cross section of expectations is the I/B/E/S database. Using data on analysts' forecasts we estimate $\boldsymbol{\beta}^{\mathcal{C}}$ and $\boldsymbol{\beta}^{\mathcal{D}}$ and assess their ability to price the cross-section of returns.

We first describe the data and how we construct $\boldsymbol{\beta}^{\mathcal{C}}$ and $\boldsymbol{\beta}^{\mathcal{D}}$ (Section 5.1). We then test the model's predictions through classical portfolio sorts along with model-implied versions of the relation in Proposition 6; we also discuss plausible magnitudes for the distortion parameter δ (Section 5.2). Finally, we provide evidence that our theory can help explain abnormal returns on Betting Against Beta (Frazzini and Pedersen, 2014) (Section 5.3).

5.1 Data and construction of main variables

From Kenneth French's data library, we obtain market excess returns and risk-free security returns at the daily frequency. From the Center for Research in Security Prices (CRSP) database, we select all stocks that are listed in the S&P 500 index, using the list of historical constituents available from Compustat. For these stocks, we obtain excess returns and market capitalizations at the daily frequency. From the AQR data library, we obtain excess returns on "Betting Against Beta" (BAB) for the U.S. at the monthly frequency. Finally, from the Institutional Brokers' Estimate System database (I/B/E/S) we obtain unadjusted data on price targets. I/B/E/S price targets data covers a period of 20 years, from 1999 to

2019, and therefore dictates the sample period used for our tests.[19]

In a first step we obtain for each individual stock n and on the last trading day of each month t: the stock's 1-year *past* excess return; the stock's 1-year *future* excess return; and the stock's 1-year *expected* excess return. Past and future excess returns are obtained directly from the CRSP database and are unique for each stock-date observation. Expected excess returns are constructed over a lookback window that spans the last 6 months that precede and include date t. For a given stock n and date t, we record all 12-months price targets issued by institution i (e.g., "Bear Stearns") over this window; in case an institution issues multiple targets over the window we select the most recent one (e.g., Engelberg, McLean, and Pontiff, 2018). We then proxy for $\mathbb{E}^i[\widetilde{R}^e_n]$ using institution i's expected excess return:

$$\mathbb{E}^i[\widetilde{R}^e_n] = \frac{\text{Price Target}^i_{n,t+12\text{ months}} - \text{Price}_{n,t}}{\text{Price}_{n,t}} - \text{RF}_t, \tag{44}$$

where RF_t denotes the risk-free rate at date t. Hence, for each stock-date observation there are as many $\mathbb{E}^i[\widetilde{R}^e_n]$'s as there are institutions issuing targets for this stock over the window.

In constructing a proxy for expected excess returns in Eq. (44) we make several choices. First, we keep track of $\mathbb{E}^i[\widetilde{R}^e_n]$ at the institution level, as opposed to the analyst level, as an institution covers many more stocks than an individual analyst does; this ensures that there are multiple forecasters covering a given pair of stocks, which is critical for the computation of dispersion betas, $\boldsymbol{\beta}^{\mathcal{D}}$. Second, choosing the length of the lookback window is not trivial but has little effect on our conclusions (see Appendix C). Third, targets that are announced when the stock market is closed are shifted to the closest, preceding business day. An analyst sometimes, although rarely, issues multiple targets for the same firm on the same day (likely by mistake), in which case we select the most recent activation time. Finally, we remove all firms that have less than median coverage, all nonpositive targets, and all expected returns below the first and above the 99th percentile; other data-cleaning details follow closely the strategy in Engelberg et al. (2018), and leave us with a total of 429,556 expected excess return datapoints issued by 585 unique forecasters from December 1999 to September 2019.

In a second step, at each month t, we compute *consensus* 1-year expected excess returns by taking the median across all forecasters for each individual stock n. This is the empirical counterpart to consensus beliefs about stock n, $\overline{\mathbb{E}}[\widetilde{R}^e_n]$. To obtain consensus beliefs about market excess returns, $\overline{\mathbb{E}}[\widetilde{R}^e_{\mathbf{M}}]$, we compute the value-weighted average of consensus beliefs across the cross-section of stocks. At this stage, the dataset necessary to compute realized

[19] Appendix C provides a description of all datasets that we use in our empirical work, together with details on all the operations we use to transform the data and obtain our main variables. We also discuss the robustness of most of the choices we have made in this section.

betas, $\widehat{\boldsymbol{\beta}}$, consensus betas, $\boldsymbol{\beta}^{\mathcal{C}}$, and dispersion betas, $\boldsymbol{\beta}^{\mathcal{D}}$, is complete. From this dataset we obtain realized and consensus betas directly: $\widehat{\beta}_n$ is the slope coefficient from regressing stock n's past excess returns on past market excess returns; similarly, $\beta_n^{\mathcal{C}}$ is the slope coefficient from regressing stock n's past *consensus* excess returns on past *consensus* excess market returns. However, we cannot run standard regressions to compute dispersion betas, because we do not directly observe investors' beliefs about market returns. Instead, we reconstruct $\boldsymbol{\beta}^{\mathcal{D}}$ directly using its definition in Proposition 6.

The central piece in obtaining $\boldsymbol{\beta}^{\mathcal{D}}$ is the matrix of "co-beliefs," $\mathrm{Var}[\mathbb{E}^{i*}[\widetilde{\mathbf{R}}^e]]$. This matrix is not the typical measure of dispersion in beliefs (e.g., Diether et al., 2002). Empiricists usually measure beliefs dispersion as cross-sectional variances, which correspond to all elements along the diagonal of this matrix. But beta is all about measuring comovement between stocks—off-diagonal terms of this matrix. We call these terms *co-beliefs* as they measure how a pair of stocks *covaries* across beliefs. In practice, since all stocks have multiple analysts covering them, computing cross-sectional variances is always possible. Computing cross-sectional covariances, however, requires that a pair of stocks has at least two analysts in common. We explain in Appendix C how to obtain the matrix of co-beliefs at each date t using a 3-year rolling window of monthly data. We then obtain $\boldsymbol{\beta}^{\mathcal{D}}$ from the matrix of co-beliefs and assets' market weights, as defined in Eq. (41).[20]

For the three sets of betas (realized betas $\widehat{\boldsymbol{\beta}}$, consensus betas $\boldsymbol{\beta}^{\mathcal{C}}$, and dispersion betas $\boldsymbol{\beta}^{\mathcal{D}}$) we now have time series of 190 end-of-month observations, ranging from December 2002 to September 2018, across an average of 410 stocks.[21] Table 1 presents summary statistics for each set of betas $\widehat{\boldsymbol{\beta}}$, $\boldsymbol{\beta}^{\mathcal{C}}$, and $\boldsymbol{\beta}^{\mathcal{D}}$, along with market excess returns, $\widetilde{R}^e_{\mathbf{M}}$, and consensus expected market excess returns, $\overline{\mathbb{E}}[\widetilde{R}^e_{\mathbf{M}}]$. All numbers are at the 1-year horizon.

[INSERT TABLE 1 HERE]

Over the sample period, the average market excess return was 9.85% per year, with volatility of 16.16%. As expected, consensus expected market excess returns, $\overline{\mathbb{E}}[\widetilde{R}^e_{\mathbf{M}}]$, are less volatile than realized market excess return (second line of the table). The last three lines of the table compute averages and standard deviations over the entire sample for the three sets

[20] There is a tradeoff in choosing the length of the rolling window used for these computations: using a shorter window yields a sparse covariance matrix (there are not many pairs of stocks that have enough forecasts to allow computation of covariances); using a longer window contaminates the cross-sectional variation in $\mathbb{E}^{i*}[\widetilde{\mathbf{R}}^e]$ with stale observations. Nevertheless, our results are robust to varying the window from two to four years.

[21] We lose 3 years of data at the beginning of the sample due to the initial window over which we estimate betas; at the end of the sample we lose another year of data to annualize future returns. All betas are winsorized at the 0.5% and 99.5% percentile levels (Bali, Engle, and Murray, 2016).

of betas; all betas have averages close to one and high degrees of variation across stocks and time, with standard deviations well above one.

In Section 2 we have stressed how central the two statistics, $\mathcal{C}^2$ and $\mathcal{D}^2$, were to our main argument; together they account for the informational gap between investors and the empiricist. We can now estimate their magnitude and find that on average $\mathcal{C}^2 = 0.0400$ and $\mathcal{D}^2 = 0.0563$.[22] The statistic $\mathcal{C}^2$, which measures how much variation in market returns is explained by aggregate variation in expectations, has received considerable attention in the literature. Our estimate is lower than what is typically found, which indicates that analysts' forecasts likely exhibit less aggregate variation than other proxies for expected returns.[23] Since variation is key to generate distortion in beta estimates, we regard this outcome as keeping our subsequent results on the conservative side.

Most importantly, the other number $\mathcal{D}^2$ measures how much variation in market returns is explained by dispersion in expectations, an estimate of which is missing in the literature. Recall that $\mathcal{D}^2$ is not a traditional measure of beliefs dispersion (e.g., Diether et al., 2002). Applied to the market such a traditional measure would actually represent a value-weighted average of cross-sectional variances along the diagonal of the matrix of co-beliefs, i.e., $\sum_{n=1}^{N} M_n \operatorname{Var}[\mathbb{E}^{i*}[\widetilde{R}^e_n]]$. In contrast, $\mathcal{D}^2$ takes a value-weighted average over the entire matrix:

$$\mathcal{D}^2 \equiv \operatorname{Var}[\widetilde{R}^e_{\mathbf{M}}]^{-1} \left(\sum_{n=1}^{N} M_n^2 \operatorname{Var}[\mathbb{E}^{i*}[\widetilde{R}^e_n]] + \sum_{n \neq m} \sum M_n M_m \operatorname{Cov}[\mathbb{E}^{i*}[\widetilde{R}^e_n], \mathbb{E}^{i*}[\widetilde{R}^e_m]] \right), \qquad (45)$$

and thus includes all *co-beliefs* about each pair of stocks, i.e., off-diagonal elements that reflect the extent to which deviations from consensus beliefs on one stock tend to covary with deviations on another, which is not the same as disagreement.

The magnitude of $\mathcal{D}^2$, which is higher than that of $\mathcal{C}^2$ in our sample, points to an important source of variation in expected returns.

5.2 Empirical tests and a plausible magnitude for δ

We start by performing classical portfolio sorts, forming five (measured) beta-sorted portfolios.[24] Using these sorts we first confirm in our sample the well-known fact that the SML looks "flat" (e.g., Black et al., 1972). Table 2 reports value-weighted averages for excess

[22] Just like betas these two numbers are computed over rolling windows and thus move over time.

[23] In Martin (2017), Table 1, $\mathcal{C}^2$ is close to 10%, depending on the horizon of option prices used for the estimation. In Cochrane (2011), Table 1, $\mathcal{C}^2$ is approximatively 11%, using return-forecasting regressions.

[24] We work with five beta-sorted portfolios due to the relatively smaller number of stocks in our sample (an average of 410 each month). Nevertheless, results are robust (but noisier) if we use ten portfolios instead.

returns, μ, CAPM alphas, $\widehat{\alpha}$, CAPM betas, $\widehat{\beta}$, consensus betas, $\beta^{\mathcal{C}}$, dispersion betas, $\beta^{\mathcal{D}}$, volatilities, $\widehat{\sigma}$, and Sharpe ratios, SR, for each portfolio.[25] In line with a vast literature, portfolios with low average betas have significantly higher average alphas.

[INSERT TABLE 2 HERE]

However, lesser known in the literature is the fact that portfolios with higher average betas also have higher consensus betas, $\beta^{\mathcal{C}}$, and higher dispersion betas, $\beta^{\mathcal{D}}$. An important implication of Proposition 6 is that $\beta^{\mathcal{C}}$ and $\beta^{\mathcal{D}}$ command negative risk premia (an implication that we confirm and discuss shortly). It is thus possible that high alphas on low-beta portfolios are in fact the consequence of a relatively weaker exposure to $\beta^{\mathcal{C}}$ and $\beta^{\mathcal{D}}$. We test this conjecture in Table 3. Panel (a) shows results of time-series regressions of each portfolio's excess returns on the market (the CAPM), whereas panel (b) adds controls for $\beta^{\mathcal{C}}$ and $\beta^{\mathcal{D}}$ for each portfolio (value weighted).

[INSERT TABLE 3 HERE]

Clearly, panel (a) confirms the results of Table 2: low-beta portfolios continue to have high alphas, which remain statistically significant. But after adding controls for $\beta^{\mathcal{C}}$ and $\beta^{\mathcal{D}}$ in panel (b), these results change substantially. Alphas on low-beta portfolios lose considerable statistical power ($\beta^{\mathcal{C}}$ and $\beta^{\mathcal{D}}$ fully eliminate statistical significance on P1, P2 and P3), whereas the alpha of the high-beta portfolio P5 becomes positive and statistically significant. Panel (b) also shows that coefficients on $\beta^{\mathcal{C}}$ and $\beta^{\mathcal{D}}$ are mainly negative, with some of them statistically significant, suggesting that $\beta^{\mathcal{C}}$ and $\beta^{\mathcal{D}}$ earn a negative premium. We test this implication next.

We first establish that the model implies a negative premium on $\beta^{\mathcal{C}}$ and $\beta^{\mathcal{D}}$. Note that the main prediction of the model is that true betas and empiricist's betas satisfy the proportionality relation in Proposition 3. Substituting this relation into the statistical relation of Proposition 6 yields an affine relation between the three sets of beta:

$$\widehat{\boldsymbol{\beta}} = \frac{\delta(1-\mathcal{C}^2-\mathcal{D}^2)}{\delta+\mathcal{C}^2+\mathcal{D}^2}\mathbf{1} + \frac{\mathcal{C}^2(1+\delta)}{\delta+\mathcal{C}^2+\mathcal{D}^2}\boldsymbol{\beta}^{\mathcal{C}} + \frac{\mathcal{D}^2(1+\delta)}{\delta+\mathcal{C}^2+\mathcal{D}^2}\boldsymbol{\beta}^{\mathcal{D}}. \tag{46}$$

Alternatively, we can express this relation in terms of expected returns, which gives:

$$\mathbb{E}[\widetilde{\mathbf{R}}^e] = \frac{\mathbb{E}[\widetilde{R}^e_{\mathbf{M}}]}{1-\mathcal{C}^2-\mathcal{D}^2}\widehat{\boldsymbol{\beta}} - \frac{\mathcal{C}^2\,\mathbb{E}[\widetilde{R}^e_{\mathbf{M}}]}{1-\mathcal{C}^2-\mathcal{D}^2}\boldsymbol{\beta}^{\mathcal{C}} - \frac{\mathcal{D}^2\,\mathbb{E}[\widetilde{R}^e_{\mathbf{M}}]}{1-\mathcal{C}^2-\mathcal{D}^2}\boldsymbol{\beta}^{\mathcal{D}}. \tag{47}$$

[25] To account for the impact of autocorrelation and heteroscedasticity, all standard errors are adjusted using the Newey and West (1987) method with four lags (Greene, 2003, p. 267).

To obtain this relation, multiply Eq. (40) by $\mathbb{E}[\widetilde{R}^e_{\mathbf{M}}]$ and substitute the true CAPM relation $\mathbb{E}[\widetilde{\mathbf{R}}^e] = \boldsymbol{\beta}\,\mathbb{E}[\widetilde{R}^e_{\mathbf{M}}]$. Thus, $\beta^{\mathcal{C}}$ and $\beta^{\mathcal{D}}$ earn a negative premium in the model.

We test the two cross-sectional relations in Eqs. (46)–(47) following a two-step method. Starting with the beta relation in Eq. (46) we estimate the cross-sectional regression of $\widehat{\boldsymbol{\beta}}$ on $\boldsymbol{\beta}^{\mathcal{C}}$ and $\boldsymbol{\beta}^{\mathcal{D}}$ at the end of each month; this yields a time series of 190 observations for each regression coefficient. We then examine whether the average of each time series differs from zero. This procedure is similar to Fama and MacBeth (1973, FM hereafter) regression analysis, with the exception that the left-hand variable is taken to be realized beta, as opposed to future excess return. We report the average coefficients in Table 4. Eq. (46) predicts a positive relation between measured betas $\widehat{\beta}$ and the two sets of betas $\boldsymbol{\beta}^{\mathcal{C}}$ and $\boldsymbol{\beta}^{\mathcal{D}}$, which the positive and strongly statistically significant estimates support.

[INSERT TABLE 4 HERE]

In Table 5 we repeat the FM regressions for Eq. (47), according to which consensus betas, $\boldsymbol{\beta}^{\mathcal{C}}$, dispersion betas, $\boldsymbol{\beta}^{\mathcal{D}}$, and realized betas, $\widehat{\boldsymbol{\beta}}$, should explain the cross-section of expected returns. In particular, this relation predicts that $\boldsymbol{\beta}^{\mathcal{C}}$ and $\boldsymbol{\beta}^{\mathcal{D}}$ earn a negative premium, which the last row of the table strongly confirms, with negative and statistically significant coefficients on $\boldsymbol{\beta}^{\mathcal{C}}$ or $\boldsymbol{\beta}^{\mathcal{D}}$. This negative relation has key implications for betting against beta (BAB), which we analyze in Section 5.3. The other rows of Table 5 present estimates when only including $\boldsymbol{\beta}^{\mathcal{C}}$ or $\boldsymbol{\beta}^{\mathcal{D}}$ as explanatory variable, along with estimates for the canonical CAPM specification. Remarkably, the CAPM slope strengthens both in magnitude and statistical significance when $\boldsymbol{\beta}^{\mathcal{C}}$ and $\boldsymbol{\beta}^{\mathcal{D}}$ are included in the regression.

[INSERT TABLE 5 HERE]

Finally, the estimates of Table 4 allow us to perform a magnitude check of the distortion coefficient, δ. Specifically, Table 4 shows that the intercept a_0 belongs to the 90% confidence interval $a_0 \in [0.78, 0.88]$. Furthermore, from Eq. (46) we have:

$$\delta = \frac{a_0(\mathcal{C}^2 + \mathcal{D}^2)}{1 - a_0 - \mathcal{C}^2 - \mathcal{D}^2}, \tag{48}$$

which in turn gives a 90% confidence interval for δ. In Figure 3 we plot this interval (shaded area) as a function of the informational gap, $\mathcal{C}^2 + \mathcal{D}^2$.

[INSERT FIGURE 3 HERE]

The plot shows that the distortion is significant, ranging from 0.5 to 3 for the point estimate, $\mathcal{C}^2 + \mathcal{D}^2 = 9.63\%$, in our sample. By comparison, the Vasicek (1973) shrinkage proposed in finance textbooks (Bodie et al., 2007; Berk and DeMarzo, 2007) and adopted by practitioners is $\delta = 0.5$, compared to our point estimate of $\delta = 1.1$. The 90% confidence interval shows that the distortion can in fact be much larger. Levi and Welch (2017) is the only reference we know that advocates for a larger shrinkage.

Another way of obtaining a rough estimate of the CAPM distortion involves computing unconditional alphas, as is customary in the literature (e.g., Lewellen and Nagel, 2006). Empiricist's alpha satisfies $\widehat{\boldsymbol{\alpha}} \equiv \delta(\mathbf{1} - \boldsymbol{\beta})\,\mathbb{E}[\widetilde{R}^e_{\mathbf{M}}]$, which replaced in Eq. (32) yields:

$$\widehat{\boldsymbol{\alpha}} = \underbrace{\left(\frac{\tau\sigma^2_{\mathbf{M}}}{\bar{\Phi}^2} - 1\right)\mathcal{C}^2 \frac{\tau_\epsilon e_1 \bar{\Phi}^2}{\tau\sigma^2_{\mathbf{M}} + \tau_\epsilon e_1 \bar{\Phi}^2}(\mathbf{1} - \boldsymbol{\beta})\,\mathbb{E}[\widetilde{R}^e_{\mathbf{M}}]}_{\text{Consensus } (\mathcal{C})} + \underbrace{\left(\frac{\tau\sigma^2_{\mathbf{M}}}{\bar{\Phi}^2} - 1\right)\mathcal{D}^2(\mathbf{1} - \boldsymbol{\beta})\,\mathbb{E}[\widetilde{R}^e_{\mathbf{M}}]}_{\text{Dispersed information } (\mathcal{D})}. \quad (49)$$

The first term in $\widehat{\boldsymbol{\alpha}}$ has been thoroughly discussed in the literature (Jagannathan and Wang, 1996; Lewellen and Nagel, 2006; Boguth et al., 2011), but the second term is new. To determine how much of $\widehat{\boldsymbol{\alpha}}$ this new channel can explain, we assume that the first term is zero. All components of the second term have empirical counterparts. In particular, $(\tau\sigma^2_{\mathbf{M}}/\bar{\Phi}^2 - 1)$ represents excess market volatility. Based on realized excess market returns volatility, 16.16% (Table 1), and $\mathcal{C}^2 + \mathcal{D}^2 = 9.63\%$, we obtain $\sigma^2_{\mathbf{M}} = 0.1616^2(1 - 0.0963) = 0.0236$. Furthermore, $\tau/\bar{\Phi}^2$ is investors' precision of information regarding fundamentals: setting volatility of fundamentals at 3%, we obtain $\tau/\bar{\Phi}^2 = 1111$ (which likely is a conservative estimate) and therefore $(\tau\sigma^2_{\mathbf{M}}/\bar{\Phi}^2 - 1) = 25.22$. Hence, combined with $\mathcal{D}^2 = 5.63\%$, an asset for which investors observe a beta of 0.5, for the empiricist has a *positive* alpha of $0.71 \times \mathbb{E}[\widetilde{R}^e_{\mathbf{M}}]$, i.e., almost 3/4 of the market risk premium.

5.3 Betting Against Beta

We now show how consensus beta $\beta^{\mathcal{C}}$ and dispersion beta $\beta^{\mathcal{D}}$ help explain the underperformance of high-beta stocks relative to low-beta stocks (Friend and Blume, 1970; Black et al., 1972). This well-known empirical anomaly is illustrated in Figure 4. This figure compares mean excess returns for the five beta-sorted portfolios of Table 2 to CAPM-implied excess returns. Clearly, high-beta stocks underperform relative to low-beta stocks.

[INSERT FIGURE 4 HERE]

Frazzini and Pedersen (2014), building on insights from Black (1972), explore the underperformance of high-beta stocks by "Betting Against Beta," and attribute the success of this investment strategy to investors' borrowing constraints. Although we do not dispute the success of the BAB strategy, our interpretation differs: we claim that part of this success is because "betting against *measured* beta is betting on *true* beta."

To bet against beta, the empiricist builds two portfolios, a low-beta portfolio and a high-beta portfolio. Denote the betas of these two portfolios, as measured by the empiricist, by $\widehat{\beta}_L < 1$ and $\widehat{\beta}_H > 1$, respectively. The empiricist takes a long, leveraged position $(1/\widehat{\beta}_L)$ in the low-beta portfolio and a short, de-leveraged position $(-1/\widehat{\beta}_H)$ in the high-beta portfolio. This strategy has zero measured beta by construction, but a flat CAPM implies it has a strictly positive alpha. Formally, Proposition 3 implies the unconditional alpha on BAB is:

$$\widehat{\alpha}_{\text{BAB}} = \left(\frac{1}{\widehat{\beta}_L} - \frac{1}{\widehat{\beta}_H}\right)\frac{\delta}{1+\delta}\,\mathbb{E}[\widetilde{R}^e_{\mathbf{M}}] > 0. \tag{50}$$

We start with a magnitude check. Our empirical estimates, $\mathcal{C}^2 + \mathcal{D}^2 = 9.63\%$, imply a point estimate for δ of 1.1 (Figure 3). In our sample $\widehat{\beta}_L$ is 0.55 on average, whereas $\widehat{\beta}_H$ is 1.93.[26] Average market excess return over the sample period is 9.9% (Table 1). Plugging these numbers in Eq. (50) yields $\widehat{\alpha}_{\text{BAB}} = 6.74\%$. In comparison, the monthly alpha on BAB for the U.S. over the same period (using data downloaded from the AQR data library, 2002-2018) is 0.64% per month (or 7.6% per year). Thus, in our sample at least, 90% of the alpha on BAB may result from beta mismeasurement. Alternatively, our claim is that alpha on BAB is in fact reward for systematic risk. Thus, the true beta on the BAB strategy needs to be as large as BAB's alpha, 7.6%, divided by average market excess return, 9.9%, that is 0.78. Eq. (50) and numbers above imply in turn a true beta of 0.68, which accounts for 90% of the beta (0.78) necessary to explain alpha on BAB.

To test whether beta mismeasurement can actually explain returns on BAB, the problem we face is that AQR implements BAB for the U.S. on a sample that covers many more firms than ours. As a result, we cannot control for $\beta^{\mathcal{C}}$ and $\beta^{\mathcal{D}}$ for all stocks on which BAB is constructed. We address this problem with two sets of tests. The main idea of the first set of tests is to construct factors associated with $\beta^{\mathcal{C}}$ and $\beta^{\mathcal{D}}$. Eq. (47) says a BAB strategy based only on $\widehat{\beta}$ has non-zero alpha, but that long-short portfolios (factors) based on $\beta^{\mathcal{C}}$ and $\beta^{\mathcal{D}}$ explain this alpha. We thus build these two factors, $\mathcal{C}$ and $\mathcal{D}$, and use their returns as controls. In the second set of tests we replicate the BAB strategy within our sample instead, which we refer to as "in-house BAB" and which allows us to control for $\beta^{\mathcal{C}}$ and $\beta^{\mathcal{D}}$ directly.

We construct the factors $\mathcal{C}$ and $\mathcal{D}$ mirroring the steps in Frazzini and Pedersen (2014) in computing standard BAB returns. At the end of each month, we rank stocks according to their beta ($\beta^{\mathcal{C}}$ or $\beta^{\mathcal{D}}$) and form two portfolios, one with high-beta stocks and another with low-beta stocks. When forming these portfolios, we follow the ranking and weighting

[26] We rank betas following Section 3.2 in Frazzini and Pedersen (2014) and compute averages using ranks as portfolio weights, exactly as in the original paper; we further follow Frazzini and Pedersen (2014) and apply the Vasicek (1973) shrinkage for betas, with $w = 0.6$.

methodology in Frazzini and Pedersen (2014). We then build returns on each factor by going long the high-beta portfolio and short the low-beta portfolio. The factor $\mathcal{C}$ provides long exposure to consensus beta $\beta^{\mathcal{C}}$; the factor $\mathcal{D}$ provides long exposure to dispersion beta $\beta^{\mathcal{D}}$.

Before testing whether factors $\mathcal{C}$ and $\mathcal{D}$ can explain returns on BAB we first verify that these factors cannot be explained by market risk. In Table 6, we regress realized excess returns of the $\mathcal{C}$ and $\mathcal{D}$ factors on the excess returns of the market. Both factors have negative and statistically significant alphas, in line with the results of Table 5. That is, exposure to consensus (dispersion) beta earns a negative premium after controlling for market risk. Although both factors have statistically significant betas, these betas are weak in economic magnitude, and thus excess returns on these two factors are not entirely explained by exposure to market risk.

[INSERT TABLE 6 HERE]

We report the first set of tests in panel (a) of Table 7. Specifically, we perform two regressions, one regressing excess returns on AQR BAB on market excess returns, the other adding to this specification excess returns on factors $\mathcal{C}$ and $\mathcal{D}$. Although weaker in our sample, the alpha on the AQR BAB strategy is positive and statistically significant, close to 4% per year. Yet both factors $\mathcal{C}$ and $\mathcal{D}$ fully eliminate this alpha, with a significant loading on factor $\mathcal{D}$, suggesting that dispersion in beliefs plays an important role in explaining abnormal returns on BAB. Remarkably, the statistical significance of the loading on the market along with the adjusted R^2 rise considerably once we include factors $\mathcal{C}$ and $\mathcal{D}$. This phenomenon strongly suggests that both factors act as omitted variables, and that BAB in reality exposes the empiricist to market risk due to beta mis-measurement.

[INSERT TABLE 7 HERE]

We now replicate the BAB strategy in our sample—"in-house." We use our realized betas at the end of each month and simply follow the guidelines in Frazzini and Pedersen (2014) (Section 3.2), with the only exception that our realized betas and returns are at the one-year horizon, as opposed to a one-month horizon. Nevertheless, our strategy does capture the BAB factor: comparing its returns with those of the original monthly BAB strategy (annualized) the correlation between the two is 0.80; regressing the returns of the AQR BAB strategy on those of in-house BAB strategy yields an intercept of 0.009 (t-stat 0.85) and a slope of 0.559 (t-stat 11.8). Further regressing in-house BAB returns on market returns delivers a positive, significant alpha of 5.21% (t-stat 2.16), indicating that in-house BAB yields abnormal returns, too (and actually performs better than AQR BAB in our sample).

In panel (b) of Table 7 we repeat the two regressions of panel (a) but on the returns of in-house BAB. Just like in our tests for BAB as implemented by AQR, the alpha on in-house BAB in the first regression is positive and statistically significant, and vanishes as we control for factors $\mathcal{C}$ and $\mathcal{D}$. In this case, we can further control for $\beta^{\mathcal{C}}$ and $\beta^{\mathcal{D}}$ directly, since we have these betas for every stock involved in the strategy. Panel (c) shows that controlling for $\beta^{\mathcal{C}}_{BAB}$ and $\beta^{\mathcal{D}}_{BAB}$ eliminates abnormal returns on the strategy, confirming that returns on BAB appear related to beta mis-measurement. Finally, going back to panel (b), we observe an effect similar to that in panel (a): controlling for the $\mathcal{C}$ and $\mathcal{D}$ factors raises adjusted R^2 and market beta, both in magnitude and significance. Overall, we obtain qualitatively identical results in both sets of tests suggesting that, without controlling for the $\mathcal{C}$ and $\mathcal{D}$ factors, market beta (alpha) on BAB appear to be mistakenly weak (strong).

Of course these conclusions are dependent on our sample, and are thus subject to its limitations. Achieving perfect inference requires observing expected returns for each individual investor in the economy, along with the market portfolio. Yet with the data at hand we cannot reject the theoretical possibility that BAB returns result from beta mismeasurement.

6 Extensions and robustness of assumptions

We discuss modeling assumptions and the generality of our results. We show that in the absence of dispersed information SML flattening always obtains in a noisy rational-expectations framework, irrespective of modeling choices. We then extend our model to a market portfolio with arbitrary weights and to multiple factors driving payoffs. Although none of these features compromise the validity of the true CAPM, they complicate the empiricist's task.

6.1 A more general proof of SML flattening

The equilibrium relation in Eq. (10) is not specific to our setup. In fact, the same relation prevails in any noisy rational-expectations (NRE) model, irrespective of the structure of assets' payoffs, the type (public or private) of information investors observe, or the structure of liquidity traders' demand, $\widetilde{\mathbf{m}}$ (e.g., Admati, 1985); provided that investors' risk aversion and the precision of their information are constant over time, it also holds period by period, $\overline{\mathbb{E}}_t[\widetilde{\mathbf{R}}^e_{t+1}] = \gamma\boldsymbol{\Sigma}(\mathbf{M} - \widetilde{\mathbf{m}}_t)$, in a dynamic NRE. Furthermore, except for being expressed in dollar excess returns, Eq. (10) also applies to more standard asset-pricing models in specific cases—it is a special case of the ICAPM (Merton, 1973) or of a standard intertemporal asset-pricing model (Campbell, 1993) when hedging demands are absent.[27]

What is specific to our NRE model is the nature of investors' covariance matrix, $\mathbf{\Sigma}$, and its relation to that of the empiricist, $\widehat{\mathbf{\Sigma}}$ (Lemma 1); they both depend on the structure of assets' payoffs (Eq. 3) and the structure of investors' signals. We now relax our assumptions regarding the structure of payoffs (Eq. 3), and adopt the general payoff structure in Admati (1985). Although we have shown in Section 4.2 that dispersed information amplifies CAPM distortion, we are unable to show that this result carries over to a general structure of private signals. We thus assume that all information is public, but allow for a general, Gaussian structure of public signals. In this context we show that the empiricist always observes a flattened SML, irrespective of the structure of payoffs or the structure of public information.

The proof builds on the observation that the slope of empiricist's SML is the slope of a regression of assets' unconditional expected excess returns on empiricist's betas. Using the true CAPM relation of Definition 1, this slope is given by:

$$\frac{\mathrm{Cov}[\mathbb{E}[\widetilde{\mathbf{R}}^e], \widehat{\boldsymbol{\beta}}]}{\mathrm{Var}[\widehat{\boldsymbol{\beta}}]} = \frac{\mathrm{Cov}[\boldsymbol{\beta}, \widehat{\boldsymbol{\beta}}]}{\mathrm{Var}[\widehat{\boldsymbol{\beta}}]} \mathbb{E}[\widetilde{R}^e_{\mathbf{M}}], \tag{51}$$

and thus empiricist's SML is flatter than the true SML when $\mathrm{Cov}[\boldsymbol{\beta}, \widehat{\boldsymbol{\beta}}] < \mathrm{Var}[\widehat{\boldsymbol{\beta}}]$. Proposition 7 says that this is always the case in an NRE model in which all information is public.

Proposition 7. *Consider the model of Section 3 with the following two modifications: (i) the structure of assets' payoffs, $\widetilde{\mathbf{D}}$, is arbitrary (e.g., as in Admati (1985)), and (ii) all information is public but of arbitrary structure. Then* $\mathrm{Cov}[\boldsymbol{\beta}, \widehat{\boldsymbol{\beta}}] < \mathrm{Var}[\widehat{\boldsymbol{\beta}}]$ *and the empiricist's SML is always flatter than that of investors.*

Intuitively, the empiricist faces higher uncertainty compared to investors, which results in greater dispersion in her betas relative to true betas, $\sigma_{\widehat{\boldsymbol{\beta}}} > \sigma_{\boldsymbol{\beta}}$. More formally, consider the eigenvalue decomposition, $\mathbf{\Sigma} \equiv \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}'$, of investors' covariance matrix, where $\mathbf{\Lambda}$ is a diagonal matrix with all positive eigenvalues on its diagonal, and $\mathbf{Q}$ is an orthogonal matrix of eigenvectors. Because all information is public the last term in Eq. (10) drops out; it follows that the empiricist's covariance matrix, $\widehat{\mathbf{\Sigma}}$, and $\mathbf{\Sigma}$ share the same eigenvectors,

$$\widehat{\mathbf{\Sigma}} = \mathbf{Q}\mathbf{\Lambda}(\mathbf{I} + \gamma^2/\tau_m \mathbf{\Lambda})\mathbf{Q}', \tag{52}$$

but not the same eigenvalues. All eigenvalues of $\widehat{\mathbf{\Sigma}}$ are an inflated version of those of $\mathbf{\Sigma}$; each eigenvalue λ is inflated by a factor $1 + \gamma^2/\tau_m \lambda$, where the ratio γ^2/τ_m captures the extent of

[27] A more general form of Eq. (10), in which γ and $\mathbf{\Sigma}$ are time-varying, has been derived by Jensen (1972), and further studied by Bollerslev, Engle, and Wooldridge (1988).

aggregate variation in expected returns. The empiricist's inflated covariance matrix results in inflated dispersion in betas around their average (which is one), leading to a flattened SML.

In the presence of dispersed information, $\mathbf{\Sigma}$ and $\widehat{\mathbf{\Sigma}}$ no longer share the same eigenvectors, and the effect of dispersed information on the SML is ambiguous. The ambiguity appears to be tied to our assumption of residual uncertainty in payoffs (i.e., $\tau_\epsilon < \infty$).[28] Without residual uncertainty, and with our specification of private signals and an arbitrary specification of payoffs, $\mathbf{\Sigma}$ and $\widehat{\mathbf{\Sigma}}$ share again the same eigenvectors, and the eigenvalues of $\widehat{\mathbf{\Sigma}}$ are now inflated by a factor, $1 + \gamma^2/\tau_m\lambda + \tau_v$. In other words, the effect of dispersed information, τ_v, simply reinforces the effect of aggregate variation, and flattening obtains systematically.

We revisit our assumption about the structure of payoffs in Section 6.3, where we examine the effect of generalizing it to multiple factors. We now relax another important assumption of our model, that the market portfolio, $\mathbf{M}$, is equally weighted. We show that relaxing this assumption may cause the empiricist's SML to be downward-sloping (in an economy in which the market risk premium is always positive).

6.2 Size effects

Assuming that assets are assigned unequal weights in the market portfolio has other, interesting effects on the empiricist's SML. For instance, the empiricist may perceive a downward-sloping SML, although the actual SML is *always* upward-sloping in this model; this situation occurs when assets that have high measured beta simultaneously have low market supply (and vice-versa).

In Section 4.1, we showed that the variance of the empiricist satisfies Lemma 1, a result that still holds under unequal supplies. The following proposition builds on this result.

Proposition 8. *In the context of Section 3 suppose the market portfolio,* $\mathbf{M}$*, is arbitrary. Then, expected returns on all assets in excess of the market satisfy the two-factor relation:*

$$\boldsymbol{\mu} - \mu_{\mathbf{M}}\mathbf{1} = \frac{\mu_{\mathbf{M}}}{1+\delta}(\widehat{\boldsymbol{\beta}} - \mathbf{1}) + \frac{\delta\mu_{\mathbf{M}}}{1+\delta}\left(\frac{\mathbf{M}}{\|\mathbf{M}\|^2} - \mathbf{1}\right), \tag{53}$$

where $\delta > 0$ *denotes the distortion coefficient associated with this new relation; this coefficient is defined as in the baseline model, except that it accounts for heterogeneous market weights.*

When the market portfolio is equally weighted, $\mathbf{M} = \mathbf{1}/N$, the second "factor," $\mathbf{M}/\|\mathbf{M}\|^2 - \mathbf{1}$, is $\mathbf{0}$ and we recover the result of Proposition 3; in contrast, when it is not, $\mathbf{M} \neq \mathbf{1}/N$, and

[28] We thank a referee for drawing our attention to this aspect.

Eq. (53) incorporates an additional factor. The sign and magnitude of this factor depend on the difference $\mathbf{M}/\|\mathbf{M}\|^2 - \mathbf{1}$, which measures the relative size of assets. To see how this second factor in Eq. (53) affects the measured slope of the SML, consider an asset that has high measured beta ($\widehat{\beta}_n > 1$) but earns *negative* returns in excess of the market portfolio. To satisfy Eq. (53) this asset must be *small*, i.e., $M_n/\|\mathbf{M}\|^2 < 1$. Thus, an economy in which high-beta assets are small can result in a downward-sloping SML, although the true CAPM holds and the true SML is upward sloping. We formalize this result in Lemma 2.

Lemma 2. *A necessary (but not sufficient) condition for empiricist's SML to be downward-sloping is that factor loadings and market weights covary negatively:*

$$\text{Cov}[\mathbf{\Phi}, \mathbf{M}] < 0. \tag{54}$$

Intuitively, there is a conflicting relation between, on the one hand, how much variation in an asset's returns the market can explain (the asset's beta) and, on the other, its weight in the market portfolio. If this tension is sufficiently strong, the empiricist uncovers a downward-sloping SML. On the contrary, if high-beta assets are large (i.e., if the covariance in Eq. (54) is sufficiently large and positive) the empiricist uncovers an SML that is even steeper than that of investors. In Figure 5 we illustrate the perplexing result of Lemma 2 whereby the empiricist perceives a negative relation between beta and expected returns. In particular, this figure depicts the investors' and empiricist's SML in an economy with three assets: assets no longer plot on a straight line, and the empiricist sees the SML as downward-sloping.

[INSERT FIGURE 5 HERE]

Of course, this size effect is merely an illusion, since the true CAPM never ceases to hold. Size appears as a priced factor because beta is mismeasured—because relative supplies in Eq. (53) scale with the distortion δ, there can be no size effect without distortion in betas.

6.3 Large economy and multiple factors

We now allow payoffs to be driven by $J \geq 1$ common factors, as opposed to just one. Furthermore, we consider a "large economy" in which the number of stocks, N, and the number of factors, J, both grow unboundedly but in a way that their relative size, $J/N \to \psi \in [0, 1]$ remains finite (e.g., Martin and Nagel (2020)). In this context, we can provide conditions under which factor multiplicity generates flattening, and an in-depth analysis of how dispersed information contributes to the SML distortion, δ.

Denote a vector of $J \leq N$ independent factors by $\widetilde{\mathbf{F}} \equiv [\widetilde{F}_1 \ \widetilde{F}_2 \ldots \ \widetilde{F}_J]'$, which we assume to be normally distributed with mean $\mathbf{0}$ and covariance matrix $(\tau_F J)^{-1}\mathbf{I}_J$. Note that we scale prior precision on common factors by J to obtain meaningful limits subsequently.[29] As in the baseline model, realized asset payoffs have a common-factor structure:

$$\widetilde{\mathbf{D}} = D\mathbf{1} + \mathbf{\Phi}\widetilde{\mathbf{F}} + \tilde{\boldsymbol{\epsilon}}, \tag{56}$$

where the $j-$th column of the vector $\mathbf{\Phi}$ contains the loadings of each stock on the $j-$th factor. We further assume that $\text{rank}(\mathbf{\Phi}) = J$ and:

$$\frac{1}{NJ}\text{tr}(\mathbf{\Phi}'\mathbf{\Phi}) = 1, \tag{57}$$

which extends the normalization introduced in Section 3 to the multiple-factor case.

Each investor i observes a vector of private signals about the J factors:

$$\widetilde{\mathbf{V}}_i = \widetilde{\mathbf{F}} + \tilde{\mathbf{v}}_i, \quad \tilde{\mathbf{v}}_i \sim \mathscr{N}\left(\mathbf{0}, (\tau_v J)^{-1}\mathbf{I}_J\right). \tag{58}$$

We also scale the precision of private signals by J to ensure that their informational content is preserved in the large-economy limit. Other than allowing multiple factors to affect payoffs, we keep the structure of the model unchanged.

Our main results rely on the following eigenvalue decomposition:

$$\frac{1}{N}\mathbf{\Phi}'\mathbf{\Phi} = \mathbf{Q\Lambda Q}', \tag{59}$$

where $\mathbf{\Lambda}$ is a diagonal matrix with all eigenvalues $\lambda_j > 0$ for $j = 1, ..., J$ on its diagonal, and $\mathbf{Q}$ is an orthogonal matrix whose columns are eigenvectors. This decomposition is possible because $\frac{1}{N}\mathbf{\Phi}'\mathbf{\Phi}$ is symmetric. All eigenvalues of this matrix are strictly positive, and the normalization of Eq. (57) implies that their sum equals 1.[30] We follow Martin and Nagel (2020)'s assumption that each eigenvalue satisfies $\lambda > \varepsilon$, for some uniform constant, ε, as $N \to \infty$. This ensures that the columns of $\mathbf{\Phi}$ never become collinear in the limit.

Following our baseline notation we write investors' precision on the J factors as:

$$\boldsymbol{\tau} = \text{Var}\left[\widetilde{\mathbf{F}}\middle|\mathscr{F}^i\right]^{-1} = (\tau_F + \tau_v)J\mathbf{I}_J + \boldsymbol{\tau}_P\mathbf{\Phi}'\mathbf{\Phi}\boldsymbol{\tau}_P, \tag{60}$$

[29]This is because $\text{Var}[\mathbf{\Phi}\widetilde{\mathbf{F}}] = \frac{\tau_F^{-1}}{J}\mathbf{\Phi}\mathbf{\Phi}'$, so that the average prior does not grow with J under Eq. (57):

$$\frac{\tau_F^{-1}}{NJ}\sum_{n=1}^{N}(\Phi\Phi')_{nn} = \tau_F^{-1}, \tag{55}$$

[30]This is because the sum of eigenvalues of matrix equals its trace.

where $\boldsymbol{\tau}_P$ is a $J \times J-$matrix, which is defined formally in the appendix. It will prove convenient to rotate this matrix using the matrix of eigenvectors, $\mathbf{Q}$, in Eq. (59) and to work with the limiting behavior of $\mathbf{Q}'\boldsymbol{\tau}/N\mathbf{Q}$, as opposed to $\boldsymbol{\tau}/N$:

$$\boldsymbol{\tau}_\infty \equiv \lim_{J\to\infty, N\to\infty, J/N\to\psi} \mathbf{Q}'\boldsymbol{\tau}/N\mathbf{Q}. \tag{61}$$

We characterize this matrix in the next lemma.

Lemma 3. *The limiting precision matrix, $\boldsymbol{\tau}_\infty$, defined in Eq. (61) is a diagonal matrix; the $j-th$ element on its diagonal corresponds to the precision on the $j-th$ factor and is uniquely identified by the eigenvalue λ on this factor according to the cubic relation:*

$$\tau_\infty(\lambda) = \psi\left(\tau_F + \tau_v + \frac{\lambda \tau_m \tau_v^2 \tau_\epsilon^2 \psi}{\gamma^2(\tau_\infty(\lambda) + \lambda\tau_\epsilon)^2}\right). \tag{62}$$

Remarkably, after rotating the precision matrix based on Eq. (59), the precision on each factor is uniquely identified by its eigenvalue. Thus, the equilibrium computation in this multiple-factor economy is not more complicated than solving the cubic equation of Proposition 1 J times, each time for a different eigenvalue identifying a factor. Interestingly, comparing Eqs. (62) and (60), shows that the ratio in Eq. (62) represents the contribution of learning from prices to factor precision. Because the effect of learning from prices is proportional to ψ^2, if ψ is small, Eq. (61) simplifies to:

$$\boldsymbol{\tau}_\infty = (\tau_F + \tau_v)\psi\mathbf{I} + O(\psi^2), \tag{63}$$

and thus all factors have identical precision, irrespective of their eigenvalue. That is, in a large economy in which the relative number of factors is small we can ignore the effect of learning from prices on precision, an observation that greatly helps for our main results.

We wish to characterize beta distortion with a single number, δ, as we did in the single-factor case. In principle, there are as many such numbers as there are factors in the economy, since factors affect the SML to different extents and in different directions. Rather, we examine how factors collectively affect the SML, which can be captured with a single number. Note that $1+\delta$ is the slope coefficient obtained from regressing true betas on measured betas:

$$1 + \delta = \text{Cov}[\boldsymbol{\beta}, \widehat{\boldsymbol{\beta}}]/\text{Var}[\boldsymbol{\beta}]. \tag{64}$$

Hence, when $\delta > 0$ ($\delta < 0$) the SML looks flatter (steeper) than it actually is.

The main difference relative to the single-factor case is that beta distortion can now create steepening, or even cause the SML to be downward-sloping. To characterize under which conditions each of these situations occur, we introduce the following assumption, which allows us to use results from random matrix theory.

Assumption 1. *The matrix of loadings can be decomposed as* $\mathbf{\Phi} \equiv \mathbf{X}'\mathbf{T}^{1/2}$*, where* $\mathbf{X}$ *is an* $J \times N-$*matrix with IID entries with mean zero, variance one and finite fourth moment, and* $\mathbf{T}$ *is a* $J \times J$ *positive-definite matrix and with* $tr(\mathbf{T}) = J$*.*

This means that loadings are on average zero (an assumption that can be relaxed) with covariance matrix $\mathbf{T}$. The relevance of this assumption is that it allows us to characterize the SML distortion in terms of the limiting variance and skewness of eigenvalues in Eq. (59):

$$\sigma_\lambda^2 \equiv \lim_{J\to\infty} \frac{1}{J}\sum_{j=1}^{J} \lambda_j^2 - \mu_\lambda^2, \tag{65}$$

$$s_\lambda \equiv \lim_{J\to\infty} \frac{1}{J}\sum_{j=1}^{J} \lambda_j^3 - 3\mu_\lambda\sigma_\lambda^2 - \mu_\lambda^3, \tag{66}$$

where $\mu_\lambda \equiv \lim_{J\to\infty} \frac{1}{J}\sum_{j=1}^{J} \lambda_j$ denotes the limiting mean of eigenvalues.

Proposition 9. (flattening of empiricist's SML) *Consider a large economy with a small ratio,* ψ*, of factors (in the sense of Eq. (63)), in which eigenvalues are not too dispersed:*

$$\sigma_\lambda^2 < \frac{1}{2}\left(\sqrt{\Delta + \frac{4s_\lambda(\mu_\lambda\tau_\epsilon + \tau_0)}{\tau_\epsilon}} + \mu_\lambda^2 + \frac{\mu_\lambda\tau_0(3-\psi)}{\tau_\epsilon} + \frac{\tau_0\psi\tau_1}{\gamma^2\tau_\epsilon^2}\right), \tag{67}$$

where τ_0*,* τ_1*,* Δ*, are strictly positive coefficients defined in the appendix. Then, if either eigenvalues are positively skewed (or exhibit little negative skewness) (Eq. (B.158)) or, on the contrary, if they exhibit strictly negative (but limited) skewness (Eq. (B.159)) and if, further, eigenvalues are not too concentrated (Eq. (B.160)), the SML will look flatter than it actually is. If, further, skewness is strictly negative (Eq. (B.162)) and if eigenvalues are sufficiently concentrated (Eq. (B.163)), the SML will be downward-sloping.*

Suppose that eigenvalues are not too dispersed (in the sense of Eq. (67)), meaning that factors have comparable predictive power. In this case, whether the distribution of eigenvalues is positively skewed or, on the contrary, exhibits negative but moderate skewness causes the SML to appear flatter than it actually is. More broadly, irrespective of skewness, low dispersion of eigenvalues induces flattening. If this dispersion is sufficiently weak (factors have sufficiently similar predictive power), the empiricist may even uncover a downward-sloping SML. Conditions for a steeper SML follow as a corollary to Proposition 9.

Corollary 9.1. (steepening of empiricist's SML) *Consider the framework of Proposition 9. If eigenvalues are sufficiently dispersed (inequality in Eq. (67) is violated) and exhibit a negative but limited skew (first inequality in Eq. (B.159) holds) or if eigenvalues are sufficiently concentrated (inequality in Eq. (B.160) is violated) and exhibit a strictly negative skew (inequality in Eq. (B.158) is violated), the SML will look steeper than it actually is.*

In conclusion, in an economy in which eigenvalues are positively skewed, heterogeneity in predictive power across factors leads to a steeper SML, whereas homogeneity thereof leads to a flatter SML. In contrast, if eigenvalues are negatively skewed, (limited) heterogeneity across factors produces a flatter SML, whereas homogeneity thereof produces a steeper SML.

We take advantage of this large economy to investigate further how information aggregation contributes to the SML distortion. We do so through the following, simplified example.

Example 1. *All entries in the loading matrix, $\mathbf{\Phi}$, have unit variance, $\mathbf{T} \equiv \mathbf{I}$.*

In this case, eigenvalues follow the Pastur-Marchenko law, so that $\mu_\lambda = 1$, $\sigma_\lambda^2 = \psi$ and $s_\lambda = \psi^2$. Proposition 9 implies that the SML looks always flatter than it really is and can never be downward-sloping. Interestingly, however, we can examine in detail how information dispersion affects SML distortion using the CIE of Section 4.2. In particular, we can write:

$$\delta = \rho(\mathcal{C}^2 + \mathcal{D}^2) + (1-\rho)\,\frac{\gamma^2}{\tau_m}\sigma_{\mathbf{M},\infty}^2\left(1 - \mathcal{C}^2 - \mathcal{D}^2\right), \tag{68}$$

where the ρ represents excess volatility and satisfies:

$$\rho \equiv \tau_\infty \sigma_{\mathbf{M},\infty}^2/\psi - 1 \equiv \lim_{\psi \to 0} \delta, \tag{69}$$

with $\tau_\infty = (\tau_F + \tau_v)\psi$ denoting the limiting precision on any given factor and $\sigma_{\mathbf{M},\infty}^2 = \lim_{N\to\infty} N\sigma_{\mathbf{M}}^2 = \tau_\epsilon^{-1} + (\tau_F + \tau_v)^{-1}$ denoting the limiting variance of the market portfolio.

To examine how information dispersion contributes to the distortion, we write an analogous relation between δ_{CIE} and $\mathcal{C}_{\text{CIE}}^2$ in the CIE, albeit in the context of a large economy:

$$\delta_{\text{CIE}} = \rho\mathcal{C}_{\text{CIE}}^2 + (1-\rho)\,\frac{\gamma^2}{\tau_m}\sigma_{\mathbf{M},\infty}^2\left(1 - \mathcal{C}_{\text{CIE}}^2\right). \tag{70}$$

Importantly, the coefficient in front of the information gap is identical in the CIE and the baseline economy ($\mathcal{C}^2 + \mathcal{D}^2$ and $\mathcal{C}_{\text{CIE}}^2$, respectively). Hence, a difference in δ across the two economies is solely attributed to a difference in information gap across the two economies.

The difference in information gap across economies is never negative, $\mathcal{D}^2 \geq \mathcal{C}_{\text{CIE}}^2 - \mathcal{C}^2 \geq 0$. From Section 4.2, we know that, for precisions to be identical across economies, the difference

$\mathcal{C}_{\text{CIE}} - \mathcal{C}$ must be positive. One would then perhaps expect this difference to be exactly accounted for by dispersion in beliefs, $\mathcal{D}^2$. However, this is typically not the case:

$$\mathcal{D}^2 = \Big(1 + \underbrace{\frac{\tau_m}{\gamma^2} \frac{\sigma^2_{\mathbf{M},\infty}}{\psi/\tau^2_\infty + \sigma^4_{\mathbf{M},\infty}}}_{d \equiv \text{information distortion}} \Big)(\mathcal{C}_{\text{CIE}} - \mathcal{C}). \tag{71}$$

The only case in which we obtain this identity is when $\psi \equiv 0$, which corresponds to the large-economy limit with finitely many factors. Eq. (71) says that the two numbers, $\mathcal{D}^2$ and $\mathcal{C}^2_{\text{CIE}} - \mathcal{C}^2$ do not add up, a distortion of the kind studied in Albagli et al. (2015). The difference in information gap across economies directly results from this distortion, d:

$$\mathcal{D}^2 + \mathcal{C}^2_{\text{CIE}} - \mathcal{C}^2 = d(\mathcal{C}^2_{\text{CIE}} - \mathcal{C}^2). \tag{72}$$

Finally, we can now show that the difference in beta distortion across economies entirely results from beliefs distortion, d, by subtracting Eq. (70) from Eq. (68):

$$\delta - \delta_{\text{CIE}} = d \left(\rho + (\rho - 1) \frac{\gamma^2}{\tau_m} \sigma^2_{\mathbf{M},\infty} \right) (\mathcal{C}^2_{\text{CIE}} - \mathcal{C}^2). \tag{73}$$

Thus, dispersed information is associated with larger SML distortion if $\psi > 0$ (it creates beliefs distortion, $d > 0$) and excess volatility, ρ, is large enough.

7 Conclusion

Why do empiricists keep rejecting the CAPM, which practitioners are not willing to abandon? We argue that the CAPM may hold from each investor's perspective, but that variations across investors' expectations causes empiricists to reject it. We thus provide a novel explanation for the empirical failure of the CAPM despite widespread practical use.

Variation in expected returns over time and across investors both contribute to the informational gap between investors and the empiricist. Whereas the literature has studied extensively how time-series variation distorts CAPM tests (e.g., Jagannathan and Wang, 1996), to our knowledge this paper is first to draw attention to dispersed information as an additional source of distortion. Based on analysts' price targets we find that the effect of dispersed information is stronger than that of time-series variation. Together these two sources of variation produce substantial CAPM distortion, sufficiently so to explain returns on BAB (Frazzini and Pedersen, 2014). The traditional interpretation (Black, 1972) is that betting against (measured) beta works because high-beta assets are overpriced; our interpretation is that betting against (measured) beta is betting on true beta.

The type of information investors observe matters. When public information dominates (e.g., FOMC meetings), dispersion in information weakens and empiricist's SML steepens. This result can explain why the CAPM appears stronger on public announcement days (Savor and Wilson, 2014), and why a compression in measured betas occurs on these occasions (Andersen et al., 2020; Eriksen and Gronborg, 2020; Bodilsen, 2019). Extensions of our model reveal puzzling situations whereby the empiricist's SML is downward sloping, although the market risk premium is always positive.

Our theory has implications for factor models in asset pricing. Some variables may appear to the empiricist as priced factors simply because betas are mis-measured. Rather than being priced factors, these variables are instruments for the measurement error in betas (Andrei, Cujean, and Fournier, 2019). Do there exist economic criteria that would allow the empiricist to distinguish variables that are economically meaningful from those that are not? This matter opens up fascinating directions for future research. Our empirical analysis exploits variation across analysts' forecasts, and demonstrates its relevance in explaining CAPM failure. An alternative approach is to exploit investors' actions (e.g., their investment decisions), which likely reveals greater variation. This approach based on "revealed preference" has caught on recently (Barber et al., 2016; Berk and Van Binsbergen, 2016).

References

Admati, A. R. (1985). A noisy rational expectations equilibrium for multi-asset securities markets. *Econometrica 53*(3), pp.629–658.

Albagli, E., C. Hellwig, and A. Tsyvinski (2015). A theory of asset pricing based on heterogeneous information. *Working Paper*.

Andersen, T. G., M. Thyrsgaard, and V. Todorov (2020). Recalcitrant betas: Intraday variation in the cross-sectional dispersion of systematic risk and expected returns.

Andrei, D., J. Cujean, and M. Fournier (2019). The low-minus-high portfolio and the factor zoo. Working paper.

Antoniou, C., J. A. Doukas, and A. Subrahmanyam (2015). Investor sentiment, beta, and the cost of equity capital. *Management Science 62*(2), 347–367.

Bacchetta, P. and E. V. Wincoop (2006, June). Can information heterogeneity explain the exchange rate determination puzzle? *American Economic Review 96*(3), 552–576.

Bai, Z. D., B. Q. Miao, and G. M. Pan (2007, 07). On asymptotics of eigenvectors of large sample covariance matrix. *Ann. Probab. 35*(4), 1532–1572.

Bai, Z. D. and J. W. Silverstein (1998, 01). No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices. *Ann. Probab. 26*(1), 316–345.

Baker, M., B. Bradley, and J. Wurgler (2011). Benchmarks as limits to arbitrage: Understanding the low-volatility anomaly. *Financial Analysts Journal 67*(1), 40–54.

Bali, T. G., N. Cakici, and R. F. Whitelaw (2011). Maxing out: Stocks as lotteries and the cross-section of expected returns. *Journal of Financial Economics 99*(2), 427–446.

Bali, T. G., R. F. Engle, and S. Murray (2016). *Empirical asset pricing: The cross section of stock returns*. John Wiley & Sons.

Banerjee, S. (2010). Learning from prices and the dispersion in beliefs. *Review of Financial Studies*.

Barber, B. M., X. Huang, and T. Odean (2016). Which factors matter to investors? evidence from mutual fund flows. *The Review of Financial Studies 29*(10), 2600.

Ben-Rephael, A., B. I. Carlin, Z. Da, and R. D. Israelsen (2017). Demand for information and asset pricing. Technical report, National Bureau of Economic Research.

Berk, J. B. and P. M. DeMarzo (2007). *Corporate finance*. Pearson Education.

Berk, J. B. and J. H. Van Binsbergen (2016). Assessing asset pricing models using revealed preference. *Journal of Financial Economics 119*(1), 1–23.

Biais, B., P. Bossaerts, and C. Spatt (2010). Equilibrium asset pricing and portfolio choice under asymmetric information. *Review of Financial Studies 23*(4), 1503–1543.

Black, F. (1972). Capital market equilibrium with restricted borrowing. *The Journal of Business 45*(3), 444–455.

Black, F., M. Jensen, and M. Scholes (1972). The capital asset pricing model: Some empirical tests. *Studies in the Theory of Capital Markets*.

Black, F. and R. Litterman (1990). Asset allocation: combining investor views with market equilibrium. Technical report, Discussion paper, Goldman, Sachs & Co.

Black, F. and R. Litterman (1992). Global portfolio optimization. *Financial analysts journal 48*(5), 28–43.

Blume, M. E. (1971). On the assessment of risk. *The Journal of Finance 26*(1), 1–10.

Bodie, Z., A. Kane, and A. Marcus (2007). *Essentials of investments.* McGraw-Hill.

Bodilsen, S. (2019). *Volatility Modeling with High-Frequency Data and News Announcements.* Ph. D. thesis, Aarhus University.

Boguth, O., M. Carlson, A. Fisher, and M. Simutin (2011). Conditional risk and performance evaluation: Volatility timing, overconditioning, and new estimates of momentum alphas. *Journal of Financial Economics 102*(2), 363–389.

Bollerslev, T., R. F. Engle, and J. M. Wooldridge (1988). A capital asset pricing model with time-varying covariances. *Journal of political Economy 96*(1), 116–131.

Buffa, A. M., D. Vayanos, and P. Woolley (2014). Asset management contracts and equilibrium prices. Technical report, National Bureau of Economic Research.

Burton, J. (1998). Revisiting the capital asset pricing model. *Dow Jones Asset Manager*, 20–28.

Campbell, J. Y. (1993). Intertemporal asset pricing without consumption data. *The American Economic Review*, 487–512.

Campbell, J. Y., S. Giglio, C. Polk, and R. Turley (2012). An intertemporal capm with stochastic volatility. Technical report, National Bureau of Economic Research.

Cochrane, J. H. (2007). The dog that did not bark: A defense of return predictability. *The Review of Financial Studies 21*(4), 1533–1575.

Cochrane, J. H. (2011). Presidential address: Discount rates. *The Journal of Finance 66*(4), 1047–1108.

Cohen, R. B., C. Polk, and T. Vuolteenaho (2005). Money illusion in the stock market: The modigliani-cohn hypothesis. *The Quarterly journal of economics 120*(2), 639–668.

Diether, K. B., C. J. Malloy, and A. Scherbina (2002). Differences of opinion and the cross section of stock returns. *The Journal of Finance 57*(5), 2113–2141.

Dybvig, P. and S. Ross (1985). Differential information and performance measurement using a security market line. *The Journal of Finance 40*(2), 383–399.

Easley, D. and M. O'Hara (2004). Information and the cost of capital. *The journal of finance 59*(4), 1553–1583.

Engelberg, J., R. D. McLean, and J. Pontiff (2018). Anomalies and news. *The Journal of Finance 73*(5), 1971–2001.

Eriksen, J. N. and N. S. Gronborg (2020). Asset pricing and fomc press conferences. Working paper.

Fama, E. F. and K. R. French (2004, September). The capital asset pricing model: Theory and evidence. *Journal of Economic Perspectives 18*(3), 25–46.

Fama, E. F. and K. R. French (2007). Disagreement, tastes, and asset prices. *Journal of Financial Economics 83*(3), 667 – 689.

Fama, E. F. and J. D. MacBeth (1973). Risk, return, and equilibrium: Empirical tests. *Journal of political economy 81*(3), 607–636.

Farboodi, M. and L. Veldkamp (2017). Long run growth of financial technology. Technical report, National Bureau of Economic Research.

Ferson, W. E. and C. R. Harvey (1991). The variation of economic risk premiums. *Journal of political economy 99*(2), 385–415.

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

Friend, I. and M. Blume (1970). Measurement of portfolio performance under uncertainty. *The American Economic Review*, 561–575.

Graham, J. R. and C. R. Harvey (2001, May). The theory and practice of corporate finance: evidence from the field. *Journal of Financial Economics 60*(2-3), 187–243.

Greene, W. H. (2003). *Econometric Analysis.* Pearson Education.

Grossman, S. J. and J. E. Stiglitz (1980). On the impossibility of informationally efficient markets. *The American Economic Review 70*(3), pp.393–408.

Hansen, L. P. and S. F. Richard (1987). The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models. *Econometrica: Journal of the Econometric Society*, 587–613.

He, H. and J. Wang (1995). Differential information and dynamic behavior of stock trading volume. *Review of Financial Studies 8*(4), 919–72.

Hendershott, T., D. Livdan, and D. Rösch (2018). Asset pricing: A tale of night and day.

Hong, H. and D. A. Sraer (2016). Speculative betas. *The Journal of Finance*.

Jagannathan, R. and Z. Wang (1996). The conditional capm and the cross-section of expected returns. *The Journal of Finance 51*(1), 3–53.

Jagannathan, R. and Z. Wang (1998). An asymptotic theory for estimating beta-pricing models using cross-sectional regression. *The Journal of Finance 53*(4), 1285–1309.

Jensen, M. C. (1972). Capital markets: Theory and evidence. *Bell Journal of Economics and Management Science 3*(2), 357–398.

Kan, R., C. Robotti, and J. Shanken (2013). Pricing model performance and the two-pass cross-sectional regression methodology. *The Journal of Finance 68*(6), 2617–2649.

Knowles, A. and J. Yin (2017). Anisotropic local laws for random matrices. *Probability Theory and Related Fields 169*(1), 257–352.

Kumar, A. (2009). Who gambles in the stock market? *The Journal of Finance 64*(4), 1889–1933.

Levi, Y. and I. Welch (2017). Best practice for cost-of-capital estimates. *Journal of Financial and Quantitative Analysis 52*(2), 427–463.

Lewellen, J. and S. Nagel (2006, November). The conditional CAPM does not explain asset-pricing anomalies. *Journal of Financial Economics 82*(2), 289–314.

Lintner, J. (1969). The aggregation of investor's diverse judgments and preferences in purely competitive security markets. *Journal of financial and quantitative analysis 4*(4), 347–400.

Markowitz, H. (1952). Portfolio selection. *The journal of finance 7*(1), 77–91.

Martin, I. (2017). What is the expected return on the market? *Quarterly Journal of Economics*, 367–433.

Martin, I. and S. Nagel (2020). Market efficiency in the age of big data. *Working Paper*.

Merton, R. (1973). An intertemporal capital asset pricing model. *Econometrica 41*(5), 867–87.

Merton, R. C. (1987). A simple model of capital market equilibrium with incomplete information. *The journal of finance 42*(3), 483–510.

Milgrom, P. and N. Stokey (1982). Information, trade and common knowledge. *Journal of Economic Theory 26*(1), 17–27.

Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*(3), 703–708.

Pesaran, M. H. and A. Timmermann (1995). Predictability of stock returns: Robustness and economic significance. *The Journal of Finance 50*(4), 1201–1228.

Roll, R. (1977). A critique of the asset pricing theory's tests part i: On past and potential testability of the theory. *Journal of financial economics 4*(2), 129–176.

Roll, R. (1978). Ambiguity when performance is measured by the securities market line. *The Journal of finance 33*(4), 1051–1069.

Roll, R. and S. A. Ross (1994). On the cross-sectional relation between expected returns and betas. *The Journal of Finance 49*(1), 101–121.

Savor, P. and M. Wilson (2014). Asset pricing: A tale of two days. *Journal of Financial Economics 113*(2), 171–201.

Shanken, J. (1992). On the estimation of beta-pricing models. *The review of financial studies 5*(1), 1–33.

Shanken, J. and G. Zhou (2007). Estimating and testing beta pricing models: Alternative methods and their performance in simulations. *Journal of Financial Economics 84*(1), 40–86.

Shiller, R. J. (1981, June). Do stock prices move too much to be justified by subsequent changes in dividends? *American Economic Review 71*(3), 421–36.

Silverstein, J. W. (1989). On the eigenvectors of large dimensional sample covariance matrices. *Journal of Multivariate Analysis 30*(1), 1 – 16.

Stambaugh, R. F. (1982). On the exclusion of assets from tests of the two-parameter model: A sensitivity analysis. *Journal of financial economics 10*(3), 237–268.

Van Nieuwerburgh, S. and L. Veldkamp (2010). Information acquisition and under-diversification. *The Review of Economic Studies 77*(2), 779–805.

Vasicek, O. A. (1973). A note on using cross-sectional information in bayesian estimation of security betas. *The Journal of Finance 28*(5), 1233–1239.

Wang, J. (1994, February). A model of competitive stock trading volume. *Journal of Political Economy 102*(1), 127–68.

Xi, H., F. Yang, and J. Yin (2020, 04). Convergence of eigenvector empirical spectral distribution of sample covariance matrices. *Ann. Statist. 48*(2), 953–982.

A Tables and Figures

Variable	Average	St. Dev.	P5%	P25%	Median	P75%	P95%
Mkt. ex. ret.	0.0985	0.1616	-0.2290	0.0353	0.1158	0.1837	0.3063
Exp. mkt. ex. ret.	0.1270	0.0281	0.0832	0.1088	0.1210	0.1445	0.1835
Realized betas	1.1826	1.2375	-0.5201	0.4631	1.0561	1.7479	3.2557
Consensus betas	1.0977	1.3389	-0.7952	0.2527	0.9886	1.7866	3.3490
Dispersion betas	0.9076	1.3842	-1.1510	0.1044	0.7913	1.6417	3.3599

Table 1: This table presents summary statistics for market excess returns $\widetilde{R}^e_{\mathbf{M}}$, consensus expected market excess returns $\overline{\mathbb{E}}[\widetilde{R}^e_{\mathbf{M}}]$, realized betas $\widehat{\boldsymbol{\beta}}$, consensus betas $\boldsymbol{\beta}^{\mathcal{C}}$, and dispersion betas $\boldsymbol{\beta}^{\mathcal{D}}$, from 2002/12/31 to 2018/9/28 (190 months). All the numbers are at 1-year horizon.

	μ	$\widehat{\alpha}$	$\widehat{\beta}$	$\beta^{\mathcal{C}}$	$\beta^{\mathcal{D}}$	$\widehat{\sigma}$	SR
P1	0.1009***	0.1587***	-0.1789***	0.5605***	0.8447***	0.1348	0.75
	(4.94)	(8.44)	(-2.68)	(9.80)	(20.56)		
P2	0.0917***	0.0759***	0.5092***	0.7566***	0.9110***	0.1341	0.68
	(4.56)	(16.74)	(12.48)	(22.84)	(36.66)		
P3	0.1004***	0.0360***	0.9796***	1.0442***	0.9976***	0.1480	0.68
	(4.60)	(8.74)	(27.59)	(29.15)	(26.30)		
P4	0.1063***	0.0054	1.5374***	1.3381***	1.0757***	0.1856	0.57
	(3.91)	(0.49)	(56.48)	(24.27)	(26.75)		
P5	0.1267***	0.0167	2.7933***	1.7411***	1.3645***	0.2452	0.52
	(3.53)	(0.48)	(33.97)	(24.90)	(29.44)		

Table 2: This table presents value-weighted averages for excess returns μ, CAPM alphas $\widehat{\alpha}$, CAPM betas $\widehat{\beta}$, consensus betas $\beta^{\mathcal{C}}$, dispersion betas $\beta^{\mathcal{D}}$, volatilities $\widehat{\sigma}$, and Sharpe ratios SR on five beta-sorted portfolios, from 2002/12/31 to 2018/9/28 (190 months). t-statistics, presented in parentheses, are computed using the Newey and West (1987) adjustment with 4 lags. Statistical significance is denoted by $^{*}p < .1$, $^{**}p < .05$, or $^{***}p < .01$.

(a) $\widetilde{R}^e_{p,t} = \widehat{\alpha}_p + \widehat{\beta}_p \widetilde{R}^e_{\mathbf{M},t} + \varepsilon_t$

	$\widehat{\alpha}_p$	$\widehat{\beta}_p$			Adj.R^2	N
P1	0.0255***	0.7647***			0.8404	190
	(2.80)	(13.48)				
P2	0.0130*	0.7991***			0.9269	190
	(1.89)	(19.38)				
P3	0.0136***	0.8805***			0.9240	190
	(2.61)	(23.34)				
P4	-0.0016	1.0943***			0.9075	190
	(-0.19)	(15.82)				
P5	-0.0106	1.3926***			0.8423	190
	(-0.62)	(11.32)				

(b) $\widetilde{R}^e_{p,t} = \widehat{\alpha}_p + \widehat{\beta}_p \widetilde{R}^e_{\mathbf{M},t} + a_{\mathcal{C}} \beta^{\mathcal{C}}_{p,t} + a_{\mathcal{D}} \beta^{\mathcal{D}}_{p,t} + \varepsilon_t$

	$\widehat{\alpha}_p$	$\widehat{\beta}_p$	$a_{\mathcal{C}}$	$a_{\mathcal{D}}$	Adj.R^2	N
P1	0.0160	0.7738***	-0.0198	0.0234	0.8445	190
	(0.71)	(13.93)	(-1.52)	(0.96)		
P2	0.0096	0.7974***	-0.0029	0.0063	0.9262	190
	(0.42)	(19.21)	(-0.13)	(0.33)		
P3	0.0335	0.8688***	0.0270	-0.0470***	0.9302	190
	(1.46)	(24.19)	(1.30)	(-3.37)		
P4	-0.0074	1.0985***	-0.0128	0.0209	0.9084	190
	(-0.13)	(15.63)	(-0.64)	(0.50)		
P5	0.1608**	1.3856***	-0.0539**	-0.0563	0.8593	190
	(2.56)	(12.95)	(-2.46)	(-1.58)		

Table 3: This table presents results from time-series regressions for five beta-sorted portfolios. The regressions are given at the top of each panel. The data is from 2002/12/31 to 2018/9/28 (190 months). t-statistics, presented in parentheses, are computed using the Newey and West (1987) adjustment with 4 lags. Statistical significance is denoted by $^{*}p < .1$, $^{**}p < .05$, or $^{***}p < .01$.

a_0	$a_{\mathcal{C}}$	$a_{\mathcal{D}}$	Adj.R^2	N
0.8292***	0.2606***	0.0531***	0.1083	410
(24.25)	(11.29)	(4.30)		

Table 4: This table presents results from Fama and MacBeth (1973) regressions of realized betas $\widehat{\boldsymbol{\beta}}$ on consensus betas $\boldsymbol{\beta}^{\mathcal{C}}$ and dispersion betas $\boldsymbol{\beta}^{\mathcal{D}}$, from 2002/12/31 to 2018/9/28 (190 months). a_0 is the time-series average of the intercept coefficient. $a_{\mathcal{C}}$ is the time-series average of the coefficient on $\boldsymbol{\beta}^{\mathcal{C}}$. $a_{\mathcal{D}}$ is the time-series average of the coefficient on $\boldsymbol{\beta}^{\mathcal{D}}$. The last two columns compute time-series averages of the adjusted R^2 and the number of firms in cross-sectional regressions. t-statistics, presented in parentheses, are computed using the Newey and West (1987) adjustment with 4 lags. Statistical significance is denoted by $^*p < .1$, $^{**}p < .05$, or $^{***}p < .01$.

a_0	a_R	$a_{\mathcal{C}}$	$a_{\mathcal{D}}$	Adj.R^2	N
0.1204***	0.0182*			0.0225	410
(5.27)	(1.75)				
0.1266***	0.0217**	-0.0098***		0.0307	410
(5.58)	(2.11)	(-2.96)			
0.1251***	0.0190*		-0.0063***	0.0255	410
(5.61)	(1.85)		(-2.93)		
0.1311***	0.0224**	-0.0095***	-0.0063***	0.0336	410
(5.91)	(2.20)	(-2.87)	(-2.92)		

Table 5: This table presents results from Fama and MacBeth (1973) regressions of mean excess returns $\mathbb{E}[\widetilde{\mathbf{R}}^e]$ on realized betas $\widehat{\boldsymbol{\beta}}$, consensus betas $\boldsymbol{\beta}^{\mathcal{C}}$ and dispersion betas $\boldsymbol{\beta}^{\mathcal{D}}$, from 2002/12/31 to 2018/9/28 (190 months). a_0 is the time-series average of the intercept coefficient. a_R is the time-series average of the coefficient on $\widehat{\boldsymbol{\beta}}$. $a_{\mathcal{C}}$ is the time-series average of the coefficient on $\boldsymbol{\beta}^{\mathcal{C}}$. $a_{\mathcal{D}}$ is the time-series average of the coefficient on $\boldsymbol{\beta}^{\mathcal{D}}$. The last two columns compute time-series averages of the adjusted R^2 and the number of firms in cross-sectional regressions. t-statistics, presented in parentheses, are computed using the Newey and West (1987) adjustment with 4 lags. Statistical significance is denoted by $^*p < .1$, $^{**}p < .05$, or $^{***}p < .01$.

Const.	Slope	Adj.R^2	N
-0.0354**	0.1792**	0.1274	190
(-2.22)	(2.00)		
-0.0286***	0.1525**	0.1960	190
(-4.04)	(2.30)		

Table 6: This table presents results from regressions of the excess returns of the $\mathcal{C}$ and $\mathcal{D}$ factors on the excess returns of the market. We construct the factor $\mathcal{C}$ ($\mathcal{D}$) by ranking stocks according to their $\beta^{\mathcal{C}}$ ($\beta^{\mathcal{D}}$) and taking a long position in high-beta stocks financed by a short position in low-beta stocks. The data is from 2002/12/31 to 2018/9/28 (190 months). t-statistics, presented in parentheses, are computed using the Newey and West (1987) adjustment with 4 lags. Statistical significance is denoted by $^{*}p < .1$, $^{**}p < .05$, or $^{***}p < .01$.

(a) AQR BAB:

$$\widetilde{R}^e_{\mathrm{BAB},t} = \widehat{\alpha}_{\mathrm{BAB}} + \widehat{\beta}_{\mathbf{M}}\widetilde{R}^e_{\mathbf{M},t} + \widehat{\beta}_{\mathcal{C}}\widetilde{R}^e_{\mathcal{C},t} + \widehat{\beta}_{\mathcal{D}}\widetilde{R}^e_{\mathcal{D},t} + \varepsilon_t$$

$\widehat{\alpha}_{\mathrm{BAB}}$	$\widehat{\beta}_{\mathbf{M}}$	$\widehat{\beta}_{\mathcal{C}}$	$\widehat{\beta}_{\mathcal{D}}$	Adj.R^2	N
0.0377*	0.4744***			0.3241	190
(1.71)	(3.93)				
-0.0028	0.6870***	-0.3580**	-0.9726***	0.5111	190
(-0.21)	(9.31)	(-2.20)	(-4.14)		

(b) In-house BAB:

$$\widetilde{R}^e_{\mathrm{BAB},t} = \widehat{\alpha}_{\mathrm{BAB}} + \widehat{\beta}_{\mathbf{M}}\widetilde{R}^e_{\mathbf{M},t} + \widehat{\beta}_{\mathcal{C}}\widetilde{R}^e_{\mathcal{C},t} + \widehat{\beta}_{\mathcal{D}}\widetilde{R}^e_{\mathcal{D},t} + \varepsilon_t$$

$\widehat{\alpha}_{\mathrm{BAB}}$	$\widehat{\beta}_{\mathbf{M}}$	$\widehat{\beta}_{\mathcal{C}}$	$\widehat{\beta}_{\mathcal{D}}$	Adj.R^2	N
0.0521**	0.9557***			0.5231	190
(2.16)	(4.48)				
-0.0185	1.3207***	-1.1975***	-0.9860***	0.7801	190
(-1.59)	(14.03)	(-7.28)	(-2.90)		

(c) In-house BAB:

$$\widetilde{R}^e_{\mathrm{BAB},t} = \widehat{\alpha}_{\mathrm{BAB}} + \widehat{\beta}_{\mathrm{BAB}}\widetilde{R}^e_{\mathbf{M},t} + a_{\mathcal{C}}\beta^{\mathcal{C}}_{\mathrm{BAB},t} + a_{\mathcal{D}}\beta^{\mathcal{D}}_{\mathrm{BAB},t} + \varepsilon_t$$

$\widehat{\alpha}_p$	$\widehat{\beta}_p$	$a_{\mathcal{C}}$	$a_{\mathcal{D}}$	Adj.R^2	N
0.0521**	0.9557***			0.5231	190
(2.16)	(4.48)				
0.0367	0.9482***	0.0534	0.0004	0.5402	190
(0.66)	(4.00)	(1.14)	(0.01)		

Table 7: This table presents results from regressions of the excess returns of the AQR BAB strategy (panel a) and of our in-house BAB strategy (panel b) on the excesss returns of the market and the excess returns of the $\mathcal{C}$ and $\mathcal{D}$ factors. We construct the factor $\mathcal{C}$ ($\mathcal{D}$) by ranking stocks according to their $\beta^{\mathcal{C}}$ ($\beta^{\mathcal{D}}$) and taking a long position in high-beta stocks financed by a short position in low-beta stocks. In panel (c) we control for $\beta^{\mathcal{C}}_{\mathrm{BAB}}$ and $\beta^{\mathcal{D}}_{\mathrm{BAB}}$ directly, since we have these betas for every stock involved in the strategy. The data is from 2002/12/31 to 2018/9/28 (190 months). t-statistics, presented in parentheses, are computed using the Newey and West (1987) adjustment with 4 lags. Statistical significance is denoted by $^{*}p < .1$, $^{**}p < .05$, or $^{***}p < .01$.

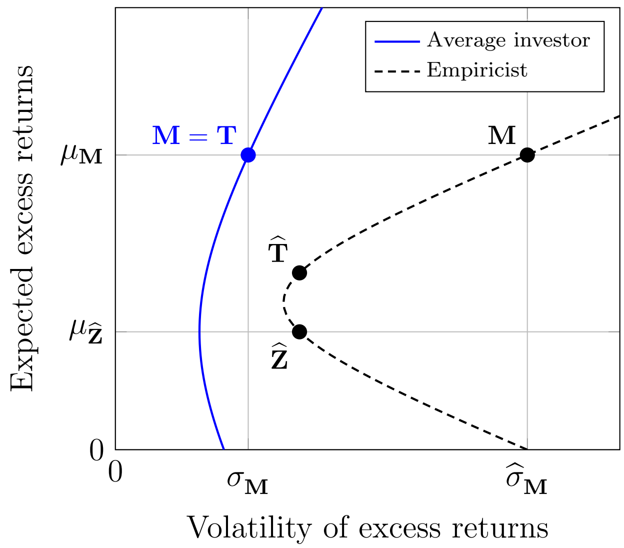

Figure 1: **CAPM rejection.** This figure compares the minimum-variance set under average unconditional beliefs (solid line) with the minimum-variance set of the empiricist (dashed line). For the average investor, the market portfolio is the tangency portfolio ($\mathbf{M} = \mathbf{T}$), but for the empiricist $\mathbf{M}$ moves upward on the minimum-variance set and is not mean-variance efficient. For the empiricist, $\widehat{\mathbf{Z}}$ is the zero-beta portfolio that has zero systematic risk.

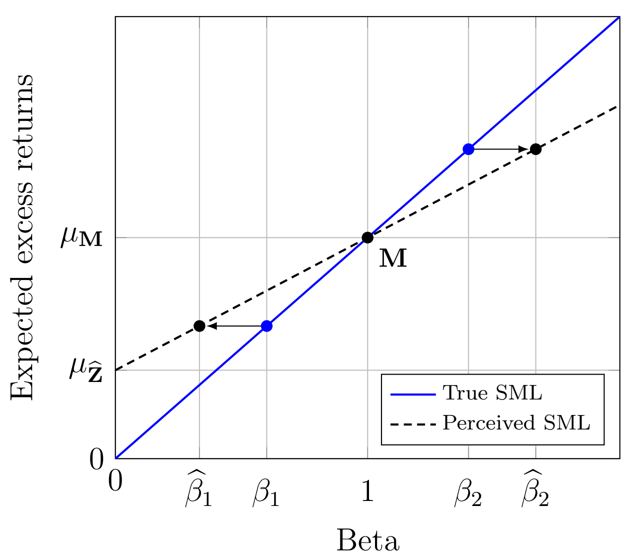

Figure 2: **CAPM distortion.** This figure illustrates the main result of the paper. The perceived SML is flatter than the actual SML in equilibrium. The dashed line and the solid line show the true and perceived SML. $\mathbf{M}$ represents the unconditional market portfolio.

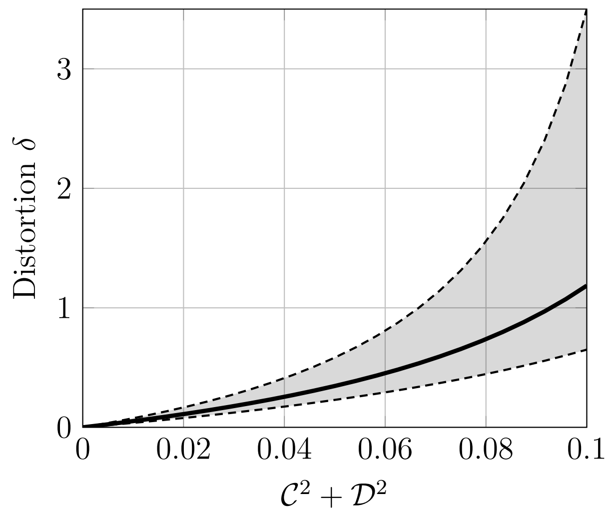

Figure 3: This figure plots the empirically plausible range for the distortion δ. The shaded area shows the 90 percent confidence region for δ based on Eq. (48) and a 90 percent confidence range for the intercept a: $a \in [0.78, 0.88]$. The distortion is plotted as a function of the informational distance between investors and the empiricist, $\mathcal{C}^2 + \mathcal{D}^2$. The data is from 2002/12/31 to 2018/9/28 (190 months).

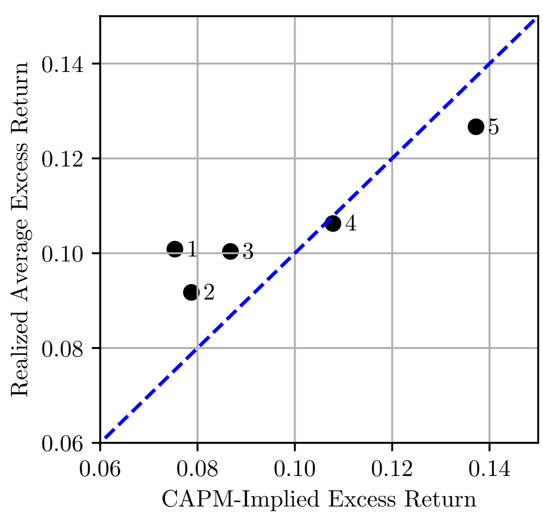

Figure 4: This figure plots the CAPM-implied excess returns versus realized average excess returns (value-weighted) on five beta-sorted portfolios, with the smallest beta stocks in portfolio 1 and the largest beta stocks in portfolio 5. The data is from 2002/12/31 to 2018/9/28 (190 months).

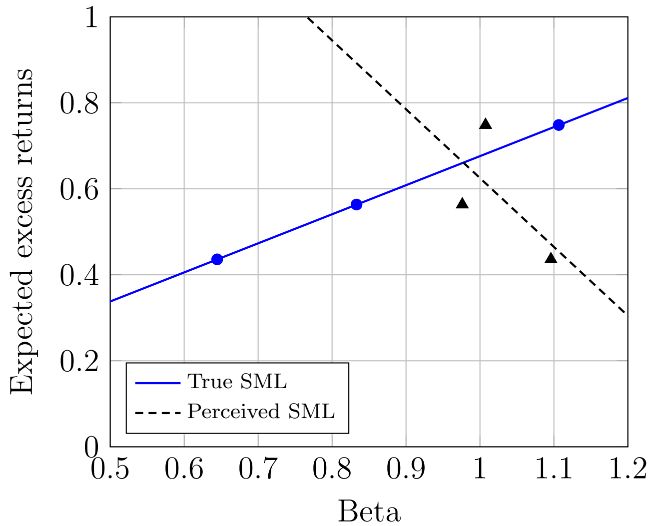

Figure 5: **Size Effects.** This figure illustrates the true SML (solid line) and observed SML (dashed line) when stocks are in heterogeneous supplies. This illustrative economy has three assets with factor loadings, $\phi_1 > \phi_2 > \phi_3 > 0$, and with supplies in the market portfolio, $0 < M_1 < M_2 < M_3$.

B Appendix (Proofs)

B.1 Proof of Proposition 1

We start by conjecturing a linear price function of the form:

$$\widetilde{\mathbf{P}} = \mathbf{1}D + \underbrace{\begin{bmatrix} \xi_{0,11} & \xi_{0,12} & \cdots & \xi_{0,1N} \\ \xi_{0,21} & \xi_{0,22} & \cdots & \xi_{0,2N} \\ \vdots & \vdots & \ddots & \\ \xi_{0,N1} & \xi_{0,N2} & \cdots & \xi_{0,NN} \end{bmatrix}}_{\boldsymbol{\xi}_0} \mathbf{M} + \underbrace{\begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_N \end{bmatrix}}_{\boldsymbol{\lambda}} \widetilde{F} + \underbrace{\begin{bmatrix} \xi_{11} & \xi_{12} & \cdots & \xi_{1N} \\ \xi_{21} & \xi_{22} & \cdots & \xi_{2N} \\ \vdots & \vdots & \ddots & \\ \xi_{N1} & \xi_{N2} & \cdots & \xi_{NN} \end{bmatrix}}_{\boldsymbol{\xi}} \widetilde{\mathbf{m}}, \tag{B.1}$$

where the undetermined coefficients multiplying the variables $\mathbf{M}$, $\widetilde{F}$, and $\widetilde{\mathbf{m}}$ will be determined by the market clearing condition.

Any investor i has two sources of information gathered in $\mathscr{F}^i$: (i) one private signal $\widetilde{V}^i$ about $\widetilde{F}$ and (ii) N public prices. We isolate the informational part of public prices:

$$\widetilde{\mathbf{P}}^a \equiv \widetilde{\mathbf{P}} - \mathbf{1}D - \boldsymbol{\xi}_0 \mathbf{M} = \boldsymbol{\lambda}\widetilde{F} + \boldsymbol{\xi}\widetilde{\mathbf{m}}, \tag{B.2}$$

and stack all information of investor i, both private and public, into a single vector

$$\begin{bmatrix} \widetilde{\mathbf{P}}^a \\ V^i \end{bmatrix} = \begin{bmatrix} \boldsymbol{\lambda} \\ 1 \end{bmatrix} \widetilde{F} + \begin{bmatrix} \boldsymbol{\xi} & \mathbf{0}_{N\times 1} \\ \mathbf{0}_{1\times N} & 1 \end{bmatrix} \begin{bmatrix} \widetilde{\mathbf{m}} \\ \tilde{v}^i \end{bmatrix} \equiv \mathbf{H}\widetilde{F} + \boldsymbol{\Theta} \begin{bmatrix} \widetilde{\mathbf{m}} \\ \tilde{v}^i \end{bmatrix}, \tag{B.3}$$

where the vector of noise in the signals, $[\widetilde{\mathbf{m}}' \ \tilde{v}^i]'$, is jointly Gaussian with covariance matrix:

$$\mathbf{C} = \begin{bmatrix} \tau_m^{-1}\mathbf{I} & \mathbf{0}_{N\times 1} \\ \mathbf{0}_{1\times N} & \tau_v^{-1} \end{bmatrix}. \tag{B.4}$$

Applying standard projection techniques we define the precision of the last term in (B.3):

$$\mathbf{r} \equiv (\boldsymbol{\Theta}\mathbf{C}\boldsymbol{\Theta}')^{-1} = \begin{bmatrix} \tau_m(\boldsymbol{\xi}\boldsymbol{\xi}')^{-1} & \mathbf{0}_{N\times 1} \\ \mathbf{0}_{1\times N} & \tau_v \end{bmatrix}, \tag{B.5}$$

and obtain that an investor i's total precision on the common factor satisfies

$$\tau \equiv \mathrm{Var}[\widetilde{F}|\mathscr{F}^i]^{-1} = \tau_F + \mathbf{H}'\mathbf{r}\mathbf{H} = \tau_F + \tau_v + \tau_m \boldsymbol{\lambda}'(\boldsymbol{\xi}\boldsymbol{\xi}')^{-1}\boldsymbol{\lambda}. \tag{B.6}$$

The precision τ is the same across investors. Furthermore, an investor i's expectation of $\widetilde{F}$ satisfies (both $\widetilde{\mathbf{P}}^a$ and $\widetilde{V}^i$ have zero unconditional means):

$$\mathbb{E}[\widetilde{F}|\mathscr{F}^i] = \frac{1}{\tau}\mathbf{H}'\mathbf{r} \begin{bmatrix} \widetilde{\mathbf{P}}^a \\ \widetilde{V}^i \end{bmatrix} = \frac{1}{\tau} \begin{bmatrix} \boldsymbol{\lambda}'(\boldsymbol{\xi}\boldsymbol{\xi}')^{-1}\tau_m & \tau_v \end{bmatrix} \begin{bmatrix} \widetilde{\mathbf{P}}^a \\ \widetilde{V}^i \end{bmatrix} \tag{B.7}$$

$$= \frac{1}{\tau} \left(\tau_m \boldsymbol{\lambda}'(\boldsymbol{\xi}\boldsymbol{\xi}')^{-1}\boldsymbol{\lambda}\widetilde{F} + \tau_m \boldsymbol{\lambda}'(\boldsymbol{\xi}\boldsymbol{\xi}')^{-1}\boldsymbol{\xi}\widetilde{\mathbf{m}} + \tau_v \widetilde{F} + \tau_v \tilde{v}^i \right). \tag{B.8}$$

Using the definition of the total precision (B.6), it follows that average market expectation of future dividends is

$$\overline{\mathbb{E}}[\widetilde{\mathbf{D}}] \equiv \int_i \mathbb{E}[\widetilde{\mathbf{D}}|\mathscr{F}^i]di = \mathbf{1}D + \mathbf{\Phi}\frac{1}{\tau}\left[(\tau - \tau_F)\widetilde{F} + \tau_m \boldsymbol{\lambda}'(\boldsymbol{\xi}\boldsymbol{\xi}')^{-1}\boldsymbol{\xi}\widetilde{\mathbf{m}}\right], \tag{B.9}$$

and individual expectations are:

$$\mathbb{E}[\widetilde{\mathbf{D}}|\mathscr{F}^i] = \overline{\mathbb{E}}[\widetilde{\mathbf{D}}] + \mathbf{\Phi}\frac{\tau_v}{\tau}\tilde{v}^i. \tag{B.10}$$

For each agent i, the uncertainty about future dividends is

$$\mathbf{\Sigma} \equiv \mathrm{Var}[\widetilde{\mathbf{D}}|\mathscr{F}^i] = \frac{1}{\tau}\mathbf{\Phi}\mathbf{\Phi}' + \frac{1}{\tau_\epsilon}\mathbf{I}. \tag{B.11}$$

Because agents hold mean-variance portfolios, the market-clearing condition implies:

$$\widetilde{\mathbf{P}} = \overline{\mathbb{E}}[\widetilde{\mathbf{D}}] - \gamma\mathbf{\Sigma}(\mathbf{M} - \widetilde{\mathbf{m}}) \tag{B.12}$$

$$= \mathbf{1}D - \gamma\left(\frac{1}{\tau}\mathbf{\Phi}\mathbf{\Phi}' + \frac{1}{\tau_\epsilon}\mathbf{I}\right)\mathbf{M} + \mathbf{\Phi}\frac{\tau - \tau_F}{\tau}\widetilde{F} + \left[\mathbf{\Phi}\frac{\tau_m}{\tau}(\boldsymbol{\xi}^{-1}\boldsymbol{\lambda})' + \gamma\left(\frac{1}{\tau}\mathbf{\Phi}\mathbf{\Phi}' + \frac{1}{\tau_\epsilon}\mathbf{I}\right)\right]\widetilde{\mathbf{m}}, \tag{B.13}$$

where we have used the simplification

$$\boldsymbol{\lambda}'(\boldsymbol{\xi}\boldsymbol{\xi}')^{-1}\boldsymbol{\xi} = (\boldsymbol{\xi}^{-1}\boldsymbol{\lambda})'. \tag{B.14}$$

The initial price conjecture then yields the following fixed point solution:

$$\boldsymbol{\xi}_0 = -\gamma\left(\frac{1}{\tau}\mathbf{\Phi}\mathbf{\Phi}' + \frac{1}{\tau_\epsilon}\mathbf{I}\right) \tag{B.15}$$

$$\boldsymbol{\lambda} = \mathbf{\Phi}\frac{\tau - \tau_F}{\tau} \tag{B.16}$$

$$\boldsymbol{\xi} = \mathbf{\Phi}\frac{\tau_m}{\tau}(\boldsymbol{\xi}^{-1}\boldsymbol{\lambda})' + \gamma\left(\frac{1}{\tau}\mathbf{\Phi}\mathbf{\Phi}' + \frac{1}{\tau_\epsilon}\mathbf{I}\right). \tag{B.17}$$

Multiply both sides of the last equation by $\boldsymbol{\xi}^{-1}\boldsymbol{\lambda}$ (to the right):

$$\boldsymbol{\lambda} = \mathbf{\Phi}\frac{\tau_m}{\tau}(\boldsymbol{\xi}^{-1}\boldsymbol{\lambda})'\boldsymbol{\xi}^{-1}\boldsymbol{\lambda} + \gamma\left(\frac{1}{\tau}\mathbf{\Phi}\mathbf{\Phi}' + \frac{1}{\tau_\epsilon}\mathbf{I}\right)\boldsymbol{\xi}^{-1}\boldsymbol{\lambda}, \tag{B.18}$$

and recognize that $\tau_m(\boldsymbol{\xi}^{-1}\boldsymbol{\lambda})'\boldsymbol{\xi}^{-1}\boldsymbol{\lambda} = \tau_m\boldsymbol{\lambda}'(\boldsymbol{\xi}\boldsymbol{\xi}')^{-1}\boldsymbol{\lambda} = \tau - \tau_F - \tau_v$ (from Eq. B.6), which can be replaced above, together with the solution for $\boldsymbol{\lambda}$ to obtain:

$$\mathbf{\Phi}\frac{\tau_v}{\tau} = \gamma\left(\frac{1}{\tau}\mathbf{\Phi}\mathbf{\Phi}' + \frac{1}{\tau_\epsilon}\mathbf{I}\right)\boldsymbol{\xi}^{-1}\boldsymbol{\lambda}, \tag{B.19}$$

which leads to an equation for $\boldsymbol{\xi}^{-1}\boldsymbol{\lambda}$:

$$\boldsymbol{\xi}^{-1}\boldsymbol{\lambda} = \frac{\tau_v}{\gamma\tau}\left(\frac{1}{\tau}\mathbf{\Phi}\mathbf{\Phi}' + \frac{1}{\tau_\epsilon}\mathbf{I}\right)^{-1}\mathbf{\Phi} = \frac{\tau_v\tau_\epsilon}{\gamma(\tau + \tau_\epsilon)}\mathbf{\Phi}. \tag{B.20}$$

The second equality results from the fact that $\mathbf{\Phi}$ is an eigenvector of the matrix $\frac{1}{\tau}\mathbf{\Phi}\mathbf{\Phi}' + \frac{1}{\tau_\epsilon}\mathbf{I}$, and the corresponding eigenvalue of $\mathbf{\Phi}$ is $1/\tau + 1/\tau_\epsilon$. (The matrix $\mathbf{\Phi}\mathbf{\Phi}' + \frac{\tau}{\tau_\epsilon}\mathbf{I}$ has two distinct eigenvalues: $1/\tau + 1/\tau_\epsilon$, of multiplicity 1; and $1/\tau_\epsilon$, of multiplicity $N-1$.)

We adopt the following notation:

$$\boldsymbol{\xi}^{-1}\boldsymbol{\lambda} = \frac{\sqrt{\tau_P}}{\sqrt{\tau_m}}\boldsymbol{\Phi}, \tag{B.21}$$

where τ_P is an unknown positive scalar such that $\frac{\sqrt{\tau_P}}{\sqrt{\tau_m}} \equiv \frac{\tau_v\tau_\epsilon}{\gamma(\tau+\tau_\epsilon)}$. Replacing Eq. (B.21) in Eq. (B.6) yields the total precision τ as a function of this scalar:

$$\tau = \tau_F + \tau_v + \tau_P, \tag{B.22}$$

which leads to a cubic equation in τ_P:

$$\tau_P(\tau_F + \tau_v + \tau_P + \tau_\epsilon)^2 = \frac{\tau_m\tau_\epsilon^2\tau_v^2}{\gamma^2}. \tag{B.23}$$

The discriminant of this cubic equation is strictly negative and thus the equation has a unique real root. Since it cannot have a negative root (the right hand side is strictly positive), it follows that τ_P is a unique positive scalar. Eq. (B.21) can now be replaced in the fixed point solution (B.17) to obtain the undetermined coefficients $\boldsymbol{\xi}$:

$$\boldsymbol{\xi} = \frac{\gamma + \sqrt{\tau_m\tau_P}}{\tau}\boldsymbol{\Phi}\boldsymbol{\Phi}' + \frac{\gamma}{\tau_\epsilon}\mathbf{I}, \tag{B.24}$$

which completes the proof of Proposition 1. □

B.1.1 Proof of Corollary 1.1

From (B.11), we know that $\boldsymbol{\Sigma} = \frac{1}{\tau}\boldsymbol{\Phi}\boldsymbol{\Phi}' + \frac{1}{\tau_\epsilon}\mathbf{I}$. Thus we can write:

$$\boldsymbol{\Sigma}\mathbf{M} = \frac{\bar{\Phi}}{\tau}\boldsymbol{\Phi} + \frac{1}{N\tau_\epsilon}\mathbf{1} \quad \text{and} \quad \sigma^2_{\mathbf{M}} = \mathbf{M}'\boldsymbol{\Sigma}\mathbf{M} = \frac{\bar{\Phi}^2}{\tau} + \frac{1}{N\tau_\epsilon}, \tag{B.25}$$

which yields

$$\boldsymbol{\beta} = \frac{\boldsymbol{\Sigma}\mathbf{M}}{\mathbf{M}'\boldsymbol{\Sigma}\mathbf{M}} = \frac{\frac{\bar{\Phi}^2}{\tau}\frac{\boldsymbol{\Phi}}{\bar{\Phi}} + \frac{1}{N\tau_\epsilon}\mathbf{1}}{\sigma^2_{\mathbf{M}}}. \tag{B.26}$$

This is a weighted average between $\mathbf{1}$ and $\boldsymbol{\Phi}/\bar{\Phi}$. Subtracting $\mathbf{1}$ on both sides yields (18). □

B.2 Proof of Proposition 2

For this proof we will make the following assumptions:

Assumption B.1. *There is no ex-ante proportionality relation between the unconditional market portfolio* $\mathbf{M}$ *and the vector of assets' loadings on the common factor* $\boldsymbol{\Phi}$*.*

Assumption B.2. $\mathbf{M}'\boldsymbol{\Phi} > 0$.

Assumption B.1 ensures that we keep the setup as general as possible, excluding pathological cases with an *exogenous* perfect relationship between stocks' market capitalizations and their exposure to the common factor. In our case, such an exogenous relation would occur when all the elements in the vector $\mathbf{\Phi}$ are equal, and thus all assets are identical. Assumption B.2 eliminates the uninteresting case $\mathbf{M}'\mathbf{\Phi} = 0$ (zero market exposure to the common factor), and is without loss of generality (if $\mathbf{M}'\mathbf{\Phi} < 0$, one can simply switch the sign of the common factor). In our case, since $\mathbf{M} = \mathbf{1}/N$, $\mathbf{M}'\mathbf{\Phi}$ represents the mean of the vector $\mathbf{\Phi}$ and equals $\bar{\Phi}$.

Setting $\mathbf{x} \equiv \widehat{\mathbf{\Sigma}}^{1/2}\mathbf{M}$ and $\mathbf{y} \equiv \widehat{\mathbf{\Sigma}}^{-1/2}\boldsymbol{\mu}$, we have $\sigma_{\mathbf{M}} = \|\mathbf{x}\|$ and $\sqrt{\boldsymbol{\mu}'\widehat{\mathbf{\Sigma}}^{-1}\boldsymbol{\mu}} = \|\mathbf{y}\|$, where $\|\cdot\|$ denotes the norm. The Cauchy-Schwartz inequality states that

$$\|\mathbf{x}\|\|\mathbf{y}\| \geq \mathbf{x}'\mathbf{y} = \mathbf{M}'\widehat{\mathbf{\Sigma}}^{1/2}\widehat{\mathbf{\Sigma}}^{-1/2}\boldsymbol{\mu} = \mu_{\mathbf{M}}, \tag{B.27}$$

where we have used the properties of symmetric positive-definite matrices for $\widehat{\mathbf{\Sigma}}$. Thus,

$$\frac{\mu_{\mathbf{M}}}{\sigma_{\mathbf{M}}} \leq \sqrt{\boldsymbol{\mu}'\widehat{\mathbf{\Sigma}}^{-1}\boldsymbol{\mu}}. \tag{B.28}$$

The relation (B.28) holds with equality *if and only if* $\mathbf{x}$ is proportional to $\mathbf{y}$, or

$$\boldsymbol{\mu} \propto \widehat{\mathbf{\Sigma}}\mathbf{M}. \tag{B.29}$$

Starting from the law of total variance (20) and replacing individual expectations from (B.10), we compute $\widehat{\mathbf{\Sigma}}$, which also proves Lemma 1 in the text:

$$\widehat{\mathbf{\Sigma}} = \mathbf{\Sigma} + \frac{\gamma^2}{\tau_m}\left(\frac{1}{\tau_\epsilon}\mathbf{\Sigma} + \frac{e_1}{\tau}\mathbf{\Phi}\mathbf{\Phi}'\right) + \frac{\tau_v}{\tau^2}\mathbf{\Phi}\mathbf{\Phi}', \tag{B.30}$$

where e_1 is the unique largest eigenvalue of $\mathbf{\Sigma}$:

$$e_1 = \frac{1}{\tau} + \frac{1}{\tau_\epsilon}. \tag{B.31}$$

By making use of Eq. (B.11), one can write $\widehat{\mathbf{\Sigma}}$ in two equivalent forms:

$$\widehat{\mathbf{\Sigma}} = c_1\mathbf{\Sigma} + c_2\mathbf{\Phi}\mathbf{\Phi}' \tag{B.32}$$

$$\widehat{\mathbf{\Sigma}} = c_3\mathbf{\Sigma} - c_4\mathbf{I} \tag{B.33}$$

where c_1, c_2, c_3, and c_4 are positive scalars:

$$c_1 = 1 + \frac{\gamma^2}{\tau_m\tau_\epsilon} > 0, \quad c_2 = \frac{\gamma^2 e_1}{\tau_m\tau} + \frac{\tau_v}{\tau^2} > 0, \tag{B.34}$$

and

$$c_3 = 1 + \frac{\gamma^2}{\tau_m\tau_\epsilon} + \frac{\gamma^2 e_1}{\tau_m} + \frac{\tau_v}{\tau} > 0, \quad c_4 = \frac{\gamma^2 e_1}{\tau_m\tau_\epsilon} + \frac{\tau_v}{\tau_\epsilon\tau} > 0. \tag{B.35}$$

Multiply Equations (B.32)-(B.33) with $\mathbf{M}$:

$$\widehat{\mathbf{\Sigma}}\mathbf{M} = c_1\mathbf{\Sigma}\mathbf{M} + c_2\bar{\Phi}\mathbf{\Phi} \tag{B.36}$$

$$\widehat{\mathbf{\Sigma}}\mathbf{M} = c_3\mathbf{\Sigma}\mathbf{M} - c_4\mathbf{M}. \tag{B.37}$$

Since $\boldsymbol{\mu} \propto \mathbf{\Sigma}\mathbf{M}$ (Definition 1), (B.29) and (B.36)-(B.37) imply that $\boldsymbol{\mu} \propto \mathbf{\Phi}$ and $\boldsymbol{\mu} \propto \mathbf{M}$. This implies $\mathbf{M} \propto \mathbf{\Phi}$, contradicting Assumption B.1. Thus, $\boldsymbol{\mu} \not\propto \widehat{\mathbf{\Sigma}}\mathbf{M}$ and empiricist's CAPM fails. □

B.3 Proof of Proposition 3

We prove first that $\mathbf{M}$ lies on empiricist's minimum-variance set. From Roll (1977, Corollary 6), we know that the betas of individual assets with respect to any portfolio are an exact linear function of individual expected excess returns *if and only if* the portfolio is minimum-variance. We can write

$$\widehat{\boldsymbol{\Sigma}}\mathbf{M} = \frac{1}{\gamma}\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} = \frac{1}{\gamma}(c_3\boldsymbol{\Sigma} - c_4\mathbf{I})\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} = \frac{c_3}{\gamma}\boldsymbol{\mu} - \frac{c_4}{N}\mathbf{1}, \tag{B.38}$$

where we have used the equilibrium relation $\boldsymbol{\mu} = \gamma\boldsymbol{\Sigma}\mathbf{M}$ for the first equality, Eq. (B.33) for the second equality, and $\mathbf{M} = \mathbf{1}/N$ for the third equality.

Using the definition of empiricist's betas in (22), it follows that empiricist's betas are an exact linear function of expected excess returns:

$$\widehat{\sigma}^2_{\mathbf{M}}\widehat{\boldsymbol{\beta}} = -\frac{c_4}{N}\mathbf{1} + \frac{c_3}{\gamma}\boldsymbol{\mu}, \tag{B.39}$$

which implies that $\mathbf{M}$ must lie on empiricist's minimum-variance set. One can further write

$$\boldsymbol{\mu} = \frac{c_4\gamma}{Nc_3}\mathbf{1} + \frac{\gamma\widehat{\sigma}^2_{\mathbf{M}}}{c_3}\widehat{\boldsymbol{\beta}}, \tag{B.40}$$

which is Eq. (23) in the text. Since $c_4\gamma/(Nc_3) > 0$, it follows that $\mu_{\widehat{\mathbf{Z}}} > 0$ and therefore $\mathbf{M}$ must lie above $\widehat{\mathbf{T}}$ on the upper limb of the minimum-variance set. □

B.4 Proof of Proposition 4

The starting point for the proof is Lemma 1, Eq. (28):

$$\widehat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma} + \underbrace{\frac{\gamma^2}{\tau_m}\left(\frac{1}{\tau_\epsilon}\boldsymbol{\Sigma} + \frac{e_1}{\tau}\boldsymbol{\Phi}\boldsymbol{\Phi}'\right)}_{\equiv \mathrm{Var}[\overline{\mathbb{E}}[\widetilde{\mathbf{R}}^e]]} + \underbrace{\frac{\tau_v}{\tau^2}\boldsymbol{\Phi}\boldsymbol{\Phi}'}_{\equiv \mathrm{Var}[\mathbb{E}^i[\widetilde{\mathbf{R}}^e] - \overline{\mathbb{E}}[\widetilde{\mathbf{R}}^e]]}, \tag{B.41}$$

which we pre-multiply with $\mathbf{M}'$ and post-multiply by $\mathbf{M}$:

$$\widehat{\sigma}^2_{\mathbf{M}} = \sigma^2_{\mathbf{M}} + \frac{\gamma^2}{\tau_m\tau_\epsilon}\sigma^2_{\mathbf{M}} + \frac{\gamma^2 e_1\bar{\Phi}^2}{\tau_m\tau} + \frac{\tau_v\bar{\Phi}^2}{\tau^2}, \tag{B.42}$$

from which we obtain $\mathcal{C}^2$ and $\mathcal{D}^2$ as defined in the text, Eqs. (30)-(31):

$$\mathcal{C}^2 \equiv \frac{\mathrm{Var}[\overline{\mathbb{E}}[\widetilde{R}^e_{\mathbf{M}}]]}{\widehat{\sigma}^2_{\mathbf{M}}} = \frac{\gamma^2}{\tau_m\tau_\epsilon\tau\widehat{\sigma}^2_{\mathbf{M}}}(\tau\sigma^2_{\mathbf{M}} + \tau_\epsilon e_1\bar{\Phi}^2) \tag{B.43}$$

$$\mathcal{D}^2 \equiv \frac{\mathrm{Var}[\mathbb{E}^i[\widetilde{R}^e_{\mathbf{M}}] - \overline{\mathbb{E}}[\widetilde{R}^e_{\mathbf{M}}]]}{\widehat{\sigma}^2_{\mathbf{M}}} = \frac{\tau_v\bar{\Phi}^2}{\tau^2\widehat{\sigma}^2_{\mathbf{M}}}. \tag{B.44}$$

Post-multiply (B.41) with $\mathbf{M}$ and divide by $\widehat{\sigma}^2_{\mathbf{M}}$ to obtain $\widehat{\boldsymbol{\beta}}$:

$$\widehat{\boldsymbol{\beta}} = \frac{\left(1 + \frac{\gamma^2}{\tau_m\tau_\epsilon}\right)\sigma^2_{\mathbf{M}}\boldsymbol{\beta} + \left(\frac{\gamma^2 e_1\bar{\Phi}^2}{\tau_m\tau} + \frac{\tau_v\bar{\Phi}^2}{\tau^2}\right)\frac{\boldsymbol{\Phi}}{\bar{\Phi}}}{\left(1 + \frac{\gamma^2}{\tau_m\tau_\epsilon}\right)\sigma^2_{\mathbf{M}} + \frac{\gamma^2 e_1\bar{\Phi}^2}{\tau_m\tau} + \frac{\tau_v\bar{\Phi}^2}{\tau^2}}, \tag{B.45}$$

which is a weighted average between $\boldsymbol{\beta}$ and $\boldsymbol{\Phi}/\bar{\Phi}$. Subtract $\mathbf{1}$ on both sides:

$$\widehat{\boldsymbol{\beta}} - \mathbf{1} = \frac{\left(1 + \frac{\gamma^2}{\tau_m \tau_\epsilon}\right) \sigma^2_{\mathbf{M}} (\boldsymbol{\beta} - \mathbf{1}) + \left(\frac{\gamma^2 e_1 \bar{\Phi}^2}{\tau_m \tau} + \frac{\tau_v \bar{\Phi}^2}{\tau^2}\right) \left(\frac{\boldsymbol{\Phi}}{\bar{\Phi}} - \mathbf{1}\right)}{\left(1 + \frac{\gamma^2}{\tau_m \tau_\epsilon}\right) \sigma^2_{\mathbf{M}} + \frac{\gamma^2 e_1 \bar{\Phi}^2}{\tau_m \tau} + \frac{\tau_v \bar{\Phi}^2}{\tau^2}}, \tag{B.46}$$

and use the definition of true betas from Corollary 1.1:

$$\widehat{\boldsymbol{\beta}} - \mathbf{1} = \frac{\left(1 + \frac{\gamma^2}{\tau_m \tau_\epsilon}\right) \sigma^2_{\mathbf{M}} (\boldsymbol{\beta} - \mathbf{1}) + \left(\frac{\gamma^2 e_1 \bar{\Phi}^2}{\tau_m \tau} + \frac{\tau_v \bar{\Phi}^2}{\tau^2}\right) \frac{\tau \sigma^2_{\mathbf{M}}}{\bar{\Phi}^2} (\boldsymbol{\beta} - \mathbf{1})}{\left(1 + \frac{\gamma^2}{\tau_m \tau_\epsilon}\right) \sigma^2_{\mathbf{M}} + \frac{\gamma^2 e_1 \bar{\Phi}^2}{\tau_m \tau} + \frac{\tau_v \bar{\Phi}^2}{\tau^2}}. \tag{B.47}$$

Thus:

$$\widehat{\boldsymbol{\beta}} - \mathbf{1} = \left[1 + \left(\frac{\tau \sigma^2_{\mathbf{M}}}{\bar{\Phi}^2} - 1\right) \left(\frac{\gamma^2 e_1 \bar{\Phi}^2}{\tau_m \tau \widehat{\sigma}^2_{\mathbf{M}}} + \mathcal{D}^2\right)\right] (\boldsymbol{\beta} - \mathbf{1}). \tag{B.48}$$

and thus the distortion δ is

$$\delta = \left(\frac{\tau \sigma^2_{\mathbf{M}}}{\bar{\Phi}^2} - 1\right) \left(\frac{\gamma^2 e_1 \bar{\Phi}^2}{\tau_m \tau \widehat{\sigma}^2_{\mathbf{M}}} + \mathcal{D}^2\right) = \left(\frac{\tau \sigma^2_{\mathbf{M}}}{\bar{\Phi}^2} - 1\right) \left(\mathcal{C}^2 \frac{\tau_\epsilon e_1 \bar{\Phi}^2}{\tau \sigma^2_{\mathbf{M}} + \tau_\epsilon e_1 \bar{\Phi}^2} + \mathcal{D}^2\right). \tag{B.49}$$

The last equality follows from (B.43). Since $\tau \sigma^2_{\mathbf{M}}/\bar{\Phi}^2 - 1 = \tau/(N \tau_\epsilon \bar{\Phi}^2) > 0$, δ is strictly positive.

Table A1 provides the limiting cases that are discussed in the paper: no idiosyncratic shocks, no risk aversion, infinitely precise private information, and no liquidity trades.

Variable	Case $\tau_\epsilon \to \infty$	Case $\gamma \to 0$	Case $\tau_v \to \infty$	Case $\tau_m \to \infty$
τ_P	$\frac{\tau_m \tau_v^2}{\gamma^2}$	∞	$\frac{\tau_m \tau_\epsilon^2}{\gamma^2}$	∞
τ	$\tau_F + \tau_v + \frac{\tau_m \tau_v^2}{\gamma^2}$	∞	∞	∞
$\boldsymbol{\Sigma}$	$\frac{1}{\tau} \boldsymbol{\Phi}\boldsymbol{\Phi}'$	$\frac{1}{\tau_\epsilon} \mathbf{I}$	$\frac{1}{\tau_\epsilon} \mathbf{I}$	$\frac{1}{\tau_\epsilon} \mathbf{I}$
$\sigma^2_{\mathbf{M}}$	$\frac{\bar{\Phi}^2}{\tau}$	$\frac{1}{N \tau_\epsilon}$	$\frac{1}{N \tau_\epsilon}$	$\frac{1}{N \tau_\epsilon}$
e_1	$\frac{1}{\tau}$	$\frac{1}{\tau_\epsilon}$	$\frac{1}{\tau_\epsilon}$	$\frac{1}{\tau_\epsilon}$
$\widehat{\boldsymbol{\Sigma}}$	$\left(\frac{1}{\tau} + \frac{\gamma^2 e_1}{\tau_m \tau} + \frac{\tau_v}{\tau^2}\right) \boldsymbol{\Phi}\boldsymbol{\Phi}'$	$\frac{1}{\tau_\epsilon} \mathbf{I}$	$\left(\frac{1}{\tau_\epsilon} + \frac{\gamma^2}{\tau_m \tau_\epsilon^2}\right) \mathbf{I}$	$\frac{1}{\tau_\epsilon} \mathbf{I}$
$\widehat{\sigma}^2_{\mathbf{M}}$	$\left(\frac{1}{\tau} + \frac{\gamma^2 e_1}{\tau_m \tau} + \frac{\tau_v}{\tau^2}\right) \bar{\Phi}^2$	$\frac{1}{N \tau_\epsilon}$	$\frac{1}{N \tau_\epsilon} + \frac{\gamma^2}{N \tau_m \tau_\epsilon^2}$	$\frac{1}{N \tau_\epsilon}$
$\mathcal{C}^2$	$\frac{\gamma^2 e_1 \bar{\Phi}^2}{\tau_m \tau \widehat{\sigma}^2_{\mathbf{M}}}$	0	$\frac{\gamma^2}{\gamma^2 + \tau_m \tau_\epsilon}$	0
$\mathcal{D}^2$	$\frac{\tau_v \bar{\Phi}^2}{\tau \widehat{\sigma}^2_{\mathbf{M}}}$	0	0	0
$\boldsymbol{\beta}$	$\frac{\boldsymbol{\Phi}}{\bar{\Phi}}$	$\mathbf{1}$	$\mathbf{1}$	$\mathbf{1}$
$\widehat{\boldsymbol{\beta}}$	$\frac{\boldsymbol{\Phi}}{\bar{\Phi}}$	$\mathbf{1}$	$\mathbf{1}$	$\mathbf{1}$
δ	0	0	$\frac{\gamma^2}{\gamma^2 + \tau_m \tau_\epsilon}$	0

Table A1: This table presents limiting cases for the distortion δ. In each case, we provide the limits for several key parameters and components of δ.

The monotonicity of δ with respect to γ, τ_m, and τ_ϵ can already be inferred from the limiting case $\tau_v \to \infty$. In this case, the distortion equals $\mathcal{C}^2$ and is strictly positive, although both true betas and empiricist's betas equal **1**. From Table A1, we notice that in the case $\tau_v \to \infty$ the distortion δ increases with the risk aversion, with the noise in idiosyncratic shocks to payoffs, and with the noise in assets' supplies. Figure A1 provides an illustration for the case of finite τ_v.

We can also characterize the monotonicity of δ with respect to γ, τ_m, and τ_ϵ for the case of diffuse priors, $\tau_F \equiv 0$. We first notice that the parameters γ and τ_m are not identified separately but only up to the ratio $h \equiv \gamma^2/\tau_m$. We thus investigate the monotonicity of δ in this ratio. In particular, we want to show that:

$$\frac{\mathrm{d}}{\mathrm{d}h}\delta > 0. \tag{B.50}$$

Differentiating δ with respect to h, and using that by the implicit function theorem:

$$\frac{\mathrm{d}}{\mathrm{d}h}\tau = -\frac{\tau_P(\tau_F + \tau_P + \tau_v + \tau_\epsilon)}{h(\tau_F + 3\tau_P + \tau_v + \tau_\epsilon)} < 0, \tag{B.51}$$

and simplifying using Eq. (17) whereby $\tau_v + \tau_P + \tau_\epsilon = \frac{\tau_\epsilon \tau_v}{\tau_P^{1/2} h^{1/2}}$, we obtain:

$$\frac{\mathrm{d}}{\mathrm{d}h}\delta = \frac{\tau_\epsilon^3 \left(\sqrt{h\tau_P} - \tau_v\right)^2}{(h\tau_P)^{3/2}} \frac{y}{h(3\tau_P + \tau_v + \tau_\epsilon)\left(\tau_\epsilon \phi^2(\tau_\epsilon(h + \tau_P + 2\tau_v) + 2h(\tau_P + \tau_v)) + (h + \tau_\epsilon)(\tau_P + \tau_v)^2\right)^2}, \tag{B.52}$$

where

$$y \equiv \left(\begin{array}{c} h^2 \tau_P \tau_v \tau_\epsilon \left(1 - \phi^2\right) + h\tau_P^2 \tau_v \tau_\epsilon \left(5 - \phi^2\right) + 2(h\tau_P)^{3/2} \tau_v \tau_\epsilon \left(1 - \phi^2\right) \\ +2(h\tau_P)^{3/2} \left((\tau_v + \tau_\epsilon)^2 - \tau_P^2\right) + h\tau_v^2 \tau_\epsilon^2 + \tau_P \tau_v^2 \tau_\epsilon^2 \end{array} \right), \tag{B.53}$$

and $\phi \equiv N^{1/2}\bar{\Phi}$. The numerator and the first ratio in Eq. (B.52) are positive, so we focus on the sign of y. Substitute from Eq. (17):

$$\tau_v + \tau_\epsilon = \frac{\tau_\epsilon \tau_v}{\tau_P^{1/2} h^{1/2}} - \tau_P, \tag{B.54}$$

into Eq. (B.53) to obtain:

$$y = \tau_v \tau_\epsilon \left(2\sqrt{h\tau_P} + h + \tau_P\right) \left(h\tau_P \left(1 - \phi^2\right) + \tau_v \tau_\epsilon\right) > 0. \tag{B.55}$$

This expression is positive because, by Hölder's inequality and the normalization $\mathbf{\Phi}'\mathbf{\Phi} \equiv 1$, $\phi \in (-1, 1)$.

B.4.1 Dollar returns vs rates of returns

The results bellow mirror the results of Banerjee (2010), who shows that the conditional CAPM holds regardless how one compute returns (dollar returns or rates of returns). Our focus is on the unconditional CAPM. Starting from

$$\mathbb{E}[\widetilde{\mathbf{R}}^e] = \gamma \mathbf{\Sigma} \mathbf{M}, \tag{B.56}$$

an decomposing the dollar returns as $\widetilde{\mathbf{R}}^e = \widetilde{\mathbf{D}} - \widetilde{\mathbf{P}}$, we obtain

$$\mathbb{E}[\widetilde{\mathbf{D}}] - \mathbb{E}[\widetilde{\mathbf{P}}] = \gamma \mathbf{\Sigma} \mathbf{M}. \tag{B.57}$$

Defining $\mathbf{P} \equiv \mathbb{E}[\widetilde{\mathbf{P}}]$ and $\text{diag}(\mathbf{P})$ as a diagonal matrix whose diagonal is $\mathbf{P}$, the unconditional expected rates of excess returns are given by:

$$\boldsymbol{\mu}^r = \text{diag}(\mathbf{P})^{-1}(\mathbb{E}[\widetilde{\mathbf{D}}] - \mathbf{P}) = \text{diag}(\mathbf{P})^{-1}\gamma \mathbf{\Sigma} \mathbf{M}. \tag{B.58}$$

(N.B. The constant parameter D plays no role in our results if we work with dollar returns, but, when working with rates of returns, a sufficiently large parameter D is necessary for preventing prices from becoming negative.)

The market portfolio weights are:

$$\mathbf{w} = \frac{\text{diag}(\mathbf{P})\mathbf{M}}{\mathbf{M}'\mathbf{P}}. \tag{B.59}$$

and thus the expected rate of excess returns on the market portfolio is:

$$\mathbf{w}'\boldsymbol{\mu}^r = \frac{\mathbf{M}'\text{diag}(\mathbf{P})}{\mathbf{M}'\mathbf{P}}\text{diag}(\mathbf{P})^{-1}\gamma \mathbf{\Sigma} \mathbf{M} = \frac{\gamma}{\mathbf{M}'\mathbf{P}}\mathbf{M}'\mathbf{\Sigma} \mathbf{M}. \tag{B.60}$$

Dividing (B.58) by (B.60) yields

$$\frac{\boldsymbol{\mu}^r}{\mathbf{w}'\boldsymbol{\mu}^r} = \mathbf{M}'\mathbf{P}\text{diag}(\mathbf{P})^{-1}\boldsymbol{\beta}. \tag{B.61}$$

We therefore recover the true unconditional CAPM, with betas modified according to the right hand side of the above. We verify that indeed these new betas have an average of 1:

$$\mathbf{w}'\mathbf{M}'\mathbf{P}\text{diag}(\mathbf{P})^{-1}\boldsymbol{\beta} = \frac{\mathbf{M}'\text{diag}(\mathbf{P})}{\mathbf{M}'\mathbf{P}}\mathbf{M}'\mathbf{P}\text{diag}(\mathbf{P})^{-1}\boldsymbol{\beta} = \mathbf{M}'\boldsymbol{\beta} = 1. \tag{B.62}$$

Moving now to the empiricist's view, realized rates of returns are computed as

$$\tilde{\mathbf{r}}^e \equiv \text{diag}(\widetilde{\mathbf{P}})^{-1}(\widetilde{\mathbf{D}} - \widetilde{\mathbf{P}}), \tag{B.63}$$

and thus are not normally distributed. Therefore, in order to obtain the CAPM as measured by the empiricist, we have to resort to simulations. We illustrate in Figure A1 the results of simulations for one particular calibration of the model (we have considered an extensive range of calibrations, and consistently obtained similar results). Each panel takes this calibration as a starting point and varies γ, τ_m, or τ_ϵ.

The three panels of the figure compare the theoretical distortion (Proposition 4) with the distortion obtained using rates of returns. The former is plotted with the solid line, and the latter with the dashed line. For each point on the dashed lines, we perform 10^7 simulations, then estimate the CAPM using realized returns and obtain δ by dividing the intercept by the slope, as dictated by Eq. (26). All the panels show that distortion with rates of returns is consistently larger than the distortion with dollar returns. These simulations therefore suggest that our main result does not depend on the use of dollar returns instead of rates of returns.

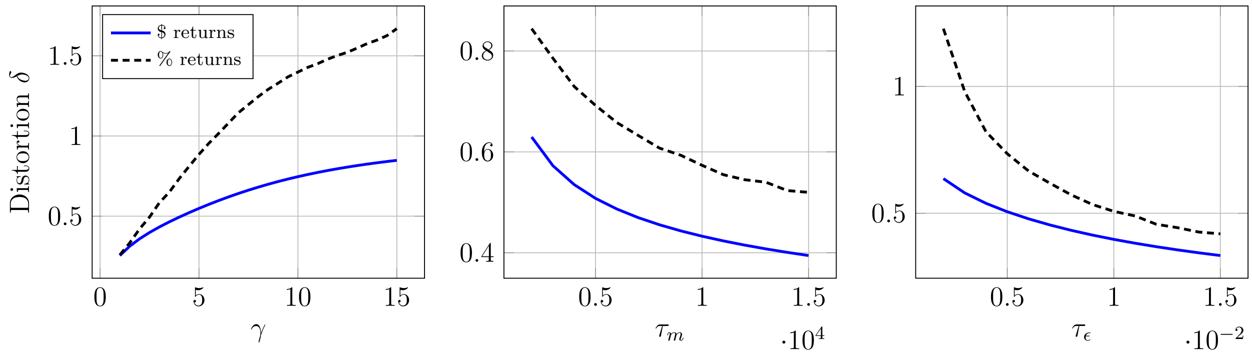

Figure A1: **Distortion with dollar returns vs. rates of return.** This figure illustrates the distortion from Proposition 4 (solid lines) and empiricist's distortion when using rates of returns instead of dollar returns (dashed lines). We plot the distortion as a function of γ, τ_m, and τ_ϵ. To obtain the dashed lines, we perform simulations of an economy with $N = 25$, $D = 200$, $\gamma = 3$, $\tau_F = 1$, $\tau_\epsilon = 0.008$, $\tau_m = 10,000$, $\mathbf{M} = \mathbf{1}/25$, and $\mathbf{\Phi} \sim \mathcal{N}(1, 1)$. For each point on the dashed lines we perform 10^7 simulations.

B.4.2 Conditioning on public information

Empiricist's rejection of the CAPM (Proposition 3) assumes the empiricist's information is limited to realized returns. In this appendix, we augment empiricist's dataset with all relevant public information (asset prices). Under this augmented information set, we show below that Proposition 3 still holds, but with a different coefficient of distortion, $\breve{\delta}$:

$$\breve{\delta} = \frac{\tau_v}{\tau_\epsilon(\tau - \tau_v)} \frac{1}{N\breve{\sigma}^2_{\mathbf{M}}} > 0, \tag{B.64}$$

where $\breve{\sigma}^2_{\mathbf{M}} \equiv \mathbf{M'} \operatorname{Var}[\widetilde{\mathbf{R}}^e|\widetilde{\mathbf{P}}]\mathbf{M}$ is the variance of excess returns on the market portfolio conditional on observing all publicly available prices, and the conditional covariance matrix $\operatorname{Var}[\widetilde{\mathbf{R}}^e|\widetilde{\mathbf{P}}]$ is

$$\operatorname{Var}[\widetilde{\mathbf{R}}^e|\widetilde{\mathbf{P}}] = \operatorname{Var}[\widetilde{\mathbf{D}}|\widetilde{\mathbf{P}}] = \frac{1}{\tau - \tau_v}\mathbf{\Phi\Phi'} + \frac{1}{\tau_\epsilon}\mathbf{I}. \tag{B.65}$$

Thus, we can write[31]

$$\operatorname{Var}[\widetilde{\mathbf{R}}^e|\widetilde{\mathbf{P}}] = \frac{\tau}{\tau - \tau_v}\mathbf{\Sigma} - \frac{\tau_v}{\tau_\epsilon(\tau - \tau_v)}\mathbf{I}. \tag{B.66}$$

The empiricist obtains a new set of betas:

$$\breve{\boldsymbol{\beta}} = \frac{\operatorname{Var}[\widetilde{\mathbf{R}}^e|\widetilde{\mathbf{P}}]\mathbf{M}}{\breve{\sigma}^2_{\mathbf{M}}} = \underbrace{\frac{\tau}{\tau - \tau_v}\frac{\widehat{\sigma}^2_{\mathbf{M}}}{\breve{\sigma}^2_{\mathbf{M}}}}_{\equiv(1+\breve{\delta})}\boldsymbol{\beta} - \frac{\tau_v}{\tau_\epsilon(\tau - \tau_v)}\frac{\mathbf{M}}{\breve{\sigma}^2_{\mathbf{M}}}. \tag{B.67}$$

[31] Notice that it is not necessary to assume here that the empiricist knows the price coefficients of Proposition 1. This is because $\operatorname{Var}[\widetilde{\mathbf{R}}^e|\widetilde{\mathbf{P}}] = \operatorname{Var}[\widetilde{\mathbf{R}}^e] - \operatorname{Var}[\mathbb{E}[\widetilde{\mathbf{R}}^e|\widetilde{\mathbf{P}}]]$. The empiricist can compute both terms on the right hand side: the first term is the covariance matrix of realized returns; the second term is the covariance matrix of expected returns obtained after regressing realized returns of each asset on the vector of prices $\widetilde{\mathbf{P}}$.

Take average on both sides by multiplying with $\mathbf{M}'$:

$$1 = (1+\breve{\delta}) - \frac{\tau_v}{\tau_\epsilon(\tau-\tau_v)}\frac{1}{N\breve{\sigma}^2_{\mathbf{M}}}, \tag{B.68}$$

and thus we obtain Eq. (B.64). Replacing $\breve{\delta}$ in (B.67) and subtracting $\mathbf{1}$ on both sides yields

$$\breve{\boldsymbol{\beta}} - \mathbf{1} = (1+\breve{\delta})(\boldsymbol{\beta}-\mathbf{1}). \tag{B.69}$$

Thus, the empiricist still observes distorted betas, even after controlling for all the available public information. We note that the new distortion $\breve{\delta}$ in (B.64) clearly depends on the precision of private information τ_v. Furthermore, it can be shown (through numerical examples) that the distortion $\breve{\delta}$ may become larger than the initial distortion obtained without conditioning, δ. This happens when the private information is sufficiently precise.

B.5 Proof of Proposition 5

We first solve for τ in the common information economy. Apply the projection theorem:

$$\tau = \tau_F + \left(\frac{\sqrt{\tau_P}}{\sqrt{\tau_m}}\boldsymbol{\Phi}\right)'\tau_g\left(\frac{\sqrt{\tau_P}}{\sqrt{\tau_m}}\boldsymbol{\Phi}\right) = \tau_F + \frac{\tau_g}{\tau_m}\tau_P. \tag{B.70}$$

Then, finding the equilibrium follows the same steps as in Appendix B.1, albeit much simpler here since there is no learning from prices. To prove Proposition 5, compare the two sets of betas:

$$\widehat{\boldsymbol{\beta}} = \frac{\left(1+\frac{\gamma^2}{\tau_m\tau_\epsilon}\right)\sigma^2_{\mathbf{M}}\boldsymbol{\beta} + \left(\frac{\gamma^2 e_1\bar{\Phi}^2}{\tau_m\tau} + \frac{\tau_v\bar{\Phi}^2}{\tau^2}\right)\frac{\boldsymbol{\Phi}}{\bar{\Phi}}}{\left(1+\frac{\gamma^2}{\tau_m\tau_\epsilon}\right)\sigma^2_{\mathbf{M}} + \frac{\gamma^2 e_1\bar{\Phi}^2}{\tau_m\tau} + \frac{\tau_v\bar{\Phi}^2}{\tau^2}} \tag{B.71}$$

$$\widehat{\boldsymbol{\beta}}_{\text{CIE}} = \frac{\left(1+\frac{\gamma^2}{\tau_m\tau_\epsilon}\right)\sigma^2_{\mathbf{M}}\boldsymbol{\beta} + \frac{\gamma^2 e_1\bar{\Phi}^2}{\tau_m\tau}\frac{\boldsymbol{\Phi}}{\bar{\Phi}}}{\left(1+\frac{\gamma^2}{\tau_m\tau_\epsilon}\right)\sigma^2_{\mathbf{M}} + \frac{\gamma^2 e_1\bar{\Phi}^2}{\tau_m\tau}}. \tag{B.72}$$

Both $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\beta}}_{\text{CIE}}$ are weighted averages of $\boldsymbol{\beta}$ and $\boldsymbol{\Phi}/\bar{\Phi}$. In the common information economy, the only variable that changes above is $\tau_v = 0$ ($\sigma^2_{\mathbf{M}}$ stays the same). Thus, the weighted average gets closer to $\boldsymbol{\beta}$: empiricist's betas move closer to the true betas, which implies $\delta_{\text{CIE}} < \delta$. □

B.6 Proof of Proposition 6

The proof starts from Eq. (27),

$$\widehat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma} + \text{Var}[\overline{\mathbb{E}}[\widetilde{\mathbf{R}}^e]] + \text{Var}[\mathbb{E}^{i*}[\widetilde{\mathbf{R}}^e]], \tag{B.73}$$

which we multiply with $\mathbf{M}$ to obtain

$$\widehat{\sigma}^2_{\mathbf{M}}\widehat{\boldsymbol{\beta}} = \sigma^2_{\mathbf{M}}\boldsymbol{\beta} + \text{Var}[\overline{\mathbb{E}}[\widetilde{R}^e_{\mathbf{M}}]]\boldsymbol{\beta}^{\mathcal{C}} + \text{Var}[\mathbb{E}^{i*}[\widetilde{R}^e_{\mathbf{M}}]]\boldsymbol{\beta}^{\mathcal{D}}, \tag{B.74}$$

and then divide by $\widehat{\sigma}^2_{\mathbf{M}}$ to obtain Eq. (40). □

B.6.1 Proof of Corollary 6.1

Start from the definition of $\boldsymbol{\beta}^{\mathcal{C}}$:

$$\boldsymbol{\beta}^{\mathcal{C}} = \frac{\frac{\gamma^2\sigma_{\mathbf{M}}^2}{\tau_m\tau_\epsilon}\boldsymbol{\beta} + \frac{\gamma^2 e_1\bar{\Phi}^2}{\tau_m\tau}\frac{\boldsymbol{\Phi}}{\bar{\Phi}}}{\frac{\gamma^2\sigma_{\mathbf{M}}^2}{\tau_m\tau_\epsilon} + \frac{\gamma^2 e_1\bar{\Phi}^2}{\tau_m\tau}}, \tag{B.75}$$

and thus we obtain Eq. (42) in the text:

$$\boldsymbol{\beta}^{\mathcal{C}} - \boldsymbol{\beta} = \frac{\frac{\gamma^2 e_1\bar{\Phi}^2}{\tau_m\tau}}{\frac{\gamma^2\sigma_{\mathbf{M}}^2}{\tau_m\tau_\epsilon} + \frac{\gamma^2 e_1\bar{\Phi}^2}{\tau_m\tau}}\left(\frac{\boldsymbol{\Phi}}{\bar{\Phi}} - \boldsymbol{\beta}\right) = \left(\frac{\tau\sigma_{\mathbf{M}}^2}{\bar{\Phi}^2} - 1\right)\frac{\tau_\epsilon e_1\bar{\Phi}^2}{\tau\sigma_{\mathbf{M}}^2 + \tau_\epsilon e_1\bar{\Phi}^2}(\boldsymbol{\beta} - \mathbf{1}), \tag{B.76}$$

where we have used Corollary 1.1 under the form $\boldsymbol{\Phi}/\bar{\Phi} - \boldsymbol{\beta} = (\tau\sigma_{\mathbf{M}}^2/\bar{\Phi}^2 - 1)(\boldsymbol{\beta} - \mathbf{1})$.

Regarding $\boldsymbol{\beta}^{\mathcal{D}}$, in our model $\boldsymbol{\beta}^{\mathcal{D}} = \boldsymbol{\Phi}/\bar{\Phi}$. Thus

$$\boldsymbol{\beta}^{\mathcal{D}} - \boldsymbol{\beta} = \frac{\boldsymbol{\Phi}}{\bar{\Phi}} - \boldsymbol{\beta} = \left(\frac{\tau\sigma_{\mathbf{M}}^2}{\bar{\Phi}^2} - 1\right)(\boldsymbol{\beta} - \mathbf{1}). \tag{B.77}$$

which is Eq. (43) in the text. □

B.7 Proof of Proposition 7

The proof starts from the definition of $\text{Cov}[\boldsymbol{\beta}, \widehat{\boldsymbol{\beta}}]$ and $\text{Var}[\widehat{\boldsymbol{\beta}}]$:

$$\text{Cov}[\boldsymbol{\beta}, \widehat{\boldsymbol{\beta}}] = \frac{1}{N}\frac{\mathbf{M}'\boldsymbol{\Sigma}\widehat{\boldsymbol{\Sigma}}\mathbf{M}}{\widehat{\sigma}_{\mathbf{M}}^2\sigma_{\mathbf{M}}^2} - \beta_{\text{avg}}\widehat{\beta}_{\text{avg}} \tag{B.78}$$

$$\text{Var}[\widehat{\boldsymbol{\beta}}] = \frac{1}{N}\frac{\mathbf{M}'\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{\Sigma}}\mathbf{M}}{\widehat{\sigma}_{\mathbf{M}}^2\widehat{\sigma}_{\mathbf{M}}^2} - \widehat{\beta}_{\text{avg}}^2, \tag{B.79}$$

where β_{avg} and $\widehat{\beta}_{\text{avg}}$ are *arithmetic* averages of true betas and empiricist's betas across stocks. If $\mathbf{M} = \mathbf{1}/N$, these arithmetic averages coincide with market-weighted averages and thus are both 1. It then follows that in the case $\mathbf{M} = \mathbf{1}/N$, $\text{Cov}[\boldsymbol{\beta}, \widehat{\boldsymbol{\beta}}] < \text{Var}[\widehat{\boldsymbol{\beta}}]$ if and only if

$$\frac{\mathbf{M}'\boldsymbol{\Sigma}\widehat{\boldsymbol{\Sigma}}\mathbf{M}}{\mathbf{M}'\boldsymbol{\Sigma}\mathbf{M}} < \frac{\mathbf{M}'\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{\Sigma}}\mathbf{M}}{\mathbf{M}'\widehat{\boldsymbol{\Sigma}}\mathbf{M}}. \tag{B.80}$$

If we further assume that there is no dispersion in beliefs and investors' information is common knowledge, then the following relation results immediately from (10):

$$\widehat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma} + \frac{\gamma^2}{\tau_m}\boldsymbol{\Sigma}\boldsymbol{\Sigma}. \tag{B.81}$$

Considering now the eigenvalue decompositions of $\boldsymbol{\Sigma}$ and $\widehat{\boldsymbol{\Sigma}}$ (it is easy to see that $\boldsymbol{\Sigma}$ and $\widehat{\boldsymbol{\Sigma}}$ have the same eigenvectors, but not the same eigenvalues),

$$\boldsymbol{\Sigma} = \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}' \tag{B.82}$$

$$\widehat{\boldsymbol{\Sigma}} = \mathbf{Q}\boldsymbol{\Lambda}(\mathbf{I} + h\boldsymbol{\Lambda})\mathbf{Q}', \tag{B.83}$$

where $h \equiv \gamma^2/\tau_m > 0$, and defining $\mathbf{O} \equiv \mathbf{Q}'\mathbf{M}$, we need to prove that

$$\frac{\mathbf{O}'\Lambda\Lambda(\mathbf{I}+h\boldsymbol{\Lambda})\mathbf{O}}{\mathbf{O}'\Lambda\mathbf{O}} < \frac{\mathbf{O}'\Lambda(\mathbf{I}+h\boldsymbol{\Lambda})\Lambda(\mathbf{I}+h\boldsymbol{\Lambda})\mathbf{O}}{\mathbf{O}'\Lambda(\mathbf{I}+h\boldsymbol{\Lambda})\mathbf{O}}, \tag{B.84}$$

or

$$\sum \frac{O_n^2\Lambda_n}{\sum O_n^2\Lambda_n}\Lambda_n(1+h\Lambda_n) < \sum \frac{O_n^2\Lambda_n(1+h\Lambda_n)}{\sum O_n^2\Lambda_n(1+h\Lambda_n)}\Lambda_n(1+h\Lambda_n). \tag{B.85}$$

This is a comparison of two weighted averages with different weights:

$$\Omega_{1n} = \frac{O_n^2\Lambda_n}{\sum O_n^2\Lambda_n} = \frac{O_n^2\Lambda_n}{A}, \quad \sum_{n=1}^{N}\Omega_{1n} = 1 \tag{B.86}$$

$$\Omega_{2n} = \frac{O_n^2\Lambda_n(1+h\Lambda_n)}{\sum O_n^2\Lambda_n(1+h\Lambda_n)} = \frac{O_n^2\Lambda_n(1+h\Lambda_n)}{A+hB}, \quad \sum_{n=1}^{N}\Omega_{2n} = 1, \tag{B.87}$$

where $A \equiv \sum O_n^2\Lambda_n > 0$ and $B \equiv \sum O_n^2\Lambda_n^2 > 0$. The difference between the weights is:

$$\Omega_{2n} - \Omega_{1n} = \frac{O_n^2\Lambda_n(1+h\Lambda_n)}{A+hB} - \frac{O_n^2\Lambda_n}{A} \tag{B.88}$$

$$= \left(\frac{1}{A+hB} - \frac{1}{A}\right)O_n^2\Lambda_n + \frac{h}{A+hB}O_n^2\Lambda_n^2 \tag{B.89}$$

$$= -\frac{hB}{A(A+hB)}O_n^2\Lambda_n + \frac{h}{A+hB}O_n^2\Lambda_n^2 \tag{B.90}$$

$$= \frac{h}{A+hB}O_n^2\Lambda_n\left(\Lambda_n - \frac{B}{A}\right). \tag{B.91}$$

This is a quadratic function of Λ_n, with two real roots: $\Lambda_n = 0$ and $\Lambda_n = B/A > 0$. Importantly, these roots do not depend on O_n. The function is strictly negative on the interval $(0, B/A)$ and strictly positive on $(B/A, \infty)$. We therefore have:

$$\begin{cases} \Omega_{2n} < \Omega_{1n}, & \text{if } \Lambda_n \in (0, B/A) \\ \Omega_{2n} > \Omega_{1n}, & \text{if } \Lambda_n > B/A, \end{cases} \tag{B.92}$$

Thus, the weighted average on the right hand side of (B.85) places higher weights on higher values, and the inequality is now verified. □

B.8 Proof of Proposition 8

The relation of Lemma 1 remains valid regardless the value of $\mathbf{M}$. Using Eq. (B.11) we can write

$$\widehat{\boldsymbol{\Sigma}} = c_3\boldsymbol{\Sigma} - c_4\mathbf{I}, \tag{B.93}$$

where c_3, and c_4 are positive scalars defined in (B.35). Multiplication with $\mathbf{M}$ yields

$$\widehat{\boldsymbol{\beta}} = \frac{c_3\sigma_{\mathbf{M}}^2}{\widehat{\sigma}_{\mathbf{M}}^2}\boldsymbol{\beta} - \frac{c_4}{\widehat{\sigma}_{\mathbf{M}}^2}\mathbf{M} \tag{B.94}$$

Multiply with $\mathbf{M}'$ and use the fact that betas must average to 1:

$$1 = \frac{c_3 \sigma^2_{\mathbf{M}}}{\widehat{\sigma}^2_{\mathbf{M}}} - \frac{c_4}{\widehat{\sigma}^2_{\mathbf{M}}} \mathbf{M}'\mathbf{M}. \tag{B.95}$$

Both elements on the right hand side are positive, and the last element is the disortion δ, generalized for an aribtrary vector $\mathbf{M}$. To see this, write again Eq. (B.94) using (B.95):

$$\widehat{\boldsymbol{\beta}} = \left(1 + \frac{c_4 \mathbf{M}'\mathbf{M}}{\widehat{\sigma}^2_{\mathbf{M}}}\right)\boldsymbol{\beta} - \frac{c_4 \mathbf{M}'\mathbf{M}}{\widehat{\sigma}^2_{\mathbf{M}}} \frac{\mathbf{M}}{\mathbf{M}'\mathbf{M}} = (1+\delta)\boldsymbol{\beta} - \delta \frac{\mathbf{M}}{\mathbf{M}'\mathbf{M}}, \tag{B.96}$$

or

$$\boldsymbol{\beta} - \mathbf{1} = \frac{1}{1+\delta}(\widehat{\boldsymbol{\beta}} - \mathbf{1}) + \frac{\delta}{1+\delta}\left(\frac{\mathbf{M}}{\mathbf{M}'\mathbf{M}} - \mathbf{1}\right). \tag{B.97}$$

where

$$\delta = \frac{\mathbf{M}'\mathbf{M}}{\widehat{\sigma}^2_{\mathbf{M}}}\left(\frac{\gamma^2 e_1}{\tau_m \tau_\epsilon} + \frac{\tau_v}{\tau_\epsilon \tau}\right). \tag{B.98}$$

Multiplying (B.97) with $\mu_{\mathbf{M}}$ and using the true CAPM relation $\boldsymbol{\mu} = \boldsymbol{\beta}\mu_{\mathbf{M}}$ yields (53). □

B.9 Proof of Lemma 2

For the SML to be downward-sloping we need that $\text{Cov}[\boldsymbol{\beta}, \widehat{\boldsymbol{\beta}}] < 0$. We first determine necessary condition for this result to obtain. Write the covariance between $\boldsymbol{\beta}$ and $\widehat{\boldsymbol{\beta}}$ as:

$$\text{Cov}[\boldsymbol{\beta}, \widehat{\boldsymbol{\beta}}] = \frac{1}{N \widehat{\sigma}^2_{\mathbf{M}} \sigma^2_{\mathbf{M}}} \underbrace{\left(\mathbf{M}'\boldsymbol{\Sigma}\widehat{\boldsymbol{\Sigma}}\mathbf{M} - \mathbf{1}'\boldsymbol{\Sigma}\mathbf{M}\mathbf{1}'\widehat{\boldsymbol{\Sigma}}\mathbf{M}/N\right)}_{\equiv \Delta}. \tag{B.99}$$

The sign of this expression depends on the term in brackets. Tedious computations show that this term can be written as:

$$\begin{aligned} \Delta \equiv\ & \tau^{-2}\left(1 + \frac{\gamma^2}{\tau_m \tau_\epsilon} + \frac{\gamma^2 e_1}{\tau_m} + \frac{\tau_n}{\tau}\right)\mathbf{M}'\boldsymbol{\Phi}\underbrace{\left(1 - N\bar{\Phi}^2\right)}_{\geq 0} + \tau_\epsilon^{-2}\left(1 + \frac{\gamma^2}{\tau_m \tau_\epsilon}\right)\underbrace{\left(\|\mathbf{M}\|^2 - 1/N\right)}_{\geq 0} \\ & + \tau^{-1}\tau_\epsilon^{-1}\left(2\left(1 + \frac{\gamma^2}{\tau_m \tau_\epsilon}\right) + \frac{\gamma^2 e_1}{\tau_m} + \frac{\tau_n}{\tau}\right)\mathbf{M}'\boldsymbol{\Phi}\,\text{Cov}[\mathbf{M}, \boldsymbol{\Phi}] \end{aligned} \tag{B.100}$$

The sign under the first curly bracket follows from the normalization, $\boldsymbol{\Phi}'\boldsymbol{\Phi} \equiv 1$, and Hölder's inequality, which together imply that $\bar{\Phi}^2 N \leq 1$; similarly, the sign under the second curly follows from that $\mathbf{1}'\mathbf{M} \equiv 1$ and Hölder's inequality, which together imply that $\|\mathbf{M}\|^2 \geq 1/N$. Hence, for $\Delta < 0$ a necessary condition is that $\text{Cov}[\mathbf{M}, \boldsymbol{\Phi}] < 0$ and the lemma follows. □

B.10 Proof of Lemma 3

We start by repeating the steps of Appendix B.1 for the multiple-factor case. As is customary, we conjecture that prices satisfy

$$\widetilde{\mathbf{P}} = \boldsymbol{\xi}_0 \mathbf{M} + \boldsymbol{\lambda} \widetilde{\mathbf{F}} + \boldsymbol{\xi} \widetilde{\mathbf{m}}, \tag{B.101}$$

for which a sufficient statistic is $\widetilde{\mathbf{P}}^a \equiv \widetilde{\mathbf{P}} - \boldsymbol{\xi}_0 \mathbf{M} = \boldsymbol{\lambda} \widetilde{\mathbf{F}} + \boldsymbol{\xi} \widetilde{\mathbf{m}}$. The projection theorem implies that

$$\boldsymbol{\tau} \equiv \operatorname{Var}\left[\widetilde{\mathbf{F}} \middle| \mathscr{F}_i\right]^{-1} = (\tau_F + \tau_v) J \mathbf{I}_J + \boldsymbol{\lambda}' (\boldsymbol{\xi}\boldsymbol{\xi}')^{-1} \boldsymbol{\lambda} \tau_m. \tag{B.102}$$

and

$$\mathbb{E}\left[\widetilde{\mathbf{F}} \middle| \mathscr{F}_i\right] = \boldsymbol{\tau}^{-1} \left((\boldsymbol{\tau} - \tau_F \mathbf{I}_J) \widetilde{\mathbf{F}} + \tau_m \boldsymbol{\lambda}' (\boldsymbol{\xi}\boldsymbol{\xi}')^{-1} \boldsymbol{\xi} \widetilde{\mathbf{m}} + \tau_v \tilde{\mathbf{v}}_i \right). \tag{B.103}$$

It follows that average expectations of future payoffs satisfy

$$\bar{\mathbb{E}}[\widetilde{\mathbf{D}}] \equiv \int_i \mathbb{E}[\widetilde{\mathbf{D}} | \mathscr{F}_i] \mathrm{d}i = D\mathbf{1} + \boldsymbol{\Phi} \boldsymbol{\tau}^{-1} \left((\boldsymbol{\tau} - \tau_F \mathbf{I}_J) \widetilde{\mathbf{F}} + \tau_m \boldsymbol{\lambda}' (\boldsymbol{\xi}\boldsymbol{\xi}')^{-1} \boldsymbol{\xi} \widetilde{\mathbf{m}} \right). \tag{B.104}$$

and the conditional covariance matrix of future payoffs satisfies:

$$\boldsymbol{\Sigma} \equiv \operatorname{Var}\left[\widetilde{\mathbf{D}} \middle| \mathscr{F}_i\right] = \boldsymbol{\Phi} \boldsymbol{\tau}^{-1} \boldsymbol{\Phi}' + \tau_\epsilon^{-1} \mathbf{I}_N. \tag{B.105}$$

The market-clearing condition then requires that $\widetilde{\mathbf{P}} = \bar{\mathbb{E}}[\widetilde{\mathbf{D}}] - \gamma \boldsymbol{\Sigma} (\mathbf{M} + \widetilde{\mathbf{m}})$, which yields

$$\widetilde{\mathbf{P}} = D\mathbf{1} + \boldsymbol{\Phi} \boldsymbol{\tau}^{-1} \left((\boldsymbol{\tau} - \tau_F \mathbf{I}_J) \widetilde{\mathbf{F}} + \tau_m (\boldsymbol{\xi}^{-1} \boldsymbol{\lambda})' \widetilde{\mathbf{m}} \right) - \gamma \boldsymbol{\Sigma} (\mathbf{M} + \widetilde{\mathbf{m}}). \tag{B.106}$$

Separating variables we obtain the following system of equations:

$$\boldsymbol{\xi}_0 = -\gamma \boldsymbol{\Sigma}, \quad \boldsymbol{\lambda} = \boldsymbol{\Phi} \boldsymbol{\tau}^{-1} (\boldsymbol{\tau} - \tau_F \mathbf{I}_J), \tag{B.107}$$

and

$$\boldsymbol{\xi} = \tau_m \boldsymbol{\Phi} \boldsymbol{\tau}^{-1} (\boldsymbol{\xi}^{-1} \boldsymbol{\lambda})' - \gamma \left(\boldsymbol{\Phi} \boldsymbol{\tau}^{-1} \boldsymbol{\Phi}' + \tau_\epsilon^{-1} \mathbf{I}_N \right). \tag{B.108}$$

To reduce the size of this system of equations, post-multiply both sides of the above by $\boldsymbol{\xi}^{-1} \boldsymbol{\lambda}$:

$$\boldsymbol{\lambda} = \tau_m \boldsymbol{\Phi} \boldsymbol{\tau}^{-1} (\boldsymbol{\xi}^{-1} \boldsymbol{\lambda})' \boldsymbol{\xi}^{-1} \boldsymbol{\lambda} - \gamma \left(\boldsymbol{\Phi} \boldsymbol{\tau}^{-1} \boldsymbol{\Phi}' + \tau_\epsilon^{-1} \mathbf{I}_N \right) \boldsymbol{\xi}^{-1} \boldsymbol{\lambda}. \tag{B.109}$$

Observing that $\tau_m \boldsymbol{\Phi} \boldsymbol{\tau}^{-1} (\boldsymbol{\xi}^{-1} \boldsymbol{\lambda})' \boldsymbol{\xi}^{-1} \boldsymbol{\lambda} = \boldsymbol{\Phi} \boldsymbol{\tau}^{-1} \left(\boldsymbol{\tau} - (\tau_F + \tau_v) \mathbf{I}_J \right) \equiv \boldsymbol{\lambda} - \tau_v \boldsymbol{\Phi} \boldsymbol{\tau}^{-1}$, we obtain

$$\tau_v \boldsymbol{\Phi} \boldsymbol{\tau}^{-1} = -\gamma \left(\boldsymbol{\Phi} \boldsymbol{\tau}^{-1} \boldsymbol{\Phi}' + \tau_\epsilon^{-1} \mathbf{I}_N \right) \boldsymbol{\xi}^{-1} \boldsymbol{\lambda}, \tag{B.110}$$

which yields an equation for the vector of signal-to-noise ratios:

$$\boldsymbol{\xi}^{-1} \boldsymbol{\lambda} = -\frac{\tau_v}{\gamma} \left(\boldsymbol{\Phi} \boldsymbol{\tau}^{-1} \boldsymbol{\Phi}' + \tau_\epsilon^{-1} \mathbf{I}_N \right)^{-1} \boldsymbol{\Phi} \boldsymbol{\tau}^{-1}. \tag{B.111}$$

Pre-multiply this equation by $\boldsymbol{\tau}^{-1}\boldsymbol{\Phi}'$ and use Woodbury matrix identity to write:

$$\boldsymbol{\tau}^{-1}\boldsymbol{\Phi}'\left(\boldsymbol{\Phi}\boldsymbol{\tau}^{-1}\boldsymbol{\Phi}'+\tau_\epsilon^{-1}\mathbf{I}_N\right)^{-1}\boldsymbol{\Phi}\boldsymbol{\tau}^{-1}=\boldsymbol{\tau}^{-1}-(\boldsymbol{\tau}+\tau_\epsilon\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1} \tag{B.112}$$

and to conclude that

$$\boldsymbol{\tau}^{-1}\boldsymbol{\Phi}'\boldsymbol{\xi}^{-1}\boldsymbol{\lambda}=-\frac{\tau_v}{\gamma}\left(\boldsymbol{\tau}^{-1}-(\boldsymbol{\tau}+\tau_\epsilon\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}\right). \tag{B.113}$$

Conjecture that $\boldsymbol{\xi}^{-1}\boldsymbol{\lambda}\equiv-\frac{1}{\sqrt{\tau_m}}\boldsymbol{\Phi}\boldsymbol{\tau}_P$, where $\boldsymbol{\tau}_P$ is a $J\times J$ symmetric matrix of $J(J+1)/2$ unknown coefficients. Replacing this conjecture in the expression for total precision in Eq. (B.102) to obtain Eq. (60). Further replacing the conjecture in Eq. (B.113) produces a matrix equation for $\boldsymbol{\tau}_P$:

$$\tau^{-1}\boldsymbol{\Phi}'\boldsymbol{\Phi}\boldsymbol{\tau}_P=\sqrt{\tau_m}\frac{\tau_v}{\gamma}\left(\boldsymbol{\tau}^{-1}-(\boldsymbol{\tau}+\tau_\epsilon\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}\right), \tag{B.114}$$

which, premultiplying by $\boldsymbol{\tau}$, can be rewritten as

$$\boldsymbol{\Phi}'\boldsymbol{\Phi}\boldsymbol{\tau}_P=\sqrt{\tau_m}\frac{\tau_v}{\gamma}\left(\mathbf{I}_J-(\mathbf{I}_J+\tau_\epsilon\boldsymbol{\tau}^{-1}\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}\right). \tag{B.115}$$

We can simplify this equation further by premultiplying by $(\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}$:

$$\boldsymbol{\tau}_P=\sqrt{\tau_m}\frac{\tau_v J}{\gamma}\left((\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}-(\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}\left((\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}+\tau_\epsilon\boldsymbol{\tau}^{-1}\right)^{-1}(\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}\right), \tag{B.116}$$

and apply Woodbury matrix identity:

$$\boldsymbol{\tau}_P=\sqrt{\tau_m}\frac{\tau_v J}{\gamma}\left(\boldsymbol{\Phi}'\boldsymbol{\Phi}+\tau_\epsilon^{-1}\boldsymbol{\tau}\right)^{-1}. \tag{B.117}$$

Now let us go back to Eq. (7), use Woodbury matrix identity, and substitute Eq. (B.117) in it:

$$\boldsymbol{\tau}^{-1}=(\tau_F+\tau_v)^{-1}J^{-1}\left(\mathbf{I}-\left(\mathbf{I}+(\tau_F+\tau_v)J\boldsymbol{\tau}_P^{-1}(\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}\boldsymbol{\tau}_P^{-1}\right)^{-1}\right) \tag{B.118}$$

$$=(\tau_F+\tau_v)^{-1}J^{-1}\left(\mathbf{I}-\left(\mathbf{I}+\frac{(\tau_F+\tau_v)\gamma^2}{\tau_m\tau_v^2 J}\left(\boldsymbol{\Phi}'\boldsymbol{\Phi}+\tau_\epsilon^{-1}\boldsymbol{\tau}\right)(\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}\left(\boldsymbol{\Phi}'\boldsymbol{\Phi}+\tau_\epsilon^{-1}\boldsymbol{\tau}\right)\right)^{-1}\right), \tag{B.119}$$

which is an explicit matrix equation for $\boldsymbol{\tau}$. Further recall that, under Eq. (57), the average eigenvalue of $\boldsymbol{\Phi}'\boldsymbol{\Phi}$ is:

$$\frac{1}{J}\text{tr}(\boldsymbol{\Phi}'\boldsymbol{\Phi})=N. \tag{B.120}$$

Hence, in the limit when $N\to\infty$, it is important to focus on $\frac{1}{N}\Phi'\Phi$ and rewrite this equation as:

$$(\boldsymbol{\tau}/N)^{-1}=((\tau_F+\tau_v)J/N)^{-1}\mathbf{I} \tag{B.121}$$

$$-\left((\tau_F+\tau_v)J/N\mathbf{I}+\frac{(\tau_F+\tau_v)^2\gamma^2}{\tau_m\tau_v^2}\left(\frac{1}{N}\boldsymbol{\Phi}'\boldsymbol{\Phi}+\tau_\epsilon^{-1}\boldsymbol{\tau}/N\right)\left(\frac{1}{N}\boldsymbol{\Phi}'\boldsymbol{\Phi}\right)^{-1}\left(\frac{1}{N}\boldsymbol{\Phi}'\boldsymbol{\Phi}+\tau_\epsilon^{-1}\boldsymbol{\tau}/N\right)\right)^{-1}. \tag{B.122}$$

Let us now introduce the eigendecomposition in Eq. (59), which further yields:

$$(\boldsymbol{\tau}/N)^{-1} = ((\tau_F+\tau_v)J/N)^{-1}\mathbf{Q}\mathbf{Q}' \tag{B.123}$$

$$-\mathbf{Q}\left((\tau_F+\tau_v)J/N\mathbf{I}+\frac{(\tau_F+\tau_v)^2\gamma^2}{\tau_m\tau_v^2}\left(\boldsymbol{\Lambda}+\tau_\epsilon^{-1}\mathbf{Q}'\boldsymbol{\tau}/N\mathbf{Q}\right)\boldsymbol{\Lambda}^{-1}\left(\boldsymbol{\Lambda}+\tau_\epsilon^{-1}\mathbf{Q}'\boldsymbol{\tau}/N\mathbf{Q}\right)\right)^{-1}\mathbf{Q}', \tag{B.124}$$

where we have used that $\mathbf{Q}\mathbf{Q}'=\mathbf{I}$. Finally, post-multiplying by $\mathbf{Q}$ and pre-multiplying by $\mathbf{Q}'$ on both sides, and taking the limit when $N\to\infty$, $J\to\infty$, and $J/N\to\psi$ we obtain a matrix equation for $\boldsymbol{\tau}_\infty$, as defined in Eq. (61):

$$\boldsymbol{\tau}_\infty^{-1} = ((\tau_F+\tau_v)\psi)^{-1}\mathbf{I}-\left((\tau_F+\tau_v)\psi\mathbf{I}+\frac{(\tau_F+\tau_v)^2\gamma^2}{\tau_m\tau_v^2}\left(\boldsymbol{\Lambda}+\tau_\epsilon^{-1}\boldsymbol{\tau}_\infty\right)\boldsymbol{\Lambda}^{-1}\left(\boldsymbol{\Lambda}+\tau_\epsilon^{-1}\boldsymbol{\tau}_\infty\right)\right)^{-1}, \tag{B.125}$$

which, using once more Woodbury identity, simplifies into:

$$\boldsymbol{\tau}_\infty = (\tau_F+\tau_v)\psi\mathbf{I}+\frac{\psi^2\tau_m\tau_v^2}{\gamma^2}\left(\boldsymbol{\Lambda}+\tau_\epsilon^{-1}\boldsymbol{\tau}_\infty\right)^{-1}\boldsymbol{\Lambda}\left(\boldsymbol{\Lambda}+\tau_\epsilon^{-1}\boldsymbol{\tau}_\infty\right)^{-1}. \tag{B.126}$$

Since all matrices are diagonal, it is natural to conjecture (and verify) that $\boldsymbol{\tau}_\infty$ is diagonal, too, with elements $\{\tau_{j,\infty}\}_{j=1}^J$ on its diagonal. Substituting this conjecture in the matrix equation decouples it into J algebraic equations for each diagonal element $\tau_{j,\infty}$, each of the form:

$$\tau_{j,\infty} = \psi\left(\tau_F+\tau_v+\frac{\lambda_j\tau_m\tau_v^2\tau_\epsilon^2\psi}{\gamma^2(\tau_{j,\infty}+\lambda_j\tau_\epsilon)^2}\right), \quad j=1,...,J. \tag{B.127}$$

This is a cubic equation in each $\tau_{j,\infty}$ that has a unique solution, and which delivers the mapping in Eq. (62).

B.11 Proof of Proposition 9

As in the single-factor case, we start from the law of total covariance to write:

$$\widehat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}+\frac{\gamma^2}{\tau_m}\boldsymbol{\Sigma}\boldsymbol{\Sigma}+\tau_vJ\boldsymbol{\Phi}\boldsymbol{\tau}^{-1}\boldsymbol{\tau}^{-1}\boldsymbol{\Phi}' \tag{B.128}$$

$$= \left(1+\frac{\gamma^2}{\tau_m\tau_\epsilon}\right)\boldsymbol{\Sigma}+\boldsymbol{\Phi}\boldsymbol{\tau}^{-1}\left(\frac{\gamma^2}{\tau_m}\left(\boldsymbol{\Phi}'\boldsymbol{\Phi}+\tau_\epsilon^{-1}\boldsymbol{\tau}\right)+\tau_vJ\mathbf{I}\right)\boldsymbol{\tau}^{-1}\boldsymbol{\Phi}'. \tag{B.129}$$

Let us now rescale by N and use the eigendecomposition in Eq. (59) and the definition of limiting precision in Eq. (61) to rewrite this expression as:

$$N\widehat{\boldsymbol{\Sigma}} = \left(1+\frac{\gamma^2}{\tau_m\tau_\epsilon}\right)N\boldsymbol{\Sigma}+\boldsymbol{\Phi}\mathbf{Q}\boldsymbol{\tau}_\infty^{-1}\left(\frac{\gamma^2}{\tau_m}\left(\boldsymbol{\Lambda}+\tau_\epsilon^{-1}\boldsymbol{\tau}_\infty\right)+\tau_v\psi\mathbf{I}\right)\boldsymbol{\tau}_\infty^{-1}\mathbf{Q}'\boldsymbol{\Phi}'. \tag{B.130}$$

We now want to use this expression to obtain an expression for δ in Eq. (64), which can be written as:

$$1+\delta = \frac{\sigma^2_{\mathbf{M}}}{\widehat{\sigma}^2_{\mathbf{M}}} \frac{\frac{1}{N}\mathbf{M}'\mathbf{\Sigma}\widehat{\mathbf{\Sigma}}\mathbf{M} - \sigma^2_{\mathbf{M}}\widehat{\sigma}^2_{\mathbf{M}}}{\frac{1}{N}\mathbf{M}'\mathbf{\Sigma}\mathbf{\Sigma}\mathbf{M} - \sigma^4_{\mathbf{M}}} = \frac{N\sigma^2_{\mathbf{M}}}{N\widehat{\sigma}^2_{\mathbf{M}}} \frac{N\mathbf{M}'\mathbf{\Sigma}\widehat{\mathbf{\Sigma}}\mathbf{M} - N\sigma^2_{\mathbf{M}}N\widehat{\sigma}^2_{\mathbf{M}}}{N\mathbf{M}'\mathbf{\Sigma}\mathbf{\Sigma}\mathbf{M} - N^2\sigma^4_{\mathbf{M}}}. \tag{B.131}$$

We first define the following vector:

$$\mathbf{Z} \equiv \mathbf{Q}'\mathbf{\Phi}'\mathbf{M}. \tag{B.132}$$

Denoting its $j-$th element by z_j, and using Eq. (B.105) and Eq. (B.130) we can write $\sigma^2_{\mathbf{M}}$ and $\widehat{\sigma}^2_{\mathbf{M}}$ as:

$$N\sigma^2_{\mathbf{M}} = \tau_\epsilon^{-1} + \sum_{j=1}^{J} z_j^2 \tau_\infty(\lambda_j)^{-1}, \tag{B.133}$$

$$N\widehat{\sigma}^2_{\mathbf{M}} = \left(1 + \frac{\gamma^2}{\tau_m\tau_\epsilon}\right) N\sigma^2_{\mathbf{M}} + \sum_{j=1}^{J} z_j^2 \tau_\infty(\lambda_j)^{-2} \left(\frac{\gamma^2}{\tau_m}(\lambda_j + \tau_\epsilon^{-1}\tau_\infty(\lambda_j)) + \tau_v\psi\right), \tag{B.134}$$

where the function $\tau_\infty(\cdot)$ is defined Lemma 3. After rearranging we can further write:

$$N\mathbf{M}'\mathbf{\Sigma}\mathbf{\Sigma}\mathbf{M} = \tau_\epsilon^{-1}N\sigma^2_{\mathbf{M}} + \sum_{j=1}^{J} z_j^2\tau_\infty(\lambda_j)^{-1}\left(\lambda_j\tau_\infty(\lambda_j)^{-1} + \tau_\epsilon^{-1}\right) \tag{B.135}$$

and thus:

$$N\mathbf{M}'\mathbf{\Sigma}\mathbf{\Sigma}\mathbf{M} - N^2\sigma^4_{\mathbf{M}} = \sum_{j=1}^{J} z_j^2\tau_\infty(\lambda_j)^{-1}\left(\tau_\infty(\lambda_j)^{-1}\lambda_j - \sum_{k=1}^{J} z_k^2\tau_\infty(\lambda_k)^{-1}\right). \tag{B.136}$$

Similarly,

$$N\mathbf{M}'\mathbf{\Sigma}\widehat{\mathbf{\Sigma}}\mathbf{M} = \left(1+\frac{\gamma^2}{\tau_m\tau_\epsilon}\right)\tau_\epsilon^{-1}N\sigma^2_{\mathbf{M}} \tag{B.137}$$

$$+ \sum_{j=1}^{J} z_j^2\tau_\infty(\lambda_j)^{-2}\left(\tau_\epsilon^{-1}\tau_\infty(\lambda_j) + \lambda_j\right)\left(1 + \frac{2\gamma^2}{\tau_m\tau_\epsilon} + \tau_\infty(\lambda_j)^{-1}\left(\frac{\gamma^2}{\tau_m}\lambda_j + \tau_v\psi\right)\right), \tag{B.138}$$

and thus

$$N\mathbf{M}'\mathbf{\Sigma}\widehat{\mathbf{\Sigma}}\mathbf{M} - N\sigma^2_{\mathbf{M}}N\widehat{\sigma}^2_{\mathbf{M}} = \tag{B.139}$$

$$\sum_{j=1}^{J} z_j^2\tau_\infty(\lambda_j)^{-2}\left(\frac{\gamma^2}{\tau_m}\lambda_j + \tau_v\psi + \left(1+\frac{2\gamma^2}{\tau_m\tau_\epsilon}\right)\tau_\infty(\lambda_j)\right)\left(\lambda_j\tau_\infty(\lambda_j)^{-1} - \sum_{k=1}^{J} z_k^2\tau_\infty(\lambda_k)^{-1}\right). \tag{B.140}$$

To obtain more transparent expressions we now use the approximation in Eq. (63) and Assumption 1. This assumption implies that $\mathbf{X}\mathbf{X}' \to N\mathbf{I}$ and thus

$$\frac{1}{NJ}\text{tr}(\mathbf{\Phi}'\mathbf{\Phi}) = \frac{1}{NJ}\text{tr}(\mathbf{T}^{1/2}\mathbf{X}\mathbf{X}'\mathbf{T}^{1/2}) \to \frac{1}{J}\text{tr}(\mathbf{T}) = 1, \tag{B.141}$$

so that Eq. (57) is satisfied in the limit. We now want to compute expressions of the form:

$$\lim_{J\to\infty}\sum_{j=1}^{J} z_j^2 f(\lambda_j), \tag{B.142}$$

for an arbitrary function f based on results in Bai, Miao, and Pan (2007) (among others). Under Assumption 1, $\mathbf{Q}$ is asymptotically Haar distributed. The main idea, which originates from Silverstein (1989), is to take any unit vector $\mathbf{x}$ and focus on $\mathbf{Q}'\mathbf{x} \equiv \mathbf{y}$, so that $\mathbf{y}$ is Uniformly distributed over $\{y \in \mathbb{R}^J : \|y\| = 1\}$. We then obtain that expressions like $\sum_{j=1}^{J} |y_j|^2 f(\lambda_j)$ converge to $\frac{1}{J}\sum_{j=1}^{J} f(\lambda_j)$ (Corollary 2 in Bai et al. (2007)). We would like our vector $\mathbf{Z}$ to share this key property of $\mathbf{y}$. However, although $\mathbf{x}$ is arbitrary, it must be nonrandom. We deal with this issue as follows. Note that, as pointed out in Bai and Silverstein (1998), the two matrices $\mathbf{A}_1 \equiv \mathbf{\Phi}'\mathbf{\Phi}/N = T^{1/2}\mathbf{X}\mathbf{X}'T^{1/2}$ and its companion $\mathbf{A}_2 \equiv \mathbf{\Phi}\mathbf{\Phi}'/N = \mathbf{X}'T\mathbf{X}/N$ share the same nonzero eigenvalues. Recalling our assumption that $N \geq J$, the remaining $N - J$ eigenvalues of $\mathbf{A}_2$ are zeroes. In particular, we can write the singular value decomposition of the vector $N^{-1/2}\mathbf{\Phi}'$ as:

$$\frac{1}{N^{1/2}}\mathbf{\Phi}' = \sum_{j=1}^{J}\sqrt{\lambda_j}\mathbf{q}_j\mathbf{v}_j', \tag{B.143}$$

where $\mathbf{q}_j$ is the $j-$th column of $\mathbf{Q}$ and $\mathbf{v}_j$ is the $j-$th column of the eigenvectors of $\mathbf{A_2}$. We can then write:

$$\frac{1}{N^{1/2}}\mathbf{Q}'\mathbf{\Phi}' = \sum_{j=1}^{J}\sqrt{\lambda_j}\mathbf{e}_j\mathbf{v}_j', \tag{B.144}$$

where $\mathbf{e}_j$ is a $J \times 1-$vector with $j-$th entry 1 and zeroes everywhere else. Choosing a nonrandom unit vector $\mathbf{x} \equiv N^{-1/2}\mathbf{1} = N^{1/2}\mathbf{M}$, we can rewrite $\mathbf{Z}$ as:

$$\mathbf{Z} = \sum_{j=1}^{J}\sqrt{\lambda_j}\mathbf{e}_j\mathbf{v}_j'\mathbf{x}, \tag{B.145}$$

with $j-$th entry:

$$z_j = \sqrt{\lambda_j}\mathbf{v}_j'\mathbf{x}. \tag{B.146}$$

Now, pick $f(\cdot)$ to be an arbitrary, bounded function. Since all last $N - J$ eigenvalues of $\mathbf{A}_2$ are zero, we can write:

$$\sum_{j=1}^{J} z_j^2 f(\lambda_j) = \sum_{j=1}^{J}\lambda_j(\mathbf{v}_j'\mathbf{x})^2 f(\lambda_j) = \sum_{j=1}^{N}\lambda_j\mathbf{1}_{j\leq J}(\mathbf{v}_j'\mathbf{x})^2 f(\lambda_j) \tag{B.147}$$

We can then apply Theorem 1.5 in Xi, Yang, and Yin (2020) (see also Knowles and Yin (2017))

$$\sum_{j=1}^{J} z_j^2 f(\lambda_j) \to \frac{1}{N}\sum_{j=1}^{N}\lambda_j\mathbf{1}_{j\leq J} f(\lambda_j) = \int f(\lambda)\lambda \mathrm{d}F^{\mathbf{A}_2}(\lambda), \tag{B.148}$$

where $F^{\mathbf{A}_2}$ denotes the empirical spectral density of $\mathbf{A}_2$. As noted in Bai and Silverstein (1998), it satisfies:

$$F^{\mathbf{A}_2} = (1-\psi)\mathbf{1}_{[0,\infty)} + \psi F^{\mathbf{A}_1}. \tag{B.149}$$

That is, the density $\mathrm{d}F^{\mathbf{A}_2}$ has an atom at 0 of size $1-\psi$, since a fraction $= 1 - J/N$ of the eigenvalues of $\mathbf{A}_2$ are zeroes. Since f is taken to be bounded, we eventually get:

$$\sum_{j=1}^{J} z_j^2 f(\lambda_j) \to \psi \int f(\lambda)\lambda \mathrm{d}F^{\mathbf{A}_1}(\lambda). \tag{B.150}$$

This result allows us to characterize the distortion, δ, in terms of the two statistics defined in Eq. (65). Using these definitions along with the approximation in Eq. (63), we obtain simpler expressions for:

$$N\sigma^2_{\mathbf{M}} \approx \tau_\epsilon^{-1} + \mu_\lambda(\tau_F+\tau_v)^{-1}, \tag{B.151}$$

$$N\mathbf{M}'\mathbf{\Sigma\Sigma M} - N^2\sigma^4_{\mathbf{M}} \approx (\tau_F+\tau_v)^{-2}\left((\sigma_\lambda^2+\mu_\lambda^2)/\psi - \mu_\lambda^2\right), \tag{B.152}$$

and

$$N\widehat{\sigma}^2_{\mathbf{M}} \approx \left(1+\frac{\gamma^2}{\tau_m\tau_\epsilon}\right)N\sigma^2_{\mathbf{M}} + (\tau_F+\tau_v)^{-2}\left(\frac{\gamma^2}{\tau_m}(\sigma_\lambda^2+\mu_\lambda^2)/\psi + \left(\frac{\gamma^2}{\tau_m\tau_\epsilon}(\tau_F+\tau_v)+\tau_v\right)\mu_\lambda\right), \tag{B.153}$$

$$N\mathbf{M}'\mathbf{\Sigma}\widehat{\mathbf{\Sigma}}\mathbf{M} - N\sigma^2_{\mathbf{M}}N\widehat{\sigma}^2_{\mathbf{M}} \tag{B.154}$$

$$\approx (\tau_F+\tau_v)^{-3}\begin{pmatrix} \left(\frac{\gamma^2}{\tau_m}\frac{\mu_\lambda^3+\mu_\lambda\sigma_\lambda^2}{\psi} + \left(\tau_v + \left(1+\frac{2\gamma^2}{\tau_m\tau_\epsilon}\right)(\tau_v+\tau_F)\right)\mu_\lambda^2\right)(1/\psi - 1) \\ +\frac{\gamma^2}{\tau_m}\frac{s_\lambda+2\mu_\lambda\sigma_\lambda^2}{\psi^2} + \left(\tau_v + \left(1+\frac{2\gamma^2}{\tau_m\tau_\epsilon}\right)(\tau_v+\tau_F)\right)\frac{\sigma_\lambda^2}{\psi} \end{pmatrix}. \tag{B.155}$$

Substituting these expressions in turn in Eq. (B.131) we can characterize δ in terms of the mean, dispersion and skewness of eigenvalues:

$$\delta = \frac{\gamma^2\tau_\epsilon\left(\sigma_\lambda^2\tau_\epsilon(\mu_\lambda-\sigma_\lambda)(\mu_\lambda+\sigma_\lambda) + \mu_\lambda s_\lambda\tau_\epsilon + \tau_0\left(\mu_\lambda^3+3\mu_\lambda\sigma_\lambda^2+s_\lambda\right)\right) + \tau_0\psi\left(\mu_\lambda^2+\sigma_\lambda^2\right)\left(\tau_1-\gamma^2\mu_\lambda\tau_\epsilon\right) - \mu_\lambda^2\tau_0\psi^2\tau_1}{\left(\mu_\lambda^2(1-\psi)+\sigma_\lambda^2\right)\left(\tau_m\tau_\epsilon\psi\left(\mu_\lambda\tau_\epsilon(\tau_F+2\tau_v)+\tau_0^2\right) + \gamma^2\left(\tau_\epsilon^2\left(\mu_\lambda^2+\sigma_\lambda^2\right)+\tau_0\psi(2\mu_\lambda\tau_\epsilon+\tau_0)\right)\right)}, \tag{B.156}$$

the denominator of which is strictly positive, and where, for convenience, we have defined $\tau_0 = \tau_F + \tau_v$ and $\tau_1 = \tau_0\gamma^2 + \tau_v\tau_m\tau_\epsilon$, and

$$\Delta = \mu_\lambda^4 + \frac{\tau_0^2\left(\gamma^2\mu_\lambda\tau_\epsilon(\psi-3)-\psi\tau_1\right)^2}{\gamma^4\tau_\epsilon^4} + \frac{2\mu_\lambda^2\tau_0\left(\gamma^2\mu_\lambda\tau_\epsilon(5-3\psi)+(3-2\psi)\psi\tau_1\right)}{\gamma^2\tau_\epsilon^2}. \tag{B.157}$$

So whether or not the SML looks steeper to the econometrician depends only on the sign of the numerator. Assume that the condition in Eq. (67) is satisfied. Then, if either the distribution of eigenvalues is positively skewed (or exhibits little negative skewness):

$$s_\lambda > \frac{\mu_\lambda^2\tau_0(\psi-1)\left(\gamma^2\mu_\lambda\tau_\epsilon+\psi\tau_1\right)}{\gamma^2\tau_\epsilon(\mu_\lambda\tau_\epsilon+\tau_0)}(<0), \tag{B.158}$$

or, on the contrary, if it exhibits strictly negative (but limited) skewness:

$$-\frac{\tau_\epsilon}{4(\tau_0+\mu_\lambda\tau_\epsilon)}\Delta \le s_\lambda < \frac{\mu_\lambda^2\tau_0(\psi-1)\left(\gamma^2\mu_\lambda\tau_\epsilon+\psi\tau_1\right)}{\gamma^2\tau_\epsilon(\mu_\lambda\tau_\epsilon+\tau_0)}, \tag{B.159}$$

and if, further, eigenvalues are not too concentrated:

$$\sigma_\lambda^2 > \frac{1}{2}\left(-\sqrt{\Delta+\frac{4s_\lambda(\mu_\lambda\tau_\epsilon+\tau_0)}{\tau_\epsilon}}+\mu_\lambda^2+\frac{\mu_\lambda\tau_0(3-\psi)}{\tau_\epsilon}+\frac{\tau_0\psi\tau_1}{\gamma^2\tau_\epsilon^2}\right), \tag{B.160}$$

the denominator is negative (the SML will look flatter than it actually is). Note that when $\psi\equiv 0$ we recover the limit we obtain in the single-factor case. Furthermore, to obtain the condition under which the SML is downward-sloping, note that the slope of the SML is:

$$\operatorname{Cov}(\widehat{\boldsymbol{\beta}},\mathbb{E}[\widetilde{\mathbf{R}}^e]) = \operatorname{Cov}(\widehat{\boldsymbol{\beta}},\boldsymbol{\beta})\mathbb{E}[\widetilde{R}^e_{\mathbf{M}}], \tag{B.161}$$

where the covariance in the second inequality is given by Eq. (B.153). If skewness is strictly negative:

$$s_\lambda < \frac{\mu_\lambda^2(\psi-1)\left(\gamma^2\mu_\lambda\tau_\epsilon+\tau_F\tau_\epsilon\tau_m\psi+2\psi\tau_1\right)}{\gamma^2\tau_\epsilon}, \tag{B.162}$$

and if eigenvalues are sufficiently concentrated:

$$\sigma_\lambda^2 < \frac{\mu_\lambda^2(\psi-1)\left(\gamma^2\mu_\lambda\tau_\epsilon+\tau_F\tau_\epsilon\tau_m\psi+2\psi\tau_1\right)-\gamma^2 s_\lambda\tau_\epsilon}{\gamma^2\mu_\lambda\tau_\epsilon(3-\psi)+\psi(\tau_F\tau_\epsilon\tau_m+2\tau_1)}, \tag{B.163}$$

then the covariance in the second inequality of Eq. (B.153) is negative (the SML is downward-sloping).

C Appendix (Data)

We describe here the data and the empirical tests that we build in Section 5. We download the market returns and the risk-free rate at daily frequency from Kenneth French's data library (https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html). We obtain daily security returns for all the S&P 500 stocks from the Center for Research in Security Prices (CRSP) database. First, we merge these two databases by date. Then, for each trading day of the sample, we compute yearly excess returns using rolling windows of 252 trading days (both into the past and into the future). Finally, to prepare the data for merging with IBES (see below), from the resulting dataset we keep only the last trading day of each month.

We download the monthly excess returns of the Betting Against Beta strategy from the AQR data library (https://www.aqr.com/Insights/Datasets). On the last trading day of each month, we compute yearly excess returns using rolling windows of 12 months (both into the past and into the future). We merge this dataset by date with the dataset of market and stock excess returns.

We download analyst forecast data from the Institutional Brokers' Estimate System database (I/B/E/S), and merge it with the excess returns database. Before merging, we clean the IBES database as follows (see also Engelberg et al., 2018): (*i*) we keep only firms that have above median coverage according to the field ESTIMID (in our sample, this median is 32); (*ii*) we select price targets with 12-month horizon; (*iii*) in case an institution issues multiple targets over the window we select the most recent one; (*iv*) we remove the 1st and 99th percentiles of price target forecasts; and (*v*) we compute expected excess returns as in Eq. (44). This results in a total of 429,556 expected excess return forecasts provided by 585 unique forecasters ("forecaster" is defined at the institution level, ESTIMID) over the period December 1999 to September 2019. Next, on each end-of-month date t, we go back in time 180 days and collect all the forecasts for each stock, which we align in time as forecasts at date t. We compute the consensus forecast for each stock as the median across forecasters. Using a simple regression of future excess returns on consensus excess returns, we show that indeed forecasts are strongly and positively correlated with future excess returns (slope coefficient 0.38 with Newey-West adjusted t-stat of 29.98).

To sum up, at the end of the above data-collection operations, we have 190 end-of-month dates (from Dec. 2002 to Sept. 2018), and for each date we obtain an average of 410 stocks, with the following data for each stock:

- past 1-year excess return;
- future 1-year excess return;
- expected 1-year excess return by forecaster;
- consensus expected 1-year excess return (median across forecasters).

At each end-of-month date t, we compute the value-weighted consensus expected 1-year excess return for the market portfolio. We then use the above data to compute three types of betas for each individual stock n: realized beta $\widehat{\beta}_n$, consensus beta $\beta_n^{\mathcal{C}}$, and dispersion beta $\beta_n^{\mathcal{D}}$. We compute these betas using a rolling window of 36 months, as follows:

1. For realized betas, we regress past excess returns of each individual stock on past excess returns on the market;

2. For consensus betas, we regress past consensus expected returns of each individual stock on past consensus expected returns on the market;

3. For dispersion betas, our starting point is to collect all individual stock/forecaster expected excess return forecasts over the past 36 months. We then remove the consensus forecasts from the individual forecasts and organize the data into a table that has forecasters (ESTIMID) as rows and individual stocks (cusip) as columns. Very often, we have several ESTIMID/cusip observations. In such cases, we take the last observation.

   As explained in Footnote 20, there is a tradeoff in choosing the length of the rolling window used for these computations: using a shorter window yields a sparse covariance matrix (there are not many pairs of stocks that have enough forecasts to allow computation of covariances); using a longer window contaminates the cross-sectional variation in $\mathbb{E}^{i*}[\widetilde{\mathbf{R}}^e]$ with stale observations, e.g., observations that are more than 3 years old.

   Once the table ESTIMID/cusip is built, we also compute average market capitalizations for all the stocks in the table over the last 36 months. Then, for each pair of stocks in the table, we compute the covariance among forecasts whenever we have at least two common forecasters available (the minimum number of forecasts needed to compute a covariance). We also compute the variance across forecasters for each individual stock. We then group all the covariances and variances into the matrix of co-beliefs $\text{Var}[\mathbb{E}^{i*}[\widetilde{\mathbf{R}}^e]]$ from Eq. (41).

To get an idea of the matrix of co-beliefs $\text{Var}[\mathbb{E}^{i*}[\widetilde{\mathbf{R}}^e]]$, we provide here a simplified example. Consider an economy with 2 stocks, $n \in \{a, b\}$ and 3 forecasters, $i \in \{1, 2, 3\}$. For this economy, the matrix ESTIMID/cusip is made of two vectors:

$$\begin{bmatrix} \mathbb{E}^{1*}[\widetilde{R}^e_a] & \mathbb{E}^{1*}[\widetilde{R}^e_b] \\ \mathbb{E}^{2*}[\widetilde{R}^e_a] & \mathbb{E}^{2*}[\widetilde{R}^e_b] \\ \mathbb{E}^{3*}[\widetilde{R}^e_a] & \mathbb{E}^{3*}[\widetilde{R}^e_b] \end{bmatrix}. \tag{C.164}$$

The matrix of co-beliefs $\text{Var}[\mathbb{E}^{i*}[\widetilde{\mathbf{R}}^e]]$ is the covariance matrix of these two vectors:

$$\begin{bmatrix} \text{Var}[\mathbb{E}^{i*}[\widetilde{R}^e_a]] & \text{Cov}[\mathbb{E}^{i*}[\widetilde{R}^e_a], \mathbb{E}^{i*}[\widetilde{R}^e_b]] \\ \text{Cov}[\mathbb{E}^{i*}[\widetilde{R}^e_a], \mathbb{E}^{i*}[\widetilde{R}^e_b]] & \text{Var}[\mathbb{E}^{i*}[\widetilde{R}^e_b]] \end{bmatrix}. \tag{C.165}$$

(Notice that covariances—of-diagonal elements—can be computed only if the two stocks have at least a pair of common forecasters.)

We keep only stocks for which we are able to measure covariance in beliefs with at least 100 other stocks (including itself).

Finally, using the average market capitalizations computed above we obtain the matrix of asset-specific market weights $\mathbf{M}$. Having obtained $\text{Var}[\mathbb{E}^{i*}[\widetilde{\mathbf{R}}^e]]$ and $\mathbf{M}$, the data necessary to compute $\boldsymbol{\beta}^{\mathcal{D}}$ at the end of each month (Eq. 41) is now complete.

We further winsorize our beta estimates at 0.5% (Bali et al., 2016). The regressions that we perform in Section 5 are standard and self-explanatory and we do not further elaborate on them here. Nevertheless, in building our data we have made several assumptions—most of them borrowed from the existing literature—and it is important to discuss here the robustness of our results. The results do not change significantly if:

1. We do not remove the 1st and 99th percentile of analyst forecasts;
2. We vary the past window of 180 calendar days over which we collect forecasts from 90 days to 240 days;
3. We compute consensus forecasts as average instead of median;
4. We use a rolling window for beta computations (realized, consensus, and dispersion betas) between 24 and 48 months (instead of 36 months);
5. We do not winsorize our beta estimates;
6. We use different approaches for computing dispersion betas over the same 36 month rolling window. Dispersion betas are new to the literature, and we have experimented several approaches, always with similar results. In particular, results do not change substantially if:
   (a) We change the threshold of 100 minimal covariances (see above) between 50 and 250.
   (b) When dealing with duplicates (ESTIMID/cusip), we take an average instead of the last observation, for forecasts or for market capitalization.
   (c) We require more than two common forecasters when computing covariances for dispersion betas (that is, between 2 and 10 forecasters).

D List of Symbols

$\mathbf{0}$	$N \times 1$	N-dimensional vector of zeros
$\mathbf{1}$	$N \times 1$	N-dimensional vector of ones
D		Unconditional mean of assets' payoffs
$\widetilde{\mathbf{D}}$	$N \times 1$	$\equiv \mathbf{1}D + \mathbf{\Phi}\widetilde{F} + \tilde{\boldsymbol{\epsilon}}$, assets' payoffs
e_1		$= 1/\tau + 1/\tau_\epsilon$, largest eigenvalue of matrix $\mathbf{\Sigma}$
$\widetilde{F}$		Unobserved fundamental
$\mathscr{F}^i$	$\{\widetilde{V}^i, \widetilde{\mathbf{P}}\}$	Information set of investor $i \in [0,1]$
$\widetilde{\mathbf{G}}$	$N \times 1$	Public signals (Section 4.2)
$\tilde{\mathbf{g}}$	$N \times 1$	Noise in public signals (Section 4.2)
$\mathbf{I}$	$N \times N$	Identity matrix of dimension N
$\widetilde{\mathbf{m}}$	$N \times 1$	Liquidity shocks (demand or liquidity traders)
$\mathbf{M}$	$N \times 1$	Market portfolio
N		Number of risky assets
$\mathcal{N}$		Normal distribution
$\widetilde{\mathbf{P}}$	$N \times 1$	$\equiv \mathbf{1}D + \boldsymbol{\xi}_0\mathbf{M} + \boldsymbol{\lambda}\widetilde{F} + \boldsymbol{\xi}\widetilde{\mathbf{m}}$, equilibrium prices
$\widetilde{\mathbf{R}}^e$	$N \times 1$	$\equiv \widetilde{\mathbf{D}} - \widetilde{\mathbf{P}}$, dollar excess returns
$\mathbf{T}$	$N \times 1$	Investors' tangency portfolio
$\widehat{\mathbf{T}}$	$N \times 1$	Empiricist's tangency portfolio
$\tilde{v}^i$		Noise in the private signal of investor $i \in [0,1]$
$\widetilde{V}^i$		$\equiv \widetilde{F} + \tilde{v}^i$, private signal of investor $i \in [0,1]$
$\mathbf{w}^i$		$= \mathbf{\Sigma}^{-1}\,\mathbb{E}^i[\widetilde{\mathbf{R}}^e]/\gamma$, optimal portfolio of investor $i \in [0,1]$
$\widehat{\mathbf{Z}}$	$N \times 1$	Empiricist's zero-beta portfolio
$\boldsymbol{\beta}$	$N \times 1$	True betas (Corollary 1.1)
$\widehat{\boldsymbol{\beta}}$	$N \times 1$	Empiricist's betas (Eq. 22)
$\boldsymbol{\beta}^{\mathcal{C}}$	$N \times 1$	Consensus betas (Proposition 6)
$\boldsymbol{\beta}^{\mathcal{D}}$	$N \times 1$	Dispersion betas (Proposition 6)
$\mathcal{C}^2$		Fraction in the variation of realized returns of the market explained by variation in consensus beliefs
$\mathcal{D}^2$		Fraction in the variation of realized returns of the market explained by dispersion in beliefs across investors
δ		CAPM distortion
$\tilde{\boldsymbol{\epsilon}}$	$N \times 1$	Idiosyncratic shocks to assets' payoffs (residual uncertainty)
$\mathbf{\Phi}$	$N \times 1$	Vector of exposures of assets' payoffs to the fundamental $\widetilde{F}$
$\bar{\Phi}$		Average of $\mathbf{\Phi}$
γ		Coefficient of absolute risk aversion
$\boldsymbol{\lambda}$	$N \times 1$	Exposure of prices to the fundamental factor $\widetilde{F}$
$\boldsymbol{\mu}$	$N \times 1$	$\equiv \mathbb{E}[\widetilde{\mathbf{R}}^e]$, unconditional expected excess returns
$\mu_{\mathbf{M}}$		$\equiv \mathbf{M}'\,\mathbb{E}[\widetilde{\mathbf{R}}^e]$, unconditional market expected excess return
$\mu_{\widehat{\mathbf{Z}}}$		Unconditional expected excess return of empiricist's zero-beta portfolio $\widehat{\mathbf{Z}}$
$\mathbf{\Sigma}$	$N \times N$	$= \mathbf{\Phi}\mathbf{\Phi}'/\tau + \mathbf{I}/\tau_\epsilon$, investors' variance matrix of future excess returns (Eq. 8)
$\sigma^2_{\mathbf{M}}$		$= \mathbf{M}'\mathbf{\Sigma}\mathbf{M} = \bar{\Phi}^2/\tau + 1/(N\tau_\epsilon)$, investors' variance of market's future excess returns
$\widehat{\mathbf{\Sigma}}$	$N \times N$	Empiricist's variance matrix of realized excess returns (Lemma 1)
$\widehat{\sigma}^2_{\mathbf{M}}$		$= \mathbf{M}'\widehat{\mathbf{\Sigma}}\mathbf{M}$, empiricist's variance of the realized excess returns on the market

$\boldsymbol{\xi}_0$	$N \times N$	Exposure of prices to the market portfolio $\mathbf{M}$
$\boldsymbol{\xi}$	$N \times N$	Exposure of prices to liquidity shocks $\widetilde{\mathbf{m}}$
τ_F		Precision of the fundamental $\widetilde{F}$
τ_ϵ		Precision of the idiosyncratic shocks
τ_g		Precision of noise in public signals
τ_v		Precision of noise in private signals
τ_m		Precision of the supply shocks
τ		$\equiv 1/\operatorname{Var}[\widetilde{F}\vert\mathscr{F}^i]$, precision of $\widetilde{F}$ conditional on the information set $\mathscr{F}^i$ of investor $i \in [0,1]$
τ_P		Price informativeness