The Lost Capital Asset Pricing Model

Daniel Andrei[*] Julien Cujean[†] Mungo Wilson[‡]

November 20, 2022

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 the Securities Market Line 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 days when public information reduces disagreement among investors.

We are grateful for comments received from three anonymous referees, Hengjie Ai, Laurent Barras, Jonathan Berk, Philip Bond, Michael Brennan, Bernard Dumas, Philip Dybvig, Ken French, Diego Garcia, Lorenzo Garlappi, Mark Grinblatt, Christian Hellwig, Philipp Illeditsch, Christian Julliard, Raymond Kan, Ralph Koijen, Pete Kyle, Hanno Lustig, Christoph Meinerding, Stijn Van Nieuwerburgh, Lasse Pedersen, Marcel Rindisbacher, Richard Roll, Guillaume Roussellet, Shri Santosh, Avanidhar Subrahmanyam, Katrin Tinn, 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, the 2018 NBER 2018 SI Asset Pricing, University of St. Andrew's, University of Nottingham, University of Rouen, and the Conference on Systemic Risk and Financial Stability, University of Freiburg, 2019. We thank Michael Tang, Nan Ma, Wenyun Qin, and Yujin Yang 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 Bern, 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, UK; 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 (SML), which becomes steeper occasionally, e.g., when public information reduces disagreement among investors.

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 investors' perspective, 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.

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

[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 (2021) document that the CAPM performs better when institutions demand more information. See also Chan and Marsh (2021) and Andrei, Friedman, and Ozel (2021).

empiricist perceives risky (high-beta) assets as riskier than they really are, and safe (low-beta) 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."

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 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]

The informational distance between investors and the empiricist originates from two sources of variation in investors' expectations. First, there is aggregate (time) variation in expected returns averaged across investors (consensus beliefs), 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 covariation in investors' expectations, which we refer to as *co-beliefs*. Co-beliefs 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 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

[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).

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 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.

We propose a new form of beta based on co-beliefs, *dispersion beta*, which measures how expected returns on a given stock and the market covary across investors. For instance, consider an investor who systematically underestimates returns (on the stock and on the market) 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 directions, their expectations on the stock and the market individually deviate from consensus in the same direction. If, further, each investor's deviation from consensus is larger on the stock than 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 shrink towards one relative to empiricist's traditional betas is confirmed in the data. Interestingly, this result corresponds to how 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) or to account for "regression towards the mean" (Blume, 1971), features absent in our model; in this paper, adjustment results entirely from the informational distance between investors and the empiricist. We can compare, however, how much shrinkage our theory and empirical analysis imply with how much shrinkage finance textbooks recommend (e.g., Berk and DeMarzo, 2007). The Bloomberg adjustment is: True Beta = $(2/3) \times$ Raw Beta + $(1/3) \times 1$. Our proposed adjustment is: True Beta = $(1/2) \times$ Raw Beta + $(1/2) \times 1$. Thus, the adjustment used in practice is likely too conservative, as suggested by Levi and Welch (2017).

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 investors) 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 investors playing a significant role in explaining returns on BAB. 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.

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. Investors' betas and the unconditional risk premium are identical across the two economies. However, in the economy where all information is public, the empiricist's SML is steeper and closer to the true SML. Because public information reduces disagreement among investors, empiricist's SML steepens through a compression in betas when public information dominates (e.g., on days with FOMC press conferences). We provide evidence supporting this result (see also Andersen, Thyrsgaard, and Todorov, 2021; Bodilsen, Eriksen, and Grønborg, 2021).

We also show that empiricist's SML can be downward-sloping, although the risk premium is always positive in this model. This puzzling outcome depends on the composition of the market portfolio: it occurs when assets with high measured beta simultaneously have little weight in the market portfolio. Finally, we study how our conclusions depend on modeling assumptions. When information is entirely public and the market portfolio is equally weighted, SML flattening always occurs, irrespective of modeling assumptions. When information is dispersed, the structure of payoffs matters; however, under assumptions on the structure of 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.

There are several established explanations for the finding that the SML is too flat, some of which go back to the 1970s.[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. Finally, whereas Albagli, Hellwig, and Tsyvinski (2022) emphasize that market aggregation of dispersed information makes prices

[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).

more sensitive to fundamental and noise trading shocks, how this mechanism affects the validity of the CAPM remains unattended in the literature.

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.

2 Background

Let $\widetilde{\mathbf{R}}^e$ be the vector of excess returns 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 an empiricist observes a dataset containing infinitely many realizations of $\widetilde{\mathbf{R}}^e$. We make the following assumptions (these assumptions hold in the main analysis; whenever relevant, we explain how they matter for derivations and our results):

A There is a large population (a continuum) of rational investors, who have common priors and update their beliefs using Bayes' rule.

B All investors observe the empiricist's dataset and private information about $\widetilde{\mathbf{R}}^e$. Private information is dispersed among investors, meaning that each investor i observes a different information set. Her posterior beliefs, $\mathbb{E}^i[\widetilde{\mathbf{R}}^e]$, are linear in a sufficient statistic for her information. Thus, individual and average beliefs, $\overline{\mathbb{E}}[\widetilde{\mathbf{R}}^e] \equiv \int \mathbb{E}^i[\widetilde{\mathbf{R}}^e]\mathrm{d}i$, satisfy:

$$\mathbb{E}^i[\widetilde{\mathbf{R}}^e] = \overline{\mathbb{E}}[\widetilde{\mathbf{R}}^e] + \tilde{\boldsymbol{\varepsilon}}^i. \tag{1}$$

We assume that the law of large numbers applies to $\tilde{\boldsymbol{\varepsilon}}^i$ across investors.[8]

C Investors work with the same conditional variance-covariance matrix of returns, $\mathrm{Var}^i[\widetilde{\mathbf{R}}^e]$, which is assumed constant over time.

By Assumption B each investor i's information set is a refinement of that of the empiricist. Under rationality (Assumption A), the law of total variance implies the following decompo-

[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 C.

[8] Appendix A.1 provides conditions under which Eq. (1) holds, and Sun (2006) provides the precise construction of a probability space where the exact law of large numbers holds in a continuum setting.

sition of the unconditional variance of $\widetilde{\mathbf{R}}^e$ the empiricist computes:

$$\operatorname{Var}[\widetilde{\mathbf{R}}^e] = \mathbb{E}\left[\operatorname{Var}^i[\widetilde{\mathbf{R}}^e]\right] + \operatorname{Var}\left[\mathbb{E}^i[\widetilde{\mathbf{R}}^e]\right]. \tag{2}$$

The vector $\overline{\mathbb{E}}[\widetilde{\mathbf{R}}^e] \equiv \int \mathbb{E}^i[\widetilde{\mathbf{R}}^e] \mathrm{d}i$ represents consensus beliefs about excess returns on the cross section of assets. Because we do not exclude the possibility of time variation in consensus beliefs, $\overline{\mathbb{E}}[\widetilde{\mathbf{R}}^e]$ should be considered a random variable. Assumption B implies that individual noises in investors' private information are independent of consensus beliefs. This allows us to decompose the last term in (2) as:

$$\operatorname{Var}\left[\mathbb{E}^i[\widetilde{\mathbf{R}}^e]\right] = \operatorname{Var}\left[\overline{\mathbb{E}}[\widetilde{\mathbf{R}}^e]\right] + \operatorname{Var}\left[\mathbb{E}^i[\widetilde{\mathbf{R}}^e] - \overline{\mathbb{E}}[\widetilde{\mathbf{R}}^e]\right]. \tag{3}$$

Assumption C, together with (2)-(3), leads to a decomposition of the empiricist's variance-covariance matrix of returns into *three* terms:

$$\operatorname{Var}[\widetilde{\mathbf{R}}^e] = \operatorname{Var}^i[\widetilde{\mathbf{R}}^e] + \operatorname{Var}\left[\overline{\mathbb{E}}[\widetilde{\mathbf{R}}^e]\right] + \operatorname{Var}\left[\mathbb{E}^i[\widetilde{\mathbf{R}}^e] - \overline{\mathbb{E}}[\widetilde{\mathbf{R}}^e]\right]. \tag{4}$$

The first term captures investor i's perception of uncertainty about $\widetilde{\mathbf{R}}^e$; the second term measures time variation in *consensus* beliefs; and the third term measures the *dispersion* in beliefs across investors.

We know that the law of total variance (2) always applies to individual, rational beliefs. When investors' information is homogeneous, their beliefs are identical to consensus beliefs and the last term in (4) drops out. But when information is dispersed, the law of total variance must incorporate 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 (4) shows is that cross-sectional variation *hides* in the variation the empiricist measures.

The last two terms in (4) create an informational distance between the empiricist and investors. How do we measure this distance—how do we aggregate the matrix relation (4) 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, which in our model is constant, and by $\widetilde{R}^e_{\mathbf{M}} \equiv \mathbf{M}'\widetilde{\mathbf{R}}^e$ its excess return. Applying (4) to this portfolio, dividing by $\operatorname{Var}[\widetilde{R}^e_{\mathbf{M}}]$ and shifting the first term to the left-hand side yields:

$$\underbrace{1 - \operatorname{Var}^i[\widetilde{R}^e_{\mathbf{M}}]/\operatorname{Var}[\widetilde{R}^e_{\mathbf{M}}]}_{\text{Informational distance}} = \underbrace{\operatorname{Var}[\overline{\mathbb{E}}[\widetilde{R}^e_{\mathbf{M}}]]/\operatorname{Var}[\widetilde{R}^e_{\mathbf{M}}]}_{\equiv \mathcal{C}^2} + \underbrace{\operatorname{Var}[\mathbb{E}^i[\widetilde{R}^e_{\mathbf{M}}] - \overline{\mathbb{E}}[\widetilde{R}^e_{\mathbf{M}}]]/\operatorname{Var}[\widetilde{R}^e_{\mathbf{M}}]}_{\equiv \mathcal{D}^2}. \tag{5}$$

The left-hand side of (5) represents the fraction of variation in market excess returns explained by investor i's information. This informational gap between the empiricist and investor i has two origins. First, time variation in consensus beliefs contributes $\mathcal{C}^2$ to this gap ($\mathcal{C}^2$ is the fraction of variation in market excess returns explained by variation in consensus beliefs). Second, because information is dispersed, there is variation in beliefs across investors, which accounts for the remaining distance, $\mathcal{D}^2$ (the fraction of variation in market excess returns explained by variation in beliefs across agents).

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). Our paper focuses mainly on $\mathcal{D}^2$, an estimate of which is missing in the literature and which has been neglected in CAPM tests. We will show that $\mathcal{D}^2$ is not a typical measure of dispersion in beliefs (e.g., Diether, Malloy, and Scherbina, 2002): $\mathcal{D}^2$ measures how expected returns on each *pair* of stocks *covary* across investors, for a large cross section of assets. In contrast, typical measures of dispersion in beliefs consider only variation across investors on single stocks but ignore covariation across stocks.

Equation (5) is a statistical decomposition of the informational distance between investors and the empiricist. It does not say how this informational distance distorts tests of the CAPM. To place economic restrictions on the resulting distortion, we build a model of how investors form expectations, imposing an equilibrium structure on excess returns.

3 Model

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, $\widetilde{\mathbf{D}} \equiv [\widetilde{D}_1 \ \ldots \ \widetilde{D}_N]'$, realized at the liquidation date (time 1). These payoffs are unobservable at the trading date (time 0) and have a common factor structure:

$$\widetilde{\mathbf{D}} = \begin{bmatrix} D \\ D \\ \vdots \\ D \end{bmatrix} + \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_N \end{bmatrix} \widetilde{F} + \begin{bmatrix} \tilde{\epsilon}_1 \\ \tilde{\epsilon}_2 \\ \vdots \\ \tilde{\epsilon}_N \end{bmatrix} = \mathbf{1}D + \mathbf{\Phi}\widetilde{F} + \tilde{\boldsymbol{\epsilon}}, \tag{6}$$

where D is a strictly positive scalar and $\mathbf{1}$ is a $N \times 1$ vector of ones. 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 (6), but do not observe the common factor, $\widetilde{F}$, nor stock-specific shocks, $\tilde{\boldsymbol{\epsilon}}$. Each investor i forms expectations about $\widetilde{F}$ based on information inferred from prices and information from a private signal, $\widetilde{V}^i$:

$$\widetilde{V}^i = \widetilde{F} + \tilde{v}^i. \tag{7}$$

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. To model uninformative price changes, we assume that an unmodeled group of agents trades for liquidity needs and/or non-informational reasons. Liquidity trading prevents prices from revealing $\widetilde{F}$ (Grossman and Stiglitz, 1980) and 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, and unobservable by investors; 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 (6) 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.

[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.

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

$$\widetilde{\mathbf{P}} = \mathbf{1}D + \boldsymbol{\xi}_0 \mathbf{M} + \boldsymbol{\lambda}\widetilde{F} + \boldsymbol{\xi}\widetilde{\mathbf{m}}, \tag{8}$$

where $\boldsymbol{\lambda}$ is an $N \times 1$ 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. We follow this convention and refer to $\widetilde{\mathbf{R}}^e \equiv \widetilde{\mathbf{D}} - \widetilde{\mathbf{P}}$, as *excess returns.*

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

$$\mathscr{F}^i = \{\widetilde{V}^i, \widetilde{\mathbf{P}}\}. \tag{9}$$

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

$$\tau \equiv \mathrm{Var}[\widetilde{F}|\mathscr{F}^i]^{-1} = \tau_F + \tau_v + \tau_P. \tag{10}$$

The last coefficient in (10), 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 $\boldsymbol{\Sigma} \equiv \mathrm{Var}[\widetilde{\mathbf{R}}^e|\mathscr{F}^i]$ as her conditional covariance matrix of returns (which in this model is constant and identical across investors—see Assumption C), investor i's optimal portfolio choice is

$$\mathbf{w}^i = \frac{1}{\gamma}\boldsymbol{\Sigma}^{-1}\,\mathbb{E}^i[\widetilde{\mathbf{R}}^e], \quad \text{where } \boldsymbol{\Sigma} = \frac{1}{\tau}\boldsymbol{\Phi}\boldsymbol{\Phi}' + \frac{1}{\tau_\epsilon}\mathbf{I}. \tag{11}$$

All investors in this model are mean-variance maximizers. However, 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 $\boldsymbol{\Sigma}$, that are the inputs of the standard Markowitz (1952) recipe.

Market clearing requires that the demand in (11) aggregated across informed investors and the demand of liquidity traders sum up to the market portfolio, $\mathbf{M}$:

$$\int_0^1 \mathbf{w}^i \mathrm{d}i + \widetilde{\mathbf{m}} = \mathbf{M}. \tag{12}$$

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 (11) into the market-clearing condition implies

$$\overline{\mathbb{E}}[\widetilde{\mathbf{R}}^e] = \gamma \mathbf{\Sigma}(\mathbf{M} - \widetilde{\mathbf{m}}), \tag{13}$$

which represents the expected rate of return that every particular asset must pay for informed 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:

$$\mathbf{w}^i = \mathbf{M} + \frac{1}{\gamma}\mathbf{\Sigma}^{-1} \underbrace{\left(\mathbb{E}^i[\widetilde{\mathbf{R}}^e] - \mathbb{E}[\widetilde{\mathbf{R}}^e]\right)}_{\text{Investor } i\text{'s private views}}, \tag{14}$$

which follows from (11) and the unconditional version of (13), where $\mathbb{E}[\widetilde{\mathbf{R}}^e] \equiv \mathbb{E}[\overline{\mathbb{E}}[\widetilde{\mathbf{R}}^e]]$. 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. (14) also implies that the investor who has *average unconditional beliefs*, $\mathbb{E}[\widetilde{\mathbf{R}}^e]$ and $\mathbf{\Sigma}$, holds the market portfolio, $\mathbf{M}$.

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). Although standard assumptions on CAPM fail, an *unconditional* version of CAPM holds if one appropriately defines expected returns and the market portfolio. Proposition 1 shows that unconditional expected returns $\mathbb{E}[\widetilde{\mathbf{R}}^e]$ are proportional to the market risk premium and to a new notion of beta, $\boldsymbol{\beta}$, which is based on investors' covariance matrix $\mathbf{\Sigma}$ and on the market portfolio $\mathbf{M}$.

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

$$\mathbb{E}[\widetilde{\mathbf{R}}^e] = \frac{\mathbf{\Sigma M}}{\sigma^2_{\mathbf{M}}}\,\mathbb{E}[\widetilde{R}^e_{\mathbf{M}}] = \boldsymbol{\beta}\,\mathbb{E}[\widetilde{R}^e_{\mathbf{M}}], \tag{15}$$

where $\mathbb{E}[\widetilde{R}^e_{\mathbf{M}}] \equiv \mathbf{M}'\,\mathbb{E}[\widetilde{\mathbf{R}}^e]$ *is the unconditional expected excess return on the market, and*

$$\sigma^2_{\mathbf{M}} \equiv \mathbf{M}'\mathbf{\Sigma M} = \frac{\bar{\Phi}^2}{\tau} + \frac{1}{N\tau_\epsilon} \tag{16}$$

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

Proof. Condition (13) down, pre-multiply by $\mathbf{M}'$, solve for γ, and substitute back. □

Eq. (15) is the *true* unconditional CAPM relation that holds in this economy. Note that one can also write a *conditional* CAPM relation (Easley and O'Hara, 2004; Fama and French, 2007; Van Nieuwerburgh and Veldkamp, 2010; Banerjee, 2010; Biais et al., 2010), which holds with respect to consensus beliefs, $\overline{\mathbb{E}}[\widetilde{\mathbf{R}}^e]$, and the market defined as informed investors' holdings, $\mathbf{M}-\widetilde{\mathbf{m}}$. Yet, no investor observes nor agrees on this conditional relation, as the investor with consensus beliefs is only a theoretical construct. Additionally, both $\overline{\mathbb{E}}[\widetilde{\mathbf{R}}^e]$ and $\mathbf{M}-\widetilde{\mathbf{m}}$ are unobservable to an empiricist, and thus this relation is not testable. In contrast, the two advantages of the unconditional relation (15) are that all investors agree on it and it is testable because the market portfolio $\mathbf{M}$ is observable.

The crux of our argument is that the notion of *true* beta, $\boldsymbol{\beta}$, that underlies the CAPM representation in (15) is not that of an unconditional beta as empiricists commonly compute,

$$\widehat{\beta}_n = \frac{\mathrm{Cov}[\widetilde{R}^e_n, \widetilde{R}^e_{\mathbf{M}}]}{\mathrm{Var}[\widetilde{R}^e_{\mathbf{M}}]}, \tag{17}$$

i.e., the slope of a regression of realized excess returns of asset n on realized excess returns of the market portfolio, $\mathbf{M}$. The origin of the difference between true betas and empiricist's betas becomes apparent once we take the perspective of an investor i. She perceives the cross-section of excess returns conditional on her own information as:

$$\mathbb{E}^i[\widetilde{\mathbf{R}}^e] = \boldsymbol{\beta}\,\mathbb{E}[\widetilde{\mathbf{R}}^e_{\mathbf{M}}] + \tilde{\boldsymbol{\varepsilon}}^i, \quad \text{where } \tilde{\boldsymbol{\varepsilon}}^i \equiv \gamma\boldsymbol{\Sigma}(\mathbf{w}^i - \mathbf{M}) \sim \mathcal{N}(\mathbf{0}, \mathrm{Var}[\mathbb{E}^i[\widetilde{\mathbf{R}}^e]]). \tag{18}$$

An investor i views expected returns as a noisy perturbation around the CAPM relation of Proposition 1. This perturbation arises because returns are predictable from her perspective. The more predictable returns are, as measured by $\mathrm{Var}[\mathbb{E}^i[\widetilde{\mathbf{R}}^e]]$, the larger this perturbation is. However, even though there is predictability at the investor level, the operation of the law of iterated expectations ensures that this perturbation vanishes on average. Thus, after conditioning down (18) investor i retrieves the CAPM relation of Proposition 1. Yet, the perturbation *leaves a mark* on the CAPM relation that the empiricist estimates: the typical betas in (17)—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).

Proposition 2 provides an explicit solution to the functional form of equilibrium prices.

Proposition 2. ***(Equilibrium)*** *There exists a unique linear equilibrium in which prices take the linear form in* (8) *and are explicitly given by*

$$\widetilde{\mathbf{P}} = \mathbf{1}D - \gamma\left(\frac{1}{\tau}\boldsymbol{\Phi}\boldsymbol{\Phi}' + \frac{1}{\tau_\epsilon}\mathbf{I}\right)\mathbf{M} + \boldsymbol{\Phi}\frac{\tau-\tau_F}{\tau}\widetilde{F} + \left(\frac{\gamma+\sqrt{\tau_m\tau_P}}{\tau}\boldsymbol{\Phi}\boldsymbol{\Phi}' + \frac{\gamma}{\tau_\epsilon}\mathbf{I}\right)\widetilde{\mathbf{m}}, \tag{19}$$

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

$$\tau_P(\tau_F + \tau_v + \tau_P + \tau_\epsilon)^2\gamma^2 = \tau_m \tau_\epsilon^2 \tau_v^2. \tag{20}$$

Corollary 2.1 delivers a simple expression for the true betas, $\boldsymbol{\beta}$, implied by the model.

Corollary 2.1. *The vector of true betas, $\boldsymbol{\beta}$, in this economy is*

$$\boldsymbol{\beta} = \mathbf{1} + \frac{\bar{\Phi}^2}{\tau\sigma_{\mathbf{M}}^2}\left(\frac{\boldsymbol{\Phi}}{\bar{\Phi}} - \mathbf{1}\right). \tag{21}$$

True betas in (21) are a weighted average of two vectors, $\mathbf{1}$ and $\boldsymbol{\Phi}/\bar{\Phi}$. Thus, the whole cross-section of stocks is spanned by just two vectors in equilibrium. This observation will prove useful for analyzing the empiricist's view of this economy.

4 The empiricist's view

We now examine how an empiricist views this economy. The empiricist observes a dataset containing a large number of realizations of $\widetilde{\mathbf{R}}^e$, and computes market excess returns using the market portfolio $\mathbf{M}$, $\widetilde{R}^e_{\mathbf{M}} = \mathbf{M}'\widetilde{\mathbf{R}}^e$. Proposition 1 implies this portfolio is mean-variance efficient for the *"average investor,"* the investor who holds average unconditional beliefs, $\mathbb{E}[\widetilde{\mathbf{R}}^e]$ and $\boldsymbol{\Sigma}$. That is, it commands the highest Sharpe ratio in the economy (Roll, 1977):

$$\frac{\mu_{\mathbf{M}}}{\sigma_{\mathbf{M}}} = \sqrt{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}, \tag{22}$$

where 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.

The difference between the empiricist and investors is that the empiricist only observes realized returns, but does not observe investors' information.[10] 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. Unlike investors, the empiricist rejects the unconditional CAPM.

Proposition 3. ***(CAPM rejection)*** *For the empiricist, who observes the economy ex post,*

[10]In Appendix A.6.1, 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.

the market portfolio, $\mathbf{M}$*, is not mean-variance efficient:*

$$\frac{\mu_{\mathbf{M}}}{\widehat{\sigma}_{\mathbf{M}}} < \sqrt{\boldsymbol{\mu}\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\mu}}, \tag{23}$$

where $\widehat{\boldsymbol{\Sigma}} \equiv \mathrm{Var}[\widetilde{\mathbf{R}}^e]$ *and* $\widehat{\sigma}_{\mathbf{M}} \equiv \sqrt{\mathbf{M}'\widehat{\boldsymbol{\Sigma}}\mathbf{M}}$.

This follows directly from arguments of Section 2. The law of iterated expectations ensures that the empiricist correctly measures $\boldsymbol{\mu}$ and $\mu_{\mathbf{M}}$. The law of total variance (2), however, implies that due to return predictability, the empiricist obtains variances and covariances of asset returns that differ from investors' own estimates. Therefore, for the empiricist, all assets have the correct unconditional expected returns but display systematically larger unconditional variances. Figure 1 illustrates this: although $\mathbf{M}$ is the correct market portfolio, it is not the tangency portfolio for the empiricist, $\widehat{\mathbf{T}}$, nor is it mean-variance efficient and she rejects the CAPM, the result of Proposition 3.

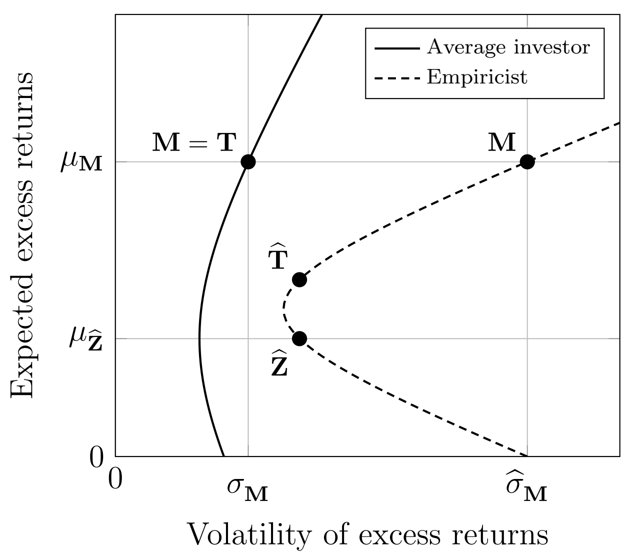

Figure 1: **CAPM rejection.** This figure compares the minimum-variance set under average unconditional beliefs (solid line) with the minimum-variance set of the empiricist (dashed line). For the average investor, the market portfolio is the tangency portfolio ($\mathbf{M} = \mathbf{T}$), but for the 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.

Because the empiricist uses different covariances and variances, she obtains different betas—empiricist's betas in (17) are mismeasured. As a result, although the empiricist finds a linear relation between expected excess returns and betas, this relation has a positive intercept.[11] This intercept is the expected excess return, $\mu_{\widehat{\mathbf{Z}}}$, of the empiricist's *zero-beta* portfolio $\widehat{\mathbf{Z}}$, the unique minimum-variance portfolio that is uncorrelated with $\mathbf{M}$. Figure 1 shows that $\widehat{\mathbf{Z}}$ lies on the lower limb of the minimum-variance set, where $\mu_{\widehat{\mathbf{Z}}}$ corresponds to the intercept of the line that is tangent to the minimum-variance set at $\mathbf{M}$. Proposition 4 describes the linear relation between betas and expected returns that the empiricist measures.

Proposition 4. ***(CAPM tests based on realized returns)*** *The empiricist observes a zero-beta CAPM (Black, 1972):*

$$\boldsymbol{\mu} = \mathbf{1}\mu_{\widehat{\mathbf{Z}}} + \widehat{\boldsymbol{\beta}}(\mu_{\mathbf{M}} - \mu_{\widehat{\mathbf{Z}}}), \tag{24}$$

in which the empiricist's vector of betas, $\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\Sigma}}\mathbf{M}/\widehat{\sigma}^2_{\mathbf{M}}$, satisfies the proportionality relation:

$$\widehat{\boldsymbol{\beta}} - \mathbf{1} = (1+\delta)(\boldsymbol{\beta} - \mathbf{1}), \tag{25}$$

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

$$\delta = \frac{\mu_{\widehat{\mathbf{Z}}}}{\mu_{\mathbf{M}} - \mu_{\widehat{\mathbf{Z}}}} \geq 0. \tag{26}$$

The coefficient δ measures the *distortion* of the CAPM relation, as estimated by the empiricist, relative to the true CAPM in (15). In our model, the zero-beta portfolio always has positive expected excess returns. Thus the empiricist perceives a flat Securities Market Line (SML) with a positive intercept. Notably, Black (1972) writes that dispersed information *"does not change the structure of capital asset prices in any significant way."*[12] But in the context of our model, this argument is incorrect since Proposition 4 shows that dispersed information, in fact, has precisely the same effect as that of borrowing constraints.

Further replacing $\mu_{\widehat{\mathbf{Z}}}$ in (24) produces the following relation:

$$\boldsymbol{\mu} = \mathbf{1}\frac{\delta}{1+\delta}\mu_{\mathbf{M}} + \widehat{\boldsymbol{\beta}}\frac{1}{1+\delta}\mu_{\mathbf{M}}, \tag{27}$$

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. The proportionality relation between betas in (25) means

[11]The empiricist's relation remains linear because the market portfolio, although not mean-variance efficient, remains minimum-variance (Roll, 1977, Corollary 6). 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 $\widehat{\mathbf{T}}$. We study these distorting effects in Section 6.

[12]He refers specifically to the work of Lintner (1969).

that the empiricist inflates all betas above the market's beta, which is 1, and deflates all others. Hence, the empiricist perceives risky (high-beta) assets as riskier than they really are, and safe (low-beta) assets as safer than they really are. As illustrated in Figure 2, the empiricist's SML rotates clockwise around the market portfolio, which flattens its slope and creates a positive intercept.

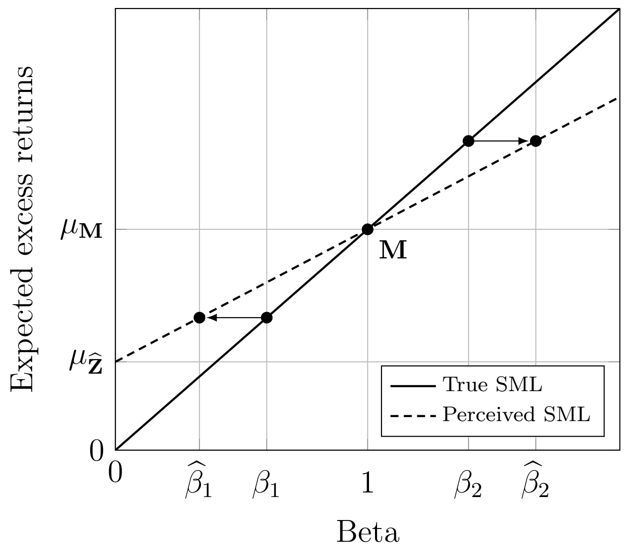

Figure 2: **CAPM distortion.** This figure illustrates the main result of the paper. The perceived SML is flatter than the actual SML in equilibrium. The dashed line and the solid line show the true and perceived SML. **M** represents the unconditional market portfolio.

In equilibrium, true betas are shrunk towards one relative to empiricist's betas.[13] The degree of shrinkage is determined by δ, which adjusts the empiricist's betas towards true betas. Interestingly, (25) is identical to the Bayesian estimator proposed by Vasicek (1973), an estimator that is popular in the financial industry ("ADJ BETA" on Bloomberg terminals).[14] We emphasize, however, that the result of Theorem 4 is not due to sampling error. Nor is this result a standard attenuation bias, which commonly plagues cross-sectional regressions in the Fama and MacBeth (1973) method.[15] Rather, in equilibrium shrinkage in betas is due to the informational distance between investors and the empiricist.

[13]Shrinkage towards one follows from the assumption that the market portfolio **M** is minimum-variance (because we assume it is equally weighted). If this were not the case, the beta of each asset would shrink towards a number determined by its relative size in the market portfolio (see Section 6.2).

[14]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.

[15]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.

We are interested in measuring the coefficient δ in terms of the informational distance between investors and the empiricist, a matter to which we turn next.

4.1 Aggregate versus cross-sectional variation

Eq. (4) in Section 2 shows that 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, $\overline{\mathbb{E}}[\widetilde{\mathbf{R}}^e]$, as can be seen from Eq. (13). Second, dispersed information generates cross-sectional dispersion in expected returns across investors, $\mathrm{Var}[\mathbb{E}^i[\widetilde{\mathbf{R}}^e] - \overline{\mathbb{E}}[\widetilde{\mathbf{R}}^e]]$. In our equilibrium framework, the statistical relation (4) becomes an endogenous relation.

Lemma 1. *The empiricist's covariance matrix of realized excess returns satisfies*

$$\widehat{\mathbf{\Sigma}} = \mathbf{\Sigma} + \underbrace{\frac{\gamma^2}{\tau_m}\left(\frac{1}{\tau_\epsilon}\mathbf{\Sigma} + \frac{e_1}{\tau}\mathbf{\Phi}\mathbf{\Phi}'\right)}_{\text{Consensus covariance matrix}} + \underbrace{\frac{\tau_v}{\tau^2}\mathbf{\Phi}\mathbf{\Phi}'}_{\text{Matrix of co-beliefs}}, \tag{28}$$

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

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

The *consensus covariance matrix* measures aggregate variation in consensus expected returns, $\mathrm{Var}[\overline{\mathbb{E}}[\widetilde{\mathbf{R}}^e]]$, whereas the *matrix of co-beliefs* measures dispersion in beliefs across investors, $\mathrm{Var}[\mathbb{E}^i[\widetilde{\mathbf{R}}^e] - \overline{\mathbb{E}}[\widetilde{\mathbf{R}}^e]]$. The diagonal elements of this latter matrix have the traditional meaning of disagreement. However, its off-diagonal elements reflect the extent to which expected returns on pairs of stocks covary across investors, which is different from disagreement. Lemma 1 further shows that a unique, endogenous scalar, e_1, determines the informational distance between $\widehat{\mathbf{\Sigma}}$ and $\mathbf{\Sigma}$. This scalar represents the largest eigenvalue of $\mathbf{\Sigma}$ (the other eigenvalues, of multiplicity $N-1$, being $1/\tau_\epsilon$).[16]

Section 2 shows how to summarize all elements of the consensus covariance and co-beliefs matrices with just two numbers, $\mathcal{C}^2$ and $\mathcal{D}^2$. They are fractions of variation in market excess returns explained by, respectively, variation in consensus beliefs and dispersion in beliefs:

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

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

[16] This uniqueness result follows from our assumption of a single common factor in payoffs, which implies that $\mathbf{\Sigma}$ has two distinct eigenvalues.

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

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

Excess market variance is key to generating distortion in beta. From investors' perspective, the ratio of market variance, $\sigma_{\mathbf{M}}^2$, to fundamental variance, $\mathrm{Var}^i[\bar{\Phi}\widetilde{F}] = \bar{\Phi}^2/\tau$, minus 1 represents market excess variance, similar to the definition of excess volatility (Shiller, 1981). There can be no distortion without excess market variance. Furthermore, 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; Lewellen and Nagel, 2006). Yet, to our knowledge, the impact of dispersion in beliefs has been neglected in CAPM tests. To the extent that $\mathcal{C}^2$ is given a weight lower than one, (32) shows that the impact of $\mathcal{D}^2$ is potentially stronger than that of $\mathcal{C}^2$.

In terms of comparative statics, we show in Appendix A.6 (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 surprising if one considers that the ratio of the two, γ^2/τ_m, captures aggregate variation in expected returns, which always increases δ (see Section 6.1). Further discussion of mathematical properties of δ may be found in Appendix A.6, and an analysis of δ in a large economy in Section 6.3.

4.2 Public versus private information

We now 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 and there is no disagreement among investors. We call this economy the *Common Information Economy* (CIE).[17] We adopt the following notation: we add "CIE" as a subscript to variables whose value differs in the common information economy. All variables without a "CIE" subscript have identical values to that in the baseline model.

[17] Bacchetta and Wincoop (2006) and Albagli et al. (2022) perform a similar exercise as we do here, with the same goal of isolating the role of dispersed information. It is perhaps tempting to create a benchmark economy with dispersion but no aggregate variation in beliefs ($\mathcal{D}^2 > 0$ and $\mathcal{C}^2 = 0$). However, in our setup this requires eliminating noise in supply (the only source of aggregate variation in expected returns). Since prices would be fully informative in that case, cross-sectional variation would only persist under additional behavioral assumptions (e.g., investors "agree to disagree").

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:

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

where $\widetilde{\mathbf{P}}^a \equiv \widetilde{\mathbf{P}} - \mathbf{1}D - \boldsymbol{\xi}_0\mathbf{M}$. To map the information structure of the main model, we assume that investors in the CIE observe a vector $\widetilde{\mathbf{G}}$ of N exogenous, public signals:

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

To maintain identical informational content about the common factor, $\widetilde{F}$, in the CIE and the baseline economy, we choose τ_g such that the precision τ is identical in both economies; the following equation determines τ_g (Appendix A.7 provides analytical details):

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

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, unlike investors in the CIE, investors in the baseline model observe information from private signals. 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, $\boldsymbol{\xi}^{-1}\widetilde{\mathbf{P}}^a$.

Because τ is identical in both economies, investors' conditional covariance matrix of future expected returns, $\boldsymbol{\Sigma}$, is identical in the CIE and the baseline model. As a result, the vector of true betas, $\boldsymbol{\beta}$, is the same as in the baseline model. Thus investors in the CIE also observe the unconditional CAPM relation in Proposition 1. But the view of the empiricist *does* change—the empiricist observes a stronger CAPM relation in the CIE.

Proposition 6. *The distortion of the SML is lower in the common information economy:*

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

In the CIE the informational gap resulting from aggregate variation is always larger, $\mathcal{C}_{\text{CIE}}^2 > \mathcal{C}^2$ (see Appendix A.7), but dispersion in investors' information is absent, $\mathcal{D}_{\text{CIE}}^2 = 0$. On balance reduced dispersion leads to lower SML distortion, $\delta_{\text{CIE}} < \delta$. Thus, empiricist's betas get closer to true betas. This result is a form of over-reaction of the kind studied in Albagli et al. (2022). Since we have concluded from (35) that the public signals $\widetilde{\mathbf{G}}$ must be more informative that market prices are in the baseline model ($\tau_g > \tau_m$), it follows that in the model with dispersed information investors act as if they were treating market (public) information as more informative than it truly is, exacerbating price sensitivity to fundamental shocks. This excess price sensitivity, or *information updating wedge* (Albagli et al., 2022), increases variation measured by the empiricist and further distorts beta estimates (see Appendix A.7).

Reduced disagreement among investors causes a *compression* in empiricist's betas. Given that in the CIE investors do not disagree and $\delta_{\text{CIE}} < \delta$, beta compression follows from Proposition 4, which links the cross-sectional dispersion in $\widehat{\boldsymbol{\beta}}_{\text{CIE}}$ to that in $\boldsymbol{\beta}$:

$$\sigma_{\widehat{\boldsymbol{\beta}}_{\text{CIE}}} = (1 + \delta_{\text{CIE}})\sigma_{\boldsymbol{\beta}}. \tag{37}$$

Since $\sigma_{\boldsymbol{\beta}}$ is identical in the CIE (true betas remain unchanged), for the empiricist $\delta_{\text{CIE}} < \delta$ translates into a beta compression. A stronger CAPM in the CIE arises solely from this compression (investors face, by construction, the same risk and the unconditional market risk premium remains the same across economies). Our prediction is that days on which public information crowds out private information and resolves disagreement are accompanied by a beta compression. We provide evidence supporting this result in Section 5.4.

5 Empirical Tests

In this section we assess quantitatively the theoretical predictions of our model. Note that these predictions are based on dollar returns, as in CARA-normal models dollar returns are normally distributed and deliver analytical expressions. However, the empirical analysis of this section uses rates of return, under which our theoretical results are intractable. We discuss this matter in Appendix A.13, where we show by means of numerical simulations and first-order approximations that all our theoretical predictions may be considered to apply equally to rates of return.

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 investors' betas, $\boldsymbol{\beta}$, and empiricist's betas, $\widehat{\boldsymbol{\beta}}$.

Proposition 7. *The empiricist's betas are a weighted average of three types of beta:*

$$\widehat{\boldsymbol{\beta}} = (1 - \mathcal{C}^2 - \mathcal{D}^2)\boldsymbol{\beta} + \mathcal{C}^2\boldsymbol{\beta}^{\mathcal{C}} + \mathcal{D}^2\boldsymbol{\beta}^{\mathcal{D}}, \tag{38}$$

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

$$\boldsymbol{\beta}^{\mathcal{C}} \equiv \frac{\operatorname{Var}[\overline{\mathbb{E}}[\widetilde{\mathbf{R}}^e]]\mathbf{M}}{\operatorname{Var}[\overline{\mathbb{E}}[\widetilde{R}^e_{\mathbf{M}}]]} \quad \textit{and} \quad \boldsymbol{\beta}^{\mathcal{D}} \equiv \frac{\operatorname{Var}[\mathbb{E}^{i*}[\widetilde{\mathbf{R}}^e]]\mathbf{M}}{\operatorname{Var}[\mathbb{E}^{i*}[\widetilde{R}^e_{\mathbf{M}}]]}. \tag{39}$$

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. 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 example illustrates the meaning of $\beta^{\mathcal{D}}_n$. 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 $\beta^{\mathcal{D}}_n = 2$.[18] 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 that captures whether or not investors' expectations on asset n and 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.

Corollary 7.1. *In the equilibrium of the model,* $\boldsymbol{\beta}^{\mathcal{C}}$ *and* $\boldsymbol{\beta}^{\mathcal{D}}$ *are given by*

$$\boldsymbol{\beta}^{\mathcal{C}} = \boldsymbol{\beta} + \left(\frac{\tau\sigma^2_{\mathbf{M}}}{\bar{\Phi}^2} - 1\right)\frac{\tau_\epsilon e_1 \bar{\Phi}^2}{\tau\sigma^2_{\mathbf{M}} + \tau_\epsilon e_1 \bar{\Phi}^2}(\boldsymbol{\beta} - \mathbf{1}) \tag{40}$$

$$\boldsymbol{\beta}^{\mathcal{D}} = \boldsymbol{\beta} + \left(\frac{\tau\sigma^2_{\mathbf{M}}}{\bar{\Phi}^2} - 1\right)(\boldsymbol{\beta} - \mathbf{1}). \tag{41}$$

The coefficients of $(\boldsymbol{\beta}-\mathbf{1})$ in (40)-(41) are identical to those of $\mathcal{C}^2$ and $\mathcal{D}^2$ in the definition of the distortion δ in Proposition 5, and thus feature the excess variance term $\tau\sigma^2_{\mathbf{M}}/\bar{\Phi}^2 - 1$. Corollary 7.1 shows that both $\boldsymbol{\beta}^{\mathcal{C}}$ and $\boldsymbol{\beta}^{\mathcal{D}}$ are more dispersed than true betas: when true betas are larger (smaller) than one, $\boldsymbol{\beta}^{\mathcal{C}}$ and $\boldsymbol{\beta}^{\mathcal{D}}$ are both larger (smaller) than $\boldsymbol{\beta}$. The extent of excess variance in the market dictates the magnitude of these effects.

The empirical challenge in testing our theory is to measure $\mathcal{C}^2$ and $\mathcal{D}^2$ and their associated betas, $\boldsymbol{\beta}^{\mathcal{C}}$ and $\boldsymbol{\beta}^{\mathcal{D}}$, for a cross-section of stocks. This, in turn, involves finding proxies for

[18]Formally, $\beta^{\mathcal{D}} = \operatorname{Cov}[\mathbb{E}^{i*}[\widetilde{R}^e_n], \mathbb{E}^{i*}[\widetilde{R}^e_{\mathbf{M}}]]/\operatorname{Var}[\mathbb{E}^{i*}[\widetilde{R}^e_{\mathbf{M}}]] = 2$, where the covariance in the numerator and the variance in the denominator are computed across investors.

consensus beliefs, $\overline{\mathbb{E}}[\widetilde{\mathbf{R}}^e]$, and individual beliefs, $\mathbb{E}^i[\widetilde{\mathbf{R}}^e]$. The objective of the next section is to use the I/B/E/S database on analysts' forecasts to estimate $\boldsymbol{\beta}^{\mathcal{C}}$ and $\boldsymbol{\beta}^{\mathcal{D}}$ and assess their ability to price the cross-section of returns.

5.1 Variable definitions and summary statistics

We obtain daily market excess returns and risk-free security returns from Kenneth French's data library. The Center for Research in Security Prices (CRSP) database provides daily excess returns and market capitalizations for all stocks listed in the S&P 500 index. The list of historical constituents is available from Compustat. The AQR data library provides monthly excess returns on "Betting Against Beta" (BAB). Finally, the Institutional Brokers' Estimate System database (I/B/E/S) provides unadjusted data on price targets from 1999 to 2019, which is the sample period we use for our tests.[19]

Using these databases, 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; 1-year *future* excess return; and 1-year *expected* excess return. Past and future excess returns are unique for each stock-date observation. We construct expected excess returns over a lookback window of 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. We then proxy for $\mathbb{E}^i[\widetilde{R}^e_n]$ using institution i's expected excess return:

$$\mathbb{E}^i[\widetilde{R}^e_n] = \frac{\text{Price Target}^i_{n,t+12\text{ months}} - \text{Price}_{n,t}}{\text{Price}_{n,t}} - \text{RF}_t, \tag{42}$$

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 the proxy for expected excess returns (42), we make several choices, which we outline in Appendix B, together with other data-cleaning details that closely follow the strategy in Engelberg, McLean, and Pontiff (2018). We obtain a total of 429,556 expected excess return data points issued by 585 unique forecasters from December 1999 to September 2019. Then, at the end of 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.

[19] Appendix B describes 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 our choices in this section.

We first verify whether the following inequality holds for each individual stock:

$$\mathrm{Var}[\widetilde{R}^e_n] > \mathrm{Var}\left[\overline{\mathbb{E}}[\widetilde{R}^e_n]\right] + \mathrm{Var}\left[\mathbb{E}^i[\widetilde{R}^e_n] - \overline{\mathbb{E}}[\widetilde{R}^e_n]\right]. \tag{43}$$

If (43) fails, the data for stock n violates at least one of the assumptions made in Section 2. This situation can arise, for instance, when there is a high level of forecast dispersion, and the last term in (43) dominates. One such case is when the common prior assumption A does not hold, i.e., analysts just add noise, rather than removing, to the common prior forecast. We find that (43) fails for 7 stocks out of 556, or 1.3% of them. Removing these 7 individual stocks has a negligible impact on our empirical analysis.

Our initial assumptions also imply that analysts' forecasts must be positively related to future excess returns (the opposite would violate assumption A, which relies on Bayes' rule). Therefore, we also examine this matter in the data: regressing future excess returns on analysts' forecasts yields positive and statistically significant coefficients for individual analyst forecasts (0.0992, $p < 0.01$) and consensus forecasts (0.3816, $p < 0.01$). These tests suggest that the assumptions made in Section 2 hold reasonably well in our data.

At this stage, the dataset necessary to compute realized betas, $\widehat{\boldsymbol{\beta}}$, consensus betas, $\boldsymbol{\beta}^{\mathcal{C}}$, and dispersion betas, $\boldsymbol{\beta}^{\mathcal{D}}$, is complete. We compute $\widehat{\beta}_n$ as the slope coefficient from regressing stock n's past excess returns on past market excess returns. Similarly, $\beta^{\mathcal{C}}_n$ 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 $\beta^{\mathcal{D}}_n$ (we do not directly observe investors' beliefs about market returns). We thus reconstruct $\boldsymbol{\beta}^{\mathcal{D}}$ directly using its definition in Proposition 7. The central piece in obtaining $\boldsymbol{\beta}^{\mathcal{D}}$ is the matrix of "co-beliefs," $\mathrm{Var}[\mathbb{E}^{i*}[\widetilde{\mathbf{R}}^e]]$, or the covariation in investors expectations for each pair of stocks. We explain in Appendix B how to obtain the matrix of co-beliefs at each date t. We then obtain $\boldsymbol{\beta}^{\mathcal{D}}$ from the matrix of co-beliefs and assets' market weights, as defined in (39).

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.[20] 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.

Over the sample period, the average market excess return was 9.85% per year, with a 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

[20] We lose 3 years of data at the beginning of the sample due to the initial window over which we estimate betas, and one year at the end of the sample to obtain future 1-year returns. All betas are winsorized at the 0.5% and 99.5% percentile levels (Bali, Engle, and Murray, 2016).

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 a 1-year horizon.

the table compute averages and standard deviations over the entire sample for the three sets of betas; all betas have averages close to one and high degrees of variation across stocks and time, with standard deviations well above one.

The average magnitudes of $\mathcal{C}^2$ and $\mathcal{D}^2$ in our sample are $\mathcal{C}^2 = 0.0400$ and $\mathcal{D}^2 = 0.0563$.[21] Our estimate of $\mathcal{C}^2$ is lower than what is typically found in the literature, which indicates that analysts' forecasts likely exhibit less aggregate variation than other proxies for expected returns.[22] An estimate of $\mathcal{D}^2$, which measures how much variation in market returns is explained by dispersion in expectations, is missing in the literature. The magnitude of $\mathcal{D}^2$, higher than that of $\mathcal{C}^2$ in our sample, points to an important source of variation in expected returns that has been neglected in CAPM tests.

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, $\mathcal{D}^2$ also includes off-diagonal elements that reflect the extent to which expected returns on pairs of stocks covary across investors, which is not the same as disagreement:

$$\mathcal{D}^2 \equiv \operatorname{Var}[\widetilde{R}^e_{\mathbf{M}}]^{-1} \left(\sum_{n=1}^{N} M_n^2 \operatorname{Var}[\mathbb{E}^{i*}[\widetilde{R}^e_n]] + \sum_{n \neq m} \sum M_n M_m \operatorname{Cov}[\mathbb{E}^{i*}[\widetilde{R}^e_n], \mathbb{E}^{i*}[\widetilde{R}^e_m]] \right). \tag{44}$$

Does our estimate of $\mathcal{D}^2$ imply empirically plausible primitive parameters for our model?

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

[22] $\mathcal{C}^2$ represents variation in expected returns on the market divided by variation in realized returns. Several estimates of this ratio exist in the literature. 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 approximately 11%, using return-forecasting regressions.

To examine this matter, we write (31) as

$$\mathcal{D}^2 = \frac{\tau_v}{\tau} \frac{\bar{\Phi}^2}{\tau \widehat{\sigma}_{\mathbf{M}}^2}. \tag{45}$$

The last term is a ratio of fundamental variance, $\mathrm{Var}^i[\bar{\Phi}\widetilde{F}] = \bar{\Phi}^2/\tau$, to market variance, $\widehat{\sigma}_{\mathbf{M}}^2$, for which a rough estimate is simple to obtain: given a realized market return volatility of 16% (Table 1) and a fundamental volatility between 3% and 6%, one obtains $\bar{\Phi}^2/(\tau\widehat{\sigma}_{\mathbf{M}}^2) \in (0.035, 0.141)$. On the other hand, we obtain an estimate of τ_v/τ directly from our data: writing the law of total variance (2) for an individual asset and assuming that all variation in expectations on the right-hand side is due to private information implies $\tau_F^{-1} = (\tau_F + \tau_v)^{-1} + \mathrm{Var}[\mathbb{E}^i[\widetilde{R}_n^e]]$, or $\tau_v/\tau = \tau_F \mathrm{Var}[\mathbb{E}^i[\widetilde{R}_n^e]] = \mathrm{Var}[\mathbb{E}^i[\widetilde{R}_n^e]]/\mathrm{Var}[\widetilde{R}_n^e]$. For 90% of firms in our sample, we obtain $\tau_v/\tau \in (0.07, 0.53)$, with an overall average of 0.24.[23] Going back to estimated values for $\bar{\Phi}^2/(\tau\widehat{\sigma}_{\mathbf{M}}^2)$ in (45), our model needs a ratio of at least $\tau_v/\tau = 0.4$ to match the empirical value of $\mathcal{D}^2 = 0.0563$. This number seems plausible and within the range above.

5.2 Empirical tests and a plausible magnitude for δ

Our analysis starts with classical portfolio sorts. We form five (measured) beta-sorted portfolios,[24] which we use to 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 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.

However, less known in the literature is that portfolios with higher average betas also have higher consensus betas, $\beta^{\mathcal{C}}$, and higher dispersion betas, $\beta^{\mathcal{D}}$. Can this explain the high alphas on low-beta portfolios? We address this question in Table 3. Panel (a) shows the 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.

[23] Alternatively, in the literature, Cho and Krishnan (2000) have estimated the primitive parameters of the Hellwig (1980) single-asset noisy rational-expectations model: the standard deviation of the prior (our τ_F^{-2}) is 5.495, and the standard deviation of the private signal (our τ_v^{-2}) ranges from 10.067 to 23.358 (see their Table 2). Assuming $\tau = \tau_F + \tau_v$, we obtain $\tau_v/\tau \in (0.05, 0.23)$. These are upper bounds for τ_v/τ, since τ may be higher than $\tau_F + \tau_v$ (due to learning from public signals such as prices). The estimate we find is slightly higher, perhaps due to the longer horizon of our data.

[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.

[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).

	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 annual 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 $^{***}p < .01$.

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), alphas on low-beta portfolios lose statistical power (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.

The negative premium earned by $\beta^{\mathcal{C}}$ and $\beta^{\mathcal{D}}$ is in fact a prediction of our model. This can be seen by substituting the relation (25) into the statistical relation of Proposition 7:

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

We can express this relation in terms of expected returns:[26]

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

Thus, $\beta^{\mathcal{C}}$ and $\beta^{\mathcal{D}}$ earn a negative premium in the model, in line with panel (b) of Table 2.

We now test the two cross-sectional relations (46) and (47) following a two-step method. For the beta relation (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

[26] To obtain (47), multiply (38) by $\mathbb{E}[\widetilde{R}^e_{\mathbf{M}}]$ and substitute the true CAPM relation $\mathbb{E}[\widetilde{\mathbf{R}}^e] = \boldsymbol{\beta}\,\mathbb{E}[\widetilde{R}^e_{\mathbf{M}}]$.

(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 $^{***}p < .01$.

coefficient. We then examine whether the average of each time series differs from zero. This procedure follows Fama and MacBeth (1973, FM hereafter), with the exception that the left-hand variable is realized beta, as opposed to future excess return. Table 4 reports the average coefficients and provides empirical support for (46).

In Table 5 we repeat the FM regressions for (47), according to which consensus betas, dispersion betas, and realized betas should explain the cross-section of expected returns. This relation also predicts that $\boldsymbol{\beta}^{\mathcal{C}}$ and $\boldsymbol{\beta}^{\mathcal{D}}$ earn a negative premium, which the first row

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 $^{***}p < .01$.

of the table confirms, with negative and statistically significant coefficients on $\boldsymbol{\beta}^{\mathcal{C}}$ and $\boldsymbol{\beta}^{\mathcal{D}}$. The other rows of Table 5 present estimates when only including $\boldsymbol{\beta}^{\mathcal{C}}$ or $\boldsymbol{\beta}^{\mathcal{D}}$ as explanatory variables, along with estimates for the canonical CAPM. The CAPM slope strengthens in magnitude and statistical significance when $\boldsymbol{\beta}^{\mathcal{C}}$ and $\boldsymbol{\beta}^{\mathcal{D}}$ are added to the regression.

a_0	a_R	$a_{\mathcal{C}}$	$a_{\mathcal{D}}$	Adj.R^2	N
0.1311***	0.0224**	-0.0095***	-0.0063***	0.0336	410
(5.91)	(2.20)	(-2.87)	(-2.92)		
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)		

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 $^{***}p < .01$.

The estimates of Table 4 allow us to perform a magnitude check of δ. Specifically, Table 4 shows that the intercept a_0 belongs to the 90% confidence interval $a_0 \in [0.77, 0.88]$. One can thus obtain a 90% confidence interval for δ using (46), which implies:

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

In Figure 3 we plot this interval (shaded area) as a function of the informational gap, $\mathcal{C}^2 + \mathcal{D}^2$.

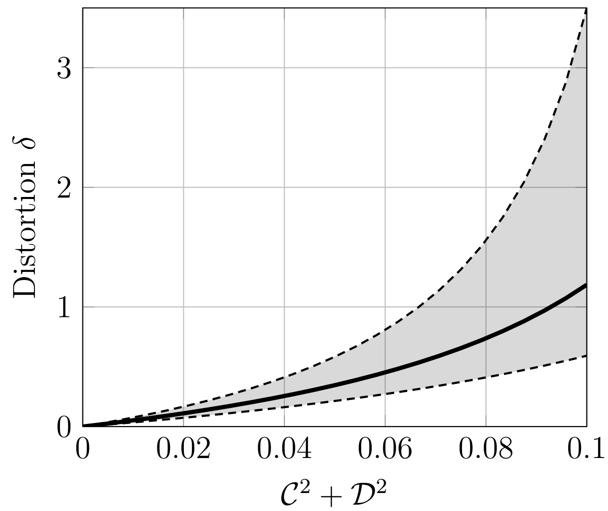

Figure 3: This figure plots the empirically plausible range for the distortion δ. The shaded area shows the 90 percent confidence region for δ based on (48) and a 90 percent confidence range for the intercept a: $a \in [0.77, 0.88]$. The distortion is plotted as a function of the informational distance between investors and the empiricist, $\mathcal{C}^2 + \mathcal{D}^2$. The data is from 2002/12/31 to 2018/9/28 (190 months).

The plot shows that δ ranges from 0.5 to 3 for $\mathcal{C}^2 + \mathcal{D}^2 = 9.63\%$ 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, $\delta = 1.1$. The 90% confidence interval shows that the distortion can be 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 (32) yields:

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

The first term in $\widehat{\boldsymbol{\alpha}}$ has been discussed in the literature.[27] The second term is new. To

[27] See, e.g., Jagannathan and Wang (1996); Lewellen and Nagel (2006); Boguth et al. (2011).

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 variance. Based on realized 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 $\tau/\bar{\Phi}^2 = 1111$ and therefore $(\tau\sigma^2_{\mathbf{M}}/\bar{\Phi}^2 - 1) = 25.22$. 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 and dispersion beta help explain the under-performance of high-beta stocks relative to low-beta stocks (Friend and Blume, 1970; Black et al., 1972). This empirical anomaly is illustrated in Figure 4, which compares mean excess returns for the five beta-sorted portfolios of Table 2 to their CAPM-implied excess returns.

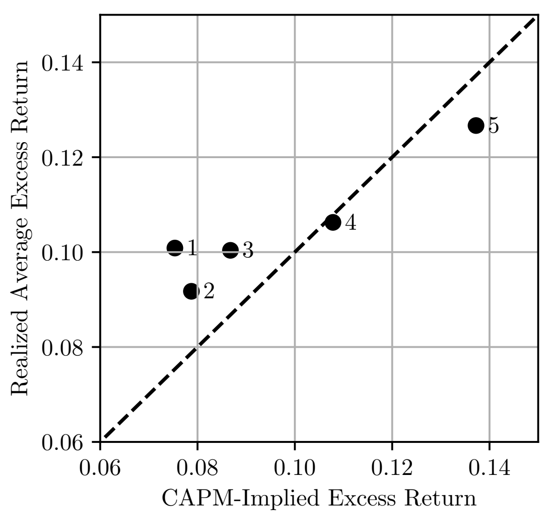

Figure 4: This figure plots the CAPM-implied excess returns versus realized average excess returns (value-weighted) on five beta-sorted portfolios, with the smallest beta stocks in portfolio 1 and the largest beta stocks in portfolio 5. The data is from 2002/12/31 to 2018/9/28 (190 months).

Frazzini and Pedersen (2014), building on insights from Black (1972), explore the under-performance of high-beta stocks by "Betting Against Beta," and attribute the success of this investment strategy to investors' borrowing constraints. 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 4 implies the unconditional alpha on BAB is:

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

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. We start our analysis 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.[28] Average market excess return over the sample period is 9.9% (Table 1). Plugging these numbers in (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 7.7% per year). Thus, at least in our sample, almost 90% of the alpha on BAB may result from beta mismeasurement. Alternatively, we claim that alpha on BAB is partly a reward for systematic risk: the true beta on the BAB strategy needs to be as large as BAB's alpha, 7.7%, divided by average market excess return, 9.9%, that is 0.78. With $\delta = 1.1$, we obtain true betas of $\beta_L = 0.79$ and $\beta_H = 1.44$, which implies, in turn, a true beta for the BAB strategy of 0.68, which accounts for 90% of the beta (0.78) necessary to explain alpha on BAB.

To test whether beta mismeasurement can 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 tests. The first test constructs 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 $\beta^{\mathcal{D}}$ may explain this alpha. We thus build these two factors, $\mathcal{C}$ and $\mathcal{D}$, and use their realized returns ($\widetilde{R}^e_{\mathcal{C},t}$ and $\widetilde{R}^e_{\mathcal{D},t}$) as controls.

We construct realized returns $\widetilde{R}^e_{\mathcal{C},t}$ and $\widetilde{R}^e_{\mathcal{D},t}$ 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 $\beta^{\mathcal{C}}$ for the $\mathcal{C}$ factor (or $\beta^{\mathcal{D}}$ for the $\mathcal{D}$ factor) and form two portfolios, one with high-$\beta^{\mathcal{C}}$ ($\beta^{\mathcal{D}}$) stocks and another with low-$\beta^{\mathcal{C}}$ ($\beta^{\mathcal{D}}$) stocks. When forming these portfolios, we follow the ranking and weighting methodology in Frazzini and Pedersen (2014).

[28] 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 $w = 0.6$.

We then obtain returns on each factor by going long the high-$\beta^{\mathcal{C}}$ ($\beta^{\mathcal{D}}$) stocks and short the low-$\beta^{\mathcal{C}}$ ($\beta^{\mathcal{D}}$) stocks. Accordingly, the factor $\mathcal{C}$ provides *long* exposure to consensus beta $\beta^{\mathcal{C}}$, and 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 ($\widetilde{R}^e_{\mathcal{C},t}$ and $\widetilde{R}^e_{\mathcal{D},t}$) on the excess returns of the market. Both factors have negative and statistically significant alphas, in line with 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.

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 time-series regressions of the excess returns of factors $\mathcal{C}$ (fist two lines) and $\mathcal{D}$ (last two lines) on the excess returns of the market:

$$\widetilde{R}^e_{\mathcal{C},t}(\widetilde{R}^e_{\mathcal{D},t}) = \alpha + \beta \widetilde{R}^e_{\mathbf{M},t} + \varepsilon_t.$$

We obtain excess returns $\widetilde{R}^e_{\mathcal{C},t}(\widetilde{R}^e_{\mathcal{D},t})$ on 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-$\beta^{\mathcal{C}}$ ($\beta^{\mathcal{D}}$) stocks financed by a short position in low-$\beta^{\mathcal{C}}$ ($\beta^{\mathcal{D}}$) 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 $^{***}p < .01$.

We report our first BAB test results in panel (a) of Table 7. 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 $\mathcal{D}$ weaken this alpha, with significant loadings on $\widetilde{R}^e_{\mathcal{C},t}$ (-0.3580, t-stat -2.20) and particularly on $\widetilde{R}^e_{\mathcal{D},t}$ (-0.9726, t-stat -4.14), suggesting that dispersion in beliefs plays an important role in explaining abnormal returns on BAB. The statistical significance of the loading on the market and the adjusted R^2 rise once we include returns on factors $\mathcal{C}$ and $\mathcal{D}$. This suggests that $\widetilde{R}^e_{\mathcal{C},t}$ and $\widetilde{R}^e_{\mathcal{D},t}$ act as omitted variables and that BAB may expose the empiricist to market risk due to beta mismeasurement.

(a) Regression using the AQR BAB factor:

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

$\widehat{\alpha}_{\mathrm{BAB}}$	$\widehat{\beta}_{\mathrm{BAB}}$	$\widehat{\beta}^{\mathcal{C}}_{\mathrm{BAB}}$	$\widehat{\beta}^{\mathcal{D}}_{\mathrm{BAB}}$	Adj.R^2	N
0.0377*	0.4744***			0.3241	190
(1.71)	(3.93)				
-0.0028	0.6870***	-0.3581**	-0.9727***	0.5111	190
(-0.21)	(9.31)	(-2.20)	(-4.14)		

(b) Regression using the in-house BAB factor:

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

$\widehat{\alpha}_{\mathrm{BAB}}$	$\widehat{\beta}_{\mathrm{BAB}}$	$\widehat{\beta}^{\mathcal{C}}_{\mathrm{BAB}}$	$\widehat{\beta}^{\mathcal{D}}_{\mathrm{BAB}}$	Adj.R^2	N
0.0521**	0.9557***			0.5231	190
(2.16)	(4.48)				
-0.0185	1.3207***	-1.1976***	-0.9861***	0.7802	190
(-1.59)	(14.03)	(-7.28)	(-2.91)		

(c) Regression using the in-house BAB factor:

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

$\widehat{\alpha}_{\mathrm{BAB}}$	$\widehat{\beta}_{\mathrm{BAB}}$	$a_{\mathcal{C}}$	$a_{\mathcal{D}}$	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: Panels (a) and (b) present results from time-series 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, $\widetilde{R}^e_{\mathcal{C},t}$ and $\widetilde{R}^e_{\mathcal{D},t}$. We obtain the excess returns $\widetilde{R}^e_{\mathcal{C},t}$ ($\widetilde{R}^e_{\mathcal{D},t}$) by ranking stocks according to their $\beta^{\mathcal{C}}$ ($\beta^{\mathcal{D}}$) and taking a long position in high-$\beta^{\mathcal{C}}$ ($\beta^{\mathcal{D}}$) stocks financed by a short position in low-$\beta^{\mathcal{C}}$ ($\beta^{\mathcal{D}}$) stocks. In panel (c), instead of using excess returns on the $\mathcal{C}$ and $\mathcal{D}$ factors, we directly control for $\beta^{\mathcal{C}}_{\mathrm{BAB}}$ and $\beta^{\mathcal{D}}_{\mathrm{BAB}}$ ($\beta^{\mathcal{C}}_{BAB}$ and $\beta^{\mathcal{D}}_{BAB}$ represent the value-weighted $\beta^{\mathcal{C}}$ and $\beta^{\mathcal{D}}$ of our in-house BAB factor). 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 $^{***}p < .01$.

In a second test, we replicate the BAB strategy within our sample ("in-house BAB"), which allows us to control for $\beta^{\mathcal{C}}$ and $\beta^{\mathcal{D}}$ directly. We use our realized betas at the end of each month and follow the guidelines in Frazzini and Pedersen (2014), with the only exception that our realized betas and returns are at the one-year horizon, as opposed to a one-month horizon. We first verify that our strategy captures the BAB factor: comparing its returns with those of the original BAB strategy (annualized), the correlation between the two is 0.74; regressing the returns of the AQR BAB strategy on those of the in-house BAB strategy yields an intercept of 0.016 (t-stat 1.13) and a slope of 0.466 (t-stat 5.47).

In panel (b) of Table 7 we repeat the two regressions of panel (a) but using the in-house BAB returns. Just like in our tests for AQR BAB, the alpha on in-house BAB in the first regression is positive and statistically significant. However, it loses statistical significance once we control for $\widetilde{R}^e_{\mathcal{C},t}$ and $\widetilde{R}^e_{\mathcal{D},t}$. We also 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. Finally, we can now directly control for $\beta^{\mathcal{C}}$ and $\beta^{\mathcal{D}}$, since we have these betas for every stock involved in the strategy. We therefore compute the value-weighted $\beta^{\mathcal{C}}$ and $\beta^{\mathcal{D}}$ of our in-house BAB factor, denoted hereafter $\beta^{\mathcal{C}}_{BAB}$ and $\beta^{\mathcal{D}}_{BAB}$. Panel (c) shows that controlling for $\beta^{\mathcal{C}}_{BAB}$ and $\beta^{\mathcal{D}}_{BAB}$ weakens abnormal returns on the strategy. Overall, Table 7 suggests that without controlling for the $\mathcal{C}$ and $\mathcal{D}$ factors, market alpha (beta) on BAB appears to be too strong (weak). With the data at hand, we cannot reject the theoretical possibility that BAB returns result from beta mismeasurement.

5.4 The crowding-out effect of public information

We conclude our empirical analysis with evidence of beta compression on days when public information "crowds out" private information (Section 4.2, Proposition 6). In testing this implication, the main difficulty is identifying days during which public information likely dominates private information. Our choice is guided by recent empirical work showing that market betas are compressed on FOMC announcement days (Andersen et al., 2021) and on days when the FOMC holds press conferences (Bodilsen et al., 2021).[29] The FOMC holds press conferences in an effort to increase public communication and provide additional transparency. Thus, press conference (PC) days are times when public information likely reduces disagreement and offer a reasonable setting for our test.

There are 38 PC days in our dataset (from Apr. 27, 2011, to Sep. 18, 2019). Following the procedure in Bodilsen et al. (2021), we form ten value-weighted portfolios sorted on betas, then separately estimate the CAPM on all days or PC days. Panel (a) of Figure 5 shows CAPM on each type of days. The estimated market risk premium is much higher. A significant beta compression occurs on PC days, extending evidence in Bodilsen et al. (2021)

[29] Before 2019, not every FOMC announcement was followed by a press conference. See Boguth, Grégoire, and Martineau (2019), who show evidence of increased investor attention on PC days.

to a longer sample period. Panel (b) further confirms the beta compression effect by plotting betas on PC days versus betas on other days.

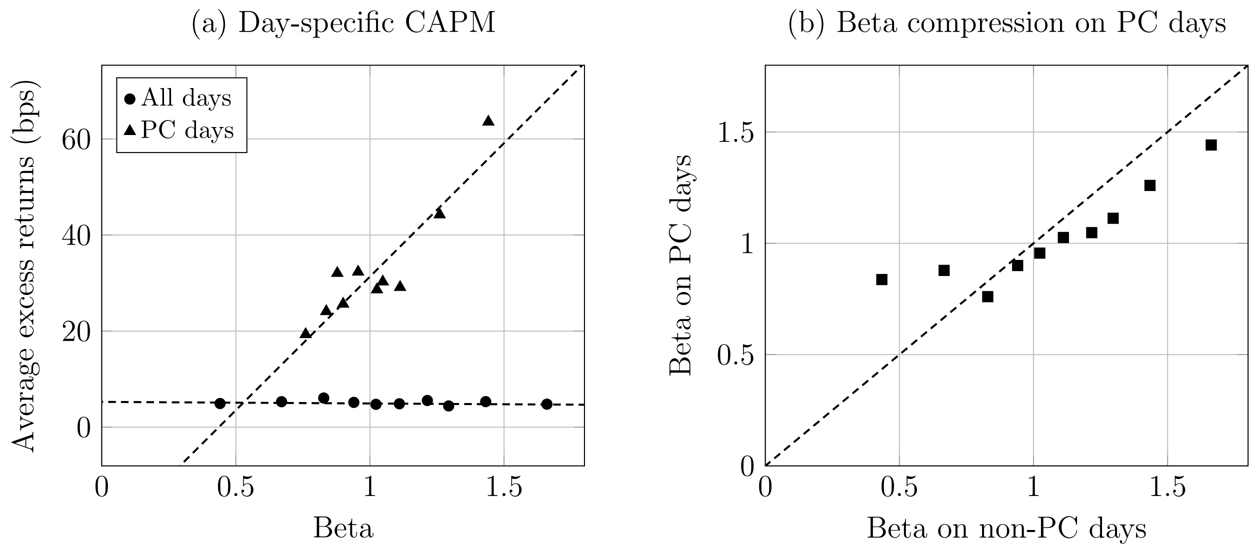

Figure 5: **The crowding-out effect of public information.** This figure provides evidence of beta compression on days when public information dominates private information. Panel (a) plots average daily excess returns in basis points (bps) against full-sample betas for ten value-weighted beta-sorted portfolios separately for all days (dots) and press conference (PC) days (triangles). The lines represent the day-specific CAPM relations. Panel (b) plots betas on PC days against betas on non-PC days, along with a 45-degree line. Beta estimates are provided in Table 8. The sample period is January 2011 to September 2019 (2200 trading days, of which 38 are scheduled PC days).

Table 8 shows results from a regression in which the intercept and portfolio beta are allowed to vary conditional on the type of day:

$$r^e_{j,t} = \alpha_{\text{Other},j} + \alpha_{\text{PC},j}\mathbf{1}_{\text{PC}} + \beta_{\text{Other},j} r^e_{M,t} + \beta_{\Delta\text{PC},j}(\mathbf{1}_{\text{PC}} \times r^e_{M,t}) + \varepsilon_t, \tag{51}$$

where $\mathbf{1}_{\text{PC}}$ is a dummy variable for PC days, $r^e_{M,t}$ is the excess return on the market, $r^e_{j,t}$ is the portfolio excess return, $\beta_{\text{Other},j}$ is the beta on other days, and $\beta_{\Delta\text{PC},j}$ measures the change in the portfolio's beta on PC days. The table demonstrates that portfolio betas compress on PC days, and that most of the differences in beta on PC days versus other days are statistically significant.

Figure 5 also shows that all portfolios (including the market) earn higher returns on PC days. This risk-premium channel is not present in our static model. However, we conjecture that in a dynamic extension the risk premium rises with the uncertainty investors face ahead

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
β_{Other}	0.43***	0.66***	0.83***	0.94***	1.02***	1.11***	1.22***	1.30***	1.43***	1.66***
	(36.06)	(75.88)	(121.52)	(156.48)	(158.59)	(163.82)	(166.89)	(149.53)	(137.16)	(103.85)
$\beta_{\Delta\text{PC}}$	0.40***	0.21***	-0.07	-0.04	-0.07	-0.08	-0.17***	-0.18***	-0.17**	-0.22*
	(4.27)	(3.12)	(-1.25)	(-0.84)	(-1.30)	(-1.58)	(-2.93)	(-2.69)	(-2.12)	(-1.76)

Table 8: This table reports estimates from the regression (51). The returns of ten value-weighted beta-sorted portfolio are computed following the approach in Bodilsen et al. (2021). The sample period is January 2011 to September 2019 (2200 trading days, of which 38 are scheduled PC days). t-statistics are presented in parentheses. Statistical significance is denoted by $^{*}p < .1$, $^{**}p < .05$, or $^{***}p < .01$.

of public announcements.[30] Finally, we view our evidence of the beta-compression channel as suggestive; additional empirical work based on high-frequency returns (e.g., Andersen et al., 2021) is perhaps needed to further confirm the validity of this result.[31]

6 Extensions and robustness of assumptions

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

6.1 A more general proof of SML flattening

The equilibrium relation (13) 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 of information investors observe (public or private), 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, (13) 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

[30] Recent theories propose various mechanisms for this channel (Savor and Wilson, 2013; Ai and Bansal, 2018; Wachter and Zhu, 2018; Hu, Pan, Wang, and Zhu, 2021; Andrei et al., 2021).

[31] In untabulated analysis, we confirm that beta dispersion decreases on PC days using the 25 size- and book-to-market sorted portfolios and using 10 industry portfolios. New studies add to evidence that the CAPM performs better when investors are more attentive to publicly available information (Ben-Rephael et al., 2021; Chan and Marsh, 2021).

asset-pricing model (Campbell, 1993) when hedging demands are absent.[32]

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. 6) and the structure of investors' signals. We now relax our assumptions regarding the structure of payoffs (Eq. 6) 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 slope of empiricist's SML is the slope of a regression of assets' unconditional expected excess returns on empiricist's betas. Proposition 1 then implies:

$$\frac{\mathrm{Cov}[\mathbb{E}[\widetilde{\mathbf{R}}^e], \widehat{\boldsymbol{\beta}}]}{\mathrm{Var}[\widehat{\boldsymbol{\beta}}]} = \frac{\mathrm{Cov}[\boldsymbol{\beta}, \widehat{\boldsymbol{\beta}}]}{\mathrm{Var}[\widehat{\boldsymbol{\beta}}]}\, \mathbb{E}[\widetilde{R}^e_{\mathbf{M}}], \tag{52}$$

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

Proposition 8. *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.*

The proof starts from the eigenvalue decomposition of investors' covariance matrix, $\mathbf{\Sigma} \equiv \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}'$, where $\mathbf{\Lambda}$ is a diagonal matrix with positive eigenvalues on its diagonal, and $\mathbf{Q}$ is an orthogonal matrix of eigenvectors. Because all information is public, the last term in (13) drops out, and the empiricist's covariance matrix, $\widehat{\mathbf{\Sigma}}$, and $\mathbf{\Sigma}$ share the same eigenvectors,

$$\widehat{\mathbf{\Sigma}} = \mathbf{Q}\mathbf{\Lambda}(\mathbf{I} + \gamma^2/\tau_m \mathbf{\Lambda})\mathbf{Q}', \tag{53}$$

but not the same eigenvalues. Each eigenvalue λ of $\widehat{\mathbf{\Sigma}}$ is an inflated version of that of $\mathbf{\Sigma}$ by a factor $1 + \gamma^2/\tau_m \lambda$, where γ^2/τ_m captures aggregate variation in expected returns. This increases the dispersion in empiricist's betas, 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

[32]A more general form of (13), in which γ and $\mathbf{\Sigma}$ are time-varying, has been derived by Jensen (1972), and further studied by Bollerslev, Engle, and Wooldridge (1988).

be tied to our assumption of residual uncertainty in payoffs (i.e., $\tau_\epsilon < \infty$).[33] 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.

6.2 Size effects

We now assume that assets have unequal weights in the market portfolio. 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 9. *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:*

$$\boldsymbol{\mu} - \mu_{\mathbf{M}}\mathbf{1} = \frac{\mu_{\mathbf{M}}}{1+\delta}(\widehat{\boldsymbol{\beta}} - \mathbf{1}) + \frac{\delta\mu_{\mathbf{M}}}{1+\delta}\left(\frac{\mathbf{M}}{\|\mathbf{M}\|^2} - \mathbf{1}\right), \tag{54}$$

where $\delta > 0$ *denotes the distortion coefficient, 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 4; in contrast, when it is not, $\mathbf{M} \neq \mathbf{1}/N$ and (54) incorporates an additional factor, whose sign and magnitude depend on the difference $\mathbf{M}/\|\mathbf{M}\|^2 - \mathbf{1}$ (i.e., the relative size of assets). To see how this second factor affects the measured slope of the SML, consider an asset that has a high measured beta ($\widehat{\beta}_n > 1$) but earns *negative* returns in excess of the market portfolio. To satisfy (54) 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. A necessary condition for a downward-sloping SML is then (see Appendix A.11):

$$\mathrm{Cov}[\mathbf{\Phi}, \mathbf{M}] < 0. \tag{55}$$

Figure 6 depicts the investors' and empiricist's SML in an economy with three assets. The empiricist perceives a negative relation between beta and expected returns. Assets no longer plot on a straight line, while the true SML is *always* upward-sloping.

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

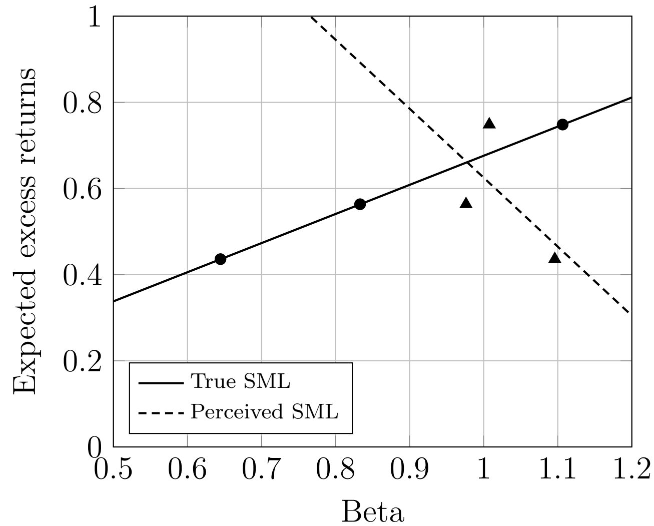

Figure 6: **Size Effects.** This figure illustrates the true SML (solid line) and observed SML (dashed line) when stocks are in heterogeneous supplies. This illustrative economy has three assets with factor loadings, $\phi_1 > \phi_2 > \phi_3 > 0$, and with supplies in the market portfolio, $0 < M_1 < M_2 < M_3$.

6.3 Large economy and multiple factors

A final extension is to allow payoffs to be driven by multiple ($J \geq 1$) common factors; we consider a "large economy" in which the number of stocks, N, and 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). We provide here conditions under which factor multiplicity generates flattening of the SML (for technical details, see Appendix A.12).

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 is normally distributed with mean $\mathbf{0}$ and covariance $(\tau_F J)^{-1}\mathbf{I}_J$. We scale prior precisions on common factors by J so that average prior $1/N \sum_{n=1}^{N} \text{Var}[\mathbf{\Phi}\widetilde{\mathbf{F}}]_{nn} = \tau_F^{-1}$ does not grow with J (see (57) below). 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 that

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

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.

Our main result relies on the following eigenvalue decomposition:

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

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 in (57) implies that their sum equals 1. We follow Martin and Nagel (2020)'s assumption that each eigenvalue satisfies $\lambda > \varepsilon$, for some uniform constant, ε, as $N \to \infty$ (i.e., the columns of $\mathbf{\Phi}$ never become collinear in the limit). Finally, we make the following assumption, allowing us to use random matrix theory results.

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

This means that loadings are on average zero with covariance matrix $\mathbf{T}$, and allows us to write SML distortion in terms of the limiting variance and skewness of eigenvalues in (59), $\sigma_\lambda^2 \equiv \lim_{J\to\infty} \frac{1}{J}\sum_{j=1}^J \lambda_j^2 - \mu_\lambda^2$ and $s_\lambda \equiv \lim_{J\to\infty} \frac{1}{J}\sum_{j=1}^J \lambda_j^3 - 3\mu_\lambda\sigma_\lambda^2 - \mu_\lambda^3$, and where $\mu_\lambda \equiv \lim_{J\to\infty} \frac{1}{J}\sum_{j=1}^J \lambda_j$ denotes the limiting mean of eigenvalues.

Proposition 10. (Flattening of empiricist's SML) *Consider a large economy with a small ratio of factors, $\psi \in (0, 1)$, in which eigenvalues are not too dispersed:*

$$\sigma_\lambda^2 < \frac{1}{2}\left(\sqrt{\Delta + \frac{4s_\lambda(\mu_\lambda\tau_\epsilon + \tau_0)}{\tau_\epsilon}} + \mu_\lambda^2 + \frac{\mu_\lambda\tau_0(3-\psi)}{\tau_\epsilon} + \frac{\tau_0\psi\tau_1}{\gamma^2\tau_\epsilon^2}\right), \tag{60}$$

where τ_0, τ_1, Δ, are strictly positive coefficients defined in Appendix A.12. Then, if eigenvalues are positively skewed (or exhibit little negative skewness) or, on the contrary, if they exhibit strictly negative (but limited) skewness and if, further, eigenvalues are not too concentrated, the SML will look flatter than it actually is. If, further, skewness is strictly negative and if eigenvalues are sufficiently concentrated, the SML will be downward-sloping.

The distribution of eigenvalues in (59) determines whether the empiricist's SML looks steep, flat, or even downward-sloping. Suppose that eigenvalues are not too dispersed in the sense of (60), meaning that factors have comparable predictive power. Low dispersion of eigenvalues induces flattening (whereas high dispersion creates steepening—see Appendix A.12, Corollary 10.1). Sufficiently low dispersion may lead to a downward-sloping SML. The distribution of eigenvalues associated with factors and the many implications it may have for asset-pricing tests opens up fascinating avenues for future research.

7 Conclusion

Why do empiricists keep rejecting the CAPM, which practitioners are unwilling to abandon? We argue that the CAPM may hold from each investor's perspective, but that variation 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. While the literature has studied extensively how time variation distorts CAPM tests (e.g., Jagannathan and Wang, 1996; Lewellen and Nagel, 2006), this paper emphasizes dispersed information as an additional source of distortion. Our empirical analysis shows that the effect of dispersed information is stronger than that of time variation. Together these two sources of variation produce substantial CAPM distortion and imply a zero-beta CAPM. Black (1972) obtains this relation in an economy with restricted borrowing. Notably, in our model it is variation in expected returns over time and across investors, as opposed to restricted borrowing, that causes the measured CAPM to appear flat. It would be interesting to explore how these two mechanisms interact.

Our theory implies that some variables may appear to the empiricist as priced factors simply because betas are mismeasured. However, rather than being priced factors, these variables are instruments for measurement error in betas (Andrei, Cujean, and Fournier, 2019). More generally, the fact that investors' information is dispersed requires careful consideration of how to measure betas correctly.

References

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

Ai, H. and R. Bansal (2018). Risk preferences and the macro announcement premium. *Econometrica*.

Albagli, E., C. Hellwig, and A. Tsyvinski (June 2022). Information aggregation and asymmetric returns. *TSE Working Paper, n. 21-1172*.

Andersen, T. G., M. Thyrsgaard, and V. Todorov (2021). Recalcitrant betas: Intraday variation in the cross-sectional dispersion of systematic risk. *Quantitative Economics 12*(2), 647–682.

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

Andrei, D., H. L. Friedman, and N. B. Ozel (2021). Economic uncertainty and investor attention. *Available at SSRN 3128673*.

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 (2021). Information consumption and asset pricing. *The Journal of Finance 76*(1), 357–394.

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., J. N. Eriksen, and N. S. Grønborg (2021). Asset pricing and fomc press conferences. *Journal of Banking & Finance 128*, 106163.

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.

Boguth, O., V. Grégoire, and C. Martineau (2019). Shaping expectations and coordinating attention: The unintended consequences of fomc press conferences. *Journal of Financial and Quantitative Analysis 54*(6), 2327–2353.

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.

Breon-Drish, B. (2015, 03). On Existence and Uniqueness of Equilibrium in a Class of Noisy Rational Expectations Models. *The Review of Economic Studies 82*(3), 868–921.

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

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.

Chan, K. F. and T. Marsh (2021). Asset pricing on earnings announcement days. *Journal of Financial Economics*.

Cho, J.-W. and M. Krishnan (2000, March). Prices as aggregators of private information: Evidence from s&p 500 Futures data. *Journal of Financial and Quantitative Analysis 35*(01), 111–126.

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.

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.

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.

Hellwig, M. F. (1980). On the aggregation of information in competitive markets. *Journal of Economic Theory 22*(3), 477–498.

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*.

Hu, G. X., J. Pan, J. Wang, and H. Zhu (2021). Premium for heightened uncertainty: Explaining pre-announcement market returns. *Journal of Financial Economics*.

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.

Kaas, R., D. Dannenburg, and M. Goovaerts (1997). Exact credibility for weighted observations. *ASTIN Bulletin 27*(2), 287?295.

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.

Rencher, A. C. and B. G. Schaalje (2008). *Linear models in statistics*. Wiley-Interscience.

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 (2013). How much do investors care about macroeconomic risk? evidence from scheduled economic announcements. *Journal of Financial and Quantitative Analysis 48*(2), 343–375.

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.

Sun, Y. (2006). The exact law of large numbers via fubini extension and characterization of insurable risks. *Journal of Economic Theory 126*(1), 31–69.

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

Vanden, J. M. (2008). Information quality and options. *The Review of Financial Studies 21*(6), 2635–2676.

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.

Wachter, J. A. and Y. Zhu (2018). The macroeconomic announcement premium. Technical report, National Bureau of Economic Research.

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 Appendix (Proofs)

A.1 Discussion of Assumption B in Section 2

In this appendix we discuss specific conditions under which (1) holds. Consider for analytical convenience excess return on a single asset, $\widetilde{R}^e_n$, and let each investor i observe K signals, $\mathbf{x}^i \equiv (x^i_1, ..., x^i_K)$, about $\widetilde{R}^e_n$. Assumption C requires that each investor has an identical number of signals and that signals have identical distribution, $p(\mathbf{x}^i|\widetilde{R}^e)$, so that investors' information is identically precise. Each investor i forms a posterior distribution $p(\widetilde{R}^e_n|\mathbf{x}^i)$ via Bayes' rule (Assumption A):

$$p(\widetilde{R}^e_n|\mathbf{x}^i) = \frac{p(\mathbf{x}^i|\widetilde{R}^e_n)p(\widetilde{R}^e_n)}{p(\mathbf{x}^i)}. \tag{A.1}$$

Assuming that distributions belong to the exponential family, Kaas, Dannenburg, and Goovaerts (1997) show that investor i's posterior expectation, $\mathbb{E}^i[\widetilde{R}^e_n] \equiv \mathbb{E}[\widetilde{R}^e_n|\mathbf{x}^i]$, is a linear, convex combination of the prior expectation and a sufficient statistic for her signals, $\mathbf{x}^i$, if the prior distribution, $p(\widetilde{R}^e_n)$, is a conjugate prior of the likelihood $p(\mathbf{x}^i|\widetilde{R}^e_n)$. A prior distribution, $p(\widetilde{R}^e_n)$, is a conjugate prior for $p(\mathbf{x}^i|\widetilde{R}^e_n)$ if it belongs to the same probability distribution family as the posterior distribution, $p(\widetilde{R}^e_n|\mathbf{x}^i)$.

The exponential family is a broad set of distributions, which includes, for instance, the Bernoulli, Gaussian, Multinomial, Dirichlet, Gamma, Poisson, Beta distributions, among others (see, e.g., Breon-Drish, 2015, and two specific examples below). The likelihood takes the following form:

$$p(x^i_k|\widetilde{R}^e_n) = \exp\left[\frac{x^i_k\widetilde{R}^e_n - b(\widetilde{R}^e_n)}{a/w_k} + c(x^i_k, a/w_k)\right], \tag{A.2}$$

where a/w_k are known parameters and $b(\cdot)$ and $c(\cdot)$ are known functions. Then, $p(\widetilde{R}^e_n)$ is a conjugate prior if it has the same functional form and the $\widetilde{R}^e_n$-dependent part is the same as in (A.2):

$$p(\widetilde{R}^e_n) = \exp\left[\frac{\widetilde{R}^e_n x_0 - b(\widetilde{R}^e_n)}{a/w_0} + d(x_0, a/w_0)\right], \tag{A.3}$$

where x_0, a/w_0 are parameters and the function $d(x_0, a/w_0)$ is chosen in such a way that the density, which ranges over some $\widetilde{R}^e_n$-interval, integrates to 1. Assume further that, conditional on $\widetilde{R}^e_n$, the random signals $x^i_1, ..., x^i_K$ are independent drawings from (A.2). Then the posterior distribution $p(\widetilde{R}^e_n|\mathbf{x}^i)$ follows from (A.1):

$$p(\widetilde{R}^e_n|\mathbf{x}^i) = C\exp\left[\frac{\widetilde{R}^e_n x_0 - b(\widetilde{R}^e_n)}{a/w_0}\right]\prod_{k=1}^{K}\exp\left[\frac{x^i_k\widetilde{R}^e_n - b(\widetilde{R}^e_n)}{a/w_k}\right] = C\exp\left[\frac{\widetilde{R}^e_n x^i_\bullet - b(\widetilde{R}^e_n)}{a/w_\bullet}\right], \tag{A.4}$$

where C is a normalizing constant, and

$$w_\bullet \equiv \sum_{k=0}^{K} w_k \quad \text{and} \quad x_\bullet \equiv \sum_{k=0}^{K}\frac{w_k}{w_\bullet}x^i_k. \tag{A.5}$$

Thus, the posterior and prior distributions, $p(\widetilde{R}^e_n|\mathbf{x}^i)$ and $p(\widetilde{R}^e_n)$, are of the same type, but with x_0 and w_0 replaced respectively by $x^i_\bullet$ and $w_\bullet$.

The consequence of the above result is that investor i's posterior expectation, $\mathbb{E}^i[\widetilde{R}_n^e] \equiv \mathbb{E}[\widetilde{R}_n^e|\mathbf{x}^i]$, is a linear, convex combination of the prior expectation and a sufficient statistic for her signals, $\mathbf{x}^i$. This is Theorem 2.1 in Kaas et al. (1997), whose proof we reproduce here using our notation. We must prove that this expression is linear in $x_1^i, ..., x_K^i$:

$$\mathbb{E}^i[\widetilde{R}_n^e] = \int \mathbb{E}[\widetilde{R}_n^e] C \exp\left[\frac{\widetilde{R}_n^e x_\bullet^i - b(\widetilde{R}_n^e)}{a/w_\bullet}\right] d\widetilde{R}_n^e, \tag{A.6}$$

where $\mathbb{E}[\widetilde{R}_n^e] = b'(\widetilde{R}_n^e)$ is the mean under the likelihood (A.2) (Eq. (4) in Kaas et al. (1997)). Replacing in the above yields:

$$\mathbb{E}^i[\widetilde{R}_n^e] = x_\bullet \int C \exp\left[\frac{\widetilde{R}_n^e x_\bullet^i - b(\widetilde{R}_n^e)}{a/w_\bullet}\right] d\widetilde{R}_n^e + \frac{Ca}{w_\bullet} \int d \exp\left[\frac{\widetilde{R}_n^e x_\bullet^i - b(\widetilde{R}_n^e)}{a/w_\bullet}\right]. \tag{A.7}$$

The first term integrates a probability density function and thus equals $x_\bullet$, whereas the second term is 0 (in the exponential family, the probability density function equals 0 at endpoints). Thus:

$$\mathbb{E}^i[\widetilde{R}_n^e] = x_\bullet = \frac{w_0}{w_\bullet} x_0 + \left(1 - \frac{w_0}{w_\bullet}\right) \frac{\sum_{k=1}^K w_k x_k^i}{\sum_{k=1}^K w_k}, \tag{A.8}$$

which is a linear, convex combination between x_0 and a sufficient statistic for investor's signals (which in this case is a weighted average of all the signals). This, together with the common prior assumption (i.e., all investors observe the same x_0), justifies the linear form in (1).

The normal-normal specification of the main model in Section 3 is a particular case of the exponential family in which (1) holds. However, we provide here two examples that go beyond the normal distribution assumption. The first example assumes that the gross return $\widetilde{R}_n^e$ is Gamma distributed with parameters α_1 and α_2 (an application in finance can be found in Vanden, 2008). Suppose further that signals are independent, Poisson with intensity $\widetilde{R}_n^e$ (for instance, these signals are the number of analysts' *buy* recommendations). Eq. (A.1) implies that the prior-to-posterior update involves adding the sufficient statistic $\sum_{k=1}^K x_k^i$ to the parameter α_1 and adding the number of signals K to the parameter α_2. As in (A.8), the posterior mean is linear in signals:

$$\mathbb{E}[\widetilde{R}_n^e|x_1^i, ..., x_K^i] = \frac{\alpha_2}{K + \alpha_2} \frac{\alpha_1}{\alpha_2} + \left(1 - \frac{\alpha_2}{K + \alpha_2}\right) \frac{1}{K} \sum_{k=1}^K x_k^i. \tag{A.9}$$

A second example that suggests we can go in fact quite far from the normal case is the beta distribution as a prior for $\widetilde{R}_n^e/z$, Be(α_1, α_2), $\alpha_1, \alpha_2 > 0$ ($z > 1$ is a scaling constant such that $\widetilde{R}_n^e/z$ belongs to the unit interval). This distribution may have multiple modes and may be skewed, among other properties that indeed make it distinct from the normal distribution. If signals are independent and Bernoulli distributed with probability $\widetilde{R}_n^e/z$, the posterior mean is linear in signals:

$$\mathbb{E}[\widetilde{R}_n^e/z|x_1^i, ..., x_K^i] = \frac{\alpha_1 + \alpha_2}{K + \alpha_1 + \alpha_2} \frac{\alpha_1}{\alpha_1 + \alpha_2} + \left(1 - \frac{\alpha_1 + \alpha_2}{K + \alpha_1 + \alpha_2}\right) \frac{1}{K} \sum_{k=1}^K x_k^i. \tag{A.10}$$

These are two examples in which investor i's posterior beliefs are linear in a sufficient statistic for her information set, as in (1).

A.2 Proof of Proposition 2

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

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

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:

$$\widetilde{\mathbf{P}}^a \equiv \widetilde{\mathbf{P}} - \mathbf{1}D - \boldsymbol{\xi}_0\mathbf{M} = \boldsymbol{\lambda}\widetilde{F} + \boldsymbol{\xi}\widetilde{\mathbf{m}}, \tag{A.12}$$

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

$$\begin{bmatrix} \widetilde{\mathbf{P}}^a \\ V^i \end{bmatrix} = \begin{bmatrix} \boldsymbol{\lambda} \\ 1 \end{bmatrix} \widetilde{F} + \begin{bmatrix} \boldsymbol{\xi} & \mathbf{0}_{N\times 1} \\ \mathbf{0}_{1\times N} & 1 \end{bmatrix} \begin{bmatrix} \widetilde{\mathbf{m}} \\ \tilde{v}^i \end{bmatrix} \equiv \mathbf{H}\widetilde{F} + \boldsymbol{\Theta} \begin{bmatrix} \widetilde{\mathbf{m}} \\ \tilde{v}^i \end{bmatrix}, \tag{A.13}$$

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

$$\mathbf{C} = \begin{bmatrix} \tau_m^{-1}\mathbf{I} & \mathbf{0}_{N\times 1} \\ \mathbf{0}_{1\times N} & \tau_v^{-1} \end{bmatrix}. \tag{A.14}$$

Applying standard projection techniques we define the precision of the last term in (A.13):

$$\mathbf{r} \equiv (\boldsymbol{\Theta}\mathbf{C}\boldsymbol{\Theta}')^{-1} = \begin{bmatrix} \tau_m(\boldsymbol{\xi}\boldsymbol{\xi}')^{-1} & \mathbf{0}_{N\times 1} \\ \mathbf{0}_{1\times N} & \tau_v \end{bmatrix}, \tag{A.15}$$

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

$$\tau \equiv \mathrm{Var}[\widetilde{F}|\mathscr{F}^i]^{-1} = \tau_F + \mathbf{H}'\mathbf{r}\mathbf{H} = \tau_F + \tau_v + \tau_m\boldsymbol{\lambda}'(\boldsymbol{\xi}\boldsymbol{\xi}')^{-1}\boldsymbol{\lambda}. \tag{A.16}$$

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):

$$\mathbb{E}[\widetilde{F}|\mathscr{F}^i] = \frac{1}{\tau}\mathbf{H}'\mathbf{r} \begin{bmatrix} \widetilde{\mathbf{P}}^a \\ \widetilde{V}^i \end{bmatrix} = \frac{1}{\tau} \left[\boldsymbol{\lambda}'(\boldsymbol{\xi}\boldsymbol{\xi}')^{-1}\tau_m \quad \tau_v \right] \begin{bmatrix} \widetilde{\mathbf{P}}^a \\ \widetilde{V}^i \end{bmatrix} \tag{A.17}$$

$$= \frac{1}{\tau}\left(\tau_m\boldsymbol{\lambda}'(\boldsymbol{\xi}\boldsymbol{\xi}')^{-1}\boldsymbol{\lambda}\widetilde{F} + \tau_m\boldsymbol{\lambda}'(\boldsymbol{\xi}\boldsymbol{\xi}')^{-1}\boldsymbol{\xi}\widetilde{\mathbf{m}} + \tau_v\widetilde{F} + \tau_v\tilde{v}^i \right). \tag{A.18}$$

Using the definition of the total precision (A.16), it follows that average market expectation of future dividends is

$$\overline{\mathbb{E}}[\widetilde{\mathbf{D}}] \equiv \int_i \mathbb{E}[\widetilde{\mathbf{D}}|\mathscr{F}^i]di = \mathbf{1}D + \boldsymbol{\Phi}\frac{1}{\tau}\left[(\tau - \tau_F)\widetilde{F} + \tau_m\boldsymbol{\lambda}'(\boldsymbol{\xi}\boldsymbol{\xi}')^{-1}\boldsymbol{\xi}\widetilde{\mathbf{m}} \right], \tag{A.19}$$

and individual expectations are:

$$\mathbb{E}[\widetilde{\mathbf{D}}|\mathscr{F}^i] = \overline{\mathbb{E}}[\widetilde{\mathbf{D}}] + \mathbf{\Phi}\frac{\tau_v}{\tau}\tilde{v}^i. \tag{A.20}$$

For each agent i, the uncertainty about future dividends is

$$\mathbf{\Sigma} \equiv \mathrm{Var}[\widetilde{\mathbf{D}}|\mathscr{F}^i] = \frac{1}{\tau}\mathbf{\Phi}\mathbf{\Phi'} + \frac{1}{\tau_\epsilon}\mathbf{I}. \tag{A.21}$$

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

$$\widetilde{\mathbf{P}} = \overline{\mathbb{E}}[\widetilde{\mathbf{D}}] - \gamma\mathbf{\Sigma}(\mathbf{M} - \widetilde{\mathbf{m}}) \tag{A.22}$$

$$= \mathbf{1}D - \gamma\left(\frac{1}{\tau}\mathbf{\Phi}\mathbf{\Phi'} + \frac{1}{\tau_\epsilon}\mathbf{I}\right)\mathbf{M} + \mathbf{\Phi}\frac{\tau - \tau_F}{\tau}\widetilde{F} + \left[\mathbf{\Phi}\frac{\tau_m}{\tau}(\boldsymbol{\xi}^{-1}\boldsymbol{\lambda})' + \gamma\left(\frac{1}{\tau}\mathbf{\Phi}\mathbf{\Phi'} + \frac{1}{\tau_\epsilon}\mathbf{I}\right)\right]\widetilde{\mathbf{m}}, \tag{A.23}$$

where we have used the simplification

$$\boldsymbol{\lambda}'(\boldsymbol{\xi}\boldsymbol{\xi}')^{-1}\boldsymbol{\xi} = (\boldsymbol{\xi}^{-1}\boldsymbol{\lambda})'. \tag{A.24}$$

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

$$\boldsymbol{\xi}_0 = -\gamma\left(\frac{1}{\tau}\mathbf{\Phi}\mathbf{\Phi'} + \frac{1}{\tau_\epsilon}\mathbf{I}\right) \tag{A.25}$$

$$\boldsymbol{\lambda} = \mathbf{\Phi}\frac{\tau - \tau_F}{\tau} \tag{A.26}$$

$$\boldsymbol{\xi} = \mathbf{\Phi}\frac{\tau_m}{\tau}(\boldsymbol{\xi}^{-1}\boldsymbol{\lambda})' + \gamma\left(\frac{1}{\tau}\mathbf{\Phi}\mathbf{\Phi'} + \frac{1}{\tau_\epsilon}\mathbf{I}\right). \tag{A.27}$$

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

$$\boldsymbol{\lambda} = \mathbf{\Phi}\frac{\tau_m}{\tau}(\boldsymbol{\xi}^{-1}\boldsymbol{\lambda})'\boldsymbol{\xi}^{-1}\boldsymbol{\lambda} + \gamma\left(\frac{1}{\tau}\mathbf{\Phi}\mathbf{\Phi'} + \frac{1}{\tau_\epsilon}\mathbf{I}\right)\boldsymbol{\xi}^{-1}\boldsymbol{\lambda}, \tag{A.28}$$

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. A.16), which can be replaced above, together with the solution (A.26) for $\boldsymbol{\lambda}$ to obtain:

$$\mathbf{\Phi}\frac{\tau_v}{\tau} = \gamma\left(\frac{1}{\tau}\mathbf{\Phi}\mathbf{\Phi'} + \frac{1}{\tau_\epsilon}\mathbf{I}\right)\boldsymbol{\xi}^{-1}\boldsymbol{\lambda}, \tag{A.29}$$

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

$$\boldsymbol{\xi}^{-1}\boldsymbol{\lambda} = \frac{\tau_v}{\gamma\tau}\left(\frac{1}{\tau}\mathbf{\Phi}\mathbf{\Phi'} + \frac{1}{\tau_\epsilon}\mathbf{I}\right)^{-1}\mathbf{\Phi} = \frac{\tau_v\tau_\epsilon}{\gamma(\tau + \tau_\epsilon)}\mathbf{\Phi}. \tag{A.30}$$

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 $\frac{1}{\tau}\mathbf{\Phi}\mathbf{\Phi'} + \frac{\tau}{\tau_\epsilon}\mathbf{I}$ 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:

$$\boldsymbol{\xi}^{-1}\boldsymbol{\lambda} = \frac{\sqrt{\tau_P}}{\sqrt{\tau_m}}\mathbf{\Phi}, \tag{A.31}$$

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 (A.31) in (A.16) yields the total precision τ as a function of this scalar:

$$\tau = \tau_F + \tau_v + \tau_P, \tag{A.32}$$

which, together with $\frac{\sqrt{\tau_P}}{\sqrt{\tau_m}} = \frac{\tau_v \tau_\epsilon}{\gamma(\tau+\tau_\epsilon)}$, leads to a cubic equation in τ_P:

$$\tau_P(\tau_F + \tau_v + \tau_P + \tau_\epsilon)^2 = \frac{\tau_m \tau_\epsilon^2 \tau_v^2}{\gamma^2}. \tag{A.33}$$

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. (A.31) can now be replaced in the fixed point solution (A.27) to obtain the undetermined coefficients $\boldsymbol{\xi}$:

$$\boldsymbol{\xi} = \frac{\gamma + \sqrt{\tau_m \tau_P}}{\tau} \boldsymbol{\Phi}\boldsymbol{\Phi}' + \frac{\gamma}{\tau_\epsilon}\mathbf{I}, \tag{A.34}$$

which completes the proof of Proposition 2. □

A.2.1 Proof of Corollary 2.1

From (A.21), we know that $\boldsymbol{\Sigma} = \frac{1}{\tau}\boldsymbol{\Phi}\boldsymbol{\Phi}' + \frac{1}{\tau_\epsilon}\mathbf{I}$. Thus,

$$\boldsymbol{\Sigma}\mathbf{M} = \frac{\bar{\Phi}}{\tau}\boldsymbol{\Phi} + \frac{1}{N\tau_\epsilon}\mathbf{1} \quad \text{and} \quad \sigma_{\mathbf{M}}^2 = \mathbf{M}'\boldsymbol{\Sigma}\mathbf{M} = \frac{\bar{\Phi}^2}{\tau} + \frac{1}{N\tau_\epsilon}, \tag{A.35}$$

which yields

$$\boldsymbol{\beta} = \frac{\boldsymbol{\Sigma}\mathbf{M}}{\mathbf{M}'\boldsymbol{\Sigma}\mathbf{M}} = \frac{\frac{\bar{\Phi}^2}{\tau}\frac{\boldsymbol{\Phi}}{\bar{\Phi}} + \frac{1}{N\tau_\epsilon}\mathbf{1}}{\sigma_{\mathbf{M}}^2}. \tag{A.36}$$

This is a weighted average between $\mathbf{1}$ and $\boldsymbol{\Phi}/\bar{\Phi}$. Subtracting $\mathbf{1}$ on both sides yields (21). □

A.3 Proof of Proposition 3

For this proof we will make the following assumptions:

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

Assumption A.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 $\boldsymbol{\Phi}$ are equal, and thus all assets are identical. Assumption A.2 eliminates the uninteresting case $\mathbf{M}'\boldsymbol{\Phi} = 0$ (zero market exposure to the common factor), and is without loss of generality (if $\mathbf{M}'\boldsymbol{\Phi} < 0$, one can simply switch the sign of the common factor). In our case, since $\mathbf{M} = \mathbf{1}/N$, $\mathbf{M}'\boldsymbol{\Phi}$ represents the mean of the vector $\boldsymbol{\Phi}$ 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

$$\|\mathbf{x}\|\|\mathbf{y}\| \geq \mathbf{x}'\mathbf{y} = \mathbf{M}'\widehat{\mathbf{\Sigma}}^{1/2}\widehat{\mathbf{\Sigma}}^{-1/2}\boldsymbol{\mu} = \mu_{\mathbf{M}}, \tag{A.37}$$

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

$$\frac{\mu_{\mathbf{M}}}{\sigma_{\mathbf{M}}} \leq \sqrt{\boldsymbol{\mu}'\widehat{\mathbf{\Sigma}}^{-1}\boldsymbol{\mu}}. \tag{A.38}$$

The relation (A.38) holds with equality *if and only if* $\mathbf{x}$ is proportional to $\mathbf{y}$, or

$$\boldsymbol{\mu} \propto \widehat{\mathbf{\Sigma}}\mathbf{M}. \tag{A.39}$$

Starting from the law of total variance (2) and replacing individual expectations from (A.20), we compute $\widehat{\mathbf{\Sigma}}$, which is Lemma 1 in the text (see Appendix A.5 for a proof):

$$\widehat{\mathbf{\Sigma}} = \mathbf{\Sigma} + \frac{\gamma^2}{\tau_m}\left(\frac{1}{\tau_\epsilon}\mathbf{\Sigma} + \frac{e_1}{\tau}\mathbf{\Phi}\mathbf{\Phi}'\right) + \frac{\tau_v}{\tau^2}\mathbf{\Phi}\mathbf{\Phi}', \tag{A.40}$$

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

$$e_1 = \frac{1}{\tau} + \frac{1}{\tau_\epsilon}. \tag{A.41}$$

By making use of (A.21), one can write $\widehat{\mathbf{\Sigma}}$ in two equivalent forms:

$$\widehat{\mathbf{\Sigma}} = c_1\mathbf{\Sigma} + c_2\mathbf{\Phi}\mathbf{\Phi}' \tag{A.42}$$

$$\widehat{\mathbf{\Sigma}} = c_3\mathbf{\Sigma} - c_4\mathbf{I} \tag{A.43}$$

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

$$c_1 = 1 + \frac{\gamma^2}{\tau_m\tau_\epsilon} > 0, \quad c_2 = \frac{\gamma^2 e_1}{\tau_m\tau} + \frac{\tau_v}{\tau^2} > 0, \tag{A.44}$$

and

$$c_3 = 1 + \frac{\gamma^2}{\tau_m\tau_\epsilon} + \frac{\gamma^2 e_1}{\tau_m} + \frac{\tau_v}{\tau} > 0, \quad c_4 = \frac{\gamma^2 e_1}{\tau_m\tau_\epsilon} + \frac{\tau_v}{\tau_\epsilon\tau} > 0. \tag{A.45}$$

Multiply Equations (A.42)-(A.43) with $\mathbf{M}$:

$$\widehat{\mathbf{\Sigma}}\mathbf{M} = c_1\mathbf{\Sigma}\mathbf{M} + c_2\bar{\Phi}\mathbf{\Phi} \tag{A.46}$$

$$\widehat{\mathbf{\Sigma}}\mathbf{M} = c_3\mathbf{\Sigma}\mathbf{M} - c_4\mathbf{M}. \tag{A.47}$$

Since $\boldsymbol{\mu} \propto \mathbf{\Sigma}\mathbf{M}$ (Proposition 1), (A.39) and (A.46)-(A.47) imply that $\boldsymbol{\mu} \propto \mathbf{\Phi}$ and $\boldsymbol{\mu} \propto \mathbf{M}$. This implies $\mathbf{M} \propto \mathbf{\Phi}$, contradicting Assumption A.1. Thus, $\boldsymbol{\mu} \not\propto \widehat{\mathbf{\Sigma}}\mathbf{M}$ and empiricist's CAPM fails. $\square$

A.4 Proof of Proposition 4

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

$$\widehat{\mathbf{\Sigma}}\mathbf{M} = \frac{1}{\gamma}\widehat{\mathbf{\Sigma}}\mathbf{\Sigma}^{-1}\boldsymbol{\mu} = \frac{1}{\gamma}(c_3\mathbf{\Sigma} - c_4\mathbf{I})\mathbf{\Sigma}^{-1}\boldsymbol{\mu} = \frac{c_3}{\gamma}\boldsymbol{\mu} - \frac{c_4}{N}\mathbf{1}, \tag{A.48}$$

where we have used the equilibrium relation $\boldsymbol{\mu} = \gamma\mathbf{\Sigma}\mathbf{M}$ for the first equality, (A.43) for the second equality, and $\gamma^{-1}\mathbf{\Sigma}^{-1}\boldsymbol{\mu} = \mathbf{M} = \mathbf{1}/N$ for the third equality.

Using the definition of empiricist's betas in (17), or $\widehat{\boldsymbol{\beta}} = \frac{\widehat{\mathbf{\Sigma}}\mathbf{M}}{\mathbf{M}'\widehat{\mathbf{\Sigma}}\mathbf{M}}$, it follows that empiricist's betas are an exact linear function of expected excess returns:

$$\widehat{\sigma}^2_{\mathbf{M}}\widehat{\boldsymbol{\beta}} = -\frac{c_4}{N}\mathbf{1} + \frac{c_3}{\gamma}\boldsymbol{\mu}, \tag{A.49}$$

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

$$\boldsymbol{\mu} = \frac{c_4\gamma}{Nc_3}\mathbf{1} + \frac{\gamma\widehat{\sigma}^2_{\mathbf{M}}}{c_3}\widehat{\boldsymbol{\beta}} = \mathbf{1}\mu_{\widehat{\mathbf{Z}}} + \widehat{\boldsymbol{\beta}}(\mu_{\mathbf{M}} - \mu_{\widehat{\mathbf{Z}}}), \tag{A.50}$$

which is (24) 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. Furthermore, since we know that true betas satisfy $\boldsymbol{\mu} = \boldsymbol{\beta}\mu_{\mathbf{M}}$, replacing in (A.50) and rearranging leads to (25) in the text, where δ is defined in (26). □

A.5 Proof of Lemma 1

The proof starts form the law of total variance (2), in which we replace individual expectations from (A.20) and consensus beliefs from (13):

$$\widehat{\mathbf{\Sigma}} = \mathbf{\Sigma} + \frac{\gamma^2}{\tau_m}\mathbf{\Sigma}\mathbf{\Sigma} + \frac{\tau_v}{\tau^2}\mathbf{\Phi}\mathbf{\Phi}' \tag{A.51}$$

$$= \mathbf{\Sigma} + \frac{\gamma^2}{\tau_m}\mathbf{\Sigma}\left(\frac{1}{\tau}\mathbf{\Phi}\mathbf{\Phi}' + \frac{1}{\tau_\epsilon}\mathbf{I}\right) + \frac{\tau_v}{\tau^2}\mathbf{\Phi}\mathbf{\Phi}' \tag{A.52}$$

$$= \mathbf{\Sigma} + \frac{\gamma^2}{\tau_m}\left(\frac{1}{\tau_\epsilon}\mathbf{\Sigma} + \frac{e_1}{\tau}\mathbf{\Phi}\mathbf{\Phi}'\right) + \frac{\tau_v}{\tau^2}\mathbf{\Phi}\mathbf{\Phi}', \tag{A.53}$$

where we have replaced $\mathbf{\Sigma}$ from (11) in the second equality, used $\mathbf{\Sigma}\mathbf{\Phi} = e_1\mathbf{\Phi}$ for the third equality, and e_1 is the unique largest eigenvalue of $\mathbf{\Sigma}$, defined in (A.41). □

A.6 Proof of Proposition 5

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

$$\widehat{\mathbf{\Sigma}} = \mathbf{\Sigma} + \underbrace{\frac{\gamma^2}{\tau_m}\left(\frac{1}{\tau_\epsilon}\mathbf{\Sigma} + \frac{e_1}{\tau}\mathbf{\Phi}\mathbf{\Phi}'\right)}_{\equiv \mathrm{Var}[\overline{\mathbb{E}}[\widetilde{\mathbf{R}}^e]]} + \underbrace{\frac{\tau_v}{\tau^2}\mathbf{\Phi}\mathbf{\Phi}'}_{\equiv \mathrm{Var}[\mathbb{E}^i[\widetilde{\mathbf{R}}^e] - \overline{\mathbb{E}}[\widetilde{\mathbf{R}}^e]]}, \tag{A.54}$$

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

$$\widehat{\sigma}^2_{\mathbf{M}} = \sigma^2_{\mathbf{M}} + \frac{\gamma^2}{\tau_m \tau_\epsilon}\sigma^2_{\mathbf{M}} + \frac{\gamma^2 e_1 \bar{\Phi}^2}{\tau_m \tau} + \frac{\tau_v \bar{\Phi}^2}{\tau^2}, \tag{A.55}$$

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

$$\mathcal{C}^2 \equiv \frac{\text{Var}[\bar{\mathbb{E}}[\widetilde{R}^e_{\mathbf{M}}]]}{\widehat{\sigma}^2_{\mathbf{M}}} = \frac{\gamma^2}{\tau_m \tau_\epsilon \tau \widehat{\sigma}^2_{\mathbf{M}}}(\tau \sigma^2_{\mathbf{M}} + \tau_\epsilon e_1 \bar{\Phi}^2) \tag{A.56}$$

$$\mathcal{D}^2 \equiv \frac{\text{Var}[\mathbb{E}^i[\widetilde{R}^e_{\mathbf{M}}] - \bar{\mathbb{E}}[\widetilde{R}^e_{\mathbf{M}}]]}{\widehat{\sigma}^2_{\mathbf{M}}} = \frac{\tau_v \bar{\Phi}^2}{\tau^2 \widehat{\sigma}^2_{\mathbf{M}}}. \tag{A.57}$$

Post-multiply (A.54) with $\mathbf{M}$ and divide by $\widehat{\sigma}^2_{\mathbf{M}}$ to obtain $\widehat{\boldsymbol{\beta}}$:

$$\widehat{\boldsymbol{\beta}} = \frac{\left(1 + \frac{\gamma^2}{\tau_m \tau_\epsilon}\right)\sigma^2_{\mathbf{M}}\boldsymbol{\beta} + \left(\frac{\gamma^2 e_1 \bar{\Phi}^2}{\tau_m \tau} + \frac{\tau_v \bar{\Phi}^2}{\tau^2}\right)\frac{\boldsymbol{\Phi}}{\bar{\Phi}}}{\left(1 + \frac{\gamma^2}{\tau_m \tau_\epsilon}\right)\sigma^2_{\mathbf{M}} + \frac{\gamma^2 e_1 \bar{\Phi}^2}{\tau_m \tau} + \frac{\tau_v \bar{\Phi}^2}{\tau^2}}, \tag{A.58}$$

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

$$\widehat{\boldsymbol{\beta}} - \mathbf{1} = \frac{\left(1 + \frac{\gamma^2}{\tau_m \tau_\epsilon}\right)\sigma^2_{\mathbf{M}}(\boldsymbol{\beta} - \mathbf{1}) + \left(\frac{\gamma^2 e_1 \bar{\Phi}^2}{\tau_m \tau} + \frac{\tau_v \bar{\Phi}^2}{\tau^2}\right)\left(\frac{\boldsymbol{\Phi}}{\bar{\Phi}} - \mathbf{1}\right)}{\left(1 + \frac{\gamma^2}{\tau_m \tau_\epsilon}\right)\sigma^2_{\mathbf{M}} + \frac{\gamma^2 e_1 \bar{\Phi}^2}{\tau_m \tau} + \frac{\tau_v \bar{\Phi}^2}{\tau^2}}, \tag{A.59}$$

and use the definition of true betas from Corollary 2.1:

$$\widehat{\boldsymbol{\beta}} - \mathbf{1} = \frac{\left(1 + \frac{\gamma^2}{\tau_m \tau_\epsilon}\right)\sigma^2_{\mathbf{M}}(\boldsymbol{\beta} - \mathbf{1}) + \left(\frac{\gamma^2 e_1 \bar{\Phi}^2}{\tau_m \tau} + \frac{\tau_v \bar{\Phi}^2}{\tau^2}\right)\frac{\tau \sigma^2_{\mathbf{M}}}{\bar{\Phi}^2}(\boldsymbol{\beta} - \mathbf{1})}{\left(1 + \frac{\gamma^2}{\tau_m \tau_\epsilon}\right)\sigma^2_{\mathbf{M}} + \frac{\gamma^2 e_1 \bar{\Phi}^2}{\tau_m \tau} + \frac{\tau_v \bar{\Phi}^2}{\tau^2}}. \tag{A.60}$$

Thus:

$$\widehat{\boldsymbol{\beta}} - \mathbf{1} = \left[1 + \left(\frac{\tau \sigma^2_{\mathbf{M}}}{\bar{\Phi}^2} - 1\right)\left(\frac{\gamma^2 e_1 \bar{\Phi}^2}{\tau_m \tau \widehat{\sigma}^2_{\mathbf{M}}} + \mathcal{D}^2\right)\right](\boldsymbol{\beta} - \mathbf{1}). \tag{A.61}$$

and thus the distortion δ is

$$\delta = \left(\frac{\tau \sigma^2_{\mathbf{M}}}{\bar{\Phi}^2} - 1\right)\left(\frac{\gamma^2 e_1 \bar{\Phi}^2}{\tau_m \tau \widehat{\sigma}^2_{\mathbf{M}}} + \mathcal{D}^2\right) = \left(\frac{\tau \sigma^2_{\mathbf{M}}}{\bar{\Phi}^2} - 1\right)\left(\mathcal{C}^2 \frac{\tau_\epsilon e_1 \bar{\Phi}^2}{\tau \sigma^2_{\mathbf{M}} + \tau_\epsilon e_1 \bar{\Phi}^2} + \mathcal{D}^2\right). \tag{A.62}$$

The last equality follows from (A.56). 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 we discuss below. It also illustrates 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). We describe here only the case of a vanishing idiosyncratic component in payoffs, $\tau_\epsilon \to \infty$. In this case, $\sigma^2_{\mathbf{M}} = \bar{\Phi}^2/\tau$, and there can be no distortion in empiricist's SML. However, the term $(\tau \sigma^2_{\mathbf{M}}/\bar{\Phi}^2 - 1)$ is in fact much larger than zero; this term measures excess variance in the market. We find that a plausible value for $(\tau \sigma^2_{\mathbf{M}}/\bar{\Phi}^2 - 1)$ 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. Section 6.3 analyzes a large economy, in which we let both the number of stocks and the number of factors grow to infinity.

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	$\tau_F + \tau_v + \frac{\tau_m \tau_v^2}{\gamma^2}$	∞	∞	∞
$\boldsymbol{\Sigma}$	$\frac{1}{\tau}\boldsymbol{\Phi}\boldsymbol{\Phi}'$	$\frac{1}{\tau_\epsilon}\mathbf{I}$	$\frac{1}{\tau_\epsilon}\mathbf{I}$	$\frac{1}{\tau_\epsilon}\mathbf{I}$
$\sigma_{\mathbf{M}}^2$	$\frac{\bar{\Phi}^2}{\tau}$	$\frac{1}{N\tau_\epsilon}$	$\frac{1}{N\tau_\epsilon}$	$\frac{1}{N\tau_\epsilon}$
e_1	$\frac{1}{\tau}$	$\frac{1}{\tau_\epsilon}$	$\frac{1}{\tau_\epsilon}$	$\frac{1}{\tau_\epsilon}$
$\widehat{\boldsymbol{\Sigma}}$	$\left(\frac{1}{\tau} + \frac{\gamma^2 e_1}{\tau_m \tau} + \frac{\tau_v}{\tau^2}\right)\boldsymbol{\Phi}\boldsymbol{\Phi}'$	$\frac{1}{\tau_\epsilon}\mathbf{I}$	$\left(\frac{1}{\tau_\epsilon} + \frac{\gamma^2}{\tau_m \tau_\epsilon^2}\right)\mathbf{I}$	$\frac{1}{\tau_\epsilon}\mathbf{I}$
$\widehat{\sigma}_{\mathbf{M}}^2$	$\left(\frac{1}{\tau} + \frac{\gamma^2 e_1}{\tau_m \tau} + \frac{\tau_v}{\tau^2}\right)\bar{\Phi}^2$	$\frac{1}{N\tau_\epsilon}$	$\frac{1}{N\tau_\epsilon} + \frac{\gamma^2}{N\tau_m \tau_\epsilon^2}$	$\frac{1}{N\tau_\epsilon}$
$\mathcal{C}^2$	$\frac{\gamma^2 e_1 \bar{\Phi}^2}{\tau_m \tau \widehat{\sigma}_{\mathbf{M}}^2}$	0	$\frac{\gamma^2}{\gamma^2 + \tau_m \tau_\epsilon}$	0
$\mathcal{D}^2$	$\frac{\tau_v \bar{\Phi}^2}{\tau \widehat{\sigma}_{\mathbf{M}}^2}$	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 $\mathbf{1}$. 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 and τ_m 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:

$$\frac{\mathrm{d}}{\mathrm{d}h}\delta > 0. \tag{A.63}$$

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

$$\frac{\mathrm{d}}{\mathrm{d}h}\tau = -\frac{\tau_P(\tau_F + \tau_P + \tau_v + \tau_\epsilon)}{h(\tau_F + 3\tau_P + \tau_v + \tau_\epsilon)} < 0, \tag{A.64}$$

and simplifying using (20) whereby $\tau_v + \tau_P + \tau_\epsilon = \frac{\tau_\epsilon \tau_v}{\tau_P^{1/2} h^{1/2}}$, we obtain:

$$\frac{\mathrm{d}}{\mathrm{d}h}\delta = \frac{\tau_\epsilon^3\left(\sqrt{h\tau_P} - \tau_v\right)^2}{(h\tau_P)^{3/2}} \frac{y}{h(3\tau_P + \tau_v + \tau_\epsilon)\left(\tau_\epsilon \phi^2(\tau_\epsilon(h + \tau_P + 2\tau_v) + 2h(\tau_P + \tau_v)) + (h + \tau_\epsilon)(\tau_P + \tau_v)^2\right)^2}, \tag{A.65}$$

where

$$y \equiv \left(\begin{array}{c} h^2\tau_P\tau_v\tau_\epsilon\left(1-\phi^2\right) + h\tau_P^2\tau_v\tau_\epsilon\left(5-\phi^2\right) + 2(h\tau_P)^{3/2}\tau_v\tau_\epsilon\left(1-\phi^2\right) \\ +2(h\tau_P)^{3/2}\left((\tau_v+\tau_\epsilon)^2-\tau_P^2\right) + h\tau_v^2\tau_\epsilon^2 + \tau_P\tau_v^2\tau_\epsilon^2 \end{array} \right), \tag{A.66}$$

and $\phi \equiv N^{1/2}\bar{\Phi}$. The numerator and the first ratio in (A.65) are positive, so we focus on the sign of y. Substitute from (20):

$$\tau_v + \tau_\epsilon = \frac{\tau_\epsilon\tau_v}{\tau_P^{1/2}h^{1/2}} - \tau_P, \tag{A.67}$$

into (A.66) to obtain:

$$y = \tau_v\tau_\epsilon\left(2\sqrt{h\tau_P} + h + \tau_P\right)\left(h\tau_P\left(1-\phi^2\right) + \tau_v\tau_\epsilon\right) > 0. \tag{A.68}$$

By Hölder's inequality and the normalization $\mathbf{\Phi'\Phi} \equiv 1$, $\phi \in (-1,1)$ and thus $y > 0$. □

A.6.1 Conditioning on public information

Empiricist's rejection of the CAPM (Proposition 4) 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 4 still holds, but with a different distortion, $\breve{\delta}$:

$$\breve{\delta} = \frac{\tau_v}{\tau_\epsilon(\tau-\tau_v)}\frac{1}{N\breve{\sigma}_{\mathbf{M}}^2} > 0, \tag{A.69}$$

where $\breve{\sigma}_{\mathbf{M}}^2 \equiv \mathbf{M'}\,\mathrm{Var}[\widetilde{\mathbf{R}}^e|\widetilde{\mathbf{P}}]\mathbf{M}$ is the variance of excess returns on the market portfolio conditional on observing all publicly available prices, and the conditional covariance matrix $\mathrm{Var}[\widetilde{\mathbf{R}}^e|\widetilde{\mathbf{P}}]$ is (which can be obtained by simply assuming the view of an agent who only observes prices and thus has zero precision of private information):[34]

$$\mathrm{Var}[\widetilde{\mathbf{R}}^e|\widetilde{\mathbf{P}}] = \mathrm{Var}[\widetilde{\mathbf{D}}|\widetilde{\mathbf{P}}] = \frac{1}{\tau-\tau_v}\mathbf{\Phi\Phi'} + \frac{1}{\tau_\epsilon}\mathbf{I}. \tag{A.70}$$

Replace (A.21) in (A.70) to obtain

$$\mathrm{Var}[\widetilde{\mathbf{R}}^e|\widetilde{\mathbf{P}}] = \frac{\tau}{\tau-\tau_v}\mathbf{\Sigma} - \frac{\tau_v}{\tau_\epsilon(\tau-\tau_v)}\mathbf{I}, \tag{A.71}$$

and

$$\breve{\sigma}_{\mathbf{M}}^2 = \frac{\overline{\Phi}^2}{\tau-\tau_v} + \frac{1}{N\tau_\epsilon}. \tag{A.72}$$

[34] Notice that it is not necessary to assume here that the empiricist knows the price coefficients of Proposition 2. This is because $\mathrm{Var}[\widetilde{\mathbf{R}}^e|\widetilde{\mathbf{P}}] = \mathrm{Var}[\widetilde{\mathbf{R}}^e] - \mathrm{Var}[\mathbb{E}[\widetilde{\mathbf{R}}^e|\widetilde{\mathbf{P}}]]$. 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 $\widetilde{\mathbf{P}}$.

When controlling for prices, the empiricist obtains a new set of betas:

$$\breve{\boldsymbol{\beta}} = \frac{\mathrm{Var}[\widetilde{\mathbf{R}}^e|\widetilde{\mathbf{P}}]\mathbf{M}}{\breve{\sigma}_{\mathbf{M}}^2} = \underbrace{\frac{\tau}{\tau-\tau_v}\frac{\widehat{\sigma}_{\mathbf{M}}^2}{\breve{\sigma}_{\mathbf{M}}^2}}_{\equiv(1+\breve{\delta})}\boldsymbol{\beta} - \frac{\tau_v}{\tau_\epsilon(\tau-\tau_v)}\frac{\mathbf{M}}{\breve{\sigma}_{\mathbf{M}}^2}. \tag{A.73}$$

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

$$1 = (1+\breve{\delta}) - \frac{\tau_v}{\tau_\epsilon(\tau-\tau_v)}\frac{1}{N\breve{\sigma}_{\mathbf{M}}^2}, \tag{A.74}$$

and thus we obtain $\breve{\delta}$ as in (A.69). Replacing $\breve{\delta}$ in (A.73) and subtracting $\mathbf{1}$ on both sides yields the main result of Proposition 4:

$$\breve{\boldsymbol{\beta}} - \mathbf{1} = (1+\breve{\delta})(\boldsymbol{\beta}-\mathbf{1}). \tag{A.75}$$

The new distortion $\breve{\delta}$ in (A.69) may be larger than the initial distortion obtained without conditioning, δ. To see this, replace (A.72) in (A.69) and use (16) to obtain

$$\breve{\delta} = \frac{\tau_v\tau_\epsilon^{-1}/N}{\tau\sigma_{\mathbf{M}}^2 - \tau_v\tau_\epsilon^{-1}/N}. \tag{A.76}$$

On the other hand we can rewrite δ in the model as:

$$\delta = \frac{\tau_\epsilon^{-1}/N}{\widehat{\sigma}_{\mathbf{M}}^2}\left(\frac{\gamma^2}{\tau_m}e_1 + \frac{\tau_v}{\tau}\right), \tag{A.77}$$

which further leads to

$$\breve{\delta} = \frac{\tau\widehat{\sigma}_{\mathbf{M}}^2}{\tau\sigma_{\mathbf{M}}^2 - \tau_v\tau_\epsilon^{-1}/N}\left(\frac{\gamma^2\tau}{\tau_m\tau_v}e_1 + 1\right)^{-1}\delta. \tag{A.78}$$

The coefficient multiplying δ on the right hand side is greater than 1 iff

$$\tau_v > \sqrt{\frac{\gamma^2(\tau_F+\tau_P)(\tau_F+\tau_P+\tau_\epsilon)}{\gamma^2+\tau_m\tau_\epsilon}}. \tag{A.79}$$

Thus, when investors' private information is sufficiently precise, the empiricist obtains a stronger CAPM distortion than when estimating a standard unconditional CAPM. Since we know from Table A1 that $\lim_{\tau_v\to\infty}\tau_P = \frac{\tau_m\tau_\epsilon^2}{\gamma^2}$, the right-hand side of (A.79) is finite as $\tau_v\to\infty$. Thus, (A.79) clearly has a solution. When τ_v is larger than this solution, the empiricist's distortion is stronger when controlling for prices.

A.7 Proof of Proposition 6

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

$$\tau = \tau_F + \left(\frac{\sqrt{\tau_P}}{\sqrt{\tau_m}}\boldsymbol{\Phi}\right)'\tau_g\left(\frac{\sqrt{\tau_P}}{\sqrt{\tau_m}}\boldsymbol{\Phi}\right) = \tau_F + \frac{\tau_g}{\tau_m}\tau_P, \tag{A.80}$$

as in (35) in the text.

Finding the equilibrium follows the same steps as in Appendix A.2, albeit much simpler here since there is no learning from prices. To prove Proposition 6, compare the two sets of betas:

$$\widehat{\boldsymbol{\beta}} = \frac{\left(1+\frac{\gamma^2}{\tau_m\tau_\epsilon}\right)\sigma_{\mathbf{M}}^2\boldsymbol{\beta} + \left(\frac{\gamma^2 e_1\bar{\Phi}^2}{\tau_m\tau} + \frac{\tau_v\bar{\Phi}^2}{\tau^2}\right)\frac{\boldsymbol{\Phi}}{\bar{\Phi}}}{\left(1+\frac{\gamma^2}{\tau_m\tau_\epsilon}\right)\sigma_{\mathbf{M}}^2 + \frac{\gamma^2 e_1\bar{\Phi}^2}{\tau_m\tau} + \frac{\tau_v\bar{\Phi}^2}{\tau^2}} \tag{A.81}$$

$$\widehat{\boldsymbol{\beta}}_{\text{CIE}} = \frac{\left(1+\frac{\gamma^2}{\tau_m\tau_\epsilon}\right)\sigma_{\mathbf{M}}^2\boldsymbol{\beta} + \frac{\gamma^2 e_1\bar{\Phi}^2}{\tau_m\tau}\frac{\boldsymbol{\Phi}}{\bar{\Phi}}}{\left(1+\frac{\gamma^2}{\tau_m\tau_\epsilon}\right)\sigma_{\mathbf{M}}^2 + \frac{\gamma^2 e_1\bar{\Phi}^2}{\tau_m\tau}}. \tag{A.82}$$

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

To show that $\mathcal{C}_{\text{CIE}}^2 > \mathcal{C}^2$, write first

$$\widehat{\boldsymbol{\Sigma}}_{\text{CIE}} = \boldsymbol{\Sigma} + \underbrace{\frac{\gamma^2}{\tau_m}\left(\frac{1}{\tau_\epsilon}\boldsymbol{\Sigma} + \frac{e_1}{\tau}\boldsymbol{\Phi}\boldsymbol{\Phi}'\right)}_{\equiv \text{Var}[\overline{\mathbb{E}}[\widetilde{\mathbf{R}}^e]]}. \tag{A.83}$$

Compared with (A.54), we notice that the last term is missing (because $\tau_v = 0$ in the CIE). Thus,

$$\mathcal{C}_{\text{CIE}}^2 = \frac{\text{Var}[\overline{\mathbb{E}}[\widetilde{R}_{\mathbf{M}}^e]]}{\mathbf{M}'\widehat{\boldsymbol{\Sigma}}_{\text{CIE}}\mathbf{M}} \tag{A.84}$$

is now larger, due to a lower denominator (the numerator is the same in both economies). □

Information updating wedge (Albagli et al., 2022). To see how dispersed information creates an *information updating wedge* in our setting, we adapt the model to make the exact same argument as that in Albagli et al. (2022), Section 2.1, Example 1. Consider a single-asset version of the baseline model in which everything else remains unchanged and $\boldsymbol{\Phi} \equiv 1$. Following the steps of Appendix A.2 (now greatly simplified) leads to

$$\mathbb{E}[\widetilde{D}|\mathscr{F}^i] = D + \frac{\tau_v}{\tau}\widetilde{V}^i + \sqrt{\frac{\tau_P}{\tau_m}}\frac{\tau_m}{\tau}\widetilde{Z}, \tag{A.85}$$

where $\widetilde{Z} \equiv \xi^{-1}\widetilde{P}^a = \sqrt{\tau_P/\tau_m}\widetilde{F} + \widetilde{m}$ is a public signal defined as in (33), informationally equivalent to the equilibrium asset price, $\widetilde{P}$.

Now, consider the investor who, at a given price $\widetilde{P}$, finds it optimal to hold exactly M units of the asset (notice that this is the same investor for which the unconditional CAPM holds in our model). Then the following must hold for this investor:

$$\frac{1}{\gamma}\Sigma^{-1}(\mathbb{E}[\widetilde{D}|\mathscr{F}^i] - \overline{\mathbb{E}}[\widetilde{D}]) = \widetilde{m}, \quad \text{where } \Sigma = \frac{1}{\tau} + \frac{1}{\tau_\epsilon}, \tag{A.86}$$

from which we conclude that the private signal of this investor must exactly be:

$$\widetilde{V}^i = \widetilde{F} + \gamma\Sigma\frac{\tau}{\tau_v}\widetilde{m} = \widetilde{F} + \frac{\gamma(\tau+\tau_\epsilon)}{\tau_v\tau_\epsilon}\widetilde{m} = \sqrt{\frac{\tau_m}{\tau_P}}\widetilde{Z}, \tag{A.87}$$

where the last equality follows from the definition of τ_P in (A.31). Thus, the private signal of the investor who finds it optimal to hold exactly M units of the asset is informationally equivalent to the public price signal, $\widetilde{Z}$. Eq. (A.85) then leads to

$$\mathbb{E}\left[\widetilde{D}\middle|\,\widetilde{V}^i=\sqrt{\frac{\tau_m}{\tau_P}}\widetilde{Z},\widetilde{Z}\right]=D+\frac{\tau_v}{\tau}\sqrt{\frac{\tau_m}{\tau_P}}\widetilde{Z}+\sqrt{\frac{\tau_P}{\tau_m}}\frac{\tau_m}{\tau}\widetilde{Z}=D+\frac{\tau_v+\tau_P}{\tau}\sqrt{\frac{\tau_m}{\tau_P}}\widetilde{Z}. \tag{A.88}$$

Therefore, one can write the equilibrium price as the risk-adjusted expectation of an investor who chooses to hold exactly M units of the asset:

$$P(\widetilde{Z})=\mathbb{E}\left[\widetilde{D}\middle|\,\sqrt{\tau_P/\tau_m}\widetilde{V}^i=\widetilde{Z},\widetilde{Z}\right]-\gamma\Sigma M. \tag{A.89}$$

Relative to the "objective" Bayesian posterior of $\widetilde{D}$ given $\widetilde{Z}$, which satisfies

$$\mathbb{E}[\widetilde{D}\mid\widetilde{Z}]=D+\frac{\tau_P}{\tau}\sqrt{\frac{\tau_m}{\tau_P}}\widetilde{Z}, \tag{A.90}$$

we notice that the price, $P(\widetilde{Z})$, responds more strongly to the market signal $\widetilde{Z}$. It follows that the equilibrium price is more sensitive to fundamental and noise trading shocks contained in $\widetilde{Z}$. In other words, in the model with dispersed information investors treat market (public) information as more informative than it truly is. This excess price sensitivity is the *information updating wedge* of Albagli et al. (2022), and is the very reason why building the *Common Information Economy* in Section 4.2 requires increasing the precision of the public signal in (35) relative to $\widetilde{Z}$ above.

A.8 Proof of Proposition 7

The proof starts from (4),

$$\widehat{\boldsymbol{\Sigma}}=\boldsymbol{\Sigma}+\mathrm{Var}[\overline{\mathbb{E}}[\widetilde{\mathbf{R}}^e]]+\mathrm{Var}[\mathbb{E}^{i*}[\widetilde{\mathbf{R}}^e]], \tag{A.91}$$

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

$$\widehat{\sigma}^2_{\mathbf{M}}\widehat{\boldsymbol{\beta}}=\sigma^2_{\mathbf{M}}\boldsymbol{\beta}+\mathrm{Var}[\overline{\mathbb{E}}[\widetilde{R}^e_{\mathbf{M}}]]\boldsymbol{\beta}^{\mathcal{C}}+\mathrm{Var}[\mathbb{E}^{i*}[\widetilde{R}^e_{\mathbf{M}}]]\boldsymbol{\beta}^{\mathcal{D}}, \tag{A.92}$$

and then divide by $\widehat{\sigma}^2_{\mathbf{M}}$ to obtain (38). □

A.8.1 Proof of Corollary 7.1

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

$$\boldsymbol{\beta}^{\mathcal{C}}=\frac{\frac{\gamma^2\sigma^2_{\mathbf{M}}}{\tau_m\tau_\epsilon}\boldsymbol{\beta}+\frac{\gamma^2e_1\bar{\Phi}^2}{\tau_m\tau}\frac{\boldsymbol{\Phi}}{\bar{\Phi}}}{\frac{\gamma^2\sigma^2_{\mathbf{M}}}{\tau_m\tau_\epsilon}+\frac{\gamma^2e_1\bar{\Phi}^2}{\tau_m\tau}}, \tag{A.93}$$

which leads to (40) in the text:

$$\boldsymbol{\beta}^{\mathcal{C}}-\boldsymbol{\beta}=\frac{\frac{\gamma^2e_1\bar{\Phi}^2}{\tau_m\tau}}{\frac{\gamma^2\sigma^2_{\mathbf{M}}}{\tau_m\tau_\epsilon}+\frac{\gamma^2e_1\bar{\Phi}^2}{\tau_m\tau}}\left(\frac{\boldsymbol{\Phi}}{\bar{\Phi}}-\boldsymbol{\beta}\right)=\left(\frac{\tau\sigma^2_{\mathbf{M}}}{\bar{\Phi}^2}-1\right)\frac{\tau_\epsilon e_1\bar{\Phi}^2}{\tau\sigma^2_{\mathbf{M}}+\tau_\epsilon e_1\bar{\Phi}^2}(\boldsymbol{\beta}-\mathbf{1}), \tag{A.94}$$

where we have used Corollary 2.1 under the form $\boldsymbol{\Phi}/\bar{\Phi}-\boldsymbol{\beta}=(\tau\sigma^2_{\mathbf{M}}/\bar{\Phi}^2-1)(\boldsymbol{\beta}-\mathbf{1})$.

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

$$\boldsymbol{\beta}^{\mathcal{D}} - \boldsymbol{\beta} = \frac{\boldsymbol{\Phi}}{\bar{\Phi}} - \boldsymbol{\beta} = \left(\frac{\tau\sigma^2_{\mathbf{M}}}{\bar{\Phi}^2} - 1\right)(\boldsymbol{\beta} - \mathbf{1}). \tag{A.95}$$

which is (41) in the text. □

A.9 Proof of Proposition 8

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

$$\mathrm{Cov}[\boldsymbol{\beta}, \widehat{\boldsymbol{\beta}}] = \frac{1}{N}\frac{\mathbf{M}'\boldsymbol{\Sigma}\widehat{\boldsymbol{\Sigma}}\mathbf{M}}{\widehat{\sigma}^2_{\mathbf{M}}\sigma^2_{\mathbf{M}}} - \beta_{\text{avg}}\widehat{\beta}_{\text{avg}} \tag{A.96}$$

$$\mathrm{Var}[\widehat{\boldsymbol{\beta}}] = \frac{1}{N}\frac{\mathbf{M}'\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{\Sigma}}\mathbf{M}}{\widehat{\sigma}^2_{\mathbf{M}}\widehat{\sigma}^2_{\mathbf{M}}} - \widehat{\beta}^2_{\text{avg}}, \tag{A.97}$$

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 $\mathbf{M} = \mathbf{1}/N$, $\mathrm{Cov}[\boldsymbol{\beta}, \widehat{\boldsymbol{\beta}}] < \mathrm{Var}[\widehat{\boldsymbol{\beta}}]$ if and only if

$$\frac{\mathbf{M}'\boldsymbol{\Sigma}\widehat{\boldsymbol{\Sigma}}\mathbf{M}}{\mathbf{M}'\boldsymbol{\Sigma}\mathbf{M}} < \frac{\mathbf{M}'\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{\Sigma}}\mathbf{M}}{\mathbf{M}'\widehat{\boldsymbol{\Sigma}}\mathbf{M}}. \tag{A.98}$$

If we further assume that there is no dispersion in beliefs and investors' information is common knowledge, then the following relation results immediately from (13):

$$\widehat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma} + \frac{\gamma^2}{\tau_m}\boldsymbol{\Sigma}\boldsymbol{\Sigma}. \tag{A.99}$$

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),

$$\boldsymbol{\Sigma} = \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}' \quad \text{and} \quad \widehat{\boldsymbol{\Sigma}} = \mathbf{Q}\boldsymbol{\Lambda}(\mathbf{I} + h\boldsymbol{\Lambda})\mathbf{Q}' \tag{A.100}$$

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

$$\frac{\mathbf{O}'\Lambda\Lambda(\mathbf{I} + h\boldsymbol{\Lambda})\mathbf{O}}{\mathbf{O}'\Lambda\mathbf{O}} < \frac{\mathbf{O}'\Lambda(\mathbf{I} + h\boldsymbol{\Lambda})\Lambda(\mathbf{I} + h\boldsymbol{\Lambda})\mathbf{O}}{\mathbf{O}'\Lambda(\mathbf{I} + h\boldsymbol{\Lambda})\mathbf{O}}, \tag{A.101}$$

or

$$\sum \frac{O_n^2\Lambda_n}{\sum O_n^2\Lambda_n}\Lambda_n(1 + h\Lambda_n) < \sum \frac{O_n^2\Lambda_n(1 + h\Lambda_n)}{\sum O_n^2\Lambda_n(1 + h\Lambda_n)}\Lambda_n(1 + h\Lambda_n). \tag{A.102}$$

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

$$\Omega_{1n} = \frac{O_n^2\Lambda_n}{\sum O_n^2\Lambda_n} = \frac{O_n^2\Lambda_n}{A}, \quad \sum_{n=1}^{N}\Omega_{1n} = 1 \tag{A.103}$$

$$\Omega_{2n} = \frac{O_n^2\Lambda_n(1+h\Lambda_n)}{\sum O_n^2\Lambda_n(1+h\Lambda_n)} = \frac{O_n^2\Lambda_n(1+h\Lambda_n)}{A+hB}, \quad \sum_{n=1}^{N}\Omega_{2n} = 1, \tag{A.104}$$

where $A \equiv \sum O_n^2\Lambda_n > 0$ and $B \equiv \sum O_n^2\Lambda_n^2 > 0$, and thus $B/A \in (\min\Lambda_n, \max\Lambda_n)$ is a weighted average of Λ_n. The difference between the weights Ω_{2n} and Ω_{1n} is:

$$\begin{aligned}
\Omega_{2n} - \Omega_{1n} &= \frac{O_n^2\Lambda_n(1+h\Lambda_n)}{A+hB} - \frac{O_n^2\Lambda_n}{A} && \text{(A.105)}\\
&= \left(\frac{1}{A+hB} - \frac{1}{A}\right)O_n^2\Lambda_n + \frac{h}{A+hB}O_n^2\Lambda_n^2 && \text{(A.106)}\\
&= -\frac{hB}{A(A+hB)}O_n^2\Lambda_n + \frac{h}{A+hB}O_n^2\Lambda_n^2 && \text{(A.107)}\\
&= \frac{h}{A+hB}O_n^2\Lambda_n\left(\Lambda_n - \frac{B}{A}\right). && \text{(A.108)}
\end{aligned}$$

This is a quadratic function of Λ_n, with two real roots: $\Lambda_n = 0$ and $\Lambda_n = B/A > 0$. Importantly, the 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 (keeping in mind that $\sum_{n=1}^{N}\Omega_{1n} = \sum_{n=1}^{N}\Omega_{2n} = 1$):

$$\begin{cases} \Omega_{2n} < \Omega_{1n}, & \text{if } \Lambda_n \in (0, B/A) \\ \Omega_{2n} > \Omega_{1n}, & \text{if } \Lambda_n > B/A, \end{cases} \tag{A.109}$$

Thus, the weighted average on the right hand side of (A.102) places strictly higher weights on higher values, and the inequality is now verified. □

A.10 Proof of Proposition 9

The relation of Lemma 1 remains valid regardless the value of $\mathbf{M}$. Writing (A.43) again,

$$\widehat{\mathbf{\Sigma}} = c_3\mathbf{\Sigma} - c_4\mathbf{I}, \tag{A.110}$$

where c_3, and c_4 are positive scalars defined in (A.45), and multiplying with $\mathbf{M}$ yields

$$\widehat{\boldsymbol{\beta}} = \frac{c_3\sigma_{\mathbf{M}}^2}{\widehat{\sigma}_{\mathbf{M}}^2}\boldsymbol{\beta} - \frac{c_4}{\widehat{\sigma}_{\mathbf{M}}^2}\mathbf{M} \tag{A.111}$$

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

$$1 = \frac{c_3\sigma_{\mathbf{M}}^2}{\widehat{\sigma}_{\mathbf{M}}^2} - \frac{c_4}{\widehat{\sigma}_{\mathbf{M}}^2}\mathbf{M}'\mathbf{M}. \tag{A.112}$$

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 (A.111) using (A.112):

$$\widehat{\boldsymbol{\beta}} = \left(1 + \frac{c_4\mathbf{M}'\mathbf{M}}{\widehat{\sigma}_{\mathbf{M}}^2}\right)\boldsymbol{\beta} - \frac{c_4\mathbf{M}'\mathbf{M}}{\widehat{\sigma}_{\mathbf{M}}^2}\frac{\mathbf{M}}{\mathbf{M}'\mathbf{M}} = (1+\delta)\boldsymbol{\beta} - \delta\frac{\mathbf{M}}{\mathbf{M}'\mathbf{M}}, \tag{A.113}$$

or

$$\boldsymbol{\beta} - \mathbf{1} = \frac{1}{1+\delta}(\widehat{\boldsymbol{\beta}} - \mathbf{1}) + \frac{\delta}{1+\delta}\left(\frac{\mathbf{M}}{\mathbf{M}'\mathbf{M}} - \mathbf{1}\right). \tag{A.114}$$

where

$$\delta = \frac{\mathbf{M}'\mathbf{M}}{\widehat{\sigma}^2_{\mathbf{M}}}\left(\frac{\gamma^2 e_1}{\tau_m \tau_\epsilon} + \frac{\tau_v}{\tau_\epsilon \tau}\right). \tag{A.115}$$

Multiplying (A.114) with $\mu_{\mathbf{M}}$ and using the true CAPM relation $\boldsymbol{\mu} = \boldsymbol{\beta}\mu_{\mathbf{M}}$ yields (54). □

A.11 Proof of Condition (55)

For the SML to be downward-sloping a necessary (but not sufficient) condition is that $\mathrm{Cov}[\boldsymbol{\beta}, \widehat{\boldsymbol{\beta}}] < 0$. Write the covariance between $\boldsymbol{\beta}$ and $\widehat{\boldsymbol{\beta}}$ as:

$$\mathrm{Cov}[\boldsymbol{\beta}, \widehat{\boldsymbol{\beta}}] = \frac{1}{N\widehat{\sigma}^2_{\mathbf{M}}\sigma^2_{\mathbf{M}}}\underbrace{\left(\mathbf{M}'\boldsymbol{\Sigma}\widehat{\boldsymbol{\Sigma}}\mathbf{M} - \mathbf{1}'\boldsymbol{\Sigma}\mathbf{M}\mathbf{1}'\widehat{\boldsymbol{\Sigma}}\mathbf{M}/N\right)}_{\equiv\Delta}. \tag{A.116}$$

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

$$\begin{aligned}\Delta \equiv{} & \tau^{-2}\left(1 + \frac{\gamma^2}{\tau_m\tau_\epsilon} + \frac{\gamma^2 e_1}{\tau_m} + \frac{\tau_n}{\tau}\right)\mathbf{M}'\boldsymbol{\Phi}\underbrace{\left(1 - N\bar{\Phi}^2\right)}_{\geq 0} + \tau_\epsilon^{-2}\left(1 + \frac{\gamma^2}{\tau_m\tau_\epsilon}\right)\underbrace{\left(\|\mathbf{M}\|^2 - 1/N\right)}_{\geq 0} \\ & + \tau^{-1}\tau_\epsilon^{-1}\left(2\left(1 + \frac{\gamma^2}{\tau_m\tau_\epsilon}\right) + \frac{\gamma^2 e_1}{\tau_m} + \frac{\tau_n}{\tau}\right)\mathbf{M}'\boldsymbol{\Phi}\,\mathrm{Cov}[\mathbf{M}, \boldsymbol{\Phi}]\end{aligned} \tag{A.117}$$

The sign under the first underbrace 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 underbrace 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 $\mathrm{Cov}[\mathbf{M}, \boldsymbol{\Phi}] < 0$. □

A.12 Proofs related to Section 6.3

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

$$\boldsymbol{\tau} = \mathrm{Var}\left[\widetilde{\mathbf{F}}\middle|\mathscr{F}^i\right]^{-1} = (\tau_F + \tau_v)J\mathbf{I}_J + \boldsymbol{\tau}_P\boldsymbol{\Phi}'\boldsymbol{\Phi}\boldsymbol{\tau}_P, \tag{A.118}$$

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 (59) and to work with the limiting behavior of $\mathbf{Q}'\boldsymbol{\tau}/N\mathbf{Q}$, as opposed to $\boldsymbol{\tau}/N$:

$$\boldsymbol{\tau}_\infty \equiv \lim_{J\to\infty, N\to\infty, J/N\to\psi} \mathbf{Q}'\boldsymbol{\tau}/N\mathbf{Q}. \tag{A.119}$$

We characterize this matrix in the next lemma. The proof is in Section A.12.1.

Lemma 2. *The limiting precision matrix, $\boldsymbol{\tau}_\infty$, defined in (A.119) 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:*

$$\tau_\infty(\lambda) = \psi\left(\tau_F + \tau_v + \frac{\lambda\tau_m\tau_v^2\tau_\epsilon^2\psi}{\gamma^2(\tau_\infty(\lambda)+\lambda\tau_\epsilon)^2}\right). \tag{A.120}$$

Remarkably, after rotating the precision matrix based on (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 2 J times, each time for a different eigenvalue identifying a factor. Interestingly, comparing (A.120) and (A.118) shows that the ratio in (A.120) 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, (A.119) simplifies to:

$$\boldsymbol{\tau}_\infty = (\tau_F + \tau_v)\psi\mathbf{I} + O(\psi^2), \tag{A.121}$$

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:

$$1+\delta = \mathrm{Cov}[\boldsymbol{\beta}, \widehat{\boldsymbol{\beta}}]/\mathrm{Var}[\boldsymbol{\beta}]. \tag{A.122}$$

Hence, when $\delta > 0$ ($\delta < 0$) the SML looks flatter (steeper) than it actually is. From this definition we obtain Proposition 10 and the corollary below; the proof is in Section A.12.2.

Corollary 10.1. (steepening of empiricist's SML) *Consider the framework of Proposition 10. If eigenvalues are sufficiently dispersed (inequality in* (60) *is violated) and exhibit a negative but limited skew (first inequality in* (A.183) *holds) or if eigenvalues are sufficiently concentrated (inequality in* (A.184) *is violated) and exhibit a strictly negative skew (inequality in* (A.182) *is violated), the SML will look steeper than it actually is.*

A.12.1 Proof of Lemma 2

We start by repeating the steps of Appendix A.2 for the multiple-factor case. As is customary, we conjecture that prices satisfy

$$\widetilde{\mathbf{P}} = \boldsymbol{\xi}_0\mathbf{M} + \boldsymbol{\lambda}\widetilde{\mathbf{F}} + \boldsymbol{\xi}\widetilde{\mathbf{m}}, \tag{A.123}$$

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

$$\boldsymbol{\tau} \equiv \mathrm{Var}\left[\widetilde{\mathbf{F}}\middle|\mathscr{F}_i\right]^{-1} = (\tau_F + \tau_v)J\mathbf{I}_J + \boldsymbol{\lambda}'(\boldsymbol{\xi}\boldsymbol{\xi}')^{-1}\boldsymbol{\lambda}\tau_m. \tag{A.124}$$

and

$$\mathbb{E}\left[\widetilde{\mathbf{F}}\middle|\mathscr{F}_i\right] = \boldsymbol{\tau}^{-1}\left((\boldsymbol{\tau} - \tau_F\mathbf{I}_J)\widetilde{\mathbf{F}} + \tau_m\boldsymbol{\lambda}'(\boldsymbol{\xi}\boldsymbol{\xi}')^{-1}\boldsymbol{\xi}\widetilde{\mathbf{m}} + \tau_v\tilde{\mathbf{v}}_i\right). \tag{A.125}$$

It follows that average expectations of future payoffs satisfy

$$\bar{\mathbb{E}}[\widetilde{\mathbf{D}}] \equiv \int_i \mathbb{E}[\widetilde{\mathbf{D}}|\mathscr{F}_i]\mathrm{d}i = D\mathbf{1} + \mathbf{\Phi}\boldsymbol{\tau}^{-1}\left((\boldsymbol{\tau} - \tau_F\mathbf{I}_J)\widetilde{\mathbf{F}} + \tau_m\boldsymbol{\lambda}'(\boldsymbol{\xi}\boldsymbol{\xi}')^{-1}\boldsymbol{\xi}\widetilde{\mathbf{m}}\right). \tag{A.126}$$

and the conditional covariance matrix of future payoffs satisfies:

$$\mathbf{\Sigma} \equiv \mathrm{Var}\left[\widetilde{\mathbf{D}}\middle|\mathscr{F}_i\right] = \mathbf{\Phi}\boldsymbol{\tau}^{-1}\mathbf{\Phi}' + \tau_\epsilon^{-1}\mathbf{I}_N. \tag{A.127}$$

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

$$\widetilde{\mathbf{P}} = D\mathbf{1} + \mathbf{\Phi}\boldsymbol{\tau}^{-1}\left((\boldsymbol{\tau} - \tau_F\mathbf{I}_J)\widetilde{\mathbf{F}} + \tau_m(\boldsymbol{\xi}^{-1}\boldsymbol{\lambda})'\widetilde{\mathbf{m}}\right) - \gamma\mathbf{\Sigma}(\mathbf{M} + \widetilde{\mathbf{m}}). \tag{A.128}$$

Separating variables we obtain the following system of equations:

$$\boldsymbol{\xi}_0 = -\gamma\mathbf{\Sigma}, \quad \boldsymbol{\lambda} = \mathbf{\Phi}\boldsymbol{\tau}^{-1}(\boldsymbol{\tau} - \tau_F\mathbf{I}_J), \tag{A.129}$$

and

$$\boldsymbol{\xi} = \tau_m\mathbf{\Phi}\boldsymbol{\tau}^{-1}(\boldsymbol{\xi}^{-1}\boldsymbol{\lambda})' - \gamma\left(\mathbf{\Phi}\boldsymbol{\tau}^{-1}\mathbf{\Phi}' + \tau_\epsilon^{-1}\mathbf{I}_N\right). \tag{A.130}$$

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

$$\boldsymbol{\lambda} = \tau_m\mathbf{\Phi}\boldsymbol{\tau}^{-1}(\boldsymbol{\xi}^{-1}\boldsymbol{\lambda})'\boldsymbol{\xi}^{-1}\boldsymbol{\lambda} - \gamma\left(\mathbf{\Phi}\boldsymbol{\tau}^{-1}\mathbf{\Phi}' + \tau_\epsilon^{-1}\mathbf{I}_N\right)\boldsymbol{\xi}^{-1}\boldsymbol{\lambda}. \tag{A.131}$$

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

$$\tau_v\mathbf{\Phi}\boldsymbol{\tau}^{-1} = -\gamma\left(\mathbf{\Phi}\boldsymbol{\tau}^{-1}\mathbf{\Phi}' + \tau_\epsilon^{-1}\mathbf{I}_N\right)\boldsymbol{\xi}^{-1}\boldsymbol{\lambda}, \tag{A.132}$$

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

$$\boldsymbol{\xi}^{-1}\boldsymbol{\lambda} = -\frac{\tau_v}{\gamma}\left(\mathbf{\Phi}\boldsymbol{\tau}^{-1}\mathbf{\Phi}' + \tau_\epsilon^{-1}\mathbf{I}_N\right)^{-1}\mathbf{\Phi}\boldsymbol{\tau}^{-1}. \tag{A.133}$$

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

$$\boldsymbol{\tau}^{-1}\mathbf{\Phi}'\left(\mathbf{\Phi}\boldsymbol{\tau}^{-1}\mathbf{\Phi}' + \tau_\epsilon^{-1}\mathbf{I}_N\right)^{-1}\mathbf{\Phi}\boldsymbol{\tau}^{-1} = \boldsymbol{\tau}^{-1} - (\boldsymbol{\tau} + \tau_\epsilon\mathbf{\Phi}'\mathbf{\Phi})^{-1} \tag{A.134}$$

and to conclude that

$$\boldsymbol{\tau}^{-1}\mathbf{\Phi}'\boldsymbol{\xi}^{-1}\boldsymbol{\lambda} = -\frac{\tau_v}{\gamma}\left(\boldsymbol{\tau}^{-1} - (\boldsymbol{\tau} + \tau_\epsilon\mathbf{\Phi}'\mathbf{\Phi})^{-1}\right). \tag{A.135}$$

Conjecture that $\boldsymbol{\xi}^{-1}\boldsymbol{\lambda} \equiv -\frac{1}{\sqrt{\tau_m}}\mathbf{\Phi}\boldsymbol{\tau}_P$, 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 (A.124) to obtain (A.118). Further replacing the conjecture in (A.135) produces a matrix equation for $\boldsymbol{\tau}_P$:

$$\tau^{-1}\mathbf{\Phi}'\mathbf{\Phi}\boldsymbol{\tau}_P = \sqrt{\tau_m}\frac{\tau_v}{\gamma}\left(\boldsymbol{\tau}^{-1} - (\boldsymbol{\tau} + \tau_\epsilon\mathbf{\Phi}'\mathbf{\Phi})^{-1}\right), \tag{A.136}$$

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

$$\boldsymbol{\Phi}'\boldsymbol{\Phi}\boldsymbol{\tau}_P = \sqrt{\tau_m}\frac{\tau_v}{\gamma}\left(\mathbf{I}_J - (\mathbf{I}_J + \tau_\epsilon \boldsymbol{\tau}^{-1}\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}\right). \tag{A.137}$$

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

$$\boldsymbol{\tau}_P = \sqrt{\tau_m}\frac{\tau_v J}{\gamma}\left((\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1} - (\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}\left((\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1} + \tau_\epsilon \boldsymbol{\tau}^{-1}\right)^{-1}(\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}\right), \tag{A.138}$$

and apply Woodbury matrix identity:

$$\boldsymbol{\tau}_P = \sqrt{\tau_m}\frac{\tau_v J}{\gamma}\left(\boldsymbol{\Phi}'\boldsymbol{\Phi} + \tau_\epsilon^{-1}\boldsymbol{\tau}\right)^{-1}. \tag{A.139}$$

Now let us go back to (10), use Woodbury matrix identity, and substitute (A.139) in it:

$$\boldsymbol{\tau}^{-1} = (\tau_F + \tau_v)^{-1}J^{-1}\left(\mathbf{I} - \left(\mathbf{I} + (\tau_F+\tau_v)J\boldsymbol{\tau}_P^{-1}(\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}\boldsymbol{\tau}_P^{-1}\right)^{-1}\right) \tag{A.140}$$

$$= (\tau_F + \tau_v)^{-1}J^{-1}\left(\mathbf{I} - \left(\mathbf{I} + \frac{(\tau_F+\tau_v)\gamma^2}{\tau_m\tau_v^2 J}\left(\boldsymbol{\Phi}'\boldsymbol{\Phi} + \tau_\epsilon^{-1}\boldsymbol{\tau}\right)(\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}\left(\boldsymbol{\Phi}'\boldsymbol{\Phi} + \tau_\epsilon^{-1}\boldsymbol{\tau}\right)\right)^{-1}\right), \tag{A.141}$$

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

$$\frac{1}{J}\mathrm{tr}(\boldsymbol{\Phi}'\boldsymbol{\Phi}) = N. \tag{A.142}$$

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

$$(\boldsymbol{\tau}/N)^{-1} = ((\tau_F+\tau_v)J/N)^{-1}\mathbf{I} \tag{A.143}$$

$$-\left((\tau_F+\tau_v)J/N\mathbf{I} + \frac{(\tau_F+\tau_v)^2\gamma^2}{\tau_m\tau_v^2}\left(\frac{1}{N}\boldsymbol{\Phi}'\boldsymbol{\Phi} + \tau_\epsilon^{-1}\boldsymbol{\tau}/N\right)\left(\frac{1}{N}\boldsymbol{\Phi}'\boldsymbol{\Phi}\right)^{-1}\left(\frac{1}{N}\boldsymbol{\Phi}'\boldsymbol{\Phi} + \tau_\epsilon^{-1}\boldsymbol{\tau}/N\right)\right)^{-1}. \tag{A.144}$$

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

$$(\boldsymbol{\tau}/N)^{-1} = ((\tau_F+\tau_v)J/N)^{-1}\mathbf{Q}\mathbf{Q}' \tag{A.145}$$

$$-\mathbf{Q}\left((\tau_F+\tau_v)J/N\mathbf{I} + \frac{(\tau_F+\tau_v)^2\gamma^2}{\tau_m\tau_v^2}\left(\boldsymbol{\Lambda} + \tau_\epsilon^{-1}\mathbf{Q}'\boldsymbol{\tau}/N\mathbf{Q}\right)\boldsymbol{\Lambda}^{-1}\left(\boldsymbol{\Lambda} + \tau_\epsilon^{-1}\mathbf{Q}'\boldsymbol{\tau}/N\mathbf{Q}\right)\right)^{-1}\mathbf{Q}', \tag{A.146}$$

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 (A.119):

$$\boldsymbol{\tau}_\infty^{-1} = ((\tau_F+\tau_v)\psi)^{-1}\mathbf{I} - \left((\tau_F+\tau_v)\psi\mathbf{I} + \frac{(\tau_F+\tau_v)^2\gamma^2}{\tau_m\tau_v^2}\left(\boldsymbol{\Lambda} + \tau_\epsilon^{-1}\boldsymbol{\tau}_\infty\right)\boldsymbol{\Lambda}^{-1}\left(\boldsymbol{\Lambda} + \tau_\epsilon^{-1}\boldsymbol{\tau}_\infty\right)\right)^{-1}, \tag{A.147}$$

which, using once more Woodbury identity, simplifies into:

$$\boldsymbol{\tau}_{\infty} = (\tau_F + \tau_v)\psi\mathbf{I} + \frac{\psi^2\tau_m\tau_v^2}{\gamma^2}\left(\boldsymbol{\Lambda} + \tau_{\epsilon}^{-1}\boldsymbol{\tau}_{\infty}\right)^{-1}\boldsymbol{\Lambda}\left(\boldsymbol{\Lambda} + \tau_{\epsilon}^{-1}\boldsymbol{\tau}_{\infty}\right)^{-1}. \tag{A.148}$$

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:

$$\tau_{j,\infty} = \psi\left(\tau_F + \tau_v + \frac{\lambda_j\tau_m\tau_v^2\tau_{\epsilon}^2\psi}{\gamma^2(\tau_{j,\infty} + \lambda_j\tau_{\epsilon})^2}\right), \quad j = 1, ..., J. \tag{A.149}$$

This is a cubic equation in each $\tau_{j,\infty}$ that has a unique solution, and which delivers the mapping in (A.120).

A.12.2 Proof of Proposition 10

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

$$\begin{aligned}\widehat{\boldsymbol{\Sigma}} &= \boldsymbol{\Sigma} + \frac{\gamma^2}{\tau_m}\boldsymbol{\Sigma}\boldsymbol{\Sigma} + \tau_v J\boldsymbol{\Phi}\boldsymbol{\tau}^{-1}\boldsymbol{\tau}^{-1}\boldsymbol{\Phi}' && \text{(A.150)}\\ &= \left(1 + \frac{\gamma^2}{\tau_m\tau_{\epsilon}}\right)\boldsymbol{\Sigma} + \boldsymbol{\Phi}\boldsymbol{\tau}^{-1}\left(\frac{\gamma^2}{\tau_m}\left(\boldsymbol{\Phi}'\boldsymbol{\Phi} + \tau_{\epsilon}^{-1}\boldsymbol{\tau}\right) + \tau_v J\mathbf{I}\right)\boldsymbol{\tau}^{-1}\boldsymbol{\Phi}'. && \text{(A.151)}\end{aligned}$$

Let us now rescale by N and use the eigendecomposition in (59) and the definition of limiting precision in (A.119) to rewrite this expression as:

$$N\widehat{\boldsymbol{\Sigma}} = \left(1 + \frac{\gamma^2}{\tau_m\tau_{\epsilon}}\right)N\boldsymbol{\Sigma} + \boldsymbol{\Phi}\mathbf{Q}\boldsymbol{\tau}_{\infty}^{-1}\left(\frac{\gamma^2}{\tau_m}\left(\boldsymbol{\Lambda} + \tau_{\epsilon}^{-1}\boldsymbol{\tau}_{\infty}\right) + \tau_v\psi\mathbf{I}\right)\boldsymbol{\tau}_{\infty}^{-1}\mathbf{Q}'\boldsymbol{\Phi}'. \tag{A.152}$$

We now want to use this expression to obtain an expression for δ in (A.122), which can be written as:

$$1 + \delta = \frac{\sigma_{\mathbf{M}}^2}{\widehat{\sigma}_{\mathbf{M}}^2}\frac{\frac{1}{N}\mathbf{M}'\boldsymbol{\Sigma}\widehat{\boldsymbol{\Sigma}}\mathbf{M} - \sigma_{\mathbf{M}}^2\widehat{\sigma}_{\mathbf{M}}^2}{\frac{1}{N}\mathbf{M}'\boldsymbol{\Sigma}\boldsymbol{\Sigma}\mathbf{M} - \sigma_{\mathbf{M}}^4} = \frac{N\sigma_{\mathbf{M}}^2}{N\widehat{\sigma}_{\mathbf{M}}^2}\frac{N\mathbf{M}'\boldsymbol{\Sigma}\widehat{\boldsymbol{\Sigma}}\mathbf{M} - N\sigma_{\mathbf{M}}^2N\widehat{\sigma}_{\mathbf{M}}^2}{N\mathbf{M}'\boldsymbol{\Sigma}\boldsymbol{\Sigma}\mathbf{M} - N^2\sigma_{\mathbf{M}}^4}. \tag{A.153}$$

We first define the following vector:

$$\mathbf{Z} \equiv \mathbf{Q}'\boldsymbol{\Phi}'\mathbf{M}. \tag{A.154}$$

Denoting its $j-$th element by z_j, and using (A.127) and (A.152) we can write $\sigma_{\mathbf{M}}^2$ and $\widehat{\sigma}_{\mathbf{M}}^2$ as:

$$\begin{aligned}N\sigma_{\mathbf{M}}^2 &= \tau_{\epsilon}^{-1} + \sum_{j=1}^{J} z_j^2\tau_{\infty}(\lambda_j)^{-1}, && \text{(A.155)}\\ N\widehat{\sigma}_{\mathbf{M}}^2 &= \left(1 + \frac{\gamma^2}{\tau_m\tau_{\epsilon}}\right)N\sigma_{\mathbf{M}}^2 + \sum_{j=1}^{J} z_j^2\tau_{\infty}(\lambda_j)^{-2}\left(\frac{\gamma^2}{\tau_m}(\lambda_j + \tau_{\epsilon}^{-1}\tau_{\infty}(\lambda_j)) + \tau_v\psi\right), && \text{(A.156)}\end{aligned}$$

where the function $\tau_\infty(\cdot)$ is defined Lemma 2. After rearranging we can further write:

$$N\mathbf{M}'\mathbf{\Sigma}\mathbf{\Sigma}\mathbf{M} = \tau_\epsilon^{-1} N\sigma_{\mathbf{M}}^2 + \sum_{j=1}^{J} z_j^2 \tau_\infty(\lambda_j)^{-1}\left(\lambda_j \tau_\infty(\lambda_j)^{-1} + \tau_\epsilon^{-1}\right) \tag{A.157}$$

and thus:

$$N\mathbf{M}'\mathbf{\Sigma}\mathbf{\Sigma}\mathbf{M} - N^2\sigma_{\mathbf{M}}^4 = \sum_{j=1}^{J} z_j^2 \tau_\infty(\lambda_j)^{-1}\left(\tau_\infty(\lambda_j)^{-1}\lambda_j - \sum_{k=1}^{J} z_k^2 \tau_\infty(\lambda_k)^{-1}\right). \tag{A.158}$$

Similarly,

$$N\mathbf{M}'\mathbf{\Sigma}\widehat{\mathbf{\Sigma}}\mathbf{M} = \left(1 + \frac{\gamma^2}{\tau_m \tau_\epsilon}\right)\tau_\epsilon^{-1} N\sigma_{\mathbf{M}}^2 \tag{A.159}$$

$$+ \sum_{j=1}^{J} z_j^2 \tau_\infty(\lambda_j)^{-2}\left(\tau_\epsilon^{-1}\tau_\infty(\lambda_j) + \lambda_j\right)\left(1 + \frac{2\gamma^2}{\tau_m\tau_\epsilon} + \tau_\infty(\lambda_j)^{-1}\left(\frac{\gamma^2}{\tau_m}\lambda_j + \tau_v\psi\right)\right), \tag{A.160}$$

and thus

$$N\mathbf{M}'\mathbf{\Sigma}\widehat{\mathbf{\Sigma}}\mathbf{M} - N\sigma_{\mathbf{M}}^2 N\widehat{\sigma}_{\mathbf{M}}^2 = \tag{A.161}$$

$$\sum_{j=1}^{J} z_j^2 \tau_\infty(\lambda_j)^{-2}\left(\frac{\gamma^2}{\tau_m}\lambda_j + \tau_v\psi + \left(1 + \frac{2\gamma^2}{\tau_m\tau_\epsilon}\right)\tau_\infty(\lambda_j)\right)\left(\lambda_j\tau_\infty(\lambda_j)^{-1} - \sum_{k=1}^{J} z_k^2\tau_\infty(\lambda_k)^{-1}\right). \tag{A.162}$$

To obtain more transparent expressions we now use the approximation in (A.121) and Assumption 1. This assumption implies that $\mathbf{X}\mathbf{X}'/N \to \mathbf{I}$ and thus

$$\frac{1}{NJ}\mathrm{tr}(\mathbf{\Phi}'\mathbf{\Phi}) = \frac{1}{NJ}\mathrm{tr}(\mathbf{T}^{1/2}\mathbf{X}\mathbf{X}'\mathbf{T}^{1/2}) \to \frac{1}{J}\mathrm{tr}(\mathbf{T}) = 1, \tag{A.163}$$

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

$$\lim_{J\to\infty}\sum_{j=1}^{J} z_j^2 f(\lambda_j), \tag{A.164}$$

for an arbitrary, bounded 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, as per Silverstein (1989), is to take an arbitrary 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 $\mathbf{A}_1 \equiv \mathbf{\Phi}'\mathbf{\Phi}/N = \mathbf{T}^{1/2}\mathbf{X}\mathbf{X}'\mathbf{T}^{1/2}$ and its companion $\mathbf{A}_2 \equiv \mathbf{\Phi}\mathbf{\Phi}'/N = \mathbf{X}'\mathbf{T}\mathbf{X}/N$ share the same non-zero 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 matrix $N^{-1/2}\mathbf{\Phi'}$ as:

$$\frac{1}{N^{1/2}}\mathbf{\Phi'} = \sum_{j=1}^{J} \sqrt{\lambda_j}\mathbf{q}_j\mathbf{v}'_j, \tag{A.165}$$

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:

$$\frac{1}{N^{1/2}}\mathbf{Q'}\mathbf{\Phi'} = \sum_{j=1}^{J} \sqrt{\lambda_j}\mathbf{e}_j\mathbf{v}'_j, \tag{A.166}$$

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:

$$\mathbf{Z} = \sum_{j=1}^{J} \sqrt{\lambda_j}\mathbf{e}_j\mathbf{v}'_j\mathbf{x}, \tag{A.167}$$

with $j-$th entry:

$$z_j = \sqrt{\lambda_j}\mathbf{v}'_j\mathbf{x}. \tag{A.168}$$

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:

$$\sum_{j=1}^{J} z_j^2 f(\lambda_j) = \sum_{j=1}^{J} \lambda_j(\mathbf{v}'_j\mathbf{x})^2 f(\lambda_j) = \sum_{j=1}^{N} \lambda_j\mathbf{1}_{j\leq J}(\mathbf{v}'_j\mathbf{x})^2 f(\lambda_j) \tag{A.169}$$

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

$$\sum_{j=1}^{J} z_j^2 f(\lambda_j) \to \frac{1}{N}\sum_{j=1}^{N} \lambda_j\mathbf{1}_{j\leq J} f(\lambda_j) = \int f(\lambda)\lambda \mathrm{d}F^{\mathbf{A}_2}(\lambda), \tag{A.170}$$

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

$$F^{\mathbf{A}_2} = (1-\psi)\mathbf{1}_{[0,\infty)} + \psi F^{\mathbf{A}_1}. \tag{A.171}$$

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:

$$\sum_{j=1}^{J} z_j^2 f(\lambda_j) \to \psi \int f(\lambda)\lambda \mathrm{d}F^{\mathbf{A}_1}(\lambda). \tag{A.172}$$

Note that in (A.155)—(A.161) all particular forms of f to which we apply this result are bounded. This is because all eigenvalues, $\lambda > 0$, are positive, f is continuous, and, using Lemma 2, for any positive integer n and any integer $m > n$:

$$\lim_{\lambda\to\infty} \lambda^n \tau_\infty(\lambda)^{-n} = \left(\frac{3\gamma^2 \sqrt[3]{\gamma^6\tau_\epsilon^3}}{\left(\sqrt[3]{\gamma^6\tau_\epsilon^3} - \gamma^2\tau_\epsilon\right)^2} \right)^n, \tag{A.173}$$

$$\lim_{\lambda\to\infty} \lambda^n \tau_\infty(\lambda)^{-m} = 0. \tag{A.174}$$

This result in (A.172) allows us to write the distortion, δ, in terms of the two statistics σ_λ and s_λ. Using these definitions along with the approximation in (A.121), we obtain simpler expressions for:

$$N\sigma_{\mathbf{M}}^2 \approx \tau_\epsilon^{-1} + \mu_\lambda(\tau_F + \tau_v)^{-1}, \tag{A.175}$$

$$N\mathbf{M}'\mathbf{\Sigma\Sigma M} - N^2\sigma_{\mathbf{M}}^4 \approx (\tau_F + \tau_v)^{-2}\left((\sigma_\lambda^2 + \mu_\lambda^2)/\psi - \mu_\lambda^2\right), \tag{A.176}$$

and

$$N\widehat{\sigma}_{\mathbf{M}}^2 \approx \left(1 + \frac{\gamma^2}{\tau_m\tau_\epsilon}\right) N\sigma_{\mathbf{M}}^2 + (\tau_F + \tau_v)^{-2}\left(\frac{\gamma^2}{\tau_m}(\sigma_\lambda^2 + \mu_\lambda^2)/\psi + \left(\frac{\gamma^2}{\tau_m\tau_\epsilon}(\tau_F + \tau_v) + \tau_v\right)\mu_\lambda\right), \tag{A.177}$$

$$N\mathbf{M}'\mathbf{\Sigma}\widehat{\mathbf{\Sigma}}\mathbf{M} - N\sigma_{\mathbf{M}}^2 N\widehat{\sigma}_{\mathbf{M}}^2 \tag{A.178}$$

$$\approx (\tau_F + \tau_v)^{-3}\left(\begin{array}{c} \left(\frac{\gamma^2}{\tau_m}\frac{\mu_\lambda^3 + \mu_\lambda\sigma_\lambda^2}{\psi} + \left(\tau_v + \left(1 + \frac{2\gamma^2}{\tau_m\tau_\epsilon}\right)(\tau_v + \tau_F)\right)\mu_\lambda^2\right)(1/\psi - 1) \\ + \frac{\gamma^2}{\tau_m}\frac{s_\lambda + 2\mu_\lambda\sigma_\lambda^2}{\psi^2} + \left(\tau_v + \left(1 + \frac{2\gamma^2}{\tau_m\tau_\epsilon}\right)(\tau_v + \tau_F)\right)\frac{\sigma_\lambda^2}{\psi} \end{array} \right). \tag{A.179}$$

Substituting these expressions in turn in (A.153) we can characterize δ in terms of the mean, dispersion and skewness of eigenvalues:

$$\delta = \frac{\gamma^2\tau_\epsilon\left(\sigma_\lambda^2\tau_\epsilon(\mu_\lambda - \sigma_\lambda)(\mu_\lambda + \sigma_\lambda) + \mu_\lambda s_\lambda\tau_\epsilon + \tau_0\left(\mu_\lambda^3 + 3\mu_\lambda\sigma_\lambda^2 + s_\lambda\right)\right) + \tau_0\psi\left(\mu_\lambda^2 + \sigma_\lambda^2\right)\left(\tau_1 - \gamma^2\mu_\lambda\tau_\epsilon\right) - \mu_\lambda^2\tau_0\psi^2\tau_1}{\left(\mu_\lambda^2(1-\psi) + \sigma_\lambda^2\right)\left(\tau_m\tau_\epsilon\psi\left(\mu_\lambda\tau_\epsilon(\tau_F + 2\tau_v) + \tau_0^2\right) + \gamma^2\left(\tau_\epsilon^2\left(\mu_\lambda^2 + \sigma_\lambda^2\right) + \tau_0\psi(2\mu_\lambda\tau_\epsilon + \tau_0)\right)\right)}, \tag{A.180}$$

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

$$\Delta = \mu_\lambda^4 + \frac{\tau_0^2\left(\gamma^2\mu_\lambda\tau_\epsilon(\psi - 3) - \psi\tau_1\right)^2}{\gamma^4\tau_\epsilon^4} + \frac{2\mu_\lambda^2\tau_0\left(\gamma^2\mu_\lambda\tau_\epsilon(5 - 3\psi) + (3 - 2\psi)\psi\tau_1\right)}{\gamma^2\tau_\epsilon^2}. \tag{A.181}$$

So whether or not the SML looks steeper to the econometrician depends only on the sign of the numerator. Assume that the condition in (60) is satisfied. Then, if either the distribution of eigenvalues is positively skewed (or exhibits little negative skewness):

$$s_\lambda > \frac{\mu_\lambda^2\tau_0(\psi - 1)\left(\gamma^2\mu_\lambda\tau_\epsilon + \psi\tau_1\right)}{\gamma^2\tau_\epsilon(\mu_\lambda\tau_\epsilon + \tau_0)}(< 0), \tag{A.182}$$

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

$$-\frac{\tau_\epsilon}{4(\tau_0 + \mu_\lambda\tau_\epsilon)}\Delta \le s_\lambda < \frac{\mu_\lambda^2\tau_0(\psi - 1)\left(\gamma^2\mu_\lambda\tau_\epsilon + \psi\tau_1\right)}{\gamma^2\tau_\epsilon(\mu_\lambda\tau_\epsilon + \tau_0)}, \tag{A.183}$$

and if, further, eigenvalues are not too concentrated:

$$\sigma_\lambda^2 > \frac{1}{2}\left(-\sqrt{\Delta + \frac{4s_\lambda(\mu_\lambda\tau_\epsilon + \tau_0)}{\tau_\epsilon}} + \mu_\lambda^2 + \frac{\mu_\lambda\tau_0(3-\psi)}{\tau_\epsilon} + \frac{\tau_0\psi\tau_1}{\gamma^2\tau_\epsilon^2}\right), \tag{A.184}$$

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:

$$\mathrm{Cov}(\widehat{\boldsymbol{\beta}}, \mathbb{E}[\widetilde{\mathbf{R}}^e]) = \mathrm{Cov}(\widehat{\boldsymbol{\beta}}, \boldsymbol{\beta})\mathbb{E}[\widetilde{R}^e_{\mathbf{M}}], \tag{A.185}$$

where the covariance in the second inequality is given by (A.177). If skewness is strictly negative:

$$s_\lambda < \frac{\mu_\lambda^2(\psi-1)\left(\gamma^2\mu_\lambda\tau_\epsilon + \tau_F\tau_\epsilon\tau_m\psi + 2\psi\tau_1\right)}{\gamma^2\tau_\epsilon}, \tag{A.186}$$

and if eigenvalues are sufficiently concentrated:

$$\sigma_\lambda^2 < \frac{\mu_\lambda^2(\psi-1)\left(\gamma^2\mu_\lambda\tau_\epsilon + \tau_F\tau_\epsilon\tau_m\psi + 2\psi\tau_1\right) - \gamma^2 s_\lambda\tau_\epsilon}{\gamma^2\mu_\lambda\tau_\epsilon(3-\psi) + \psi(\tau_F\tau_\epsilon\tau_m + 2\tau_1)}, \tag{A.187}$$

then the covariance in the second inequality of (A.177) is negative (the SML is downward-sloping).

A.13 Dollar returns vs rates of returns

In our model with CARA preferences and normally distributed cash flows, returns are expressed in dollars per share. While this allows us to derive analytical expressions for all our results, it also raises two concerns. First, does our main result—that the CAPM holds for investors but fails for the empiricist—survive if we use rates of returns instead of dollar returns? Second, dollar returns are not entirely consistent with the data analysis of Section 5, where we use rates of returns, as in the empirical literature. In this appendix we deal with these two concerns.

A.13.1 True versus empiricist's CAPM with rates of return

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

$$\mathbb{E}[\widetilde{\mathbf{R}}^e] = \gamma\mathbf{\Sigma}\mathbf{M}, \tag{A.188}$$

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

$$\mathbb{E}[\widetilde{\mathbf{D}}] - \mathbb{E}[\widetilde{\mathbf{P}}] = \gamma\mathbf{\Sigma}\mathbf{M}. \tag{A.189}$$

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

$$\boldsymbol{\mu}^r = \mathrm{diag}(\mathbf{P})^{-1}(\mathbb{E}[\widetilde{\mathbf{D}}] - \mathbf{P}) = \mathrm{diag}(\mathbf{P})^{-1}\gamma\mathbf{\Sigma}\mathbf{M}. \tag{A.190}$$

(N.B. The constant parameter D plays no role in our results when we work with dollar returns, but when working with rates of returns, a sufficiently large parameter D ensures that the probability of obtaining negative prices remains negligible. The parameter D is also necessary for a realistic calibration of the model—see below.)

The market portfolio weights are:

$$\mathbf{w} = \frac{\text{diag}(\mathbf{P})\mathbf{M}}{\mathbf{M}'\mathbf{P}}, \tag{A.191}$$

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

$$\mathbf{w}'\boldsymbol{\mu}^r = \frac{\mathbf{M}'\text{diag}(\mathbf{P})}{\mathbf{M}'\mathbf{P}}\text{diag}(\mathbf{P})^{-1}\gamma\boldsymbol{\Sigma}\mathbf{M} = \frac{\gamma}{\mathbf{M}'\mathbf{P}}\mathbf{M}'\boldsymbol{\Sigma}\mathbf{M}. \tag{A.192}$$

Dividing (A.190) by (A.192) yields

$$\frac{\boldsymbol{\mu}^r}{\mathbf{w}'\boldsymbol{\mu}^r} = \mathbf{M}'\mathbf{P}\text{diag}(\mathbf{P})^{-1}\boldsymbol{\beta}. \tag{A.193}$$

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 average 1:

$$\mathbf{w}'\mathbf{M}'\mathbf{P}\text{diag}(\mathbf{P})^{-1}\boldsymbol{\beta} = \frac{\mathbf{M}'\text{diag}(\mathbf{P})}{\mathbf{M}'\mathbf{P}}\mathbf{M}'\mathbf{P}\text{diag}(\mathbf{P})^{-1}\boldsymbol{\beta} = \mathbf{M}'\boldsymbol{\beta} = 1. \tag{A.194}$$

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

$$\tilde{\mathbf{r}}^e \equiv \text{diag}(\widetilde{\mathbf{P}})^{-1}(\widetilde{\mathbf{D}} - \widetilde{\mathbf{P}}), \tag{A.195}$$

and thus are not normally distributed. Therefore, we resort to simulations in order to obtain the CAPM as measured by the empiricist.

Calibration We calibrate an economy with 50 risky assets such that: (i) the annual average rate of excess return for the market portfolio is $\sim 6\%$; (ii) the range of the betas of the 50 assets, computed with rates of returns, is from ~ 0.5 to ~ 1.5; and (iii) The values $\mathcal{C}^2$ and $\mathcal{D}^2$ computed with rates of return are both ~ 0.05, consistent with our empirical findings (page 23). The resulting calibration, which will be used for all our simulations below, is $N = 50$, $D = 5000$, $\gamma = 40$, $\tau_F = 0.05$, $\tau_\epsilon = 1$, $\tau_v = 1$, $\tau_m = 22,500$, $\mathbf{M} = \mathbf{1}/50$, and $\boldsymbol{\Phi} = \mathbf{z}/\text{norm}(\mathbf{z})$ where $\mathbf{z}$ is a $N \times 1$ vector normally distributed with mean 1 and variance 1. (We have experimented with an extensive range of calibrations, and consistently obtained similar results.)

Figure A1 depicts the first simulation results. Each panel takes the above calibration as a starting point and varies γ, τ_m, or τ_ϵ. The three panels compare the theoretical distortion (Proposition 5) 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 one simulation of the economy at daily frequency, consisting in generating 10^7 returns for the 50 assets. Then, we estimate the CAPM using realized rates of return and obtain δ by dividing the intercept by the slope according to (27). All the panels show that the distortion with rates of returns is consistently larger than the distortion with dollar returns.

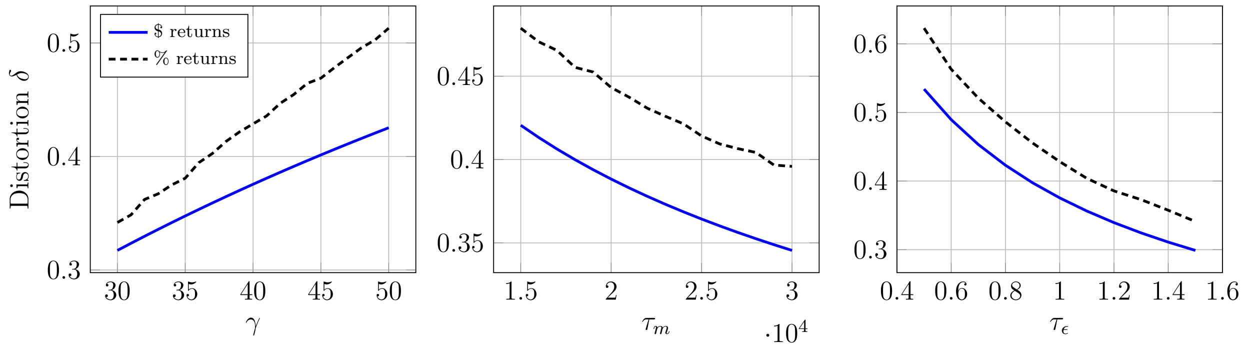

Figure A1: **Distortion with dollar returns vs. rates of return.** This figure illustrates the distortion from Proposition 5 (solid lines) and empiricist's distortion when using rates of returns instead of dollar returns (dashed lines). We plot the distortion as a function of γ, τ_m, and τ_ϵ. For each point on the dashed lines we perform one simulation of our economy at daily frequency, using the calibration below. Each simulation consists in generating 10^7 returns for the 50 assets in the economy.

A.13.2 Validity of our model predictions when using rates of returns

We now verify that the main testable implications of our model, which we derive in Section 5 using dollar returns, remain valid when based on rates of return. In a first step we plot histograms for δ, $\mathcal{C}^2$, and $\mathcal{D}^2$ computed from simulated data and compare them with their true values from (32), (30), and (31). In the analysis that follows, when we say *"one simulation"* we mean that we have generated 10^5 random returns for all the 50 assets in the economy using the calibration from page 70, then used the resulting dataset to compute all of our variables of interest with rates of return. When we say *"x simulations"* we mean that we have done this exercise x times.

Figure A2 shows one typical simulation. Panel (a) presents our main result obtained in closed form with dollar returns (as in Figure 2 in the main text). Panel (b) shows the same result obtained with rates of return from simulated data (returns are annualized). In both panels, the CAPM holds for investors, but it appears flatter to the empiricist.

Figure A3 plots histograms from 100 simulations of the model. In each panel, the vertical dashed lines show the true values for δ, $\mathcal{C}^2$, and $\mathcal{D}^2$ obtained using (32), (30), and (31). The histograms confirm that δ, $\mathcal{C}^2$, and $\mathcal{D}^2$ computed from simulated data and using rates of return are close to the numbers obtained in closed form. Although the $\mathcal{C}^2$ and $\mathcal{D}^2$ that result from rates of return are on average lower than the model-implied $\mathcal{C}^2$ and $\mathcal{D}^2$, their magnitudes remain very close to the true theoretical values. We have noticed that this pattern—$\mathcal{C}^2$ and $\mathcal{D}^2$ from rates of return being lower on average than their theoretical values—is not consistent across simulations, but can revert for different calibrations of the model. In all cases, however, $\mathcal{C}^2$ and $\mathcal{D}^2$ from rates of return continue to stay close to their theoretical counterparts.

Next, we check if Equations (46) and (47), which form the basis of our empirical work (Tables 4 and 5), are verified when using rates of return instead of dollar returns. For (46) we proceed as follows. Using data from one simulation, we compute the rates of return for all assets and for the value-weighted market portfolio. With this dataset at hand, we compute $\widehat{\boldsymbol{\beta}}$ (empiricist's betas). We then compute consensus expected rates of return for the assets (using (13) and dividing by the current prices of assets) and for the market (using the market weights that result from simulations). This allows us to compute $\boldsymbol{\beta}^{\mathcal{C}}$ as in (39). Finally, we simulate one agent's time series of private signals, which allows us to compute $\boldsymbol{\beta}^{\mathcal{D}}$ as in (39) using rates of return. (NB: Simulating the private

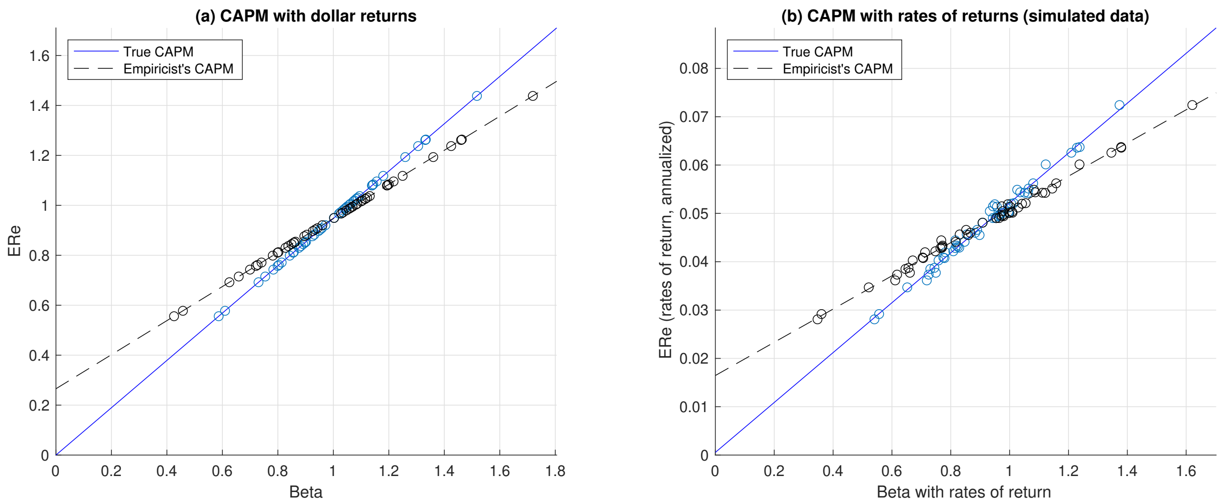

Figure A2: **The CAPM with dollar returns vs. rates of return.** The left panel plots our main result with dollar returns: the CAPM holds for investors, but it appears flat to the empiricist. The right panel confirms this result with an economy simulated in which 10^5 returns are generated for each of the 50 assets.

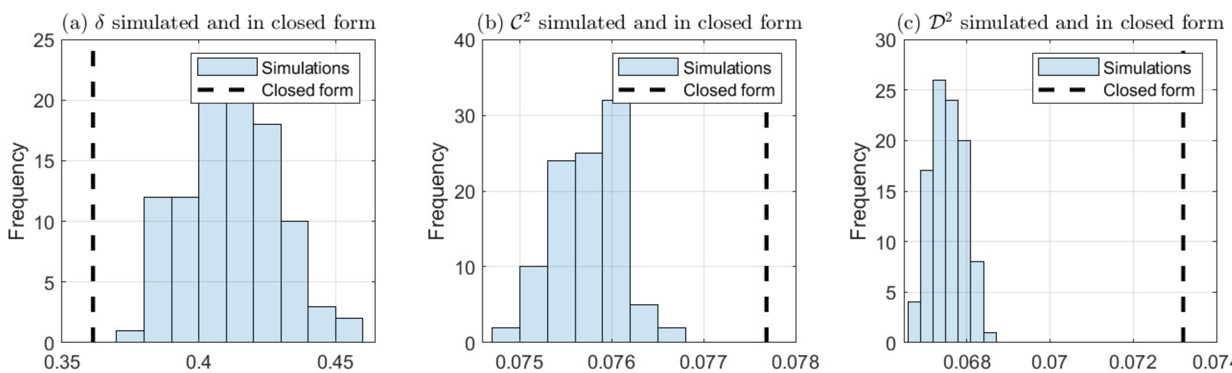

Figure A3: **δ, $\mathcal{C}^2$, and $\mathcal{D}^2$: dollar returns vs. rates of return.** Each panel shows a histogram of the parameter δ (left panel), $\mathcal{C}^2$ (center panel), and $\mathcal{D}^2$ (right panel), obtained from 100 simulations of the model, and computed using rates of return. The vertical dashed lines show the true values of δ, $\mathcal{C}^2$, and $\mathcal{D}^2$ obtained using (32), (30), and (31).

information of only one agent is sufficient for our purpose, thanks to Assumption C, page 5). We now have all the necessary information to compute both the left-hand side and the right-hand side of (46). This yields two vectors of dimension 50×1, which should be identical if the relationship holds for rates of return: a regression of any of these two vectors on the other should yield an intercept of 0, a slope of 1, and a R^2 of 1.

Figure A4 presents the intercept (left), slope (center), and R^2 (right) that result from 100 simulations. The histograms show that all numbers are very close to what they should be, allowing us to conclude that (46) is indeed verified based on rates of return. In particular we notice that the R^2 from simulations is almost 1.

Next, we check Equation (47) using the same procedure as for (46): based on simulated data and using rates of return, we compute both the left-hand and right-hand side of (47) and obtain

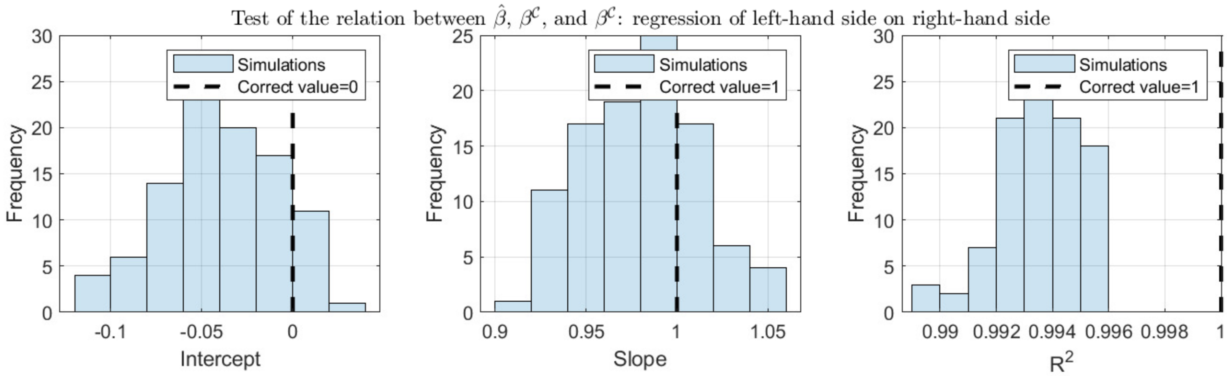

Figure A4: **Verification of Equation (46) with rates of return.** For each simulation of the model, we use rates of returns to compute the left-hand and right-hand side of (46). We then regress one vector on the other. If (46) holds with rates of returns, the regression should yield an intercept of 0, a slope of 1, and a R^2 of 1. The figure shows histograms of the intercept (left), slope (center), and R^2 (right), from 100 simulations of the model.

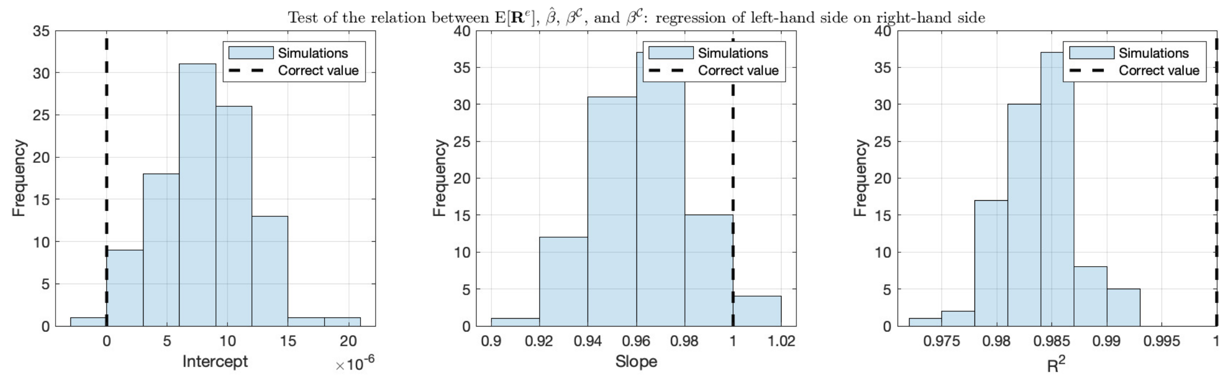

Figure A5: **Verification of Equation (47) with rates of return.** For each simulation of the model, we use rates of returns to compute the left-hand and right-hand side of (47). We then regress one vector on the other. If (47) holds with rates of returns, the regression should yield an intercept of 0, a slope of 1, and a R^2 of 1. The figure shows histograms of the intercept (left), slope (center), and R^2 (right), from 100 simulations of the model.

two 50×1 vectors, then regress one vector on the other. If (47) holds with rates of return, the regression should yield an intercept of 0, a slope of 1, and a R^2 of 1. Figure A5 shows that indeed the intercept of these regressions is very close to 0, and the slope and R^2 are both very close to 1. Thus, (47) holds with rates of return, justifying the use of rates of return in our empirical analysis.

Finally, we also check how closely related are the betas computed based on rates of return to our theoretical betas based on dollar returns. In Figure A6, we plot histograms for the correlation coefficients between $\widehat{\beta}$ and $\widehat{\beta}_r$ (left); $\beta^{\mathcal{C}}$ and $\beta^{\mathcal{C}}_r$ (center); and $\beta^{\mathcal{D}}$ and $\beta^{\mathcal{D}}_r$ (right). (We use the subscript r to denote "rates of return.") The three plots confirm that betas computed based on rates of return are extremely close to their theoretical counterparts.

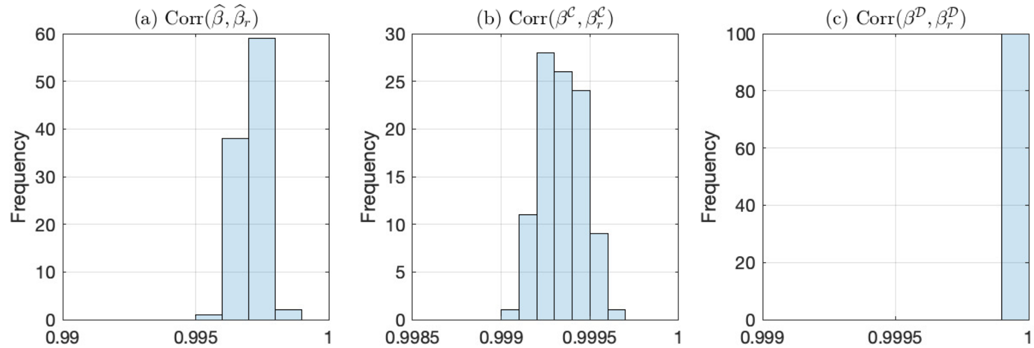

Figure A6: **Correlations between theoretical betas and betas computed with rates of return.** For each simulation of the model, we compute the correlations between $\widehat{\beta}$ and $\widehat{\beta}_r$ (left); $\beta^{\mathcal{C}}$ and $\beta^{\mathcal{C}}_r$ (center); and $\beta^{\mathcal{D}}$ and $\beta^{\mathcal{D}}_r$ (right). We use the subscript r to denote "rates of return." The figure shows histograms resulting from 100 simulations of the model.

A.13.3 First-order approximation of rates of return

In this section we introduce a first-order approximation of rates of returns, which allows us to obtain many expressions relevant to our analysis in closed-form.[35] We examine the two main theoretical relations (46) and (47) used for empirical tests based on this approximation. We show that the effect of using first-order approximations of rates of returns, as opposed to dollar returns, leave the two relation in (46) and (47) virtually unaffected.

For each asset n, we introduce the following first-order approximation of simple rates of returns, defined as $\widetilde{R}^e_n = \widetilde{D}_n/\widetilde{P}_n - 1$, in the price $\widetilde{P}_n$ around its unconditional average, $\mathbb{E}[\widetilde{P}_n]$:

$$\begin{aligned}\widetilde{R}^e_n &= \widetilde{D}_n(2\,\mathbb{E}[\widetilde{P}_n] - \widetilde{P}_n)/\,\mathbb{E}[\widetilde{P}_n]^2 - 1 + O((\widetilde{P}_n - \mathbb{E}[\widetilde{P}_n])^2) &\text{(A.196)}\\ &\approx \widetilde{D}_n(2\,\mathbb{E}[\widetilde{P}_n] - \widetilde{P}_n)/\,\mathbb{E}[\widetilde{P}_n]^2 - 1. &\text{(A.197)}\end{aligned}$$

Thereafter, whenever we write "≈" we mean that the relevant expressions is computed based on returns ignoring terms of order higher than one. Based on this approximation we want to compute $\mathcal{C}^2$, $\mathcal{D}^2$ and δ, along with $\beta^{\mathcal{C}}_n$, $\beta^{\mathcal{D}}_n$, and $\widehat{\beta}_n$.

In a first step we consider a first-order approximation of market returns. In particular, using market weights as defined in (A.191) we write $\widetilde{R}^e_{\mathbf{M}} = \mathbf{w}'\mathrm{diag}(\widetilde{\mathbf{P}})^{-1}\widetilde{\mathbf{D}} - 1 = \mathbf{M}'\widetilde{\mathbf{D}}/\mathbf{M}'\widetilde{\mathbf{P}} - 1$, which satisfies the first-order approximation:

$$\widetilde{R}^e_{\mathbf{M}} \approx \mathbf{M}'\widetilde{\mathbf{D}}(2\mathbf{M}'\,\mathbb{E}[\widetilde{\mathbf{P}}] - \mathbf{M}'\widetilde{\mathbf{P}})/\,\mathbb{E}[\mathbf{M}'\widetilde{\mathbf{P}}]^2 - 1, \tag{A.198}$$

where, using the result of Proposition 2, we have:

$$\mathbb{E}[\mathbf{M}'\widetilde{\mathbf{P}}] = D - \gamma\left(\tau^{-1}\bar{\Phi}^2 + \tau_\epsilon^{-1}/N\right). \tag{A.199}$$

We then re-express approximated market returns, consensus beliefs about them, along with deviations from consensus beliefs in quadratic forms (in some Gaussian vector to be specified). Specifically, using the result of Proposition 2 we can write:

$$\widetilde{R}^e_{\mathbf{M}} + 1 \approx \mathbf{Z}_1'\mathbf{\Lambda}_1\mathbf{Z}_1/\,\mathbb{E}[\mathbf{M}'\widetilde{\mathbf{P}}]^2, \tag{A.200}$$

[35] We thank an anonymous referee for suggesting this approach.

where the $5 \times 5-$matrix $\mathbf{\Lambda}_1$ satisfies:

$$\mathbf{\Lambda}_1 \equiv \begin{pmatrix} D\left(\frac{\gamma\bar{\Phi}^2}{\tau} - D + \frac{\gamma}{N\tau_\epsilon}\right) & \frac{1}{2}\left(\frac{\gamma\bar{\Phi}^3}{\tau} - \frac{D\tau_F\bar{\Phi}}{\tau} + \frac{\gamma\bar{\Phi}}{N\tau_\epsilon}\right) & \frac{D\bar{\Phi}\left(\gamma+\sqrt{\tau_m\tau_P}\right)}{2\tau} & \frac{D\gamma}{2\tau_\epsilon} & \frac{1}{2}\left(\frac{\gamma\bar{\Phi}^2}{\tau} - D + \frac{\gamma}{N\tau_\epsilon}\right) \\ \frac{1}{2}\left(\frac{\gamma\bar{\Phi}^3}{\tau} - \frac{D\tau_F\bar{\Phi}}{\tau} + \frac{\gamma\bar{\Phi}}{N\tau_\epsilon}\right) & \frac{\bar{\Phi}^2(\tau-\tau_F)}{\tau} & \frac{\bar{\Phi}^2\left(\gamma+\sqrt{\tau_m\tau_P}\right)}{2\tau} & \frac{\gamma\bar{\Phi}}{2\tau_\epsilon} & \frac{1}{2}\left(\bar{\Phi} - \frac{\bar{\Phi}\tau_F}{\tau}\right) \\ \frac{D\bar{\Phi}\left(\gamma+\sqrt{\tau_m\tau_P}\right)}{2\tau} & \frac{\bar{\Phi}^2\left(\gamma+\sqrt{\tau_m\tau_P}\right)}{2\tau} & 0 & 0 & \frac{\bar{\Phi}\left(\gamma+\sqrt{\tau_m\tau_P}\right)}{2\tau} \\ \frac{D\gamma}{2\tau_\epsilon} & \frac{\gamma\bar{\Phi}}{2\tau_\epsilon} & 0 & 0 & \frac{\gamma}{2\tau_\epsilon} \\ \frac{1}{2}\left(\frac{\gamma\bar{\Phi}^2}{\tau} - D + \frac{\gamma}{N\tau_\epsilon}\right) & \frac{1}{2}\left(\bar{\Phi} - \frac{\bar{\Phi}\tau_F}{\tau}\right) & \frac{\bar{\Phi}\left(\gamma+\sqrt{\tau_m\tau_P}\right)}{2\tau} & \frac{\gamma}{2\tau_\epsilon} & 0 \end{pmatrix}, \tag{A.201}$$

and where

$$\mathbf{Z}_1 \equiv \begin{pmatrix} 1 & \widetilde{\mathbf{F}} & \mathbf{\Phi}'\widetilde{\mathbf{m}} & \mathbf{M}'\widetilde{\mathbf{m}} & \mathbf{M}'\widetilde{\boldsymbol{\epsilon}} \end{pmatrix}' \tag{A.202}$$

is a Gaussian vector with mean $\boldsymbol{\mu}_1 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \end{pmatrix}'$ and covariance matrix

$$\mathbf{\Sigma}_1 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{\tau_F} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\tau_m} & \frac{\bar{\Phi}}{\tau_m} & 0 \\ 0 & 0 & \frac{\bar{\Phi}}{\tau_m} & \frac{1}{N\tau_m} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{N\tau_\epsilon} \\ 0 & 0 & \frac{\Phi_n}{\tau_m} & \frac{1}{N\tau_m} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{N\tau_\epsilon} \end{pmatrix}. \tag{A.203}$$

Furthermore, using (A.19) we can write consensus beliefs about approximated market returns as:

$$\overline{\mathbb{E}}[\widetilde{R}^e_{\mathbf{M}}] + 1 \approx \mathbf{M}'\overline{\mathbb{E}}[\widetilde{\mathbf{D}}](2\mathbf{M}'\,\mathbb{E}[\widetilde{\mathbf{P}}] - \mathbf{M}'\widetilde{\mathbf{P}})/\,\mathbb{E}[\mathbf{M}'\widetilde{\mathbf{P}}]^2 \tag{A.204}$$

$$= \mathbf{Z}_2'\mathbf{\Lambda}_2\mathbf{Z}_2/\,\mathbb{E}[\mathbf{M}'\widetilde{\mathbf{P}}]^2, \tag{A.205}$$

where the $4 \times 4-$matrix $\mathbf{\Lambda}_2$ satisfies:

$$\mathbf{\Lambda_2} \equiv \begin{pmatrix} D\gamma\left(\frac{\bar{\Phi}^2}{\tau} + \frac{1}{N\tau_\epsilon}\right) - D^2 & \frac{\gamma\bar{\Phi}(\tau-\tau_F)\left(\frac{\bar{\Phi}^2}{\tau}+\frac{1}{N\tau_\epsilon}\right)}{2\tau} & \frac{\gamma\bar{\Phi}\left(N\sqrt{\tau_m\tau_P}\tau_\epsilon\bar{\Phi}^2+\tau\left(DN\tau_\epsilon+\sqrt{\tau_m\tau_P}\right)\right)}{2N\tau^2\tau_\epsilon} & \frac{D\gamma}{2\tau_\epsilon} \\ \frac{\gamma\bar{\Phi}(\tau-\tau_F)\left(\frac{\bar{\Phi}^2}{\tau}+\frac{1}{N\tau_\epsilon}\right)}{2\tau} & \frac{\bar{\Phi}^2(\tau-\tau_F)^2}{\tau^2} & \frac{\bar{\Phi}^2(\tau-\tau_F)\left(\gamma+2\sqrt{\tau_m\tau_P}\right)}{2\tau^2} & \frac{\gamma\bar{\Phi}(\tau-\tau_F)}{2\tau\tau_\epsilon} \\ \frac{\gamma\bar{\Phi}\left(N\sqrt{\tau_m\tau_P}\tau_\epsilon\bar{\Phi}^2+\tau\left(DN\tau_\epsilon+\sqrt{\tau_m\tau_P}\right)\right)}{2N\tau^2\tau_\epsilon} & \frac{\bar{\Phi}^2(\tau-\tau_F)\left(\gamma+2\sqrt{\tau_m\tau_P}\right)}{2\tau^2} & \frac{\bar{\Phi}^2\left(\sqrt{\tau_m\tau_P}\gamma+\tau_m\tau_P\right)}{\tau^2} & \frac{\gamma\bar{\Phi}\sqrt{\tau_m\tau_P}}{2\tau\tau_\epsilon} \\ \frac{D\gamma}{2\tau_\epsilon} & \frac{\gamma\bar{\Phi}(\tau-\tau_F)}{2\tau\tau_\epsilon} & \frac{\gamma\bar{\Phi}\sqrt{\tau_m\tau_P}}{2\tau\tau_\epsilon} & 0 \end{pmatrix}, \tag{A.206}$$

and where

$$\mathbf{Z}_2 \equiv \begin{pmatrix} 1 & \widetilde{\mathbf{F}} & \mathbf{\Phi}'\widetilde{\mathbf{m}} & \mathbf{M}'\widetilde{\mathbf{m}} \end{pmatrix}' \tag{A.207}$$

is a Gaussian vector with mean $\boldsymbol{\mu}_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \end{pmatrix}'$ and covariance matrix

$$\boldsymbol{\Sigma}_2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \tau_F^{-1} & 0 & 0 \\ 0 & 0 & \tau_m^{-1} & \tau_m^{-1}\bar{\Phi} \\ 0 & 0 & \tau_m^{-1}\bar{\Phi} & \tau_m^{-1}/N \end{pmatrix}. \tag{A.208}$$

Similarly, using (A.18) we can rewrite individual deviations from consensus as:

$$\mathbb{E}^i[\widetilde{R}^e_{\mathbf{M}}] - \overline{\mathbb{E}}[\widetilde{R}^e_{\mathbf{M}}] \approx \mathbf{M}'(\mathbb{E}^i[\widetilde{\mathbf{D}}] - \overline{\mathbb{E}}[\widetilde{\mathbf{D}}])(2\mathbf{M}'\,\mathbb{E}[\widetilde{\mathbf{P}}] - \mathbf{M}'\widetilde{\mathbf{P}})/\,\mathbb{E}[\mathbf{M}'\widetilde{\mathbf{P}}]^2 \tag{A.209}$$

$$= \mathbf{Z}_3'\boldsymbol{\Lambda}_3\mathbf{Z}_3/\,\mathbb{E}[\mathbf{M}'\widetilde{\mathbf{P}}]^2, \tag{A.210}$$

where the $5\times 5-$matrix $\boldsymbol{\Lambda}_3$ satisfies:

$$\boldsymbol{\Lambda_3} \equiv \begin{pmatrix} 0 & 0 & 0 & \frac{\bar{\Phi}\tau_v\left(\gamma\left(N\tau_\epsilon\bar{\Phi}^2+\tau\right)-DN\tau\tau_\epsilon\right)}{2N\tau^2\tau_\epsilon} \\ 0 & 0 & 0 & \frac{\bar{\Phi}^2(\tau-\tau_F)\tau_v}{2\tau^2} \\ 0 & 0 & 0 & \frac{\bar{\Phi}\tau_v}{2\tau} \\ \frac{\bar{\Phi}\tau_v\left(\gamma\left(N\tau_\epsilon\bar{\Phi}^2+\tau\right)-DN\tau\tau_\epsilon\right)}{2N\tau^2\tau_\epsilon} & \frac{\bar{\Phi}^2(\tau-\tau_F)\tau_v}{2\tau^2} & \frac{\bar{\Phi}\tau_v}{2\tau} & 0 \end{pmatrix}, \tag{A.211}$$

and where

$$\mathbf{Z}_3 \equiv \begin{pmatrix} 1 & \widetilde{\mathbf{F}} & \boldsymbol{\Phi}'\widetilde{\mathbf{m}} & \mathbf{M}'\widetilde{\mathbf{m}} & \widetilde{v}^i \end{pmatrix}' \tag{A.212}$$

is a Gaussian vector with mean $\boldsymbol{\mu}_1$ and covariance matrix

$$\boldsymbol{\Sigma}_3 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & \tau_F^{-1} & 0 & 0 & 0 \\ 0 & 0 & \tau_m^{-1} & \tau_m^{-1}\bar{\Phi} & 0 \\ 0 & 0 & \tau_m^{-1}\bar{\Phi} & \tau_m^{-1}/N & 0 \\ 0 & 0 & 0 & 0 & \tau_v^{-1} \end{pmatrix}. \tag{A.213}$$

We now repeat these steps to compute consensus beliefs and individual deviations thereof for approximated returns on stock n. Using (A.197) we start by rewriting realized returns on individual assets as:

$$\widetilde{R}^e_n + 1 \approx \mathbf{Z}_4'\boldsymbol{\Lambda}_4\mathbf{Z}_4/\,\mathbb{E}[\widetilde{P}_n]^2, \tag{A.214}$$

where the $7\times 7-$matrix $\boldsymbol{\Lambda}_4$ satisfies:

$$\boldsymbol{\Lambda_4} \equiv \tag{A.215}$$

$$
\begin{pmatrix}
D\left(\frac{\gamma\bar{\Phi}^2}{\tau}-D+\frac{\gamma}{N\tau_\epsilon}\right) & \frac{1}{2}\left(\frac{\gamma}{N\tau_\epsilon}+\frac{\gamma\bar{\Phi}^2-D\tau_F}{\tau}\right)\Phi_n & \frac{D\left(\gamma+\sqrt{\tau_m\tau_P}\right)\Phi_n}{2\tau} & 0 & 0 & \frac{D\gamma}{2\tau_\epsilon} & \frac{1}{2}\left(\frac{\gamma\bar{\Phi}^2}{\tau}-D+\frac{\gamma}{N\tau_\epsilon}\right) \\
\frac{1}{2}\left(\frac{\gamma}{N\tau_\epsilon}+\frac{\gamma\bar{\Phi}^2-D\tau_F}{\tau}\right)\Phi_n & \frac{(\tau-\tau_F)\Phi_n^2}{\tau} & \frac{\left(\gamma+\sqrt{\tau_m\tau_P}\right)\Phi_n^2}{2\tau} & 0 & 0 & \frac{\gamma\Phi_n}{2\tau_\epsilon} & \frac{1}{2}\left(\Phi_n-\frac{\tau_F\Phi_n}{\tau}\right) \\
\frac{D\left(\gamma+\sqrt{\tau_m\tau_P}\right)\Phi_n}{2\tau} & \frac{\left(\gamma+\sqrt{\tau_m\tau_P}\right)\Phi_n^2}{2\tau} & 0 & 0 & 0 & 0 & \frac{\left(\gamma+\sqrt{\tau_m\tau_P}\right)\Phi_n}{2\tau} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\frac{D\gamma}{2\tau_\epsilon} & \frac{\gamma\Phi_n}{2\tau_\epsilon} & 0 & 0 & 0 & 0 & \frac{\gamma}{2\tau_\epsilon} \\
\frac{1}{2}\left(\frac{\gamma\bar{\Phi}^2}{\tau}-D+\frac{\gamma}{N\tau_\epsilon}\right) & \frac{1}{2}\left(\Phi_n-\frac{\tau_F\Phi_n}{\tau}\right) & \frac{\left(\gamma+\sqrt{\tau_m\tau_P}\right)\Phi_n}{2\tau} & 0 & 0 & \frac{\gamma}{2\tau_\epsilon} & 0
\end{pmatrix},
\tag{A.216}
$$

and where

$$
\mathbf{Z}_4 \equiv \begin{pmatrix} 1 & \widetilde{\mathbf{F}} & \mathbf{\Phi}'\widetilde{\mathbf{m}} & \mathbf{M}'\widetilde{\mathbf{m}} & \mathbf{M}'\widetilde{\boldsymbol{\epsilon}} & \widetilde{m}_n & \widetilde{\epsilon}_n \end{pmatrix}'
\tag{A.217}
$$

is a Gaussian vector with mean $\boldsymbol{\mu}_4 \equiv \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}'$ and covariance matrix

$$
\mathbf{\Sigma}_4 = \begin{pmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \frac{1}{\tau_F} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \frac{1}{\tau_m} & \frac{\bar{\Phi}}{\tau_m} & 0 & \frac{\Phi_n}{\tau_m} & 0 \\
0 & 0 & \frac{\bar{\Phi}}{\tau_m} & \frac{1}{N\tau_m} & 0 & \frac{1}{N\tau_m} & 0 \\
0 & 0 & 0 & 0 & \frac{1}{N\tau_\epsilon} & 0 & \frac{1}{N\tau_\epsilon} \\
0 & 0 & \frac{\Phi_n}{\tau_m} & \frac{1}{N\tau_m} & 0 & \frac{1}{\tau_m} & 0 \\
0 & 0 & 0 & 0 & \frac{1}{N\tau_\epsilon} & 0 & \frac{1}{\tau_\epsilon}
\end{pmatrix}.
\tag{A.218}
$$

We then rewrite consensus beliefs on individual, approximated returns as:

$$
\overline{\mathbb{E}}[\widetilde{R}_n^e] + 1 \approx \overline{\mathbb{E}}[\widetilde{D}_n](2\,\mathbb{E}[\widetilde{P}_n] - \widetilde{P}_n)/\,\mathbb{E}[\widetilde{P}_n]^2 \tag{A.219}
$$

$$
= \mathbf{Z}_5'\mathbf{\Lambda}_5\mathbf{Z}_5/\,\mathbb{E}[\widetilde{P}_n]^2, \tag{A.220}
$$

where the $5 \times 5-$matrix $\mathbf{\Lambda}_5$ satisfies:

$$
\mathbf{\Lambda_5} \equiv \tag{A.221}
$$

$$
\begin{pmatrix}
D\left(D-\gamma\left(\frac{\bar{\Phi}\Phi_n}{\tau}+\frac{1}{N\tau_\epsilon}\right)\right) & -\frac{\gamma(\tau-\tau_F)\Phi_n\left(\tau+N\bar{\Phi}\tau_\epsilon\Phi_n\right)}{2N\tau^2\tau_\epsilon} & -\frac{\gamma\Phi_n\left(\tau\left(DN\tau_\epsilon+\sqrt{\tau_m\tau_P}\right)+N\bar{\Phi}\sqrt{\tau_m\tau_P}\tau_\epsilon\Phi_n\right)}{2N\tau^2\tau_\epsilon} & 0 & -\frac{D\gamma}{2\tau_\epsilon} \\
-\frac{\gamma(\tau-\tau_F)\Phi_n\left(\tau+N\bar{\Phi}\tau_\epsilon\Phi_n\right)}{2N\tau^2\tau_\epsilon} & -\frac{(\tau-\tau_F)^2\Phi_n^2}{\tau^2} & -\frac{(\tau-\tau_F)\left(\gamma+2\sqrt{\tau_m\tau_P}\right)\Phi_n^2}{2\tau^2} & 0 & \frac{\gamma(\tau_F-\tau)\Phi_n}{2\tau\tau_\epsilon} \\
-\frac{\gamma\Phi_n\left(\tau\left(DN\tau_\epsilon+\sqrt{\tau_m\tau_P}\right)+N\bar{\Phi}\sqrt{\tau_m\tau_P}\tau_\epsilon\Phi_n\right)}{2N\tau^2\tau_\epsilon} & -\frac{(\tau-\tau_F)\left(\gamma+2\sqrt{\tau_m\tau_P}\right)\Phi_n^2}{2\tau^2} & -\frac{\left(\sqrt{\tau_m\tau_P}\gamma+\tau_m\tau_P\right)\Phi_n^2}{\tau^2} & 0 & -\frac{\gamma\sqrt{\tau_m\tau_P}\Phi_n}{2\tau\tau_\epsilon} \\
0 & 0 & 0 & 0 & 0 \\
-\frac{D\gamma}{2\tau_\epsilon} & \frac{\gamma(\tau_F-\tau)\Phi_n}{2\tau\tau_\epsilon} & -\frac{\gamma\sqrt{\tau_m\tau_P}\Phi_n}{2\tau\tau_\epsilon} & 0 & 0
\end{pmatrix},
\tag{A.222}
$$

and where

$$
\mathbf{Z}_5 \equiv \begin{pmatrix} 1 & \widetilde{\mathbf{F}} & \mathbf{\Phi}'\widetilde{\mathbf{m}} & \mathbf{M}'\widetilde{\mathbf{m}} & \widetilde{m}_n \end{pmatrix}'
\tag{A.223}
$$

is a Gaussian vector with mean $\boldsymbol{\mu}_3$ and covariance matrix

$$\boldsymbol{\Sigma}_5 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{\tau_F} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\tau_m} & \frac{\bar{\Phi}}{\tau_m} & \frac{\Phi_n}{\tau_m} \\ 0 & 0 & \frac{\bar{\Phi}}{\tau_m} & \frac{1}{N\tau_m} & \frac{1}{N\tau_m} \\ 0 & 0 & \frac{\Phi_n}{\tau_m} & \frac{1}{N\tau_m} & \frac{1}{\tau_m} \end{pmatrix}. \tag{A.224}$$

Finally, we can write deviations from consensus on individual stocks as:

$$\mathbb{E}^i[\widetilde{R}^e_n] - \overline{\mathbb{E}}[\widetilde{R}^e_n] \approx (\mathbb{E}^i[\widetilde{D}_n] - \overline{\mathbb{E}}[\widetilde{D}_n])(2\,\mathbb{E}[\widetilde{P}_n] - \widetilde{P}_n)/\,\mathbb{E}[\widetilde{P}_n]^2 \tag{A.225}$$

$$= \mathbf{Z}_6'\boldsymbol{\Lambda}_6\mathbf{Z}_6/\,\mathbb{E}[\widetilde{P}_n]^2, \tag{A.226}$$

where the $6 \times 6-$matrix $\boldsymbol{\Lambda}_6$ satisfies:

$$\boldsymbol{\Lambda_6} \equiv \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & \frac{\tau_v\Phi_n\left(DN\tau\tau_\epsilon - \gamma\left(\tau + N\bar{\Phi}\tau_\epsilon\Phi_n\right)\right)}{2N\tau^2\tau_\epsilon} \\ 0 & 0 & 0 & 0 & 0 & \frac{(\tau_F-\tau)\tau_v\Phi_n^2}{2\tau^2} \\ 0 & 0 & 0 & 0 & 0 & -\frac{\left(\gamma+\sqrt{\tau_m\tau_P}\right)\tau_v\Phi_n^2}{2\tau^2} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -\frac{\gamma\tau_v\Phi_n}{2\tau\tau_\epsilon} \\ \frac{\tau_v\Phi_n\left(DN\tau\tau_\epsilon - \gamma\left(\tau + N\bar{\Phi}\tau_\epsilon\Phi_n\right)\right)}{2N\tau^2\tau_\epsilon} & \frac{(\tau_F-\tau)\tau_v\Phi_n^2}{2\tau^2} & -\frac{\left(\gamma+\sqrt{\tau_m\tau_P}\right)\tau_v\Phi_n^2}{2\tau^2} & 0 & -\frac{\gamma\tau_v\Phi_n}{2\tau\tau_\epsilon} & 0 \end{pmatrix}, \tag{A.227}$$

and where

$$\mathbf{Z}_6 \equiv \begin{pmatrix} 1 & \widetilde{\mathbf{F}} & \boldsymbol{\Phi}'\widetilde{\mathbf{m}} & \mathbf{M}'\widetilde{\mathbf{m}} & \widetilde{m}_n & \widetilde{v}^i \end{pmatrix}' \tag{A.228}$$

is a Gaussian vector with mean $\boldsymbol{\mu}_4 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}'$ and covariance matrix

$$\boldsymbol{\Sigma}_6 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{\tau_F} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\tau_m} & \frac{\bar{\Phi}}{\tau_m} & \frac{\Phi_n}{\tau_m} & 0 \\ 0 & 0 & \frac{\bar{\Phi}}{\tau_m} & \frac{1}{N\tau_m} & \frac{1}{N\tau_m} & 0 \\ 0 & 0 & \frac{\Phi_n}{\tau_m} & \frac{1}{N\tau_m} & \frac{1}{\tau_m} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{\tau_v} \end{pmatrix}. \tag{A.229}$$

In a second step, we use the following result.

Lemma 3. *Let $\mathbf{Z}$ be a $N-$dimensional Gaussian vector with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, and let $\boldsymbol{\Lambda}_a$ and $\boldsymbol{\Lambda}_b$ be two $N \times N-$symmetric matrices. Then:*

$$\mathbb{V}[\mathbf{Z}'\boldsymbol{\Lambda}.\mathbf{Z}] = 2tr(\boldsymbol{\Lambda}.\boldsymbol{\Sigma}\boldsymbol{\Lambda}.\boldsymbol{\Sigma}) + 4\boldsymbol{\mu}'\boldsymbol{\Lambda}.\boldsymbol{\Sigma}\boldsymbol{\Lambda}.\boldsymbol{\mu} \tag{A.230}$$

and

$$Cov(\mathbf{Z}'\boldsymbol{\Lambda}_a\mathbf{Z}, \mathbf{Z}'\boldsymbol{\Lambda}_b\mathbf{Z}) = 2tr(\boldsymbol{\Lambda}_a\boldsymbol{\Sigma}\boldsymbol{\Lambda}_b\boldsymbol{\Sigma}) + 4\boldsymbol{\mu}'\boldsymbol{\Lambda}_a\boldsymbol{\Sigma}\boldsymbol{\Lambda}_b\boldsymbol{\mu}. \tag{A.231}$$

Proof. See, e.g., Rencher and Schaalje (2008). □

Using this result we can obtain analytical expressions for $\mathcal{C}^2$, $\mathcal{D}^2$, $\beta_n^{\mathcal{C}}$, $\beta_n^{\mathcal{D}}$, and $\widehat{\beta}_n$ based on approximated returns. To do so, denote by $\mathbf{X}$ some matrix and by $\widehat{\mathbf{X}}$ the matrix obtained by bordering $\mathbf{X}$ with zeroes so that it is of conformable size. We can then write:

$$\mathcal{C}^2 = \frac{2\mathrm{tr}(\mathbf{\Lambda}_2\mathbf{\Sigma}_2\mathbf{\Lambda}_2\mathbf{\Sigma}_2) + 4\boldsymbol{\mu}_2'\mathbf{\Lambda}_2\mathbf{\Sigma}_2\mathbf{\Lambda}_2\boldsymbol{\mu}_2}{2\mathrm{tr}(\mathbf{\Lambda}_1\mathbf{\Sigma}_1\mathbf{\Lambda}_1\mathbf{\Sigma}_1) + 4\boldsymbol{\mu}_1'\mathbf{\Lambda}_1\mathbf{\Sigma}_1\mathbf{\Lambda}_1\boldsymbol{\mu}_1}, \tag{A.232}$$

$$\mathcal{D}^2 = \frac{2\mathrm{tr}(\mathbf{\Lambda}_3\mathbf{\Sigma}_3\mathbf{\Lambda}_3\mathbf{\Sigma}_3) + 4\boldsymbol{\mu}_1'\mathbf{\Lambda}_3\mathbf{\Sigma}_3\mathbf{\Lambda}_3\boldsymbol{\mu}_1}{2\mathrm{tr}(\mathbf{\Lambda}_1\mathbf{\Sigma}_1\mathbf{\Lambda}_1\mathbf{\Sigma}_1) + 4\boldsymbol{\mu}_1'\mathbf{\Lambda}_1\mathbf{\Sigma}_1\mathbf{\Lambda}_1\boldsymbol{\mu}_1}, \tag{A.233}$$

$$\widehat{\beta}_n = \frac{\mathbb{E}[\mathbf{M}'\widetilde{\mathbf{P}}]^2}{\mathbb{E}[\widetilde{P}_n]^2}\frac{2\mathrm{tr}(\widehat{\mathbf{\Lambda}}_1\mathbf{\Sigma}_4\mathbf{\Lambda}_4\mathbf{\Sigma}_4) + 4\boldsymbol{\mu}_4'\widehat{\mathbf{\Lambda}}_1\mathbf{\Sigma}_4\mathbf{\Lambda}_4\boldsymbol{\mu}_4}{2\mathrm{tr}(\mathbf{\Lambda}_1\mathbf{\Sigma}_1\mathbf{\Lambda}_1\mathbf{\Sigma}_1) + 4\boldsymbol{\mu}_1'\mathbf{\Lambda}_1\mathbf{\Sigma}_1\mathbf{\Lambda}_1\boldsymbol{\mu}_1}, \tag{A.234}$$

$$\beta_n^{\mathcal{C}} = \frac{\mathbb{E}[\mathbf{M}'\widetilde{\mathbf{P}}]^2}{\mathbb{E}[\widetilde{P}_n]^2}\frac{2\mathrm{tr}(\widehat{\mathbf{\Lambda}}_2\mathbf{\Sigma}_5\mathbf{\Lambda}_5\mathbf{\Sigma}_5) + 4\boldsymbol{\mu}_3'\widehat{\mathbf{\Lambda}}_2\mathbf{\Sigma}_5\mathbf{\Lambda}_5\boldsymbol{\mu}_3}{2\mathrm{tr}(\mathbf{\Lambda}_2\mathbf{\Sigma}_2\mathbf{\Lambda}_2\mathbf{\Sigma}_2) + 4\boldsymbol{\mu}_2'\mathbf{\Lambda}_2\mathbf{\Sigma}_2\mathbf{\Lambda}_2\boldsymbol{\mu}_2}, \tag{A.235}$$

$$\beta_n^{\mathcal{D}} = \frac{\mathbb{E}[\mathbf{M}'\widetilde{\mathbf{P}}]^2}{\mathbb{E}[\widetilde{P}_n]^2}\frac{2\mathrm{tr}(\widehat{\mathbf{\Lambda}}_3\mathbf{\Sigma}_6\mathbf{\Lambda}_6\mathbf{\Sigma}_6) + 4\boldsymbol{\mu}_4'\widehat{\mathbf{\Lambda}}_3\mathbf{\Sigma}_6\mathbf{\Lambda}_6\boldsymbol{\mu}_4}{2\mathrm{tr}(\mathbf{\Lambda}_3\mathbf{\Sigma}_3\mathbf{\Lambda}_3\mathbf{\Sigma}_3) + 4\boldsymbol{\mu}_1'\mathbf{\Lambda}_3\mathbf{\Sigma}_3\mathbf{\Lambda}_3\boldsymbol{\mu}_1}. \tag{A.236}$$

Finally, we obtain δ by applying (A.122) using $\widehat{\beta}_n$ as per (A.234) and constructing true betas based on approximated rates of returns. In this construction, it is important to note that conditional betas based on rates of returns are random, unlike those based on dollar returns. Hence, the construction of true betas is based the unconditional average of the conditional covariance matrix, as appearing in (2):

$$\beta_n = \frac{\mathbb{E}[\mathrm{Var}^i[\widetilde{\mathbf{R}}^e]]\mathbf{M}}{\mathbf{M}'\,\mathbb{E}[\mathrm{Var}^i[\widetilde{\mathbf{R}}^e]]\mathbf{M}} \approx \frac{\mathbb{E}[\mathbf{M}'\widetilde{\mathbf{P}}]^2\,\mathbb{E}[\widetilde{P}_n\mathbf{M}'\widetilde{\mathbf{P}}]}{\mathbb{E}[\widetilde{P}_n]^2\,\mathbb{E}[(\mathbf{M}'\widetilde{\mathbf{P}})^2]}\underbrace{\frac{\mathbf{\Sigma}_n\mathbf{M}}{\mathbf{M}'\mathbf{\Sigma}\mathbf{M}}}_{=\beta_n \text{ in (15)}}, \tag{A.237}$$

where $\mathbf{\Sigma}_n$ denotes the $n-$th row of $\mathbf{\Sigma}$, as defined in (11), and where:

$$\begin{aligned}\mathbb{E}[\widetilde{P}_n\mathbf{M}'\widetilde{\mathbf{P}}] &= \mathbb{E}[\widetilde{P}_n]\,\mathbb{E}[\mathbf{M}'\widetilde{\mathbf{P}}] + \gamma^2\tau_m^{-1}\tau_\epsilon^{-1}/N \qquad \text{(A.238)}\\ &\quad + \bar{\Phi}\Phi_n\tau^{-1}\left(\tau_F^{-1}\tau^{-1}(\tau-\tau_F)^2 + \tau_m^{-1}(\gamma+\sqrt{\tau_m\tau_P})(\tau^{-1}(\gamma+\sqrt{\tau_m\tau_P}) + 2\gamma\tau_\epsilon^{-1})\right),\\ \mathbb{E}[(\mathbf{M}'\widetilde{\mathbf{P}})^2] &= \mathbb{E}[\mathbf{M}'\widetilde{\mathbf{P}}]^2 + \gamma^2\tau_m^{-1}\tau_\epsilon^{-1}/N \qquad \text{(A.239)}\\ &\quad + \bar{\Phi}^2\tau^{-1}\left(\tau_F^{-1}\tau^{-1}(\tau-\tau_F)^2 + \tau_m^{-1}(\gamma+\sqrt{\tau_m\tau_P})(\tau^{-1}(\gamma+\sqrt{\tau_m\tau_P}) + 2\gamma\tau_\epsilon^{-1})\right).\end{aligned}$$

Although analytic (except for δ) these expressions, after substitutions of the relevant matrices, are unintuitive. There is one limiting case, however, in which these expressions are intuitive. This limiting case corresponds to $D \to +\infty$, that is when the unconditional average of dividends is infinite. This limiting case is interesting for two reasons. First, note that the unconditional expectation of the average market price satisfies:

$$\mathbb{E}[\mathbf{M}'\widetilde{\mathbf{P}}] = D - \gamma(\tau^{-1}\bar{\Phi}^2 + \tau_\epsilon^{-1}/N). \tag{A.240}$$

Hence, unless D is sufficiently large, due to the liquidity discount the unconditional expected average market price is negative, causing the market risk premium to be either negative or excessively large. Thus, a first condition to obtain a realistic market risk premium is that D be sufficiently large. Second, a popular interpretation of the simple rate of return, $\widetilde{R}_n^e = (\widetilde{D}_n - \widetilde{P}_n)/\widetilde{P}_n$, on asset n is as a first-order approximation of its log-return, $\log(\widetilde{D}_n/\widetilde{P}_n)$, around $\widetilde{P}_n$, which in this limiting case is well-defined given that we can ensure that $\widetilde{D}_n$ nor $\widetilde{P}_n$ never become negative. In this case, the coefficients $\mathcal{C}^2$ and $\mathcal{D}^2$ in (A.232)-(A.233) simplify to:

$$\lim_{D\to+\infty} \mathcal{C}^2 \approx \frac{\gamma^2\left(N\bar{\Phi}^2\tau_\epsilon\left(2\tau+\tau_\epsilon\right)+\tau^2\right)}{N\bar{\Phi}^2\tau_\epsilon\left(\gamma^2\left(2\tau+\tau_\epsilon\right)+\tau_m\left(\tau+\tau_v\right)\tau_\epsilon\right)+\tau^2\left(\gamma^2+\tau_m\tau_\epsilon\right)}, \tag{A.241}$$

$$\lim_{D\to+\infty} \mathcal{D}^2 \approx \frac{N\bar{\Phi}^2\tau_m\tau_v\tau_\epsilon^2}{N\bar{\Phi}^2\tau_\epsilon\left(\gamma^2\left(2\tau+\tau_\epsilon\right)+\tau_m\left(\tau+\tau_v\right)\tau_\epsilon\right)+\tau^2\left(\gamma^2+\tau_m\tau_\epsilon\right)}. \tag{A.242}$$

These two limiting expressions exactly coincide with $\mathcal{C}^2$ and $\mathcal{D}^2$ obtained based on dollar returns in (30) and (31). Hence, we expect expressions based on first-order approximations to be close to tehir counterparts based on dollar returns when D is taken to be large. Under the calibration of page A.13.1, $\mathcal{C}^2$, $\mathcal{D}^2$ and δ are virtually identical for the range of parameters considered in Figure A1 and are thus not reported here.

We now examine the relations in (46) and (47), which consistute the basis of the empirical analysis in Section 5. Due to the nonlinearity in $\beta_n^{\mathcal{C}}$ and $\beta_n^{\mathcal{D}}$ implied by the division by $\mathbb{E}[\widetilde{P}_n]$ we proceed with simulations. Specifically, when evaluating the relation in (46) we simulate the vector Φ and use the expressions for $\beta_n^{\mathcal{C}}$ and $\beta_n^{\mathcal{D}}$ in (A.235) and (A.236), respectively, and the expression for $\widehat{\beta}_n$ in (A.234). In each simulation we also obtain $\mathcal{C}^2$, $\mathcal{D}^2$ based on (A.232)-(A.233) (the only random element being the average loading, $\bar{\Phi}$) and δ in (A.122). We then compute the right-hand side of (46) and estimate the specification:

$$\widehat{\beta}_n = a_0 + a_1\left(\frac{\delta(1-\mathcal{C}^2-\mathcal{D}^2)}{\delta+\mathcal{C}^2+\mathcal{D}^2}\mathbf{1} + \frac{\mathcal{D}^2(1+\delta)}{\delta+\mathcal{C}^2+\mathcal{D}^2}\beta_n^{\mathcal{C}} + \frac{\mathcal{D}^2(1+\delta)}{\delta+\mathcal{C}^2+\mathcal{D}^2}\beta_n^{\mathcal{D}}\right). \tag{A.243}$$

For each draw of Φ we obtain an estimate for the intercept, a_0, and the slope, a_1, of this specification, which reiterates the exercise of Figure A4 but based on approximated rates of return.

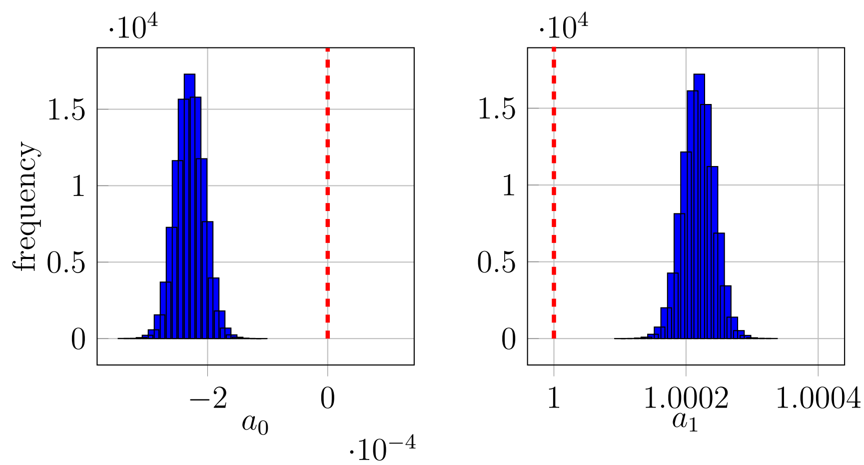

Figure A7: Verification of Equation (46). This figure plots the histogram of the intercept, a_0, (left panel) and slope a_1, (right panel) of the specification in (A.243) across 10^5 simulations. It reports the theoretical value of these coefficients when based on dollar returns (red dashed line). Parameter values correspond to those on page 70.

Based on dollar returns the intercept of the relation in (A.243) is exactly zero and the slope is exactly 1. Figure A7 shows that for approximated rates of return virtually the same relation holds. This exercise suggests that the relation in (46) also applies to approximated rates of return.

We now repeat this exercise for the relation in (47), the second main relation used in empirical tests. In particular, when evaluating the relation in (47) we simulate the vector Φ and use the expressions for $\widehat{\beta}_n$, $\beta_n^{\mathcal{C}}$ and $\beta_n^{\mathcal{D}}$ in (A.234), (A.235) and (A.236), respectively. We also need the unconditional average approximated rate of return on asset n, which satisfies:

$$\mathbb{E}[\widetilde{R}_n^e] = 2\,\mathbb{E}[\widetilde{D}_n]/\,\mathbb{E}[\widetilde{P}_n] - \mathbb{E}[\widetilde{D}_n\widetilde{P}_n]/\,\mathbb{E}[\widetilde{P}_n]^2 - 1 \tag{A.244}$$

$$= \frac{D}{D - \gamma(\tau^{-1}\bar{\Phi}\Phi_n + \tau_\epsilon^{-1}/N)} - \frac{\Phi_n^2\tau^{-1}(\tau - \tau_F)\tau_F^{-1}}{(D - \gamma(\tau^{-1}\bar{\Phi}\Phi_n + \tau_\epsilon^{-1}/N))^2} - 1, \tag{A.245}$$

which we recompute for every draw of Φ. In each simulation we also obtain $\mathcal{C}^2$ and $\mathcal{D}^2$ based on (A.232)-(A.233) and $\mathbb{E}[\widetilde{R}_{\mathbf{M}}^e]$. We then compute the right-hand side of (47) and estimate the specification:

$$\mathbb{E}[\widetilde{R}_n^e] = b_0 + b_1\left(\frac{\mathbb{E}[\widetilde{R}_{\mathbf{M}}^e]}{1-\mathcal{C}^2-\mathcal{D}^2}\widehat{\beta}_n + \frac{\mathcal{C}^2\,\mathbb{E}[\widetilde{R}_{\mathbf{M}}^e]}{1-\mathcal{C}^2-\mathcal{D}^2}\beta_n^{\mathcal{C}} + \frac{\mathcal{D}^2\,\mathbb{E}[\widetilde{R}_{\mathbf{M}}^e]}{1-\mathcal{C}^2-\mathcal{D}^2}\beta_n^{\mathcal{D}}\right). \tag{A.246}$$

We report simulated estimates of the intercept, b_0, and the slope, b_1, in the figure below.

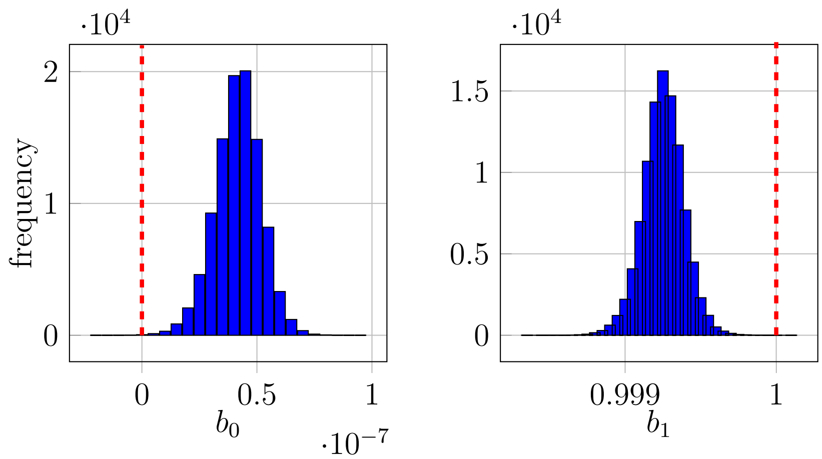

Figure A8: Verification of Equation (47). This figure plots the histogram of the intercept, b_0, (left panel) and slope b_1, (right panel) of the specification in (A.246) across 10^5 simulations. It reports the theoretical value of these coefficients when based on dollar returns (red dashed line). Parameter values correspond to those on page 70.

Based on dollar returns the intercept of the relation in (A.246) is exactly zero and the slope is exactly 1. Figure A8 shows that for approximated rates of return virtually the same relation holds. This exercise suggests that the relation in (A.246) also applies to approximated rates of return.

B 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 (42). 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).

In constructing a proxy for expected excess returns in (42) we make several choices. First, we keep track of $\mathbb{E}^i[\widetilde{R}_n^e]$ 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 point 3 below). 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).

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.

   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 $\mathrm{Var}[\mathbb{E}^{i*}[\widetilde{\mathbf{R}}^e]]$ from (39).

   To get an idea of the matrix of co-beliefs $\mathrm{Var}[\mathbb{E}^{i*}[\widetilde{\mathbf{R}}^e]]$, 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:

$$\begin{bmatrix} \mathbb{E}^{1*}[\widetilde{R}^e_a] & \mathbb{E}^{1*}[\widetilde{R}^e_b] \\ \mathbb{E}^{2*}[\widetilde{R}^e_a] & \mathbb{E}^{2*}[\widetilde{R}^e_b] \\ \mathbb{E}^{3*}[\widetilde{R}^e_a] & \mathbb{E}^{3*}[\widetilde{R}^e_b] \end{bmatrix}. \tag{B.247}$$

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

$$\begin{bmatrix} \mathrm{Var}[\mathbb{E}^{i*}[\widetilde{R}^e_a]] & \mathrm{Cov}[\mathbb{E}^{i*}[\widetilde{R}^e_a], \mathbb{E}^{i*}[\widetilde{R}^e_b]] \\ \mathrm{Cov}[\mathbb{E}^{i*}[\widetilde{R}^e_a], \mathbb{E}^{i*}[\widetilde{R}^e_b]] & \mathrm{Var}[\mathbb{E}^{i*}[\widetilde{R}^e_b]] \end{bmatrix}. \tag{B.248}$$

   (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 $\mathrm{Var}[\mathbb{E}^{i*}[\widetilde{\mathbf{R}}^e]]$ and $\mathbf{M}$, the data necessary to compute $\boldsymbol{\beta}^{\mathcal{D}}$ at the end of each month (Eq. 39) 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:

(a) We do not remove the 1st and 99th percentile of analyst forecasts;

(b) We vary the past window of 180 calendar days over which we collect forecasts from 90 days to 240 days;

(c) We compute consensus forecasts as average instead of median;

(d) We use a rolling window for beta computations (realized, consensus, and dispersion betas) between 24 and 48 months (instead of 36 months);

(e) We do not winsorize our beta estimates;

(f) 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:

   i. We change the threshold of 100 minimal covariances (see above) between 50 and 250.

   ii. When dealing with duplicates (ESTIMID/cusip), we take an average instead of the last observation, for forecasts or for market capitalization.

   iii. We require more than two common forecasters when computing covariances for dispersion betas (that is, between 2 and 10 forecasters).

C 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 2.1)
$\widehat{\boldsymbol{\beta}}$	$N \times 1$	Empiricist's betas (Eq. 17)
$\boldsymbol{\beta}^{\mathcal{C}}$	$N \times 1$	Consensus betas (Proposition 7)
$\boldsymbol{\beta}^{\mathcal{D}}$	$N \times 1$	Dispersion betas (Proposition 7)

$\mathcal{C}^2$		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)
$\boldsymbol{\Phi}$	$N \times 1$	Vector of exposures of assets' payoffs to the fundamental $\widetilde{F}$
$\bar{\Phi}$		Average of $\boldsymbol{\Phi}$
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 $\widehat{\mathbf{Z}}$
$\boldsymbol{\Sigma}$	$N \times N$	$= \boldsymbol{\Phi}\boldsymbol{\Phi}'/\tau + \mathbf{I}/\tau_\epsilon$, investors' variance matrix of future excess returns (Eq. 11)
$\sigma^2_{\mathbf{M}}$		$= \mathbf{M}'\boldsymbol{\Sigma}\mathbf{M} = \bar{\Phi}^2/\tau + 1/(N\tau_\epsilon)$, investors' variance of market's future excess returns
$\widehat{\boldsymbol{\Sigma}}$	$N \times N$	Empiricist's variance matrix of realized excess returns (Lemma 1)
$\widehat{\sigma}^2_{\mathbf{M}}$		$= \mathbf{M}'\widehat{\boldsymbol{\Sigma}}\mathbf{M}$, 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		$\equiv 1/\operatorname{Var}[\widetilde{F}\|\mathscr{F}^i]$, precision of $\widetilde{F}$ conditional on the information set $\mathscr{F}^i$ of investor $i \in [0,1]$
τ_P		Price informativeness