Distribution Agreement

Three Essays in Financial Economics

By

Joonki Noh
Doctor of Philosophy
in Business

Tarun Chordia, Ph.D.
Advisor

Narasimhan Jegadeesh, Ph.D.
Advisor

Francisco Barillas, Ph.D.
Committee Member

Shivaram Rajgopal, Ph.D.
Committee Member

Jay Shanken, Ph.D.
Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

Date

Three Essays in Financial Economics

By

Joonki Noh
B.S., Seoul National University, 2003
M.S., University of Michigan, 2005
M.A., University of Michigan, 2007
Ph.D., University of Michigan, 2007

Advisors:

Tarun Chordia, Ph.D.
Narasimhan Jegadeesh, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirement for the degree of
Doctor of Philosophy
in Business
2015

Abstract

Three Essays in Financial Economics

By Joonki Noh

This dissertation covers the information dissemination in financial markets and the risk-return relationship in the cross-section of stocks. The first essay (Industry Networks and the Speed of Information Flow) explores the information diffusion to equity markets in the framework of networks. I investigate whether the number of connections that an industry has in the network of inter-industry trade affects the speed of information flow to the industry. I find that the information flows substantially more slowly to central industries (which have more connections to other industries) from their related (=customer and supplier) industries than to peripheral industries from their related industries. The strong return predictability to central industries from related industries leads to highly profitable trading strategies whose risk-adjusted returns are 7.0% to 7.9% per annum. To explain this finding, I argue that investors who invest in central industries need to process more complicated information about related industries, slowing down the information flow to central industries. I find evidence that the sell-side analysts of central industries also face more complicated information about related industries, slowing down their processing new information about related industries. The network framework helps to identify an unknown and unique anomaly inherent in the industry network and to better understand potential sources of anomalies in financial markets. The second essay (Empirical Tests of Asset Pricing Models with Individual Stocks), co-authored with Narasimhan Jegadeesh, develops an instrumental variables methodology to test asset pricing models using individual stocks as test assets. We obtain consistent estimates of risk premiums, and simulation evidence indicates that the associated tests are well specified even in small samples. When testing three asset pricing models known to be successful in the literature when they were tested with characteristics-sorted portfolios as test assets, we find weak evidence that their factor risks are reliably priced in the cross-section of individual stock returns. The third essay (Information in CEOs' Facial Expressions: A First Look), co-authored with Narasimhan Jegadeesh and Jingran Zhao, investigates whether the visual cues such as facial expressions in CEOs' televised interviews can convey value-relevant information about firms to investors in financial markets and whether investors understand and react to it. We find evidence that negative facial expressions are correlated with cumulative abnormal returns and turnover over the next one to two days after air dates. We also find that negative facial expressions are

associated with firms’ one-quarter-ahead earnings. Taken together, this essay presents the first evidence in financial economics that CEOs’ facial expressions in their televised interviews can be a channel through which value-relevant information is disseminated.

Three Essays in Financial Economics

By

Joonki Noh
B.S., Seoul National University, 2003
M.S., University of Michigan, 2005
M.A., University of Michigan, 2007
Ph.D., University of Michigan, 2007

Advisors:

Tarun Chordia, Ph.D.
Narasimhan Jegadeesh, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirement for the degree of
Doctor of Philosophy
in Business
2015

Acknowledgement

This dissertation owes its existence to many people. First and foremost, I am deeply grateful to my two advisors: Narasimhan Jegadeesh and Tarun Chordia and committee members: Jay Shanken, Shivaram Rajgopal, and Francisco Barillas for their guidance, advices, encouragement and patience in the development of my dissertation. I would like to express my sincere gratitude to Clifton Green, Jeffrey Busse, Breno Schmidt, and Oliver Randall for their valuable suggestions and insightful comments on my research agenda, presentation, and writing. I am also thankful for discussion with Yakov Amihud, Hank Bessembinder, Michael Brennan, John Cochrane, Kent Daniel, Byung-Hyoun Hwang, Russell Jame, Juhani Linnainmaa, Avanidhar Subrahmanyam, Quan Wen, Dexin Zhou, Guofu Zhou, and seminar participants at Emory University, Texas Christian University, Syracuse University, Case Western Reserve University, Purdue University, Tulane University, State University of New York at Buffalo, Virginia Tech, University of Melbourne, University of New South Wales, University of Technology at Sydney, York University, and 2014 FMA Annual Meeting. I would like to extend my thanks to my colleagues, Kiseo Chung, Badrinath Kottimukalur, Shikha Jaiswal, and Hyeyoung Hah for their feedback on my presentation and writing. Last but not the least, I am immensely indebted to my parents and elder sister for their sacrifice and support during my Ph.D. journey. Without them, I would not be the person who I am in all aspects.

Table of Contents

List of Tables

List of Figures

Industry Networks and the Speed of Information Flow

Joonki Noh†

Abstract

I investigate whether an industry's position in the network of inter-industry trade affects the speed of information flow. I find that return predictability to central industries from their related (=customer and supplier) industries is substantially stronger than that to peripheral industries from their related industries. Long-short portfolios of central industries yield risk-adjusted returns of 7.0% to 7.9% per annum, which are 3.6% to 5.3% higher than those of peripheral industries. To explain this finding, I argue that investors who invest in central industries need to process more complicated information about related industries, making the prices of central industries slower to incorporate all the information. I find that sell-side analysts of central industries also face more complicated information about related industries, as their earnings forecast revisions of related industries predict their future revisions of central industries more strongly and for a longer period. In addition, I present evidence that our finding is not explained by existing anomalies.

†Joonki Noh is a doctoral student at the Goizueta Business School, Emory University, Atlanta, GA 30322, USA. Email: *Joonki.Noh@emory.edu*. I am grateful to Narasimhan Jegadeesh, Tarun Chordia, Jay Shanken, Clifton Green, Jeffrey Busse, Francisco Barillas, and Oliver Randall for their encouragement and insightful comments. I am also thankful for valuable suggestions from and discussions with Yakov Amihud, Kent Daniel, Juhani Linnainmaa, Byoung-Hyoun Hwang, Russell Jame, Quan Wen, Dexin Zhou, and seminar participants at Emory University, Texas Christian University, Syracuse University, Case Western Reserve University, Purdue University, Tulane University, State University of New York at Buffalo, and Virginia Tech.

1. Introduction

Financial economists have long recognized the importance of understanding how value-relevant information disseminates to stock markets and how market participants incorporate this information into stock prices. Classical asset pricing theories posit that value-relevant information diffuses immediately in a complete and frictionless market. However, considerable empirical evidence has been accumulated indicating that information can disseminate with sizable delay to financial markets.[1] The gradual information dissemination can be caused by many different sources, e.g., including asymmetric information, investors' limited cognitive resources, trading costs, institutional constraints, and other types of market frictions.

My paper explores this fundamental research topic, i.e., information flow to stock markets, in the framework of networks.[2] I question whether a node's (industry or firm) position in the network affects the complexity of value-relevant information that investors need to process and thus influences the speed of information flow through the network. For example, the wholesale and fishery industries are central and peripheral, respectively, and the former is related (or connected) to more industries than the latter by definition. Suppose that the wholesale industry buys a non-negligible amount of canned fish from the fishery industry and a hurricane hits southern coastal fisheries and damages them. The stock price of the fishery industry would reflect this negative shock immediately. How would the stock price of the wholesale industry react to this shock and the price drop of the fishery industry? To price the wholesale industry, investors need to understand not only shocks to other related industries and their price

[1] E.g., among others, Lo and MacKinlay (1990), Brennan et al. (1993), Badrinath et al. (1995), Chordia and Swaminathan (2000), Cohen and Frazzini (2008), Menzly and Ozbas (2010), and Cohen and Lou (2012).

[2] An emerging literature in finance and economics emphasizes the importance of direct and indirect connections through networks and investigates their economic and financial implications in different contexts, e.g., Acemoglu et al. (2012); Buraschi and Porchia (2012); Ahern (2013); Kelly et al. (2013); Ahern and Harford (2014); Aobdia et al. (2014); Anjos and Fracassi (2014); and Wu and Birge (2014).

movements but also how important canned fish is in the total revenue of the wholesale industry. Processing more complicated information (about related industries) can slow down the information flow to the wholesale industry.

I utilize the network of inter-industry trade to answer our question. Extant studies have documented that the market is segmented along the boundaries of industries (e.g., Menzly and Ozbas (2010)). Furthermore, the industry-level supply chain provides clear economic links through which shocks and relevant information can propagate (Menzly and Ozbas (2010) and Chen et al. (2014)). I thus gauge the speed of information flow to different positions in the industry network by measuring the strength of lead-lag relations of returns and that of earnings forecast revisions (by sell-side analysts) among economically related (=customer and supplier) industries.

The specific research question that I want to answer is as follows: does the value-relevant information flow to central industries from their related industries more slowly than to peripheral industries from their related industries? The answer to this question is not obvious ex-ante since (at least) two conflicting economic factors can affect the speed of information flow to central industries.[3] The first economic factor can slow down the information flow to central industries. Since central industries, by definition, are connected to more related industries than peripheral industries,[4] investors (with bounded rationality[5]) are required to process more complicated information about related industries (Recall the example of the wholesale and fisher industries above).[6] Thus it takes them longer to incorporate all the information into the prices of central industries, producing more gradual information dissemination.[7]

[3]Examples of central industries are the finance (real estate and banking), automobile, construction, and wholesale trade industries.

[4]In our empirical analyses, I use eigenvector centralities. For more details, see Section 3.1.

[5]See, e.g., Simon (1955) and Jensen and Meckling (1992).

[6]This can be interpreted to mean that investors who invest in central industries have higher information collection or processing costs than those in peripheral industries.

[7]When new information arrives, investors with limited cognitive resources do not perform the rational expectations inference to recover the information from observed prices and thus they do not adjust their demand fully as in Grossman and Stiglitz (1980). Similarly to our logic, Cohen and Lou

I call this the information complexity effect.[8] The slower information flow to central industries from their related industries implies that the return predictability and the predictability of earnings forecast revisions to central industries from their related industries are stronger than those to peripheral industries from their related industries. Figure 1 demonstrates the directions of information flow, return predictability, and the predictability of earnings forecast revisions when the industry in question is located at two different locations: the center (left) and the periphery (right) of the network. Note that related industries can be located in any positions in the network.

The second economic factor can accelerate the information flow to central industries. If investors understand the importance of central industries in shock propagation as shown in recent studies,[9] they may pay more attention or allocate more cognitive resources to central industries than to peripheral industries, making the information dissemination to central industries less gradual.[10] The faster information flow to central industries from their related industries implies that the return predictability and the predictability of earnings forecast revisions to central industries are weaker than those to peripheral industries. I call this the investors' attention effect.

In empirical analyses, I find that the information complexity effect substantially dominates the investors' attention effect in the center of industry network. More

(2012) find that the prices of conglomerates are slower to reflect the same piece of information than those of standalone firms.

[8] My information complexity effect is related to the literature of limited-information models and empirical tests of their predictions. A long line of extant work belongs to this literature. Examples of recent studies include Hong and Stein (1999), Hong et al. (2007a), Hong et al. (2007b), Cohen and Frazzini (2008), Menzly and Ozbas (2010), and Cohen and Lou (2012).

[9] Acemoglu et al. (2012) show that sectoral shocks from central industries, e.g., banking, automobile, and wholesale trade industries, are more likely to become macro-level fluctuations. Buraschi and Porchia (2012) and Ahern (2013) argue that central firms and industries are riskier due to their higher exposure to systematic risks.

[10] Among early theoretical studies, Holden and Subrahmanyam (1992) and Foster and Viswanathan (1993) show that the price reflects new information more rapidly as the number of informed investors increases. Among early empirical papers, using various proxies for investors' attention, Lo and MacKinlay (1990), Brennan et al. (1993), and Badrinath et al. (1995) document that the returns of stocks with high investors' attention lead those with low investors' attention.

specifically, I first document robust evidence that returns of related industries can predict future returns of central industries significantly more strongly than those of peripheral industries. I then quantify the economic magnitude of stronger return predictability to central industries from their related industries. For each month, I form quintile portfolios of central industries sorted on one-month lagged returns of their related industries, go long the quintile portfolio with the highest past return of related industries, and go short the quintile portfolio with the lowest past return of related industries. This self-financing trading strategy is rebalanced every month and involves the buying and selling of central industries. The self-financing trading strategies that trade central industries provide significant and economically large risk-adjusted returns, ranging from 7.0% (VW) to 7.9% (EW) per annum, after controlling for the exposure to five known factors. In contrast, the same self-financing trading strategies that trade peripheral industries produce risk-adjusted returns of 1.7% (VW) to 4.3% (EW) per annum.[11]

In the next set of tests, I conduct an in-depth investigation to better understand potential underlying economic mechanisms that drive the stronger return predictability to central industries from their related industries. My hypothesis is that investors who invest in central industries need to process more complicated information about related industries. If these investors have limited information processing capabilities, it would take them longer to process all the information about related industries and to incorporate it into the prices of central industries fully. I utilize the earnings forecast revisions of sell-side analysts to examine whether the speed of their responses to new information about related industries differs substantially in the center and periphery of the industry network. I uncover strong evidence that the sell-side analysts of central industries also need to process more complicated information about related industries, as their earnings forecast revisions of related industries predict their one-

[11] 1.7% (VW) is statistically insignificant at any conventional levels.

month-ahead and two-month-ahead revisions of central industries substantially more strongly than those of peripheral industries.[12] In contrast, the earnings forecast revisions of related industries fail to predict the future revisions of peripheral industries. This evidence is consistent with our information complexity effect.

My information complexity effect is distinct from the conglomerate effect documented by Cohen and Lou (2012). They find that the prices of conglomerates are slower to reflect the same piece of information than the prices of standalone firms that operate in the same industry. The conglomerate effect explains return predictability within an industry, while our information complexity effect explains the information flow across industries.[13] Investors who invest in standalone firms that belong to central industries can face more complicated information about related industries without having the conglomerate effect. In contrast, investors who invest in conglomerates that belong to peripheral industries can have less complicated information about related industries but experience the conglomerate effect. In empirical analyses, I present evidence that information flows still significantly more slowly to standalone firms in central industries from their related industries, and the magnitude of return predictability to these standalone firms and that to all firms in central industries are similar, implying that our findings are not explained by the conglomerate effect.

In addition, I test whether other anomalies that previous studies have documented can explain our findings. I examine four representative anomalies: (1) the limits to arbitrage effect captured by idiosyncratic volatility, (2) the institutional ownership effect by Badrinath et al. (1995) and Menzly and Ozbas (2010),[14] (3) the trading

[12]I find that, when new informative signals about related industries arrive in a given month, the sell-side analysts of central industries process 51%, 30%, and 19% of the new information in the same month, in the next month, and two months later, respectively. For more details, see Section 4.3.

[13]Another major difference is that in the conglomerate effect, the returns of standalone firms always lead those of conglomerates, while in our information complexity effect, the directions of information flow can change depending on which industries are the information sources.

[14]In these studies, institutional ownership is used as a proxy for investors' attention.

volume effect by Chordia and Swaminathan (2000), and (4) the illiquidity effect by Bali et al. (2014).[15] To control for each of these anomalies, I partition the stock universe into three sub-groups sorted on each of corresponding characteristics, e.g., idiosyncratic volatility. I uncover evidence that value-relevant information flows to central industries substantially more slowly within sub-groups, which indicates that our findings are not explained by these existing anomalies.

The remainder of this paper is organized as follow. Section 2 situates this paper in the extant literature. Section 3 presents the methodology to construct the US industry network and preliminary analyses. Section 4 tests whether the speed of information flow differs at different positions in the industry network. Section 4 also investigates potential underlying economic mechanisms that drive the difference in the speed of information flow across network positions. Section 5 concludes the paper. Throughout this paper, **return cross-predictability** signifies that the returns of customer and supplier industries predict the future returns of the industry in question.

2. Related Literatures

My paper is related to a large and well-established literature in finance on how value-relevant information diffuses to stock markets and what frictions induce return predictability. One strand of this literature investigates the lead-lag relations of returns among stocks. For example, Lo and MacKinlay (1990), Brennan et al. (1993), Badrinath et al. (1995), Chordia and Swaminathan (2000), Hou (2007), and Cohen and Lou (2012) document evidence that one group of stocks always react to common information faster than another group of stocks. More recent empirical studies present evidence that return predictability along the supply chain is a pervasive phe-

[15] I do not argue that I can disentangle our information complexity effect completely from underlying economic mechanisms that those existing studies argue. I aim to present evidence that our findings are not explained by the effects captured by the variables that those studies employed.

nomenon caused by investors' inattention and resulting market segmentation. Cohen and Frazzini (2008) provide evidence of return predictability along the firm-level supply chain. Menzly and Ozbas (2010) and Chen et al. (2014) show that the return predictability along the supply chain exists even at industry-level and in different asset classes: stocks and corporate bonds, respectively. My paper contributes to this literature by exploring the information flow to stock markets in the framework of networks and by identifying an unknown and unique friction inherent in the industry network, through which strong return predictability arises.

An emerging literature in finance and economics emphasizes the importance of the positions of industries in the US industry network in understanding the mechanisms of shock propagation and innovation transfer across industries. The seminal paper by Acemoglu et al. (2012) proposes a model in which idiosyncratic shocks from the central sectors of the economy can produce macro-economic fluctuations in outputs. Ahern (2013) investigates the relationship between industries' positions in the network and the cross-section of stock returns. He argues that companies in the center of the economy have higher market risk and thus their expected returns are higher as compensation for the risk. Aobdia et al. (2014) find evidence that the financial and accounting performance of central industries can be explained by systematic components to a larger extent than that of peripheral industries.[16] My paper contributes to this burgeoning literature by presenting evidence that an industry's position in the network affects the complexity of information that investors process and thus the speed of information dissemination through the network. I believe that this is an important advance in understanding the asset pricing implications of complex networks

[16] Aobdia et al. (2014) also find evidence that the changes in ROA of central industries can predict those of related industries more strongly. However, they fail to find evidence that the returns of central industries can predict those of related industries more strongly. This might seem contradictory to our findings. I emphasize that the opposite directions of lead-lag relations and information flow in these two papers lead to completely different economic effects and thus test outcomes.

and potential sources of anomalies in financial markets.

3. Methodology, Data, and Preliminary Analyses

3.1. Methodology for the Industry Network

Constructing the Industry Network: To construct an industry network, I exploit the Detailed Input-Output (IO) Tables produced by the Bureau of Economic Analysis (BEA).[17] As of February 2014, twelve BEA reports are available. Among these BEA reports, eliminating the earliest four reports that span from January 1947 through December 1971,[18] I use eight BEA reports for our empirical analyses. My entire sample period is January 1972 through December 2012. The BEA reports are published roughly every five years and they contain two tables: the MAKE and USE tables. The MAKE table records the dollar values of commodities that each industry produces and the USE table presents the dollar values of commodities that are consumed by each industry as inputs or by final user. In the BEA reports, a commodity means any good or service produced or provided by industries. The detailed BEA reports provide the MAKE and USE tables that record this information for roughly 400 to 500 industries and commodities,[19] whose exact numbers change across different BEA reports.

I interpret the MAKE and USE tables as matrices and perform several matrix algebraic operations.[20] If the (i, j) entry is negative, I move it and add its absolute

[17]These IO-tables are available at *http://www.bea.gov/industry/index.htm#benchmark_io*.

[18]The earliest four BEA reports do not have accounts for the "compensation for employees" or "(total) value added" that reflects the labor costs in their USE tables. Thus the use of these BEA reports might distort industry networks by overestimating the labor-intensive industries compared to the capital-intensive industries.

[19]In our sample period, the BEA employs its own six-digit industry codes (IO-codes), which roughly correspond to six-digit SIC codes. The exact definitions of the IO-codes are provided in the BEA reports.

[20]The (i, j) entry of the MAKE table is the dollar value of commodity j produced by industry i and the (i, j) entry of the USE table corresponds to the dollar value of commodity i consumed by industry j.

value to the (j, i) entry, facilitating the interpretations of the MAKE and USE tables. Following Becker and Thomas (2010) and Ahern and Harford (2014), I construct a matrix that records industry-to-industry trades, which I call REVSHARE, by combining the MAKE and USE tables.[21] Normalizing REVSHARE along each row produces a matrix, which I call CUST, recording the fraction of industry i's sales consumed by industry j. Therefore the (i, j) entry in CUST shows how important industry j is as a customer to industry i. Normalizing REVSHARE along each column produces a matrix, which I call SUPP, recording the fraction of industry j's purchases produced by industry i. Thus the (i, j) entry in SUPP shows how important industry i is as a supplier to industry j.

After creating these SUPP and CUST, I exclude the rows and columns that correspond to miscellaneous industry accounts, e.g., households, governments, special industries, and final users (including imports and exports), since the economic activities in these industries do not seem relevant to the theme of our paper. I then combine SUPP and CUST by averaging them into one square matrix, which I call COMB, and symmetrize the COMB by taking the maximum of (i, j) and (j, i) entries.[22] I relegate more detailed discussions on the construction of the industry network into Appendix 6.1.

Determining an Industry's Position in the Network: Interpreting the COMB as an adjacency matrix that defines the strength of links among nodes in

[21] I first normalize the MAKE table along each column and multiply it by the USE table, producing the REVSHARE. As in Ahern (2013), to properly account for the labor costs in combining the MAKE and USE tables, I generate an artificial industry in the MAKE table, i.e., the account for the compensation for employees. In the same spirit, I generate sets of artificial industries, which vary across different BEA reports. For example, the artificial industries for the 1997 BEA report include the accounts for the compensation for employees, non-comparable imports, used and secondhand goods, rest of world adjustment to final users, indirect business tax and non-tax liability, and other value added. For other BEA reports, sets of artificial industries are similarly determined based on the accounts for adjustments in the USE tables.

[22] This symmetrization has two purposes. First, it facilitates the economic interpretation of COMB as its (i, j) entry defines the strength of connection between industries i and j. However, I can not determine the directions of connections (either customers or suppliers) in COMB. Second, the symmetrization also prevents the eigenvector centralities from being complex-valued.

a network,[23] I can determine the positions of all industries in the US industry network. For each BEA report, the eigenvector centrality is defined as follows: Setting $A_{i,i} = 0$ for all is,

$$c_i = \frac{1}{\lambda} \sum_j A_{i,j} c_j, \tag{1.1}$$

where $A_{i,j}$ denotes the (i, j) entry of the adjacency matrix $\mathbf{A}$ and λ is a scaling constant (see Bonacich (1972)). Equation (1) is intuitively appealing because, to be more central in the network, an industry needs to be connected to more industries and/or connected strongly to other more central industries. Thus the eigenvector centrality captures not only the number of connections but also their strength in the industry network. In a matrix form, equation (1.1) is $\lambda\mathbf{c} = \mathbf{Ac}$, meaning that the principal eigenvector of $\mathbf{A}$ with the highest eigenvalue defines the eigenvector centralities of the industry network. Ahern (2013) argues that the eigenvector centrality is the most appropriate centrality measure for the network of inter-industry trade.

When assigning the eigenvector centralities to individual stocks, I use two different types of concordance tables provided by the BEA and US Census Bureau to map the BEA IO-codes into the standard SIC/NAICS codes.[24] Each BEA concordance table enables us to map the BEA IO-codes (unique in the corresponding BEA report) to the most recent SIC/NAICS codes available at its release date. The concordance tables by the Census Bureau enable us to convert SIC/NAICS codes in one BEA report to those in another BEA report.[25]

Dynamic versus Static Industry Networks: Assuming that the US industrial structure does not change over time substantially, Ahern and Harford (2014) and

[23] As in Ahern and Harford (2014), it is also possible to interpret SUPP and CUST as adjacency matrices of supplier and customer networks, respectively. However, in this case, the economic interpretations of network centralities may not be clear after symmetrizing the adjacency matrices.

[24] The concordance tables provided by the US Census Bureau are available at *http://www.census.gov/eos/www/naics/concordances/concordances.html*.

[25] I convert the 1997- and 2002-versions of NAICS codes to 1997-version of SIC codes since the NAICS codes in CRSP and COMPUSTAT are not well populated until 2004.

Aobdia et al. (2014) employ the 1997 BEA IO-tables located in the middle of their sample periods to construct the industry networks. Ahern and Harford (2014) present evidence that their findings in merger waves are robust to choosing a different BEA report. However, assuming that the industry network is static over time might not be valid for longer sample periods and/or in different applications.[26] For example, Carvalho and Gabaix (2013) provide evidence that the relative importance of US industries has changed substantially over their sample period: 1960 to 2008. They document that the US economy had experienced a decreasing share of manufacturing industries between 1975 and 1985 and the importance of financial industries has increased in recent years. They also argue that this change in the US industrial structure can explain the swings in the macroeconomic volatility such as the Great Moderation and its undoing.

Unlike the extant studies in which static networks are analyzed, I allow the industry network to change across different BEA reports. My main tests of return cross-predictability involve Fama-MacBeth (FM) cross-sectional regressions both in firm- and industry-levels, and thus they do not suffer from the time-inconsistency of the definitions of IO-codes across different BEA reports.[27] When testing the predictability of earnings forecast revisions among economically related industries, I perform firm-level pooled regressions. Industry-level pooled regressions can be unreliable due to the time-inconsistency in the definitions of IO-codes.

3.2. Other Data Sources and Variables

Other Data Sources: Empirical analyses in this paper are based on the intersection of the BEA IO-tables and two standard databases: daily and monthly

[26] The sample period of our paper is from January 1972 through December 2012, thus over 40 years.

[27] To reflect the changes in the US industrial structure properly, the BEA defines IO-codes for each BEA report and their definitions vary across different BEA reports.

financial data from the Center for Research in Security Prices (CRSP) and quarterly and annual accounting data available on the COMPUSTAT. The entire stock universe from NYSE/AMEX and NASDAQ is employed for the subsequent analyses after the following individual stock screenings. Since the trading characteristics of common stocks (with the CRSP share code of 10 or 11) might be different from those of other asset classes listed in exchanges, as in Chordia et al. (2000), I expunge assets that belong to the following categories from the sample universe: certificates, American depository receipts (ADRs), shares of beneficial interest, units, companies incorporated outside the US, Americus Trust components, close-end funds, preferred stocks, and real estate investment trusts (REITs). Similar to Acharya and Pedersen (2005) and Korajczyk and Sadka (2008), stocks are required to have prices above \$1.[28] Stocks are also required to have either historical SIC or NAICS codes since this information is utilized to combine the BEA IO-tables and the intersection of CRSP and COMPUSTAT.[29] After these screenings, I merge the stock universe with the BEA reports. My sample universe has the average cross-sectional size of 1608 stocks per month. The institutional ownership data are obtained from Thomson Financial's 13F Holdings database, which are combined with CRSP through historical CUSIP codes. The analysts coverage and earnings forecast data come from the I/B/E/S database.

Variables: I here explain the variables frequently used in the subsequent analyses. Following Fama and French (1992) and Davis et al. (2000), I compute the book-to-market (BM) ratios at the end of every calendar year. I then merge them with monthly financial items, allowing a six-month delay to ensure that market participants are fully aware of firms' accounting information released to markets. BM ratios are cross-sectionally winsorized at the 0.5% and 99.5% levels each month. The size (SIZE)

[28]Choosing different cutoffs, e.g., \$3 and \$5, does not affect our main results.

[29]When merging the BEA IO-tables and the intersection of CRSP and COMPUSTAT, I use the primary SIC/NAICS codes available in COMPUSTAT. When this information is missing in COMPUSTAT, I use SIC/NAICS codes available in CRSP.

of a firm is defined as the natural logarithm of its market capitalization (=number of shares outstanding $\times$ price). The return on asset (ROA) is defined as the ratio of the income before extraordinary items (IBQ in quarterly COMPUSTAT) to total assets (ATQ in quarterly COMPUSTAT). To avoid the survivorship bias in returns that might be induced by delisting, I adjust the daily and monthly CRSP returns for delistings as suggested by Shumway (1997). The one-month treasury-bill rate from Ibbotson Associates is used as a proxy for the risk-free rate.

The following firm characteristics are included in monthly FM cross-sectional regressions: lagged SIZE and BM ratio by one month, momentum (MOM) over the past eleven months defined as a summation of monthly returns from month $t-2$ to $t-12$, and short-term return-reversal (REV) defined as a one-month lagged return. To control for the exposure to known risk factors in computing the risk-adjusted returns of self-financing trading strategies, I include the Fama-French three factors (downloaded from Ken French's homepage[30]), i.e., the market return in excess of the risk-free rate (MKTRF), small-minus-big (SMB), and high-minus-low (HML); the momentum factor, i.e., up-minus-down (UMD); and the Pastor and Stambaugh tradable liquidity (PS-LIQ) factor.

3.3. Preliminary Analyses

Before starting our main empirical analyses, I here perform some preliminary analyses. Panel A of Table 1 provides the summary statistics for eigenvector centralities obtained from the 1992 Detailed BEA IO-tables.[31] It shows that the distribution of eigenvector centralities is positively skewed. Based on the network centralities, Panel B of Table 1 presents the twenty most and least central industries in 1992. Examples of the most central industries include the wholesale trade, construction,

[30] *http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html*

[31] I choose a BEA report located in the middle of our sample period.

finance, utility, and auto industries, while those of the least central industries contain the tobacco, hosiery, leather goods, and jewelery industries. These lists are largely consistent with our prior notion about which industries are likely to be central and with previous studies such as Ahern (2013) and Ahern and Harford (2014).[32]

Table 2 provides various average characteristics of centrality-sorted quintile portfolios: the market capitalization (MKTCAP) in billion dollars, BM ratio, return volatility (VOL), idiosyncratic return volatility (IVOL), share turnover (TURN),[33] the number of analysts following on a given month (ANALFOLL), and the percentage of institutional holdings (IHP). For each month, VOL is defined as the standard deviation of daily returns. To compute IVOL each month, I run a time-series regression of monthly firm returns on the Fama-French three factors over the past 24 months, and I define the standard deviation of monthly residuals as IVOL. TURN is defined as the ratio of share trading volume to the number of shares outstanding for a given month. These characteristics are based on the constituent stocks of each quintile portfolio. The two rightmost columns in Table 2 are based on the following industry-level characteristics: the Herfindahl-Hirschman indices (HHIs) of customer and supplier industries for the industry in question. I call them the customer-HHI and supplier-HHI, respectively. The customer-HHI for industry i is defined as, excluding s_{ii},

$$\text{Customer-HHI}_i = \sum_{j \neq i} s_{ij}^2, \tag{1.2}$$

where s_{ij} is the fraction of industry i's sales that industry j consumes and $\sum_j s_{ij} = 1$.

[32]The industry network that I constructed in this paper is different from the social accounting matrix (SAM) in Ahern (2013) in that mine models the US industries and their inter-connections as a network, while the SAM models the entire US economy as a network, thus including not only industries, but also government and household sectors.

[33]TURN is based on the NYSE/AMEX stock universe.

The supplier-HHI for industry j is defined in a similar way as, excluding p_{jj},

$$\text{Supplier-HHI}_j = \sum_{i \neq j} p_{ij}^2, \tag{1.3}$$

where p_{ij} is the fraction of industry j's purchases that industry i produces and $\sum_i p_{ij} = 1$. All portfolio characteristics are averaged over the sample period in which they are available. Table 2 indicates that the more central industries are, the higher BM ratios they have, the less volatile their idiosyncratic returns are, the more frequently traded they are, and the more analysts follow them. For MKTCAP, it is hard to find a clear monotonic pattern across centralities although the most central industries have the largest firms on average. For ANALFOLL, the highest quintile portfolio entertains about 0.5 more analysts following than the lowest quintile portfolio. This spread in the number of analysts following might seem small economically. For Customer-HHIs, the most and least central industries tend to have more concentrated customer industries and the values of Customer-HHIs are generally considered by the Antitrust Division of the US Department of Justice as being moderately concentrated.[34] In contrast, for Supplier-HHIs, although a similar pattern exists, their magnitudes and the spread across quintile portfolios are substantially smaller than those of Customer-HHIs. The values of Supplier-HHIs in Table 2 are generally considered by the Antitrust Division of the US Department of Justice as being unconcentrated.

4. Empirical Results

In this section, I conduct various tests to investigate the effect of an industry's position in the network on the speed of information flow. I first examine whether the returns of related industries can predict the future returns of central industries more

[34] *http://www.justice.gov/atr/public/guidelines/hhi.html.*

strongly than those of peripheral industries. By constructing self-financing trading strategies, I next quantify the economic magnitude of the difference in return cross-predictability between central and peripheral industries. I then perform an in-depth investigation to better understand potential underlying economic mechanisms that drive our findings.

4.1. Return Cross-predictability

I here test whether the returns of related industries can predict the future returns of central industries more strongly than those of peripheral industries. I consider various special cases of the following predictive FM cross-sectional regression: for firm (or industry) i in month t,[35]

$$r^e_{i,t} = \alpha_t + \gamma^T_{Z,t} Z_{i,t} + \sum_{l \in (C,M,P)} \gamma^{related}_{l,t} D_{i,l,t-1} r^{related}_{i,t-1} + \varepsilon_{i,t}, \tag{1.4}$$

where $r^e_{i,t}$ is the excess return of firm i, $Z_{i,t}$ contains the characteristics of interest, e.g., SIZE, BM, MOM, and REV, and the industry location subscript l is chosen from $(C, M, P) = (Center, Middle, Periphery)$ which correspond to centrality-sorted tercile portfolios. $D_{i,l,t-1}$ is an indicator variable defining the location of firm i in the industry network. For example, if the industry which firm i belongs to is located in the center of the industry network, $D_{i,C,t-1}$ is 1 in month $t-1$. Otherwise, it is 0. $r^{related}_{i,t-1}$ denotes the one-month lagged aggregate return of related (=customer and supplier) industries. To compute $r^{related}_{i,t}$, I employ the following three steps. For each industry, I first compute the industry-level return by averaging the returns of its constituent stocks. For each industry, I next calculate $r^{customer}_{i,t}$ ($r^{supplier}_{i,t}$) by weighting the industry-level returns of its customer (supplier) industries with the

[35]My tests for return cross-predictability are conducted both in firm- and industry-levels.

relative importance as a customer (supplier).[36] Then I average $r_{i,t}^{customer}$ and $r_{i,t}^{supplier}$ to obtain $r_{i,t}^{related}$. After running FM cross-sectional regressions, I take the time-series averages of slope coefficients. To make statistical inferences, I employ the heteroskedasticity and autocorrelation consistent (HAC) covariance matrix estimator proposed by Newey and West (1987), in which the number of lags is determined as suggested by Newey and West (1994).[37]

Panels A and B in Table 3 provide the slope coefficient estimates from regression (1.4) in firm- and industry-levels, respectively. In Column (1), I assign separate dummy variables to the center, middle, and periphery of the industry network. In Column (2), I assign dummy variables only to the center and middle of the industry network to formally test whether $\gamma_{C,t}^{related}$ and $\gamma_{M,t}^{related}$ are reliably different from $\gamma_{P,t}^{related}$. Column (1) in Panel A indicates that when the lagged return of related industries is used as the information source, $\widehat{\gamma}_l^{related}$s are monotonically aligned in network centralities, i.e., the highest ($\widehat{\gamma}_C^{related}$=0.182 with t-statistic=5.56) is in the center and the lowest ($\widehat{\gamma}_P^{related}$=0.095 with t-statistic=3.94) is in the periphery. Column (2) in Panel A confirms the finding by showing that the difference between $\widehat{\gamma}_C^{related}$ and $\widehat{\gamma}_P^{related}$ is 0.088 and statistically significant at the 5% level (t-statistic=2.36). In unreported tests, I find that controlling for industry-level momentum as in Moskowitz and Grinblatt (1999) and Amihud illiquidity level as in Amihud (2002) do not alter these results.

In Panel B, when running FM regressions in industry-level, I obtain similar results with the smaller magnitudes of $\widehat{\gamma}_l^{related}$s for all ls. For example, the industry-level and firm-level FM regressions produce the $\widehat{\gamma}_C^{related}$ of 0.140 and 0.182, respectively. However, the differences between $\widehat{\gamma}_C^{related}$ and $\widehat{\gamma}_P^{related}$ stay similar in industry- and firm-

[36]For industry k, the relative importance of industry j as a customer is s_{kj} in equation (1.2) and the relative importance of industry j as a supplier is p_{jk} in equation (1.3).

[37]The number of lags is determined as $\lfloor 4(T/100)^{\frac{2}{9}} \rfloor$, where T is the number of time-series observations and the operator $\lfloor x \rfloor$ extracts the integer portion of a real number x.

levels (0.076 and 0.088, respectively) and they are reliably different from zero at the 5% level. In summary, both firm- and industry-level FM regressions provide strong evidence that value-relevant information flows substantially more slowly to central industries from their customer and supplier industries than to peripheral industries.

4.2. Quantifying the Economic Magnitudes

Once finding that the returns of central industries are more predictable by their related industries, one may ask how much more attractive the trading strategy that invests exclusively in central industries can be to investors than the trading strategy that invests exclusively in peripheral industries. If central industries have higher exposure to systematic risk factors as documented by Ahern (2013) and Aobdia et al. (2014) and the higher profit of trading strategy that invests exclusively in central industries is mainly driven by this higher exposure, the difference in risk-adjusted returns between those two types of trading strategies might be economically negligible (after removing the portion explained by systematic risks). For self-financing trading strategies that invest in central (peripheral) industries, at the beginning of each month, I form quintile portfolios of central (peripheral) industries sorted on the average return of customer and supplier industries in the previous month. I then go long the quintile portfolio with the highest past return of related industries and go short the one with the lowest past return of related industries. These trading strategies are rebalanced every month and involve the buying and selling of central (peripheral) industries.

To determine central and peripheral industries, as before, I use centrality-sorted tercile portfolios. To compute risk-adjusted returns, I control for the exposure to the following five factors: the Fama-French three (MKTRF, SMB, HML), the momentum (UMD), and the Pastor-Stambaugh tradable liquidity (PS-LIQ) factors.[38] To avoid

[38]The original liquidity factor proposed by Pastor and Stambaugh (2003) is non-tradable. PS-LIQ

any potential forward-looking biases, I delay each BEA report until the end of the year in which it becomes publicly available. The different BEA reports became publicly available in 2013, 2007, 2002, 1997, 1994, 1991, 1984, and 1979, respectively.[39]

Excess and Risk-adjusted Returns: Panels A1 and B1 in Table 4 yield the annualized excess returns of value- and equal-weighted self-financing trading strategies, respectively. The rows labeled as "Central industries" ("Peripheral industries") correspond to the trading strategies that invest exclusively in central (peripheral) industries. For quintile portfolios that invest in central industries, in Panel A1, annualized excess returns and Sharpe ratios tend to decrease monotonically from the highest (in High (1)) to the lowest (in Low (5)) quintile portfolios. Excess returns range from 5.6% to 12.7% per annum and Sharpe ratios range from 0.299 to 0.686. Their annualized standard deviations are fairly stable across quintile portfolios. The long-short hedging portfolio (in High-Low) provides an annualized return of 7.1% with a Sharpe ratio of 0.544. For quintile portfolios that invest in peripheral industries, the long-short hedging portfolio produces an annualized return of 2.7% and a Sharpe ratio of 0.156, respectively, which are significantly lower than their counterparts that invest in central industries.

Controlling for the exposure to known risk factors, in Panel A2, the risk-adjusted return of the trading strategy that invests in central industries is 7.0% per annum and remains statistically significant at the 1% level (t-statistic=3.55). However, the risk-adjusted return of the trading strategy that invests in peripheral industries is 1.7% per annum and not reliably different from zero (t-statistic=0.53). The trading strategy that invests in peripheral industries has significant loadings on UMD and PS-LIQ. The difference in risk-adjusted returns between the trading strategies that invest in central industries and peripheral industries is 5.3% per annum. Thus, as

can be constructed based on factor-mimicking portfolios.

[39]When analyzing self-financing trading strategies, I do not use the 2007 BEA report since it was released at the end of 2013.

an investment strategy, the trading strategy that invests in central industries seems more attractive than the other trading strategy (before transaction costs).

I repeat the same analyses with equal-weighted quintile portfolios and report the results in Panels B1 and B2. For the trading strategy that invests in central industries, its annualized excess return is 8.1% with a Sharpe ratio of 0.877. Its risk-adjusted return is 7.9% per annum and remains statistically significant at the 1% level (t-statistc=4.64). The risk-adjusted return of the trading strategy that invests in peripheral industries is 4.3% per annum and becomes significant at the 5% level (t-statistic=2.15). The difference in risk-adjusted returns between the trading strategies that invest in central industries and peripheral industries is 3.6% per annum and significant at the 5% level.[40]

In summary, I find evidence that self-financing trading strategies that exclusively invest in central industries produce significant and economically large risk-adjusted returns, ranging from 7.0% to 7.9% per annum. In contrast, the risk-adjusted returns of self-financing trading strategies that exclusively invest in peripheral industries range from 1.7% to 4.3% per annum. This evidence indicates that the economic magnitude of the slower information flow to central industries than to peripheral industries is large.

4.3. Potential Economic Mechanism

In this section, I conduct an in-depth investigation to better understand what potential economic mechanisms drive the stronger return predictability to central industries from their related industries. My hypothesis is stated as follows. By definition, central industries are connected to more related industries than peripheral industries. Investors with limited information processing capabilities who invest in

[40] I run seemingly unrelated regressions (SURs) of monthly returns of these two trading strategies on five risk factors and test whether their intercepts are statistically different.

central industries thus need to process more complicated information about their related industries. It can take these investors longer to incorporate all the information about related industries into the prices of central industries, inducing the slower information flow to central industries from related industries. I call it the information complexity effect.

To examine whether the information complexity effect drives our findings, as in equation (1.4), I consider the following predictive FM cross-sectional regressions for multi-periods: for firm i in month t,

$$r^e_{i,t} = \alpha_t + \gamma^T_{Z,t} Z_{i,t} + \sum_{l \in (C,M,P)} \gamma^{related}_{l,t,k} D_{i,l,t-k} r^{related}_{i,t-k} + \varepsilon_{i,t,k}, \tag{1.5}$$

where $r^{related}_{i,t-k}$ is the k-month lagged aggregate return of related industries and all variables are standardized cross-sectionally. Table 5 presents the results of FM regressions (1.5) and evidence that the returns of related industries predict the future returns of central industries substantially more strongly and for a longer time period, up to two months. In Columns (5) and (6), when the predictability of two-month-ahead returns is tested (i.e., when $k = 2$), $\widehat{\gamma}^{related}_{C,2}$ is 0.013 and statistically significant at the 5% level (t-statistic=2.43), while $\widehat{\gamma}^{related}_{P,2}$ is 0.004 and insignificant at any conventional levels (t-statistic=0.87). Examining the comovement of returns of the industry in question and its related industries in Columns (1) and (2), I find that the returns of related industries comove with the returns of central industries more strongly than those of peripheral industries. The difference between $\widehat{\gamma}^{related}_{C,0}$ and $\widehat{\gamma}^{related}_{P,0}$ is 0.025 and is highly significant (t-statistic=3.94). For $k = 0, 1, 2$, although $\widehat{\gamma}^{related}_{C,k}$s and $\widehat{\gamma}^{related}_{M,k}$s support the information complexity effect, $\widehat{\gamma}^{related}_{P,k}$ can suggest alternative explanations for our findings. In particular, the stronger economic ties of central industries with their related industries might drive the stronger return predictability to central industries from their related industries than to peripheral industries.

I here attempt to disentangle the information complexity effect from the alternative explanation based on economic ties. I utilize the earnings forecast revisions of sell-side analysts to test whether the speed of their responses to new information about related industries differs substantially in the center and periphery of the industry network. If the sell-side analysts of central industries also face more complicated information about related industries as investors, then (1) their earnings forecast revisions of related industries would predict their future revisions of central industries more strongly than those of peripheral industries and (2) this stronger predictability of earnings forecast revisions to central industries would stay significant for a longer time period than that to peripheral industries.

For firm i in month t, as in Cohen and Frazzini (2008), I define monthly analysts' earnings forecast revision as $\text{AREV}_{i,t} = (\text{UP}_{i,t} - \text{DOWN}_{i,t})/\text{NUMEST}_{i,t}$, where $\text{NUMEST}_{i,t}$ is the number of available earnings forecast revisions for the current fiscal quarter end, and $\text{UP}_{i,t}$ ($\text{DOWN}_{i,t}$) is the number of upward (downward) earnings forecast revisions.[41] I emphasize that $\text{AREV}_{i,t}$ picks up the **signs** (instead of the levels) of earnings forecast revisions by sell-side analysts and thus $\text{AREV}_{i,t}$ is not affected by the strength of economic ties of the industry in question with its related industries. I consider the following pooled regressions for k-month lag ($k = 0, 1, 2$):

$$\text{AREV}_{i,t} = \alpha_i + \beta_t + \gamma_Z^T Z_{i,t} + \sum_{l \in (C,M,P)} \psi_{l,k}^{related} D_{i,l,t-k} \text{AREV}_{i,t-k}^{related} + \varepsilon_{i,t,k}, \tag{1.6}$$

where α_i and β_t denote firm- and time-fixed effects, respectively, and $Z_{i,t}$ contains control variables: lagged AREV, the aggregate return of related industries of firm i lagged by one month, i.e., $r_{i,t-1}^{related}$, and the industry-level analysts' revision (AREVIND) lagged by one month.[42] Note that all control variables are lagged by one month even

[41] Using analysts' earnings forecast revisions for the current fiscal year end produces similar results.

[42] Running monthly Fama-MacBeth regressions as in Cohen and Frazzini (2008) yields similar results to those in Table 6.

for $k = 2$. $\text{AREV}_{i,t-k}^{related}$ is the aggregate analysts' revision of related industries of firm i lagged by k months and it is computed in the same way as $r_{i,t-1}^{related}$. The location dummy variable $D_{i,l,t-k}$ in month $t-k$ is defined in the same way as before. To report t-statistics, I compute robust standard errors by double-clustering by firm and year-month. To facilitate the economic interpretation of slope coefficient estimates, I standardize all variables by subtracting the means and dividing by the standard deviations of all observations.

Table 6 presents the results of panel regression (1.6) and evidence that the sell-side analysts of central industries also face more complicated information about related industries, which slows down their processing new information about related industries. From Columns (3) and (4), I find that analysts' earnings forecast revisions of related industries predict their one-month-ahead revisions of central industries substantially more strongly than those of peripheral industries. In Column (3), $\widehat{\psi}_{C,1}^{related}$ is 0.049 and highly significant at the 1% level (t-statistic=6.06), while $\widehat{\psi}_{P,1}^{related}$ is 0.009 and insignificant at any conventional level (t-statistic=1.34). When testing whether the difference between $\widehat{\psi}_{C,1}^{related}$ and $\widehat{\psi}_{P,1}^{related}$ differs from zero reliably, in Column (4), I find that the difference is 0.040, more than four-times larger than $\widehat{\psi}_{P,1}^{related}$, and highly significant at the 1% level (t-statistic=4.34). From Columns (5) and (6), I find that analysts' earnings forecast revisions of related industries can also predict two-month-ahead revisions of central industries substantially more strongly than those of peripheral industries although the strength of predictability is reduced.[43] In contrast, I discover evidence that analysts' earnings forecast revisions of related industries fail to predict their future revisions of peripheral industries, i.e., $\widehat{\psi}_{P,1}^{related}$ and $\widehat{\psi}_{P,2}^{related}$ are insignificant at any conventional level. Columns (1) and (2) show that the comovement of $\text{AREV}_{i,t}$ and $\text{AREV}_{i,t}^{related}$ does not significantly differ in the center and periphery of the industry

[43] These results of two-month-ahead revision predictability survive controlling for $\text{AREV}_{i,t-1}^{related}$ additionally.

network, indicating that our findings in Columns (3) to (6) are not driven by the staleness in analysts' earnings forecast revisions.

Table 6 indicates that the sell-side analysts of central industries indeed respond to new information about related industries substantially more slowly than those of peripheral industries, while the sell-side analysts of peripheral industries process new information about related industries immediately (i.e., in the same month). By interpreting the slope coefficients, i.e., $\widehat{\psi}_{C,k}^{related}$, across columns in Table 6, I can have a better understanding of information processing by the sell-side analysts of central industries. When new value-relevant information about related industries becomes available in a given month, the sell-side analysts of central industries process 51%, 30%, and 19% of this new information in the same month, in the next month, and two months later, respectively.[44] In contrast, when new informative signals about related industries arrive in a given month, the sell-side analysts of peripheral industries process almost 100% of this new information about related industries in the same month. This evidence supports our information complexity effect strongly.

4.4. Distinction from Existing Anomalies

I here examine whether our findings can be explained by anomalies that previous studies have documented. I do not attempt to disentangle our information complexity effect completely from underlying economic mechanisms that those anomalies are potentially based on. I aim to present evidence that our findings are not entirely explained by effects captured by particular variables used in previous studies. I consider the following five anomalies: (1) the conglomerate effect by Cohen and Lou (2012), (2) the limits to arbitrage effect captured by idiosyncratic volatility, (3) the institutional ownership effect by Badrinath et al. (1995) and Menzly and Ozbas (2010), (4)

[44] I have $51\% = 0.085/(0.085 + 0.049 + 0.032) \times 100$ in the same month, $30\% = 0.049/(0.085 + 0.049 + 0.032) \times 100$ in the next month, and $19\% = 0.032/(0.085 + 0.049 + 0.032) \times 100$ two months later.

the trading volume effect by Chordia and Swaminathan (2000), and (5) the illiquidity effect by Bali et al. (2014).

Conglomerate Effect: Cohen and Lou (2012) find that conglomerates are slower to incorporate the same piece of information into their prices than standalone firms that operate in the same industry. If central industries have significantly more conglomerates than peripheral industries, ceteris paribus, the conglomerate effect might drive our findings. To disentangle our information complexity effect from the conglomerate effect, I choose standalone firms from the entire stock universe and then test whether value-relevant information still flows more slowly to standalone firms in central industries from their related industries. For each firm, I determine industry segments based on two-digit SIC codes and compute the HHI of segment sales. As in Cohen and Lou (2012), I define standalone firms as those that have HHIs greater than 0.64.[45]

Table 7 reports the testing results with standalone firms only. As in Table 3, Column (1) shows that the slope coefficients, i.e., $\widehat{\gamma}_l^{related}$s, of interaction terms between the lagged returns of related industries and location dummies are nicely aligned in network centralities. Column (2) confirms that the difference between $\widehat{\gamma}_C^{related}$ and $\widehat{\gamma}_P^{related}$ is 0.085 and significantly different from zero at the 5% level (t-statistic=2.13). With standalone firms which are free from the conglomerate effect by definition, I find that the magnitudes of $\widehat{\gamma}_C^{related}$ and $\widehat{\gamma}_P^{related}$ are slightly smaller than those with all firms (Table 3), implying that our findings are not explained by the conglomerate effect by Cohen and Lou (2012).

Limits to Arbitrage Effect: To test whether our findings can be explained by the limits to arbitrage captured by idiosyncratic volatility, I partition the entire stock universe into three sub-groups sorted on IVOL and then run FM cross-sectional

[45] Cohen and Lou (2012) define standalone firms as those that operate in one industry and whose segment sales account for more than 80% of the total sales. Altering the cutoff for standalone firms and using one-digit SIC codes to determine industry segments do not change the results in Table 7.

regressions in equation (1.4) within each sub-group. Table 8 presents the results of FM cross-sectional regressions for three sub-groups. Within sub-groups that have high and medium levels of IVOL, the slope coefficients, i.e., $\widehat{\gamma}_l^{related}$s, are nicely aligned in network centralities. For these sub-groups, the differences between $\widehat{\gamma}_C^{related}$ and $\widehat{\gamma}_P^{related}$ differ from zero significantly at the 5% level, indicating that information still flows significantly more slowly to central industries from related industries even after controlling for IVOL levels. When analyzing the sub-group that has low IVOL level, I find that the difference between $\widehat{\gamma}_C^{related}$ and $\widehat{\gamma}_P^{related}$ is not reliably different from zero, although $\widehat{\gamma}_l^{related}$ still increases monotonically in network centralities. Across three sub-groups, regardless of network positions, return cross-predictability becomes stronger monotonically as the level of IVOL increases, implying that the more binding limits to arbitrage are, the slower the information flow is.

Institutional Ownership Effect: Theories in the limited-information models predict that the higher level of investors' inattention makes information dissemination more gradual (see, e.g., Merton (1987); Hong and Stein (1999); Hirshleifer and Teoh (2003); Hong et al. (2007b)). Among empirical studies, e.g., Badrinath et al. (1995) and Menzly and Ozbas (2010) employ institutional ownership as a proxy for investors' inattention. To disentangle our information complexity effect from the institutional ownership effect, I partition the entire stock universe into three sub-groups sorted on the percentage ownership of institutional investors, i.e., IHP. I then test whether information still flows significantly more slowly to central industries from related industries within each sub-group.

Panel A in Table 9 reports the results of FM cross-sectional regressions when I control for the institutional ownership effect. Within all sub-groups, $\widehat{\gamma}_l^{related}$s increase monotonically in network centralities and the differences between $\widehat{\gamma}_C^{related}$ and $\widehat{\gamma}_P^{related}$ are significantly different from zero at the 5% level. Within the sub-groups of high and medium IHP levels, $\widehat{\gamma}_P^{related}$s are insignificant at the 5% level (t-statistics are 0.83 and 1.85, respectively). This implies that the weakest return cross-predictability to peripheral industries disappears first as IHP level increases from Column (6) to Column (1). Overall, these results strongly suggest that our information complexity effect is not explained by the institutional ownership effect.

Turnover Effect: Chordia and Swaminathan (2000) document that the returns of firms with high trading volume lead those with low trading volume. To test whether our findings can be explained by this turnover effect to a certain extent, I partition the NYSE/AMEX stock universe into three sub-groups sorted on share turnover, i.e., TURN. Panel B in Table 9 reports the results of FM cross-sectional regressions. Within sub-groups that have medium and low TURN levels, the differences between $\widehat{\gamma}_C^{related}$ and $\widehat{\gamma}_P^{related}$ stay significant at the 5% level (t-statistics are 2.11 and 2.09, respectively). For the sub-group that has high TURN level, this difference becomes insignificant although $\widehat{\gamma}_l^{related}$s are increasingly monotonic in network centralities.

Illiquidity Effect: Bali et al. (2014) document evidence that illiquidity contributes to the short-term stock market under-reaction (up to six months) and thus price discovery can be delayed following liquidity shocks. To test whether our findings can be explained by this illiquidity effect, I partition the NYSE/AMEX stock universe into three sub-groups sorted on Amihud illiquidity (ILLIQ). Panel C in Table 9 reports the results of FM cross-sectional regressions. Within sub-groups that have high and medium levels of ILLIQ, the differences between $\widehat{\gamma}_C^{related}$ and $\widehat{\gamma}_P^{related}$ stay significant at the 5% level (t-statistics are 2.07 and 2.14, respectively). For the sub-group that has low ILLIQ level, this difference is marginally significant at the 10% level (t-statistic=1.68). Overall, the evidence suggests that our findings are not explained by the illiquidity effect by Bali et al. (2014).

In summary, all the evidence in this section supports that our findings are distinct from existing anomalies that previous studies have documented in the literature and makes our information complexity effect more compelling as a potential explanation for our findings.

4.5. Change in Institutional Co-ownership

Institutional investors are likely to be more efficient in processing value-relevant information obtained from related industries or to have lower information collection costs than retail investors. In addition, self-financing trading strategies that invest in central industries might be fairly attractive to institutional investors in several respects. For example, Table 4 shows that the risk-adjusted returns of trading strategies with top centrality are significantly higher than those of the other two trading strategies. It is thus possible that institutional investors might exploit the stronger return cross-predictability from related industries to central industries for their investment.

I here test whether institutional investors increase (decrease) their positions more in central industries when they increase (decrease) the positions in related industries. I consider the following panel regression: for firm i in quarter q,

$$\Delta \mathrm{IHP}_{i,q} = \alpha_i + \beta_q + \sum_{l \in (C,M,P)} \theta_l^{related} D_{i,l,q} \Delta \mathrm{IHP}_{i,q}^{related} + \varepsilon_{i,q}, \tag{1.7}$$

where $\Delta \mathrm{IHP}_{i,q}$ denotes the change in the percentage ownership of institutional investors in firm i from quarter $q-1$ to quarter q and $\Delta \mathrm{IHP}_{i,q}^{related}$ is the aggregate change in the percentage ownership of institutional investors in related industries of firm i from quarter $q-1$ to quarter q.[46] I compute $\Delta \mathrm{IHP}_{i,q}^{related}$ in the same way as $\mathrm{ROA}_{i,q}^{related}$ and $\Delta \mathrm{ROA}_{i,q}^{related}$. To control for any potential biases that might be induced by unobserved heterogeneity across firms and for systematic fund inflows

[46] I winsorize the percentage ownership of each firm by institutional investors into 100% when it is over 100%. Winsorizing or eliminating these observations produces similar results.

and outflows of institutional investors, firm (α_i) and year-quarter (β_q) fixed effects are included in regression (1.7), respectively. To report t-statistics, robust standard errors are computed by double-clustering by firm and year-quarter.

Table 10 presents the results of panel regression (1.7) and shows that all $\widehat{\theta}_l^{related}$s have similar values to each other (in Column (1)) and their differences are not significant (in Column (2)). For example, $\widehat{\theta}_C^{related}$ is not reliably different from $\widehat{\theta}_P^{related}$ with a low t-statistic of 0.19. The finding that $\widehat{\theta}_l^{related}$s are all significant with high t-statistics indicates that institutional investors exploit the value-relevant information disseminated from customer and supplier industries when they rebalance their portfolio positions. However, the insignificant differences among $\widehat{\theta}_l^{related}$s do not support that institutional investors invest (disinvest) more in central industries than in peripheral industries when they invest (disinvest) in related industries. This implies that even sophisticated and potentially more informed institutional investors as a whole do not exploit the difference in return cross-predictability across network positions.

Overall, the finding in this section is consistent with the small spread (=0.5 analysts) in the number of analysts following across centrality-sorted quintile portfolios presented in Table 2. These two findings imply that investors as a whole do not pay significantly more attention nor allocate by far more cognitive resources to central industries although they need to process more complicated information about related industries.

5. Conclusions

I interpret the US industries and their inter-industry trades as a network over 40 years and identify an unknown and unique anomaly inherent in the US industry network. I find robust and strong evidence that value-relevant information flows substantially more slowly to central industries from their economically related (=customer and supplier) industries than to peripheral industries. Accordingly, I uncover evidence that the returns of related industries predict the future returns of central industries substantially more strongly. Long-short portfolios of central industries formed on the past returns of their related industries yield significant and economically large risk-adjusted returns of 7.0% to 7.9% per annum.

To better understand a potential economic mechanism that drives our findings, I conduct an in-depth investigation into the differential information complexity that investors at different positions in the industry network face. The information complexity effect is that investors who invest in central industries need to process more complicated information about economically related industries, and thus it take them longer to process all the information and to incorporate it into the prices of central industries fully. To support the information complexity argument, I document strong evidence that the sell-side analysts of central industries respond to new information about related industries substantially more slowly than those of peripheral industries. I also present evidence that our information complexity effect is not explained by various anomalies that previous studies have documented.

6. Technical Appendix

This section covers the detailed materials not provided in the main text of this paper: the construction of the industry network from the BEA reports.

6.1. Details on Constructing the Industry Network

For each pair of the MAKE and USE tables, I aggregate the IO-codes with the same SIC/NAICS codes since they will be assigned to the same stocks when I merge the BEA report and the standard databases: CRSP and COMPUSTAT. The lists of aggregated IO-codes are different across different BEA reports. I also expunge the IO-codes that do not have SIC/NAICS codes in the BEA concordance tables.

Following Ahern (2013), I generate sets of artificial industries to take the adjustment accounts in the USE table into consideration. After adding the rows and columns for these artificial industries to the MAKE table, I combine the MAKE and USE tables, producing the REVSHARE matrix. The list of artificial industries is as follows: (1) non-comparable imports, (2) used and second-hand goods, (3) rest of the world adjustment to final users, (4) compensation for employees, (5) taxes on production and imports less subsidies, (6) gross operating surplus, (7) indirect business tax and non-tax liability, (8) other or total value added, (9) commodity credit corporation, and (10) profit-type income. Subsets of these accounts show up in different BEA reports and these artificial industries are not included when I compute the eigenvector centralities of the industry network.

Empirical Tests of Asset Pricing Models with Individual Stocks

Narasimhan Jegadeesh and Joonki Noh[☆]

Abstract

We develop an instrumental variables methodology to obtain consistent estimates of risk premiums using individual stocks as test assets. Simulation evidence indicates that this methodology yields nearly unbiased estimates of risk premiums and that the associated tests are well specified even in small samples. We test a number of asset pricing models proposed in the literature using this approach. We find that the CAPM market risk, SMB and HML factors risks, investment and ROE factors risks under the production-based asset pricing model, and the liquidity-adjusted market risk under the LCAPM are not priced in the cross-section of individual returns.

[☆]Narasimhan Jegadeesh is the Dean's Distinguished Professor at the Goizueta Business School, Emory University, Atlanta, GA 30322, phone: 404-727-4821, email: Jegadeesh@emory.edu. Joonki Noh is a Doctoral student at Emory University and the corresponding author, phone: 734-276-3162, email: Joonki.Noh @emory.edu. We would like to thank Francisco Barillas, Hank Bessembinder, Michael Brennan, Tarun Chordia, John Cochrane, Jay Shanken, Avanidhar Subrahmanyam, Guofu Zhou, and seminar participants at Emory University, University of Melbourne, University of New South Wales, University of Technology at Sydney, York University, and 2014 FMA Annual Meeting.

1. Introduction

One of the fundamental concepts of financial economics is that capital market investors are compensated for higher systemic risks through higher returns. While this basic concept is well accepted, there is very little agreement on the specific risk factors that indeed command risk premiums. The Sharpe-Lintner CAPM laid the theoretical foundation for this concept but the empirical support for CAPM is weak at best. A number of recent papers propose a variety of other risk factors that in theory should explain the cross-sectional differences in expected returns. Some of these factors are the Fama and French size and book-to-market factors, human capital risk (Jagannathan and Wang, 1996), productivity and capital investment risk (Cochrane, 1996; Chen, Novy-Marx and Zhang, 2011), different components of consumption risk (Lettau and Ludvigson, 2001; Ait-Sahalia, Parker, and Yogo, 2004; Li, Vassalou, and Xing, 2006), cash flow risk and discount rate risk (Campbell and Vuolteenaho, 2004) and illiquidity risk (Pastor and Stambaugh, 2003; Acharya and Pedersen, 2005).

These studies also typically report empirical evidence supporting the hypotheses that the loadings on the risk factors that they propose have ability to explain the cross-sectional differences in expected returns. Their empirical tests use selected portfolios as test assets, and in many instances these portfolios are sorted by size and book-to-market ratios to obtain the cross-sectional variation in average returns. The ability of so many different factor loadings to explain the cross-sectional differences in expected returns have generated some skepticism in the literature. Lewellen, Nagel, and Shanken (2010) show that the strong factor structure inherent in the test portfolios can enable any factors that are even weakly correlated with the characteristics used to sort the test portfolios to explain the differences in average returns across them regardless of the economic merits of the underlying theory. Daniel and Titman (2012) find that the factors proposed in various papers exhibit low time-series correlation with one another and argue that it is unlikely that the loadings of all these factors can simultaneously explain the cross-section of expected returns.

There are many other papers that propose a variety of other risk factors. Harvey, Liu and Zhu (2013) survey this literature and report that hundreds of factors have been proposed. Some of them are derived from theoretical models and others are based on empirical observations. To understand the economic importance of these factors, we should examine whether the associated risks are priced in the market, but portfolio based tests are not well suited to address this issue because of the low dimensionality problems. Cochrane's (2011) AFA presidential address notes this problem and states that "we must address the factor zoo, and I do not see how to do it by a high-dimensional portfolio sort." (p. 1063)

We develop a procedure to test asset pricing models using individual stocks as tests assets, and this procedure is not exposed to the low dimensionality problems inherent in tests using portfolios. The main reason why the literature typically uses portfolios as test assets is that the errors-in-variables (EIV) problem is less severe with portfolios than with individual stocks. Our approach uses the instrumental variables technique to address the EIV problem. We refer to the estimator that we propose as the IV estimator.[1]

The EIV problem arises because the standard approach estimates factor sensitivities in the first stage and uses these estimates as independent variables in the second stage cross-sectional regressions. The estimation error in factor sensitivities biases the factor risk premium estimates when the standard Fama-MacBeth approach (the "FM" estimator) is used. We propose an instrumental variables approach to estimate the second stage regression. Specifically, we use factor sensitivities estimated in even months as independent variables and the corresponding sensitivities estimated in odd months as their instruments or vice versa.

The IV estimator is N-consistent in the number of stocks (N) in the cross-section for a finite T under the mild technical assumptions that have been used in the literature,

[1] After circulating earlier drafts of this paper, we came across Pukthuanthong and Roll (2014) that also proposes an IV approach, which it suggests could be used to identify common factors. That paper does not address the consistency or the rate of convergence properties of the IV estimator. There are also a number of other differences between our paper and theirs in the issues we address and the details of the implementation of the IV approach.

unlike the FM estimator. For a sufficiently large N, the IV estimator is also T-consistent as the length T of the sample period used in the first stage regression to estimate factor sensitivity tends to infinity. For a sufficiently large N, we show that the rate of convergence "in probability" for the IV estimator is exponential in T, while the rate of convergence "in probability" for the standard FM estimator is linear in T.

While large sample properties can provide some guidance, it is important to examine the small sample performance of various estimators for practical applications. To do so, we conduct a number of simulation experiments. We choose simulation parameters to be similar to those in the actual data. Simulations with a single factor model find that the FM estimator is significantly biased in finite T due to the EIV problem. The risk premium estimates are biased even when factor sensitivities are estimated with about 2640 daily observations. In contrast, the IV estimator yields nearly unbiased estimates even when only about 264 daily observations are used to estimate factor sensitivities. We also find that the conventional t-tests are well specified in small samples with the IV estimator. We find similar results with multifactor models as well.

We apply the IV approach to estimate the risk premiums for several factors proposed in the literature. The models that we test with individual stocks are the CAPM, the Fama-French three-factor model, a production-based asset pricing model by Chen, Novy-Marx and Zhang (2011), and the liquidity-adjusted CAPM by Acharya and Pedersen (2005). The rest of the paper is organized as follows: Section 2 presents our methodology. Section 3 examines the finite sample properties using simulation experiments. Section 4 presents the empirical tests of existing asset pricing models, and Section 5 concludes the paper. Appendices provide the details of simulation experiments and mathematical proofs.

2. Methodology

A number of asset pricing models predict that expected returns on risky assets are linearly related to their covariance with certain risk factors. A general specification of a K-factor asset pricing model can be written as:

$$E(r_i) = \gamma_o + \sum_{k=1}^{K} \beta_{i,k} \times \gamma_k \tag{2.1}$$

where $E(r_i)$ is the expected excess return on stock i, $\beta_{i,k}$ is the sensitivity of stock *i* to factor k, and γ_k is the risk premium on factor k. γ_o is the excess return on the zero-beta asset. If riskless borrowing and lending are allowed, then the zero-beta asset earns the risk-free rate and its excess return is zero, i.e. $\gamma_o = 0$. The CAPM predicts that only the market risk will be priced in the cross-section. Several multifactor models identify additional risk factors based on empirical findings or based on variations of models such as the ICAPM by Merton (1973).

Empirical tests of asset pricing models typically use the Fama-MacBeth two-stage regression procedure to estimate factor risk premiums. The first stage estimates factor sensitivities using the following time-series regressions with *T* periods of data:

$$r_{i,t} = a_i + \sum_{k=1}^{K} \beta_{i,k} \times f_{k,t} + u_{i,t}, \tag{2.2}$$

where $f_{k,t}$ is the realization of factor *k* in time *t*. The time series estimates of factor sensitivities, say $\hat{\beta}_{i,k}$, are the independent variables in the following second stage cross-sectional regressions used to estimate factor risk premiums:

$$r_{i,t} = \gamma_{o,t} + \sum_{k=1}^{K} \hat{\beta}_{i,k} \times \gamma_{k,t} + \varphi_{i,t}, \tag{2.3}$$

where realized excess return $r_{i,t}$ is the dependent variable. The standard FM approach fits OLS regression to estimate the parameters of cross-sectional regression (2.3). These OLS estimates are biased due to the EIV problem since $\hat{\beta}_{i,k}$s are estimated with errors. To mitigate such bias, the literature typically uses selected portfolios as test assets rather than individual stocks since portfolio betas can be estimated more precisely than individual stock betas.

The use of test portfolios, however, presents a different set of problems. The test portfolios are typically sorted on a few characteristics such as size and book-to-market. Sorting on characteristics that are known to predict returns helps generate a reasonable variation in average returns across test assets. However, Lewellen, Nagel, and Shanken (2010) point out sorting on characteristics also imparts a strong factor structure across the test portfolios. Lewellen *et al.* (2010) show that as a result even factors that are weakly correlated with the sorting characteristics would explain the differences in average returns across test portfolios regardless of the economic merits of the theories that underlie the factors.

Moreover, the statistical significance and economic magnitudes of risk premiums estimated using regression (2.3) could critically depend on the choice of test portfolios. For example, the Fama-French SMB and HML risk factors are significantly priced when test portfolios are sorted based on size and book-to-market, but they do not command significant risk premiums if test portfolios are sorted only based on momentum.

This paper proposes a methodology that uses individual stocks as test assets which addresses these problems. The use of individual assets preserves the dimensionality of the variation in expected returns that we observe in the stock market. Also, since our tests use all listed stocks individually, the asset pricing test results are not dependent on subjective choices made to construct the test portfolios.

We propose an instrumental variables regression to estimate the risk premiums in regression (2.3). To describe our IV estimator, rewrite regression (2.3) in a matrix form as:

$$R_t = \widehat{\mathrm{B}}\,\Gamma_t + \Phi_t$$

where R_t is the $N \times 1$ vector of realized excess returns in month t, $\widehat{\mathrm{B}}$ is the $N \times (K+1)$ matrix containing the intercept and K factor loadings, and Γ_t is the $(K+1) \times 1$ vector that includes K factor risk premiums if N stocks are used. For each month t, we propose the following instrumental variables estimator:

$$\hat{\Gamma}_{IV,t} = \left(\widehat{\mathrm{B}}'_{odd}\,\widehat{\mathrm{B}}_{even}\right)^{-1}\hat{B}'R_t, \tag{2.4}$$

where:

$\widehat{\mathrm{B}}_{odd}$ is the matrix of factor sensitivities estimated using data in odd months (“odd-month betas” or “odd-month factor sensitivities”),

$\widehat{\mathrm{B}}_{even}$ is the matrix of factor sensitivities estimated using data in even months (“even-month betas” or “even-month factor sensitivities”), and

$\hat{B}$ is $\widehat{\mathrm{B}}_{odd}$ if month t is even and $\widehat{\mathrm{B}}_{even}$ otherwise.

We use odd-month beta estimates as instruments when month t is even and the even-month beta estimates otherwise. In our empirical applications (see section III), we estimate factor sensitivities with daily data available over the past three years before month t for regression (2.4). We then slide forward the three-year rolling window by one month and repeat the estimations of factor sensitivities and risk premiums. In this section, for simplicity, we omit the subscript t of factor sensitivities. The IV estimator is consistent under the conditions in the proposition below:

Proposition 1: *Suppose stock returns follow an approximate factor structure. If the number of stocks in the cross-section is sufficiently large then under mild regularity conditions, the IV estimator given by equation (2.4) is consistent in the number of observations used to estimate factor sensitivities* (T-consistent).[2]

Proof: See Appendix 1.

Corollary: *The IV estimator is N-consistent for any finite T.*

As Proposition 1 shows, for sufficiently large N, the IV estimator is T-consistent as T grows without bound. For a fixed T, the IV estimator is N-consistent as we allow the number of stocks in the cross-section to increase indefinitely. In contrast, the FM estimator is biased under this assumption. The proposition below shows that the IV estimator converges faster in probability than the FM estimator.[3]

Proposition 2: *Suppose stock returns follow an approximate factor structure. If the number of stocks in the cross-section is sufficiently large, under mild regularity conditions, then:*

---

[2] Shanken (1992) defines a risk-premium estimator as N-consistent if for a finite T the estimator converges to the realized risk premium in the sample as N, the number of stocks in the cross-section, increases indefinitely. An estimator is T-consistent if the estimator converges to the population risk premium as T increases indefinitely.

[3] Note that the convergence in probability is not about the IV estimator itself. Throughout the paper, the rate of convergence in probability denotes the rate of convergence of the probability bound, i.e., $f(T)$, in the right-hand side of the following inequality: For any $\varepsilon > 0$, when $x_1, x_2, \dots, x_T$ are samples,

$$\Pr(|\bar{x}_T - \mu| > \varepsilon) < f(T),$$

where $\bar{x}_T$ is the sample mean, μ is the population mean, and $f(T)$ is a function of T whose value is positive.

a. *The rate of convergence in probability of the IV estimator given by equation (2.4) is exponential in T; and*

b. *The rate of convergence in probability of the FM estimator by regression (2.3) is linear in T.*

Proof: See Appendix 1.

To see the intuition behind this proposition, consider the sources of *T*-inconsistencies for the IV and the FM estimators when *T* is finite. In the case of the FM estimator, the EIV problem arises because of a term related to the variance of the measurement error in beta estimates, which decreases linearly in *T*. Since variance is always positive, *T* should be sufficiently large to make the positive number converge to zero. This implies that the probability bound of the FM estimator decays linearly in *T*. On the other hand, the *T*-inconsistency of the IV estimator arises due to a term related to average factor surprise and its expected value is zero.[4] So for the IV estimator to be *T*-consistent, *T* should be large enough so that the sample mean converges to its population mean, whose probability bound decays at an exponential rate in *T* (See Bahadur and Rao, 1960).

The literature also offers an alternative approach to address the EIV problems in the second stage regressions. Since the measurement error in factor loadings is the source of the EIV bias in the second stage regression (2.3), one could in principle undo the EIV bias using a factor that is the appropriate function of the measurement error. Litzenberger and Ramaswamy (1979) propose such a correction to estimate the CAPM risk premium with individual stocks. In the case of the CAPM or any single factor model, the measurement error in betas bias the slope coefficient estimate in regression (2.3) by a factor equal to the variance of true betas divided by the variance of estimated betas (which equals the variance of the true betas plus the variance of measurement errors). So if one can obtain a consistent estimate of the variance of measurement errors in beta estimates, then one could analytically correct for the EIV bias and obtain the *N*-consistent estimates of risk premiums in regression (2.3). The Litzenberger and Ramaswamy

[4] In addition to the EIV term, the FM estimator also has a term related to average factor surprise, whose probability bound decays exponentially in *T*. However, the rate of convergence in probability of the FM estimator is determined by the EIV term, whose probability bound decays linearly in *T*. For detailed discussion, see Appendix 1.

correction assumes that residual returns from the market model regressions are asymptotically weakly correlated across stocks. Shanken (1992) generalizes this EIV correction to multifactor asset pricing models. Brennan *et al.* (1998) propose another alternative method to avoid the EIV bias by employing risk-adjusted returns as the dependent variable in the second stage regressions. However, their methodology is not allowed to estimate the risk premiums of factors.

Adjustment for Finite Sample Moments: As the number of assets (*N*) grows infinitely for fixed *T*, it is trivial to show that the IV estimator is asymptotically normally distributed. However, since $\widehat{\mathrm{B}}'_{odd}\,\widehat{\mathrm{B}}_{even}$ in equation (2.4) might not be positive definite for large but finite *N*, there is small but still positive probability that the IV estimtator in equation (2.4) has an extremely large value due to $\widehat{\mathrm{B}}'_{odd}\,\widehat{\mathrm{B}}_{even}$ being near-singularity, which can make the finite sample moments of the IV estimator not exist. To avoid this ill-behaved finite sample property, we truncate the IV estimator based on the sample means and standard deivations of realizations of risk factors. Shanken and Zhou (2007) employ a similar approach for maximum-likelihood estimator based on portfolios. Specially, suppose that $\hat{\gamma}_{k,t}$ is the risk premium estimate of factor k in month *t*. We treat $\hat{\gamma}_{k,t}$ as a missing value if the deviation of $\hat{\gamma}_{k,t}$ from the sample mean of factor k realizations is greater than six times of their sample standard deviation. Excluding the extreme values of $\hat{\gamma}_{k,t}$ ensures that all finite sample moments of the IV estimator exist. In our simulations (see section 3), we find that the chance of thruncation binding is negligeable, less than 0.1% on average, for all choices of *N* and *T* and all risk factors, and find that this chance decreases as either *N* and *T* increases. In our real data analyses (see section 4), the chance of truncation binding becomes higher but is below 3% for all asset pricing tests. For example, when testing the CAPM and Fama-French three-factor model, the chance of $\hat{\gamma}_{HML,t}$ truncation binding tends to be the highest among the three Fama-French factors and it is close to 3% over the entire sample period (January 1956 through December 2012) having 684 months in total. In the case of HML, this 3% corresponds to about 20 months and most of the truncation binding occur in early years when the numbers of individual stocks range from 300 to 400, which are substantially smaller than the average size of cross-sections (=1934 stocks) over the entire sample period.

3. Small Sample Properties

The results so far indicate that the rate of convergence in probability of the IV estimator is faster than that of the standard FM estimator as the number of observation used to estimate factor sensitivities grows without bound. While our analytical results based on large sample properties could point us in the right direction, the small sample performance of an estimator would critically influence its adoption. This section examines the small sample properties of the following two-stage approaches to estimating the factor risk premiums under the CAPM and the Fama-French three-factor model:

(i) FM: The standard Fama-MacBeth methodology that estimates the risk premiums in the second stage regression using OLS methodology

(ii) IV: The instrumental variables methodology that we propose.

We employ individual stocks as tests assets in the simulation experiments.

3.1. Return Generating Processes

We consider a general setting where K common factors are priced. Under this model, we can specify realized daily stock returns as:

$$r_{i,\tau} = \alpha_i + \sum_{k=1}^{K} \beta_{i,k} \times f_{k,\tau} + \varepsilon_{i,\tau}, \tag{2.5}$$

where :[5]

$r_{i,\tau} \equiv$ Day τ excess return of stock *i;*

$\alpha_i \equiv$ Expected excess return of stock *i*

$\beta_{i,k} \equiv$ Sensitivity of stock *i* to the priced factor *k;*

$f_{k,\tau} \equiv$ Day τ realization of the priced factor *k*; $E(f_{k,\tau}) = 0$; and

$\varepsilon_{i,\tau} \equiv$ Day τ residual return of stock *i*.

[5] For clarity, we use τ to denote data for a particular day and t to denote data for a particular month.

Under the *K*-factor model, the expected excess returns are given by:

$$E(r_i) \quad = \gamma_0 + \sum_{k=1}^{K} \beta_{i,k} \times \gamma_k \tag{2.6}$$

where γ_k is the risk premium on the priced factor k. The *K*-factor asset pricing model imposes no restriction on the covariance structure of residual returns across assets. In general not all factors would be priced and hence residual returns can be correlated across assets.

3.2. Simulation Experiments: Parameters and Methodology

We choose the simulation parameters based on the corresponding statistics in the actual data over the sample period January 1956 through December 2012. For the single factor model, we match the simulation parameters to the average market risk premium, the risk free rate, the distribution of betas, and volatility of firm-specific returns from the actual data.

We use the CRSP value-weighted index as the market index and the short-term T-bill rate as the risk free rate to estimate these parameters. We run the market model regression for each stock to compute betas and residual returns. We then conduct the simulation with a sample of *N* stocks. We randomly generate daily returns using the following procedure:

1) Randomly generate beta and $\sigma_{\varepsilon,i}$ for each stock from normal distributions with means and standard deviations equal to the corresponding sample means and standard deviations from the actual data. For $\sigma_{\varepsilon,i}$, we take the absolute values of the random draws to make it positive. Thus $\sigma_{\varepsilon,i}$ follows a folded normal distribution.

2) Generate market excess return for each day as a random draw from a normal distribution with mean and standard deviation equal to the sample mean and standard deviation from the data.

3) Generate $\varepsilon_{i,\tau}$ for each stock from independent normal distributions with mean zero and standard deviation corresponding to the value generated in step 1).

For each stock i, we compute excess return in day τ as:

$$r_{i,\tau} = \beta_i \times r_{m,\tau} + \varepsilon_{i,\tau}. \tag{2.7}$$

We repeat the simulation 1,000 times each for N = 1000 and 2000, and T=264, 528, 792, 1320, and 2640 days. Each "month" in the simulation is 22 days long and we denote month with the subscript t. Therefore, the sample periods we use in the simulations range from 12 to 120 months.

For the first stage regression in the simulations, we fit the following market model regression with daily excess returns data for each stock to estimate beta:

$$r_{i,\tau} = \alpha_i + \beta_i \times r_{m,\tau} + e_{i,\tau}. \tag{2.8}$$

We use returns at a daily frequency rather than at a monthly frequency to get more precise beta estimates (see Merton, 1972). For the standard Fama-MacBeth approach, we fit the time-series regression over all T days. For the IV approach, we fit the time-series regressions separately with daily returns data from odd and even months.

We fit the second stage regression with monthly return data since this is the common practice in the literature. We could have fit the second stage regression with daily data as well, but there is no practical difference between fitting the regressions with monthly and daily returns. To see this intuitively, compare fitting one cross-sectional regression for month t with fitting 22 separate daily regressions for the month and averaging the daily regression estimates over the month. With the same set of firms in both regressions and same betas for the month, the slope coefficient of the monthly regression would be exactly 22 times the average slope coefficient of the daily regressions and the standard error of the monthly regression would also be 22 times the standard error of average daily regression coefficient. As a result, both specifications would yield exactly the same t-statistic for the slope coefficient. There would be some differences between the two specifications if daily returns are compounded to compute monthly returns but such differences are likely small.

We compute stock and factor returns for month t by aggregating the corresponding daily returns within the month. We then fit the following cross-sectional regression with monthly data

$$r_{i,t} = \gamma_{0,t} + \hat{\beta}_i \times \gamma_{1,t} + e_{i,t}. \tag{2.9}$$

where $\hat{\beta}_i$ is the estimate from regression (2.8). For the Fama-MacBeth approach, we estimate the OLS parameters of $\gamma_{0,t}$ and $\gamma_{1,t}$ each month. For the IV approach, we use equation (2.4) to estimate the parameters of $\gamma_{0,t}$ and $\gamma_{1,t}$ each month. We compute Fama-MacBeth standard errors for both approaches.

We carry out the three-factor model simulation experiments analogously, but in addition to market returns and market betas, we also generate risk factors and their sensitivities corresponding to the Fama-French SMB and HML factors. We match means and standard deviations for these parameters in the simulations to what we observe in the actual data. We then carry out the two-stage procedure to estimate $\gamma_0, \gamma_m, \gamma_{smb}$, and γ_{hml}. Appendix 2 presents the simulation parameters and the simulation experiment design in more detail.

3.3. EIV Bias

We run the simulation experiments with N = 1000 and 2000, and T=264, 528, 792, 1320, and 2640 days. For each simulation, we compare the true factor risk premiums used to generate data and the corresponding sample estimates. The average of this difference over the 1000 replications is the EIV-induced ex-ante bias. We also report the corresponding ex-ante root-mean-squared error (RMSE). In addition to the ex-ante bias and RMSE, we also investigate the ex-post bias and RMSE, which compare the sample means of simulated risk factors and the corresponding sample estimates.

Panel A of Table 11 presents the EIV-induced biases and RMSEs for the single factor model under the IV approach.[6] We report ex-ante and ex-post biases and RMSEs as percentages of the true parameter. For a given *T*, the odd and even month betas are estimated separately using *T*/2 observations. For both *N*=1000 and *N*=2000, ex-ante and ex-post biases are less than 1% for all *T*s and ex-ante and ex-post RMSEs are less than 10% for all *Ts* greater than or equal to 792 days.

To get a perspective on the EIV difficulty, Figure 2 presents the ex-ante and ex-post biases (Panel A) and RMSEs (Panel B) under the FM and IV estimators against the number of daily observations in the sample period with 2000 stocks. The vertical axis reports the biases and RMSEs as percentages of the true risk premium used in the simulations. In Panel A, the downward biases under the FM procedure are greater than 5% even for 2640 daily observations. The biases are less that 1% for the IV approach even with 264 daily observations, which is about 12 months of data. In Panel B, both ex-ante and ex-post RMSEs induced by EIV problem in the FM approach is substantially larger than those in the IV approach across all *T*s, especially, the ex-post RMSEs. For example, the ex-post RMSE is less than 10% for the IV approach even with 264 daily observations and it is less than 2% with 2640 daily observations. In contrast, the ex-post RMSE is about 7% for the FM approach with 2640 daily observations. The faster rate of convergence for the IV approach is consistent with the analytical results in Proposition 2.

Panel B of Table 11 presents the results for the three-factor model. The EIV problem always biases the slope coefficient estimates towards zero in univariate regressions, but the direction of the bias is in general indeterminate in multivariate regressions. As the table shows, the multivariate (three-factor) IV estimator bias is small for all sample sizes, and the ex-ante and ex-post biases are insignificant for all risk premiums estimated with at least 528 daily observations. The maximum bias is about 4% with 264 daily observations (when *N*=1000), which is approximately the number of trading days in a single year.

---

[6] The slope coefficient estimates using the FM procedure (unreported) are biased towards zero for all values of *N* and *T* due to the EIV problem.

3.4. Small Sample Distribution of the Test Statistic

We are primarily interested in using the IV estimator to test whether risk premiums associated with various factors are significantly different from zero. We propose to use the conventional *t*-statistic to test this hypothesis. We compute the *t*-statistic as follows:

$$t_\gamma = \frac{\hat{\gamma}}{\hat{\sigma}_\gamma} \tag{2.10}$$

where $\hat{\gamma}$ is the time-series average of monthly estimates of risk premiums and $\hat{\sigma}_\gamma$ is the corresponding Fama-MacBeth standard error.

This subsection examines the small sample distribution of the *t*-statistics in equation (2.10). We follow the same steps as in the last section to generate simulated data, but we set all true risk premiums equal to zero according to the null hypothesis. We then examine the percentage of simulations when the *t*-statistics are significant at the various levels (two-sided) using critical values based on the asymptotic normal distribution of the *t*-test statistic.

Table 12 presents the test sizes for the CAPM and the Fama-French three-factor model. We present the results for T=792 days, since we use a three-year rolling window for the first stage regression in our empirical applications. The results indicate that the tests are well specified. For example, in both single factor model and multifactor model, the test sizes for all factor risk premiums at the 5% level of significance are between 4.7% and 5.6% and those for 10% level of significance are between 9.6% and 10.4%, respectively. We found that the tests are well specified even when T=264 days, which corresponds to one year of daily data. Therefore, conventional *t*-test statistics based on the IV approach can be used to draw reliable inferences about the significance of risk premium estimates in finite samples.

4. Empirical Tests

4.1. Data

We obtain stock return and market capitalization data from the CRSP files and balance sheet data from COMPUSTAT during the sample period from January 1956 through December 2012. We exclude American depository receipts (ADRs), shares of beneficial interest, Americus Trust components, close-end funds, preferred stocks, and real estate investment trusts (REITs). We include only common stocks in our sample (CRSP share codes of 10 or 11). We also exclude stocks with prices below $1 and market capitalization less than $500,000 at the end of each month from the sample for the following month. We include all stocks that meet these criteria for which returns and book values are available. Since we employ daily returns data to estimate betas, we restrict the sample to stocks with at least 200 daily returns per year during the estimation period, i.e., the past three years.[7]

Table 13 presents the summary statistics of the stocks in the sample. There are a total of 7508 distinct stocks which enter the sample at different points in time. There are 1934 stocks per month in the sample on average.

4.2. CAPM and Fama-French Three-Factor Model

This section tests the CAPM and the Fama-French three-factor model. We first test whether the estimated factor risk premiums under the CAPM and the Fama-French three-factor models are different from zero using individual stocks as test assets. We also examine whether the risk premiums after controlling for stock characteristics are priced.

Early empirical tests of the CAPM by Fama and MacBeth (1973) and others find strong support for the CAPM. However several subsequent papers show that market betas are not priced after controlling for other factors. For instance, Jegadeesh (1992) and Fama and French (1993) find that after controlling for the size effect, the market risk premium is not significantly different from zero.

---

[7] We repeat the asset pricing tests with different thresholds for the number of daily observations per year, i.e., 100 and 150 daily observations per year, and find that our conclusions on the asset pricing tests do not change.

The inability of the CAPM to account for any of the cross-sectional differences in expected returns reinvigorated the search for alternative asset pricing models. The Fama-French three-factor model is perhaps the most widely used alternative. This model proposes size (i.e., SMB) and book-to-market (i.e., HML) factors as additional risk factors along with the market factor.

The empirical support for the Fama-French three-factor model is mixed. Fama and French (1992) estimate factor risk premiums using the portfolios sorted on size and book-to-market and show that both premiums are significantly positive. But the loadings on these risk factors are highly correlated with the size and book-to-market characteristics of the test portfolios. Therefore, as Lewellen *et al*. (2010) show, it is hard to reliably conclude that these risk premiums are indeed compensation for systematic risks rather than for portfolio characteristics because of the low dimensionality problems when portfolios are used as test assets.

The conflicting results of the empirical tests in Daniel and Titman (1997) and Davis, Fama, and French (2000) further illustrate the difficulty in making reliable inferences with portfolios as test assets. Daniel and Titman argue that the differences of average returns in size and book-to-market sorted portfolio returns are due to their characteristics. However, Davis, Fama, and French (2000) extend the sample period back to 1925 and find that in this extended sample period, the SMB and HML factor risks are priced and they argue that the differences in average returns across the test portfolios are due to factor risks and not due to their characteristics.

This subsection uses individual stocks in the tests and avoids the low dimensionality problems inherent in the tests that employ characteristics-sorted portfolios as test assets. We use rolling windows from month t-36 to month t-1 to estimate betas for month t. In untabulated tests, we found similar asset pricing test results when we used betas estimated within 60-, 24-, and 12-month rolling windows as well.

We use daily returns data to estimate betas and use one day lead and lag of the independent variables to adjust for non-synchronous trading effects (Dimson, 1979).

Specifically, we use the following regressions to estimate factor sensitivities under the CAPM:

$$r_{i,\tau} = a_i + \sum_{\kappa=-1}^{1} \beta_{i,m,\kappa} \times r_{m,\tau-\kappa} + u_{i,\tau}, \qquad (2.11)$$
$$\hat{\beta}_{i,m} = \hat{\beta}_{i,m,-1} + \hat{\beta}_{i,m,0} + \hat{\beta}_{i,m,1}.$$

We estimate odd- and even-month betas separately using returns on days belonging to odd and even months, respectively. Because we adjust for possible non-synchronous trading effects in regression (2.11), we exclude the first and the last days of each month to avoid any overlap.[8] We estimate betas for each month using returns data over the previous 36 months. We use an analogous multivariate regression to estimate the three factor sensitivities for the Fama-French three-factor model.

To get the characteristics for each stock for each month, we compute Size as the natural logarithm of market capitalization at the end of the previous month. BM is the book-to-market ratio, or the ratio of book value to market value. We compute book value as the sum of book value of equity plus deferred taxes and credits minus book value of preferred stock. We compute correlations between each pair of firm-specific variables each month and Table 14 presents the average cross-sectional correlations among factor sensitivities and characteristics. The CAPM beta estimated using the market model is negatively correlated with both Size and BM. In the Fama-French model, the correlation between market betas and the betas on the other two factors are positive. The correlation between Size and SMB factor sensitivities is negative, and the correlation between HML factor and BM is positive, which reflect the fact that the SMB and HML factors are constructed using these characteristics.

For direct comparison, Table 14 also presents the average cross-sectional correlations between each pair of portfolio-specific variables for 25 Fama-French size and book-to-market sorted portfolios. For each portfolio, we compute Size and BM each month as the

[8] We found almost identical results when we included the first and last days of each month. Also, the results were qualitatively similar to those we report when we set $\hat{\beta}_{i,m} = \hat{\beta}_{i,m,0}$.

value-weighted averages across all stocks in the portfolios. The correlation between SMB risk and Size is -.97 and that between HML risk and BM is .88 for these portfolios.

We estimate factor risk premiums using the IV methodology next. Table 15 presents the factor risk premium estimates for a number of different specifications of the second stage regressions. We first test the CAPM using betas estimated with the univariate regression (2.11). In column (1), the market risk premium estimate is −.189%, which is not reliably different from zero. Therefore, we do not find any support for the CAPM with individual returns.

We consider the Fama-French three-factor model next. We now estimate market betas and the other factor sensitivities using multivariate time-series regressions with all three factors. As a preliminary step, we estimate SMB and HML risk premiums in univariate specifications.[9] The SMB risk premium is .227% and not reliably different from zero at any conventional significance level, while the HML risk premium is .483% and is significant at the 1% level. Table 15 also reports factor risk premium estimates when we include all three factor sensitivities in the cross-sectional regressions. In column (4), the market risk premium estimate is now -.315% and the SMB and HML risk premiums are .311% and .504%, respectively. The SMB and HML risk premiums are now significant at the 5% level.

The statistical significance of the SMB and HML risk premiums in some of the above regressions may suggest that these risks are priced, but it is also possible that these estimates are merely a result of the correlation between factors sensitivities and the underlying characteristics. To examine this issue, our next set of tests includes Size and BM as additional independent variables in the second stage regressions. In column (5), the Size and BM coefficients in the regression which also includes market betas are -.152% and .163%, which are both statistically significant at the 1% level. The market risk premium estimate is .010%, which is not different from zero. In the regression that includes both SMB factor sensitivity and Size as independent variables (i.e., column (6)), SMB risk premium is -.025%, which is not statistically significant. In the regression that

[9] In untabulated results, we found that the market risk premium using market beta estimates from the Fama-French three-factor model specification was similar to the CAPM result in column (1) of Table 15.

includes HML factor sensitivity and book-to-market as independent variables (i.e., column (7)), the HML risk premium is .289%, which is still significant at the 5% level although its magnitude is substantially reduced compared to those in columns (3) and (4).

The final regression includes all factor sensitivities and characteristics as independent variables. In this regression (i.e., column (8)), the market risk premium is .113%, SMB risk premium is -.077%, and HML risk premium is .259%. All these estimates are not significantly different from zero at the 5% level. The point estimate of HML risk premium is the largest in this specification and it is marginally significant at the 10% level. The slope coefficients on Size and BM are highly significant at any conventional levels.

Table 15 also presents the results for two roughly equal subperiods. The factor risk premiums are not significant in most of the regressions that include characteristics in these subperiods as well. The only exception is column (7) of Panel B showing that the premium for HML risk in the first subperiod is .352% and marginally significant at the 5% level. However, this significance disappears when the other two factor sensitivities and Size are included in column (8). In contrast, the slope coefficients on Size and BM are generally significant in both subperiods.

Overall, the results indicate that factor risk premium estimates are significant in some specifications that have only factor sensitivities as independent variables, but become insignificant when the corresponding characteristics are also included in the regressions. Therefore, the variation in risk premium estimates across regression specifications seems to be driven by the statistical correlations across included and omitted variables rather than by economic phenomenon.

4.3. Production-based Asset Pricing Model

Several recent asset pricing models build on Merton's (1973) ICAPM to identify risk factors that are grounded in theory. Merton shows that when investment opportunity set varies over time, risks related to changing opportunity set will also be priced in addition to the market risk under the CAPM. Cochrane (1991) and Liu, Whited and Zhang (2009) present production-based asset pricing models in which productivity shocks are tied to the changes in the investment opportunity set. Since shocks to productivity are difficult to accurately measure, Chen, Novy-Marx and Zhang (2011) propose an investment factor and an ROE factor to capture productivity shocks. In this production-based asset pricing model, expected returns in excess of the risk-free rate ($r_{i,t}$) are given by:

$$E[r_{i,t}] = \beta_{M,i} \times \gamma_M + \beta_{I,i} \times \gamma_I + \beta_{ROE,i} \times \gamma_{ROE} \qquad (2.12)$$

where $\beta_{M,i}$, $\beta_{I,i}$, and $\beta_{ROE,i}$ are the betas with respect to market, investment and ROE factors, and γ_M, γ_I and γ_{ROE} are the corresponding risk premiums, respectively.

The investment factor captures the level of investments and the ROE factor captures the return on investments, i.e., profitability. The investment factor is constructed as the return difference between firms with low levels of investment and firms with high levels of investment and the ROE factor is constructed as the return difference between firms with high return on investment and firms with low return on investment. Intuitively, the levels of investments and rates of return on investments are likely to reflect the sensitivity to productivity shocks, and these factors are constructed to capture the price impact of the shocks. Chen, Novy-Marx and Zhang (2011) report that these factors better explain the cross-sectional return differences across portfolios constructed based on book-to-market, size, momentum, SUE, and net stock issues than the Fama-French three-factor model.

The Chen, Novy-Marx and Zhang (2011) model is appealing because it identifies risk factors based on an underlying theory rather than based on empirical anomalies. Also, their empirical tests use a variety of different portfolios and these test portfolios and the common factors are constructed using different characteristics. For instance, the test for the size and book-to-market effects uses 25 Fama-French size and book-to-market sorted portfolios, the test of the momentum effect uses 25 portfolios formed based on size and momentum, and the test for the SUE effect uses 10 SUE portfolios. However, these tests are still exposed to the low dimensionality problem because each test uses portfolios constructed using one or two anomalies.

This subsection tests the production-based asset pricing model using individual stocks as test assets. We use the same procedure as in Chen, Novy-Marx and Zhang (2011) to construct daily investment and ROE factors. We independently sort firms based on firm sizes, investments to total assets and ROEs. We classify stocks into three categories based on each of these variables, where we assign the top 30% and bottom 30% of the stocks to the high and low categories and the middle 40% to the medium category. We form 27 value-weighted portfolios with stocks in the intersections of these categories. The investment factor is long the equally weighted portfolio of the nine low investments to total assets portfolios and short the nine high investments to total assets portfolios. We construct the ROE factor as the difference between the high ROE and low ROE portfolios. We use the last announced quarterly balance sheet data to compute the level of investments and ROE each month.[10] Since earnings announcement dates are available on COMPUSTAT only after 1972, the sample period for the tests in this subsection is from 1972 to 2012.

Table 16 reports the average cross-sectional correlations among factor sensitivities and firm characteristics. The sensitivities to investment and ROE factors are positively correlated across stocks. Investment factor sensitivity is negatively correlated with Size and positively correlated with BM, and the ROE factor sensitivity is positively correlated with Size and negatively correlated with BM. The correlations between these factor sensitivities and the characteristics are smaller than those for the SMB and HML factors in Table 14. Table 16 also presents these correlations for the 25 Fama-French size and book-to-market sorted portfolios. For these portfolios, the correlation between investment factor sensitivity and BM is .88 and the correlation between ROE factor sensitivity and Size is .74. These high correlations suggest that the low dimensionality issues discussed in Lewellen *et al.* (2010) could affect the results of the tests that use the 25 Fama-French portfolios.

[10] Following Chen, Novy-Marx and Zhang (2011), the investment to total assets is defined as the annual change in gross property, plant, and equipment plus the annual change in inventories divided by the book value of assets lagged by one quarter. ROE is defined as income before extraordinary items divided by book equity lagged by one quarter.

Table 17 presents the results of the asset pricing tests with individual stocks. When we consider each of the factor sensitivities individually, we find that the result for market risk premium is similar to what we found in Table 15. In column (2), investment risk premium is .375% and marginally significant at the 5% level. In column (3), ROE risk premium is negative and insignificant at any conventional level. When we include all three factor sensitivities simultaneously in the second stage regression (i.e., column (4)), none of the factor risks have significant slope coefficients. In comparison, the mean of the ROE factor during our sample period is .7% per month, which is significantly positive. So if the ROE factor reflected risk, the risk premium on this factor should be positive as well. Table 17 also reports the regression estimates when we include Size and BM in the regressions (columns (5) to (7)). Once again, we find that none of the investment and ROE risk factors are priced when characteristics are included. Table 17 also presents the results for the two subperiods and these factors are priced in neither of them when Size and BM are included in regressions.

4.4. Liquidity-adjusted CAPM

This subsection tests the liquidity-adjusted capital asset pricing model (LCAPM) proposed by Acharya and Pedersen (2005). In contrast to the models such as Chen, Novy-Marx and Zhang (2011), Campbell and Vuolteenaho (2004), and others that identify risk factors based on specialized variations of the ICAPM, the LCAPM models the effect of illiquidity-based trading frictions on asset pricing. According to the LCAPM, the level of illiquidity and the covariances of return and illiquidity innovation with the market return and illiquidity innovation vary across assets. The unconditional expected return in excess of the risk-free rate ($r_{i,t}$) under the LCAPM is:

$$E[r_{i,t}] = E[c_{i,t}] + \lambda(\beta_{1,i} + \beta_{2,i} - \beta_{3,i} - \beta_{4,i}) \,, \tag{2.13}$$

where $c_{i,t}$ is the illiquidity cost, the risk premium is the market excess return minus aggregate illiquidity cost (i.e. $\lambda = E[r_{m,t} - c_{m,t}]$), and the betas are

$$\beta_{1,i} = \frac{Cov(r_{i,t},\ r_{m,t} - E_{t-1}(r_{m,t}))}{Var(r_{m,t} - E_{t-1}(r_{m,t}) - [c_{m,t} - E_{t-1}(c_{m,t})])}, \tag{2.14}$$

$$\beta_{2,i} = \frac{Cov(c_{i,t} - E_{t-1}(c_{i,t}),\ c_{m,t} - E_{t-1}(c_{m,t}))}{Var(r_{m,t} - E_{t-1}(r_{m,t}) - [c_{m,t} - E_{t-1}(c_{m,t})])},$$

$$\beta_{3,i} = \frac{Cov(r_{i,t},\ c_{m,t} - E_{t-1}(c_{m,t}))}{Var(r_{m,t} - E_{t-1}(r_{m,t}) - [c_{m,t} - E_{t-1}(c_{m,t})])},$$

$$\beta_{4,i} = \frac{Cov(c_{i,t} - E_{t-1}(c_{i,t}),\ r_{m,t} - E_{t-1}(r_{m,t}))}{Var(r_{m,t} - E_{t-1}(r_{m,t}) - [c_{m,t} - E_{t-1}(c_{m,t})])}.$$

The term $E[c_{i,t}]$ is the reward for firm-specific illiquidity level, which is the compensation for holding an illiquid asset as in Amihud and Mendelson (1986). Acharya and Pederson define illiquidity-adjusted net beta as:

$$\beta_{LMKT,i} = \beta_{1,i} + \beta_{2,i} - \beta_{3,i} - \beta_{4,i}\,. \tag{2.15}$$

The LCAPM implies that the linear relation between risk and return applies for liquidity-adjusted beta but not for the standard CAPM beta. The LCAPM also implies that the linearity between risk and return applies to excess return net of firm-specific illiquidity cost.

Acharya and Pedersen (2005) test the LCAPM using two sets of test portfolios formed based on illiquidity and standard deviation of illiquidity. They sort stocks based on Amihud (2002) illiquidity measures during each year and form 25 value-weighted illiquidity test portfolios for the subsequent year. They also form 25 σ(illiquidity) portfolios similarly by sorting based on the standard deviation of illiquidity.

We examined the correlations between β_{LMKT} and the value-weighted average of Size and BM for these portfolios. We found that correlations of β_{LMKT} with Size for illiquidity and σ(illiquidity) portfolios are -.96 and -.97, and those with BM are .71 and .74, respectively. These particularly high correlations between liquidity-adjusted betas and Size suggest that it would be hard to empirically differentiate whether differences in returns across test portfolios are due to Size or illiquidity-adjusted betas. This situation parallels that in Chan and Chen (1988) who use 20 size-sorted portfolios as test assets and find strong support for the CAPM. The correlations between betas and Size for Chan and Chen's test portfolios range from -.988 to -.909 over different periods, and the corresponding correlations in the case of illiquidity and σ(illiquidity) portfolios are within this range. Jegadeesh (1992) shows that when test portfolios are constructed so that size and beta have low correlations, the market risk is not priced and that the significant market risk premium found using size-sorted portfolios is due to the high correlation between size and beta.

We investigate whether the LCAPM beta is priced using the IV estimator with individual stocks. To facilitate comparability, we follow the same approach as in Acharya and Pederson (2005) in all other respects. Because of the differences in the market structures of the NYSE/AMEX and NASDAQ, the trading volumes reported in these markets are not comparable and hence we exclude NASDAQ stocks for this subsection. Also, we exclude any stock that does not trade for at least 200 days over the previous three years.

Acharya and Pederson define illiquidity cost as follows:[11]

$$ILLIQ_{i,\tau} = \frac{|r_{i,\tau}|}{v_{i,\tau}}, \tag{2.16}$$

$$c_{i,\tau} = \min\left(0.25 + 0.3\, ILLIQ_{i,\tau}\, P_{m,\tau-1}, 30\right), \tag{2.17}$$

where $r_{i,\tau}$ is the return on day τ, $v_{i,\tau}$ is the dollar volume (in millions) and $P_{m,\tau-1}$ is the day $\tau-1$ value of $1 invested in the market portfolio as of the end of July 1962. Equation (2.16) is based on Amihud's (2002) illiquidity measure. Acharya and Pederson use equation (2.17) as a measure of illiquidity costs where $P_{m,\tau-1}$ is used to adjust for inflation and the illiquidity cost is capped at 30%. Market illiquidity cost $c_{m,\tau}$ is the value-weighted average of the illiquidity costs of the individual stocks.

As in Acharya and Pederson (2005), we estimate innovations in illiquidity using an AR model and we then estimate each individual component of betas in equation (2.14) using a time-series GMM approach.[12] We then fit the following cross-sectional regression each month:

$$r_{i,t} = \alpha_t + \gamma_{ILLIQ,t} \times c_{i,t} + \gamma_{LMKT,t} \times \hat{\beta}_{LMKT,i} + \epsilon_{i,t}. \tag{2.18}$$

where $c_{i,t}$ is the average illiquidity for stock *i* in month *t*.

The IV estimator for month *t* is:

---

[11] Acharya and Pederson (2005) use illiquidity costs at monthly frequency but we use them at daily frequency.

[12] Appendix 3 presents the AR model we use to estimate expected and unexpected components of illiquidity.

$$\hat{\Gamma}_{even,t} = \left(\hat{\Psi}'_{odd,t}\ \hat{\Psi}_{even,t}\right)^{-1}\hat{\Psi}'_{odd,t}R_{even,t},$$

$$\hat{\Gamma}_{odd,t} = \left(\hat{\Psi}'_{even,t}\ \hat{\Psi}_{odd,t}\right)^{-1}\hat{\Psi}'_{even,t}R_{odd,t},$$

where:

$\hat{\Psi}_{even,t}$ ≡ *N*x3 matrix of independent variables with unit vector as the first column, $c_{i,t}$, and estimated even-month LMKT betas for N individual stocks as the second and third columns, respectively. We estimate the even-month LMKT betas using daily data in even months in the period of month *t*-36 to month *t*-1.

$\hat{\Psi}_{odd,t}$ ≡ Analogous to $\hat{\Psi}_{even,t}$, estimated using all daily data in odd months

We use the Fama-MacBeth approach to compute the point estimates and standards errors of risk premium estimates.

Table 18 presents the regression estimates with individual stocks. The slope coefficient on Amihud illiquidity measure is .184%, which is significantly positive at the 1% level. However, the liquidity-adjusted market risk premium estimates (the slope coefficients on β_{LMKT}) are .140% and .075%, which are not reliably different from zero. These results indicate that firm-specific illiquidity level, which is a firm characteristic, is positively related to returns, but a stock's liquidity-adjusted market beta, which is systematic risk, does not earn a risk premium. We also find that this risk is not priced in either of the subperiods.

In comparison, Acharya and Pederson (2005) report liquidity-adjusted market risk premium estimates of about 2.5% per month using the value-weighted index (see Panel B of Table 5 in Acharya and Pedersen), which is about 30% per year.[13] The equity risk premium puzzle literature argues that even an annual risk premium of about 6% observed in the data is hard to justify with realistic levels of risk aversion, and larger risk premiums would be harder to justify. The large estimates obtained using portfolios as test assets seem likely to be a result of the correlation between β_{LMKT} and portfolios characteristics rather than a true depiction of rewards to risk.

---

[13] The liquidity-adjusted market risk premium equals market risk premium minus expected illiquidity costs and hence it is smaller than the unadjusted market risk premium.

Our findings further illustrate the problems that arise when portfolios are used as test assets. In the earlier size versus beta debate, portfolios were formed based on size ranks and hence it would be fairly natural to check the correlation between size and beta and discover the problem. In the case of illiquidity-sorted portfolios, size was not explicitly used as a sorting variable and hence it is not readily apparent that one should check the correlation with this variable, but such correlations could lead to mistaken inferences. Our tests with individual stocks avoid such confounding issues.

4.5. Strength of instruments

An important issue to consider in instrumental variable regressions is the correlation between the instrumental variables and the corresponding independent variables. The cross-product of the vector of instrumental variables and independent variables could be close to singularity if the correlation is low. Nelson and Startz (1990) show that if the instruments are sufficiently weak then the expected value of the IV estimator may not exist. The intuition behind this result can be seen in a univariate regression with weak instruments. If the covariance between the independent variable and the instrument is close to zero then the sample covariance could be small and be either negative or positive, resulting in large variations in both the sign and magnitude of the slope coefficient estimates in finite samples. However, if the covariance and the sample size are sufficiently large, then the likelihood that the sample estimate of the covariance is close to zero becomes negligibly small, and the IV estimator is well behaved.

Nelson and Stratz show that weak instruments would be a concern if

$$\frac{1}{\hat{\rho}_{xz}^2} \gg N, \qquad (2.19)$$

where $\hat{\rho}_{xz}$ is the correlation between the independent variable and the corresponding instrument (which in our context is the correlation between odd- and even-month betas) and N is the number of observations in the cross-sectional regression. There are 1934 stocks per month in our sample, and the minimum number of stocks is 305. The critical value from equation (2.19) based on the minimum (average) number of stocks is 0.057 (0.023) in absolute value.

Table 19 presents the average correlation between the odd and even month factor sensitivity estimates. The correlation for market beta under the CAPM is .67. The market beta of Fama-French three-factor model is less precisely estimated and the correlation is smaller at .52. The market beta in the production-based asset pricing model and the LCAPM betas also exhibit similar levels of correlation as the three-factor market betas. The average correlations for SMB, HML, INV, and ROE betas range from .26 to .44. Although these correlations are smaller than those for market betas, they are all comfortably above the Nelson and Startz (1990) critical value.

Nelson and Startz (1990) and Staiger and Stock (1997) also show that the conventional IV standard error estimator based on asymptotic theory will not be reliable in small samples if the instruments are weak. However, this concern is not relevant in our application because we use the Fama-MacBeth approach to estimate standard errors and we do not use the analytic estimator derived using asymptotic theory. Nevertheless, we further examined the strength of the instruments using tests proposed by Staiger and Stock (1997) as well and in untabulated results we found that the tests statistics were all well above the critical values.[14]

To provide further insights into the strength of the instruments, we also estimate the correlation between the instruments that we use and the corresponding true but unobservable factor sensitivities. Although the true factor sensitivity is unobservable, we can estimate this correlation based on the correlation between the odd- and even-month betas as we show in the proposition below:

Proposition 3*: Let $\beta_{i,k}$ be stock i's true unobservable sensitivity to factor k and let $\hat{\beta}_{i,k}^{odd}$ and $\hat{\beta}_{i,k}^{even}$ be the odd and even month estimates of the corresponding factor sensitivities, respectively. Then:*

[14] Staiger and Stock (1997) regress the independent variable against the instrumental variables and develop a test based on the goodness of fit of this regression. In untabulated results, we found that the test statistics in our applications were well above the critical values for all instruments for all months.

$$Correlation\left(\beta_{i,k}, \hat{\beta}_{i,k}^{even}\right) = Correlation\left(\beta_{i,k}, \hat{\beta}_{i,k}^{odd}\right) = \sqrt{Correlation\left(\hat{\beta}_{i,k}^{odd}, \hat{\beta}_{i,k}^{even}\right)} \quad (2.20)$$

Proof: See Appendix 4.

Table 19 presents the mean correlation between estimated factor sensitivities and true factor sensitivities.[15] The average correlation between even- and odd-month market betas is .67 and the average correlation between estimated market beta and the unobserved true market beta is .82. We find smaller correlations for SMB and HML betas, but even for HML the average correlation between estimated beta and unobservable true beta is .54. The correlations for the investment and ROE factor sensitivities are about the same as that for the HML factor sensitivity and the correlations are bigger for other factors sensitivities. All these estimates are significantly above the cutoff prescribed by Nelson and Startz (1990).

5. Conclusion

Empirical tests of asset pricing models typically use portfolios rather than individual stocks as test assets to mitigate the errors-in-variables problems. This problem arises because the sensitivities to risk factors specified by the asset pricing models are estimated from the data and they contain sampling errors. Since factor sensitivities for portfolios are estimated more precisely than for individual stocks, the factor risk premium estimates in the second stage regressions will be less biased due to the errors-in-variables problems if one were to use portfolios as test assets rather than individual stocks. However, a problem with using test portfolios is that they limit the number of dimensions

[15] To compute the mean correlation between estimated factor sensitivities and true factor sensitivities, we first compute the square root of the correlation between odd- and even-month factor sensitivities each month and then compute the average across months. Because the variability of correlation between odd- and even-month factor sensitivities is relatively small, the square root of average correlation is about the same as the mean of the square root of the correlation.

along which expected returns vary. As a result even factors that are not important from an economic perspective may mistakenly appear to command a risk premium simply because of their correlations with the variables used to construct the test portfolios.

This paper develops an instrumental variables methodology to obtain consistent estimates of risk premiums using individual stocks as test assets. This approach overcomes the low dimensionality problem associated with using portfolios as test assets and it also removes the subjectivity associated with the choice of test portfolios. Our simulation evidence indicates that this methodology yields nearly unbiased estimates in finite samples and that the associated tests are well specified even in small samples.

We empirically test several asset pricing models with individual stocks using the instrumental variables methodology. The models that we test are the CAPM, the Fama-French three-factor model, a production-based asset pricing model proposed by Chen, Novy-Marx and Zhang (2011), and the liquidity-adjusted CAPM proposed by Acharya and Pedersen (2005). Earlier empirical tests in the literature using portfolios as tests assets find support for these models, but as Lewellen, Nagel and Shanken (2010) caution, these results could be misleading because of the low dimensionality problems. Contrary to these supportive results, we find that none of the factor risks in these models command a risk premium.

Technical Appendix 1: Proofs of Propositions

In this appendix, we provide the detailed mathematical proofs for Propositions 1 and 2 presented in the main body of the paper. The following data generating process for individual stock return $r_{i,t}$ is assumed: for stock i in month t,

$$r_{i,t} = \gamma_0 + \beta_i^{'} \left(f_t - E[f_t] + \gamma_1\right) + \varepsilon_{i,t}, \tag{A.1}$$

where γ_0 and γ_1 are the risk-free rate and factor risk premium, respectively. f_t is a $K \times 1$ vector containing risk factors, β_i ($K \times 1$ column vector) consists of the associated factor loadings, and $E[\varepsilon_{i,t}] = 0$. We assume the balanced panel defined by the number of stocks N and the time-series length T, where T is assumed an even number. For simplicity, we posit that $T = 2T_e = 2T_o$, where T_e (T_o) is the total number of even (odd) months. To investigate the rate of convergence of the IV estimator, we first let N go to ∞ and then let T go to ∞.

List of Technical Assumptions:

TA.1. Laws of Large Number for Cross-section

As N increases without bound, we assume the following probability limits:

$$\frac{1}{N}\sum_i \beta_i \rightarrow \mu_\beta, \qquad \frac{1}{N}\sum_i \varepsilon_{i,t} \rightarrow 0 \quad \text{for all } t, \tag{A.2}$$

$$\frac{1}{N}\sum_i \beta_i\beta_i^{'} \rightarrow \mu_\beta\mu_\beta^{'} + \Sigma_\beta, \qquad \frac{1}{N}\sum_i \varepsilon_{i,t}\beta_i \rightarrow \mathbf{0}_K \quad \text{for all } t, \tag{A.3}$$

$$\frac{1}{N}\sum_i \varepsilon_{i,t}\varepsilon_{i,s} \rightarrow 0 \quad \text{for all } (t,s) \text{ with } t \neq s, \tag{A.4}$$

where the cross-sectional summations are taken over $i = 1, \ldots, N$. $\mathbf{0}_K$ is a $K \times 1$ zero vector.

TA.2. Laws of Large Number for Time

As T increases without bound, we assume the following probability limits:

$$\text{Even months:} \qquad \frac{1}{T_e}\sum_{t_e}(f_{t_e}-\bar{f}_{T_e})(f_{t_e}-\bar{f}_{T_e})' \to \Sigma_f, \tag{A.5}$$

$$\frac{1}{T_e}\sum_{t_e} f_{t_e} = \bar{f}_{T_e} \to E[f_t], \tag{A.6}$$

$$\text{Odd months:} \qquad \frac{1}{T_o}\sum_{t_o}(f_{t_o}-\bar{f}_{T_o})(f_{t_o}-\bar{f}_{T_o})' \to \Sigma_f, \tag{A.7}$$

$$\frac{1}{T_o}\sum_{t_o} f_{t_o} = \bar{f}_{T_o} \to E[f_t], \tag{A.8}$$

where t_e and t_o denote the indices for even and odd months, respectively. $\bar{f}_{T_e}$ ($\bar{f}_{T_o}$) are the sample mean of f_t taken over even (odd) months, respectively.

Beta Estimates from Even- and Odd-months (Time-series Regressions): We run time-series regressions to estimate even- and odd-month betas separately. The betas estimated from even months for stocks i are given by

$$\begin{aligned}
\widehat{\beta}_{i,T_e} &= \beta_i + \underbrace{\widehat{\Sigma}_{f,T_e}^{-1}\left[\frac{1}{T_e}\sum_{t_e}\left(\varepsilon_{i,t_e}-\bar{\varepsilon}_{i,T_e}\right)\left(f_{t_e}-\bar{f}_{T_e}\right)\right]}_{=\xi_{i,T_e}}, \qquad \text{(A.9)}\\
\widehat{\Sigma}_{f,T_e} &= \frac{1}{T_e}\sum_{t_e}\left(f_{t_e}-\bar{f}_{T_e}\right)\left(f_{t_e}-\bar{f}_{T_e}\right)',
\end{aligned}$$

where the subscript T_e denotes that the associated variable is computed based on even months. The measurement errors are written as

$$\xi_{i,T_e} = \widehat{\Sigma}_{f,T_e}^{-1}\widehat{\Sigma}_{f\varepsilon,T_e}^{i}. \tag{A.10}$$

In the same way, the betas estimated from odd months and their measurement errors for stocks i are given by

$$\widehat{\beta}_{i,T_o} = \beta_i + \underbrace{\widehat{\Sigma}^{-1}_{f,T_o}\left[\frac{1}{T_o}\sum_{t_o}\left(\varepsilon_{i,t_o} - \bar{\varepsilon}_{i,T_o}\right)\left(f_{t_o} - \bar{f}_{T_o}\right)\right]}_{=\xi_{i,T_o}}, \tag{A.11}$$

$$\widehat{\Sigma}_{f,T_o} = \frac{1}{T_o}\sum_{t_o}\left(f_{t_o} - \bar{f}_{T_o}\right)\left(f_{t_o} - \bar{f}_{T_o}\right)',$$

$$\xi_{i,T_o} = \widehat{\Sigma}^{-1}_{f,T_o}\widehat{\Sigma}^{i}_{f\varepsilon,T_o}.$$

These expressions for beta estimates from even and odd months will be important building blocks for the subsequent analyses. Thus we here collect them as follows:

$$\widehat{\beta}_{i,T_e} = \beta_i + \widehat{\Sigma}^{-1}_{f,T_e}\widehat{\Sigma}^{i}_{f\varepsilon,T_e}, \qquad \widehat{\beta}_{i,T_o} = \beta_i + \widehat{\Sigma}^{-1}_{f,T_o}\widehat{\Sigma}^{i}_{f\varepsilon,T_o}. \tag{A.12}$$

Cross-sectional Regressions: the IV Estimator

CSR.1. Probability Limit of Bottom Term of the IV Estimator

We define the following $N \times (1+K)$ matrices:

$$\widehat{\mathbf{X}}_{T_e} = \begin{bmatrix} \mathbf{1}_N & \widehat{\mathbf{B}}_{T_e} \end{bmatrix}, \qquad \widehat{\mathbf{X}}_{T_o} = \begin{bmatrix} \mathbf{1}_N & \widehat{\mathbf{B}}_{T_o} \end{bmatrix}, \tag{A.13}$$

where, e.g. $\widehat{\mathbf{B}}_{T_e}$ is composed of stacked $\widehat{\beta}'_{i,T_e}$ over all stocks. For odd months, in which even-month betas are employed as the instruments for odd-month betas, the bottom term of the IV Estimator is defined as follows:

$$\frac{1}{N}\widehat{\mathbf{X}}'_{T_e}\widehat{\mathbf{X}}_{T_o} = \begin{bmatrix} 1 & \frac{1}{N}\sum_i \widehat{\beta}'_{i,T_o} \\ & \\ \frac{1}{N}\sum_i \widehat{\beta}_{i,T_e} & \frac{1}{N}\sum_i \widehat{\beta}_{i,T_e}\widehat{\beta}'_{i,T_o} \end{bmatrix}, \tag{A.14}$$

We deal with each sub-block matrix one by one.

$$\begin{aligned}\frac{1}{N}\sum_i \widehat{\beta}_{i,T_e} &= \frac{1}{N}\left[\sum_i \beta_i + \widehat{\Sigma}^{-1}_{f,T_e}\sum_i \widehat{\Sigma}^i_{f\varepsilon,T_e}\right] \\ &\to \mu_\beta \qquad \text{as } N \text{ goes to } \infty,\end{aligned} \tag{A.15}$$

due to the technical assumptions. In the same way, it is straightforward to show that

$$\frac{1}{N}\sum_i \widehat{\beta}'_{i,T_o} \to \mu_\beta' \qquad \text{as } N \text{ goes to } \infty. \tag{A.16}$$

Now we investigate the sub-block matrix in the right and bottom corner, i.e. $\frac{1}{N}\sum_i \widehat{\beta}_{i,T_e}\widehat{\beta}'_{i,T_o}$.

$$\begin{aligned}\frac{1}{N}\sum_i \widehat{\beta}_{i,T_e}\widehat{\beta}'_{i,T_o} &= \frac{1}{N}\left[\sum_i \beta_i\beta_i' + \widehat{\Sigma}^{-1}_{f,T_e}\widehat{\Sigma}^i_{f\varepsilon,T_e}\beta_i' + \beta_i\widehat{\Sigma}^{i'}_{f\varepsilon,T_o}\widehat{\Sigma}^{-1}_{f,T_o} + \widehat{\Sigma}^{-1}_{f,T_e}\widehat{\Sigma}^i_{f\varepsilon,T_e}\widehat{\Sigma}^{i'}_{f\varepsilon,T_o}\widehat{\Sigma}^{-1}_{f,T_o}\right] \\ &= \frac{1}{N}\sum_i \beta_i\beta_i' + \widehat{\Sigma}^{-1}_{f,T_e}\left(\frac{1}{N}\sum_i \widehat{\Sigma}^i_{f\varepsilon,T_e}\beta_i'\right) + \left(\frac{1}{N}\sum_i \beta_i\widehat{\Sigma}^{i'}_{f\varepsilon,T_o}\right)\widehat{\Sigma}^{-1}_{f,T_o} \\ &+ \widehat{\Sigma}^{-1}_{f,T_e}\left(\frac{1}{N}\sum_i \widehat{\Sigma}^i_{f\varepsilon,T_e}\widehat{\Sigma}^{i'}_{f\varepsilon,T_o}\right)\widehat{\Sigma}^{-1}_{f,T_o} \qquad \text{(A.17)} \\ &\to \mu_\beta\mu_\beta' + \Sigma_\beta \qquad \text{as } N \text{ goes to } \infty, \qquad \text{(A.18)}\end{aligned}$$

due to the technical assumptions.

Now combining all intermediate results for the bottom term of the IV estimator shown above, we have the following probability limit:

$$\left(\frac{\widehat{\mathbf{X}}'_{T_e}\widehat{\mathbf{X}}_{T_o}}{N}\right)^{-1} \to \begin{bmatrix} 1 & \mu_\beta' \\ & \\ \mu_\beta & \mu_\beta\mu_\beta' + \Sigma_\beta \end{bmatrix}^{-1} \tag{A.19}$$

as N goes to ∞.

For even months, in which odd-month betas are employed as the instruments for even-month betas, the bottom term of the IV estimator is defined as $\frac{1}{N}\widehat{\mathbf{X}}_{T_o}^{'}\widehat{\mathbf{X}}_{T_e}$. By following the same procedure as above, we can show that as N goes to ∞,

$$\left(\frac{\widehat{\mathbf{X}}_{T_o}^{'}\widehat{\mathbf{X}}_{T_e}}{N}\right)^{-1} \rightarrow \begin{bmatrix} 1 & \mu_\beta^{'} \\ & \\ & \\ \mu_\beta & \mu_\beta\mu_\beta^{'} + \Sigma_\beta \end{bmatrix}^{-1} \tag{A.20}$$

CSR.2. Probability Limit of Top Term of the IV Estimator

We now consider the top term of the IV estimator, which is defined as follows: For odd months, in which even-month betas are employed as the instruments for odd-month betas,

$$\frac{1}{N}\widehat{\mathbf{X}}_{T_e}^{'}\bar{\mathbf{R}}_{T_o} \quad = \quad \begin{bmatrix} \frac{1}{N}\sum_i \bar{r}_{i,T_o} \\ \\ \frac{1}{N}\sum_i \widehat{\beta}_{i,T_e}\bar{r}_{i,T_o} \end{bmatrix}, \tag{A.21}$$

where $\bar{r}_{i,T_o}$ is the sample mean of $r_{i,t}$ taken over odd months for stock i. From equations (A.1), we have the following expression for average return:

$$\bar{r}_{i,T_o} \quad = \quad \gamma_0 + \beta_i^{'}\left(\bar{f}_{T_o} - E[f_t] + \gamma_1\right) + \bar{\varepsilon}_{i,T_o}, \tag{A.22}$$

implying that due to the technical assumptions,

$$\frac{1}{N}\sum_i \bar{r}_{i,T_o} \quad \rightarrow \quad \gamma_0 + \mu_\beta^{'}\left(\bar{f}_{T_o} - E[f_t] + \gamma_1\right) \tag{A.23}$$

as N goes to ∞.

We now consider the term in the bottom of equation (A.21), i.e. $\frac{1}{N}\sum_i \widehat{\beta}_{i,T_e}\bar{r}_{i,T_o}$.

$$\begin{aligned} \frac{1}{N}\sum_i \widehat{\beta}_{i,T_e}\bar{r}_{i,T_o} &= \gamma_0\left(\frac{1}{N}\sum_i \beta_i\right) + \left(\frac{1}{N}\sum_i \beta_i\beta_i'\right)\left(\bar{f}_{T_o} - E[f_t] + \gamma_1\right) + \ldots, \\ &\to \gamma_0\mu_\beta + \left(\mu_\beta\mu_\beta' + \Sigma_\beta\right)\left(\bar{f}_{T_o} - E[f_t] + \gamma_1\right), \end{aligned} \tag{A.24}$$

as N goes to ∞, where the extra terms vanish as N goes to ∞ due to the technical assumptions.

Collecting all these probability limits produces the following probability limit:

$$\frac{1}{N}\widehat{\mathbf{X}}_{T_e}'\bar{\mathbf{R}}_{T_o} \to \begin{bmatrix} \gamma_0 + \mu_\beta'\left(\bar{f}_{T_o} - E[f_t] + \gamma_1\right) \\ \\ \gamma_0\mu_\beta + \left(\mu_\beta\mu_\beta' + \Sigma_\beta\right)\left(\bar{f}_{T_o} - E[f_t] + \gamma_1\right) \end{bmatrix} \tag{A.25}$$

as N increases without bound.

For even months, in which odd-month betas are employed as the instruments for even-month betas, by following the same procedure as above, it is straightforward to show that

$$\frac{1}{N}\widehat{\mathbf{X}}_{T_o}'\bar{\mathbf{R}}_{T_e} \to \begin{bmatrix} \gamma_0 + \mu_\beta'\left(\bar{f}_{T_e} - E[f_t] + \gamma_1\right) \\ \\ \gamma_0\mu_\beta + \left(\mu_\beta\mu_\beta' + \Sigma_\beta\right)\left(\bar{f}_{T_e} - E[f_t] + \gamma_1\right) \end{bmatrix} \tag{A.26}$$

as N increases without bound.

CSR.3. Combining the Probability Limits of Top and Bottom Terms of the IV Estimator:

For odd months, we have

$$\left(\widehat{\mathbf{X}}_{T_e}'\widehat{\mathbf{X}}_{T_o}\right)^{-1}\widehat{\mathbf{X}}_{T_e}'\bar{\mathbf{R}}_{T_o} \rightarrow \begin{bmatrix} 1 & \mu_\beta' \\ & \\ & \\ \mu_\beta & \mu_\beta\mu_\beta' + \Sigma_\beta \end{bmatrix}^{-1} \begin{bmatrix} \gamma_0 + \mu_\beta'\left(\bar{f}_{T_o} - E[f_t] + \gamma_1\right) \\ \\ \gamma_0\mu_\beta + \left(\mu_\beta\mu_\beta' + \Sigma_\beta\right)\left(\bar{f}_{T_o} - E[f_t] + \gamma_1\right) \end{bmatrix} \tag{A.27}$$

as N goes to ∞. For even months, we have

$$\left(\widehat{\mathbf{X}}_{T_o}'\widehat{\mathbf{X}}_{T_e}\right)^{-1}\widehat{\mathbf{X}}_{T_o}'\bar{\mathbf{R}}_{T_e} \rightarrow \begin{bmatrix} 1 & \mu_\beta' \\ & \\ & \\ \mu_\beta & \mu_\beta\mu_\beta' + \Sigma_\beta \end{bmatrix}^{-1} \begin{bmatrix} \gamma_0 + \mu_\beta'\left(\bar{f}_{T_e} - E[f_t] + \gamma_1\right) \\ \\ \gamma_0\mu_\beta + \left(\mu_\beta\mu_\beta' + \Sigma_\beta\right)\left(\bar{f}_{T_e} - E[f_t] + \gamma_1\right) \end{bmatrix} \tag{A.28}$$

as N goes to ∞. The IV estimate of risk premiums is then defined as

$$\widehat{\Gamma}_{IV} = \frac{1}{2}\left(\left(\widehat{\mathbf{X}}_{T_e}'\widehat{\mathbf{X}}_{T_o}\right)^{-1}\widehat{\mathbf{X}}_{T_e}'\bar{\mathbf{R}}_{T_o} + \left(\widehat{\mathbf{X}}_{T_o}'\widehat{\mathbf{X}}_{T_e}\right)^{-1}\widehat{\mathbf{X}}_{T_o}'\bar{\mathbf{R}}_{T_e}\right). \tag{A.29}$$

N-Consistency of the IV Estimator (Corollary of Proposition 1): For a fixed T, as N goes to ∞,

$$\begin{aligned}
\widehat{\Gamma}_{IV} &\to \begin{bmatrix} 1 & \mu_\beta^{'} \\ \mu_\beta & \mu_\beta\mu_\beta^{'} + \Sigma_\beta \end{bmatrix}^{-1} \begin{bmatrix} \gamma_0 + \mu_\beta^{'}\left(\bar{f}_T - E[f_t] + \gamma_1\right) \\ \gamma_0\mu_\beta + \left(\mu_\beta\mu_\beta^{'} + \Sigma_\beta\right)\left(\bar{f}_T - E[f_t] + \gamma_1\right) \end{bmatrix} \qquad \text{(A.30)} \\
&= \begin{bmatrix} 1 + \mu_\beta^{'}\Sigma_\beta^{-1}\mu_\beta & -\mu_\beta^{'}\Sigma_\beta^{-1} \\ -\Sigma_\beta^{-1}\mu_\beta & \Sigma_\beta^{-1} \end{bmatrix} \begin{bmatrix} \gamma_0 + \mu_\beta^{'}\left(\bar{f}_T - E[f_t] + \gamma_1\right) \\ \gamma_0\mu_\beta + \left(\mu_\beta\mu_\beta^{'} + \Sigma_\beta\right)\left(\bar{f}_T - E[f_t] + \gamma_1\right) \end{bmatrix} \\
&= \begin{bmatrix} \gamma_0 \\ \bar{f}_T - E[f_t] + \gamma_1 \end{bmatrix} = \widetilde{\Gamma},
\end{aligned}$$

where $\widetilde{\Gamma}$ is the ex-post risk premium and the first equality is obtained by the matrix inversion lemma.

T-Consistency of the IV Estimator (Proposition 1):

Finally, by Slutsky's theorem, we have the following probability limit of the IV estimator:

$$\begin{aligned}
\widehat{\Gamma}_{IV} &\to \begin{bmatrix} 1 & \mu_\beta^{'} \\ \mu_\beta & \mu_\beta\mu_\beta^{'} + \Sigma_\beta \end{bmatrix}^{-1} \begin{bmatrix} \gamma_0 + \mu_\beta^{'}\gamma_1 \\ \gamma_0\mu_\beta + \left(\mu_\beta\mu_\beta^{'} + \Sigma_\beta\right)\gamma_1 \end{bmatrix} \qquad \text{(A.31)} \\
&= \begin{bmatrix} 1 + \mu_\beta^{'}\Sigma_\beta^{-1}\mu_\beta & -\mu_\beta^{'}\Sigma_\beta^{-1} \\ -\Sigma_\beta^{-1}\mu_\beta & \Sigma_\beta^{-1} \end{bmatrix} \begin{bmatrix} \gamma_0 + \mu_\beta^{'}\gamma_1 \\ \gamma_0\mu_\beta + \left(\mu_\beta\mu_\beta^{'} + \Sigma_\beta\right)\gamma_1 \end{bmatrix} = \begin{bmatrix} \gamma_0 \\ \gamma_1 \end{bmatrix} = \Gamma
\end{aligned}$$

as N first goes to ∞ and then T goes to ∞, proving the T-consistency of the IV estimator.

The Rates of Convergence in Probability for the IV and FM Estimators (Proposition 2):

The above consistency proof of the IV estimator is silent about its rate of convergence, i.e. how fast $\widehat{\Gamma}_{IV}$ converges to its true value as T increases. Our discussion on the rate of convergence in probability addresses this issue. Under mild regularity conditions, Bahadur and Rao (1960) showed that the probability that sample means deviate from their corresponding population means for a pre-specified amount decays exponentially in T. In more technical terms, they showed that the probability bound $f(T)$ in the following inequality decays exponentially in T: for any $\varepsilon > 0$, when $x_1, \ldots, x_T$ are samples,

$$\Pr\left(|\bar{x}_T - \mu| > \varepsilon\right) < f(T),$$

where $\bar{x}_T$ denotes the sample mean and μ is the population mean.[a] Therefore the rates of convergence in probability of all sample means in equations (A.27) and (A.28), i.e., $\bar{f}_{T_o}$ and $\bar{f}_{T_e}$, are exponential in T. Since all terms in equations (A.27) and (A.28) consist of continuous mappings of sample means, for sufficiently large N, the probability that $\widehat{\Gamma}_{IV}$ deviates from the true value of Γ for a given amount also decays exponentially in T by the Contraction Principle (see Theorem 1.3.2 in Dupuis and Ellis (1997)). This means that the rate of convergence in probability of $\widehat{\Gamma}_{IV}$ is exponential in T.

In contrast, the FM estimator provides a different rate of convergence in probability due to the EIV problem. Following the same procedure as in the IV estimator, we can show that the FM estimator has the following probability limit as N goes to

[a]Brillinger (1962) provided weaker conditions which imply that the rate of convergence of $f(T)$ is a power in T.

∞:

$$\widehat{\Gamma}_{FM} = \left(\widehat{\mathbf{X}}_T'\widehat{\mathbf{X}}_T\right)^{-1}\widehat{\mathbf{X}}_T'\bar{\mathbf{R}}_T \rightarrow \begin{bmatrix} 1 & \mu_\beta' \\ \mu_\beta & \mu_\beta\mu_\beta' + \Sigma_\beta + \underbrace{\frac{\bar{\sigma}^2}{T}\widehat{\Sigma}_{f,T}^{-1}}_{=\text{EIV term}} \end{bmatrix}^{-1} \begin{bmatrix} \gamma_0 + \mu_\beta'\left(\bar{f}_T - E[f_t] + \gamma_1\right) \\ \gamma_0\mu_\beta + \left(\mu_\beta\mu_\beta' + \Sigma_\beta\right)\left(\bar{f}_T - E[f_t] + \gamma_1\right) \end{bmatrix}, \tag{A.32}$$

where, e.g., $\widehat{\mathbf{X}}_T$ is defined similarly to $\widehat{\mathbf{X}}_{T_e}$ (or $\widehat{\mathbf{X}}_{T_o}$) but for the entire sample period regardless of even and odd months. The cross-sectional mean of idiosyncratic variances, say $\bar{\sigma}^2$, is defined as the probability limit of $\frac{1}{N}\sum_i \sigma_i^2$ as N goes to ∞, where $\sigma_i^2 = \mathrm{Var}(\varepsilon_{i,t})$.[b] Note that the rates of convergence in probability of all sample means in equation (A.32) are still exponential in T. However, the EIV term $(=\frac{\bar{\sigma}^2}{T}\widehat{\Sigma}_{f,T}^{-1})$ in equation (A.32) determines the overall rate of convergence in probability of $\widehat{\Gamma}_{FM}$. Due to this EIV term, which did not arise in the IV estimator (see equations (A.27) and (A.28)), the rate of convergence in probability of the FM estimator is now linear in T (also by the Contraction Principle). In other words, for sufficiently large N, the probability that $\widehat{\Gamma}_{FM}$ deviates from the true value of Γ for a given amount decays linearly in T.

[b]For the derivation of equation (A.32), additional technical assumptions for $\varepsilon_{i,t}$ are needed. As in Jagannathan et al. (2009), defining $v_{i,t} = \varepsilon_{i,t}^2 - \sigma_i^2$, we assume the following laws of large numbers for cross-section:

$$\begin{aligned} \frac{1}{N}\sum_i v_{i,t} &\rightarrow 0 \quad \text{for all } t, \\ \frac{1}{N}\sum_i \sigma_i^2 &\rightarrow \bar{\sigma}^2 \end{aligned}$$

as N goes to ∞.

Technical Appendix 2: Simulation Parameters

This appendix describes how we choose simulation parameters based on the corresponding statistics in the data. We first determine the mean risk premiums and the covariance structure of the common factors based on the realizations of the three Fama-French factors over the 1956 to 2012 sample period. Table A.1 presents the summary statistics from the data, which we use in the simulations. The simulation uses a risk-free rate of 0.9996% per annum.

TABLE A.1: Simulation Parameters

Panel A: Time-series Means and Standard Deviations of Common Factors

		CAPM		Fama-French Three-factor Model	
		Mean (%)	StdDev (%)	Mean (%)	StdDev (%)
Factors	MKT	5.8008	15.2929	5.8008	15.2929
(per annum)	SMB			2.6388	10.3331
	HML			4.3620	9.6153

Panel B: Cross-sectional Means and Standard Deviations of factor sensitivities

		CAPM		Fama-French Three-factor Model	
		Mean	StdDev	Mean	StdDev
Factor Loadings	MKT	1.17	0.35	1.04	0.36
	SMB			1.05	0.68
	HML			0.59	0.55
Idiosyncratic Volatility (per annum)		0.125	0.053	0.129	0.053

Technical Appendix 3: Innovations in Illiquidity Costs

We follow Acharya and Pedersen (2005) and fit the following time-series regression to estimate expected and unexpected components of market-wide illiquidity cost $\left(\tilde{c}_{m,\tau} = c_{m,\tau} - E_{\tau-1}[c_{m,\tau}] \right)$:

$$0.25 + 0.3\, \widetilde{ILLIQ}_{m,\tau}\, P_{m,\tau-1} = a_0 + \sum_{l=1}^{L} a_l \times (0.25 + 0.3\, \widetilde{ILLIQ}_{m,\tau-l}\, P_{m,\tau-1}) + \tilde{c}_{m,\tau}, \quad (A.33)$$

where $\widetilde{ILLIQ}_{m,\tau}$ is a value-weighted average of $\min\left(ILLIQ_{i,\tau}, \frac{30-0.25}{0.30 P_{m,\tau-1}} \right)$, which Acharya and Pederson define as un-normalized illiquidity, truncated for outliers. We could not reject the hypothesis that the residuals were white noise based on the Durbin-Watson tests for L=2. The results we report uses the AR(2) model to estimate expected and unexpected components of illiquidity for the market as well as for individual stocks. We repeated the tests with L ranging from 2 to 6 and we found that the results were not sensitive to the choice of L.

Technical Appendix 4: Proof of Proposition 3

This appendix presents the proof of proposition 3. For expositional convenience, we assume that the even-month beta is the independent variable and odd-month beta is its instrument. We need to show that the correlation of true beta (x) and estimated beta (x^*) from even months is the square root of the correlation of estimated beta (x^*) and its instruments (z), i.e.,

$$Correlation(x, x^*) = \sqrt{Correlation\left(x^*, z\right)} \qquad (A.34)$$

where

$$x^* = x + u_{even}$$

$$z = x + u_{odd}$$

and x, u_{even}, and u_{odd} are mutually independent and $\sigma_u^2 = \sigma_{u_{even}}^2 = \sigma_{u_{odd}}^2$.

By the definition of correlation, we then have

$$Correlation\left(x^*, z\right) = \frac{Cov(x^*, z)}{\sqrt{var(x^*)var(z)}} = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_u^2} \qquad (A.35)$$

$$Correlation(x, x^*) = \frac{Cov(x, x^*)}{\sqrt{var\left(x\right)var(x^*)}} = \frac{\sigma_x^2}{\sqrt{\sigma_x^2\ (\sigma_x^2 + \sigma_u^2)}}$$

$$= \sqrt{Correlation\left(x^*, z\right)} \qquad (A.36)$$

completing the proof of proposition 3.

Information in CEOs' Facial Expressions: A First Look

Narasimhan Jegadeesh, Joonki Noh, and Jingran Zhao☆

Abstract

We investigate whether CEOs' facial expressions during their CNBC and Bloomberg interviews convey value-relevant information to financial markets. We employ a commercial software to quantify CEOs' basic emotions revealed through their facial expressions. We find that negative emotions are strongly correlated with cumulative abnormal stock returns and share turnover over the next one to two days after air dates. We also find that negative emotions are strongly associated with firms' one-quarter-ahead earnings. Taken together, our evidence supports that CEOs' basic emotions captured by their facial expressions in televised interviews with national TVs can convey value-relevant information about their firms to financial markets, and investors understand and react to it.

☆ Narasimhan Jegadeesh is the Dean's Distinguished Chair of Finance at the Goizueta Business School, Emory University, Atlanta, GA 30322, phone: 404-727-4821, email: Jegadeesh@emory. edu. Joonki Noh is a Doctoral student in Finance at Emory University and the corresponding author, phone: 734-276-3162, email: Joonki.Noh@emory.edu. Jingran Zhao is a Doctoral student in Accounting at Emory University, phone: 423-488-0269, email: Jingran.Zhao@emory.edu.

1. Introduction

Investors' ability to obtain and process information affects their portfolio performance critically. The source of the best information about any company is corporate insiders, but average investors do not have direct access to their inside information. Therefore, investors attempt to gauge the nature of insiders' information through a variety of channels including corporate visits, investor conferences, CEO interviews, and conference calls.

During televised interviews and personal interactions, investors observe managers' facial expressions. It is well established in the psychology literature that facial expressions can convey underlying basic emotions (e.g., Ekman, 1970). Since good news evokes positive emotions and bad news evokes negative emotions, an observer can potentially infer the nature of the subjects' private information by observing their facial expressions. This paper takes a first look at whether managers' facial expressions in their televised interviews can convey value-relevant information to financial markets and whether investors in markets react to it.

We analyze CEOs' facial expressions in a sample of 1,079 CEO interviews on CNBC and Bloomberg. We use FaceReader, a commercial software, to quantify the nature and intensity of CEOs' facial expressions during these interviews. FaceReader measures the intensities of the six basic and universal emotions that Ekman (1970) identifies: happy, sad, angry, surprised, scared, and disgusted. These six emotions are widely accepted as being basic and universal in the psychology literature (e.g., Ekman, 1993 and Busso, 2004).[1] 'Happy' is the only positive emotion and the other five are negative emotions. The emotional valence is defined as the difference between the intensity of positive emotion and that of the negative emotion that has the highest value.

We investigate the relation between CEOs' emotions captured by their facial expressions and stock market reactions in two dimensions, i.e., cumulative abnormal returns and cumulative abnormal turnover over one to two days after interviews are aired.

[1] Ekman (1993) defines *basic* emotions as ones that (1) differ from one another in important ways and (2) evolution played an important role in shaping their unique and the common features.

First, we find that the emotional valence is positively related to cumulative abnormal returns over the next one day after interviews are aired, which is mainly driven by negative emotions. Second, we find that cumulative abnormal turnover over the next one to two days after air date is related to the emotional valence. This relation is driven by both positive and negative emotions, but it is stronger for negative emotions. This asymmetric effect of CEOs' positive and negative facial expressions on cumulative abnormal returns and turnovers over the next one to two days might be due to the inconsistency with investors' expectations about CEOs' interviews. Investors might expect these CEOs to have at least neutral (if not happy) emotions when they are interviewed on national TVs.[2] If CEOs express negative emotions, this information might be inconsistent with investors' (implicit) expectation and therefore, become salient to investors. This might make investors react more strongly to CEOs' negative facial expressions.

An alternative explanation for the asymmetric effect of CEOs' positive and negative facial expressions might be that only the negative facial expression contains value-relevant information about firms' fundamentals, while its positive counterpart does not. Accordingly, in an additional empirical test, we investigate the relation between CEOs' emotions revealed through their facial expressions and one-quarter-ahead earnings after interviews are aired. We find evidence that CEOs' negative facial expression is strongly associated with one-quarter-ahead earnings, while CEOs' positive facial expression is not. This finding suggests that CEOs express negative emotions during their televised interviews when they expect that their firms would not perform well in the near future. The negative firm performance shows up in earnings one quarter later. This evidence of future earnings supports the alternative explanation for the asymmetric reactions by financial markets to CEOs' positive and negative facial expressions during interviews.

Taken together, our results suggest that CEOs' facial expressions contain value-relevant information about their firms. We expect that investors would be able to glean

[2] The descriptive statistics in Panel A of Table 20 are consistent with this argument. The mean and median of the 'Face' variable, which captures the emotional valence, are close to zero.

more information by personally observing facial expressions of corporate managers during investor conferences and corporate visits than what they can gain by observing televised interviews on national TVs. Although statutory restrictions such as Reg-FD prohibit managers from disclosing value-relevant information selectively to certain investors, our evidence suggests that those who are able to interact with managers personally would likely have an informational edge.

2. Quantifying Emotions based on Facial Expressions

Facial expressions can convey individuals' underlying emotions. Charles Darwin first theorizes that emotions are biologically determined and universal to human culture (Darwin, Ekman, and Phillip 1998). Ekman (1972) and others (e.g., Ekman and Dacher, 1970; Ekman and Friesen, 1975 & 1978; Ekman, 1992) systematically document evidence that shows the facial expressions associated with basic emotions are universal. Ekman (1972) defines facial expressions of emotions as being discrete, innate, and culturally independent.

Ekman and Friesen (1997) develop a coding system to classify the basic facial expressions into seven categories (i.e., happy, sad, angry, disgusted, scared, surprised, and neutral). They classify the expressions of emotions with semi-universal sequences of facial muscle contractions. For example, raising cheeks and lip corners are classified as being happy; raising inner brows, lowering brows, and depressing lip corners are classified as being sad (Ekman and Friesen, 1978). This coding system is referred to as the Facial Action Coding System (i.e., FACS). It is widely used in scientific research to describe visible movements of facial muscles. As technology advances, machines can learn to automatically recognize these emotions using digital images or video clips. In Appendix 3, we provide screenshots of CEOs who have the strongest facial expressions for each of the classified emotions other than neutral (i.e., facial expressions with the highest intensities). We also provide video clips in our sample that have the highest intensity for positive and negative emotions. The URL links to these video clips are provided in Appendix 2.

We measure CEOs' facial expressions with a software called FaceReader. FaceReader (version 5.0) is a commercial software developed by Noldus Information Technology, a Dutch company. Broadly speaking, FaceReader operates in the following three steps. First, FaceReader uses the Viola-Jones algorithm to detect the presence of a face in a video or in an image (Viola and Jones, 2001). Second, FaceReader models the face using an algorithm based on the Active Appearance Method proposed by Cootes and Taylor (2000). FaceReader identifies over 500 points on a face and analyzes the facial texture. These points on the face enable FaceReader to recognize the frame and components of the face, e.g., lip, eyebrows, nose, and eyes. The texture of the face includes the presence of wrinkles and the shape of eye brows. These are important cues for classifying facial expressions. Lastly, FaceReader classifies facial expressions into the seven pre-determined basic emotions using one of widely used machine learning algorithms, a neural network (for more details, see Bishop, 1995).[3]

FaceReader generates a numeric score for each emotion in every frame of a given video. The frame rate is three per second, which means FaceReader samples each frame to analyze every 0.33 second. Each emotion has a value between 0 and 1, indicating its intensity. '0' means that the emotion is absent in the frame, and '1' means that the emotion is fully present in the frame. For each video, we compute the intensity of each emotion using the median score for that emotion across all frames in a given video.

The Psychology literature defines emotional valence as the difference between the intensities of positive and negative emotions (e.g., see Bradley and Lang 2000). Happy is the only positive emotion in our study and hence we use the score on happy as the intensity of positive emotion. Negative emotions include 'sad', 'angry', 'scared', and 'disgusted'. Since 'surprise' can be either positive or negative, it is excluded from the calculation of the emotional valence, which we call 'Face'. We define the Face as the emotion captured by facial expressions. It is calculated as the intensity of 'happy' minus the intensity of the negative emotion with the highest intensity. For example, if the

[3] Over 10,000 manually annotated images were used to train the neural network algorithm. The Karolinska Directed Emotional Faces (KDEF) data were used for the original version of FaceReader, but additional training data were used in subsequent versions. KDEF data contain 4,900 pictures of 70 individuals, each displaying seven different emotions and each emotion is photographed twice from five different angles. These images have been annotated by human experts.

intensity of ‘happy’ is 0.7 and the intensities of ‘sad’, ‘angry’, ‘scared’, and ‘disgusted’ are 0.1, 0.5, 0.6, and 0.2, respectively, then Face is 0.1. It is calculated as 0.7 from ‘happy’ minus 0.6 from ‘scared’.

Several studies test the reliability of FaceReader classifications (e.g., Bijlstra and Dotsch 2011; Benta et al. 2009; Terzis et al 2010). Bijlstra and Dotsch (2011) employ images of the Radbound Faces Database (i.e., RaFD) in their reliability test. RaFD is a standardized database that contains a set of pictures with eight intended emotional expressions. The test persons in the RaFD pictures are trained to pose the intended emotions, and the images have been labeled by researchers accordingly. Bijlstra and Dotsch (2011) feed the RaFD images to FaceReader and compare the results with the “true” classifications in the RaFD database. Figure 3 reproduces the results of the validity test. FaceReader is the best at identifying ‘happy’ emotion with the 95.9% accuracy. The overall accuracy of identifying emotions is about 90%.

3. Video Sample Selection

We construct our sample universe by collecting CEO interviews that were aired on Bloomberg from 2010 to April 2014 (We stopped the data collection at the end of April 2014.), and that were aired on CNBC from 2012 to April 2014 (When we started to collect the CNBC videos, we could only retrieve videos back to 2012 from CNBC’s website.[4]). CNBC and Bloomberg are the only sources where we could find sufficient numbers of CEO interviews. From the Bloomberg terminal, we collect the videos of firms with different sizes. We use Fama-French market equity (ME) breakpoints[5] to categorize firms into small, medium, and big categories. We select 100 firms in each of these three size categories. We are able to obtain 291 videos from the Bloomberg terminal that were aired during trading hours. FaceReader processes 241 of these videos, and we obtain transcripts for 204 videos out of them. We also collect Bloomberg CEO

[4] Source: http://www.cnbc.com/id/100004032
[5] Source: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

interviews from the internet.[6] We obtain all available CEO interviews from the CNBC website.

A full video include frames that have faces of only the CEO, only the interviewer, and both the CEO and the interviewer. Since our focus is the CEOs' facial expressions, we manually edit these CEO interview videos to keep the frames in which only the CEOs are on the screen. The length of edited videos ranges from one (=150 frames) to five minutes (=900 frames). We use FaceReader to generate numeric scores for the basic emotions of a CEO during his/her interview. We also examine the informational content of words that the CEO uses in these segments using the bag of words approach used by Jegadeesh and Wu (2014). To get the transcripts of edited videos, we upload the edited videos to YouTube and manually download the automated transcripts to local drives to process for empirical analyses. In total, we have 1,645 videos that could be processed by FaceReader and had transcripts.

We obtain stock market data from CRSP and accounting data from COMPUSTAT. We obtain the analyst expectations of earnings data from I/B/E/S. After combing the videos with these databases, we have the final sample that is comprised of 1,079 videos for the subsequent analyses. All variables are winsorized at the 1% and 99% levels to avoid that outliers drive our empirical results.

4. Empirical Results

4.1. Preliminary Results

Various descriptive statistics are presented in Table 20. The definitions of all variables used in our empirical analyses are provided in Appendix 1. Panel A of Table 20 summarizes the characteristics of CEO interview videos that belong to the sample universe. The mean (median) of the valence of facial expressions (i.e., Face) is -3.85% (-1.40%), which means that CEOs' emotions captured by their facial expressions stay neutral during CNBC and Bloomberg interviews. To a certain extent, this is not

[6] Source: http://www.bloomberg.com/video/

surprising since CEOs might not want to express too much emotion during their interviews. Similarly, we find that on average 'neutral' has the highest intensity (mean=59.50% and median=65.53%). The results in Panel C indicate that CEOs whose firms had positive earnings surprises in the most recent fiscal quarter before interviews are substantially more likely to have interviews with CNBC and Bloomberg (71% versus 29%). Therefore, it seems less likely that CEOs would show negative facial expressions during CNBC and Bloomberg interviews.

Panel B of Table 20 summarizes the characteristics of firms whose CEOs are interviewed with national TVs. On average, the firms being interviewed have market value of equity of $26.30 billion. The median market value of equity in the sample is $7.48 billion, which is more than three times larger than the size of the median firm ($2.10 billion) in the Fama-French market equity breakpoint as of January 2012. This indicates that the firms being interviewed by CNBC and Bloomberg tend to be large firms. The sample firms have an average (median) book-to-market ratio of 0.84 (0.49).

Panel C of Table 20 presents the performance of firms whose CEOs are interviewed before the interviews are conducted. In the sample universe, 71% (29%) of the firms have standardized unexpected earnings (SUE) equal or larger (less) than zero before interviews. In unreported table, we also find similar patterns with CARs and raw returns over the past three months before the air dates, i.e., substantially more firms experience positive CARs and raw returns. These findings seem to make sense since CEOs whose firms experience superior performances, e.g., compared to earnings expectations, are more likely to have incentive to show up in televised interviews with national TVs. This is consistent with that CEOs speak more positive words than negative words in interviews as shown in Panel A.

Panel C of Table 20 also indicates that CEO interviews largely focus on their firms. We manually group all interviews into four categories based on the contents of videos: earnings-related, firm-related,[7] industry-related, and economy-related interviews. We find that 70% of videos have interviews about either earnings-related or firm-related

[7] Firm-related videos mean that interviews are about other aspects of firms rather than earnings.

topics, while 30% of videos have interviews about either industries that firms belong to or the economy. The remaining 3% of videos do not belong to these four categories and remain unclassified.

In Figure 5, we investigate the timing of CEOs' interviews relative to the earnings announcements of their firms. According to Panel C of Table 20, about 45% of interviews (=491 videos) are conducted within 20 trading days after earnings announcements, while the remaining 55% of interviews (=610 videos) are conducted outside the 20 trading days window. We find that substantially more earnings-related interviews are aired on the days of earnings announcements. Specifically, we find that 54% of earnings-related interviews (=103 videos) are conducted on the days of earnings announcements and that 28% of them (=53 videos) are aired on the next trading days. For other categories of interviews, the histograms are spread out over different dates.

Table 21 presents the Pearson and Spearman pair-wise correlations of facial expression variables, transcript tones, and other firm characteristics. Several observations are worth noting. First, transcript tones are not strongly correlated with facial expression variables. For example, the emotional valence, i.e., Face, has correlations of about 0.06 with both positive and negative transcript tones, i.e., PosWords and NegWords. This implies that transcript and facial expression variables can capture different information, and that facial expressions in CEOs' interviews can convey additional information to the viewers. Second, from the correlations between Ne and the four other negative facial expression variables, we find that Sad, Angry, and Disgusted contribute the most to Ne and Face. The correlations between Ne and Sad and Angry are the strongest. The Pearson (Spearman) correlation between Ne and Sad is 0.64 (0.60) and the Pearson (Spearman) correlation between Ne and Angry is 0.51 (0.47). The correlation between Ne and Scared is the lowest. The Pearson (Spearman) correlation between Ne and Scared is 0.24 (0.18). Third, we find that positive and negative emotions are correlated with each other. Pearson (Spearman) correlation between Po and Ne is -0.11 (-0.14). This result do not seem surprising because CEOs who stay positive during their interviews show less negative facial expressions. This negative relationship exists not only for Ne, but also for other negative emotions, e.g., Sad, Angry, and Disgusted. Lastly, many of emotional variables are correlated with each other. This makes sense in that facial expressions can be often a mixture of different emotions. Two or more facial expressions can occur simultaneously with high intensities.

In Figure 4, we provide preliminary evidence showing that the financial market reacts to the information contained in CEO interviews and facial expressions can convey value-relevant information to the market. We identify clear peaks in returns around the air dates of interviews when 60 trading days are examined around the air dates. We sort all videos into tercile portfolios based on Po (Ne) scores and the top tercile portfolio is denoted as Po (Ne) High, middle tercile portfolio is as Po (Ne) Med, and the bottom tercile portfolio is as Po (Ne) Low. In the top panel, we find that the return is the highest for firms that have the highest Po values on the interview air dates. This evidence suggests that the stock market reacts positively to positive facial expressions during CEOs' interviews. In the bottom panel, we find that the return is the lowest (highest) for firms that have the highest (lowest) Ne values on the interview air dates. This evidence suggests that the stock market reacts negatively to negative facial expressions during CEOs' interviews. In the bottom panel, it is worthwhile to emphasize that, for the tercile portfolio with the highest Ne value (blue graph), the peak remains almost intact one day after interview air dates, indicating that the impact of Ne on stock markets stays longer than its Po counterpart, i.e., the tercile portfolio with the highest Po value (green graph in the top panel).

4.2. Facial Expressions and Cumulative Abnormal Returns

In this section, we investigate whether stock market participants understand CEOs' facial expressions revealed during their televised interviews and react to them through cumulative abnormal returns (CARs). We first test the relationship between facial expression variables and variables that are related to CEOs' interviews. We consider the following panel regression. In day *d* (=0, .., 180),

$$CARd = \alpha + \beta_{Face}Face + \beta_{Po}Po + \beta_{Ne}Ne + TranscriptTone + \varepsilon, \qquad (3.1)$$

where *TranscriptTone* contains PosWords and NegWords. PosWords (NegWords) is the number of positive (negative) words in interview transcripts. The list of negative words comes from Jegadeesh and Wu (2014). To capture CEOs' face fixed effects in regression (3.1), we sort all videos into three sub-groups based on the facial expression scores of the first 30 frames in the edited videos and we then assign three dummy variables to constituent videos within sub-groups. Note that CAR*d* denotes the cumulative abnormal return from day 1 to day *d* and CAR0 is the abnormal return on day 0, i.e., the interview air date. To facilitate the interpretation of slope coefficients, we standardize them by subtracting the mean and dividing by the standard deviation of the corresponding variables. For *t*-statistics, to account for cross-sectional correlation across firms, we compute the robust standard errors by clustering by year-quarter.

Table 22 presents the results of panel regression (3.1). In the columns labeled as CAR*d*, we report the regression results when CAR*d* is employed as dependent variable. For CAR0, we find insignificant slope coefficients for facial expression variables. In terms of transcript tones, the slope coefficient of PosWords is 0.180% and significant at the 10% level and that of NegWords is -0.244% and significant at the 5% level. This evidence indicates that the stock market reacts to the number of negative words that CEOs spoke in their interviews, which is consistent with the findings in previous studies such as Tetlock (2007) and Tetlock et al. (2008).

For CAR1, we find evidence that the stock market participants strongly react to facial expression variables, especially Ne. The slope coefficient of Face is 0.175% and statistically significant at the 5% level. The slope coefficients of Ne range from -0.168% to -0.158% and are significant at the 5% level even when transcript tones are included in regression (3.1). The slope coefficient of Ne is not affected by controlling for positive and negative transcript tones in terms of its magnitude and statistical significance. These results imply that investors understand CEOs' emotions revealed through their facial expressions in CNBC and Bloomberg interviews and react to them through CAR1. In contrast, the slope coefficients of Po are not reliably different from zero at any conventional levels, which indicates that the significant slope coefficient of Face is mainly driven by Ne, not Po. The one-day delayed response to facial expression variables in CAR1 regressions could be due to the fact that the information contained in facial expressions is softer than that in transcript tones and that it thus take longer investors to digest it than its spoken counterpart, i.e., transcript tones. In CAR1, we find that the slope coefficient of NegWords is -0.191% and its statistical significance is weakened compared to CAR0.

To control for other known variables that potentially explains CAR*d*, we now run panel regressions with standard control variables. We consider the following pooled regressions. In day *d* (=0, .., 180),

$$CARd = \alpha + \beta_{Face}Face + \beta_{Po}Po + \beta_{Ne}Ne\ + Control + \varepsilon, \tag{3.2}$$

where *Control* contains the firm characteristics and transcript tones: PosWords, NegWords, Size, LogB2M, Momem, PEAD. and SUE. CEOs' face fixed effects are also included in regression (3.2). Size is the natural logarithm of market value of equity at one month before air date. LogB2M is the natural logarithm of book-to-market ratios. The book-to-market ratio is calculated as the book value of shareholders equity scaled by the market value of equity at the end of the previous year. Momem is the cumulative daily return over the past 100 days trading window, i.e., [–125, one day prior to the most recent earnings announcement before air date]. PEAD is the cumulative daily return over the trading window from the most recent earnings announcement date to one day before air date. We employ PEAD to control for the information contained in the earnings announcement released before the air date. SUE is the difference between actual earnings per share and the consensus (median) earnings forecast for the quarter prior to the video air date scaled by the price at 2 days before the most recent earnings announcement date. All other specifications are the same as in regression (3.1).

Table 23 provides the results of pooled regression (3.2). For CAR0, as in Table 22, we find insignificant slope coefficients for facial expression variables. In terms of transcript tones, the slope coefficient of PosWords is 0.224% and becomes significant at the 5% level, while that of NegWords loses its significance.

For CAR1, we find evidence that the stock market participants strongly react to facial expression variables, especially Ne, as in Table 22. The slope coefficient of Face is

0.19% and statistically significant at the 5% level. The slope coefficients of Ne range from -0.172% to -0.162% and are significant at the 5% level even when all control variables are included in regression (3.2). The significant slope coefficient of Ne is not affected by controlling for positive and negative transcript tones. None of control variables have consistently strong relationship with CARs across different specifications, which indicates that Ne has a unique explanatory power for CARs in regression (3.2). Overall, the regression results are similar to those in equation (3.1).

Different from Table 22, we find weak evidence that Po is positively associated with a longer-term CAR (i.e., CAR180). We find that the slope coefficient of Po is 0.866% and significant at the 10% level without controlling for Ne or transcript tones, which is not affected by controlling for positive and negative tones.

4.3. Facial Expressions and Cumulative Abnormal Turnover

In this section, we investigate whether investors understand CEOs' facial expressions revealed during their interviews with national TVs and react to them through trading activity, i.e., cumulative abnormal turnover (CAT). The abnormal turnover is the share turnover measured on day *d* minus the average share turnover over the past three months, i.e., (-90 days, -5 days). Share turnover is measured as number of shares traded divided by number of shares outstanding. Note that CAT*d* denotes the cumulative abnormal turnover from day 1 to day *d* and CAT0 is accordingly defined as the abnormal turnover on day 0, i.e., the interview air date. We consider the following pooled regressions. In day *d* (=0, 1,2),

$$CATd = \alpha + \beta_{Face}Face + \beta_{Po}Po + \beta_{Ne}Ne + Control + \epsilon, \qquad (3.3)$$

where *Control* contains the characteristics of interest: PosWords, NegWords, Size, LogB2M, SUE, LogAna, and LogAnaStd. LogAna is 1 plus the natural logarithm of number of analysts' forecasts issued for the current fiscal quarter end before the air date. LogAnaStd is the natural logarithm of the standard deviation of analysts' forecasts. All other control variables are defined as in regression (3.2). CEOs' face fixed effects and year fixed effects are also included in regression (3.3). To facilitate the interpretation of slope coefficients, we standardize them with the mean and the standard deviation of all available observations. For *t*-statistics, to account for cross-sectional correlation across firms and autocorrelation over time, we compute the robust standard errors by clustering by firm and year-quarter.

Table 24 provides the results of the pooled regression (3.3). In the columns labeled as CAT*d*, we report the regression results when CAT*d* is employed as dependent variable. For CAT0, we find no evidence that the stock market reacts to facial expression variables through trading activity. For CAT1, we find evidence that facial expression variables are strongly correlated with abnormal turnover over the next day after air dates. The slope coefficient of Face is 0.14% and is highly significant at the 1% level. When decomposing Face into Po and Ne, we find almost equally strong responses to positive and negative facial expressions in terms of their economic magnitudes. The slope coefficients of Po and Ne are 0.105% (significant at the 1% level) and -0.0985% (significant at the 10% level), respectively, without controlling for transcript tones. These slope coefficients stay almost intact when both Po and Ne are included or when transcript tones are controlled in regressions.

For CAT2, we again find strong evidence that investors respond to facial expression variables through CATs. The slope coefficient of Face is 0.207% and statistically significant at the 1% level. When decomposing Face into Po and Ne, we uncover evidence that the cumulative responses to Ne are stronger than those to Po regardless of specifications. For example, the slope coefficients of Po and Ne are 0.126% and -0.165%, respectively, and they are significant at the 5% level when transcript tones are excluded from regressions. Their magnitudes and statistical significance stay intact even when we include transcript tones as controls in regression (3.3). This evidence is consistent with the stronger price reactions to Ne in Table 23. In unreported table, we find that only Ne has significant slope coefficients at the 1% level and their values range from -0.071% to -0.066% when abnormal turnover on day 2 is employed as dependent variable, while the slope coefficients of Po do not differ from zero reliably and their magnitudes are substantially smaller than those of Ne.

In contrast, we find that the tones of interview transcripts do not have significant slope coefficients in any specifications, which indicates that investors do not react to the contents of spoken transcripts during CEOs' interviews through trading activity. Among other control variables, Size is the only variable that consistently has strong relationship with CATs across different specifications. It has negative relation with CATs, indicating that investors trade small stocks more when their CEOs' have interviews with national TVs.

4.4. Facial Expressions and Future Earnings

In this section, we examine whether CEOs' facial expressions contain information about the fundamentals of their firms, especially, earnings. Following prior literature (e.g., Li 2010), we consider the following pooled regressions:

$$Earn1 = \alpha + \beta_{Face}Face + \beta_{Po}Po + \beta_{Ne}Ne + Control + \epsilon, \tag{3.4}$$

where *Earn1* is one-quarter-ahead earnings after air date and *Control* contains the characteristics of interest: PosWords, NegWords, *Earn_1*, Size, LogB2M, and Momem. *Earn_1* is the most recent quarterly earnings before air date and it is computed as earnings before the extraordinary item (IBQ in COMPUSTAT) scaled by the book value. For example, when an interview is conducted in quarter *t*, *Earn_1* is the earnings in quarter *t-1* and *Earn1* is the earnings in quarter *t+1*. All other control variables are constructed in the same way as in CAR regressions (3.1) and (3.2). To facilitate the interpretation of slope coefficients, we standardize each variable with the mean and the standard deviation of all available observations. To control for unobserved heterogeneity across CEO faces and over time, we also include CEO face fixed effects and Year-quarter fixed effects. For *t*-statistics, to account for the time-series correlation over time, we compute the robust standard errors by clustering by industry.

Table 25 provides the results of the future earnings regressions in (3.4). We find that the slope coefficients of Ne range from -0.018% to -0.017% and they are significant at the 5% level. Their magnitudes and statistical significance stay unaffected by adding transcript tones into the regressions. In contrast, the slope coefficients of Po are insignificant at any conventional level. The slope coefficients of *Earn_1* are about 0.1% and significant at the 5% level across different specifications, which is consistent with the fact that earnings are persistent over time. None of other control variables have consistently strong relationships with future earnings across different specifications. The slope coefficient of NegWords is -0.008% and marginally significant at the 10% level. Taken together, the results in Table 25 suggest that the CEOs' facial expressions, especially negative ones, are strongly associated with one-quarter-ahead earnings of their firms.

4.5. Discussions

Overall, the evidence in Tables 22 to 25 indicates that investors in financial markets understand and react to the economic implications of CEOs' facial expressions during their televised interviews with national TVs. Combining these regression results, we can obtain better understanding of the reactions by investors to relatively more "soft" information contained in CEOs' facial expressions than the "hard" information contained in other traditional channels such as earnings announcements. First, the CAR regressions in Tables 22 and 23 present strong evidence that CAR1 can be predicted by CEOs' negative facial expressions. Second, the CAT regressions in Table 24 provide evidence that abnormal turnover can be predicated by both positive and negative facial expressions, but the CAT predictability by negative facial expression is stronger than that by positive facial expression. Combining CAR and CAT predictability regressions indicates that investors' reactions to CEOs' facial expressions are asymmetric to positive and negative facial expressions. However, these regressions are silent about why investors' reactions are asymmetric to Po and Ne. The future earnings regressions in Table 25 can provide us an (at least partial) answer for why investors in financial markets react to Ne substantially more strongly than to Po. The evidence in Table 25 indicates that Ne has strong association with one-quarter-ahead earnings and thus it contains value-relevant information about firm's fundamentals, while we find fairly weak evidence that Po also has such association with one-quarter-ahead earnings. To the extent that the stock markets are efficient, investors would quickly reflect the new value-relevant information contained in Ne to stock prices and thus Ne predicts the future abnormal return on the next day after air date. Since Po does not contain the information about firms' fundamentals, Po does not have such predictability of future abnormal returns.

5. Conclusion

For the first time in finance and economics, we empirically test whether CEOs' facial expressions revealed through CNBC and Bloomberg interviews convey value-relevant information to financial markets and how market participants react to it. To quantify the intensities of CEOs' facial expressions, we employ a commercial software that maps CEOs' facial expressions in their televised interviews into seven basic emotions. We uncover evidence that negative facial expressions are correlated with cumulative abnormal stock returns and cumulative abnormal turnover over the next one to two days after air dates. We also find that negative facial expressions are strongly associated with firms' one-quarter-ahead earnings. Taken together, our evidence suggests that CEOs' emotions captured by their facial expressions in their interviews with CNBC and Bloomberg convey value-relevant information about their firms to investors in financial markets, and investors understand and react to it.

Technical Appendix 1: Variable Definition

Variable	Definitions
Angry	Angry emotion measured as the median of angry emotion scores during an interview
B2M	Book-to-market ratio, calculated as the book value of shareholders equity scaled by the market value of equity at the end of the year before air date
CAR0	Abnormal returns measured on air date
CAR30 (CAR180)	30-day (180-day) cumulative abnormal returns measured after air date (start from the next trading day of the interview)
CAT0 (CAT1, CAT2)	Cumulative abnormal turnover on the air date minus the average turnover measured over (-5 days, -90 days). Turnover is measured as number of shares traded divided by number of shares outstanding.
Disgusted	Disgusted emotion measured as the median of disgusted emotion scores during an interview
Earn_1	Most recent quarterly earnings before air date and it is computed as earnings before the extraordinary item (IBQ in COMPUSTAT) scaled by the book value
Earn1	One quarter-ahead earnings after air date
Face	Facial expression score defined as Po-Ne
Po	Positive emotion measured as the median of happy emotion scores during an interview
LagRec	Level of analyst recommendations prior to air date
LogAna	The natural logarithm of number of analysts following in the year of air date plus 1.
LogB2M	Natural logarithm of book-to-market ratio
MktCap	Market capitalization defined as stock price times number of shares outstanding at the end of prior month to air date
Momem	Cumulative daily returns over a past 100 trading days window [–125, one day prior to the most recent earnings announcement before air date]
Ne	Score of the most salient negative emotion during an interview. Negative emotions identified by FaceReader include angry, disgusted, sad, and scared.
NegWords	Number of negative words spoken by a CEO during his/her interview. The list of negative words is from Jegadeesh and Wu (2014).
Nu	Neutral emotion measured as the median of neutral emotion scores during an interview
PosWords	Number of positive words spoken by a CEO during his/her interview. The list of positive words is from Jegadeesh and Wu (2014).
PEAD	Cumulative daily returns over the trading window from the most recent earnings announcement till one day before air date
Recommendation Revision	Difference between the average consensus recommendation immediately after and before video air date
Sad	Sad emotion measured as the median of sad emotion scores during an interview
Scared	Scared emotion measured as the median of scared emotion scores during an interview

Technical Appendix 1 ***(continued)***

Surprised	Surprised emotion measured as the median of surprised emotion scores during an interview
Size	Natural logarithm of market value of equity at the end of prior month to air date
SUE	Difference between actual earnings per share and the summary consensus median earnings forecast for the fiscal quarter end prior to video air date scaled by the price at 2 days before the earnings announcement date
Volat	The standard deviation of daily stock returns over the past 125 trading days before air date

Technical Appendix 2: Links to Sample Videos

Highest Face Videos:
https://www.youtube.com/watch?v=oEewdFapSII&list=PLhs3gTH3HCXU9RZaVScAEoLgsK1KF8YpL

Lowest Face Videos:

https://www.youtube.com/watch?v=xjC7SU64zhY&list=PLhs3gTH3HCXXa_vOdbStRbHy5LxlDWALw

Technical Appendix 3: Screenshots from the Videos with the Most Extreme Facial Expression Scores

Disgusted

Scared

Technical Appendix 3 ***(continued)***

Sad

Angry

Technical Appendix 3 ***(continued)***

Surprise

Happy

Appendices

Table 1.

Summary Statistics for the Industry Network Centrality (in 1992)

This table provides the summary statistics for the industry network centralities (Panel A) and the lists of the twenty most and twenty least central industries (Panel B) in the 1992 BEA report. 465 disaggregate industries are analyzed in the 1992 Detailed Input-Output Tables from the Bureau of Economic Analysis (BEA). The reported centralities are the eigenvector centralities based on their inter-industry trades.

Panel A: Summary Statistics for Centralities

Statistics & percentiles	Eigenvector centrality of the industry network
Mean	0.038
Standard deviation	0.027
Minimum	0.010
5th percentile	0.018
10th percentile	0.021
25th percentile	0.025
Median	0.033
75th percentile	0.040
90th percentile	0.056
95 percentile	0.068
Maximum	0.347
Number of observations	465

Panel B: Lists of the Twenty Most and Twenty Least Central Industries

Twenty most central industries	Twenty least central industries
Wholesale trade	Petroleum, natural gas, solid mineral exploration
Eating and drinking places	Racing, including track operation
Construction industries	Cigars
Real estate agents, managers, and operators	Boot and shoe cut stock and findings
Commercial construction industries	Women's hosiery, except socks
Trucking and courier services, except air	Hosiery
Blast furnaces and steel mills	Tobacco stemming and redrying
Electric services (utilities)	Manufactured ice
Miscellaneous plastics products	Chewing and smoking tobacco and snuff
Motor vehicles and passenger car bodies	Schiffli machine embroideries
Banking	Professional sports clubs and promoters
Petroleum refining	Women's handbags and purses
Retail trade, except eating and drinking	Burial caskets
Industrial inorganic and organic chemicals	Leather gloves and mittens
Paper and paperboard mills	Personal leather goods
Paperboard containers and boxes	Special product sawmills
Motor vehicle parts and accessories	Jewelers' materials and lapidary work
Telephone, telgraph communications, and communications	X-ray apparatus and tubes
Hospitals	Leather goods
Bread, cake, and related products	Costume jewelry
Automotive repair shops and services	Nonferrous metal ores, except copper

Table 2.

Characteristics of Centrality-sorted Quintile Portfolios

This table provides the averages of various characteristics of centrality-sorted quintile portfolios over January 1972 through December 2012. Eight detailed Input-Output tables from the Bureau of Economic Analysis (BEA) are merged with CRSP and COMPUSTAT data through SIC/NAICS codes. For each BEA report, eigenvector centralities are computed and assigned to individual stocks based on their industry memberships. Within each quintile portfolio, the following characteristics are computed by equal-weighting constituent stocks: the market capitalization (MKTCAP) in billion dollars, book-to-market (BM) ratio, return volatility (VOL), idiosyncratic return volatility (IVOL), share turnover (TURN) based on the NYSE/AMEX stock universe, the number of analysts following on a given month (ANALFOLL), and the percentage of institutional holdings (IHP). Customer-HHI and Supplier-HHI (two rightmost columns) are the Herfindahl-Hirschman indices (HHIs) for sales and purchases per industry, respectively, which are aggregated across industries. Characteristics of quintile portfolios are averaged over the sample period.

Panel: Average Characteristics of Centrality-sorted Quintile Portfolios

Centrality rank	MKTCAP (Bill$)	BM ratio	VOL	IVOL	TURN*	ANAL-FOLL	IHP (%)	Customer-HHI	Supplier-HHI
High (1)	2.563	1.094	0.026	0.085	0.085	3.10	42.1	0.292	0.117
(2)	1.986	0.934	0.026	0.087	0.083	2.88	44.5	0.157	0.099
(3)	1.979	0.938	0.026	0.089	0.082	2.93	45.1	0.170	0.092
(4)	2.104	0.866	0.028	0.094	0.080	2.74	44.0	0.203	0.102
Low (5)	2.070	0.895	0.028	0.096	0.079	2.58	43.2	0.259	0.127

*based on the NYSE/AMEX stock universe

Table 3.

Return Cross-predictability and Network Positions

I consider special cases of the following FM cross-sectional regression: for firm i in month t,

$$r^e_{i,t} = \alpha_t + \gamma^T_{Z,t} Z_{i,t} + \sum_{l \in (C,M,P)} \gamma^{related}_{l,t} D_{i,l,t-1} r^{related}_{i,t-1} + \varepsilon_{i,t},$$

where $r^e_{i,t}$ is the excess return of firm i which belongs to the industry in question, $Z_{i,t}$ contains characteristics of interest: SIZE, BM, MOM, and REV, and l is chosen from $(C, M, P) = (Center, Middle, Periphery)$. $D_{i,l,t-1}$ is an indicator variable defining the location of firm i in the industry network in month $t-1$. For example, if the industry which firm i belongs to is central, $D_{i,C,t-1}$ is 1 otherwise 0. $r^{related}_{i,t-1}$ is the lagged aggregate return of customer and supplier industries of firm i by one month. Panels A and B conduct the firm-level and industry-level FM cross-sectional regressions, respectively. To report t-statistics in parentheses (bold if significant at the 5% level), the Newey-West HAC covariance matrix estimators are employed.

Panel A: Firm-level

	(1) RET(t)	(2) RET(t)
SIZE	-0.001	-0.001
	(-1.43)	(-1.43)
BM	**0.002**	**0.002**
	(2.99)	**(2.99)**
MOM	**0.006**	**0.006**
	(3.06)	**(3.06)**
REV	**-0.050**	**-0.050**
	(-10.45)	**(-10.45)**
RET(Related,$t-1$) × Dummy(Center)	**0.182**	
	(5.56)	
RET(Related,$t-1$) × Dummy(Middle)	**0.156**	
	(5.73)	
RET(Related,$t-1$) × Dummy(Periphery)	**0.095**	
	(3.94)	
RET(Related,$t-1$) × Dummy(Center)		**0.088**
		(2.36)
RET(Related,$t-1$) × Dummy(Middle)		0.062
		(1.84)
RET(Related,$t-1$)		**0.095**
		(3.94)
Dummy(Center)	**0.020**	**0.020**
	(2.39)	**(2.39)**
Dummy(Middle)	**0.020**	**0.020**
	(2.33)	**(2.33)**
Dummy(Periphery)	**0.019**	**0.019**
	(2.31)	**(2.31)**

Table 3.
Return Cross-Predictability and Network Positions (Continued)

Panel B: Industry-level

	(1) RET(t)	(2) RET(t)
SIZE	0.000	0.000
	(-0.71)	(-0.71)
BM	**0.002**	**0.002**
	(2.18)	**(2.18)**
MOM	**0.008**	**0.008**
	(3.27)	**(3.27)**
REV	**-0.035**	**-0.035**
	(-5.39)	**(-5.39)**
RET(Related,$t-1$) $\times$ Dummy(Center)	**0.140**	
	(4.75)	
RET(Related,$t-1$) $\times$ Dummy(Middle)	**0.115**	
	(4.50)	
RET(Related,$t-1$) $\times$ Dummy(Periphery)	**0.064**	
	(2.58)	
RET(Related,$t-1$) $\times$ Dummy(Center)		**0.076**
		(2.09)
RET(Related,$t-1$) $\times$ Dummy(Middle)		0.051
		(1.44)
RET (Related,$t-1$)		**0.064**
		(2.58)
Dummy(Center)	0.012	0.012
	(1.27)	(1.27)
Dummy(Middle)	0.014	0.014
	(1.51)	(1.51)
Dummy(Periphery)	0.014	0.014
	(1.52)	(1.52)

Table 4.
Self-financing Trading Strategies based on Central and Peripheral Industries

Panels A1 (B1) and A2 (B2) provide the value-weighted (equal-weighted) excess and risk-adjusted returns of self-financing trading strategies that invest exclusively in central or peripheral industries, respectively. For example, to implement a trading strategy that invests exclusively in central industries, quintile portfolios are formed based on the average one-month lagged return of customer and supplier industries and a long-short hedging portfolio (labeled as "High-Low") is constructed by going long the highest quintile (labeled as "High (1)") and going short the lowest quintile (labeled as "Low (5)") portfolios. Quintile portfolios are rebalanced every month. Returns and their standard deviations are annualized in Panels A1 and B1. Monthly risk-adjusted returns (alphas) are obtained after controlling for the exposure to Fama-French three (MKTRF, SMB, HML), momentum (UMD), and Pastor and Stambaugh tradable liquidity (PS-LIQ) factors. In Panels A2 and B2, t-statistics are presented in parentheses (bold if significant at the 5% level).

Panel A1: Excess Returns (Value-weighted)

		High (1)	(2)	(3)	(4)	Low (5)	High-Low
Central	Mean excess return	0.127	0.091	0.073	0.078	0.056	0.071
industries	Standard deviation	0.185	0.186	0.187	0.182	0.189	0.131
	Sharpe ratio	0.686	0.489	0.391	0.426	0.299	0.544
Peripheral	Mean excess return	0.122	0.110	0.077	0.081	0.095	0.027
industries	Standard deviation	0.195	0.200	0.178	0.207	0.214	0.173
	Sharpe ratio	0.627	0.549	0.432	0.390	0.447	0.156

Panel A2: Risk-adjusted Returns (Value-weighted)

Trading Strategy based on	Alpha	MKTRF	SMB	HML	UMD	PS-LIQ	Adj-R^2
Central	**0.006**	0.007	-0.040	-0.022	0.043	0.008	-0.007
industries	**(3.55)**	(0.17)	(-0.76)	(-0.40)	(1.27)	(0.30)	
Peripheral	0.001	-0.081	-0.153	0.034	**0.162**	**0.103**	0.037
industries	(0.53)	(-1.30)	(-1.77)	(0.38)	**(2.93)**	**(2.32)**	

Panel B1: Excess Returns (Equal-weighted)

		High (1)	(2)	(3)	(4)	Low (5)	High-Low
Central	Mean excess return	0.141	0.110	0.093	0.079	0.060	0.081
industries	Standard deviation	0.184	0.184	0.183	0.181	0.186	0.092
	Sharpe ratio	0.770	0.596	0.510	0.435	0.325	0.877
Peripheral	Mean excess return	0.131	0.104	0.090	0.089	0.086	0.045
industries	Standard deviation	0.180	0.184	0.177	0.185	0.207	0.117
	Sharpe ratio	0.727	0.567	0.508	0.481	0.414	0.384

Panel B2: Risk-adjusted Returns (Equal-weighted)

Trading Strategy based on	Alpha	MKTRF	SMB	HML	UMD	PS-LIQ	Adj-R^2
Central	**0.007**	0.031	-0.034	0.013	0.012	-0.024	-0.008
industries	**(4.64)**	(0.92)	(-0.72)	(0.26)	(0.40)	(-0.99)	
Peripheral	**0.004**	**-0.113**	-0.087	0.039	**0.088**	0.013	0.045
industries	**(2.15)**	**(-2.70)**	(-1.50)	(0.63)	**(2.36)**	(0.44)	

Table 5.

Return Cross-predictability and Network Positions for Multi-periods

I consider special cases of the following FM cross-sectional regression: for firm i in month t,

$$r_{i,t}^{e} = \alpha_t + \gamma_{Z,t}^{T} Z_{i,t} + \sum_{l \in (C,M,P)} \gamma_{l,t,k}^{related} D_{i,l,t-k} r_{i,t-k}^{related} + \varepsilon_{i,t,k},$$

where $r_{i,t}^{e}$ is the excess return of firm i which belongs to the industry in question, $Z_{i,t}$ contains characteristics of interest: SIZE, BM, MOM, and REV, and l is chosen from $(C, M, P) = (Center, Middle, Periphery)$. $D_{i,l,t-k}$ is an indicator variable defining the location of firm i in the industry network in month $t-k$. For example, if the industry which firm i belongs to is central, $D_{i,C,t-k}$ is 1 otherwise 0. $r_{i,t-k}^{related}$ is the lagged aggregate return of customer and supplier industries of firm i by k months. For each month, all variables are standardized cross-sectionally. To report t-statistics in parentheses (bold if significant at the 5% level), the Newey-West HAC covariance matrix estimators are employed.

Panel: Firm-level

	(1) RET(t) $k=0$	(2) RET(t) $k=0$	(3) RET(t) $k=1$	(4) RET(t) $k=1$	(5) RET(t) $k=2$	(6) RET(t) $k=2$
SIZE	-0.005	-0.005	-0.005	-0.005	-0.004	-0.004
	(-0.93)	(-0.93)	(-0.85)	(-0.85)	(-0.68)	(-0.68)
BM	**0.014**	**0.014**	**0.014**	**0.014**	**0.013**	**0.013**
	(3.66)	**(3.66)**	**(3.67)**	**(3.67)**	**(3.48)**	**(3.48)**
MOM	**0.019**	**0.019**	**0.020**	**0.020**	**0.019**	**0.019**
	(3.78)	**(3.78)**	**(3.77)**	**(3.77)**	**(3.60)**	**(3.60)**
REV	**-0.049**	**-0.049**	**-0.049**	**-0.049**	**-0.048**	**-0.048**
	(-11.07)	**(-11.07)**	**(-11.02)**	**(-11.02)**	**(-10.67)**	**(-10.67)**
RET(Related,k)×Dummy(C)	**0.110**		**0.031**		**0.013**	
	(24.57)		**(6.53)**		**(2.43)**	
RET(Related,k)×Dummy(M)	**0.109**		**0.025**		0.008	
	(18.71)		**(5.39)**		(1.64)	
RET(Related,k)×Dummy(P)	**0.085**		**0.020**		0.004	
	(14.58)		**(4.51)**		(0.87)	
RET(Related,k)×Dummy(C)		**0.025**		**0.011**		0.009
		(3.94)		**(2.12)**		(1.86)
RET(Related,k)×Dummy(M)		**0.024**		0.005		0.004
		(3.68)		(0.87)		(0.61)
RET(Related,k)		**0.085**		**0.020**		0.004
		(14.58)		**(4.51)**		(0.87)
Dummy(Center)	-0.003	-0.003	-0.002	-0.002	-0.002	-0.002
	(-1.45)	(-1.45)	(-0.9)	(-0.9)	(-1.14)	(-1.14)
Dummy(Middle)	0.004	0.004	0.005	0.005	0.005	0.005
	(1.60)	(1.60)	(1.50)	(1.50)	(1.76)	(1.76)
Dummy(Periphery)	0.006	0.006	0.006	0.006	0.005	0.005
	(1.87)	(1.87)	(1.72)	(1.72)	(1.21)	(1.21)

Table 6.

Cross-predictability of Analysts' Earnings Forecast Revisions and Network Positions

I consider the following pooled regression: for firm i in month t and k-month lag ($k = 0, 1, 2$),

$$\text{AREV}_{i,t} = \alpha_i + \beta_t + \gamma_Z^T Z_{i,t} + \sum_{l \in (C,M,P)} \psi_{l,k}^{related} D_{i,l,t-k} \text{AREV}_{i,t-k}^{related} + \varepsilon_{i,t,k},$$

where $\text{AREV}_{i,t}$ is the analysts' revision of earnings forecast of firm i in month t. I define the analysts' revision as $\text{AREV}_{i,t} = (\text{UP}_{i,t} - \text{DOWN}_{i,t})/\text{NUMEST}_{i,t}$, where $\text{NUMEST}_{i,t}$ is the number of estimates of firm i's earnings for the current fiscal quarter end, and $\text{UP}_{i,t}$ ($\text{DOWN}_{i,t}$) is the number of upward (downward) earnings forecast revisions. $Z_{i,t}$ contains control variables: lagged AREV, the aggregate return of related industries of firm i lagged by one month, and the industry-level analysts' revision (AREVIND) lagged by one month. l is chosen from $(C, M, P) = (Center, Middle, Periphery)$. $D_{i,l,t-k}$ is an indicator variable defining the location of firm i in month $t-k$. For example, if the industry which firm i belongs to is central, $D_{i,C,t-k}$ is 1 otherwise 0. $\text{AREV}_{i,t-k}^{related}$ is the aggregate analysts' revision of the related industries of firm i in month $t-k$. To report t-statistics in parentheses (bold if significant at the 5% level), robust standard errors are computed by double-clustering by firm and year-month. All variables are standardized.

Table 6.
Cross-predictability of Analysts' Earnings Forecast Revisions and Network Positions (Continued)

Panel: Revisions of Analysts' Earnings Forecast

	(1) AREV(t) $k=0$	(2) AREV(t) $k=0$	(3) AREV(t) $k=1$	(4) AREV(t) $k=1$	(5) AREV(t) $k=2$	(6) AREV(t) $k=2$
AREV($t-1$)			**0.063**	**0.063**	**0.063**	**0.063**
			(12.67)	**(12.67)**	**(12.88)**	**(12.88)**
RET(Related,$t-1$)			**0.044**	**0.044**	**0.048**	**0.048**
			(3.84)	**(3.84)**	**(3.99)**	**(3.99)**
AREVIND($t-1$)			**0.042**	**0.042**	**0.044**	**0.044**
			(10.51)	**(10.51)**	**(10.83)**	**(10.83)**
AREV(Related,$t-k$) × Dummy(Center)	**0.085**		**0.049**		**0.032**	
	(8.70)		**(6.06)**		**(4.83)**	
AREV(Related,$t-k$) × Dummy(Middle)	**0.092**		**0.033**		**0.020**	
	(9.54)		**(5.17)**		**(3.40)**	
AREV(Related,$t-k$) × Dummy(Periphery)	**0.079**		0.009		0.006	
	(8.70)		(1.34)		(0.77)	
AREV(Related,$t-k$) × Dummy(Center)		0.005		**0.040**		**0.026**
		(0.45)		**(4.34)**		**(2.80)**
AREV(Related,$t-k$) × Dummy(Middle)		0.012		**0.024**		0.014
		(1.18)		**(3.24)**		(1.69)
AREV(Related,$t-k$)		**0.079**		0.009		0.006
		(4.42)		(1.34)		(0.77)
Dummy(Center)	-0.014	-0.014	-0.002	-0.002	-0.003	-0.003
	(-0.98)	(-0.98)	(-0.12)	(-0.12)	(-0.18)	(-0.18)
Dummy(Middle)	-0.006	-0.006	-0.003	-0.003	-0.003	-0.003
	(-0.48)	(-0.48)	(-0.19)	(-0.19)	(-0.21)	(-0.21)
Constant	**0.064**	**0.064**	**0.092**	**0.092**	-0.043	-0.043
	(4.42)	**(4.42)**	**(3.94)**	**(3.94)**	(-1.46)	(-1.46)
Firm/Time-fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Clustered S.E.	Firm/ Year-Month	Firm/ Year-Month	Firm/ Year-Month	Firm/ Year-Month	Firm/ Year-Month	Firm/ Year-Month

Table 7.
Testing Existing Anomalies: the Conglomerate Effect

Using standalone firms, I consider special cases of the following FM cross-sectional regression: for firm i in month t,

$$r^e_{i,t} = \alpha_t + \gamma^T_{Z,t} Z_{i,t} + \sum_{l \in (C,M,P)} \gamma^{related}_{l,t} D_{i,l,t-1} r^{related}_{i,t-1} + \varepsilon_{i,t},$$

where $r^e_{i,t}$ is the excess return of firm i which belongs to the industry in question, $Z_{i,t}$ contains characteristics of interest: SIZE, BM, MOM, and REV, and l is chosen from $(C, M, P) = (Center, Middle, Periphery)$. $D_{i,l,t-1}$ is an indicator variable defining the location of firm i in the industry network in month $t-1$. For example, if the industry which firm i belongs to is central, $D_{i,C,t-1}$ is 1 otherwise 0. $r^{related}_{i,t-1}$ is the lagged aggregate return of customer and supplier industries of firm i by one month. To report t-statistics in parentheses (bold if significant at the 5% level), the Newey-West HAC covariance matrix estimators are employed.

Panel: Standalone Firms

	(1) RET(t)	(2) RET(t)
SIZE	**-0.001**	**-0.001**
	(-2.43)	**(-2.43)**
BM	**0.002**	**0.002**
	(2.75)	**(2.75)**
MOM	**0.005**	**0.005**
	(2.57)	**(2.57)**
REV	**-0.043**	**-0.043**
	(-10.53)	**(-10.53)**
RET(Related,$t-1$) × Dummy(Center)	**0.162**	
	(4.05)	
RET(Related,$t-1$) × Dummy(Middle)	**0.104**	
	(2.78)	
RET(Related,$t-1$) × Dummy(Periphery)	**0.078**	
	(2.43)	
RET(Related,$t-1$) × Dummy(Center)		**0.085**
		(2.13)
RET(Related,$t-1$) × Dummy(Middle)		0.026
		(0.58)
RET(Related,$t-1$)		**0.078**
		(2.43)
Dummy(Center)	**0.033**	**0.033**
	(3.28)	**(3.28)**
Dummy(Middle)	**0.034**	**0.034**
	(3.35)	**(3.35)**
Dummy(Periphery)	**0.034**	**0.034**
	(3.47)	**(3.47)**

Table 8.
Testing Existing Anomalies: Limits to Arbitrage

I partition the entire stock universe into three sub-groups sorted on firm-level idiosyncratic volatility. For example, columns titled H-IVOL mean that the sub-group with high level of idiosyncratic volatility are analyzed. Within each sub-group, I consider special cases of the following FM cross-sectional regression: for firm i in month t,

$$r^e_{i,t} = \alpha_t + \gamma^T_{Z,t} Z_{i,t} + \sum_{l \in (C,M,P)} \gamma^{related}_{l,t} D_{i,l,t-1} r^{related}_{i,t-1} + \varepsilon_{i,t},$$

where $r^e_{i,t}$ is the excess return of firm i which belongs to the industry in question, $Z_{i,t}$ contains characteristics of interest: SIZE, BM, MOM, and REV, and l is chosen from $(C, M, P) = (Center, Middle, Periphery)$. $D_{i,l,t-1}$ is an indicator variable defining the location of firm i in the industry network in month $t-1$. For example, if the industry which firm i belongs to is central, $D_{i,C,t-1}$ is 1 otherwise 0. $r^{related}_{i,t-1}$ is the lagged aggregate return of customer and supplier industries of firm i by one month. To report t-statistics in parentheses (bold if significant at the 5% level), the Newey-West HAC covariance matrix estimators are employed.

Panel: Subsets sorted on Idiosyncratic Volatility

	(1) RET(t) H-IVOL	(2) RET(t) H-IVOL	(3) RET(t) M-IVOL	(4) RET(t) M-IVOL	(5) RET(t) L-IVOL	(6) RET(t) L-IVOL
SIZE	**-0.001**	**-0.001**	**-0.001**	**-0.001**	**-0.001**	**-0.001**
	(-3.09)	**(-3.09)**	**(-2.36)**	**(-2.36)**	**(-2.32)**	**(-2.32)**
BM	**0.002**	**0.002**	**0.001**	**0.001**	**0.001**	0.001
	(3.22)	**(3.22)**	**(2.02)**	**(2.02)**	**(1.73)**	(1.73)
MOM	**0.005**	**0.005**	**0.010**	**0.010**	**0.008**	**0.008**
	(2.78)	**(2.78)**	**(4.28)**	**(4.28)**	**(3.07)**	**(3.07)**
REV	**-0.050**	**-0.050**	**-0.054**	**-0.054**	**-0.058**	**-0.058**
	(-8.61)	**(-8.61)**	**(-10.49)**	**(-10.49)**	**(-11.65)**	**(-11.65)**
RET(Related,1)×Dummy(C)	**0.301**		**0.170**		**0.125**	
	(4.70)		**(5.37)**		**(4.14)**	
RET(Related,1)×Dummy(M)	**0.161**		**0.162**		**0.122**	
	(3.15)		**(4.25)**		**(4.64)**	
RET(Related,1)×Dummy(P)	**0.151**		0.055		**0.079**	
	(3.25)		(1.87)		**(3.28)**	
RET(Related,1)×Dummy(C)		**0.149**		**0.115**		0.046
		(2.00)		**(2.87)**		(1.35)
RET(Related,1)×Dummy(M)		0.009		**0.107**		0.043
		(0.14)		**(2.55)**		(1.29)
RET(Related,1)		**0.151**		0.055		**0.079**
		(3.25)		(1.87)		**(3.28)**
Dummy(Center)	**0.035**	**0.035**	**0.024**	**0.024**	**0.020**	**0.020**
	(4.08)	**(4.08)**	**(3.07)**	**(3.07)**	**(3.41)**	**(3.41)**
Dummy(Middle)	**0.034**	**0.034**	**0.025**	**0.025**	**0.022**	**0.022**
	(3.55)	**(3.55)**	**(3.31)**	**(3.31)**	**(3.47)**	**(3.47)**
Dummy(Periphery)	**0.029**	**0.029**	**0.026**	**0.026**	**0.025**	**0.025**
	(3.43)	**(3.43)**	**(3.37)**	**(3.37)**	**(4.06)**	**(4.06)**

Table 9.
Testing Existing Anomalies: the Institutional Ownership, Turnover, and Illiquidity Effects

I partition the entire stock universe into three sub-groups sorted on the percentage ownership of institutional investors (IHP), share turnover (TURN), or Amihud illiquidity (ILLIQ). For example, in Panel A (Panel B), columns titled H-IHP (H-TURN) mean that the sub-group with high level of IHP (TURN) are analyzed. Within each sub-group, I consider special cases of the following FM cross-sectional regression: for firm i in month t,

$$r^e_{i,t} = \alpha_t + \gamma^T_{Z,t} Z_{i,t} + \sum_{l\in(C,M,P)} \gamma^{related}_{l,t} D_{i,l,t-1} r^{related}_{i,t-1} + \varepsilon_{i,t},$$

where $r^e_{i,t}$ is the excess return of firm i which belongs to the industry in question, $Z_{i,t}$ contains characteristics of interest: SIZE, BM, MOM, and REV, and l is chosen from $(C, M, P) = (Center, Middle, Periphery)$. $D_{i,l,t-1}$ is an indicator variable defining the location of firm i in the industry network in month $t-1$. For example, if the industry which firm i belongs to is central, $D_{i,C,t-1}$ is 1 otherwise 0. $r^{related}_{i,t-1}$ is the lagged aggregate return of customer and supplier industries of firm i by one month. To report t-statistics in parentheses (bold if significant at the 5% level), the Newey-West HAC covariance matrix estimators are employed.

Panel A: Subsets sorted on IHP

	(1) RET(t) H-IHP	(2) RET(t) H-IHP	(3) RET(t) M-IHP	(4) RET(t) M-IHP	(5) RET(t) L-IHP	(6) RET(t) L-IHP
SIZE	**0.001**	**0.001**	0.000	0.000	**-0.001**	**-0.001**
	(2.68)	**(2.68)**	(-0.66)	(-0.66)	**(-3.08)**	**(-3.08)**
BM	0.000	0.000	**0.002**	**0.002**	**0.002**	**0.002**
	(0.39)	(0.39)	**(2.94)**	**(2.94)**	**(3.35)**	**(3.35)**
MOM	**0.005**	**0.005**	0.003	0.003	**0.006**	**0.006**
	(1.98)	**(1.98)**	(1.08)	(1.08)	**(2.95)**	**(2.95)**
REV	**-0.037**	**-0.037**	**-0.046**	**-0.046**	**-0.038**	**-0.038**
	(-7.37)	**(-7.37)**	**(-10.30)**	**(-10.30)**	**(-7.17)**	**(-7.17)**
RET(Related,1)×Dummy(C)	**0.152**		**0.219**		**0.223**	
	(3.17)		**(4.85)**		**(4.78)**	
RET(Related,1)×Dummy(M)	**0.118**		**0.164**		**0.161**	
	(2.87)		**(3.40)**		**(3.50)**	
RET(Related,1)×Dummy(P)	0.029		0.069		**0.136**	
	(0.83)		(1.85)		**(3.34)**	
RET(Related,1)×Dummy(C)		**0.123**		**0.149**		**0.088**
		(2.24)		**(3.16)**		**(2.29)**
RET(Related,1)×Dummy(M)		0.088		0.095		0.025
		(1.64)		(1.80)		(0.42)
RET(Related,1)		0.029		0.069		**0.136**
		(0.83)		(1.85)		**(3.34)**
Dummy(Center)	-0.018	-0.018	0.017	0.017	**0.032**	**0.032**
	(-1.75)	(-1.75)	(1.70)	(1.70)	**(3.62)**	**(3.62)**
Dummy(Middle)	**-0.020**	**-0.020**	0.019	0.019	**0.035**	**0.035**
	(-2.08)	**(-2.08)**	(1.73)	(1.73)	**(4.02)**	**(4.02)**
Dummy(Periphery)	-0.019	-0.019	0.017	0.017	**0.034**	**0.034**
	(-1.86)	(-1.86)	(1.72)	(1.72)	**(3.84)**	**(3.84)**

Table 9.
Testing Existing Anomalies: the Institutional Ownership, Turnover, and Illiquidity Effects (Continued)

Panel B: Subsets sorted on TURN (NYSE/AMEX)

	(1) RET(t) H-TURN	(2) RET(t) H-TURN	(3) RET(t) M-TURN	(4) RET(t) M-TURN	(5) RET(t) L-TURN	(6) RET(t) L-TURN
SIZE	**-0.001**	**-0.001**	**-0.001**	**-0.001**	0.000	0.000
	(-2.82)	**(-2.82)**	**(-1.99)**	**(-1.99)**	(-0.49)	(-0.49)
BM	0.002	0.002	**0.002**	**0.002**	**0.002**	**0.002**
	(1.65)	(1.65)	**(2.58)**	**(2.58)**	**(3.54)**	**(3.54)**
MOM	**0.005**	**0.005**	**0.006**	**0.006**	**0.009**	**0.009**
	(2.26)	**(2.26)**	**(2.94)**	**(2.94)**	**(4.57)**	**(4.57)**
REV	**-0.029**	**-0.029**	**-0.067**	**-0.067**	**-0.084**	**-0.084**
	(-5.76)	**(-5.76)**	**(-10.96)**	**(-10.96)**	**(-14.48)**	**(-14.48)**
RET(Related,1)×Dummy(C)	**0.156**		**0.198**		**0.211**	
	(3.06)		**(5.37)**		**(6.29)**	
RET(Related,1)×Dummy(M)	**0.140**		**0.163**		**0.181**	
	(3.11)		**(4.89)**		**(5.11)**	
RET(Related,1)×Dummy(P)	**0.078**		**0.098**		**0.111**	
	(2.04)		**(3.44)**		**(3.23)**	
RET(Related,1)×Dummy(C)		0.078		**0.101**		**0.100**
		(1.43)		**(2.11)**		**(2.09)**
RET(Related,1)×Dummy(M)		0.062		0.066		0.069
		(1.20)		(1.50)		(1.52)
RET(Related,1)		**0.078**		**0.098**		**0.111**
		(2.04)		**(3.44)**		**(3.23)**
Dummy(Center)	**0.036**	**0.036**	**0.028**	**0.028**	0.011	0.011
	(3.15)	**(3.15)**	**(2.77)**	**(2.77)**	(1.25)	(1.25)
Dummy(Middle)	**0.032**	**0.032**	**0.027**	**0.027**	0.016	0.016
	(2.89)	**(2.89)**	**(2.62)**	**(2.62)**	(1.65)	(1.65)
Dummy(Periphery)	**0.035**	**0.035**	**0.027**	**0.027**	0.012	0.012
	(3.07)	**(3.07)**	**(2.74)**	**(2.74)**	(1.35)	(1.35)

Table 9.
Testing Existing Anomalies: the Institutional Ownership, Turnover, and Illiquidity Effects (Continued)

Panel C: Subsets sorted on Amihud Illiquidity (NYSE/AMEX)

	(1) RET(t) H-ILLIQ	(2) RET(t) H-ILLIQ	(3) RET(t) M-ILLIQ	(4) RET(t) M-ILLIQ	(5) RET(t) L-ILLIQ	(6) RET(t) L-ILLIQ
SIZE	-0.001	-0.001	**-0.002**	**-0.002**	-0.001	-0.001
	(-1.95)	(-1.95)	**(-2.58)**	**(-2.58)**	(-1.22)	(-1.22)
BM	**0.002**	**0.002**	0.001	0.001	**0.002**	**0.002**
	(3.74)	**(3.74)**	(1.11)	(1.11)	**(2.24)**	**(2.24)**
MOM	**0.005**	**0.005**	**0.008**	**0.008**	0.004	0.004
	(2.37)	**(2.37)**	**(3.50)**	**(3.50)**	(1.59)	(1.59)
REV	**-0.065**	**-0.065**	**-0.031**	**-0.031**	**-0.027**	**-0.027**
	(-8.93)	**(-8.93)**	**(-5.01)**	**(-5.01)**	**(-4.56)**	**(-4.56)**
RET(Related,1)×Dummy(C)	**0.220**		**0.172**		**0.107**	
	(5.73)		**(4.08)**		**(2.86)**	
RET(Related,1)×Dummy(M)	**0.133**		**0.127**		**0.100**	
	(3.41)		**(3.73)**		**(2.86)**	
RET(Related,1)×Dummy(P)	**0.114**		0.060		0.029	
	(3.06)		(1.63)		(0.98)	
RET(Related,1)×Dummy(C)		**0.105**		**0.112**		0.078
		(2.07)		**(2.14)**		(1.68)
RET(Related,1)×Dummy(M)		0.018		0.067		0.072
		(0.33)		(1.46)		(1.64)
RET(Related,1)		**0.114**		0.060		0.029
		(3.06)		(1.63)		(0.98)
Dummy(Center)	**0.029**	**0.029**	**0.044**	**0.044**	0.021	0.021
	(2.69)	**(2.69)**	**(3.04)**	**(3.04)**	(1.77)	(1.77)
Dummy(Middle)	**0.026**	**0.026**	**0.042**	**0.042**	0.020	0.020
	(2.39)	**(2.39)**	**(2.81)**	**(2.81)**	(1.71)	(1.71)
Dummy(Periphery)	**0.025**	**0.025**	**0.043**	**0.043**	0.022	0.022
	(2.35)	**(2.35)**	**(2.93)**	**(2.93)**	(1.86)	(1.86)

Table 10.
Changes in Institutional Co-ownership

I consider special cases of the following pooled regression: for firm i in quarter q,

$$\Delta \mathrm{IHP}_{i,q} = \alpha_i + \beta_q + \sum_{l \in (C,M,P)} \theta_l^{related} D_{i,l,q} \Delta \mathrm{IHP}_{i,q}^{related} + \varepsilon_{i,q},$$

where $\Delta\mathrm{IHP}_{i,q}$ is the change in the percentage ownership of institutional investors in firm i from quarter $q-1$ to quarter q. l is chosen from $(C, M, P) = (Center, Middle, Periphery)$. $D_{i,l,q}$ is an indicator variable defining the location of firm i in quarter q. For example, if the industry which firm i belongs to is central, $D_{i,C,q}$ is 1 otherwise 0. $\Delta\mathrm{IHP}_{i,q}^{related}$ denotes the aggregate change in the percentage ownership of institutional investors in the customer and supplier industries of firm i from quarter $q-1$ to quarter q. To report t-statistics in parentheses (bold if significant at the 5% level), robust standard errors are computed by double-clustering by firm and year-quarter.

Panel: Changes in Institutional Co-ownership

	(1) ΔIHP(q)	(2) ΔIHP(q)
ΔIHP(Related,q) $\times$ Dummy(Center)	**0.203**	
	(6.18)	
ΔIHP(Related,q) $\times$ Dummy(Middle)	**0.195**	
	(6.06)	
ΔIHP(Related,q) $\times$ Dummy(Periphery)	**0.200**	
	(5.66)	
ΔIHP(Related,q) $\times$ Dummy(Center)		0.004
		(0.19)
ΔIHP(Related,q) $\times$ Dummy(Middle)		-0.004
		(-0.21)
ΔIHP(Related,q)		**0.205**
		(6.02)
Dummy(Center)	0.056	0.068
	(1.29)	(1.58)
Dummy(Middle)	0.055	0.064
	(1.43)	(1.68)
Constant	-0.073	-0.075
	(-0.83)	(-0.86)
Firm/Time-fixed Effects	Yes	Yes
Clustered S.E.	Firm/Year-Quarter	Firm/Year-Quarter

Table 11.
Small Sample Properties of Risk Premium Estimates: Simulation Evidence

This table presents the small sample properties of risk premium estimates based on the Instrumental Variable (IV) estimator. Panel A presents the results for the CAPM and Panel B presents the results for the Fama-French three-factor model. Appendix 3 describes the details of the simulation. *T* is the number of observations in days and *T/2* observations each are used to estimate the independent variable and the instrumental variable for the IV estimator. *N* is the number of stocks in the cross-section. The results are based on 1,000 replications for each *N* and *T*. Ex-ante bias is the difference between the mean estimate and the corresponding true parameter. Ex-ante root-mean-squared error (RMSE) is also reported. Ex-post bias is the difference between the mean estimate and the sample mean of the corresponding simulated risk factor. Ex-post RMSE is also reported. All biases and RMSEs are expressed as percentages of the true parameters.

Panel A: CAPM

Number of Stocks (*N*)	Risk Premium		Sample Length (*T* in days)				
			264	528	792	1320	2640
			Percentage Biases				
1000	Market	Ex-ante	0.95	0.73	-0.04	-0.16	0.11
		Ex-post	0.12	0.50	0.29	-0.05	0.02
2000		Ex-ante	-0.52	0.34	0.15	0.00	-0.03
		Ex-post	0.15	-0.04	0.11	-0.18	0.00
			Percentage RMSEs				
1000	Market	Ex-ante	17.71	11.03	8.56	6.43	4.43
		Ex-post	12.77	7.03	5.01	3.44	2.13
2000		Ex-ante	15.06	9.88	7.97	5.87	4.12
		Ex-post	8.57	4.65	3.55	2.51	1.53

Panel B: Fama-French Three-factor Model

Number of Stocks (*N*)	Risk Premium		Sample Length (*T* in days)				
			264	528	792	1320	2640
			Percentage Biases				
1000	Market	Ex-ante	3.84	0.17	0.10	0.47	0.11
		Ex-post	3.48	0.22	0.11	0.18	0.25
2000		Ex-ante	1.45	0.53	0.25	0.11	-0.02
		Ex-post	1.38	0.07	0.29	0.10	0.08
1000	SMB	Ex-ante	-2.31	0.47	1.22	0.48	0.57
		Ex-post	-1.84	-0.02	1.04	0.47	0.46
2000		Ex-ante	-0.20	0.59	-0.35	-0.26	-0.14

		Ex-post	0.05	0.75	-0.07	0.00	-0.14
1000	HML	Ex-ante	-1.53	0.54	1.56	-0.19	0.49
		Ex-post	-1.36	0.58	1.43	0.02	0.37
2000		Ex-ante	0.55	0.68	0.30	0.21	0.15
		Ex-post	0.62	0.82	0.25	0.20	0.12
			Percentage RMSEs				
1000	Market	Ex-ante	31.72	14.84	10.95	8.52	5.38
		Ex-post	29.68	12.34	8.40	6.02	3.68
2000		Ex-ante	19.78	11.64	9.71	6.91	4.60
		Ex-post	15.55	7.87	6.22	4.00	2.57
1000	SMB	Ex-ante	45.52	29.52	20.28	14.24	8.79
		Ex-post	42.86	27.48	18.68	12.57	7.56
2000		Ex-ante	38.29	20.93	15.37	10.46	6.80
		Ex-post	35.51	17.86	13.56	8.57	5.07
1000	HML	Ex-ante	28.37	18.51	13.61	9.11	5.72
		Ex-post	26.98	17.45	12.83	8.25	5.22
2000		Ex-ante	24.20	14.23	10.35	7.06	4.26
		Ex-post	23.20	13.08	9.26	6.15	3.44

Table 12.
Small Sample Distribution of the Test Statistic using the IV Estimator

This table presents the size of the test of the null hypothesis that the risk premiums equal zero using the *t*-statistic of the corresponding slope coefficients. We estimate the slope coefficients using IV estimator and compute *t*-statistics using Fama-MacBeth standard errors. Panel A presents the results for the CAPM and Panel B presents the results for the Fama-French three-factor model. Appendix 3 describes the details of the simulation experiments. *N* is the number of stocks in the cross-section. The results are based on 1,000 replications for each *N*. The simulations are based on 792 time-series observations (*T* in days) for each stock.

Panel A: CAPM

Number of Stocks (*N*)	Risk Premium	Theoretical Percentiles 1%	2.5%	5%	7.5%	10%
1000	Market	0.009	0.025	0.049	0.075	0.104
2000		0.009	0.024	0.048	0.074	0.097

Panel B: Fama-French Three-factor Model

Number of Stocks (*N*)	Risk Premium	Theoretical Percentiles 1%	2.5%	5%	7.5%	10%
1000	Market	0.010	0.026	0.054	0.072	0.102
2000		0.011	0.024	0.050	0.077	0.099
1000	SMB	0.010	0.026	0.050	0.076	0.097
2000		0.010	0.024	0.054	0.073	0.098
1000	HML	0.012	0.027	0.056	0.072	0.103
2000		0.011	0.024	0.047	0.076	0.096

Table 13.
Summary Statistics

This table presents sample summary statistics (mean, median, standard deviation, first and third quartiles). *N* is the number of stocks per month, *T* is the life span of firm in months. For the first row, the time-series statistics of the number of stocks per month are shown. For the second row, the cross-sectional statistics of the length of firm time-series are presented. For all other rows, the time-series averages of cross-sectional statistics are provided. Market capitalization is defined as price multiplied by the number of shares outstanding. Book-to-market is defined following Davis *et al*. (2000). Excess return is the return in the excess of risk-free rate. Return volatility is defined as the standard deviation of daily returns. The sample period is from January 1956 through December 2012.

	Mean	Median	Standard Deviation	Q1	Q3
Number of Stocks (*N*)	1934	1980	900	1368	2697
Time-series Length (*T*)	176	136	134	76	231
Market Capitalization ($ billion)	1.498	0.191	6.509	0.052	0.751
Book-to-market Ratio	0.904	0.750	0.674	0.463	1.151
Excess Return (%)	0.888	0.044	11.202	-5.391	5.996
Return Volatility (%)	2.711	2.336	1.652	1.628	3.351

Table 14.
Correlations among Factor Sensitivities and Size and Book-to-market Ratios: CAPM and Fama-French Three-Factor Model

This table presents the average cross-sectional correlations among factor sensitivities and size and book-to-market (BM) ratios. Factor sensitivities are estimated for each month using daily returns data from the prior 36 months. Size is the natural logarithm of market capitalization and BM is the book-to-market ratio. Panel A reports the result for the CAPM and Panels B and C report the results for the Fama-French three-factor model. The sample period is from January 1956 to December 2012.

Panel A: CAPM

		Size	BM
Individual Stocks	MKT	-0.18	-0.20
25 Fama-French Size and BM sorted Portfolios	MKT	-0.56	-0.44

Panel B: Fama-French Three-factor Model: Individual Stocks

	MKT	SMB	HML	Size	BM
MKT	1				
SMB	0.35	1			
HML	0.14	0.13	1		
Size	0.15	-0.44	-0.15	1	
BM	-0.12	0.06	0.28	-0.35	1

Panel C: Fama-French Three-factor Model: 25 Size and BM sorted Portfolios

	MKT	SMB	HML	Size	BM
MKT	1				
SMB	-0.08	1			
HML	-0.08	-0.15	1		
Size	0.19	-0.97	-0.01	1	
BM	0.07	-0.03	0.88	-0.08	1

Table 15.
Risk Premium Estimates with Individual Stocks: CAPM and Fama-French Three-factor Model

This table presents factor risk premium estimates using individual stocks as test assets. The units for the slope coefficients are percentages per month. The table reports *t*-statistics in parentheses (bold if significant at the 5% level). The rows titled MKT, SMB and HML are factor risk premium estimates for market, SMB and HML risks, respectively. Size is the natural logarithm of market capitalization and BM is the book-to-market ratio at the end of the previous month. Factor sensitivities for each month are estimated using daily returns data over the previous 36 months. Panels A, B and C report the results using the IV estimator for the second-stage regression for the entire sample period and two subperiods, respectively. The entire sample period is from January 1956 through December 2012. The row titled Avg *N* presents the average number of stocks per month.

Panel A: IV Estimates: 1956 to 2012

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Const	**0.011**	**0.006**	**0.007**	**0.008**	**0.036**	**0.044**	**0.005**	**0.036**
	(7.80)	**(4.28)**	**(3.15)**	**(5.66)**	**(5.07)**	**(6.91)**	**(2.00)**	**(5.77)**
MKT	-0.189			-0.315	0.010			0.113
	(-1.00)			(-1.65)	(0.05)			(0.62)
SMB		0.227		**0.311**		-0.025		-0.077
		(1.52)		**(2.09)**		(-0.18)		(-0.71)
HML			**0.483**	**0.504**			**0.289**	0.259
			(3.24)	**(3.22)**			**(2.11)**	(1.77)
Size					**-0.152**	**-0.188**		**-0.161**
					(-4.31)	**(-6.03)**		**(-5.19)**
BM					**0.163**		**0.330**	**0.134**
					(3.50)		**(5.62)**	**(3.13)**
Avg *N*					1936			

Panel B: IV Estimates: 1956 to 1985

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Const	**0.012**	**0.005**	**0.006**	**0.008**	**0.034**	**0.045**	0.003	**0.037**
	(6.39)	**(2.60)**	**(2.02)**	**(4.25)**	**(3.41)**	**(4.65)**	(1.06)	**(3.94)**
MKT	-0.386			-0.394	-0.163			0.061
	(-1.67)			(-1.55)	(-0.70)			(0.24)
SMB		0.254		0.358		-0.042		-0.082
		(1.39)		(1.75)		(-0.28)		(-0.60)
HML			**0.625**	**0.594**			**0.352**	0.317
			(3.13)	**(2.64)**			**(1.97)**	(1.56)
Size					**-0.144**	**-0.196**		**-0.167**
					(-2.80)	**(-4.16)**		**(-3.52)**
BM					**0.204**		**0.423**	**0.171**
					(3.00)		**(4.68)**	**(2.65)**
Avg *N*				1239				

Panel C: IV Estimates: 1986 to 2012

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Const	**0.009**	**0.008**	**0.008**	**0.008**	**0.037**	**0.044**	0.006	**0.035**
	(4.57)	**(3.62)**	**(2.46)**	**(3.96)**	**(3.77)**	**(5.20)**	(1.77)	**(4.24)**
MKT	0.030			-0.254	0.201			0.301
	(0.10)			(-0.85)	(0.62)			(1.07)
SMB		0.205		0.322		-0.005		-0.155
		(0.84)		(1.37)		(-0.02)		(-0.84)
HML			0.377	0.321			0.277	0.248
			(1.65)	(1.47)			(1.28)	(1.19)
Size					**-0.160**	**-0.177**		**-0.157**
					(-3.32)	**(-4.43)**		**(-3.91)**
BM					0.116		**0.224**	0.092
					(1.84)		**(3.08)**	(1.67)
Avg *N*				2710				

Table 16.

Correlations among Factor Sensitivities and Size and Book-to-market Ratios: Production-based Asset Pricing Model

This table presents the average cross-sectional correlations among investment and ROE factor sensitivities and size and book-to-market (BM) ratios. Factor sensitivities are estimated for each month using daily returns data from the previous 36 months. Size is the natural logarithm of market capitalization. Panel A reports the results for individual stocks and Panel B reports the results for 25 Fama-French Size and BM sorted portfolios. The sample period is from January 1972 to December 2012.

Panel A: Individual Stocks

	MKT	INV	ROE	Size	BM
MKT	1				
INV	0.04	1			
ROE	-0.03	0.33	1		
Size	0.24	-0.05	0.12	1	
BM	-0.17	0.09	-0.07	-0.32	1

Panel B: 25 Fama-French Size and BM sorted portfolios

	MKT	INV	ROE	Size	BM
MKT	1				
INV	-0.70	1			
ROE	-0.69	0.52	1		
Size	-0.44	0.04	0.74	1	
BM	-0.48	0.88	0.29	-0.08	1

Table 17.
Risk Premium Estimates with Individual Stocks: Production-Based Asset Pricing Model

This table presents factor risk premium estimates for market, investment and ROE risk factors proposed by production-Based asset pricing model using individual stocks as test assets. The slope coefficients are reported in percentages per month. The table reports *t*-statistics in parentheses (bold if significant at the 5% level). The rows titled MKT, INV and ROE are risk premium estimates for market, investment and ROE factors, respectively. Size is the natural logarithm of market capitalization and BM is the book-to-market ratio at the end of the previous month. Factor sensitivities for each month are estimated using daily returns data over the previous 36 months. The sample period is from January 1972 through December 2012. The row titled Avg *N* presents the average number of stocks per month.

Panel A: 1972 to 2012

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Const	**0.011**	**0.009**	**0.007**	**0.008**	**0.007**	**0.044**	**0.038**
	(6.36)	**(3.62)**	**(2.91)**	**(4.18)**	**(2.43)**	**(5.47)**	**(4.71)**
MKT	-0.168			-0.356			0.372
	(-0.63)			(-1.44)			(1.53)
INV		**0.375**		0.297	0.247		0.342
		(1.99)		(1.62)	(1.33)		(1.92)
ROE			-0.113	-0.217		-0.030	-0.125
			(-0.49)	(-0.96)		(-0.13)	(-0.63)
Size						**-0.189**	**-0.183**
						(-5.05)	**(-4.53)**
BM					**0.297**		**0.112**
					(4.48)		**(2.35)**
Avg *N*				2431			

Panel B: 1972 to 1992

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Const	**0.011**	**0.008**	0.005	**0.006**	0.005	**0.037**	**0.028**
	(4.01)	**(2.19)**	(1.45)	**(2.13)**	(1.29)	**(2.92)**	**(2.37)**
MKT	-0.105			-0.250			0.129
	(-0.33)			(-0.76)			(0.37)
INV		**0.650**		0.351	0.452		0.233
		(2.46)		(1.22)	(1.77)		(0.86)
ROE			-0.167	-0.043		-0.032	-0.027
			(-0.55)	(-0.14)		(-0.12)	(-0.10)
Size						**-0.166**	**-0.137**
						(-2.76)	**(-2.27)**
BM					**0.271**		0.109
					(2.79)		(1.70)
Avg N				2082			

Panel C: 1993 to 2012

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Const	**0.012**	**0.010**	**0.007**	**0.011**	0.007	**0.048**	**0.045**
	(5.03)	**(2.77)**	**(2.62)**	**(4.04)**	(1.85)	**(4.90)**	**(4.14)**
MKT	-0.229			-0.490			0.363
	(-0.54)			(-1.34)			(1.08)
INV		0.072		0.442	0.034		0.259
		(0.26)		(1.83)	(0.13)		(1.12)
ROE			-0.280	-0.569		-0.110	-0.163
			(-0.80)	(-1.62)		(-0.32)	(-0.53)
Size						**-0.204**	**-0.201**
						(-4.46)	**(-3.90)**
BM					**0.307**		0.101
					(3.51)		(1.47)
Avg N				2768			

Table 18.
Risk Premium Estimates with Individual Stocks: Liquidity-adjusted CAPM

This table presents the results of asset pricing tests of the liquidity-adjusted CAPM (LCAPM) by Acharya and Pedersen (2005) using individual stocks. The regression coefficients are estimated using the instrumental variables approach. The slope coefficients on Amihud illiquidity measure and LMKT beta are expressed in percentages. The table reports *t*-statistics in parentheses (bold if significant at the 5% level). The row titled Avg *N* presents the average number of stocks per month.

	Sample Period					
	1956-2012		1956-1985		1986-2012	
Constant	**0.006**	**0.006**	**0.008**	**0.007**	0.004	0.004
	(4.64)	**(4.07)**	**(4.38)**	**(3.84)**	(1.83)	(1.92)
LMKT Beta	0.140	0.075	-0.086	-0.136	0.462	0.300
	(0.63)	(0.34)	(-0.27)	(-0.44)	(1.47)	(0.97)
Amihud Illiquidity		**0.184**		**0.310**		0.040
		(3.89)		**(3.53)**		(1.91)
Avg *N*	1265		1192		1344	

Table 19.
Instrument Strength

This table presents the average correlations between odd- and even-month estimates of factor loadings under various models. Panel A reports correlation for the CAPM, and MKT denotes market beta. Panel B reports correlations for the Fama-French three-factor model. MKT, SMB and HML denote sensitivities to the market, SMB and HML factors, respectively. Panel C reports the correlations for the production-based asset pricing model and MKT, INV and ROE denote sensitivities to the market, investment and ROE factors, respectively. Panel D reports the correlation for liquidity-adjusted market beta under the LCAPM. The correlation critical value for the weak instruments tests proposed by Nelson and Startz (1990) is .06, based on the smallest number of stocks in the sample in any month. The square root of the odd- and even-month correlation is the correlation between the unobservable "true" factor sensitivities and the corresponding factor sensitivity estimates.

Panel A: CAPM

Sample period	Odd- and even-month correlation	Correlation between "true" and estimated factor sensitivities
	MKT	MKT
1956-2012	0.67	0.82

Panel B: Fama-French Three-factor Model

Sample period	Odd- and even-month correlation			Correlation between "true" and estimated factor sensitivities		
	MKT	SMB	HML	MKT	SMB	HML
1956-2012	0.52	0.44	0.30	0.71	0.66	0.54

Panel C: Production-based Asset Pricing Model

Sample period	Odd- and even-month correlation			Correlation between "true" and estimated factor sensitivities		
	MKT	INV	ROE	MKT	INV	ROE
1972-2012	0.52	0.29	0.26	0.73	0.53	0.51

Panel D: Liquidity-adjusted CAPM

Sample period	Odd- and even-month correlation	Correlation between "true" and estimated factor sensitivities
	LMKT	LMKT

1956-2012	0.58	0.76

Table 20.
Descriptive Statistics

Panel A: Videos Characteristics

Variable	N	Mean	Std Dev	Min	p25	p50	p75	Max
Face	1,101	-3.85%	13.09%	-56.59%	-6.97%	-1.40%	0.86%	38.14%
Negative	1,101	7.23%	10.40%	0.00%	1.33%	3.27%	8.41%	56.59%
Happy	1,101	3.35%	6.73%	0.00%	0.22%	0.92%	3.13%	41.00%
Neutral	1,101	59.50%	28.48%	0.00%	40.55%	65.53%	83.50%	99.52%
Sad	1,101	4.16%	8.01%	0.00%	0.29%	1.34%	4.53%	84.09%
Angry	1,101	2.70%	6.14%	0.00%	0.26%	0.79%	2.10%	62.63%
Surprised	1,101	5.03%	11.44%	0.00%	0.29%	0.99%	3.44%	93.45%
Scared	1,101	0.60%	3.19%	0.00%	0.01%	0.05%	0.21%	71.62%
Disgusted	1,101	1.96%	7.39%	0.00%	0.01%	0.07%	0.42%	89.97%
PosWords	1,101	6.01	5.84	0.00	2.00	4.00	8.00	28.00
NegWords	1,101	3.50	4.23	0.00	1.00	2.00	5.00	22.00

Panel B: Firm Characteristics

Variable	N	Mean	Std Dev	Min	p25	p50	p75	Max
MkCap	1101	26,300	45,900	47	2,520	7,480	30,700	500,000
B2M	1101	0.84	1.35	0.05	0.27	0.49	0.86	9.86
SUE	1096	0.00	0.00	-0.02	0.00	0.00	0.00	0.02
Momem	1101	0.00	0.00	0.00	0.00	0.00	0.00	0.01
PEAD	1101	0.00	0.01	-0.04	0.00	0.00	0.00	0.05
LogAna	1101	2.75	0.58	1.10	2.48	2.83	3.18	3.91
CAR0	1088	0.006	0.033	-0.097	-0.008	0.002	0.015	0.142
CAR1	1075	0.001	0.020	-0.061	-0.009	-0.001	0.008	0.083
CAR180	1101	0.003	0.193	-0.567	-0.102	0.000	0.110	0.565
AbnTurn0	1101	0.011	0.031	-0.012	-0.001	0.002	0.010	0.211
AbnTurn1	1075	0.004	0.013	-0.016	-0.002	0.001	0.005	0.075
AbnTurn2	1055	0.002	0.011	-0.016	-0.002	0.000	0.003	0.064
Earn1(Earn_1)	978	0.040	0.198	-1.470	0.014	0.031	0.052	2.481
Mk-Breakpoint as of 1/2012					687	2,098	5,920	401,383

Panel C: Video Classifications

Video Classification	# of Videos	Percentage
Interview about earnings	191	17%
Interview about the firm	586	53%
Interview about the industry	169	15%
Interview about the economy	133	12%
SUE before the air date >= 0	784	71%
SUE before the air date < 0	317	29%
Interview that has Earn Ann within 20 trading days before the air date	491	45%
Interview that has Earn Ann outside 20 trading days before the air date	610	55%

This table presents various descriptive statistics: video characteristics (Panel A), firm characteristics (Panel B), and video classification (Panel C).

Table 21.
Pair-wise Correlations

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	Face	1	-0.7642*	0.6257*	0.0299	-0.4021*	-0.4404*	-0.0819*	-0.0555*	-0.3325*	0.0464	0.0564*	-0.0109	-0.004	0.045	0.0009	0.0405	-0.0615*
2	Ne	-0.8513*	1	-0.1420*	-0.1218*	0.5988*	0.4656*	-0.0901*	0.1768*	0.3900*	-0.0362	-0.0336	-0.01	0.0572*	-0.0647*	0.0073	-0.0202	0.0336
3	Ha	0.6151*	-0.1121*	1	0.0453	-0.0459	-0.1385*	-0.1516*	0.1273*	-0.0476	0.046	0.0743*	-0.0331	0.0603*	0.0054	0.0042	0.0569*	-0.0284
4	Neutral	0.1186*	-0.2749*	-0.1892*	1	-0.0953*	0.1413*	0.2169*	-0.0337	0.0165	0.0028	-0.0285	-0.0186	-0.0549*	-0.0314	-0.0098	-0.0824*	0.0074
5	Sad	-0.5496*	0.6385*	-0.0843*	-0.2440*	1	0.1009*	-0.1108*	0.2804*	0.0203	0.0432	0.0770*	-0.0414	0.1228*	-0.0264	-0.0268	0.0143	0.0169
6	Angry	-0.4542*	0.5059*	-0.1017*	-0.0650*	0.0195	1	0.0874*	-0.2111*	0.2994*	0.0001	0.0166	-0.0207	0.0550*	-0.0626*	0.0252	-0.0691*	0.0122
7	Surprised	0.022	-0.1084*	-0.1229*	-0.0403	-0.0806*	-0.0462	1	0.1196*	0.0766*	-0.0169	-0.0564*	-0.0011	-0.0975*	0.0132	0.0083	0.0299	0.017
8	Scared	-0.1900*	0.2387*	-0.0033	-0.1059*	0.0835*	-0.048	-0.0015	1	0.0514	-0.0376	-0.0475	-0.0329	-0.0315	-0.0369	0.0219	0.0389	0.0098
9	Disgusted	-0.4831*	0.5706*	-0.0550*	-0.1848*	0.1670*	0.0381	-0.0846*	-0.0232	1	-0.0525	-0.1044*	0.0714*	-0.0847*	-0.0084	0.0798*	-0.0622*	0.1137*
10	PosWords	0.0596*	-0.0760*	0.0002	0.016	-0.0522*	-0.0371	-0.0578*	-0.0486	-0.0544*	1	0.4954*	0.2049*	0.0472	-0.0208	0.0196	0.0194	0.0124
11	NegWords	0.0622*	-0.0651*	0.023	-0.0137	0.0006	-0.0541*	-0.0777*	-0.0638*	-0.0560*	0.5080*	1	0.1895*	0.1461*	-0.0332	-0.0646*	0.0304	-0.0449
12	Size	0.0145	-0.015	0.0049	-0.0226	-0.008	-0.0081	-0.018	-0.0615*	0.0035	0.2081*	0.1733*	1	0.0623*	-0.0236	-0.1245*	-0.0242	0.2429*
13	LogB2M	-0.0012	0.0338	0.0537*	-0.0422	0.049	0.0604*	-0.0756*	-0.0475	-0.043	0.0466	0.1110*	-0.0002	1	-0.0269	-0.0526	0.0813*	-0.3712*
14	Momem	0.0406	-0.0229	0.0433	-0.0257	-0.0297	-0.0345	0.0337	0.0069	-0.0161	0.0072	-0.0005	-0.0713*	-0.0246	1	0.0686*	0.1517*	-0.0458
15	PEAD	-0.0467	0.0457	-0.0195	0.0017	0.0176	0.0231	-0.0177	-0.0257	0.0267	0.028	-0.0174	-0.0859*	-0.0393	0.1754*	1	0.0942*	0.0012
16	SUE	-0.0349	0.0192	-0.0329	-0.0206	0.0116	-0.0066	0.0629*	0.0612*	-0.0152	0.0610*	0.0961*	0.0039	-0.0119	0.1537*	0.0406	1	0.0778*
17	Earn1(Earn_1)	-0.0657*	0.0701*	-0.0199	-0.0243	0.1134*	-0.0192	-0.0291	-0.0008	0.0065	0.0251	-0.0072	0.1316*	-0.2231*	-0.0459	-0.0136	0.1451*	1

This table presents Pearson correlations on the left bottom corner and Spearman correlations on the right top corner.
* indicates significance level equal or smaller than 0.05.

Table 22.
Return Regressions with Interview Variables

	CAR0					CAR1					CAR180				
Face	-0.111					0.175**					0.069				
	(-0.96)					(2.26)					(0.10)				
Ha		-0.011		-0.005	0.004		0.076		0.069	0.069		0.82		0.847	0.887
		(-0.11)		(-0.05)	(0.03)		(1.37)		(1.25)	(1.25)		(1.58)		(1.62)	(1.69)
Ne			0.128	0.127	0.13			-0.161**	-0.158**	-0.168**			0.47	0.516	0.607
			(1.19)	(1.19)	(1.14)			(-2.26)	(-2.22)	(-2.46)			(0.73)	(0.81)	(0.92)
PosWords					0.180*					-0.02					1.173*
					(1.78)					(-0.25)					(2.06)
NegWords					-0.244**					-0.191*					0.21
					(-2.28)					(-1.92)					(0.26)
Intercept	0.006***	0.006***	0.006***	0.006***	0.006***	0.001	0.001	0.001	0.001	0.001	0.003	0.003	0.003	0.003	0.003
	(5.74)	(5.79)	(5.67)	(5.66)	(5.76)	(1.51)	(1.60)	(1.53)	(1.51)	(1.67)	(0.39)	(0.40)	(0.39)	(0.40)	(0.41)
N	1088	1088	1088	1088	1088	1075	1075	1075	1075	1075	1101	1101	1101	1101	1101
Adj R^2	-0.002	-0.003	-0.001	-0.002	0.000	0.003	-0.001	0.003	0.003	0.011	0.007	0.009	0.008	0.009	0.011

This table presents the results of CAR regressions with interview variables and CEO face fixed effects. All regressions are clustered by year-quarter. The slope coefficients are expressed in percentages. T-stats are presented in parentheses. ∗∗∗, ∗∗, ∗: significant at 0.01, 0.05, and 0.10 level, respectively.

Table 23.
Return Regressions with Control Variables

	CAR0					CAR1					CAR180				
Face	-0.086					0.190**					0.052				
	(-0.74)					(2.35)					(0.08)				
Ha		0.017		0.023	0.032		0.095		0.088	0.086		0.866*		0.893*	0.937*
		(0.17)		(0.23)	(0.32)		(1.54)		(1.43)	(1.43)		(1.85)		(1.90)	(1.97)
Ne			0.115	0.116	0.123			-0.166**	-0.162**	-0.172**			0.522	0.566	0.654
			(1.08)	(1.09)	(1.09)			(-2.27)	(-2.22)	(-2.46)			(0.82)	(0.90)	(1.01)
PosWords					0.224**					-0.02					1.095*
					(2.18)					(-0.23)					(1.80)
NegWords					-0.206					-0.172*					0.298
					(-1.62)					(-1.76)					(0.38)
Size	-0.277***	-0.281***	-0.277***	-0.277***	-0.289***	-0.074	-0.069	-0.073	-0.075	-0.041	0.251	0.236	0.269	0.252	-0.034
	(-3.07)	(-3.11)	(-3.01)	(-3.05)	(-3.58)	(-1.13)	(-1.06)	(-1.10)	(-1.13)	(-0.51)	(0.39)	(0.37)	(0.41)	(0.39)	(-0.05)
LogB2M	-0.126	-0.127	-0.129	-0.131	-0.117	-0.115	-0.119	-0.11	-0.114	-0.093	-0.286	-0.326	-0.3	-0.343	-0.435
	(-1.43)	(-1.44)	(-1.49)	(-1.47)	(-1.43)	(-1.51)	(-1.53)	(-1.43)	(-1.46)	(-1.34)	(-0.34)	(-0.38)	(-0.36)	(-0.40)	(-0.51)
Momem	0.072	0.067	0.072	0.071	0.072	-0.024	-0.017	-0.02	-0.024	-0.023	0.857	0.826	0.879	0.845	0.845
	(0.58)	(0.54)	(0.58)	(0.57)	(0.58)	(-0.31)	(-0.22)	(-0.26)	(-0.31)	(-0.30)	(1.51)	(1.48)	(1.60)	(1.53)	(1.46)
Pead	0.05	0.055	0.049	0.05	0.041	0.067	0.06	0.065	0.067	0.068	0.092	0.107	0.071	0.087	0.036
	(0.42)	(0.45)	(0.41)	(0.42)	(0.35)	(0.83)	(0.72)	(0.82)	(0.84)	(0.86)	(0.09)	(0.11)	(0.07)	(0.09)	(0.04)
SUE	0.031	0.036	0.031	0.032	0.039	-0.019	-0.024	-0.023	-0.019	-0.003	0.897**	0.932**	0.876*	0.913**	0.821*
	(0.20)	(0.24)	(0.21)	(0.21)	(0.24)	(-0.18)	(-0.24)	(-0.22)	(-0.19)	(-0.03)	(2.14)	(2.24)	(2.08)	(2.16)	(1.95)
Intercept	0.006***	0.006***	0.006***	0.006***	0.006***	0.001	0.001	0.001	0.001	0.001*	0.003	0.003	0.003	0.003	0.003
	(5.69)	(5.69)	(5.64)	(5.60)	(5.70)	(1.60)	(1.71)	(1.62)	(1.60)	(1.75)	(0.44)	(0.45)	(0.44)	(0.45)	(0.46)
N	1083	1083	1083	1083	1083	1070	1070	1070	1070	1070	1096	1096	1096	1096	1096
Adj R^2	0.003	0.003	0.004	0.003	0.005	0.005	0.000	0.004	0.004	0.010	0.008	0.010	0.009	0.010	0.012

This table presents the results of CAR regressions with control variables and CEO face fixed effects. All regressions are clustered by year-quarter. The slope coefficients are expressed in percentages. T-stats are presented in parentheses. ∗∗∗, ∗∗, ∗: significant at 0.01, 0.05, and 0.10 level, respectively.

Table 24.
Turnover Regressions

	CAT0					CAT1					CAT2				
Face	0.146					0.140***					0.207***				
	(1.08)					(2.89)					(2.64)				
Ha		0.165		0.162	0.169		0.105***		0.101***	0.104***		0.126**		0.120**	0.127**
		(1.45)		(1.45)	(1.49)		(2.81)		(2.75)	(2.89)		(2.08)		(2.02)	(2.21)
Ne			-0.067	-0.058	-0.051			-0.098*	-0.094*	-0.090*			-0.165**	-0.159**	-0.149**
			(-0.56)	(-0.51)	(-0.45)			(-1.95)	(-1.91)	(-1.90)			(-2.37)	(-2.36)	(-2.38)
PosWords					0.001					0.001					0.001
					(0.87)					(1.23)					(1.33)
NegWords					0.000					0.000					0.001
					(0.14)					(-0.02)					(0.46)
Size	-0.797***	-0.789***	-0.792***	-0.794***	-0.824***	-0.265***	-0.256***	-0.262***	-0.264***	-0.276***	-0.349***	-0.337***	-0.345***	-0.348***	-0.383***
	(-4.38)	(-4.51)	(-4.33)	(-4.38)	(-4.46)	(-3.99)	(-4.00)	(-3.88)	(-3.95)	(-3.74)	(-3.33)	(-3.35)	(-3.24)	(-3.32)	(-3.42)
LogB2M	-0.157	-0.165	-0.156	-0.163	-0.172	-0.044	-0.049	-0.042	-0.046	-0.049	-0.029	-0.033	-0.024	-0.03	-0.041
	(-1.33)	(-1.38)	(-1.31)	(-1.37)	(-1.47)	(-0.81)	(-0.88)	(-0.77)	(-0.83)	(-0.90)	(-0.33)	(-0.38)	(-0.27)	(-0.34)	(-0.47)
SUE	0.016	0.016	0.011	0.017	0.008	-0.042	-0.043	-0.045	-0.041	-0.044	-0.032	-0.034	-0.034	-0.032	-0.042
	(0.08)	(0.08)	(0.05)	(0.08)	(0.04)	(-0.42)	(-0.45)	(-0.46)	(-0.42)	(-0.45)	(-0.22)	(-0.24)	(-0.23)	(-0.22)	(-0.29)
LogAna	0.296**	0.288**	0.292**	0.292**	0.281**	0.076*	0.07	0.075	0.076*	0.07	0.085	0.073	0.083	0.084	0.077
	(2.34)	(2.39)	(2.29)	(2.38)	(2.32)	(1.65)	(1.52)	(1.56)	(1.65)	(1.51)	(1.24)	(1.09)	(1.16)	(1.24)	(1.13)
LogAnaStd	0.006	0.000	0.007	0.002	-0.007	-0.016	-0.021	-0.014	-0.017	-0.02	0.021	0.016	0.024	0.02	0.008
	(0.12)	(-0.01)	(0.13)	(0.03)	(-0.10)	(-0.70)	(-0.89)	(-0.64)	(-0.76)	(-0.77)	(0.48)	(0.37)	(0.53)	(0.47)	(0.16)
Intercept	-0.003	-0.003	-0.003	-0.003	-0.003	-0.003***	-0.003***	-0.003***	-0.003***	-0.004***	-0.004**	-0.004*	-0.004**	-0.004**	-0.005**
	(-1.22)	(-1.07)	(-1.23)	(-1.10)	(-1.28)	(-3.00)	(-2.76)	(-3.11)	(-2.89)	(-2.74)	(-2.04)	(-1.86)	(-2.18)	(-2.01)	(-2.11)
N	1096	1096	1096	1096	1096	1070	1070	1070	1070	1070	1024	1024	1024	1024	1024
Adj R^2	0.061	0.062	0.060	0.061	0.062	0.053	0.050	0.049	0.053	0.054	0.041	0.036	0.038	0.040	0.044

This table presents the results of cumulative abnormal turnover regressions with CEO face fixed effects and year fixed effects. The standard errors are clustered by firm and year-quarter. The slope coefficients are expressed in percentages. T-stats are presented in parentheses. ∗∗∗, ∗∗, ∗: significant at 0.01, 0.05, and 0.10 level, respectively.

Table 25.
Earnings Regressions

	Earn1				
Face	0.092				
	(1.21)				
Ha		-0.002		-0.003	-0.003
		(-0.21)		(-0.31)	(-0.32)
Ne			-0.017**	-0.017**	-0.018**
			(-2.06)	(-2.11)	(-2.13)
PosWords					0.000
					(-0.88)
NegWords					-0.008*
					(-1.69)
Earn_1	0.108**	0.108**	0.109**	0.109**	0.109**
	(2.30)	(2.26)	(2.33)	(2.32)	(2.32)
Size	0.005	0.006	0.005	0.005	0.008
	(0.80)	(0.88)	(0.78)	(0.77)	(1.12)
LogB2M	-0.015	-0.015	-0.015	-0.015	-0.014
	(-1.52)	(-1.50)	(-1.51)	(-1.49)	(-1.42)
Momem	-0.003	-0.002	-0.003	-0.003	-0.003
	(-0.29)	(-0.25)	(-0.31)	(-0.31)	(-0.31)
Intercept	0.043***	0.033***	0.042***	0.041***	0.041***
	(3.43)	(4.26)	(5.14)	(4.35)	(4.33)
N	978.00	978.00	978.00	978.00	978.00
Adj R^2	0.329	0.326	0.332	0.331	0.332

This table presents the results of future earnings regressions. Earnt1 is one-quarter-ahead earnings and it is computed as earnings before the extraordinary item (IBQ in COMPUSTAT) scaled by the book value. The regressions include CEO face fixed effects and year-quarter fixed effects with robust standard errors clustered by industry. The slope coefficients are expressed in percentages. T-stats are presented in parentheses. ***, **, *: significant at 0.01, 0.05, and 0.10 level, respectively.

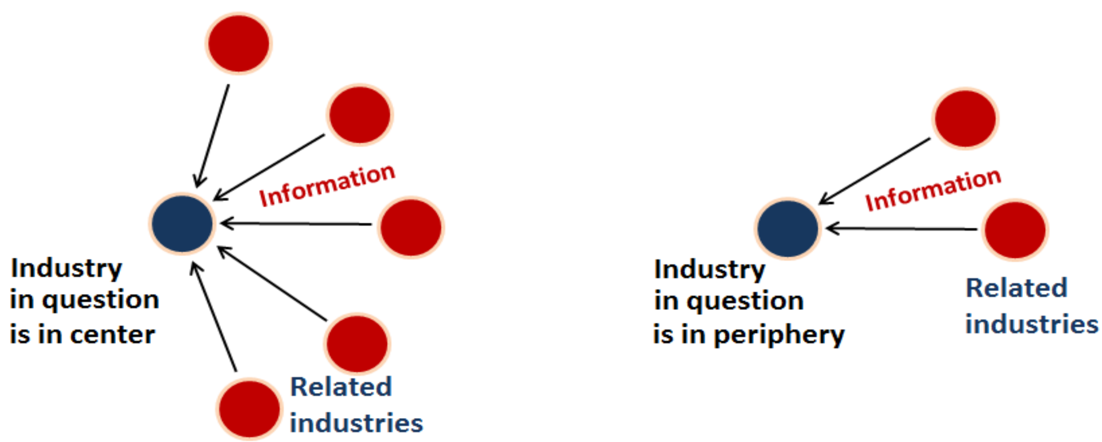
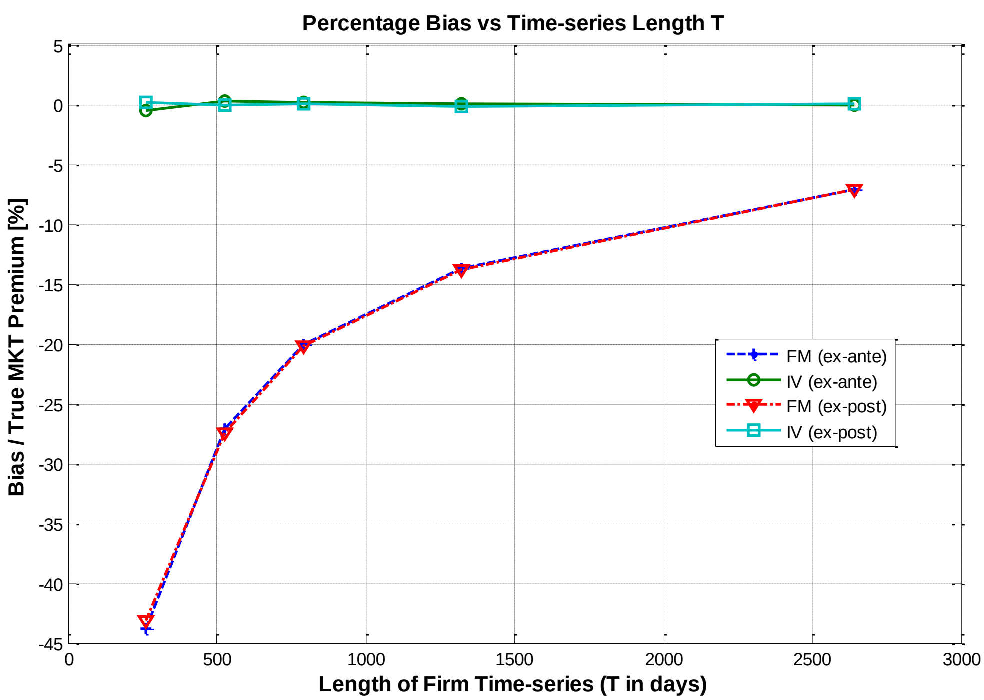


Figure 1. Directions of Information Flow, Return Predictability, and the Predictability of Earnings Forecast Revisions when the Industry in Question is Located at Different Locations. In the left (right) diagram, the industry in question is located in the center (periphery) of the industry network and related industries can be located anywhere in the network. In both diagrams, arrows indicate the directions of information flow, return predictability, and the predictability of earnings forecast revisions to the industry in question from its related industries. When the industry in question is located in the center (periphery), it is connected to more (less) related industries.

Figure 2: Percentage Bias and RMSE versus Time-series Length *T*

This figure presents the ex-ante and ex-post biases (Panel A) and RMSEs (Panel B) in the market risk premium estimates as percentages of true market risk premium in simulations when risk premiums are estimated using Fama-MacBeth (FM) and the Instrumental Variable (IV) estimators. The simulation uses a risk-free rate of 0.9996%, a market risk premium of 5.8008% per annum, and 2000 individual stocks. Appendix 3 describes the details of the simulation experiments. *T*, the number of time-series observations in days, is allowed to vary across simulations and is plotted on the horizontal axis. The results are based on 1,000 replications for each simulated model and each *T*.

Panel A: Ex-ante and Ex-post Percentage Biases



Panel B: Ex-ante and Ex-post Percentage RMSEs

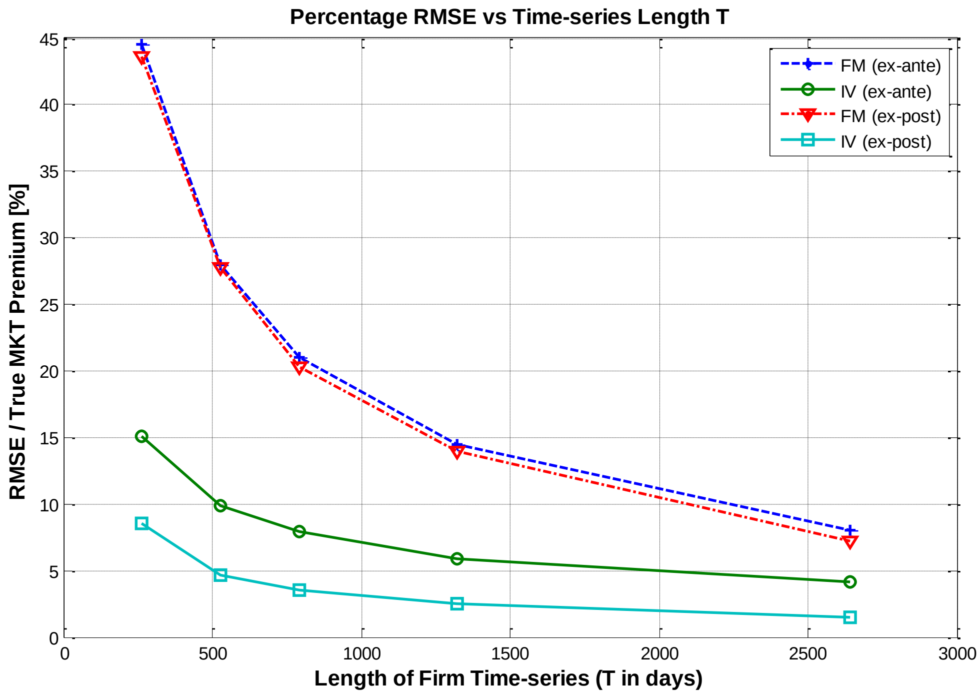
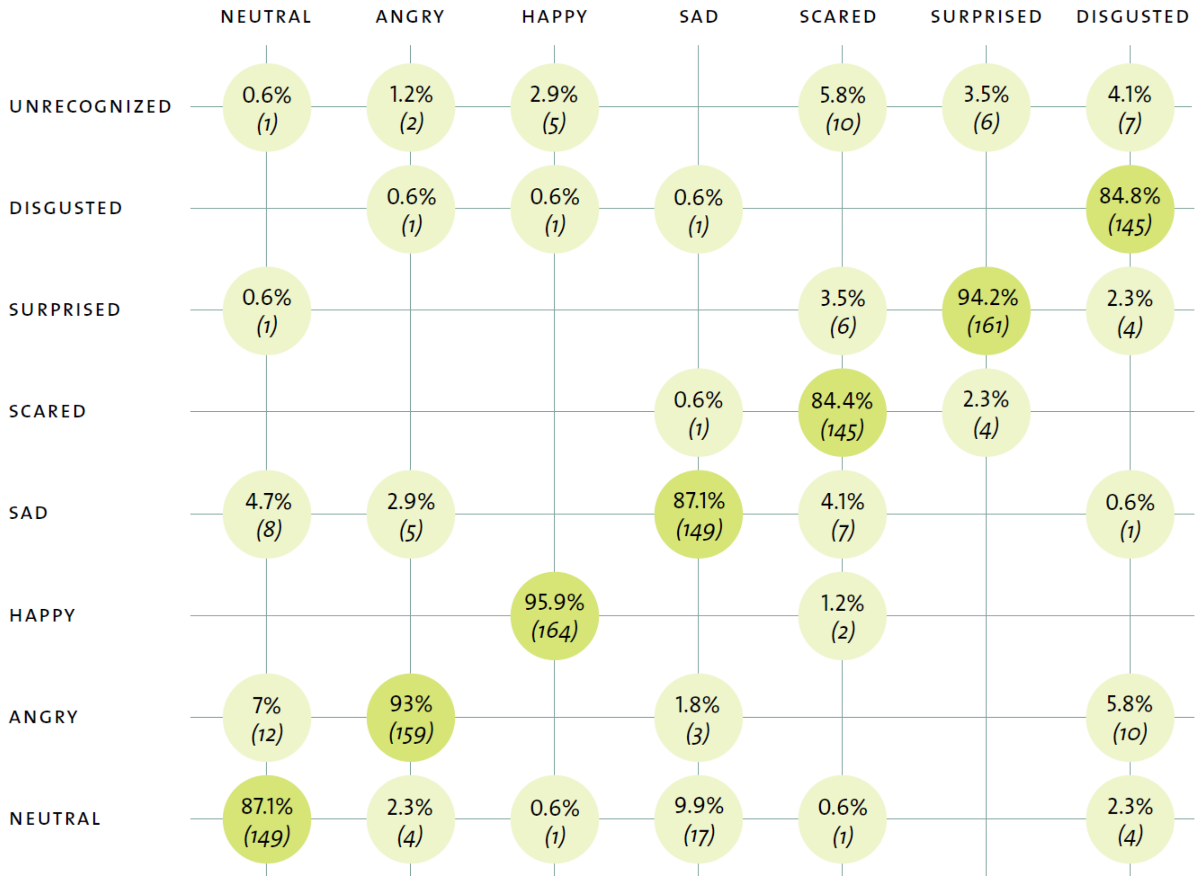


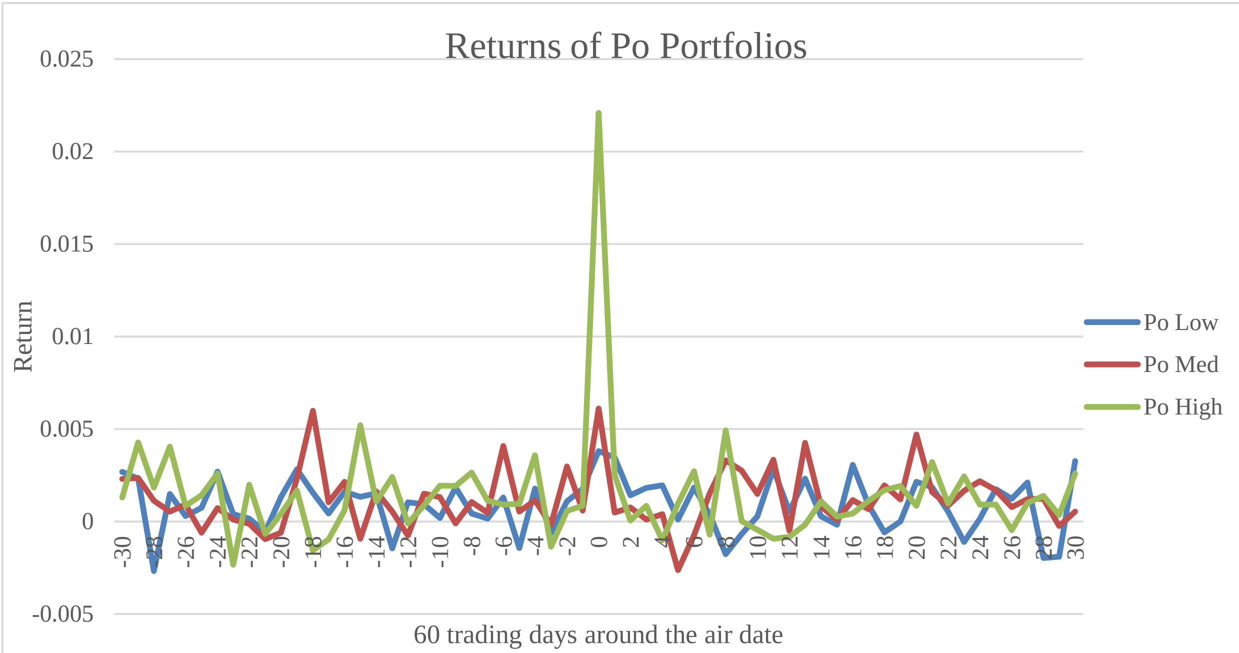
Figure 3: Reliability of FaceReader Software

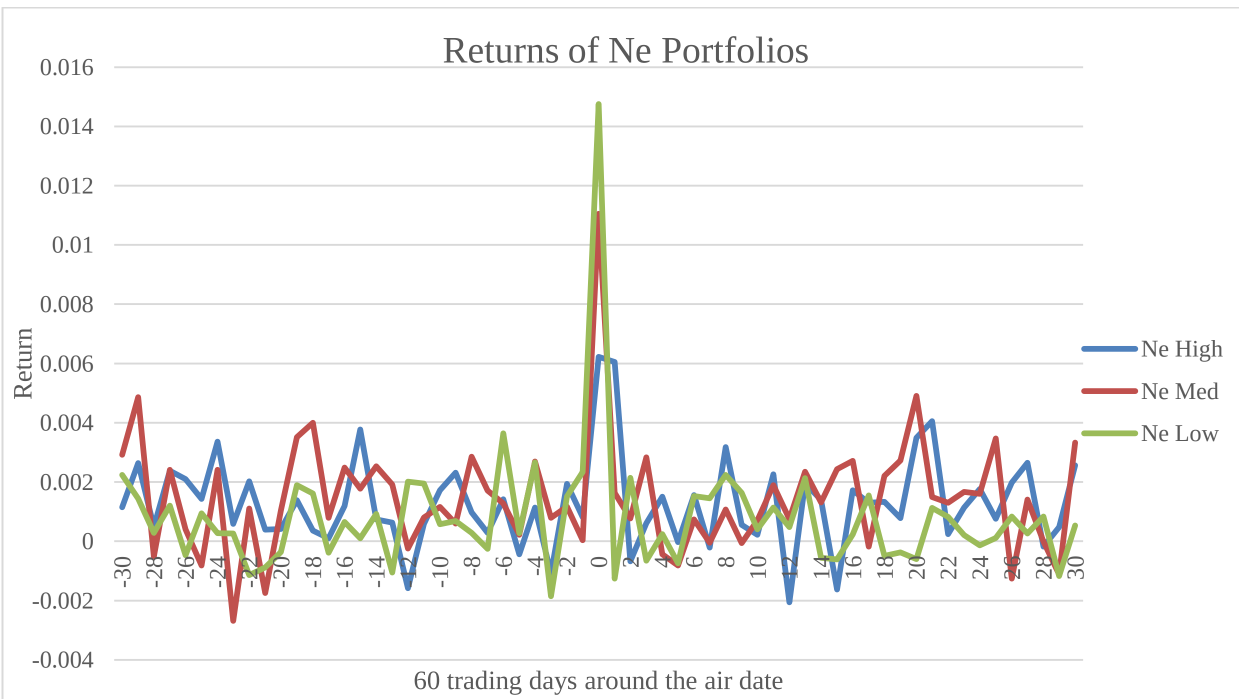


Source: Langner et al. (2010)

This figure presents the proportion of agreement between the facial expressions scored manually by the annotators of the Radbound Faces Database (horizontally) and the facial expressions scored by FaceReader (vertically).

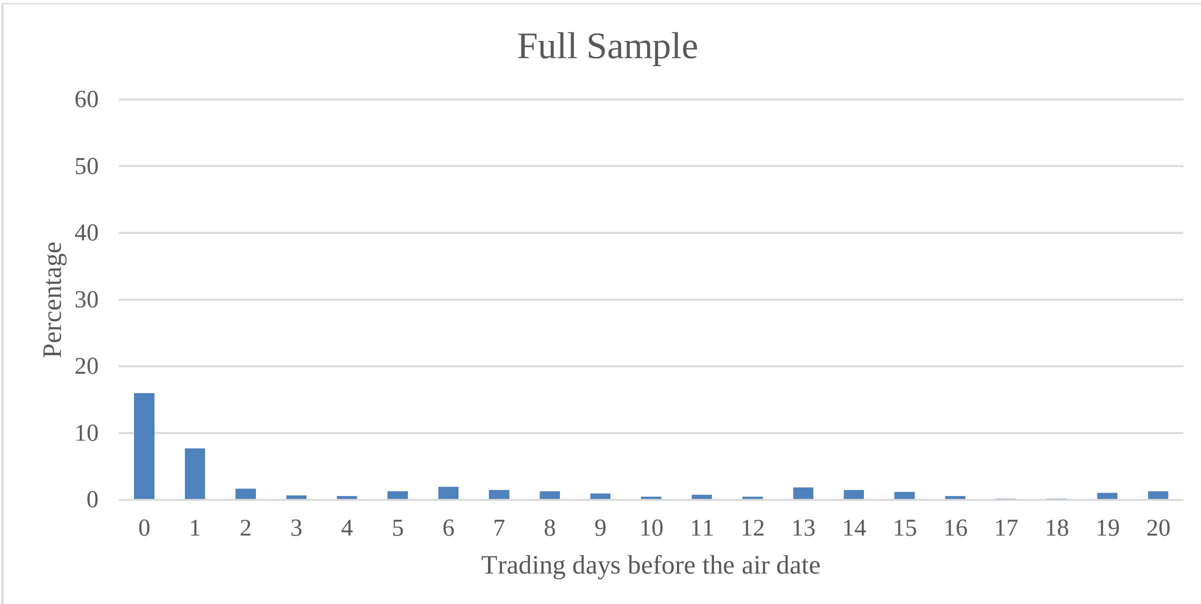
Figure 4: Returns of Tercile Portfolios sorted on Po and Ne





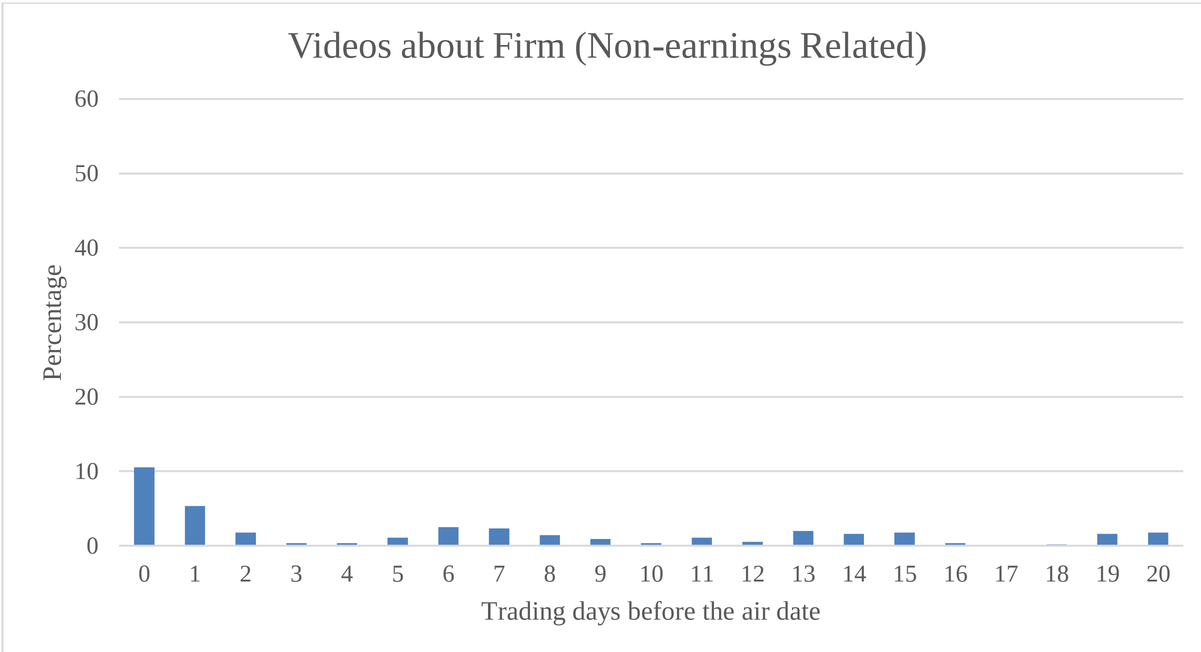
This figure presents the returns of tercile portfolios sorted on positive (top panel) and negative (bottom panel) facial expression scores, i.e., Po and Ne, respectively.

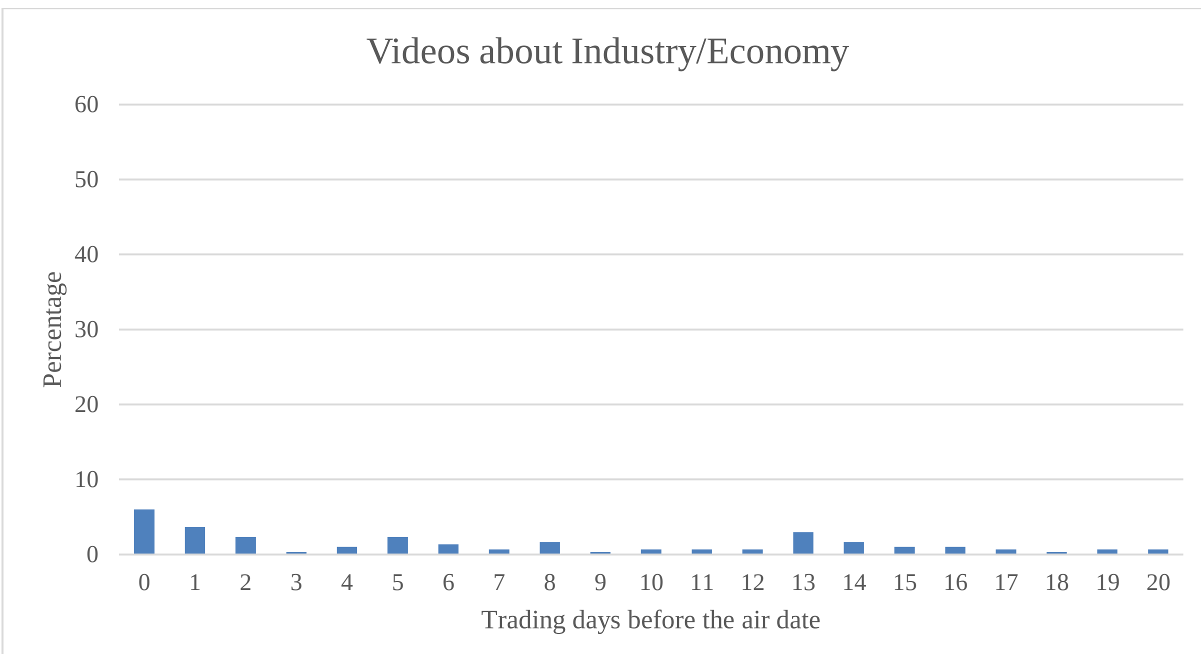
Figure 5: Percentages of Videos that have Earnings Announcements *d* trading days before CEOs' Interviews



This figure presents the percentages of videos that have earnings announcements *d* (=1,…20) trading days before CEOs' interviews. The top plot is based on all available videos in the sample universe and the subsequent three plots are based on three categories of interview topics: earnings-related, firm-related (non-earnings-related), and industry/ economy-related interviews.

Figure 5 ***(continued)***





References

D. Acemoglu, V. M. Carvalho, A. Ozdaglar, and A. Tahbaz-Salehi, "The Network Origins of Aggregate Fluctuations", *Econometrica*, 80(5):1977-2016, Sept. 2012.

V. V. Acharya and L. H. Pedersen, "Asset Pricing with Liquidity Risk", *Journal of Financial Economics* 77, 375-410, Aug. 2005.

K. R. Ahern, "Network Centrality and the Cross-section of Stock Returns", SSRN Scholarly Paper, Dec. 2013.

K. R. Ahern and J. Harford, "The Importance of Industry Links in Merger Waves", *The Journal of Finance*, 69(2):527-576, Apr. 2014.

Y. Ait-Sahalia, J.A. Parker, and M. Yogo, "Luxury Goods and the Equity Premium", *The Journal of Finance* 59, 2959-3004, 2004.

Y. Amihud, "Illiquidity and Stock Returns: Cross-section and Time-series Effects", *Journal of Financial Markets* 5, 31-56, Jan. 2002.

Y. Amihud and H. Mendelson, "Asset Pricing and the Bid-ask Spread", *Journal of Financial Economics* 17, 223-249, Dec. 1986.

A. Ang, J. Liu, and K. Schwarz, "Using Stocks or Portfolios in Tests of Factor Models", Working paper, Columbia University, 2010.

F. Anjos and C. Fracassi, "Shopping for Information? Diversification and the Network of Industries", Technical report, Social Science Research Network, June 2014.

D. Aobdia, J. Caskey, and N. B. Ozel, "Inter-industry Network Structure and the Cross-predictability of Earnings and Stock Returns", *Review of Accounting Studies*, pages 1-34, Apr. 2014.

S. G. Badrinath, J. R. Kale, and T. H. Noe, "Of Shepherds, Sheep, and the Cross-autocorrelations in Equity Returns", *Review of Financial Studies*, 8(2):401-430, Apr. 1995.

R.R. Bahadur and R.R. Rao, "On Deviations of the Sample Mean", *The Annals of Mathematical Statistics* 31(4):1015-1027, Dec. 1960.

J. Bai and S. Ng, "Determining the Number of Factors in Approximate Factor Models", *Econometrica*, 70(1):191-221, Jan. 2002.

T. G. Bali, L. Peng, Y. Shen, and Y. Tang, "Liquidity Shocks and Stock Market Reactions", *Review of Financial Studies*, 27(5):1434-1485, May 2014.

M. J. Becker and S. Thomas, "The Spillover Effects of Changes in Industry Concentration", Unpublished Working paper, University of Pittsburgh, 2010.

K. Benţa, K. Hans, X. Ulises, U. Marten, C. Marcel, H. Amalia, and C. Octavian, "Evaluation of a System for Real Time Valence Assessment of Spontaneous Facial Expressions", In *Distributed Environments Adaptability, Semantics and Security Issues International Romanian-French Workshop*, 2009.

G. Bijlstra and R. Dotsch, “FaceReader 4 Emotion Classification Performance on Images from the Radboud Faces Database”, *Unpublished manuscript, Department of Social and Cultural Psychology, Radboud University Nijmegen*, 2011.

Bishop, Christopher M., “Neural Networks for Pattern Recognition”, Clarendon Press. (1995): 5.

F. Black, M.C. Jensen, and M. Scholes, “The Capital Asset Pricing Model: Some Empirical Tests”, in M. C. Jensen, ed., *Studies in the Theory of Capital Markets*, 79–121, 1972.

P. Bonacich, “Factoring and Weighting Approaches to Status Scores and Clique Identification”, *Journal of Mathematical Sociology*, 2(1):113-120, Jan. 1972.

M. Bradley and P. Lang, “Measuring Emotion: Behavior, Feeling and Physiology”, *In Lane R., Nadel L. (Eds.), Cognitive Neuroscience of Emotion* (pp. 242-276). New York, NY: Oxford University Press, 2000.

M. Bradshaw, “How Do Analysts Use Their Earnings Forecasts in Generating Stock Recommendations?” *The Accounting Review* 79 (1): 25–50, 2004.

M. Brennan, T. Chordia, and A. Subrahmanyam, “Alternative Factor Specifications, Security Characteristics and the Cross-Section of Expected Stock Returns”, *Journal of Financial Economics* 49, 345–373, 1998.

M. J. Brennan, N. Jegadeesh, and B. Swaminathan, “Investment Analysis and the Adjustment of Stock Prices to Common Information”, *Review of Financial Studies*, 6(4):799-824, Oct. 1993.

A. Buraschi and P. Porchia, “Dynamic Networks and Asset Pricing”, Social Science Research Network, Dec. 2012.

C. Busso, Z. Deng, S. Yildirim, M. Bulut, C.M. Lee, A. Kazemzadeh, S. Lee, U. Neumann, and S. Narayanan, “Analysis of Emotion Recognition using Facial Expressions, Speech and Multimodal Information”, In *Proceedings of the 6th international conference on Multimodal interfaces*, 205-211, 2004.

J.Y. Campbell and T. Vuolteenaho, “Bad Beta, Good Beta”, *American Economic Review* 94, 1249-1275, 2004.

V. Carvalho and X. Gabaix, “The Great Diversification and Its Undoing”, *American Economic Review*, 103(5):1697-1727, Aug. 2013.

K.C. Chan and N. Chen, “An Unconditional Asset Pricing Test and the Role of Firm Size as an Instrumental Variable for Risk”, *Journal of Finance* 43, 309-325, 1988.

L. Chen, R. Novy-Marx, and L. Zhang, “An Alternative Three-Factor Model”, Working paper, 2011.

L. Chen, G. Zhang, and W. Zhang, “Return Predictability in Corporate Bond Market Along the Supply Chain”, SSRN Scholarly Paper, Apr. 2014.

T. Chordia, A. Goyal, and J. Shanken, “Cross-sectional Asset Pricing with Individual Stocks: Betas versus Characteristics”, Working Paper, Emory University, 2012.

T. Chordia, R. Roll, and A. Subrahmanyam, "Commonality in Liquidity", *Journal of Financial Economics* 56, 3-28, Apr. 2000.

T. Chordia and B. Swaminathan, "Trading Volume and Cross-autocorrelations in Stock Returns", *The Journal of Finance*, 55(2):913-935, Apr. 2000.

J.H. Cochrane, "Production-based Asset Pricing and the Link Between Stock Returns and Economic Fluctuations", *Journal of Finance* 46, 209-237, 1991.

J.H. Cochrane, "A Cross-Sectional Test of an Investment-based Asset Pricing Model", *Journal of Political Economy* 104, 572-621, 1996.

J.H. Cochrane, "Presidential Address: Discount Rates", *Journal of Finance* 66(4), 1047-1108, 2011.

L. Cohen and A. Frazzini, "Economic Links and Predictable Returns", *The Journal of Finance*, 63(4):1977-2011, Aug. 2008.

L. Cohen and D. Lou, "Complicated Firms", *Journal of Financial Economics*, 104(2):383-400, May 2012.

G. Connor and R.A. Korajczyk, "Performance Measurement with the Arbitrage Pricing Theory: A New Framework for Analysis", *Journal of Financial Economics* 15, 373-394, Mar. 1986.

G. Connor and R.A. Korajczyk, "Estimating Pervasive Economic Factors with Missing Observations", Working Paper, 1987.

G. Connor and R.A. Korajczyk, "A Test for the Number of Factors in an Approximate Factor Model", The *Journal of Finance* 48, 1263-1291, 1993.

T.F. Cootes and C.J. Taylor, "Statistical Models of Appearance for Computer Vision", 2004.

K. Daniel and S. Titman, "Evidence on the Characteristics of Cross Sectional Variation in Stock Returns", *Journal of Finance* 52 (1), 1-33, 1997.

K. Daniel and S. Titman, "Testing Factor-Model Explanations of Market Anomalies", *Critical Finance Review* 1, 103-139, 2012.

C. Darwin, P. Ekman, P. Prodger, "The Expression of the Emotions in Man and Animals", 3rd Edition, London: Harper Collins.

J. Davis, E.F. Fama, and K.R. French, "Characteristics, Covariances, and Average Returns: 1929-1997", *Journal of Finance* 55, 389–406, 2000.

E. Dimson, "Risk Measurement When Shares Are Subject to Infrequent Trading", *Journal of Financial Economics* 7, 197-226, 1979.

P. Ekman, "Universals and Cultural Differences in Facial Expression of Emotion", *In J. Cole (Ed.), Nebraska Symposium on Motivation*, vol.19, pp.207-282, 1971.

P. Ekman, "An Argument for Basic Emotions", *Cognition and Emotion*, 6(3-4), 169-200,1992.

P. Ekman and K. Dacher, "Universal Facial Expressions of Emotion", *California Mental Health Research Digest* 8, no. 4: 151-158, 1970.

P. Ekman and W. V. Friesen, “Unmasking the Face: A Guide to Recognizing Emotions from Facial Clues”, *Oxford, England: Prentice-Hal,* 1975.

P. Ekman and W. V. Friesen, “Facial Action Coding System: A Technique for the Measurement of Facial Movement”, *Consulting Psychologists Press,* 1978.

P. Ekman, “Facial Expression and Emotion”, *American psychologist* 48(4): 384-392, 1993.

E.F. Fama and K.R. French, “The Cross-Section of Expected Stock Returns”, *Journal of Finance* 47, 427–465, 1992.

E.F. Fama and K.R. French, “Common Risk Factors in the Returns on Stocks and Bonds”, *Journal of Financial Economics* 33, 3–56, 1993.

E.F. Fama and J. MacBeth, “Risk, Return and Equilibrium: Empirical Tests”, *Journal of Political Economy* 81, 607–636, 1973.

F. D. Foster and S. Viswanathan, “The Effect of Public Information and Competition on Trading Volume and Price Volatility”, *Review of Financial Studies*, 6(1):23-56, Jan. 1993.

A. Goyal, C. Perignon, and C. Villa, “How Common are Common Return Factors across the NYSE and NASDAQ?”, *Journal of Financial Economics* 90, 252-271, 2008.

S. J. Grossman and J. E. Stiglitz, “On the Impossibility of Informationally Efficient Markets”, *The American Economic Review*, 70(3):393-408, June 1980.

C.R. Harvey, Y. Liu, and H. Zhu, “… And the Cross-section of Expected Returns”, Working Paper, Duke University, 2013.

D. Hirshleifer and S. H. Teoh, “Limited Attention, Information Disclosure, and Financial Reporting”, *Journal of Accounting and Economics*, 36:337-386, Dec. 2003.

C. W. Holden and A. Subrahmanyam, “Long-lived Private Information and Imperfect Competition”, *The Journal of Finance*, 47(1):247-270, Mar. 1992.

H. Hong and J. C. Stein, “A Unified Theory of Under-reaction, Momentum Trading, and Over-reaction in Asset Markets”, *The Journal of Finance*, 54(6):2143-2184, Dec. 1999.

H. Hong, J. C. Stein, and J. Yu, “Simple Forecasts and Paradigm Shifts”, *The Journal of Finance*, 62(3):1207-1242, June 2007a.

H. Hong, W. Torous, and R. Valkanov, “Do Industries Lead Stock Markets?”, *Journal of Financial Economics*, 83(2):367-396, Feb. 2007b.

K. Hou, “Industry Information Diffusion and the Lead-lag Effect in Stock Returns”, *Review of Financial Studies*, 20(4):1113-1138, July 2007.

R. Jagannathan, G. Skoulakis, and Z. Wang. “The Analysis of the Cross-section of Security Returns”, *Handbook of Financial Econometrics*, Y. Ait-Sahalia and L. Hansen (Eds.), Vol.2: 73-134, 2009.

R. Jagannathan and Z. Wang, “The Conditional CAPM and the Cross-section of Expected Returns”, *Journal of Finance* 51, 3–53, 1996.

R. Jagannathan and Z. Wang, “An Asymptotic Theory for Estimating Beta-Pricing Models Using Cross-Sectional Regressions”, *Journal of Finance* 53, 1285–1309, 1998.

N. Jegadeesh, “Does Market Risk Really Explain the Size Effect?”, *Journal of Financial and Quantitative Analysis* 27, 337-351, 1992.

M. Jensen and W. Meckling, “ Specific and General Knowledge and Organizational Structure”, *in Lars Werin and Hans Wijkander eds.: Contract Economics*, Blackwell, Oxford, UK, 1992.

B. Kelly, H. Lustig, and S. V. Nieuwerburgh, “Firm Volatility in Granular Networks”, Working paper, National Bureau of Economic Research, Sept. 2013.

D. Kim, “The Errors in the Variables Problem in the Cross-Section of Expected Stock Returns”, *Journal of Finance* 50, 1605–1634, 1995.

D. Kim, “A Reexamination of Firm Size, Book-to-Market, and Earnings Price in the Cross-Section of Expected Stock Returns”, *Journal of Financial and Quantitative Analysis* 32, 463–489, 1997.

R. A. Korajczyk and R. Sadka, “Pricing the Commonality across Alternative Measures of Liquidity”, *Journal of Financial Economics*, 87(1):45-72, Jan. 2008.

M. Lettau and S. Ludvigson, “Resurrecting the (C)CAPM: A Cross-Sectional Test When Risk Premia Are Time-Varying”, *Journal of Political Economy* 109, 1238–1287, 2001.

J. Lewellen, J. Shanken, and S. Nagel, “A Skeptical Appraisal of Asset Pricing Tests”, *Journal of Financial Economics* 96, 175–194, 2010.

F. Li, “The Information Content of Forward-Looking Statements in Corporate Filings—A Naïve Bayesian Machine Learning Approach”, *Journal of Accounting Research* 48 (5): 1049–1102, 2010.

Q. Li, M. Vassalou, and Xing, “Sector Investment Growth Rates and the Cross-Section of Equity Returns”, *Journal of Business* 79, 1637-1665, 2006.

R. Litzenberger and K. Ramaswamy, “The Effect of Personal Taxes and Dividends on Capital Asset Prices: Theory and Empirical Evidence”, *Journal of Financial Economics* 7, 163–196, 1979.

L.X. Liu, T.M. Whited, and L. Zhang, “Investment-based Expected Stock Returns”, *Journal of Political Economy* 117, 1105-1139, 2009.

A. W. Lo and A. C. MacKinlay, “When are Contrarian Profits due to Stock Market Overreaction?”, *Review of Financial Studies*, 3(2):175-205, Apr. 1990.

A. Lo and A.C. MacKinlay, “Data-Snooping Biases in Tests of Financial Asset Pricing Models”, *Review of Financial Studies* 3, 431–468, 1990.

W. Mayew and M. Venkatachalam, “The Power of Voice: Managerial Affective States and Future Firm Performance”, *The Journal of Finance*, 2011.

L. Menzly and O. Ozbas, “Market Segmentation and Cross-predictability of Returns”, *The Journal of Finance*, 65(4):1555-1580, Aug. 2010.

R.C. Merton, "A Simple Model of Capital Market Equilibrium with Incomplete Information", *The Journal of Finance*, 42(3):483-510, July 1987.

R.C. Merton, "An Intertemporal Capital Asset Pricing Model", *Econometrica 41*, 867–887, 1973.

R.C. Merton, "On Estimating the Expected Return on the Market: An Exploratory Investigation", *Journal of Financial Economics* 8, 323–361, 1980.

T. J. Moskowitz and M. Grinblatt, "Do Industries Explain Momentum?", *The Journal of Finance*, 54(4):1249-1290, Aug. 1999.

C.R. Nelson and R. Startz, "The Distribution of the Instrumental Variables Estimator and Its t-Ratio When the Instrument is a Poor One", *Journal of Business* 63, 125-140, 1990.

W.K. Newey and K.D. West, "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix", *Econometrica* 55, 703-708, May 1987.

W.K. Newey and K.D. West, "Automatic Lag Selection in Covariance Matrix Estimation", *Review of Economic Studies* 61, 631-653, Oct. 1994.

Noldus, "Client List FaceReader.", Accessed August 11, 2014.http://www.noldus.com/facereader/clients.

L. Pastor and R.F. Stambaugh, "Liquidity Risk and Expected Stock Returns", *Journal of Political Economy* 111, 642-685, June 2003.

K. Pukthuanthong and R. Roll, "A Protocol for Factor Identification", Working Paper, UCLA, 2014.

J. Shanken, "On the estimation of Beta-Pricing models", *Review of Financial Studies* 5, 1-55, Jan. 1992.

J. Shanken and G. Zhou, "Estimating and Testing Beta-Pricing Models: Alternative Methods and Their Performance in Simulations", *Journal of Financial Economics* 84, 40–86, 2007.

T. Shumway, "The Delisting Bias in CRSP data", *Journal of Finance* 52, 327-340, Mar. 1997.

H. A. Simon, "A Behavioral Model of Rational Choice", *The Quarterly Journal of Economics*, 69(1): 99-118, Feb. 1955

D. Staiger and J.H. Stock, "Instrumental Variables Regression with Weak Instruments", *Econometrica 65*, 557–586, 1997.

V. Terzis, N. Christos, and S. Anastasios, "Measuring Instant Emotions during a Self-assessment Test: the Use of FaceReader", In *Proceedings of the 7th International Conference on Methods and Techniques in Behavioral Research*, p. 18. ACM, 2010.

P. Tetlock, "Giving Content to Investor Sentiment: The Role of Media in the Stock Market", *The Journal of Finance* 62, no. 3: 1139–68, 2007

P. Tetlock, M. Saar-Tsechansky, and S. Macskassy, "More Than Words: Quantifying Language to Measure Firms' Fundamentals", *The Journal of Finance* 63, no. 3: 1437–67, 2008.

C. Trzcinka, “On the Number of Factors in the Arbitrage Pricing Model”, *Journal of Finance* 41, 347-368, 1986.

P. Viola and J. Michael, "Rapid Object Detection Using a Boosted Cascade of Simple Features", In *Computer Vision and Pattern Recognition,* CVPR 2001.

J. Wu and J. R. Birge, “Supply chain network structure and firm returns”, SSRN scholarly paper, Social Science Research Network, Jan. 2014.