Granular Asset Pricing[*]

Link to current version

Abstract

The market capitalization distribution of US firms has a fat tail populated by the largest firms. I refer to this fat tail as granularity and quantify it by the Pareto distribution to study asset pricing implications. Granularity breaks the diversification of idiosyncratic risks assumed by factor models. The size-adjusted idiosyncratic risk explains the expected returns such that only large firms have their idiosyncratic risks un-diversified to generate positive risk premiums. This finding explains the negative relation between idiosyncratic risk and stock returns, known as the ″idiosyncratic risk puzzle.″ The level of granularity, measured by the Pareto coefficient of firm size, explains market expected returns since it determines the under-diversification of idiosyncratic risk at the aggregate level.

JEL-Classification:
Keywords: Granularity, Fat Tail Distribution, Pareto Distribution, Arbitrage Asset Pricing

[*]

1 Introduction

A few publicly traded firms account for a significant fraction of the overall US stock market valuation. In 2020, the ten largest firms accounted for over a quarter of the total market value of the around 4,000 publicly traded firms in the Center for Research in Security Prices (CRSP) database as shown in **Figure 1**. Further, this striking market value concentration in large firms persisted over time. I plot the market weight of the largest 1 percent firms per month during 1926-2020 in **Figure 2**, . This rough proxy for market concentration displays an increasing trend over time and reaches a maximum of more than 50 percent during the 2000-2001 ″Internet Bubble″ period. These stylized facts suggest that the distribution of firms' market value has a fat tail populated by a few giant firms, which is consistent with the fat-tailed distribution of firms' fundamental values documented in the literature (number of employees in Axtell (2001), production value in Gabaix (2011), etc.). I refer to the fat-tailed distribution of market capitalization as stock market granularity and study its asset pricing implication.

The granularity in stock market data challenges a crucial assumption in asset pricing theory to imply the classical multi-factor model for expected returns. The multi-factor model assumes that the stock market is well-diversified, which requires a thin tail distribution of firm size such that no asset has large enough weight in the market portfolio. With diversification, only factors explain the expected returns, and idiosyncratic risks relative to factors are diversified away. In contrast, when the granularity is significant, as shown in **Figure 1**, the diversification argument fails, and the factor model does not obtain.

The first contribution of this paper is developing a granular APT (GAPT) theoretical framework to illustrate how granularity generates deviations from the factor models by making idiosyncratic risks explain expected returns. I apply the risk structure employed by the arbitrage pricing theory (APT) to derive factor and idiosyncratic risks as two independent components by decomposing risks in asset returns. The derivation of factor and idiosyncratic risks is based on a statistical criteria, which is independent of the market portfolio composition and distribution of firm size (e.g.,Chamberlain and Rothschild (1983),Connor and Korajczyk (1986)).[1] This risk structure allows APT models to derive a linear factor model for expected returns by adding the diversification assumption as a regulating condition from the perspective of firm size distribution to rule out the impact of idiosyncratic risk. Intuitively, my framework adds granularity to the same risk structure to illustrate how idiosyncratic risk explains the expected returns due to the failure of diversification.

I quantify the level of granularity by fitting the fat-tailed distribution of firms' market value with the Pareto distribution, which is frequently used in macroeconomic literature (see Gabaix (2011)). It describes the fat tail parsimoniously with a single parameter, the Pareto coefficient ζ. In the asset pricing context, ζ quantifies the granularity level and determines the magnitude of idiosyncratic risks under-diversified to generate deviations from the classical APT factor models. When ζ is small ($\zeta < 2$), the distribution has a fat tail, such that there are large firms with non-negligible market weight, and their idiosyncratic shocks generate size-related abnormal returns relative to APT factors. Granularity becomes smaller as ζ increases, and my analytical framework reverts to the conventional APT factor model when $\zeta > 2$. In this way, a thin-tail distribution of firm size invokes the law of large numbers and diversifies idiosyncratic shocks sufficiently to have a negligible impact on expected returns.

As the second contribution, I test a novel relation between idiosyncratic risk and expected returns in the cross-section implied by my model. With granularity, large firms with high market weights break the diversification and have their idiosyncratic risk explain expected returns. Specifically, I find that, the size-adjusted idiosyncratic risk (product of an asset's market weight and idiosyncratic variance) positively explains

[1] Chamberlain and Rothschild (1983),Chamberlain (1983),Connor and Korajczyk (1986) define the factor by a portfolio formed based on the eigenvectors of the covariance. The factors map to the unbounded eigenvalue of covariance, which captures a pervasive pattern in the covariance.

the expected returns in the cross-section, with various factors and characteristics controlled.[2] This result explains the "idiosyncratic risk puzzle" (IRP hereafter) that there is a very robust negative relationship between idiosyncratic risk and expected returns in the cross-section, investigated in Ang et al. (2006) and Ang et al. (2009).[3] As a feature of data found in the cited papers and my empirical exploration, large firms tend to have low idiosyncratic risk. This negative relation between firm size and idiosyncratic risk, combined with the granularity, explains the IRP. When the granularity is significant, large firms that populate the fat tail will account for most of the market valuation, as shown in **Figure 1**. Consequently, large firms have low idiosyncratic risks but have a significant risk premium tied to their idiosyncratic risks. Conversely, firms with high idiosyncratic risks tend to have negligible market weights and non-significant risk premiums raised by idiosyncratic risks.

The third contribution of my analysis is to test the aggregate impact of granularity on market returns. I estimate the Pareto coefficient ζ by fitting the fat-tail in firm size distribution each month and find that ζ is time-varying with an average value around 1. This finding suggests a granular channel of aggregate variation in the stock market since a lower Pareto coefficient ζ (higher granularity) indicates less diversified idiosyncratic risks in the market portfolio and more risk premium on aggregate. This result relates to whether time-variation of idiosyncratic risk explains the market expected returns in literature (see Goyal and Santa-Clara (2003), Bali et al. (2005)). I reconcile this literature with my model implication to test whether ζ generates additional time-variation of market risk premium, controlling the magnitude of idiosyncratic risk. My tests show that the Pareto coefficient successfully predicts the market return, especially in longer time horizons, even when controlling for additional predictors.

Related Literature

[2] As in Ang et al. (2006) and Ang et al. (2009), I use Fama-French factors to control the pervasive correlation and identify the idiosyncratic shocks. To be consistent with my theoretical approach, I also use principal components of returns as factors to measure idiosyncratic risk and find the same positive relation.

[3] Hou and Loh (2016) gives a thorough survey of explanations in published papers for this puzzling negative risk-return relation and concludes that none of them is sufficiently satisfying.

The paper relates to the massive amount of APT literature starting from Ross (1976), which is one of the major topics in asset pricing research (see Chamberlain and Rothschild (1983), Chamberlain (1983), Dybvig (1983), Connor and Korajczyk (1986), Connor and Korajczyk (1993), Huberman (2005)). I take the definition of diversification, factors, and idiosyncratic risk from Chamberlain and Rothschild (1983), and Chamberlain (1983). Based on these definitions, I show how granularity breaks the diversification and link it to the risk premium. Independently, there has been interesting research to better identify the factors based on the APT framework and improve the associating tests (see Feng, Giglio, and Xiu (2020), Kelly, Pruitt, and Su (2020), Giglio, Xiu, and Zhang (2021) Giglio and Xiu (2021),Giglio, Kelly, and Xiu (2022)).

The advantage of applying the APT framework is to set factor and idiosyncratic risk as two independent components in asset returns. The independence is attractive for the empirical test since it ensures the exogenous condition in estimating the factor model by linear regressions. Alternative factor framework may not ensure this advantage for the empirical test yet give similar risk-return relation to what's derived in this paper. For example, Byun and Schmidt (2020) argue that the granularity induces an endogenous relationship between the value-weighted returns and idiosyncratic shocks of large firms, potentially biasing the estimates of the CAPM risk exposure (”beta”) of large firms. Gabaix and Koijen (2020) develop a ”granular instrumental variable” to solve a similar endogenous bias issue in identifying supply and demand elasticity in a granular market.

My research relates to economic literature that studies the impact of large firms on aggregate fluctuation, e.g. , Gabaix (2011), Acemoglu et al. (2012), Acemoglu, Ozdaglar, and Tahbaz-Salehi (2015). From the macroeconomic perspective, they measure firm size by fundamental values such as production value and the number of employees. To study the asset pricing implication, I measure firm size by weight account in the market portfolio and link it to the classical diversification assumption employed by factor models. Another inspiring paper that studies the asset pricing implication of a fat-tailed distribution is Kelly and Jiang (2014), which measures the tail distribution of asset returns instead of firm size.

My analysis relates also to those studies that examine the relationship between asset prices and idiosyncratic risks, such as Campbell et al. (2001), Xu and Malkiel (2003), Goyal and Santa-Clara (2003) and Herskovic et al. (2016). Specifically, I reconcile the idiosyncratic puzzle posited by Ang et al. (2006) and Ang et al. (2009). Hou and Loh (2016) surveyed the existing explanations in the literature and found none of them is sufficiently convincing. My analysis contributes to this strand of literature by highlighting how any cross-sectional test relating to idiosyncratic risks must account for the size-related exposure caused by market granularity.

2 A Granular APT

I start with a standard APT model[4]. Firstly, I decompose risks in asset returns into two independent components: factor and idiosyncratic risks. Then I apply a simple competitive equilibrium in the stock market to derive how these two components explain the expected returns. This approach is applied in literature, such as Dybvig (1983), Connor and Korajczyk (1995), to derive the specific format of expected returns tied to idiosyncratic risks and hence show how this term of expected returns changes as the firm size distribution. As documented in the literature, I show that a thin-tailed distribution satisfies the diversification assumption and leads to a linear factor model by ruling out the impact of idiosyncratic risks. On the other hand, this theoretical framework allows me to study the expected return in an equilibrium where the distribution of market values is granular.

I only present necessary components here and attach the APT derivations in the **Appendix Section I**. There are n assets in the market; each asset return is r_i:

[4] Since most of the APT material is known, I leave out the cluster of citations here. The primary reference of this subsection is Connor and Korajczyk (1995), Chamberlain and Rothschild (1983), Chamberlain (1983)

$$
\begin{aligned}
r_i &= E[r_i] + \sum_{s=1}^{k} \beta_{i,s} f_s + \epsilon_i && (1)\\
E[\epsilon_i | f] &= 0, \forall i && (2)
\end{aligned}
$$

There are k common factors $f_s, s = 1...k$ with factor loadings $\beta_{i,s}$. The idiosyncratic shocks ϵ_i are independent of factors, treated as the "residual" or "firm-specific shock" of each asset return. A representative investor holds a portfolio described by the weights $\{w_i\}, i = 1...n$ such that $\sum_i^n w_i = 1$ and maximize a constant absolute risk aversion (CARA) utility base on the portfolio return $u(\sum_i^n w_i r_i)$. Under this classic APT setup, the expected returns are determined by the shocks of the pricing kernel, which equal to:

$$
-\gamma(\sum_i^n w_i \left(\beta_{i,s} f_s + \epsilon_i\right))
$$

γ is the risk aversion coefficient of the CARA utility. The shocks of the pricing kernel are proportional to shocks of the aggregate portfolio return $\sum_i^n w_i r_i$, which contains the weighted average of f and ϵ. An asset's expected return is determined by its covariance with the shocks of the pricing kernel. As a result, an asset's risk premium is a constant risk-free rate μ_0 plus a linear span of factor risk premiums $\mu_s, s = 1...k$ and a granular term determined by w_i and ϵ_i:

$$
E[r_i] = \mu_0 + \sum_{s=1}^{k} \beta_{i,s} \mu_s + \gamma COV(\epsilon_i, \sum_i^n w_i \epsilon_i) \tag{3}
$$

γ is the risk aversion coefficient of the utility. μ_s is the risk premium tied to factor f_s and $\beta_{i,s}, s = 1...k$ are the asset's exposures to each factor. μ_0 is a constant equal to the expected return of a zero factor exposure portfolio.

The granular shocks, $\sum_i^n w_i \epsilon_i$, are equal to the sum of firm-specific shocks and are weighted by each asset's relative weight in the market w_i. As a part of the pricing kernel, $\sum_i^n w_i \epsilon_i$ drives the expected return of an asset in (3) by its covariance with the idiosyncratic components of the asset's return ϵ_i. Additionally, the market expected return $E[r_m]$ is the weighted-average of $E[r_i]$ such that $E[r_m] = E[\sum_i^n w_i r_i]$ and equals to:

$$E[r_m] = \mu_0 + \sum_i^n w_i \left(\sum_{s=1}^k \beta_{i,s} \mu_s \right) + \gamma VAR(\sum_i^n w_i \epsilon_i) \tag{4}$$

The granular covariance terms $COV(\epsilon_i, \sum_i^n w_i \epsilon_i)$ in individual assets are compounded into the variance of the granular shocks $VAR(\sum_i^n w_i \epsilon_i)$ in expected market returns.

2.1 APT, diversification, and thin tail distribution

The APT models make assumptions about the distribution of w_i to rule out the idiosyncratic risk's impact on expected returns as in (3) and (4). Specifically, the APT models decompose asset returns into factors and idiosyncratic components by the covariance matrix. Let the covariance matrix of ϵ_i be $\boldsymbol{\Sigma\epsilon}$ and $\rho_i(\boldsymbol{\Sigma\epsilon}), i = 1...n$ be the eigenvalues of it, sorted in descending order. The idiosyncratic shocks ϵ_i are weakly correlated such that the covariance matrix among them has bounded eigenvalues as $n \to \infty$:

$$\lim_{n \to \infty} \rho_i(\boldsymbol{\Sigma\epsilon}) \leq C, \forall i$$

On the opposite, the common factors f_i are the principal components of asset returns that have a strong correlation with sufficiently many assets such that the eigenvalues of factor covariance approach infinite as $n \to \infty$.

Based on this definition, all the APT papers (including but not limited to my main references Ross (1976), Chamberlain (1983), Chamberlain (1983), Dybvig (1983), Connor and Korajczyk (1995)) assume the same diversification condition to rule out the impact of idiosyncratic shocks on expected returns. They assume that the market portfolio $\{w_i\}, i = 1...n$ is well-diversified, such that:

$$\lim_{n\to\infty} \sum w_i^2 = 0 \tag{5}$$

This definition of diversification implies no firm size dispersion as the number of assets approaches infinity. It is trivial to observe that with the diversification assumption, all the assets would have negligible weight in a market with sufficiently many assets. I formalize this argument in the following lemma:

Lemma 1. *If the market is well-diversified such that:*

$$\lim_{n\to\infty} \sum w_i^2 = 0$$

then all the firms must have their market weight converge to zero as $n \to \infty$:

$$\lim_{n\to\infty} w_i = 0, \forall i$$

The negligible market weight of an asset, implied by the diversification assumption, makes its idiosyncratic risk fail to impact expected returns. Intuitively, with diversification, idiosyncratic shocks have a negligible impact on the pricing kernel due to the weak correlation. In consequence, the idiosyncratic risk terms $COV(\epsilon_i, \sum_i^n w_i \epsilon_i)$ in expected returns, as derived in (3), converge to zero as number of assets approach infinity. In contrast, common factors in the asset covariance are not diversified away and explain the expected return in a linear structure as shown in the following lemma:

Lemma 2. *Suppose the market portfolio is well-diversified such that $\lim_{n\to\infty} \sum w_i^2 = 0$ and the risk structure among asset returns follow an APT model. In that case, the expected returns have a linear factor structure as $n \to \infty$:*

$$\lim_{n\to\infty} E[r_i] = \mu_0 + \sum_{s=1}^{k} \beta_{i,s}\mu_s$$

where $\mu_s, s = 1...k$ is the risk premium tied to each factor and $\beta_{i,s}$ is the asset i's exposure to *factors.*

In the **Appendix Section I**, I give a rigorous proof of **Lemma 2**, which describes the classic APT result: With diversification, the expected return of each asset converges to a linear function of the pervasive factors among asset returns. This simple and elegant structure is probably one of the most important results in asset pricing research. Researchers place a massive amount of effort on determining the right number of factors k as the number of assets n approaches infinity and, more importantly, on identifying the pervasive factors $f_s, s = 1...k$ and the associating risk premiums $\mu_s, s = 1...k$.

I show that the measure of diversification $\lim_{n\to\infty} \sum w_i^2$ relies on firm size distribution. Further, a thin-tailed distribution of firm size induces the diversification assumed in (5). Since the market weight w_i is scaled by the total market value to make $\sum_i^n w_i = 1$, I work on the un-scaled firm size X_i distribution instead and assume i.i.d distribution of X_i for simplicity. The portfolio in the market portfolio is:

$$w_i = X_i / \sum_{i=1}^{n} X_i$$

The diversification measure depends on the mean and variance of X_i such that:

$$\lim_{n\to\infty} \sum w_i^2 = \lim_{n\to\infty} \sum \frac{(X_i)^2}{(\sum X_i)^2} = \lim_{n\to\infty} \frac{1}{n} \frac{1/n \sum (X_i)^2}{(1/n \sum X_i)^2} \tag{6}$$

A thin-tail distribution has finite mean and variance, which invokes the Law of Large numbers (LLN hereafter) to meet the diversification condition assumed by APT in (5). I formalize this argument in the following lemma:

Lemma 3. *The distribution of market value X_i has a thin tail if its first and second moments are finite as the number of firms approaches infinity. A market portfolio with the thin tail distribution defined is well-diversified since:*

$$\lim_{n\to\infty} \sum w_i^2 = \lim_{n\to\infty} \frac{1}{n} \frac{E[(X_i)^2]}{E[X_i]^2} = 0$$

Lemma 3 reveals that the converge rate of the diversification measure $\sum w_i^2$ is $1/n$. A thin-tail firm size distribution implies a well-diversified market portfolio in (5) and further the linear factor model. With a thin-tail distribution, no firm-specific shock matters for the pricing kernel since every asset has negligible weight in the market. Therefore, only pervasive factors in the covariance drive the risk premium regardless of the portfolio composition, as concluded in APT models.

2.2 Pareto distribution and violation of APT

In contrast to the classic case assumed by APT models, when firm size distribution has a fat tail, the probability of having giant firms becomes non-trivial. The fat tail makes the first and second moments of X_i explode to infinity, and the diversification measurement in (6) does not converge to zero. In other words, with granularity, the presence of large firms would break the diversification and generate risk premium from the idiosyncratic risk in the format of $COV(\epsilon_i, \sum_i^n w_i \epsilon_i)$ as derived. Conceivably, the violation of APT raises a granularity effect in the stock market that the level of the fat tail should determine the magnitude of granular risk premiums in the expected returns. Thus, the granularity effect is an implication derived directly from the violation of the APT assumption. It is a crucial topic to explore, considering the vital status of APT models in asset pricing research.

I quantify this granular channel of expected returns by fitting the distribution of firms' market value X_i using Pareto distribution and measure the level of granularity by the Pareto coefficient ζ. The Pareto distribution has a survival function equal to:

$$P(X_i > x) = \left(\frac{x}{x_m}\right)^{-\xi}, x > x_m \tag{7}$$

A firm's portfolio weight w_i is the market value divided by the total value in the portfolio $X_i / \sum^i X_i$ as mentioned. The elegance of a Pareto distribution is that it parsimoniously describes the level of a fat tail by a single parameter $\xi > 0$. The Pareto coefficient ξ determines how fast the probability of a firm's size larger than a threshold x_m decreases as x approaches infinity. Therefore, a high Pareto coefficient ξ implies a low level of granularity. When $\xi > 2$, the distribution has a thin tail: The first and second moments of X are finite such that the diversification in (6) holds. A small $\xi < 2$ implies a high probability of firms with huge value in the distribution and means a high level of the fat tail. Specifically, the i moments of X are:

$$\begin{aligned} E[X^i] &= \infty, \zeta \leq i \\ &= \frac{\zeta x_m^i}{\zeta - i}, \zeta > i \end{aligned} \tag{8}$$

Using the Pareto distribution, I derive the limit of the diversification measure $\lim_{n\to\infty} \sum w_i^2$. The idea is that when the first and second moments of X_i approach infinite, the convergence rate of w_i^2 starts to decrease as the level of granularity increases, instead of being $1/n$ shown in **Lemma 3**.

2.2.1 Pareto distribution and failure of diversification

The failure of diversification occurs when $\zeta < 2$: the convergence of sample mean, and variance in (6) does not follow LLN since it could explode to infinite. To derive the convergence under this case, one needs to apply a generalized "stable law" theorem for infinite-variance random variables (see Durrett (2019), Theorem 3.8.2.). I derive the limit of the diversification measure $\lim_{n\to\infty} \sum w_i^2$ in the following lemma:

Lemma 4. *If the firm size X_i follows an i.i.d Pareto distribution defined in* (7) *and $\zeta < 2$, then the convergence in equation* (6) *is determined by ζ as follows.*

$$\begin{aligned} \lim_{n\to\infty} \sum w_i^2 &= c_1 \frac{Y_2+1}{(Y_1)^2}, \zeta < 1 & (9) \\ &= c_2 n^{2/\zeta-2} \frac{Y_2+1}{E[X]^2}, \zeta > 1 & (10) \\ &= c_3 \frac{Y_2+1}{(Y_1+\log n)^2}, \zeta = 1 & (11) \end{aligned}$$

where c_1, c_2, c_3 are constants under different range of ζ. Y_1 and Y_2 are two random variables that have stable distributions.

The derivation of **Lemma 4** is in **Appendix Section II**. Similar to ζ measured by firm fundamentals (Axtell (2001), Gabaix (1999), Gabaix (2011), Gabaix and Ibragimov (2011)), I found ζ estimated from stock market value is around 1. This estimate maps to the granularity case in **Lemma 4**: When $1 < \zeta < 2$, the diversification rate of idiosyncratic shocks is $n^{2/\zeta-2}$, which is way slower than $1/n$ under the thin tail case when $\zeta > 2$. For example, let $n = 10^5$ and $\zeta = 1.1$. Under this case, $2/\zeta - 2 \approx -0.2$ and the convergence rate of diversification is roughly $n^{-1/5} = 1/10$ instead of $1/n = 1/10000$. When $\zeta = 1$, the convergence rate is $1/(\log n)^2$, which is a limitation of the $n^{2/\zeta-2}$ as ζ approaches 1, as a special case. When the level of granularity is more extreme ($\zeta < 1$), the diversification measure $\sum w_i^2$ does not depend the number of asset n but on the non-degenerate random variables Y_1 and Y_2. When the fat-tail is extreme, large values would drive the infinite moments and break the convergence in LLN. These large values dominate the size variation in $\sum w_i^2$ and produce convergence terms Y_1 and Y_2, which follow stable distributions.

2.2.2 Pareto distribution and large firms

Given the heuristic argument that large values would dominate the size variation of w_i, large firms in a fat-tailed distribution of size would account for a significant fraction of the total market value. Specifically, The non degenerate convergence term Y_1 in **Lemma**

4 comes from the extremely large values such that $X_i > a_n$

$$a_n = \inf\{x : P(X_i > x) \leq n^{-1}\} = n^{1/\xi}$$

In the stock market context, a_n represents the value of the largest firms, which appears at a frequency around $1/n$. Large firms with a market value more enormous than a_n would break the diversification since it drives the average market value to infinite. As a result, a firm with extreme size a_n would have dominating weights in the market portfolio since it drives most of the total market value. I formalize this argument as follows:

Lemma 5. *If the firm size X_i follows an i.i.d Pareto distribution defined in* (7) *and $\zeta < 2$, then a large firm with size equals to a_n would have its market weight $\frac{a_n}{\sum X_i}$ converges to*

$$\begin{aligned} \lim_{n \to \infty} \frac{a_n}{\sum X_i} &= 1/(1 + Y_1), \xi < 1 && (12) \\ &= n^{1/\zeta - 1}(1 - 1/\zeta), \xi > 1 && (13) \end{aligned}$$

Y_1 is a random variable that has a stable distribution.

Using the same example, let $n = 10^5$ and $\zeta = 1.1$, then the weight of large firms converges to around 0.03. Therefore, my model uses a Pareto distribution to show that with a fat tail, the largest firms have non-negligible market weight.

In summary, I quantify the level of granularity by a Pareto distribution and show how a fat-tailed distribution violates the APT assumption. Precisely, the employment of Pareto distribution quantifies two violations of APT assumption in the market portfolio composition. In cross-section, Large firms have non-negligible market weights $\lim_{n\to\infty} w_i \neq 0$. On aggregate, the firm size variation is non-trivial, which breaks the diversification of APT such that $\lim_{n\to\infty} \sum w_i^2 \neq 0$. These two results give immediate asset pricing implications, making idiosyncratic risk explain the expected returns in cross-section and aggregate.

2.3 Asset pricing implications of granularity

I now combine the results from the Pareto distribution with the asset pricing equations in (3) and (4) to produce testable results for expected returns.

2.3.1 granularity and the idiosyncratic risk puzzle

I use the result in **Lemma 5** to establish asset pricing implications in the cross-section. Idiosyncratic risks of large firms such that $\lim_{n\to\infty} w_i \neq 0$ should not be diversified and generate risk premiums in the format of $COV(\epsilon_i, \sum_i^n w_i\epsilon_i)$ as derived in (3). To emphasize the impact of large market weight w_i, I further assume that idiosyncratic shocks among assets are independent, which gives the following result:

Proposition 6. *With granularity, there exist large firms s.t.* $\lim_{n\to\infty} w_i \neq 0$ *as shown in* ***Lemma 5****. If the idiosyncratic shocks are independent of each other with variance* θ_i*, then the expected return for each asset converges to:*

$$\lim_{n\to\infty} E[r_i] = \mu_0 + \sum_{s=1}^{k} \beta_{i,s}\mu_s + \gamma w_i \theta_i \tag{14}$$

Assuming independence among ϵ in **Proposition 6** simplifies the empirical test of my model implication. Identifying the idiosyncratic shocks ϵ_i and testing whether the covariance in $COV(\epsilon_i, \sum_i^n w_i\epsilon_i)$ explains the expected returns of assets might suffer from omitted factor bias (see Giglio and Xiu (2021)), or the lack of power due to weakly identified factor models (Giglio, Xiu, and Zhang (2021)). Instead, measuring the variance of idiosyncratic shocks θ_i provides convenience and robustness relative to the selection of factor models. From this perspective, most of the variance in the asset returns is idiosyncratic. Hence the magnitude of θ measured relative to various factor models must not change dramatically. Further, the analysis based on (14) only requires measuring the relative ranking of θ_i and $w_i\theta_i$ in the cross-section, which avoids the issue of miss-measuring the magnitude of idiosyncratic variance due to improper factor model selection.

In terms of theoretical insight, **Proposition 6** points out that it should be the size-adjusted idiosyncratic risk $w_i\theta_i$ instead of itself θ_i that explains expected returns. This insight suggests that only large firms $\lim_{n\to\infty} w_i \neq 0$ could have their idiosyncratic shocks un-diversified to generate expected returns such that $\lim_{n\to\infty} w_i\theta_i \neq 0$. The product of firm size and idiosyncratic variance determines the magnitude of abnormal returns relative to APT factor models, or a ″granular alpha″:

$$\alpha_i = \gamma w_i \theta_i$$

Notably, an asset's market weight determines the marginal impact of idiosyncratic risk on expected returns. Large firms have a high alpha per unit of idiosyncratic variance since being″ large″ must require compensation in terms of pricing and make the expected returns exhibit more of the idiosyncratic risk premium. This effect is different from a size factor in Fama and French (1992), which states that small firms commonly have higher expected returns due to a higher variance of returns than large firms. In my framework, a ″small minus big″ portfolio can be interpreted as an APT-defined factor since it captures the pervasive pattern in the return covariance.

More importantly, **Proposition 6** explains the ″idiosyncratic risk puzzle″ (IRP hereafter) that there is a very robust negative relationship between idiosyncratic variance and future returns, investigated in Ang et al. (2006) and Ang et al. (2009). As in their papers, a typical test of whether idiosyncratic risks matter in the cross-section is to estimate a linear regression between α_i (expected returns unexplained by factors) and the idiosyncratic risk θ_i:

$$\alpha_i = constant + \eta\theta_i$$

The estimate of $\hat{\eta}$ is documented to be negative, which seems puzzling since there should not be a negative risk-return relation in asset prices.

If the expected returns follow the structure implied by my model, the estimate of η would capture the correlation between the size-adjusted idiosyncratic risk $w_i\theta_i$ and the risk itself θ_i instead of the relation between risk and return. In other words, the estimate η in IRP is proportional to the correlation $corr(w_i\theta_i, \theta_i)$:

$$\eta \propto corr(w_i\theta_i, \theta_i)$$

Accordingly, it is possible that performing cross-sectional tests for whether idiosyncratic risk explains the expected returns without adjusting for w_i can generate model miss-specifications. With a thin-tailed distribution of firm size, this miss-specification does not induce a misleading empirical conclusion since there is no significant size difference in the cross-section. For example, if all the assets have the same market weight such that $w_i = 1/n, \forall i$, then the estimate of η equals:

$$\eta = \frac{1}{n}\gamma > 0$$

However, when the granularity is significant, large firms that populate the fat tail account for most of the market valuation, and small firms have negligible market weights. This huge size difference due to granularity makes the correlation between $w_i\theta_i$ and θ_i dominated by the correlation between w_i and θ_i. This correlation $corr(w_i, \theta_i)$ is negative as a feature of data, which is found in the cited papers and my empirical test. Consequently,

$$\eta \propto corr(w_i, \theta_i) < 0$$

With granularity, large firms (low idiosyncratic risk) have a significantly higher risk premium tied to their idiosyncratic risks than small firms (high idiosyncratic risk). In other words, firms with high idiosyncratic risks tend to have negligible market weights and low risk premiums raised by idiosyncratic risks, which drives the puzzling empirical results in IRP.

2.3.2 granularity and the market risk premium

As the extension of the cross-sectional implication, large firms populate the fat tail and violate the diversification in (5), which makes the level of granularity increase idiosyncratic risks un-diversified on aggregate and hence affect the market risk premium $E[r_m]$. I formalize this intuition in **Proposition 7**:

Proposition 7. *If the idiosyncratic shocks are independent of each other with variance θ_i, then the expected return for the aggregate market converges to:*

$$\lim_{n\to\infty} E[r_m] = \mu_0 + \sum_i^n w_i \left(\sum_{s=1}^k \beta_{i,s}\mu_s \right) + \gamma \lim_{n\to\infty} \sum w_i^2\theta_i \tag{15}$$

The diversification assumption ensures the aggregate impact of idiosyncratic risk $\sum w_i^2\theta_i$ converges to zero since all the assets should have bounded variance such that $\theta_i \leq \theta_{\max}$, hence:

$$\lim_{n\to\infty} \sum w_i^2\theta_i \leq \theta_{\max} \lim_{n\to\infty} \sum w_i^2 = 0$$

In contrast, granularity fails the diversification and affects the magnitude of the market expected returns tied to idiosyncratic risks.

I decompose the granular term $\sum w_i^2\theta_i$ into two parts to emphasize the aggregate impact of granularity, such that:

$$\sum w_i^2\theta_i = \sum w_i^2 \left(\sum \frac{w_i^2}{\sum w_i^2}\theta_i \right)$$

This decomposition reveals that two channels determine the market expected return tied to idiosyncratic risk: The level of granularity captured in $\sum w_i^2$ as an indicator of the under-diversification, and the level of idiosyncratic risk captured in $\left(\sum \frac{w_i^2}{\sum w_i^2}\theta_i \right)$ as a weighted-average of idiosyncratic risk. My derivations using the Pareto distribution highlight the first channel, which derives the convergence of $\sum w_i^2$ as a function of ζ. As shown in **Lemma 4**, a lower Pareto coefficient ζ (higher granularity) indicates less diversified idiosyncratic risks in the market portfolio and more risk premium on aggregate. The second channel relates to whether time-variation of idiosyncratic risk explains the market expected returns in literature (see Goyal and Santa-Clara (2003), Bali et al. (2005)). I estimate the Pareto coefficient ζ by fitting the fat-tail in firm size distribution each month and find that ζ is time-varying with an average value around 1. This finding suggests a granular channel of market variation besides the time-varying idiosyncratic risk documented in the literature.

Therefore, **Proposition 7** motivates a time-series implication to test whether ζ generates additional time-variation of market risk premium, controlling the magnitude of idiosyncratic risk. Taking log of the granular term $\sum w_i^2 \theta_i$, by the decomposition, gives a linear relation:

$$\log(r_{m,t+1}) = \text{constant} + \text{controls} + \mathrm{A} \log \zeta_{\mathrm{t}} \tag{16}$$

My model implies $A < 0$ since ζ decreases the magnitude of the market expected returns tied to idiosyncratic risks.

3 Empirical Test

3.1 Data

My cross-sectional test is at the monthly frequency from June 1963 to December 2020. Specifically, the test relies on an estimate of each firm's idiosyncratic variance and factor exposures each month. I run daily return data on the first three principal components/Fama-French three factors to estimate these variables per month. I use return and firm size data in the CRSP and other characteristic data in COMPUSTAT for control variables. I merge the monthly CRSP return data and annually/quarterly COMPUSTAT characteristics data. I use a standard timing convention of leaving a six-month lag between the quarter end of characteristics and the monthly returns to ensure the constructed variables are available. Fama-French factors and other sorted portfolios are from the Kenneth French data library.

As additional controls in the time-series test, I include the predictors from Welch and Goyal (2008), available from 1945 to 2020. I test whether $\log \zeta_t$ captures the time variation of the market expected returns in this sample period.

3.2 Cross Section Test

My result in **Proposition 6** states that the alpha relative to factor models should depend on size-adjusted idiosyncratic risk:

$$\alpha_i = \gamma w_i \theta_i$$

Intuitively, I conduct empirical tests to study the cross-sectional relation between α_i, w_i and θ_i. Furthermore, since this result explains the IRP (as in Ang et al. (2006) and Ang et al. (2009)), I construct my tests based on the same measurement of θ_i and α_i. To start with, I replicate their findings as a benchmark result to document that performing cross-sectional tests for whether idiosyncratic risk explains the expected returns without adjusting for w_i can generate misleading empirical results. Then I add the size adjustment implied by my model to show that granularity helps identify a positive relation between idiosyncratic risk and returns.

3.2.1 Portfolio level test using 5 portfolios sorted by idiosyncratic risk

Like Ang et al. (2006), I sort all the assets by their idiosyncratic risk θ measured by daily returns in each month using Fama-French 3 factors (FF3 hereafter). Then I split all the assets into five quintiles to construct five value-weighted portfolios sorted by the idiosyncratic variance measured in the last month $\theta_{i,t-1}$.

I report results using the five idiosyncratic risk sorting portfolios in **Table 1**. First, I report the mean and volatility (annualized, in percent) of excess returns in each portfolio, together with the total market weight of assets in each portfolio as a measure of the average size in Panel A. I found the same pattern as documented in Ang et al. (2006), the lowest risk portfolio r_L tends to have a significantly higher return than the highest r_H:

$$E[r_L - r_H] > 0$$

The annualized return spread between the lowest and the highest equals 7.23 percent with significance. Furthermore, assets in the portfolio with the lowest risk account for roughly 60 percent of the total market value, which indicates a significant size difference in the cross-section due to granularity. In addition, as the idiosyncratic risk increases from the lowest row to the highest, the size of firms in each quintile decreases. As I explained in the theoretical derivations, this negative relationship between risk and size is an essential feature of data to reconcile the IRP.

To further test the granularity's impact on expected returns, I examine the relation between α_i, w_i, and θ_i in the five portfolios. In Panel B, I measure the post-sample alpha and idiosyncratic volatility relative to FF3 as the benchmark model. The alpha spread between the lowest and the highest is 12.6 percent with significance. The negative return spread observed in Panel A is not explained by factors. From the granularity perspective, assets with low idiosyncratic risk θ_i but have high market weight w_i, which suggests a high ratio of alpha to idiosyncratic variance since the model implies:

$$\frac{\alpha_i}{\theta_i} = \gamma w_i$$

To verify the model implication, I find a decreasing α/θ ratio from the first row to the last. For robustness, I also present the same test using the CAPM in Panel B, using the three principal components of asset returns (PCA) as factors in Panel C. These results reveal the same pattern: As θ_i increases, both the alpha α_i and the market weight w_i decrease. In terms of the granular alpha implied by my model, the α_i/θ_i also decreases due to decreasing w_i. This result depends on large firms having non-negligible market weight and the high marginal impact of idiosyncratic risk on expected returns.

Therefore, the cross-sectional results above suggest that large firms provide more compensation for the investor to bear each unit of idiosyncratic risk. An immediate implication of this argument is to take advantage of the high marginal risk-payoff due to high market weight and construct a long-short trading strategy accordingly. I construct the "bet on granularity" portfolio by leveraging a long position of the lowest θ portfolio with excess return $r_L - r_f$ (large firms) and short the highest θ portfolio with excess return $r_H - r_f$ (small firms). The long-short strategy is constructed as follows:

$$r_{L-H,t} = \frac{1/\theta_{L,t-1}}{1/\theta_{L,t-1} - 1/\theta_{H,t-1}}(r_{L,t} - r_f) - \frac{1/\theta_{H,t-1}}{1/\theta_{L,t-1} - 1/\theta_{H,t-1}}(r_{H,t} - r_f) \tag{17}$$

This portfolio leverages the large firms (lowest θ) by the inverse of θ to capture the high marginal impact of their idiosyncratic risk. I update the portfolio per month and estimate the $\theta_{H,t-1}$ and $\theta_{L,t-1}$ by the average idiosyncratic variance within each quintile. The resulting denominator $1/\theta_{L,t-1} - 1/\theta_{H,t-1}$ is positive and normalizes the portfolio return to be dollar-neutral. Given the negative relation between firm size w_i and idiosyncratic variance θ_i, the "bet on granularity" portfolio should generate a positive return spread unexplained by the factor model such that:

$$\alpha_{L-H} = \frac{w_L(large) - w_H(small)}{1/\theta_L - 1/\theta_H} > 0$$

The positive alpha captures the size spread between portfolios with low and high idiosyncratic risk such that $w_L(\text{large size}) - w_H(\text{small size}) > 0$. The long-short strategy using portfolios summarized in **Table 1** has an annualized average return equal 7.36 percent and a volatility equal 13.60 percent. In addition, this positive return is not explained by factor models used as controls. The long-short strategy has a 1.49 percent alpha relative to FF3 factors with significance and a similar magnitude of alpha relative to CAPM and PCA factors.

The cross-sectional tests using these five portfolios are robust to a longer measurement window of θ and alternative factor models to measure θ. This robustness using the five-portfolios setting is not a surprise: The analysis only requires measuring the relative ranking of θ_i in the cross-section, which avoids the issue of miss-measuring the magnitude of idiosyncratic variance due to improper measurement window or factor model selection.

I construct the five portfolios using idiosyncratic risk measured by the three principal components of daily returns in each month and report similar results in **Table 2**. Also, I construct portfolios using the idiosyncratic variance measured by the daily returns in the past 3,6, and 12 months and find the same pattern. In **Table 3**, I summarize the "bet on granularity" portfolio constructed by the idiosyncratic variance measured by the daily returns in the past 1,3,6 and 12 months. The results using Fama-French 3 factors and three principal components of daily returns in the estimation window are listed in Panel A and B, respectively. All the long-short portfolios formed by estimation of the past 1,3,6 and 12 months generate positive alphas relative to the benchmark models, which verifies the insight that large firms have high marginal impacts of idiosyncratic risk on expected returns.

3.2.2 Portfolio level test using 100 portfolios sorted by idiosyncratic risk

The above results replicate findings in Ang et al. (2006) and test my theoretical insight by constructing a long-short portfolio. I further explain the IRP by estimating the cross-sectional relation between α_i, θ_i, and w_i, which requires a bigger cross-section for statistical power. Therefore, I extend the 5-portfolio setting to split all the assets by percentiles of θ to construct 100 value-weighted portfolios. For each portfolio $i = 1, .., 100$, I estimate a FF3 factor model to compute the post-sample α_i, θ_i (annualized, in percent) and also the summed market weight w_i of assets in the portfolio. I use the 100 portfolios to present the ability of size-adjusted idiosyncratic variance $w_i\theta_i$ to explain alphas and reconcile the idiosyncratic risk puzzle.

I start with estimating a typical test of risk-return relation in IRP:

$$\alpha_i = constant + \eta\sqrt{\theta_i}$$

Follow Ang et al. (2006) and Ang et al. (2009), I use the idiosyncratic volatility as the explanatory variable, which is the square root of θ_i. The estimate of $\hat{\eta} = -0.74$ with a standard error $std(\hat{\eta}) = 0.039$. This significantly negative estimate confirms the IRP that there is a negative relation between θ_i and α_i in the cross-section. I compare the IRP specification to the granular alpha implied by my model:

$$\alpha_i = constant + \gamma w_i \theta_i$$

The estimate of $\hat{\gamma} = 5.17$ with a standard error $std(\hat{\gamma}) = 0.59$. This estimate is consistent with what the model implies since a positive estimate of $\hat{\gamma}$ represents the risk-aversion coefficient.

Furthermore, I use the 100 portfolios to illustrate how the granular impact of idiosyncratic risk explains the IRP. If the expected returns follow the structure implied by my model, the estimate of η would capture the correlation between the size-adjusted idiosyncratic risk $w_i\theta_i$ and the risk itself $\sqrt{\theta_i}$ instead of the relation between risk and return. The correlation estimated in the 100 portfolios indicates a negative relation between the size-adjusted idiosyncratic risk $w_i\theta_i$ and the risk itself $\sqrt{\theta_i}$ such that

$$corr(w_i\theta_i, \sqrt{\theta_i}) = -0.72$$

Intuitively, this negative correlation is driven by the relation between market weights w_i and idiosyncratic risk $\sqrt{\theta_i}$. The correlation between size and risk, under this context, equals to:

$$corr(w_i, \sqrt{\theta_i}) = -0.68$$

As explained in my theoretical derivations, the negative size-risk relation, combined with granularity, explains the IRP. Without the significant size difference in the cross-section, the impact of w_i would be negligible. In contrast, with granularity, the huge size difference in w_i makes the correlation between $w_i\theta_i$ and θ_i dominated by the correlation between w_i and θ_i. With granularity, large firms (low idiosyncratic risk) have a significantly higher risk premium tied to their idiosyncratic risks than small firms (high idiosyncratic risk). In other words, firms with high idiosyncratic risks tend to have negligible market weights and low risk premiums raised by idiosyncratic risks, which drives the puzzling empirical results in IRP.

I examine the robustness of my 100-portfolio results for different lengths of measurement window and using different factor models. In **Table 4**, I summary the estimate of η, γ and the estimated correlations $corr(w_i\theta_i, \sqrt{\theta_i})$, $corr(w_i, \sqrt{\theta_i})$ using portfolios formed by the idiosyncratic variance measured by the daily returns in the past 1,3,6 and 12 months. The results using Fama-French 3 factors and three principal components of daily returns in the estimation window are listed in Panel A and B, respectively. All the estimates using different formation periods are significant and consistent with granular alpha channels for idiosyncratic risk to explain the expected returns of my model.

Interestingly, the magnitude of $\hat{\eta}$ and $\hat{\gamma}$ decreases as the measurement window increases. Also, both the two correlations $corr(w_i\theta_i, \sqrt{\theta_i})$ and $corr(w_i, \sqrt{\theta_i})$ decrease as the measurement window increases. Ang et al. (2009) tested the IRP at the individual asset level and found a similar pattern that the significance of η decreases as window length increases. They explain this issue by stating that the rankings of idiosyncratic volatility (relative magnitude in the cross-section) change across longer sample periods. The cross-sectional relation between α_i and θ_i may not remain over longer periods since an asset with high θ in a certain month may not continue to have high θ in the next 3, 6, or 12 months. Compared to the individual asset level test, the portfolio setting is more regulated since the portfolio-sorting forces the ranking of idiosyncratic variance among portfolios to be fixed. From this perspective, it is useful also to test the stability of my model implication at the individual asset level and compare the results to IRP tests in Ang et al. (2009).

3.2.3 Individual asset level test for explaining idiosyncratic puzzle

The portfolio level tests extend the results in Ang et al. (2006) and explain the IRP. I generalize the portfolio level test to individual asset levels following the same construction in Ang et al. (2009). I replicate their specification:

$$r_{i,t} = \mu_0 + \sum_{s=1}^{k} \beta_{i,s,t} \left(f_{s,t} + \mu_s\right) + \eta \sqrt{\theta_{i,t-1}} + \epsilon_{i,t} \tag{18}$$

They test the cross-sectional relation between expected returns and idiosyncratic risk with time-varying parameters and apply a Fama-Macbeth regression using monthly data to estimate $\hat{\eta} < 0$.[5]

To compare to the test in Ang et al. (2009), I generalize (14) to be time-varying and

[5] The negative return spread between the highest and the lowest portfolios sorted by $\sqrt{\theta_{i,t-1}}$ in Ang et al. (2006) implicitly confirms a negative estimate of $\hat{\eta} < 0$.

estimate:

$$r_{i,t} = \mu_0 + \sum_{s=1}^{k} \beta_{i,s,t} \left(f_{s,t} + \mu_s \right) + \gamma w_{i,t-1} \theta_{i,t} + \epsilon_{i,t} \tag{19}$$

This specification originates from extending the single period competitive equilibrium derived in my model to multiple periods similar to Merton (1973). I assume a special case that parameters $\beta_{i,s,t}, ..., \theta_{i,t}$ (from conditional covariance among asset returns) change over time with i.i.d distribution not driven by any state variable, which leads to the cross-sectional specification in (19). The size-adjusted idiosyncratic risk $w_{i,t-1}\theta_{i,t}$, in this context, approximates the time-varying covariance between idiosyncratic shocks $\epsilon_{i,t}$ and the weighted average $\sum_{i=1}^{n} w_{i,t-1}\epsilon_{i,t}$, which is similar to the time-varying factor loading $\beta_{i,s,t}$.

My setup is the same with Ang et al. (2009) in (18) except that they use the past idiosyncratic volatility $\sqrt{\theta_{i,t-1}}$ as the explanatory variable to document the IRP. I estimate $\hat{\eta} < 0$ to replicate the IRP results and compare it to the estimate of $\hat{\gamma} > 0$ in my model. The comparison between $\hat{\gamma}$ and $\hat{\eta}$ emphasizes that one should include both the idiosyncratic risk and marginal impact of idiosyncratic risk determined by w_i to test the risk-return relation in the cross-section.

As in theirs, I apply the two-step Fama-Macbeth estimation procedure. In the first step, I run factor regressions (FF3 as in Ang et al. (2009)) to the daily returns of each asset in each month. This procedure gives estimates of factor exposures $\beta_{i,s,t}$ and the size-adjusted idiosyncratic variance $\theta_{i,t}$ per month. Then in the second step, I use the factor exposures and the size-adjusted idiosyncratic risk of each asset $w_{i,t-1}\theta_{i,t}$ estimated to explain the cross-sectional variation of expected returns. The second step gives an estimate of $\hat{\gamma}_t$ in each month, and the estimate of $\hat{\gamma}$ the average value of all the estimates in each sample period, such that:

$$\hat{\gamma} = 1/T \sum_{t=1}^{T} \hat{\gamma}_t$$

As in typical Fama-Macbeth regressions, I use the simultaneous risk exposure $\hat{\beta}_{i,s,t}$ and $w_{i,t-1}\hat{\theta}_{i,t}$ estimated from the first step to identify factor risk premium μ_s and the risk aversion coefficient γ. I use the lagged weight $w_{i,t-1}$ to avoid the mechanical correlation between the holding period return $r_{i,t}$ and the market weight at the end of each month $w_{i,t}$. Further, I control the lagged characteristics since they also tend to explain the cross-sectional variation of expected returns suggested by Daniel and Titman (1997). I control the lagged book-to-market ratio and the momentum factor computed by the sum of returns in the last six months as in Jegadeesh and Titman (1993).

In **Table 5**, I report the cross-sectional regression estimates $\hat{\eta}$ (using $\sqrt{\theta}_{i,t-1}$) and $\hat{\gamma}$ (using $w_{i,t-1}\theta_{i,t}$) separately. In column 1, I estimate a significant negative coefficient $\hat{\eta} = -0.01$, which is consistent with the Ang et al. (2009) result. Conversely, the main result in column 4 shows a significantly positive estimate of $\hat{\gamma} = 9.15$, which suggests the importance of using size-adjusted idiosyncratic variance to identify a positive risk-return relation. For robustness, I also report several other specifications. In the second specification reported in columns 2, I use both $w_{i,t-1}$ and $\sqrt{\theta_{i,t-1}}$ as two variables to explain the returns. The coefficient for $\sqrt{\theta_{i,t-1}}$ is still negatively significant with the size controlled. In column 3, I use the firm size $w_{i,t-1}$ as the only explanatory variable besides the factor exposures and characteristics. The estimate in column 3 shows an insignificantly positive coefficient for $w_{i,t-1}$ since it does not control the magnitude of idiosyncratic risk θ_i but only uses the marginal impact of θ_i as suggested by my model. The specifications in column 2 and 3 does not identify a positive risk-return relation either, which emphasize the importance of using the right functional form $w_{i,t-1}\theta_{i,t}$ since it is a proxy for the covariance with the granular shocks in the pricing kernel. In column5, I test a specification using both the $\sqrt{\theta_{i,t-1}}$ and $w_{i,t-1}\theta_{i,t}$. The estimates for this specification show the same significance of $\hat{\eta} < 0$ and $\hat{\gamma} > 0$, which suggests the robustness of using size-adjusted idiosyncratic variance to identify a positive risk-return relation.

In **Table 6**, I show similar results using the the three principle components to measure the factor exposures and idiosyncratic variance, which confirms the sign of $\hat{\gamma}$ and $\hat{\eta}$. Specifically, I estimate the three principal components from all the asset returns available each month and then use them to perform the two-step estimation. The $\hat{\eta}$ using PCA has a similar magnitude of around -0.01, and the $\hat{\gamma}$ is 6.73. The principal components method is consistent with my theoretical definition of factors and idiosyncratic risk in the APT models. I report results using both Fama-French factors and PCA to reconcile the method in Ang et al. (2009) and verify my factor definition's robustness.

For the robustness of my result to explain the idiosyncratic volatility puzzle, I expand the estimation window of idiosyncratic variance similar to the results in Ang et al. (2009). Specifically, I use a rolling-window estimation each month to include all the daily returns for the past 3,6 and 12 months to estimate factor models in the first step of the FM regression. Then I repeat the second step to estimate $\hat{\eta}$ and $\hat{\gamma}$. In **Table 7**, I present the results of measuring $\theta_{i,t}$ by different estimation windows and running the same specification of column 5 in **Table 5**. The results using Fama-French 3 factors and three principal components of daily returns in the estimation window are listed in Panel A and B, respectively. Similar to findings in Ang et al. (2009), the significance of $\hat{\eta}$ and $\hat{\gamma}$ decreases as window length increases. The 3-month estimation results in both Panel A and B still show a significantly negative η and positive γ. However, as the window length increases to 6/12 months, the significance of η and γ decreases. It is not a surprise that the individual asset level result is less robust to a longer measurement window than the portfolio level since the rankings of idiosyncratic volatility change across longer sample periods. My model promotes this argument to be more concrete by linking the alpha to size-adjusted variance:

$$\alpha_{i,t} = \gamma w_{i,t-1} \theta_{i,t}$$

The cross-sectional ranking of α_i depends on both the size w_i and the idiosyncratic variance θ. The $\theta_{i,t}$ has time-series persistence. Yet, it also has big fluctuations due to random shocks over time, which could make the cross-sectional ranking of idiosyncratic risk unstable over a longer sample period, as argued in Ang et al. (2009). However, the ranking of size $w_{i,t-1}$ should be more persistent over time since large firms tend to stay large over 12 months or longer. Based on this point, results in **Table 7** show that the lag firm size $w_{i,t-1}$ has an increasing significance to explain expected returns as the window length expands to 12 months.

3.3 Time-Series Test

3.3.1 Estimate the Pareto coefficient

My paper's main results hinge on the Pareto coefficient ζ value, which quantifies the asset pricing implication of granularity. I estimate the tail parameter ζ of the Pareto distribution using the Hill estimator (see Hill (1975)). At each month t, I sort all the n_t firm sizes in a descending order $X_{i=1...n_t,t}$ and select a threshold value $X_{k_t,t}$ to use the largest k_t firms for estimating ζ_t. The Hill estimator is:

$$\zeta_t = \left\{ 1/k_t \sum_{i=1}^{k_t} (\log X_{i,t} - \log X_{k_t,t}) \right\}^{-1} \tag{20}$$

When $X_{i=1...n}$ are i.i.d and follows the exact Pareto distribution in (7), this estimator can be interpreted as a maximum likelihood estimator of ζ conditioning on a known minimum threshold $X_{k_t,t}$, which has a simple to derive asymptotic inference property as $k_t \to \infty$. Therefore, the literature typically selects the threshold position k by fixing a cutoff ratio $k/n = 5\%, 10\%...$ to make k proportional to the total number of assets n and conduct the statistical inference by the asymptotic property of the estimator as $n \to \infty$.

For time-series implication in my paper, the cutoff selection affects the predictability of ζ on market returns as motivated in (16):

$$\log(r_{m,t+1}) = \text{constant} + \text{controls} + \text{A} \log \zeta_{\text{t}}$$

A low cutoff ratio k/n would include more firms and reduce the estimator's variance for better statistical power of my time-series test. However, a low cutoff also generates a downward bias of ζ since it could include small firms in the sample that may not follow the Pareto distribution. To illustrate this point, I use the largest 20% firms in December 2020 to fit a Pareto distribution. The Pareto distribution implies a linear logged rank-size relation in (7) since:

$$\log(i/n) \approx \log\left(\left(\frac{X_i}{X_k}\right)^{-\zeta}\right) = -\zeta\left(\log X_i - \log X_k\right)$$

If these largest 20% firms follow a Pareto distribution, there must be a linear relation between their logged rank and size. Therefore, to check the goodness of fitting, I plot the logged rank-size plot of the largest 20% firms in the December of 2020 in **Figure 3**. I fit the linear relationship in the red dash line using the Hill estimator of ζ and find a slight deviation from the straight line with concavity. The concavity comes from including firms smaller than the size implied by the Pareto distribution, which induces a downward bias of the Hill estimator.

Due to the downward bias, a time-series estimate of ζ would be non-stationary since its variance and magnitude depend on the number of assets n. I estimate ζ using the largest 20% firms in each month to form a time-series of ζ_t and plot it in **Figure 4**. I plot the estimate of ζ_t in the blue line, together with the confidence interval (+/- two times the standard errors of ζ_t as a maximum likelihood estimator) in the two red lines below and above. **Figure 4** shows that the estimates of $\hat{\zeta}_t$ have higher standard errors at the beginning of the sample period due to fewer observations. As the number of firms included increases over time, the standard errors decrease, but the downward bias increases due to more small firms included in the estimation. Notably, there are two downside jumps of ζ_t in June 1962 and January 1973 due to the merging of AMEX-listed and NASDAQ-listed firms. In summary, I find that the average estimate of $\hat{\zeta}_t$ using the largest 20 % firms is around 1, which verifies the significant level of granularity used in my asset pricing results. However, the time-series estimate tends to have downward biases and hence a decreasing trend due to the increasing n in the sample period.

To construct a stationary estimate of ζ_t, I firstly test a co-integration relation between $\log \hat{\zeta}_t$ and the logged number of firms $\log n_t$ in the data each month presented in **Figure 5**. I apply an augmented Engle-Granger two-step co-integration test and find a p-value equal to 0.023. I take advantage of the co-integration relation between $\log \hat{\zeta}_t$ and $\log n_t$ and subtract the non-stationary trend due to an increasing number of firms over the sample period and then take the de-trended ζ_t into (16) to estimate:

$$\log(r_{m,t+1}) = \text{constant} + \text{controls} + \text{A} \log \zeta_\text{t}(\text{debias})$$

In this time-series context, I use a 10-fold cross validation to select an optimal cutoff ratio k/n, which gives the best out-of-sample predictability of the de-biased $\log \zeta_t$.

To adjust for the bias-variance issue, a vast amount of papers assume a more general class of fat tail distribution to develop the bias-correction methods accordingly (see Hall and Welsh (1985), Diebold, Schuermann, and Stroughair (1998), Peng (1998), Beirlant et al. (1999), Feuerverger and Hall (1999), Gomesa and Martins (2002), Alves, Gomes, and de Haan (2003)). Instead of applying these bias-correction methods for $\hat{\zeta}_t$ at each time separately, my "de-bias" procedure takes advantage of the co-integration and intends to improve the power of testing whether the level of fat tail predicts the market returns.

3.3.2 Single variable prediction results

In this section, I test whether the Pareto coefficient predicts market return at a monthly frequency:

$$\log(r_{m,t+1}) = \text{constant} + \text{controls} + \text{A} \log \zeta_{\text{t}}(\text{debias})$$

The hypothesized predictive coefficient A should be significantly negative since a low ζ_t indicates a high level of granularity and high risk premium in the market returns.

In **Table 8**, I present the single variable regression that the granular predictor $\log \zeta_t$ predicts the logged excess market return $r_{m,t+k}$ at various horizon and different sub-samples. I use this single variable regression as a benchmark result and control other predictors later for comparison. In the first panel of **Table 8**, I report the results using the whole sample at various horizon $k = 1, 12, 60$: The one-period ahead predictive coefficient is -0.28 with a significant t-stat value of -2.11. I also report the coefficient to correct the Stambaugh bias due to high serial correlation in $\log \zeta_t$ (see Stambaugh (1999)). The prediction significance remains in the long horizon for $k = 12, 60$. To check the predictive power of ζ out of the sample, I also compute the out-of-sample R^2 by comparing the predictive error of $\log \zeta$ to the historical mean computed by a rolling window. The out-of-sample R^2 reaches 1.50 percent at the 12-month horizon and 13.34 percent at the 60-month horizon, which indicates a robust predictive power of ζ in the long period.

In addition to the full sample result, I compare the predictive power of ζ_t in various sub-periods. In Panel B of **Table 8**, I compare the predictive power in the NBER recessions versus non-NBER recessions. The predictive coefficients (-1.05 v.s -0.28) and T-stat values (-2.59 v.s. -2.11) are more significant in the recession period than in the whole sample. In contrast, the predictive power in the non-recession period is weaker. This difference provides additional evidence that the granularity explains the aggregate risk premium, which is highly counter-cyclical and drives significant stock market downturns. My sub-sample result is consistent with the argument in Cujean and Hasler (2017) that the predictive power is concentrated in bad times.

3.3.3 Control for time-varying idiosyncratic risks

My empirical test above is motivated by the granular channel of market variation that ζ reflects how much of the idiosyncratic risks are un-diversified. This channel relates to whether time-variation of idiosyncratic risk explains the market expected returns in literature (see Goyal and Santa-Clara (2003), Bali et al. (2005)). Therefore, I further test whether ζ generates additional time-variation of market risk premium, controlling the magnitude of idiosyncratic risk. I measure the level of idiosyncratic risk at the aggregate level by the weighted average of θ. Specifically, I use FF3 factors or the three principal components of daily returns, to measure each asset's idiosyncratic variance $\theta_{i,t}$ in each month and compute the sum of all weighted by market weight $w_{i,t-1}$. I also consider a measure of idiosyncratic risk relative to the CAPM model introduced in Campbell et al. (2001). I plot these three idiosyncratic risk measures in **Figure 6** and find a very similar magnitude of idiosyncratic risk changing over time.

In **Table 9**, I report the results controlling the idiosyncratic risk under the three measures above. The magnitude of the coefficient almost does not change, controlling for the idiosyncratic risk, yet the significance of predictability is generally weaker. All three sets of results remain positive out-of-sample R^2 at the 60-months horizon.

In **Figure 7**, I plot the de-trended tail estimator ζ_t together with the weighted average of idiosyncratic variance $\theta_{i,t}$ relative to the Fama-French 3 factor models. The Pareto coefficient tends to reach the bottom value at the shaded area, marking the NBER recession. The tail predictor has a weakly negative correlation (-0.17) with the level of idiosyncratic risk since the aggregate risk is counter-cyclical and increases with the market risk premium. The evidence shown in this plot consists of the intuition that a low Pareto coefficient implies a high risk premium and hence high future market returns.

3.3.4 Control for alternative predictors

I use predictors listed in Welch and Goyal (2008) as controls for other systematic risks to better identify the granular channel of risk premium. In **Table 10**, I provide a summary of predictors, including their definitions, AR1 coefficients, and their correlation coefficients with the main predictor $\log \zeta_t$. I normalize all the predictors to zero-mean and unit variance. Further, I adjust heteroskedasticity and serial correlation in residuals in all of our predictive regressions using the Newey-West standard error. The correlations between published predictors and the granular predictor are weak: Besides the default spread, which has a 0.27 correlation, all the other predictors have absolute correlations with ζ close to or less than 0.1. The weak correlation suggests that existing predictors in literature do not capture the granular effect.

In **Table 11**, I report results controlling for other predictors investigated in Welch and Goyal (2008). I add each predictor to the single variable regression and present bi-variate regression results. The granular predictor $\log \zeta_t$ negatively predicts the market returns with all the predictors controlled at all horizons. The bi-variate results highlight the stability of coefficients on $\log \zeta_t$ at all horizons: At monthly frequency, the coefficient is between -0.34 and -0.25. The 12-month-ahead coefficient is between -2.69 and -1.65, and the 60-month-ahead is between -11.21 and -8.27. The stability of coefficients suggests that the granular part of the market expected return is independent of other resources in the literature, which is consistent with the weak correlation between the Pareto coefficient and controlling variables. The significance remains in the long horizon at $k = 12, 60$, especially for the 60-month ahead.

In summary, I show that the Pareto coefficient negatively predicts the market returns. The results confirm the economic intuition that a low ζ indicates a high risk premium due to failure of diversification and high future market returns. Further, the results verify the time-series implication of my model: The level of granularity increases the un-diversified idiosyncratic risks in the market and explains the time-variation of the market expected returns.

4 Conclusion

I contribute to the existing asset pricing research by documenting a granular channel of idiosyncratic risk to explain expected returns. The fat-tailed distribution of firm size breaks the market diversification assumed by APT, making idiosyncratic risk matters for asset prices. I show that the size-adjusted idiosyncratic risk positively explains the cross-sectional variation of expected returns. This finding of mine explains the puzzling finding in Ang et al. (2006) and Ang et al. (2009). With granularity, only large firms have their idiosyncratic risks to explain expected returns. In contrast, small firms have their idiosyncratic risks diversified away due to their negligible weights in the market portfolio. This finding points out the potential bias of cross-sectional tests for identifying the relation between idiosyncratic risk and return without controlling the distribution of market weight across firms. This result is supported when running multiple sets of robustness checks as well. For implication at the aggregate level, I use a Pareto distribution to measure the level of granularity and show that the Pareto coefficient explains the market variation while controlling for time-varying idiosyncratic risk and alternative predictors in literature.

My theoretical model is based on a static APT model and treats the degree of market granularity as a feature of data to explore potential deviations from factor models. It would be interesting to combine the asset pricing study in this paper with dynamic growth models that endogenously generate a fat-tailed distribution of firm size (see Champernowne (1953), Wold and Whittle (1957), Gabaix (1999), Beare and Toda (2022))). Further, a dynamic framework may include the existing features in the asset pricing study: An asset pricing model that includes the factor risk structure as in APT, or an equilibrium mechanism to generate factor structures in expected returns, with the negative relation between firm size and volatility incorporated, must produce fruitful understandings of the dynamic interaction between granularity and asset returns.

References

Acemoglu, Daron, Vasco M Carvalho, Asuman Ozdaglar, and Alireza Tahbaz-Salehi, 2012, The network origins of aggregate fluctuations, *Econometrica* 80, 1977–2016.

Acemoglu, Daron, Asuman Ozdaglar, and Alireza Tahbaz-Salehi, 2015, Systemic risk and stability in financial networks, *American Economic Review* 105, 564–608.

Alves, MI Fraga, M Ivette Gomes, and Laurens de Haan, 2003, A new class of semi-parametric estimators of the second order parameter, *Portugaliae Mathematica* 60, 193–214.

Ang, Andrew, Robert J Hodrick, Yuhang Xing, and Xiaoyan Zhang, 2006, The cross-section of volatility and expected returns, *The journal of finance* 61, 259–299.

Ang, Andrew, Robert J Hodrick, Yuhang Xing, and Xiaoyan Zhang, 2009, High idiosyncratic volatility and low returns: International and further us evidence, *Journal of Financial Economics* 91, 1–23.

Axtell, Robert L, 2001, Zipf distribution of us firm sizes, *science* 293, 1818–1820.

Bali, Turan G, Nusret Cakici, Xuemin Yan, and Zhe Zhang, 2005, Does idiosyncratic risk really matter?, *The Journal of Finance* 60, 905–929.

Beare, Brendan K, and Alexis Akira Toda, 2022, Determination of pareto exponents in economic models driven by markov multiplicative processes, *Econometrica* 90, 1811–1833.

Beirlant, Jan, Goedele Dierckx, Yuri Goegebeur, and Gunther Matthys, 1999, Tail index estimation and an exponential regression model, *Extremes* 2, 177–200.

Byun, Sung Je, and Lawrence Schmidt, 2020, Real risk or paper risk? mis-measured factors, granular measurement errors, and empirical asset pricing tests, *Mis-Measured Factors, Granular Measurement Errors, and Empirical Asset Pricing Tests (Febrary 2020)* .

Campbell, John Y, Martin Lettau, Burton G Malkiel, and Yexiao Xu, 2001, Have individual stocks become more volatile? an empirical exploration of idiosyncratic risk, *The journal of finance* 56, 1–43.

Chamberlain, Gary, 1983, Funds, factors, and diversification in arbitrage pricing models, *Econometrica: Journal of the Econometric Society* 1305–1323.

Chamberlain, Gary, and Michael Rothschild, 1983, Arbitrage, factor structure, and mean-variance analysis on large asset markets, *Econometrica: Journal of the Econometric Society* 1281–1304.

Champernowne, David G, 1953, A model of income distribution, *The Economic Journal* 63, 318–351.

Connor, Gregory, and Robert A Korajczyk, 1986, Performance measurement with the arbitrage pricing theory: A new framework for analysis, *Journal of financial economics* 15, 373–394.

Connor, Gregory, and Robert A Korajczyk, 1993, A test for the number of factors in an approximate factor model, *the Journal of Finance* 48, 1263–1291.

Connor, Gregory, and Robert A Korajczyk, 1995, The arbitrage pricing theory and multifactor models of asset returns, *Handbooks in operations research and management science* 9, 87–144.

Cujean, Julien, and Michael Hasler, 2017, Why does return predictability concentrate in bad times?, *The Journal of Finance* 72, 2717–2758.

Daniel, Kent, and Sheridan Titman, 1997, Evidence on the characteristics of cross sectional variation in stock returns, *the Journal of Finance* 52, 1–33.

Diebold, Francis X, Til Schuermann, and John D Stroughair, 1998, Pitfalls and opportunities in the use of extreme value theory in risk management, in *Decision technologies for computational finance*, 3–12 (Springer).

Durrett, Rick, 2019, *Probability: theory and examples*, volume 49 (Cambridge university press).

Dybvig, Philip H, 1983, An explicit bound on individual assets' deviations from apt pricing in a finite economy, *Journal of Financial Economics* 12, 483–496.

Fama, Eugene F, and Kenneth R French, 1992, The cross-section of expected stock returns, *the Journal of Finance* 47, 427–465.

Feng, Guanhao, Stefano Giglio, and Dacheng Xiu, 2020, Taming the factor zoo: A test of new factors, *The Journal of Finance* 75, 1327–1370.

Feuerverger, Andrey, and Peter Hall, 1999, Estimating a tail exponent by modelling departure from a pareto distribution, *The Annals of Statistics* 27, 760–781.

Gabaix, Xavier, 1999, Zipf's law for cities: an explanation, *The Quarterly journal of economics* 114, 739–767.

Gabaix, Xavier, 2011, The granular origins of aggregate fluctuations, *Econometrica* 79, 733–772.

Gabaix, Xavier, and Rustam Ibragimov, 2011, Rank- 1/2: a simple way to improve the ols estimation of tail exponents, *Journal of Business & Economic Statistics* 29, 24–39.

Gabaix, Xavier, and Ralph SJ Koijen, 2020, Granular instrumental variables, Technical report, National Bureau of Economic Research.

Giglio, Stefano, Bryan Kelly, and Dacheng Xiu, 2022, Factor models, machine learning, and asset pricing, *Annual Review of Financial Economics* 14.

Giglio, Stefano, and Dacheng Xiu, 2021, Asset pricing with omitted factors, *Journal of Political Economy* 129, 000–000.

Giglio, Stefano, Dacheng Xiu, and Dake Zhang, 2021, Test assets and weak factors, Technical report, National Bureau of Economic Research.

Gomesa, M Ivette, and M João Martins, 2002, "asymptotically unbiased" estimators of the tail index based on external estimation of the second order parameter, *Extremes* 5, 5–31.

Goyal, Amit, and Pedro Santa-Clara, 2003, Idiosyncratic risk matters!, *The journal of finance* 58, 975–1007.

Hall, Peter, and Alan H Welsh, 1985, Adaptive estimates of parameters of regular variation, *The Annals of Statistics* 331–341.

Herskovic, Bernard, Bryan Kelly, Hanno Lustig, and Stijn Van Nieuwerburgh, 2016, The common factor in idiosyncratic volatility: Quantitative asset pricing implications, *Journal of Financial Economics* 119, 249–283.

Hill, Bruce M, 1975, A simple general approach to inference about the tail of a distribution, *The annals of statistics* 1163–1174.

Hou, Kewei, and Roger K Loh, 2016, Have we solved the idiosyncratic volatility puzzle?, *Journal of Financial Economics* 121, 167–194.

Huberman, Gur, 2005, A simple approach to arbitrage pricing theory, in *Theory of Valuation*, 289–308 (World Scientific).

Jegadeesh, Narasimhan, and Sheridan Titman, 1993, Returns to buying winners and selling losers: Implications for stock market efficiency, *The Journal of finance* 48, 65–91.

Kelly, Bryan, and Hao Jiang, 2014, Tail risk and asset prices, *The Review of Financial Studies* 27, 2841–2871.

Kelly, Bryan T, Seth Pruitt, and Yinan Su, 2020, Instrumented principal component analysis, *Available at SSRN 2983919* .

Merton, Robert C, 1973, An intertemporal capital asset pricing model, *Econometrica: Journal of the Econometric Society* 867–887.

Peng, L, 1998, Asymptotically unbiased estimators for the extreme-value index, *Statistics & Probability Letters* 38, 107–115.

Ross, Stephen, 1976, The arbitrage theory of capital asset pricing, *Journal of Economic Theory* 13, 341–360.

Stambaugh, Robert F, 1999, Predictive regressions, *Journal of financial economics* 54, 375–421.

Welch, Ivo, and Amit Goyal, 2008, A comprehensive look at the empirical performance of equity premium prediction, *The Review of Financial Studies* 21, 1455–1508.

Wold, Herman OA, and Peter Whittle, 1957, A model explaining the pareto distribution of wealth, *Econometrica, Journal of the Econometric Society* 591–595.

Xu, Yexiao, and Burton G Malkiel, 2003, Investigating the behavior of idiosyncratic volatility, *The Journal of Business* 76, 613–645.

5 Tables and Figures

Figure 1: **Firm Market Weight Sorted in Year 2020**. This figure displays the fat right tail of firm size. I measure the firm size by each asset's relative weight in the market portfolio. The 10 largest firms are highlighted and accounts for about 25 percents of the whole CRSP data in 2020 contains about 4,000 firms. The red dash line marks the portfolio weight under a uniform firm size distribution, as a comparison to the "1/N" benchmark.

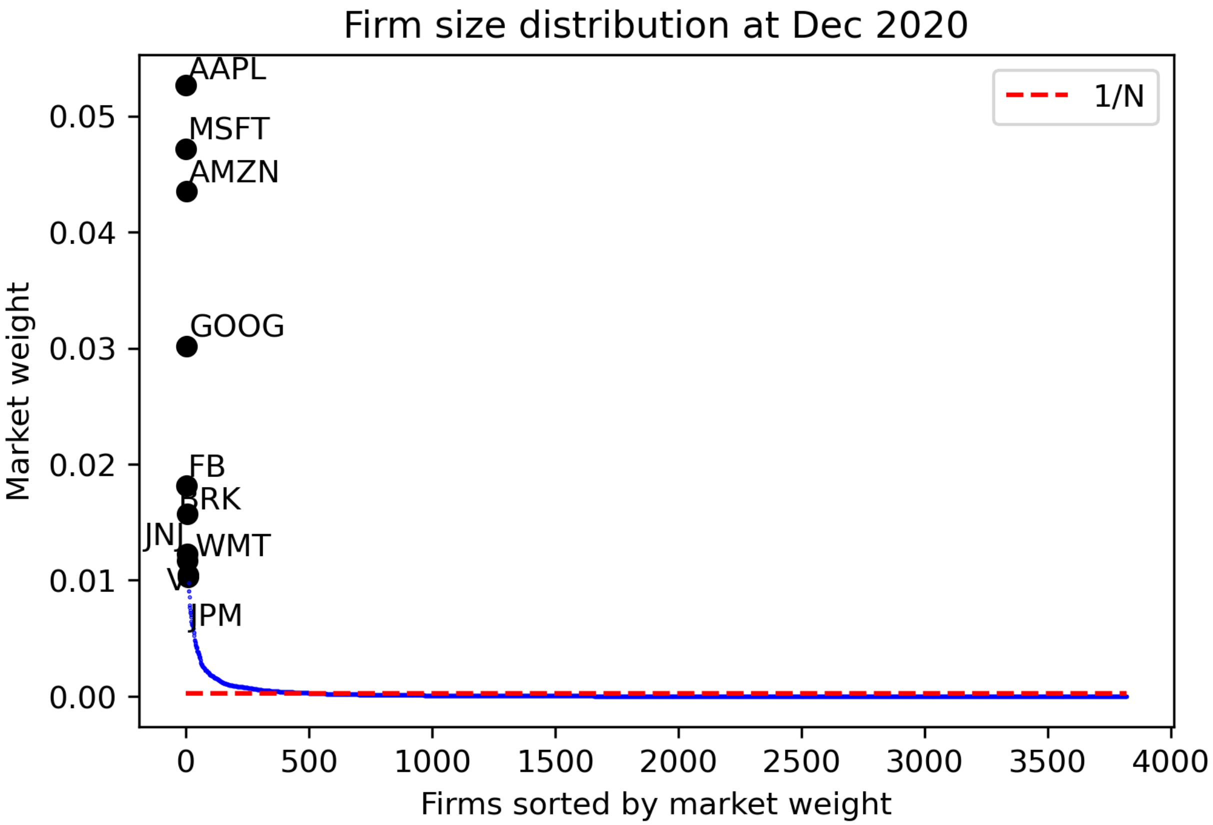

Figure 2: **Market Weight of the Largest 1 Percent Firms (12 month moving-average)**. In this figure, I plot monthly total market weight of the top 1 percent firms, moving-averaged by a 12 month window. The shaded areas are NBER recession periods.

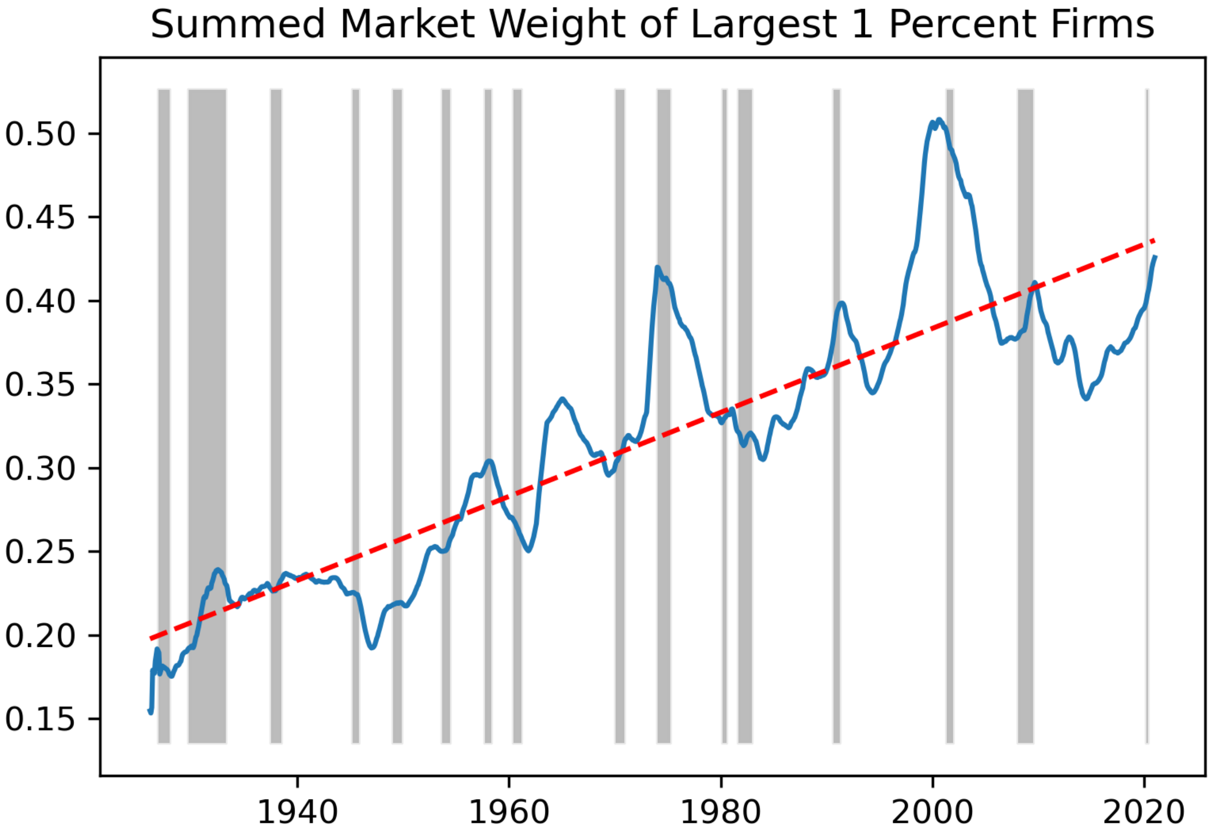

Figure 3: **Logged rank-size plot in December 2020** In this figure, I plot the logged rank-size plot of the largest 20% firms in December 2020. The red dash line show the fitted relation implied by the Pareto distribution. The 10 largest firms are highlighted.

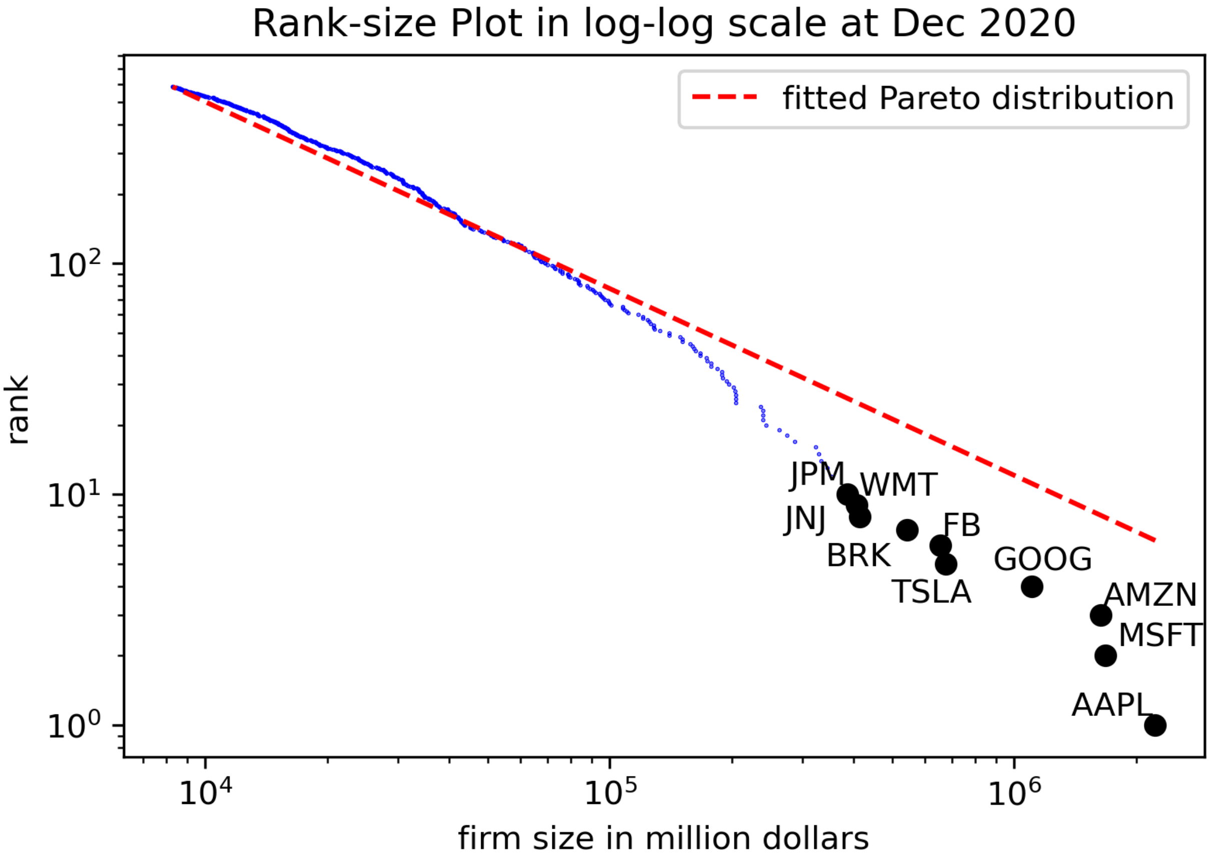

Figure 4: **Pareto Coefficient Estimate of Market Value per Month** At the end of each month, I estimate the tail parameter ζ of Pareto distribution using the Hill estimator (see Hill (1975)) at a monthly frequency. I use the largest 20 % firms to illustrate a trade-off between bias and variance of the Hill estimator. I plot the estimate of ζ_t in blue line, together with the confidence interval (+/- two times the standard errors of ζ_t as a maximum likelihood estimator) in the two red lines below and above. The two vertical dash lines in the plot mark the expansion of n due to merging of security exchanges: AMEX in June 1962 and NASDAQ in January 1973. The shaded areas are NBER recession periods.

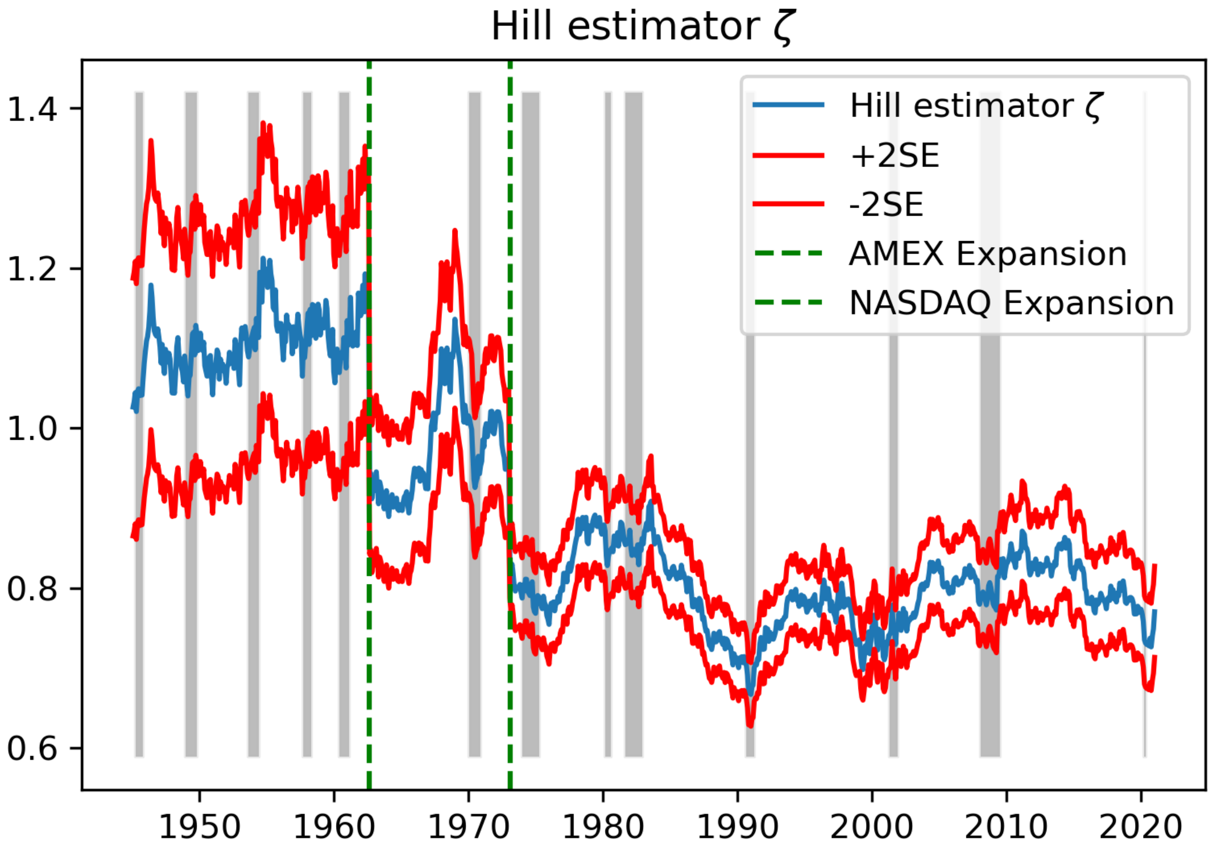

Figure 5: **Pareto Coefficient Estimate of Market Value per Month** I plot the cointegration relation between the logged Pareto coefficient $\log \zeta$ (estimated from the largest 20 % firms) and logged number of firms n. Both the time series are normalized to have mean zero and unit variance with their raw magnitudes displayed on two separate sets of ticks on y axis.

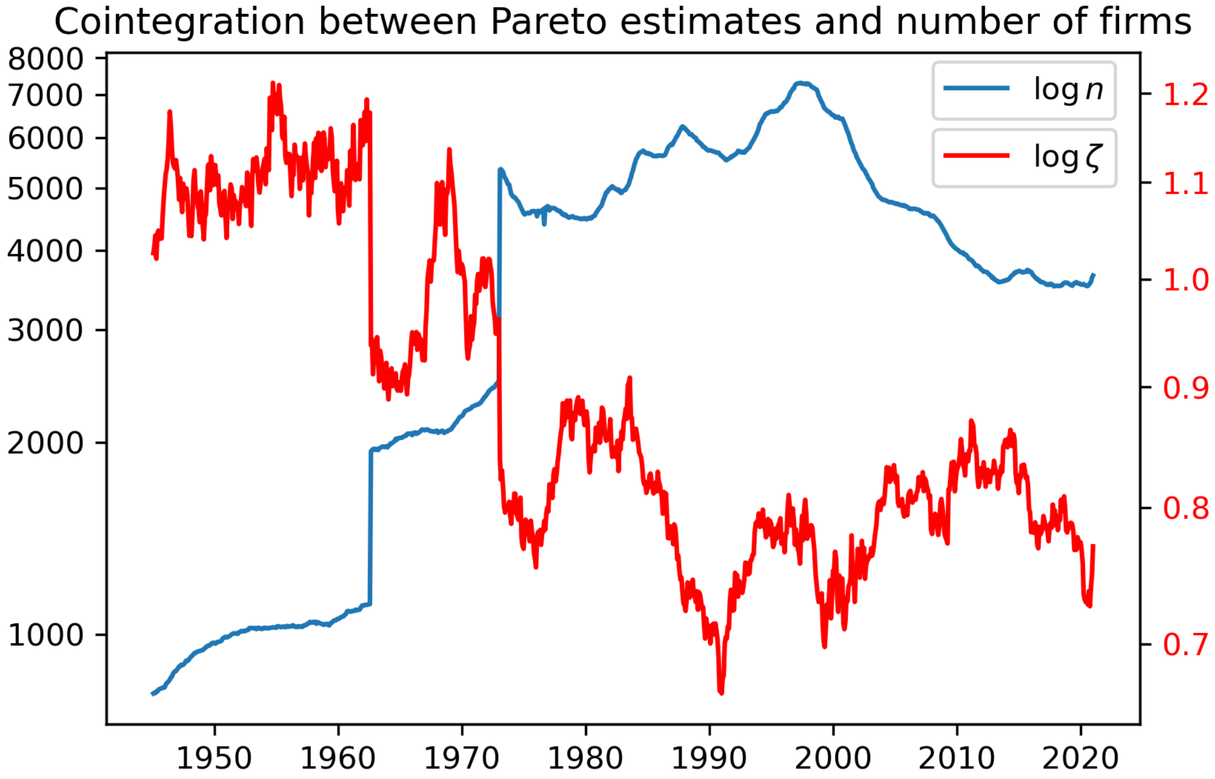

Figure 6: **Three measures of idiosyncratic risk** . I measure the level of idiosyncratic risk at the aggregate level by the weighted average of θ. Specifically, I use FF3 factors, or the three principal components of daily returns, to measure each asset's idiosyncratic variance $\theta_{i,t}$ in each month and compute the sum of all weighted by market weight $w_{i,t-1}$. I also consider a measure of idiosyncratic risk relative to the CAPM model introduced in Campbell et al. (2001). The shaded areas are NBER recessions.

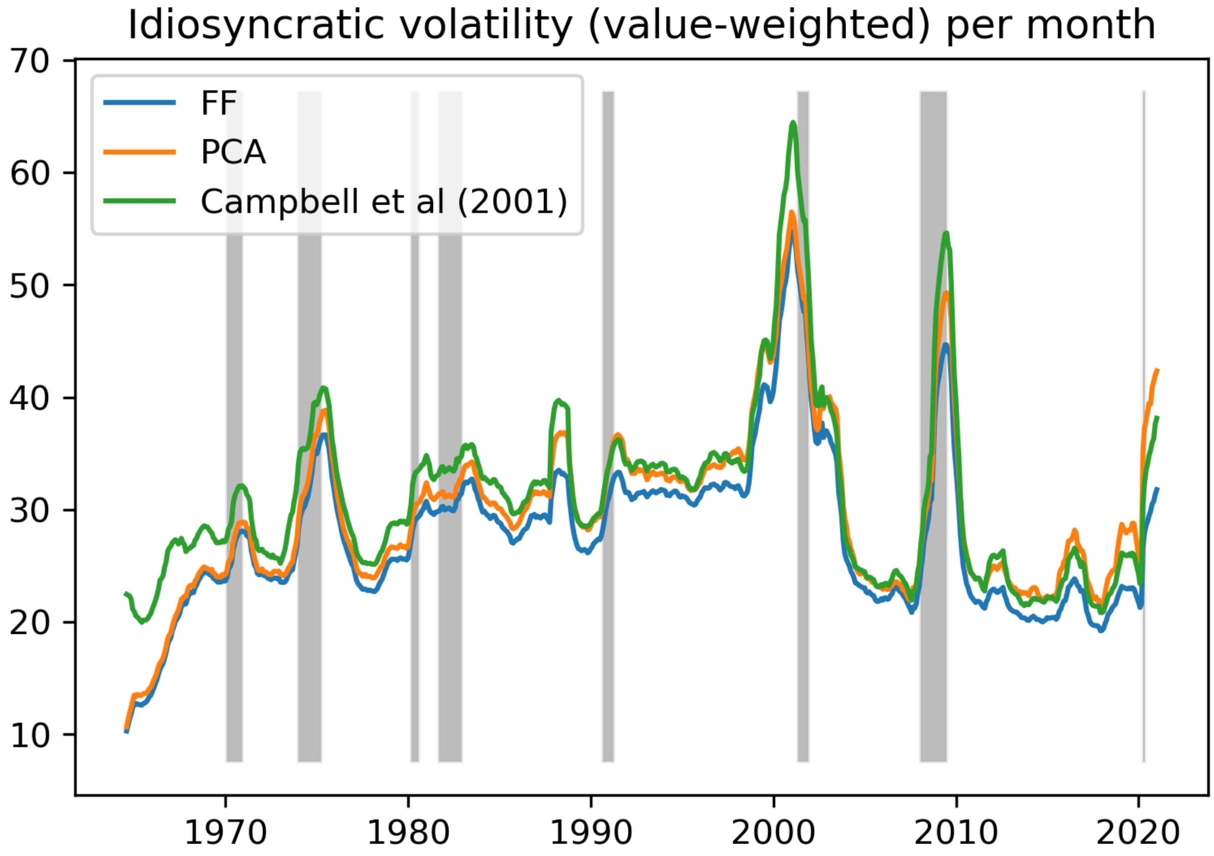

Figure 7: **Pareto coefficient (de-trended) v.s. idiosyncratic risk**. I measure the level of idiosyncratic risk at the aggregate level by the weighted average of idiosyncratic risk relative to Fama-French 3 factors. I plot the two series together where the blue line is the Pareto predictor and the yellow line is the weighted average of idiosyncratic variance. The shaded areas are NBER recessions.

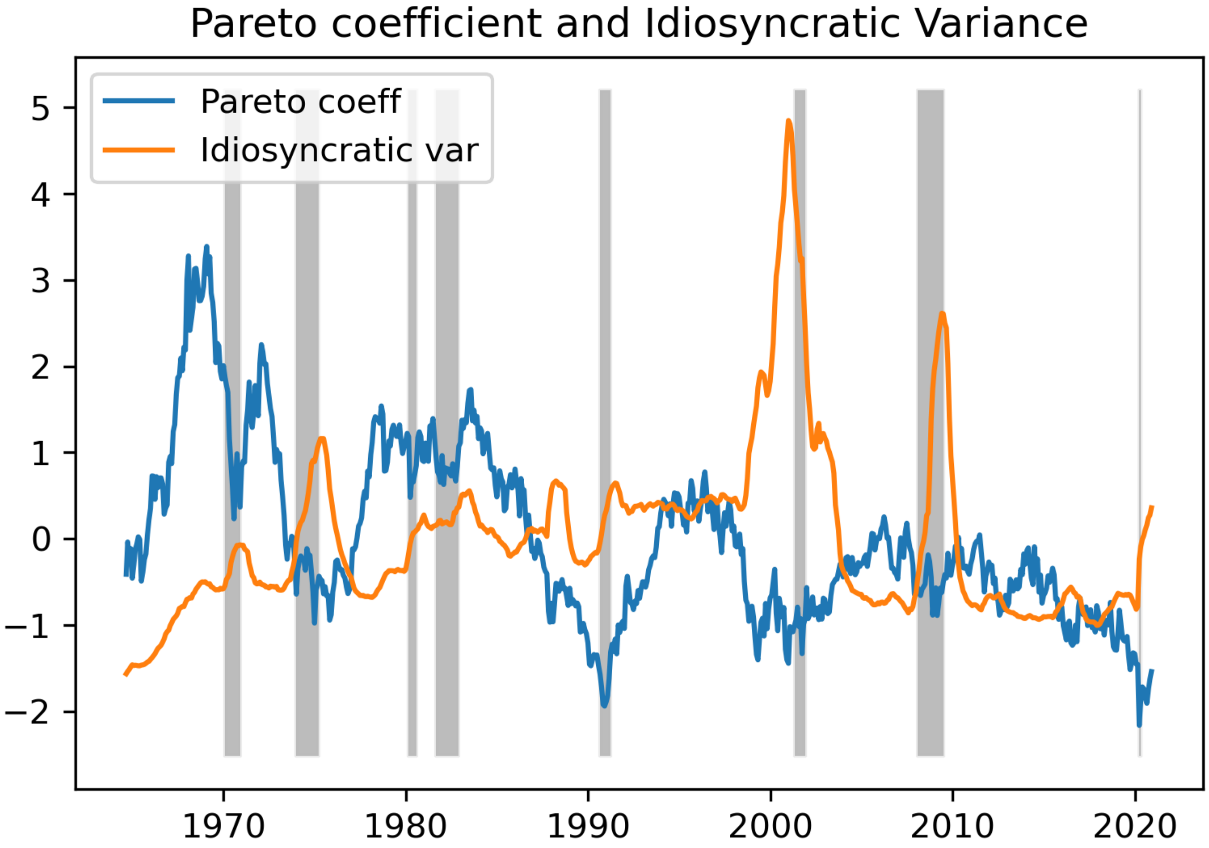

Table 1: **Portfolios sorted by idiosyncratic variance estimated by Fama-French 3 factors per month.**
Like Ang et al. (2006), I sort all the assets by their idiosyncratic risk θ measured per month using FF3 factors. Then I split all the assets into five quintiles to construct five value-weighted portfolio sorted by the idiosyncratic variance measured in the last month $\theta_{i,t-1}$. I report the mean (annualized, in percent), volatility (annualized, in percent) and market weight of each portfolio in Panel A. I also examine the alpha and idiosyncratic volatility (both annualized, in percent) of these portfolios relative to several benchmark models. I report results using Fama-French 3 factors in Panel B as the benchmark case, CAPM in Panel C and a factor model including the three principal components of all the available asset returns in Panel D.

Panel A: Summary of portfolios sorted by idiosyncratic variance						
	L	2	3	4	H	L-H
Mean	7.17	7.42	8.38	4.75	-0.06	7.23
Volatility	13.73	17.35	21.35	26.05	30.12	23.65
w_i	0.60	0.23	0.11	0.05	0.02	
Panel B: alpha relative to FF3						
α_{FF3}	1.18	-0.20	-0.44	-5.29	-11.42	12.60
T-stat	2.91	-0.38	-0.53	-3.88	-6.54	6.34
$\sqrt{\theta}_{FF3}$	2.82	3.97	5.80	8.96	13.85	
$\alpha_{FF3}/\theta_{FF3}$	14.85	-1.29	-1.31	-6.60	-5.95	
Panel C: alpha relative to CAPM						
α_{CAPM}	1.34	-0.02	-0.50	-5.39	-10.59	11.92
T-stat	2.52	-0.04	-0.48	-2.98	-4.35	4.20
$\sqrt{\theta_{CAPM}}$	3.67	4.00	7.16	12.29	18.38	
$\alpha_{CAPM}/\theta_{CAPM}$	9.94	-0.14	-0.97	-3.57	-3.13	
Panel D: alpha relative to PCA factors						
α_{PC}	5.90	5.29	5.06	0.11	-5.88	11.79
T-stat	3.45	2.66	2.33	0.04	-2.16	5.29
$\sqrt{\theta}_{PC}$	12.83	15.28	17.24	19.31	20.11	
α_{PC}/θ_{PC}	3.59	2.27	1.70	0.03	-1.45	

Table 2: **Portfolios sorted by idiosyncratic variance estimated by three principal components per month**
Like Ang et al. (2006), I sort all the assets by their idiosyncratic risk θ. Different from the main result in Table 1, I first estimate the three principal components per month for all the available daily returns and then I estimate the idiosyncratic variance $\theta_{i,t}$ by running daily returns on the three PCs per month. Then I split all the assets into five quintiles to construct five value-weighted portfolio sorted by the idiosyncratic variance measured in the last month $\theta_{i,t-1}$. I report the mean (annualized, in percent), volatility (annualized, in percent) and market weight of each portfolio in Panel A. I also examine the alpha and idiosyncratic volatility (both annualized, in percent) of these portfolios relative to several benchmark models. I report results using Fama-French 3 factors in Panel B as the benchmark case, CAPM in Panel C and a factor model including the three principal components of all the available asset returns in Panel D.

Panel A: Summary of portfolios sorted by idiosyncratic variance						
	L	2	3	4	H	L-H
Mean	7.25	7.34	7.76	5.45	-0.39	7.64
Volatility	13.36	17.10	21.03	25.49	30.26	23.96
w_i	0.55	0.26	0.12	0.05	0.02	
Panel B: alpha relative to FF3						
α_{FF3}	1.41	-0.18	-0.99	-4.45	-11.89	13.31
T-stat	3.05	-0.39	-1.15	-3.51	-6.01	5.92
$\sqrt{\theta}_{FF3}$	3.31	3.85	5.87	8.43	14.13	
$\alpha_{FF3}/\theta_{FF3}$	12.91	-1.20	-2.86	-6.26	-5.95	
Panel C: alpha relative to CAPM						
α_{CAPM}	1.62	-0.01	-1.03	-4.56	-11.07	12.69
T-stat	2.66	-0.01	-1.00	-2.68	-4.44	4.31
$\sqrt{\theta_{CAPM}}$	3.96	3.87	6.74	11.62	18.16	
$\alpha_{CAPM}/\theta_{CAPM}$	10.32	-0.04	-2.27	-3.38	-3.36	
Panel D: alpha relative to PCA factors						
α_{PC}	6.02	5.28	4.61	0.97	-6.14	12.16
T-stat	3.55	2.73	2.05	0.40	-2.11	5.15
$\sqrt{\theta}_{PC}$	12.51	15.17	17.36	19.11	20.26	
α_{PC}/θ_{PC}	3.84	2.30	1.53	0.27	-1.50	

Table 3: **Robustness check for "bet on granularity portfolios" under different formation periods.**

Like Ang et al. (2006), I sort all the assets by their idiosyncratic risk θ measured by the past 1,3,6 and 12 months. I split all the assets into five quintiles to construct five value-weighted portfolio sorted by the idiosyncratic variance measured in the last month $\theta_{i,t-1}$. I construct the "bet on granularity" portfolio by leveraging a long position of the lowest θ portfolio r_L and short the highest θ portfolio r_H. The long-short strategy is constructed as follows:

$$r_{L-H,t} = \frac{1/\theta_{L,t-1}}{1/\theta_{L,t-1} - 1/\theta_{H,t-1}}(r_{L,t} - r_f) - \frac{1/\theta_{H,t-1}}{1/\theta_{L,t-1} - 1/\theta_{H,t-1}}(r_{H,t} - r_f)$$

I examine the alpha and idiosyncratic volatility (both annualized, in percent) of these portfolios relative to several benchmark models. I report results using CAPM, Fama-French 3 factors and a factor model including the three principal components of portfolios sorted by various characteristics. The results using Fama-French 3 factors and three principal components of daily returns in the estimation window are listed in Panel A and B respectively.

Panel A: Measured by FF 3 factors					**Panel B: Measured by 3 principal components**				
window length	1	3	6	12		1	3	6	12
	r_{L-H}	r_{L-H}	r_{L-H}	r_{L-H}		r_{L-H}	r_{L-H}	r_{L-H}	r_{L-H}
Mean	7.36	7.29	7.25	7.03	Mean	7.40	7.39	7.29	6.97
Volatility	13.60	13.67	13.66	13.74	Volatility	13.22	13.41	13.45	13.51
α_{FF3}	1.49	1.48	1.57	1.46	α_{FF3}	1.69	1.77	1.82	1.58
T-stat	3.67	3.62	3.79	3.54	T-stat	3.58	3.92	4.05	3.45
α_{CAPM}	1.64	1.60	1.62	1.47	α_{CAPM}	1.89	1.87	1.81	1.57
T-stat	3.04	2.79	2.76	2.45	T-stat	3.28	3.05	2.87	2.41
α_{PC}	6.20	6.25	6.27	6.04	α_{PC}	6.28	6.44	6.40	6.08
T-stat	3.65	3.62	3.62	3.45	T-stat	3.79	3.79	3.73	3.50

Table 4: **Cross-sectional results using 100 portfolios sorted by idiosyncratic variance, robustness check for measurement window of idiosyncratic risk**
Like Ang et al. (2006), I sort all the assets by their idiosyncratic risk θ measured per month using FF3 factors/three principal components of daily returns. I examine the robustness of my 100-portfolio results for different measurement window length. As in Section 3.2.2, I report estimate of $\alpha_i = constant + \eta\sqrt{\theta_i}$

$$\alpha_i = constant + \gamma w_i \theta_i$$

I summary the estimate of η, γ and the estimated correlations $corr(w_i\theta_i, \sqrt{\theta_i})$, $corr(w_i, \sqrt{\theta_i})$ using portfolios formed by the idiosyncratic variance measured by the daily returns in the past 1,3,6 and 12 months. The results using Fama-French 3 factors and 3 principal components of daily returns are listed in Panel A and B respectively.

Panel A: FF 3 factors				
estimates \ window length	1	3	6	12
η	-0.74	-0.75	-0.71	-0.57
	-18.77	-18.70	-23.46	-22.94
γ	5.17	4.67	3.90	3.16
	8.72	7.78	7.88	7.60
$corr(w_i\theta_i, \sqrt{\theta_i})$	-0.72	-0.68	-0.64	-0.63
$corr(w_i, \sqrt{\theta_i})$	-0.68	-0.63	-0.59	-0.56
Panel B: 3 principal components				
estimates \ window length	1	3	6	12
η	-0.71	-0.75	-0.67	-0.57
	-18.89	-19.67	-21.91	-17.91
γ	6.22	6.03	4.85	3.87
	10.21	10.75	9.84	9.28
$corr(w_i\theta_i, \sqrt{\theta_i})$	-0.80	-0.77	-0.71	-0.69
$corr(w_i, \sqrt{\theta_i})$	-0.74	-0.69	-0.65	-0.63

Table 5: **Fama-MacBeth results, individual asset level using FF3**
In this table, I report the individual asset level test of granular risk premium by running $r_{i,t}$ on the size-adjusted idiosyncratic variance $w_{i,t-1}\theta_{i,t}$. The goal is to compare my estimate to estimate in Ang et al. (2009):

$$r_{i,t} = \text{constant} + \text{controls} + \sum_{s=1}^{k} \hat{\beta}_{i,s,t}\mu_s + \eta\sqrt{\hat{\theta}_{i,t-1}} + \epsilon_{i,t}$$

to my model:

$$r_{i,t} = \text{constant} + \text{controls} + \sum_{s=1}^{k} \hat{\beta}_{i,s,t}\mu_s + \gamma w_{i,t-1}\hat{\theta}_{i,t} + \epsilon_{i,t}$$

I estimate $\hat{\eta}$ in the columns 1 to replicate the results in Ang et al. (2009) and compare it to the estimate of $\hat{\gamma}$ from my model in the column 4. I estimate the idiosyncratic variance $\hat{\theta}_{i,t}$ by running daily returns on the FF3 factors per month. The controlling variables are the FF3 factor loadings and the lagged characteristics suggested by Daniel and Titman (1997). As in their paper, I control the lagged book-to-market ratio $b/m_{i,t-1}$ and the momentum factor $mom_{i,t-1}$ computed by the sum of returns in the last six months as in Jegadeesh and Titman (1993).

	Cross-sectional Regression, Stock Level				
	$r_{i,t}$	$r_{i,t}$	$r_{i,t}$	$r_{i,t}$	$r_{i,t}$
const	0.62	0.63	0.58	0.49	0.55
	3.68	3.69	2.98	2.58	3.28
$\hat{\beta}_{i,t}^{Mkt-RF}$	0.01	0.01	-0.00	-0.01	0.00
	0.17	0.17	-0.08	-0.19	0.07
$\hat{\beta}_{i,t}^{SMB}$	0.05	0.05	0.04	0.04	0.04
	1.84	1.87	1.66	1.54	1.75
$\hat{\beta}_{i,t}^{HML}$	-0.01	-0.01	-0.00	0.00	-0.01
	-0.45	-0.46	-0.02	0.10	-0.34
$b/m_{i,t-1}$	0.24	0.24	0.24	0.25	0.25
	8.79	8.74	8.57	8.84	9.05
$mom_{i,t-1}$	-0.57	-0.58	-0.45	-0.48	-0.61
	-2.89	-2.92	-2.01	-2.14	-3.09
$\sqrt{\hat{\theta}_{i,t-1}}$	**-0.01**	**-0.01**			**-0.01**
	-1.98	**-2.10**			**-2.16**
$w_{i,t-1}$		-0.10	-0.11	-1.86	-1.78
		-0.63	-0.59	-5.05	-5.26
$w_{i,t-1}\hat{\theta}_{i,t}$				**9.15**	**8.95**
				8.99	**8.87**

Table 6: **Fama-MacBeth results, individual asset level using PCA**

In this table, I report the individual asset level test of granular risk premium by running $r_{i,t}$ on the size-adjusted idiosyncratic variance $w_{i,t-1}\theta_{i,t}$. I estimate $\hat{\eta}$ in the columns 1 to replicate the results in Ang et al. (2009) and compare it to the estimate of $\hat{\gamma}$ from my model in the column 4. Different from the main result in Table 5, I first estimate the three principal components per month for all the available daily returns and then I estimate the idiosyncratic variance $\theta_{i,t}$ by running daily returns on the three PCs per month. The controlling variables are the three PC loadings and the lagged characteristics suggested by Daniel and Titman (1997). As in their paper, I control the lagged book-to-market ratio and the momentum factor computed by the sum of returns in the last six months as in Jegadeesh and Titman (1993).

Cross-sectional Regression, Stock Level					
	$r_{i,t}$	$r_{i,t}$	$r_{i,t}$	$r_{i,t}$	$r_{i,t}$
const	0.71	0.72	0.65	0.59	0.66
	4.14	4.16	3.30	2.98	3.86
$\hat{\beta}_{i,t}^{PCA1}$	8.10	8.09	8.10	8.03	8.03
	7.84	7.83	7.69	7.66	7.81
$\hat{\beta}_{i,t}^{PCA2}$	5.98	5.98	6.01	5.99	5.96
	6.42	6.43	6.30	6.28	6.41
$\hat{\beta}_{i,t}^{PCA3}$	5.23	5.23	5.32	5.34	5.26
	6.85	6.85	6.80	6.85	6.90
$\mathrm{b/m}_{\mathrm{i,t-1}}$	0.23	0.23	0.23	0.24	0.24
	8.64	8.57	8.40	8.67	8.87
$\mathrm{mom}_{\mathrm{i,t-1}}$	-0.56	-0.56	-0.40	-0.42	-0.58
	-2.85	-2.88	-1.83	-1.93	-3.00
$\sqrt{\hat{\theta}_{i,t-1}}$	**-0.01**	**-0.01**			**-0.01**
	-1.88	**-2.01**			**-2.12**
$w_{i,t-1}$		-0.16	-0.15	-2.00	-1.97
		-1.00	-0.81	-4.73	-5.03
$w_{i,t-1}\hat{\theta}_{i,t}$				**6.73**	**6.62**
				7.97	**7.86**

Table 7: **Fama-MacBeth results, robustness check for measurement window of idiosyncratic risk**
This table examine the robustness of results (column 4) in Table 5 by estimating the idiosyncratic variance per month using daily returns in the past 1,3,6,12 months. The results using Fama-French 3 factors and three principal components of daily returns in the estimation window are listed in Panel A and B respectively. The controlling variables are the factor loadings and the lagged characteristics suggested by Daniel and Titman (1997). As in their paper, I control the lagged book-to-market ratio, firm size, and the momentum factor computed by the sum of returns in the last six months as in Jegadeesh and Titman (1993).

Panel A: FF 3 factors					**Panel B: 3 principal components**				
window length	1	3	6	12	window length	1	3	6	12
	$r_{i,t}$	$r_{i,t}$	$r_{i,t}$	$r_{i,t}$		$r_{i,t}$	$r_{i,t}$	$r_{i,t}$	$r_{i,t}$
const	0.55	0.56	0.59	0.57	const	0.66	0.64	0.63	0.59
	3.28	3.84	4.18	4.20		3.86	3.99	3.99	3.81
$\hat{\beta}_{i,t}^{Mkt-RF}$	0.00	-0.02	-0.07	-0.11	$\hat{\beta}_{i,t}^{PCA1}$	8.03	1.84	0.35	-0.10
	0.07	-0.18	-0.72	-0.99		7.81	1.84	0.42	-0.17
$\hat{\beta}_{i,t}^{SMB}$	0.04	0.06	0.12	0.12	$\hat{\beta}_{i,t}^{PCA2}$	5.96	1.95	3.75	1.17
	1.75	1.05	1.74	1.54		6.41	1.59	4.12	1.90
$\hat{\beta}_{i,t}^{HML}$	-0.01	-0.04	-0.05	-0.06	$\hat{\beta}_{i,t}^{PCA3}$	5.26	2.02	-0.17	0.12
	-0.34	-0.69	-0.70	-0.75		6.90	1.96	-0.22	0.19
$\mathrm{b/m}_{\mathrm{i,t-1}}$	0.25	0.23	0.23	0.23	$\mathrm{b/m}_{\mathrm{i,t-1}}$	0.24	0.23	0.22	0.22
	9.05	9.87	10.18	9.50		8.87	9.69	9.55	9.12
$\mathrm{mom}_{\mathrm{i,t-1}}$	-0.61	-0.04	-0.10	-0.46	$\mathrm{mom}_{\mathrm{i,t-1}}$	-0.58	-0.17	-0.24	-0.47
	-3.09	-0.25	-0.59	-2.63		-3.00	-0.99	-1.48	-2.77
$\sqrt{\hat{\theta}_{i,t-1}}$	-0.01	-0.01	-0.01	-0.01	$\sqrt{\hat{\theta}_{i,t-1}}$	-0.01	-0.01	-0.01	0.00
	-2.16	-1.98	-1.74	-1.05		-2.12	-1.39	-1.10	-0.18
$w_{i,t-1}$	-1.78	0.10	0.63	0.74	$w_{i,t-1}$	-1.97	-0.17	0.48	0.57
	-5.26	0.42	2.79	3.52		-5.03	-0.53	1.51	1.80
$w_{i,t-1}\hat{\theta}_{i,t}$	8.95	1.51	-0.86	-1.53	$w_{i,t-1}\hat{\theta}_{i,t}$	6.62	1.71	0.00	-0.96
	8.87	1.83	-1.11	-2.12		7.86	2.21	-0.60	-1.32

Table 8: **Single Variable Prediction**
In this table, I reports the monthly single variable regression results for the Pareto coefficient ζ to predict log excess-return for the aggregate market. In Panel A, I check the prediction results at multiple-horizons at various horizon $k = 1, 12, 60$.

$$\log(r_{m,t+k}) = \text{constant} + \mathrm{A} \log \xi_{\mathrm{t}}(\text{debias})$$

In Panel B, I compare the predictive power in the whole sample, the NBER recessions ,and non-NBER recessions. The Pareto coefficient is de-trended and a lower ξ implies a fatter tail, the hypothesize predictive relation should be negative $A < 0$.

Panel A: Single variable prediction, multiple-horizon results			
	$\log r_{m,t \to t+1}$	$\log r_{m,t \to t+12}$	$\log r_{m,t \to t+60}$
$\log \zeta_t$	**-0.28**	**-2.03**	**-10.81**
T-stat	**-2.11**	**-1.70**	**-3.42**
$R^2(\%)$	0.43	1.67	9.61
$\log \zeta_t$ (de Stambaugh-bias)	-0.27	-2.04	-10.78
T-stat	-1.87	-3.61	-8.77
$Out-of-sampleR^2(\%)$	-0.17	1.50	13.34

Panel B: Single variable prediction, sub-sample results			
	Whole Sample	NBER Recession	Non-NBER Recession
$\log \zeta_t$	**-0.28**	**-1.05**	**-0.10**
T-stat	**-2.11**	**-2.59**	**-0.76**
$R^2(\%)$	0.43	5.00	0.05
$\log \zeta_t$ (de Stambaugh-bias)	-0.27	-1.05	-0.09
T-stat	-1.88	-4.18	-0.59

Table 9: **Prediction results controlling for idiosyncratic risks**
In this table, I reports the monthly regression results for the Pareto coefficient ζ to predict log excess-return for the aggregate market controlling for three measures of idiosyncratic risk. I run the prediction results at multiple-horizons at various horizon $k = 1, 12, 60$. I measure the level of idiosyncratic risk at the aggregate level by the weighted average of θ. Specifically, I use FF3 factors, or the three principal components of daily returns, to measure each asset's idiosyncratic variance $\theta_{i,t}$ in each month and compute the sum of all weighted by market weight $w_{i,t-1}$. I also consider a measure of idiosyncratic risk relative to the CAPM model introduced in Campbell et al. (2001). The results controlling for these three measures are in Panel A, B and C, respectively.

$$\log(r_{m,t+k}) = \text{constant} + \mathrm{A}\log\xi_{\mathrm{t}}(\text{debias}) + \sum \mathrm{w}_{\mathrm{i,t-1}}\theta_{\mathrm{i,t}}$$

The Pareto coefficient is de-trended and a lower ξ implies a fatter tail, the hypothesize predictive relation should be negative $A < 0$.

Panel A: control $\sum w_i\theta_i$(FF3)			
	$\log r_{m,t\to t+1}$	$\log r_{m,t\to t+12}$	$\log r_{m,t\to t+60}$
$\log\zeta_t$	**-0.27**	**-1.70**	**-6.17**
T-stat	**-1.91**	**-1.31**	**-1.91**
$\sum w_i\theta_i$(FF3)	-0.20	-1.69	0.80
T-stat	-0.99	-0.91	0.18
$R^2(\%)$	0.48	1.79	3.34
$Out-of-sampleR^2(\%)$	-2.45	-8.12	5.85

Panel B: control $\sum w_i\theta_i$(PCA)			
	$\log r_{m,t\to t+1}$	$\log r_{m,t\to t+12}$	$\log r_{m,t\to t+60}$
$\log\zeta_t$	**-0.26**	**-1.64**	**-5.83**
T-stat	**-1.81**	**-1.23**	**-1.78**
$\sum w_i\theta_i$(PCA)	-0.10	-1.11	1.53
T-stat	-0.47	-0.56	0.33
$R^2(\%)$	0.33	1.12	3.47
$Out-of-sampleR^2(\%)$	-2.85	-10.20	4.72

Panel C: control $\sum w_i\theta_i$(Campbell et al)			
	$\log r_{m,t\to t+1}$	$\log r_{m,t\to t+12}$	$\log r_{m,t\to t+60}$
$\log\zeta_t$	**-0.28**	**-1.78**	**-6.52**
T-stat	**-1.94**	**-1.38**	**-2.06**
$\sum w_i\theta_i$ (Campbell et al)	-0.23	-2.18	-0.21
T-stat	-1.11	-1.20	-0.05
$R^2(\%)$	0.55	2.55	3.29
$Out-of-sampleR^2(\%)$	-2.69	-7.11	3.04

Table 10: **Summary of Predictors**

In this table, I report the AR1 coefficient of all the predictors used. Also I include the correlation coefficient between each controlling predictors and the Pareto coefficients. The controlling predictors in Welch and Goyal (2008) are defined as follows: bm is the book to market ratio, dspr is the default spread, dp is the dividend price ratio, ep is the earning prices ratio, ltr is the long term government bond return, ntis is the net equity expansion ratio, svar is the stock variance, tspr is the term spread, corpr is the corporate bond return.

Summary of Predictors			
	Description	AR1	Corr with ξ_t
ξ_t	granularity measure	0.97	1.00
bm	book to market ratio	0.99	0.07
dspr	default spread	0.97	0.27
dp	dividend price ratio	0.99	-0.11
ep	earning price ratio	0.99	0.04
ltr	long term government bond return	0.05	0.01
ntis	net equity expansion ratio	0.98	0.08
svar	stock variance	0.40	-0.07
tspr	term spread	0.96	0.11
corpr	corporate bond return	0.11	0.01

Table 11: **Bi-Variable Prediction**

In this table, I report the double variable regression results for the logged excess market return $\log(r_{m,t+k})$ at various horizon $k = 1, 12, 60$.

$$\log(r_{m,t+k}) = \text{constant} + \text{A} \log \zeta_{\text{t}}(\text{debias}) + \text{predictor}$$

In Panel A,B,C, I report the results at different horizons and the other predictors are controlled in each column for a bi-variate regression. The Pareto coefficient is detrended and a lower ξ implies a fatter tail, the hypothesize predictive relation should be negative $A < 0$.

The controlling predictors in Welch and Goyal (2008) are defined as follows: bm is the book to market ratio, dspr is the default spread, dp is the dividend price ratio, ep is the earning prices ratio, ltr is the long term government bond return, ntis is the net equity expansion ratio, svar is the stock variance, tspr is the term spread, corpr is the corporate bond return.

Panel A: Predictors Controlled, 1 Month Horizon

	bm	dspr	dp	ep	ltr	ntis	svar	tspr	corpr
$\log \zeta_t$	**-0.29**	**-0.34**	**-0.25**	**-0.29**	**-0.28**	**-0.27**	**-0.28**	**-0.30**	**-0.29**
T-stat	**-2.13**	**-2.38**	**-1.86**	**-2.13**	**-2.22**	**-2.09**	**-2.14**	**-2.31**	**-2.30**
predictor	0.13	0.20	0.26	0.22	0.37	-0.07	-0.16	0.26	0.52
T-stat	0.85	0.83	1.82	1.14	2.68	-0.36	-0.48	1.73	3.42
$R^2(\%)$	0.53	0.62	0.80	0.70	1.21	0.46	0.57	0.82	1.92

Panel B: Predictors Controlled, 12 Month Horizon

	bm	dspr	dp	ep	ltr	ntis	svar	tspr	corpr
$\log \zeta_t$	**-2.15**	**-2.69**	**-1.65**	**-2.11**	**-2.07**	**-2.03**	**-2.02**	**-2.21**	**-2.08**
T-stat	**-1.74**	**-2.12**	**-1.34**	**-1.78**	**-1.76**	**-1.71**	**-1.69**	**-1.90**	**-1.78**
predictor	2.01	1.96	3.49	2.82	1.63	-0.43	0.68	3.04	1.96
T-stat	1.40	1.48	2.59	1.77	3.44	-0.23	0.94	2.45	4.00
$R^2(\%)$	3.31	3.07	6.57	4.89	2.74	1.74	1.80	5.46	3.22

Panel C: Predictors Controlled, 60 Month Horizon

	bm	dspr	dp	ep	ltr	ntis	svar	tspr	corpr
$\log \zeta_t$	**-10.78**	**-13.31**	**-8.27**	**-10.63**	**-10.85**	**-10.95**	**-10.75**	**-11.21**	**-10.88**
T-stat	**-3.56**	**-3.80**	**-3.06**	**-3.85**	**-3.44**	**-3.39**	**-3.40**	**-3.62**	**-3.46**
predictor	6.32	7.11	14.24	8.86	1.84	-1.56	2.56	10.52	2.43
T-stat	1.97	2.47	5.79	2.03	1.56	-0.48	1.39	3.56	1.96
$R^2(\%)$	12.92	13.56	25.99	16.22	9.89	9.79	10.03	19.23	10.10

Appendices

Appendix I Factor,Idiosyncratic Risk and Diversification in a Standard APT

I first present definitions of important elements in arbitrage pricing theory, which are kept throughout the appendix. Based on the factor model in APT, we can decompose the covariance among n returns $\boldsymbol{\Sigma^n}$ into two parts, strong covariance from factors $\boldsymbol{\Sigma^n_f}$ and covariance among idiosyncratic risk $\boldsymbol{\Sigma^n_\epsilon}$. We define factor and idiosyncratic risk by covariance as in Chamberlain (1983):

Definition 1. *A portfolio described in vector form $\boldsymbol{w} = [w_1, .., w_n]$ is well-diversified if:*

$$\lim_{n\to\infty} \|\boldsymbol{w}\|_2^2 = \lim_{n\to\infty} \sum_i^n w_i^2 = 0 \tag{I.1}$$

$\|\boldsymbol{w}\|_2^2 = \sum_i^n w_i^2$ is the portfolio weight vector's "2-norm" that measures the dispersion, or variance among portfolio weights. A well-diversified portfolio has zero weight dispersion at the limit of infinite assets, which means that all assets have the roughly same size. For example, an equal-weighted portfolio is well-diversified since its size dispersion has order $O(1/n)$: $\sum_i^n w_i^2 = 1/n$. In a market with n asset, let $\rho_i(\Sigma), i = 1...n$ be the eigenvalues of a covariance matrix Σ, sorted in descending order.

Definition 2. *Σ^n_f have a factor structure if:*

$$\exists k \leq n, s.t. \lim_{n\to\infty} \rho_{i=1..k}(\boldsymbol{\Sigma^n}) = \infty \tag{I.2}$$

The factor structure is defined by unbounded eigenvalues of the covariance, or "pervasive" components among returns. If there is one portfolio that correlates with sufficiently many assets, then it is a factor. Idiosyncratic risk is defined by the complement:

Definition 3. *Σ^n_ϵ is idiosyncratic if:*

$$\lim_{n\to\infty} \rho_i(\boldsymbol{\Sigma^n}) \leq C, \forall i \tag{I.3}$$

In other words, covariance among assets can be decomposed into two parts, a strongly correlated factor structure and an idiosyncratic "residual" variance. Our definition is same as Chamberlain (1983) and Chamberlain and Rothschild (1983), which generalize the assumptions in Ross (1976). I hybridize these general definitions with a standard APT model in the textbook of Connor and Korajczyk (1995) and present the perspective that when diversification fails, idiosyncratic risk produces aggregate risk premium. The definition implies that there is no portfolio that contains only idiosyncratic risk that could have a strong correlation with all the assets. We repeat the basic setup in the **Section ??** in matrix form and present the derivation of APT. There are n firms in the whole asset space, each has a return:

$$\begin{aligned} \boldsymbol{r} &= E[\boldsymbol{r}] + \boldsymbol{B}\boldsymbol{f} + \boldsymbol{\epsilon} \qquad \text{(I.4)} \\ E[\boldsymbol{\epsilon}|\boldsymbol{f}] &= 0 \qquad \text{(I.5)} \end{aligned}$$

this leads to a variance decomposition:

$$\boldsymbol{\Sigma^n} = \boldsymbol{B}\boldsymbol{\Sigma^n_f}\boldsymbol{B'} + \boldsymbol{\Sigma^n_\epsilon} \qquad \text{(I.6)}$$

the term $\boldsymbol{B}\boldsymbol{\Sigma^n_f}\boldsymbol{B'}$ is a variation of our factor definition: I perform an eigenvalue decomposition to the defined factor covariance, where the factor loading is the eigenvector of covariance matrix. I further assume that there is a representative investor who has a quadratic utility base on the aggregate return $u(\boldsymbol{w'r})$ such that $u'' < 0,$ constant. The Euler equation:

$$E[u'(\boldsymbol{w'r})\boldsymbol{r}] = \boldsymbol{1}\gamma_0 \qquad \text{(I.7)}$$

where γ_0 is the reciprocal of the investor's subjective discount. Inserting the return equation (I.4) into the pricing formula gives:

$$E[\boldsymbol{r}] = \boldsymbol{1}\gamma_0 - \boldsymbol{B}\frac{E[u'(\boldsymbol{w'r})\boldsymbol{f}]}{E[u']} - \frac{E[u'(\boldsymbol{w'r})\boldsymbol{\epsilon}]}{E[u']} \qquad \text{(I.8)}$$

Use taylor expansion to $u'(\boldsymbol{w'r})$ at point $u'(\boldsymbol{w'}(\boldsymbol{E}[\boldsymbol{r}] + \boldsymbol{Bf}))$ gives:

$$u'(\boldsymbol{w'r}) \approx u'(\boldsymbol{w'}(\boldsymbol{E}[\boldsymbol{r}] + \boldsymbol{Bf})) + u''\boldsymbol{w'\epsilon} \qquad \text{(I.9)}$$

where the u'' is a constant due to the quadratic utility we assumed. We can approximate the last term $u'(\boldsymbol{w'r})\boldsymbol{\epsilon}$ by inserting the Taylor expansion result. Given the assumption that factor is independent from ϵ, the last term $E[u'(\boldsymbol{w'r})\boldsymbol{\epsilon}]$ is simplified to:

$$\begin{aligned} E[u'(\boldsymbol{w'r})\boldsymbol{\epsilon}] &\approx \boldsymbol{\Sigma^n_\epsilon}\boldsymbol{w}u'' \qquad \text{(I.10)} \\ & \qquad \text{(I.11)} \end{aligned}$$

Define the factor risk premium $\tau = \frac{E[u'(\boldsymbol{w'r})\boldsymbol{f}]}{E[u']}$ as the factor risk premium and $\gamma = -\frac{u''}{E[u']} > 0$ we can have:

$$E[\boldsymbol{r}] = \boldsymbol{1}\gamma_0 + \boldsymbol{B\tau} + \boldsymbol{\Sigma^n_\epsilon}\boldsymbol{w}\gamma \qquad \text{(I.12)}$$

The market risk premium is:

$$E[r_m] = \boldsymbol{w'}E[\boldsymbol{r}] = \gamma_0 + \boldsymbol{w'B\tau} + \boldsymbol{w'\Sigma^n_\epsilon w}\gamma \qquad \text{(I.13)}$$

When the market portfolio is well-diversified, the granular risk premium $e^g(n) = \gamma\boldsymbol{w'\Sigma^n_\epsilon w}$

converge to zero as n approaching infinity:

$$\lim_{n\to\infty} e^g(n) = \lim_{n\to\infty} \gamma \boldsymbol{w}'\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}^{\boldsymbol{n}}\boldsymbol{w} \leq \gamma \lim_{n\to\infty} \|\boldsymbol{w}\|_2^2 \rho_1(\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}^{\boldsymbol{n}}) = 0 \tag{I.14}$$

Appendix II Derivation using a Pareto Distribution

Now I use a Pareto distribution to derive a non-zero limitation of the diversification measure. The main tool I use here is a generalized convergence theorem for infinite-variance random variables (Durrett (2019),Theorem 3.8.2.):

Theorem 8. *Suppose X1,X2, . . . are i.i.d. with a distribution that satisfies*
(i) $\lim_{x\to\infty} P(X_1 > x)/P(|X_1| > x) = \theta \in [0,1]$
(ii)$P(|X_1| > x) = x^{-\alpha}L(x)$
where $\alpha < 2$ *and L is slowly varying. Let* $S_n = \sum_{i=1}^n X_i$
$a_n = \inf\{x : P(|X_1| > x) \leq n^{-1}\}$ *and* $b_n = nE(X_1 1_{|X_1|\leq a_n})$
As $n \to \infty, (S_n - b_n)/a_n \Rightarrow Y$ *where Y has a nondegenerate distribution.*

For a Pareto distribution such that $X_i, i.i.d$:

$$P(X_i > x) = x^{-\zeta}, x > 1$$

The Pareto coefficient $\zeta < 2$ for a fat tail distribution. In this case, $\theta = 1$,$\alpha = \zeta$ and $L(x) = 1$.

Now, let's figure out the convergence rate for $1/n \sum X_i$ in the denominator. For X_i:

$$a_n = n^{1/\zeta}$$

and

$$b_n = n \int_1^{n^{1/\zeta}} \zeta x^{-\zeta} dx$$

Now, there are some subtle difference for ζ at different ranges to calculate b_n

$$\begin{aligned} b_n &= n\left(n^{1/\zeta-1} - \frac{\zeta}{1-\zeta}\right) \approx n^{1/\zeta} = a_n, \zeta < 1 && \text{(II.1)} \\ &= n\left(n^{1/\zeta-1} - \frac{\zeta}{1-\zeta}\right) \approx n\frac{\zeta}{\zeta-1} = nE[X], \zeta > 1 && \text{(II.2)} \\ &= n \log n, \zeta = 1 && \text{(II.3)} \end{aligned}$$

Based on the calculation, I derive the convergence of the sample mean $1/n \sum X_i$:

$$1/n \sum X_i \to 1/n(a_n Y_1 + b_n)$$

$$
\begin{aligned}
1/n \sum X_i &\rightarrow n^{1/\zeta - 1}(Y_1 + 1), \zeta < 1 && \text{(II.4)} \\
&\rightarrow n^{1/\zeta - 1} Y_1 + E[X] \rightarrow E[X], \zeta > 1 && \text{(II.5)} \\
&\rightarrow Y_1 + \log n, \zeta = 1 && \text{(II.6)}
\end{aligned}
$$

where the characteristic function of Y_1 is a stable distribution with shape parameter ζ:

$$\varphi_{Y_1} = \exp\{t\mu i - \sigma|t|\zeta \left(1 + sign(t) w_\zeta(t) i\right)\}$$

Now, for the numerator, observe that X_i^2 also follow a Pareto distribution with index $\zeta/2 < 1$ so that:

$$1/n \sum X_i^2 \rightarrow n^{2/\zeta - 1}(Y_2 + 1) \tag{II.7}$$

where the characteristic function of Y_2 is a stable distribution with shape parameter $\zeta/2$:

$$\varphi_{Y_2} = \exp\{t\mu i - \sigma|t|^{\zeta/2} \left(1 + sign(t) w_{\zeta/2}(t) i\right)\}$$

The convergence in equation (6) is hence:

$$
\begin{aligned}
\lim_{n\rightarrow\infty} \sum w_i^2 &= \frac{Y_2 + 1}{(Y_1)^2}, \zeta < 1 && \text{(II.8)} \\
&= n^{2/\zeta - 2} \frac{Y_2 + 1}{E[X]^2}, \zeta > 1 && \text{(II.9)} \\
&= \frac{Y_2 + 1}{(Y_1 + \log n)^2}, \zeta = 1 && \text{(II.10)}
\end{aligned}
$$

Appendix III Evidence of granularity over decades

I provide the evidence of granularity in US stock market over decades in **Table IV.1**. Specifically, I compute the average market weight of all firms available in each decade from 1940-2020. I report the names of the ten largest firms and the summed market weight of these firms. This number is a rough measure of granularity populated by large firms used in **Figure 1** in December 2020. I present the results for the 40s-70s in Panel A and decades 80s-2020 in Panel B. The ten largest firms constantly account for around 25 percent of the total market valuation before the merging of NASDAQ-listed firms in the 70s. This number decreases to a minimum of 14 percent as the number of assets explodes to over 10000 during the 80s-90s. After the "internet bubble" period, the number of assets decreased, and the market weight of the ten largest firms reverted to around 25 percent in 2020.

In summary, the fat tail of firm size distribution (granularity) phenomenon has persisted over decades, though the list of large firms that populate this fat tail varies as production technology evolves. The largest firms in the economy were industrial or energy companies like General Motors and Stand Oil in decades the 40s-60s and transferred to big technology companies nowadays, like Apple, Microsoft, etc.

Appendix IV Additional tables and figures

Table IV.1: **Evidence of granularity over decades**

This table presents the names of the ten largest firms and their market weight in each decade.

Panel A: Summary of the 10 largest firms, 40s-70s

	1940	1950	1960	1970
1	GENERAL MOTORS (0.05)	STANDARD OIL NJ(0.05)	IBM (0.05)	IBM (0.05)
2	STANDARD OIL NJ(0.04)	GENERAL MOTORS (0.05)	GENERAL MOTORS (0.04)	STANDARD OIL NJ(0.03)
3	DUPONT (0.04)	DUPONT (0.04)	STANDARD OIL NJ(0.04)	GENERAL MOTORS (0.02)
4	GENERAL ELECTRIC (0.03)	GENERAL ELECTRIC(0.03)	TEXACO INC(0.02)	EASTMAN KODAK(0.02)
5	TEXASCO(0.02)	TEXASCO(0.02)	GENERAL ELECTRIC(0.02)	GENERAL ELECTRIC(0.02)
6	STANDARD OIL IND(0.01)	STANDARD OIL CAL(0.02)	DUPONT (0.02)	TEXACO(0.01)
7	STANDARD OIL CAL(0.01)	GULF OIL (0.02)	EASTMAN KODAK(0.01)	PROCTER & GAMBLE(0.01)
8	COCA COLA(0.01)	IBM (0.01)	GULF OIL (0.01)	MINNESOTA MINING & MFG(0.01)
9	GULF OIL (0.01)	SOCONY VACUUM OIL(0.01)	STANDARD OIL CAL(0.01)	DUPONT (0.01)
10	KENNECOTT COPPER (0.01)	STANDARD OIL IND(0.01)	MINNESOTA MINING & MFG(0.01)	STANDARD OIL CO IND(0.01)
Total weight	**0.24**	**0.26**	**0.24**	**0.19**
Number of assets	1019	1215	2995	6718

Panel B:Summary of the 10 largest firms, 80s-2020

	1980	1990	2000	2010	2020
1	IBM(0.04)	GE(0.02)	XOM(0.03)	AAPL(0.03)	AAPL(0.05)
2	XON(0.02)	XON(0.02)	GE(0.03)	GOOG(0.02)	MSFT(0.05)
3	GE(0.02)	KO(0.02)	MSFT(0.02)	MSFT(0.02)	AMZN(0.04)
4	SUO(0.01)	WMT(0.01)	WMT(0.02)	XOM(0.02)	GOOG(0.03)
5	SN(0.01)	IBM(0.01)	C(0.02)	BRK(0.02)	FB(0.02)
6	GM(0.01)	MSFT(0.01)	PFE(0.02)	BRK(0.02)	BRK(0.02)
7	MOB(0.01)	MRK(0.01)	JNJ(0.01)	AMZN(0.01)	JNJ(0.01)
8	SD(0.01)	PG(0.01)	INTC(0.01)	JNJ(0.01)	WMT(0.01)
9	BLS(0.01)	BMY(0.01)	CSCO(0.01)	WMT(0.01)	V(0.01)
10	DD(0.01)	JNJ(0.01)	IBM(0.01)	JPM(0.01)	JPM(0.01)
Summed weight	**0.15**	**0.14**	**0.17**	**0.18**	**0.25**
Number of assets	10428	12477	9040	6060	3823