Granular Asset Pricing

Junxiong Gao

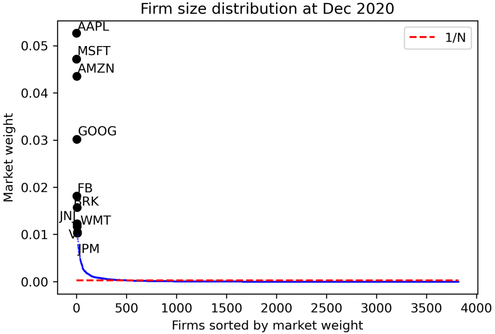
October 20, 2022

Granularity in macroeconomics

- Granularity: fat tail distribution of firm size. Large firms.
- Firm size is measured by fundamentals (value of product, number of employee, etc).
- For example, Nokia in 2000. 1.6 % of Finland's GDP growth.
- Implication: Idiosyncratic shocks of large firms
  - Impact on the aggregate output. (Gabaix (2011),Acemoglu et al. (2012))
  - Help identify price elasticity of aggregate demand/supply. (Gabaix and Koijen (2020)).

Granularity in stock market

- Nokia in 2000: 60% of market cap in the Finnish market index.
- Market weight of US firms in 2020:



The 10 largest firms account for over 25 percents of the total market value (about 4,000 firms).

- Similar finding over time. evidence

Theoretical motivation

- Classical view in asset pricing:
  - Common risk factors in asset returns.
  - Idiosyncratic risks are defined relative to factors.
  - Only common factors explain expected returns. Idiosyncratic risks do not.
- Key assumption: diversification in market.
  - No large firms in a market with sufficiently many assets.
  - Idiosyncratic risks are diversified away.
- **This paper: a granular asset pricing model**
  - Granularity/Fat tail distribution $\Rightarrow$ Large firms. Failure of diversification.
  - Idiosyncratic risk explains expected returns.

Theoretical framework: APT+granularity

- APT: two independent components in shocks of asset returns:

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

Factors $f_{s=1\dots k} \perp$ Idiosyncratic shocks ϵ_i discussion

- Definition of f and ϵ is independent of firm size distribution.
  - Statistical criteria based on covariance among shocks.
  - f (strong correlation) v.s. ϵ_i (weak/no correlation)
- Add regulating condition from firm size distribution perspective.
  - Thin tail. Diversification. Factor model results.
  - With granularity. Variance of ϵ_i explains the expected returns.

Model Implication

- In the cross-section, the impact of ϵ_i depends on firm size.
  - Large firms have their idiosyncratic risk explain expected returns.
  - Small firms do not.
- On aggregate, Idiosyncratic risk of large firms also explains market variation.
- Quantify these implications of granularity by fitting Pareto distribution of firm size
  - Widely used in macroeconomic literature to fit fat tail distribution. Gabaix (1999),Gabaix (2011),Acemoglu, Ozdaglar, and Tahbaz-Salehi (2015), etc.

Related literature

- **Granularity in macroeconomic:** Gabaix (2011), Acemoglu et al. (2012), Acemoglu, Akcigit, and Kerr (2016), Gabaix and Koijen (2020)
- **Asset pricing theory/test**
  - APT models
    Chamberlain and Rothschild (1983), Chamberlain (1983), Dybvig (1983), Connor and Korajczyk (1993), Connor and Korajczyk (1995),Huberman (2005)
  - Test of factors: Feng, Giglio, and Xiu (2020), Kelly, Pruitt, and Su (2020), Giglio, Xiu, and Zhang (2021), Giglio and Xiu (2021),Giglio, Kelly, and Xiu (2022)
  - Idiosyncratic risk and expected returns:
    Ang et al. (2006), Ang et al. (2009), Hou and Loh (2016) Campbell et al. (2001), Xu and Malkiel (2003), Goyal and Santa-Clara (2003) and Herskovic et al. (2016).
- **This paper: granularity's asset pricing implication.**

Key result with granularity

- Expected returns implied by model:

$$E[r_i] - r_f = \text{factor terms}_i + \gamma w_i \theta_i$$

- γ risk aversion coefficient.
- w_i market weight.
- θ_i variance of idiosyncratic shocks ϵ_i.
- $w_i \theta_i$ size-adjusted idiosyncratic risk explains the expected returns.
  - Large firms. High w_i, high impact of idiosyncratic risk.
  - Small firms. Negligible w_i. $w_i \theta_i \to 0$. No impact.

Implication: Granularity and idiosyncratic risk puzzle

- Idiosyncratic risk puzzle in Ang et al. (2006) and Ang et al. (2009) (IRP hereafter). Estimate:

$$E[r_i] - r_f = \text{factor terms} + \eta\sqrt{\theta_i}$$

- $\hat{\eta} < 0$. Negative relation between idiosyncratic risk and return.
    - No satisfying enough explanation in literature (Hou and Loh (2016)).
- This paper:
    - Explain the puzzle. Misspecification if ignoring the size difference among firms.
    - Identify a positive risk-return relation using $w_i\theta_i$. Risk aversion coefficient $\hat{\gamma} > 0$.

Implication: Granularity and aggregate market variation

- Market expected return. $E[r_m] = \sum_i w_i E[r_i]$

$$E[r_m] - r_f = \sum w_i(w_i\theta_i) + \text{factor terms}$$

- Relate to whether idiosyncratic risk matters for market returns in literature.
    - $\theta \uparrow$ overall, more aggregate risk, more expected returns.
    - Time-series implication in Campbell et al. (2001),Goyal and Santa-Clara (2003),Bali et al. (2005).
- This paper:
    - Granularity $\uparrow$. Idiosyncratic risks are less diversified on aggregate.
    - Time-series implication. Controlling for the magnitude of idiosyncratic risk, the level of granularity explains market expected returns.

Expected returns: APT+granularity

- n assets. w_i market weight for each. A representative agent maximizes expected utility based on portfolio return $r_m = \sum w_i r_i$.
- CARA utility for simplicity. Shocks of the pricing kernel:

$$-\gamma \sum w_i \left(\beta_i f + \epsilon_i\right) = -\gamma \sum_i^n w_i \beta_i f - \gamma \sum_i^n w_i \epsilon_i$$

- If firm size is granular in equilibrium, what's the expected returns?

$$E[r_i] - r_f = \underbrace{\sum_{s=1}^{k} \beta_{i,s}\mu_s}_{\text{APT}} + \underbrace{\gamma COV(\epsilon_i, \sum_i^n w_i \epsilon_i)}_{\text{granularity}} \tag{1}$$

μ_s risk premium for factor f_s, γ risk aversion coefficient.

Diversification and factor models

- Diversification assumed by APT models:

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

  - No granularity. All firms have market weight $w_i \to 0$
  - The granularity terms $COV(\epsilon_i, \sum_i^n w_i\epsilon_i)$ converge to zero
- Classical implication. A multi-factor model for expected returns?

$$E[r_i] - r_f = \sum_{s=1}^{k} \beta_{i,s}\mu_s$$

- The diversification $\lim_{n\to\infty} \sum w_i^2$ depends on firm size distribution.

Thin tail size distribution and diversification

- Work on the market value X_i, assume i.i.d for simplicity. $w_i = X_i / \sum_{i=1}^{n} X_i$
- The convergence of $\sum w_i^2$ depends on the first and second moments of X_i:

$$\lim_{n\to\infty} \sum w_i^2 = \lim_{n\to\infty} \sum \frac{(X_i)^2}{(\sum X_i)^2} = \lim_{n\to\infty} \left(\frac{1}{n}\right) \frac{1/n \sum (X_i)^2}{(1/n \sum X_i)^2} \tag{3}$$

- A thin-tail distribution of X_i. Finite mean and variance

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

- $\sum w_i^2$ scales as $1/n$

Granularity and failure of diversification

- Quantify granularity by the Pareto distribution (Gabaix (2011)). Two implications:
  - With granularity. Infinite moments of X_i. Failure of diversification.

$$\lim_{n\to\infty} \sum w_i^2 = \lim_{n\to\infty} \frac{1}{n} \frac{1/n \sum (X_i)^2}{(1/n \sum X_i)^2} \neq 0$$

  - Large firms have non-negligible market weight.

$$\lim_{n\to\infty} w_i \neq 0$$

- Asset pricing result: Idiosyncratic risk explains the expected returns in cross-section and on aggregate.

Quantify the granularity by the Pareto distribution

- Pareto distribution with survival probability:

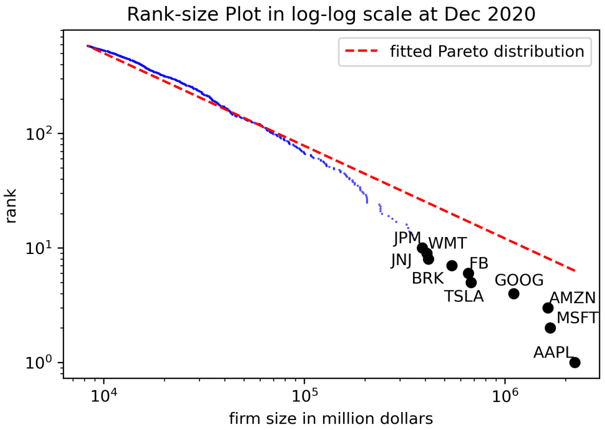
$$P(X_i > x) = \left(\frac{x}{X_m}\right)^{-\xi}, x > X_m \tag{4}$$

  fit for large firms over a threshold x_m
- The Pareto coefficient ξ determines thickness of the tails (granularity). High $\xi \rightarrow$ low level of granularity.
  - ζ is estimated to be around 1 in literature (Zipf's law)
  - $\zeta < 2$, granularity, failure of diversification.
  - $\zeta > 2$, finite first and second moments, diversification.
- A linear relation between i (rank) and X_i (size)

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

Fit of the Pareto distribution: log-log plot

As an example, fit the largest 20% firms at Dec 2020 with a Pareto distribution. Check the linear relation between rank and size.



Estimate of ζ is around 1. Slight concavity.

Pareto distribution and failure of diversification

- Pareto distribution $\zeta < 2$

$$\lim_{n\to\infty} \sum w_i^2 = c_1 \frac{Y_2 + 1}{(Y_1)^2}, \zeta < 1 \tag{5}$$

$$= c_2 n^{2/\zeta - 2} \frac{Y_2 + 1}{E[X]^2}, \zeta > 1 \tag{6}$$

where c_1, c_2 are constants under different range of ζ.

- Y_1, Y_2 are the non-degenerate terms due to infinite moments
- Failure of diversification. For example, let $n = 10^5$ and $\zeta = 1.1$.
  - $\sum w_i^2$ scales as $n^{2/\zeta - 2} = 1/10$
  - Instead of $1/n = 1/10000$ in thin tail case.

Pareto distribution and large firms

- Y_1 comes from the extreme values. Leads to infinite first moment, such that $X_i > a_n$

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

- Stable law: need to adjust for these extreme values to regulate the convergence.
- A firm with extreme size a_n have weights in market portfolio equals to $\frac{a_n}{\sum X_i}$

$$\lim_{n\to\infty} \frac{a_n}{\sum X_i} = 1/(1+Y_1), \xi < 1 \quad (7)$$

$$= n^{1/\zeta - 1}(1 - 1/\zeta), \xi > 1 \quad (8)$$

- Same example. $n = 10^5$ and $\zeta = 1.1$.
  - The market weight of a large firm scales as a_n is roughly 0.03.

Asset pricing implications

- For a firm large enough $\lim_{n\to\infty} w_i \neq 0$. Assume ϵ_i independent simplifies:

$$\lim_{n\to\infty} COV(\epsilon_i, \sum_i^n w_i\epsilon_i) \underset{\epsilon_i independent}{\approx} \lim_{n\to\infty} w_i\theta_i \neq 0$$

- θ_i variance of idiosyncratic shocks. w_i market weight
- A "granular alpha", abnormal return relative to factors:

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

- On aggregate, if diversification fails $\lim_{n\to\infty} \sum w_i^2 \neq 0$ $\sum w_i(w_i\theta_i)$ matters for market expected return.

5 portfolios sorted by idiosyncratic risk

- Replicate Ang et al. (2006) to form five portfolios sorted by θ
    - FF 3 factors as benchmark. Measure θ_i by daily returns in each month.
    - Compute w_i total market weight of assets in each portfolio.

Panel A: alpha relative to FF3						
	L	2	3	4	H	L-H
$\sqrt{\theta}_{FF3}$	2.82	3.97	5.80	8.96	13.85	
w_i	0.60	0.23	0.11	0.05	0.02	
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
$\alpha_{FF3}/\theta_{FF3}$	14.85	-1.29	-1.31	-6.60	-5.95	

From the lowest θ to highest:

- Decreasing size w_i, decreasing α_i (as found in literature)
- Model implies $\alpha_i = \gamma w_i \theta_i$
- Decreasing α_i/θ_i as $w_i \downarrow$ since $\alpha_i/\theta_i = \gamma w_i$

5 portfolios sorted by idiosyncratic risk (Cont'd)

Similar pattern. Measure α relative to CAPM/PCA factors.

Panel A: alpha relative to CAPM						
	L	2	3	4	H	L-H
$\sqrt{\theta}_{CAPM}$	3.67	4.00	7.16	12.29	18.38	
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
$\alpha_{CAPM}/\theta_{CAPM}$	9.94	-0.14	-0.97	-3.57	-3.13	
Panel C: alpha relative to three PCA factors						
$\sqrt{\theta}_{PC}$	12.83	15.28	17.24	19.31	20.11	
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
α_{PC}/θ_{PC}	3.59	2.27	1.70	0.03	-1.45	

Evidence in 100 portfolios sorted

- Cross-sectional test of the relation between α, θ, w
  - Requires a big enough cross-section for regression.
  - Same measure and portfolio construction. Extend to 100 portfolios.
  - Estimate α_i, θ_i of the 100 portfolios. w_i total market weight of assets in each portfolio.
- Same as in IRP.

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

$\hat{\eta} = -0.74$, $\text{std}(\hat{\eta}) = 0.039$.

- Compare to the granular channel of alpha

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

$\hat{\gamma} = 5.17$, $\text{std}(\hat{\gamma}) = 0.59$.

A granular explanation for IRP

- If the data generating process is as model implied.

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

- The negative relation between risk $\sqrt{\theta_i}$ and return α_i is because of:

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

  driven by
  - Negative relation between size and risk. (feature in data) $corr(\sqrt{\theta_i}, w_i) = -0.68$
  - Granularity $\Rightarrow$ a few large firms. Other firms have negligible w_i.
- High θ_i firms. Low w_i. Low $w_i\theta_i$. Low α_i
- Low θ_i firms. High w_i. High $w_i\theta_i$. High α_i.

Individual asset level test

- Replicate Ang et al. (2009). $\hat{\eta} < 0$ in:

$$r_{i,t} - r_f = controls + \eta \sqrt{\theta_{i,t-1}} + \epsilon_{i,t}$$

- Use the size-adjusted idiosyncratic risk instead, $\hat{\gamma} > 0$

$$r_{i,t} - r_f = controls + \gamma w_{i,t-1} \theta_{i,t} + \epsilon_{i,t}$$

- As in Ang et al. (2009). Fama-Macbeth approach.
  - Estimate the Fama French 3 factors model using daily returns to measure $\theta_{i,t}$ and factor exposures in each month.
  - Cross-sectional regression in each month. Take the average estimates as $\hat{\eta}$ and $\hat{\gamma}$.

Results of individual asset level test

Cross-sectional Regression, Stock Level					
	$r_{i,t}$ controls	$r_{i,t}$ controls	$r_{i,t}$ controls	$r_{i,t}$ controls	$r_{i,t}$ controls
$\sqrt{\hat{\theta}_{i,t-1}}$	**-0.01**	**-0.01**			**-0.01**
	-1.98	**-2.10**			**-2.16**
$w_{i,t-1}$		-0.10	-0.11	-1.86	-1.78
		-0.63	-0.59	-5.05	-5.26
$w_{i,t-1}\hat{\theta}_{i,t}$				**9.15**	**8.95**
				8.99	**8.87**

- Same controls in each column: factor exposures to FF3, lagged book to market value ratio, size, momentum (past six month returns) as in Ang et al. (2009).
- In column 1, $\hat{\eta} < 0$. Simply controlling for size w_i does not change the sign in columns 2.
- Functional form is important. Use size-adjusted risk gives $\hat{\gamma} > 0$.

Robustness checks for cross-sectional results

- Benchmark results. Portfolio level and individual asset level.
  - Take FF3 as factor model.
  - Measure θ using daily returns in the past month.
- Robust for longer measurement window, using daily returns in the past 3,6,12 months.
- Robust for using PCA factor models.
- The tests relies on cross-sectional difference of θ among firms. Should be insensitive to factor model selection. For example:
  - Firm1 $\theta_1 >$ Firm2 θ_2
  - Different shocks/residuals from fitting different factor models. (omitted factors in Giglio and Xiu (2021), weakly tested factors in Giglio, Xiu, and Zhang (2021), etc.)
  - But similar θ. $\theta_1 > \theta_2$ for various factor models.

Time-series tests

- The market expected return $E[r_m] = \sum w_i E[r_i]$

$$\lim_{n\to\infty} E[r_m] - rf = \text{factor terms} + \gamma \lim_{n\to\infty} \sum w_i^2 \theta_i$$

- Time-series implication.
  - Two components matter for market expected returns.

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

  - Control the time-variation of idiosyncratic risk. Does level of granularity matter?
  - Pareto coefficient measures the level of granularity

$$\zeta \downarrow \Rightarrow \lim_{n\to\infty} \sum w_i^2 \uparrow \Rightarrow \lim_{n\to\infty} E[r_m] - rf \uparrow ?$$

- A time-series test:

$$\log r_{m,t+1} = controls + A \log \zeta_t$$

$H_0 : A = 0,\ H_1 : A < 0$

Estimate of the Pareto coefficient ζ

- Hill estimator (see Hill (1975)).Sort all the n firm sizes in a descending order. Use the largest k firms $X_{1,2\ldots,k}$ to estimate:

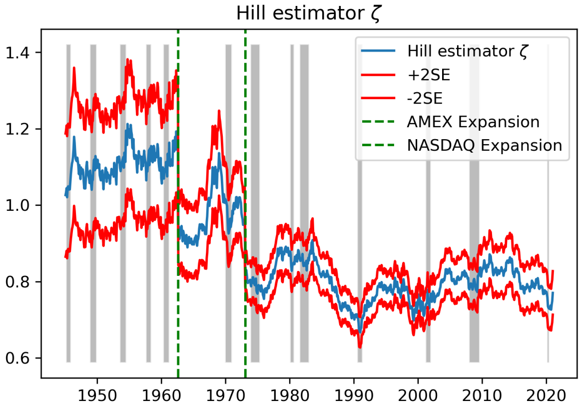
$$\zeta = \left\{ 1/k \sum_{i=1}^{k} (\log X_i - \log X_k) \right\}^{-1} \tag{9}$$

- Maximum likelihood estimator of ζ conditioning on a known minimum threshold X_k (simple inference)
- Pick cutoff k proportional to number of total assets n.

$$k/n = 1\%, 5\%, 10\%, 20\% \ldots$$

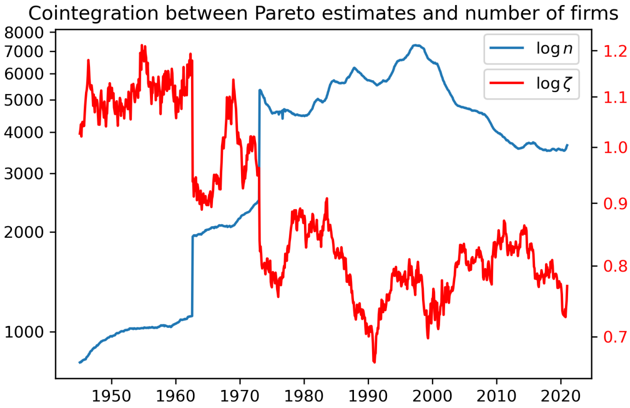
- Select k/n faces a trade-off between bias and variance. More observations. Less variance. Yet, more bias (more small firms included). discussion

Example: Hill estimator use the largest 20 % firms



- Dataset expands in June 1962 (AMEX) and December 1973 (NASDAQ). Over time, n increases such that:
  - Variance $\downarrow$. the estimation error decreases.
  - Bias $\uparrow$. downward bias increases due to more small firms.

De-bias by co-integration



For time-series test purpose:

- Form a "de-biased" predictor by subtracting the non-stationary trend due to increasing n_t
- Select optimal cutoff ratio k/n for best out-of-sample results (10-fold cross-validation)

Single variable prediction results

$$r_{m,t+1} - r_f = constant + A \log \zeta_t (debias)$$

$H_0 : A = 0,\ H_1 : A < 0$

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

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

- **Panel A:** Significance at all predictive horizons. Positive out-of-sample R^2 at the long horizon (12, 60 months).
- **Panel B:** More predictive power during bad times.

Results controlling for idiosyncratic risk

- Three measures using Fama French 3 factors, PCA and Campbell et al. (2001). plot

Panel A: control $\sum w_i\theta_i$(FF3)			
	log $r_{m,t\to t+1}$	log $r_{m,t\to t+12}$	log $r_{m,t\to t+60}$
log ζ_t	**-0.27**	**-1.70**	**-6.17**
T-stat	**-1.91**	**-1.31**	**-1.91**
$\sum w_i\theta_i$(FF3)	-0.20	-1.69	0.80
T-stat	-0.99	-0.91	0.18
$R^2(\%)$	0.48	1.79	3.34
$Out-of-sampleR^2(\%)$	-2.45	-8.12	5.85
Panel B: control $\sum w_i\theta_i$(PCA)			
	log $r_{m,t\to t+1}$	log $r_{m,t\to t+12}$	log $r_{m,t\to t+60}$
log ζ_t	**-0.26**	**-1.64**	**-5.83**
T-stat	**-1.81**	**-1.23**	**-1.78**
$\sum w_i\theta_i$(PCA)	-0.10	-1.11	1.53
T-stat	-0.47	-0.56	0.33
$R^2(\%)$	0.33	1.12	3.47
$Out-of-sampleR^2(\%)$	-2.85	-10.20	4.72
Panel C: control $\sum w_i\theta_i$(Campbell et al)			
	log $r_{m,t\to t+1}$	log $r_{m,t\to t+12}$	log $r_{m,t\to t+60}$
log ζ_t	**-0.28**	**-1.78**	**-6.52**
T-stat	**-1.94**	**-1.38**	**-2.06**
$\sum w_i\theta_i$ (Campbell et al)	-0.23	-2.18	-0.21
T-stat	-1.11	-1.20	-0.05
$R^2(\%)$	0.55	2.55	3.29
$Out-of-sampleR^2(\%)$	-2.69	-7.11	3.04

Results controlling for other predictors

Panel A: Predictors Controlled, 1 Month Horizon									
	bm	dspr	dp	ep	ltr	ntis	svar	tspr	corpr
log ζ_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 ζ_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 ζ_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

Stable coefficient controlling for other predictors. predictor defintion

Conclusion

- APT+granularity framework
- Treat granularity as a feature of DGP, study the asset pricing implication.
- Granularity implies
  - Failure of diversification
  - Large firms in cross-section
- Asset pricing implications:
  - Size-adjusted idiosyncratic risk explains the expected returns in cross-section.
  - Granularity explains a puzzling risk-return relation in the cross-section.
  - Level of granularity explains the time-variation of market expected returns.

Future Research

- Future research question: reasons for an under-diversified stock market.
- An asset pricing model to generate endogenous granularity
- Endogenous granularity in firm size: Champernowne (1953), Wold and Whittle (1957), Gabaix (1999), Beare and Toda (2022)
- Combine with asset pricing features:
  - negative relation between size and variance
  - factor structure in covariance.
  - joint process of size and return

Thanks!

Acemoglu, Daron, Ufuk Akcigit, and William Kerr, 2016, Networks and the macroeconomy: An empirical exploration, *NBER Macroeconomics Annual* 30, 273–335.

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.

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.

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.

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, 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.

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.

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 Ralph SJ Koijen, 2020, Granular instrumental variables, Technical report, National

Summary of the 10 largest firms, 40s-70s

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

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

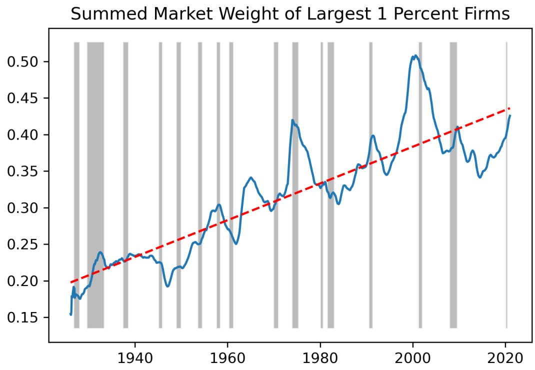
Summary of the 10 largest firms, 80s-2020

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

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

- Existence of large firms over decades.
- In each decade, commonality of large firms
- Can be merged as one larger firm. Granularity is not overestimated.

Granularity increases over time



- Fat tail persists over time.
- Measure the market weight of the largest 1% firms.
- **Implication:** The market portfolio is more "concentrated" over time intro

Granularity and other factor models

- For example, CAPM states a relation between expected returns by assuming multiple conditions

$$E[r_i] - r_f = \beta_i E[r_m - r_f]$$

- CAPM+APT risk structure: $w_i\theta_i$ in β_i
- APT assumes $f \perp \epsilon$. Simplicity for empirical test.
- Econometric issue: bias in measuring factor exposure β_i. OLS:

$$r_i - rf = \beta_i\,(r_m - rf) + \varepsilon_i$$

  $r_m - r_f$ includes $\sum w_i\varepsilon_i$. Endogenous when w_i is non-negligible.
- More complicated to identify shocks propagation in an equilibrium. Loop feedback effect: $\varepsilon_i \to r_m \to \varepsilon_i$
- Different instrumental variables to use depending on what equilibrium mechanism to identify.

main

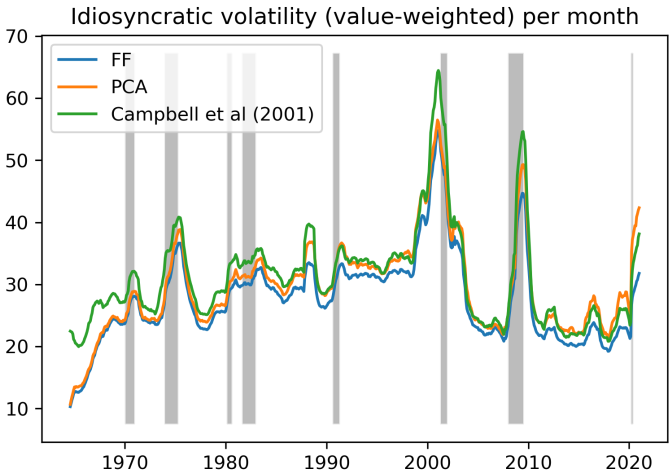
Bias v.s. Variance in the ζ estimation

- A trade-off between bias and variance. More observations. Less variance. Yet, more concave deviations and more bias (more small firms included).
- Vast literature to correct the bias by assuming a distribution to adjust for the concavity.
- For example, Feuerverger and Hall (1999) assumes a survival function as:

$$P(X_i > x) = ax^{-\zeta}(1 + bx^{-\rho} + o(bx^{-\rho}))$$

- Derive the bias by the distribution assumed.
- Generally, the bias depends on number of observations n and level of the concavity. estimate

Time-variation of idiosyncratic risk



⏩ table

Definition of predictors controlled

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

results