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. In **Table 1**, I listed the ten largest firms over decades from the 1940s to the 2010s.[1] Although the list of these large firms varies as production technology evolves, they constantly account for a considerable fraction of the total market value. 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), sales as a proxy of 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 in-

[1] Specifically, I compute the average market weight of all firms available in each decade from the 1940s to 2010s.

dependent 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)).[2] 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

[2]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.

risk (product of an asset's market weight and idiosyncratic variance) positively explains the expected returns in the cross-section, with various factors and characteristics controlled[3]. 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).[4] 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 the 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 explains the time-variation of the expected returns on aggregate, especially in longer time horizons, even when controlling for additional predictors

[3]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.

[4]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.

surveyed in Welch and Goyal (2008).

Related Literature

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 exciting 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 also relates 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

My Granular APT model is a combination of using APT risk structure[5] to define idiosyncratic and factor risks and a Pareto distribution of firm size. The Pareto distribution brings tractability to capture the stylized facts shown in **Figure 1** and **Table 1**: Large firms have non-negligible weights in the market and hence breaks the diversification of idiosyncratic risks. Furthermore, the theoretical derivations using Pareto distribution holds in a finite economy with n assets and also in a sequence of economies as n approaches infinity.

I apply a simple competitive equilibrium in the stock market to derive how factors and idiosyncratic risks 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. This framework allows me to show how

[5]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)

the impact of idiosyncratic risks on 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 the 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:

$$r_i = E[r_i] + \sum_{s=1}^{k} \beta_{i,s} f_s + \epsilon_i; \tag{1}$$

$$E[\epsilon_i|f] = 0, \forall i. \tag{2}$$

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 the expectation of a constant absolute risk aversion (CARA) utility based 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.

Intuitively, the variance of $\sum w_i\epsilon_i$ depends on the composition of the market portfolio. If there are no large firms in the market such that all w_i are negligible, then the impact of idiosyncratic risks must be negligible due to the weak correlation among ϵ_i. Furthermore, as the number of asset n approaches infinity, the impact of idiosyncratic risk in (3) and (4) should converge to zero. APT models illustrate this intuition formally by making the diversification assumption of w_i. In the following section, I introduce the diversification assumption in APT and link it to the firm size distribution. Furthermore, I show that a thin-tailed distribution induces diversification in w_i.

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 $\mathbf{\Sigma}\boldsymbol{\epsilon}$ and $\rho_i(\mathbf{\Sigma}\boldsymbol{\epsilon}), i = 1...n$ 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 $n \to \infty$:

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

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 the number of assets approaches 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 such that the covariance matrix among* ϵ_i *has bounded eigenvalues as* $n \to \infty$*:*

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

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 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. Empirical works in the literature take the finite but sufficiently many assets observed in data as a good proxy of the theoretical results of $n \to \infty$. The fundamental assumption behind is that the diversification measure $\sum w_i^2$ converges to zero at a fast speed so that even with a finite n, the impact of idiosyncratic risk is negligible. Based on this assumption, researchers place a massive amount of effort on determining the correct 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 $\sum w_i^2$ relies on firm size distribution. Moreover, 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. I assume firms' market values X_i are independent and follow the same distribution. The weight 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 of X 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 extreme values is non-trivial, and the diversification assumption of APT models does not hold. The large firms that populate the fat tail have a dominant size. Hence their market weights would not converge to zero when n approaches infinity. In addition, the presence of these extremely large firms makes the first and second moments of X_i explode to infinity. Hence the diversification measure $\sum w_i^2$ does not converge to zero. Conceivably, the violation of APT raises a granularity effect in the expected returns in the format of $COV(\epsilon_i, \sum_i^n w_i \epsilon_i)$ as derived. These violations, even in a finite but large n economy, are crucial and cannot be ignored in the empirical works.

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. Specifically, the i moments of X are:

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

A small $\xi < 2$ implies a high probability of firms with extremely large values in the distribution and means a high level of the fat tail. As a result, the moments of firm size explode to infinity, and the sample average of X_i and X_i^2 in (6) does not converge to a finite value.

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, , which suggests a significant level of fat tail. In **Section 3**, I estimate the value of ζ using the firm size each month and find the estimation of Pareto distribution also fits the firm size in data well. Therefore, I use the Pareto distribution to drive violations of the APT models, which induces testable asset pricing implications. For simplicity, I focus on the fat tail case that $\zeta < 2$.

2.2.1 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. I illustrate this phenomenon by firstly solving the market weight of the maximum firm size in a sample of i.i.d Pareto distribution $X_{\max} = \max\{X_{1,\ldots,n}\}$. The maximum market weight w_{max} equals

$$w_{\max} = X_{\max} / \sum_{i=1}^{n} X_i.$$

In the thin-tailed case, the probability of extreme values converge to zero at a fast speed as n increases. As a result, $X_{\max}$ increases with n slowly as the largest value of a random draw from the Pareto distribution with n assets. On the other hand, the numerator $\sum_{i=1}^{n} X_i$ converges to $nE[X]$ and drives the market weight $w_{\max}$ to be negligible as n increases. When the fat tail is significant (ζ <2), the $X_{\max}$ becomes dominating and increases with n at a fast rate to make $w_{\max}$ significant. I formalize the result 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 maximum value $X_{\max} = \max\{X_{1,\ldots,n}\}$ would have its market weight $w_{\max} = X_{\max}/\sum_{i=1}^{n} X_i$ converge to*

$$\lim_{n\to\infty} w_{\max} = X_{\max}/\sum_{i=1}^{n} X_i = \begin{cases} \dfrac{F_\zeta}{Y_\zeta + 1} & \xi < 1 \\ \lim\limits_{n\to\infty} \dfrac{F_\zeta}{Y_\zeta + \log n} & \xi = 1 \\ \lim\limits_{n\to\infty} \dfrac{F_\zeta}{Y_\zeta + n^{1-1/\zeta} E[X]} & \xi > 1 \end{cases} \tag{9}$$

F_ζ is a random variable following the Frechet distribution with cumulative density function $e^{-x^{-\zeta}}, x > 0$. Y_ζ is a random variable following a stable distribution with the shape parameter equals ζ.

I show proof for **Lemma 4** in the **Appendix Section II**. I give heuristic explanations here to highlight the role of the fat tail in generating non-negligible market weights. With the fat tail, the scale of extreme values increases with n such that its appearance probability is around $1/n$ (the largest firm). Specifically, the extremely large values such that $X_i > a_n$, which is defined by

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

The largest firm value X_{max} is random depending on the realization, yet it has a scale around $a_n = n^{1/\zeta}$. Intuitively, I show that $X_{\max}/a_n$ converges to a random variable F_ζ with Frechet distribution (an implication of the Fisher–Tippett–Gnedenko theorem, see Gnedenko (1943)), which is also a fat-tail distribution. In other words, the extreme values increase with n at the rate of $n^{1/\zeta}$ and can be presented as $n^{1/\zeta}$ times a random variable F_ζ. Similarly, the convergence of $\sum X_i$ is stated by a "stable law" (see Durrett (2019), Theorem 3.8.2.) such that $\sum X_i/a_n$ converges to a stable distribution $Y_\zeta > 0$, which also have a fat tail with shape parameter ζ.

Combining the convergence of $X_{\max}$ and $\sum X_i$ gives the results in **Lemma 4**. When $1 < \zeta < 2$, the first moment of X is finite and $\sum X_i$ converges to $n^{1/\zeta} Y_\zeta + nE[X]$, which scale as n since $n^{1/\zeta} < n$. Consequently, large firms with a scale of $n^{1/\zeta}$ would have their market weight converge to zero at a rate of $n^{1/\zeta - 1}$. When the tail is heavy ($\zeta < 1$), large values around $n^{1/\zeta}$ would dominate the variation of $\sum X_i$ such that both the $X_{\max}$ and $\sum X_i$ increases with n at the same rate. Consequently, the market weight of the largest firm w_{max} does not converge to zero but converges to a positive random variable $\frac{F_\zeta}{Y_\zeta + 1}$. The case when $\zeta = 1$ is simply a limiting scenario of $\zeta > 1$ such that the rate of $w_{\max}$ converging to zero is $1/\log n$.

I verify the results in **Lemma 4** using simulation of the Pareto distribution to see how w_{max} changes with n in **Figure 2**. In the first subplot, $\zeta = 0.9 < 1$, the $w_{\max}$ does not converge to zero even when $n = 10^6$, yet it fluctuates as a random variable with non-negligible magnitude depending on the realization of $X_{\max}$. When $\zeta = 1.5$, the w_{max} also fluctuate as X_{max}, but converge to zero at the rate of $n^{1/\zeta - 1}$ as fitted by the red dash line. As another example, I also simulate the thin tail case $\zeta = 2.5$. With thin tail, $\sum X_i$ simply converges to $nE[X]$ by LLN, and the maximum value $X_{\max}$ can also be presented by $n^{1/\zeta} F_\zeta$. Consequently, the $w_{\max}$ converges to zero faster, as implied by my theoretical results, and the magnitude is negligible (around 0.1 percent).

Since ζ is estimated to be around 1, **Lemma 4** states a violation of APT that there are large firms with non-negligible weight in the market portfolio. When $\zeta < 1$, the market weight of the largest firm converges to a positive random variable independent of n. It could be several percent as in **Figure 1**, or even more than 80 percent as in the simulation results shown by **Figure 2**. In a finite economy with n assets, the significant magnitude of $w_{\max}$ exists even when $\zeta > 1$ since the convergence rate $n^{1/\zeta-1}$ is slow, which is a weak version of APT violation in a finite economy. For example, let $n = 10^5$ and $\zeta = 1.1$. Under this case, the deterministic term of n is $w_{\max}$ is calibrated to be:

$$\frac{1}{n^{1-1/\zeta}E[X]} = n^{1/\zeta-1}\frac{\zeta-1}{\zeta} = n^{1/1.1-1}\frac{1.1-1}{1.1} \approx 0.03,$$

which matches with the magnitude in **Figure 1**. The convergence rate of diversification is around $n^{-1/10}$ instead of $1/n = 1/10000$. In addition, the results for $w_{\max}$ hold for the few largest firms. The k largest firm X_k would have a magnitude such that,

$$P(X_k > x) \approx k/n$$

and scale as $n^{1/\zeta}k^{-1/\zeta} = a_n k^{-1/\zeta}$. In other words, the second-largest firm would have a market weight such that

$$w_2 \approx w_{\max} * 2^{-1/1.1}.$$

Similarly, the largest ten firms would have their summed market weight approximately equal $w_{\max} * \sum_{k=1}^{10} k^{-1/1.1} \approx 3.2 * w_{\max}$. Using the same example as in **Figure 1**, the largest firm has roughly 6 percent of the market weight, and this calibration suggests the summed weight of the ten largest firms is approximately equal to 20 percent. In other words, the fat tail distribution, in a finite but large n economy, creates large market weights of individual assets. This granular effect violates the APT assumption and must make the idiosyncratic risks of these large firms explain the expected return considerably. As a comparison of the maximum result, I derive the limiting convergence of $X_{\min} = \min\{X_{1,\ldots,n}\}$ in **Appendix Section II** to illustrate how fast small firms in the Pareto distribution would have their market converge to zero. The minimum weight of a small firm $w_{\min}$ converges to zero at a rate faster than $1/n$, which indicates that small firms do not violate the APT assumption.

The violation of APT models does not only appear in the cross-section such that there are large w_i. On aggregate, the fat tail breaks the diversification assumption that $\lim_{n\to\infty}\sum w_i^2 = 0$ as well. Using the Pareto distribution, I derive the limit of the diversification measure $\sum w_i^2$. Similar to the infinite value of the $\sum X_i$ for the first moment, the fat tail also breaks the LLN convergence of the $\sum X_i^2$. As a result, 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**.

2.2.2 Pareto distribution and failure of diversification

I derive the limit of the diversification measure $\lim_{n\to\infty}\sum w_i^2$ in the following lemma:

Lemma 5. *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.*

$$\lim_{n\to\infty}\sum w_i^2 = \begin{cases} \dfrac{Y_{\zeta/2}}{(Y_\zeta)^2} & \zeta < 1 \\ \lim\limits_{n\to\infty}\dfrac{Y_{\zeta/2}}{(Y_\zeta + \log n)^2} & \zeta = 1 \\ \lim\limits_{n\to\infty}\dfrac{Y_{\zeta/2}}{(Y_\zeta + n^{1-1/\zeta}E[X])^2} & \zeta > 1 \end{cases} \tag{10}$$

Y_ζ is a random variable following a stable distribution with the shape parameter equals ζ. Similarly, $Y_{\zeta/2}$ follows the stable distribution with shape parameter $\zeta/2$.

The derivation of **Lemma 5** is in **Appendix Section II**. The heuristic explanation of **Lemma 5** is simply an application of the "stable law ."Recall that,

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

The convergence of $\sum w_i^2$ hence depends on the convergence the sample average of X_i and X_i^2. The convergence of $1/n \sum X_i$ used in the last section is given by the stable law. The convergence of $1/n \sum X_i^2$ is solved similarly since X_i^2 also follows a Pareto distribution with the tail parameter $\zeta/2$.

I verify the results in **Lemma 5** using simulation of the Pareto distribution to see how $\sum w_i^2$ changes with n in **Figure 3**. In the first subplot, $\zeta = 0.9 < 1$, the $\sum w_i^2$ does not converge to zero even when $n = 10^6$, yet it fluctuates as a random variable with non-negligible magnitude depending on the realization of large firms. When $\zeta = 1.5$, the $\sum w_i^2$ also fluctuates as the appearance of large values but converges to zero at the rate of $n^{2/\zeta-2}$ as fitted by the red dash line. Intuitively, the convergence rate of $\sum w_i^2$ is simply the square power of $n^{1/\zeta-1}$, as the convergence rate of $w_{\max}$. For the thin tail case, both the first and second moments of X_i are finite, and the LLN holds. Therefore, in the last subplot ($\zeta = 2.5$), the $\sum w_i^2$ converges to zero at the rate of $1/n$ as fitted by the red dash line. Furthermore, the random realization of large values does not affect the convergence of $\sum w_i^2$ due to the LLN.

Lemma 5 suggests the constant failure of the diversification assumption in APT models. When $\zeta < 1$, the diversification measure $\sum w_i^2$ converges to a positive random variable independent of n. As shown in **Figure 3**, this large variation of $\sum w_i^2$ is driven by the large values of X_i. In a finite economy with n assets, the significant magnitude of $\sum w_i^2$ exists even when $\zeta > 1$ since the convergence rate $n^{2/\zeta-2}$ is slow, which is a weak version of APT violation in a finite economy. Using the same 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 $1/n = 1/10000$. Therefore, the granularity of firm size must also have a strong impact on the aggregate market fluctuation in a finite n economy.

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 $\lim_{n\to\infty} \sum w_i^2 \neq 0$. In addition, these two results hold well in a finite economy with sufficiently many assets, as observed in the data. 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. As discussed in the last section, my derivations when $n \to \infty$ is also well approximated by the results when n is sufficiently large enough in data. Therefore, I directly use the limiting case to discuss the associating asset pricing tests.

2.3.1 granularity and the idiosyncratic risk puzzle

I use the result in **Lemma 4** 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 4****. If the idiosyncratic shocks are independent of each other with variance* θ_i*, then the expected return for each asset converges to:*

$$\lim_{n\to\infty} E[r_i] = \mu_0 + \sum_{s=1}^{k} \beta_{i,s}\mu_s + \theta_i \gamma \lim_{n\to\infty} w_i. \tag{11}$$

The idiosyncratic variance θ_i, by definition, is bounded and hence the limitation of $w_i\theta_i$ is determined by the convergence of w_i. 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 (11) 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 $corr(w_i\theta_i, \theta_i)$, such that

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

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. Consequently, the magnitude of $w_i\theta_i$ is mainly driven by the granularity in w_i. I plot the $w_i\theta_i$ of individual assets at the end of 2020 in **Figure 4**. Comparing this plot to **Figure 1** shows that the large firms tend to have high $w_i\theta_i$ and model-implied alpha relative to factor models. Moreover, the magnitude of $w_i\theta_i$ shown in **Figure 4** is empirically reasonable. Assuming a risk aversion coefficient $\gamma = 5$ gives 2.5 percent of α annually for the largest $w_i\theta_i$ firm in **Figure 4**.

To summarize, my model suggests that large firms (low idiosyncratic risk) have a significantly higher risk premium tied to their idiosyncratic risks than small firms (high idiosyncratic risk). As a result, the 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,

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

Therefore, 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:*

$$\lim_{n\to\infty} E[r_m] = \mu_0 + \sum_i^n w_i \left(\sum_{s=1}^k \beta_{i,s}\mu_s \right) + \gamma \lim_{n\to\infty} \sum w_i^2 \theta_i. \tag{12}$$

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 $\theta_{\min} \leq \theta_i \leq \theta_{\max}$, hence,

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

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 5**, 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 the 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:

$$\log(r_{m,t+1}) = \text{constant} + \text{controls} + \mathrm{A} \log \zeta_{\mathrm{t}}. \tag{13}$$

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 quarterly COMPUSTAT characteristics data (replaced with annual data if not available). 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

The theoretical derivation suggests that large firms violate the APT assumption as $n \to \infty$. In a dataset with finite but sufficiently many assets, the tests for my theoretical results are still valuable since it highlights the granular channel of idiosyncratic risk premium, given the non-negligible market weights of large firms observed in **Figure 1** and **Table 1**.

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.

Notably, the functional format $\alpha_i = \gamma w_i \theta_i$ is derived at the individual asset level and is not invariant to a test using portfolios sorted by θ_i. Therefore, I also perform the cross-sectional test at the firm level to be consistent with my theoretical derivations. Nevertheless, tests at the portfolio level are useful benchmark to compare my model implication to typical tests in the cited papers, which emphasizes the economic mechanism of idiosyncratic risks to explain expected returns due to granularity.

3.2.1 Portfolio level test using five 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 2**. 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:

$$r_{L-H,t} = \frac{1/\theta_{L,t-1}}{1/\theta_{L,t-1} - 1/\theta_{H,t-1}}(r_{L,t} - r_f) - \frac{1/\theta_{H,t-1}}{1/\theta_{L,t-1} - 1/\theta_{H,t-1}}(r_{H,t} - r_f). \tag{14}$$

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 2** has an annualized average return equal to 7.36 percent and volatility equal to 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 (for example, see Giglio and Xiu (2021), Giglio, Xiu, and Zhang (2021)).

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 3**. 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 4**, 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 an 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:

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

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 $\hat{\eta} = -1.78$ with a significant T-value. 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 significant T-value. This estimate is consistent with what the model implies since a positive estimate of $\hat{\gamma}$ represents the risk-aversion coefficient. In addition, to understand whether the size-adjusted risk $w_i\theta_i$ has more explanatory power than θ_i , I normalize $w_i\theta_i$ and θ_i to make their standard deviation equal one and estimate a constrained regression,

$$\alpha_i = constant + \lambda w_i \theta_i + (1 - \lambda)\theta_i.$$

The estimate of $\hat{\lambda} = 3.13$ with a significant T-value. The estimate suggests that the granular channel of the idiosyncratic risk premium is the dominating force to explain α_i.

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 θ_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 θ_i such that

$$corr(w_i\theta_i, \theta_i) = -0.61.$$

Intuitively, the negative correlation must be driven by the relationship between market weights w_i and idiosyncratic risk θ_i. The correlation between size and risk, under this context, equals to:

$$corr(w_i, \theta_i) = -0.56.$$

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 dominantly drives the correlation between $w_i\theta_i$ and θ_i to negative, due to the negative 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.

To better illustrate this idea, I plot the relationship between size w_i and θ_i of the 100 portfolios in **Figure 5**. I find this negative relationship can be well approximated by:

$$\log \theta_i \approx constant + a \log w_i.$$

I plot this close to a linear relation between logged θ_i and w_i. This relation is an interesting pattern in the data, which is worthy of further investigation. Nevertheless, the granular explanation of IRP relies on the dominance of size effect in w_i to make large firms have high $w_i\theta_i$. In **Figure 6**, I plot the relationship between $w_i\theta_i$ and θ_i of the 100 portfolios. The dot size in this plot is scaled by the total market weight of each portfolio w_i. The granularity in w_i dominantly drives the distribution of $w_i\theta_i$ and hence explains the IRP as explained since only low θ_i portfolios have non-negligible w_i and $w_i\theta_i$. In contrast, the high θ_i portfolios have close to zero w_i and $w_i\theta_i$.

I examine the robustness of my 100-portfolio results for different lengths of measurement window and using different factor models. In **Table 5**, I summarize the estimate of η, γ, and the constrained estimate λ, together with the estimated correlations $corr(w_i\theta_i, \theta_i)$, $corr(w_i, \theta_i)$ 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 $corr(w_i\theta_i, \theta_i)$ and $corr(w_i, \theta_i)$ 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. The magnitude of λ, as the constrained estimate of $w_i\theta_i$, does not change much as the measurement window increases. This finding suggests the robustness of my model implication: The size adjustment is essential to understand the relationship between idiosyncratic risks and expected returns.

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:

$$r_{i,t} = \mu_0 + \sum_{s=1}^{k} \beta_{i,s,t} \left(f_{s,t} + \mu_s\right) + \eta\theta_{i,t-1} + \epsilon_{i,t}. \tag{15}$$

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 $\hat{\eta} < 0$.[6]

To compare to the test in Ang et al. (2009), I generalize (11) to be time-varying and estimate:

$$r_{i,t} = \mu_0 + \sum_{s=1}^{k} \beta_{i,s,t} \left(f_{s,t} + \mu_s\right) + \gamma w_{i,t-1}\theta_{i,t} + \epsilon_{i,t}. \tag{16}$$

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

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 (16). 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 (15) except that they use the past idiosyncratic variance $\theta_{i,t-1}$ 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. Similarly, to emphasize the importance of size adjustment, I estimate a constrained model:

$$r_{i,t} = \mu_0 + \sum_{s=1}^{k} \beta_{i,s,t}\left(f_{s,t} + \mu_s\right) + \lambda w_{i,t-1}\theta_{i,t} + (1-\lambda)\theta_{i,t-1} + \epsilon_{i,t}. \tag{17}$$

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 6**, I report the cross-sectional regression estimates $\hat{\eta}$ (using $\theta_{i,t-1}$) and $\hat{\gamma}$ (using $w_{i,t-1}\theta_{i,t}$) separately. I also report $\hat{\lambda}$ in the constrained model as in (17) using both $w_{i,t-1}\theta_{i,t}$ and $\theta_{i,t-1}$. In column 1, I estimate a significant negative coefficient $\hat{\eta} = -2.23$, 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 column 2, I use both $w_{i,t-1}$ and $\theta_{i,t-1}$ as two variables to explain the returns. The coefficient for $\theta_{i,t-1}$ is still negatively significant with the size controlled, and the magnitude of the coefficient does not change. 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 $\theta_{i,t-1}$ 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 addition, I estimate the constrained regression using both the $\theta_{i,t-1}$ and $w_{i,t-1}\theta_{i,t}$ as in (17) to emphasize the granular effect in expected returns. The estimate of $\hat{\lambda}$ is 0.71 with significance. Also, this constrained model, in the time-varying setup, helps to identify a positive relationship between $\theta_{i,t-1}$ and $r_{i,t}$. This finding concretely highlights the importance of using size adjustment to test whether idiosyncratic risks explain risk premiums, as implied by my model.

In **Table 7**, I show similar results using 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 -2.22, and the $\hat{\gamma}$ is 6.73. The estimate of $\hat{\lambda}$ is 0.72, also similar to the estimate using FF factors, which indicates the importance of controlling granularity. 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 $\hat{\lambda}$ as in (17),

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

In **Table 8**, I present the results of measuring $\theta_{i,t}$ by different estimation windows and estimating the constrained model in (17) (column 6 in **Table 6**). 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 estimate of λ is stable across specifications using different measurement windows and factor models. These results suggest the robustness of the granular channel for idiosyncratic risk to explain expected returns.

3.3 Time-Series Test

3.3.1 Estimate the Pareto distribution

The main results of this paper hinge on the Pareto coefficient ζ value, which quantifies the level of granularity and the associating asset pricing implication. When $X_{i=1\ldots n}$ are i.i.d and follows the exact Pareto distribution in (7) such that

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

The Pareto distribution implicitly assumes that only firms with market values larger than x_m follow a Pareto distribution. Selecting a threshold to estimate the Pareto distribution excludes the small firms in the sample, which is consistent with the theoretical motivation that large firms induce violations of the APT models.

I estimate the tail parameter ζ of the Pareto distribution using the Hill estimator (see Hill (1975)). At each month, I sort all the n firm sizes in a descending order $X_{i=1,\ldots,n}$ and select a threshold value $x_m = X_k$ to use the largest k firms for estimating ζ. The Hill estimator is:

$$\zeta = \left\{1/k \sum_{i=1}^{k} (\log X_i - \log X_k)\right\}^{-1}. \tag{18}$$

this estimator can be interpreted as a maximum likelihood estimator of ζ conditioning on a known minimum threshold $x_m = X_k$, which has a simple to derive asymptotic inference property as $k \to \infty$. Therefore, the literature typically selects the threshold position k by fixing a cutoff ratio $k/n = 5\%, 10\%\ldots$ to make k proportional to the total number of assets n and conduct the statistical inference by the asymptotic property of the estimator as $n \to \infty$.

I find that the large firms in the stock market are fitted well by the Hill estimator of the Pareto distribution, which justifies my theoretical derivations[7]. Specifically, matching the survival probability in (7) with the frequency in data gives,

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

The logarithm of this equation implies a linear relationship between logged rank i and size X_i in (7) since:

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

Therefore, to check the goodness of fitting, I plot the logged rank-size plot of the largest 10% firms in the December of 2020 in **Figure 7**. I fit the linear relationship in the red dash line using the Hill estimator of $\hat{\zeta} = 0.94$, which suggests a significant level of the fat tail and the APT violations as implied by my model. Meanwhile, I 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 might induce a downward bias of the Hill estimator.

For time-series implication in my paper, the cutoff selection affects the predictability of ζ on market returns as motivated in (13):

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

A loose cutoff ratio k/n (large k) would include more firms and reduce the estimator's variance for better statistical power of my time-series test. However, a loose cutoff also

[7]This assumption should not affect the theoretical results in **Section 2.2** since small firms only account for a tiny fraction of the total value. Furthermore, I can derive the same theoretical results when the whole sample is drawn from a mixture of the Pareto distribution and a thin tail distribution. The proof is available upon request.

generates a downward bias of ζ since it could include small firms in the sample that may not follow the Pareto distribution. In the **Appendix Section III**, I use the largest 10% firms as a benchmark estimator and find that the magnitude of ζ is around 1. In addition, 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 discuss methods to exclude the non-stationary impact of n on the estimate of ζ in the **Appendix Section III** and proceed to the time-series results. 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 $\log \zeta_t$.

3.3.2 Time-series results

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

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

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. 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.

I summarize the main results in **Table 9**. In Panel A, 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 horizons 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 9**, I report the results using the whole sample at various horizons $k = 1, 12, 60$: 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 $k = 12, 60$.

Meanwhile, 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 the 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 8** and find a very similar magnitude of idiosyncratic risk changing over time. In Panel B, C, and D of **Table 9**, I report the results controlling the idiosyncratic risk under the three measures above, respectively. The magnitude of the coefficient almost does not change, controlling for the idiosyncratic risk, yet the significance of predictability is generally weaker.

In **Figure 9**, I plot the time-series estimator $\log \zeta_t$ 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.

3.3.3 Control for alternative predictors

I use predictors listed in Welch and Goyal (2008) as controls for other systematic risks to identify the granular channel of risk premium better. In **Table 10**, I provide a summary of predictors, including their definitions, AR1 coefficients, and their correlation coefficients with the main predictor $\log \zeta_t$. 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 explains the time-variation of the market capital value. The results confirm the economic intuition that a low ζ indicates a high risk premium due to the 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's expected returns.

For robustness, 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. I summarize the out-of-sample results in the **Appendix Section IV**.

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.

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.

Gnedenko, Boris, 1943, Sur la distribution limite du terme maximum d'une serie aleatoire, *Annals of mathematics* 423–453.

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 at the end of 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 over 25 percents of the whole CRSP data in 2020 contains about 4,000 firms.

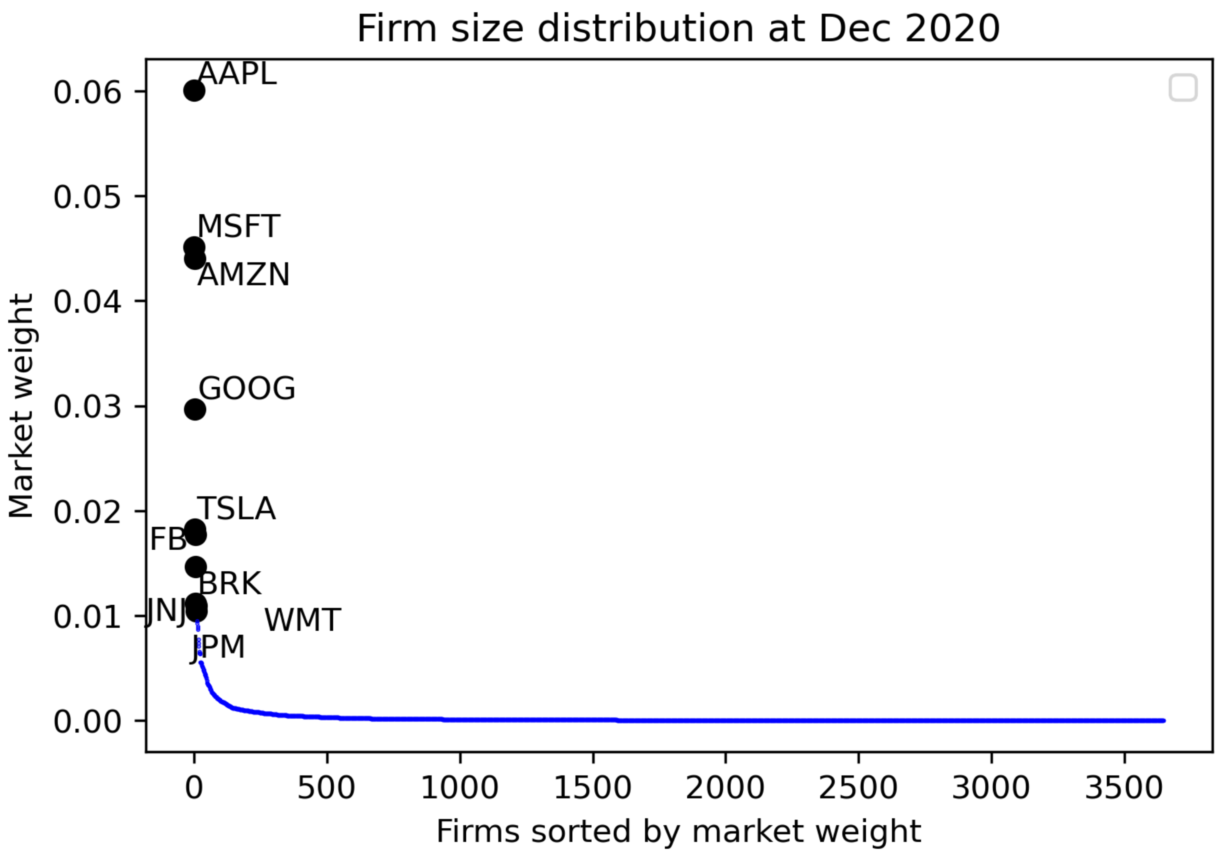
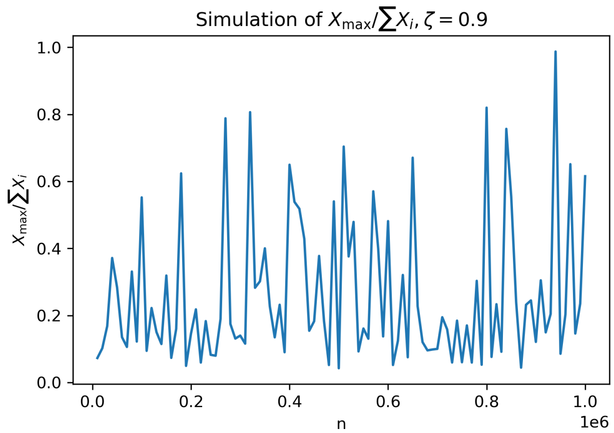
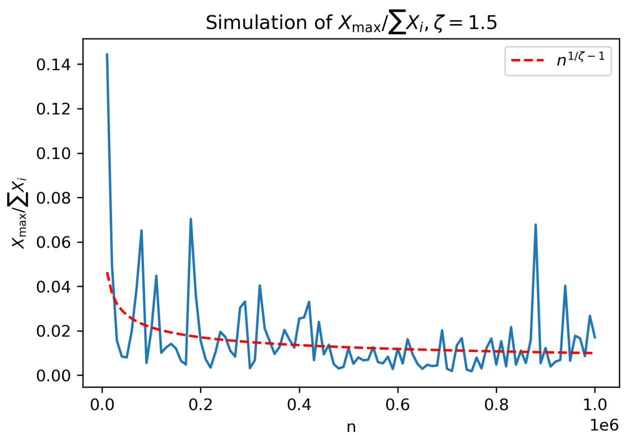


Figure 2: **Simulation of the largest firm's market weight**. In this figure, I use simulation of Pareto distribution with $\zeta = 0.9, 1.5$ and 2.5 to study how the market weight of the largest firm $w_{\max} = \frac{X_{\max}}{\sum X_i}$ changes as n increases.

(a) $\zeta = 0.9$



(b) $\zeta = 1.5$



(c) $\zeta = 2.5$

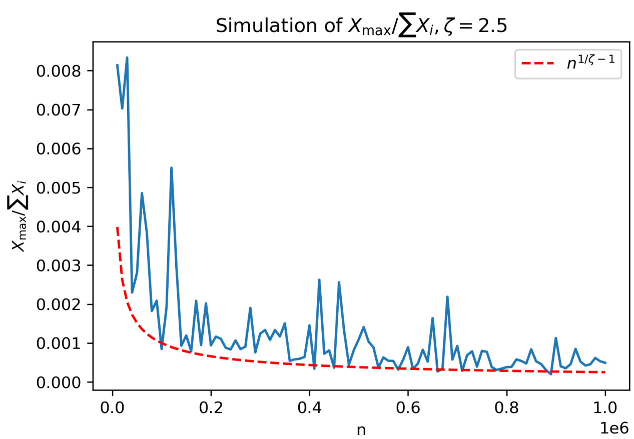
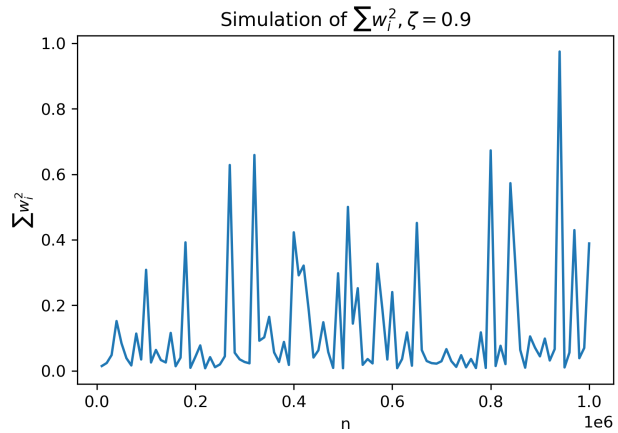
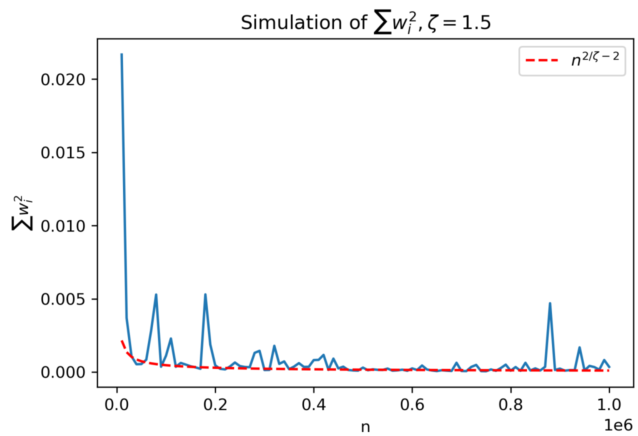


Figure 3: **Simulation of the $\sum_i^n w_i^2$ as n increases** . In this figure, I use simulation of Pareto distribution with $\zeta = 0.9, 1.5$ and 2.5 to study how the diversification measure $\sum_i^n w_i^2$ changes as n increases.

(a) $\zeta = 0.9$



(b) $\zeta = 1.5$



(c) $\zeta = 2.5$

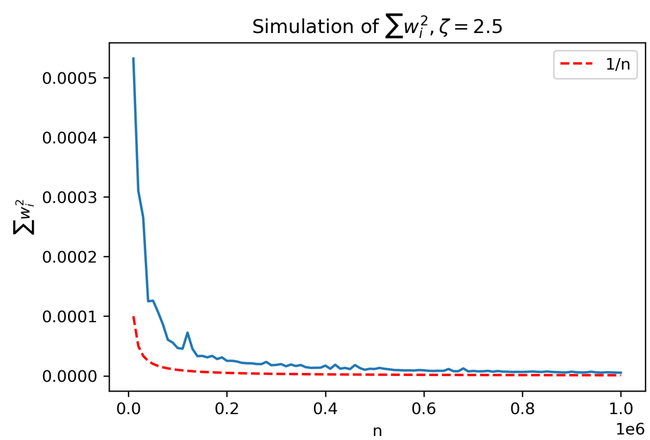


Figure 4: **Size-adjusted idiosyncratic risk of individual assets**. In this figure, I plot the $w_i\theta_i$ of individual assets sorted by market weight at the end of 2020. The dot size is scaled by the total market weight of each portfolio w_i.

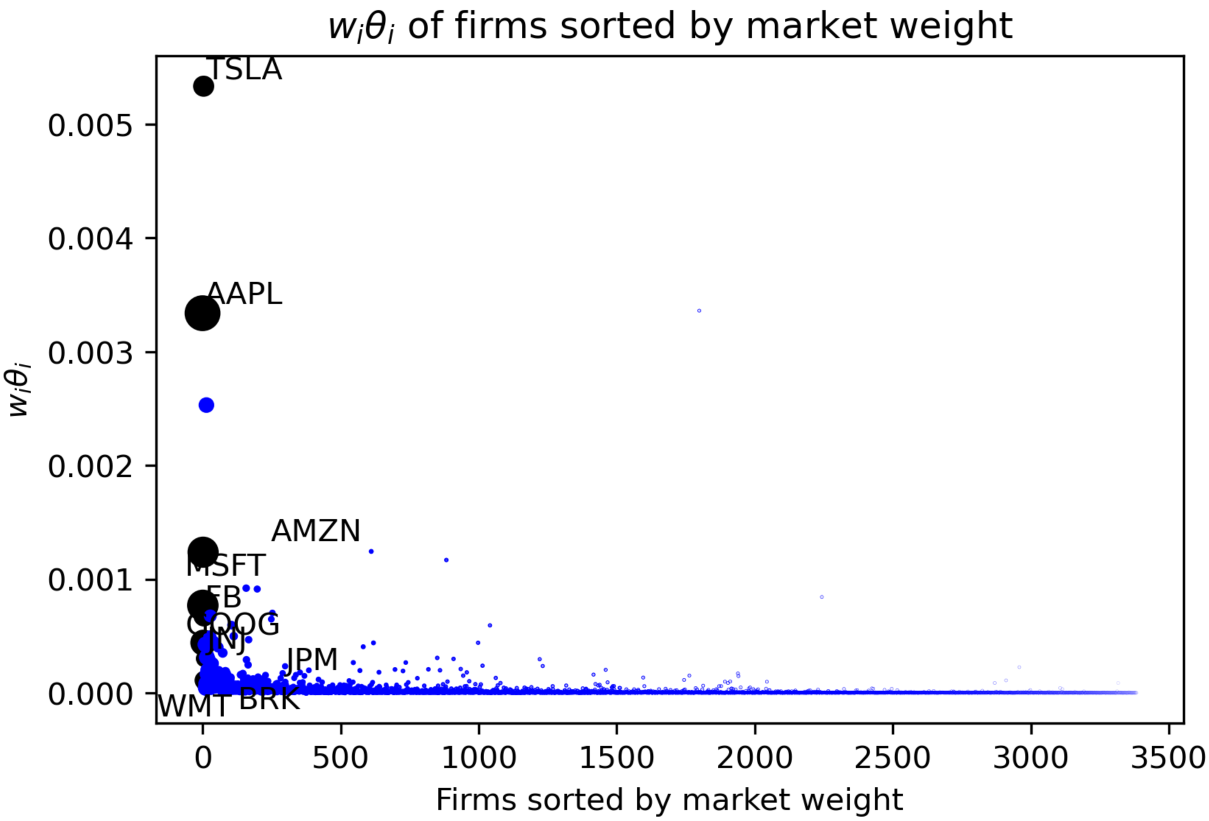


Figure 5: **Size and idiosyncratic risk of the 100 sorted portfolios in log scale**. In this figure, I plot the relation between θ_i and w_i of the 100 portfolios sorted by θ_i in log scale. The dot size is scaled by the total market weight of each portfolio w_i.

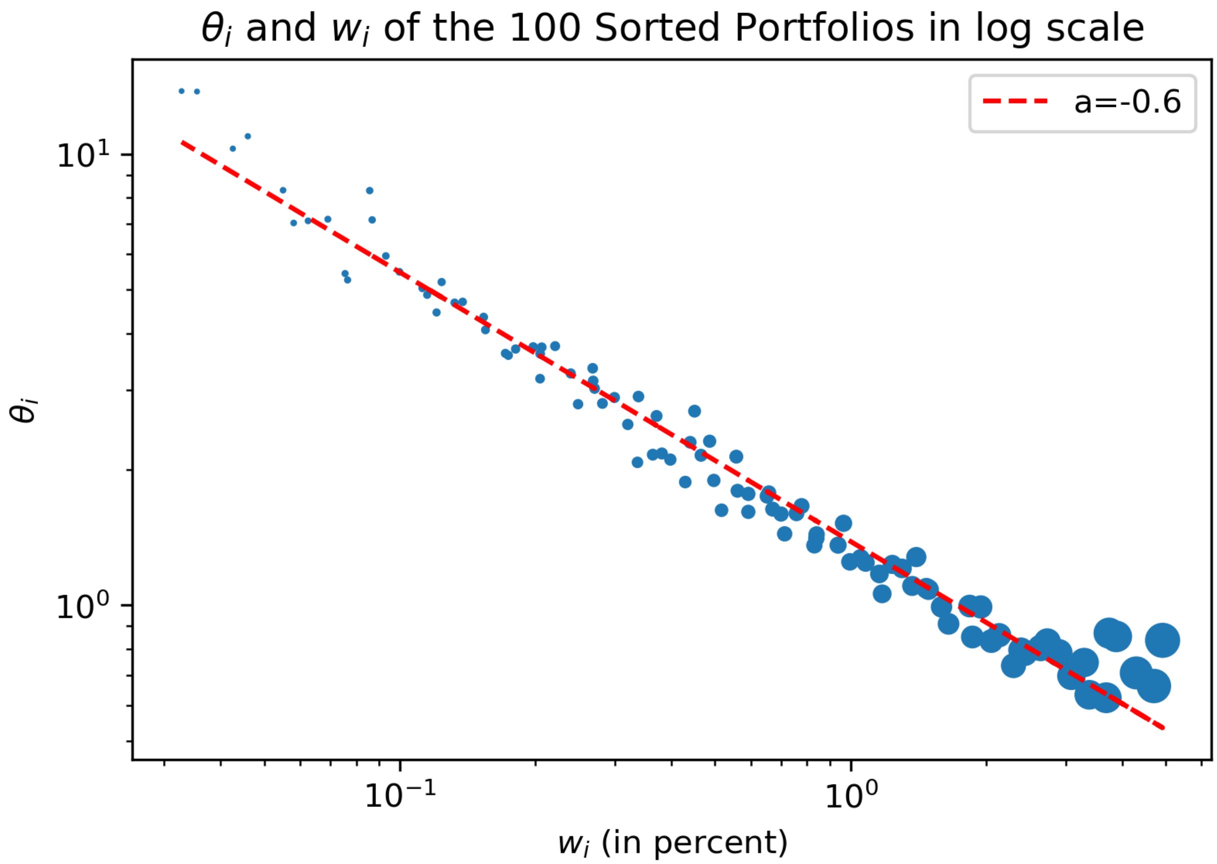


Figure 6: **Size-adjusted idiosyncratic risk of the 100 sorted portfolios**. In this figure, I plot the relation between $w_i\theta_i$ and θ_i of the 100 portfolios sorted by θ_i. The dot size is scaled by the total market weight of each portfolio w_i.

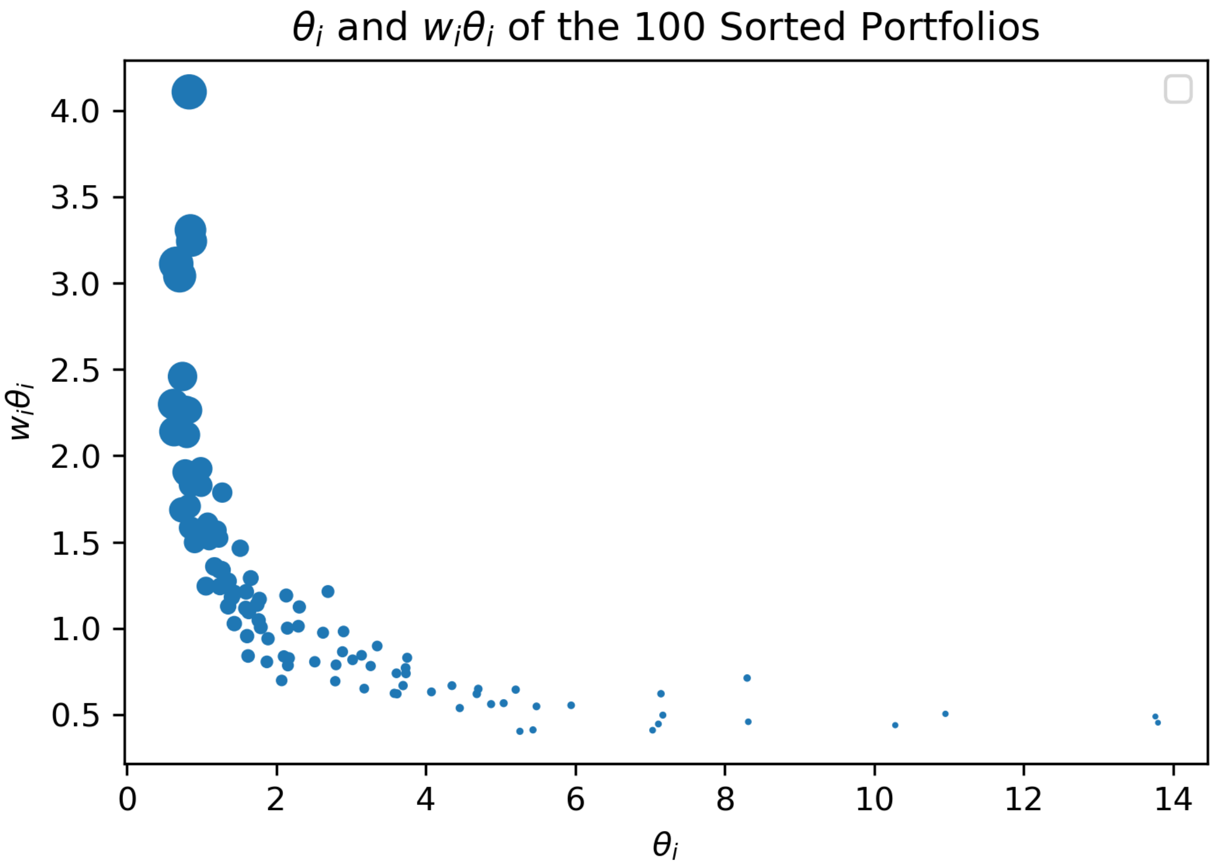


Figure 7: **Logged rank-size plot in December 2020** In this figure, I plot the logged rank-size plot of the largest 20% firms in December 2020. The red dash line show the fitted relation implied by the Pareto distribution. The 10 largest firms are highlighted.

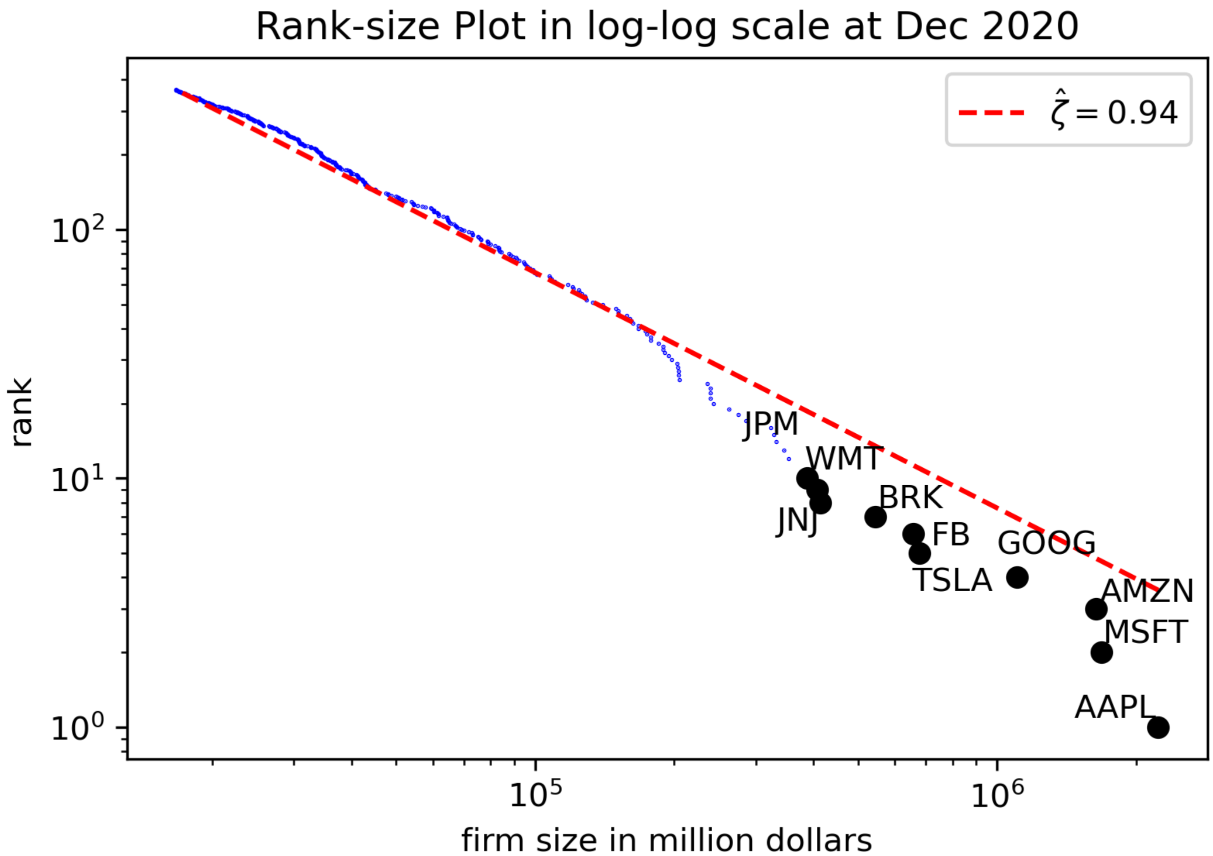


Figure 8: **Three measures of idiosyncratic risk** . I measure the level of idiosyncratic risk at the aggregate level by the weighted average of θ. Specifically, I use FF3 factors, or the three principal components of daily returns, to measure each asset's idiosyncratic variance $\theta_{i,t}$ in each 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.

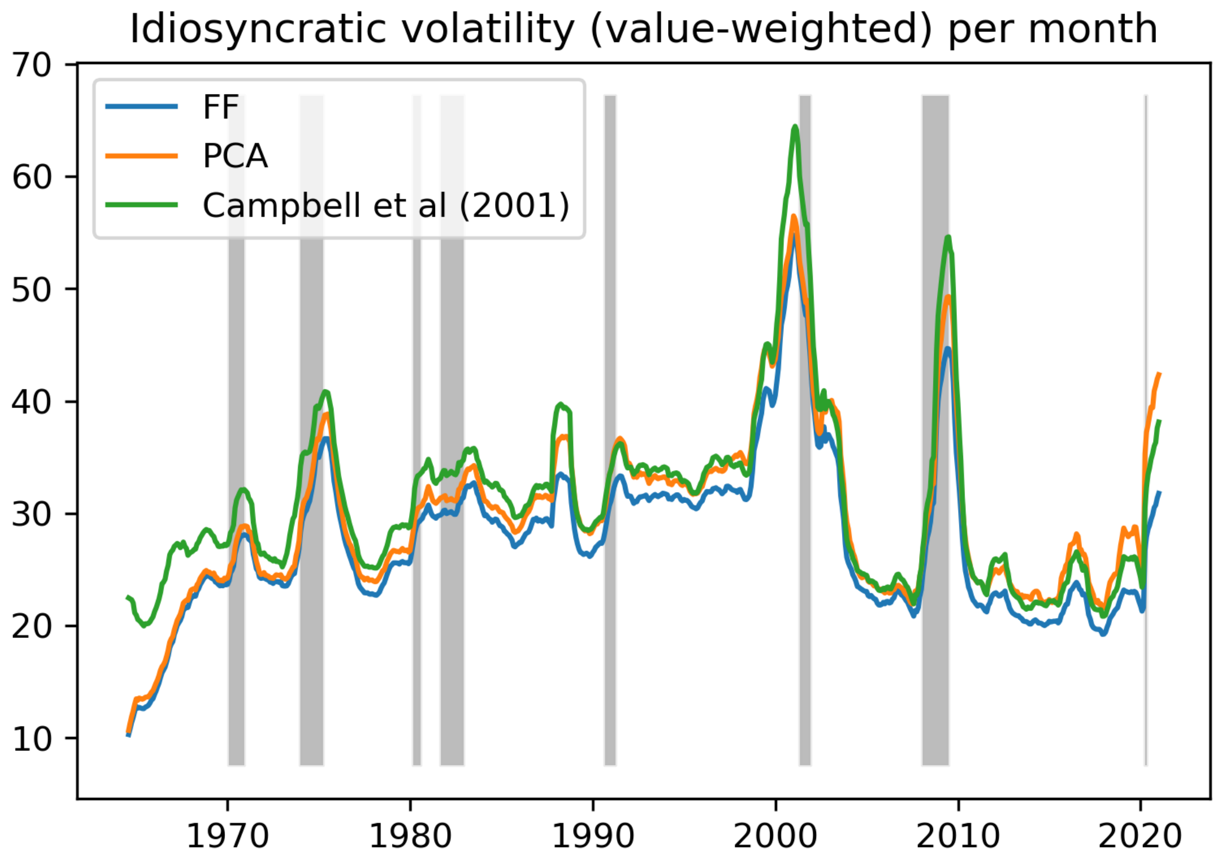


Figure 9: **Pareto coefficient (de-trended) v.s. idiosyncratic risk**. I measure the level of idiosyncratic risk at the aggregate level by the weighted average of idiosyncratic risk relative to Fama-French 3 factors. I plot the two series together where the blue line is the Pareto predictor and the yellow line is the weighted average of idiosyncratic variance. The shaded areas are NBER recessions.

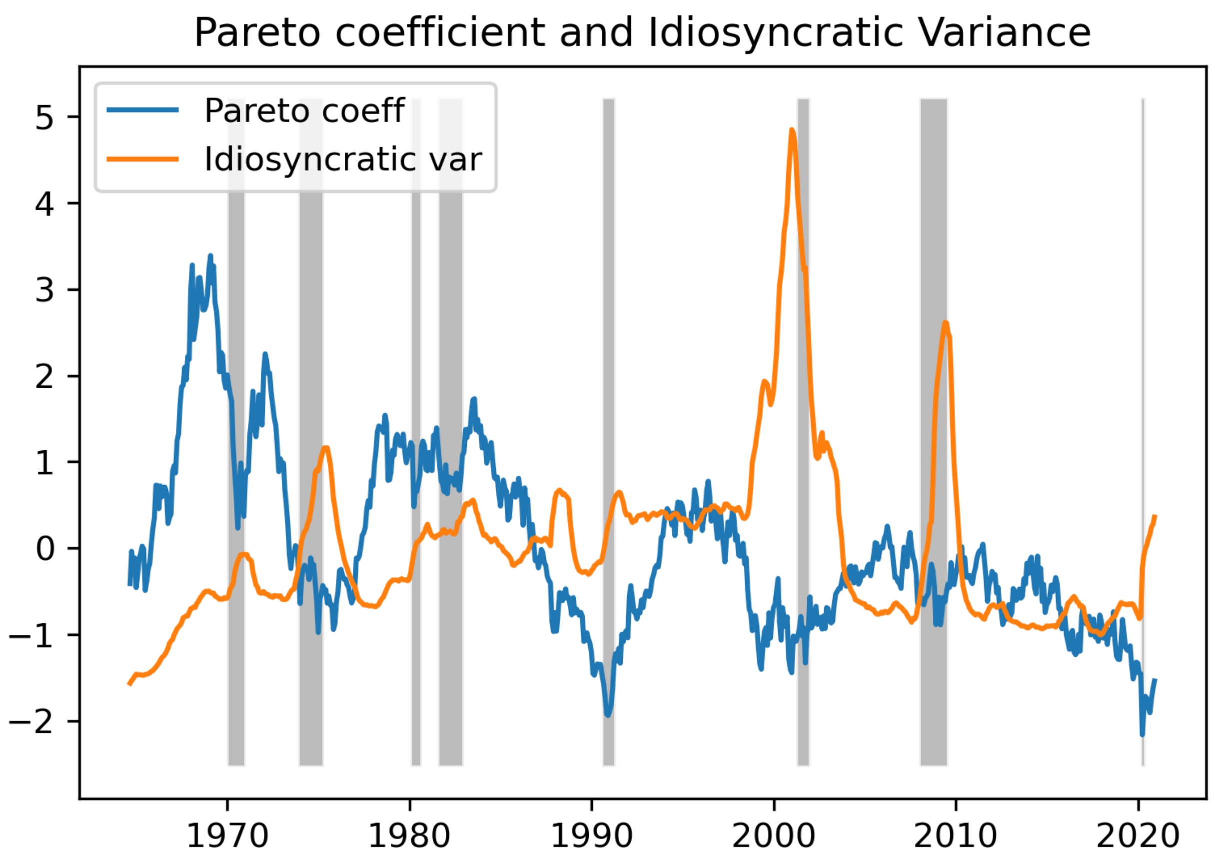


Table 1: **Evidence of granularity over decades**

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, 1940s-1970s				
	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, 1980s-2010s				
	1980	1990	2000	2010
1	IBM(0.04)	GE(0.02)	XOM(0.03)	AAPL(0.03)
2	XON(0.02)	XON(0.02)	GE(0.03)	GOOG(0.02)
3	GE(0.02)	KO(0.02)	MSFT(0.02)	MSFT(0.02)
4	SUO(0.01)	WMT(0.01)	WMT(0.02)	XOM(0.02)
5	SN(0.01)	IBM(0.01)	C(0.02)	BRK(0.02)
6	GM(0.01)	MSFT(0.01)	PFE(0.02)	BRK(0.02)
7	MOB(0.01)	MRK(0.01)	JNJ(0.01)	AMZN(0.01)
8	SD(0.01)	PG(0.01)	INTC(0.01)	JNJ(0.01)
9	BLS(0.01)	BMY(0.01)	CSCO(0.01)	WMT(0.01)
10	DD(0.01)	JNJ(0.01)	IBM(0.01)	JPM(0.01)
Summed weight	**0.15**	**0.14**	**0.17**	**0.18**
Number of assets	10428	12477	9040	6060

Table 2: **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 3: **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 2, 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 4: **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 5: **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

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

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

In addition, I normalize $w_i\theta_i$ and θ_i to make their standard deviation equals 1 and estimate a constrained regression

$$\alpha_i = constant + \lambda w_i \theta_i + (1-\lambda)\theta_i$$

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	-1.78	-1.76	-1.64	-1.28
T-stat	-15.90	-15.63	-18.32	-18.09
γ	5.17	4.67	3.90	3.16
T-stat	8.72	7.78	7.88	7.60
λ	3.13	3.42	3.37	2.87
T-stat	17.21	15.85	17.66	17.47
$corr(w_i\theta_i, \theta_i)$	-0.61	-0.58	-0.54	-0.52
$corr(w_i, \theta_i)$	-0.56	-0.51	-0.47	-0.44
Panel B: 3 principal components				
estimates \ window length	1	3	6	12
η	-1.71	-1.77	-1.51	-1.29
T-stat	-16.42	-16.50	-17.03	-14.69
γ	6.22	6.03	4.85	3.87
T-stat	10.21	10.75	9.84	9.28
λ	2.99	3.40	3.28	2.88
T-stat	18.53	19.73	19.74	18.10
$corr(w_i\theta_i, \theta_i)$	-0.70	-0.67	-0.61	-0.59
$corr(w_i, \theta_i)$	-0.62	-0.57	-0.53	-0.51

Table 6: **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. To emphasize the importance of size-adjustment, I estimate a constrained model $r_{i,t} = \mu_0 + \sum_{s=1}^{k} \beta_{i,s,t}\left(f_{s,t} + \mu_s\right) + \lambda w_{i,t-1}\theta_{i,t} + (1-\lambda)\theta_{i,t-1} + \epsilon_{i,t}$ in the column 6. 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						
	$r_{i,t}$	$r_{i,t}$	$r_{i,t}$	$r_{i,t}$	$r_{i,t}$	$r_{i,t}$
const	0.56	0.57	0.58	0.49	0.49	0.38
	3.00	2.98	2.98	2.58	2.60	1.96
$\hat{\beta}_{i,t}^{Mkt-RF}$	0.00	0.00	-0.00	-0.01	-0.00	-0.03
	0.05	0.05	-0.08	-0.19	-0.05	-0.53
$\hat{\beta}_{i,t}^{SMB}$	0.04	0.04	0.04	0.04	0.04	0.04
	1.65	1.70	1.66	1.54	1.59	1.53
$\hat{\beta}_{i,t}^{HML}$	-0.01	-0.01	-0.00	0.00	-0.00	0.00
	-0.19	-0.20	-0.02	0.10	-0.08	0.12
$b/m_{i,t-1}$	0.24	0.24	0.24	0.25	0.25	0.25
	8.65	8.60	8.57	8.84	8.90	8.78
$mom_{i,t-1}$	-0.45	-0.46	-0.45	-0.48	-0.49	-0.36
	-2.09	-2.15	-2.01	-2.14	-2.30	-1.66
$\hat{\theta}_{i,t-1}$	-2.23	-2.23			-2.24	0.29 (constrained)
	-1.98	-2.04			-2.09	7.79
$w_{i,t-1}$		-0.08	-0.11	-1.86	-1.73	-3.18
		-0.47	-0.59	-5.05	-4.80	-13.17
$w_{i,t-1}\hat{\theta}_{i,t}$				9.15	8.77	0.71 (constrained)
				8.99	8.73	19.34

Table 7: **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. To emphasize the importance of size-adjustment, I estimate a constrained model $r_{i,t} = \mu_0 + \sum_{s=1}^{k} \beta_{i,s,t} \left(f_{s,t} + \mu_s\right) + \lambda w_{i,t-1}\theta_{i,t} + (1 - \lambda)\theta_{i,t-1} + \epsilon_{i,t}$ in the column 6. Different from the main result in Table 6, 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						
	$r_{i,t}$	$r_{i,t}$	$r_{i,t}$	$r_{i,t}$	$r_{i,t}$	$r_{i,t}$
const	0.65	0.66	0.65	0.59	0.59	0.48
	3.41	3.40	3.30	2.98	3.09	2.41
$\hat{\beta}_{i,t}^{PCA1}$	8.09	8.08	8.10	8.03	8.02	7.89
	7.75	7.74	7.69	7.66	7.71	7.54
$\hat{\beta}_{i,t}^{PCA2}$	5.94	5.95	6.01	5.99	5.93	5.94
	6.32	6.34	6.30	6.28	6.32	6.29
$\hat{\beta}_{i,t}^{PCA3}$	5.23	5.23	5.32	5.34	5.25	5.27
	6.77	6.78	6.80	6.85	6.83	6.80
$\mathrm{b/m_{i,t-1}}$	0.23	0.23	0.23	0.24	0.24	0.24
	8.60	8.53	8.40	8.67	8.83	8.66
$\mathrm{mom_{i,t-1}}$	-0.44	-0.45	-0.40	-0.42	-0.47	-0.33
	-2.08	-2.14	-1.83	-1.93	-2.25	-1.57
$\hat{\theta}_{i,t-1}$	-2.22	-2.23			-2.25	0.28 (constrained)
	-2.01	-2.08			-2.13	7.97
$w_{i,t-1}$		-0.14	-0.15	-2.00	-1.93	-4.04
		-0.79	-0.81	-4.73	-4.66	-16.55
$w_{i,t-1}\hat{\theta}_{i,t}$				6.73	6.52	0.72 (constrained)
				7.97	7.77	20.99

Table 8: **Fama-MacBeth results, robustness check for measurement window of idiosyncratic risk**
This table examine the robustness of results (column 6) in Table 6, $r_{i,t} = \mu_0 + \sum_{s=1}^{k} \beta_{i,s,t} \left(f_{s,t} + \mu_s\right) + \lambda w_{i,t-1}\theta_{i,t} + (1-\lambda)\theta_{i,t-1} + \epsilon_{i,t}$ **I estimate 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	0.38	0.31	0.33	0.39	const	0.48	0.42	0.41	0.42
	1.96	1.97	2.25	2.89		2.41	2.36	2.44	2.36
$\hat{\beta}_{i,t}^{Mkt-RF}$	-0.03	-0.09	-0.18	-0.25	$\hat{\beta}_{i,t}^{PCA1}$	7.89	1.43	-0.09	1.43
	-0.53	-0.96	-1.69	-2.22		7.54	1.40	-0.11	1.40
$\hat{\beta}_{i,t}^{SMB}$	0.04	0.05	0.10	0.09	$\hat{\beta}_{i,t}^{PCA2}$	5.94	1.80	3.57	1.80
	1.53	0.82	1.40	1.15		6.29	1.44	3.88	1.44
$\hat{\beta}_{i,t}^{HML}$	0.00	0.00	0.01	0.01	$\hat{\beta}_{i,t}^{PCA3}$	5.27	1.90	-0.39	1.90
	0.12	-0.04	0.12	0.19		6.80	1.81	-0.48	1.81
$\mathrm{b/m_{i,t-1}}$	0.25	0.23	0.23	0.23	$\mathrm{b/m_{i,t-1}}$	0.24	0.23	0.22	0.23
	8.78	9.71	9.88	9.35		8.66	9.57	9.41	9.57
$\mathrm{mom_{i,t-1}}$	-0.36	0.19	0.03	-0.41	$\mathrm{mom_{i,t-1}}$	-0.33	0.08	-0.10	0.08
	-1.66	1.06	0.16	-2.22		-1.57	0.44	-0.60	0.44
$1-\lambda\ (\hat{\theta}_{i,t-1})$	0.29	0.38	0.37	0.36	$1-\lambda\ (\hat{\theta}_{i,t-1})$	0.28	0.33	0.30	0.33
	7.79	10.25	9.48	9.00		7.97	8.14	7.02	8.14
$w_{i,t-1}$	-3.18	-3.15	-3.39	-3.65	$w_{i,t-1}$	-4.04	-4.11	-4.50	-4.11
	-13.17	-13.61	-13.61	-13.59		-16.55	-15.18	-15.39	-15.18
$\lambda\ (w_{i,t-1}\hat{\theta}_{i,t})$	0.71	0.62	0.63	0.64	$\lambda\ (w_{i,t-1}\hat{\theta}_{i,t})$	0.72	0.67	0.70	0.67
	19.34	16.47	16.35	15.81		20.99	16.82	16.12	16.82

Table 9: **Time-series results**

In this table, I reports the monthly times-series results for the Pareto coefficient ζ to predict log excess-return for the aggregate market. A lower ξ implies a fatter tail, the hypothesize predictive relation should be negative $A < 0$. In Panel A, I check the prediction results at multiple-horizons at various horizon $k = 1, 12, 60$, $\log(r_{m,t+k}) = \text{constant} + \text{A} \log \zeta_\text{t}$. Furthermore, I control for the level of idiosyncratic risk in $\log(r_{m,t+k}) = \text{constant} + \text{A} \log \zeta_\text{t} + \sum \text{w}_{\text{i,t}-1} \theta_{\text{i,t}}$. 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 B, C and D, respectively.

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
Panel B: 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
Panel C: 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
Panel D: 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

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 $k = 1, 12, 60$.

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

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 de-trended 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

In this section, I give a proof of **Lemma 1** and **Lemma 2**. 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:*

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

$\sum_i^n w_i^2$ measures the dispersion of the portfolio weights or variance among portfolio weights. A well-diversified portfolio has zero weight dispersion at the limit of infinite assets, meaning that all assets are roughly the same size. For example, an equal-weighted portfolio is well-diversified since its size dispersion scales as $1/n$: $\sum_i^n w_i^2 = 1/n$. Based on this definition, a proof of **Lemma 1** is straightforward:

Proof of Lemma 1: If there exists an asset such that $\lim_{n\to\infty} w_i \neq 0$, then the diversification measure

$$\lim_{n\to\infty} w_i^2 \neq 0.$$

Therefore, to satisfy the diversification condition, it must be $\lim_{n\to\infty} w_i = 0, \forall i$.

Now I proceed to the APT derivation. To prove the **Lemma 2**, We repeat the basic setup in the **Section 2.1** in matrix form and present the derivation of APT. There are n firms in the whole asset space; each has a return:

$$\boldsymbol{r} = E[\boldsymbol{r}] + \boldsymbol{B}\boldsymbol{f} + \boldsymbol{\epsilon}, \tag{I.2}$$
$$E[\boldsymbol{\epsilon}|\boldsymbol{f}] = 0. \tag{I.3}$$

this leads to a variance decomposition:

$$\boldsymbol{\Sigma^n} = \boldsymbol{B}\boldsymbol{\Sigma^n_f}\boldsymbol{B'} + \boldsymbol{\Sigma^n_\epsilon}. \tag{I.4}$$

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 the covariance matrix. This method is consistent with Chamberlain (1983) and Chamberlain and Rothschild (1983), which generalize the assumptions in Ross (1976). Precisely, they define factors and idiosyncratic risks by the eigenvalue of the covariance matrix. 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:*

$$\exists k \le n, s.t. \lim_{n\to\infty} \rho_{i=1..k}(\mathbf{\Sigma^n}) = \infty. \tag{I.5}$$

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_ϵ^n *is idiosyncratic if:*

$$\lim_{n\to\infty} \rho_i(\mathbf{\Sigma^n}) \le C, \forall i. \tag{I.6}$$

In other words, covariance among assets can be decomposed into two parts, a strongly correlated factor structure, and an idiosyncratic "residual" variance. 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.

I further assume that there is a representative investor who has a CARA utility base on the aggregate return $u(\boldsymbol{w'r})$ such that $u'' < 0,$ constant. The Euler equation:

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

where γ_0 is the reciprocal of the investor's subjective discount. Inserting the return equation (I.2) into the pricing formula gives:

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

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

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

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:

$$E[u'(\boldsymbol{w'r})\boldsymbol{\epsilon}] \;\;\approx\;\; \gamma\boldsymbol{\Sigma_\epsilon^n w}E[u']. \tag{I.10}$$

where the risk aversion coefficient is $\gamma = -\frac{u''}{u'} > 0$.

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

$$E[\boldsymbol{r}] = \mathbf{1}\gamma_0 + \boldsymbol{B\tau} + \boldsymbol{\Sigma_\epsilon^n w}\gamma. \tag{I.11}$$

The covariance term $COV(\epsilon_i, \sum_i^n w_i\epsilon_i)$ in (3) is stacked into the vector $\boldsymbol{\Sigma_\epsilon^n w}$. The market risk premium is:

$$E[r_m] = \boldsymbol{w'}E[\boldsymbol{r}] = \gamma_0 + \boldsymbol{w'B\tau} + \boldsymbol{w'\Sigma_\epsilon^n w}\gamma. \tag{I.12}$$

When the market portfolio is well-diversified, the granular risk premium $e^g(n) = \gamma VAR(\sum_i^n w_i \epsilon_i) = \gamma \boldsymbol{w'}\boldsymbol{\Sigma}_\epsilon^n \boldsymbol{w}$ converge to zero as n approaching infinity, which gives the proof of **Lemma 2**.
Proof of Lemma 2: With diversification,

$$\lim_{n\to\infty} e^g(n) = \lim_{n\to\infty} \gamma \boldsymbol{w'}\boldsymbol{\Sigma}_\epsilon^n \boldsymbol{w} \leq \gamma \lim_{n\to\infty} \sum_{i=1}^n w_i^2 \rho_1(\boldsymbol{\Sigma}_\epsilon^n) = 0. \tag{I.13}$$

Furthermore, the vector term $\boldsymbol{\Sigma}_\epsilon^n \boldsymbol{w}$ in the expected return of each asset is smaller or equal to $\boldsymbol{w'}\boldsymbol{\Sigma}_\epsilon^n \boldsymbol{w}$, and hence converge to zero. As a result,

$$\lim_{n\to\infty} E[\boldsymbol{r}] = \mathbf{1}\gamma_0 + \boldsymbol{B\tau}.$$

Appendix II Derivation using a Pareto Distribution

I show the proof of **Lemma 3** as the case of the thin-tail distribution.
Proof of Lemma 3: Recall that,

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

If the first and second moments of X_i is finite, then:

$$\lim_{n\to\infty} 1/n \sum (X_i)^2 = E[X^2],$$

$$\lim_{n\to\infty} 1/n \sum X_i = E[X].$$

Therefore, the diversification measure converges to:

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

Now I use a Pareto distribution to derive the violation of the APT assumption when $\zeta < 2$. I start with the proof of **Lemma 4**. Recall that the maximum market weight w_{max} equals:

$$w_{\max} = X_{\max} / \sum_{i=1}^n X_i$$

The derivation for $w_{\max}$ is invariant to re-scale of X_i. Therefore, for simplicity, I normalize the lower bound of Pareto distribution x_m to equal one such that:

$$P(X_i > x) = x^{-\zeta}, x > 1 \tag{II.2}$$

The limiting distribution of the maximum value from an i.i.d sample following any distribution is derived by the Fisher–Tippett–Gnedenko theorem (see Gnedenko (1943)).

I use this theorem on the Pareto distribution to show that the $X_{\max}$ converges to a Frechet distribution in the following lemma:

Lemma 8. *If the firm size X_i follows an i.i.d Pareto distribution such that*

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

Define $a_n = n^{1/\zeta}$, then the maximum value $X_{\max} = \max\{X_{1,\ldots,n}\}$ has a limiting distribution such that:

$$\lim_{n\to\infty} P(X_{max}/a_n \le x) = \lim_{n\to\infty} F^n(a_n x) = e^{-x^{-\zeta}}.$$

$X_{\max}/a_n$ converges to a random variable F_ζ that follows a Frechet distribution with tail parameter ζ.

Proof of Lemma 8: The proof is an implication of the Fisher–Tippett–Gnedenko theorem. By definition,

$$F(a_n x) = 1 - (a_n x)^{-\zeta} = 1 - \frac{1}{n} x^{-\zeta}.$$

The limiting distribution of $X_{\max}/a_n$ is given by:

$$\lim_{n\to\infty} P(X_{max}/a_n \le x) = \lim_{n\to\infty} F^n(a_n x) = \lim_{n\to\infty} \left(1 - \frac{1}{n} x^{-\zeta}\right)^n = e^{-x^{-\zeta}}.$$

The convergence of $\sum X_i$ when $\zeta < 2$ is given by the stable law, which is a generalized convergence theorem for infinite-variance random variables (Durrett (2019), Theorem 3.8.2.):

Theorem. *(Stable Law) 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 $S_n = \sum_{i=1}^n X_i$
$a_n = \inf\{x : P(|X_1| > x) \le n^{-1}\}$ *and* $b_n = nE(X_1 1_{|X_1| \le a_n})$
As $n \to \infty$, $(S_n - b_n)/a_n \Rightarrow Y$ where Y has a non-degenerate distribution.

I apply this theorem to the Pareto distribution. The random variable Y, in this context, have the shape parameter ζ. I denote the convergence to be Y_ζ and specify how the characteristic function of Y_ζ in the following derivations. For the Pareto distribution in (II.2), $\theta = 1$, $\alpha = \zeta$ and $L(x) = 1$, such that

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

and

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

The magnitude of b_n depends on the range of ζ such that:

$$b_n = \begin{cases} n\left(n^{1/\zeta-1} - \dfrac{\zeta}{1-\zeta}\right) \approx n^{1/\zeta} = a_n & \zeta < 1 \\ n\left(n^{1/\zeta-1} - \dfrac{\zeta}{1-\zeta}\right) \approx n\dfrac{\zeta}{\zeta-1} = nE[X] & \zeta > 1 \\ n\log n & \zeta = 1 \end{cases} \tag{II.3}$$

With these calculations, I derive the convergence of $\sum X_i$:

$$\lim_{n\to\infty} \sum X_i = \lim_{n\to\infty} (a_n Y_\zeta + b_n),$$

such that,

$$\lim_{n\to\infty} \sum X_i = \begin{cases} \lim\limits_{n\to\infty} n^{1/\zeta}(Y_\zeta + 1) & \zeta < 1 \\ \lim\limits_{n\to\infty} Y_\zeta + \log n & \zeta = 1 \\ \lim\limits_{n\to\infty} n^{1/\zeta} Y_\zeta + nE[X] & \zeta > 1 \end{cases} \tag{II.4}$$

where the characteristic function of Y_ζ, $\varphi_{Y_\zeta}(t)$, is a stable distribution with shape parameter ζ:

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

where $sign(t)$ is the sign function and w_t is a function determined by ζ:

$$\begin{aligned} w_\zeta(t) &= \tan(\pi\zeta/2), \zeta \neq 1 && \text{(II.5)} \\ &= \pi/2 \log|t|, \zeta \neq 1 && \text{(II.6)} \end{aligned}$$

A distribution with this type of characteristic function is known as a stable distribution. μ and σ are the location and scale parameters, and the shape parameter is determined by ζ, the Pareto coefficient of X.

Combining the results above gives the convergence of $w_{\max} = X_{\max}/\sum X_i$ as in **Lemma 4**:

$$\lim_{n\to\infty} w_{\max} = X_{\max}/\sum_{i=1}^{n} X_i = \begin{cases} \dfrac{F_\zeta}{Y_\zeta + 1} & \xi < 1 \\ \lim\limits_{n\to\infty} \dfrac{F_\zeta}{Y_\zeta + \log n} & \xi = 1 \\ \lim\limits_{n\to\infty} \dfrac{F_\zeta}{Y_\zeta + n^{1-1/\zeta} E[X]} & \xi > 1 \end{cases} \tag{II.7}$$

As a comparison of the maximum result, I derive the limiting convergence of $X_{\min} = \min\{X_{1,\ldots,n}\}$ to illustrate how fast small firms in the Pareto distribution would have their market converge to zero and hence does violate the APT assumption.

Lemma 9. *If the firm size X_i follows an i.i.d Pareto distribution such that*

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

The minimum value $X_{\min} = \min\{X_{1,\ldots,n}\}$ has a limiting distribution such that:

$$\lim_{n\to\infty} P(n(X_{min} - 1) > x) = \lim_{n\to\infty} P^n(X > x/n + 1) = e^{-x\zeta}.$$

Therefore, $n(X_{min} - 1)$ converges to a random variable $\exp_\zeta$ that follows a exponential distribution with shape parameter ζ.

Proof of Lemma 9: The proof of this lemma is quite straightforward since:

$$\lim_{n\to\infty} P(n(X_{min} - 1) > x) = \lim_{n\to\infty} P^n(X > x/n + 1) = \lim_{n\to\infty} \left[(1/nx + 1)^n\right]^{-\zeta} = e^{-x\zeta}.$$

Therefore, the cumulative density function of $n(X_{min} - 1)$ is $1 - e^{-x\zeta}$ as n approaches infinity, which is an exponential distribution. In other words, the minimum value $X_{\min}$ decreases with n at the rate of $1/n$. As a result, one can show that the minimum market weight $w_{\min}$ converges to:

$$\lim_{n\to\infty} w_{\min} = X_{\min} / \sum_{i=1}^{n} X_i = \begin{cases} \lim\limits_{n\to\infty} \dfrac{1 + \exp_\zeta / n}{n^{1/\zeta}(Y_\zeta + 1)} & \xi < 1 \\ \lim\limits_{n\to\infty} \dfrac{1 + \exp_\zeta / n}{Y_\zeta + \log n} & \xi = 1 \\ \lim\limits_{n\to\infty} \dfrac{1 + \exp_\zeta / n}{n^{1/\zeta} Y_\zeta + nE[X]} & \xi > 1 \end{cases} \tag{II.8}$$

As a comparison of the maximum results, the minimum market weight always converges to zero faster than $1/n$, which indicates that small firms do not violate the APT assumption.

The proof of **Lemma 5** is another implication of the stable law to derive the convergence of $\sum X_i^2$. Now, since X_i^2 also follow a Pareto distribution with index $\zeta/2 < 1$, the convergence is:

$$\lim_{n\to\infty} \sum X_i^2 = \lim_{n\to\infty} n^{2/\zeta}(Y_{\zeta/2} + 1). \tag{II.9}$$

Similarly, the characteristic function of $Y_{\zeta/2}$ is a stable distribution with shape parameter $\zeta/2$:

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

Proof of Lemma 5: Combining the results in (II.4) and (II.9) gives the convergence of $\sum_i^n w_i^2$:

$$\lim_{n\to\infty}\sum w_i^2 = \begin{cases} \dfrac{Y_{\zeta/2}+1}{(Y_\zeta+1)^2} & \zeta<1 \\ \lim\limits_{n\to\infty}\dfrac{Y_{\zeta/2+1}}{(Y_\zeta+\log n)^2} & \zeta=1 \\ \lim\limits_{n\to\infty}\dfrac{Y_{\zeta/2}+1}{(Y_\zeta+n^{1-1/\zeta}E[X])^2} & \zeta>1 \end{cases} \tag{II.10}$$

The proof of **Proposition 6** and **Proposition 7** is derived by (I.11) and (I.12) with assuming independence among ϵ_i. Given the value of ζ is around 1, the results in **Lemma 5** and **Lemma 4** induces asset pricing implications in **Proposition 6** and **Proposition 7**.

Appendix III Estimation of the Pareto distribution

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 10% firms in each month to form a time-series of ζ_t and plot it in **Figure 10**. 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 10** 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 10 % 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 the relation between $\log\hat{\zeta}_t$ and the logged number of firms $\log n_t$ in the data each month presented in **Figure 11**. I take advantage of the 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 (13) to estimate:

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

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.

Appendix IV Out-of-sample predictive results

I check the out-of-sample predictive power of my model and report results in **Table V.1**. I estimate the single variable case using $\log \zeta_t$,and bi-variate cases adding time-varying idiosyncratic risk and other predictors surveyed in Welch and Goyal (2008) at horizon $k = 1, 12, 60$. For each set of predictors I test, I compute the Out-of-sample R^2 (Oos R^2) by comparing the predictive error of each set of predictors to the historical mean computed by a 240-month rolling window. I also perform the Diebold-Mariano test (DM) to check whether my predictive model outperform the historical mean. The lag number h used for DM tests in different horizon k is computed by the rule of thumb $h = k^{1/3} + 1$. For using $\log \zeta_t$ only, the out-of-sample R^2 reaches 1.50 percent at the 12-month horizon and 13.34 percent (with a significant T-stat 2.07) at the 60-month horizon, which indicates a robust predictive power of ζ in the long period. Combining $\log \zeta_t$ with other predictors also displays out-of-sample predictive power at the long-horizon. I highlight the list of predictors that have positive out-of-sample R^2 at $k = 60$ ahead with a significant DM test T-stat.

Appendix V Additional figures and tables

Figure 10: **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 10 % 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.

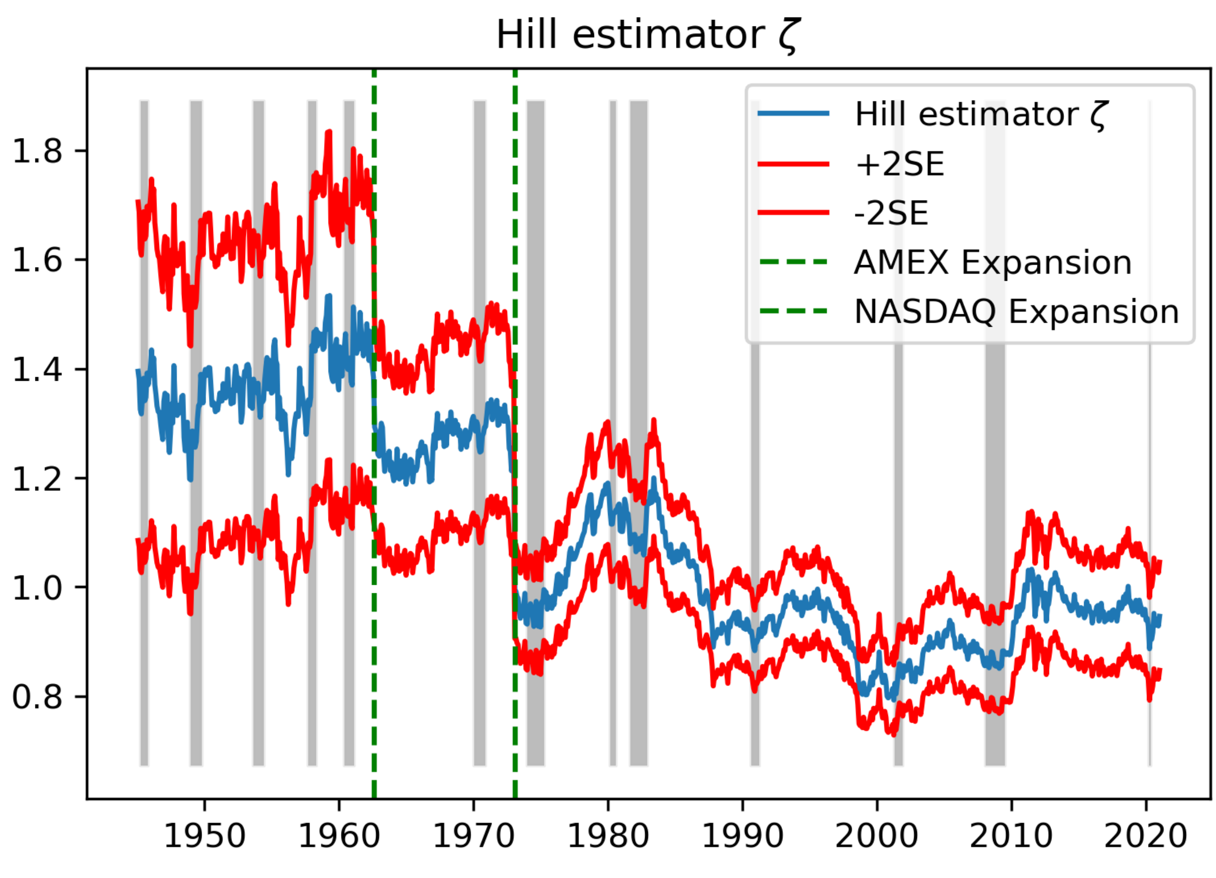


Figure 11: **Pareto Coefficient Estimate of Market Value per Month** I plot the co-integration relation between the logged Pareto coefficient $\log \zeta$ (estimated from the largest 10 % firms) and logged number of firms n. Both the time series are normalized to have mean zero and unit variance with their raw magnitudes displayed on two separate sets of ticks on y axis.

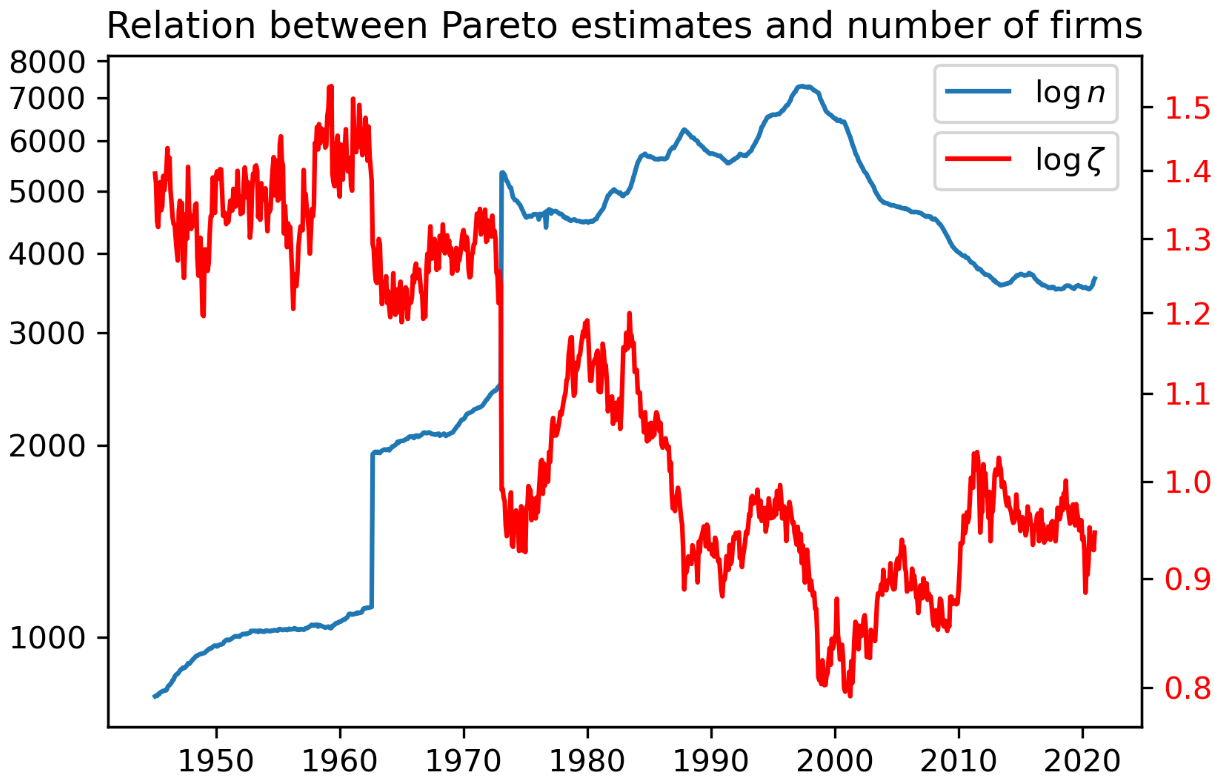


Table V.1: **Out-of-sample prediction results.**
I summarize the out-of-sample predictive power of all the sets of predictor I test in this paper. I compute the Out-of-sample R^2 by comparing the predictive error of each set of predictors to the historical mean computed by a 240-month rolling window. I also perform the Diebold-Mariano test to check whether my predictive model outperform the historical mean.

predictor \ horizon		$\log r_{m,t\to t+1}$	$\log r_{m,t\to t+12}$	$\log r_{m,t\to t+60}$
$\log\zeta$	$OosR^2$	-0.17	1.50	13.34
	DM	-0.17	0.31	2.07
$\log\zeta, \sum w_i\theta_i$(FF3)	$OosR^2$	-0.20	-1.69	0.80
	DM	-0.99	-0.91	0.18
$\log\zeta, \sum w_i\theta_i$(PCA)	$OosR^2$	-2.85	-10.20	4.72
	DM	-1.73	-1.11	0.47
$\log\zeta, \sum w_i\theta_i$(Campbell et al)	$OosR^2$	-2.69	-7.11	3.04
	DM	-1.55	-0.76	0.38
$\log\zeta$, bm	$OosR^2$	-1.47	1.04	**17.60**
	DM	-1.43	0.17	**1.98**
$\log\zeta$, dspr	$OosR^2$	-1.79	-3.27	**15.48**
	DM	-0.63	-0.58	**1.89**
$\log\zeta$, dp	$OosR^2$	0.05	12.92	**40.60**
	DM	0.05	2.19	**4.16**
$\log\zeta$, ep	$OosR^2$	-2.24	-3.48	10.52
	DM	-0.91	-0.48	1.12
$\log\zeta$, ltr	$OosR^2$	0.30	1.81	**13.41**
	DM	0.21	0.38	**2.05**
$\log\zeta$, ntis	$OosR^2$	-1.04	3.17	10.26
	DM	-0.67	0.35	1.40
$\log\zeta$, svar	$OosR^2$	-7.28	-14.22	-15.09
	DM	-1.82	-0.96	-0.63
$\log\zeta$, tspr	$OosR^2$	0.06	6.64	**22.64**
	DM	0.04	0.89	**2.51**
$\log\zeta$, corpr	$OosR^2$	0.66	2.01	**13.44**
	DM	0.43	0.41	**2.06**