Title
Essays in Asset Pricing

Author
Gao, Junxiong

Publication Date
2023

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Essays in Asset Pricing

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Management

by

Junxiong Gao

Committee in charge:

Professor Rossen Valkanov, Chair
Professor Jun Liu, Co-Chair
Professor James Hamilton
Professor Allan Timmermann

2023

The Dissertation of Junxiong Gao is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2023

DEDICATION

I dedicate this dissertation to my family, friends, and academic advisors for their support.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I am indebted to my advisors Rossen Valkanov and Jun Liu for their unlimited help throughout my Ph.D. journey. I would like to express my deepest appreciation to Yan Xu for his support and advice and also as a good co-author. I am grateful to my committee members, Allan Timmermann, and James Hamilton for their valuable insights on my reasearch. I also wish to thank Rady faculty members for their helpful suggestions on both research and teaching, especially Snehal Banerjee, William Mullins, Michael Reher, Bradyn Breon-Drish, and Richard Townsend. Last but definitely not least, my family and friends are always there for me along the way.

Chapter 1 is currently being prepared for submission of publication of the material. It is solely authored by the dissertation author.

Chapter 2 is currently being prepared for submission of publication of the material, and is coauthored with Rossen Valkanov and Yan Xu. The dissertation author is a primary investigator of this material.

Chapter 3 is currently being prepared for submission of publication of the material. It is solely authored by the dissertation author.

VITA

2015	B.S. in Statistics, University of Science and Technology of China, China
2016	Master of Finance, University of California San Diego
2023	Ph.D. in Management, University of California San Diego

ABSTRACT OF THE DISSERTATION

Essays in Asset Pricing

by

Junxiong Gao

Doctor of Philosophy in Management

University of California San Diego, 2023

Professor Rossen Valkanov, Chair
Professor Jun Liu, Co-Chair

This dissertation comprises three papers examining questions in asset pricing, investigating the implications of new asset pricing theories on the cross-section and time series of asset prices. The papers are as follows:

Chapter 1 studies how the fat-tailed distribution of US firm size generates extra risk premiums compared to the classical theory. The author refers to this fat tail as ″granularity″ and shows that it breaks the diversification of idiosyncratic risks assumed by arbitrage pricing theory (APT) to imply factor models. In the cross-section, large firms have higher idiosyncratic risk premiums than small firms despite having a lower

level of risk. This finding explains the negative relation between idiosyncratic risk and risk premium, known as the ”idiosyncratic risk premium puzzle.” On aggregate, the level of granularity, measured by the Pareto distribution, explains market expected returns since it determines the under-diversification of idiosyncratic risk.

Chapter 2 (joint work with Rossen Valkanov and Yan Xu) investigates the joint dynamics and predictability of asset returns for the equity, treasury, and foreign asset investment sectors, utilizing their respective valuation ratios constructed from their intertemporal budget constraints. We propose a new framework that enforces an aggregate accounting identity of the three sectors using a constrained estimation by the GMM method, which accounts for the cyclical movement of the whole economy. Our key finding shows that the government surplus-to-debt ratio negatively predicts the risk premium in the equity and foreign asset investment sectors. Our results suggest that incorporating data from all three sectors and imposing aggregate budget constraints can help to better identify how the fiscal policy adjustment channel propagates throughout the economy.

Chapter 3 presents a model for modeling the correlation dynamics of stock returns using a conditional factor model. In this model, the employment of factors helps to reduce the estimation dimension by presenting the asset returns' covariance matrix as a quadratic function of the conditional covariance with factors. The factor structure allows for a closed-form solution for the inverse and determinant of the covariance matrix, which is convenient for computing the likelihood function and allocating a minimum variance portfolio. The model accurately fits the realized correlation among S&P 500 stocks computed from 5-minute data. It also generates out-of-sample minimum variance portfolios with a higher information ratio.

Chapter 1

Granular Asset Pricing

1.1 Introduction

According to the arbitrage pricing theory (APT), only a few common factors in asset returns are tied to risk premiums, while idiosyncratic risks, as the "residuals" relative to these factors, are diversified away. This classical view is widely accepted in asset pricing theory as it implies a tested factor structure in expected returns. However, this may not hold true as the diversification of idiosyncratic risks does not always occur in practice. The diversification assumption requires a thin-tailed distribution of firm size, meaning that no firm is large enough for its firm-specific shocks to have a systematic impact and be tied to the risk premium.

Contrary to the assumption of diversification, I have found evidence of firms that are significantly larger than others in the US stock market, with high weights in the market portfolio and a fat-tailed distribution of stock market values. I refer to this phenomenon as the stock market granularity and theoretically show it breaks the diversification of idiosyncratic risks in the market portfolio.[1] Furthermore, large firms have their idiosyncratic risks less diversified than small firms and have more risk premiums tied to idiosyncratic risks.

The evidence of stock market granularity is striking and persistent over time. In 2020, the ten largest firms accounted for over a quarter of the total US stock market value as shown in **Figure 1.1**. In addition, I listed the ten largest firms over decades from the 1940s to the 2010s in **Table 1.1** to show a level of granularity similar to **Figure 1.1** over time.[2] Although the list of these large firms varies as production technology evolves, they constantly have dominantly large market weights and break

[1]The stock market granularity 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.) For my paper, measuring granularity in the stock market is natural since it shows how firm-specific shocks can have systematic impacts by generating fluctuations in the market portfolio.

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

the diversification of idiosyncratic risks. Therefore, I develop a theoretical framework to study how the granular channel of under-diversified idiosyncratic risks affects asset prices.

The first contribution of this paper is to demonstrate that granularity breaks the diversification of idiosyncratic risks assumed in APT theory and generates an idiosyncratic risk premium in expected returns in addition to the factor risk premiums. The intuition behind this result is based on the classical view in APT theory that there are two types of risks in asset returns: factors, which are the common components of asset returns that drive the strong correlation among assets, and idiosyncratic risks, which are firm-specific and have a weak correlation. I incorporate granularity into this risk structure and use a competitive equilibrium approach (see Dybvig (1983), Grinblatt and Titman (1983), Connor and Korajczyk (1995)), where a representative agent holds the market portfolio. With granularity, both these two types of risks are tied to the risk premium since they all affect the wealth fluctuation of the representative investor, but they have different economic meanings and empirical patterns.

The second contribution of this paper is to provide a novel and simple-to-test relation between idiosyncratic risk and expected returns in the cross-section. The size-adjusted idiosyncratic risk (product of an asset's market weight and variance of idiosyncratic shock) positively explains the expected returns, with various factors and characteristics controlled. With granularity, large firms have market weights significantly higher than small firms and therefore have more idiosyncratic risk premiums. This granular channel of risk compensation in expected return is ignored by the factor models that assume diversification of idiosyncratic risks and is empirically different from a factor risk premium, which is proportional to the factor risk exposure ("beta"). Specifically, a factor driven by size states the opposite of my results, such that small firms have high factor risk premiums due to high exposure to the factor risk.

Furthermore, my result of using the product of market weight and variance of idiosyncratic shock (Ivar hereafter) reconciles tests in the literature that use Ivar only to explain how idiosyncratic risks affect asset returns. As a leading example, it explains the ″idiosyncratic risk premium puzzle″ (IRP hereafter) that high Ivar firms have low risk-compensation in expected returns in the cross-section, investigated in Ang et al. (2006) and Ang et al. (2009).[3] Small firms have high levels of Ivar due to an inverse relation between the firm size and the level of risk. When the granularity is significant, the size difference among firms is substantial such that large firms account for most of the market valuation, as shown in **Figure 1.1** and small firms have negligible market weights. Consequently, firms with high idiosyncratic risks tend to have negligible idiosyncratic risk premiums due to low impacts on the market. Conversely, firms with low idiosyncratic risks are large firms with high idiosyncratic risk premiums due to high market weights. I find that the granular explanation of IRP is robust to measuring Ivar by various factor models and works within groups of firms separated by size.

The third contribution of my analysis is to test the aggregate impact of granularity on market returns. Tests in literature (see Campbell et al. (2001), Goyal and Santa-Clara (2003), Bali et al. (2005)) measure the aggregate level of idiosyncratic risk and use it to explain the aggregate variation of the stock market using a time-series approach. As a separate channel, with the level of idiosyncratic risks controlled, a high level of granularity implies less diversified idiosyncratic risk and hence should increase the aggregate expected returns of the market portfolio. I measure the level of granularity by a Pareto distribution and find this measure explains the time variation of market risk premium, especially in longer time horizons, controlling for the measures of idiosyncratic risks in the cited papers and additional predictors surveyed in Welch and Goyal (2008).

Specifically, I fit the fat-tailed distribution of firms' market values with the

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

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 my time-series tests, ζ measures the level of granularity and determines the magnitude of idiosyncratic risks under-diversified to affect expected returns. 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, or "alpha" 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.

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 et al. (2020), Kelly et al. (2020), Giglio et al. (2021) Giglio and Xiu (2021),Giglio et al. (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 et al. (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.

1.2 Granular APT

My theoretical framework is a granular APT model, which is a combination of using APT risk structure[4] to define idiosyncratic and factor risks and granularity in the market portfolio quantified by a Pareto distribution. The Pareto distribution brings tractability to capture the stylized facts shown in **Figure 1.1** and **Table 1.1**: Large firms have non-negligible weights in the market and hence breaks the diversification of idiosyncratic risks.

I apply a competitive equilibrium approach (see Dybvig (1983), Grinblatt and Titman (1983), Connor and Korajczyk (1995)) to derive how idiosyncratic risks are tied to risk premium. Specifically, a representative investor holds the market portfolio to maximize its utility by allocating the weights in the market. Notably, many other APT papers do not need to specify the preference nor a competitive equilibrium but only need to assume no-arbitrage and a well-diversified market portfolio since their goal is only to derive a factor model of expected returns by showing the "pricing errors" relative to factors is negligible instead of to show an economic origin of the pricing errors. My framework explicitly links the risk premium unexplained by factors to un-diversified idiosyncratic risks, which is a function of an asset's market weight and level of idiosyncratic risks. Therefore, it illustrates how the impact of idiosyncratic risks changes as the market portfolio composition and firm size distribution.

As the benchmark case, I show that a thin-tailed distribution of firm size implies a well-diversified market portfolio. In consequence, an investor who holds the market portfolio is only exposed to factor risks that drive the common co-movement among asset returns, and the impact of idiosyncratic risks is ruled out.

On the other hand, this theoretical framework allows me to study the expected

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

returns in an equilibrium where the distribution of market values is granular, and a representative investor chooses to hold an un-diversified market portfolio. To justify this portfolio allocation, large firms must have high risk premiums tied to their idiosyncratic risks.

I only present the necessary components here and attach the APT derivations in the **Appendix Section A.1**. There are n assets in the market; each asset return is r_i:

$$r_i = E[r_i] + \sum_{s=1}^{k} \beta_{i,s} f_s + \epsilon_i; \tag{1.1}$$

$$E[\epsilon_i|f] = 0, \forall i. \tag{1.2}$$

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 is approximated by

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

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:

$$E[r_i] = \mu_0 + \sum_{s=1}^{k} \beta_{i,s}\mu_s + \gamma COV(\epsilon_i, \sum_i^n w_i \epsilon_i). \tag{1.3}$$

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 (1.3) by its covariance with the idiosyncratic components of the asset's return ϵ_i. Additionally, it also explains the market expected return $E[r_m]$ such that $E[r_m] = E[\sum_i^n w_i r_i]$ and equals to:

$$E[r_m] = \mu_0 + \sum_i^n w_i \left(\sum_{s=1}^{k} \beta_{i,s}\mu_s \right) + \gamma VAR(\sum_i^n w_i \epsilon_i). \tag{1.4}$$

Intuitively, if there are no large firms in the market such that all w_i are close to zero, then the impact of idiosyncratic risks is diversified away due to the weak correlation among ϵ_i such that

$$\lim_{n \to \infty} \sum_{i=1}^{n} w_i \epsilon_i \to 0.$$

In other words, the impact of idiosyncratic risks converges to zero as the number of assets n approach infinity. In practice, a finite but large n is a good proxy of the limiting case and implies a negligible idiosyncratic risk premium in (1.3) and (1.4) when idiosyncratic shocks are diversified away.

APT models illustrate this intuition formally by making the diversification assumption of w_i. In **Section 1.2.1**, I introduce the diversification assumption in APT and link it to the firm size distribution. As a theoretical result, I show that a thin-tailed distribution induces diversification in w_i. On the opposite, I quantify the level of granularity by a Pareto distribution and show it breaks the diversification and makes the idiosyncratic shocks $\sum_{i=1}^{n} w_i \epsilon_i$ priced in terms of risk premium in **Section 1.2.2**. I then discuss the asset pricing implications of my theoretical results to emphasize the importance of granularity in asset pricing tests in **Section 1.2.3**.

1.2.1 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 (1.3) and (1.4). Specifically, the APT models decompose asset returns into factors and idiosyncratic components by the covariance matrix. Let the covariance matrix of ϵ_i be $\Sigma_{\boldsymbol{\epsilon}}$ and $\rho_i(\Sigma_{\boldsymbol{\epsilon}}), i = 1...n$ 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 $n \to \infty$:

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

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

$$\lim_{n\to\infty} \sum w_i^2 = 0. \tag{1.5}$$

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 (1.3), 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 in* (1.1) *such that the covariance matrix among ϵ_i has bounded eigenvalues as $n \to \infty$:*

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

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 A.1**, 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 this 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 into 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 (1.5). 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:

$$\lim_{n\to\infty} \sum w_i^2 = \lim_{n\to\infty} \sum \frac{(X_i)^2}{(\sum X_i)^2} = \lim_{n\to\infty} \frac{1}{n} \frac{1/n \sum (X_i)^2}{(1/n \sum X_i)^2}. \tag{1.6}$$

A thin-tailed 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 (1.5). 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-tailed firm size distribution implies a well-diversified market portfolio in (1.5) and further the linear factor model. With a thin-tailed 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.

1.2.2 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:

$$P(X_i > x) = \left(\frac{x}{x_m}\right)^{-\xi}, x > x_m. \tag{1.7}$$

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 (1.6) holds. Specifically, the i moments of X are:

$$\begin{aligned} E[X^i] &= \infty, \zeta \leq i; \\ &= \frac{\zeta x_m^i}{\zeta - i}, \zeta > i. \end{aligned} \tag{1.8}$$

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 (1.6) 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 **Appendix Section A.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 $\zeta < 2$.

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 $X_{\max} = \max\{X_{1,\dots,n}\}$. 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 dominant 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* (1.7) *and $\zeta <$ 2,, then the maximum value $X_{\max} = \max\{X_{1,\dots,n}\}$ would have its market weight $w_{\max} = X_{\max} / \sum_{i=1}^{n} X_i$ converge to*

$$\lim_{n\to\infty} w_{\max} = X_{\max} / \sum_{i=1}^{n} X_i = \begin{cases} \dfrac{F_\zeta}{Y_\zeta + 1} & \xi < 1 \\ \lim\limits_{n\to\infty} \dfrac{F_\zeta}{Y_\zeta + \log n} & \xi = 1 \\ \lim\limits_{n\to\infty} \dfrac{F_\zeta}{Y_\zeta + n^{1-1/\zeta} E[X]} & \xi > 1 \end{cases} \tag{1.9}$$

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 A.2**. 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 1.2**. 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 1.1**, or even more than 80 percent as in the simulation results shown by **Figure 1.2**. 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 in $w_{\max}$ 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 1.1**. 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 1.1**, 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, generates large market weights of individual assets. This result is consistent with the feature of data and cannot be ignored in the empirical tests. 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 A.2** 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**.

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* (1.7) *and $\zeta < 2$, then the convergence in equation* (1.6) *is determined by ζ as follows.*

$$\lim_{n\to\infty}\sum w_i^2 = \begin{cases} \dfrac{Y_{\zeta/2}}{(Y_\zeta)^2} & \zeta < 1 \\ \lim\limits_{n\to\infty}\dfrac{Y_{\zeta/2}}{(Y_\zeta+\log n)^2} & \zeta = 1 \\ \lim\limits_{n\to\infty}\dfrac{Y_{\zeta/2}}{(Y_\zeta+n^{1-1/\zeta}E[X])^2} & \zeta > 1 \end{cases} \tag{1.10}$$

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 A.2**. 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 1.3**. 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 dashed 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 1.3**, 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.

1.2.3 Asset pricing implications of granularity

I now combine the results from the Pareto distribution with the asset pricing equations in (1.3) and (1.4) to produce testable results for expected returns. As discussed in the last section, my derivations when $n \to \infty$ are also well approximated by the results when n is sufficiently large enough in data. Therefore, I use the limiting case to discuss the associating asset pricing tests.

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 (1.3). 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 1. *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:*

$$\lim_{n\to\infty} E[r_i] = \mu_0 + \sum_{s=1}^{k} \beta_{i,s}\mu_s + \theta_i \gamma \lim_{n\to\infty} w_i. \tag{1.11}$$

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 1** 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 et al. (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 (1.11) 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 1** points out that it should be the size-adjusted idiosyncratic risk $w_i\theta_i$ instead of itself θ_i that explains expected returns. In the limiting case when n approaches infinity, only large firms $\lim_{n\to\infty} w_i \neq 0$ could have their idiosyncratic shocks un-diversified to generate expected returns such that $\lim_{n\to\infty} w_i\theta_i \neq 0$. For a finite n market, the granularity drives big size differences in the cross-section such that large firms have higher idiosyncratic risk premiums than small firms. 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.

Controlling for the factor risk premiums, 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.

More importantly, **Proposition 1** 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 will 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 $\hat{\eta}$ in IRP is proportional to the correlation $corr(w_i\theta_i, \theta_i)$, such that

$$\hat{\eta} \propto corr(w_i\theta_i, \theta_i).$$

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 $\hat{\eta}$ equals:

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

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 1.4**. Comparing this plot to **Figure 1.1** 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 1.4** 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 1.4**.

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,

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

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.

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 (1.5), 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 2**:

Proposition 2. *If the idiosyncratic shocks are independent of each other with variance θ_i, then the expected return for the aggregate market converges to:*

$$\lim_{n\to\infty} E[r_m] = \mu_0 + \sum_i^n w_i \left(\sum_{s=1}^k \beta_{i,s}\mu_s \right) + \gamma \lim_{n\to\infty} \sum w_i^2 \theta_i. \tag{1.12}$$

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 $\sum w_i^2$ is an indicator of the under-diversification such that if it is negligible, then there is no aggregate impact of idiosyncratic risk. The $\left(\sum \frac{w_i^2}{\sum w_i^2} \theta_i \right)$ is a weighted-average of idiosyncratic risk. My derivations use the Pareto distribution to 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 2** 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:

$$\log(r_{m,t+1}) = \text{constant} + \text{controls} + \mathrm{A} \log \mathrm{1_t}. \tag{1.13}$$

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

1.3 Empirical Test

1.3.1 Data

My cross-sectional test is at the monthly frequency from June 1963 to December 2020. I use monthly return and firm size data in the CRSP and other characteristic data in COMPUSTAT for control variables. I merged the monthly CRSP 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 sorted variables are available for constructing portfolios. Fama-French factors 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 the Pareto coefficient, as a measure of the level of granularity, captures the time variation of the market expected returns in this sample period.

1.3.2 Cross Section Test

My result in **Proposition 1** 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, I derive a linear relation between α_i and $w_i\theta_i$ at the firm level. The same linear relationship may not hold perfectly in a portfolio-level test since If we treat a portfolio as an asset, its alpha α_p is simply a linear combination of each asset's alpha but its size-adjusted idiosyncratic risk $w_p\theta_p$ does not equal the linear combination of each asset's. Therefore,

$$\alpha_p \neq \gamma w_p \theta_p.$$

From this perspective, I still use the portfolio level test as a benchmark (compared to Ang et al. (2006)) to illustrate the economic insight of my model and further use the firm level test (compared to Ang et al. (2009))) to justify my model implication.

Portfolio level tests

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

I report results using the five idiosyncratic risk sorting portfolios in **Table 1.2**. 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 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 factor 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:

$$r_{L-H,t} = \frac{1/\theta_{L,t-1}}{1/\theta_{L,t-1} - 1/\theta_{H,t-1}}(r_{L,t} - r_f) - \frac{1/\theta_{H,t-1}}{1/\theta_{L,t-1} - 1/\theta_{H,t-1}}(r_{H,t} - r_f). \tag{1.14}$$

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 $w_L(\text{large size}) - w_H(\text{small size}) > 0$.

As a benchmark, the portfolio constructed by θ measured by the past month 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 patterns in these five portfolios replicate findings in Ang et al. (2006) and verify the insight that large firms have high impacts of idiosyncratic risk on expected returns. A robustness check in the cited paper is to construct sorted portfolios by a longer measurement window of idiosyncratic variance θ, which is a reasonable way to test whether the IRP is sensitive to the time-varying level of idiosyncratic risks.

To ease the concern in this perspective, I apply the same method to construct the long-short portfolio using θ measured by the past 3,6 and 12 months and summarize its performance in **Table 1.3**. The patterns showed by the five sorted portfolios are robust to the longer measurement window of the idiosyncratic variance θ. The long-short portfolios formed by estimation of the past 3,6, and 12 months also generate positive alphas relative to the benchmark models.

The above results replicate findings in Ang et al. (2006) and test my theoretical insight by constructing a long-short portfolio. To further explore the cross-sectional relation between α_i, θ_i, and w_i, I extend the 5-portfolio setting to split all the assets by percentiles of θ to construct 100 value-weighted portfolios. The 100 portfolios are constructed following the same steps and provide a larger cross-section to test my model implications. 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:

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

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. To minimize the total estimation error, the constrained estimation picks a explanatory variable that fits the cross-sectional variation of expected returns with more precision. The estimate suggests that the granular channel of the idiosyncratic risk premium has more explanatory power than θ itself 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 will 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 1.5**. 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 1.6**, 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 $w_i\theta_i$.

Similar to the five-portfolio case in Ang et al. (2006), I also examine the robustness of my 100-portfolio results for different lengths of the measurement window. In **Table 1.4**, 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. 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.

Individual asset level test

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:

$$r_{i,t} = \mu_0 + \sum_{s=1}^{k} \beta_{i,s,t} \left(f_{s,t} + \mu_s \right) + \eta \theta_{i,t-1} + \epsilon_{i,t}. \tag{1.15}$$

To incorporate the time-varying magnitude of risks in asset returns, they test the cross-sectional relation between expected returns and idiosyncratic risk with time-varying parameters and apply a Fama-Macbeth regression using monthly data to estimate $\hat{\eta} < 0$.[5]

To compare to the test in Ang et al. (2009), I generalize (1.11) to be time-varying and estimate:

$$r_{i,t} = \mu_0 + \sum_{s=1}^{k} \beta_{i,s,t} \left(f_{s,t} + \mu_s \right) + \gamma w_{i,t-1} \theta_{i,t} + \epsilon_{i,t}. \tag{1.16}$$

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 (1.16). 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 (1.15) 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

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

test the risk-return relation in the cross-section. Similarly, to emphasize the importance of size adjustment, I estimate a constrained model:

$$r_{i,t} = \mu_0 + \sum_{s=1}^{k} \beta_{i,s,t} \left(f_{s,t} + \mu_s \right) + \lambda w_{i,t-1}\theta_{i,t} + (1-\lambda)\theta_{i,t-1} + \epsilon_{i,t}. \tag{1.17}$$

A large $\hat{\lambda}$ with significance suggests that the size-adjusted idiosyncratic risk explains the cross-sectional variation of expected returns with more precision.

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 1.5**, 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 (1.17) 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 (1.17) 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.

Robustness check for the cross-section tests

The first robustness check is to reconcile my results with the double-sorting tests in the literature, which separate firms into several groups by size and then construct portfolios sorted by idiosyncratic risks using firms within each group. Similarly, I separate firms into three groups by size (the largest 30 percent, the smallest 30 percent, and the left 40 percent in the middle) and apply the firm-level tests as in (1.15), (1.16) and (1.17) to examine how the size-adjusted idiosyncratic risk explains expected returns within each group of firms.

I summarize the results in **Table 1.6**. The η estimates in (1.15) from each group are all significantly negative, which is consistent with the results in the literature that the double-sorting does not resolve the IRP. This finding is not a surprise under the granular explanation of IRP since the significant size difference due to granularity exists in all groups of firms separated by size. In other words, even in the group of firms with the smallest market value, the relative relation implied by my model still holds such that firms with higher market weights have more idiosyncratic risk premiums.

Consequently, the γ estimates using $w_i\theta_i$ in (1.16) are all significantly positive in the three groups, which further verifies the model's cross-sectional implication. In addition, the constrained estimates of λ in (1.17) are all significantly positive. Meanwhile, the magnitudes of $\hat{\gamma}$ in the three groups are quite different since the $w_i\theta_i$ of firms in the smallest group is way smaller than firms in the largest group. This difference in cross-section verifies the argument of my paper: Large firms have market weights significantly higher than small firms and hence have high idiosyncratic risk premiums captured by $\gamma w_i\theta_i$.

The second robustness check is to use other factor models to measure the idiosyncratic variance θ to examine whether my empirical tests are sensitive to the factor model selection. As discussed in my theoretical derivations, measuring the variance of idiosyncratic shocks θ_i should be robust to factor model selection since various models give the same cross-sectional ranking of θ_i among firms. From this perspective, my tests avoid issues in the identification of idiosyncratic shocks from improper factor model selection (see Feng et al. (2020), Giglio et al. (2021) Giglio and Xiu (2021),Giglio et al. (2022)).

Therefore, I extend my benchmark results using FF3 factors with the same firm-level tests but using FF5 factors, PCA factors (the three principal components of all asset returns), and the Q5 factors (see Hou et al. (2015), Hou et al. (2021)) and summarize the results in **Table 1.7**. The portfolio-level results using other factor models show the same pattern and are available upon request. Notably, the estimates of η are still negative but less significant using the FF5 and Q5 factor models, which is consistent with the findings in the literature. However, the size-adjusted idiosyncratic risk always positively explains the expected returns with significance and a similar magnitude of $\hat{\gamma}$ and $\hat{\lambda}$ to the benchmark results.

In summary, my empirical results are robust to factor model selection and tests using firms grouped by size. The driving force of these results is the significant size difference in the cross-section, which makes large firms have higher idiosyncratic risk premiums than small firms, as captured by $\gamma w_i \theta_i$. In practice, the size difference among firms changes over time. Especially, there were lots of small firms in the market from the decade 70s-90s as shown in **Table 1.1**, which could reduce the overall cross-sectional variance of w_i and hence weaken the IRP results and the explanatory power of $w_i \theta_i$ on expected returns.

Therefore, the last robustness check is to examine whether the granular explanation of IRP changes over sub-sample periods. I separate the whole sample by decades to run the same tests and summarize the results in **Table 1.8**. The total number of firms jumped from 2995 in the 1960s to 6718 in the 1970s due to the emergence of the

NASDAQ exchange, which kept increasing in the 80s and 90s. Most of these firms were emerging technology companies and hence reduced the overall size difference among firms. Consequently, the IRP estimates of η are not significantly negative in the 70s-90s. This result is consistent with my theoretical derivation since if all firms have the same size, then IRP would not exist since

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

On the opposite, the size-adjusted idiosyncratic risk $w_i\theta_i$ positively explains the risk premium with significance except in the 90s due to a notably high number of small firms during the "Internet bubble" period. The constrained estimate of λ is always positive and significant, which suggests that large firms constantly exist in the market and have high idiosyncratic risk premiums. The sub-sample results for the 100 portfolios show the same pattern and is available upon request.

1.3.3 Time-Series Test

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. I estimate the tail parameter ζ of the Pareto distribution using the Hill estimator (see Hill (1975)). I introduce the details to estimate a monthly time-series of ζ_t in **Appendix Section A.3** and present the result of ζ_t explaining the time-variation of market risk premium in the following section.

Time-series results

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

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

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 1.9**. 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 1.9**, 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 1.7** and find a very similar magnitude of idiosyncratic risk changing over time. In Panel B, C, and D of **Table 1.9**, 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 1.8**, 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.

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 1.10**, 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 1.11**, 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 A.4**.

1.4 Conclusion

I contribute to the existing asset pricing research by documenting a granular channel of idiosyncratic risk to explain expected returns. I theoretically show that the fat-tailed distribution of firm size breaks the market diversification assumed by APT, making idiosyncratic risk matters for asset prices.

Moreover, my results highlight a novel asset pricing pattern: Low risk level firms do not always have to generate low risk premiums. With granularity, large firms have higher idiosyncratic risk premium than small firms, in spite of having a lower level of idiosyncratic risks. This result is supported when running multiple sets of robustness checks as well. Furthermore, this finding of mine explains the influential "idiosyncratic risk premium puzzle" in Ang et al. (2006) and Ang et al. (2009). 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.

1.5 Acknowledgements

Chapter 1, in full, is currently being prepared for submission for publication of the material. Junxiong Gao, the dissertation author, is the primary investigator and author of this paper.

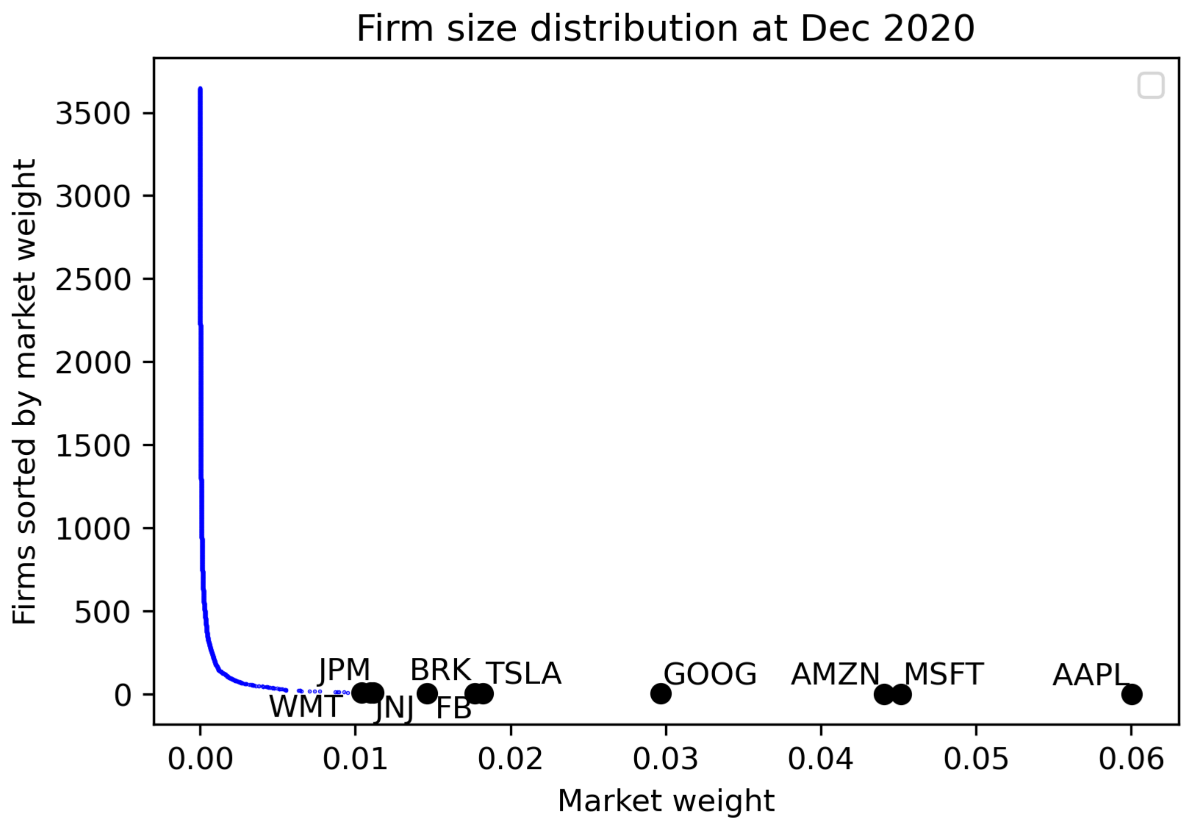
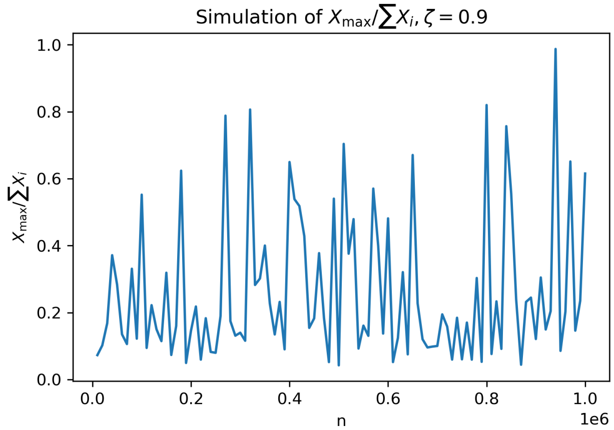
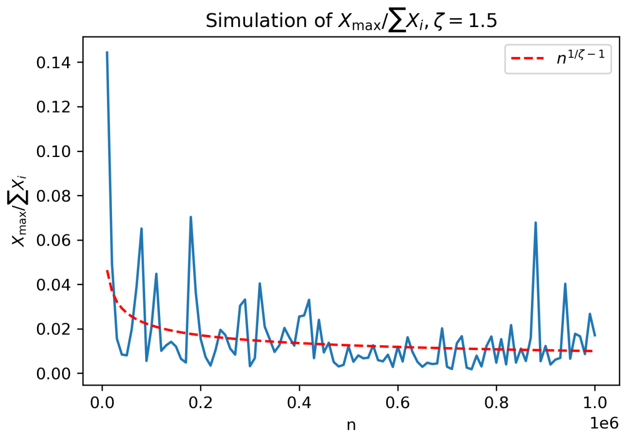


Figure 1.1. Firm Market Weight Sorted at the end of 2020.

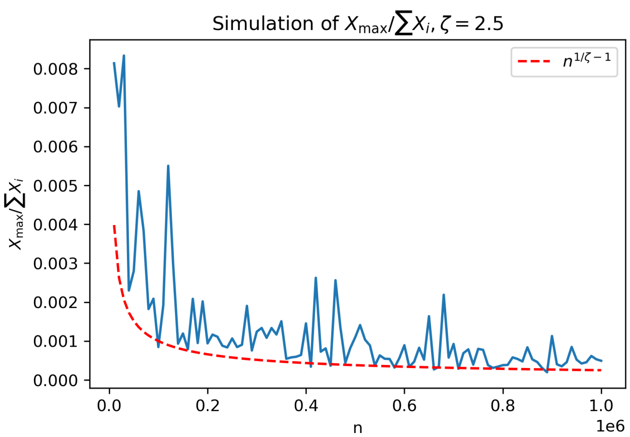
This figure displays the fat right tail of firm size. I measure the firm size by each asset's relative weight in the market portfolio. The 10 largest firms are highlighted and account for over 25 percent of the whole CRSP data in 2020 containing about 4,000 firms.



(a) $\zeta = 0.9$



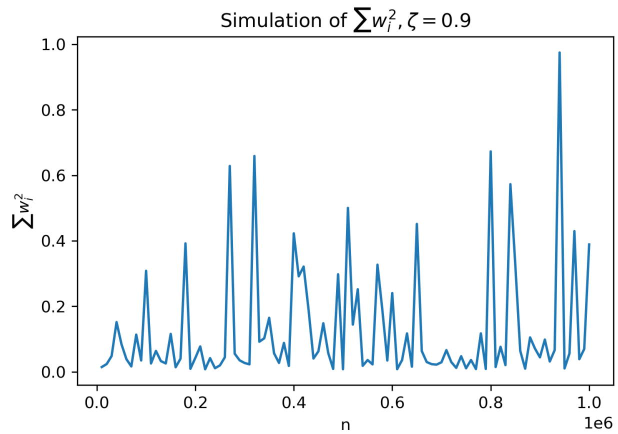
(b) $\zeta = 1.5$



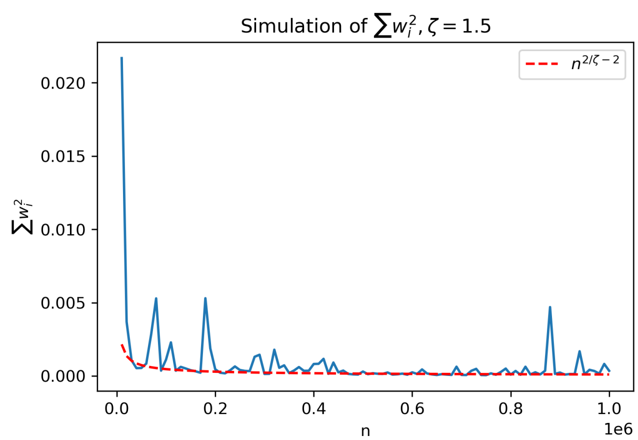
(c) $\zeta = 2.5$

Figure 1.2. Simulation of the largest firm's market weight

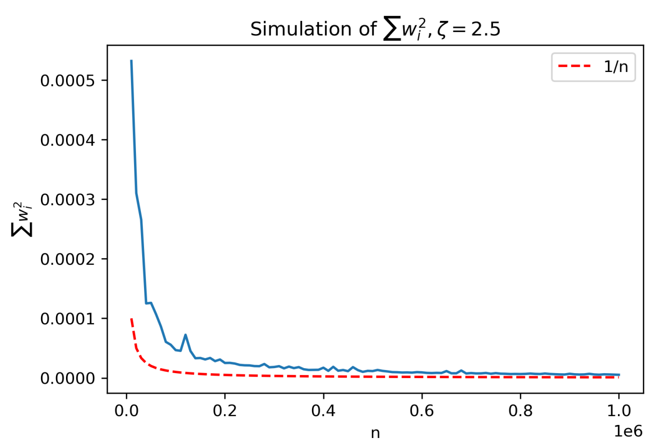
In this figure, I use simulation of Pareto distribution with $\zeta = 0.9, 1.5$ and 2.5 to study how the market weight of the largest firm $w_{\max} = \frac{X_{\max}}{\sum X_i}$ changes as n increases.



(a) $\zeta = 0.9$



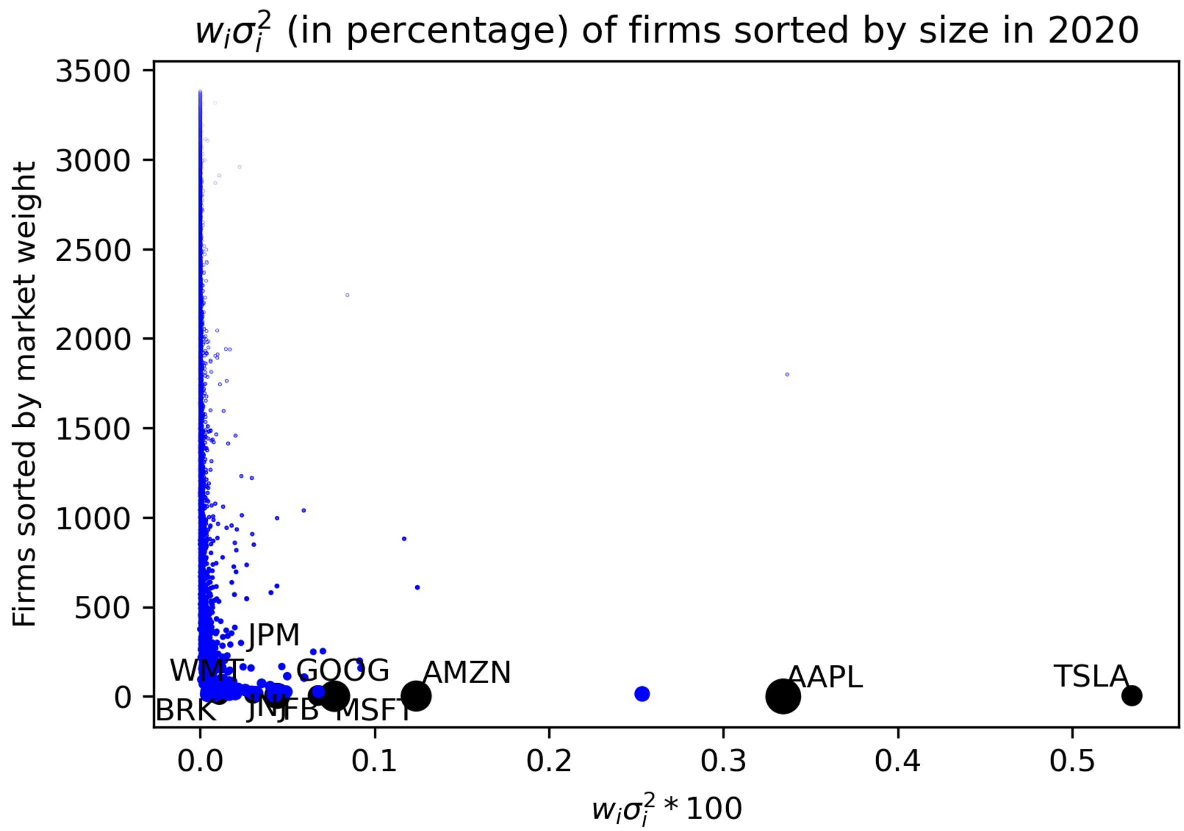
(b) $\zeta = 1.5$



(c) $\zeta = 2.5$

Figure 1.3. 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.



.

Figure 1.4. 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.

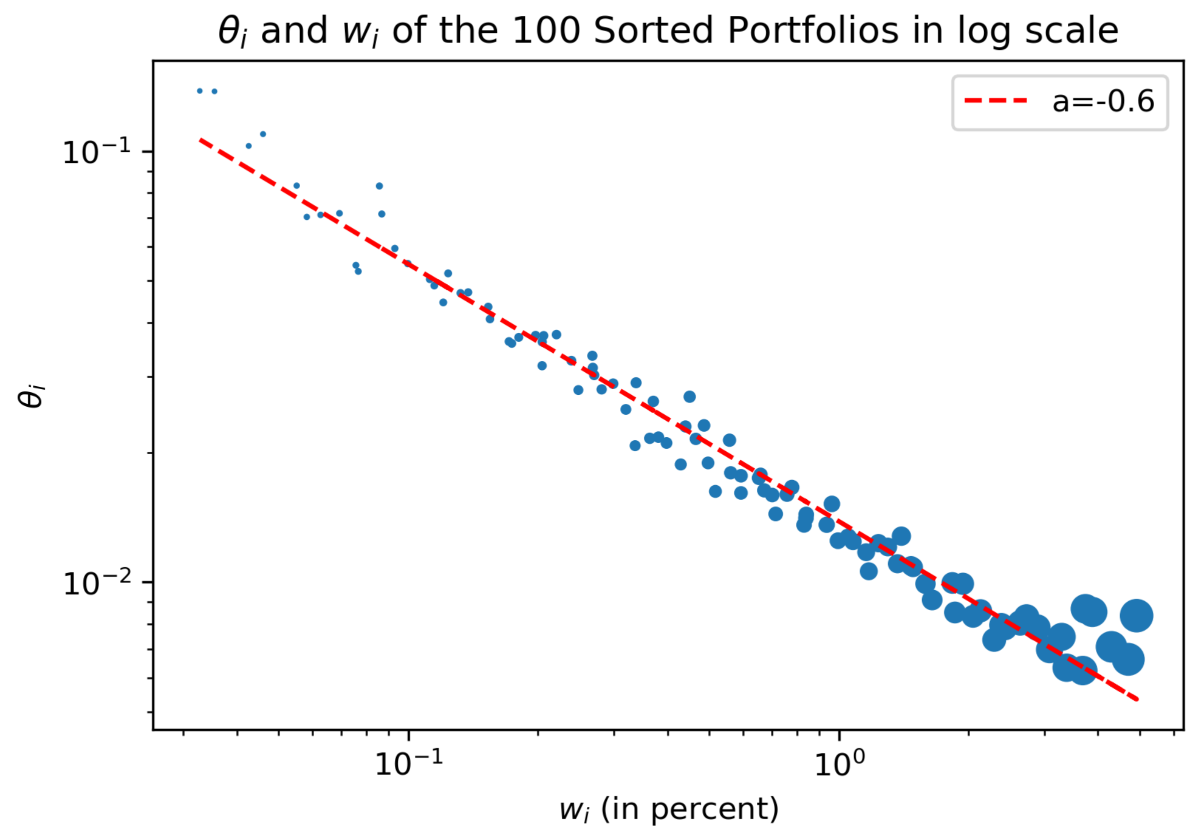


Figure 1.5. Size and idiosyncratic risk of the 100 sorted portfolios in log scale

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.

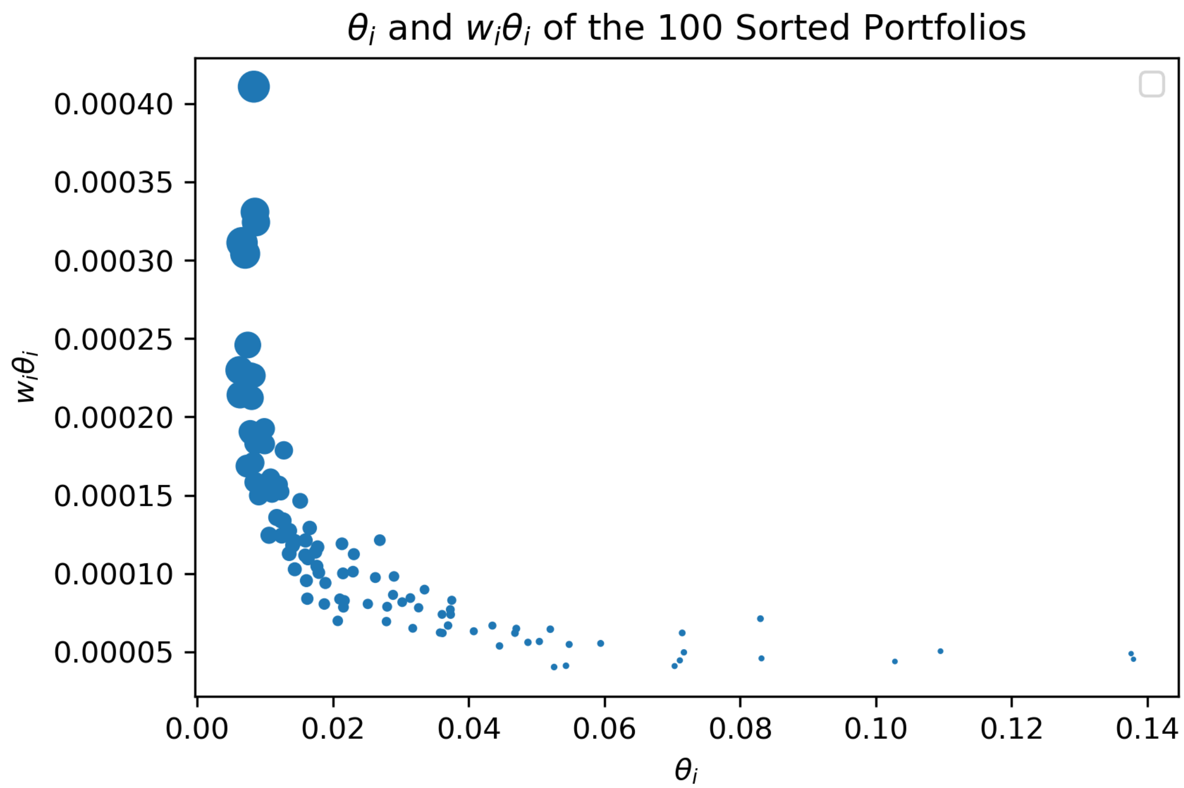
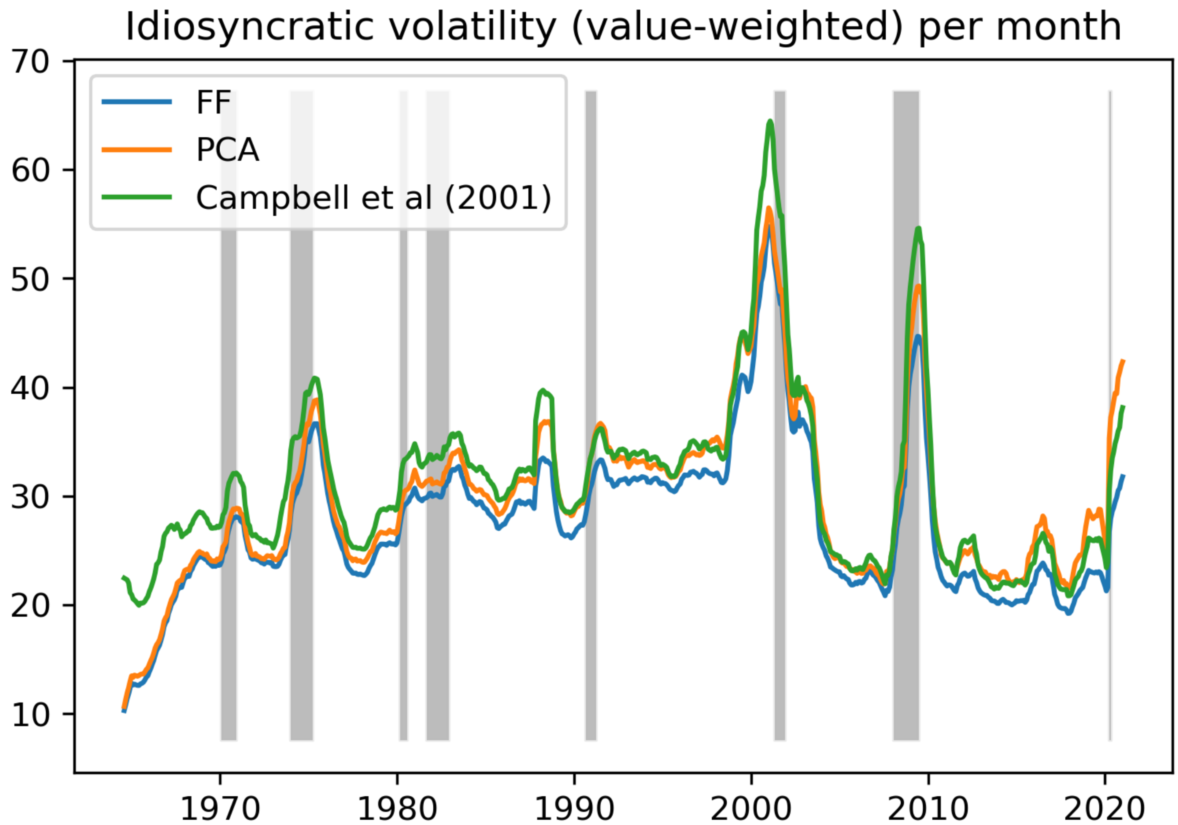


Figure 1.6. Size-adjusted idiosyncratic risk of the 100 sorted portfolios

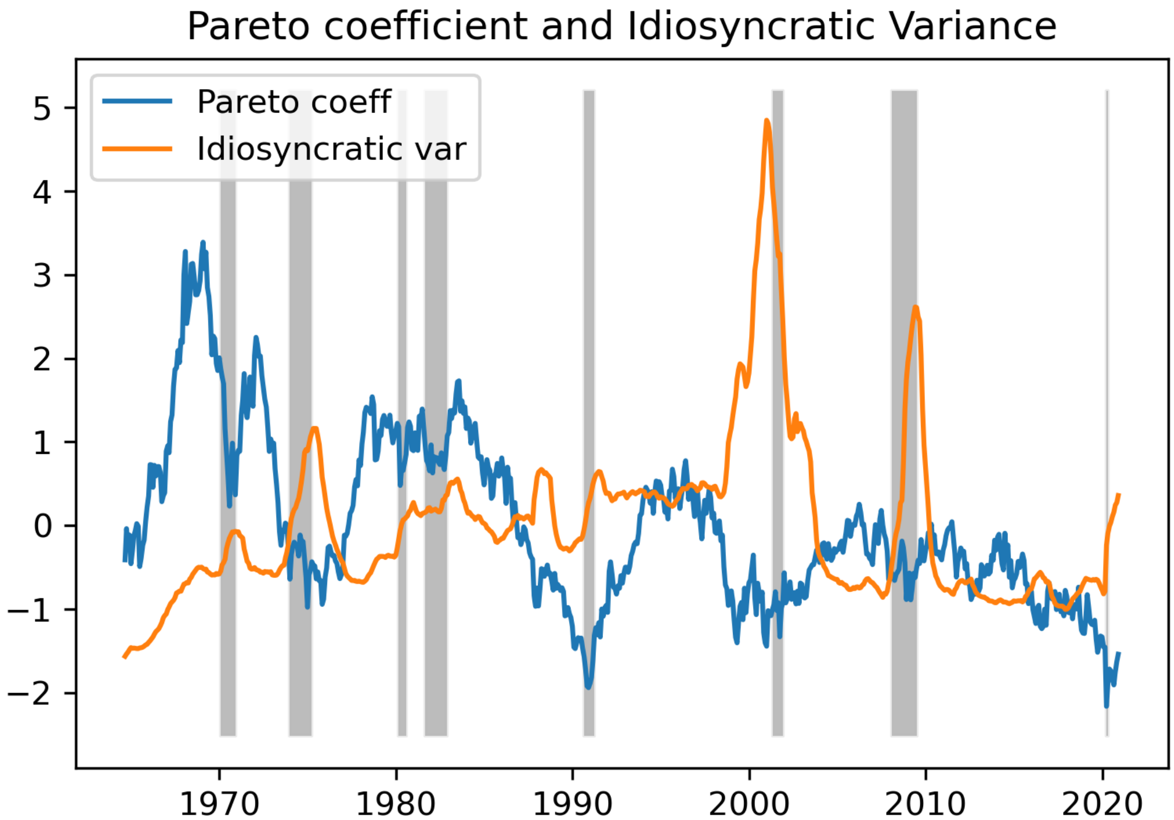
In this figure, I plot the relation between $w_i\theta_i$ and θ_i of the 100 portfolios sorted by θ_i. The dot size is scaled by the total market weight of each portfolio w_i.



.

Figure 1.7. 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 1.8. Pareto predictor 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 this series together with the Pareto predictor. 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 1.1. Evidence of granularity over decades

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)
Weight	**0.24**	**0.26**	**0.24**	**0.19**
Firm number	1019	1215	2995	6718

Panel B:Summary of the 10 largest firms, 1980s-2010s and 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)
Weight	**0.15**	**0.14**	**0.17**	**0.18**	**0.25**
Firm number	10428	12477	9040	6060	3823

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

Table 1.2. Portfolios sorted by idiosyncratic variance estimated by Fama-French 3 factors per month.

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	

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 portfolios 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.

Table 1.3. Performance of "bet on granularity portfolios" under different formation periods.

Bet on granularity portfolios				
window length	1	3	6	12
	r_{L-H}	r_{L-H}	r_{L-H}	r_{L-H}
Mean	7.36	7.29	7.25	7.03
Volatility	13.60	13.67	13.66	13.74
α_{FF3}	1.49	1.48	1.57	1.46
T-stat	3.67	3.62	3.79	3.54
α_{CAPM}	1.64	1.60	1.62	1.47
T-stat	3.04	2.79	2.76	2.45
α_{PC}	6.20	6.25	6.27	6.04
T-stat	3.65	3.62	3.62	3.45

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 using the Fama French three factors. 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 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.

Table 1.4. Cross-sectional results using 100 portfolios sorted by idiosyncratic variance, robustness check for measurement window of idiosyncratic risk

Cross-sectional results using 100 portfolios				
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

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 lengths. I report estimate of

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

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

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

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

I summarize the estimate of η, γ and 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.

Table 1.5. Fama-MacBeth results, individual asset level

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
$\mathrm{b/m}_{i,t-1}$	0.24	0.24	0.24	0.25	0.25	0.25
	8.65	8.60	8.57	8.84	8.90	8.78
$\mathrm{mom}_{i,t-1}$	-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 (η)**			-2.24	0.29 (1-λ)
	-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 (γ)**	8.77	**0.71 (λ)**
				8.99	8.73	19.34

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), $r_{i,t} = \text{constant} + \text{controls} + \sum_{s=1}^{k} \hat{\beta}_{i,s,t}\bar{\lambda}_s + \hat{\eta}_{i,t-1} + \epsilon_{i,t}$, to my model: $r_{i,t} = \text{constant} + \text{controls} + \sum_{s=1}^{k} \hat{\beta}_{i,s,t}\bar{\lambda}_s + \gamma w_{i,t-1}\hat{\theta}_{i,t} + \epsilon_{i,t}$. I estimate $\hat{\eta}$ in the columns 2 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 $\mathrm{b/m}_{i,t-1}$ and the momentum factor $\mathrm{mom}_{i,t-1}$ computed by the sum of returns in the last six months as in Jegadeesh and Titman (1993).

Table 1.6. Fama-MacBeth results, individual asset level, with three size groups.

Panel A: Cross-sectional regression, large-cap firms						
	$r_{i,t}$ controls	$r_{i,t}$ controls	$r_{i,t}$ controls	$r_{i,t}$ controls	$r_{i,t}$ controls	$r_{i,t}$ controls
$\hat{\theta}_{i,t-1}$	-16.47	**-17.35** **(η)**			-18.30	0.26 (1-λ)
	-4.40	-4.56			-4.53	9.51
$w_{i,t-1}$		-0.23	-0.14	-0.09	-0.25	-2.03
		-2.49	-1.39	-0.44	-1.28	-15.91
$w_{i,t-1}\hat{\theta}_{i,t}$				**1.52** **(γ)**	1.75	**0.74** **(λ)**
				2.34	2.74	27.01
Panel B: Cross-sectional regression, middle-cap firms						
	$r_{i,t}$ controls	$r_{i,t}$ controls	$r_{i,t}$ controls	$r_{i,t}$ controls	$r_{i,t}$ controls	$r_{i,t}$ controls
$\hat{\theta}_{i,t-1}$	-10.89	**-11.24** **(η)**			-21.49	-0.39 (1-λ)
.	-5.52	-5.80			-9.46	-6.91
$w_{i,t-1}$		-8.86	30.05	-37.24	-101.43	-107.23
		-0.38	1.24	-1.09	-3.01	-3.59
$w_{i,t-1}\hat{\theta}_{i,t}$				**56.81** **(γ)**	69.73	**1.39** **(λ)**
				7.38	9.01	24.48
Panel C: Cross-sectional regression, small-cap firms						
	$r_{i,t}$ controls	$r_{i,t}$ controls	$r_{i,t}$ controls	$r_{i,t}$ controls	$r_{i,t}$ controls	$r_{i,t}$ controls
$\hat{\theta}_{i,t-1}$	-2.63	**-3.10** **(η)**			-10.29	-1.36 (1-λ)
	-2.51	-3.24			-7.22	-12.65
$w_{i,t-1}$		-3062.46	-3076.17	-4643.75	-4913.00	-5131.17
		-9.18	-8.67	-11.11	-11.93	-11.80
$w_{i,t-1}\hat{\theta}_{i,t}$				**595.42** **(γ)**	619.36	**2.36** **(λ)**
				15.79	16.21	21.97

In this table, I report the same individual asset level test within three groups of firms sorted by size. I report the results of the same tests in **Table 1.5**. 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 $\mathrm{b/m_{i,t-1}}$ and the momentum factor $\mathrm{mom_{i,t-1}}$ computed by the sum of returns in the last six months as in Jegadeesh and Titman (1993).

Table 1.7. Fama-MacBeth results, individual asset level, using other factor models.

Panel A: Cross-sectional regression, controlling for FF5 factors						
	$r_{i,t}$ controls	$r_{i,t}$ controls	$r_{i,t}$ controls	$r_{i,t}$ controls	$r_{i,t}$ controls	$r_{i,t}$ controls
$\hat{\theta}_{i,t-1}$	-1.68	**-1.70** (η)			-1.74	0.30 (1-λ)
	-1.76	-1.83			-1.85	8.88
$w_{i,t-1}$		-0.17	-0.19	-1.55	-1.48	-3.09
		-0.98	-1.05	-4.50	-4.35	-14.44
$w_{i,t-1}\hat{\theta}_{i,t}$				**8.35** (γ)	8.16	**0.70** (λ)
				8.22	7.95	20.87
Panel B: Cross-sectional regression, controlling for PCA factors						
	$r_{i,t}$ controls	$r_{i,t}$ controls	$r_{i,t}$ controls	$r_{i,t}$ controls	$r_{i,t}$ controls	$r_{i,t}$ controls
$\hat{\theta}_{i,t-1}$	-2.22	**-2.23** (η)			-2.25	0.28 (1-λ)
	-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** (γ)	6.52	**0.72** (λ)
				7.97	7.77	20.99
Panel C: Cross-sectional regression, controlling for Q5 factors						
	$r_{i,t}$ controls	$r_{i,t}$ controls	$r_{i,t}$ controls	$r_{i,t}$ controls	$r_{i,t}$ controls	$r_{i,t}$ controls
$\hat{\theta}_{i,t-1}$	-1.17	**-1.19** (η)			-1.19	0.30 (1-λ)
	-1.40	-1.47			-1.46	8.91
$w_{i,t-1}$		-0.19	-0.22	-1.66	-1.57	-3.26
		-1.00	-1.12	-4.57	-4.42	-14.79
$w_{i,t-1}\hat{\theta}_{i,t}$				**8.30** (γ)	8.10	**0.70** (λ)
				8.47	8.27	20.56

In this table, I report the same individual asset level test using FF5 factors, PCA factors (the three principal components of all asset returns) and the Q5 factors (see Hou et al. (2015), Hou et al. (2021)). I report the results of the same tests in **Table 1.5**. I estimate the idiosyncratic variance $\hat{\theta}_{i,t}$ by running daily returns on the selected factors per month. The controlling variables are the estimated factor loadings and the lagged characteristics suggested by Daniel and Titman (1997). As in their paper, I control the lagged book-to-market ratio $\mathrm{b/m_{i,t-1}}$ and the momentum factor $\mathrm{mom_{i,t-1}}$ computed by the sum of returns in the last six months as in Jegadeesh and Titman (1993).

Table 1.8. Fama-MacBeth results, individual asset level, sub-sample results separated by decades.

Fama-Macbeth regression per decade, Firm level						
	1960	1970	1980	1990	2000	2010
η	-13.50	-0.10	0.25	1.01	-2.35	-1.61
	-1.77	-0.04	0.43	1.66	-3.29	-2.02
γ	10.81	11.64	11.45	-1.53	7.80	15.17
	3.44	4.92	5.42	-0.71	3.07	5.96
λ	0.93	0.61	0.69	0.25	0.80	1.05
	9.37	6.78	8.53	2.24	12.00	15.59
Number of firms	2995	6718	10428	12477	9040	6060

In this table, I report the same individual asset level test in sub-samples separated by decades. I report the results of the same tests in **Table 1.5**. 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 $\mathrm{b/m_{i,t-1}}$ and the momentum factor $\mathrm{mom_{i,t-1}}$ computed by the sum of returns in the last six months as in Jegadeesh and Titman (1993).

Table 1.9. Time-series results

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

In this table, I report 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 hypothesized predictive relation should be negative $A < 0$. In Panel A, I check the prediction results at multiple horizons at various horizons $k = 1, 12, 60$, $\log(r_{m,t+k})$ = constant + A log ¸t. Furthermore, I control for the level of idiosyncratic risk in $\log(r_{m,t+k})$ = constant + A log ¸t + ∑wi,t−1`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.

Table 1.10. Summary of Predictors

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

In this table, I report the AR1 coefficient of all the predictors used. Also, I include the correlation coefficient between each controlling predictor 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.

Table 1.11. Time series results controlling for other predictors.

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, 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

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

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

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 hypothesized 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.

Chapter 2

Fiscal Imbalances, Foreign Account Imbalances, and Asset Returns

2.1 Introduction

The dynamics of US stock returns, US Treasury returns, and foreign asset investment returns are of central interest in finance and macroeconomic research, as they represent the private, public, and external sectors of the economy, respectively. These sectors are not isolated from each other; in fact, they are intricately connected, as the aggregate output of the economy equals the sum of the cash flows of the three sectors. Understanding the joint dynamics of these sectors is therefore crucial for investors and policymakers, as it can help answer questions such as: What are the implications of a contractionary fiscal policy for equity and foreign investment returns? How do changes in net imports affect domestic asset valuation?

In this paper, we present a framework that allows us to explore the joint dynamics of the three sectors by imposing the aggregate budget constraint of the economy. Specifically, our objective is to examine the predictability of asset returns within each sector and across sectors and to investigate whether imposing the aggregate level accounting identity can improve the results of return predictability. Naturally, our framework integrates three significant strands of literature that examine the predictability of returns within each sector separately. The literature, beginning with the pioneering paper of Campbell and Shiller (1988b), uses each sector's intertemporal budget constraint to demonstrate that the sector's current cash flow can predict its future expected returns or cash flow growth rate.[1] For instance, the dividend-to-price ratio can predict future expected returns or the future dividend growth rate of a firm in the private sector. Similarly, the literature has used the government surplus-to-debt ratio as a predictor for the public sector and the import-to-foreign asset value ratio as

[1] Since the literature on the predictability of returns within each sector separately is vast, we discuss it further in the related literature section. Some notable references on budget constraints within each sector include Campbell and Shiller (1988b), Larrain and Yogo (2008), Berndt et al. (2012), and Gourinchas and Rey (2007b), among others.

a predictor for the external sector.

The key insight of our framework is to use the aggregate budget constraint to tie the three predictors, motivated by separate sector budget constraints, together. Specifically, we use a GMM estimation to enforce the accounting identity that total consumption equals the sum of cash flows in the three sectors: dividends, government surplus (tax minus spending), and net imports, respectively. Our approach can help identify how changes in one sector propagate to the others, conditioning on its contribution to the cyclical movement of the whole economy. We apply this framework to test the cross-predictability among the three sectors and investigate whether a value adjustment or policy shifting in one sector could predict a corresponding change in the expected returns or growth rate of the other sectors. This test of cross-predictability is of great importance, as it can shed light on how different sectors of the economy are interconnected and how shocks in one sector could affect the others. For example, it can help us understand how fiscal policy may affect risk premiums in equity and foreign asset investments.

Turning to our empirical analysis, we first replicate the univariate results of each sector and find that our results align with the predictability literature for each sector. Specifically, the dividend-to-price ratio predicts equity returns, with more significance at longer horizons (e.g., see Campbell and Shiller (1988a), Cochrane (2008), and Pástor and Stambaugh (2012)). On the other hand, the predictability of Treasury bond returns is relatively weaker, as documented in recent papers such as Jiang et al. (2021). For the external sector, we also find that the import-foreign asset ratio predicts foreign investment returns, which is similar to the findings in Gourinchas and Rey (2007b) and others. [2]

[2] We use the import-to-foreign asset ratio instead of the export ratio measured in Gourinchas and Rey (2007b) to express the consumption as the sum of three cash flows. While they computed the net export-to-net asset ratio and found a non-stationary trend due to the increasing import of the US over time, we applied a non-forward-looking filter (see Hamilton (2018)) to subtract the non-stationary trend rather than the Hodrick–Prescott filter in their analysis.

Next, we run a multivariate regression to test the cross-predictability between sectors. Our key finding is that the government surplus-to-debt ratio negatively predicts future equity and foreign asset returns. Furthermore, by imposing the aggregate budget constraint using the GMM estimator to estimate all predictive regressions jointly, we find that the constrained estimate has smaller estimation error and more predictive power in the longer horizon, which amplifies our key finding that the government surplus-to-debt ratio is a significant predictor of equity and foreign asset returns.

Our empirical results underscore the importance of considering budget constraints when examining the joint dynamics of asset returns. Moreover, they highlight the impact of fiscal policy on the other two sectors. The negative predictability of the government surplus-to-debt ratio on the asset returns in the two other sectors is anchored on the classical view of the impact of fiscal policy, the Ricardian Equivalence (Barro (1979)). If investors believe in the notion that the government has to either ”tax now” or ”tax later”, a low surplus today could lead them to expect higher taxes in the future. This, in turn, may result in higher risk premiums for investing in firms, as they anticipate increased fixed costs imposed by the government.

In summary, our empirical results contribute to the literature by testing whether fiscal decisions could predict the future expected returns or payouts growth rate in the private and external sector, which is of importance yet remains unknown before our general framework on cross-sector return predictability. Overall, our paper contributes to the understanding of the joint dynamics of asset returns in the US economy by examining the cross-predictability among the three major sectors - equity, Treasury bonds, and foreign assets. Although our primary focus is on return predictability, the general framework we develop to impose the aggregate budget constraint can help to deepen our understanding of the regime switching among sectors and the structural reasons behind their balance from each other.

Related Literature

Our approach to decomposing the entire economy into three sectors and imposing the aggregate budget constraint relates to three strands of literature on dynamic value-adjustment based on the separate budget constraints of each sector, respectively: the equity, public debt, and foreign asset account.

For equities, a large literature beginning with Campbell and Shiller (1988b) and Campbell (1993) analyzes the dynamics of stock returns based on the budget constraint of a representative investor. Research using this framework demonstrates that the dividend-price ratio is driven by a time-varying risk premium of equity returns (see Campbell (1991), Cochrane (2008), Cochrane (2011), Pástor and Stambaugh (2012)). The predictability of dividend growth rate is also investigated by Menzly et al. (2004) and Lettau and Ludvigson (2005). Additionally, there is an extensive body of research focusing on understanding the drivers of stock return predictability and testing the stability of this framework with richer econometric or structural model specifications, e.g., Timmermann (1993), Ang et al. (2007), Campbell and Thompson (2008), Koijen and Van Nieuwerburgh (2011). The vast amount of literature cited primarily focuses on the equity sector and contributes extensively to understanding the drivers of stock return predictability. For our purpose, we use the dividend-to-price ratio as the predictor of the equity sector and employ it to test the joint dynamics of the three sectors.

Our paper also relates to Lettau and Ludvigson (2001) who use an aggregate budget constraint to forecast equity returns. Our study is different in focus and results in several important aspects. First, Lettau and Ludvigson (2001) use a consumption-wealth ratio ("cay") to predict equity returns only, by decomposing aggregate wealth into human capital and investable assets. They do not decompose assets into the three sectors and do not look at across-sector predictability, which is the focus of our decomposition. Second, they use aggregate consumption without decomposing it. We use the national accounts identity to decompose consumption into flow variables for the three sectors. The novelty of the across-sector decomposition is precisely what allows us to ask whether there is across-sector predictability in the data. Naturally, the labor-non-labor decomposition and our three-sector decomposition are complementary. The private, public, and foreign sectors can further be decomposed into labor and non-labor income components, as done by Lettau and Ludvigson (2001).[3]

Similarly, the government budget constraint is extensively used in macroeconomics and finance literature to tackle questions related to the effects of fiscal policy on the real economy and inflation. Recent examples of this literature include Berndt et al. (2012), Berndt and Yeltekin (2015), and Giannitsarou and Scott (2006). It also underlies the Ricardian equivalence literature (Barro (1979)). This literature uses log-linearization of the government budget constraint, similar to the equity sector. However, the net surplus (tax minus government spending) can be negative, unlike the cash flows of equities. To avoid the technical issue of log-linearization due to negative surplus, these papers assume a stable co-integration of tax and government spending, measuring fiscal imbalance by a linear combination of the logged tax-to-debt ratio and the spending-to-debt ratio. Our measure of fiscal policy imbalance differs from this literature, as we derive a new linear identity that allows for using net government surplus directly. Our measure is more general, as it does not require a stable co-integration of tax, government spending, and debt. Our measure captures the high volatility of net surplus in the data and better fits the high government spending sample periods. Recent papers also derive a linear identity without using negative net surplus, but by using the logged debt-to-total output (GDP) ratio to predict treasury bond returns (see Cochrane (2022) and Jiang et al. (2021)). This approach differs from ours, as it

[3]We do not emphasize labor income, as our focus is on the three sectors and across-sector predictability. Conseptually, one can introduce labor income more explicitly into our framework by further decomposing each of the three sector's assets into parts due to human capital and investable income. As labor income is non-observable, one would have to use the approach in Lettau and Ludvigson (2001) and come up with equivalent measure of "cay" for each sector.

does not impose the same budget constraint[4]. In addition, our approach of using the budget constraint to predict treasury bond returns differs from Fama and Bliss (1987) and Cochrane and Piazzesi (2005), which use the expectation of future interest rates to predict realized bond returns.

In international finance, the valuation channel refers to the impact of capital gains and losses on the external balance sheet. Lane and Milesi-Ferretti (2001) pioneered studies on the valuation effects. Our measure of foreign asset account imbalance is closely related to that of Gourinchas and Rey (2007b), who laid the foundations using a similar budget constraint for the foreign sector to test the dynamic value-adjustment based on the imbalance of the foreign asset account. They documented that external imbalance strongly predicts returns on net foreign assets. Gourinchas (2008) suggested that short-term movements in external asset positions appear to be driven by the valuation component in recent periods. Gourinchas and Rey (2014) concluded that valuation effects have gained growing importance for both developed and emerging countries. In particular, Gourinchas and Rey (2007b) cast doubt on the notion that exchange rates may well be random walks (Meese and Rogoff (1983)) and also documented strong predictability of exchange rates, overshadowing conventional predictors such as cross-country differences in fundamentals (Rossi (2013)). Lane and Shambaugh (2010) further studied the role of exchange rate movements in the valuation channel by constructing a comprehensive dataset of a large number of countries. Corte et al. (2016) showed that external imbalance can be useful in developing profitable cross-country currency trading strategies.

[4]Liu (2023) also uses debt-to-GDP ratio to predict stock returns.

2.2 Three Budget Constraints and One Aggregate Budget Constraint

The starting point is the aggregate accounting identity of GDP:

$$Y_t = C_t + I_t + G_t + X_t - M_t \tag{2.1}$$

where Y_t is GDP, C_t is aggregate consumption, I_t is the aggregate investment for future production, G_t is the government spending, and X_t, M_t are export and import. We can re-write this identity as

$$C_t = D_t + NS_t + NM_t \tag{2.2}$$

where $D_t = Y_t - I_t - T_t$ is the dividend. This definition of dividend equals to the output minus investment and tax, which is typical in literature (see Larrain and Yogo (2008)), as the net payout to the firm owner. $NS_t = T_t - S_t$ is the net government tax surplus, and $NM_t = M_t - X_t$ is the net imports. Conceptually, aggregate consumption is driven by flows variables from three sectors: the private sector, the public sector, and the external sector. In the private sector flow, the flow is the after-tax aggregate dividend. In the public sector, it is the net government surplus. In the foreign sector, the flow is net imports.

In asset pricing, it is often assumed that $C_t = D_t$ (Lucas Jr (1978), Campbell (1996)), which implicitly side-steps the role of fiscal policy or net imports on asset returns. For each of the three sectors, we can define their own intertemporal budget constraints. For the private sector, it is

$$A_{t+1} + D_{t+1} = (1 + R^A_{t+1})A_t \tag{2.3}$$

where A_t is the value of private wealth in the economy, and R^A_{t+1} is the return to private wealth.

The public sector's budget constraint is:

$$B_{t+1} + NS_{t+1} = (1 + R^B_{t+1})B_t \tag{2.4}$$

where B_t is the total value of government debt outstanding, and R^B_{t+1} is the return of that debt.

Similarly, for the external sector, the budget constraint is:

$$F_{t+1} + NM_{t+1} = (1 + R^X_{t+1})F_t \tag{2.5}$$

where F_t is the net value (assets minus liabilities) of foreign assets, and R^X_{t+1} is the return on those foreign assets.

These budget constraints have been used in isolation in the literature (for example, Campbell and Shiller (1988a) for the equity sector, Berndt et al. (2012) for the public sector, and Gourinchas and Rey (2007b) for the external sector).

We can define the total domestic wealth of the economy W_t as:

$$W_t = A_t + B_t + F_t \tag{2.6}$$

and using total outflows, C_t, can write the aggregate budget constraint as

$$W_{t+1} + C_{t+1} = (1 + R^W_{t+1})W_t \tag{2.7}$$

where the gross return of the economy is equal to the weighted average of the returns in each sector, or

$$1 + R^W_{t+1} = (1 + R^A_{t+1})(A_t/W_t) + (1 + R^B_{t+1})(B_t/W_t) + (1 + R^X_{t+1})(F_t/W_t) \tag{2.8}$$

2.2.1 Three BCs and Within-Sector Predictability

We use a variation of a common log-linearization approach (Campbell and Shiller (1988b), Larrain and Yogo (2008), Gourinchas and Rey (2007b)) to obtain a log-linear relation between state variables, future log returns, and future dividend growth rates. For the three sector BCs in (2.3), (2.4), and (2.5) we obtain

$$DA_t = \kappa_{0,DA} + \kappa_{1,DA} \sum_{j=0}^{\infty} \rho_{DA}^{j} \left(r^{E}_{t+j+1} - \Delta d_{t+j+1} \right) \tag{2.9}$$

$$NSB_t = \kappa_{0,NSB} + \kappa_{1,NSB} \sum_{j=0}^{\infty} \rho_{NSB}^{j} \left(r^{D}_{t+j+1} - \Delta ns_{t+j+1} \right) \tag{2.10}$$

$$NMA_t = \kappa_{0,NMA} + \kappa_{1,NMA} \sum_{j=0}^{\infty} \rho_{NMA}^{j} \left(r^{X}_{t+j+1} - \Delta nm_{t+j+1} \right) \tag{2.11}$$

where $DA_t = D_t/A_t$, $NSB_t = NS_t/B_t$, $NMA_t = NM_t/F_t$ are the valuation ratios in each sector. The κs and ρs are constants determined by the steady-state values of each ratio. Their functional form is specified in the **Appendix Section B.1**.

It is worth highlighting that, in expressions (2.9), the dividend yield DA_t is not in logs, and neither are the state variables in the other sectors, NSB_t and NMA_t. This is because, as outlined in the appendix, we use a modification of the standard log-linearization to accommodate negative and volatile cash flows. Specifically, the modification is necessary because, in the data, net government surplus is negative and volatile for extended periods, which challenges some of the existing log-linearization approaches (Larrain and Yogo (2008), Berndt et al. (2012)). Similarly, in the foreign sector, net imports are negative in a part of our sample period, and net foreign assets can be close to zero when net foreign assets and liabilities are of similar magnitudes. These features present a difficulty for the typical log-linearization for the external sector (Gourinchas and Rey (2007a)). But conceptually, we obtain linear relations between the state variables on the left-hand-side and future log returns and log growth of fundamentals which then allow us to test for predictability in returns and growth in fundamentals.

The three intertemporal relations (2.9-2.11) have been the foundation of separate strands of literature that analyze the predictability of equity returns, Treasury returns, and foreign asset returns. The conditioning variables used are either the within-sector conditioning variables or macroeconomic quantities, such as inflation, interest rates, term spread, and default spread.

2.2.2 The ABC and Across-Sector Predictability

The log-linearized version of the ABC obtains similarly to the three sector BCs. Namely, linearizing and iterating forward equation (2.7) yields the following expression

$$CW_t = \kappa_{0,CW} + \kappa_{1,CW} \sum_{j=0}^{\infty} \rho_{CW}^{j} \left(r_{t+j+1}^{W} - \Delta c_{t+j+1} \right). \tag{2.12}$$

The state variable in this expression is the consumption-to-wealth ratio, CW_t, which is a weighted average of the state variables of the individual sectors, or

$$CW_t = \alpha_0 + w_1 DA_t + w_2 NSB_t + w_3 NMA_t. \tag{2.13}$$

Similarly, the aggregate the log return of the economy, r_{t+j+1}^{W}, is a weighted average of the log returns in the three sectors,

$$r_{t+j+1}^{W} = \alpha_1 + w_1 r_{t+j+1}^{E} + w_2 r_{t+j+1}^{D} + w_3 r_{t+j+1}^{X}. \tag{2.14}$$

And similarly, for the consumption growth rate,

$$\Delta c_t = \alpha_2 + w_1 \Delta d_t + w_2 \Delta ns_t + w_3 \Delta nm_t. \tag{2.15}$$

The weights of the three sectors are $w_1 = \frac{AW/\rho_{DA}}{AW/\rho_{DA}+BW/\rho_{NSB}+FW/\rho_{NMA}}$, $w_2 = \frac{BW/\rho_{NSB}}{AW/\rho_{DA}+BW/\rho_{NSB}+FW/\rho_{NMA}}$, and $w_3 = \frac{FW/\rho_{NMA}}{AW/\rho_{DA}+BW/\rho_{NSB}+FW/\rho_{NMA}}$. Intuitively, w_1, w_2, and w_3 are each sector's weight in the aggregate wealth at the steady state, adjusted by the log linearization coefficients ρs.

The aggregate budget constraint presented in equation (2.13-2.15) ties the predictability of each sector together by enforcing the weighted average of the three predictors, i.e. DA, NSB and NMA, respectively, equal to the weighted average of the expected future returns and cash flow growth rates of the three sectors.

Therefore, if there is an imbalance in one sector, such as $NSB = \frac{T-G}{B}$, which measures fiscal policy, this change will affect the consumption-to-wealth ratio and could generate correspondence to future asset returns and cash flow growth rates in all the sectors. We identify this channel by enforcing the coefficient of predicting the consumption growth rate to be equal to the weighted average of the coefficients of predicting all three cash flow growth rates, as indicated by (2.15). Furthermore, to link to the return predictability, we also impose within-sector constraints such that the coefficients of predicting returns and cash flow growth rate satisfies the separate budget constraints in (2.9-2.11). We apply the GMM estimation method to enforce all the budget constraints in a multivariate predictive system that combines all the sectors to study cross-sector predictability. We provide details of our estimation strategy in **Appendix Section B.2**.

The study of cross-sector predictability provides valuable insights into the interplay between different sectors' returns and cash flow growth rates, shedding light on the mechanisms that drive asset pricing in the economy. By decomposing total consumption into dividends, government surplus, and net imports, our approach allows us to investigate how fiscal policies and other macroeconomic shocks impact the returns of different asset classes. For example, if a tax increase has a real impact on the equity market, we would expect the government surplus to predict equity returns in our framework. Moreover, by studying the predictability of returns and cash flow growth rates across multiple sectors simultaneously, we can also gain a better understanding of the factors that drive asset prices and the transmission of shocks across different sectors of the economy. Ultimately, our goal is to develop a more comprehensive and accurate framework for asset pricing that takes into account the complex interdependencies between different sectors of the economy.

Our approach differs significantly from that of Lettau and Ludvigson (2001) (hereafter LL) in several ways. LL introduced a consumption-to-wealth ratio, known as cay, to analyze equity return predictability. However, they focused only on the private sector and divided income into labor and non-labor income. Their objective was not to examine cross-sector predictability but to develop a state variable that could reliably capture expected equity return variation while controlling for the effect of labor income. In contrast, our approach utilizes the same consumption-wealth dynamic as LL to explore interactions among different sectors of the economy and analyze predictability across sectors. By doing so, we aim to gain insights into the drivers of asset pricing and the interplay between different sectors' returns and cash flow growth rates.

2.3 Data

We use quarterly data for the private sector (US equity), public sector (US public debt), and external sector (US foreign asset account). All data start from 1947 on a quarterly basis and are updated until the end of 2021, unless they are not available at the beginning of our sample period. In such cases, we provide specific instructions on where the original series started and how we interpolate back to 1947.

We collect dividend and price data for the S&P 500 from Robert Shiller's website. We compute the summed market capitalization of all the firms in CRSP to scale the dividends and market value of the private sector. To enforce the aggregate budget constraint, we gather the total net worth and personal consumption expenditure (PCE) data from the collection of economic data series on the FRED website. The quarterly data of total net worth begin in 1952. We use annual data from 1947-1952 and assume equal quarterly growth within the same year to recover quarterly values.

Our public sector data primarily comes from the NIPA Table, which is available on the FRED or BEA websites. We collect the total public debt series from FRED. Following Berndt et al. (2012), we compute tax and government spending from terms in NIPA Table 3.2. Tax equals total receipts (line 37) plus seigniorage revenue, which is calculated as $(M_t - M_{t-1})/CPI_t$. M_t represents the money base, and CPI_t is the consumer price index obtained from FRED. Government spending equals current expenditures (line 41) plus gross government investment (line 42) plus capital transfer payments (line 43) minus consumption of fixed capital (line 45) minus debt interest payments (line 29).

For the external sector, we rely on the methodology of Gourinchas and Rey (2007b) to construct measures on the gross positions, flows, and returns. As Gourinchas et al. (2010) updated certain data sources and computations of these variables, we closely follow them in obtaining the variables from numerous databases, including Flow of Funds Accounts, Board of Governors of the Federal Reserve System, and Bureau of Economic Advisors. Our calculation includes data from 13 countries: Australia, Canada, Denmark, France, Germany, Ireland, Italy, Japan, Mexico, Netherlands, Spain, Sweden, and the United Kingdom. The raw data for the external sector is not as well-organized as for the other two sectors, and we interpolate the data to match the length and frequency of our sample when the raw data is unavailable. Detailed documentation is available upon request.

2.4 Results

The summary statistics of the returns of the three sectors (r^E, r^D, r^X), cash flows of the three sectors (Δd, Δns, Δnm), and valuation ratios DA, NSB, and NMA are in Panel A of Table 2.1. The average annual return of the US stock market, adjusted for inflation, is 8% with a standard deviation of 16%. The Treasury return adjusted for inflation is 2%, while the return of foreign assets adjusted for inflation is 6%. The real returns of the three sectors are not correlated, as evidenced by the correlation matrix presented in Panel B of Table 2.1. The weak correlations among these returns suggest that they represent three distinct asset classes, each potentially attracting a different group of investors. This further highlights the importance of considering these sectors separately when examining their dynamics and interactions.

We would also like to draw attention to the high volatility of the growth rate of net government surplus, which stands at an annual level of 146%. This significant volatility is primarily due to shifts in fiscal policy. This stylized fact led us to develop a linearization technique that captures this high volatility, utilizing the raw government surplus-to-debt ratio rather than a log-linearized approximation employed in Larrain and Yogo (2008) and Berndt et al. (2012). In **Figure 2.1**, we compare the two time-series to highlight the differences: During extreme historical periods, such as the oil crisis in 1973, the global financial meltdown of 2008-09, and the unprecedented COVID-19 pandemic in 2020, the rich dynamics of $NSB = \frac{T-G}{B}$ cannot be accurately captured by its log-linearization. These extraordinary events led to dramatic shifts in government policies and fiscal interventions, which in turn resulted in considerable fluctuations in the net government surplus.

Interestingly, the correlations between the valuation ratios and the cash-flow growth of the three sectors throughout the whole sample are not significant, with the strongest correlation being -0.12 between dividends and government surplus. We plot the three ratios, which serve as state variables driving the cyclical movement of the whole economy, together in **Figure 2.2**. In the plot, we observe temporary correlation regimes among the three sectors. For example, the dividend-to-price ratio kept going down during the 1990s, and both the net surplus and import increased simultaneously. However, these correlations are not permanent, and there isn't always one sector's valuation ratio going down while the other two go up. The time-varying mechanism of the three sectors balancing against each other cancels out throughout the entire sample, leading to a weak unconditional correlation. From this perspective, our framework, which studies the joint dynamics of the three sectors and enforces the aggregate budget constraint, provides a solid starting point. Although our primary focus in this paper is on asset return predictability, exploring the conditional correlations and regime-switching of the joint dynamics of the three sectors would be an interesting topic for further investigation.

We now begin by replicating the predictability results of each sector using our data. The univariate predictive regressions, which use each sector's valuation ratio, are presented in Table 2.2. This table replicates the results for the predictive regressions of the equity sector (columns 1 and 2), Treasury sector (columns 3 and 4), and foreign asset investment sector (columns 5 and 6). The table consists of three forecasting horizons in Panels A, B, and C. Additionally, we report unbiased estimates of our coefficients, as suggested in Stambaugh (1999).

Our findings align with the predictability literature for each sector. The dividend-to-price ratio, DA, predicts the equity return r^E at long horizons, as evidenced by studies such as Campbell and Shiller (1988a), Cochrane (2008), and Pástor and Stambaugh (2012). Although we construct the dividend-price ratio without logging it, resulting in greater volatility than the logged ratio, the return predictability result is quite similar. Most of the predictability stems from equity returns rather than the dividend growth rate. In other words, equity valuations are primarily driven by time-varying equity risk premiums.

The predictability of treasury bond returns r^B by NSB is less strong and only at 8-period-ahead horizons. At shorter horizon, we observe no predictability, which is similar to the public debt valuation puzzle in Jiang et al. (2021). In other words, the expected treasury bond returns in the future, due to potential mis-valuation of the public debt, do not change according to the current fiscal policy status, captured by $NSB = \frac{T-G}{B}$. This valuation puzzle, within our framework, further motivates us to investigate the cross-predictability of one sector's valuation ratio on the other sector's returns.

We ask whether the expected returns of equity and foreign assets correspond to fiscal policy. If investors subscribe to Ricardian Equivalence (Barro (1979)) – the notion that the government has to either "tax now" or "tax later" – a low surplus today could lead them to expect higher taxes in the future. This, in turn, may result in higher risk premiums for investing in firms, as they anticipate increased fixed costs imposed by the government.

For the external sector, we similarly find that NMA significantly predicts future foreign investment returns, while cash-flow predictability remains weak, suggesting that the valuation adjustment in foreign assets is primarily driven by the time-varying risk premium. Although our construction of the predictor differs from that of Gourinchas and Rey (2007b) as discussed in **Appendix Section B.1**, our findings align well with their results.

Table 2.3 presents the multivariate results, where the returns of each sector are regressed on their own predictors and the predictors of the other two sectors. Serving as a benchmark, these regressions are unrestricted, meaning no constraints are imposed during the estimation process. Motivated by the univariate results and our intention to understand whether the expected returns in one sector correspond to imbalances in other sectors, our primary focus is on cross-predictability. Specifically, we examine the influence of the net surplus-to-debt ratio (NSB), the dividend-to-price ratio (DA), and the net import-to-foreign asset ratio (NMA) in predicting returns across different markets. In line with the univariate analysis, the multivariate results are presented across three forecasting horizons in Panels A, B, and C, allowing for a comprehensive examination of the joint return dynamics in both the short and long terms.

In summary, the multivariate analysis presented in Table 2.3 confirms that the predictability of each sector's own returns remains largely intact even after accounting for state variables in other sectors. The coefficients for each predictor are not significantly different from their univariate counterparts in Table 2.2, given the low correlation among the predictors as indicated in Table 2.1. Regarding cross-predictability, we observe a significant negative relationship between NSB and returns in the private and foreign sectors, confirming our intuition that a high government surplus now suggests a lower tax in the future and lower risk premiums for investing in equity or foreign assets. Meanwhile, neither DA nor NMA exhibit significant predictability on returns in the other two sectors. Notably, the cross-predictability of NSB is weaker in longer horizons in the unrestricted regressions. This result could possibly relates to the regime switching of how sectors balance from each other in the long horizon.

To better understand the joint dynamics of the three sectors, following our unconstrained estimation in Table 2.3, we impose the sector budget constraints and the aggregate budget constraint and estimate all predictive regressions jointly with GMM.[5] A detailed explanation of the estimation strategy is provided in **Appendix Section B.2**. With the imposition of constraints, we observe some notable changes in our results reported in 2.4 . Firstly, the predictive power of NSB is stronger in predicting public debt returns at longer horizons. Secondly, the negative cross-predictability of NSB is also more pronounced in both the short and long horizons. These results highlight the importance of considering budget constraints when examining the joint

[5]The standard errors of this constrained estimation are computed by bootstrapping for 1,000 times.

dynamics of asset returns. If the Ricardian Equivalence holds in the long run, a lower government surplus-to-debt ratio relative to its long-term average would signal a lower cost of borrowing or a higher net surplus in the future, which could be achieved by increasing taxes or decreasing government spending. In this sense, the NSB ratio negatively predicts the risk premium in the equity and foreign asset sector, as a lower government surplus could signal a contractionary fiscal policy in the future and higher risk premium for investing in equity or foreign assets. However, this mechanism may be difficult to observe, particularly in the short term using only public sector data. Our results suggest that incorporating data from the other two sectors and imposing aggregate budget constraints can help to better identify how this fiscal policy adjustment channel propagates throughout the economy.

All of our results are based on asset returns adjusted for inflation. In **Appendix Section B.3**, we present robustness checks of our results using nominal returns and excess returns relative to risk-free rates. We find that the predictability of each sector's own returns and cross-sector predictability remains largely similar regardless of whether we use real, nominal, or excess returns. These results suggest that our findings are robust to different measures of asset returns.

2.5 Conclusion

In this study, we investigate the joint dynamics and predictability of asset returns for the equity, treasury, and foreign asset investment sectors, as each represents a distinct asset class with different dynamics and potential investors. We find that the predictability of each sector's own returns remains largely intact even after accounting for state variables in other sectors.

To better understand the joint dynamics of the three sectors, we impose the aggregate budget constraint and estimate all predictive regressions jointly. The results of this estimation highlight the importance of considering the aggregate cyclical movement of the economy as a whole when examining the joint dynamics of asset returns. Based on our multivariate analysis, we find that the net surplus-to-debt ratio negatively predicts the risk premium in the equity and foreign asset investment sectors, as a lower government surplus could signal a contractionary fiscal policy in the future and higher risk premiums for investing in equity or foreign assets. However, this mechanism may be difficult to observe in the short term based solely on public sector data. Our results suggest that considering data from all three sectors and imposing aggregate budget constraints can help to better identify how this fiscal policy adjustment channel propagates throughout the economy.

Overall, our study contributes to the literature on asset return predictability by examining the joint dynamics of asset returns across different markets, considering their respective valuation ratios and cash flows, and incorporating budget constraints to better understand their interactions. Further research could explore the conditional correlations and regime-switching of the joint dynamics of the three sectors, as well as the implications of our findings for portfolio diversification and asset allocation.

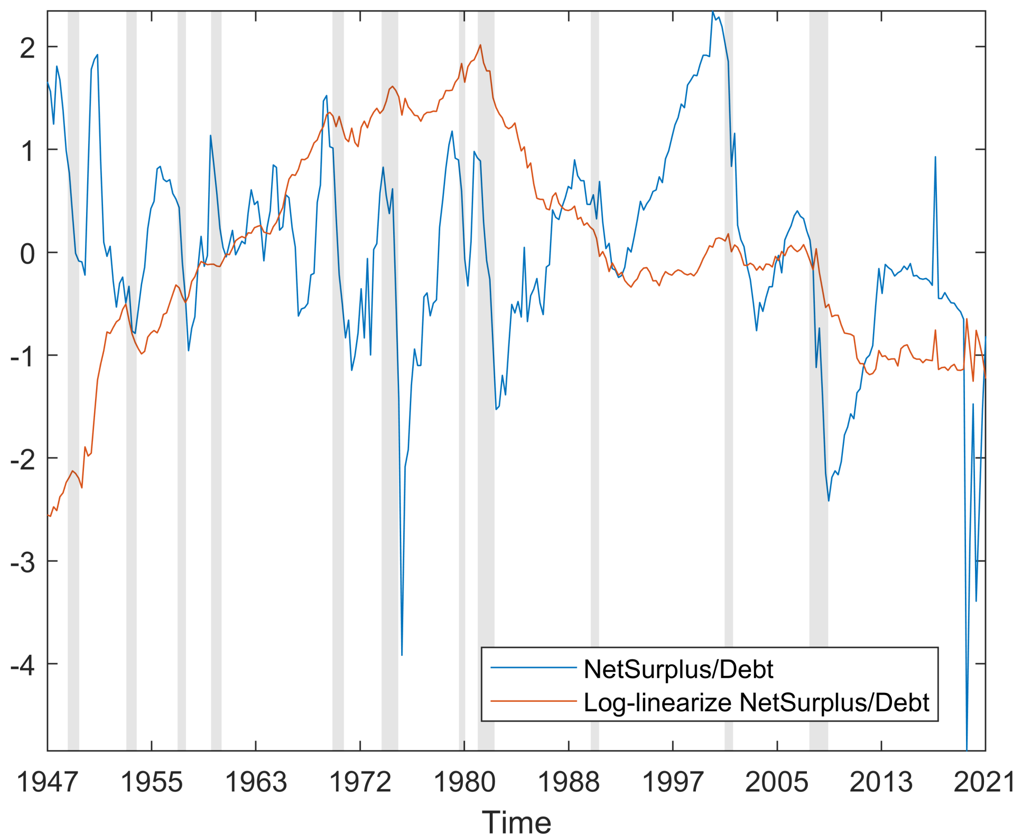


Figure 2.1. Comparison between Government Surplus-to-Debt Ratio and Its Log-linearization.

In this figure, I plot the raw government surplus-to-debt ratio and its log-linearized approximation following the methods in Larrain and Yogo (2008) and Berndt et al. (2012). Both ratios are normalized to have zero mean and unit variance.

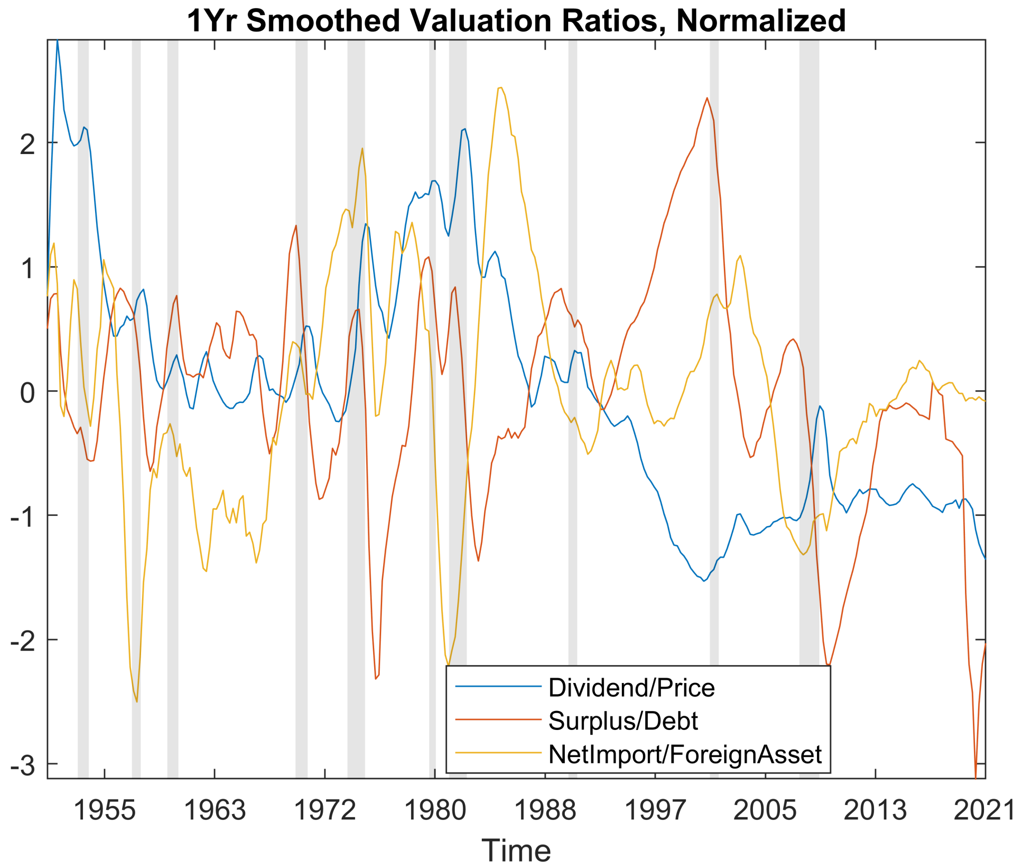


Figure 2.2. The Valuation Ratios of the Three Sectors

In this figure, we plot the valuation ratios of the three sectors. All the ratios are normalized to zero-mean and unit-variance.

Table 2.1. Summary Statistics

Table 1 Panel A: Summary Statistics									
	r^E	r^D	r^X	Δd	Δns	Δnm	DA	NSB	NMA
Mean	0.08	0.02	0.06	0.01	-0.04	0.02	0.03	-0.01	0.05
Std	0.16	0.05	0.23	0.01	1.46	0.06	0.01	0.02	0.09
AR1	0.06	0.04	0.13	0.64	-0.07	0.06	0.96	0.86	0.86
T-stat	0.90	0.44	1.99	7.41	-2.63	0.99	66.86	21.65	21.77

Table 1 Panel B: Correlation matrix among variables								
Corr(ret)			Corr(Δcf)			Corr(ratio)		
1.00	0.03	0.08	1.00	-0.12	0.00	1.00	-0.03	0.11
0.03	1.00	0.03	-0.12	1.00	0.02	-0.03	1.00	0.02
0.08	0.03	1.00	0.00	0.02	1.00	0.11	0.02	1.00

This table displays summary statistics of three sector log returns (real returns), cash flow growth, and valuation ratios. We construct real return by logged nominal return minus logged inflation(growth of the CPI index). Our data includes 300 observations for each time series at a quarterly frequency between 1947 Q1 and 2021 Q4. Panel A firstly shows annualized mean and standard deviation of each time series, then the AR1 coefficients, and the T-stats of the coefficients. Panel B shows correlation between returns, cash flow growth, and valuation ratios of the three sectors. r^E, r^D, r^X are the logged real return of equity, debt, and external asset, respectively. $\Delta d, \Delta ns, \Delta nm$ are the cash flow growth of private, debt, and external sectors, respectively. DA is the dividend price ratio, NSB is the net government surplus (tax minus spending) to debt ratio, and NMA is the net import to foreign asset ratio.

Table 2.2. Single Variable Predictive Regression, OLS estimated

Table 2 Panel A: Single Variable Predictive Regression, One-period ahead								
	r^E_{t+1}	Δd_{t+1}		r^B_{t+1}	Δns_{t+1}		r^X_{t+1}	Δnm_{t+1}
DA_t	0.63	-0.18	NSB_t	0.02	-0.84	NMA_t	0.11	-0.05
	1.74	-1.26		0.55	-0.89		2.26	-0.76
R^2	0.01	0.02	R^2	0.00	0.00	R^2	0.03	0.00
unbiased	0.28	-0.15	unbiased	0.02	-1.28	unbiased	0.11	-0.03
	0.67	-1.70		0.43	-0.27		2.88	-0.79
$\tilde{R}^2$	0.00	0.01	$\tilde{R}^2$	0.00	0.00	$\tilde{R}^2$	0.03	0.00
Table 2 Panel B: Single Variable Predictive Regression, Four-period ahead								
	$\sum_{k=1}^4 \rho^k r^E_{t+k}$	$\sum_{k=1}^4 \rho^k \Delta d_{t+k}$		$\sum_{k=1}^4 \rho^k r^D_{t+k}$	$\sum_{k=1}^4 \rho^k \Delta ns_{t+k}$		$\sum_{k=1}^4 \rho^k r^X_{t+k}$	$\sum_{k=1}^4 \rho^k \Delta nm_{t+k}$
DA_t	2.70	-0.63	NSB_t	0.15	1.12	NMA_t	0.38	-0.09
	2.17	-1.27		1.27	0.18		2.30	-0.64
R^2	0.05	0.02	R^2	0.01	0.00	R^2	0.08	0.00
unbiased	2.36	-0.61	unbiased	0.15	0.50	unbiased	0.38	-0.08
	3.00	-2.22		1.71	0.06		4.65	-0.89
$\tilde{R}^2$	0.04	0.02	$\tilde{R}^2$	0.01	0.00	$\tilde{R}^2$	0.08	0.00
Table 2 Panel C: Single Variable Predictive Regression, Eight-period ahead								
	$\sum_{k=1}^8 \rho^k r^E_{t+k}$	$\sum_{k=1}^8 \rho^k \Delta d_{t+k}$		$\sum_{k=1}^8 \rho^k r^D_{t+k}$	$\sum_{k=1}^8 \rho^k \Delta ns_{t+k}$		$\sum_{k=1}^8 \rho^k r^X_{t+k}$	$\sum_{k=1}^8 \rho^k \Delta nm_{t+k}$
DA_t	4.62	-0.70	NSB_t	0.32	3.86	NMA_t	0.51	-0.08
	2.36	-0.79		1.50	0.28		1.89	-0.50
R^2	0.08	0.01	R^2	0.02	0.00	R^2	0.07	0.00
unbiased	4.38	-0.71	unbiased	0.32	2.97	unbiased	0.52	-0.07
	4.18	-1.53		2.20	0.22		4.31	-0.67
$\tilde{R}^2$	0.07	0.01	$\tilde{R}^2$	0.02	0.00	$\tilde{R}^2$	0.07	0.00

This table shows the prediction results of three sectors using each valuation ratio in a single variable regression, estimated by OLS. r^E, r^D, r^X are logged real returns of equity, debt, and external asset, respectively. $\Delta d, \Delta ns, \Delta nx$ are cash flow growth of private, debt, and external sectors, respectively. DA is the dividend price ratio, NSB is the net government surplus (tax minus spending) to debt ratio and NMA is the net import to foreign asset ratio. We correct for Stambaugh bias and compute the unbiased coefficient and R-squared $\tilde{R}^2$ accordingly. The standard error of the unbiased estimate is generated by bootstrapping 10,000 draws. In Panel A, we forecast the one-period ahead variables. In Panels B and C, we compute the summed returns and cash-flow growth in the long horizon.

Table 2.3. Multiple Variables Predictive Regression

Table 3 Panel A, Multiple Variable Predictive Regression, One-period ahead						
	r^E_{t+1}	r^D_{t+1}	r^X_{t+1}	Δd_{t+1}	Δns_{t+1}	Δnx_{t+1}
DA_t	0.61	0.07	-0.03	-0.19	21.81	-0.48
	1.78	0.41	-0.06	-1.24	1.74	-0.57
NSB_t	-0.27	0.02	-0.41	0.05	-0.64	-0.40
	-2.16	0.56	-2.52	0.66	-0.45	-1.62
NMA_t	-0.01	0.00	0.12	0.01	0.39	-0.04
	-0.18	0.36	2.40	1.21	0.33	-0.75
R^2	0.03	0.00	0.05	0.04	0.01	0.02
Table 3 Panel B, Multiple Variable Predictive Regression, Four-period ahead						
	$\sum_{k=1}^4 \rho^k r^E_{t+k}$	$\sum_{k=1}^4 \rho^k r^D_{t+k}$	$\sum_{k=1}^4 \rho^k r^X_{t+k}$	$\sum_{k=1}^4 \rho^k \Delta d_{t+k}$	$\sum_{k=1}^4 \rho^k \Delta ns_{t+k}$	$\sum_{k=1}^4 \rho^k \Delta nm_{t+k}$
DA_t	2.55	0.17	0.07	-0.71	89.03	1.24
	2.20	0.29	0.05	-1.42	2.06	0.58
NSB_t	-0.86	0.16	-1.75	-0.08	2.96	-1.24
	-2.31	1.26	-3.27	-0.32	0.40	-1.73
NMA_t	-0.02	-0.01	0.39	0.04	4.03	-0.09
	-0.20	-0.16	2.50	1.35	1.11	-0.70
R^2	0.09	0.01	0.15	0.04	0.06	0.05
Table 3 Panel C, Multiple Variable Predictive Regression, Eight-period ahead						
	$\sum_{k=1}^8 \rho^k r^E_{t+k}$	$\sum_{k=1}^8 \rho^k r^D_{t+k}$	$\sum_{k=1}^8 \rho^k r^X_{t+k}$	$\sum_{k=1}^8 \rho^k \Delta d_{t+k}$	$\sum_{k=1}^8 \rho^k \Delta ns_{t+k}$	$\sum_{k=1}^8 \rho^k \Delta nm_{t+k}$
DA_t	4.43	0.30	0.73	-1.09	114.60	4.72
	2.52	0.34	0.27	-1.42	2.05	1.70
NSB_t	-0.90	0.34	-3.48	-0.81	8.28	-0.49
	-1.36	1.49	-4.03	-2.30	0.57	-0.41
NMA_t	-0.06	-0.04	0.51	0.09	7.31	-0.12
	-0.48	-0.45	2.03	1.72	1.50	-0.75
R^2	0.10	0.03	0.20	0.14	0.07	0.05

This table shows the OLS estimates of using three valuation ratios in one regression predicting logged real returns and cash flow growths of all three sectors. r^E, r^D, r^X are logged real returns of equity, debt, and external asset, respectively. $\Delta d, \Delta ns, \Delta nx$ are cash flow growth of private, debt, and external sectors, respectively. DA is the dividend price ratio, NSB is the net government surplus (tax minus spending) to debt ratio, and NMA is the net import to foreign asset ratio.

Table 2.4. Multiple Variables Prediction, using GMM to Enforce Aggregate Budget Constraint

Table 3 Panel A, Multiple Variable Prediction, One-period Ahead						
	ret^E_{t+1}	ret^D_{t+1}	ret^X_{t+1}	Δd_{t+1}	Δns_{t+1}	Δnx_{t+1}
DA_t	0.48	-0.06	-0.03	-1.25	17.51	0.07
	2.44	-0.64	-0.08	-2.49	2.43	0.16
NSB_t	-0.28	0.02	-0.54	0.03	-0.10	-0.13
	-2.58	0.37	-2.80	0.19	-0.04	-0.63
NMA_t	-0.03	0.01	0.10	0.00	0.21	0.04
	-1.19	1.23	2.39	-0.10	0.38	0.81
Table 3 Panel B, Multiple Variable Prediction, Four-period Ahead						
	$\sum_{k=1}^4 \rho^k r^E_{t+k}$	$\sum_{k=1}^4 \rho^k r^D_{t+k}$	$\sum_{k=1}^4 \rho^k r^X_{t+k}$	$\sum_{k=1}^4 \rho^k \Delta d_{t+k}$	$\sum_{k=1}^4 \rho^k \Delta ns_{t+k}$	$\sum_{k=1}^4 \rho^k \Delta nm_{t+k}$
DA_t	1.92	0.04	0.47	-5.67	78.04	1.85
	3.52	0.19	0.51	-4.70	4.61	1.91
NSB_t	-0.70	0.22	-2.41	-0.22	3.04	-0.59
	-3.11	2.11	-6.49	-0.56	0.55	-1.46
NMA_t	-0.08	0.01	0.33	-0.15	2.89	0.03
	-1.51	0.25	3.86	-1.68	2.30	0.38
Table 3 Panel C, Multiple Variable Prediction, Eight-period Ahead						
	$\sum_{k=1}^8 \rho^k r^E_{t+k}$	$\sum_{k=1}^8 \rho^k r^D_{t+k}$	$\sum_{k=1}^8 \rho^k r^X_{t+k}$	$\sum_{k=1}^8 \rho^k \Delta d_{t+k}$	$\sum_{k=1}^8 \rho^k \Delta ns_{t+k}$	$\sum_{k=1}^8 \rho^k \Delta nm_{t+k}$
DA_t	3.48	0.51	1.57	-8.02	112.85	4.48
	4.70	1.56	1.25	-4.82	4.80	3.73
NSB_t	-0.88	0.47	-4.12	-0.84	7.12	0.22
	-2.67	2.76	-6.99	-1.32	0.77	0.40
NMA_t	-0.11	-0.02	0.43	-0.34	6.32	-0.03
	-1.56	-0.63	3.60	-2.36	3.12	-0.30

This table shows the GMM estimates of three valuation ratios predicting logged real returns and cash flow growths of all three sectors. We use the GMM method to enforce the aggregate budget constraint and the budget constraint of each sector. r^E, r^D, r^X are logged real returns of equity, debt, and external asset, respectively. The standard errors of this constrained estimation are computed by bootstrapping 1,000 times. $\Delta d, \Delta ns, \Delta nx$ are cash flow growth of private, debt, and external sectors, respectively. DA is the dividend price ratio, NSB is the net government surplus (tax minus spending) to debt ratio, and NMA is the net import to foreign asset ratio.

2.6 Acknowledgements

Chapter 2 is currently being prepared for submission for publication and is coauthored with Rossen Valkanov and Yan Xu. Junxiong Gao, the dissertation author, is the primary investigator and author of this paper.

Chapter 3

Conditional Correlation with Factor Structure

3.1 Introduction

Precise estimation of the covariance matrix among assets is essential for finance practitioners in both portfolio allocation and risk management. However, as the number of assets increases, the estimation of the covariance matrix loses accuracy due to the high-dimensional curse.

In terms of cross-sectional estimation, the shrinkage method and factor models are widely used to address the high-dimensional dilemma. The shrinkage method corrects the bias of the covariance matrix eigenvalues using a shrinkage estimator, as demonstrated by Ledoit and Wolf (2004) and Ledoit et al. (2012). Factor models, such as those proposed by Chan et al. (1999) and Ledoit and Wolf (2003), present the asset returns' covariance matrix as a quadratic function of the conditional covariance with factors.

$$Q = B\Sigma B' + Q_\varepsilon$$

, where B is the N by K factor loading, Σ is the covariance among factors, and Q_ε is a diagonal covariance matrix among the residuals. This factor structure reduces the estimation dimension from N to K, where K is the number of factors.

In this paper, we propose a conditional factor model to address the high-dimensional dilemma in time-series estimation using the Dynamic Conditional Correlation (DCC) model, as documented by Engle and Sheppard (2001). The DCC model is a multivariate generalized autoregressive conditional heteroskedasticity-type (GARCH) model that decomposes the covariance to volatility and correlation matrix R_t using the "DeGarch" returns. The covariance likelihood is decomposed into the single-assets GARCH and conditional correlation likelihood. However, as the number of assets increases, the optimization over R_t^{-1} and $|R_t|$ becomes imprecise.

As the primary contribution, we solve R_t^{-1} and $|R_t|$ in closed form by inserting a conditional factor structure into the DCC model, which simplifies the likelihood estimation. Our factor model shrinks the $N \times N$ correlation matrix to a factor loading matrix $X_{N \times K}$, as in cross-sectional factor models. We show that the quadratic form of the correlation matrix provides more dynamism than the linear model used in factor models, providing more accurate estimates of the covariance matrix.

Our proposed model combines techniques from two strands of literature: cross-sectional factor models and time-series DCC models. We highlight the advantages of our model, including its ability to reduce the estimation dimension, its quadratic form for modeling the correlation dynamics, and its well-known inference property, as documented by White (1996).

To test the performance of our model, we compare it against other variations of the DCC model, both in-sample and out-of-sample. We construct the minimum variance portfolio and tangency portfolio based on the out-of-sample forecast of the covariance matrix, as in the conventional literature. Our method outperforms the classical DCC model both in and out-of-sample, and we show that combining our model with the nonlinear shrinkage method in Ledoit et al. (2012) yields better estimates of the unconditional covariance.

Our proposed model relates to recent methods to solve the high-dimensional issue in dynamic conditional correlation models, such as the DECO model by Engle and Kelly (2012). However, our method assumes a different structure from the DECO model, which assumes a pairwise equal correlation structure in the covariance matrix. Our method also differs from the factor models proposed by Engle et al. (1990) and Engle (2009), which assume a constant factor structure.

The remainder of the paper is organized as follows. Section 3.2 presents the proposed model and estimation method. Section 3.3 describes the data and benchmark models used and reports the in- and out-of-sample performance

3.2 Model and Estimation Method

This section introduces the model specification and estimation steps. Within the same quasi-maximum likelihood estimate approach, the estimators' asymptotic distribution is included in Appendix C.2.

3.2.1 Model Specification

I begin by taking a conditional factor setting. Assume there are K factors for all N assets. The pricing kernel m_t is modeled as:

$$m_t = a_{t-1} - b'_{t-1} f_t, \tag{3.1}$$

where a_{t-1} is a scalar, b_{t-1} is a K by 1 rotation vector, and f_t is the K by 1 vector of factors. Each stock r_{it} and factor f_{kt} satisfies the pricing equation:

$$E_{t-1}[r_{it} m_t] = 1. \tag{3.2}$$

Proposition 2.1 Excess Return

The expected excess return under conditions in (3.1) and (3.2) is:

$$E_{t-1}[r_{it} - r_t^f] = \beta'_{it} E_{t-1}[f_t]. \tag{3.3}$$

$$\beta_{it} = COV_{t-1}(r_{it}, f_t) VAR_{t-1}(f_t)^{-1}. \tag{3.4}$$

$E_{t-1}[f_t]$ is a K by 1 vector of each factor risk premium and β_{it} is the K by 1 dynamic factor loading vector. In this equation, $COV_{t-1}(r_{it}, f_t)$ is the K by 1 vector of the conditional covariance between $r_[it]$ and each factor; $VAR_{t-1}(f_t)$ is the conditional variance matrix among K factors.

Proposition 2.1 can be written in matrix format as follows:

$$E_{t-1}[r_t - r_t^f] = \beta_t E_{t-1}[f_t]. \tag{3.5}$$

$$\beta_t = COV_{t-1}(r_t, f_t) VAR_{t-1}(f_t)^{-1}, \tag{3.6}$$

where $COV_{t-1}(r_t, f_t)$ is the N by K combination of all the assets' factor covariance. Based on proposition 2.1, I model the asset returns as follows.

Factor Model Assumption

$$r_t - r_t^f = \beta_t f_t + \varepsilon_t \tag{3.7}$$

2a. f is exogenous: $E_{t-1}(f_t \epsilon_{it}) = 0$ for any i; and

2b. Idiosyncratic returns are independent cross assets: $E_{t-1}(\epsilon_{it}\epsilon_{jt}) = 0, i \neq j$.

The conditional factor setting in this section generalizes the static CAPM model to the extent that it allows for time-varying betas. Further, it is correctly specified if all the risk factors are added in the model. We then derive the conditional distribution of $r_{t|t-1}$:

Proposition 2.2 Conditional Distribution

$$r_{t|t-1} \sim N(\mu_t, H_t) \tag{3.8}$$

$$\mu_t = \beta_t E_{t-1}[f_t] + r_t^f \tag{3.9}$$

$$H_t = COV_{t-1}(r_t, f_t) VAR_{t-1}(f_t)^{-1} COV_{t-1}(r_t, f_t)' + VAR_{t-1}(\varepsilon_t) \tag{3.10}$$

$$\beta_t = COV_{t-1}(r_t, f_t) VAR_{t-1}(f_t)^{-1} \tag{3.11}$$

Proposition 2.2 indicates that my method is robust for combining with the enormous class of forecasting literature modeling $E_{t-1}[f_t]$. As a guide to practitioners, this method can have both the conditional mean and the variance forecast as its inputs for constructing the optimal portfolio. Because I emphasize conditional correlation, I treat the conditional mean as constants and directly work with demeaned returns.

In terms of the second moment, proposition 2.2 shows that the N by N dynamic covariance matrix can be shrunk to the N by K covariance matrix between asset returns and factors. Under assumption 2a and 2b, by decomposing the conditional covariance $H_t = D_t R_t D_t$, where D_t is the diagonal matrix of each asset's conditional volatility ("DeGarch"), I derive R_t such that I obtain the closed-form solution of its inverse and determinant. This allows me to construct the likelihood function, which, in turn, can be feasibly optimized.

Proposition 2.3 Conditional Correlation

Let $K_t = Corr_{t-1}(f_t)$ be the conditional correlation matrix among K factors. Let X_t be the N by K matrix written in partition form $X_t = [\rho'_{1,t} \cdots \rho'_{n,t}]$, where $\rho_{i,t}$ is a K by 1 vector representing the conditional correlation between r_i and each factor.

$$R_t = X_t K_t^{-1} X_t' + diag\{1 - \rho_{i,t}' K_t^{-1} \rho_{i,t}\} \tag{3.12}$$

$$\begin{aligned} R_t^{-1} &= diag\{\frac{1}{1 - \rho_{i,t}' K_t^{-1} \rho_{i,t}}\} \\ &- diag\{\frac{1}{1 - \rho_{i,t}' K_t^{-1} \rho_{i,t}}\} X_t (K_t + X_t' diag\{\frac{1}{1 - \rho_{i,t}' K_t^{-1} \rho_{i,t}}\} X_t)^{-1} X_t' diag\{\frac{1}{1 - \rho_{i,t}' K_t^{-1} \rho_{i,t}}\} \end{aligned} \tag{3.13}$$

$$|R_t| = \det(K_t + X_t' diag\{\frac{1}{1-\rho_{i,t}' K_t^{-1}\rho_{i,t}}\} X_t) \det(K_t^{-1}) \prod_i (1-\rho_{i,t}' K_t^{-1}\rho_{i,t}) \tag{3.14}$$

Proposition 2.3 works as a dynamic multi-factor extension of the toy model in the Introduction. If only a market risk exists, then the correlation among the factors is 1. This makes the construction of the likelihood function in this study trivially simple. The spirit of these tricks is to shrink an infeasible optimization problem, including the dynamic adjustment of an N by N matrix inverse, to a K by K case. Thus, we set a conditional factor model and show important tools for likelihood simplification. The next section introduces the estimation method based on these tools.

3.2.2 Estimation Method

Based on propositions 2.2 and 2.3, the estimation procedure is designed as follows:

Step 1: DCC Normalization

Set a joint DCC among the demeaned assets and factors:

$$\begin{aligned} r^1_{t|t-1} &= [r_{t|t-1} \quad f_{t|t-1}] \\ r^1_{t|t-1} &\sim N(0, D_t R^1_t D_t) \end{aligned} \tag{3.15}$$

The joint DCC is specified as follows:

$$
\begin{aligned}
&\text{GARCH for each } r_i^1\text{:}\\
&D_t^2 = diag\{w_i\} + diag\{\kappa_i\} \circ r_{t-1}r_{t-1}' + diag\{\lambda_i\} \circ D_{t-1}^2\\
&\text{DeGARCH:}\\
&\varepsilon_t^1 = D_t^{-1} r_t^1\\
&Q_t = S(1-\alpha-\beta) + \alpha\varepsilon_{t-1}^1 {\varepsilon_{t-1}^1}' + \beta Q_{t-1}\\
&R_t^1 = diag\{Q_t\}^{-1} Q_t diag\{Q_t\}^{-1}
\end{aligned}
\tag{3.16}
$$

Step 2: Reconstruct the Correlation

Partition the joint correlation R_t^1:

$$
R_t^1 = \begin{Bmatrix} R_t & X_t \\ X_t' & K_t \end{Bmatrix}
\tag{3.17}
$$

Where R_t and K_t are correlation matrixes among N assets and K factors, respectively, and $X_t = \{\rho_{i,k,t}\}_{N\times K} = [\rho_{1,t}' \cdots \rho_{n,t}']$ contains the correlations between each asset and factor. This partition allows me to construct the likelihood function with the factor model-implied correlation matrix R_t^L in lieu of R_t from the traditional DCC.

$$
R_t^L = X_t K_t^{-1} X_t' + diag\{1 - \rho_{i,t}' K_t^{-1} \rho_{i,t}\}
\tag{3.18}
$$

Step 3: Maximize the Likelihood

Construct the likelihood function based on R_t^L. The likelihood is generated by parameters S, α, β and, thus, optimized by the selection of these parameters.

$$R_t^{L^{-1}} = diag\{\frac{1}{1-\rho'_{i,t}K_t^{-1}\rho_{i,t}}\}$$
$$- diag\{\frac{1}{1-\rho'_{i,t}K_t^{-1}\rho_{i,t}}\}X_t(K_t + X'_t diag\{\frac{1}{1-\rho'_{i,t}K_t^{-1}\rho_{i,t}}\}X_t)^{-1}X'_t diag\{\frac{1}{1-\rho'_{i,t}K_t^{-1}\rho_{i,t}}\} \tag{3.19}$$

$$\det(R_t^L) = \det(K_t + X'_t diag\{\frac{1}{1-\rho'_{i,t}K_t^{-1}\rho_{i,t}}\}X_t)\det(K_t^{-1})\prod_i(1-\rho'_{i,t}K_t^{-1}\rho_{i,t}). \tag{3.20}$$

The correlation likelihood is given as follows:

$$L_c(\alpha,\beta,S) = -\frac{1}{2}\sum_{t=1}^{T}(-\varepsilon'_t\varepsilon_t + \log\det(R_t^L) + \varepsilon'_t R_t^{L^{-1}}\varepsilon_t). \tag{3.21}$$

Equations (3.18), (3.19), and (3.20) build a joint dynamic for R_t^L as a function of X_t. As mentioned in the introduction, the model presents the conditional correlation in a quadratic form as it estimates a simple linear forecast model. The estimator's asymptotic distribution can be derived similar to DCC, as shown in Appendix C.2.

3.2.3 Alternative Estimation Method: MacGyver

A joint likelihood estimate is more efficient, though it requires high computation power. To address this drawback, Engle (2009) proposed a "MacGyver" method to separately estimate $\frac{N(N-1)}{2}$ pairwise DCC, and then take the α and β from the average among all the pairwise α_i and β_i. Though this method requires lower computation power, it is not as efficient.

Similarly, one can always implement the method herein in an easier manner: In lieu of building the full matrix dynamic, separately estimate N by K pairs of the bivariate

DCC model to obtain each $\rho_{i,k,t}$, and then build R_t by (3.18):

$$\forall i, \forall k, V_{i,k} = [r_i; F_k], \text{build } n \times k \text{ pairs of DCC for } V_{i,k} \tag{3.22}$$

$$q_{i,k,t} = s_{i,k} + \alpha_{i,k}\varepsilon_{i,k,t-1}\varepsilon^F_{k,t-1} + \beta_i q_{i,k,t-1} \tag{3.23}$$

$$\rho_{i,k,t} = \frac{q_{i,k,t}}{\sqrt{\sigma^2_{i,t} h^F_{k,t}}} \tag{3.24}$$

$$X_t = \{\rho_{i,k,t}\}_{N\times K} = [\rho'_{1,t} \cdots \rho'_{n,t}] \tag{3.25}$$

$$R_t = X_t K_t^{-1} X_t' + diag\{1 - \rho'_{i,t} K_t^{-1} \rho_{i,t}\} \tag{3.26}$$

$$\forall k, \alpha_k = f(\alpha_{i,k}), \beta_k = f(\beta_{i,k}) \tag{3.27}$$

Like the MacGyver DCC, the estimates from this method have an unknown inference property. Rather than $\frac{N(N-1)}{2}$ pairs for the DCC, my model only requires N times K pairs of estimation.

3.3 Empirical Results

In this section, two datasets of the constituents returns from $S\&P500$ index are used to test the model performance for measuring correlation. The first is the 5-minutes price data from 2010 to 2017. After cleaning up the data, 377 stocks returns are available in the full sample. I use this high-frequency dataset to compute the realized correlation as the benchmark, and then compare it with the model-implied under certain loss functions. The second is the daily returns of $S\&P500$ index constituents from 2000 to 2018[1] , used to form out-of-sample portfolios.

I attempt to include some representative models as candidate estimators:

- DCC: original Dynamic Conditional Correlation Model
- DECO: Dynamic Equicorrelation Model, as in Engle and Kelly, 2012

[1]The original data is from 2000 to 2015, with 395 assets that with full history during this period. We then extend our sample by combining new data from 2015 to 2018, with 449 assets.

- DCC–NLS: DCC with nonlinear shrinkage, as in Engle et al. (2017)
- Model: The model proposed in this study; conditional factor structured.
- Model–NLS: The model proposed in this study with nonlinear shrinkage.

Without loss of generality, I use the $S\&P500$ index value-weighted return as the only factor. The trade-off between single and multiple factors is clear: More factors ensure the validity of exogeneity and independence among idiosyncratic returns. The single factor model is more parsimonious for estimation and, thus, less noisy.

3.3.1 Loss against Realized Correlation

Based on the high-frequency covariance matrix theory (see Barndorff-Nielsen and Shephard (2004)), I compute the realized correlation $R_t^{realized}$ and the average correlation of a matrix as follows:

$$R_t^{realized} = \frac{1}{n_t} \sum_{i=0}^{n_t} r_{t+i\Delta} r'_{t+i\Delta} \tag{3.28}$$

$$\bar{\rho}_t = \sum_{i \neq j} \sum_{j} R_t(i,j) \frac{2}{n(n-1)} \tag{3.29}$$

where n_t is the number of small-time intervals in each unit time.

I further define several loss functions to compare the accuracy of the model-implied correlation:

- Squared Error for the Average Correlation $\bar{\rho}_t$:

$$\overline{SE}_t = (\overline{\rho_t} - \overline{\rho_t^{realized}})^2. \tag{3.30}$$

- Mean Squared Error:

$$MSE_t = \frac{2}{N(N-1)} \sum_{i \neq j} \sum_{j} (R_t(i,j) - R_t^{realized}(i,j))^2. \tag{3.31}$$

- Mean Absolute Error:

$$MAE_t = \frac{2}{N(N-1)} \sum_{i \neq j} \sum_{j} \mid (R_t(i,j) - R_t^{realized}(i,j) \mid \tag{3.32}$$

- Quasi-Normal Likelihood Error:

$$QL_t = \det R_t - \log(\det R_t) - \det R_t^{realized} \tag{3.33}$$

R_t in these functions denotes the model-implied correlations and $R_t^{realized}$ is measured by 5-minute returns.

The loss of the models against the benchmark over time is reported in **Table 3.1**. Panel A shows the average loss function values of different models and Panel B reports the t-statistics of each model's loss against the DCC model. In general, a nonlinear shrinkage intercept does improve the accuracy of the models. The DCC–NLS and DECO models win over the original DCC under certain loss functions. My model and its combination with the nonlinear shrinkage has the smallest loss and wins over the original DCC model compared with the other models.

In addition, my model naturally yields a dynamic factor loading β_t. Like the partition of conditional correlation matrix, I derive β_t by decomposing the conditional covariance matrix $H_t^1 = VAR_{t-1}([r_t \quad f_t])$:

$$H_t^1 = \begin{Bmatrix} H_t & COV_{t-1}(r_t, f_t) \\ COV_{t-1}(r_t, f_t)' & VAR_{t-1}(f_t) \end{Bmatrix} \tag{3.34}$$

$$\beta_t = COV_{t-1}(r_t, f_t) VAR_{t-1}(f_t)^{-1} \tag{3.35}$$

Equation (3.34) eases the practical applications in the following sections.

3.3.2 Model Implication

Risk Management: Marginal Value at Risk

I now demonstrate how my method applies to risk management for allocating the value at risk of a portfolio. Unlike an allocation problem that determines the optimal weight, a risk management task takes the portfolio weights as given and analyzes the effect of altering asset positions marginally on the portfolio risk.

For any portfolio with N assets:

$$r_{pt} = \sum_{i=1}^{N} w_{it} r_{it} \tag{3.36}$$

The joint relationship between assets and the portfolio returns is depicted by a conditional factor model:

$$r_{it} = \beta_{it} r_{pt} + \varepsilon_{it} \tag{3.37}$$

$$\beta_{it} = COV_{t-1}(r_{it}, r_{pt}) VAR_{t-1}(r_{pt})^{-1} \tag{3.38}$$

This leads to:

$$\sum_{i=1}^{N} w_{it} \beta_{it} = 1, \forall t \tag{3.39}$$

Equation (3.39) allows us to link each asset's contribution to the total risk through

β_{it}. This argument falls in line with the definition of the marginal value at risk: $\Delta VaR_{it} = \frac{\partial VaR_{pt}}{\partial w_{it} W_t}$ is a partial derivative that measures how a dollar position adjustment shifts the total value at risk. A marginal value at risk (with confidence level θ) is then defined as follows:

$$\Delta VaR_{it} = \frac{\partial VaR_{pt}}{\partial w_{it} W_t} = Z_\theta \frac{\partial \sigma_{pt}}{\partial w_{it}} = Z_\theta \beta_{it} \sigma_{pt} \tag{3.40}$$

where Z_θ is the critical value for normal distribution, W_t is the total wealth, σ_{pt} is the conditional total volatility, and β_{it} is the dynamic portfolio risk loading. Equation (3.39) and (3.40) indicate that:

$$\frac{VaR_{pt}}{W_t} = Z_\theta \sigma_{pt} = \sum_{i=1}^{N} w_{it} \beta_{it} Z_\theta \sigma_{pt} = \sum_{i}^{N} w_{it} \Delta VaR_{it} \tag{3.41}$$

Equation (3.41) gives a clear decomposition of each asset's contribution to the portfolio's unite money value at risk. I take the $S\&P500$ 5-minutes data to construct a portfolio weighted by their market capital, and then estimate the joint correlation among all the stocks and the portfolio. The performance of the models is measured comparing it with the value at risk from the realized covariance matrix:

$$\begin{aligned} &\Delta VaR_{it}^{realized} = Z_\theta \beta_{it}^{realized} \sigma_{pt}^{realized} \\ &\beta_{it}^{realized} = COV_{t-1}(r_{it}, r_{pt})^{realized} \{VAR_{t-1}(r_{pt})^{realized}\}^{-1} \end{aligned} \tag{3.42}$$

I compute the marginal value at risk for each asset with my model, the combination of my model with nonlinear shrinkage and the factor DCC model with the MacGyver method (which also yields dynamic betas by pairwise estimation). Then, the capital-weighted square errors, absolute errors, are computed as follows:

$$
\begin{aligned}
MSE_t^{VaR} &= (\sum_i^N w_{it}\Delta VaR_{it} - \sum_i^N w_{it}\Delta VaR_{it}^{realized})^2 \\
MAE_t^{VaR} &= \mid \sum_i^N w_{it}\Delta VaR_{it} - \sum_i^N w_{it}\Delta VaR_{it}^{realized} \mid
\end{aligned}
\tag{3.43}
$$

The average error over time is reported in **Table 3.1**, Panel C. Unlike the loss measured in Panels A and B, the combination of nonlinear shrinkage improves the accuracy substantially. This is because I measure the average accuracy of the N asset dynamics in lieu of all the $\frac{N(N-1)}{2}$ elements in the correlation matrix. The benefit of a precise intercept estimation is often important for reliable practical work.

Portfolio Allocation

Based on the mean-variance utility developed in Markowitz (1952), the conditional mean and variance is specified as in proposition 2.2. I also assume a constant risk-free rate for convenience.

$$r_{t|t-1} \sim N(\mu_t, H_t) \tag{3.44}$$

$$\mu_t = \beta_t E_{t-1}[f_t] + r^f \tag{3.45}$$

$$H_t = COV_{t-1}(r_t, f_t) VAR_{t-1}(f_t)^{-1} COV_{t-1}(r_t, f_t)' + VAR_{t-1}(\varepsilon_t) \tag{3.46}$$

$$\beta_t = COV_{t-1}(r_t, f_t) VAR_{t-1}(f_t)^{-1} \tag{3.47}$$

By construction, my model can yield the first moment forecast by estimate $E_{t-1}[f_t]$ in lieu of the entire cross-section of expected returns $E_{t-1}[r_t]$. One can simply combine a risk premium forecast model with my method to derive better portfolios. Because I focus on forecasting the conditional variance, I simply take the historical average risk premium times dynamic factor loading as the expected return, $\hat{\mu}_t = \beta_t \overline{f_t}$. For all my models, because there is no dynamic factor loading, the expected return is given as historical means $\hat{\mu}_t = \overline{r_t}$.

Then, three target portfolios are defined as:

- Global Minimum Variance Portfolio:

$$\widehat{w_{GMV,t}} = \mathbf{argmin} w' H_t w \quad s.t \quad w'\vec{1} = 1 \tag{3.48}$$

$$w_{G\hat{M}V,t} = \frac{H_t^{-1}\vec{1}}{\vec{1}' H_t^{-1}\vec{1}} \tag{3.49}$$

- Minimum Variance Portfolio (with a required return q, set q = 0.1) (MV):

$$w_{\hat{M}V,t} = \mathbf{argmin} w' H_t w \quad s.t \quad w'\vec{1} = 1 \quad w'\mu \geqslant q. \tag{3.50}$$

$$w_{\hat{M}V,t} = \frac{C - qB}{AC - B^2} H_t^{-1}\vec{1} + \frac{qA - B}{AC - B^2} H_t^{-1}\mu_t \tag{3.51}$$

where $A = \vec{1}' H_t^{-1}\vec{1}, B = \vec{1}' H_t^{-1}\mu_t, C = \mu_t' H_t^{-1}\mu_t$.

- Tangency Portfolio:

$$w_{T\hat{G}C,t} = \arg\max w'(\mu_t - r_f) - \frac{1}{2}\gamma w' H_t w \quad s.t \quad w'\vec{1} = 1 \tag{3.52}$$

$$w_{T\hat{G}C,t} = \frac{H_t^{-1}(\mu_t - r_f)}{\vec{1}' H_t^{-1}(\mu_t - r_f)} \tag{3.53}$$

I then construct monthly updating portfolios for all the candidates listed before. The procedure is set as follows:

Step 1 : Take the end of t month's last day forecast μ_{t+1}, H_{t+1} to construct portfolios $w_{\widehat{GMV,t+1}}, w_{\widehat{MV,t+1}}, w_{\widehat{TGC,t+1}}$.

Step 2 : Pseudo out-of-sample test the performance and use the realized daily returns in the $t+1$ month (averagely 22 days per month) to compute all the portfolios returns, standard deviations, and Sharpe ratios.

Step 3: Include the $t+1$ month returns in the sample based on a 1,000-day rolling window. This yields the forecast μ_{t+2}, H_{t+2}, and then return to step 1.

Following this recursive pseudo-out-of-sample method, I report the average performance over time of the three portfolios generated by the models in **Table 3.2**. I find that, on average, my model and the model with nonlinear shrinkage produce lower volatility portfolios and higher information ratio/Sharpe ratio. Consistent with the literature, both DCC–NLS and DECO produce better results than DCC does.

3.4 Conclusions

The study applies the conditional factor model to multivariate covariance models. It combines both cross-section and time-series features of the covariance models to solve the high-dimensional curse.

To improve the methodology on mean-variance allocation, my method incorporates the literature on conditional mean forecasting. Under my framework, the expected return forecasting task is reduced to risk premium forecasting. For professions with strong intuitions on what creates a risk premium, the method is sound for practical use.

As shown in the empirical work, a single factor specification, conditional CAPM setting generates a robust in and out-of-sample covariance fit. However, in terms of asset pricing research, it is important to study further what factors drive correlation among assets with the proposed method. The model should be extended to a vaster class of risk factor literature as a test of this factor's power to create cross-sectional co-movement. My method thus opens a new channel to test conditional asset pricing models within the second moment.

This study emphasizes the method's performance with a single factor. What factors to be added to the model and to what extent can adding factors increase the prediction power of the model remain open questions for future research.

3.5 Acknowledgements

Chapter 3 is currently being prepared for submission for publication. Junxiong Gao, the dissertation author, is the only investigator and author of this paper.

Table 3.1. Average Loss of each model and T-stats against DCC

Panel A :Loss Under Different Measures					
	DCC	DCC_NLS	DECO	Model	Model_NLS
SE_t	0.010	0.010	0.010	0.007	0.007
MSE_t	0.078	0.078	0.109	0.065	0.065
MAE_t	0.222	0.222	0.222	0.199	0.199
QL_t	-74.2	-68.8	-71.72	-17.8	-18.4
Panel B: T-stats of Loss(DCC)-Loss(models)					
		DCC_NLS	DECO	Model	Model_NLS
SE_t		-9.1	1.7	41.0	39.2
MSE_t		-16.4	-55.9	40.4	42.8
MAE_t		1.5	0.91	50.5	53.4
QL_t		22.2	16.4	22.3	22.3
Panel C: Loss of Portfolio VaR					
			DCC(MacGyver)	Model	Model_NLS
MSE_t^{VaR}			0.29	0.24	0.89
MAE_t^{VaR}			0.49	0.47	0.12

In this table, I report the estimation of my model, my model combined with the non-linear shrinkage (Ledoit et al. (2012), Engle et al. (2017)), and variations of the DCC model.

- DCC: original Dynamic Conditional Correlation Model
- DECO: Dynamic Equicorrelation Model, as in Engle and Kelly, 2012
- DCC-NLS: DCC with nonlinear shrinkage, as in Engle et al. (2017)
- Model: The model proposed in this study; conditional factor structured.
- Model-NLS: The model proposed in this study with nonlinear shrinkage.

Table 3.2. Portfolio Allocation Results

Panel A :Global Minimum Variance Portfolio (GMV)					
	Deco	DCC	DCC_NLS	Model	Model_NLS
Mean %	20.1	22.7	21.6	15.6	15.4
Std %	15.2	13.8	12.5	11.8	11.7
IR	0.32	0.59	0.68	0.99	0.98
Panel B :Minimum Variance Portfolio(MV)					
	Deco	DCC	DCC_NLS	Model	Model_NLS
Mean %	24.1	21.5	19.9	14.8	12.2
Std %	14.7	13.4	12.1	12.7	11.3
IR	0.68	0.87	0.96	1.43	1.51
Panel C :Tagency Portfolio(TGC)					
	Deco	DCC	DCC_NLS	Model	Model_NLS
Mean %	-20.2	-36.9	-20.9	11.0	11.2
Std %	36.3	112.3	28.4	15.4	15.6
IR	0.6	0.012	0.88	1.45	1.45

In this table, I report the estimation of my model, my model combined with the non-linear shrinkage (Ledoit et al. (2012), Engle et al. (2017)), and variations of the DCC model.

- •DCC: original Dynamic Conditional Correlation Model
- •DECO: Dynamic Equicorrelation Model, as in Engle and Kelly, 2012
- •DCC-NLS: DCC with nonlinear shrinkage, as in Engle et al. (2017)
- •Model: The model proposed in this study; conditional factor structured.
- •Model-NLS: The model proposed in this study with nonlinear shrinkage.

Appendix A

Appendix for Chapter 1

A.1 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 $\Sigma^{\boldsymbol{n}}$ into two parts, strong covariance from factors $\Sigma_{\boldsymbol{f}}^{\boldsymbol{n}}$ and covariance among idiosyncratic risk $\Sigma_{\epsilon}^{\boldsymbol{n}}$. 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:*

$$\lim_{n\to\infty} \sum_{i}^{n} w_i^2 = 0. \tag{A.1}$$

$\sum_i^n w_i^2$ 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 1.2.1** in matrix form and present the derivation of APT. There are n firms in the whole asset space; each has a return:

$$\boldsymbol{r} = E[\boldsymbol{r}] + \boldsymbol{B}\boldsymbol{f} + \boldsymbol{\epsilon}, \tag{A.2}$$

$$E[\boldsymbol{\epsilon}|\boldsymbol{f}] = 0. \tag{A.3}$$

this leads to a variance decomposition:

$$\Sigma^{\boldsymbol{n}} = \boldsymbol{B}\Sigma_{\boldsymbol{f}}^{\boldsymbol{n}}\boldsymbol{B}' + \Sigma_{\epsilon}^{\boldsymbol{n}}. \tag{A.4}$$

the term $\boldsymbol{B}\Sigma_{\boldsymbol{f}}^{\boldsymbol{n}}\boldsymbol{B}'$ 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_f^n *have a factor structure if:*

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

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:*

$$\lim_{n\to\infty} \rho_i(\Sigma^{\boldsymbol{n}}) \leq C, \forall i. \tag{A.6}$$

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:

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

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

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

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

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

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:

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

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:

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

The covariance term $COV(\epsilon_i, \sum_i^n w_i\epsilon_i)$ in (1.3) is stacked into the vector $\Sigma_{\epsilon}^{\boldsymbol{n}}\boldsymbol{w}$. The market risk premium is:

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

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

Proof of Lemma 2: With diversification,

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

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

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

A.2 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,

$$\lim_{n\to\infty}\sum w_i^2 = \lim_{n\to\infty}\sum\frac{(X_i)^2}{(\sum X_i)^2} = \lim_{n\to\infty}\frac{1}{n}\frac{1/n\sum(X_i)^2}{(1/n\sum X_i)^2}. \tag{A.14}$$

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:

$$P(X_i > x) = x^{-\zeta}, x > 1 \tag{A.15}$$

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 6. *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 6: 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 $S_n = \sum_{i=1}^{n} X_i$
$a_n = \inf\{x : P(|X_1| > x) \leq n^{-1}\}$ and $b_n = nE(X_1 1_{|X_1| \leq a_n})$
As $n \to \infty, (S_n - b_n)/a_n \Rightarrow Y$ where Y has a 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 (A.15), $\theta = 1$, $\alpha = \zeta$ 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:

$$b_n = \begin{cases} n\left(n^{1/\zeta - 1} - \dfrac{\zeta}{1-\zeta}\right) \approx n^{1/\zeta} = a_n & \zeta < 1 \\ n\left(n^{1/\zeta - 1} - \dfrac{\zeta}{1-\zeta}\right) \approx n\dfrac{\zeta}{\zeta - 1} = nE[X] & \zeta > 1 \\ n \log n & \zeta = 1 \end{cases} \tag{A.16}$$

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,

$$\lim_{n\to\infty} \sum X_i = \begin{cases} \lim_{n\to\infty} n^{1/\zeta}(Y_\zeta + 1) & \zeta < 1 \\ \lim_{n\to\infty} Y_\zeta + \log n & \zeta = 1 \\ \lim_{n\to\infty} n^{1/\zeta} Y_\zeta + nE[X] & \zeta > 1 \end{cases} \tag{A.17}$$

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:

$$\begin{aligned} w_\zeta(t) &= \tan(\pi\zeta/2), \zeta \neq 1 & \text{(A.18)} \\ &= \pi/2 \log|t|, \zeta \neq 1 & \text{(A.19)} \end{aligned}$$

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**:

$$\lim_{n\to\infty} w_{\max} = X_{\max} / \sum_{i=1}^{n} X_i = \begin{cases} \frac{F_\zeta}{Y_\zeta + 1} & \xi < 1 \\ \lim_{n\to\infty} \frac{F_\zeta}{Y_\zeta + \log n} & \xi = 1 \\ \lim_{n\to\infty} \frac{F_\zeta}{Y_\zeta + n^{1-1/\zeta} E[X]} & \xi > 1 \end{cases} \tag{A.20}$$

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 7. *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 $X_{\min} = \min\{X_{1,\dots,n}\}$ 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 7: 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:

$$\lim_{n\to\infty} w_{\min} = X_{\min} / \sum_{i=1}^{n} X_i = \begin{cases} \lim\limits_{n\to\infty} \dfrac{1 + \exp_\zeta / n}{n^{1/\zeta}(Y_\zeta + 1)} & \xi < 1 \\ \lim\limits_{n\to\infty} \dfrac{1 + \exp_\zeta / n}{Y_\zeta + \log n} & \xi = 1 \\ \lim\limits_{n\to\infty} \dfrac{1 + \exp_\zeta / n}{n^{1/\zeta} Y_\zeta + nE[X]} & \xi > 1 \end{cases} \tag{A.21}$$

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:

$$\lim_{n\to\infty} \sum X_i^2 = \lim_{n\to\infty} n^{2/\zeta}(Y_{\zeta/2} + 1). \tag{A.22}$$

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 (A.17) and (A.22) gives the convergence of $\sum_i^n w_i^2$:

$$\lim_{n\to\infty} \sum w_i^2 = \begin{cases} \dfrac{Y_{\zeta/2}+1}{(Y_\zeta+1)^2} & \zeta < 1 \\ \lim\limits_{n\to\infty} \dfrac{Y_{\zeta/2+1}}{(Y_\zeta + \log n)^2} & \zeta = 1 \\ \lim\limits_{n\to\infty} \dfrac{Y_{\zeta/2}+1}{(Y_\zeta + n^{1-1/\zeta}E[X])^2} & \zeta > 1 \end{cases} \tag{A.23}$$

The proof of **Proposition 1** and **Proposition 2** is derived by (A.11) and (A.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 1** and **Proposition 2**.

A.3 Estimation of 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

$X_{i=1...n}$ are i.i.d and follows the exact Pareto distribution in (1.7) 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 $X_{i=1,...,n}$ and select a threshold value $x_m = X_k$ to use the largest k firms for estimating ζ. The Hill estimator is:

$$\zeta = \left\{1/k\sum_{i=1}^{k}(\log X_i - \log X_k)\right\}^{-1}. \tag{A.24}$$

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 $k/n = 5\%, 10\%...$ 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[1]. Specifically, matching the survival probability in (1.7) with the frequency in data gives,

[1]This assumption should not affect the theoretical results in **Section 1.2.2** 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.

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

The logarithm of this equation implies a linear relationship between logged rank i and size X_i in (1.7) 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 A.1**. 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 (1.13):

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

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 generates a downward bias of ζ since it could include small firms in the sample that may not follow 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 A.2**. 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 A.2** 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 A.3**. 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 (1.13) to estimate:

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

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 et al. (1998), Peng (1998), Beirlant et al. (1999), Feuerverger and Hall (1999), Gomesa and Martins (2002), Alves et al. (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.

A.4 Out-of-sample predictive results

I check the out-of-sample predictive power of my model and report results in **Table A.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.

A.5 Additional figures and tables

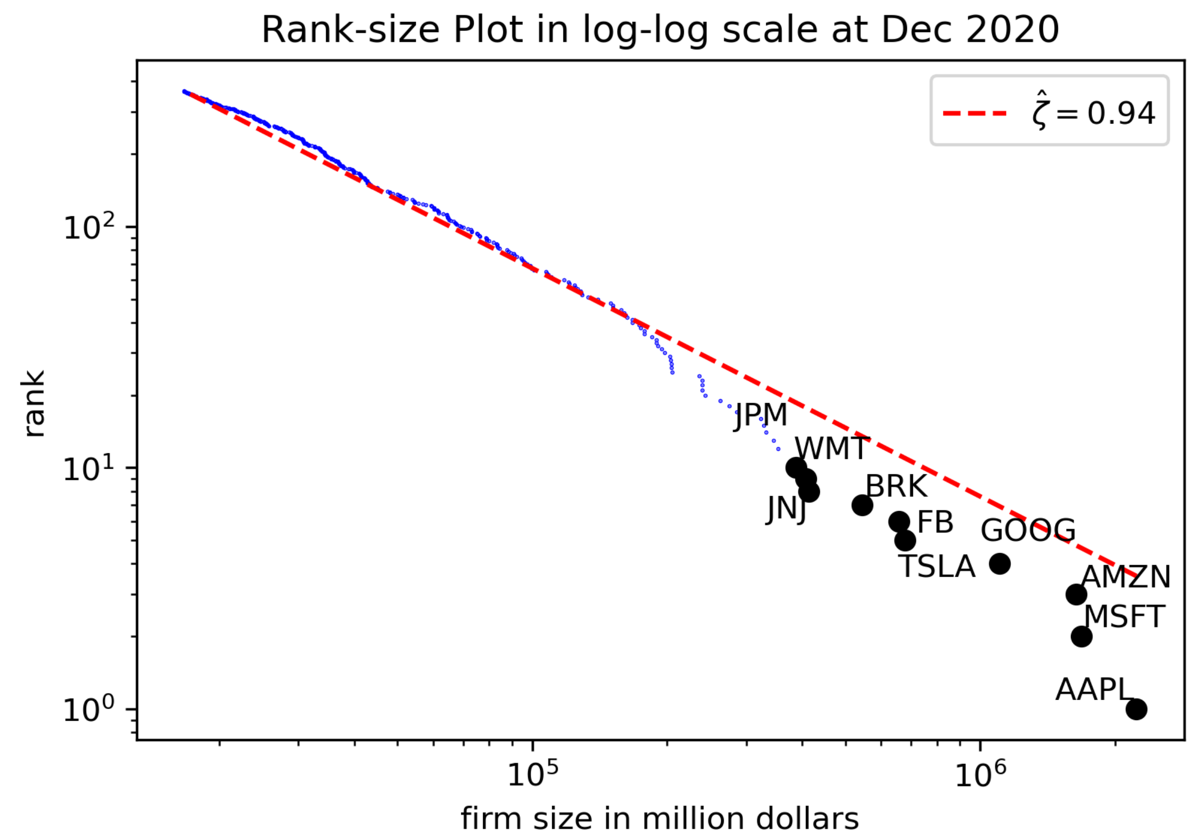


Figure A.1. Logged rank-size plot in December 2020

In this figure, I plot the logged rank-size plot of the largest 10% firms in December 2020. The red dashed line shows the fitted relation implied by the Pareto distribution. The ten largest firms are highlighted.

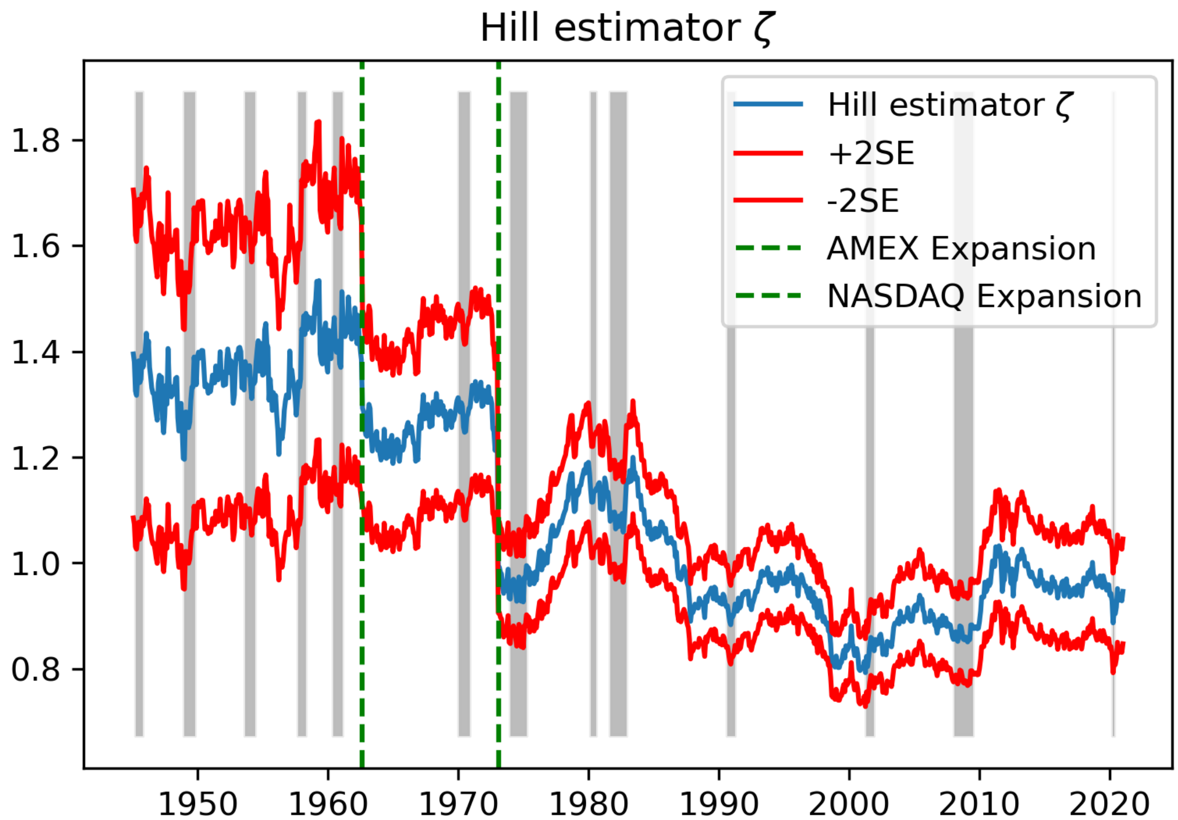


Figure A.2. Pareto Coefficient Estimate of Market Value per Month

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 ten % firms to illustrate a trade-off between bias and variance of the Hill estimator. 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. The two vertical dash lines in the plot mark the expansion of n due to the merging of security exchanges: AMEX in June 1962 and NASDAQ in January 1973. The shaded areas are NBER recession periods.

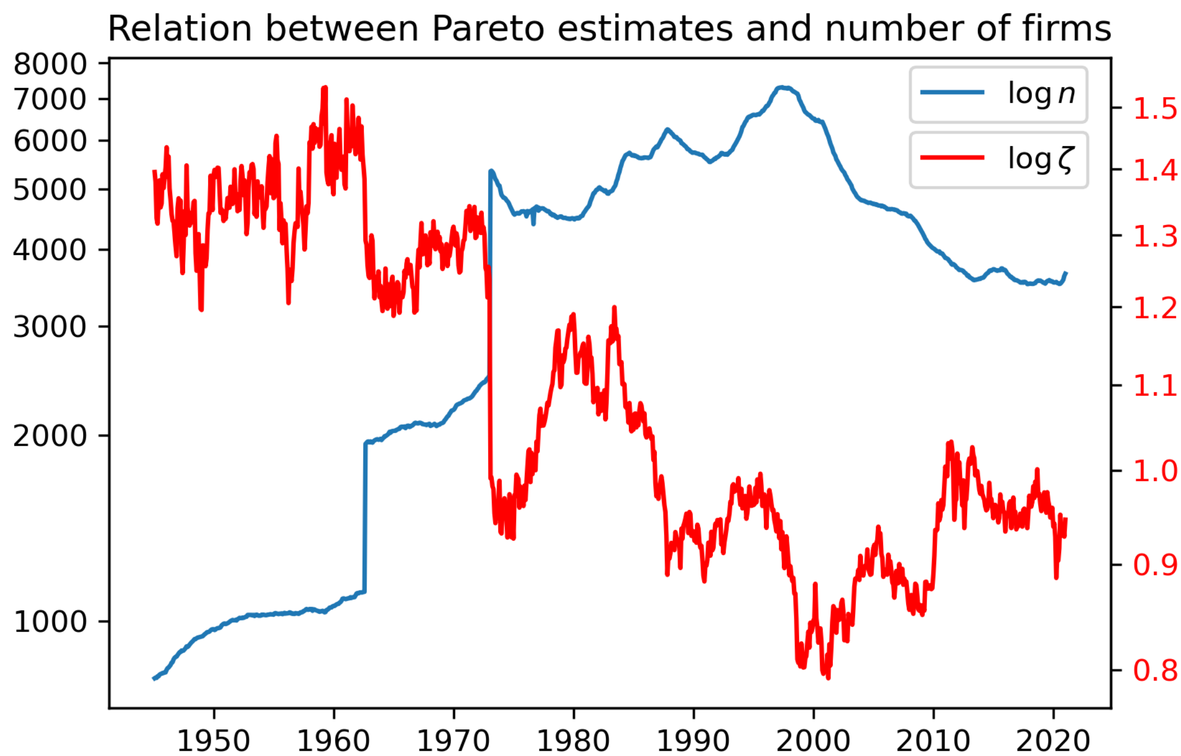


Figure A.3. Pareto Coefficient Estimate and Number of Firms

I plot the co-integration relation between the logged Pareto coefficient $\log \zeta$ (estimated from the largest ten % firms) and the logged number of firms n. Both the time series are normalized to zero-mean and unit-variance with their raw magnitudes displayed on two separate sets of ticks on the y-axis.

Table A.1. Out-of-sample prediction results.

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**

I summarize the out-of-sample predictive power of all the sets of predictors I test. 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 outperforms the historical mean.

Appendix B

Appendix for Chapter 2

B.1 Derivation of the linearized identities

In this section, we show the derivation of linearized identities for the three sectors.

$$DA_t = \kappa_{0,DA} + \kappa_{1,DA} \sum_{j=0}^{\infty} \rho_{DA}^{j} \left(r_{t+j+1}^{E} - \Delta d_{t+j+1} \right) \tag{B.1}$$

$$NSB_t = \kappa_{0,NSB} + \kappa_{1,NSB} \sum_{j=0}^{\infty} \rho_{NSB}^{j} \left(r_{t+j+1}^{D} - \Delta ns_{t+j+1} \right) \tag{B.2}$$

$$NMA_t = \kappa_{0,NMA} + \kappa_{1,NMA} \sum_{j=0}^{\infty} \rho_{NMA}^{j} \left(r_{t+j+1}^{X} - \Delta nm_{t+j+1} \right) \tag{B.3}$$

In addition, these three linear identities are weighted-summed to derive an aggregate identity of the economy such that

$$CW_t = \kappa_{0,CW} + \kappa_{1,CW} \sum_{j=0}^{\infty} \rho_{CW}^{j} \left(r_{t+j+1}^{W} - \Delta c_{t+j+1} \right) \tag{B.4}$$

We further show that the aggregate budget constraint is approximately equal to a weighted average of each sector budget constraint. Specifically,

$$CW_t = \alpha_0 + w_1 DA_t + w_2 NSB_t + w_3 NMA_t,$$

$$r_t^W = \alpha_1 + w_1 r_t^E + w_2 r_t^D + w_3 r_t^X.$$

And according to the budget constraint, the cash-flow growth also follows,

$$\Delta c_t = \alpha_2 + w_1 \Delta d_t + w_2 \Delta ns_t + w_3 \Delta nm_t.$$

We derive the weight of each sector $w_{1,2,3}$ based on our calculation and further use it to impose the constraint in our estimation.

B.1.1 The Aggregate Accounting Identity

The accounting identity of GDP is:

$$Y_t = C_t + I_t + G_t + X_t - M_t \tag{B.5}$$

which leads to a decomposition of consumption:

$$C_t = Y_t - I_t - T_t + T_t - G_t + M_t - X_t \tag{B.6}$$

Define the net surplus as:

$$NS_t = T_t - G_t \tag{B.7}$$

and the net import as:

$$NM_t = M_t - X_t \tag{B.8}$$

$$C_t = Y_t - I_t - T_t + NS_t + NM_t \tag{B.9}$$

Typical finance literature like Larraine and Yogo (2008) take the firm as production technology and its after-tax revenue as a dividend (or total payout to the household), i.e. $D_t = Y_t - I_t - T_t$. This leads to:

$$C_t = D_t + NS_t + NM_t \tag{B.10}$$

Equation (B.10) is an aggregate budget constraint for the whole economy. It decomposes the total consumption into flows of three sectors: The private sector (financed by the after-tax revenue D_t), the Public sector (financed by the government surplus $NS_t = T_t - G_t$), and external sector adjustment.

The aggregate budget constraint wraps budget constraints for the three sectors together:

Budget constraint for the private sector is:

$$A_{t+1} + D_{t+1} = (1 + R^A_{t+1})A_t \tag{B.11}$$

For the public sector is:

$$B_{t+1} + NS_{t+1} = (1 + R^B_{t+1})B_t \tag{B.12}$$

For the external sector is:

$$F_{t+1} + NM_{t+1} = (1 + R^X_{t+1})F_t \tag{B.13}$$

The stock value of the three sectors is also linked together as in (B.10) if we define the total domestic wealth of economy W_t as:

$$W_t = A_t + B_t + F_t \tag{B.14}$$

One can treat the total asset value as net wealth, matched to the equity value of each firm. The total wealth, as accumulated value produced by the economy, is kept by firms, borrowed by the government, or traded out for foreign assets. Therefore, the total wealth equals the sum of domestic asset value, government debt, and foreign asset value.

The total wealth W_t follows a similar dynamic to each sector, with total outflows C_t:

$$W_{t+1} + C_{t+1} = (1 + R^W_{t+1})W_t \tag{B.15}$$

Identity defined in equation (B.10) and (B.14) allow us to express the total wealth dynamic (B.15) by the sum of each sector. The total wealth return R^W, as a portfolio return includes the three classes of assets, is the weighted average of the three sectors' return.

$$1 + R^W_{t+1} = (1 + R^A_{t+1})A_t/W_t + (1 + R^B_{t+1})B_t/W_t + (1 + R^X_{t+1})F_t/W_t \tag{B.16}$$

The portfolio return relation wraps each sectors budget constraint into aggregate:

$$1 + R^W_{t+1} = \frac{W_{t+1} + C_{t+1}}{W_t} \tag{B.17}$$

$$= \frac{A_{t+1} + D_{t+1}}{A_t}A_t/W_t + \frac{B_{t+1} + S_{t+1}}{B_t}B_t/W_t + \frac{F_{t+1} + NM_{t+1}}{F_t}F_t/W_t \tag{B.18}$$

Based on this intuition, we apply the linearization technique to each sector and the aggregate account and then show that the aggregate identity is a weighted average of the three sectors.

B.1.2 Separate Budget Constraints

We now linearize the sector budget constraint. Unlike the equity sector case solved by Campbell and Shiller (1988a) (CS hereafter), the cash flows considered in our setting could be negative, i.e. when the government is in deficit or the external sector has a net export. Therefore, a logged flow-value ratio would not be available and hence we develop a comparable technique to handle the negative cash flow.

For saving the notation, we express all the sector budget constraints in a general form,

$$V_t(1+R_{t+1}) = V_{t+1} + CF_{t+1}.$$

A sector has the cumulative account value V_t and cash flow CF_t, and its return is defined by,

$$1+R_{t+1} = \frac{V_{t+1}+CF_{t+1}}{V_t} = \frac{V_{t+1}+CF_{t+1}}{CF_{t+1}}\frac{CF_{t+1}}{CF_t}\frac{CF_t}{V_t} = (1+\frac{1}{CFV_{t+1}})\Delta CF_{t+1}CFV_t. \tag{B.19}$$

We denote the un-logged cash flow growth by $\Delta CF_{t+1} = \frac{CF_{t+1}}{CF_t}$ and the cash flow value ratio by $CFV_t = \frac{CF_t}{V_t}$. Taking log for both sides of the equation gives:

$$r_{t+1} = \log(\frac{V_{t+1}+CF_{t+1}}{V_t}) = \log\left((1+\frac{1}{CFV_{t+1}})\Delta CF_{t+1}CFV_t\right) \tag{B.20}$$

We approximate the right side equation as a function of CFV_{t+1},ΔCF_{t+1},CFV_t by the first-order Taylor expansion around a steady state. This presentation is valid in the sense it ensures the product of the three terms to be positive as long as $V_t > 0$ and $V_t + CF_t > 0$ at any time t. Let us present the function $\log(\frac{V_{t+1}+CF_{t+1}}{V_t})$ in a general form,

$$f(X_1, X_2, X_3) = \log((1 + \frac{1}{X_1})X_2 X_3).$$

Taking first-order derivative and combining with the fact that $\Delta CF = 1$ and the two valuation ratio equals the same CFV at steady state leads to:

$$\log(\frac{V_{t+1} + CF_{t+1}}{V_t}) \approx \log(1 + CFV) - \frac{V}{V + CF} CFV^{-1}(CFV_{t+1} - CFV) + (\Delta CF_{t+1} - 1) + CFV^{-1}(CFV_t - CFV) \tag{B.21}$$

Using the approximate relation that $\log(1 + x) \approx x$, we define the simple cash flow growth mapped to the log cash flow growth in CS:

$$\Delta cf_{t+1} = \Delta CF_{t+1} - 1 = \frac{CF_{t+1}}{CF_t} - 1 \tag{B.22}$$

Let $\rho = \frac{V}{V+CF} = 1/(1 + CFV)$, we can write:

$$r_{t+1} = \kappa - \rho(CFV)^{-1} CFV_{t+1} + \Delta cf_{t+1} + (CFV)^{-1} CFV_t \tag{B.23}$$

where the constant term is

$$\kappa = \log(1 + CFV) + \rho - 1 \tag{B.24}$$

At steady state, the logged expected return equals

$$r = \log(1 + CFV) + \rho - 1 - \rho + 1 = \log(1 + CFV) \tag{B.25}$$

Re-organize the term and makes

$$(CFV)^{-1}CFV_t = -\kappa + r_{t+1} - \Delta cf_{t+1} + \rho(CFV)^{-1}CFV_{t+1} \tag{B.26}$$

or we can write

$$CFV_t = -\kappa CFV + CFV\left(r_{t+1} - \Delta cf_{t+1}\right) + \rho CFV_{t+1}. \tag{B.27}$$

Solving this equation $K = 2,...,\infty$ periods in forward gives

$$CFV_t = \kappa_0 + \kappa_1 \sum_{j=0}^{K-1} \rho^j \left(r_{t+j+1} - \Delta cf_{t+j+1}\right) + \rho^K (CFV)^{-1} CFV_{t+K}. \tag{B.28}$$

κ_0 and κ_1 are constants computed from the steady state CFV. In the long-horizon, when K approached zeros, and given a positive steady state CFV such that $\rho < 1$,

$$CFV_t = \kappa_0 + \kappa_1 \sum_{j=0}^{\infty} \rho^j \left(r_{t+j+1} - \Delta cf_{t+j+1}\right). \tag{B.29}$$

In the main body of the paper, we apply this general form of linear identity to the three sectors and aggregate consumption-wealth relations. Accordingly, we use the notation $\kappa_{0,DA}$, $\kappa_{1,DA}$ and ρ_{DA} to present the constant in the linearization of DA_t, and r^E, Δd for the return and cash-flow growth in the equity sector. And, we use a similar notation for the other budget constraints.

B.1.3 Aggregate Budget Constraint as a Weighted Average of Separate Budget Constraints

As we showed in the first section, the aggregate budget constraint is a sum of all the sector budget constraints, we now show that the linearized A.B.C is approximately a weighted average of all the linearized B.C. As shown in (B.16), the aggregate account return r^W is a weighted average of each sector. We start with using a general format to present this relation here to save notations. The aggregate portfolio contains sectors $i = 1...n$ with each sector's return, cash flow, and stock value denoted by R_i, CF_i, V_i, respectively. Each sector has a budget constraint, which is approximated by:

$$r_{i,t+1} = \Delta cf_{i,t+1} - \rho_i (CFV_i)^{-1} CFV_{i,t+1} + (CFV_i)^{-1} CFV_{i,t} \tag{B.30}$$

The total portfolio return under this general setup equals:

$$1 + R^W_{t+1} = \sum_i^n V_{i,t}/W_t (1 + R_{i,t+1})$$

Mapping with the general solution in the last subsection, the consumption wealth dynamic in (B.15) is approximated by:

$$r^w_{t+1} = \Delta c_{t+1} + (CW)^{-1} CW_t - \rho (CW)^{-1} CW_{t+1} \tag{B.31}$$

We proxy the aggregate log return r^W_{t+1} as a weighted average of each log return using a textbook log-linearization technique in the macroeconomic literature.

$$\frac{1 + R^w_{t+1}}{1 + R^w} - 1 = \sum_i^n \frac{VW^i (1 + R_i)}{\sum VW^i (1 + R_i)} \left(\frac{1 + R_{i,t+1}}{1 + R_i} - 1 \right) \tag{B.32}$$

Therefore, we define the weight

$$w_i = \frac{VW_i (1 + R_i)}{\sum VW_i (1 + R_i)},$$

and use the approximate relation $\log(1 + x) \approx x$ for a small x to show that,

$$r^w_{t+1} \approx \text{constant} + \sum_i^n w_i r_{i,t+1}. \tag{B.33}$$

Similarly, for the consumption wealth ratio CW_t, we decompose the consumption as follows,

$$C_t = \sum CF_{i,t}.$$

Dividing this equation by W_t, we have

$$CW_t = \sum CFV_{i,t} \frac{V_{i,t}}{W_t}.$$

Use the steady state condition in (B.25), we can show,

$$1 + R_i = 1 + CFV_i = 1/\rho_i. \tag{B.34}$$

Consequently,

$$(CW)^{-1} CW_t = \text{constant} + \sum w_i (CFV^i)^{-1} CFV_t^i. \tag{B.35}$$

At last, we combine the relation between asset returns and the linearized identity.

$$r_{t+1}^W = \sum w_i r_{i,t+1} = \sum w_i \left(\Delta cf_{i,t+1} - \rho^i (CFV_i)^{-1} CFV_{i,t+1} + (CFV_i)^{-1} CFV_{i,t} \right) \tag{B.36}$$

Matching terms in this equation with the A.B.C gives,

$$\Delta c_{t+1} = \text{constant} + \sum w_i \Delta cf_{t+1}^i. \tag{B.37}$$

Therefore, our derivation in this section shows that the aggregate budget constraint is a weighted average of each sector by solving the relation between the aggregate portfolio return r^W and returns in each sector. Mapping with our three-sector setting, the weights are

$$w_1 = \frac{AW/\rho_{DA}}{AW/\rho_{DA} + BW/\rho_{NSB} + FW/\rho_{NMA}},$$

$$w_2 = \frac{BW/\rho_{NSB}}{AW/\rho_{DA} + BW/\rho_{NSB} + FW/\rho_{NMA}},$$

$$w_3 = \frac{FW/\rho_{NMA}}{AW/\rho_{DA} + BW/\rho_{NSB} + FW/\rho_{NMA}}.$$

Intuitively, the weights are each account's weight in the aggregate wealth at the steady state, adjusted by a linearization coefficient ρ. These calculation showed that the log-linearized A.B.C is a weighted average of log-linearized B.C.

B.1.4 Combine the Foreign Asset and Liability

The linearization we develop in the above is very general yet it requires a positive account value V_t and a well-defined log return such that $1 + R_{t+1} = \frac{V_{t+1}+CF_{t+1}}{V_t} > 0$. We find that this assumption does not hold for the external sector with the return equals,

$$1 + R^x_{t+1} = \frac{F_{t+1} + NM_{t+1}}{F_t}.$$

The net foreign asset value F equals to the difference between foreign asset A^x and liability L^x. During certain period, the net foreign asset $F_t = A^x_t - L^x_t$ is close to zero and the net import is negative, which drives the gross return to be extremely negative so that we cannot define a log return.

We handle this problem by separately approximating the dynamic of foreign assets and liability:

$$A^x_t(1 + R^{ax}_{t+1}) = A^x_{t+1} + M_{t+1} \tag{B.38}$$

$$L_t^x(1+R_{t+1}^{lx}) = L_{t+1}^x + X_{t+1} \tag{B.39}$$

Given $F_t = A_t^x - L_t^x$, we use the portfolio conclusion in the last section to derive a proxy for the external sector B.C. as a weighted average of the B.C. for the asset and liability, where we assign the weight to be:

$$\varphi^a = \frac{VW^{ax}(1+R^{ax})}{VW^{ax}(1+R^{ax}) - VW^{al}(1+R^{al})} \tag{B.40}$$

$$\varphi^l = \frac{VW^{al}(1+R^{al})}{VW^{ax}(1+R^{ax}) - VW^{al}(1+R^{al})} \tag{B.41}$$

These exercise allows us to directly work on the net foreign asset as a single sector. We thus derive the separate linear identity for each sector, and the aggregate wealth of the economy.

B.2 Empirical Setting

B.2.1 Sector Budget Constraints

Using the technique in the last section, we first approximate each sector's budget constraint:

$$(DA)^{-1}DA_t = r_{t+1}^A - \Delta d_{t+1} + \rho^A (DA)^{-1} DA_{t+1} \tag{B.42}$$

$$(NSB)^{-1}NSB_t = r_{t+1}^B - \Delta ns_{t+1} + \rho^B (NSB)^{-1} NSB_{t+1} \tag{B.43}$$

$$(NMA)^{-1}NMA_t = r_{t+1}^X - \Delta nm_{t+1} + \rho^X (NMA)^{-1} NMA_{t+1} \tag{B.44}$$

where the notation for each sector maps with (B.1),(B.2),(B.3).

Due to the steady state relation in (B.34), we estimate the value of DA, NSB, and NMA and ρ^A, ρ^B, ρ^X using the average log return for each sector $\bar{r}^i$:

$$CFV_i = exp(\bar{r}_i) - 1 \tag{B.45}$$

$$\rho_i = 1/(1 + CFV_i) \tag{B.46}$$

The A.B.C is:

$$(CW)^{-1}CW_t = r^w_{t+1} - \Delta c_{t+1} + \rho(CW)^{-1}CW_{t+1} \tag{B.47}$$

which is a weighted average of the three-sector B.C. We estimate the weight by the average log return for each sector $\bar{r}^i$ and average log V/W ratio $v\bar{w}^i$:

$$w_i = \frac{e^{v\bar{w}_i}e^{\bar{r}_i}}{\sum e^{v\bar{w}_i}e^{\bar{r}_i}} \tag{B.48}$$

B.2.2 A Predictive Regression System

We now set a predictive regression framework in matrix format. Define:

$$\boldsymbol{ratio_t} = [DA_t; NSB_t; NMA_t]$$

as a column vector and similarly:

$$\boldsymbol{r_t} = [r^A_t; r^B_t; r^X_t]$$

$$\Delta \boldsymbol{cf_t} = [\Delta d_t; \Delta ns_t; \Delta nx_t]$$

We then have:

$$\boldsymbol{r}_{t+1} = \boldsymbol{\alpha_r} + \boldsymbol{b_r ratio_t} + \boldsymbol{\epsilon}^{\boldsymbol{r}}_{t+1} \tag{B.49}$$

$$\Delta \boldsymbol{cf}_{t+1} = \boldsymbol{\alpha}_{cf} + \boldsymbol{b}_{cf}\boldsymbol{ratio_t} + \boldsymbol{\epsilon}^{cf}_{t+1} \tag{B.50}$$

$$\Delta c_{t+1} = \alpha_c + \boldsymbol{b}'_c \boldsymbol{ratio_t} + \boldsymbol{\epsilon}^{c}_{t+1} \tag{B.51}$$

$$\boldsymbol{ratio}_{t+1} = \boldsymbol{c} + \boldsymbol{\phi ratio_t} + \boldsymbol{\epsilon}^{ratio}_{t+1} \tag{B.52}$$

The sector b.c. says for each sector:

$$1 = CFV^i(b^i_r - b^i_{cf}) + \rho^i \phi^i \tag{B.53}$$

in matrix format, we have the sector B.C:

$$\boldsymbol{I} = \boldsymbol{diag}(\boldsymbol{DA}, \boldsymbol{NSB}, \boldsymbol{NMA})(\boldsymbol{b_r} - \boldsymbol{b}_{cf}) + \boldsymbol{diag}(\boldsymbol{\rho^A}, \boldsymbol{\rho^B}, \boldsymbol{\rho^X})\boldsymbol{\phi} \tag{B.54}$$

Same as we did in the unconstrained case, we estimate the $\boldsymbol{b}'_c$ by running three regressions and impose

$$\boldsymbol{b}'_c = \boldsymbol{\varphi}'\boldsymbol{b}_{CF} \tag{B.55}$$

Using the multi-period relation, if we run the sum of K-period ahead return and cash flow growth on the ratios, we can get similar constraints:

$$\boldsymbol{I} = \boldsymbol{diag}(\boldsymbol{DA}, \boldsymbol{NSB}, \boldsymbol{NMA})(\boldsymbol{b_r}(\boldsymbol{k}) - \boldsymbol{b}_{cf}(\boldsymbol{k})) + \boldsymbol{diag}(\boldsymbol{\rho^A}, \boldsymbol{\rho^B}, \boldsymbol{\rho^X})^K \boldsymbol{\phi}(\boldsymbol{k}) \tag{B.56}$$

B.3 Results of Forecasting Nominal and Excess Returns

Table B.1. Predictive Regression for Nominal and Excess Returns, Single-Variable

Panel A: Single Variable Predictive Regression, One-period ahead								
	Nominal	Excess		Nominal	Excess		Nominal	Excess
	r^E_{t+1}	r^E_{t+1}		r^B_{t+1}	r^B_{t+1}		r^X_{t+1}	r^X_{t+1}
DA_t	0.80	0.54	NSB_t	0.04	-0.03	NMA_t	0.12	0.12
	2.28	1.48		1.26	-0.96		2.47	2.32
R^2	0.02	0.01	R^2	0.00	0.00	R^2	0.04	0.03
unbiased	0.46	0.19	unbiased	0.04	-0.03	unbiased	0.12	0.11
	1.13	0.46		0.90	-0.78		3.08	2.92
$\tilde{R}^2$	0.01	0.00	$\tilde{R}^2$	0.00	0.00	$\tilde{R}^2$	0.03	0.03
Panel B: Single Variable Predictive Regression, Four-period ahead								
	Nominal	Excess		Nominal	Excess		Nominal	Excess
	$\sum_{k=1}^4 \rho^k r^E_{t+k}$	$\sum_{k=1}^4 \rho^k r^E_{t+k}$		$\sum_{k=1}^4 \rho^k r^D_{t+k}$	$\sum_{k=1}^4 \rho^k r^D_{t+k}$		$\sum_{k=1}^4 \rho^k r^X_{t+k}$	$\sum_{k=1}^4 \rho^k r^X_{t+k}$
DA_t	3.24	2.31	NSB_t	0.20	-0.03	NMA_t	0.41	0.40
	2.78	1.85		1.82	-0.34		2.63	2.44
R^2	0.07	0.04	R^2	0.02	0.00	R^2	0.09	0.08
unbiased	2.93	1.98	unbiased	0.20	-0.04	unbiased	0.41	0.40
	4.03	2.52		2.46	-0.50		5.08	4.80
$\tilde{R}^2$	0.06	0.03	$\tilde{R}^2$	0.02	0.00	$\tilde{R}^2$	0.09	0.08
Panel C: Single Variable Predictive Regression, Eight-period ahead								
	Nominal	Excess		Nominal	Excess		Nominal	Excess
	$\sum_{k=1}^8 \rho^k r^E_{t+k}$	$\sum_{k=1}^8 \rho^k r^E_{t+k}$		$\sum_{k=1}^8 \rho^k r^D_{t+k}$	$\sum_{k=1}^8 \rho^k r^D_{t+k}$		$\sum_{k=1}^8 \rho^k r^X_{t+k}$	$\sum_{k=1}^8 \rho^k r^X_{t+k}$
DA_t	5.52	3.82	NSB_t	0.34	-0.01	NMA_t	0.54	-0.08
	3.15	1.93		1.67	-0.03		2.06	-0.51
R^2	0.13	0.06	R^2	0.03	0.00	R^2	0.07	0.00
unbiased	5.31	3.59	unbiased	0.33	-0.01	unbiased	0.55	-0.07
	5.71	3.44		2.50	-0.11		4.54	-0.69
$\tilde{R}^2$	0.12	0.05	$\tilde{R}^2$	0.03	0.00	$\tilde{R}^2$	0.08	0.00

This table shows the prediction results for nominal and excess returns of three sectors using each valuation ratio in a single variable regression, estimated by OLS. For both excess and nominal returns, r^E, r^D, r^X are the log returns of equity, debt, and external asset, respectively. DA is the dividend price ratio, NSB is the net government surplus (tax minus spending) to debt ratio, and NMA is the net import to foreign asset ratio. We correct for Stambaugh bias and compute the unbiased coefficient and R-squared $\tilde{R}^2$ accordingly. The standard error of the unbiased estimate is generated by bootstrapping 10,000 draws. In Panel A, we forecast the one-period ahead variables. In Panels B and C, we compute the summed returns and cash-flow growth in the long horizon.

Table B.2. Predictive Regression for Nominal and Excess Returns, Multiple-Variable

Panel A, Multiple Variable Predictive Regression, One-period ahead						
	Nominal	Nominal	Nominal	Excess	Excess	Excess
	r^E_{t+1}	r^D_{t+1}	r^X_{t+1}	r^E_{t+1}	r^D_{t+1}	r^X_{t+1}
DA_t	0.78	0.24	0.14	0.52	-0.03	-0.13
	2.31	1.38	0.30	1.47	-0.20	-0.28
NSB_t	-0.25	0.04	-0.40	-0.32	-0.03	-0.46
	-2.06	1.33	-2.50	-2.57	-0.99	-2.85
NMA_t	0.00	0.01	0.12	0.00	0.01	0.12
	-0.01	0.92	2.59	-0.08	0.76	2.49
R^2	0.03	0.03	0.06	0.03	0.01	0.06
Panel B, Multiple Variable Predictive Regression, Four-period ahead						
	Nominal	Nominal	Nominal	Excess	Excess	Excess
	$\sum_{k=1}^{4}\rho^k r^E_{t+k}$	$\sum_{k=1}^{4}\rho^k r^D_{t+k}$	$\sum_{k=1}^{4}\rho^k r^X_{t+k}$	$\sum_{k=1}^{4}\rho^k r^E_{t+k}$	$\sum_{k=1}^{4}\rho^k r^D_{t+k}$	$\sum_{k=1}^{4}\rho^k r^X_{t+k}$
DA_t	3.06	0.73	0.62	2.10	-0.29	-0.38
	2.82	1.27	0.43	1.82	-0.65	-0.26
NSB_t	-0.79	0.21	-1.67	-1.06	-0.04	-1.95
	-2.20	1.95	-3.30	-2.81	-0.42	-3.63
NMA_t	0.01	0.02	0.41	0.00	0.01	0.40
	0.10	0.53	2.83	-0.02	0.29	2.68
R^2	0.11	0.06	0.16	0.10	0.01	0.17
Panel C, Multiple Variable Predictive Regression, Eight-period ahead						
	Nominal	Nominal	Nominal	Excess	Excess	Excess
	$\sum_{k=1}^{8}\rho^k r^E_{t+k}$	$\sum_{k=1}^{8}\rho^k r^D_{t+k}$	$\sum_{k=1}^{8}\rho^k r^X_{t+k}$	$\sum_{k=1}^{8}\rho^k r^E_{t+k}$	$\sum_{k=1}^{8}\rho^k r^D_{t+k}$	$\sum_{k=1}^{8}\rho^k r^X_{t+k}$
DA_t	5.25	1.25	1.63	3.46	-0.75	-0.26
	3.38	1.37	0.66	1.97	-1.17	-0.10
NSB_t	-0.82	0.39	-3.32	-1.26	-0.04	-3.85
	-1.34	1.84	-4.13	-1.91	-0.20	-4.48
NMA_t	0.01	0.04	0.57	-0.03	0.00	0.55
	0.05	0.52	2.49	-0.21	0.07	2.26
R^2	0.16	0.08	0.22	0.10	0.02	0.22

This table shows the OLS estimates of using three valuation ratios in one regression predicting logged nominal and excess returns of all three sectors. For both excess and nominal returns, r^E, r^D, r^X are the log returns of equity, debt, and external asset, respectively. DA is the dividend price ratio, NSB is the net government surplus (tax minus spending) to debt ratio, and NMA is the net import to foreign asset ratio. We correct for Stambaugh bias and compute the unbiased coefficient and R-squared $\tilde{R}^2$ accordingly. The standard error of the unbiased estimate is generated by bootstrapping 10,000 draws. In Panel A, we forecast the one-period ahead variables. In Panels B and C, we compute the summed returns and cash-flow growth in the long horizon.

Table B.3. Predictive Regression for Nominal and Excess Returns, using GMM to Enforce Aggregate Budget Constraint

Table 3 Panel A, Multiple Variable Prediction, One-period Ahead

	Nominal	Nominal	Nominal	Excess	Excess	Excess
	r^E_{t+1}	r^D_{t+1}	r^X_{t+1}	r^E_{t+1}	r^D_{t+1}	r^X_{t+1}
DA_t	0.87	0.18	0.19	0.31	-0.17	-0.11
	4.86	2.37	0.52	1.50	-2.15	-0.29
NSB_t	-0.26	0.05	-0.50	-0.30	-0.04	-0.59
	-2.45	1.04	-2.70	-2.86	-1.03	-3.14
NXA_t	-0.02	0.02	0.11	-0.03	0.01	0.10
	-1.06	2.23	2.85	-1.02	1.73	2.35
R^2	0.04	0.04	0.06	0.03	0.02	0.06

Table 3 Panel B, Multiple Variable Prediction, Four-period Ahead

	Nominal	Nominal	Nominal	Excess	Excess	Excess
	$\sum_{k=1}^4 \rho^k r^E_{t+k}$	$\sum_{k=1}^4 \rho^k r^D_{t+k}$	$\sum_{k=1}^4 \rho^k r^X_{t+k}$	$\sum_{k=1}^4 \rho^k r^E_{t+k}$	$\sum_{k=1}^4 \rho^k r^D_{t+k}$	$\sum_{k=1}^4 \rho^k r^X_{t+k}$
DA_t	2.93	0.63	0.84	1.20	-0.49	-0.04
	6.05	3.40	1.00	2.08	-2.71	-0.04
NSB_t	-0.66	0.30	-2.33	-0.85	-0.04	-2.64
	-3.21	3.20	-6.52	-3.47	-0.48	-6.87
NMA_t	-0.03	0.04	0.38	-0.07	0.02	0.34
	-0.68	2.23	4.49	-1.20	1.22	3.98
R^2	0.09	0.10	0.22	0.06	0.02	0.22

Table 3 Panel C, Multiple Variable Prediction, Eight-period Ahead

	Nominal	Nominal	Nominal	Excess	Excess	Excess
	$\sum_{k=1}^8 \rho^k r^E_{t+k}$	$\sum_{k=1}^8 \rho^k r^D_{t+k}$	$\sum_{k=1}^8 \rho^k r^X_{t+k}$	$\sum_{k=1}^8 \rho^k r^E_{t+k}$	$\sum_{k=1}^8 \rho^k r^D_{t+k}$	$\sum_{k=1}^8 \rho^k r^X_{t+k}$
DA_t	4.81	1.50	1.82	2.20	-0.64	0.40
	7.73	4.94	1.46	3.05	-2.55	0.32
NSB_t	-0.82	0.57	-4.01	-1.08	0.05	-4.60
	-2.82	3.72	-7.29	-3.35	0.35	-7.86
NMA_t	-0.02	0.07	0.53	-0.09	0.00	0.43
	-0.26	2.74	4.68	-1.29	0.16	3.73
R^2	0.14	0.17	0.27	0.06	0.02	0.26

This table shows the constrained GMM estimates of using three valuation ratios in one regression predicting logged nominal and excess returns of all three sectors. For both excess and nominal returns, r^E,r^D,r^X are the log returns of equity, debt, and external asset. DA is the dividend price ratio, NSB is the net government surplus (tax minus spending) to debt ratio, and NMA is the net import to foreign asset ratio. We correct for Stambaugh bias and compute the unbiased coefficient and R-squared $\tilde{R}^2$ accordingly. The standard error of the unbiased estimate is generated by bootstrapping 10,000 draws. In Panel A, we forecast the one-period ahead variables. In Panels B and C, we compute the summed returns and cash-flow growth in the long horizon.

Appendix C

Appendix for Chapter 3

C.1 Proof of the Conditional Factor Model

Proof of Proposition 2.1

Equation (3.2) leads to:

$$r_t^f = 1/E_{t-1}[m_t] \tag{C.1}$$

$$1 = E_{t-1}[r_{it}m_t] = COV_{t-1}(r_{it}, m_t) + E_{t-1}[r_{it}]E_{t-1}[m_t] \tag{C.2}$$

$$\Rightarrow E_{t-1}[r_{it} - r_t^f] = -COV_{t-1}(r_{it}, m_t)/E_{t-1}[m_t] \tag{C.3}$$

Combine with the pricing kernel in (3.1):

$$E_{t-1}[r_{it} - r_t^f] = b'_{t-1}COV_{t-1}(r_{it}, f_t)/E_{t-1}[m_t] \tag{C.4}$$

By assuming each factor satisfies the pricing equation:

$$E_{t-1}[f_t - r_t^f] = b'_{t-1}VAR_{t-1}(f_t)/E_{t-1}[m_t] \tag{C.5}$$

As an extension of proof 2.1, let $H_t^K = VAR_{t-1}(f_t)$ be the conditional covariance matrix of factors and $COV_{t-1}(r_{it}, f_t)$ be the k by 1 vector containing all the conditional covariance between r_{it} and each f_{kt}. Then, based on assumptions 2a and 2b:

Proof of Proposition 3.2

$$\forall i \neq j, r_{i,t} = f_t' \beta_{it} + \varepsilon_{i,t}, r_{j,t} = f_t' \beta_{jt} + \varepsilon_{j,t} \tag{C.6}$$

$$COV_{t-1}(r_{it}, r_{jt}) = \beta_i' VAR_{t-1}(f_t) \ = \beta_i' H_t^K \beta_j \tag{C.7}$$

$$COV_{t-1}(r_{it}, f_t)' = \beta_i' VAR_{t-1}(f_t) = \beta_i' H_t^K \tag{C.8}$$

$$COV_{t-1}(r_{jt}, f_t)' = \beta_j' VAR_{t-1}(f_t)) = \beta_j' H_t^K \tag{C.9}$$

$$\Rightarrow COV_{t-1}(r_{it}, f_t)'(H_t^K)^{-1} COV_{t-1}(r_{jt}, f_t) = COV_{t-1}(r_{it}, r_{jt}) \tag{C.10}$$

Decompose H_t^K to $D_t^K K_t D_t^K$, and then divide (C.10) by $\sigma_{i,t}\sigma_{j,t}$:

$$\frac{1}{\sigma_{i,t}} COV_{t-1}(r_{it}, f_t)'(D_t^K)^{-1} K_t^{-1} (D_t^K)^{-1} COV_{t-1}(r_{jt}, f_t) \frac{1}{\sigma_{j,t}} = \rho_{i,j,t} \tag{C.11}$$

Note that, by definition, $(D_t^K)^{-1} COV_{t-1}(r_{it}, f_t) \frac{1}{\sigma_{i,t}} = \rho_{i,t}$ is the K by 1 vector containing all the correlations between r_{it} and each f_{kt}. Thus,

$$\rho_{i,j,t} = \rho_{i,t}' K_t^{-1} \rho_{j,t}. \tag{C.12}$$

Equation (C.12) can be extended to matrix form:

Define X_t as the $N \times K$ matrix containing every $Corr_{t-1}(r_{it}, f_{kt})$. Then, R_t can be presented by $X_t K_t^{-1} X_t'$ plus an adjustment term for diagonal elements:

$$R_t = X_t K_t^{-1} X_t' + diag\{1 - \rho_{i,t}' K_t^{-1} \rho_{i,t}\}$$

Derive the closed-form solution of R_t^{-1} and $|R_t|$ by applying the following theorem:

Theorem Woodbury Identity

$$(A + UCV')^{-1} = A^{-1} - A^{-1}U(C^{-1} + V'A^{-1}U)^{-1}V'A^{-1} \tag{C.13}$$

$$\det(A + UCV') = \det(C^{-1} + V'A^{-1}U)\det(C)\det(A) \tag{C.14}$$

Proof of Proposition 3.3

$$R_t = X_t K_t^{-1} X_t' + diag\{1 - \rho_{i,t}' K_t^{-1} \rho_{i,t}\} \tag{C.15}$$

Set $A = diag\{1 - \rho_{i,t}' K_t^{-1} \rho_{i,t}\}, U = V = X_t, C = K_t^{-1}$:

$$A^{-1} = diag\{\tfrac{1}{1-\rho_{i,t}' K_t^{-1} \rho_{i,t}}\} \tag{C.16}$$

$$\det(A) = \textstyle\prod_i (1 - \rho_{i,t}' K_t^{-1} \rho_{i,t}) \tag{C.17}$$

$$(C^{-1} + V'A^{-1}U) = K_t + X_t' diag\{\tfrac{1}{1-\rho_{i,t}' K_t^{-1} \rho_{i,t}}\} X_t \tag{C.18}$$

$$\tag{C.19}$$

which leads to:

$$R_t^{-1} = diag\{\frac{1}{1 - \rho_{i,t}' K_t^{-1} \rho_{i,t}}\}$$
$$- diag\{\frac{1}{1 - \rho_{i,t}' K_t^{-1} \rho_{i,t}}\} X_t (K_t + X_t' diag\{\frac{1}{1 - \rho_{i,t}' K_t^{-1} \rho_{i,t}}\} X_t)^{-1} X_t' diag\{\frac{1}{1 - \rho_{i,t}' K_t^{-1} \rho_{i,t}}\} \tag{C.20}$$

$$\det(R_t) = \det(K_t + X_t' diag\{\frac{1}{1 - \rho_{i,t}' K_t^{-1} \rho_{i,t}}\} X_t) \det(K_t^{-1}) \prod_i (1 - \rho_{i,t}' K_t^{-1} \rho_{i,t}) \tag{C.21}$$

C.2 Statistical Inference

The likelihood is decomposed into a GARCH part and a correlation part:

$$
\begin{aligned}
L_1(\theta) &= -\frac{1}{2}\sum_t \sum_{i=1}^{n}(\log(2\pi) + \log(h_{i,t}) + \frac{r_{i,t}^2}{h_{i,t}}) - \frac{1}{2}\sum_t(\log(2\pi) + \log(h_{m,t}) + \frac{r_{m,t}^2}{h_{m,t}}) \\
L_2(\theta,\phi) &= L_c(\theta,\phi) = -\frac{1}{2}\sum_{t=1}^{T}(-\varepsilon_t'\varepsilon_t + \log|R_t| + \varepsilon_t' R_t^{-1}\varepsilon_t) \\
\log f_{1,t} &= -\frac{1}{2}\sum_{i=1}^{n}(\log(2\pi) + \log(h_{i,t}) + \frac{r_{i,t}^2}{h_{i,t}}) - \frac{1}{2}(\log(2\pi) + \log(h_{m,t}) + \frac{r_{m,t}^2}{h_{m,t}}) \\
\log f_{2,t} &= \frac{1}{2}(\log|R_t| + \varepsilon_t' R_t^{-1}\varepsilon_t)
\end{aligned}
\tag{C.22}
$$

White (1996) Theorem 6.1

Under assumptions C.1–C.6,

$$\sqrt{T}(\hat{\gamma} - \gamma^*) \sim^A N(0, A^{*-1} B A^{*-1})$$

,

where $A^* = \begin{Bmatrix} E[\nabla_{\theta\theta} L_1(r_t,\theta^*)] & 0 \\ .E[\nabla_{\theta\phi} L_2(r_t,\theta^*,\phi^*)] & E[\nabla_{\theta\theta} L_2(r_t,\theta^*,\phi^*)] \end{Bmatrix}$.

and

$B^* = var(\, T^{-\frac{1}{2}} \sum_t (s_{1,t}^{*\,\prime}, s_{2,t}^{*\,\prime})\,)$,

where $s_{1,t}^* = E[\nabla_\theta L_1(r_t,\theta^*)]$ and $s_{2,t}^* = E[\nabla_\phi L_1(r_t,\theta^*,\phi^*)]$.

Write assumptions in White (1996) and Engle and Kelly (2012):

Assumptions C.1

(a) For all $\theta \in \Theta, \phi \in \Phi$, $E[\log f_{1,t}(r_t,\theta)]$ and $E[\log f_{2,t}(r_t,\theta,\phi)]$ exist and are finite, $\forall t$;

(b) $E[\log f_{1,t}(r_t,\theta)]$ and $E[\log f_{2,t}(r_t,\theta,\phi)]$ are continuous on Θ and Φ, $\forall t$; and

(c) $\{\log f_{1,t}(r_t,\theta)\}$ and $\{\log f_{2,t}(r_t,\theta,\phi)\}$ each obey the strong uniform law of large number.

Assumptions C.2

$f_{1,t}$ and $f_{2,t}$ are each twice continuously differentiable on Θ and Φ, $\forall t$.

Assumptions C.3

For all $\theta \in \Theta, \phi \in \Phi$, $E[\nabla_{\theta} L_1(r_t, \theta)] < \infty$ and $E[\nabla_{\phi} L_2(r_t, \theta, \phi)] < \infty$, $\forall t$.

Assumptions C.4

(a) For all $\theta \in \Theta, \phi \in \Phi$, $E[\nabla_{\theta\theta} L_1(r_t, \theta)] < \infty$ and $E[\nabla_{\phi\phi} L_2(r_t, \theta, \phi)] < \infty$;

(b) $E[\nabla_{\theta\theta} L_1(r_t, \theta)]$ and $E[\nabla_{\phi\phi} L_2(r_t, \theta, \phi)]$ are continuous on Θ and Φ;

(c) $\{\nabla'_{\theta} s_{1,t}(r_t) = \nabla_{\theta\theta} \log f_1(r_t, \theta)\}$ and $\{\nabla'_{\phi} s_{2,t}(r_t) = \nabla_{\phi\phi} \log f_2(r_t, \theta, \phi)\}$; and

(d) A^* is negative definite.

Assumptions C.5

$E[L_1(r_t, \theta)]$ is uniquely maximized by θ^* interior to Θ, and $E[L_2(r_t, \theta, \phi)]$ is uniquely maximized by ϕ^* interior to Φ.

Assumptions C.6

$\{(T^{-\frac{1}{2}} {s^*_{1,t}}', T^{-\frac{1}{2}} {s^*_{2,t}}')\} \equiv \{(T^{-\frac{1}{2}} \nabla'_{\theta} L_1(r_t, \theta^*), T^{-\frac{1}{2}} \nabla'_{\phi} L_2(r_t, \theta^*, \phi^*))\}$ obeys the central limit theorem.

Like DECO, the studied model ensures identification as long as each pairwise DCC is properly identified.

Bibliography

Acemoglu, D., V. M. Carvalho, A. Ozdaglar, and A. Tahbaz-Salehi (2012). The network origins of aggregate fluctuations. *Econometrica 80*(5), 1977–2016.

Acemoglu, D., A. Ozdaglar, and A. Tahbaz-Salehi (2015). Systemic risk and stability in financial networks. *American Economic Review 105*(2), 564–608.

Alves, M. F., M. I. Gomes, and L. de Haan (2003). A new class of semi-parametric estimators of the second order parameter. *Portugaliae Mathematica 60*(2), 193–214.

Ang, A., G. Bekaert, and M. Wei (2007). Do macro variables, asset markets, or surveys forecast inflation better? *Journal of Monetary Economics 54*(4), 1163 – 1212.

Ang, A., R. J. Hodrick, Y. Xing, and X. Zhang (2006). The cross-section of volatility and expected returns. *The journal of finance 61*(1), 259–299.

Ang, A., R. J. Hodrick, Y. Xing, and X. Zhang (2009). High idiosyncratic volatility and low returns: International and further us evidence. *Journal of Financial Economics 91*(1), 1–23.

Axtell, R. L. (2001). Zipf distribution of us firm sizes. *science 293*(5536), 1818–1820.

Bali, T. G., N. Cakici, X. Yan, and Z. Zhang (2005). Does idiosyncratic risk really matter? *The Journal of Finance 60*(2), 905–929.

Barndorff-Nielsen, O. E. and N. Shephard (2004). Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics. *Econometrica 72*(3), 885–925.

Barro, R. J. (1979). On the determination of the public debt. *Journal of Political Economy 87*(5), 940–971.

Beare, B. K. and A. A. Toda (2022). Determination of pareto exponents in economic models driven by markov multiplicative processes. *Econometrica 90*(4), 1811–1833.

Beirlant, J., G. Dierckx, Y. Goegebeur, and G. Matthys (1999). Tail index estimation and an exponential regression model. *Extremes 2*(2), 177–200.

Berndt, A., H. Lustig, and S. Yeltekin (2012). How does the us government finance fiscal shocks? *American Economic Journal: Macroeconomics 4*(1), 69–104.

Berndt, A. and S. Yeltekin (2015). Monetary policy, bond returns and debt dynamics. *Journal of Monetary Economics 73*, 119 – 136.

Byun, S. J. and L. 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, J. Y. (1991). A variance decomposition for stock returns. *The Economic Journal 101*(405), 157–179.

Campbell, J. Y. (1993). Intertemporal asset pricing without consumption data. *The American Economic Review 83*(3), 487–512.

Campbell, J. Y. (1996). Understanding risk and return. *Journal of Political Economy 104*(2), pp. 298–345.

Campbell, J. Y., M. Lettau, B. G. Malkiel, and Y. Xu (2001). Have individual stocks become more volatile? an empirical exploration of idiosyncratic risk. *The journal of finance 56*(1), 1–43.

Campbell, J. Y. and R. J. Shiller (1988a). The dividend-price ratio and expectations of future dividends and discount factors. *The Review of Financial Studies 1*(3), 195–228.

Campbell, J. Y. and R. J. Shiller (1988b). Stock prices, earnings, and expected dividends. *The Journal of Finance 43*(3), pp. 661–676.

Campbell, J. Y. and S. B. Thompson (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? *The Review of Financial Studies 21*(4), pp. 1509–1531.

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

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

Champernowne, D. G. (1953). A model of income distribution. *The Economic Jour-*

nal 63(250), 318–351.

Chan, L. K., J. Karceski, and J. Lakonishok (1999). On portfolio optimization: Forecasting covariances and choosing the risk model. *The review of Financial studies 12*(5), 937–974.

Cochrane, J. H. (2008). The Dog That Did Not Bark: A Defense of Return Predictability. *Review of Financial Studies 21*(4), 1533–1575.

Cochrane, J. H. (2011). Presidential address: Discount rates. *The Journal of Finance 66*(4), 1047–1108.

Cochrane, J. H. (2022). The fiscal roots of inflation. *Review of Economic Dynamics 45*, 22–40.

Cochrane, J. H. and M. Piazzesi (2005). Bond risk premia. *The American Economic Review 95*(1), 138–160.

Connor, G. and R. A. Korajczyk (1986). Performance measurement with the arbitrage pricing theory: A new framework for analysis. *Journal of financial economics 15*(3), 373–394.

Connor, G. and R. A. Korajczyk (1993). A test for the number of factors in an approximate factor model. *the Journal of Finance 48*(4), 1263–1291.

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

Corte, P. D., S. J. Riddiough, and L. Sarno (2016). Currency premia and global imbalances. *The Review of financial studies 29*(8), 2161–2193.

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

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

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

Dybvig, P. H. (1983). An explicit bound on individual assets' deviations from apt pricing in a finite economy. *Journal of Financial Economics 12*(4), 483–496.

Engle, R. and B. Kelly (2012). Dynamic equicorrelation. *Journal of Business & Economic Statistics 30*(2), 212–228.

Engle, R. F. (2009). High dimension dynamic correlations. *The methodology and practice of econometrics: A festschrift in honour of David F. Hendry*, 122–148.

Engle, R. F., O. Ledoit, and M. Wolf (2017). Large dynamic covariance matrices. *Journal of Business & Economic Statistics*, 1–13.

Engle, R. F., V. K. Ng, and M. Rothschild (1990). Asset pricing with a factor-arch covariance structure: Empirical estimates for treasury bills. *Journal of Econometrics 45*(1-2), 213–237.

Engle, R. F. and K. Sheppard (2001). Theoretical and empirical properties of dynamic conditional correlation multivariate garch. Technical report, National Bureau of Economic Research.

Fama, E. F. and R. R. Bliss (1987). The information in long-maturity forward rates. *The American Economic Review*, 680–692.

Fama, E. F. and K. R. French (1992). The cross-section of expected stock returns. *the Journal of Finance 47*(2), 427–465.

Feng, G., S. Giglio, and D. Xiu (2020). Taming the factor zoo: A test of new factors. *The Journal of Finance 75*(3), 1327–1370.

Feuerverger, A. and P. Hall (1999). Estimating a tail exponent by modelling departure from a pareto distribution. *The Annals of Statistics 27*(2), 760–781.

Gabaix, X. (1999). Zipf's law for cities: an explanation. *The Quarterly journal of economics 114*(3), 739–767.

Gabaix, X. (2011). The granular origins of aggregate fluctuations. *Econometrica 79*(3), 733–772.

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

Gabaix, X. and R. S. Koijen (2020). Granular instrumental variables. Technical report, National Bureau of Economic Research.

Giannitsarou, C. and A. Scott (2006). Inflation implications of rising government debt. *NBER International Seminar on Macroeconomics*, 393–439.

Giglio, S., B. Kelly, and D. Xiu (2022). Factor models, machine learning, and asset pricing. *Annual Review of Financial Economics 14*.

Giglio, S. and D. Xiu (2021). Asset pricing with omitted factors. *Journal of Political Economy 129*(7), 000–000.

Giglio, S., D. Xiu, and D. Zhang (2021). Test assets and weak factors. Technical report, National Bureau of Economic Research.

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

Gomesa, M. I. and M. J. Martins (2002). "asymptotically unbiased" estimators of the tail index based on external estimation of the second order parameter. *Extremes 5*(1), 5–31.

Gourinchas, P.-O. (2008). Valuation effects and external adjustment: a review. In S. E. Kevin Cowan and R. Valdes (Eds.), *Current Account and External Financing*, Volume XII of *Series on Central Banking, Analysis and Economic Policy*, pp. 195–236. Banco Central de Chile.

Gourinchas, P.-O. and H. Rey (2007a). From world banker to world venture capitalist: Us external adjustment and the exorbitant privilege. In *G7 Current Account Imbalances: Sustainability and Adjustment*, pp. 11–66. University of Chicago Press.

Gourinchas, P.-O. and H. Rey (2007b). International financial adjustment. *Journal of Political Economy 115*(4), 665–703.

Gourinchas, P.-O. and H. Rey (2014). Chapter 10 - external adjustment, global imbalances, valuation effects. In G. Gopinath, E. Helpman, and K. Rogoff (Eds.), *Handbook of International Economics*, Volume 4 of *Handbook of International Economics*, pp. 585–645. Elsevier.

Gourinchas, P.-O., H. Rey, and N. Govillot (2010). Exorbitant privilege and exorbitant duty. IMES Discussion Paper Series 10-E-20, Institute for Monetary and Economic Studies, Bank of Japan.

Goyal, A. and P. Santa-Clara (2003). Idiosyncratic risk matters! *The journal of finance 58*(3), 975–1007.

Grinblatt, M. and S. Titman (1983). Factor pricing in a finite economy. *Journal of Financial Economics 12*(4), 497–507.

Hall, P. and A. H. Welsh (1985). Adaptive estimates of parameters of regular variation.

The Annals of Statistics, 331–341.

Hamilton, J. D. (2018). Why you should never use the hodrick-prescott filter. *Review of Economics and Statistics 100*(5), 831–843.

Herskovic, B., B. Kelly, H. Lustig, and S. Van Nieuwerburgh (2016). The common factor in idiosyncratic volatility: Quantitative asset pricing implications. *Journal of Financial Economics 119*(2), 249–283.

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

Hou, K. and R. K. Loh (2016). Have we solved the idiosyncratic volatility puzzle? *Journal of Financial Economics 121*(1), 167–194.

Hou, K., H. Mo, C. Xue, and L. Zhang (2021). An augmented q-factor model with expected growth. *Review of Finance 25*(1), 1–41.

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

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

Jegadeesh, N. and S. Titman (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of finance 48*(1), 65–91.

Jiang, Z., H. Lustig, S. Van Nieuwerburgh, and M. Z. Xiaolan (2021). What drives variation in the us debt/output ratio? the dogs that didn't bark. Technical report, National Bureau of Economic Research.

Kelly, B. and H. Jiang (2014). Tail risk and asset prices. *The Review of Financial Studies 27*(10), 2841–2871.

Kelly, B. T., S. Pruitt, and Y. Su (2020). Instrumented principal component analysis. *Available at SSRN 2983919*.

Koijen, R. S. and S. Van Nieuwerburgh (2011). Predictability of returns and cash flows. *Annu. Rev. Financ. Econ. 3*(1), 467–491.

Lane, P. R. and G. M. Milesi-Ferretti (2001). The external wealth of nations: measures of foreign assets and liabilities for industrial and developing countries. *Journal of international economics 55*(2), 263–294.

Lane, P. R. and J. C. Shambaugh (2010). Financial exchange rates and international currency exposures. *The American Economic Review 100*(1), 518–540.

Larrain, B. and M. Yogo (2008). Does firm value move too much to be justified by subsequent changes in cash flow. *Journal of Financial Economics 87*(1), 200 – 226.

Ledoit, O. and M. Wolf (2003). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of empirical finance 10*(5), 603–621.

Ledoit, O. and M. Wolf (2004). Honey, i shrunk the sample covariance matrix. *The Journal of Portfolio Management 30*(4), 110–119.

Ledoit, O., M. Wolf, et al. (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices. *The Annals of Statistics 40*(2), 1024–1060.

Lettau, M. and S. Ludvigson (2001). Consumption, aggregate wealth, and expected stock returns. *The Journal of Finance 56*(3), 815–849.

Lettau, M. and S. C. Ludvigson (2005). Expected returns and expected dividend growth. *Journal of Financial Economics 76*(3), 583–626.

Liu, Y. (2023). Government debt and risk premia. *Journal of monetary economics*.

Lucas Jr, R. E. (1978). Asset prices in an exchange economy. *Econometrica: Journal of the Econometric Society*, 1429–1445.

Markowitz, H. (1952). Portfolio selection. *The journal of finance 7*(1), 77–91.

Meese, R. A. and K. Rogoff (1983). Empirical exchange rate models of the seventies: Do they fit out of sample? *Journal of International Economics 14*(1 - 2), 3 – 24.

Menzly, L., T. Santos, and P. Veronesi (2004). Understanding predictability. *Journal of Political Economy 112*(1), 1–47.

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

Pástor, L. and R. F. Stambaugh (2012). Are stocks really less volatile in the long run? *The Journal of Finance 67*(2), 431–478.

Peng, L. (1998). Asymptotically unbiased estimators for the extreme-value index. *Statistics & Probability Letters 38*(2), 107–115.

Ross, S. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory 13*(3), 341–360.

Rossi, B. (2013). Exchange rate predictability. *Journal of Economic Literature 51*(4), 1063–1119.

Stambaugh, R. F. (1999). Predictive regressions. *Journal of Financial Economics 54*(3), 375 – 421.

Timmermann, A. G. (1993). How learning in financial markets generates excess volatility and predictability in stock prices. *The Quarterly Journal of Economics 108*(4), 1135–1145.

Welch, I. and A. Goyal (2008). A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies 21*(4), 1455–1508.

White, H. (1996). *Estimation, inference and specification analysis*. Number 22. Cambridge university press.

Wold, H. O. and P. Whittle (1957). A model explaining the pareto distribution of wealth. *Econometrica, Journal of the Econometric Society*, 591–595.

Xu, Y. and B. G. Malkiel (2003). Investigating the behavior of idiosyncratic volatility. *The Journal of Business 76*(4), 613–645.