The Low-Minus-High Portfolio and the Factor Zoo

Daniel Andrei[*] Julien Cujean[†] Mathieu Fournier[‡]

May 23, 2021

Abstract

Regardless of whether the CAPM is rejected for valid reasons or by mistake, a single long-short portfolio will always explain, together with the market, 100% of the cross-sectional variation in returns. Yet, this portfolio, which we coin the "Low-Minus-High (LMH) portfolio," need not proxy for fundamental risk. We show theoretically how factors based on valuation ratios (e.g, book-to-market), or on investment rates, can be proxies for the LMH portfolio. More generally, the empiricist can uncover an infinity of proxies for the LMH portfolio, thus unleashing the factor zoo.

We thank Patrick Augustin, Pierluigi Balduzzi, Snehal Banerjee, Laurent Barras, Matthieu Bouvard, Bradyn Breon-Drish, Kent Daniel, Adolfo de Motta, Michael Hasler, Rawley Heimer, Dmitry Livdan (NFA discussant), Alexandre Jeanneret, Serhiy Kozak (AFA discussant), Alessio Saretto, Sergei Sarkissian, Allan Timmermann, and conference and seminar participants at Baruch College, the Bank of Canada, McGill University, Tilburg University, Goethe University, Deutsche Bundesbank, the 2019 Asset Pricing Gerzensee Meeting, the 2019 NFA Meeting, UT Dallas, and UCSD for comments and suggestions.

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

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

[‡]HEC Montréal, Office 4230, 3000 Chemin de la Côte-Sainte-Catherine, Montréal, Québec, H3T 2B1, Canada, mathieufournier.net, mathieu.fournier@hec.ca.

"We really don't know whether to believe the theory or the data... In the light of the uncertainty about the reasons for the difference between the theory and the data, the safest course may be to assume that the theory is correct."

— *Black and Scholes (1974, p. 405)*

1 Introduction

Perhaps the most tested hypothesis in empirical asset pricing is the linear relation between expected returns and betas—the main prediction of the Capital Asset Pricing Model. Tests of this relation led first to its partial empirical validation,[1] then to its repeated demise.[2] Not only returns and betas are not related as the CAPM predicts, but empiricists have also uncovered a "zoo" of anomalies—*hundreds* of factors now explain the cross section of returns[3], so many that the empirical asset pricing field is drowning in anomalies.

Although the interpretation of this vast number of anomalies is debated in the literature, everyone agrees that they are clear evidence against the CAPM—"the CAPM is dead in its tracks" (Fama and French, 2004, p. 36). But before we adopt this general view too heartily, we should consider first what the CAPM rejection means. Suppose we reject the CAPM, then what is the alternative hypothesis? We argue that when the CAPM is rejected—for valid reason(s) or by mistake—there will always be a long-short portfolio (i.e., a factor) that explains, together with the market, the cross-section of returns. This portfolio will become empiricist's strongest ally against the CAPM: even when this portfolio has no economic meaning, the empiricist will *always* fail to reject the alternative hypothesis of a two-factor model of returns. This portfolio has the potential to cause great confusion—any observable variable (macroeconomic factor, or firm characteristic) that covaries with it becomes a contender to join the zoo. Empiricists may thus uncover an infinity of anomalies, forever "adding epicycles."

We base our theoretical argument on a counterexample to prove that finding priced factors other than the market does not necessarily imply that the CAPM fails. Our starting point is an equilibrium model in which the CAPM holds—in the words of Fischer Black and Myron Scholes, we "assume that the theory is correct" (Black and Scholes, 1974, p. 405). In this equilibrium model, investors trade based on private and public information and, on aggregate, hold the market portfolio. The empiricist, who does not observe the information of investors, mis-measures betas and rejects the CAPM (Andrei, Cujean, and Wilson, 2018). Thus, our counterexample builds on the premise that the CAPM is rejected by mistake.

[1] Blume and Friend (1973), Fama and MacBeth (1973).

[2] Reinganum (1981), Lakonishok and Shapiro (1986), Fama and French (1992, 1993). See Fama and French (2004) for a comprehensive review.

[3] Harvey, Liu, and Zhu (2016), Hou, Xue, and Zhang (2018).

In this equilibrium model, an empiricist can build a long-short portfolio that, together with the market, explains 100% of the cross-sectional variation of returns. This portfolio represents the difference between the full-sample mean-variance efficient portfolio with the highest Sharpe ratio (i.e., the *tangency* portfolio) and the market portfolio. For the empiricist who observes data ex-post, this long-short portfolio represents a way to improve efficiency of the market portfolio: assets with positive weights in this portfolio are *under*-invested (cheap) and assets with negative weights are *over*-invested (expensive). Hence, we call the difference between the tangency and the market portfolios the *Low-Minus-High* (LMH) portfolio.

Using the market and the LMH portfolio, the empiricist fails to reject a two-factor model of returns. Of course, one can always find a tangency portfolio that generates exact linearity between betas and expected returns in-sample (Roll, 1977). Should we build the LMH portfolio *ex-post*, it would tautologically benefit from this perfect hindsight. However, the role of the equilibrium model is to identify *ex-ante* variables that covary with the LMH portfolio. We show that observable characteristics, such as market-to-book ratios or investment rates, are good proxies for the LMH portfolio. For the empiricist, firms with low market-to-book ratios or firms with low investment rates command a positive risk premium, in addition to the premium earned from exposure to the market alone. Yet, in the model there is no economic reason for firms with low market-to-book ratio or low investment rates to appear relatively riskier than the CAPM predicts. Rather, as opposed to being priced factors in the model, market-to-book ratio and corporate investment are instruments for beta mis-measurement. More broadly, an empiricist who performs a Principal Component Analysis of the LMH portfolio will find as many principal components as there are factors driving payoffs in the model. The myriad of "factors" the empiricist may uncover thus promptly turns into a factor zoo.

This counterexample illustrates how the hunt for new factors may constitute a methodological trap. We do not debate the importance of these factors, nor do we contest their high risk-adjusted returns. Rather, we, as many others, question their interpretation. In our simple counterexample, factors instrument for beta mis-measurement, thus luring empiricists into believing that these factors are priced. In other words, because CAPM mispricing is possibly spurious, this approach of looking for factors may be one of looking for instruments. And because this approach will always lead to the discovery of new factors—economically meaningful or not—it will never explain why asset-pricing models fail.

We provide an empirical illustration of our theoretical argument, in a set of portfolios that has become the "playing field" of empirical asset pricing: the 25 size and book/market sorted portfolios (Fama and French, 1993). The spectacular failure of the CAPM in this set of portfolios is now a textbook example (e.g., Cochrane, 2009; Campbell, 2017); one can perhaps regard this 5×5 portfolio space as the coffin of the CAPM.

In this portfolio space, we build a long-short portfolio that, together with the market, explains 89% of the cross-sectional variation of returns (in comparison, the Fama and French (1993) three-factor model explains 63% of the variation, whereas the Fama and French (2015) five-factor model explains 73%). This portfolio is the empirical counterpart of the LMH portfolio that we constructed in our theoretical exercise.

As we previously emphasized, in *any* sample of returns, one can *always* find a tangency portfolio that generates exact linearity between betas and expected returns (Roll, 1977); the LMH portfolio, which is built based on the tangency portfolio, will tautologically benefit from this perfect hindsight. In our example, however, we build the LMH portfolio the same way any other risk factor is commonly built, with data that are available only at the time of portfolio formation. Yet, we show that none of the five risk factors from Fama and French (2015) or the momentum factor from Carhart (1997) can explain the returns of the LMH portfolio: its alpha is above 1% per month, independently of the factors used as control variables. Instead, the alphas of existing factors mostly disappear when we regress their returns on the returns of the LMH portfolio. Furthermore, the LMH portfolio commands a positive and strongly statistically significant risk premium, alone or when controlling for the market and/or any other factor(s). And, as our theory predicts, the LMH portfolio is strongly positively correlated with the value and investment factors. Finally, the LMH portfolio has significant explanatory power in other portfolio sorts (e.g., sorts based on past returns).

Have we found, yet again, a better factor, a mighty inhabitant of the zoo? According to our theory—No. While the LMH portfolio does capture risk, its economic interpretation remains elusive. Because we observe realized betas, as opposed to ex-ante measures of betas, we do not know the origin of the CAPM rejection, and the LMH portfolio is not helpful in this matter. Instead, the LMH portfolio only captures what we do not observe.

A large and growing empirical literature attempts to tame the factor zoo.[4] Our approach is theoretical, and, unfortunately, results in a theory of the factor zoo. Roll (1977) has argued that the CAPM will perhaps never be tested. Berk (1995) has argued that the size anomaly cannot be regarded as evidence against any asset pricing theory.[5] We argue that anomalies *in general* should not be regarded as evidence against the CAPM, because they do not reveal the true cause of the CAPM rejection. We conclude that, upon rejection of the CAPM, the factor zoo is ineluctable.

[4] Barillas and Shanken (2018), Bryzgalova (2015), Chen and Zimmermann (2018), Chinco, Neuhierl, and Weber (2019), Chordia, Goyal, and Saretto (2017), Engelberg, McLean, and Pontiff (2018), Feng, Giglio, and Xiu (2019), Giglio and Xiu (2018), Giglio, Liao, and Xiu (2018), Harvey and Liu (2018), Kan and Zhang (1999), Lewellen, Nagel, and Shanken (2010), Linnainmaa and Roberts (2018), McLean and Pontiff (2016), Harvey (2017), Harvey et al. (2016), Hou et al. (2018), Romano and Wolf (2005), Smith (2018), Yan and Zheng (2017).

[5] See also Ferson, Sarkissian, and Simin (1999), MacKinlay and Pástor (2000).

2 The Low-Minus-High Portfolio and the Factor Zoo

In this section we characterize the factor zoo in an equilibrium model. We build a model in which a true CAPM relationship holds for investors, but fails for the empiricist (Andrei et al., 2018). In the model, although there is no economic reason for firms' characteristics (e.g., market-to-book ratios, or investment rates) to be priced, the empiricist concludes that these characteristics yield an additional risk premium beyond what the CAPM can justify, and consequently fails to reject a multifactor model of returns. We show that rejecting the CAPM leaves the empiricist lost in the factor zoo. We use the simplest possible model to make these points; in Section 2.4, we discuss our modeling assumptions.

Consider a one-period economy in which the market consists of one risk-free asset with gross return normalized to 1 and N firms indexed by $n = 1,...,N$. We assume the value of assets in place to be the same for all firms, and denote this value by K. Firms are heterogeneous with respect to the productivity of their assets. Specifically, firm productivities are unobservable at time 0 and have a single-factor structure:

$$\widetilde{\mathbf{Z}} = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_N \end{bmatrix} (F + \widetilde{F}) + \begin{bmatrix} \tilde{\epsilon}_1 \\ \tilde{\epsilon}_2 \\ \vdots \\ \tilde{\epsilon}_N \end{bmatrix} \equiv \mathbf{\Phi}(F + \widetilde{F}) + \tilde{\boldsymbol{\epsilon}}. \tag{1}$$

The common productivity shock $\widetilde{F}$ and each firm-specific shock $\tilde{\epsilon}_n$ are independently normally distributed with means 0 and precisions τ_F and τ_ϵ. Without loss of generality, we assume that the cross-sectional average of firms' loadings on the common productivity shock is positive: $\bar{\Phi} \equiv N^{-1} \sum_{n=1}^{N} \phi_n > 0$. The final values of firms depend on their assets in place and their productivities:

$$\widetilde{\mathbf{D}} = K\widetilde{\mathbf{Z}}. \tag{2}$$

We assume that N claims on these final values (one for each firm) are traded in financial markets.

The economy is populated by a continuum of investors indexed by $i \in [0,1]$, who choose their portfolio at time 0 and derive utility from terminal wealth with constant absolute risk aversion coefficient γ:

$$\begin{aligned} &\max_{\boldsymbol{\omega}_i} \mathbb{E}[-e^{-\gamma \widetilde{W}_i} | \mathscr{F}_i] \\ \text{s.t.} \quad &\widetilde{W}_i = W_{i,0} R_f + \boldsymbol{\omega}_i' (\widetilde{\mathbf{D}} - \widetilde{\mathbf{P}} R_f), \end{aligned} \tag{3}$$

where $W_{i,0}$ is investor i' initial wealth, $\boldsymbol{\omega}_i$ is investor i's portfolio (in units of assets), $R_f \geq 1$ is the gross interest rate, $\widetilde{\mathbf{P}}$ is the vector of equilibrium prices, and $\mathscr{F}_i$ is investor i's information set that we describe in more details below. Without loss of generality, we fix $W_{i,0} = 0$ and $R_f = 1$. Finally, because in this framework rates of returns are not normally distributed, we follow the convention (e.g., Dybvig and Ross, 1985) in the literate of working with dollar returns, $\widetilde{\mathbf{R}}^e \equiv \widetilde{\mathbf{D}} - \widetilde{\mathbf{P}} R_f$, which we refer to as excess returns.

Investors know the structure of realized payoffs in Eqs. (1)-(2), but do not observe the common productivity shock $\widetilde{F}$ and firms' specific productivity shocks $\tilde{\boldsymbol{\epsilon}}$ separately. Each investor i forms expectations about $\widetilde{F}$ based on information inferred from prices, and both a private signal $\widetilde{V}_i = \widetilde{F} + \tilde{v}_i$ and a public signal $\widetilde{G} = \widetilde{F} + \tilde{v}$. The signal noises $\tilde{v}$ and $\tilde{v}_i \perp \tilde{v}$, $\forall i$, are unbiased and independently normally distributed with precisions τ_G and τ_v, respectively. These signals, together with the vector of prices, account for the information set of investor i, $\mathscr{F}_i = \{\widetilde{V}_i, \widetilde{G}, \widetilde{\mathbf{P}}\}$.

Equilibrium prices do not fully reveal investors' private information about the common factor, $\widetilde{F}$. Prices change to reflect new information about firm values, but they also change for reasons unrelated to information, e.g., endowments shocks, preference shocks, or private investment opportunities. To model uninformative price changes, we assume that an unmodeled group of agents trade for liquidity needs and/or for non-informational reasons. Liquidity trading prevents prices from revealing $\widetilde{F}$ (Grossman and Stiglitz, 1980), and prevents investors from refusing to trade (Milgrom and Stokey, 1982).[6]

The total number of shares for all firms is $\mathbf{M} = [M_1 \ldots M_N]'$ (hereafter the *market portfolio*), a vector with strictly positive values that sum up to one. Liquidity traders have inelastic demands of $\widetilde{\mathbf{m}}$ shares, a vector whose elements are normally and independently distributed with precision τ_m, i.e., $\widetilde{\mathbf{m}} \sim \mathcal{N}(\mathbf{0}, \tau_m^{-1}\mathbf{I})$ (where $\mathbf{0}$ denotes a vector of zeros and $\mathbf{I}$ the identity matrix, both of dimension N); 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).

We solve for a linear equilibrium of the economy. Proposition 1 characterizes equilibrium prices, and shows that equilibrium market-to-book ratios reflect both the information of market participants, $\widetilde{F}$ and $\widetilde{G}$, and liquidity needs $\widetilde{\mathbf{m}}$.

Proposition 1. *There exists a partially revealing rational expectations equilibrium in which*

[6] 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.

the vector of market-to-book ratios, $\widetilde{\mathbf{P}}/K$, *is given by*

$$\frac{\widetilde{\mathbf{P}}}{K} = \mathbf{\Phi}F + \boldsymbol{\xi}_0\mathbf{M} + \boldsymbol{\alpha}\widetilde{F} + \mathbf{g}\widetilde{G} + \boldsymbol{\xi}\widetilde{\mathbf{m}}, \tag{4}$$

where the coefficients $\boldsymbol{\xi}_0$ $(N \times N)$, $\boldsymbol{\alpha}$ $(N \times 1)$, $\mathbf{g}$ $(N \times 1)$, *and* $\boldsymbol{\xi}$ $(N \times N)$ *solve*

$$\boldsymbol{\xi}_0 = -\gamma\frac{\mathbf{\Sigma}}{K}, \quad \boldsymbol{\alpha} = \mathbf{\Phi}\frac{\tau - \tau_F - \tau_G}{\tau}, \quad \mathbf{g} = \mathbf{\Phi}\frac{\tau_G}{\tau}, \quad \boldsymbol{\xi} = -\frac{\gamma K + \sqrt{\tau_m \tau_P}}{\tau}\mathbf{\Phi}\mathbf{\Phi}' - \frac{\gamma K}{\tau_\epsilon}\mathbf{I}, \tag{5}$$

τ_P *represents an endogenous scalar linked to price informativeness,* $\tau \equiv \mathrm{Var}^{-1}[\widetilde{F}|\mathscr{F}_i]$ *is a scalar identified as the unique positive solution to a cubic equation, and* $\mathbf{I}$ *is the identity matrix of dimension* $N \times N$.

Proof. See Appendix A.1. □

Define by $\boldsymbol{\mu}_i \equiv \mathbb{E}[\widetilde{\mathbf{R}}^e|\mathscr{F}_i]$ the vector of expected returns that investor i builds for all firms based on her information $\mathscr{F}_i$. Similarly, denote by $\mathbf{\Sigma} \equiv \mathrm{Var}[\widetilde{\mathbf{R}}^e|\mathscr{F}_i]$ the conditional covariance matrix of returns for all firms, which is identical across investors (because they have identical precision, τ, over $\widetilde{F}$; see Proposition 1). We can then write investor i's optimal portfolio choice as

$$\boldsymbol{\omega}_i = \frac{1}{\gamma}\mathbf{\Sigma}^{-1}\boldsymbol{\mu}_i. \tag{6}$$

All investors are mean-variance maximizers. However, unlike in a standard CAPM framework, they all operate under their own information set. In particular, as parameters of the conditional mean-variance set, they all use the same $\mathbf{\Sigma}$ but different $\boldsymbol{\mu}^i$.

Market clearing requires that investors' individual demand and that of liquidity traders sum up to the market portfolio, M:

$$\int_i \boldsymbol{\omega}_i di + \widetilde{\mathbf{m}} = \mathbf{M}. \tag{7}$$

Let *consensus beliefs* of investors be $\bar{\boldsymbol{\mu}} = \int_i \boldsymbol{\mu}_i di$. Substituting individual portfolios in Eq. (6) into the market-clearing condition in Eq. (7) gives:

$$\bar{\boldsymbol{\mu}} = \gamma\mathbf{\Sigma}(\mathbf{M} - \widetilde{\mathbf{m}}), \tag{8}$$

which represents the expected rate of return that every particular asset must pay for investors to be willing to hold the supplies of the N assets, net of liquidity traders' demand. A more general form of Eq. (8), in which γ and $\mathbf{\Sigma}$ are time-varying, has been derived by Jensen

(1972), and further studied by Bollerslev, Engle, and Wooldridge (1988).[7]

Note that, although Eq. (8) will be shortly shown to imply an unconditional CAPM relation on which all investors agree, this model is not a standard CAPM framework. Because each investor holds different expectation, $\boldsymbol{\mu}^i$, she does not find it optimal conditionally to hold the market portfolio, $\mathbf{M}$. Only an investor who has *average unconditional beliefs*, $\mathbb{E}[\widetilde{\mathbf{R}}^e]$ amd $\boldsymbol{\Sigma}$, holds the market portfolio. In this context, there is CAPM pricing in the sense that unconditional betas are conditional on information that investors know. Formally, conditioning down Eq. (8) expected returns are proportional to a new notion of $\boldsymbol{\beta}$ based on investors' covariance matrix.

Corollary 1.1. *In this economy, an unconditional CAPM relation holds:*

$$\boldsymbol{\mu} = \frac{\boldsymbol{\Sigma}\mathbf{M}}{\sigma_M^2}\mu_M = \boldsymbol{\beta}\mu_M, \tag{9}$$

where $\boldsymbol{\mu} \equiv \mathbb{E}[\bar{\boldsymbol{\mu}}]$, $\sigma_M^2 \equiv \mathbf{M}'\boldsymbol{\Sigma}\mathbf{M}$ *is the variance of excess returns for the market portfolio, and* $\mu_M \equiv \mathbf{M}'\boldsymbol{\mu}$ *is the unconditional expected excess return on the market.*

Proof. Take unconditional expectation of Eq. (8) to obtain $\boldsymbol{\mu} = \gamma\boldsymbol{\Sigma}\mathbf{M}$. Multiplication with $\mathbf{M}'$ yields $\mu_M = \gamma\sigma_M^2$. Divide $\boldsymbol{\mu} = \gamma\boldsymbol{\Sigma}\mathbf{M}$ by $\mu_M = \gamma\sigma_M^2$ to obtain (9). □

Since all investors observe the market portfolio, $\mathbf{M}$, they all agree on the unconditional relation in Eq. (9). Therefore, from investors' perspective $\boldsymbol{\beta}$ is the vector of *true betas* and Eq. (9) is the *true* CAPM. An equivalent way of stating Corollary 1.1 is that the market portfolio $\mathbf{M}$ is mean-variance efficient under average unconditional beliefs and thus commands the highest Sharpe ratio in the economy (Roll, 1977).

Corollary 1.2. *Based on the observation available in the market, the Sharpe ratio of the market portfolio* $\mathbf{M}$ *reaches its maximum attainable level in the economy:*

$$\frac{\mu_M}{\sigma_M} = \sqrt{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}. \tag{10}$$

Proof. The proof follows from efficient set mathematics. Define $B \equiv \mathbf{1}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$, $C \equiv \boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$, where $\mathbf{1}$ is a vector of ones of conformable dimension. The tangency (market) portfolio has an expected excess return of $\mu_M = C/B$ and a variance of excess returns of $\sigma_M^2 = C/B^2$. Thus, $\mu_M/\sigma_M = \sqrt{C}$, which yields (10). □

[7] Eq. (8), common in noisy rational expectation models (Admati, 1985), is a special case of the ICAPM without hedging terms (Merton, 1973), or of a standard intertemporal asset pricing model (Campbell, 1993).

The empiricist observes realized returns—as opposed to investors' expected returns—on all assets and on the market portfolio, $\mathbf{M}$. The law of iterated expectations implies that the empiricist correctly measures $\boldsymbol{\mu}$ and μ_M. But, because empiricist's information set is coarser than of any individual investor, the law of total variance implies that the covariance matrix of excess returns of the empiricist, $\widehat{\boldsymbol{\Sigma}} \equiv \mathrm{Cov}[\widetilde{\mathbf{R}}^e]$, differs from $\boldsymbol{\Sigma}$. As a result, the empiricist rejects the CAPM.

Corollary 1.3. *Under the information set of the empiricist, the unconditional market portfolio is not mean-variance efficient:*

$$\frac{\mu_M}{\widehat{\sigma}_M} < \sqrt{\boldsymbol{\mu}' \widehat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\mu}}, \tag{11}$$

where $\widehat{\boldsymbol{\Sigma}}$ is the unconditional covariance matrix of realized excess returns and $\widehat{\sigma}_M \equiv \sqrt{\mathbf{M}' \widehat{\boldsymbol{\Sigma}} \mathbf{M}}$ is the volatility of excess returns of the market. Thus, the empiricist rejects the CAPM.

Proof. See Appendix A.2. □

Empiricist's rejection of the CAPM—despite using the correct market portfolio $\mathbf{M}$—can be understood in two ways. Under average unconditional beliefs, the market portfolio is the tangency portfolio. For the empiricist, all assets have the correct unconditional expected returns, but display systemtically larger unconditional variance (due to variation in expected returns, which the empiricist does not observe). Hence, all assets—including the market—in the mean-variance space move to the right. But, since Corollary 1.3 implies that the market portfolio cannot possibly be the tangency portfolio, the market portfolio in the eyes of the empiricist moves *inside* the mean-variance frontier.

Another way of understanding the CAPM rejection is from an econometric perspective. Conditional on her own information each investor i sees expected returns as a noisy perturbation around the CAPM relation in Eq. (9):

$$\boldsymbol{\mu}^i = \boldsymbol{\beta} \mu_M + \boldsymbol{\epsilon}^i, \quad \text{where } \boldsymbol{\epsilon}^i \sim \mathcal{N}(\mathbf{0}, \mathrm{Var}[\boldsymbol{\mu}^i]). \tag{12}$$

This perturbation arises because returns are predictable from an investor's perspective (Ferson and Harvey, 1991; Pesaran and Timmermann, 1995; Cochrane, 2007), (i.e., $\mathrm{Var}[\boldsymbol{\mu}^i]$ is not zero). Even though there may be substantial predictability at investors' level, the law of iterated expectations ensures that this perturbation vanishes when conditioning down. However, for the empiricist, who uses realized returns as opposed to expected returns, the law of total variance:

$$\widehat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma} + \mathrm{Var}[\boldsymbol{\mu}^i], \tag{13}$$

ensures that this pertubation leaves a mark on the CAPM relation she estimates. Typical betas, which are computed using the covariance matrix of realized returns, do depend on $\mathrm{Var}[\boldsymbol{\mu}^i]$, the extent to which returns are predictable (Andrei et al., 2018).

Two sources of variation together lead the empiricist to reject the CAPM. In particular, we can rewrite the second term in the law of total variance in Eq. (13) as:

$$\mathrm{Var}[\boldsymbol{\mu}^i] = \mathrm{Var}[\bar{\boldsymbol{\mu}}] + \mathrm{Var}[\boldsymbol{\mu}^i - \bar{\boldsymbol{\mu}}]. \tag{14}$$

First, there is aggregate (time-series) variation in consensus expected returns, $\bar{\boldsymbol{\mu}}$. The empiricist, who observes realized returns but does not observe $\bar{\boldsymbol{\mu}}$, can compute realized beta but cannot compute ex-ante measures of beta. Second, there is cross-sectional variation in expected returns across agents, $\mathrm{Var}[\boldsymbol{\mu}^i - \bar{\boldsymbol{\mu}}]$. This term is identical for any investor, and thus represents the variance of expected returns computed across the population of investors. Thus, this decomposition shows that, although the empiricist observes time variation only, cross-sectional variation "hides" in the variation the empiricist measures. Because the empiricist does not observe investors' individual information, she cannot control for cross-sectional variation either (see also Section 2.4). Andrei et al. (2018) show that cross-sectional variation generates substantially more distortion in the CAPM relation relative to what time-series variation can justify (Jagannathan and Wang, 1998).

Corollary 1.3 implies that, for the empiricist, there exists a tangency portfolio $\mathbf{T} \neq \mathbf{M}$, which, based on observed realized returns, is mean-variance efficient (i.e., the portfolio with the maximum attainable Sharpe ratio). Assume now that the empiricist constructs a portfolio based on deviations between T and M:[8]

$$\boldsymbol{\Delta} = \mathbf{T} - \mathbf{M}. \tag{15}$$

Since both $\mathbf{T}$ and $\mathbf{M}$ sum up to one, $\boldsymbol{\Delta}$ sums up to zero and is therefore a long-short portfolio. For an empiricist who observes data ex-post, $\boldsymbol{\Delta}$ represents a way to improve efficiency of the market portfolio (with perfect hindsight). Assets with positive weights in this portfolio ($\boldsymbol{\Delta}^+$ assets) are *under*-invested; assets with negative weights in this portfolio ($\boldsymbol{\Delta}^-$ assets) are *over*-invested. Since under-invested assets appear relatively cheaper than over-invested assets, the portfolio $\boldsymbol{\Delta}$ is a *Low-Minus-High* (LMH) portfolio.

[8]The portfolio T likely involves short positions, but this does not change the argument.

The following Proposition shows that the LMH portfolio, together with the market, explains 100% of the cross-sectional variation in realized excess returns.

Proposition 2. *For the empiricist, the portfolio* $\mathbf{T} = \widehat{\mathbf{\Sigma}}^{-1}\boldsymbol{\mu}/\widehat{B}$*, where* $\widehat{B} \equiv \mathbf{1}'\widehat{\mathbf{\Sigma}}^{-1}\boldsymbol{\mu}$*, is ex-post efficient (Roll, 1977). Since* $\mathbf{T} \neq \mathbf{M}$ *(from Corollary 1.3), the empiricist can define* $\mathbf{\Delta} \equiv \mathbf{T} - \mathbf{M}$ *and write the expected returns on all assets as*

$$\boldsymbol{\mu} = \frac{\mu_T \widehat{\sigma}_M^2}{\widehat{\sigma}_T^2}\widehat{\boldsymbol{\beta}} + \frac{\mu_T \widehat{\sigma}_\Delta^2}{\widehat{\sigma}_T^2}\widehat{\boldsymbol{\beta}}_\Delta, \tag{16}$$

where $\mu_T \equiv \mu' T$*,* $\widehat{\sigma}_T^2 \equiv \mathbf{T}'\widehat{\mathbf{\Sigma}}\mathbf{T}$*,* $\widehat{\sigma}_\Delta^2 \equiv \mathbf{\Delta}'\widehat{\mathbf{\Sigma}}\mathbf{\Delta}$*,* $\widehat{\boldsymbol{\beta}} \equiv \widehat{\mathbf{\Sigma}}\mathbf{M}/\widehat{\sigma}_M^2$*, and* $\widehat{\boldsymbol{\beta}}_\Delta \equiv \widehat{\mathbf{\Sigma}}\mathbf{\Delta}/\widehat{\sigma}_\Delta^2$*.*

Proof. Since the portfolio $\mathbf{T}$ is ex-post efficient, efficient set mathematics imply

$$\boldsymbol{\mu} = \frac{\mu_T}{\widehat{\sigma}_T^2}\widehat{\mathbf{\Sigma}}\mathbf{T}. \tag{17}$$

Replacing $\mathbf{T} = \mathbf{M} + \mathbf{\Delta}$ and using the definitions of $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\beta}}_\Delta$ yields (16). □

The set of market betas that the empiricist computes, $\widehat{\boldsymbol{\beta}}$, are based on the covariance matrix of realized excess returns $\widehat{\mathbf{\Sigma}}$ and thus differ from the set of true betas, which are based on $\mathbf{\Sigma}$. Hence, upon rejecting the CAPM, Proposition 2 implies that the empiricist can always build two sets of betas, one with respect to the market portfolio and one with respect to the LMH portfolio that together explain ex-post 100% of cross-sectional variation in returns. In other words, since for the empiricist beta does not capture all the risk, a second factor picks up the associated mispricing, i.e., the remaining variation.

When the empiricist rejects the CAPM, there will always be a portfolio $\mathbf{\Delta}$ that, when added as a second factor, helps explain the entire cross-sectional variation in excess returns. In particular, recall that there always exists a portfolio that is ex-post efficient (Roll, 1977). In this one-period model, the LMH portfolio can only be constructed ex-post. As a result, it will be tautologically true that adding LMH formed ex-post to the CAPM relation produces an R^2 of 1, which is the result of Proposition 2. In contrast, in our subsequent empirical exercise we will form LMH ex-ante, and thus this tautological result will no longer apply.

The theoretical merit of forming LMH ex-post (Proposition 2) is that we can identify factors observable ex-ante that are proxies for the LMH portfolio. In this simplest version of the model with just one factor, the whole cross-section of stocks is spanned by just two vectors in equilibrium: the market portfolio, $\mathbf{M}$, and the vector of loadings, $\mathbf{\Phi}$. Since $\widehat{\boldsymbol{\beta}}$ is a linear combination of two, adding any factor that depends either on $\mathbf{\Phi}$ or $\mathbf{M}$ or a combination of the two will perfectly explain along with the market portfolio the entire cross-section of returns; we can quickly identify two such factors.

2.1 The Value Factor

We will now provide a theoretical link between the LMH portfolio and the High-Minus-Low (HML, or value) factor. We directly state this result in the following Proposition.

Proposition 3. *In this economy, the cross-sectional variation in excess returns is entirely explained by empiricist's betas together with the vector of market-to-book ratios,* $\mathbb{E}[\widetilde{\mathbf{P}}]/K$*:*

$$\boldsymbol{\mu} = \lambda_1 \widehat{\boldsymbol{\beta}} + \lambda_2 \frac{\mathbb{E}[\widetilde{\mathbf{P}}]}{K}, \text{ with } \lambda_1 > 0 \text{ and } \lambda_2 < 0. \tag{18}$$

Proof. See Appendix A.3. □

Proposition 3 follows from that empiricist's betas capture only partially variations in returns generated by exposure to the common factor $\widetilde{F}$. The remaining variation is captured by the vector of sensitivities $\boldsymbol{\Phi}$. Since firms' productivities are driven by one factor only in the model (in other words, the covariance matrix of returns has a strong factor structure), in equilibrium the vector $\boldsymbol{\Phi}$ correlates perfectly with the vector of market-to-book ratios, $\mathbb{E}[\widetilde{\mathbf{P}}]/K$. It follows that $\mathbb{E}[\widetilde{\mathbf{P}}]/K$ is a *characteristic* that proxies perfectly for $\boldsymbol{\Phi}$.

The negative sign of λ_2 admits the following interpretation. Firms with low market-to-book ratios (value firms) have higher expected excess returns after controlling for market beta. Because we have not assumed ex-ante that these firms are inherently riskier, in this economy a value premium cannot possibly be reward for fundamental risk. Rather, the value premium simply reflects mis-measurement in beta estimates. Similarly, that value firms command a risk premium that is not explained by exposure to the market cannot be regarded as evidence against the CAPM. Corollary 1.1 shows that a true CAPM holds in this economy. Yet, because the empiricist conducts inference under a coarser information set than that of investors, mis-measurement in betas not only leads to the rejection of the CAPM, but also to the creation of a value factor.

This result should not be taken to refute the validity of the value factor as a determinant of returns, as it *does* capture risk. Rather, what this result questions is the interpretation of value as a risk factor, as well as its use against the CAPM. After all, Proposition 3 shows that in this model (with one common factor driving payoffs) the value factor explains all the remaining variation not captured by empiricist's betas. From a pragmatic perspective, having found an instrument that improves cross-sectional fit may be good enough. From a theoretical perspective, however, all empirical attempts at interpreting value as a risk factor will be subject to the tautological result of Proposition 2, and hence inconclusive.

As we point out above, that the value factor captures all the remaining variation depends entirely on our assumption of a single common productivity factor $\widetilde{F}$. We analyze the case of multiple productivity factors in Section 2.3, and show that the value factor captures only partially the remaining risk—even adopting a pragmatic perspective, the value factor is not a panacea. Furthermore, according to our model, the data should exhibit a positive relation between the returns of the LMH portfolio and the returns of the value factor, at least in portfolio spaces that indeed have a strong factor structure (e.g., Fama and French, 1993). We turn to this empirical implication in Section 3.

2.2 The Investment Factor

The q theory of investment (Jorgenson, 1963; Tobin, 1969; Lucas and Prescott, 1971; Hayashi, 1982) predicts a strong relationship between firms' market values and their investment rates. Because firms' valuations are driven by expectations of their future cash-flows, high valuations must indicate profitable opportunities and therefore highly-valued firms should invest more aggressively. Recent data lend support for this positive relationship (Andrei, Mann, and Moyen, 2018).

This theoretical link between the valuation ratio and the investment rate, in conjunction with our previous result that the LMH portfolio is a good proxy for the value factor (and a perfect proxy when payoffs are driven by a single common factor), suggest that factors built based on firms' investment rates must be related with the LMH portfolio. One such factor is the investment factor, whose returns represent the difference between the returns on diversified portfolios of the stocks of low and high investment firms. Fama and French (2015) and Hou, Xue, and Zhang (2015) document that the investment factor explains a substantial amount of variation in the cross section of returns.

In this section, we provide a theoretical link between the LMH portfolio and the investment factor. Our argument is heuristic, in that we consider a minimal extension of our setup in which market equilibrium is reached as in Proposition 1, but we incorporate firms' decision to invest. More precisely, let the ex-post profit of a firm n be

$$\Pi_n = \left[\phi_n(F + \widetilde{F}) + \tilde{\epsilon}_n\right](K + I_n) - I_n - \frac{a}{2}\left(\frac{I_n}{K}\right)^2 K, \tag{19}$$

where I_n represents the investment decision of firm n. The last term represents adjustment costs, which are strictly convex ($a > 0$) and linear homogeneous in I and K (Hayashi, 1982).

We assume that the firms-specific component, $\tilde{\epsilon}_n$, is perfectly observed by the insider of the firm (hereafter "the manager"). Furthermore, the manager observes a private signal about $\widetilde{F}$, $V_m = \widetilde{F} + v_m$, where $v_m \sim N(0, 1/\tau_{v,m})$. Because this signal is imperfect, the manager also uses public prices to learn about $\widetilde{F}$, as investors do. Maximization of (19) yields the optimal investment decision

$$\frac{I_n^*}{K} = -\frac{1}{a} + \frac{1}{a}\left(\phi_n \mathbb{E}[F + \widetilde{F}|\mathscr{F}_m] + \tilde{\varepsilon}_n\right), \tag{20}$$

where $\mathscr{F}_m$ is the information set of the manager. This yields a direct relationship between the investment rate of firm n and the beliefs of the manager. Taking unconditional expectation and writing this relationship for all N firms yields

$$\frac{\mathbb{E}[\mathbf{I}^*]}{K} = -\frac{1}{a}\mathbf{1} + \frac{1}{a}\mathbf{\Phi}F. \tag{21}$$

Eq. (21) provides a direct link between firms' average investment rates and firms' exposure to the common factor. Let us assume, for the sake of the argument, that equilibrium prices preserve the form given in Proposition 1.[9] The following Proposition, then, draws directly from Proposition 3 and Eq. (21).

Proposition 4. *In this economy, the cross-sectional variation in excess returns is entirely explained by firms' betas together with the vector of investment ratios,* $\mathbb{E}[\mathbf{I}^*]/K$*:*

$$\boldsymbol{\mu} = \eta_0\mathbf{1} + \eta_1\boldsymbol{\beta} + \eta_2\frac{\mathbb{E}[\mathbf{I}^*]}{K}, \text{ with } \eta_0 < 0,\ \eta_1 > 0 \text{ and } \eta_2 < 0. \tag{22}$$

Proof. See Appendix A.4 □

Firms with low investment rates (conservative firms) have higher returns after controlling for market beta. The q theory of investment implies that these firms also have low valuations, which, in the context of our model, means low market-to-book ratios. Because these firms earn a positive risk premium after controlling for their market beta (Proposition 3), the investment factor must command a positive risk premium, hence $\eta_2 < 0$. We conclude that firms with lower investment rates earn a risk premium beyond the premium required by their exposure to the market. Our model therefore predicts a positive relationship between the returns of the LMH portfolio and the returns of the investment factor, a prediction that we will confirm empirically in Section 3. However, the conclusion from the previous section still holds—that low investment firms command a risk premium that is not explained by exposure to the market cannot be regarded as evidence against the CAPM.

[9] We do not revisit here the equilibrium of Proposition 1. Because the manager has information about $\widetilde{F}$ and $\tilde{\varepsilon}_n$, the investment decision (20) is public information from which investors can learn. This would require the addition of N public signals to the learning problem of each agent. Furthermore, the optimal investment decision will introduce quadratic terms in the firms' final payoffs, which breaks the linearity of the CARA-normal setup. Overcoming this technical issue would require a different model (Albagli, Hellwig, and Tsyvinski, 2011a,b; David, Hopenhayn, and Venkateswaran, 2016), which is beyond the scope of our paper. Alternatively, one can focus on first-order terms in firms' payoffs (Bai, Philippon, and Savov, 2016), which would preserve the linearity of the model.

2.3 The Factor Zoo

In the previous sections, we have shown how three factors (LMH, the value factor, and the investment factor) can account separately for the mispricing the empiricist perceives. Beta mis-measurement creates the illusion that factors other than the market are asset-pricing relevant. The empiricist can find several such factors, even if in our theoretical model asset payoffs are driven by a single common factor $\widetilde{F}$. We now allow payoffs to be driven by $J \geq 1$ common factors. Although the unconditional CAPM still holds, the empiricist may now uncover a myriad of factors, thus unleashing the factor zoo.

To link the LMH portfolio to the factor zoo, recall that adding LMH formed ex-post to the CAPM relation always produces an R^2 of 1. But then the J additional factors can be identified as principal components of LMH formed ex-post. We now examine this idea in the context of a model with J factors driving payoffs. We then examine a "large economy" in which the number of stocks and the number of factors grow unboundedly but in a way that their relative size, $J/N \to \psi \in [0,1]$, remains finite (e.g., Martin and Nagel (2020)). In this context, we can examine in further details how factors and the LMH portfolio relate in terms of the distribution of eigenvalues of factor loadings; this distribution of eigenvalues specifies the "demographics" of the factor zoo.

2.3.1 LMH in the factor zoo

Denote a vector of $J \leq N$ independent factors by $\widetilde{\mathbf{F}} \equiv [F_1+\widetilde{F}_1\ F_2+\widetilde{F}_2 \ldots\ F_J+\widetilde{F}_J]'$. Let this vector be normally distributed with mean $\mathbf{F} \equiv [F_1\ F_2 \ldots F_J]'$ and covariance matrix $(J\tau_F)^{-1}\mathbf{I}_J$. We scale prior precision on factors by J to obtain meaningful limits subsequently.[10] Let the vector of realized asset payoffs have the structure:

$$\widetilde{\mathbf{D}} = K(\mathbf{\Phi}\widetilde{\mathbf{F}} + \tilde{\boldsymbol{\epsilon}}), \tag{23}$$

where we decompose the matrix of loadings as $\mathbf{\Phi} \equiv [\mathbf{\Phi}_1\ \mathbf{\Phi}_2\ \ldots\ \mathbf{\Phi}_J]$. That is, the vector $\mathbf{\Phi}_j$ contains the loadings of each stock on the $j-$th factor. We further assume that $\text{rank}(\mathbf{\Phi}) = J$ and:

$$\frac{1}{NJ}\text{tr}(\mathbf{\Phi}'\mathbf{\Phi}) = 1. \tag{24}$$

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

$$\widetilde{\mathbf{V}}_i = \widetilde{\mathbf{F}} + \tilde{\mathbf{v}}_i, \quad \tilde{\mathbf{v}}_i \sim N\left(\mathbf{0}, (J\tau_v)^{-1}\mathbf{I}_J\right), \tag{25}$$

[10] This is because $\text{Var}[\mathbf{\Phi}\widetilde{\mathbf{F}}] = \frac{\tau_F^{-1}}{J}\mathbf{\Phi}\mathbf{\Phi}'$, so that the average prior does not grow with J under Eq. (24).

as well as a common public signal,

$$\widetilde{\mathbf{G}} = \widetilde{\mathbf{F}} + \tilde{\mathbf{v}}, \quad \tilde{\mathbf{v}} \sim N\left(\mathbf{0}, (J\tau_G)^{-1}\mathbf{I}_J\right). \tag{26}$$

We also scale the precision of both kinds of signal by J to ensure that they retain informational content in the large-economy limit we consider subsequently.

Other than allowing multiple factors to affect payoffs, we keep the structure of the model unchanged. Proposition 5 characterizes equilibrium prices in this economy.

Proposition 5. *There exists a partially revealing rational expectations equilibrium in which the vector of market-to-book ratios is given by*

$$\frac{1}{K}\widetilde{\mathbf{P}} = \mathbf{\Phi}\mathbf{F} - \gamma\frac{1}{K}\mathbf{\Sigma}\mathbf{M} + \boldsymbol{\alpha}\widetilde{\mathbf{F}} + \mathbf{\Phi}\tau_G J\boldsymbol{\tau}^{-1}\widetilde{\mathbf{G}} + \boldsymbol{\xi}\widetilde{\mathbf{m}}, \tag{27}$$

where the coefficients $\boldsymbol{\alpha}$ $(N \times J)$*, and* $\boldsymbol{\xi}$ $(N \times N)$ *solve*

$$\boldsymbol{\alpha} = \mathbf{\Phi}\boldsymbol{\tau}^{-1}(\boldsymbol{\tau} - \tau_F J\mathbf{I}_J - \tau_G\mathbf{I}_J), \quad \boldsymbol{\xi} = -\sqrt{\tau_m}\mathbf{\Phi}\boldsymbol{\tau}^{-1}\boldsymbol{\tau}_P'\mathbf{\Phi}' - \gamma K\left(\mathbf{\Phi}\boldsymbol{\tau}^{-1}\mathbf{\Phi}' + \tau_\epsilon^{-1}\mathbf{I}_N\right), \tag{28}$$

and the $J \times J$ *matrix* $\boldsymbol{\tau} \equiv \mathrm{Var}^{-1}[\widetilde{\mathbf{F}}|\mathscr{F}_i]$ *denotes total precision on the vector of* J *factors:*

$$\boldsymbol{\tau} \equiv (\tau_F + \tau_v + \tau_G)J\mathbf{I}_J + \boldsymbol{\tau}_P\mathbf{\Phi}'\mathbf{\Phi}\boldsymbol{\tau}_P, \tag{29}$$

where $\boldsymbol{\tau}_P$ *is a* $J \times J-$*matrix, which is defined in the appendix.*

Proof. See Appendix A.5. □

In the presence of multiple factors the CAPM still holds under average unconditional beliefs (see Corollary 1.1 for a proof). We proceed directly with empiricist's perception of the CAPM, which we characterize below.

Proposition 6. ***(Factor Zoo)*** *In the eyes of the empiricist, expected returns satisfy the equilibrium relation:*

$$\boldsymbol{\mu} = \underbrace{\frac{\widehat{\sigma}_M^2}{\sigma_M^2}\left(1 + \frac{\gamma^2K^2}{\tau_m\tau_\epsilon}\right)^{-1}\widehat{\boldsymbol{\beta}}\mu_M}_{\text{distorted CAPM relation}} - \underbrace{\frac{\gamma\tau_m\tau_\epsilon}{\gamma^2K^2 + \tau_m\tau_\epsilon}\sum_{k=1}^{J}\left(\sum_{j=1}^{J}\bar{\phi}_j c_{kj}\right)\mathbf{\Phi}_k}_{\text{Factor Zoo}}, \tag{30}$$

where $\bar{\phi}_j \equiv \mathbf{\Phi}_j'\mathbf{M}$ *denotes the average loading on factor* j *and the coefficients* c_{jk} *are defined* *in the appendix. In this context, the betas on the LMH portfolio satisfies the relation:*

$$\widehat{\boldsymbol{\beta}}_{\Delta} = \frac{\widehat{\sigma}_T^2}{\mu_T \widehat{\sigma}_{\Delta}^2} \frac{\gamma \tau_m \tau_\epsilon}{\gamma^2 K^2 + \tau_m \tau_\epsilon} \sum_{k=1}^{J} \widehat{B}_k \left(\sum_{j=1}^{J} \bar{\phi}_j c_{kj} \right) \widehat{\boldsymbol{\Sigma}} \left(\mathbf{M} - \widehat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\Phi}_k / \widehat{B}_k \right), \tag{31}$$

where $\widehat{B}_k \equiv \boldsymbol{\Phi}_k' \widehat{\boldsymbol{\Sigma}}^{-1} \mathbf{1}_N$ *and thus* $\widehat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\Phi}_k / \widehat{B}_k$ *is the efficient portfolio fully invested in Factor k.*

Proof. See Appendix A.6. □

Eq. (30) shows how beta mis-measurement opens the gate to the factor zoo. Although none of the J factors is relevant for asset-pricing tests—the CAPM holds—the empiricist concludes that all J factors should be added to the (distorted) CAPM relation (this is the best possible scenario, in which the empiricist would observe $\boldsymbol{\Phi}_j$, $\forall j$). In other words, although some factors may appear more relevant than others, they each separately increase the R^2 of the relation. In fact, if we added them all, the fit would be perfect.

What each factor does is to correct the market portfolio to bring it closer to the empiricist's perceived tangency portfolio. Eq. (31) shows how. Recall that the LMH portfolio, $\boldsymbol{\Delta}$, represents the gap between the market portfolio and the perceived tangency portfolio. If these two portfolios coincided, the CAPM would no longer be rejected. Suppose the empiricist has successfully identified the J factors driving payoffs. Based on mean-variance analysis, she could form J efficient portfolios, each fully invested in each factor. Formally, the mathematics of the efficient frontier imply that the efficient portfolio invested in the j-th factor would take the form, $\widehat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\Phi}_j / \widehat{B}_j$. The portfolio LMH compares each of these efficient portfolios to the market portfolio (the term in bracket in Eq. (31)) and weighs the difference between the two according to the importance of each factor j (i.e., its average loading $\bar{\phi}_j$).

It is worth mentioning that the LMH portfolio takes us to the empirical tangency portfolio in one step. However, adding one factor after another would get us gradually toward the tangency portfolio. The issue with this *partial* approach is that it creates an omitted variable bias, for there remains unexplained variation that would be certainly captured by the remaining factors. These factors act as omitted variables and distort the risk-premium estimates. We will elaborate on this issue in Section 3.

Using the vectors $\boldsymbol{\Phi}_j$ as factors is only one way for the empiricist to account for the CAPM mispricing. An alternative is to turn to firm characteristics. In this case, none of the characteristics alone can account for all the missing factors. That is, Propositions 3 and 4 do not hold: because the matrix $\boldsymbol{\Phi}$ has now several columns, the value factor and the investment factor are not proportional to the "Factor Zoo" term in Eq. (30). Proposition 7 provides the proof of this result for the value factor.

Proposition 7. *The market-to-book ratio alone in the factor zoo (or any other firm characteristic that does not span the zoo entirely) is no longer sufficient to account for the CAPM mispricing. The empiricist now finds that J other factors (one of which is redundant), beyond the market and market-to-book ratio, explain cross-sectional variation in excess returns:*

$$\boldsymbol{\mu} = \lambda_1 \widehat{\boldsymbol{\beta}} + \lambda_2 \frac{\mathbb{E}[\widetilde{\mathbf{P}}]}{K} + \sum_{k=1}^{J} \lambda_{k+2} \boldsymbol{\Phi}_k. \tag{32}$$

Proof. See Appendix A.7. □

Proposition 7 shows that the statistical significance of a given firm characteristic is diluted away in the factor zoo. Furthermore, as soon as more than a single factor drives asset payoffs, the sign of the loading λ_2 on the market-to-book ratio becomes arbitrary, in contrast to the result of Proposition 3. In other words, anything goes—the empiricist may find infinitely many combinations of factors that help improve the fit of asset-pricing relationships, and the sign of the loading on a given factor may switch depending on the combination of factors she has selected. The framework of Proposition 6, therefore, creates a fertile playing field for empirical asset pricing.

2.3.2 The factor zoo in a large economy

We now consider the large version of this economy. Our main results rely on the following eigenvalue decomposition:

$$\frac{1}{N} \boldsymbol{\Phi}' \boldsymbol{\Phi} = \mathbf{Q} \boldsymbol{\Lambda} \mathbf{Q}', \tag{33}$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix with all eigenvalues $\lambda_j > 0$ for $j = 1, \ldots, J$ on its diagonal, and $\mathbf{Q}$ is an orthogonal matrix whose columns are eigenvectors. This decomposition is possible because $\frac{1}{N}\boldsymbol{\Phi}'\boldsymbol{\Phi}$ is symmetric. All eigenvalues of this matrix are strictly positive, and the normalization of Eq. (24) implies that their sum equals 1.[11] We follow Martin and Nagel (2020)'s assumption that each eigenvalue satisfies $\lambda > \varepsilon$, for some uniform constant, ε, as $N \to \infty$. This ensures that the columns of $\boldsymbol{\Phi}$ never become collinear in the limit.

It will prove convenient to rotate the precision matrix, $\boldsymbol{\tau}$, using the matrix of eigenvectors, $\mathbf{Q}$, in Eq. (33) and to work with the limiting behavior of $\mathbf{Q}'\boldsymbol{\tau}/N\mathbf{Q}$, as opposed to $\boldsymbol{\tau}/N$:

$$\boldsymbol{\tau}_\infty \equiv \lim_{J\to\infty, N\to\infty, J/N\to\psi} \mathbf{Q}'\boldsymbol{\tau}/N\mathbf{Q}. \tag{34}$$

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

This limiting precision matrix is a diagonal matrix, with the $j-$th element on its diagonal corresponding to the precision on the $j-$th factor; each of these element is uniquely identified by the eigenvalue λ on this factor according to the cubic relation (see Andrei et al. (2018)):

$$\tau_\infty(\lambda) = \psi\left(\tau_F + \tau_v + \tau_G + \frac{\lambda \tau_m \tau_v^2 \tau_\epsilon^2 \psi}{\gamma^2(\tau_\infty(\lambda) + \lambda\tau_\epsilon)^2}\right). \tag{35}$$

After rotating the precision matrix based on Eq. (33), the precision on each factor is uniquely identified by its eigenvalue. The ratio in Eq. (35) represents the contribution of learning from prices to factor precision. Andrei et al. (2018) note that, because the effect of learning from prices is proportional to ψ^2, if ψ is small, Eq. (34) simplifies to:

$$\boldsymbol{\tau}_\infty = (\tau_F + \tau_v + \tau_G)\psi\mathbf{I} + O(\psi^2), \tag{36}$$

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 simplifies matters.

To test a given factor asset-pricing model, the empirical asset-pricing literature commonly tests the null hypothesis that the intercept of the model is zero. Rejecting this hypothesis means that additional risk factors must be included in the model. By construction adding LMH as a factor in the CAPM relation fully eliminates its intercept; instead adding factors that do not span entirely the zoo will leave a nonzero intercept. To simplify matters, suppose that the empiricist does not add any factor beyond the market to the model. We can test *ex post* how far from zero the intercept on each asset is using the multivariate test of Gibbons, Ross, and Shanken (1989) (GRS, henceforth).

Let the empiricist run the following CAPM regression:

$$\widetilde{\mathbf{R}}_t^e = \widehat{\boldsymbol{\alpha}} + \widehat{\boldsymbol{\beta}}\widetilde{R}_{M,t}^e + \widetilde{\boldsymbol{\epsilon}}_t, \quad t = 1, ..., T, \tag{37}$$

where $\widehat{\boldsymbol{\alpha}}$ denotes the vector of CAPM intercepts on each asset; the "time" index t represents an i.i.d. draw from the model, e.g., throughout the paper we have assumed that the empiricist observes infinitely many draws (realizations) of returns by simulating the model infinitely many times. In our model, even observing infinite time series of returns the empiricist's estimate of the covariance matrix $\widehat{\boldsymbol{\Sigma}}$ does not coincide with that of investors, $\boldsymbol{\Sigma}$; we can thus assume that the empiricist knows $\widehat{\boldsymbol{\Sigma}}$ (she does not need to estimate it) when conducting the test. As a result, she obtains $\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\Sigma}}\mathbf{M}/\widehat{\sigma}_M^2$ and:

$$\widehat{\boldsymbol{\alpha}}_T = \overline{\mathbf{R}}_T^e - \widehat{\boldsymbol{\beta}}\overline{R}_{M,T}^e, \tag{38}$$

where $\overline{\mathbf{R}}^e_T = \frac{1}{T}\sum_{t=1}^T \widetilde{\mathbf{R}}^e_t$ denotes sample average returns and $\overline{R}^e_{M,T} = \mathbf{M}'\overline{\mathbf{R}}^e_T$. As $T \to \infty$ this estimates converges in probability to:

$$\widehat{\boldsymbol{\alpha}} = \boldsymbol{\mu} - \widehat{\boldsymbol{\beta}}\mu_M = (\mathbf{I} - \widehat{\boldsymbol{\beta}}\mathbf{M}')\boldsymbol{\mu}. \tag{39}$$

We want to test whether all elements of $\widehat{\boldsymbol{\alpha}}_T$ are jointly zero, which can be done through the (rescaled) GRS statistic:

$$\text{GRS}_{N,T} = \widehat{\boldsymbol{\alpha}}'_T \mathbb{V}[\widehat{\boldsymbol{\alpha}}_T]^{-1}\widehat{\boldsymbol{\alpha}}_T = T\frac{\theta^{\star}_{N,T} - \theta^M_{N,T}}{1+\theta^M_{N,T}}, \tag{40}$$

where $\theta^{\star}_{N,t} = (\overline{\mathbf{R}}^e_T)'\widehat{\boldsymbol{\Sigma}}^{-1}\overline{\mathbf{R}}^e_T$ denotes the (sample) squared Sharpe ratio on the empiricist's perceived tangency portfolio and $\theta^M_{N,t} = \left(\overline{R}^e_{M,T}/\widehat{\sigma}_M\right)^2$ is the squared Sharpe ratio on the market portfolio. In the words of our model, the GRS statistic captures the squared Sharpe ratio on the LMH portfolio formed ex-post. The magnitude of this statistic, which measures how badly we need to add factors to the CAPM equation, measures equivalently how well the LMH portfolio performs in the model.

Under the empiricist's null hypothesis that the CAPM holds (i.e., the mean of $\widehat{\boldsymbol{\alpha}}_T$ is $\mathbf{0}$), conditional on $\widetilde{R}^e_{M,t}$, $t = 1,...,T$, the GRS statistic is distributed according to:[12]

$$\text{GRS}_{N,T} \sim \chi^2_N \quad \text{under the null.} \tag{41}$$

But, in fact, in the model the mean of $\widehat{\boldsymbol{\alpha}}_T$ is $\widehat{\boldsymbol{\alpha}} \neq \mathbf{0}$ and thus:

$$(\widehat{\boldsymbol{\alpha}}_T - \widehat{\boldsymbol{\alpha}})'\mathbb{V}[\widehat{\boldsymbol{\alpha}}_T]^{-1}(\widehat{\boldsymbol{\alpha}}_T - \widehat{\boldsymbol{\alpha}}) \sim \chi^2_N \quad \text{under the true model.} \tag{42}$$

Note first that, because we treat the covariance matrix of residuals as known, the GRS test in this model corresponds to a stronger version of the Wald test, one in which the distributional properties of the statistic hold even in finite samples (for any T). Note further that in GRS the statistic depends on realizations of market returns under the alternative (but not under the null). In our case, if we assume that the market portoflio is equally weigthed, $\mathbf{M} \equiv \mathbf{1}/N$, the Law of Large Numbers implies that $\overline{R}^e_{M,T} \to \mu_M$ *for any T (i.e., finite or not)*. As a result, in the large-economy limit for any fixed T the GRS statistic is no longer conditional on realizations of market returns. Furthermore, since we want to consider the large-economy limit ($N \to \infty$), following Martin and Nagel (2020), we standardize the GRS

[12]Note that this statistic in Gibbons et al. (1989) is distributed acccording to a non central F distribution; this is because they treat the covariance matrix of residuals as an unknown quantity (distributed according to a Wishart distribution); we instead treat it as a known quantity.

statistic in a way that allows us to apply the Central Limit Theorem:

$$\widehat{W}_{N,T} \equiv \frac{GRS_{N,T} - N}{\sqrt{2N}} \xrightarrow{\text{d}} \mathcal{N}(0,1) \quad \text{under the null.} \tag{43}$$

We can then rewrite the asymptotic distribution of the statistic in the true model as:

$$\widehat{W}_{N,T} + \ell_{N,T} \xrightarrow{\text{d}} \mathcal{N}(0,1), \tag{44}$$

where $\ell_{N,T}$ denotes the mean of the statistic under the true model:

$$\ell_{N,T} = \frac{T}{N^{3/2}} \frac{N}{\sqrt{2}(1+\theta^M_{N,T})} \left((\overline{\mathbf{R}}^e_T)' \left(\widehat{\boldsymbol{\Sigma}}^{-1} - \frac{1}{\widehat{\sigma}^2_M} \mathbf{MM}' \right) \boldsymbol{\mu} + \boldsymbol{\mu}' \left(\widehat{\boldsymbol{\Sigma}}^{-1} - \frac{1}{\widehat{\sigma}^2_M} \mathbf{MM}' \right) (\overline{\mathbf{R}}^e_T - \boldsymbol{\mu}) \right). \tag{45}$$

In words, in the true model $\widehat{W}_{N,T}$ is normally distributed but its mean is not zero.

As $T \to \infty$, $\overline{\mathbf{R}}^e_T \xrightarrow{\text{p}} \boldsymbol{\mu}$; this suggests that if we increase T appropriately fast (specifically, in a way that $T/N^{3/2} \to \varphi > 0$) as we take the large economy limit, the mean of $\widehat{W}_{N,T}$ under the true model will remain finite and the outcome of the test will remain nontrivial. The last step is to compute the mean ℓ_∞ of the statistic in the large-economy limit. To do so we make the following, simplifying assumption.

Assumption 1. *The matrix of loadings, $\boldsymbol{\Phi}$, has IID entries with mean zero, variance one and finite fourth moment.*

Under Assumption 1, eigenvalues λ_j, $j = 1,...,J$ of the loadings matrix follow the Pastur-Marchenko law. The Pastur-Marchenko law determines the demographics of the zoo. This distribution is a function of a single parameter, $J/N \to \psi$; this parameter dictates how dispersed and skewed the zoo is. The relevance of this assumption is that it allows us to characterize the unconditional probability of rejection of the GRS test in terms of the demographics of the zoo (the limiting distribution of the loadings' eigenvalues). The next proposition formalizes the likelihood of rejection in the large-economy limit.

Proposition 8. *(Asymptotic power of the GRS test in a large economy.) Let the market portfolio be equally weighted. In a large economy in which $N, J \to \infty$ with $N \to \infty$ and letting the sample size grow as $T \to N^{3/2}\varphi$, for a critical value c_α the test is rejected with probability:*

$$p - value = \lim_{N,J\to\infty, J/N\to\psi, T/N^{3/2}\to\varphi} \mathbb{P}\left[\widehat{W}_{N,T} < c_\alpha\right] = 1 - \Phi(c - \ell_\infty), \tag{46}$$

where Φ denotes the cdf of the standard normal and $\ell_\infty > 0$ denotes the asymptotic mean of $\widehat{W}_{N,T}$ under the true model.

Under Assumption 1, $\ell_\infty > 0$, and thus the GRS test is always rejected (i.e., p−value < α). This outcome is specific to the zoo shape (dispersion and skewness) implied by the Pastur-Marchenko law. Andrei et al. (2018) show that in a large economy in which eigenvalues are positively skewed and sufficiently dispersed, heterogeneity in predictive power across factors leads to a steeper SML; a similar zoo demographics could also lead the test not to be rejected for certain values of the asymptotic ratio $J/N \to \psi$ of the number of factor to the number of assets. Interestingly, however, Figure 1 shows that the relation between the $p-$value of the test and this ratio is nonmonotonic, and thus that the shape of the zoo does matter for the outcome of the test.

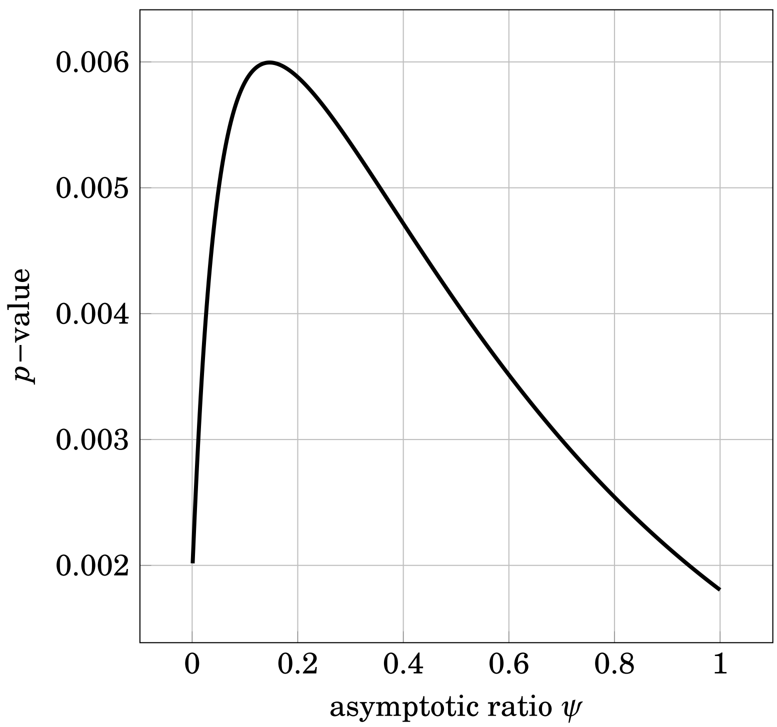

Figure 1: $p-$value of GRS test as a function of the zoo demographics. This figure plots the asymptotic $p-$value of the GRS test as a function of the asymptotic ratio ψ of number of factors to number of assets. The test is at a 1% confidence level, and is based on naive parameter values, $\tau_F = \tau_m = \tau_\epsilon = 1$, $\gamma = 2$ and $\tau_v = 3$.

.

Figure 1 plots the $p-$value of the test as a function of the asymptotic ratio ψ, which under Assumption 1 completely summarizes the demographics of the zoo; it assumes a 1% confidence level for the test and is based on naive parameter values. The GRS test is strongly rejected for values of ψ either close to 0 or 1. When ψ is close to 1, the zoo is populated with a large majority of factors that have negligible predictive power; this implies that many factors in the zoo are approximately collinear. Instead, when ψ is close to zero, the zoo is concentrated and populated by factors that have comparatively high predictive power.

As ψ increases away from 0, the distribution of eigenvalues becomes increasingly skewed with more and more factors having little predictive power. It is in this situation when the distribution of eigenvalues is positively but moderately skewed that the GRS test is close to success. The point is that it is not so much the existence of a factor zoo that is problematic for asset-pricing tests—it is the demographics of the zoo that really matters.

2.4 Discussion of Assumptions

No financial theory is a perfect representation of reality, and ours is no exception. We have made several simplifying assumptions, which have allowed us to tell a cautionary tale on the merits of multifactor models of returns. We now discuss how some of these assumptions can be relaxed.

First, our static model offers the convenience of a closed form solution in Proposition 1. But the same results will hold in a dynamic model, at the expense of losing some tractability. Specifically, in a dynamic model with overlapping generations (Spiegel, 1998; Watanabe, 2008), Eq. (8) becomes

$$\int_i \mathbb{E}_t^i[\widetilde{\mathbf{R}}_{t+1}^e]di = \gamma \mathbf{\Sigma}(\mathbf{M} + \widetilde{\mathbf{m}}_t). \tag{47}$$

As in the static model, this equation shows that expected returns vary both over time and cross-sectionally among agents. This variation in expected returns once again distorts the view of the empiricist, who can write, as in Proposition 2, a two-factor model of returns (although a true unconditional CAPM continues to hold). The dynamic model offers the potential benefit of time variation in γ and/or $\mathbf{\Sigma}$, which will further distort empiricist's view through well-known conditional effects (Jagannathan and Wang, 1998).[13]

Second, we do not impose in our baseline model any assumptions of the vector of unconditional market weights, $\mathbf{M}$, except that all its components are strictly positive. Assuming a link between $\mathbf{M}$ and the vector of sensitivities $\mathbf{\Phi}$ offers new insights. One can show that in an economy where small firms tend to have high exposure to $\widetilde{F}$, the empiricist observes a negative alpha for small-growth firms—a puzzling observation, according to Fama and French (1993, 1996, 2015). In more extreme cases, the empiricist can observe a downward sloping Security Market Line (see also Andrei et al., 2018).

Third, the information structure of our baseline model is stylized, but any information structure ultimately boils down to the equilibrium condition (8). This condition leads to the CAPM rejection and to the two-factor model of Proposition 2. With a general information

[13] Hasler and Martineau (2019a,b) provide recent evidence for the conditional relationship (47). See also Boguth, Carlson, Fisher, and Simutin (2011).

structure, however, the task of interpreting the LMH portfolio and linking it to firm characteristics becomes more difficult, as we have shown in Section 2.3.

Fourth, it is not crucial for our argument that the empiricist knows *less* or *more* than investors. What matters is that the empiricist has a *different* information set; this is what leads to the CAPM rejection. For instance, in our baseline model, let us assume that the empiricist knows the average expected excess returns on all assets, $\bar{\boldsymbol{\mu}}$. Because prices are only imperfect aggregators of information, individual agents cannot possibly know $\bar{\boldsymbol{\mu}}$. Thus, neither the empiricist nor individual agents knows more than the other. The law of total covariance from the perspective of the empiricist writes:

$$\widehat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma} + \mathrm{Var}[\bar{\boldsymbol{\mu}}] + \frac{K^2 \tau_v}{\tau^2} \boldsymbol{\Phi}\boldsymbol{\Phi}', \tag{48}$$

where the last term is the *disagreement* matrix (Banerjee, 2010). Unless the empiricist observes the subjective beliefs of individual investors, this term is unobservable, and the CAPM is rejected. Moreover, even if the empiricist could observe returns in continuous time—arguably a difficult task—the last term in Eq. (48) does not vanish.

3 Empirical Illustration

We have shown theoretically that the empiricist can always build a two-factor model that explains 100% of the cross section of realized excess returns. The CAPM rejection is a sufficient condition for this result. Although we have started from the premise that the CAPM is rejected by mistake, this condition is not necessary; regardless of the reason for which the CAPM is rejected, the empiricist fails to reject a two-factor model of returns.

The tautological nature of this result is a shaky base for empirical asset pricing. Any multi-factor model that is used to explain the cross section of returns may come uncomfortably close to this tautology. In this section, we offer an empirical illustration using the 25 size and book/market sorted portfolios (Fama and French, 1993). We emphasize that our purpose is *not* to find a better factor of returns, nor to dismiss the previously found factors. Instead, we try to understand how close to the LMH portfolio can empiricist get *without perfect hindsight*, i.e., by forming LMH based on data that is available at the time of portfolio formation. Equally importantly, we attempt to evaluate empirically the theoretical implications of the previous section.

Figure 2 presents mean excess returns and volatilities for 25 size and book/market sorted portfolios, in monthly data from 7/1963 to 12/2018, 666 observations. Panel (a) plots these portfolios in a mean-standard deviation diagram, together with the full-sample efficient fron-

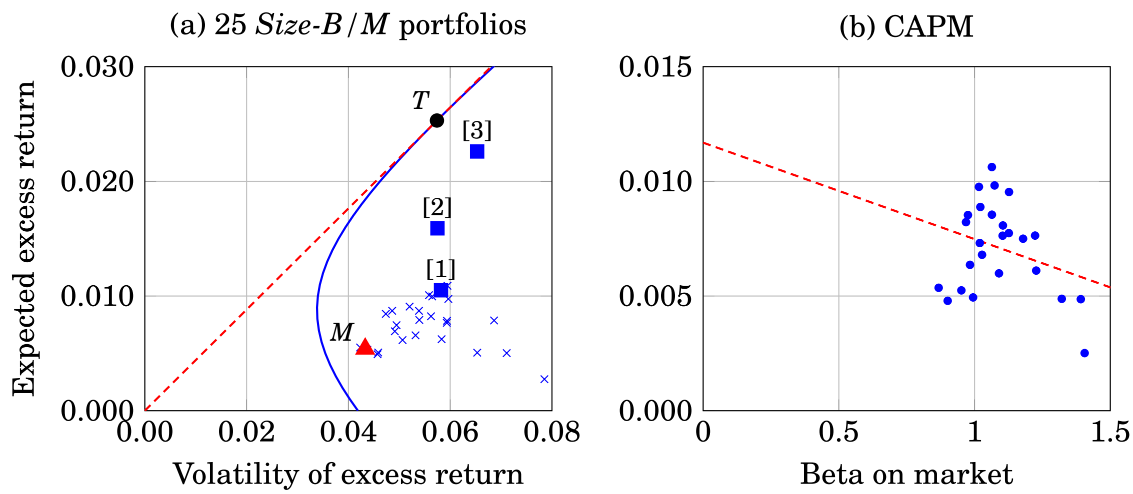

Figure 2: **25 *Size-B/M* portfolios and the CAPM**
Panel (a): Monthly mean excess returns and volatilities for 25 *Size-B/M* sorted portfolios (Fama and French, 1993), July 1963-December 2018, 666 months. The solid line is the (ex-post) minimum-variance frontier and the dashed line is the CML. The squares labeled [1], [2], and [3] are defined in the text. Panel (b): Mean excess returns against beta on market (the market portfolio is depicted in panel (a) with the red triangle labeled M). The dashed line is the SML, which has an intercept of 0.0117 (t-stat 3.01), a slope of -0.0042 (t-stat -1.00), and a coefficient R^2 of 4.8%.

tier. The triangle labeled M is the market portfolio, whereas the dot labeled T is the tangency portfolio, computed based on the entire sample. Panel (b) illustrates the empirical failure of the CAPM. It plots the expected excess returns of the above 25 portfolios against market betas. Return and beta are not related as the CAPM suggests—the Security Market Line (SML) has a negative slope and a statistically significant positive intercept (numbers are provided in the caption of the figure).

The rejection of the CAPM can also be interpreted in terms of the geometric distance between the points M and T. It is clearly apparent from panel (a) that this distance is large: the expected excess return of portfolio T is orders of magnitude larger than the expected excess return of the market. Thus, a test based on this distance (Gibbons et al., 1989) will likely reject the CAPM, which panel (b) confirms.

The rejection of the CAPM, as illustrated above, has led to a decades-long quest for alternative factors. This can be best understood in the familiar mean-variance diagram: when factors are long-short portfolios with zero net investment—e.g., “Small-Minus-Big” (SMB), “High-Minus-Low” (HML)—one can simply add these portfolios to the market and obtain a *tilted* portfolio, with the hope that the resulting portfolio does a better job at explaining asset returns. Consider, for instance, the square labeled [1] in panel (a) of Figure 2. It represents the tilted portfolio that results from adding the Fama and French (1993) SMB and HML factors to the market. The square labeled [2] adds two more factors, "Robust-Minus-Weak" (RMW) and "Conservative-Minus-Aggressive" (CMA) (Fama and French, 2015). Finally, the square labeled [3] further adds the momentum factor (MOM) (Carhart, 1997). As more factors are added, there is a clear tendency of the tilted portfolio to move towards the portfolio with the maximum Sharpe ratio attainable theoretically, T. One may therefore think of the search for factors as an effort to improve the Sharpe ratio.

Instead of trying to move gradually towards the tangency portfolio T, we follow the theoretical result of Proposition 2 and start directly from T. Let M and T be the vectors of portfolio weights for the market and the tangency portfolios, and consider the *Low-Minus-High* portfolio of deviations between T and M, $\Delta = T - M$. Let the covariance matrix of excess returns for the 25 *Size-B/M* portfolios be $\widehat{\Sigma}$ and their expected excess returns be μ. Proposition 2 then implies

$$\mu = \frac{\mu_T \widehat{\sigma}_M^2}{\widehat{\sigma}_T^2}\widehat{\beta} + \frac{\mu_T \widehat{\sigma}_\Delta^2}{\widehat{\sigma}_T^2}\widehat{\beta}_\Delta, \tag{49}$$

where $\mu_T \equiv \mu' T$, $\widehat{\sigma}_M^2 \equiv M'\widehat{\Sigma}M$, $\widehat{\sigma}_T^2 \equiv T'\widehat{\Sigma}T$, $\widehat{\sigma}_\Delta^2 \equiv \Delta'\widehat{\Sigma}\Delta$, $\widehat{\beta} \equiv \widehat{\Sigma}M/\widehat{\sigma}_M^2$, and $\widehat{\beta}_\Delta \equiv \widehat{\Sigma}\Delta/\widehat{\sigma}_\Delta^2$.

Two sets of betas, one computed on the market portfolio M and one computed on the Δ portfolio, explain ex-post 100% of the variation in expected excess returns. A similar point has been previously made by Roll (1977, p. 138, emphasis his): "there will always be *some* portfolio which is ex-post efficient and will bring about exact observed linearity among ex-post sample mean returns and ex-post sample betas." In our context, this portfolio is T. Notice, however, that we are not interested in computing betas based on T. Instead, we compute two sets of betas ($\widehat{\beta}$ and $\widehat{\beta}_\Delta$), because this allows us to understand how the CAPM is rejected when considering the market and the LMH portfolios together.

According to Eq. (49), assets earn a positive risk premium $\mu_T\widehat{\sigma}_M^2/\widehat{\sigma}_T^2$ per unit of $\widehat{\beta}$ on the market and a positive risk premium $\mu_T\widehat{\sigma}_\Delta^2/\widehat{\sigma}_T^2$ per unit of $\widehat{\beta}_\Delta$ on the Δ portfolio.[14] Interestingly, there is a positive relationship between $\widehat{\beta}_\Delta$ and Δ:

$$\widehat{\beta}_\Delta = \bar{\beta}_\Delta + \frac{1}{\Delta'\Delta}\Delta + u, \tag{50}$$

where $\bar{\beta}_\Delta$ is the arithmetic average over $\widehat{\beta}_\Delta$ and u is uncorrelated with Δ.[15] Together with

[14]The two sets of betas in Eq. (49) result from two univariate first-pass regressions, as opposed to one multivariate regression of asset returns on the market and the Δ portfolio. A multivariate regression will yield linear combinations of $\widehat{\beta}$ and $\widehat{\beta}_\Delta$, with different slopes, but same R^2 (see Proposition 9).

[15]Eq. (50) results from a least squares calculation. Consider the $N \times 2$ matrix $X = [\mathbf{1}_N \Delta]$ and compute the estimated coefficients as $(X'X)^{-1}X'\widehat{\beta}_\Delta$. Since Δ is a long-short portfolio, we have $\mathbf{1}_N'\Delta = 0$ which simplifies the algebra and yields (50). Notice also that the arithmetic average $\bar{\beta}_\Delta$ is different from 1 (instead, the weighted average $\Delta'\widehat{\beta}_\Delta$ equals one).

Eq. (49), Eq. (50) implies that under-valued assets—in empiricist's view—earn a positive risk premium relatively to over-valued assets, risk premium which is not explained by the market factor. This result hints at a possible positive correlation between the returns of the the value factor and the returns of the LMH portfolio.

The relation (49) is tautological in the sense that it is uninformative about the validity of the CAPM. The only assumptions necessary for Proposition 2 are that the covariance matrix $\widehat{\Sigma}$ is non-singular and that at least one asset has a different sample mean return from others—which, incidentally, are assumptions (A.1) and (A.2) in Roll (1977). Aside from these two assumptions, we do not need to impose assumptions about investor preferences or to characterize the dynamics of assets' excess returns.

How close to the portfolio Δ can an empiricist get without perfect hindsight? To answer this question, we build a time-varying LMH portfolio that is not based on the full sample of returns, but on the data that is available at the time of portfolio formation only. Doing so, we build the LMH portfolio in real time, and this portfolio is thus as much "ex-ante" as any other risk factor.

The procedure is as follows. At every time t starting from 6/1963, we use the past 30 years of monthly excess returns up to and including time t, to compute mean excess returns, μ_t, and the covariance matrix of excess returns, $\widehat{\Sigma}_t$. Relying on efficient set mathematics, we compute the tangency portfolio $T_t = \widehat{\Sigma}_t^{-1}\mu_t/B_t$, where $B_t \equiv \mathbf{1}_N'\widehat{\Sigma}_t^{-1}\mu_t$ and $N = 25$. Then we compute Δ_t as the difference between T_t and the mean market capitalizations for the 25 portfolios over the past 30 years of monthly data.[16] Using Δ_t, we compute the one-month ahead excess return of the low-minus-high portfolio (from t to $t+1$). This yields a time series of monthly excess returns from 7/1963 to 12/2018. This approach of building the LMH factor, which draws directly from mean-variance theory, optimally uses the covariance information from past returns, as opposed to the common practice of creating factor portfolios by sorting on characteristics (see also Daniel, Mota, Rottke, and Santos, 2017).

During the period from 7/1963 to 12/2018, the LMH portfolio produces a mean monthly excess return of 1.49% and a monthly standard deviation of 7.6%. The annualized Sharpe ratio for the monthly returns of the LMH portfolio is thus 0.68. The correlations of the LMH portfolio returns with the Fama and French (2015) five factors (MKT, HML, SMB, RMW, CMA) and the Carhart (1997) momentum factor (MOM) are shown in Table 1 (Pearson product-moment correlations below-diagonal; Spearman rank correlations above-diagonal). The LMH portfolio returns are strongly negatively correlated with the MKT factor returns and strongly positively correlated with the HML and CMA factor returns.

[16] We choose a window of 30 years of monthly data in order to reliably estimate the covariance matrix $\widehat{\Sigma}$. Our results are robust to this choice, as long as we use more than 20 years of past data. We have also estimated the covariance matrix using a DCC-GARCH model, with similar results.

	LMH	MKT	HML	SMB	RMW	CMA	MOM
LMH		-0.40	0.45	-0.06	0.06	0.38	0.11
MKT	-0.40		-0.24	0.26	-0.19	-0.32	-0.09
HML	0.46	-0.24		-0.14	-0.20	0.68	-0.16
SMB	-0.02	0.27	-0.19		-0.27	-0.16	0.01
RMW	0.08	-0.21	0.06	-0.39		-0.20	0.17
CMA	0.43	-0.37	0.70	-0.17	-0.04		-0.08
MOM	0.17	-0.14	-0.19	0.00	0.11	-0.03	

Table 1: **Correlations**
This table presents the time-series correlations between the returns of the LMH portfolio and returns of the MKT, HML, SMB, RMW, CMA, and MOM factors, July 1963-December 2018, 666 months. The below-diagonal entries show Pearson product-moment correlations. The above-diagonal entries show Spearman rank correlations.

In Table 2, we regress the LMH portfolio returns on the Fama and French (2015) five factors and the MOM returns. Columns (1) to (6) in Table 2 show that none of these factors is able to explain the returns of the LMH portfolio, whose risk-adjusted alphas range between 1.03% and 1.85% per month, with t-statistics ranging from 3.89 to 6.48. Regressing the LMH portfolio returns on all the six factors' returns (column 7) indicates that the LMH portfolio has a risk-adjusted alpha of 0.94% per month with a Newey and West (1987) t-statistic of 3.74, adjusted for six autocorrelation lags. The returns of the LMH portfolio, therefore, are not explained by exposure to the MKT, HML, SMB, RMW, CMA, or MOM factors. We also notice the strong negative sensitivity of the LMH portfolio to the MKT factor (-0.69, with a t-statistic of -7.04) and the strong positive sensitivity to the HML factor (1.23, with a t-statistic of 6.28) and the CMA factor (1.62, with a t-statistic of 8.61). Finally, the sensitivity to the MOM factor is 0.31, with a significant t-statistic of 2.40.

Table 3 presents the results of regressions of the Fama-French-Carhart six factors on the LMH portfolio. Although in Table 2 the alpha of the LMH portfolio returns remains strongly economically and statistically significant, and is thus unexplained by any of these factors (separately or together), Table 3 indicates that, with the exception of the market factor and the momentum factor, the alphas of all other factors become either statistically insignificant (HML, SMB, CMA) or weakly economically significant (RMW). In particular, the alpha from regressing the HML factor returns on the LMH portfolio returns is 0.07% per month, and the alpha of the CMA factor is 0.11% per month. These values are small in both practical and statistical terms.

We take the results of Tables 2 and 3 as evidence for our theoretical implication that the LMH portfolio should be strongly positively related to the value factor and to the investment factor. Furthermore, Table 3 shows that the LMH portfolio is able to price other factors, in

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Intercept	0.0185	0.0109	0.0150	0.0142	0.0103	0.0128	0.0094
	(6.48)	(3.89)	(4.74)	(4.80)	(3.98)	(4.47)	(3.75)
MKT	-0.6937						-0.4812
	(-7.04)						(-4.97)
HML		1.2319					1.0812
		(6.28)					(6.40)
SMB			-0.0502				0.3952
			(-0.28)				(2.82)
RMW				0.2714			0.1374
				(0.80)			(0.55)
CMA					1.6169		0.3026
					(8.61)		(1.29)
MOM						0.3130	0.3759
						(2.40)	(3.86)
Adj. R^2	0.1575	0.2070	-0.0011	0.0046	0.1814	0.0284	0.3569
Obs.	666	666	666	666	666	666	666

Table 2: **Regression results: LMH on the Fama-French-Carhart six factors**
This table presents the results of regressions of returns of the LMH portfolio on the MKT, HML, SMB, RMW, CMA, and MOM factors, July 1963-December 2018, 666 months. The columns labeled (1), (2), (3), (4), (5), and (6) present results for univariate specifications using only MKT, HML, SMB, MOM, RMW, and CMA, respectively, as the independent variable. The column labeled (7) presents results from the multivariate specification using all six factors as independent variables. t-statistics, adjusted following Newey and West (1987) using six lags, are presented in parentheses.

	(1) MKT	(2) HML	(3) SMB	(4) RMW	(5) CMA	(6) MOM
Intercept	0.0086	0.0007	0.0022	0.0022	0.0011	0.0052
	(5.87)	(0.72)	(1.78)	(2.61)	(1.53)	(2.71)
LMH	-0.2288	0.1690	-0.0082	0.0224	0.1129	0.0954
	(-7.49)	(7.21)	(-0.28)	(0.83)	(6.88)	(2.17)
Adj. R^2	0.1575	0.2070	-0.0011	0.0046	0.1814	0.0284
Obs.	666	666	666	666	666	666

Table 3: **Regression results: Fama-French-Carhart six factors on LMH**
This table presents the results of regressions of returns of the MKT, HML, SMB, RMW, CMA, and MOM factors on the LMH portfolio, July 1963-December 2018, 666 months. t-statistics, adjusted following Newey and West (1987) using six lags, are presented in parentheses.

particular HML and CMA. Put differently, following logic from Barillas and Shanken (2017), the HML and CMA factors are redundant for describing average returns, as they appear to be explained by exposure to the LMH portfolio alone.

Before testing Proposition 2, we also verify the prediction of Eq. (50), which states that assets deemed by the empiricist as being in low-demand (Δ^{+} assets) have a higher β_{Δ}. Indeed, the correlation between β_{Δ} and the average of Δ over the full sample ($\bar{\Delta}$) is 54%. Regressing β_{Δ} on $\bar{\Delta}$ yields a positive slope of 0.083 (t-statistic 2.94), but not statistically different from $(\bar{\Delta}'\bar{\Delta})^{-1} = 0.12$, as Eq. (50) predicts. Consistent with Eq. (50), low-demand assets have higher β_{Δ} which, according to Proposition 2, should command a positive risk premium.

The following Proposition transforms Eq. (49) into an equivalent relation that is testable using standard two-stage multivariate regressions.

Proposition 9. *Eq.* (49) *can be equivalently written as*

$$\mu = \mu_M \widetilde{\beta} + \mu_{\Delta} \widetilde{\beta}_{\Delta}, \tag{51}$$

where $\mu_M \equiv \mu' M$, $\mu_{\Delta} \equiv \mu' \Delta$, and $\widetilde{\beta}$ and $\widetilde{\beta}_{\Delta}$ are jointly estimated from multivariate time-series regressions of assets' excess returns on the excess returns of the market and of the LMH portfolio. Furthermore, $\widetilde{\beta}$ and $\widetilde{\beta}_{\Delta}$ are linear combinations of $\widehat{\beta}$ and $\widehat{\beta}_{\Delta}$:

$$\widetilde{\beta} = \frac{\widehat{\sigma}_{\Delta}^2}{\widehat{\sigma}_{\Delta}^2 \widehat{\sigma}_M^2 - \widehat{\sigma}_{M\Delta}^2} (\widehat{\sigma}_M^2 \widehat{\beta} - \widehat{\sigma}_{M\Delta} \widehat{\beta}_{\Delta}) \quad \textit{and} \quad \widetilde{\beta}_{\Delta} = \frac{\widehat{\sigma}_M^2}{\widehat{\sigma}_{\Delta}^2 \widehat{\sigma}_M^2 - \widehat{\sigma}_{M\Delta}^2} (\widehat{\sigma}_{\Delta}^2 \widehat{\beta}_{\Delta} - \widehat{\sigma}_{M\Delta} \widehat{\beta}), \tag{52}$$

where $\widehat{\sigma}_{M\Delta}$ is the covariance between the returns of the market portfolio and the returns of the LMH portfolio.

Proof. See Appendix A.8. □

Testing Proposition 2 is therefore equivalent to testing Eq. (51). We first regress the time series returns of each of the 25 *Size-B/M* portfolios on the excess returns of MKT and LMH. Table 4 shows the intercepts and the two factor slopes for these time-series regressions. Interestingly, none of the intercepts, which range from -0.26% to 0.21%, are statistically significant, with t-stats between -1.30 and 1.71. This is the case even for extreme growth stocks, which are a typical problem for traditional factor models (Fama and French, 2015). The slopes of the market factor are positive and strongly statistically significant, with t-stats ranging from 27.75 to 66.71. Small stocks tend to have higher betas than large stocks, whereas there is no clear pattern on the B/M dimension. The slopes on the LMH portfolio are mostly negative for the extreme growth stocks and positive for the extreme value stocks, and a large majority are statistically significant. Because the average excess return on the LMH portfolio is positive, the negative slopes of growth stocks lowers their average excess returns and the positive slopes of value stocks increases their average excess returns. Finally, the R^2 coefficients of the 25 regressions (unreported here) range from 0.60 to 0.89.

$$R_n(t) = \alpha_n + \widetilde{\beta}_n R_{MKT}(t) + \widetilde{\beta}_{\Delta,n} R_{LMH}(t) + \epsilon_n(t), \quad n = 1, \ldots, 25$$

	Low	2	3	4	High	Low	2	3	4	High
			$\alpha_n(\%)$					t-stat(α_n)		
Small	-0.26	0.06	0.07	0.19	0.21	-1.30	0.32	0.47	1.35	1.42
2	-0.04	0.08	0.13	0.10	0.07	-0.29	0.62	1.21	1.00	0.56
3	-0.07	0.08	0.08	0.12	0.20	-0.57	0.87	0.92	1.31	1.63
4	0.05	0.04	0.01	0.15	0.08	0.56	0.51	0.14	1.71	0.68
Big	0.07	0.06	-0.03	-0.06	0.07	1.08	1.02	-0.45	-0.61	0.53
			$\widetilde{\beta}_n$					t-stat($\widetilde{\beta}_n$)		
Small	1.32	1.25	1.15	1.11	1.18	27.75	29.33	33.63	33.42	33.38
2	1.32	1.20	1.13	1.11	1.24	36.56	40.73	42.26	43.94	38.87
3	1.27	1.17	1.06	1.06	1.16	42.10	52.54	49.44	47.69	39.45
4	1.19	1.09	1.08	1.03	1.16	51.74	60.39	54.36	49.41	40.69
Big	0.96	0.94	0.91	0.93	1.00	64.48	66.71	48.86	39.86	32.13
			$\widetilde{\beta}_{\Delta,n}$					t-stat($\widetilde{\beta}_{\Delta,n}$)		
Small	-0.12	0.04	0.06	0.14	0.16	-4.53	1.47	3.25	7.27	7.95
2	-0.11	0.03	0.09	0.13	0.16	-5.21	1.79	5.78	9.21	8.57
3	-0.07	0.06	0.06	0.12	0.12	-4.13	4.41	5.05	9.40	7.10
4	-0.05	-0.01	0.07	0.09	0.08	-3.51	-0.52	6.19	7.51	4.89
Big	-0.05	-0.02	0.06	0.04	0.03	-5.94	-2.22	5.84	2.67	1.58

Table 4: **Regressions for 25 *Size-B/M* portfolios**
This table shows two-factor intercepts, slopes for MKT and LHM, and t-statistics for these coefficients, July 1963-December 2018, 666 months. The two-factor regression equation is provided above the table. t-statistics, adjusted following Newey and West (1987) using three lags, are presented in parentheses.

We then use the estimates from Table 4 in cross-sectional regressions of average excess returns on betas. For the sake of comparison, column (1) of Table 5 shows a direct test of the CAPM. When betas on the market are used alone, we obtain the typical failure of the CAPM: a strong positive intercept and a slope not significantly different from zero, in this case slightly negative (see also panel (b) in Figure 2).

Column (2) presents the direct test of Proposition 9. When both betas on the market and on the LMH portfolio are used as explanatory variables, the slope on β_{MKT} changes sign from negative to positive, although it remains statistically insignificant (t-stat 0.91). The intercept is virtually zero, the slope on β_{LMH} is strongly statistically significant (t-stat 4.56), and the two-factor model explains 89% of the variation in average returns. The values of the two slopes are to be compared with their historical counterparts: over the period

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Intercept	0.0117	0.0005	0.0122	-0.0007	0.0098	0.0007	0.0028	0.0007
	(3.01)	(0.10)	(4.63)	(-0.18)	(3.31)	(0.17)	(0.66)	(0.18)
β_{MKT}	-0.0042	0.0049	-0.0067	0.0064	-0.0047	0.0049	0.0025	0.0048
	(-1.00)	(0.92)	(-2.16)	(1.43)	(-1.37)	(1.13)	(0.55)	(1.13)
β_{LMH}		0.0241		0.0284		0.0263		0.0264
		(4.57)		(5.56)		(6.25)		(5.96)
β_{HML}			0.0034	0.0033	0.0030	0.0032	0.0033	0.0032
			(3.09)	(2.91)	(2.70)	(2.88)	(2.98)	(2.87)
β_{SMB}			0.0015	0.0021	0.0020	0.0021	0.0021	0.0021
			(1.23)	(1.72)	(1.63)	(1.75)	(1.71)	(1.75)
β_{RMW}					0.0049	0.0012	0.0057	0.0011
					(2.88)	(0.71)	(2.83)	(0.60)
β_{CMA}					-0.0007	-0.0003	-0.0013	-0.0003
					(-0.38)	(-0.19)	(-0.63)	(-0.18)
β_{MOM}							0.0254	0.0015
							(3.47)	(0.18)
Adj. R^2	0.0479	0.8852	0.6272	0.8898	0.7270	0.8900	0.8126	0.8835
GLS R^2	0.1087	0.6457	0.2838	0.6595	0.3352	0.6781	0.4636	0.6801
Obs.	25	25	25	25	25	25	25	25

Table 5: **Cross-sectional regressions: Average excess returns on factor betas**
This table presents the results of regressions of average excess returns for 25 size and book/market sorted portfolios (Fama and French, 1993), on various combinations of seven factors: MKT, LMH, HML, SMB, RMW, CMA, and MOM, July 1963-December 2018, 666 months. Standard errors and t-statistics (presented in parentheses) are corrected for cross-sectional correlations in alphas and for errors in estimating betas (Shanken, 1992).

7/1963-12/2018, the average monthly excess return on the market was 0.52% and on the LMH portfolio 1.49%. The slope on β_{MKT} (0.49%) is thus not significantly different from the historical market risk premium of 0.52% (t-stat -0.06), whereas the slope on β_{LMH} (2.41%) is marginally statistically different from 1.49% (t-stat 1.75). The Wald test that the intercept equals zero, the first slope equals 0.52%, and the second slope equals 1.49% is not rejected at 5% significance level (p-value 35%), suggesting that the two-factor model in column (2) is a reasonable description of expected returns.

Figure 3, panel (a) depicts the performance of the CAPM and panel (b) depicts the performance of the two-factor specification MKT+LMH. The vertical axis plots the unconditional expected excess returns for the 25 portfolios, whereas the horizontal axis plots the predicted values from columns (1) and (2) of Table 5. The points lie closely to a 45° line (dashed line) only in panel (b). Overall, column (2) of Table 5 and panel (b) of Figure 3 show support for Proposition 9: the market portfolio and the LMH portfolio explain a significant fraction of cross-sectional variation in expected returns and thus the two-factor model (51) does a good

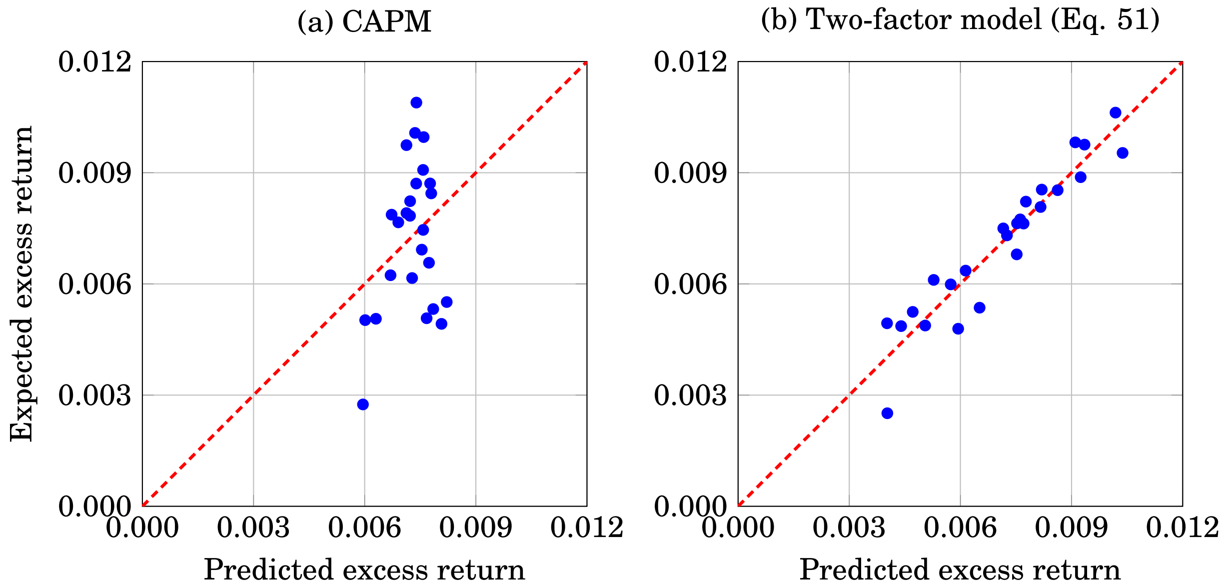

Figure 3: **Average excess returns vs. prediction**
Expected excess monthly returns for the 25 *Size-B/M* portfolios (y-axis) vs predicted monthly excess returns (x-axis) from the regression specifications in columns (1) and (2) of Table 5, July 1963-December 2018, 666 months.

job at capturing the returns of the 25 *Size-B/M* portfolios.

Column (3) in Table 5 shows the results for the Fama and French (1993) three-factor model, whereas column (4) adds the LMH portfolio. The striking difference between columns (3) and (4) is the sign of the slope with respect to the market factor. Whereas in column (3) the estimated premium is negative and statistically significant, in column (4) the estimated premium is positive and close to the average excess market return of 0.52%. Yet, the intercept is strongly statistically significant for the Fama and French (1993) model, and the R^2 coefficient is considerably lower than in column (3). This suggests that the LMH portfolio acts like an omitted variable in column (3), distorting regression coefficients. Indeed, as discussed in Section 2.3, the LMH portfolio will always be able to explain the remaining variation in returns, and thus acts as an omitted variable in cross-sectional regressions.

Column (5) shows the results for the Fama and French (2015) five-factor model, whereas column (6) adds the LMH portfolio. Once again, the market estimated premium changes sign and the alpha becomes statistically insignificant when adding LMH. Columns (7) and (8) further add the momentum factor. Curiously, when the LMH is added to the regression (column 8), the momentum factor loses significance. This suggests that the LMH portfolio may be able to explain momentum returns, a finding that we will verify later in this section. Finally, the slope on the LMH portfolio returns is strongly statistically significant in all the specifications (t-stats ranging from 4.57 to 6.25), whereas with the exception of the HML factor, all other factors lose their statistical significance in presence of LMH.

In the mean-variance space high cross-sectional R^2s are not necessarily indicative of a good fit (Roll and Ross, 1994; Kandel and Stambaugh, 1995). As proposed by Lewellen et al. (2010), we also report GLS R^2s, which measure the proximity of a given model's portfolio to the minimum-variance frontier (mean-variance efficiency is obtained when the GLS R^2 is 1). As Table 5 shows, according to this metric, the two-factor model of column (2) is the closest to mean-variance efficiency (GLS R^2 0.65) when compared to the Fama and French (1993) three-factor model (GLS R^2 0.28), to the Fama and French (2015) five-factor model (GLS R^2 0.34), and the Fama and French (2015)-Carhart (1997) six-factor model (GLS R^2 0.46).

We turn to the ability of the the LMH portfolio to price additional portfolios sorts, such as portfolios based on past performance, on alternative price multiples, and on investment. An interesting result of this section is the relatively high positive correlation between the returns of the LMH portfolio and the returns of the momentum factor (0.17, Table 1). Although the LMH portfolio did not entirely eliminate the alpha of the momentum factor (Table 3, column 6), it did eliminate its statistical significance in cross-sectional regressions (Table 5).

That LMH partially eliminates the statistical significance of momentum suggests that the LMH portfolio could price stocks sorted based on past performance, and Table 6 confirms. In columns (1)-(3) of panel A, we present cross-sectional regressions on 10 portfolios sorted based on their past performance (from Professor Ken French's website). The LMH portfolio explains, together with the market, 95% of the cross-section of returns. Its estimated premium is positive and statistically significant (0.1058, with a t-stat of 2.22), and the intercept of the regression is not statistically different from zero. In contrast, the intercept of the Fama and French (1993) three-factor model is economically large and strongly statistical significant (0.0251, with a t-stat of 3.17), and none of the three factors are statistically significant. Overall, columns (1)-(3) of Table 6, panel A, attest to the ability of the LMH portfolio to price stocks sorted based on momentum.

Columns (4)-(9) of panel A further show results when stocks are sorted based on short-term reversal and long-term reversal. While in both cases the CAPM performs particularly badly, the LMH portfolio earns a positive risk premium, statistically significant only in one case. The Fama and French (1993) three-factor model performs relatively better in these portfolio sorts. Finally, as shown in panel B, the LMH portfolio performs extremely well in sorts based on investment, earnings/price ratios, or cashflow/price ratios. In particular, we notice that the intercepts are all statistically indistinguishable from zero when the MKT and LMH are both used in the regressions.

Overall, the powerful explanatory power of the LMH portfolio in the 25 *Size*-B/M portfolio set carries over to other portfolio sorts. But, as the results above show, the LMH portfolio

Panel A: Sorts involving prior returns

	Momentum			ST Reversal			LT Reversal		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Intercept	0.0170	-0.0035	0.0251	0.0022	-0.0048	-0.0003	0.0052	0.0035	0.0195
	(4.75)	(-0.42)	(3.17)	(0.68)	(-0.81)	(-0.03)	(1.68)	(0.97)	(1.47)
MKT	-0.0101	0.0114	-0.0169	0.0030	0.0099	0.0049	0.0010	0.0023	-0.0136
	(-2.55)	(1.32)	(-1.96)	(0.82)	(1.63)	(0.57)	(0.29)	(0.60)	(-1.04)
LMH		0.1058			0.0442			0.0124	
		(2.22)			(2.15)			(1.62)	
HML			-0.0093			0.0200			-0.0016
			(-1.44)			(2.20)			(-0.40)
SMB			0.0036			-0.0011			0.0064
			(0.61)			(-0.23)			(1.29)
Adj. R^2	0.1817	0.9502	0.8351	-0.0251	0.1233	0.7987	-0.1076	0.7976	0.9063
Obs.	10	10	10	10	10	10	10	10	10

Panel B: Sorts involving Investment, E/P, and CF/P

	Investment			Earnings/Price			Cashflow/Price		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Intercept	0.0096	-0.0042	0.0104	0.0102	-0.0006	0.0088	0.0114	-0.0009	0.0115
	(3.59)	(-0.73)	(1.54)	(2.27)	(-0.14)	(0.99)	(2.47)	(-0.17)	(1.54)
MKT	-0.0037	0.0096	-0.0048	-0.0042	0.0064	-0.0031	-0.0055	0.0067	-0.0058
	(-1.19)	(1.61)	(-0.70)	(-0.88)	(1.29)	(-0.35)	(-1.14)	(1.17)	(-0.76)
LMH		0.0441			0.0239			0.0252	
		(2.59)			(2.66)			(2.19)	
HML			0.0028			0.0023			0.0012
			(1.31)			(1.44)			(0.72)
SMB			0.0038			0.0071			0.0087
			(1.27)			(1.56)			(2.06)
Adj. R^2	0.0784	0.6816	0.4318	-0.0532	0.9071	0.8698	0.0585	0.6843	0.8627
Obs.	10	10	10	10	10	10	10	10	10

Table 6: **Regression results: Various portfolio sorts**
This table presents the results of regressions of average excess returns for six different portfolio sorts on various combinations of four factors: MKT, LMH, HML, and SMB, July 1963-December 2018, 666 months. t-statistics, adjusted following Shanken (1992), are presented in parentheses.

is hardly a panacea. We emphasize, nevertheless, that the LMH portfolio need not be built exclusively from the 25 *Size-B/M* portfolio set. One can in fact build one LMH portfolio for each of the portfolio sorts considered here. Any combination of these portfolios—always a zero-investment portfolio—would yield a global LMH portfolio.

In sum, we have shown theoretically that, when the CAPM fails, a low-minus-high portfolio explains, together with the market, 100% of the cross-sectional variation in returns. We have confirmed empirically these properties of the LMH portfolio within the Fama and

French (1993) *Size-B/M* portfolio space. We emphasize that this factor appears significant *regardless* of the reason behind the CAPM rejection. In particular, if the CAPM is rejected by mistake—because the empiricist does not use the correct market portfolio (Roll, 1977), or has information that differs from that of investors (Andrei et al., 2018)—then the LMH portfolio becomes a proxy for this mistake. In this case, not only will the empiricist reject the CAPM, but she will also fail to reject a multifactor model. We have illustrated this possibility in an equilibrium model of stock returns in Section 2, and in this section have found empirical support for it. The credibility of the claim "anomalies are evidence against the CAPM" stands on slippery ground.

4 Conclusion

Empirical asset pricing has identified hundreds of anomalies, and interpreted them as evidence against the CAPM. Admittedly, their sheer number makes the case against the theory compelling. Yet, we argue that anomalies are not, by themselves, evidence that the theory is wrong. On the contrary, their large number could be the greatest weakness of the case against theory.

Regardless of whether the CAPM is rejected for valid reasons or by mistake, finding anomalies is the unavoidable symptom of the rejection. Theoretically, one can always build a long-short portfolio that explains, together with the market, the cross section of returns. The moment a factor based on firm characteristics or on macroeconomic fundamentals covaries with this portfolio, it becomes an anomaly, although in fact it need not be. This situation resembles a statistical mirage in which empiricists are lured into accepting multifactor models of returns, and theorists feel compelled to interpret these models. This hypothetical state of affairs raises legitimate concerns regarding p-hacking (Simmons, Nelson, and Simonsohn, 2011; Chordia et al., 2017) and HARKing (Kerr, 1998).

More problematically, though, anomalies are silent about the true cause of the CAPM rejection. Because the potential number of anomalies is unlimited, multifactor models of returns do not reveal the true reasons for the difference between the theory and the data. In our theoretical model, for instance, the CAPM is rejected by mistake. But even if the CAPM is rejected for valid reasons, finding anomalies will likely not identify these reasons. Anomalies are, at best, uninformative and, at worst, misleading.

Several fascinating questions arise. Can our economic setup identify instrumental variables that would help improve cross-sectional asset-pricing tests? What does our model say about the return of anomalies on days with public announcements (Savor and Wilson, 2014; Engelberg et al., 2018) or during non-trading hours (Hendershott, Livdan, and Rösch, 2018)?

Is there a possible theoretical link between the return of the LMH portfolio and momentum? We leave these questions for future research, and conclude with William F. Sharpe's words in response to the Fama and French (1992) empirical results (according to Eric Berg of *The New York Times*, February 18, 1992, emphasis ours):

> "It is a remarkable set of empirical results about what happened *in the past*, but I am not willing to make investment decisions based on the theory that there is no relationship between beta, *properly measured*, and expected returns."
>
> — *William F. Sharpe*

References

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

Albagli, E., C. Hellwig, and A. Tsyvinski (2011a). Information aggregation, investment, and managerial incentives. Technical report, National Bureau of Economic Research.

Albagli, E., C. Hellwig, and A. Tsyvinski (2011b). A theory of asset pricing based on heterogeneous information. Technical report, National Bureau of Economic Research.

Andrei, D., J. Cujean, and M. I. Wilson (2018). The lost capital asset pricing model.

Andrei, D., W. Mann, and N. Moyen (2018). Why did the q theory of investment start working? *Journal of Financial Economics Forthcoming*.

Bai, J., T. Philippon, and A. Savov (2016). Have financial markets become more informative? *Journal of Financial Economics 122*(3), 625–654.

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.

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

Barillas, F. and J. Shanken (2017). Which alpha? *The Review of Financial Studies 30*(4), 1316–1338.

Barillas, F. and J. Shanken (2018). Comparing asset pricing models. *The Journal of Finance 73*(2), 715–754.

Berk, J. B. (1995). A critique of size-related anomalies. *The Review of Financial Studies 8*(2), 275–286.

Black, F. and M. Scholes (1974). From theory to a new financial product. *The Journal of Finance 29*(2), 399–412.

Blume, M. E. and I. Friend (1973). A new look at the capital asset pricing model. *The journal of finance 28*(1), 19–34.

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

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

Bryzgalova, S. (2015). Spurious factors in linear asset pricing models. *LSE manuscript 1*.

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

Campbell, J. Y. (2017). *Financial decisions and markets: A course in asset pricing*. Princeton University Press.

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of finance 52*(1), 57–82.

Chen, A. Y. and T. Zimmermann (2018). Publication bias and the cross-section of stock returns.

Chinco, A., A. Neuhierl, and M. Weber (2019). Estimating the anomaly baserate. *Chicago Booth Research Paper* (19-10).

Chordia, T., A. Goyal, and A. Saretto (2017). p-hacking: Evidence from two million trading strategies.

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

Cochrane, J. H. (2009). *Asset Pricing:(Revised Edition)*. Princeton university press.

Daniel, K., L. Mota, S. Rottke, and T. Santos (2017). The cross-section of risk and return. Technical report, National Bureau of Economic Research.

David, J. M., H. A. Hopenhayn, and V. Venkateswaran (2016). Information, misallocation, and aggregate productivity. *The Quarterly Journal of Economics 131*(2), 943–1005.

DeGroot, M. H. (2005). *Optimal statistical decisions*, Volume 82. John Wiley & Sons.

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

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 (1992). The cross-section of expected stock returns. *the Journal of Finance 47*(2), 427–465.

Fama, E. F. and K. R. French (1993). Common risk factors in the returns on stocks and bonds. *Journal of financial economics 33*(1), 3–56.

Fama, E. F. and K. R. French (1996, March). Multifactor Explanations of Asset Pricing Anomalies. *Journal of Finance 51*(1), 55–84.

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

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

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.

Feng, G., S. Giglio, and D. Xiu (2019). Taming the factor zoo: A test of new factors. Technical report, National Bureau of Economic Research.

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

Ferson, W. E., S. Sarkissian, and T. Simin (1999). The alpha factor asset pricing model: A parable. *Journal of Financial Markets 2*(1), 49–68.

Gibbons, M. R., S. A. Ross, and J. Shanken (1989). A test of the efficiency of a given portfolio. *Econometrica: Journal of the Econometric Society*, 1121–1152.

Giglio, S., Y. Liao, and D. Xiu (2018). Thousands of alpha tests. *Working Paper* (18-09).

Giglio, S. and D. Xiu (2018). Asset pricing with omitted factors. *Working Paper* (16-21).

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

Harvey, C. R. (2017). Presidential address: The scientific outlook in financial economics. *The Journal of Finance 72*(4), 1399–1440.

Harvey, C. R. and Y. Liu (2018). Lucky factors. *Available at SSRN 2528780*.

Harvey, C. R., Y. Liu, and H. Zhu (2016). ... and the cross-section of expected returns. *The Review of Financial Studies 29*(1), 5–68.

Hasler, M. and C. Martineau (2019a). The capm holds. Technical report, UT Dallas and University of Toronto.

Hasler, M. and C. Martineau (2019b). Does the capm predict returns? Technical report, UT Dallas and University of Toronto.

Hayashi, F. (1982). Tobin's marginal q and average q: A neoclassical interpretation. *Econometrica: Journal of the Econometric Society*, 213–224.

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

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

Hou, K., C. Xue, and L. Zhang (2015). Digesting anomalies: An investment approach. *The Review of Financial Studies 28*(3), 650–705.

Hou, K., C. Xue, and L. Zhang (2018). Replicating anomalies. *The Review of Financial Studies*.

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.

Jorgenson, D. W. (1963). Capital theory and investment behavior. *American Economic Review (Papers and Proceedings) 53*(2), 247–59.

Kan, R. and C. Zhang (1999). Two-pass tests of asset pricing models with useless factors. *the Journal of Finance 54*(1), 203–235.

Kandel, S. and R. F. Stambaugh (1995). Portfolio inefficiency and the cross-section of expected returns. *The Journal of Finance 50*(1), 157–184.

Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review 2*(3), 196–217.

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

Lakonishok, J. and A. C. Shapiro (1986). Systematic risk, total risk and size as determinants of stock market returns. *Journal of Banking & Finance 10*(1), 115–132.

Lewellen, J., S. Nagel, and J. Shanken (2010). A skeptical appraisal of asset pricing tests. *Journal of Financial economics 96*(2), 175–194.

Linnainmaa, J. T. and M. R. Roberts (2018). The history of the cross-section of stock returns. *The Review of Financial Studies 31*(7), 2606–2649.

Lucas, R. E. J. and E. C. Prescott (1971). Investment under uncertainty. *Econometrica: Journal of the Econometric Society*, 659–681.

MacKinlay, C. and L. Pástor (2000). Asset pricing models: Implications for expected returns and portfolio selection. *The Review of financial studies 13*(4), 883–916.

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

McLean, R. D. and J. Pontiff (2016). Does academic research destroy stock return predictability? *The Journal of Finance 71*(1), 5–32.

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

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

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

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

Reinganum, M. R. (1981). A new empirical perspective on the capm. *Journal of financial and quantitative analysis 16*(4), 439–462.

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

Romano, J. P. and M. Wolf (2005). Stepwise multiple testing as formalized data snooping. *Econometrica 73*(4), 1237–1282.

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.

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

Simmons, J. P., L. D. Nelson, and U. Simonsohn (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological science 22*(11), 1359–1366.

Smith, S. (2018). Learning, time-variation, and the factor zoo. *Available at SSRN 3286613*.

Spiegel, M. (1998). Stock price volatility in a multiple security overlapping generations model. *Rev 11*, 419–447.

Tobin, J. (1969). A general equilibrium approach to monetary theory. *Journal of money, credit and banking 1*(1), 15–29.

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

Watanabe, M. (2008, February). Price volatility and investor behavior in an overlapping generations model with information asymmetry. *Journal of Finance 63*(1), 229–272.

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.

Yan, X. and L. Zheng (2017). Fundamental analysis and the cross-section of stock returns: A data-mining approach. *The Review of Financial Studies 30*(4), 1382–1423.

A Appendix

A.1 Proof of Proposition 1

We solve for a linear equilibrium of the economy in which market-to-book ratios satisfy

$$\frac{P}{K} = \alpha_0 F + \xi_0 M + \alpha \widetilde{F} + g\widetilde{G} + \xi \widetilde{m}, \tag{A.1}$$

where α_0, α, and g are N-dimensional vectors and ξ_0 and ξ are $N \times N$ matrices, all of which will be determined in equilibrium by imposing the market clearing condition (7).

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

$$\mathscr{F}_i = \{V_i, \widetilde{G}, P\}. \tag{A.2}$$

It will be convenient to isolate the informational part of market-to-book ratios by writing

$$\frac{P^a}{K} \equiv \frac{P}{K} - \alpha_0 F - \xi_0 M - g\widetilde{G} = \alpha \widetilde{F} + \xi \widetilde{m}, \tag{A.3}$$

This equation shows that each market-to-book ratio is a noisy signal on $\widetilde{F}$. The precision of each one of these signals is endogenously determined in equilibrium.

We use the *Projection Theorem* (see, e.g., DeGroot, 2005), which we restate here for convenience.

Projection Theorem. *Consider the n-dimensional normal random variable*

$$(\theta, s) \sim \mathcal{N}\left(\begin{bmatrix} \mu_\theta \\ \mu_s \end{bmatrix}, \begin{bmatrix} \Sigma_{\theta,\theta} & \Sigma_{\theta,s} \\ \Sigma_{s,\theta} & \Sigma_{s,s} \end{bmatrix} \right). \tag{A.4}$$

Provided $\Sigma_{s,s}$ is non-singular, the conditional density of θ given s is normal with conditional mean and conditional variance-covariance matrix:

$$\mathbb{E}[\theta|s] = \mu_\theta + \Sigma_{\theta,s}\Sigma_{s,s}^{-1}\left(s - \mu_s\right) \tag{A.5}$$

$$\mathrm{Var}[\theta|s] = \Sigma_{\theta,\theta} - \Sigma_{\theta,s}\Sigma_{s,s}^{-1}\Sigma_{s,\theta}. \tag{A.6}$$

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

$$S_i = \begin{bmatrix} P^a/K \\ \widetilde{V}_i \\ \widetilde{G} \end{bmatrix} = \begin{bmatrix} \alpha \\ 1 \\ 1 \end{bmatrix} \widetilde{F} + \begin{bmatrix} \xi & \mathbf{0}_{N\times 1} & \mathbf{0}_{N\times 1} \\ \mathbf{0}_{1\times N} & 1 & 0 \\ \mathbf{0}_{1\times N} & 0 & 1 \end{bmatrix} \begin{bmatrix} \widetilde{m} \\ \tilde{v}_i \\ \tilde{v} \end{bmatrix} \equiv H\widetilde{F} + \Theta \begin{bmatrix} \widetilde{m} \\ \tilde{v}_i \\ \tilde{v} \end{bmatrix}, \tag{A.7}$$

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

$$\Sigma = \begin{bmatrix} \tau_m^{-1}\mathbf{I}_N & \mathbf{0}_{N\times 1} & \mathbf{0}_{N\times 1} \\ \mathbf{0}_{1\times N} & \tau_v^{-1} & 0 \\ \mathbf{0}_{1\times N} & 0 & \tau_G^{-1} \end{bmatrix}. \tag{A.8}$$

We define

$$r \equiv (\Theta\Sigma\Theta')^{-1} = \begin{bmatrix} \tau_m(\xi\xi')^{-1} & \mathbf{0}_{N\times 1} & \mathbf{0}_{N\times 1} \\ \mathbf{0}_{1\times N} & \tau_v & 0 \\ \mathbf{0}_{1\times N} & 0 & \tau_G \end{bmatrix}, \tag{A.9}$$

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

$$\tau \equiv \mathrm{Var}[\widetilde{F}|\mathscr{F}_i]^{-1} = \tau_F + H'rH = \tau_F + \tau_G + \tau_v + \tau_m \alpha'(\xi\xi')^{-1}\alpha. \tag{A.10}$$

To obtain (A.10), first replace $\Sigma_{\theta,\theta} = 1/\tau_F$, $\Sigma_{\theta,s} = H'/\tau_F$, $\Sigma_{s,\theta} = H/\tau_F$, and $\Sigma_{s,s} = HH'/\tau_F + \Theta\Sigma\Theta'$ in Eq. (A.6), then use the Woodbury matrix identity.

The precision τ is the same across investors. Furthermore, investor i's expectation of $\widetilde{F}$ satisfies:

$$\mathbb{E}[\widetilde{F}|\mathscr{F}_i] = \frac{1}{\tau}H'rS_i = \frac{1}{\tau}\begin{bmatrix}\tau_m \alpha'(\xi\xi')^{-1} & \tau_v & \tau_G\end{bmatrix} S_i. \tag{A.11}$$

To obtain (A.11), start from (A.5):

$$\mathbb{E}[\widetilde{F}|\mathscr{F}_i] = \frac{H'}{\tau_F}\left(\frac{HH'}{\tau_F} + \Theta\Sigma\Theta'\right)^{-1} S_i \tag{A.12}$$

$$= \frac{1}{\tau_F}H'r - \frac{1}{\tau\tau_F}H'rHH'r, \tag{A.13}$$

and replace $H'rH = \tau - \tau_F$ in the last term on the right hand side. Replacing S_i in (A.11) yields

$$\mathbb{E}[\widetilde{F}|\mathscr{F}_i] = \frac{1}{\tau}\left[(\tau - \tau_F - \tau_G)\widetilde{F} + \tau_G G + \tau_m \alpha'(\xi\xi')^{-1}\xi\widetilde{m} + \tau_v v_i\right], \tag{A.14}$$

where we have used the definition of the total precision (A.10) for the term that multiplies $\widetilde{F}$. It follows that average market expectation of future payoffs is

$$\bar{\mathbb{E}}[D] \equiv \int_i \mathbb{E}[D|\mathscr{F}_i]di = K\Phi F + K\Phi\frac{1}{\tau}\left[(\tau - \tau_F - \tau_G)\widetilde{F} + \tau_G G + \tau_m \alpha'(\xi\xi')^{-1}\xi\widetilde{m}\right], \tag{A.15}$$

and the covariance matrix of future payoffs is

$$\Sigma \equiv \mathrm{Var}[D|\mathscr{F}_i] = K^2\left(\frac{1}{\tau}\Phi\Phi' + \frac{1}{\tau_\epsilon}\mathbf{I}_N\right). \tag{A.16}$$

The market-clearing condition (7) implies

$$P = \bar{\mathbb{E}}[D] - \gamma\Sigma(M + \widetilde{m}). \tag{A.17}$$

Thus

$$\frac{P}{K} = \Phi F + \Phi\frac{1}{\tau}\left[(\tau - \tau_F - \tau_G)\widetilde{F} + \tau_G G + \tau_m(\xi^{-1}\alpha)'\widetilde{m}\right] - \gamma\frac{\Sigma}{K}(M + \widetilde{m}) \tag{A.18}$$

where we have used the simplification $\alpha'(\xi\xi')^{-1}\xi = (\xi^{-1}\alpha)'$. This yields

$$\alpha_0 = \Phi, \quad \xi_0 = -\gamma\frac{\Sigma}{K}, \quad \alpha = \Phi\frac{\tau - \tau_F - \tau_G}{\tau}, \quad g = \Phi\frac{\tau_G}{\tau}, \tag{A.19}$$

and

$$\xi = \Phi\frac{\tau_m}{\tau}(\xi^{-1}\alpha)' - \gamma K\left(\frac{1}{\tau}\Phi\Phi' + \frac{1}{\tau_\epsilon}\mathbf{I}_N\right). \tag{A.20}$$

Multiply both sides of Eq. (A.20) by $\xi^{-1}\alpha$ (to the right):

$$\alpha = \Phi\frac{\tau_m}{\tau}(\xi^{-1}\alpha)'\xi^{-1}\alpha - \gamma K\left(\frac{1}{\tau}\Phi\Phi' + \frac{1}{\tau_\epsilon}\mathbf{I}_N\right)\xi^{-1}\alpha, \tag{A.21}$$

and then recognize that $\tau_m(\xi^{-1}\alpha)'\xi^{-1}\alpha = \tau_m\alpha'(\xi\xi')^{-1}\alpha = \tau - \tau_F - \tau_G - \tau_v$ (from Eq. A.10), which can be replaced above, together with the solution for α to obtain (after multiplication with τ):

$$\Phi\tau_v = -\gamma K\left(\Phi\Phi' + \frac{\tau}{\tau_\epsilon}\mathbf{I}_N\right)\xi^{-1}\alpha, \tag{A.22}$$

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

$$\xi^{-1}\alpha = -\frac{\tau_v}{\gamma K}\left(\Phi\Phi' + \frac{\tau}{\tau_\epsilon}\mathbf{I}_N\right)^{-1}\Phi. \tag{A.23}$$

Multiply both sides with Φ' (to the left):

$$\Phi'\xi^{-1}\alpha = -\frac{\tau_v}{\gamma K}\Phi'\left(\Phi\Phi' + \frac{\tau}{\tau_\epsilon}\mathbf{I}_N\right)^{-1}\Phi = -\frac{\tau_v\tau_\epsilon\Phi'\Phi}{\gamma K(\tau + \tau_\epsilon\Phi'\Phi)}, \tag{A.24}$$

where the second equality follows from the Woodbury matrix identity. Conjecture

$$\xi^{-1}\alpha \equiv -\frac{\sqrt{\tau_P}}{\sqrt{\tau_m}}\Phi, \tag{A.25}$$

where τ_P is an unknown positive scalar. Replacing Eq. (A.25) in Eq. (A.10) yields the total precision τ as a function of this scalar:

$$\tau = \tau_F + \tau_G + \tau_v + \tau_P\Phi'\Phi. \tag{A.26}$$

Furthermore, replacing the conjecture (A.25) in Eq. (A.24) yields

$$\frac{\sqrt{\tau_P}}{\sqrt{\tau_m}} = \frac{\tau_v\tau_\epsilon}{\gamma K(\tau + \tau_\epsilon\Phi'\Phi)} \tag{A.27}$$

which leads to a cubic equation in τ_P:

$$\tau_P\left[\tau_F + \tau_v + \tau_G + (\tau_P + \tau_\epsilon)\Phi'\Phi\right]^2 = \frac{\tau_m\tau_\epsilon^2\tau_v^2}{\gamma^2K^2}. \tag{A.28}$$

The discriminant of this equation is strictly negative and thus it 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. The conjecture (A.25) can now be replaced in (A.20) to obtain the undetermined coefficients ξ:

$$\xi = -\frac{\gamma K + \sqrt{\tau_m\tau_P}}{\tau}\Phi\Phi' - \frac{\gamma K}{\tau_\epsilon}\mathbf{I}_N. \tag{A.29}$$

This completes the proof of Proposition 1. □

A.2 Proof of Corollary 1.3

For this proof we will make the following assumptions:

Assumption 2. *There is no ex-ante proportionality relation between the unconditional market portfolio M and the vector of firms' loadings on the common productivity factor Φ.*

Assumption 2 ensures that we keep the setup as general as possible, excluding pathological cases with an *exogenous* perfect relationship between firms' market capitalizations and their exposure to the common productivity factor.

Assumption 3. $M'\Phi > 0$.

Assumption 3 eliminates the uninteresting case $M'\Phi = 0$ (zero market exposure to the common factor), and is without loss of generality (if $M'\Phi < 0$, one can simply switch the sign of the common factor).

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

$$\|x\|\|y\| \geq x'y = M'\widehat{\Sigma}^{1/2}\widehat{\Sigma}^{-1/2}\mu = \mu_M. \tag{A.30}$$

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

$$\frac{\mu_M}{\sigma_M} \leq \sqrt{\mu'\widehat{\Sigma}^{-1}\mu}. \tag{A.31}$$

The relation (A.31) holds with equality if and only if x is proportional to y, or

$$\mu \propto \widehat{\Sigma}M. \tag{A.32}$$

To show that the proportionality relation (A.32) is cannot hold for the empiricist, we first compute $\widehat{\Sigma}$ by using the law of total variance:

$$\begin{aligned}\widehat{\Sigma} &= \Sigma + \text{Var}[\mathbb{E}[D - P|\mathscr{F}_i]] && \text{(A.33)}\\ &= \Sigma + \text{Var}\left[\bar{\mathbb{E}}[D] + K\Phi\frac{\tau_v}{\tau}v_i\right] && \text{(A.34)}\\ &= \Sigma + \frac{\gamma^2}{\tau_m}\Sigma^2 + \frac{K^2\tau_v}{\tau^2}\Phi\Phi' && \text{(A.35)}\end{aligned}$$

Replace (A.16) above to obtain

$$\widehat{\Sigma} = \left(\frac{K^2(\tau+\tau_v)}{\tau^2} + \frac{K^4\gamma^2\Phi'\Phi}{\tau^2\tau_m} + \frac{2K^4\gamma^2}{\tau\tau_m\tau_\epsilon}\right)\Phi\Phi' + \left(\frac{K^2}{\tau_\epsilon} + \frac{K^4\gamma^2}{\tau_m\tau_\epsilon^2}\right)\mathbf{I}_N. \tag{A.36}$$

Replace $\mathbf{I}_N$ from (A.16) or $\Phi\Phi'$ from (A.16) to write $\widehat{\Sigma}$ in two equivalent forms:

$$\begin{aligned}\widehat{\Sigma} &= c_1\Sigma + c_2\Phi\Phi' && \text{(A.37)}\\ \widehat{\Sigma} &= c_3\Sigma - c_4\mathbf{I}_N && \text{(A.38)}\end{aligned}$$

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

$$c_1 = 1 + \frac{K^2\gamma^2}{\tau_m\tau_\epsilon} > 0, \quad c_2 = \frac{K^2\tau_v}{\tau^2} + \frac{\gamma^2K^4(\tau+\tau_\epsilon\Phi'\Phi)}{\tau^2\tau_m\tau_\epsilon} > 0, \tag{A.39}$$

and

$$c_3 = 1 + \frac{\tau_v}{\tau} + \frac{K^2\gamma^2(2\tau + \tau_\epsilon \Phi'\Phi)}{\tau\tau_m\tau_\epsilon} > 0, \quad c_4 = \frac{K^2\tau_v}{\tau\tau_\epsilon} + \frac{\gamma^2 K^4(\tau + \tau_\epsilon\Phi'\Phi)}{\tau\tau_m\tau_\epsilon^2} > 0. \tag{A.40}$$

Multiply Equations (A.37)-(A.38) with M:

$$\widehat{\Sigma}M = c_1\Sigma M + c_2(\Phi' M)\Phi \tag{A.41}$$

$$\widehat{\Sigma}M = c_3\Sigma M - c_4 M. \tag{A.42}$$

Since ΣM and μ are proportional (Corollary 1.1), (A.32) implies that $\mu \propto \Phi$ and $\mu \propto M$. This implies $M \propto \Phi$, contradicting Assumption 2, which completes the proof of Corollary 1.3. □

A.3 Proof of Proposition 3

Start from the true CAPM relation (Corollary 1.1):

$$\mu = \frac{\mu_M}{\sigma_M^2}\Sigma M, \tag{A.43}$$

and replace Σ from (A.37):

$$\mu = \frac{\mu_M \widehat{\sigma}_M^2}{c_1\sigma_M^2}\beta - \frac{\mu_M c_2 M'\Phi}{c_1\sigma_M^2}\Phi. \tag{A.44}$$

where $\widehat{\sigma}_M^2 \equiv M'\widehat{\Sigma}M$. We further know that average market-to-book ratios are

$$\frac{\mathbb{E}[P]}{K} = \Phi F - \frac{\mu}{K}, \tag{A.45}$$

from which we can replace Φ in (A.44) and solve for μ. This yields

$$\mu = \lambda_1\beta + \lambda_2\frac{\mathbb{E}[P]}{K}, \tag{A.46}$$

with

$$\lambda_1 = \frac{FK\mu_M\widehat{\sigma}_M^2}{c_1FK\sigma_M^2 + c_2\mu_M M'\Phi} \quad \text{and} \quad \lambda_2 = -\frac{c_2K\mu_M M'\Phi}{c_1FK\sigma_M^2 + c_2\mu_M M'\Phi}. \tag{A.47}$$

Assumption 3 ensures that $\lambda_1 > 0$ and $\lambda_2 < 0$. This completes the proof of Proposition 3. □

A.4 Proof of Proposition 4

Replacing Φ from (21) in (A.44) yields (22), with

$$\eta_0 = -\frac{c_2M'\Phi\mu_M}{c_1F\sigma_M^2} < 0, \quad \eta_1 = \frac{\mu_M\widehat{\sigma}_M^2}{c_1\sigma_M^2} > 0, \quad \text{and} \quad \eta_2 = -\frac{ac_2M'\Phi\mu_M}{c_1F\sigma_M^2} < 0. \tag{A.48}$$

This completes the proof of Proposition 4. □

A.5 Proof of Proposition 5

This appendix solves an extension of the model when assets' payoffs are driven by multiple factors. We start by conjecturing that market-to-book ratios satisfy

$$\frac{P}{K} = \alpha_0 F + \xi_0 M + \alpha \widetilde{F} + g\widetilde{G} + \xi \widetilde{m}. \tag{A.49}$$

Since agents observe $\widetilde{G}$, M, and F the effective price signal is

$$\frac{P^a}{K} \equiv \frac{P}{K} - g\widetilde{G} - \alpha_0 F - \xi_0 M = \alpha \widetilde{F} + \xi \widetilde{m}. \tag{A.50}$$

Regrouping all signals in a vector we obtain

$$S^i = \begin{bmatrix} P^a/K \\ \widetilde{V}_i \\ \widetilde{G} \end{bmatrix} = \begin{bmatrix} \alpha \\ \mathbf{I}_J \\ \mathbf{I}_J \end{bmatrix} \widetilde{F} + \begin{bmatrix} \xi & \mathbf{0}_{N\times J} & \mathbf{0}_{N\times J} \\ \mathbf{0}_{J\times N} & \mathbf{I}_J & \mathbf{0}_{J\times J} \\ \mathbf{0}_{J\times N} & \mathbf{0}_{J\times J} & \mathbf{I}_J \end{bmatrix} \begin{bmatrix} \widetilde{m} \\ \tilde{v}_i \\ \tilde{v} \end{bmatrix} \tag{A.51}$$

with

$$\begin{bmatrix} \widetilde{m} \\ \tilde{v}_i \\ \tilde{v} \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \tau_m^{-1}\mathbf{I}_N & \mathbf{0}_{N\times J} & \mathbf{0}_{N\times J} \\ \mathbf{0}_{J\times N} & \tau_v^{-1}\mathbf{I}_J & \mathbf{0}_{J\times J} \\ \mathbf{0}_{J\times N} & \mathbf{0}_{J\times J} & \tau_G^{-1}\mathbf{I}_J \end{bmatrix}\right). \tag{A.52}$$

Using these matrices we now define a $(N+2J)\times(N+2J)$ matrix:

$$R \equiv (\Theta\Sigma\Theta')^{-1} = \begin{bmatrix} (\xi\xi')^{-1}\tau_m & \mathbf{0}_{N\times J} & \mathbf{0}_{N\times J} \\ \mathbf{0}_{J\times N} & \mathbf{I}_J\tau_v & \mathbf{0}_{J\times J} \\ \mathbf{0}_{J\times N} & \mathbf{0}_{J\times J} & \mathbf{I}_J\tau_G \end{bmatrix}. \tag{A.53}$$

The projection theorem then implies that

$$\tau \equiv \mathrm{Var}[\widetilde{F}|\mathscr{F}_i]^{-1} = \mathbf{I}_J\tau_F + H'RH = (\tau_F + \tau_v + \tau_G)\mathbf{I}_J + \alpha'(\xi\xi')^{-1}\alpha\tau_m. \tag{A.54}$$

The projection theorem also yields

$$\mathbb{E}[\widetilde{F}|\mathscr{F}_i] = \tau^{-1}H'RS_i = \tau^{-1}\left[\tau_m\alpha'(\xi\xi')^{-1}(\alpha\widetilde{F} + \xi\widetilde{m}) + \tau_v(\widetilde{F} + \tilde{v}_i) + \tau_G(\widetilde{F} + \tilde{v})\right] \tag{A.55}$$

$$= \tau^{-1}\left[(\tau - \tau_F\mathbf{I}_J - \tau_G\mathbf{I}_J)\widetilde{F} + \tau_m\alpha'(\xi\xi')^{-1}\xi\widetilde{m} + \tau_v\tilde{v}_i + \tau_G\widetilde{G}\right]. \tag{A.56}$$

It follows that average expectations of future payoffs satisfy

$$\bar{\mathbb{E}}[\widetilde{D}] \equiv \int_i \mathbb{E}[\widetilde{D}|\mathscr{F}_i]\mathrm{d}i = \Phi F + \Phi\tau^{-1}\left[(\tau - \tau_F\mathbf{I}_J - \tau_G\mathbf{I}_J)\widetilde{F} + \tau_m\alpha'(\xi\xi')^{-1}\xi\widetilde{m} + \tau_G\widetilde{G}\right]. \tag{A.57}$$

and the covariance matrix of future payoffs satisfies:

$$\Sigma \equiv \mathrm{Var}[\widetilde{D}|\mathscr{F}_i] = K^2(\Phi\tau^{-1}\Phi' + \tau_\epsilon^{-1}\mathbf{I}_N). \tag{A.58}$$

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

$$\frac{P}{K} = \Phi F + \Phi\tau^{-1}\left[(\tau - \tau_F\mathbf{I}_J - \tau_G\mathbf{I}_J)\widetilde{F} + \tau_m(\xi^{-1}\alpha)'\widetilde{m} + \tau_G\widetilde{G}\right] - \gamma\frac{\Sigma}{K}(M + \widetilde{m}). \tag{A.59}$$

Separating variables we obtain the following system of equations:

$$\alpha_0 = \Phi, \quad \xi_0 = -\gamma\frac{\Sigma}{K}, \quad g = \tau_G\Phi\tau^{-1}, \quad \alpha = \Phi\tau^{-1}(\tau - \tau_F\mathbf{I}_J - \tau_G\mathbf{I}_J), \tag{A.60}$$

and

$$\xi = \tau_m\Phi\tau^{-1}(\xi^{-1}\alpha)' - \gamma K\left(\Phi\tau^{-1}\Phi' + \tau_\epsilon^{-1}\mathbf{I}_N\right). \tag{A.61}$$

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

$$\alpha = \tau_m\Phi\tau^{-1}(\xi^{-1}\alpha)'\xi^{-1}\alpha - \gamma K\left(\Phi\tau^{-1}\Phi' + \tau_\epsilon^{-1}\mathbf{I}_N\right)\xi^{-1}\alpha \tag{A.62}$$

Observing that $\tau_m\Phi\tau^{-1}(\xi^{-1}\alpha)'\xi^{-1}\alpha = \Phi\tau^{-1}[\tau - (\tau_F + \tau_G + \tau_v)\mathbf{I}_J] \equiv \alpha - \tau_v\Phi\tau^{-1}$, we obtain

$$\tau_v\Phi\tau^{-1} = -\gamma K\left(\Phi\tau^{-1}\Phi' + \tau_\epsilon^{-1}\mathbf{I}_N\right)\xi^{-1}\alpha, \tag{A.63}$$

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

$$\xi^{-1}\alpha = -\frac{\tau_v}{\gamma K}\left(\Phi\tau^{-1}\Phi' + \tau_\epsilon^{-1}\mathbf{I}_N\right)^{-1}\Phi\tau^{-1}. \tag{A.64}$$

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

$$\tau^{-1}\Phi'\left(\Phi\tau^{-1}\Phi' + \tau_\epsilon^{-1}\mathbf{I}_N\right)^{-1}\Phi\tau^{-1} = \tau^{-1} - (\tau + \tau_\epsilon\Phi'\Phi)^{-1} \tag{A.65}$$

to conclude that

$$\tau^{-1}\Phi'\xi^{-1}\alpha = -\frac{\tau_v}{\gamma K}\left(\tau^{-1} - (\tau + \tau_\epsilon\Phi'\Phi)^{-1}\right). \tag{A.66}$$

Conjecture that $\xi^{-1}\alpha \equiv -\frac{1}{\sqrt{\tau_m}}\Phi\tau_P$, where τ_P is a $J \times J$ symmetric matrix of $J(J+1)/2$ unknown coefficients. Replacing this conjecture in the expression for total precision in Eq. (A.54):

$$\tau \equiv (\tau_F + \tau_v + \tau_G)\mathbf{I}_J + \tau_P\Phi'\Phi\tau_P. \tag{A.67}$$

Further replacing the conjecture in Eq. (A.66) produces a matrix equation for τ_P:

$$\tau^{-1}\Phi'\Phi\tau_P = \sqrt{\tau_m}\frac{\tau_v}{\gamma K}\left(\tau^{-1} - (\tau + \tau_\epsilon\Phi'\Phi)^{-1}\right), \tag{A.68}$$

which, premultiplying by τ, can be rewritten as

$$\Phi'\Phi\tau_P = \sqrt{\tau_m}\frac{\tau_v}{\gamma K}\left(\mathbf{I}_J - (\mathbf{I}_J + \tau_\epsilon\tau^{-1}\Phi'\Phi)^{-1}\right). \tag{A.69}$$

Once we have a solution for τ_P we can subsitute the conjecture we obtain the matrix ξ as:

$$\xi = -\sqrt{\tau_m}\Phi\tau^{-1}\tau_P'\Phi' - \gamma K\left(\Phi\tau^{-1}\Phi' + \tau_\epsilon^{-1}\mathbf{I}_N\right). \tag{A.70}$$

completing the proof of Proposition 5. □

A.6 Proof of Proposition 6

As in the single factor case, the market-clearing condition implies

$$\bar{\mathbb{E}}[R^e] = \gamma\Sigma(M + \widetilde{m}), \tag{A.71}$$

a relation that can be represented in the traditional CAPM form. To construct this relation as measured by the empiricist, we use the law of total covariance:

$$\widehat{\Sigma} = \Sigma + \text{Var}\left[\bar{\mathbb{E}}[R] + \tau_v \Phi\tau^{-1} v_i\right] = \Sigma + \frac{\gamma^2}{\tau_m}\Sigma^2 + \tau_v \Phi\tau^{-1}\tau^{-1}\Phi', \tag{A.72}$$

where the second line follows from substituting the market-clearing condition above. Furthermore, the conditional covariance matrix of returns satisfies:

$$\Sigma = K^2\left(\Phi\tau^{-1}\Phi' + \tau_\epsilon^{-1}\mathbf{I}_N\right). \tag{A.73}$$

Substituting one relation into the other we obtain

$$\widehat{\Sigma} = \Sigma + \frac{\gamma^2 K^4}{\tau_m}(\Phi\tau^{-1}\Phi' + \tau_\epsilon^{-1}\mathbf{I}_N)(\Phi\tau^{-1}\Phi' + \tau_\epsilon^{-1}\mathbf{I}_N) + \tau_v \Phi\tau^{-1}\tau^{-1}\Phi' \tag{A.74}$$

$$= \left(1 + \frac{\gamma^2 K^2}{\tau_m \tau_\epsilon}\right)\Sigma + \frac{\gamma^2 K^4}{\tau_m \tau_\epsilon}\Phi\tau^{-1}\Phi' + \Phi\tau^{-1}\left(\frac{\gamma^2 K^4}{\tau_m}\Phi'\Phi + \tau_v \mathbf{I}_J\right)\tau^{-1}\Phi' \tag{A.75}$$

Let us further write:

$$\tau^{-1} = \begin{bmatrix} \omega_{11} & \omega_{12} & \dots & \omega_{1J} \\ \omega_{12} & \omega_{22} & \dots & \omega_{2J} \\ \vdots & \vdots & \vdots & \vdots \\ \omega_{1J} & \omega_{2J} & \dots & \omega_{JJ} \end{bmatrix}, \tag{A.76}$$

and

$$\Phi \equiv \begin{bmatrix} \Phi_1 & \Phi_2 & \dots & \Phi_J \end{bmatrix}. \tag{A.77}$$

We can then write:

$$\Phi\tau^{-1}\Phi' = \sum_{j=1}^{J}\sum_{k=1}^{J}\omega_{kj}\Phi_k\Phi_j', \tag{A.78}$$

$$\Phi\tau^{-1}\tau^{-1}\Phi' = \sum_{j=1}^{J}\sum_{k=1}^{J}\left(\sum_{n=1}^{J}\omega_{kn}\omega_{nj}\right)\Phi_k\Phi_j'. \tag{A.79}$$

We further need to compute:

$$\Phi\tau^{-1}\Phi'\Phi\tau^{-1}\Phi' = \sum_{n=1}^{J}\sum_{i=1}^{J}\sum_{j=1}^{J}\sum_{k=1}^{J}\omega_{in}\omega_{kj}\Phi_i\Phi_n'\Phi_k\Phi_j'. \tag{A.80}$$

We first note that:

$$\Phi_i\Phi_n'\Phi_k\Phi_j' = (\Phi_n'\Phi_k)\Phi_i\Phi_j'. \tag{A.81}$$

Substituting back we get:

$$\Phi\tau^{-1}\Phi'\Phi\tau^{-1}\Phi' = \sum_{n=1}^{J}\sum_{i=1}^{J}\sum_{j=1}^{J}\sum_{k=1}^{J}\omega_{in}\omega_{kj}(\Phi_n'\Phi_k)\Phi_i\Phi_j' \tag{A.82}$$

$$= \sum_{j=1}^{J}\sum_{i=1}^{J}\left(\sum_{k=1}^{J}\sum_{n=1}^{J}\omega_{in}\omega_{kj}(\Phi_n'\Phi_k)\right)\Phi_i\Phi_j' \tag{A.83}$$

For convenience, we relabel the indices to obtain:

$$\Phi\tau^{-1}\Phi'\Phi\tau^{-1}\Phi' = \sum_{j=1}^{J}\sum_{k=1}^{J}\left(\sum_{i=1}^{J}\sum_{n=1}^{J}\omega_{kn}\omega_{ij}(\Phi_n'\Phi_i)\right)\Phi_k\Phi_j' \tag{A.84}$$

We can then compute:

$$\Phi\tau^{-1}\left(\frac{\gamma^2K^4}{\tau_m}\Phi'\Phi+\tau_v\mathbf{I}_J\right)\tau^{-1}\Phi'+\frac{\gamma^2K^4}{\tau_m\tau_\epsilon}\Phi\tau^{-1}\Phi' \tag{A.85}$$

$$= \sum_{j=1}^{J}\sum_{k=1}^{J}\left(\frac{\gamma^2K^4}{\tau_m\tau_\epsilon}\omega_{kj}+\tau_v\sum_{n=1}^{J}\omega_{nj}\omega_{kn}+\frac{\gamma^2K^4}{\tau_m}\sum_{i=1}^{J}\sum_{n=1}^{J}\omega_{kn}\omega_{ij}(\Phi_n'\Phi_i)\right)\Phi_k\Phi_j' \tag{A.86}$$

$$= \sum_{j=1}^{J}\sum_{k=1}^{J}\underbrace{\left(\frac{\gamma^2K^4}{\tau_m\tau_\epsilon}\omega_{kj}+\sum_{n=1}^{J}\omega_{kn}\left(\tau_v\omega_{nj}+\frac{\gamma^2K^4}{\tau_m}\sum_{i=1}^{J}\omega_{ij}\Phi_n'\Phi_i\right)\right)}_{\equiv c_{kj}}\Phi_k\Phi_j', \tag{A.87}$$

and thus

$$\widehat{\Sigma} = \left(1+\frac{\gamma^2K^2}{\tau_m\tau_\epsilon}\right)\Sigma+\sum_{j=1}^{J}\sum_{k=1}^{J}c_{kj}\Phi_k\Phi_j'. \tag{A.88}$$

We can write expected returns as

$$\mu = \frac{\mu_M}{\sigma_M^2}\left(1+\frac{\gamma^2K^2}{\tau_m\tau_\epsilon}\right)^{-1}\left(\widehat{\sigma}_M^2\widehat{\beta}-\sum_{k=1}^{J}\left(\sum_{j=1}^{J}\bar{\phi}_jc_{kj}\right)\Phi_k\right), \tag{A.89}$$

where $\bar{\phi}_j \equiv \Phi_j'M$ denotes the average loading on factor j. Similarly, using Proposition 2, we can obtain a relation between $\widehat{\beta}_\Delta$ and the J factors:

$$\widehat{\beta}_\Delta = \frac{\widehat{\sigma}_T^2}{\mu_T\widehat{\sigma}_\Delta^2}\left(\widehat{\sigma}_M^2\left(\frac{\mu_M}{\sigma_M^2}\left(1+\frac{\gamma^2K^2}{\tau_m\tau_\epsilon}\right)^{-1}-\frac{\mu_T}{\widehat{\sigma}_T^2}\right)\widehat{\beta}-\frac{\mu_M}{\sigma_M^2}\left(1+\frac{\gamma^2K^2}{\tau_m\tau_\epsilon}\right)^{-1}\sum_{j=1}^{J}\left(\bar{\phi}_j\sum_{k=1}^{J}c_{kj}\right)\Phi_k\right) \tag{A.90}$$

$$= \frac{\widehat{\sigma}_T^2}{\mu_T\widehat{\sigma}_\Delta^2}\left(\widehat{\sigma}_M^2\left(\frac{\gamma\tau_m\tau_\epsilon}{\tau_m\tau_\epsilon+\gamma^2K^2}-\widehat{B}\right)\widehat{\beta}-\frac{\mu_M}{\widehat{\sigma}_M^2}\left(1+\frac{\gamma^2K^2}{\tau_m\tau_\epsilon}\right)^{-1}\sum_{j=1}^{J}\left(\bar{\phi}_j\sum_{k=1}^{J}c_{kj}\right)\Phi_k\right), \tag{A.91}$$

where we have used $\mu_T/\widehat{\sigma}_T^2 \equiv \widehat{B} = \mu'\widehat{\Sigma}^{-1}\mathbf{1}$. Premultiply Eq. (A.88) by $\widehat{\Sigma}^{-1}$ and rearrange to obtain:

$$\frac{\tau_m\tau_\epsilon}{\gamma^2K^2+\tau_m\tau_\epsilon}\left(\mathbf{I}_N-\sum_{j=1}^{J}\sum_{k=1}^{J}c_{kj}\widehat{\Sigma}^{-1}\Phi_k\Phi_j'\right) = \widehat{\Sigma}^{-1}\Sigma, \tag{A.92}$$

which substituted back in $\widehat{B} = \mathbf{1}'\widehat{\Sigma}^{-1}\mu = \gamma\mathbf{1}'\widehat{\Sigma}^{-1}\Sigma M$ yields:

$$\widehat{B} = \frac{\gamma\tau_m\tau_\epsilon}{\gamma^2K^2+\tau_m\tau_\epsilon}\left(\underbrace{M'\mathbf{1}}_{\equiv 1} - \sum_{j=1}^{J}\sum_{k=1}^{J} c_{kj}\mathbf{1}'\widehat{\Sigma}^{-1}\Phi_k \underbrace{\Phi_j' M}_{\equiv\bar{\phi}_j}\right) \tag{A.93}$$

$$= \frac{\gamma\tau_m\tau_\epsilon}{\gamma^2K^2+\tau_m\tau_\epsilon}\left(1 - \sum_{j=1}^{J}\bar{\phi}_j\left(\sum_{k=1}^{J} c_{kj}\right)\mathbf{1}'\widehat{\Sigma}^{-1}\Phi_k\right) \tag{A.94}$$

Finally, substitute back in the relation for β_Δ and get:

$$\widehat{\beta}_\Delta = \frac{\widehat{\sigma}_T^2}{\mu_T\widehat{\sigma}_\Delta^2}\frac{\gamma\tau_m\tau_\epsilon}{\gamma^2K^2+\tau_m\tau_\epsilon}\sum_{j=1}^{J}\bar{\phi}_j\sum_{k=1}^{J}c_{kj}\left(\underbrace{\mathbf{1}'\widehat{\Sigma}^{-1}\Phi_k}_{\equiv\widehat{B}_k}\widehat{\sigma}_M^2\widehat{\beta} - \Phi_k\right) \tag{A.95}$$

$$= \frac{\widehat{\sigma}_T^2}{\mu_T\widehat{\sigma}_\Delta^2}\frac{\gamma\tau_m\tau_\epsilon}{\gamma^2K^2+\tau_m\tau_\epsilon}\sum_{k=1}^{J}\widehat{B}_k\left(\sum_{j=1}^{J}\bar{\phi}_j c_{kj}\right)\widehat{\Sigma}\left(M - \widehat{\Sigma}^{-1}\Phi_k/\widehat{B}_k\right) \tag{A.96}$$

which completes the Proof of Proposition 6. □

A.7 Proof of Proposition 7

Following the steps of Appendix A.3 we start from average market-to-book ratios, yet in the $J-$factors case:

$$\frac{\mathbb{E}[P]}{K} = \sum_{k=1}^{J}\Phi_k F_k - \frac{\mu}{K}. \tag{A.97}$$

We then use the asset-pricing relation in Eq. (30) to write the vector of loadings on the $k-$th factor as:

$$\Phi_k = \left(\sum_{j=1}^{J}\bar{\phi}_j c_{k,j}\right)^{-1}\left(\frac{\widehat{\sigma}_M^2}{\sigma_M^2}\frac{\mu_M}{\gamma}\widehat{\beta} - \frac{\gamma^2K^2+\tau_m\tau_\epsilon}{\gamma\tau_m\tau_\epsilon}\mu - \sum_{l\neq k}\left(\sum_{j=1}^{J}\bar{\phi}_j c_{lj}\right)\Phi_l\right). \tag{A.98}$$

We then subtitute it back in the above equation, isolating unconditional expected excess returns, μ, on the left-hand side:

$$\mu = \left(\frac{1}{K} + \frac{\gamma^2K^2+\tau_m\tau_\epsilon}{\gamma\tau_m\tau_\epsilon}\sum_{k=1}^{J}F_k\left(\sum_{j=1}^{J}\bar{\phi}_j c_{k,j}\right)^{-1}\right)^{-1}\begin{pmatrix} \frac{\widehat{\sigma}_M^2}{\sigma_M^2}\frac{\mu_M}{\gamma}\sum_{k=1}^{J}F_k\left(\sum_{j=1}^{J}\bar{\phi}_j c_{k,j}\right)^{-1}\widehat{\beta} - \frac{\mathbb{E}[P]}{K} \\ -\sum_{k=1}^{J}F_k\left(\sum_{j=1}^{J}\bar{\phi}_j c_{kj}\right)^{-1}\sum_{l\neq k}\left(\sum_{j=1}^{J}\bar{\phi}_j c_{lj}\right)\Phi_l \end{pmatrix}. \tag{A.99}$$

To rewrite the second line in the second bracket we use that, for arbitrary coefficients a and b:

$$\sum_{k=1}^{J}a_k\sum_{l\neq k}b_l\Phi_l = \sum_{k=1}^{J}b_k\left(\sum_{l\neq k}a_l\right)\Phi_k. \tag{A.100}$$

This yields the following asset-pricing relation (where we have relabled the indices for convenience):

$$\mu = \left(\frac{1}{K} + \frac{\gamma^2 K^2 + \tau_m \tau_\epsilon}{\gamma \tau_m \tau_\epsilon} \sum_{k=1}^{J} F_k \left(\sum_{j=1}^{J} \bar{\phi}_j c_{k,j} \right)^{-1} \right)^{-1} \left(\begin{array}{c} \frac{\widehat{\sigma}_M^2}{\sigma_M^2} \frac{\mu_M}{\gamma} \sum_{k=1}^{J} F_k \left(\sum_{j=1}^{J} \bar{\phi}_j c_{k,j} \right)^{-1} \widehat{\beta} - \frac{\mathbb{E}[P]}{K} \\ - \sum_{k=1}^{J} \left(\sum_{l \neq k} F_l \frac{\sum_{j=1}^{J} \bar{\phi}_j c_{kj}}{\sum_{j=1}^{J} \bar{\phi}_j c_{lj}} \right) \Phi_k \end{array} \right), \tag{A.101}$$

which in turn delivers the relation in Eq. (32), with:

$$\lambda_2 \equiv - \left(\frac{1}{K} + \frac{\gamma^2 K^2 + \tau_m \tau_\epsilon}{\gamma \tau_m \tau_\epsilon} \sum_{k=1}^{J} F_k \left(\sum_{j=1}^{J} \bar{\phi}_j c_{k,j} \right)^{-1} \right)^{-1}, \tag{A.102}$$

$$\lambda_1 \equiv -\lambda_2 \frac{\widehat{\sigma}_M^2}{\sigma_M^2} \frac{\mu_M}{\gamma} \sum_{k=1}^{J} F_k \left(\sum_{j=1}^{J} \bar{\phi}_j c_{k,j} \right)^{-1}, \tag{A.103}$$

$$\lambda_k \equiv \lambda_2 \sum_{l \neq k} F_l \frac{\sum_{j=1}^{J} \bar{\phi}_j c_{kj}}{\sum_{j=1}^{J} \bar{\phi}_j c_{lj}}, \quad k = 1, \ldots, J. \tag{A.104}$$

The statement that one of the J factors is redundant follows from that we can extract one factor, say factor l, arbitrarily from Eq. (A.97):

$$\frac{\mathbb{E}[P]}{K} = \Phi_l F_l + \sum_{k \neq l} \Phi_k F_k - \frac{\mu}{K}. \tag{A.105}$$

Mirroring the computations above, we can then use Eq. (A.98) to write:

$$\mu = \left(\frac{1}{K} + \frac{\gamma^2 K^2 + \tau_m \tau_\epsilon}{\gamma \tau_m \tau_\epsilon} F_l \left(\sum_{j=1}^{J} \bar{\phi}_j c_{l,j} \right)^{-1} \right)^{-1} \left(\begin{array}{c} \frac{\widehat{\sigma}_M^2}{\sigma_M^2} \frac{\mu_M}{\gamma} F_l \left(\sum_{j=1}^{J} \bar{\phi}_j c_{l,j} \right)^{-1} \widehat{\beta} - \frac{\mathbb{E}[P]}{K} \\ + \sum_{k \neq l} \left(F_k - F_l \frac{\sum_{j=1}^{J} \bar{\phi}_j c_{kj}}{\sum_{j=1}^{J} \bar{\phi}_j c_{lj}} \right) \Phi_k \end{array} \right) \tag{A.106}$$

$$\equiv \lambda_1' \widehat{\beta} + \lambda_2' \frac{\mathbb{E}[P]}{K} + \sum_{k \neq l} \lambda_{k+2}' \Phi_k. \tag{A.107}$$

This completes the proof of Proposition 7. □

A.8 Proof of Proposition 9

Consider the following multivariate time-series relation (stacked for all assets):

$$R^e = \widetilde{\alpha} + \widetilde{\beta} R_M^e + \widetilde{\beta}_\Delta R_\Delta^e + u, \tag{A.108}$$

where R^e is the vector of excess returns on all assets, R_M^e is the excess return on the market, and R_Δ^e is the excess return on the Δ portfolio. Let the univariate time-series relation be

$$R^e = \widehat{\alpha} + \widehat{\beta} R_M^e + \varepsilon, \tag{A.109}$$

where the coefficient β is the same as in Proposition 2. Comparing Eqs. (A.108) and (A.109), there is an omitted variable (R_Δ^e) in Eq. (A.109). This yields

$$\widehat{\beta} = \widetilde{\beta} + \frac{\widehat{\sigma}_{M\Delta}}{\widehat{\sigma}_M^2} \widetilde{\beta}_\Delta. \tag{A.110}$$

By the same logic, writing $R^e = \widehat{\alpha}_\Delta + \widehat{\beta}_\Delta R^e_M + \varepsilon_\Delta$ and comparing with Eq. (A.108) yields

$$\widehat{\beta}_\Delta = \widetilde{\beta}_\Delta + \frac{\widehat{\sigma}_{M\Delta}}{\widehat{\sigma}^2_\Delta}\widetilde{\beta}. \tag{A.111}$$

Eqs. (A.110)-(A.111) can then be replaced in Proposition 1:

$$\mu = \frac{\mu_T \widehat{\sigma}^2_M}{\widehat{\sigma}^2_T}\left(\widetilde{\beta} + \frac{\widehat{\sigma}_{M\Delta}}{\widehat{\sigma}^2_M}\widetilde{\beta}_\Delta\right) + \frac{\mu_T \widehat{\sigma}^2_\Delta}{\widehat{\sigma}^2_T}\left(\widetilde{\beta}_\Delta + \frac{\widehat{\sigma}_{M\Delta}}{\widehat{\sigma}^2_\Delta}\widetilde{\beta}\right). \tag{A.112}$$

This relation can be further simplified by eliminating $\widehat{\sigma}_{M\Delta}$ and μ_T. In order to do this, multiply Eq. (49) with M' and with Δ' to obtain two equations with two unknowns:

$$\mu_M = \frac{\mu_T \widehat{\sigma}^2_M}{\widehat{\sigma}^2_T} + \frac{\mu_T}{\widehat{\sigma}^2_T}\widehat{\sigma}_{M\Delta} \tag{A.113}$$

$$\mu_\Delta = \frac{\mu_T}{\widehat{\sigma}^2_T}\widehat{\sigma}_{M\Delta} + \frac{\mu_T \widehat{\sigma}^2_\Delta}{\widehat{\sigma}^2_T}, \tag{A.114}$$

which yields

$$\widehat{\sigma}_{M\Delta} = \frac{\mu_\Delta \widehat{\sigma}^2_M - \mu_M \widehat{\sigma}^2_\Delta}{\mu_M - \mu_\Delta} \tag{A.115}$$

$$\mu_T = \frac{(\mu_M - \mu_\Delta)\widehat{\sigma}^2_T}{\widehat{\sigma}^2_M - \widehat{\sigma}^2_\Delta}. \tag{A.116}$$

Replacing $\widehat{\sigma}_{M\Delta}$ and μ_T from above in Eq. (A.112) yields

$$\mu = \mu_M \widetilde{\beta} + \mu_\Delta \widetilde{\beta}_\Delta. \tag{A.117}$$

which is Eq. (51) of Proposition 9. Furthermore, solving for $\widetilde{\beta}$ and $\widetilde{\beta}_\Delta$ in (A.110)-(A.111) yields Eq. (52). This completes the proof of Proposition 9. □

A.9 Proof of Proposition 8

As $T \to \infty$, $\overline{\mathbf{R}}^e_T \xrightarrow{\text{p}} \boldsymbol{\mu}$; this suggests that if we increase T appropriately fast (specifically, in a way that $T/N^{3/2} \to \varphi > 0$) as we take the large economy limit, the mean of $\widehat{W}_{N,T}$ under the true model will remain finite:

$$\ell_\infty \equiv \lim_{N,J\to\infty, J/N\to\psi, T/N^{3/2}\to\varphi} \ell_{N,T} = \lim_{N,J\to\infty, J/N\to\psi} \frac{\varphi}{\sqrt{2}(1+\mu^2_M/\widehat{\sigma}^2_M)} N\boldsymbol{\mu}'\left(\widehat{\boldsymbol{\Sigma}}^{-1} - \frac{1}{\widehat{\sigma}^2_M}\mathbf{M}\mathbf{M}'\right)\boldsymbol{\mu}. \tag{A.118}$$

Substituting the equilibrium expression for $\boldsymbol{\mu}$ further yields:

$$\ell_\infty = \lim_{N,J\to\infty, J/N\to\psi} \frac{\varphi\gamma^2}{\sqrt{2}(1+\gamma^2\sigma^4_M/\widehat{\sigma}^2_M)}\left(N\mathbf{M}'\boldsymbol{\Sigma}\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}\mathbf{M} - \frac{(N\sigma^2_M)^2}{N\widehat{\sigma}^2_M}\right). \tag{A.119}$$

We start from the law of total covariance to write:

$$\widehat{\mathbf{\Sigma}} = \mathbf{\Sigma} + \frac{\gamma^2}{\tau_m}\mathbf{\Sigma}\mathbf{\Sigma} + \tau_v J \mathbf{\Phi}\boldsymbol{\tau}^{-1}\boldsymbol{\tau}^{-1}\mathbf{\Phi}' \tag{A.120}$$

$$= \left(1 + \frac{K^2\gamma^2}{\tau_m\tau_\epsilon}\right)\mathbf{\Sigma} + \mathbf{\Phi}\boldsymbol{\tau}^{-1}\left(\frac{K^4\gamma^2}{\tau_m}\left(\mathbf{\Phi}'\mathbf{\Phi} + \tau_\epsilon^{-1}\boldsymbol{\tau}\right) + \tau_v J\mathbf{I}\right)\boldsymbol{\tau}^{-1}\mathbf{\Phi}', \tag{A.121}$$

where

$$\mathbf{\Sigma} = \mathbf{\Phi}\boldsymbol{\tau}^{-1}\mathbf{\Phi} + \tau_\epsilon^{-1}\mathbf{I}. \tag{A.122}$$

Let us now rescale by N and use the eigendecomposition in Eq. (33) and the definition of limiting precision in Eq. (34) to rewrite this expression as:

$$N\widehat{\mathbf{\Sigma}} = \left(1 + \frac{K^2\gamma^2}{\tau_m\tau_\epsilon}\right)N\mathbf{\Sigma} + \mathbf{\Phi}\mathbf{Q}\underbrace{\boldsymbol{\tau}_\infty^{-1}\left(\frac{K^4\gamma^2}{\tau_m}\left(\mathbf{\Lambda} + \tau_\epsilon^{-1}\boldsymbol{\tau}_\infty\right) + \tau_v\psi\mathbf{I}\right)\boldsymbol{\tau}_\infty^{-1}}_{\equiv \mathbf{B}\text{ (diagonal matrix)}}\mathbf{Q}'\mathbf{\Phi}'. \tag{A.123}$$

We then define the following vector:

$$\mathbf{Z} \equiv \mathbf{Q}'\mathbf{\Phi}'\mathbf{M}. \tag{A.124}$$

Denoting its $j-$th element by z_j, and using Eq. (A.122) and Eq. (A.123) we can write σ_M^2 and $\widehat{\sigma}_M^2$ as:

$$N\sigma_M^2 = \tau_\epsilon^{-1} + \sum_{j=1}^{J} z_j^2\tau_\infty(\lambda_j)^{-1}, \tag{A.125}$$

$$N\widehat{\sigma}_M^2 = \left(1 + \frac{K^2\gamma^2}{\tau_m\tau_\epsilon}\right)N\sigma_M^2 + \sum_{j=1}^{J} z_j^2\tau_\infty(\lambda_j)^{-2}\left(\frac{K^4\gamma^2}{\tau_m}(\lambda_j + \tau_\epsilon^{-1}\tau_\infty(\lambda_j)) + \tau_v\psi\right), \tag{A.126}$$

where the function $\tau_\infty(\cdot)$ is defined in Eq. (35). Taking now care of the term in brackets in Eq. (A.119), we first use Woodbury matrix identity to write:

$$(N\widehat{\mathbf{\Sigma}})^{-1} = \left(1 + \frac{\gamma^2K^2}{\tau_m\tau_\epsilon}\right)^{-1}(N\mathbf{\Sigma})^{-1}\left(\mathbf{I} - \left(1 + \frac{\gamma^2K^2}{\tau_m\tau_\epsilon}\right)^{-1}\mathbf{\Phi}\mathbf{Q}\left(\mathbf{B}^{-1} + \mathbf{Q}'\mathbf{\Phi}'(N\mathbf{\Sigma})^{-1}\mathbf{\Phi}\mathbf{Q}\left(1 + \frac{\gamma^2K^2}{\tau_m\tau_\epsilon}\right)^{-1}\right)^{-1}\mathbf{Q}'\mathbf{\Phi}'(N\mathbf{\Sigma})^{-1}\right) \tag{A.127}$$

and thus we have:

$$N\mathbf{\Sigma}\widehat{\mathbf{\Sigma}}^{-1}\mathbf{\Sigma} = \left(1 + \frac{\gamma^2K^2}{\tau_m\tau_\epsilon}\right)^{-1}\left(N\mathbf{\Sigma} - \left(1 + \frac{\gamma^2K^2}{\tau_m\tau_\epsilon}\right)^{-1}\mathbf{\Phi}\mathbf{Q}\left(\mathbf{B}^{-1} + \mathbf{Q}'\mathbf{\Phi}'(N\mathbf{\Sigma})^{-1}\mathbf{\Phi}\mathbf{Q}\left(1 + \frac{\gamma^2K^2}{\tau_m\tau_\epsilon}\right)^{-1}\right)^{-1}\mathbf{Q}'\mathbf{\Phi}'\right). \tag{A.128}$$

Similarly, we obtain:

$$(N\mathbf{\Sigma})^{-1} = N^{-1}\tau_\epsilon\left(\mathbf{I} - N^{-1}\tau_\epsilon\mathbf{\Phi}\mathbf{Q}(\boldsymbol{\tau}_\infty + \tau_\epsilon\mathbf{\Lambda})^{-1}\mathbf{Q}'\mathbf{\Phi}'\right), \tag{A.129}$$

which gives:

$$\mathbf{Q}'\mathbf{\Phi}'(N\mathbf{\Sigma})^{-1}\mathbf{\Phi}\mathbf{Q} = \tau_\epsilon\left(\mathbf{\Lambda} - \tau_\epsilon\mathbf{\Lambda}(\boldsymbol{\tau}_\infty + \tau_\epsilon\mathbf{\Lambda})^{-1}\mathbf{\Lambda}\right), \tag{A.130}$$

which is a diagonal matrix. Substituting back we get:

$$N\boldsymbol{\Sigma}\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma} = \left(1+\frac{\gamma^2K^2}{\tau_m\tau_\epsilon}\right)^{-1} N\tau_\epsilon^{-1}\mathbf{I} + \left(1+\frac{\gamma^2K^2}{\tau_m\tau_\epsilon}\right)^{-1} \boldsymbol{\Phi}\mathbf{Q}\boldsymbol{\tau}_\infty^{-1}\mathbf{Q}'\boldsymbol{\Phi}' \tag{A.131}$$

$$-\boldsymbol{\Phi}\mathbf{Q}\left(1+\frac{\gamma^2K^2}{\tau_m\tau_\epsilon}\right)^{-2}\left(\mathbf{B}^{-1}+\left(1+\frac{\gamma^2K^2}{\tau_m\tau_\epsilon}\right)^{-1}\tau_\epsilon\left(\boldsymbol{\Lambda}-\tau_\epsilon\boldsymbol{\Lambda}(\boldsymbol{\tau}_\infty+\tau_\epsilon\boldsymbol{\Lambda})^{-1}\boldsymbol{\Lambda}\right)^{-1}\right)\mathbf{Q}'\boldsymbol{\Phi}'. \tag{A.132}$$

Finally, using the definition of the vector $\mathbf{Z}$ and substituting back $\mathbf{B}$ we obtain:

$$N\mathbf{M}'\boldsymbol{\Sigma}\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}\mathbf{M} = \left(1+\frac{\gamma^2K^2}{\tau_m\tau_\epsilon}\right)\tau_\epsilon^{-1} \tag{A.133}$$

$$+\sum_{j=1}^{J} z_j^2 \left(\begin{array}{c} \tau_\infty(\lambda_j)^{-1} - \left(1+\frac{\gamma^2K^2}{\tau_m\tau_\epsilon}\right)^{-1}\left(\tau_\infty(\lambda_j)^2\left(\frac{\gamma^2K^4}{\tau_m\tau_\epsilon}(\tau_\infty(\lambda_j)+\tau_\epsilon\lambda_j)\right)+\tau_v\psi\right)^{-1} \\ +\left(1+\frac{\gamma^2K^2}{\tau_m\tau_\epsilon}\right)^{-1}\tau_\epsilon\left(\lambda_j-\tau_\epsilon\lambda_j^2(\tau_\infty(\lambda_j)+\tau_\epsilon\lambda_j)^{-1}\right)^{-1} \end{array}\right). \tag{A.134}$$

To obtain more transparent expressions we now use the approximation in Eq. (36) and Assumption 1. This assumption implies that $\boldsymbol{\Phi}\boldsymbol{\Phi}' \to N\mathbf{I}$ and thus

$$\frac{1}{NJ}\mathrm{tr}(\boldsymbol{\Phi}'\boldsymbol{\Phi}) = 1, \tag{A.135}$$

so that Eq. (24) is satisfied in the limit. We now want to compute expressions of the form:

$$\lim_{J\to\infty}\sum_{j=1}^{J} z_j^2 f(\lambda_j), \tag{A.136}$$

for an arbitrary function f based on results in Bai, Miao, and Pan (2007) (among others). Under Assumption 1, $\mathbf{Q}$ is asymptotically Haar distributed. The main idea, which originates from Silverstein (1989), is to take any unit vector $\mathbf{x}$ and focus on $\mathbf{Q}'\mathbf{x} \equiv \mathbf{y}$, so that $\mathbf{y}$ is Uniformly distributed over $\{y \in \mathbb{R}^J : \|y\| = 1\}$. We then obtain that expressions like $\sum_{j=1}^J |y_j|^2 f(\lambda_j)$ converge to $\frac{1}{J}\sum_{j=1}^J f(\lambda_j)$ (Corollary 2 in Bai et al. (2007)). We would like our vector $\mathbf{Z}$ to share this key property of $\mathbf{y}$. However, although $\mathbf{x}$ is arbitrary, it must be nonrandom. We deal with this issue as follows. Note that, as pointed out in Bai and Silverstein (1998), the two matrices $\mathbf{A}_1 \equiv \boldsymbol{\Phi}'\boldsymbol{\Phi}/N$ and its companion $\mathbf{A}_2 \equiv \boldsymbol{\Phi}\boldsymbol{\Phi}'/N$ 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 vector $N^{-1/2}\boldsymbol{\Phi}'$ as:

$$\frac{1}{N^{1/2}}\boldsymbol{\Phi}' = \sum_{j=1}^{J}\sqrt{\lambda_j}\mathbf{q}_j\mathbf{v}_j', \tag{A.137}$$

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:

$$\frac{1}{N^{1/2}}\mathbf{Q}'\boldsymbol{\Phi}' = \sum_{j=1}^{J}\sqrt{\lambda_j}\mathbf{e}_j\mathbf{v}_j', \tag{A.138}$$

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:

$$\mathbf{Z} = \sum_{j=1}^{J} \sqrt{\lambda_j}\mathbf{e}_j\mathbf{v}_j'\mathbf{x}, \tag{A.139}$$

with $j-$th entry:

$$z_j = \sqrt{\lambda_j}\mathbf{v}_j'\mathbf{x}. \tag{A.140}$$

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:

$$\sum_{j=1}^{J} z_j^2 f(\lambda_j) = \sum_{j=1}^{J} \lambda_j(\mathbf{v}_j'\mathbf{x})^2 f(\lambda_j) = \sum_{j=1}^{N} \lambda_j \mathbf{1}_{j\leq J}(\mathbf{v}_j'\mathbf{x})^2 f(\lambda_j) \tag{A.141}$$

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

$$\sum_{j=1}^{J} z_j^2 f(\lambda_j) \to \frac{1}{N}\sum_{j=1}^{N} \lambda_j \mathbf{1}_{j\leq J} f(\lambda_j) = \int f(\lambda)\lambda \mathrm{d}F^{\mathbf{A}_2}(\lambda), \tag{A.142}$$

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

$$F^{\mathbf{A}_2} = (1-\psi)\mathbf{1}_{[0,\infty)} + \psi F^{\mathbf{A}_1}. \tag{A.143}$$

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:

$$\sum_{j=1}^{J} z_j^2 f(\lambda_j) \to \psi \int f(\lambda)\lambda \mathrm{d}F^{\mathbf{A}_1}(\lambda). \tag{A.144}$$

This result allows us to characterize the mean ℓ_∞ of the GRS statistic in terms of the distribution of eigenvalues, λ_j. Under Assumption 1 this distribution is the Marchenko-Pastur law. Using this distribution along with the approximation in Eq. (36), we obtain simpler expressions for:

$$N\sigma_M^2 \approx \tau_\epsilon^{-1} + (\tau_F + \tau_G + \tau_v)^{-1}, \tag{A.145}$$

and

$$N\widehat{\sigma}_M^2 \approx \left(1 + \frac{K^2\gamma^2}{\tau_m\tau_\epsilon}\right)N\sigma_M^2 + (\tau_F + \tau_G + \tau_v)^{-2}\left(\frac{\gamma^2}{\tau_m}(\psi+1)/\psi + \frac{\gamma^2}{\tau_m\tau_\epsilon}(\tau_F + \tau_G + \tau_v) + \tau_v\right), \tag{A.146}$$

$$N\mathbf{M}'\mathbf{\Sigma}\widehat{\mathbf{\Sigma}}^{-1}\mathbf{\Sigma}\mathbf{M} \approx \left(1 + \frac{\gamma^2K^2}{\tau_m\tau_\epsilon}\right)\tau_\epsilon^{-1} \tag{A.147}$$

$$+\psi\int_{(1-\sqrt{\psi})^2}^{(1+\sqrt{\psi})^2} \lambda \begin{pmatrix} -\left(1+\frac{\gamma^2K^2}{\tau_m\tau_\epsilon}\right)^{-1}\left((\tau_F+\tau_G+\tau_v)^2\psi^2\left(\frac{\gamma^2K^4}{\tau_m\tau_\epsilon}((\tau_F+\tau_G+\tau_v)\psi+\tau_\epsilon\lambda)\right)+\tau_v\psi\right)^{-1} \\ (\tau_F+\tau_G+\tau_v)i^{-1}\psi^{-1}+\left(1+\frac{\gamma^2K^2}{\tau_m\tau_\epsilon}\right)^{-1}\tau_\epsilon\left(\lambda-\tau_\epsilon\lambda^2((\tau_F+\tau_G+\tau_v)\psi+\tau_\epsilon\lambda)^{-1}\right)^{-1}\end{pmatrix}\mathrm{d}\nu(\lambda), \tag{A.148}$$

where ν denotes the Pastur-Marchenko law. Under particular parameter values this integral has an analytic solution. For instance, under the naive calibration we use in the main text, this quantity is available in closed form. Finally, to complete the computation of the asymptotic mean ℓ_∞, we need to worry about the ratio, $\mu_M^2/\widehat{\sigma}_M^2$, which appears in the denominator. Because μ_M^2 is of order $O(1/N^2)$ and $\widehat{\sigma}_M^2$ is of order $O(1/N)$, $\mu_M^2/\widehat{\sigma}_M^2 \to 0$. With this information we can compute the $p-$value of the GRS test.

□