Industry Concentration and Average Stock Returns Revisited

R. Jacob Applin

May 2019

INDUSTRY CONCENTRATION AND AVERAGE STOCK RETURNS REVISTED

A Thesis
Presented to
The Faculty of the Department of Economics
Western Kentucky University
Bowling Green, Kentucky

In Partial Fulfillment
Of the Requirements for the Degree
Master of Arts in Applied Economics

By
R. Jacob Applin

May 2019

INDUSTRY CONCENTRATION AND AVERAGE STOCK RETURNS REVISTED

Date Recommended: ____________________

Dr. Alex Lebedinsky, Advisor of Thesis

Dr. Brian Goff

Dr. Mohsen Mollagholamali

Dean, Graduate Studies and Research   Date

Acknowledgements

Although this paper is technically the work of one person, there are numerous individuals who deserve credit for their advice and help along the way.

First, and foremost, a thank you is necessary to all of the faculty and students I have had the pleasure to work with at Western Kentucky University. The frustrations I experienced in attempting to replicate Hou and Robinson were lessened quite a bit by the humor from my friends and peers in the GA office.

Second, I would like to acknowledge Dr. Brian Goff and Dr. Mohsen Mollagholamali for taking the time to serve as readers. Dr. Moolagholamali also deserves recognition for assisting with data collection.

Most importantly, I would like to thank Dr. Alex Lebedinsky, my main advisor. Not only did he offer excellent insight and ideas - in particular, the idea to try a nonlinear model, - this project simply would not have been possible without the SAS expertise he bestowed upon me. He endured weeks of frustration with coding up Hou and Robinson's methodology right along with me.

Finally, a thank you to my parents, for always providing love and support for my endeavors is necessary. Without them, I quite literally would not be where I am today. Their dogs, Pascal and Dewey, also deserve a shoutout for providing much needed entertainment during visits home.

Abstract

In 2006 Kewei Hou and David Robinson documented a competition premium: a higher cost of common equity for firms operating in less concentrated industries compared to firms operating in more concentrated industries. This paper attempts to replicate this result and test its robustness against over a decade's worth of new data. Although certain results were unable to be replicated, their main result of a statistically significant competition premium between 0.20% and 0.40% at both the industry-level and firm-level is replicated for their July 1963-December 2001 sample period. Following this, this study finds evidence of a statistically significant competition premium of around 0.40% for the January 2002-December 2018 period. However, the result is called into question as it is not significant for the entire 1963-2018 sample period. The issue appears to be one of specification and not robustness, though, as this study finds evidence for a non-monotonic competition premium.

Contents

List of Tables

List of Figures

1 Introduction

In 2006, Hou and Robinson provided empirical evidence that industry concentration constitutes a significant factor for stock returns.[1] Specifically, they find that firms in more concentrted industries post, on average, lower stock returns - roughly 0.30% lower for the July 1963-December 2001 period. The authors docuemnt this finding at both the industry-level and the firm-level. Hence, Hou and Robinson provide convincing evidence for the existence of a *competition premium*: investors appear to require a premium for investing in firms that operate in more competitive enviornments relative to firms operating in more monopolsitic and oligopolistic environments.

Hou and Robinson's work proved to be a catalyst for a wide range of research at the intersection of industrial organization and finance focusing on the relationship between the cost of equity capital (as well as other types of capital) and market structure[2] The literature review treats a large portion of this scholarship. Given that it has been over a decade since Hou and Robinson's work, there exists an ample amount of new data to use in extending and testing their original findings. As discussed in the literature review, a number of papers - for example, Bustamante and Donangelo (2017) - have produced results that would seem to contradict Hou and Robison's finding that more monopolistic industries require lower stock returns. Hence, taking advantage of new data to extend Hou and Robinson's results may clear up this ambiguity regarding the relationship between the cost of equity and market structure (or whether such a relationship even exists in the first place).

While some difficulties are encountered in replicating some of Hou and Robinson's findings, this study is able to successfully reproduce their competition premium of approximately 0.30% for the 1963-2001 sample period. In extending their model to new data, this study finds that the competition premium persists throughout the 2002-2018 time period at the firm-level only: the industry-level premium fails to reach statistical significance for all but one or two partiuclar specifications for the 2002-2018 time period. Perhaps most importantly, this study finds the competition premium paradoxically fails to be significant - at both the industry-level and firm-level - when examining the entire 1963-2018 time period. This finding raises questions regarding the robustness of the competition premium. As a further investigation to the premium's robustness, this study examines the behavior of the premium within and across concentration quintiles at both the industry-level and the firm-level. This excercise gives reason to believe that the issue is not one of robustness but rather of specification: when explictly testing for a non-monotonic premium, this study finds significant results that suggest the premium may exhibit parabolic behavior. Such a finding implies that risk pertaining to market structure is itself non-monotonic and likely exhibits similar parabolic behavior.

Broadly speaking, this paper is part of a larger body of finance literature (including Harvey, Liu, and Zhu (2015), McLean and Pontiff (2015), and Hou, Xue, and Zhang (2018)[3] that aims to test whether prior research can hold up to new data and new specifications. Replication is a core part of the scientific method (if not the most important part). In the context of economic and financial research, this would ideally be particularly true. However, it seems that business disciplines have only in the past decade or so started to recognize the importance of replicability.

The paper proceeds as follows: the next section provides the requiste financial and industrial organization background as well as a review of relevant literature that fuses industrial organization with finance. Section 3 provides a summary of the main portion of Hou and Robinson's original paper. In turn, section 4 discusses the process of replicating their results. Moreover, section 4 presents the main replication results. Following this, section 5 presents extensions. In doing so, this paper presents trends in industry concentration since 2001. Additionally, informal bivariate analysis and visualization

[1] Hou, Kewei, and David T. Robinson. 2006. Industry Concentration and Average Stock Returns.

[2] As of March 2019, Google Scholar lists 726 citations of Hou and Robinson's work.

[3] Given Hou is co-author, it is interesting to see (when digging into the appendices of "Replicating Anomalies") that Hou and Robinson's concentration variable ends up with mixed results. Not that it is alone - their factor is only one of what seems like millions in a sea of factors with mixed evidence.

is conducted on concentration and stock returns. From here the paper moves into more formal tests. Specifically, the latter part this section examines 1) the existence of the competition premium for the January 2002-December 2018 period; 2) the existence of the competition premium over the entire July 1963-December 2018 time period; and 3) the question of whether the competition premium exhibits non-monotonic behavior.

2 Framework and Literature Review

2.1 Financial Background and Context

In order to value an investment (whether it be an internal project or an entire firm), an analyst must decide on a *discount rate* for the purposes of determining the *present value* of the investment. A commonly used measure for a risk-appropriate discount rate is a firm's *weighted average cost of capital*, which is simply the "weighted average of the expected after-tax rates of return of [a] firm's various sources of capital" (Titman and Martin 2016). The weighted average cost of capital, often abbreviated as WACC, can be interpreted as the *opportunity cost of capital*. Titman and Martin provide an excellent illustration of this:

> "Because investors can invest their money elsewhere, providing money to a firm by purchasing its securities (bonds and shares of stock) has an opportunity cost. That is, if an investor puts her money into stock from Google (GOOG), she gives up (forgoes) the return she could have earned by investing in Microsoft (MSFT) stock. What this means is that if Google and Microsoft have equivalent risks, the expected rate of return on Microsoft stock can be viewed as the opportunity cost of capital for Google, and the expected return on Google stock can be viewed as the opportunity cost of capital for Microsoft."
>
> — p. 100

Assuming three sources of capital - interest-bearing debt, preferred equity, and common equity - as Titman and Martin do, WACC may expressed mathematically as below:

$$\text{WACC} = k_d(1-T)w_d + k_p w_p + k_e w_e \tag{1}$$

Where k_d, k_p and k_e are the *required rates of return* for interest-bearing debt, preferred equity, and common equity respectively; w_d, w_p and w_e are the weights for interest-bearing debt, preferred equity, and common equity respectively; and T is the corporate tax rate.[4] Hou and Robinson's research pertains only to the required rate of return on common equity - in other words, to common stock returns. However, as detailed in section 2.3, there is literature that attempts to examine the relationship between market structure and the other sources of capital. For the purposes of replicating and extending Hou and Robinson's work, this study keeps the focus on the required rate of return on common equity.

From an empirical perspective, there are two popular methodologies for estimating the required rate of return on common equity. The first is the Capital Asset Pricing Model (CAPM), which regresses excess common stock returns for a firm on the *market-risk premium* (the difference between the return on the market portfolio and the return on a riskless asset.[5]). While theoretically pleasing in the sense that investors should be most concerned with how risk contributes to stock returns, empirical evidence for CAPM has tended to be lacking (Titman and Martin 2016, 124).[6] This fact leads into the second popular approach: factor models. In a statistical sense, factor models tend to be regressions of stock returns on various variables (i.e., "factors") that are argued to contribute to the risk of a security (Titman and Martin 2016, 125). As will become readily apparent in section 3, Hou and Robinson's main model may be described as a factor model: with one of the risk factors being industry concentration. While arguably extremely popular, it is worth noting that the significance of many

[4] k_d is scaled by the factor $(1-T)$ to account for the fact that interest expenses are tax-deductible.

[5] Typically, the yield to maturity on government bonds is used.

[6] Fama and French (1992) give particularly convincing evidence highlighting CAPM's shortcomings.

supposedly significant factors may be dubious. In particular, Harvey, Liu, and Zhu (2015) argue that the significance of many "discovered" factors is the result of data mining. In response, the authors propose increasing the common threshold for factor significance from a *t*-statistic of 2.0 to a *t*-statistic of 3.0.[7] The potentially dubious nature of factors further underscores the importance this paper will hold in the literature: by extending Hou and Robinson's results to more than ten years of new data, it will become more clear whether industry concentration truely consistutes a statistically significant and robust risk factor.

2.2 Industrial Organization Background and Context

Hou and Robinson's work mainly aims to establish a link between industry concentration and the required rate of return on common equity. Industry concentration may be taken as a proxy for the level of competition in a market. It is important to note that it fails to be a true indicator of the level of competition as the economic definitions of competition hinge on the level of product heterogeneity in addition to the sheer number of firms in the market. Even so, industry concentration likely serves as a decent approximation

Throughout the literature, the *Herfindahl Index*, also known as the *Herfindahl-Hirschman Index*, serves as the predominant metric for industry concentration and is commonly abbreviated as *HHI*. It is simply the sum of squared market shares of the firms within a market. That is,

$$\text{HHI} = \sum_{i=1}^{N} s_i^2 \tag{2}$$

where N is the number of firms in the market and s_i is the market share of the i^{th} firm. Mathematically, the HHI is bounded below by zero and bounded above by one. Although not a perfect metric (see below) by any means, it is worth noting that it is used by the Department of Justice and the Federal Trade Commission to make legal decisions regarding firm mergers and antitrust issues.[8]

The HHI as a metric for industry concentration (and in turn as a metric for the level of competition in a market) has attracted criticism. For example, Berger (2014) gives evidence that the HHI failed to respond to the exogenous competition shock imposed by the US-Canada Free Trade Agreement. In response, Berger proposes the Boone Indicator (see Boone 2008) as an alternative. While the Boone Indicator may be potentially more attractive relative to the HHI, it is more computationally intensive to estimate.[9]

An additional issue to keep in mind with the HHI corresponds to a general issue regarding various concentration measures: such measures do not always correlate perfectly with one another. Hence, results may be sensitive to the choice of concentration measure. For example, Scanlon et al. (2006) find that measures of concentration (including the HHI) within the health care industry are only modestly correlated with one another. Specifically, they note that the level of correlation is low enough to the point that choice of concentration metric may indeed matter (45S). In fact, in the context of a larger study concerning limitations of concentration metrics, Ali, Klasa, and Yeung (2009) *find results directly contradicting Hou and Robinson's findings (which are based on the usage of*

[7] In regards to Hou and Robinson's industry concentration factor, only two of their models meet this proposed threshold on average. (See Table 4.)

[8] See https://www.justice.gov/atr/herfindahl-hirschman-index.

[9] It requires estimating the following: $\pi_{i,j} = \alpha_i + \sum_{k=1}^{T} \beta_k I_{kt} \ln c_{i,j} + \sum_{k=1}^{T-1} \gamma_k I_{kt} + \epsilon_{i,j}$ where "$\pi_{i,j}$ represents the profits of firm i in industry j and $c_{i,j}$ is the marginal cost of each firm in industry j. T is the total number of periods and the time indicator Ikt equals 1 if $k = t$ and zero otherwise... The parameter β is the Boone index measure" (Berger 2014, 10).

COMPUSTAT data to calculate HHI - see section 3.) when using alternative measures[10] (3853-3857). This is yet another factor contributing to the importance that replication and extension studies such as this one hold in the literature.

2.3 Literature on Market Structure and Common Equity Returns

Hou and Robinson note that their paper is, to their knowledge, "the [first] to link expected stock returns to industry product market characteristics through the channel [they] propose" (1929). They do note that there had been a series of scholarship preceding their work that examined links between capital structure and market structure. Specifically, they cite Titman (1984); MacKay and Phillips (2005); Almazan and Molino (2001); Asness and Stevens (1996); Moskowitz and Grinblatt (1999); Cohen, Polk, and Vuolteenaho (2003); and Hou (2003) as examples.

In addition to the literature Hou and Robinson cite, this study found that linking the cost of equity to market structure has been a theoretical curiosity since at least the 1980s. In particular, Booth (1981) built off Aivazian and Callen (1979) (who argued investment decisions by monopolies can impact future returns) and constructed a comparative statics model to theoretically demonstrate that monopolistic industries should post lower returns relative to competitive industries.[11] Conine Jr. (1983) reaches a similar conclusion with a model that focuses on price elasticity of demand. Roughly a decade later Lee, Liaw, and Rahman (1990) constructed yet another model that again concludes with more monopolistic industries requiring lower returns. Perhaps most notably, Sullivan (1978) presents what may be one of the first empirical investigations[12] into the link between capital structure and market power and finds that "a powerful firm. . . seems to be confronted by lower costs to attract capital than a non-powerful firm" (215).[13]

The interesting common thread throughout all of these early works pertains to the fact that they all independently conclude that investors require a *competition premium* - that is, investors require a higher rate of return from competitive firms relative to less competitive firms. From an intuitive and elementary economic perspective, such a conclusion should not be all too surprising.

Consider the two extremes: A perfectly competitive market and a pure monopoly. The former is infamously doomed to earn zero economic profits in the long-run. In contrast, the latter may earn positive economic profits in the long-run. However, firms operating in a perfectly competitive market may earn positive economic profits *in the short-run.* Indeed, consider a positive shock to aggregate demand. Then, in the short-run, firms in the competitive market will earn positive profits. However, because the market is perfectly competitive, these positive profits will attract firm entry. Specifically, entry will occur until economic profits have been driven back down to zero. All else equal, this does not occur for the monopoly. In contrast, the monopoly will be in position to either raise its prices and/or output, and thus capture even more profits. This observation forms the crux of Hou and Robinson's *Barriers to Entry Hypothesis* (BEH), which states that "firms in highly concentrated industries earn lower returns because, all else equal, they are better insulated from undiversifiable, aggregate demand shocks" (1932). The "insulation" comes from the fact that positive demand shocks will leave more monopolistic firms better off relative to more competitive firms. As a result, monopolistic firms may be able to "weather economic downturns without facing industry exit" (1931). In contrast, economic

[10] Specifically, they find that industry concentration is not a significant factor when using alternative measures. Moreover, if their results were statistically significant, the effect would be in the opposite direction that Hou and Robinson find. That is, they find industry concentration to be positively related to stock returns whereas Hou and Robinson find industry concentration to be inversely related to stock returns.

[11] Which is what Hou and Robinson find empirically.

[12] This does not contradict Hou and Robinson's claim that their work is the first to link stock returns to product market characteristics. Sullivan only examines whether more monopolistic firms appear to be post significantly lower returns relative to more competitive firms whereas Hou and Robinson's work actually examines whether industry concentration is a risk factor.

[13] Also see Sullivan (1982). The program Sullivan wrote to conduct his analysis was found to have a logical error. The results after fixing the error were qualitatively similar.

downturns may threaten some firms in competitive environments. Thus, overall, it is reasonable to conjecture that monopolistic firms have less distress risk relative to competitive firms.

In addition to the BEH, Hou and Robinson also offer what they deem the *Creative Destruction Hypothesis* (CDH) as a (related) alternative. The notion of *creative destruction* originated with Schumpeter (1912) and refers to the argument that "innovation and technological progress involve unseating incumbent firms in industries" (Hou and Robinson 2006, 1930). Hou and Robinson argue that this could be the channel through which industry concentration influences stock returns on the basis that "firms in more concentrated industries engage in less innovation" (1931). As an opinion, the current author finds the CDH to be the weaker of the two hypotheses. In particular, it is a basic exercise in game theory to demonstrate that the *threat of competition can be competition in and of itself.* In the sense that failure to consistently innovate in a concentrated industry may - over time - weaken the "barriers" and thus give openings for potential competitors, it is plausible to think that some, if not all, concentrated industries face just as much risk of creative destruction as do less concentrated industries. Even so, Hou and Robinson's original results find concentration to be negatively related to research and development (R&D) spending - a finding they argue to be consistent with the argument that concentrated industries innovate less than less concentrated industries (1936).[14]

As noted in the introduction, Hou and Robinson's work spawned a plethora of empirical researching aimed at bridging market structure to the rate of return on various types of capital. Regarding the cost of common equity, there have been numerous studies that have produced findings in line with Hou and Robinson's findings. Such studies include Loualiche (2016), Jory and Ngo (2017), Chiu (2010), Hashem (2011), Datta and Chakraborty (2018), Mouselli and Jaafar (2018), O'Brien (2011) and Abadi, Bozorgmehrian, and Javadi (2015). However, there have also been numerous studies that have found industry concentration and stock returns to be positively related (the opposite of what Hou and Robinson find). These studies include Grullon, Larkin, and Michaely (2018), Bustamante and Donangelo (2017) and Gallagher, Ignatieva, and McCulloch (2015). Finally, many papers have failed to establish a significant link between industry concentration and stock returns - for example, Corstjens and Vanderheyden (2010), Du, Chen, and Jarrett (2014), Mazali (2017), and Ali, Klasa, and Yeung (2009). Thus, as the literature stands now, the evidence for the relationship between industry and stock returns is mixed. With over a decade of new data available, attempting to replicate and extend Hou and Robinson's work now may help to clarify the degree of support for the authors' original findings.

Although this paper (in line with Hou and Robinson) keeps the focus on examining the relationship between market structure and the required rate of return on common equity, it is worth mentioning that there has been research that has examined the relationship between market structure and other types of capital. Namely, there have been plenty of studies that have linked market structure to the cost of debt. For example, Valta (2012) and Platt (2015) both document a negative relationship between the cost of debt and industry concentration. Even so, Choi and Kim (2018) seem to find evidence suggesting the opposite in that credit ratings appear to be negatively related to industry concentration.

[14] Of course, while R&D spending is sure to correlate with innovation, R&D spending is not a metric for innovation itself.

3 Hou and Robinson Revisited

3.1 The Data and Sample Selection

Hou and Robinson take securities at the intersection of the CRSP monthly returns file and the COMPUSTAT annual industrial file[15] between July 1963 and December 2001 that have share codes of 10 or 11[16] traded on either the NYSE, AMEX[17], or NASDAQ. In line with the standard set by Fama and French, Hou and Robinson merge CRSP stock data from July of year t to June of year $t+1$ with COMPUSTAT data for fiscal year $t-1$ (1932). The idea is to assure enough time for accounting information to be reflected in stock returns. Like Fama and French, Hou and Robinson exclude financial firms (SIC codes $6000-6999$ inclusive).[18] Additionally, Hou and Robinson also exclude regulated industries. These industries are taken from Barclay and Smith (1995) and are as follows: "railroads (SIC code 4011) through 1980, trucking (4210 and 4213) through 1980, airlines (4512) through 1978, telecommunications (4812 and 4813) through 1982, and gas and electric utilities (4900 to 4939)" (618). Hou and Robinson remove these industries as "regulated industries may face lower costs of capital because they have lower operating leverage (due to regulated entry and exit), or because their capital structure and/or capital charges are legally constrained" (1933). Finally, Hou and Robinson adjust stock returns for delisting bias in accordance with Shumway (1997). In addition to stock returns and measures for industry concentration (discussed below), the variables Hou and Robinson extract from this data are summed up in Table 1.

3.2 Measuring Industry Concentration

Hou and Robinson measure concentration via the HHI. For the results presented in their paper, Hou and Robinson use 3-digit SIC codes from CRSP for the purposes of assigning firms to industries. SIC codes are four-digit codes used to classify industries. The 3-digit codes are obtained by truncating the fourth digit. As an aside, there appears to be disagreement within the literature on the question of whether to obtain SIC does from CRSP or from COMPUSTAT. Indeed, the choice is not necessarily trivial as there do exist some inconsistencies between the two databases' listed SIC codes.[19] Since Hou and Robinson use CRSP SIC codes, all further mentions of SIC codes in this paper are to be taken as referring to CRSP SIC codes.

[15] CRSP, which stands for "Center for Research in Security Prices", is a database containing historical stock market data. COMPUSTAT is a database containing financial and accounting information for firms throughout the world. Both are operated by Wharton Research Data Services (WRDS) at the University of Pennsylvania and are among the most used databases in financial research.

[16] Share codes indicate the class of a security. For the first digit, a 1 indicates ordinary common shares. For the second digit, a 0 indicates securities which have not been further defined and a 1 indicates securities which need not be further defined. This boils down to share codes of 10 and 11 representing U.S. common stocks. Thus, stocks that originated outside of the states are not included and various types of funds - e.g., exchange traded funds - are also not included.

[17] Since Hou and Robinson, the AMEX has gone through several name changes. In 2008, NYSE Euronext acquired AMEX and renamed it the NYSE Alternext U.S. In 2009, the name changed once more to NYSE AMEX Equities. By 2012 the name had changed to NYSE MKT LLC. Finally, in 2016, it was renamed as NYSE American. For simplicity, this paper refers to it by its original name, AMEX, so as not to cause any confusion when detailing Hou and Robinson's methodology.

[18] This exclusion is made due to financial firms typically having much higher operating leverage than firms in other industries.

[19] See https://wrds-support.wharton.upenn.edu/hc/en-us/articles/115003440892-Explaining-inconsistencies-between-SIC-code-in-CRSP-and-SICH-Standard-Industrial-Classification-Historical-in-Compustat.

To actually compute the HHI, Hou and Robinson use three metrics: net sales, total assets, and book equity. They denote these variants as H(Sales), H(Assets), and H(Equity) respectively. To account for potential data errors, Hou and Robinson take a 3-year moving average of each industry's HHI. Thus, from a computational perspective, these measures are as follows:

$$\text{H(Var)}_{j,t=0} = \frac{1}{3} \sum_{t=-2}^{0} \left[\sum_{i=1}^{N_j} \left(\frac{\text{Var}_{i,j,t}}{\sum_{i=1}^{N_j} \text{Var}_{i,j,t}} \right)^2 \right] \tag{3}$$

where N_j is the number of firms in industry j and $\text{Var}_{i,j,t}$ is the variable (sales, assets, or equity) used for the purposes of calculating market share for firm i in industry j at time t where $t = 0$ denotes the current year. Hou and Robinson that H(Sales), H(Assets), and H(Equity) correlate quite well with another - specifically, all correlations within their correlation matrix exceed 0.95 (Table I Summary Statistics, 1935). Not surprisingly, Hou and Robinson find that their main results are qualitatively similar regardless of the variable used for the purposes of constructing the HHI. As a result, they only present results using H(Sales).

3.3 Methodology

To get a sense of potential links between industry concentration and stock returns, Hou and Robinson first compute industry averages of variables by year. Then, for each year, each securitiy is placed into one of five portfolios based on quintile breakpoints of H(Sales). Hou and Robinson argue that the picture that starts to form from this exercise is "consistent with prior literature" (1936). For instance, (as alluded to earlier), various measures using R&D expenditures are lower for concentrated industries relative to less concentrated industries. Overall, Hou and Robinson argue that their statistics "paint a picture of concentrated industries as innovation-poor, profit-rich industries with high barriers to entry" (1936). Hou and Robinson's original summary statistics as originally published appear in Table 2.

The authors examine these potential channels further by regressing H(Sales) on various industry average characteristics. That is, they estimate equations of form

$$\text{H(Sales)}_{j,t} = \alpha_t + \sum_{n=1}^{N} \lambda_{n,t} X_{j,t} + \epsilon_{j,t} \tag{4}$$

where $X_{j,t}$ denotes industry average characteristics. More precisely, "regressions are run for every year t from 1963 to 2001, and the time-series means of annual cross-sectional coefficient estimates are reported with the time-series t-statistics" (1936). One of their main findings from these regressions includes the observation that "[measures] of profitability are positively correlated with industry concentration" (1936). All together, they take these results as evidence that "concentrated industries have large asset bases and high unit profitability." (1936). The author's original Fama-Macbeth results appear in Table 3.

Panel A of Table 3 is read column by column. For example, the third column of Panel, labeled $\ln(\text{Sales})$, gives estimates for the univariate equation $\text{H(Sales)} = \alpha_t + \lambda_t \ln(\text{Sales})_{j,t} + \epsilon_{j,t}$ where $\ln(\text{Sales})_{j,t}$ is the average of the natural log of sales for industry j at time t. This regression produces a *time-series* of regressions estimates. Specifically, there are regression results for each year t. Thus, if there m years, then there are m regressions. For the July 1963-December 2001 sample period, there are 38 years (462 months) of data. The regressions on H(Sales) are annual, so $m = 38$. The first reported number is the *average* of the coefficients from the 38 regressions. Using the $\ln(\text{Sales})$ as an example again, the reported $\hat{\lambda} = 0.213$ was obtained from $\frac{\sum_{i=1}^{38} \boldsymbol{\lambda}}{38}$ where $\boldsymbol{\lambda}$ denotes the $m \times 1$ column vector of regression coefficients estimated from regression i where $i = 1, 2, \ldots, 38$. The second row of Panel A provides the time-series t-statistic for λ_t. This is obtained by first calculating the time-series standard error per Fama and Macbeth (1973): $\sigma = \sqrt{\frac{1}{38-1} \sum_{i=1}^{38} (\hat{\lambda} - \hat{\lambda}_t)^2}$. From here the t-statistic is computed as usual: $t = \frac{\hat{\lambda}}{\sigma/\sqrt{38}}$. Some readers may be skeptical of computing standard errors "as usual" without any adjustments (namely, Newey-West adjustments) to account for correlated error terms common in time series. Regarding the Fama-Macbeth procedure, the literature appears to be divided as to whether an adjustment is needed. Hou and Robinson state that "[the Fama-Macbeth] procedure allows for multivariate correlation analysis, and [is] robust to cross-correlated error terms. Thus, the resulting coefficients can be interpreted as simple or conditional correlations between concentration and industry-average characteristics, and appropriate statistical inferences can be drawn about the magnitude of these relations" (1936) This study takes this language to mean that the authors *did not* use Newey-West standard errors - meaning this study gathers that Hou and Robinson calculated standard errors as described above.

In contrast, Panel B of Table 3 should be read by rows. Note that each specification has two rows: the first row reports the average of the regression time-series coefficients and the second row gives the time-series t-statistic for the average coefficient. Blank cells indicate that the variable was not part of the particular specification. For example, rows 3 and 4 are the results for the equation H(Sales) = $\alpha_t + \lambda_{1,t} \ln(\text{Assets})_{j,t} + \lambda_{2,t}\text{E/A}_{j,t} + \lambda_{3,t}\text{R\&D/A}_{j,t} + \lambda_{4,t}\text{Leverage}_{j,t} + \lambda_{5,t}\ln(\text{B/M})_{j,t} + \lambda_{6,t}\text{Beta}_{j,t} + \epsilon_t$. The average time-series coefficient and t-statistic for each λ are obtained as described above in the description of Panel A.

To test the relationship between stock returns and industry concentration, the authors form ten portfolios: five portfolios at the *industry* level and five portfolios at the *firm* level based on the HHI. For raw returns, Hou and Robinson observe a significantly negative spread of -0.26% between the least concentrated quintile and the most concentrated quintile for both the industry-level portfolios and the firm-level portfolios.[20] (1938).

To further test the apparent fact that industry concentration appears to be negatively related to stock returns, Hou and Robinson again resort to Fama-MacBeth (1973) style regressions. Specifically, they estimate and present results for monthly regressions of the form:

$$R_{j,t} = \alpha_t + \sum_{n=1}^{N} \lambda_{n,t} X_{j,t} + \epsilon_{j,t} \tag{5}$$

where $R_{j,t}$ is the equal-weighted industry average return for industry j for month t and $X_{j,t}$ is a vector of industry average values of various variables and the HHI.

Hou and Robinson also estimate this regression at the firm level as well:

$$R_{i,t} = \alpha_t + \sum_{n=1}^{N} \lambda_{n,t} X_{i,t} + \epsilon_{i,t} \tag{6}$$

where the i denotes firm i (as opposed to industry j). The main conclusion from these regressions (and arguably the main conclusion from Hou and Robinson's entire paper in general) "is that not only do

[20]The authors demonstrate the robustness of this spread by varying the sample period; using adjusted returns; and by using alternate concentration measures. The High-Low Concentration spread varies between -0.20% and -3.32% for the industry level portfolios and between -0.15% and -3.83% at the firm level across their various specifications and sample periods.

industry returns vary with industry concentration, but so do individual returns: Firms in concentrated industries earn lower stock returns than firms in more competitive industries". As observed in their original results (presented in Table 4), this conclusion is robust across various specifications.

Both panels of Table 4 are read in a similar fashion. Like Panel B of table 3, both panels of Table 4 should be read by rows. Again, each specification gets two rows: the first row reports the average time-series coefficient and the second row reports the t-statistic for this time-series average. The difference between Panel A and Panel B of Table 4 is that Panel A reports results for industry-level regressions (equation (5)) while Panel B reports results for firm-level regressions (equation (6)). Like before, blanks for a particular cell indicate that the variable is not included in that particular specification. For example, rows 13 and 14 of Panel A give the estimates for the industry-level equation $R_{j,t} = \alpha_t + \lambda_{1,t}\ln(\text{Size})_{j,t} + \lambda_{2,t}\ln(\text{B/M})_{j,t} + \lambda_{3,t}\text{Momentum}_{j,t} + \lambda_{4,t}\text{Beta}_{j,t} + \lambda_{5,t}\text{Leverage}_{j,t} + \epsilon_t$.

3.4 The Potential Issue of Cash Flow Surprises

Hou and Robinson express concern that the true driver behind findings may pertain to "persistent differences in cash flow surprises across industries with different market structure" (1942). Although this paper does not provide replication of their investigation into this issue,[21] it is worth summarizing their methodology they use to examine this issue.

To test the possibility that cash flow surprises (which need not persist in the future) is the reason for their results, Hou and Robinson estimate and extend the Fama and French (2000) profitability model by following Vuolteenaho (2002). That is, the authors estimate the following:

$$\frac{E_t}{A_t} = \alpha_0 + \alpha_1 \frac{V_t}{A_t} + \alpha_2 DD_t + \alpha_3 \frac{D_t}{B_t} + \alpha_4 \frac{E_{t-1}}{A_{t-1}} + \epsilon_t \quad (7)$$

where "E/A is earnings scaled by total assets, V/A is the ratio of market value of assets to book assets, DD is a dummy variable for non-dividend-paying firms and D/B is the ratio of dividend payments to book equity. Expected profitability is the fitted value from this regression and unexpected profitability is regression error." (1943).

Using the results from estimates of this model, Hou and Robinson then relate the regression errors to industry concentration. They note that if "[their] results were driven by cash flow shocks, then we should expect to see large positive average profitability shocks" for the least concentrated industries and "large negative shocks" for the most concentrated industries (1943). Hou and Robinson actually observe the opposite, noting that "Concentrated industries have experienced better-than-expected profitability over the 1963 to 2001 period, while competitive industries have experienced poorer-than expected profitability" (1943). As a result, the authors conclude that cash flow surprises do not explain their findings. Furthermore, given that the shocks occur "in the opposite direction of the return spread", Hou and Robinson argue "that the true spread in expected returns is more pronounced than the spread that we observe in the data" (1946).

[21] This paper is in and of itself a robustness check of their findings.

Table 1: Variable Definitions

Variable	Source	Definition	COMPUSTAT Item Numbers
Size	CRSP	$(PRC_{\text{June}})(SHROUT_{\text{June}})$	
Momentum	CRSP	Past-year cumulative stock return	
Total Assets (AT)	COMPUSTAT	AT	6
Sales (SALE)	COMPUSTAT	$SALE$	12
Research/Development Expenditure (XRD)	COMPUSTAT	XRD	46
Research to Assets (R&D/A)	COMPUSTAT	XRD/AT	46/6
Earnings (EARN)	COMPUSTAT	$IB + XINT + TXDI$	$118 + 15 + 50$
Earnings to Assets (E/A)	COMPUSTAT	$EARN/AT$	$(118 + 15 + 50)/6$
Earnings to Sales (E/S)	COMPUSTAT	$EARN/SALE$	$(118 + 15 + 50)/12$
Shareholder's Equity (SHE)	COMPUSTAT	SEQ OR if not available,	216
		$CEQ + PSTK$ OR if not available	$60 + 130$
		$AT - (LT + MIB)$	$6 - (181 + 38)$
Preferred Stock Par Value (PS)	COMPUSTAT	$PSTKRV$ OR if not available	56
		$PSTKL$ OR if not available	10
		$PSTK$	130
Book Equity (BE)		$SHE - PS + TXDITC - PRBA$	$SHE - PS + 35 - 330$
COMPUSTAT Value of Equity (CCAP)	COMPUSTAT	$(PRCC)(CSHO)$	(24)(25)
COMPUSTAT Value of Firm (CVAL)	COMPUSTAT	$CCAP + AT - BE$	$(24)(25) + 6 - BE$
Book to Market (B/M)	COMPUSTAT	$BE/CCAP$	$BE/((25)(24))$
Dividends (DIV)	COMPUSTAT	$DVC + DVP$	$21 + 19$
Dividends to Book (D/B)	COMPUSTAT	DIV/BE	$(21 + 19)/BE$
Value to Assets (V/A)	COMPUSTAT	$CVAL/AT$	$((24)(25) + 6 - BE)/6$
Leverage (Lev)	COMPUSTAT	$(AT - BE)/(CVAL)$	$(6 - BE)/((24)(25) + 6 - BE)$
Beta		Fama and French (1992)	

Definition Variables: PRC_{June} is stock price in June. $SHROUT_{\text{June}}$ is shares outstanding in June. IB is income before extraordinary items. XINT is interest expense. TXDI is income statement deferred taxes. SEQ is stockholder's total equity. CEQ is common equity. PSTK is total preferred stock. LT is total liabilities. MIB is minority interest. PSTKRV is redemption value of preferred stock. PSTKL is liquidating value of preferred stock. TXDITC is balance sheet deferred taxes and investment tax credit. PRBA is post-retirement benefit asset. DVC is common dividends. DVP is preferred dividends. PRCC is annual close price. CSHO is common shares outstanding. Beta is post-ranking beta as in Fama and French (1992).

Table 2: Hou and Robinson's Summary Statistics As Published (1963-2001)

Panel A: Summary of Industry Concentration Measures

	Mean	Median	SD	Max	Min	20%	40%	60%	80%	H(Sales)	H(Assets)	H(Equity)
H(Sales)	0.544	0.490	0.310	1.000	0.025	0.231	0.385	0.611	0.944	1.000	0.976	0.951
H(Assets)	0.539	0.499	0.307	1.000	0.024	0.233	0.397	0.618	0.936	0.976	1.000	0.964
H(Equity)	0.546	0.502	0.308	1.000	0.024	0.230	0.405	0.609	0.931	0.953	0.966	1.000

Panel B: Characteristics of H(Sales) Sorted Quintile

Rank	H(Sales)	Newlist	Delists	Size	Asset	Sales	E/A	E/S	V/A	D/B	R&D	R&D/A	Lev.	B/M	Beta
Low	0.133	267.40	214.60	531.3	1200.4	582.5	0.013	0.110	1.293	0.026	35.293	0.075	0.437	0.798	1.579
2	0.287	126.21	84.70	527.8	645.1	509.6	0.029	0.111	1.257	0.024	21.226	0.060	0.399	0.742	1.632
3	0.470	60.47	42.70	607.4	1204.8	786.7	0.036	0.116	1.327	0.031	21.759	0.040	0.432	0.809	1.595
4	0.745	41.51	23.82	606.4	1087.9	629.3	0.038	0.124	1.558	0.041	17.164	0.037	0.428	0.787	1.606
High	0.982	20.13	8.68	431.3	1604.9	717.6	0.037	0.136	1.696	0.036	13.059	0.027	0.421	0.767	1.609

"The sample includes all NYSE/AMEX/NASAQ-listed securities with share codes 10 or 11 that are contained in the intersection of the CRSP monthly returns file and the COMPUSTAT industrial annual file between July 1963 and December 2001. Panel A reports summary statistics of industry concentration measures for three-digit SIC industries. The H(Sales) for an industry is formed by first calculating the sum of squared sales-based market shares of all firms in that industry in a given year and then averaging over the past 3 years. H(Assets) and H(Equity) are computed analogously, using total assets and book equity in place of sales. The right-most columns present Spearman and Pearson correlations between industry concentration measures. Spearman (rank) correlations are presented below the main diagonal, Pearson above. Panel B reports average characteristics of quintile portfolios sorted by H(Sales). Quintile 1 corresponds to the 20% of industries with the lowest concentration, while Quintile 5 corresponds to the 20% of industries with the highest concentration." - Hou and Robinson (1935)

Table 3: Hou and Robinson's Fama-Macbeth Regressions of H(Sales) As Published

Panel A: Simple Regressions

ln(Size)	ln(Assets)	ln(Sales)	E/A	E/S	V/A	D/B	R&D/A	Leverage	ln(B/M)	Beta
-0.040	-0.032	-0.043	0.179	0.213	0.014	-0.002	-0.984	-0.033	-0.028	0.091
-13.45	-17.97	-22.81	4.77	7.15	3.71	-0.02	-4.97	-1.73	-5.34	4.92

Panel B: Multiple Regressions

ln(Size)	ln(Assets)	ln(Sales)	E/A	E/S	V/A	D/B	R&D/A	Leverage	ln(B/M)	Beta
-0.034			0.522				-1.433	-0.056	-0.057	-0.011
-3.89			3.28				-6.99	-0.84	-6.08	-0.23
	-.027		0.525				-1.527	0.026	-0.044	0.021
	-3.26		3.32				-7.22	0.44	-3.75	0.46
		-0.27	0.489				-1.514	0.012	-0.036	0.021
		-4.09	2.50				-7.26	0.20	-3.35	0.55
-0.039				0.580			-1.399	-0.017	-0.056	-0.016
4.36				7.57			-6.81	-0.24	-5.80	-0.38
-0.023					0.024		-1.487	-0.114		0.019
-1.91					4.23		-7.05	-1.80		0.36
-0.34			0.542			0.570	-1.361	-0.043	-0.044	0.011
-3.83			3.21			2.97	-6.75	-0.64	-3.83	0.22

"This table presents Fama-Macbeth regressions of the H(Sales) index with other industry characteristics. The variables are defined according to Table 1. Every year, a cross-section regression is estimated. The time series mean of the annual regression coefficients and the time-series *t*-statistics (appearing below) are reported. In Panel A, each coefficient is obtained from a simple (univariate) regression of H(Sales) on each characteristic alone. Panel B reports the results of multiple (multivariate) regressions of H(Sales) on a series of industry characteristics." - Hou and Robinson (1937)

Table 4: Hou and Robinson's Fama-Macbeth Regressions of Returns (1963-2001) As Published

Panel A: Simple Regressions

H(Sales)	ln(Size)	ln(B.M)	Momentum	Beta	Leverage
-0.30					
-2.41					
	-0.12				
	-1.54				
		0.39			
		4.16			
			1.03		
			4.21		
				-0.18	
				-0.43	
					0.98
					2.96
	-0.24	0.28	0.95	-0.95	0.08
	-2.81	2.77	4.40	-2.56	0.24
-0.30	-0.12	0.29	0.90		
-2.58	-1.57	3.07	3.93		
-0.31	-0.25	0.27	0.94	-1.00	0.04
-2.85	-2.98	2.68	4.36	-2.73	0.12

Panel B: Firm-Level Regressions

H(Sales)	ln(Size)	ln(B.M)	Momentum	Beta	Leverage
-0.35					
-2.41					
	-0.14	0.35	0.56		
	-2.62	4.55	3.34		
-0.44	-0.14	0.35	0.55		
-3.75	-2.63	4.62	3.32		
				0.26	
				0.90	
					0.76
					2.81
	-0.18	0.38	0.60	-0.39	-0.30
	-3.78	6.41	3.81	-1.87	-1.56
-0.42	-0.18	0.39	0.59	-0.41	-0.33
-3.42	-3.81	6.62	3.78	-1.95	-1.70

Panel A presents industry-level monthly regressions of industry average returns on industry average values and Panel B presents firm-level monthly regressions of individual returns on individual firm characteristics from July 1963 to December 2001.

4 Replicating Hou and Robinson

This section is dedicated to detailing this study's attempt at replicating some of Hou and Robinson's original results. The overall goal of this study is to test whether Hou and Robinson's main results (i.e., those presented in Table IV) will hold up to over a decade's worth of new out-of-sample data. Thus, it is pertinent to replicate Hou and Robinson's original results as closely as possible - otherwise, extensions would be meaningless.

4.1 The Data and Sample Selection

This study attempts to follow Hou and Robinson as closely as possible. The hardest part of this task concerns the job of constructing a sample that approximates their sample as closely as possible. In general, comments on data work tend to be disregarded as not interesting. For replication transparency, though, it is worth it to spend a page or two describing the process this study conducted as well as the challenges this study faced in replicating Hou and Robinson. The latter part of this crucial: although the authors likely did not intend for ambiguity, there are plenty of points throughout their description that are not completely clear.

To start, like Hou and Robinson, all securities at the intersection of COMPUSTAT and CRSP with share codes of either 10 or 11 and traded on either the NYSE, NASDAQ, or AMEX are taken. The key word here is "intersection": merging COMPUSTAT data with CRSP data is not a trivial task. Based upon informal research done on the internet[22], the "kosher" methodology for merging COMPUSTAT and CRSP appears to be to use a "link table" provided by WRDS. However, this study did not have access to this link table.[23] Hence, this study resorted to the seemingly second-best approach: Using CUSIPs (nine-character alphanumeric identifiers for North American financial securities) to create one's own link table between COMPUSTAT's GVKEY (COMPUSTAT's unique stock identifier) and PERMNO (CRSP's unique stock identifier).

For the uninitiated, a question may be something along the lines of "if it is possible to get a link between COMPUSTAT's identifier and CRSP's identifier with CUSIPs, then why not just use CUSIP as the merging variable?" The issue pertains to the fact that CUSIP will not provide a one-to-one match over time nor across various releases of the data. Hence, getting a match with PERMNO is the ideal solution.

Although the CUSIP-linking method works, there are reasons why it is not the *best* method for merging. Namely, whereas CRSP CUSIPs are historical, COMPUSTAT CUSIPs are "header" CUSIPs in the sense that the listed CUSIP is the most recent CUSIP belonging to a particular GVKEY. Moreover, CUSIPs did not exist until 1968. Hence, matches between CUSIPs, GVKEYs, and PERMNOs prior to 1968 may not be as reliable. Consequently, the CUSIP-GVKEY-PERMNO (CGP) matching method will almost always produce a lower number of unique securities than will a merge via the linking table that WRDS provides. Indeed, it appears that the CGP method will tend to result in roughly 500-1,000 less unique firms compared to a merge via the link table provided by WRDS.[24] Hou and Robinson do not mention the merge methodology they use (not that failure to do so is atypical). Furthermore, Hou and Robinson do not report the average number of unique securities by year. Thus, it is difficult to know whether this study's merge closely approximates their sample or not.

As noted in section 3.1, an additional wrinkle in the merging process concerns the time gap in the merge. Hou and Robinson use the approach made popular by Fama and French (1992): they merge

[22]Because the internet is never wrong (Citation needed).

[23]It is also worth noting that WRDS provides an already merged COMPUSTAT-CRSP dataset. Hence, the ideal methodology would probably be to convince your institution to pay thousands of dollars to subscribe to the merged dataset instead of paying thousands of dollars for access to each database individually.

[24]See http://www.ruidaiwrds.info/data/linking-crsp-and-compustat for a good discussion on these two methodologies.

COMPUSTAT data for fiscal year $t-1$ with CRSP data for July of year t to June of year $t+1$. Hence, years "start" on July 1st and end on June 31st. Thus, for a CRSP observation in January of 2000, 2000 is actually equal to $t+1$ (because the year started during July of 1999). Hence, such an observation would get merged with COMPUSTAT data for fiscal year 1998. In turn, for a CRSP observation in December of 2000, 2000 is equal to t (because the year started during July of 2000). So, this observation would be merged with COMPUSTAT data for fiscal year 1999. To summarize this algorithmically, for a CRSP observation in calendar year t, if the month of the observation is greater than or equal to July, then the observation is merged with its corresponding COMPUSTAT data with fiscal year $t-1$. Otherwise, if the month of the observation is less than July, then the observation is merged with its corresponding COMPUSTAT data with fiscal year $t-2$.

There also appear to be issues with construction of some of Hou and Robinson's variables. In particular, the replication results (see Table 5) suggest a discrepancy concerning variables constructed from book equity (which is itself a variable constructed with COMPUSTAT data - see Table 1). First, it is worth noting that this study used a WRDS research macro[25] that constructed book equity *exactly* as Hou and Robinson construct book equity. However, the COMPUSTAT dataset this study had access to did not appear to contain the post-retirement benefit variable that the authors subtract off. So, it would seem that this study's book equities should be slightly upward biased compared to Hou and Robinson's. Even so, some of the results suggest that there may be more than basic upward bias going on. Part of this can likely be chalked up to this study using the CGP merging method (which again results in fewer unique firms per year). Thus, the sample used for replication is likely smaller compared to Hou and Robinson's original sample. Additionally, it is also possible that discrepancies are due to COMPUSTAT and CRSP backfilling data.[26] That said, it is hard to be sure what the main source of the discrepancies are (more than likely a combination of all of these mentioned potential problems) as Hou and Robinson do not indicate the average number of firms in any of their results.

The final source of uncertainty regarding variable construction concerns the calculation of post-ranking beta. As noted in Table 1, post-ranking beta is constructed according to Fama and French (1992). However, following Fama and French (1992) is not a trivial task. Thus, it is unfortunate that Hou and Robinson do not provide any replications of some of the results from Fama and French (1992).[27] Of course, a full blown appendix of replications would not be expected - but, one simple table would have proved useful to gauge whether one is on the right track.

For readers without much experience with financial topics, post-ranking beta is the CAPM beta. It measures the sensistivity of a stock to movements in entire markets. As noted in the literature review, CAPM and beta have recieved quite a bit of criticism for failing to hold empirically. The "post-ranking" terminology refers to the fact that it has been calculated after sorting stocks based on market size *and* pre-ranking beta where pre-ranking beta is calculated without any sorting or ranking. This two-step methodology is to allow for "variation in beta unrelated to size" (Fama and French 1992).

The full details of the estimation process for post-ranking betas are provided in Appendix A. However, the gist of the process is this: first, one calculates "pre-ranking betas." In order for a security to be included in the pre-ranking beta estimation for a given year, the security must have data on book equity, total assets, and earnings. Additionally, the security must have data on CRSP market equity.[28]

[25]See Glushkov, Denys. "Market-to-Book (M/B) Ratio." WRDS Research Macros, 2018. Wharton Research Data Services, the Wharton School, University of Pennsylvania, https://wrds-www.wharton.upenn.edu/pages/support/applications/risk-and-valuation-measures/market-book-mb-ratio/.

[26]Indeed, as Sursock notes, "[COMPUSTAT] systematically add[s] companies to [its] database and backfill the data. . . Therefore, researchers using the database in 1989 and 1991 will observe different companies in the 'same' database for 1988" (34). This study of course is taking place in 2019. Hou and Robinson's work was published in 2006 - which likely means the bulk of the work took place sometime between 2004 and 2006. Thus, it is certainly possible this study's COMPUSTAT data may vary significantly from the COMPUSTAT data Hou and Robinson used given this time gap.

[27]This study *does* provide such replications in Appendix A. In fact, one could succinctly summarize the main portion of Hou and Robinson's paper as Fama and French (1992) with industry concentration thrown in. Hence, successful replications of Fama and French (1992) are a good indication that the process is being done correctly.

[28]In other words, securities that do not meet these requirements get filtered out.

Once pre-ranking betas are estimated, securities are sorted into ten size portfolios that begin in July of year t and end in June of year $t+1$. Then, these ten size portfolios are sorted in ten-sub portfolios each based off of pre-ranking beta. Thus, one ends up with 100 portfolios. Post-ranking is calculated by regressing equal-weighted portfolio returns on the CRSP value-weighted market portfolio returns.

The main variable of interest, industry concentration, actually proved to be relatively simple to calculate. Like and Hou and Robinson, this study constructed three versions of HHI based on COMPUSTAT net sales, COMPUSTAT total assets, and COMPUSTAT book equity. Also in line with Hou and Robinson, this study calculated the three-year moving average of HHI for all three definitions. Thus, from a computational perspective, this study followed equation (3) for the purposes of constructing the HHI.

A larger issue related to all of the above discussion concerns the sequence of data wrangling - which Hou and Robinson are particularly ambivalent in describing. To see why, for a particular stock to be included in post-ranking beta estimation for year t, it must have COMPUSTAT data on earnings, assets, and book equity for fiscal year $t-1$. Thus, in regards to post-ranking beta estimation, stocks that do not meet these data requirements get filtered out. This begs the question: does one permanently eliminate these stocks for years in which they do not meet COMPUSTAT data requirements? The answer is not clear. While these stocks certainly cannot be used in regression models containing post-ranking betas, they could potentially have the necessary data to be included in regression models that do not include beta. Furthermore, suppose the answer to the former question is yes - i.e., stocks that are not included in post-beta estimation for year t get filtered out. Then, a related question is as follows: should HHI's be based on stocks that met the data requirements (i.e., have a post-ranking beta assigned to them?) Or should HHI's be based on all stocks. Computationally, this question boils down to a question of whether to calculate HHI's first and then estimate betas or whether to calculate betas first and then calculate HHI's. The former results in a larger universe for HHI's while the latter results in a smaller universe. To address these questions, this study tried various permutations of sequencing of data wrangling. However, the core results mostly stayed the same qualitatively - the only exception being that calculating betas first and then calculating HHI's based on the smaller universe results in HHI's that are particularly larger on average compared to Hou and Robinson's original results. At the end of the day, this study felt it prudent to calculate HHI's with a larger universe. That is, HHI's were calculated before any filtering took place.

4.2 Summary Statistics

Table 5 presents this study's replication of Hou and Robinson's Table 1. The first noticeable difference pertains to the fact that this study ended up with lower values of the HHI constructed with sales and assets across the board. For book equity, the values are higher across the board - this may pertain to the fact that this study's book equity values are likely higher than Hou and Robinson due to this study's dataset not having post-retirement benefit assets to subtract off. Regarding the lower values for the sales and assets versions, the most likely reason for the lower values is sample selection. Unlike other variables, HHI is directly dependent on the sample used. More broadly speaking, HHI depends on how a market is defined. Redefining a market ever so slighty in order to include an additional firm could have large implications if the additional firm will command a particularly high or low proportion of the total market share. Once again, knowing the true reasons for discrepancies is difficult given that Hou and Robinson fail to provide information of average number of firms (whether it be per year, per industry, etc.). As for the higher values for HHI constructed with book equity, this is likely to (at least in part) be due to the apparent divergence in this study's book equities versus Hou and Robinson's values for book equities.

Moving on to Panel B, there are even more discrepancies to consider. While the basic trends in many variables are the same as the trends observed in Hou and Robinson's Table I - e.g., H(Sales) increases with concentration; newlists and delists decrease with concentration; assets and sales mostly increases with concentration; E/A and E/S mostly increase with concentration; and R&D/A increases with concentration - there are a few noticeable differences. In particular, whereas size (ln(June Market Equity)) and R&D appear to decrease with concentration in Hou and Robinson's table, the results using this study's sample suggest the opposite. Yet again, this issue can likely be attributed to sample selection. First, note that the results in Panel B are averages for quintiles. Any considerable difference in sample could very well skew averages. This is indeed what this study believes is the reason for the seemingly "incorrect" results. Furthermore, note that variables constructed as ratios tend to perform much better in matching up with Hou and Robinson. For example, while the replicated panel B might suggest that R&D spending increases with concentration, the ratio R&D spending to Assets decreases (as it does in Hou and Robinson's table). Furthermore, if industry concentration is truly negatively related to stock returns, then it could be argued that some of the results in the replicated panel B of table 5 make more sense compared to Hou and Robinson. For instance, there is a well documented size effect (see Fama and French (1992)). Specifically, it is generally agreed upon that larger firms require lower returns compared to smaller firms. Thus, seeing size increase with concentration (as the replicated table suggests) would not be at all surprising.

4.3 Fama-Macbeth Annual Regressions of H(Sales) on Industry Average Characteristics

Given that the discrepancies in some of the summary statistics, it is not surprising that this study ended up with some discrepancies in the results for the Fama-Macbeth regressions of H(Sales) - which are presented in Table 6 which may be interpreted and read analogous to how Table 3 is interpreted. However, Table 6 also provides the average amount of securities in the cross-sectional regressions as well as the average R^2 of the cross-sectional regressions - of which there are 38 since one regression is run of each year and the sample period is 1963-2001). For the simple regressions in Panel A, average n is reported in the third row and average R^2 is reported in the fourth row. This n *is not* the number of coefficients from the time-series of regression results (again, there are 38 regressions so there are 38 “observations” of regression results). Rather, each of the 38 regressions has an n. Letting $\mathbf{n}$ denote the 38×1 column vector of regression ns, then the reported n is obtained from $\frac{\sum_{i=1}^{38} \mathbf{n}}{38}$ where $i = 1, 2, \ldots, 38$. The reported R^2 is computed analogously with a vector of regression R^2s replacing n. While averaging R^2s may sound questionable, this technique has been performed in the literature - for example, Lewellen (2015) reports average R^2s of Fama-Macbeth style regressions. Panel B also reports average n and average R^2 - however, they are given their own columns, labeled n and R^2 respectively.

The variables that appear most robust to sample are E/A, E/S, and R&D/A. While there are certainly discrepancies in their coefficients compared to Hou and Robinson’s results (which were reported in Table 6), the main interpretation from them is qualitatively the same. Unsurprisingly, values relating to book equity do not perform well (likely due to reasons already discussed above). Additionally, ln(Size), ln(Assets) and leverage all have the “wrong” sign compared to Hou and Robinson. However, the coefficients in Table 6 table are statistically insignificant - a fact that gives even further credence to the theory that this study’s sample must differ significantly from Hou and Robinson’s sample for reasons discussed before.

The worst offender compared to Hou and Robinson’s original results concerns beta. Whereas Hou and Robinson find that beta seems to be unrelated to industry concentration, this study finds that it is significantly negatively related. This in and of itself is not a particularly surprising result, though. At the end of the day, stocks that receive higher returns should be stocks with higher risk. Thus, if one believes industry concentration to be negatively related to stock returns, then one should believe that less concentrated industries entail more risk. In turn, finding that less concentrated industries appear to have significantly higher betas compared to more concentrated industries should not provide a shock.

4.4 Fama-Macbeth Monthly Regressions of Stock Returns

Given the discrepancies discussed above, the replications of Hou and Robinson’s cross-sectional regressions on industry-level and firm-level returns (presented in Table 7) are reassuring. Table 7 may be interpreted and read the same as Table 4. However, just as Table 6 added in average n and average R^2, so does Table 7. For these averages, the number of regressions ran totals 462 (the regressions are run monthly and there are 462 months of data in the July 1963 to December 2001 sample period). In Panel A, n represents the average number of *industries* present in the 462 regressions. In turn, in Panel B, n represents the average number of unique *firms* present in the 462 regressions.

Ignoring H(Sales) for the moment, for the most part, coefficients on ln(Size), ln(B/M), and leverage are in line with expectations and are qualitatively similar to Hou and Robinson’s original results for both the industry and firm level - even if their is some discrepancy regarding magnitudes. Results for beta are also qualitatively similar in that beta is insignificant. Momentum has coefficients in the correct direction, however the magnitudes seem a bit off. Furthermore, the replication results suggest momentum is statistically significant whereas Hou and Robinson’s results suggest that momentum is not significant.

The main column of interest, of course, is the H(Sales) column. And, while there are some issues with magnitude (albeit relatively small ones), the replicated H(Sales) column suggests that Hou and Robinson's original results may have been no fluke. Even despite the obvious discrepancy in samples that must exist, this study was able to replicate a competition premium that appears to hover in realm of 0.20% to 0.40%. That is, like Hou and Robinson, the results of Table 7 suggest that common equity, will, on average, cost about 0.20% to 0.40% more for firms operating in less concentrated (i.e., more competitive) industries. Of course, this result is only for the July 1963 to December 2001 period.

Since the 462 Fama-Macbeth regressions produce a time-series of regression results, it is possible to examine the estimated competition premium month-by-month. Each of these regressions that includes H(Sales) produces a coefficient for H(Sales) for the corresponding month. Denote this 462×1 column vector of regression coefficients on H(Sales) as $\boldsymbol{\lambda}_{HHI}$. Based on Hou and Robinson, the expectation is that the average value of $\boldsymbol{\lambda}_{HHI}$ is less than 0 - meaning that an increase in concentration results in a decrease in the cost of common equity (average stock returns). Hence, left as is, $\boldsymbol{\lambda}_{HHI}$ is effectively a vector of concentration discounts (at least on average). To achieve the competition premium interpretation (an increase in competition results in an increase in the cost of equity), one simply multiplies the vector by -1. Denote this transformed column vector as $\boldsymbol{\rho}$ where $\boldsymbol{\rho} = -1 \cdot \boldsymbol{\lambda}_{HHI}$. The average value of $\boldsymbol{\rho}$ should be greater than 0 and thus may be interpreted as a competition premium. Figure 1 plots the values for the industry-level $\boldsymbol{\rho}$ over time using the industry-level $\boldsymbol{\lambda}_{HHI}$ produced by the specification presented in the final two rows of Panel A (the specification that includes all the controls). In turn, Figure 2 plots the values for the firm-level $\boldsymbol{\rho}$ over time using the firm-level $\boldsymbol{\lambda}_{HHI}$ produced by the specification presented in the final two rows of Panel B (the specification that includes all the controls).

Based on these figures, the industry-level premium appears to be more volatile than the firm-level premium. An additional observation is that the premium (whether at the industry-level or firm-level) is not perpetually positive. Rather, it fluctuates around zero quite a bit. Of course, for the 1963-2001 period the competition premium is significantly positive based on both Hou and Robinson's original results and the replicated results. Even so, the visual that figures 1 and 2 provide may suggest that the result may not have much practical significance.

Table 5: Replication of Hou and Robinson's Summary Statistics

Panel A: Summary of Industry Concentration Measures

	Mean	Median	SD	Max	Min	20%	40%	60%	80%	H(Sales)	H(Assets)	H(Equity)
H(Sales)	0.471	0.405	0.275	1.000	0.034	0.217	0.334	0.487	0.746	1.000	0.971	0.794
H(Assets)	0.478	0.418	0.274	1.000	0.043	0.220	0.347	0.501	0.754	0.967	1.000	0.807
H(Equity)	0.622	0.610	0.281	1.000	0.059	0.333	0.515	0.716	0.962	0.814	0.825	1.000

Panel B: Characteristics of H(Sales) Sorted Quintile

Rank	H(Sales)	Newlist	Delists	Size	Asset	Sales	E/A	E/S	V/A	D/B	R&D	R&D/A	Lev.	B/M	Beta
Low	0.137	151.18	161.51	897.3	626.9	676.5	0.068	0.111	2.672	2.005	26.220	0.091	0.521	0.402	1.200
2	0.274	73.23	68.07	714.2	610.5	667.7	0.074	0.075	2.344	1.241	23.539	0.062	0.554	0.430	1.212
3	0.400	36.38	34.03	657.3	898.8	858.8	0.078	0.095	2.032	1.383	36.443	0.032	0.584	0.498	1.219
4	0.601	23.12	23.74	1003.6	1238.7	969.1	0.076	0.083	2.047	1.492	42.886	0.039	0.581	0.516	1.182
High	0.878	12.69	9.25	1486.1	1727.9	1305.1	0.082	0.105	2.029	2.470	35.398	0.027	0.589	0.750	1.162

All variables constructed as defined in Table 1 with the exception of Book Equity, which does include post-retirement asset. Note that this difference may also affect values of Dividends-to-Book (D/B), Leverage (Lev), and Book-to-Market (B/M).

Table 6: Replication of Hou and Robinson's Fama-Macbeth Regressions of H(Sales)

Panel A: Simple Regressions

ln(Size)	ln(Assets)	ln(Sales)	E/A	E/S	V/A	D/B	R&D/A	Leverage	ln(B/M)	Beta
-0.005	0.004	-0.006	0.282	0.144	-0.029	003	-1.241	0.136	0.030	-0.126
-1.04	0.78	-1.18	1.68	4.34	-6.39	2.22	-9.21	3.74	6.70	-3.33
258	258	258	258	258	254	228	177	256	224	213
0.03	0.02	0.03	0.01	0.004	0.01	0.01	0.03	0.01	0.01	0.02

Panel B: Multiple Regressions

Spec	ln(Size)	ln(Assets)	ln(Sales)	E/A	E/S	V/A	D/B	R&D/A	Leverage	ln(B/M)	Beta	n	R^2
1	0.004			0.576				-0.758	0.108	0.007	-0.178		
	0.85			2.91				-6.02	1.80	0.89	-4.31	155	0.09
2		0.004		0.575				-0.759	0.091	0.006	-0.180		
		1.00		2.97				-5.85	1.68	0.83	-4.45	155	0.08
3			-0.012	0.800				-0.771	0.145	0.009	-0.217		
			-2.42	3.93				-5.91	2.58	1.39	-5.58	155	0.09
4	0.003				0.345			-0.780	0.091	0.005	-0.178		
	0.64				2.84			-4.98	1.32	0.62	-4.19	155	0.09
5	0.006					-0.027		-0.800	0.029		-0.169		
	1.33					-2.00		-6.05	0.28		-3.74	158	0.08
6	0.004			0.434			0.006	-0.747	0.077	0.015	-0.161		
	0.71			2.36			3.98	-6.15	1.23	1.91	-3.81	154	0.10

All variables are as defined in Table I with the exception of Book Equity, which does include post-retirement asset. Note that this difference may also affect values of Dividends-to-Book (D/B), Leverage (Lev), and Book-to-Market (B/M). In Panel A, the first row gives the coefficient. The second row gives the time-series t-statistic. The third statistic gives average n. Finally, the fourth row provides average R^2. In Panel B, the first row gives the average coefficient and the second row gives the t-statistic along with average n and average R^2 in the two right-most columns.

Table 7: Replication of Hou and Robinson's Fama-Macbeth Regressions of Returns

Panel A: Industry-Level Regressions

Spec	H(Sales)	ln(Size)	ln(B/M)	Momentum	Beta	Leverage	n	R^2
1	-0.223							
	-1.60						208	0.01
2		-0.120						
		-2.12					208	0.03
3			0.154					
			3.51				199	0.01
4				0.16				
				0.72			207	0.02
5					0.32			
					0.07		208	0.02
6						1.42		
						2.89	207	0.02
7		-0.11	0.084	0.239	-0.460	0.302		
		-2.12	1.72	1.72	-1.07	0.68	198	0.09
8	-0.310	-0.111	0.136	0.232				
	-2.39	-2.08	3.22	1.19			198	0.06
9	-0.348	-0.107	0.095	0.225	-0.490	0.267		
	-2.70	-2.08	1.96	1.23	-1.14	0.56	198	0.09

Panel B: Firm-Level Regressions

Spec	H(Sales)	ln(Size)	ln(B/M)	Momentum	Beta	Leverage	n	R^2
1	-0.23							
	-1.43						1306	0.003
2		-0.14	0.09	0.18				
		-3.20	2.77	1.23			1111	0.03
3	-0.32	-0.13	0.09	0.18				
	-2.270	-3.16	2.93	1.22			1111	0.04
4					0.24			
					0.65		1306	0.01
5						1.73		
						4.55	1279	0.01
6		-0.11	0.02	0.18	-0.16	0.88		
		-2.91	0.62	1.37	-0.51	2.84	1111	0.05
7	-0.35	-0.11	0.02	0.18	-0.16	0.90		
	-2.68	-2.84	0.68	1.36	-0.54	2.94	1191	0.05

All variables are as defined in Table I with the exception of Book Equity, which does include post-retirement asset. Note that this difference may also affect values of Dividends-to-Book (D/B), Leverage (Lev), and Book-to-Market (B/M). Average n and average R^2 appear in the two right-most columns.

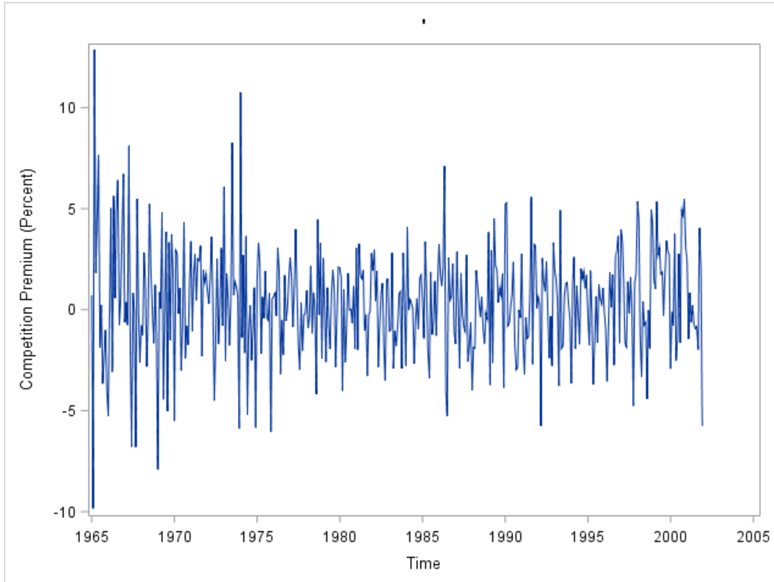

Figure 1: The Competition Premium (Industry Level) over Time: Hou and Robinson Period

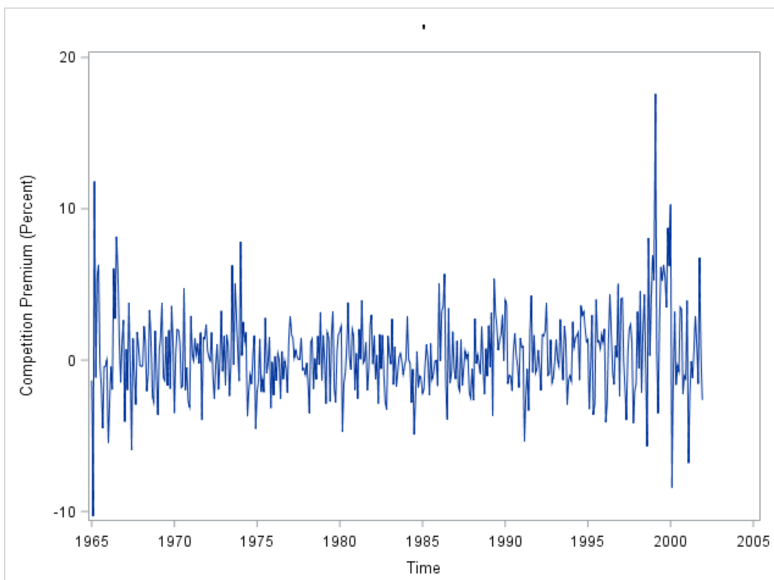

Figure 2: The Competition Premium (Firm Level) over Time: Hou and Robinson Period

5 Extending Hou and Robinson

This section presents extensions of Hou and Robinson's original paper. First, trends in the HHI are examined. Following this, an informal bivariate analysis and visualization is performed to hypothesize whether the competition premium has persisted since the Hou and Robinson period. From there the study formally extends Hou and Robinson's main model to new data. To finish up, competition premiums across quintiles are examined.

5.1 Trends in the HHI

Prior to doing any formal tests to examine whether a competition premium still persists, it will be useful to first examine how industry concentration has changed over time - especially since the Hou and Robinson 1963-2001 time period. It is worth noting that Grullon, Larkin, and Michaely (2018) find that "[m]ore than 75% of US industries have experienced an increase in concentration levels over the past two decades" (2).

This study finds a similar result: Figure 3 plots out the mean, median, and minimum of the HHI (constructed with H(Sales) at the SIC-3 level) over time. While the minimum HHI has remained flat over time, the mean and medians have gradually increased since the 1980s. Additionally, Figure 4 plots the behavior of the mean HHI within quintiles across time. The most concentrated quintile (quintile 5) has remained relatively flat (hovering close to 1.0). Flatness is also seen in the least concentrated quintile (quintile 1) as well - which has hovered just below 0.20). However, the middle three quintiles all exhibit the gradual increase since the 1980s that is evident in Panel B.

Figure 3: HHI Mean, Median, and Minimum over Time

That average concentration has been increasing since the 1980s - especially since the Hou and Robinson period (which ends in 2001) - may have ramifications for the competition premium. The exact direction, however, is not immediately obvious. On one hand, increasing average concentration may diminish the competition premium as investors may struggle to perceive differences in risk as industries converge towards concentration. At the same time, increasing average concentration may magnify the competition premium as the difference between the fewer remaining less concentrated industries may be even more stark compared to the concentrated norm.

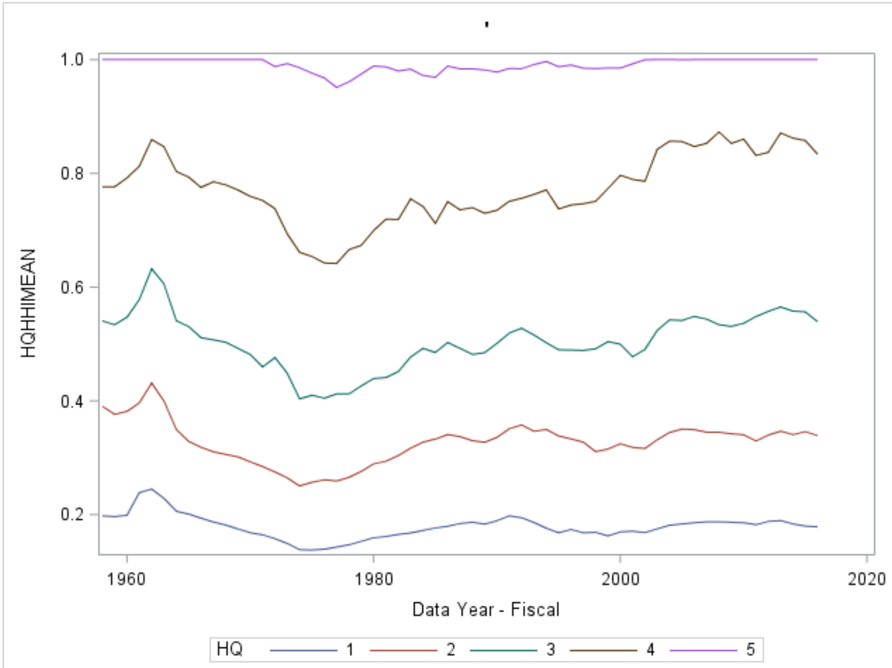

Figure 4: HHI Mean Within Quintiles Over Time

5.2 Visualizing the HHI and Stock Returns

To get a sense of whether or not the competition premium diminished (or magnified) since the Hou and Robinson time period, visualizations of the HHI and stock returns are performed. This basic bivariate visualization exercise, while informal, may provide the grounds to make an educated guess as to whether the competition premium has persisted.

Figure 5 presents a contour plot of stock returns (restricted between -30% and 30%) and H(Sales) HHI[29] for stocks within an industry with an HHI less than 0.5 for the Hou and Robinson period. In turn, Figure 6 presents a contour plot of stock returns (restricted between -30% and 30%) and HHI for stocks within an industry with an HHI greater than 0.5 for the Hou and Robinson period. Figures 7 and 8 present corresponding contour plots for data from 2001 through 2018.

Two things immediately stick out upon glancing through Figures 5-10. First, it appears (as would be expected if a competition premium does exist) that returns for securities in less concentrated are slightly more dispersed compared to securities in more concentrated industries. Second, given that this observation is evident regardless of the sample time-period, these figures provide evidence that the competition premium should persist throughout the 2000-2018 period. In particular, figures 7 and 8 suggest the competition premium may have magnified during the 2002-2018 period: the returns of stocks in low-concentration industries are especially dispersed compared to returns of stocks in high-concentration industries. These informal observations are tested explicitly in the following section.

[29]From here on out, HHI should be taken to refer to the version constructed with H(Sales) unless stated otherwise.

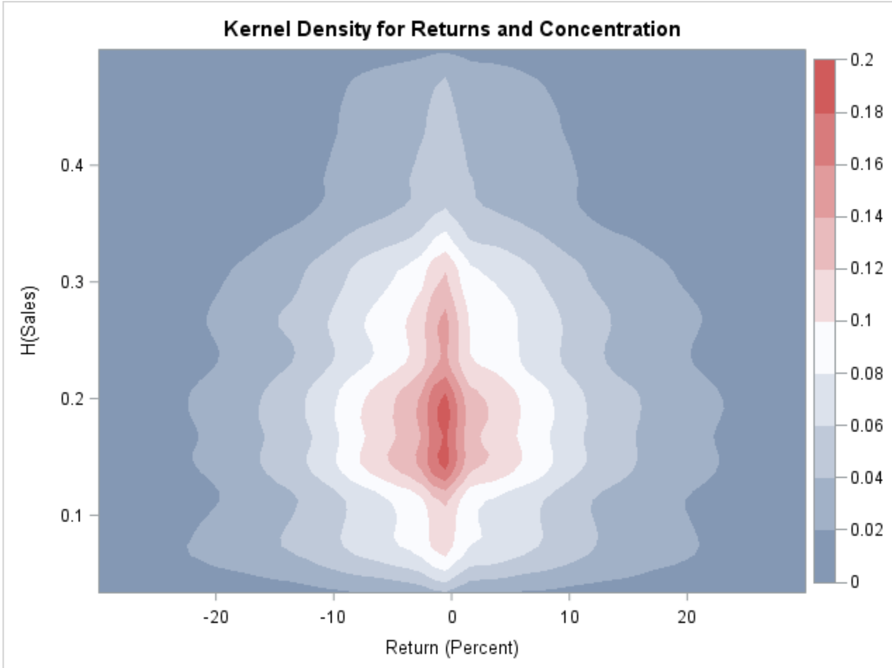

Figure 5: Returns and HHI where HHI < 0.5 for 1963-2001

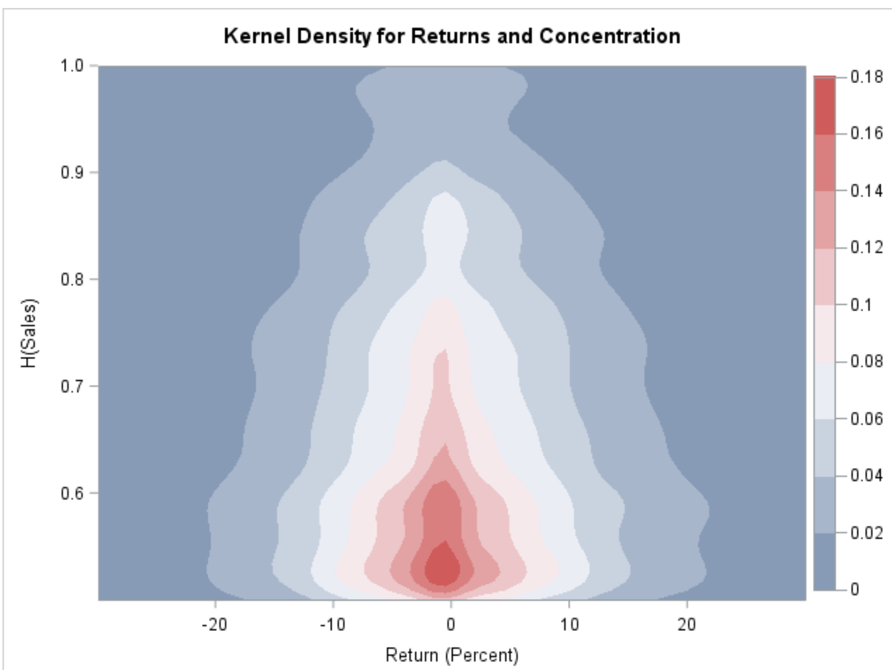

Figure 6: Returns and HHI where HHI > 0.5 for 1963-2001

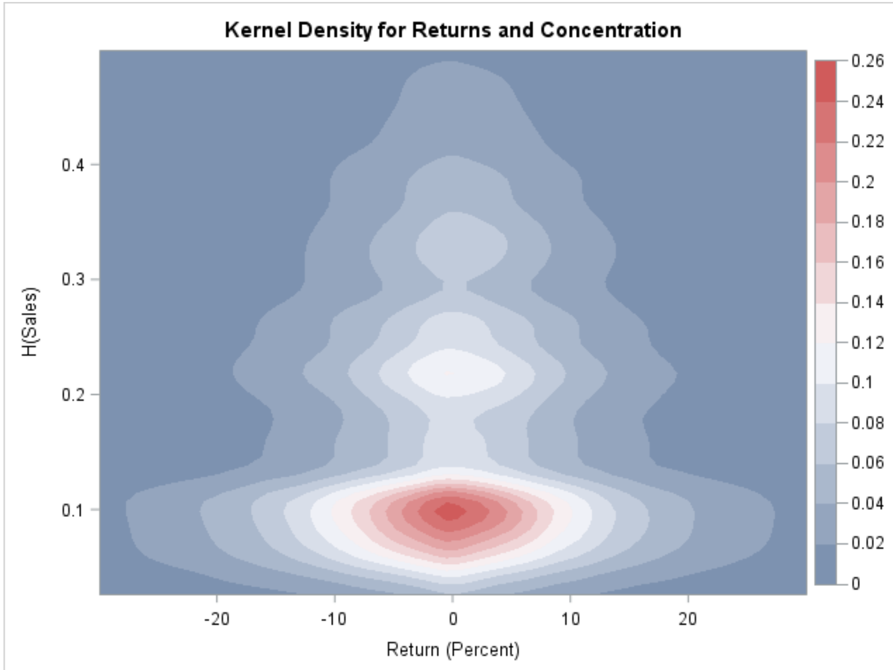

Figure 7: Returns and HHI where HHI < 0.5 for 2002-2018

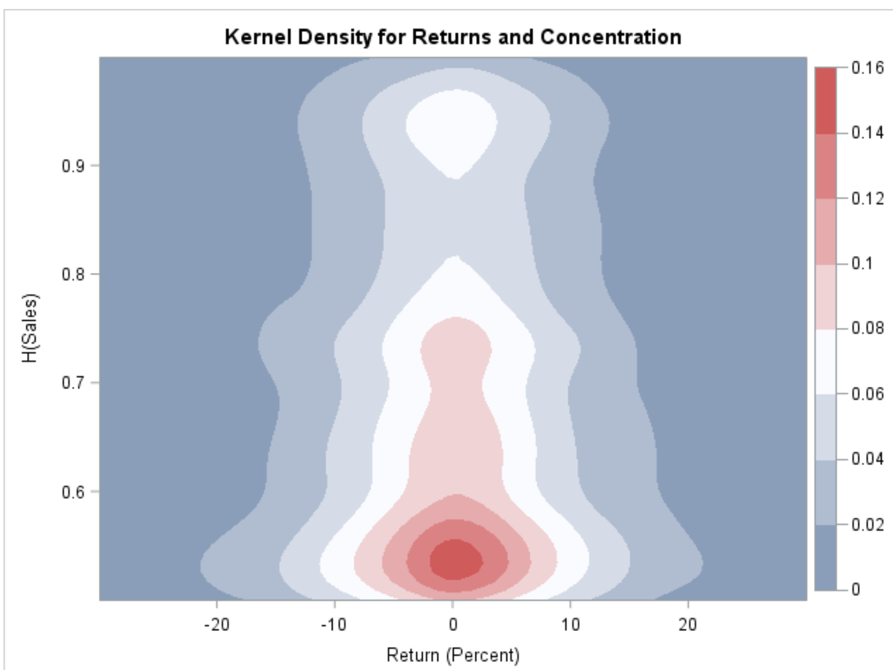

Figure 8: Returns and HHI where HHI > 0.5 for 2002-2018

5.3 The Competition Premium From 2002 to 2018

This section tests the Hou and Robinson model for the 2002-2018 time period. All variables are as in the replication for the 1963-2001 period. Note that summary statistics and Fama-Macbeth regressions of H(Sales) on industry average characteristics are presented in Appendix A. Table 8 presents the results for industry level regressions and firm-level regressions and is read just as Tables 4 and 7 are read. Ignoring the variable of interest (H(Sales)), some side observations pertaining to Table 8 include the following: beta remains remarkably insignificant[30], the size effect still remains in effect (although seemingly insignificant at the industry-level), and the book-to-market effect appears to have diminished in significance.

The results of table 8 suggest an interesting dynamic concerning the competition premium over the 2002-2018 period: at the industry level, the competition premium is statistically insignificant - all the coefficients in all the specifications are negative, but have t-statistics much lower than the -1.96 that would be "comfortable". In contrast, the competition premium at the firm level is not only significant, but it has apparently magnified since the 1963-2001 period. Whereas the average premium across specifications hovered between 0.20% and 0.40% during the 1963-2001 period, the average competition premium suggested by Table 8 appears to be comfortably above and around 0.40%. Moreover, even though none of the coefficients on the HHI achieve a t-statistic of -3.00 (as recommended by Harvey, Liu, and Zhu (2015)), that the premium has appeared to increase while staying significant at the 5% level suggests the premium is robust at the firm level and across time.

Again, it is possible to examine the behavior of the competition premium over time. Figures 9 and 10 are constructed analogously to figures 1 and 2. When examined closely, the plots appear to suggest that over the 2002-2018 period, the firm-level competition premium tends to persist in the positives slightly longer than the industry-level premium. Furthermore, the "spikes" in the premium at the firm-level appear slightly higher compared to the spikes in the industry-level premium over the time-period. Another interesting observation that can be picked up from these plots is the fact that the volatility of the competition premium appears to have increased around the time of the 2007-2008 financial crisis and the resulting Great Recession.

5.4 The Competition Premium for the Entire 1963-2018 time period

As a final test of the competition premium, the entire 1963-2018 sample is used. Once again, all variables are constructed as in the replication of the 1963-2001 period. Summary statistics and Fama-Macbeth regressions of H(Sales) for this period are presented in Appendix A. Table 9 presents the results for the industry-level and firm-level cross-sectional regressions of stock returns. Again, the size effect and the book-to-market premium still stay alive while beta still continues to disappoint.

More importantly (for this paper), Table 9 presents strong evidence against the existence of a competition premium at the at the industry level - the coefficient on H(Sales) in the univariate model is *positive* - suggesting an increase in stock returns when concentration increases! While the coefficients in other models exhibit the expected negative sign, the t-statistics are dismally small in magnitude. However, once again the firm-level regressions appear to redeem the existence of a competition premium. All the average coefficients across specifications are negative. Furthermore, with the exception of the univariate model, the t-statistics are large enough in magnitude to reject the null of no competition premium at the firm level.

[30]Sorry, CAPM and beta fans.

Table 8: Fama-Macbeth Regressions of Returns (2002-2018)

Panel A: Industry-Level Regressions

Spec	H(Sales)	ln(Size)	ln(B/M)	Momentum	Beta	Leverage	n	R^2
1	-0.14							
	-0.89						227	0.01
2		-0.07						
		-1.42					227	0.01
3			-0.07					
			1.45				215	0.01
4				-0.43				
				-1.59			227	0.02
5					0.40			
					0.89		227	0.02
6						1.12		
						1.72	227	0.02
7		-0.05	0.03	-0.30	-0.04	0.65		
		-0.86	0.50	-1.41	-0.08	0.89	215	0.07
8	-0.12	-0.08	0.06	-0.36				
	-0.73	-1.58	1.29	-1.37			215	0.05
9	-0.09	-0.05	0.03	-0.30	-0.04	0.65		
	-0.59	-0.89	0.50	-1.38	-0.09	0.90	215	0.07

Panel B: Firm-Level Regressions

Spec	H(Sales)	ln(Size)	ln(B/M)	Momentum	Beta	Leverage	n	R^2
1	-0.44							
	-2.22						1384	0.002
2		-0.10	0.07	-0.422				
		-2.60	1.81	-2.50			1150	0.02
3	-0.40	-0.11	0.08	-0.42				
	-2.22	-2.62	1.92	1-2.47			1150	0.02
4					0.57			
					1.49		1384	0.01
5						1.47		
						2.37	1351	0.01
6		-0.09	0.04	-0.38	-0.09	0.47		
		-2.00	1.24	-2.65	-0.05	0.90	1150	0.04
7	-0.42	-0.09	0.04	-0.38	-0.47	0.55		
	-2.34	-2.01	1.25	-2.63	-0.13	1.03	1150	0.04

All variables are as defined in Table I with the exception of Book Equity, which does include post-retirement asset. Note that this difference may also affect values of Dividends-to-Book (D/B), Leverage (Lev), and Book-to-Market (B/M).

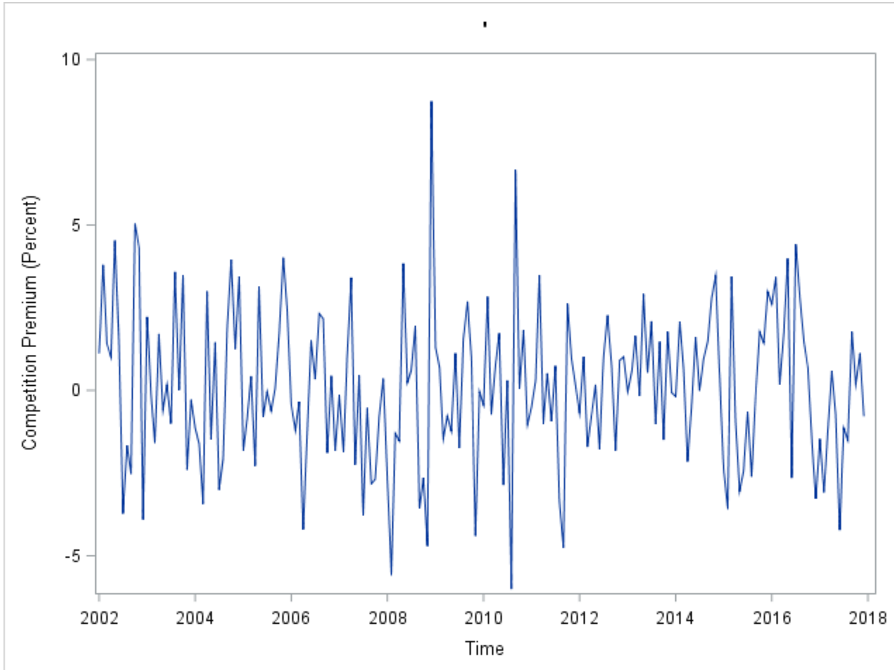

Figure 9: The Competition Premium (Industry Level) over Time: 2002-2018 Period

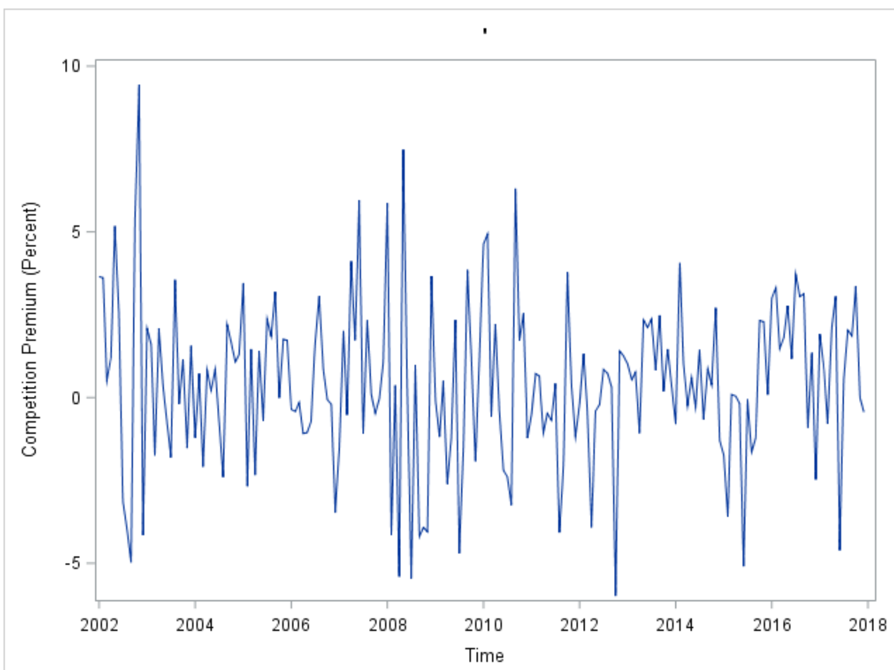

Figure 10: The Competition Premium (Firm Level) over Time: 2002-2018 Period

Table 9: Fama-Macbeth Regressions of Returns (1963-2018)

Panel A: Industry-Level Regressions

Spec	H(Sales)	ln(Size)	ln(B/M)	Momentum	Beta	Leverage	n	R^2[0pt]
1	0.040							
	0.36						227	0.01
2		-0.12						
		-2.64					227	0.01
3			0.14					
			4.22				215	0.01
4				0.031				
				0.18			227	0.02
5					0.02			
					0.07		227	0.02
6						1.57		
						3.94	227	0.02
7		-0.10	0.08	0.20	-0.34	0.49		
		-2.32	2.04	1.47	-1.08	1.27	215	0.07
8	-0.03	-0.11	0.11	0.14				
	-0.22	-2.65	3.53	0.94			215	0.05
9	-0.03	-0.10	0.07	0.18	-0.36	0.50		
	-0.22	-2.27	1.94	1.36	-1.14	1.33	215	0.07

Panel B: Firm-Level Regressions

Spec	H(Sales)	ln(Size)	ln(B/M)	Momentum	Beta	Leverage	n	R^2
1	-0.156							
	-1.25						1384	0.002
2		-0.13	0.09	0.07				
		-4.10	3.70	0.65			1150	0.02
3	-0.216	-0.13	0.09	0.07				
	-1.98	-4.07	3.80	0.65			1150	0.02
4					0.27			
					1.03		1384	0.01
5						1.76		
						5.56	1351	0.01
6		-0.11	0.03	0.09	-0.14	0.78		
		-3.75	1.45	0.93	-0.66	3.03	1150	0.04
7	-0.24	-0.11	0.03	0.09	-0.17	0.81		
	-2.26	-3.69	1.42	0.93	-0.75	3.18	1150	0.04

All variables are as defined in Table I with the exception of Book Equity, which does include post-retirement asset. Note that this difference may also affect values of Dividends-to-Book (D/B), Leverage (Lev), and Book-to-Market (B/M).

Like with the other sample periods, figures 11 and 12 depict the competition premium over time and are constructed analogously to figures 1, 2, 9, and 10. Despite the results from Table 9, the picture that emerges from these plots (both the industry-level and firm-level plots) is that the competition premium hugs zero quite closely. Of course, the implication is that everything taken to together, the supposedly documented competition premium may be just an anomaly.[31] Regarding the premium at the firm-level specifically, figure 12 surely gives cause as to wonder whether there is much practical significance to the evidently statistically significant premium observed. Indeed, given it swings about zero on a regular basis, it is hard to imagine exploiting the result to build a successful investment strategy based on the statistically significant premium.

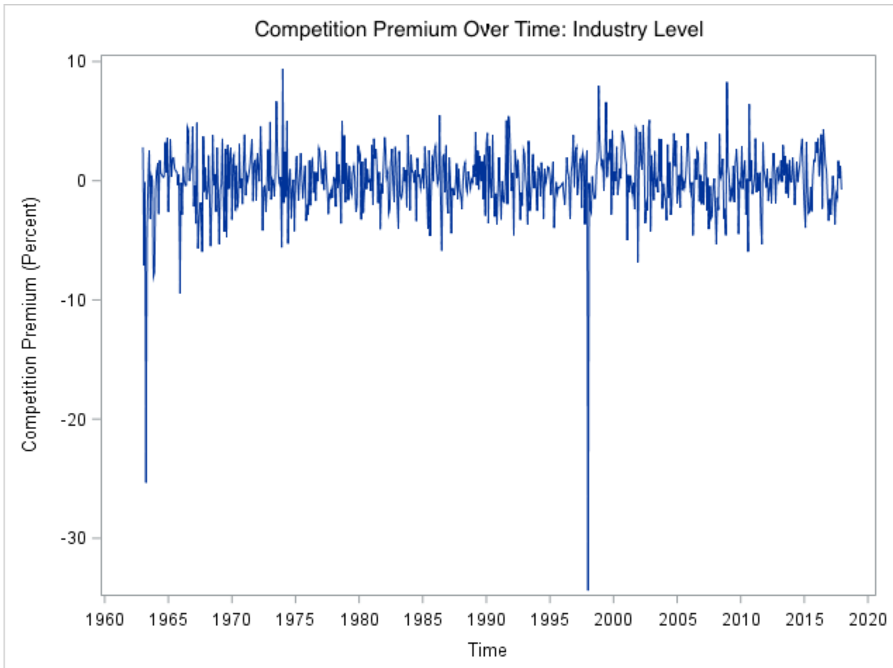

Figure 11: The Competition Premium (Industry Level) over Time: 1963-2018 Period

[31] A close reader may note two outliers are evident (both at the industry level and at the firm level) - specifically, there are two considerably large and negative competition premiums (i.e., positive concentration premiums) during the early 1960s and during the late 1990s. A valid question might concern whether or not removing these outliers from the time-series is enough to “save” the competition premium - particularly at the industry level. The answer is no: removing these outliers still results in a time-series that consistently fluctuates around zero quite closely.

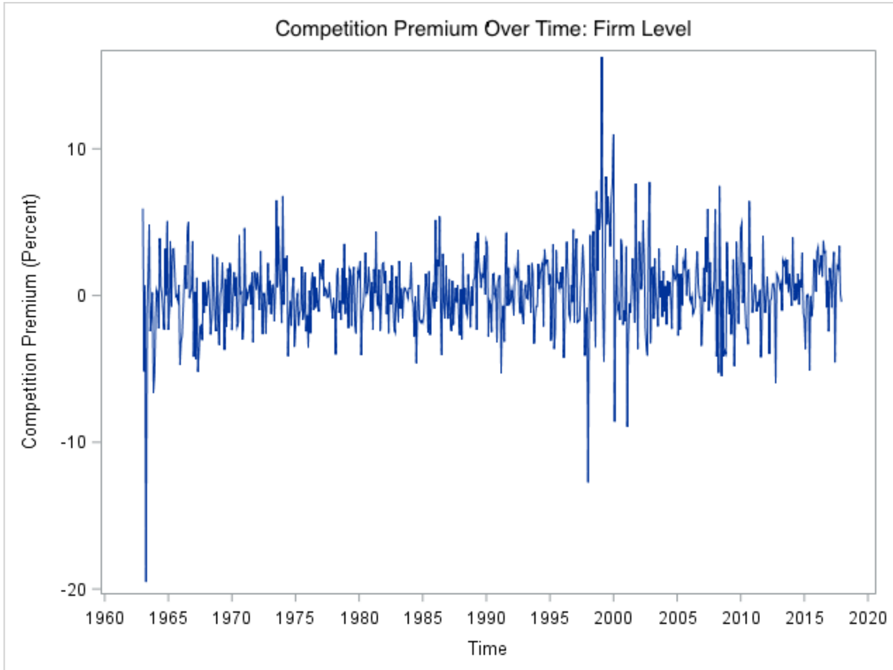

Figure 12: The Competition Premium (Firm Level) over Time: 1963-2018 Period

5.5 The Competition Premium Across Quintiles

One topic that Hou and Robinson do not address concerns the behavior of the competition premium within and across concentration quintiles. This may be an interesting area to explore for various reasons. The following thought excercise illustrates the idea: consider a firm in a particular concentration quintile. Furthermore, suppose that the HHI of the industry this firm operates in is at the lower portion of the quintile. For instance, if the bounds of the quintile are an HHI of 0.2 to an HHI of 0.4, say that the firm's industry has an HHI close to 0.2. Does the competition premium within this quintile appear to differ significantly from the competition premium evident in other quintiles? If so, do competition premiums decrease as one moves up quintiles (i.e., as one increases concentration)? Or do they increase?

This study posits that the latter is more likely - i.e., that competition premiums decreasing as one moves up quintiles is the more likely result. The idea is simple. Consider a firm operating in an industry with the minimum HHI. Such a firm not only operates in the most competitive quintile, but it operates in one of the most competitive industries overall. Assuming that the existence of a competition premium is due to competitive firms exhibiting more risk, it is reasonable to believe that competition risk would decrease dramatically as a firm's industry moves up within its own quintile.

This theory is tested at both the industry-level and at the firm-level. To do so, this study estimates model of form:

$$R_{j,q,t} = \alpha_t + \gamma_{j,t} \text{H(Sales)}_{j,t} + \sum_{n=1}^{N} \lambda_{n,t} X_{j,q,t} + \epsilon_{j,q,t} \tag{8}$$

$$R_{i,q,t} = \alpha_t + \gamma_{i,t}\text{H(Sales)}_{i,t} + \sum_{n=1}^{N} \lambda_{n,t} X_{i,q,t} + \epsilon_{i,q,t} \quad (9)$$

Where $R_{j,q,t}$ and $R_{i,q,t}$ are the returns for industry j and firm i in concentration quintile q at time t respectively; $\text{H(Sales)}_{j,t}$ and $\text{H(Sales)}_{i,t}$ are the HHI of industry j and the HHI for the industry of firm i at time time t respectively; and $X_{j,q,t}$ and $X_{i,q,t}$ are industry average characteristics for industry j and firm characteristics for firm i in concentration quintile q at time t. Panel A presents results for these models at the industry level for the competition premium (the negative of $\beta_{j,t}$) for various time periods using the variables specified in rows 17/18 from Tables 7, 8, and 9. Panel B presents analogous results for the firm-level.

The results appear to be yet another strike against the robustness of the competition premium. Indeed, within some quintiles - both at the industry-level *and* at the firm-level (which has been more "well-behaved" in previous tests), the coefficient on H(Sales) is *positive*. In other words, there is a *concentration premium* as opposed to a competition premium! Save for one estimate (the estimate for the least concentrated quintile during the 2002-2018 period at the firm-level), all of the reported coefficients are insignificant at the 5% level. In defense of the existence of a competition premium, the variance of the HHI within quintiles is not particularly high - especially within the highest quintile. To address this fact, the regressions are run a second time. This time around though, regressions are run for the bottom 40% (quintiles 1 and 2), the middle 60% (quintiles 3, 4, and 5), and the top 40% (quintiles 4 and 5). Results for this exercise appear in Table 11.

Table 10: Fama-Macbeth Regressions of Returns Across Quintiles

Panel A: Industry-Level Regressions

	Low	2	3	4	High
1963-2018	-0.56	-0.05	-0.57	2.79	NA
	-0.92	-0.06	-0.67	1.26	NA
1963-2001	-0.34	-0.89	-1.06	3.75	NA
	-0.44	-0.81	-0.96	1.21	NA
2002-2018	-1.11	1.20	0.59	0.45	NA
	-1.18	1.24	0.47	0.40	NA

Panel B: Firm-Level Regressions

	Low	2	3	4	High
1963-2018	-0.81	1.36	-0.76	3.14	NA
	-1.39	1.60	-1.03	1.43	NA
1963-2001	-0.39	0.65	-1.01	4.07	NA
	-0.53	0.65	-1.09	1.32	NA
2002-2018	-1.82	3.12	-0.15	0.89	NA
	-2.77	1.86	-0.13	0.92	NA

Each month, regressions at the industry-level and firm-level (equations (8) and (9) respectively) are run within each quintile. The estimated models include H(Sales), ln(Size), ln(B/M), Momentum, Beta, and Leverage. Only the time-series average coefficient of H(Sales) is reported along with its time-series t-statistic below. NA is reported for the most concentrated quintile due to their being a dearth of observations.

Table 11: Fama-Macbeth Regressions of Returns Across the Bottom 40%, Middle 60%, and Top 40%

Panel A: Industry-Level Regressions

	Bottom 40%	Middle 60%	Top 40%
1963-2018	-0.82	-0.23	3.23
	-3.00	-1.51	1.48
1963-2001	-0.63	-0.14	3.35
	-1.72	-0.73	1.58
2002-2018	-0.75	-0.38	2.27
	-1.34	-1.34	1.21

Panel B: Firm-Level Regressions

	Bottom 40%	Middle 60%	Top 40%
1963-2018	-0.69	-0.14	3.54
	-1.87	-0.99	0.95
1963-2001	-0.70	-0.20	3.64
	-2.05	-1.02	-1.67
2002-2018	-0.62	-0.09	3.04
	-1.14	-0.36	1.38

Each month, regressions at the industry-level and firm-level (equations (8) and (9) respectively) are run for the bottom 40include H(Sales), ln(Size), ln(B/M), Momentum, Beta, and Leverage. Only the time-series average coefficient of H(Sales) is reported along with its time-series *t*-statistic below.

Based on Table 11, the results certainly become more stable in a sense. However, once again, the average coefficient on the HHI is not significant for the most part. Even so, were the results significant, the pattern would fall in line with the hypothesis this study presented: there would exist a higher competition premium within the bottom 40%. Next, there would still be a competition premium within the middle 60%, however the premium would be considerably smaller. Finally, once within the top 40%, the competition premium actually turns negative - i.e., there is a *concentration premium*. Figures 13 and 14 (constructed analogously as Figures 1 and 2) plot the competition premium over time at the industry- and firm-level for the bottom 40%, middle 60% and top 40%. The picture that emerges lines up with the results from the Table 11. Save for a few obvious outliers (likely due to data issues), the premium is flat across quintiles and fluctuates about zero.

Although the results of this section may give further credence to the no-competition-premium side of things, the issue may not be one of robustness. Rather, the issue may be one of misspecification. The results for Table 11 suggest that the competition premium may not be monotonic. The following section examines whether there is evidence for this being the case.

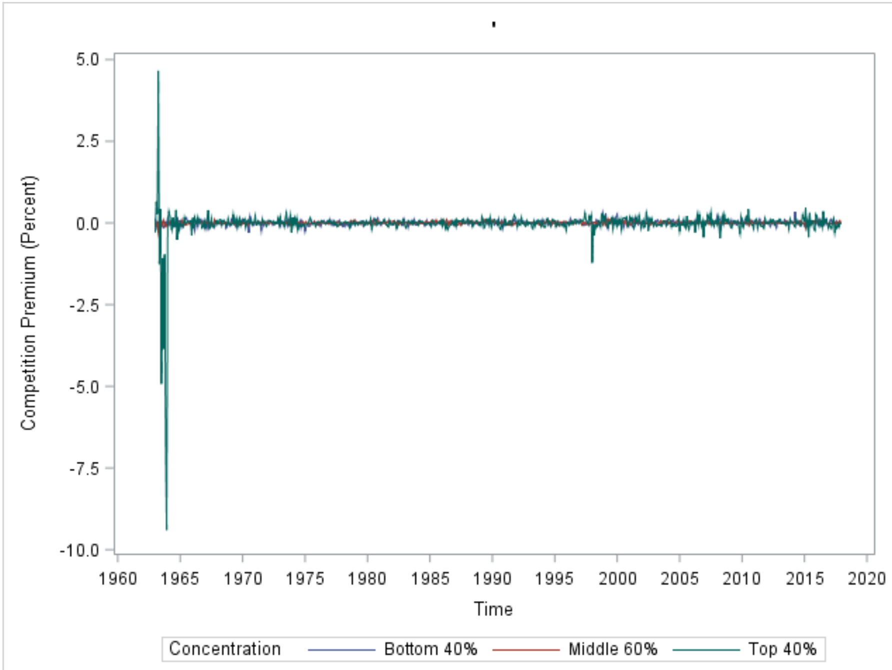

Figure 13: The Competition Premium (Industry Level) Across Quintiles

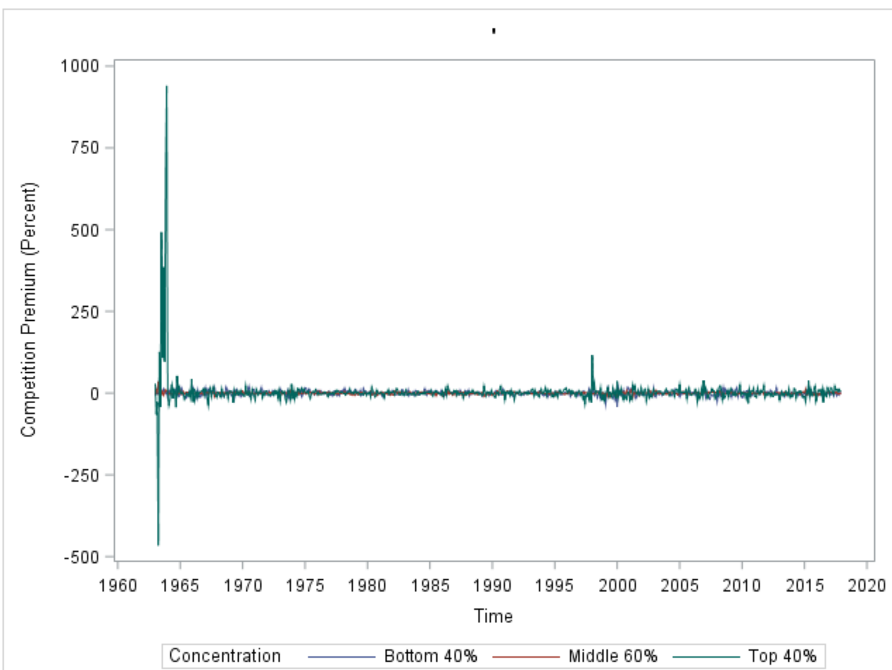

Figure 14: The Competition Premium (Firm Level) Across Quintiles

5.6 Evidence for a Non-Monotonic Competition Premium

There exists good reason to believe that the competition premium may not be monotonic. To see why, first note that *if* a competition premium truly does exist, then it should exist because of *competition risk* that is not being picked up by other controls. Hence, a non-monotonic premium would seem to suggest that competition risk is not monotonic. Although such a claim is difficult to test for, one can easily imagine why competition risk would exhibit non-linear and non-monotonic behavior. Recall that regulated industries are not under consideration in this study. Thus, the highly concentrated industries left under consideration are those highly concentrated industries that likely do not have the "blessing" of government. This fact underlies this study's theory for why competition risk may not be monotonic: once an industry becomes too concentrated it (along with its firms) may find itself at risk of facing scrutiny from government agencies such as the Federal Trade Commission. Thus, while the distress risk suggested by Hou and Robinson surely decreases with concentration, at some point industries and their firms begin to face an increasing amount of *antitrust risk*: the risk associated with becoming concentrated to the point that it catches the eyes of government agencies charged with keeping monopolies and trusts in check.

Although difficult to test the monotonicity of competition risk itself, assuming that the behavior of the competition premium reflects the behavior of competition risk, testing the monotonicity of the premium itself should be enough to gain insight. To do this, this study estimates the following equations for the various sample periods.

$$R_{j,t} = \alpha_t + \gamma_{1,j,t}\text{H(Sales)}_{j,t} + \gamma_{2,j,t}\text{H(Sales)}^2_{j,t} + \sum_{n=1}^{N} \lambda_{n,t} X_{j,t} + \epsilon_t \tag{10}$$

$$R_{i,t} = \alpha_t + \gamma_{1,i,t}\text{H(Sales)}_{i,t} + \gamma_{2,i,t}\text{H(Sales)}^2_{i,t} + \sum_{n=1}^{N} \lambda_{n,t} X_{i,t} + \epsilon_t \tag{11}$$

where $R_{j,t}$ and $R_{i,t}$ are returns for industry j and firm i respectively at time t; $\text{H(Sales)}_{j,t}$ and $\text{H(Sales)}_{i,t}$ are the HHI of industry j and the HHI of the industry firm i operates in respectively at time t; and $X_{j,t}$ and $X_{i,t}$ are a vector of industry average characteristics for industry j and a vector of firm characteristics for firm i at time t.

Table 12 reports the results coefficients on the linear H(Sales) and on the quadratic H(Sales)^2 terms from the equations using the controls from specifications 9 and 7 from Table 7. That is, these average coefficients were produced from regressions where
$X_{j,t} \equiv X_{i,t} \equiv \{\ln(\text{Size}), \ln(\text{B/M}), \text{Momentum}, \text{Beta}, \text{Leverage}\}$.

The results appear to confirm the suspicion that the reason mixed results have been obtained (both in this paper and across the literature) pertains to specification. Save for the 1963-2001 period - where the results are marginally significant - the coefficients on the linear and quadratic H(Sales) terms obtain t-statistics well above the traditional 2.00 cutoff. This holds at *both* the industry- and firm-level!

These results may be exploited to estimate the *critical concentration point*: the point at which both a decrease and an increase in concentration will result in an increase in stock returns (presumably because risk related to market structure will increase for movements in both directions at this point as well). Holding other variables constant, the results for the entire 1963-2018 sample period allow for estimating stock returns at the industry-level and firm-level with the following equations:

$$\hat{R}_j = C - 1.09(\text{HHI}) + 0.99(\text{HHI})^2 \tag{12}$$

$$\hat{R}_i = C - 1.05(\text{HHI}) + 0.91(\text{HHI})^2 \tag{13}$$

where $\hat{R}_j$ and $\hat{R}_i$ denote estimated values for industry-level and firm-level returns respectively and C denotes a constant term consisting of the intercept and the impact of variables held constant at some fixed point. Taking derivatives yields:

$$\frac{\mathrm{d}\hat{R}_j}{\mathrm{dHHI}} = -1.09 + 1.98(\text{HHI}) \tag{14}$$

$$\frac{\mathrm{d}\hat{R}_i}{\mathrm{dHHI}} = -1.05 + 1.82(\text{HHI}) \tag{15}$$

From these equations it is possible to obtain estimates for the critical value of HHIs at the industry- and firm-level. Elementary calculus techniques yield $\text{HHI}^*_{\text{Industry}} \approx 0.55$ and $\text{HHI}^*_{\text{Firm}} \approx 0.58$. Hence, the critical point of concentration appears to occur around the median point of concentration. This finding makes intuitive sense. At and around the median point of concentration, industries and their firms are likely at a sweet spot when it comes to risk related to market structure. On one hand, firms command enough market power to where their distress risk is likely minimal. However, they are not *too* concentrated. Being at the median likely means that there are many firms with similar levels of concentration. As a result, such firms are likely safe from government regulation. If government really wanted to it, it could “make an example” out of a firm or two at a concentration level near the median level. However, firms that operate at levels well above the median are likelier easier to prosecute and make against. In other words, it would not be surprising to find that the government prefers to target “low-hanging fruit” as opposed to taking on the onslaught of firms and industries operating at median levels of concentration.

An alternative, yet related, story could pertain to the fact that in the most concentrated industries, industry-risk is firm-risk. This theory is best illustrated with monopoly: in a monopoly, the firm is the industry and the industry is the firm. Hence, not only may a monopoly find itself exposed to firm-level risk that may be experienced regardless of concentration, but it also bears the burden of the entirety of any inherent industry-level risk. Contrast this with more competitive industries. Firms in competitive enviornments still may face firm-level risk. However, industry-level risk may be dispersed throughout the entire market. In other words, individual firms do not have to carry the weight of the entire market’s inherent risks.

Regardless of the exact reason for the evident quadratic behavior of market structure risk, this finding puts previous literature in a better perspective. As discussed in the literature review, some have found Hou and Robinson’s to hold up, others have found opposite results, and some have found insignificant results. If market structrure risk truely is quadratic (or at the least, non-monotonic), then in one sense *all* of this previous work has been correct. The reason for the difference in findings may, at least partially, be attributed to this previous work attempting to force a non-linear relationship to behave in a linear fashion. Thus, for those that have been able to find similar results to Hou and Robinson, it is possible that the respective samples used just happened to exist on the "left side of the parabola." In turn, studies that have found evidence for concentration premiums may have achieved a sample that sits on the "right side of the parabola." Finally, for the studies that have found insignificant results, their samples may have sit in the "middle of the parabola." This is, of course speculation. However, at the end of the day, there certainly appears to be good evidence for the issue being not one of robustness, but rather one of specification.

Table 12: Fama-Macbeth Regressions of Returns with a Quadratic H(Sales) Term

Panel A: Industry-Level Regressions

	H(Sales)	H(Sales)^2	n	R^2
1963-2018	-1.09	0.99		
	-2.47	2.14	203	0.09
1963-2001	-0.87	0.82		
	-1.55	1.40	198	0.10
2002-2018	-1.60	1.32		
	-2.52	2.23	215	0.08

Panel B: Firm-Level Regressions

	H(Sales)	H(Sales)^2	n	R^2
1963-2018	-1.05	0.91		
	-2.84	2.48	1122	0.05
1963-2001	-0.85	0.76		
	-1.85	1.63	1110	0.05
2002-2018	-1.52	1.25		
	-2.49	2.28	1150	0.07

6 Conclusion

Does common equity cost less for firms operating in more concentrated environments? If one takes Hou and Robinson at their word, the answer would appear to yes. However, as documented in previous literature, this paper finds that the answer is not necessarily clear cut. Indeed, some of the results presented throughout this paper suggest that the existence of a competition premium may be sensitive to time-period and universe of firms examined. Additionally, the distinction between industry-level and firm-level appears to be important. With that said, the actual issue may be one of attempting to force a non-linear relationship to be linear.

This paper will likely not be the end of the story for the competition premium. Indeed, the mixed results documented here underscore the importance of future empirical research that tests the robustness and significance of the competition premium.

As discussed in section 5.6, the issue may not actually be an issue of robustness. Rather, the issue appears to pertain specification. Section 5.6 provides strong evidence for a non-linear relationship between stock returns and industry concentration. This finding suggests that there should be a non-linear relationship between market structure risk and industry concentration. This study barely scratches the surface of this partiular finding. Thus, there are many avenues for future research. First, further testing of the robustness of the non-linear relationship will be important for assessing whether this study's finding was a fluke or not. Second, assuming at the moment that the relationship truely is non-linear, there is a need for modern work that finds the theoretical links between stock returns, market structure risk, and industry concentration. Specifically, there is a need to deconstruct the factors driving the quadratic relationship.

Finally, as this this study is a replication paper at its core, the author offers the following comments on replicability in general. In retrospect, Hou and Robinson's paper is not necessarily difficult to replicate. However, there are certainly portions that could have been more clear. Specifically, as discussed, one of the largest sources of ambiguity within their paper concerned the sequencing of the empirical process. It was rare to encounter ambiguity in *what* they did. In contrast, it was quite common to be puzzled as to *when* they did what they did in relation to other steps. Thus, future research (in general - this need not just apply to economic and financial research) should always strive to be as explicit as humanely possible. In the author's opinion, Sursock (1995) represents the gold standard: his description of the Fama and French (1992) procedure is second to none (and is arguably more clear than Fama and French's own words). Indeed, not only does Appendix C (which describes the post-ranking beta caluclation process) directly borrow from Sursock, but the entireity of Appendix B (which provides a step-by-step process for replicating Hou and Robinson) was inspired by Sursock's paper.

7 References

Abadi, Hassan Heidary Soltan, Shahrokh Bozorgmehrian, and Roholah Javadi. 2015. "The Relationship Between Product Market Competition and Stock Returns in Tehran Stock Exchange Companies." *European Journal of Natural and Social Sciences* 4 (1): 154–64.

Abdoh, Hussein. 2016. "Product Market Competition, Corporate Investments and Risk." PhD thesis, The University of Texas at El Paso.

Aivazian, Varouj A., and Jeffrey L. Callen. 1979. "Investment, Market Structure, and the Cost of Capital." *The Journal of Finance* 34 (1). Wiley: 85–92. https://doi.org/https://doi.org/10.1111/j.1540-6261.1979.tb02072.x.

Alexander, Donald L., and Paul D. Thistle. 1999. "Market Power, Efficiency and the Dispersion of Systematic Risk." *Review of Industrial Organization* 14 (4): 377–90. https://doi.org/10.1023/A:1007721630527.

Ali, Ashiq, Sandy Klasa, and Eric Yeung. 2009. "The Limitations of Industry Concentration Measures Constructed with Compustat Data: Implications for Finance Research." *The Review of Financial Studies* 22 (10): 3839–71. https://doi.org/10.1093/rfs/hhn103.

Almazan, Andres, and Carlos Molino. 2001. "Intra-Industry Capital Structure Dispersion." University of Texas.

Asness, Clifford, and Ross Stevens. 1996. "Intra-Industry and Inter-Industry Factors in the Cross-Section of Expected Stock Returns." Goldman Sachs Asset Management.

Aw, Edward N.W., Christopher Dornick, John Q. Jiang, and Gregory Y. Sivin. n.d. "Industry Concentration and the Cross Section of Expected Stock Returns: A Global Perspective." *The Journal of Investing* 27 (1). Euromoney Institutional Investor: 43–51. https://doi.org/https://doi.org/10.3905/joi.2018.27.1.043.

Babenko, Ilona, Oliver Boguth, and Yuri Tserlukevich. 2018. "Competition, No-Arbitrage, and Systematic Risk." Working Paper. Finance Department, Arizona State University. https://foster.uw.edu/wp-content/uploads/2016/01/Competitive_AP_July2018_Jul15.pdf.

Banzdresch, Santiago. 2011. "Product Differentiation and Systematic Risk: Theory and Empirical Evidence." Working Paper. Bank of Mexico.

Barclay, Michael J., and Clifford W. Smith. 1995. "The Maturity Structure of Corporate Debt." *Journal of Finance* 50 (02): 609–31. https://doi.org/https://doi.org/10.1111/j.1540-6261.1995.tb04797.x.

Bergbrant, Mikael, Delory M. Hunter, and Patrick J. Kelly. 2018. "Product Market Competition, Capital Constraints and Firm Growth." Working Paper. Centre for Economic and Financial Research at New Economic School.

Berger, Elizabeth. 2014. "Survival of the Fittest: An Assessment of the Herfindahl Index and Product Market Competition." Working Paper. Cornell University.

Bernier, Gilles. 1987. "Market Power and Systematic Risk: An Empirical Analysis Using Tobin's Q Ratio." *Journal of Economics and Business* 39 (2): 91–99. https://doi.org/https://doi.org/10.1016/0148-6195(87)90009-9.

Binder, John J. 1992. "Beta, Firm Size, and Concentration." *Economic Inquiry* 30 (3). Wiley-Blackwell: 556–63. https://doi.org/10.1111/j.1465-7295.1992.tb01981.x.

Boone, Jan. 2008. "A New Way to Measure Competition." *The Economic Journal* 118 (531): 1245–61. https://doi.org/10.1111/j.1468-0297.2008.02168.x.

Booth, Laurence D. 1981. "Market Structure Uncertainty and the Cost of Equity Capital." *Journal of Banking and Finance* 5 (4): 467–82. https://doi.org/https://doi.org/10.1016/S0378-4266(81)80002-7.

Bustamante, M. Cecilia, and Andres Donangelo. 2017. "Product Market Competition and Industry Returns." *The Review of Financial Studies* 30 (12). Oxford University Press: 4216–66. https://doi.org/https://doi.org/10.1093/rfs/hhx033.

Chen, Chen, Li Li, and Mary L. Z. Ma. 2014. "Product Market Competition and the Cost of Equity Capital: Evidence from China." *Asia-Pacific Journal of Accounting & Economics* 21 (3): 227–61. https://doi.org/http://dx.doi.org/10.1080/16081625.2014.893197.

Chiu, Yen-Chen. 2010. "Industry Concentration and Cash Flow Risk." *Economics Bulletin* 30 (1): 405–13. https://EconPapers.repec.org/RePEc:ebl:ecbull:eb-09-00341.

Choi, Heeick, and Myungsun Kim. 2018. "Product Market Competition and Credit Ratings." *Quarterly Journal of Finance and Accounting* 56 (3-4): 103–47.

Chou, Pin-Huang, Po-Hsin Ho, and Kuan-Cheng Ko. 2012. "Do Industries Matter in Explaining Stock Returns and Asset-Pricing Anomalies." *Journal of Banking and Finance* 36 (2): 355–70. https://doi.org/https://doi.org/10.1016/j.jbankfin.2011.07.016.

Cohen, Randolph, Christopher Polk, and Tuomo Vuolteenaho. 2003. "The Value Spread." *Journal of Finance* 58 (2): 609–41.

Conine Jr., Thomas E. 1983. "On the Theoretical Relationship Between Systematic Risk and Price Elasticity of Demand." *Journal of Banking & Finance* 10 (2): 173–82. https://doi.org/https://doi.org/10.1111/j.1468-5957.1983.tb00421.x.

Corstjens, Marcel, and Ludo Vanderheyden. 2010. "Competition, Risk and Return in the Us Grocery Industry." *Review of Marketing Science* 8 (2). De Gruyter: 1–28. https://EconPapers.repec.org/RePEc:bpj:revmkt:v:8:y:2010:i:1:p:1-28:n:1.

Datta, Smita, and Anindita Chakraborty. 2018. "Industry Concentration and Stock Returns: Indian Evidence." *JIM Quest* 14 (1): 93–101. http://jaipuria.edu.in/jim/wp-content/uploads/2018/09/JIMQUEST_Second_Issue_2018.pdf.

Diop, Moussa. 2017. "Real Estate Investments, Product Market Competition and Stock Returns." *Real Estate Economics* 46 (2): 291–333. https://doi.org/https://doi.org/10.1111/1540-6229.12201.

Du, Jiang, Shaw K. Chen, and Jeffrey E. Jarrett. 2014. "The Relationship Between the Intensity of Competition in China's Real Estate Industry and Its Return on Equity." *Journal of Asia-Pacific Business* 15 (4). Routledge: 324–34. https://doi.org/10.1080/10599231.2014.965961.

Fama, Eugene F., and Kenneth R. French. 1992. "The Cross-Section of Expected Stock Returns." *Journal of Finance* 47 (02): 427–65. https://doi.org/0.2307/2329112.

———. 2000. "Forecasting Profitability and Earnings." *The Journal of Business* 73 (2): 161–75.

Fama, Eugene F., and James D. MacBeth. 1973. "Risk, Return, and Equilibrium: Empirical Tests." *The Journal of Political Economy* 81 (03): 607–36.

Gallagher, David R., Katja Ignatieva, and James McCulloch. 2015. "Industry Concentration, Excess Returns and Innovation in Australia." *Accounting and Finance* 55 (2): 443–66.

https://doi.org/http://dx.doi.org/10.1111/acfi.12074.

Gaspar, Jose-Miguel, and Massimo Massa. 2006. "Idiosyncratic Volatility and Product Market Competition." *The Journal of Business* 79 (6). The University of Chicago Press: 3125–52. https://doi.org/10.1086/505251.

Giroud, Xavier, and Holger M. Mueller. 2011. "Corporate Governance, Product Market Competition, and Equity Prices." *The Journal of Finance* 66 (2): 563–600. https://doi.org/https://doi.org/10.1111/j.1540-6261.2010.01642.x.

Grullon, Gustavo, Yelena Larkin, and Roni Michaely. 2018. "Are U.s. Industries Becoming More Concentrated?"

Gschwandtner, Adelina, and Michael Hauser. 2016. "Profit Persistence and Stock Returns." *Applied Economics* 48 (37): 3538–49. https://doi.org/https://doi.org/10.1080/00036846.2016.1142652.

Gu, Lifeng. 2016. "Product Market Competition, R&D Investment, and Stock Returns." *Journal of Financial Economics* 119 (2): 441–55. https://doi.org/https://doi.org/10.1016/j.jfineco.2015.09.008.

Harvey, Campbell R., Yan Liu, and Heqing Zhu. 2015.".. and the Cross-Section of Expected Returns." *The Review of Financial Studies* 29 (1): 5–68. https://doi.org/10.1093/rfs/hhv059.

Hashem, Nawar. 2011. "Industry Concentration and Stock Returns: Evidence from Publicly Listed Firms in the U.k." PhD thesis, University of Greenwich.

Hou, Kewei. 2003. "Industry Information Diffusion and the Lead-Lag Effect in Stock Returns." Ohio State University.

Hou, Kewei, and David T. Robinson. 2006. "Industry Concentration and Average Stock Returns." *The Journal of Finance* 61 (4): 1927–56. https://doi.org/https://doi.org/10.1111/j.1540-6261.2006.00893.x.

Hou, Kewei, Chen Xue, and Lu Zhang. 2017. "A Comparison of Factor Models." Working Paper. Fisher College of Business, Ohio State University.

———. 2018. "Replicating Anomalies." *The Review of Financial Studies* hhy131 (December). https://doi.org/https://doi.org/10.1093/rfs/hhy131.

Hrazdil, Karel, and Ray Zhang. 2012. "The Importance of Industry Classification in Estimating Concentration Ratios." *Economics Letters* 114 (2): 224–27. https://doi.org/https://doi.org/10.1016/j.econlet.2011.10.001.

Irvine, Paul J., and Jeffrey Pontiff. 2008. "Idiosyncratic Return Volatility, Cash Flows, and Product Market Competition." *The Review of Financial Studies* 22 (3). Oxford University Press: 1149–77. https://doi.org/https://doi.org/10.1093/rfs/hhn039.

Jaspersen, Stefan. 2019. "Betting on Market Power: Product Market Competition and Mutual Funds' Information Production." Working Paper. University of Cologne - Centre for Financial Research (CFR).

Jory, Surendranath, and Thanh Ngo. 2017. "Firm Power in Product Market and Stock Returns." *The Quarterly Review of Economics and Finance* 65 (August): 182–93. https://doi.org/https://doi.org/10.1016/j.qref.2016.09.008.

Laksmana, Indrarini, and Ya-wen Yang. 2015. "Product Market Competition and Corporate Investment Decisions." *Review of Accounting and Finance* 14 (2): 128–48. https://doi.org/https://doi.org/10.1108/ RAF-11-2013-0123.

Lee, Cheng-Few, Thomas Liaw, and Shafiqur Rahman. 1990. "Impacts of Market Power and Capital-Labor Ratio on Systematic Risk: A Cobbs-Douglass Approach." *Journal of Economics and Business* 42 (3): 237–41. https://doi.org/https://doi.org/10.1016/0148-6195(90)90049-I.

Lewellen, Jonathan. 2015. "The Cross-Section of Expected Stock Returns." *Critical Finance Review* 4 (1): 1–44. https://doi.org/10.1561/104.00000024.

Lin, Wei, and Xu Li-wen. 2011. "Research on Product Market Competition and Risk of Assets." Conference Paper. 2011 International Conference on Management Science & Engineering (18th).

Liu, Chung-Shin. 2011. "Impact of Product Market Competition on Expected Returns." PhD thesis, Department of Finance, University of Oregon. https://scholarsbank.uoregon.edu/xmlui/bitstream/handle/1794/12143/Liu_Chung-Shin_phd2011fa.pdf;sequence=1.

Loualiche, Erik. 2016. "Asset Pricing with Entry and Imperfect Competition." Working Paper. Department of Economics, Northwestern University.

MacKay, Peter, and Gordon Phillips. 2005. "How Does Industry Affect Firm Financial Structure?" *Review of Financial Studies* 18 (04): 1433–66.

Mazali, Rogerio. 2017. "How Are Industry Concentration and Risk Factors Related? Evidence from Brazilian Stock Markets." *New Zealand Economic Papers* 51 (2). Taylor and Francis: 148–76. https://doi.org/https://doi.org/10.1080/00779954.2017.1314317.

McLean, R. D., and J. Pontiff. 2015. "Does Academic Research Destroy Stock Market Predictability?" *Journal of Finance* 71 (1): 5–32. https://doi.org/https://doi.org/10.1111/jofi.12365.

Moghadam, Shahram Amini. 2018. "Two Essays on Competition, Corporate Investment, and Corporate Earnings." PhD thesis, Virginia Polytechnic Institute and State University.

Moskowitz, Tobias J., and Mark Grinblatt. 1999. "Do Industries Explain Momentum?" *Journal of Finance* 54 (4): 1249–90.

Mouselli, Sulaiman, and Aziz Jaafar. 2018. "Industry Concentration, Stock Returns, and Asset Pricing: The Uk Evidence." *Cogent Economics & Finance*, January. https://doi.org/https://doi.org/10.1080/23322039.2019.1576350.

Novy-Marx, Robert. 2009. "Competition, Productivity, Organization and the Cross Section of Expected Returns." Working Paper. University of Chicago and NBER. http://rnm.simon.rochester.edu/research/ROCatXR.pdf.

O'Brien, Thomas J. 2011. "Managerial Economics and Operating Beta." *Managerial and Decision Economics* 32 (3). Wiley: 175–91. https://doi.org/https://doi.org/10.1002/mde.1525.

Pham, Thu A. T. 2018. "Industry Concentration, Firm Efficiency and Average Stock Returns: Evidence from Australia." *Asia-Pacific Financial Markets* 25 (3). Springer Japan: 221–47. https://doi.org/https://doi.org/10.1007/s10690-018-9246-5.

Platt, Katarzyna. 2015. "Product Market Competition and the Cost of Corporate Debt." Working Paper. Department of Management, Marketing and Finance, SUNY College at Old Westbury.

Ryu, Doowon, Doojin Ryu, and Joon Ho Hwang. n.d. "Corporate Governance, Product-Market Competition and Stock Returns: Evidence from the Korean Market." *Asian Business & Management* 16 (1-2). Palgrave Macmillan UK: 50–91. https://doi.org/https://doi.org/10.1057/s41291-017-0014-6.

Scanlon, Dennis P., Michael Chernew, Shailender Swaminathan, and Woolton Lee. 2006. "Competition in Health Insurance Markets: Limitations of Current Measures for Policy Analysis."

Medical Care Research and Review 63 (6). Sage Publications: 37S–55S. https://doi.org/10.1177/1077558706293834.

Schumpeter, Joseph. 1912. *The Theory of Economic Development*. Cambridge, Mass: Harvard University Press.

———. 1942. *Capitalism, Socialism, and Democracy*. New York and London: Harper & Row.

Sharma, Vivek. 2011. "Stock Returns and Product Market Competition: Beyond Industry Concentration." *Review of Quantitative Finance and Accounting* 37 (3): 283–99. https://doi.org/https://doi.org/10.1007/s11156-010-0205-0.

Shu, Jinghon. 2015. "Market Concentration and Cost of Equity in Emerging Market - the Case of China." *International Research Journal of Applied Finance* 6 (2): 61–74. https://login.libsrv.wku.edu/login?url=https://search.proquest.com/docview/1960272856?accountid=15150.

Shumway, Tyler. 1997. "The Delisting Bias in Crsp Data." *Journal of Finance* 52 (1): 327–40. https://doi.org/10.2307/2329566.

Sullivan, Timothy G. 1978. "The Cost of Capital and the Market Power of Firms." *The Review of Economics and Statistics* 60 (2): 209–17. https://doi.org/10.2307/1924974.

———. 1982. "The Cost of Capital and the Market Power of Firms: Reply and Correction." *The Review of Economics and Statistics* 64 (3): 523–25. https://doi.org/10.2307/1925955.

Titman, Sheridan. 1984. "The Effects of Capital Structure and a Firm's Liquidation Decision." *Journal of Financial Economics* 13 (1): 137–51.

Titman, Sheridan, and John D. Martin. 2016. *Valuation: The Art and Science of Corporate Investment Decisions*. 3rd ed. The Pearson Series in Finance. Pearson.

Valta, Philip. 2012. "Competition and the Cost of Debt." *Journal of Financial Economics* 105 (3): 661–82. https://doi.org/https://doi.org/10.1016/j.jfineco.2012.04.004.

Wang, Ching-Chang, and Chilulien C. Venezia. n.d. "Market Competition Structure and Mututal Fund Performance." *International Journal of Science & Informatics* 2 (1): 1–7.

8 Appendix A: Summary Statistics and Fama-Macbeth Regressions of H(Sales) for the 2002-2018 Period and the 1963-2018 Period

Table 13: Summary Statistics (2002-2018)

Panel A: Summary of Industry Concentration Measures

	Mean	Median	SD	Max	Min	20%	40%	60%	80%	H(Sales)	H(Assets)	H(Equity)
H(Sales)	0.578	0.516	0.313	1.000	0.026	0.265	0.413	0.675	0.995	1.000	0.978	0.810
H(Assets)	0.582	0.525	0.310	1.000	0.032	0.032	0.272	0.429	0.996	0.979	1.000	0.820
H(Equity)	0.688	0.721	0.284	1.000	0.069	0.069	0.387	0.595	1.000	0.834	0.844	1.000

Panel B: Characteristics of H(Sales) Sorted Quintile

Rank	H(Sales)	Newlist	Delists	Size	Asset	Sales	E/A	E/S	V/A	D/B	R&D	R&D/A	Lev.	B/M	Beta
Low	0.129	258.88	247.71	3505.4	2537.5	2039.2	0.053	0.132	2.958	4.659	113.08	0.144	0.452	0.316	1.392
2	0.332	52.58	65.53	5281.7	4929.2	4592.2	0.067	0.077	2.288	6.319	97.781	0.051	0.522	0.750	1.339
3	0.517	33.65	30.35	4157.5	4770.5	4454.3	0.069	0.087	2.440	5.497	70.965	0.044	0.511	2.296	1.362
4	0.869	18.41	22.06	4037.5	3278.4	3278.5	0.066	0.100	2.456	4.787	60.244	0.041	0.510	0.416	1.347
High	0.995	11.00	5.00	1244.3	1138.4	1000.6	0.065	0.073	2.866	2.353	8.031	0.025	0.495	0.374	1.485

All variables constructed as defined in Table 1 with the exception of Book Equity, which does include post-retirement asset. Note that this difference may also affect values of Dividends-to-Book (D/B), Leverage (Lev), and Book-to-Market (B/M).

Table 14: Fama-Macbeth Regressions of H(Sales) (2002-2018)

Panel A: Simple Regressions

ln(Size)	ln(Assets)	ln(Sales)	E/A	E/S	V/A	D/B	R&D/A	Leverage	ln(B/M)	Beta
-0.023	-0.032	-0.030	-0.102	-0.074	0.012	-0.0001	-0.560	-0.127	-0.017	-0.071
-12.04	-18.79	-18.74	-1.35	-1.61	4.41	-0.12	-4.77	-6.23	-5.65	-3.61

Panel B: Multiple Regressions

ln(Size)	ln(Assets)	ln(Sales)	E/A	E/S	V/A	D/B	R&D/A	Leverage	ln(B/M)	Beta
-0.071			0.135				-0.829	-0.090	-0.029	-0.167
-3.52			0.64				-5.63	-1.31	-4.84	-7.13
	-0.021		0.145				-0.854	-0.006	-0.028	-0.176
	-4.14		0.68				-5.84	-0.09	-4.56	-7.19
		-0.032	0.348				-0.978	0.070	-0.301	-0.196
		-8.19	1.74				-6.12	1.02	-5.11	-9.08
-0.017				0.093			-0.831	-0.084	-0.031	-0.174
-3.39				0.96			-5.80	-1.77	-5.37	-7.69
-0.024					0.136		-0.890	-0.183		-0.148
-4.44					1.91		-6.30	-2.46		-5.69
-0.017			0.151			-0.003	-0.809	-0.077	-0.033	-0.170
-3.60			0.74			-2.21	-5.59	-1.14	-5.38	-7.22

All variables constructed as defined in Table 1 with the exception of Book Equity, which does include post-retirement asset. Note that this difference may also affect values of Dividends-to-Book (D/B), Leverage (Lev), and Book-to-Market (B/M).

Table 15: Summary Statistics (1963-2018)

Panel A: Summary of Industry Concentration Measures

	Mean	Median	SD	Max	Min	20%	40%	60%	80%	H(Sales)	H(Assets)	H(Equity)
H(Sales)	0.554	0.498	0.306	1.000	0.026	0.251	0.399	0.616	0.964	1.000	0.978	0.793
H(Assets)	0.558	0.505	0.303	1.000	0.032	0.255	0.411	0.620	0.949	0.978	1.000	0.804
H(Equity)	0.669	0.687	0.287	1.000	0.059	0.369	0.569	0.824	0.998	0.817	0.826	1.000

Panel B: Characteristics of H(Sales) Sorted Quintile

Rank	H(Sales)	Newlist	Delists	Size	Asset	Sales	E/A	E/S	V/A	D/B	R&D	R&D/A	Lev.	B/M	Beta
Low	0.142	138.61	146.17	1812.2	1312.6	1173.1	0.062	0.116	2.740	2.864	62.522	0.110	0.499	0.364	1.264
2	0.310	54.88	53.37	1677.3	1547.5	1501.9	0.073	0.075	2.303	2.701	38.316	0.057	0.546	0.491	1.258
3	0485	26.02	25.02	1601.9	1968.9	1817.3	0.075	0.093	2.178	2.538	49.415	0.036	0.558	1.128	1.244
4	0.776	20.32	17.42	1866.2	2367.6	1503.3	0.073	0.087	2.197	2.632	45.142	0.037	0.558	0.469	1.238
High	0.975	8.64	5.38	761.6	1112.7	1042.6	0.104	0.081	2.467	2.828	23.208	0.021	0.568	0.566	1.222

All variables constructed as defined in Table 1 with the exception of Book Equity, which does include post-retirement asset. Note that this difference may also affect values of Dividends-to-Book (D/B), Leverage (Lev), and Book-to-Market (B/M).

Table 16: Fama-Macbeth Regressions of H(Sales) (1963-2018)

Panel A: Simple Regressions

ln(Size)	ln(Assets)	ln(Sales)	E/A	E/S	V/A	D/B	R&D/A	Leverage	ln(B/M)	Beta
-0.036	-0.039	-0.041	0.157	0.052	-0.0004	0.002	-1.204	-0.018	0.004	-0.065
-13.74	-19.46	-18.96	1.68	2.08	-0.10	2.98	-7.09	-0.59	1.21	-2.50

Panel B: Multiple Regressions

ln(Size)	ln(Assets)	ln(Sales)	E/A	E/S	V/A	D/B	R&D/A	Leverage	ln(B/M)	Beta
-0.007			0.557				-0.913	-0.001	0.001	-0.125
-1.54			3.86				-8.95	-0.02	0.14	-5.34
	-0.008		0.550				-0.933	0.032	0.001	-0.128
	-1.97		3.85				-9.04	0.76	0.16	-5.59
		-0.022	0.768				-0.967	0.107	0.001	-0.153
		-6.61	5.43				-9.61	2.54	0.18	-7.18
-0.007				0.265			-0.969	-0.027	-0.003	-0.135
-2.04				3.67			-8.52	-0.55	-0.50	-5.96
-0.006					-0.013		-0.950	-0.064		-0.119
-1.54					-1.01		-9.96	-74		-5.21
-0.007			0.605			0.001	-0.914	0.015	-0.001	-0.116
-1.53			4.47			0.49	-9.18	0.27	-0.17	-4.95

All variables constructed as defined in Table 1 with the exception of Book Equity, which does include post-retirement asset. Note that this difference may also affect values of Dividends-to-Book (D/B), Leverage (Lev), and Book-to-Market (B/M).

9 Appendix B: Replication Information

This appendix presents the step-by-step process[32] necessary to follow Hou and Robinson. Additionally, it provides a time-series plot of the number of firms within this study's sample at various stages of data-wrangling.

9.1 Steps to Follow Hou and Robinson

1. From the COMPUSTAT annual file, extract and/or calculate the following variables according to Table 1: Total Assets, Net Sales, Research and Development Expenditure, Research to Assets Ratio, Earnings, Earnings to Assets Ratio, Earnings to Sales Ratio, Book Equity, COMPUSTAT Market Value of Equity, COMPUSTAT value of firm, Book Equity to Market Equity Ratio, Dividends, Dividents to Book Equity Ratio, Value to Assets Ratio, and Leverage.
2. From CRSP, calculate the following variables according to Table 1: Size and Momentum.
3. If not already using the pre-merged COMPUSTAT-CRSP database from WRDS, merge COMPUSTAT and CRSP. As discussed in the main text, the merge will ideally be done with the linktable that WRDS provides. Otherwise, the second-best solution is to get a link between GVKEY (COMPUSTAT's unique identifier) and PERMNO (CRSP's unique identifier) for each year via CUSIP. Do not just perform a merge via CUSIP itself. Note that COMPUSTAT data for fiscal year $t-1$ gets merged with CRSP data for July of year t to June of year $t+1$.
4. Calculate HHI for each industry based on CRSP's SIC code according to equation (3): $\text{H(Var)}_{j,t=0} = \frac{1}{3}\sum_{t=-2}^{0}\left[\sum_{i=1}^{N_j}\left(\frac{\text{Var}_{i,j,t}}{\sum_{i=1}^{N_j}\text{Var}_{i,j,t}}\right)^2\right]$. Note that CRSP only provides 4-digit SIC codes. The corresponding 3-digit and 2-digit codes may be obtained if desired: The 3-digit code will be the first three digits from the 4-digit code (for example, if the 4-digit code is 1234, the 3-digit code is 123). Likewise, the 2-digit code will be the first two digits from the 4-digit code (for example, if the 4-digit code is 1234, the 2-digit code is 12). Stocks that are missing CRSP's 4-digit SIC code get filtered out.
5. Calculate post-ranking beta according to Fama-French (1992). See Appendix C for a step-by-step process (partially adapted from Sursock (1996)) on how to go about this process. Note that post-ranking betas are time-period dependent. I.e., calculating post-ranking betas for 1963-2001, for 2001-2018, and for 1963-2018 will differ (at least slightly). This is because post-ranking betas come from a regression that uses the *entire* sample - meaning, post-ranking betas ARE NOT calculated on a rolling basis. Stocks that do not meet the data-requirements to be assigned a post-ranking beta get filtered out.
6. For the entire sample (i.e., at the "aggegate" level), calculate summary statistics for the HHI - mean, median, standard deviation, max, min, quintile breakpoints, and correlations with other metrics used to calculate HHI. This constitutes the replication of Hou and Robinson Table I Panel A.
7. Sort stocks into concentration quintiles based on HHI and calculate within each quintile the following: mean HHI, mean number of newly listed firms each year, mean number delisted firms each year, mean size, mean assets, mean E/A, mean E/S, mean V/A, mean D/B, mean R&D spending, mean R&D/A, mean leverage, mean B/M, and mean post-ranking beta. This constitutes the replication of Hou and Robinson Table I: Panel B.
8. Calculate the mean of ln(Size), ln(Assets), ln(Sales), E/A, E/S, V/A, D/B, R&D/A, Leverage, ln(B/M), and Beta for each industry for each year.
9. For each industry j (based on SIC code) for each year t, run regressions according to equation (4) for various covariates: $\text{H(Sales)}_{j,t} = \alpha_t + \sum_{n=1}^{N}\lambda_{n,t}X_{j,t} + \epsilon_{j,t}$. For λ_i for i in $1, 2, \ldots, n$ (the coefficient on variable x_i from $X_{j,t}$), a time-series of λ_is is produced. Take the average of this time series to obtain the reported λ_i. Calculate the t-statistic by computing the time-series

[32]Partially inspired by Sursock (1995) - who presents an excellent step-by-step process to replicate Fama-French (1996).

standard error and calculating the t-statistic as usual.

10. Adjust CRSP monthly stock returns for delisting. WRDS offers a research macro that includes a method for this delisting adjustment.[33]
11. Run Fama-Macbeth (1973) style regressions according to equations (5): $R_{j,t} = \alpha_t + \sum_{n=1}^{N} \lambda_{n,t} X_{j,t} + \epsilon_{j,t}$; and (6): $R_{i,t} = \alpha_t + \sum_{n=1}^{N} \lambda_{n,t} X_{i,t} + \epsilon_{i,t}$. Note that (5) is for industry-level regressions and (6) is for firm-level regressions. For each λ_i where i in $1, 2, \ldots, n$, a time-series of λ_is is produced. Take the average of this time series to obtain the reported λ_i. Calculate the t-statistic by computing the time-series standard error and calculating the t-statistic as usual.

The following steps are specifically for reproducing/extending some of this paper's new extensions.

12. To reproduce a time-series plot of the competition premium, first from one of the specifications of (5) or of (6) that includes H(Sales), multiply the time series of λ_{HHI}s on H(Sales) by negative one. This transformation is made since it is expected that λ_{HHI} will usually be neagtive: meaning, as concentration (HHI) increases, stock returns decrease. Thus, this trnasformation allows for the "premium" interpretation: an increase in competition will result in an increase in stock returns.
13. Plot the transformed time-series $-1 \cdot \lambda_{HHI}$.
14. To reproduce Table 10, first group stocks into yearly quintiles based on HHI.
15. Run regressions according to (8): $R_{j,q,t} = \alpha_t + \beta_{j,t} \text{H(Sales)}_{j,t} + \sum_{n=1}^{N} \lambda_{n,t} X_{j,q,t} + \epsilon_{j,q,t}$; and (9): $R_{i,q,t} = \alpha_t + \beta_{i,t} \text{H(Sales)}_{i,t} + \sum_{n=1}^{N} \lambda_{n,t} X_{i,q,t} + \epsilon_{i,q,t}$. Note that (8) corresponds to industry-level and (9) corresponds to firm-level.
    For each λ_i for i in $1, 2, \ldots, n$ a time-series of λ_is is produced. Take the average of this time series to obtain the reported λ_i. Calculate the t-statistic by calculating the time-series standard error and calculating the t-statistic as usual.
16. To reproudce Table 11, first group stocks into bottom 40%, middle 60%, and top 40% each year based on HHI. Note that the bottom 40% corresponds to quintiles 1 and 2. The middle 60% corresponds to quintiles 2, 3, and 4. Finally, the top 40% corresponds to quintiles 4 and 5.
17. Run regressions akin to (8) and (9) again. However, this time around, instead of quintiles, stocks are grouped into bottom 40%, middle 60%, and top 40%. Reported coefficients and t-statistics are calculated similarly as described above.
18. Run regressions akin to equations (5) and (6) again. However, in addition to an HHI term, also include a quadratic HHI term to test for non-linearity.

Admittedly, it can be difficult to succinctly describe research methodology: once one is an expert in what they have done, it can be far too easy to realize that what might be obvious to the expert, may not actually be obvious. Hence, for researchers looking to reproduce any aspect of this paper who finds something here ambigious or unclear, the author will be glad to answer any questions via email at `jacob.applin@outlook.com`.

[33] Vora, Premal and Luis Palacios. "Fama-French Factors". WRDS Research Macros. 2010. Wharton Research Data Services, the Wharton School, University of Pennsylvania.

9.2 Number of Firms Per Year

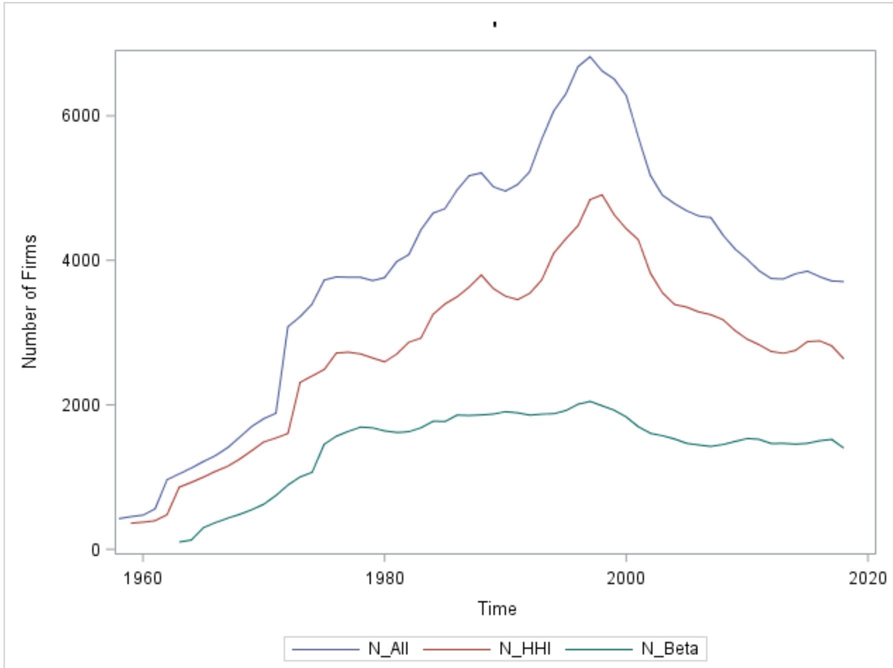

Figure 15: Number of Unique Firms per Year At Various Stages of Data Wrangling

In the above figure, **N__All** denotes the number of unique firms at the intersection of COMPUSTAT and CRSP with share-codes of 10 or 11 and traded on the NYSE, NASDAQ, and/or AMEX. Hence, **N__All** is calculated right after the merging of the two datasets take place with no further data-filtering. **N__HHI** denotes the number of unique firms that had a 4-digit SIC code listed in CRSP - meaning that this firm was able to contributue to the calculation of HHIs. Finally, **N__Beta** denotes the number of unique firms that *both* were able to contribute HHI calculations *and* met the data-requirements to be assigned a post-ranking beta per Fama and French (1992).

10 Appendix C: Fama-French Replications

10.1 The Fama-French (1992) Procedure

The following gives the step-by-step process for following Fama and French (1992) to calculate post-ranking betas. It is taken directly from Sursock (1995) - who gives what this author believes to be the absolute best explanation of the process.

" 1. Merge the CRSP Monthly Stock File with the COMPUSTAT Industrial Annual File. The COMPUSTAT data is merged to the CRSP data with a 6-month gap. More precisely, COMPUSTAT accounting data from year $t-1$ is merged to CRSP return data from July of year t to June of year $t+1$. For example, returns for CRSP data between July 1991 and June 1992 get merged with COMPUSTAT data from 1990.

2. Delete COMPUSTAT data before 1962.
3. Delete all financial services firms. Financial services firms have SIC codes between 6000 and 6999 inclusive.
4. The first pre-ranking estimation period is defined for the 60 months from July 1958 to June 1963.
5. To qualify for the pre-ranking estimation period, securities must satisfy *all* of the following requirements:
   (a) a CRSP stock *price* in December 1962 *and* June 1963
   (b) monthly CRSP *returns* for at least 24 of the 60 months
   (c) COMPUSTAT data on total assets, book equity, and earnings as defined in Table 3.1 (*Table 1 in this paper*) for its fiscal year ending in any month of 1962
6. For each stock in the pre-ranking period run the following regression

$$R_{it} = \alpha_i + \beta_{i1} R_{mt} + \beta_{i2} R_{mt-1} + \epsilon_{it} \tag{16}$$

where R_{it} is the return (including dividends) on stock i in month t, R_{mt} and R_{mt-1} are the CRSP value-weighted returns in months t and $t-1$ on the NYSE, AMEX, and NASDAQ market portfolio. "

7. Calculate the pre-ranking CAPM β of stock i as the sum of the regression coefficient estimates $\hat{\beta}_i = \widehat{\beta_{i1}} + \widehat{\beta_{i2}}$.
8. Calculate portfolio decile breakpoints for size from all stocks on the NYSE based on their market equity in June of 1963.
9. Within each of the size deciles calculate portfolio decile breakpoints from all stocks that meet the CRSP/COMPUSTAT requirements for their pre-ranking beta, $\hat{\beta}_i$.
10. For each of the 100 portfolios created above (10 beta-sorted portfolios for each of the 10 size-sorted portfolios) calculate the equally-weighted return for the 12 months starting July 1963 and ending June 1964. Use the ranks assigned to each security in June 1963.
11. Shift the window of observation for the data one year and repeat steps (4)-(10). In other words, the next period used for the pre-ranking betas is July 1959 to June 1964 and the corresponding portfolio returns are calculated for July 1964 to June 1964. Continue this procedure until 1990 (or the last year of data). Since the last year has data until December of 1990 instead of June 1991 portfolio returns can only be computed for 6 months.

12. For each portfolio p run the following regression across all 330 months of data (June 1963 to December 1990)

$$R_{pt} = \alpha_p + \beta_{p1} R_{mt} + \beta_{p2} R_{mt-1} + \epsilon_{pt} \tag{17}$$

where R_{pt} is the monthly return of portfolio p and R_{mt} and R_{mt-1} are defined as before in equation (3.1) (*equation (16) in this paper*)

13. Calculate the post-ranking CAPM β of portfolio p as the sum of the regression coefficient estimates $\hat{\beta_p} = \hat{\beta_{p1}} + \hat{\beta_{p2}}$.
14. Assign the post-ranking beta of each portfolio to each of the stocks within the portfolio.
15. For each of the accounting variables (except market-equity) set the values of the top and bottom 0.5 percent equal to the top and bottom 0.5 percentile respectively.
16. Run the Fama-Macbeth regressions for the linear asset pricing model using the full 330 months of return data. In other words, run the following regression across all stocks i for each for each month t:

$$R_{it} = \gamma_{0t} + \gamma_{1t}\hat{\beta_{it}} + \gamma_{2t} \ln{(ME)}_{it} + \gamma_{3t} \ln{(BE/ME)}_{it} + \cdots + \eta_{it} \tag{18}$$

where R_{it} is the return on stock i and $\hat{\beta_{it}}$ is the post-ranking beta for stock i at month t. Other explanatory variables may be chosen to test other pricing models.

17. Regression equation (3.3) (*equation (18) in this paper*) produces a time series of regression $\hat{\gamma_{0t}}$, $\hat{\gamma_{1t}}, \ldots$, which cover 330 months. Compute the mean, variance, and t-statistic of each gamma as explained in the description of the Fama-MacBeth procedure (*See Sursock's paper for this description*)."

$$\hat{\gamma_1} = \frac{1}{T} \sum_{i=1}^{T} \hat{\gamma_{1t}} \tag{19}$$

$$\sigma^2_{\hat{\gamma_{1t}}} = \frac{1}{T-1} \sum_{i=1}^{T} (\hat{\gamma_{1t}} - \hat{\gamma_1}) \tag{20}$$

$$t_{\hat{\gamma_1}} = \frac{\hat{\gamma_1}}{\sigma^2_{\hat{\gamma_{1t}}} / \sqrt{T}} \tag{21}$$

— p. 29-31

10.2 Fama-French Replications

	All	Low-β	β-2	β-3	β-4	β-5	β-6	β-7	β-8	β-9	High-β
					Panel A: Average Monthly Returns (in Percent)						
All	1.25	1.34	1.29	1.36	1.31	1.33	1.28	1.24	1.21	1.25	1.14
Small-ME	1.52	1.71	1.57	1.79	1.61	1.50	1.50	1.37	1.63	1.50	1.42
ME-2	1.29	1.25	1.42	1.36	1.39	1.65	1.61	1.37	1.31	1.34	1.11
ME-3	1.24	1.12	1.31	1.17	1.70	1.29	1.10	1.31	1.36	1.26	0.76
ME-4	1.25	1.27	1.13	1.54	1.06	1.34	1.06	1.41	1.17	1.35	0.98
ME-5	1.29	1.34	1.42	1.39	1.48	1.42	1.18	1.13	1.27	1.18	1.08
ME-6	1.17	1.08	1.53	1.27	1.15	1.20	1.21	1.18	1.04	1.07	1.02
ME-7	1.07	0.95	1.21	1.26	1.09	1.18	1.11	1.24	0.62	1.32	0.76
ME-8	1.10	1.09	1.05	1.37	1.20	1.27	0.98	1.18	1.02	1.01	0.94
ME-9	0.95	0.98	0.88	1.02	1.14	1.07	1.23	0.94	0.82	0.88	0.59
Large-ME	0.89	1.01	0.93	1.10	0.94	0.93	0.89	1.03	0.71	0.74	0.56

Table I—*Continued*

	All	Low-β	β-2	β-3	β-4	β-5	β-6	β-7	β-8	β-9	High-β
					Panel B: Post-Ranking βs						
All		0.87	0.99	1.09	1.16	1.26	1.29	1.35	1.45	1.52	1.72
Small-ME	1.44	1.05	1.18	1.28	1.32	1.40	1.40	1.49	1.61	1.64	1.79
ME-2	1.39	0.91	1.15	1.17	1.24	1.36	1.41	1.43	1.50	1.66	1.76
ME-3	1.35	0.97	1.13	1.13	1.21	1.26	1.28	1.39	1.50	1.51	1.75
ME-4	1.34	0.78	1.03	1.17	1.16	1.29	1.37	1.46	1.51	1.64	1.71
ME-5	1.25	0.66	0.85	1.12	1.15	1.16	1.26	1.30	1.43	1.59	1.68
ME-6	1.23	0.61	0.78	1.05	1.16	1.22	1.28	1.36	1.46	1.49	1.70
ME-7	1.17	0.57	0.92	1.01	1.11	1.14	1.26	1.24	1.39	1.34	1.60
ME-8	1.09	0.53	0.74	0.94	1.02	1.13	1.12	1.18	1.26	1.35	1.52
ME-9	1.03	0.58	0.74	0.80	0.95	1.06	1.15	1.14	1.21	1.22	1.42
Large-ME	0.92	0.57	0.71	0.78	0.89	0.95	0.92	1.02	1.01	1.11	1.32
					Panel C: Average Size (ln(ME))						
All	4.11	3.86	4.26	4.33	4.41	4.27	4.32	4.26	4.19	4.03	3.77
Small-ME	2.24	2.12	2.27	2.30	2.30	2.28	2.29	2.30	2.32	2.25	2.15
ME-2	3.63	3.65	3.68	3.70	3.72	3.69	3.70	3.69	3.69	3.70	3.68
ME-3	4.10	4.14	4.18	4.12	4.15	4.16	4.16	4.18	4.14	4.15	4.15
ME-4	4.50	4.53	4.53	4.57	4.54	4.56	4.55	4.52	4.58	4.52	4.56
ME-5	4.89	4.91	4.91	4.93	4.95	4.93	4.92	4.93	4.92	4.92	4.95
ME-6	5.30	5.30	5.33	5.34	5.34	5.33	5.33	5.33	5.33	5.34	5.36
ME-7	5.73	5.73	5.75	5.77	5.76	5.73	5.77	5.77	5.76	5.72	5.76
ME-8	6.24	6.26	6.27	6.26	6.24	6.24	6.27	6.24	6.24	6.24	6.26
ME-9	6.82	6.82	6.84	6.82	6.82	6.81	6.81	6.81	6.81	6.80	6.83
Large-ME	7.93	7.94	8.04	8.10	8.04	8.02	8.02	7.94	7.80	7.75	7.62

Figure 16: Fama and French (1992) Table 1 As Originally Published

Per Fama and French (1992): "Portfolios are formed yearly. The breakpoints for the size (ME, price times shares outstanding) deciles are determined in June of year t ($t = 1963 - 1990$) using all NYSE stocks on CRSP. All NYSE, AMEX, and NASDASQ stocks that meet the CRSP-COMPUSTAT data requirements are allocated to the 10 size portfolios using the NYSE breakpoints. Each size decile is subdivided into 10 β portfolios using pre-ranking *beta* of individual stocks, estimated with 2 to 5 years of monthly returns (as available) ending in June of year t. [Only] NYSE stocks that meet the CRSP-COMPUSTAT data requirements to establish the β breakpoints. The equal-weighted monthly returns on the resulting 100 portfolios are then calculated for July of year t to June of year $t+1$. The post-ranking βs use the full (July 1963) to December 1990) sample of post-ranking returns for each portfolio. The pre- and post-ranking $\beta\ldots$ are the sum of the slopes from a regression of monthly returns on the current and prior month's returns on the value-weighted portfolio returns, in percent. The average size of a portfolio is the time-series average of monthly averages of ln(ME) for stocks in the portfolio at the end of June of each year, with ME denominated in millions of dollars."

Table 17: Fama-French (1992) Table 1 Replication

.	All	Low B	B-2	B-3	B-4	B-5	B-6	B-7	B-8	B-9	High B
Panel A: Average Monthly Returns (in Percent)											
All	1.38	1.38	1.38	1.33	1.37	1.36	1.46	1.16	1.00	1.00	0.79
Small-ME	1.59	1.45	1.49	1.75	1.31	1.71	1.88	1.29	1.93	1.34	1.77
ME-2	1.33	1.33	1.45	1.50	1.43	1.43	1.62	1.44	0.93	1.17	1.00
ME-3	1.29	1.13	1.10	1.17	1.84	1.41	1.71	1.63	1.20	1.13	0.54
ME-4	1.28	1.36	1.54	1.27	1.48	1.57	1.67	1.00	1.17	0.88	0.82
ME-5	1.25	1.46	1.64	1.50	1.43	1.38	1.41	1.44	0.56	1.02	0.67
ME-6	1.31	1.65	1.31	1.53	1.29	1.36	1.28	1.34	1.16	1.11	1.04
ME-7	1.18	1.38	1.31	1.16	1.43	1.55	1.23	0.92	0.79	1.35	0.67
ME-8	1.08	1.31	1.16	1.29	1.29	1.22	1.48	0.93	0.70	1.04	0.43
ME-9	1.05	1.32	1.51	1.03	1.38	1.26	1.34	0.86	0.78	0.63	0.35
Large-ME	0.89	1.44	1.29	1.12	0.86	0.74	0.99	0.76	0.76	0.37	0.60
Panel B: Post-Ranking Betas											
All		0.94	1.04	1.10	1.16	1.21	1.27	1.30	1.34	1.45	1.59
Small-ME	1.35	1.04	1.17	1.25	1.25	1.29	1.36	1.37	1.55	1.52	1.74
ME-2	1.35	1.01	1.17	1.24	1.21	1.32	1.45	1.41	1.39	1.59	1.71
ME-3	1.34	1.06	1.06	1.16	1.29	1.28	1.32	1.52	1.48	1.57	1.68
ME-4	1.35	1.08	1.15	1.18	1.25	1.25	1.31	1.44	1.57	1.54	1.72
ME-5	1.32	1.04	1.15	1.19	1.17	1.27	1.29	1.40	1.44	1.61	1.67
ME-6	1.22	0.88	0.98	1.01	1.19	1.31	1.28	1.30	1.25	1.37	1.65
ME-7	1.23	0.96	1.03	1.16	1.17	1.24	1.26	1.23	1.27	1.38	1.61
ME-8	1.16	0.87	0.97	1.01	1.03	1.19	1.24	1.23	1.21	1.39	1.47
ME-9	1.09	0.77	0.90	0.97	1.06	1.00	1.24	1.19	1.15	1.31	1.28
Large-ME	0.98	0.64	0.82	0.85	1.00	0.97	0.91	0.96	1.10	1.21	1.33
Panel C: Average Size (ln(ME))											
All	10.76	12.16	12.24	12.23	12.23	12.22	12.21	12.21	12.19	12.21	12.21
Small-ME	9.22	8.99	9.23	9.27	9.25	9.29	9.26	9.31	9.25	9.25	9.15
ME-2	10.72	10.63	10.72	10.74	10.77	10.65	10.68	10.76	10.77	10.74	10.78
ME-3	11.16	11.07	11.20	11.12	11.23	11.19	11.16	11.05	11.15	11.24	11.20
ME-4	11.58	11.49	11.58	11.63	11.59	11.58	11.58	11.60	11.55	11.54	11.67
ME-5	11.97	11.89	11.97	11.89	12.06	12.04	11.95	12.00	11.93	12.00	12.01
ME-6	12.36	12.24	12.34	12.35	12.29	12.31	12.41	12.39	12.42	12.43	12.42
ME-7	12.78	12.70	12.82	12.85	12.82	12.76	12.77	12.75	12.73	12.77	12.82
ME-8	13.29	13.29	13.27	13.27	13.30	13.30	13.29	13.29	13.24	13.30	13.37
ME-9	13.90	13.92	13.91	13.91	13.87	13.90	13.87	13.94	13.90	13.91	13.90
Large-ME	15.12	15.41	15.38	15.28	15.17	15.21	15.10	15.03	14.94	14.91	14.77

Table 18: Fama-French (1992) Table 3 As Originally Published

B	ln(Size)	ln(BE/ME)	ln(AT/ME)	ln(AT/BE)	E/P Dummy	E(+)/P
0.15						
(0.46)						
	-0.15					
	(-2.58)					
-0.37	-0.17					
(-1.21)	(-3.41)					
		0.46				
		(5.09)				
			0.44	-0.43		
			(4.93)	(-3.54)		
					0.57	0.4.72
					(2.28)	(4.57)
	-0.11	0.32				
	(-1.99)	(3.92)				
	(-0.11)		0.31	-0.39		
	(-2.06)		(3.75)	(-3.08)		
	-0.16				0.06	2.99
	(-3.06)				(0.38)	(3.04)
	-0.13	0.33			-0.14	-0.87
	(-2.47)	(4.46)			(-0.90)	(1.23)
	-0.13		0.32	-0.46	-0.08	1.15
	(-2.47)		(4,28)	(-4.45)	(-0.56)	(1.57)

Each security is assigned the post-ranking β of its portfolio at the end of June of year t. If earnings are positive, then E(+)/P is the ratio of earnings to market equity and E/P dummy is 0. If earnings are negative, E(+)/P is 0 and E/P dummy is 1. The average slope is the time-series average of the monthly regression slope coefficients from July 1963 to December 1990 and the t-statistic is the average slope divided by its time-series standard error.

Table 19: Fama-French (1992) Table 3 Replicaiton

B	ln(Size)	ln(BE/ME)	ln(AT/ME)	ln(AT/BE)	E/P Dummy	E(+)/P
0.15						
(0.42)						
	-0.14					
	(-2.40)					
-0.36	-0.16					
(-1.05)	(-2.77)					
		0.46				
		(5.09)				
			0.44	-0.43		
			(4.93)	(-3.54)		
					0.39	2.90
					(1.95)	(3.78)
	-0.10	0.32				
	(-1.63)	(3.92)				
	(-0.10)		0.31	-0.39		
	(-1.70)		(3.75)	(-3.08)		
	-0.16				0.04	1.69
	(-2.76)				(0.31)	(2.22)
	-0.11	0.31			-0.10	0.367
	(-2.02)	(3.74)			(-0.80)	(0.61)
	-0.11		0.29	-0.34	-0.09	0.427
	(-2.04)		(3.60)	(2.81	(-0.73)	(0.67)

Constructed analogously to Fama and French's original table 3.