Essays on Technological Innovation, Uncertainty and Firm Behavior

by

J. Frank Li

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Business Administration)
in The University of Michigan
2020

Doctoral Committee:

Associate Professor Kyle Handley, Chair
Associate Professor Ying Fan
Professor Jagadeesh Sivadasan
Associate Professor Yue Maggie Zhou

J. Frank Li

jiazeli@umich.edu

ORCID iD: 0000-0001-5191-4900

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisor Prof. Kyle Handley for the continuous support and encouragement of my doctoral study and research. His patience, motivation, and immense knowledge allow me to grow as a researcher. His advice on research as well as on career development have been priceless. I could not have imagined having a better advisor and mentor.

Many thanks also to the rest of my dissertation committee, Prof. Ying Fan, Prof. Jagadeesh Sivadasan, and Prof. Yue Maggie Zhou for their insightful comments and suggestions. I greatly acknowledge those who are thanked in the working paper versions of my dissertation chapters. Sincere thanks also goes to my Ph.D. fellows and friends for stimulating discussions, comfort and all the fun we had.

Last but not the least, I want to express my deepest gratitude to my mother, my father, and my fiancee Xin for their love and spiritual support during trying times and my life in general.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

LIST OF APPENDICES

Appendix

ABSTRACT

Technological innovation is critical to long-run productivity and economic growth. In Schumpeter's "Creative Destruction", firms play a central role in technological innovation, which in return shapes industry structure through the firm entry, exit, and resource reallocation. Uncertainty, on the other hand, has been identified as the culprit of slow economic recovery, reduced investment and hiring, and cautious innovative behavior. This dissertation consists of three studies that explore how technological innovation and uncertainty shape firm conduct, such as growth opportunities, capital investment, merger decisions, as well as labor reallocation.

The first chapter, "Firm performance, Mergers, and the Technological Frontier, A Text-Analysis Approach" studies the interplay between technological innovation, firm performance, and industry dynamics especially M&A activities. To overcome the difficulties in measuring technological innovation, I propose a novel measure of firm-level proximity to the technological frontier (PTF) by exploiting textual data from USPTO patents and SEC EDGAR filings. The measure strongly predicts firm future growth on multiple dimensions including output, investment, employment, sales, total factor productivity, and Tobin's Q. It also positively correlates with firm previous R&D and patenting activities. I find that technology influences firms' activities in the M&A markets and efficiency gain under resource reallocation through a positive assortative matching pattern. In a sample of all possible public mergers, empirical investigation shows that firms with different levels of PTF, regardless of advanced or not, and similar

technology portfolios tend to merge.

The second chapter, "Measuring the Effects of Firm Uncertainty on Investment: New Evidence on One Million Documents", joint with Kyle Handley, measures time-varying uncertainty at the firm level and its impact on corporate investment. We construct a firm-specific company reported uncertainty index (CRUX) by measuring the frequency of variations of the word "uncertainty" relative to the total number of words in the business context of SEC filings. We find sizable delay and caution effects of uncertainty on the aggregate level, firm-level, and establishment-level investment rate, and our results are not driven by aggregate fluctuations or firm-specific characteristics.

We construct a new time-varying measure of *firm-specific* uncertainty from analyzing the text of company reports filed with the U.S. Securities and Exchange Commission. We explore the implications of variation uncertainty with firm-level and aggregate data. We find the new measure is negatively correlated with growth in aggregate investment, GDP, and employment even after controlling for other measures of first moment shocks and aggregate uncertainty. The effect of our firm-level measure on aggregate data is comparable to alternative uncertainty measures such as the VIX. Using firm-level panel data on investment and employment with a rich set of controls, we find our measure of firm-specific uncertainty has reasonably large effects on investment and employment even after controlling for aggregate and industry time-varying shocks. Firm uncertainty shocks (1) reduce investment rates by 0.5% and attenuate the response to positive sales shock by 50% and (2) reduce employment growth rates by 1.4% and responsive to positive sales shocks by 30%. Most of the employment growth reductions operate through diminished gross job creation at new plants and continuing establishments. Moreover, we find firms are less responsive to demand shocks at the firm level and across establishments within a firm, even after controlling for any unobservable firm-year shocks.

The third chapter, "Uncertainty and Firm Labor Reallocation", joint with Kyle Handley, investigates the impact of uncertainty on labor growth and reallocation at the firm level. By matching CRUX constructed in chapter two with U.S. Census microdata, we find that uncertainty reduces firm-level employment growth and firms are less responsive to positive demand shocks. Furthermore, we successfully identify firm-level employment growth on different margins, such as establishment birth, death, acquisition, divestiture, and continuers. The empirical results show that the effect of uncertainty is most salient on the margins with higher adjustment costs, such as birth and acquisition.

CHAPTER I

Firm performance, Mergers, and the Technological Frontier, A Text-Analysis Approach

1.1 Introduction

This paper examines the role of innovation and technology in enhancing economic performance and shaping capital reallocation decisions (especially M&A activities) of U.S. public firms. Economists believe technological innovation is critical to long-run total factor productivity (TFP) growth, economic growth, and standards of living. But since the 2000s, particularly after the financial crisis, low productivity growth seems to have become the new normal of U.S. economy along with slowing rates of innovation (Bloom et al. 2017).

According to Schumpeter's (1942) "Creative Destruction" argument, firms play a central role in technological innovation, which in return shapes industry structure through firm entry, exit, and resource reallocation, such as mergers and acquisitions, to gain efficiency. Schumpeter also defined entrepreneurial activities as "the carrying out of new combinations" such as new products, new production process, opening new markets, and "new organization of any industry, like the creation of a monopoly position (for example through trustification) or the breaking up of a monopoly position".

However, the interplay between technological innovation and industry dynamics, especially M&A activities, has not been fully understood. One major challenge is the difficulty in measuring firm-level innovation and technology adoption.

A large and growing body of literature has investigated the measurement of innovation and the impact of innovation on economic behavior. Early work by Romer (1990) endogenizes technological changes into a model of economic growth. Klette and Kortum (2004), using patents as an innovation measure, study innovation and industry dynamics with firm entry and exit. Acemoglu et al. (2018) further explores R&D, reallocation, and productivity growth with endogenous firm entry and exit. Acemoglu, Zilibotti, and Aghion (2006) tackle distance to innovation frontier and economic growth at the aggregate level. Kogan et al. (2017) estimate the economic value of patents by linking patents to firm-level stock returns data in a long time panel. Works based on traditional measures of technological innovation, such as patents (counts and citations) and R&D expenditures, provide lots of insights, but the drawbacks are also apparent.

Measuring (firm-level) innovation and technology adoption is difficult for several reasons. First, not all innovation is directly observed through measures like patents and R&D expenditure because a firm may choose not to patent, it may not report R&D expenditure for tax reasons, or the firm is simply combining existing innovations in new ways. Second, innovation does not necessarily remain with the individual or firm because it may be sold, licensed, or modified by other firms. Third, only successful innovations, especially those that push technology frontier, are observed, which creates selection bias. Finally, R&D and patent measures are highly skewed towards manufacturing and IT sectors, thus the coverage of other major sectors, like services and retail, is limited.

To overcome these difficulties, in this paper I propose a novel measure of firm level technological innovation that captures both innovation at the technological frontier and technology adoption, which is named after the firm-level proximity to the technological frontier (PTF). I further investigate how technology affects firm-level performance including investment, employment, output and productivity growth, and firm matching patterns in the capital reallocation (M&A) market.

Following a simple assumption that firms discuss cutting-edge and established technologies that are beneficial to their operations when communicating with investors, I construct the measure in the following steps. First, identify technologies from patent data. Innovating entities (individuals, firms, and public sectors) file patents to protect intellectual property rights and the economic return of technological innovations. I take the textual information from all patent documents from the United States Patent and Trademark Office (USPTO) and extract key phrases associated with innovation using computational linguistics techniques. Executed by each year, this approach generates lists of phrases that describe technologies that appear in the patents within the year. The second step is to identify technology frontiers based on the generated lists. I explore the rise and decline of the usage of each phrase by calculating the annual growth rate adjusted by both phrase frequencies and document-wise distribution differences. That is, the phrase about more popular technology should have a higher level of usage under this calculation. The rising phrases (phrases with positive growth rate) represent relatively new technologies (technology at the frontier) than declining phrases since more patents on such technology are filed in the current year than the previous year. Third, to obtain firm-level exposure to the technology frontier, I exploit textual content in corporate annual and quarterly filings from EDGAR database, which covers all public domestic firms' reports to the U.S. Securities and Exchange Commission on a regular basis. Firms are required to disclose business operations, risk concerns, financial information, and managerial discussions and analysis to their investors. It is not uncommon for firms to discuss cutting-edge and/or mature technologies in their filings that are critical to their business. Converting patent data and EDGAR corporate filing data into vectors allows me to compute the similarity of the technology vector and the firm-level corporate filing vector. The computed similarity is the measure of proximity of technology frontier (PTF).

Compared with traditional measures of firm-level innovation, such as (citation-weighted) patent counts and R&D expenditures, PTF differs in two major ways. One is that PTF extends the coverage of firms to the entire space of publicly traded firms, while only a small subset of firms can be matched with patents and are reporting their R&D expenditures. A firm can be assigned a PTF measure as long as it is required to submit annual and quarterly filings. The second difference is that PTF allows technologies with different proximity to the frontier to have different weights. PTF captures both mature and cutting-edge technologies and grants more recent technologies a higher score. Furthermore, since PTF is a time-varying firm-level measure, we are allowed to exercise cross-sectional and time-series comparisons. To test the validity of PTF, I correlate it with alternative measures of innovation and firm performance. I find that PTF is strongly positively correlated with firms' previous R&D investment and patenting activities measured by economic value (Kogan et. al 2017). Firms with higher R&D expenditure and more profitable patenting activities tend to have higher PTF measures in the following year with or without a set of control variables. PTF also predicts better firm performance on multiple dimensions including output growth, investment rate, employment, sales, and TFP (OLS and Olley-Pakes) growth, and Tobin's Q in both short-run and long-run. Firms with higher PTF tend to grow faster in all the dimensions mentioned above, even for firms without patenting behavior. The results are robust to firm-level size and idiosyncratic uncertainty controls, which suggests that PTF provides valuable information about a firm's technology capability and it is consistent with classical growth theory where technological innovation plays a critical role.

Besides firm growth, technological innovation also shapes resource reallocation patterns in the market. Firms participate in merger and acquisition activities to restructure their business and gain efficiency through better matching of technology space, management, and market-driven characteristics. This paper focuses on the impact of the impact of technology on firm's activities in the M&A market. I draw M&A data from SDC platinum and match them with COMPUSTAT and PTF to investigate how different levels of technology (PTF) and how similarity in terms of technology space affect firm's merger and acquisition decisions. Traditional wisdom offers two explanations of M&A patterns: high buys low (Jovanovic and Rousseau 2002, 2008) and similar buys similar (Rhodes-Kropf and Robinson 2008) The empirical results with the new measure suggest that firms with different levels of PTF, that is, firms with different level of technological advancement, tend to merge while the impact will be mitigated by larger differences in firm size and smaller differences of profitability. Based on firm-level PTF, a firm pair-wise technology similarity can be constructed using the same natural language processing method. By taking a Cartesian product, I identify all possible merger pairs and find that firms with similar technologies, regardless of advanced or not, tend to merge. These results offer additional evidence on how technology interacts with the firm restructuring market.

This paper contributes to the literature threefold. First, it contributes to the literature of measuring technological innovation and technology adoption and its impact on firm level activity. I provide a novel way to measure technology and extend the coverage to all publicly traded firms. Early works such as Griliches (1990), Klette and Kortum (2004), Hall, Jaffe, and Trajtenberg (2005) mainly focus on patents or R&D expenditures while Balasubramanian and Sivadasan (2011) suggest that less than 5% of all U.S. firms (public and private) ever file a patent. Even within publicly traded firms, compared with other common measures of innovation such as patents, the PTF measure extends the coverage of number of firms by at least five times among publicly traded firms. To show this, I match the PTF measure with COMPUSTAT and two carefully constructed patent-firm data sources: Kogan, Papaniokolaou, Seru, and Stoffman (2017) and Autor, Dorn, Hanson, Pisano, and Shu (2019). I plot the ratio of patenting firms to all firms in different margins in Figure 1.1. It shows that when considering patent grant year in Figure 1.1(a), patenting firms account for about 15% of the entire sample. Note that they are larger firms in terms of employment, sales, total assets, and market capitalization as patenting firms have larger weights in size. Patenting firms account for up to 50% of total market capitalization. I find a similar pattern when considering application year in Figure 1.1(b). Thus PTF significantly extends the coverage, which allows me to conduct the Cartesian product when constructing the M&A sample.

Second, the paper uses a simple yet useful method to contribute to the means of textual analysis in economics. With the rise of machine learning and big data, text analysis (or natural language processing techniques) has become more popular in economic research. Loughran and McDonald (2011) cleans SEC corporate reports and uses it to study firm IPO behavior. Hoberg and Phillips (2016) constructs a firm-to-firm product market competition measure based on processing text in 10-Ks. Baker, Bloom, and Davis (2016) mines newspapers to construct an economic policy uncertainty index. Hassan et al. (2018) and Handley and Li (2019) generate firm-level uncertainty measures from conference calls and EDGAR 10-Ks and 10-Qs. Hansen, McMahon, and Prat (2018) mine Federal Open Market Committee transcripts to study monetary policy. Similarly, Bellsttam, Bhagat, and Cookson (2019) apply LDA (latent Dirichlet Allocation) to measure innovation using conference calls. PTF differs from LDA-generated measures as it measures proximity to the technology frontier rather than the intensity of innovation, and it covers more firms since not all public firms hold conference calls. Third, the paper seeks to shed light on the literature on the determinants of mergers. There are two streams of argument among many others. Jovanovic and Rousseau (2002, 2008) argue a high Q buys low Q pattern, while Rhodes-Kropf and Robinson (2008) suggest a similar buys similar pattern. Our empirical results suggest that the determinants can be more subtle when controlling for alternative firm characteristics. Firms with different levels of technology are more likely to merge, and this effect can be attenuated by size and profitability difference. Other papers including Hoberg and Phillips (2010a, 2010b), Phillips and Zhdanov (2013), and Bena and Li (2014) examine the effect of innovation on mergers, which provide a potential future research agenda for PTF measures.

The rest of the paper is organized as follows. In Section 1.2, I describe the details of the construction of proximity to technology frontier (PTF) at the firm level. Section 1.3 investigates the impact of PTF on firm performance and its correlation with firm's previous innovative activities. The interplay between merger and acquisition activities and technology and other firm characteristics is explored in Section 1.4. Section 1.5 concludes.

1.2 Construction of Proximity to Technology Frontier (PTF)

The main objective of this section is to obtain firm level measure of proximity to technology to frontier (PTF) with both cross-sectional and time-series variation. The PTF measure is constructed by combining textual information extracted from USPTO patents and EDGAR 10-K documents. This section consists of discussing the data sources and steps to create the measure in detail.

1.2.1 Data Description

To construct the proximity to technology frontier (PTF) at the firm level, I draw text data from two main sources: USPTO patent documents and SEC EDGAR 10-K documents. Both USPTO patent and EDGAR 10-K documents are standardized and consistently collected by government agencies. The U.S. Patent and Trademark Office (USPTO) has recorded all successfully applied patents and converted them into downloadable electronic versions since 1976. Vintage archives are available in a much longer time series, starting at least in the 1800s. However, this paper will focus on more recent patent files due to the time constraint of EDGAR data.

To protect and extract economic value of innovations, especially in the North American market, innovators from all around the world have strong motivation to apply for patents through USPTO, which makes USPTO patents a natural candidate to identify the frontier of technologies. Thus, I download all patent applications from 1976-2015 from the USPTO official website and transform bulk digital patent files into machine readable individual documents by each patent. Keeping utility patents only, there are almost six million patents available for analysis. Because it takes 29 months on average to get patent applications officially approved and available online, more recent patent applications are not included in the sample to prevent selection bias. Second, to best gauge the technology information, I keep only the abstracts and brief descriptions sections of each patent and remove the other parts. Detailed descriptions are not included because they usually involve unnecessarily technical terms which are undesired in later steps when matching with 10-K documents.[1] Other patent sections such as figures and citations are also excluded since they can not be matched with 10-K documents.

The second part of the data source, firm-level descriptions, relies on the annual

[1]It is very rare that firms use very technical terms in their annual reports and words and phrases in patent abstracts and brief descriptions are enough to measure technolgy frontier.

reports of U.S. public firms. The U.S. Securities and Exchange Commission (SEC) requires all publicly-traded companies to file an annual comprehensive report (10-K) about their business and products, organizational structure, financial performance, executive compensation, managerial discussion and analysis, and any other relevant data and statements. Similar to Handley and Li (2019), I take advantage of the rich information available in the text of the filings to obtain key words of interest. Specifically, I download all 10-K documents (along with the derivatives of 10-Ks, such as 10-KSB, 10-KA, etc.) from SEC EDGAR database that maintains the metadata as well as all filings in text format. The documents are cleaned according to the method in Handley and Li (2019). Then I take the entire document as unit of analysis instead of parts of the documents, such as business, risk factors, or manager's discussion and analysis. The ultimate goal is to extract technology related words and measure their level of importance to the firm, and it is unclear in which part of 10-K documents such words appear. There is good reason for these technology phrases to appear in the product and business description, risk factors, MD&A, or even financial statement section. The EDGAR database contains around 4,000-10,000 firms (with entry and exit) that regularly make filings to the SEC every year.

1.2.2 Three Steps to Construct PTF

In this section, I will discuss the steps of measure construction in detail.

1.2.2.1 USPTO Patent Documents

As described in section 1.2.1, each (utility) patent document is a collection of words, equations, and figures, and only abstracts and brief descriptions are kept. I clean each document by removing stop words as in Handley and Li (2019), exclusive terms very specific to particular technologies, such as DNA or RNA sequencing, and words that appear in only one single patent document among all patent applications within a particular year. To match with COMPUSTAT and EDGAR 10-K sample, I keep patents with application year from 1985-2015, which yields 5,270,133 patents with 693,798 distinct words spanning 31 years. Next, I pool all patents within the same application year[2] to create a bag of words by year and transform the bag of words into vectors. Therefore, all technology related words from patent documents are converted into vectors in an annual basis. Each vector has a length of 693,798 elements, each of which represents a distinct word.[3] This gives a potential representation of the technology frontier of the year.[4]

For each element in a vector, I make two adjustments. First, empirical regularity suggests that, in general, natural English language follows Zipf's law, which states that the frequency of a particular term in a corpus is inversely proportional to its rank (usually log transformed in both frequency and rank). This means that a small number of terms will have very large counts while many others will remain a small number of counts. Therefore, simple number-of-counts biases towards the frequently used phrases may not be able to convey much information with a high probability. It is necessary to reassign weights to each term, in this case, words in patent documents of any given year. To do so, I adopt a popular technique and calculate the TF-IDF (term frequency-inverse document frequency) index of each word in each year by

$$\mathrm{pTFIDF}_{wt} = (1 + \log(\mathrm{pTF}_{wt}))\log(\frac{\mathrm{pD}_t}{\mathrm{pDF}_{wt}}).$$

[2]I am using application year rather than grant year because if granted the average application-grant gap is 29 months, thus application year is in lead for about two years and a half better representing the technology frontier.

[3]It does not matter whether the word appears in any patent in the corresponding year.

[4]Representing each patent document into the vector will generate a huge sparse matrix for each year and will not be particularly helpful for identifying the technology frontier in the aggregate. It might be useful for patent classification and identifying technologies in a particular area, which is not the focus of the current exercise.

pTF_{wt} is the total frequency of word w in all patent documents in year t. pD_t is the total number of documents in year t, and pDF_{wt} is the number of documents that contain word w in year t. The TF-IDF index captures the relative importance of each word w in the corpus of all patent documents in year t. The larger the total frequency (TF), the larger the TF-IDF index, which means more frequent words are relatively more important in the corpus with a log transformation to dampen the weight, especially for ultra-high frequency words. The second inverse document frequency (IDF) term awards higher weight for words in fewer documents (with log dampening their weights) since these words are possibly critical to the patents where they belong. The product of the two terms allows the TF-IDF index to assign higher weights to popular words yet also prevent words that are less frequent, but critical, to specific technologies from diminishing.

Second, the relative importance of words in a given year is not sufficient to capture the rise and fall of technologies. In other words, TF-IDF is a static measure that ignores the dynamics of word usage that represents the dynamics of technological progress. If a technology is new and/or emerging, one should observe the words describing such technology to become more salient in terms of TF-IDF measure. Similarly, words describing declining, mature, or less used technologies should have a more stable or decreasing TF-IDF measure. To examine the usage dynamics of each word, I compute the usage growth rate of each word as

$$\text{pTFIDF\_GROWTH}_{wt} = \frac{\text{pTFIDF}_{wt} - \frac{1}{5}\sum_{i=1}^{5}\text{pTFIDF}_{wt-i}}{\text{pTFIDF}_{wt} + \frac{1}{5}\sum_{i=1}^{5}\text{pTFIDF}_{wt-i}} \in [-1, 1].$$

The TF-IDF growth index takes (half of) the mid-point growth of the TF-IDF index of current year and the average TF-IDF index of the past five years.[5] This measure

[5]Taking the five-year average smooths the growth measure. The results are largely robust to one-, three-, and ten-year averages.

tracks the appearance, disappearance, growth, and decline of each specific technological vocabulary. If the word appears in year t_0 but never appears in year $t_{-5}, t_{-4}, t_{-3}, t_{-2}$, and t_{-1}, the TF-IDF growth measure is 1, which means that the technology the word represents does not show up in patent documents for at least five years before it occurs. I argue that such words represent new (emerging) technology. A TF-IDF growth measure being -1 means that some patent documents contain the word in the past five years but the word disappears in the current year; that is, patent filing inventors stop using the word, which suggests that their research output on the technology the word represents has fallen. If the TF-IDF growth measure is positive, the word has been increasingly used in patents, and I assume rapid progress in the technology the word represents. If the TF-IDF growth measure is negative, the word usage has decreased, and I assume the technology the word represents is mature or has been abandoned.

Therefore, a vector to identify technology frontier is created, $\mathbf{pTFIDF\_GROWTH}_t$ with arguments $\text{pTFIDF\_GROWTH}_{wt}$, the TF-IDF growth rate of each word. The main idea of this approach is that words used in patent documents are technology related and the relative importance of the words represents the status of the technological progress the words describe. Another advantage of this approach is that it mechanically eliminates frequently used words (such as stopwords) and words whose TF-IDF measure does not change much over time, such as "invent" and "sum". by assigning them a weight of zero. Some of these words are related to technology or innovation but do not represent specific technologies. Thus, informative features (words that describe specific technologies) are kept in the measure.

To help understand and illustrate the source of validity of TF-IDF growth, I plot TF-IDF growth dynamics of some technological vocabularies in Figure 1.2. In panel (a), I pick three words in information technology. In the late 1980s, internet was the cutting edge technology and it first showed up in patent applications, which has a TF-IDF growth measure of 1 as shown in the figure. The measure goes down in the following years as the internet technology matured and more relative innovations followed. It became stable in recent years, which indicates the rapid spread of the internet technology; it is no longer a cutting-edge frontier technology but has not been abandoned by innovators, otherwise the measure will become negative and towards zero. The other two technologies in panel (a), bluetooth and OFDMA (Orthogonal Frequency-Division Multiple Access) coincide in the growth measure. Inventors started patenting technologies in bluetooth in the late 1990s and at the same time were equipped with an information technology OFDMA, which is widely used in communication industry. The two technology-representing words co-moved in the following years with the same pattern as the internet. Thus, $\mathbf{pTFIDF\_GROWTH}_t$ manages to track technologies that co-appear and co-develop in a consistent way. Panel (b) implies that the TF-IDF growth measure also captures words at different levels of category. Pixel is the unit of a minute area of illumination on a display screen of electronic devices. Semiconductor is a solid substance that conducts between an insulator and most metals, but also stands for a broader category of technology or an industry. Roughly speaking, pixel belongs to a sub-category of semiconductor technology. The measures of two technologies fluctuates together from 1980-2015, which suggests that the measure is valid and consistent across relevant technologies. The next two examples are drawn from biomedical technology. PCR (Polymerase Chain Reaction) method is critical when molecular biologists make copies of specific DNA sequences. It is trivial to observe that the two words PCR and polymerase co-move in panel (c). They are also mature technologies compared with siRNA, which was introduced in the early 2000s. siRNA known as short interfering RNA, operates to knock out or interfere with the expression of specific genes. Similar to patterns of internet, bluetooth, and OFDMA in panel (a), TF-IDF growth measure of siRNA starts with 1 when first introduced and declines as the technology matures.

1.2.2.2 EDGAR 10-K Documents

The second step is to clean firm-level EDGAR 10-K documents and construct a vector of word counts measure at the firm level. The entire 10-K (and its amendments) reports from 1994 till 2015 are downloaded from SEC EDGAR database. The cleaning document step closely follows Handley and Li (2019). Then repeating TF-IDF methods on firm's SEC EDGAR 10-K filings yields the firm-level measure

$$\mathrm{fTFIDF}_{wft} = (1 + \log(\mathrm{fTF}_{wft})) \log(\frac{\mathrm{fD}_t}{\mathrm{fDF}_{wt}}).$$

fTF_{wft} is the total frequency of word w in document (10-K filing) of firm f in year t. fD_t is the total number of 10-K documents in year t and fDF_{wt} is the number of 10-Ks that contain word w in year t. The rationale to implement TF-IDF rather than frequency count is the same as discussed in the previous section. Firm information can be expressed as a vector $\mathbf{fTFIDF}_{ft}$ with fTFIDF_{wft} as its arguments. Note that $\mathbf{fTFIDF}_{ft}$ is document-time (or firm-time) specific rather than a technology frontier measure, which is time specific only.

1.2.2.3 PTF

The final step to construct the firm-level proximity to the technology frontier (PTF) is to combine the two measures created in section 1.2.2.1 and 1.2.2.2 by taking the cosine similarity of the technology vector and firm vector:

$$\mathrm{PTF}_{ft} = \frac{< \mathbf{fTFIDF}_{ft}, \mathbf{pTFIDF\_GROWTH}_t >}{\|\mathbf{fTFIDF}_{ft}\| \cdot \|\mathbf{pTFIDF\_GROWTH}_t\|}.$$

$< \mathbf{fTFIDF}_{ft}, \mathbf{pTFIDF\_GROWTH}_t >$ is the inner product of the two vectors and PTF_{ft} is the inner product normalized by the L_2 norm of the vectors. The larger the PTF_{ft}, the closer the technology and firm vector are in terms of distance, which indicates that the firm is closer to the technology frontier. An important feature of this measure is that if the patent word TF-IDF growth is zero or close to zero, the respective word in $\mathbf{fTFIDF}_{ft}$ will carry zero weight when calculating the inner product. Therefore, firms that state words like "innovation", "technology", "invent" often without elaboration on specific technologies will not be rewarded nor punished in the measure since those words typically have zero TF-IDF growth rates. Only words with positive or negative growth matter in the PTF measure. Words that show up only in technology vectors or only in firm vectors are automatically dropped. The measure captures the aggregate level of proximity to technology frontier not in any specific industry. Firms documenting high tech in some aspects but low tech in other aspects in their 10-K filings do not necessarily receive a high PTF meaure. Assume that a firm discusses two technologies in its report. One technology is new and rising and one is mature and declining. The rising technology-related words grant positive weights in the $< \mathbf{fTFIDF}_{ft}, \mathbf{pTFIDF\_GROWTH}_t >$ inner product while the declining ones yield negative weights. The PTF of the firm might not be high in this case. It is also worth noting that firms are more motivated to mention cutting-edge technologies than mature technologies, which could create bias in our construction of the measure. However, when assuming all firms behave in a consistent way, the bias can be mitigated when controlling for firm fixed effects and performing cross-sectional comparisons.

More specifically, Hall, Jaffe and Trajtenberg (2001) classify patent classes into six large technological categories: chemical; computers and communications; drugs and medical; electrical and electronic; mechanical, and others. Using these categories allows me to deal with the issue that PTF_{ft} aggregates all technologies, which can be particularly problematic on conglomerates or firms operating in multiple industries. I can construct firm-level PTF for each specific technological category c, PTF_{fct}. The construction of PTF_{fct} is exactly the same as in section 1.2.1.1 except the corpus is the documents of patents in one of the six categories. I can also construct a PTF breadth measure at the firm level (similar to construction of HHI) which captures how broad a firm's technologies as when taking proximity to frontier into account

$$\text{PTF\_Breadth}_{ft} = \sqrt{\sum_{c=1}^{6} (\text{PTF}_{fct} - \text{avgPTF}_{ft})^2},$$

where $\text{avgPTF}_{ft} = \frac{1}{6} \sum_{c=1}^{6} \text{PTF}_{fct}$.

1.3 Proximity to Technology Frontier (PTF) and Firm Performance

To test the validity of the measure of proximity to technology frontier at the firm-level and to investigate its impact on short and long-run firm behavior, I match PTF with firm characteristics obtained from COMPUSTAT and run empirical models on both firm performance measures and firm level input in innovation on PTF.

1.3.1 Data Description

Firm information is drawn from COMPUSTAT-Capital IQ North America Fundamentals Annual from WRDS. It consists of firm balance sheet, cash flow, and income statements. We match PTF measure with COMPUSTAT through firm identifier central index key (CIK) and year. Missing variables are replaced by taking the average of the value of the variable of the previous and later year given they are not missing. Then the sample is matched with idiosyncratic uncertainty measure CRUX constructed in Handley and Li (2019). After removing missing values in PTF, CRUX, firm capital stock measure total gross property, plant, and equipment (ppegt) and firm total employment (emp), we have over 101,000 observations.

Firm level investment opportunity is captured by Tobin's Q (logged), which is computed by

$$\text{Tobin's Q} = \frac{\text{Market Capitalization} + \text{Market Value of Liability}}{\text{Total Asset Value}} = \frac{csho \times prcc\_f + at - ceq}{at},$$

where market capitalization is common shares outstanding (csho) $\times$ price closed at fiscal year (prcc_f) and market value of liability is approximated by book value of liability which is total asset (at) $-$ total common/ordinary equity (ceq). Firm size is approximated by total asset (logged) as well.

The measures of firm performance include investment rate (growth of constructed firm capital stock), employment growth, output growth, sales growth, total factor productivity growth (estimated by OLS and Olley and Pakes, 1996) in one, two, three, four, and five years from the base year in logged terms, that is $\log(X_{ft+\tau}) - \log(X_{ft})$, $\tau = 1, 2, 3, 4, 5$.

Capital stock is computed through the perpetual inventory method $K_{ft} = \pi_t((1 - \delta)K_{ft-1} + I_{ft-1})$ with π_t = ratio of producer price index by commodity for final demand private capital equipment (WPSFD41312) between year t and $t-1$ and $I_{ft} = capx$ (capital investment) and initial total capital stock = $ppegt$. Employment and sales are values as reported in COMPUSTAT. Total output is computed as sum of sales and inventory (sale + invt). Total factor productivity (TFP) is estimated in two ways in a sample where utility (SIC code 4900-4999) and financial (SIC code 6000-6999) firms are excluded. In the OLS version, TFP is the predicted residual of regressing output on firm size (capital and labor), inventory materials and firm and year fixed effects (specifically log(output) on log(ppegt), log(emp), log(invrm)). In the OP version, TFP is estimated by applying production function using Olley and Pakes' (1996) technique. Exit is the year a firm drops out of the COMPUSTAT sample. State variables include firm age and log(ppegt). The variable to proxy for unobserved productivity is log of capital investment = log(capx). Additional variables used in the second stage are log employment and log inventory materials. The estimation is bootstrapped with 100 repetitions. PTF is winsorized by three standard deviations from mean by year. All other variables are winsorized at 1% and 99% level by year. Table 1.1 provides summary statistics of the constructed sample.

To evaluate the relationship between PTF and firm's other innovation behavior, such as research and development expenditure and patenting activities, I match the PTF-COMPUSTAT sample with Kogan, Papanikolaou, Seru, and Stoffman (2017) economic value of patents sample. Firm innovation activities R&D expenditure and economic value of patents are normalized by lagged total asset. Control variables include lagged sales growth ($\log(sales_{ft-1}) - \log(sales_{ft-2})$), lagged leverage (short and long term debt over equity = (dlc + dltt) / seq), lagged log employment, lagged log Tobin's Q and lagged CRUX. Normalized R&D expenditure is lagged by one, two, and three years because it takes time for research and development input to realize the gain and push up technology closer to the frontier. Economic value of patents is lagged by one year and firms without patents granted are replaced with zero. PTF is winsorized by three standard deviations from mean by year. All other variables are winsorized at 1% and 99% level by year. To alleviate the concern that R&D is self-reported and patenting behavior is self-selected, I further restrict the sample to positive innovation measures. Table 1.5 provides summary statistics of the constructed sample.

1.3.2 PTF vs Tobin's Q and Firm Size

More advanced technology induces higher productivity, which achieves more investment opportunities and growth. Productivity literature also documents that measures of total factor productivity (TFP) highly correlates with firm size. If the PTF measure really captures technology beyond merely total factor productivity, then it should be positively correlated with investment opportunities and growth and not strongly monotonically correlated with TFP measures. One of the threats to the validation of the PTF measure is that it might be simply capturing firm size and makes the positive correlation with productivity and growth mechanical. It is not rare to observe that smaller and younger firms are keen on technological innovation and thus score high in PTF measure.

To address this concern, I regress Tobin's Q (investment opportunity) and total asset (firm size) on PTF controlling for capital stock, employment, idiosyncratic uncertainty, and year fixed effects. To visualize, I plot the binscatter of the regression results according to Cattaneo, Crump, Farrell, and Feng (2019) as shown in Figure 1.3. The observations are clustered into bins and the model is fitted with a second order polynomial. In panel (a), I find that PTF is highly positively (not perfectly) correlated with log Tobin's Q, which suggests that higher PTF predicts higher investment opportunity. The result does not imply that PTF can be simply replaced by Tobin's Q as PTF is measuring technology, which is beyond Tobin's Q. Also note that the Tobin's Q measure is average Q but not marginal Q, which, in theory, should summarize all possible investment opportunities. Panel (b) presents a U-shape correlation between PTF and firm size, thus large firms are not guaranteed a high PTF score. In fact, some firms with large assets are low in the technology measure. The figures show that PTF is associated with better investment opportunities yet not monotonically associated with firm size. Our concern that PTF is positively biased towards large firms is rejected.

1.3.3 PTF and Firm Growth

In this section, I study the effects of PTF on firm long-run and short-run growth. I run the following empirical model:

$$\log(X_{ft+\tau}) - \log(X_{ft}) = \lambda \text{PTF}_{ft} + \boldsymbol{\beta}\text{Controls}_{ft} + G_{it} + \epsilon_{ft+\tau}.$$

PTF is standardized by subtracting mean and dividing standard deviation. X takes capital stock, total employment, total output, total sales, and TFP (OLS and OP, 1996). Control variables include firm size (ppegt and emp) to control for the effects of size on growth, idiosyncratic uncertainty index CRUX to control for the effect of uncertainty on growth, industry-year[6] fixed effects to account for industry level fluctuations and demand shocks. Standard errors are clustered at industry-year level to account for possible serial correlation. $\tau = 1, 2, 3, 4, 5$.

Table 1.2 provides the baseline results. The odd numbered columns show the impact of PTF on firm performance without control variables, while the even numbered columns show the results of the full model. In all panels (a)-(f), PTF has a positive and sizable impact on all firm growth measures and the impact is, in general, larger in a longer time horizon. The impact is generally larger in the long-run than in the short-run.[7] Figure 1.4 plots the coefficients and confidence interval of PTF in all full-model regressions. In panel (a), the impact of PTF on investment rate is positive and significant and keeps growing as the time horizon goes up. To quantify the impact, one standard deviation increase in PTF increases investment rate by 2%, 3.5%, 4.5%, 5.4% and 6.4% when the time horizon ranges from one to five years. Similarly, in panel (b), the impact of one standard deviation increase in PTF on employment growth is around 0.8% to 1.2% in different year horizons. The marginal increase in the impact becomes

[6] Industry is defined as 4-digit NAICS code.

[7] The growth of the impact diminishes in time but does not fall to or below zero.

smaller as time horizon becomes larger but the impact is still significant. Similar patterns can be found in panels (c) and (d) where one standard deviation increase in PTF predicts 1.3% - 2.2% (or 2.4%) increase in total sales growth (or output growth) and the marginal change of the impact becomes flatter as τ goes up. Finally, in panel (e) and (f), the impact of PTF on TFP (OLS and OP, 1996) ranges from 0.7% - 1.6% (OLS) or 0.75% - 1.8% (OP, 1996). Firms with higher PTF consistently outperform by a wide margin those with lower PTF in terms of input (investment and employment), output (total output and sales), and productivity growth.

Perhaps all the results in Table 1.2 and Figure 1.4 are driven by patenting firms only. To alleviate this concern, I include an indicator variable (= 1 if granted patent and = 0 otherwise). The results are reported in Table 1.3. It shows that patent indicator has positive and significant impact on all firm performance margins, especially in the long run and when control variables are included. That is, patenting firms grow faster in inputs, outputs, and productivity. However, the effect of PTF is not completely erased and the PTF maintains significant and sizable impact on multiple margins.

To further compare patenting firms with non-patenting firms, I conduct sub sample analysis in Table 1.4, and Figure 1.5 displays the comparison of the coefficients and confidence intervals of the PTF. By running the full model on both patenting sub sample and non-patenting sub sample, it is trivial to observe (especially in the figure) that PTF has a larger impact on patenting firms than non-patenting firms, but the impact on non-patenting firms remains significant and sizable. As shown in Figure 1.5, the spread of TFP on performance between patenting and non-patenting firms grows in the long-run. This is reasonable because patenting firms obtain the leading position in the market since they are one of the first innovators. In the meantime, patenting firms can also acquire profits from selling and licensing intellectual property rights. Firms that do not patent but adopt technologies also enjoy the benefits from the innovation, but less than the patenting firms do. It is less obvious in the TFP sample since the number of non-patenting firms in relatively small as TFP estimation is mainly concentrated in IT and manufacturing firms where most of them file patents.

1.3.4 PTF and Innovation

Firms' past R&D input and patent output should also affect firms' current level of technology. To test the hypothesis, I run the following regression

$$\text{PTF}_{ft} = \alpha Z_{ft-\nu} + \boldsymbol{\eta}\text{Controls}_{ft} + H_{it} + F_f + \epsilon_{ft-\nu}.$$

Z takes firm R&D expenditure of KPSS market value of patents.[8] Control variables include common factors that could impact firm innovation activity and technology: Tobin's Q, leverage, sales growth, and uncertainty. Industry-year fixed effects are included to control for industry level fluctuations and demand shocks. Firm level fixed effects are included to remove the firm invariant effects on technology. Standard errors are clustered at firm level. $\nu = 1, 2, 3$.

Table 1.6 shows strong evidence that higher innovation input and output predicts higher PTF. In panel (a), columns (1) - (6) show that previous R&D expenditure is positively associated with current PTF. The impact is larger in a longer time horizon, which indicates that it takes time for innovation input to be effective and realized. Previous patenting activity is also positively correlated with current PTF as shown in columns (7) - (8). In panel (b), I restrict the sample to positive R&D expenditure and/or patenting firms only to deal with a selection problem since R&D expenditure is usually self-reported and some firms might strategically choose not to file patent to protect their technology secrets. The results are robust with a large set of controls.

[8] Note that number of patents are not used since not all patents are equally valuable.

1.4 Proximity to Technology Frontier (PTF) and Merger Decisions

Technology (both innovation and adoption) not only influences firm growth, but also shapes industry structure mainly through entry, exit, and reallocation. In this section, I focus on the impact of technology, specifically PTF, on firm capital reallocation decisions, i.e. mergers and acquisitions.

1.4.1 Data and Sample Construction

In addition to data sets described in previous sections, SDC Platinum data is accessed to obtain merger and acquisition information. SDC Platinum is an online historical financial-transactions database. It provides detailed financial transaction information on M&A activities, as well as private equity, bonds, new issues, and loans among others. The SDC M&A data tracks domestic and international merger deals from 1976 (1985 if cross-border deals). It records more than 100,000 deals with detailed deal information, such as time frame, deal value, deal type[9], third party (legal and auditing services), and detailed information about the target and the acquirer including firm identifier, income and financial statements, and other information. I mainly rely on the detailed deal information (announcement and completion date, deal type, and status) and firm identifiers to track firm behavior in resource reallocation. After removing duplicates (e.g., same firm ID, same date, and multiple rounds), I match SDC Platinum M&A database with COMPUSTAT through the SDC deal numbers - GVKEY link provided by Ewens, Peters, and Wang (2019). The link connects SDC's M&A database to COMPUSTAT GVKEYs for both the acquirer and target (if they are traded/public). Based on Phillips and Zhdanov (2013), who created the first major mapping between

[9]Such as mergers and acquisitions, stake purchases, LBOs, tender offers, privatizations, and spinoffs.

firms in SDC and Compustat using a combination of name and date matching, Ewens et al. (2019) fills gaps using this website that replicates the GVKEY search in WRDS. The link extends the sample to more recent years and improves coverage and accuracy.

Next, I create the merger and acquisition sample. Public mergers and acquisitions are usually a small sample. Looking at only the announced or successfully completed deals misses the point that more firms can potentially match and make a deal. To overcome this small sample issue, I extend the coverage and include deals that could possibly had happened in the following three steps. (1) Identify industry-year (industry defined as 4-digit NAICS code[10]) groups that contain at least one firm announced as acquirer or target. (2) Fully interact (Cartesian product) all possible mergers based on the sample created in (1), remove acquirer-target pair duplicates and self-self pairs, keep only industry-year pairs with at least one merger deal announcement. For example, if in year 2000, some merger deal was announced between industry 5112 (software publishers) with N_1 firms and industry 3332 (industrial machinery manufacturing) with N_2 firms, then there were $\binom{N_1}{1}\binom{N_2}{1} = N_1 \times N_2$ number of possible mergers and each one became a created observation. If the merger occurs within the same industry with N firms, then number of possible mergers is $\binom{N}{2} = N(N-1)/2$. In short, firms from industries with firms conducting M&A deals are defined as potential target or acquirer in a potential M&A deal. (3) Define variables of interest including PTF related variables and those suggested in Bena and Li (2014). The merger deal variable is simply an indicator (= 1 if announces merger, = 0 otherwise). Other variables focus on the similarity and/or closeness (distance) of the acquirer and the target. These variables include PTF, PTF breadth as defined in section 1.2, log total asset, log sales growth, log Tobin's Q, leverage as defined in section 1.3, and cash holding (cash/lagged total asset = che/lag at) and return to asset (ROA = oibdp / at). PTF is winsorized

[10]Results are robust when industry is defined by 4-digit SIC code as shown in the appendix.

by three standard deviations from mean by year. All other variables are winsorized at 1% and 99% level by year. The distance variable is defined as the absolute value of the respective variables between the acquirer and the target. The smaller the absolute value, the closer the distance between the merger firms is.

Besides the distance of PTF among firms, it is also critical to understand how firms are similar in terms of technology portfolio. It is less meaningful to compare a low PTF difference generated from IT firm pair and a high PTF difference generated from a pharmaceutical and auto-producing firm. Thus, I construct PTF similarity by taking cosine similarity of TF-IDF growth rate weighted SEC EDGAR 10-K vector between the acquirer and the target firms. Specifically,

$$\text{PTF Similarity}_{fg} = \frac{< \mathbf{fTFIDF}_{ft} \circ \mathbf{pTFIDF\_GROWTH}_t, \mathbf{fTFIDF}_{gt} \circ \mathbf{pTFIDF\_GROWTH}_t >}{\|\mathbf{fTFIDF}_{ft} \circ \mathbf{pTFIDF\_GROWTH}_t\| \cdot \|\mathbf{fTFIDF}_{gt} \circ \mathbf{pTFIDF\_GROWTH}_t\|},$$

where f and g represents firms and $\circ$ denotes element-by-element multiplication. The calculation is the same as PTF but the inner product is taken by vectors drawn from firm 10-K text documents weighted by the same technology vector. Larger PTF Similarity indicates the closer technology portfolio between the two firms. I also match the sample with Hoberg and Phillips (2016) pairwise measure of product competition, which is also based on text analysis, as a robustness check since the construction of PTF similarity is close to how the Hoberg and Phillips product competition measure is created. This yields a sample of almost 16 million observations. In addition, a sample where all measures are adjusted by industry average is also constructed. Summary statistics are provided in Table 1.7.

This approach is implementable because every publicly-traded firm in COMPUSTAT has a PTF measure that significantly extends the coverage of technology and innovation measure. Also, note that this approach does not distinguish between the acquirer and the target firms. In quite a few merger deals, it is unclear which firm is the target and which firm is the acquirer. The role can often change and adjust due to tax or financial issues or other issues. One famous example is the Porsche and Volkswagen merger, which started with Porsche buying Volkswagen shares but ended with Volkswagen buying up Porsche. Thus it is reasonable to put aside acquirer vs target characterizations and focus on the merger pair only.

1.4.2 PTF and Mergers

Determinants of mergers and acquisitions have been extensively studied, but the literature has not yet come to agreement about what the determinants are. There are theories on high Q buys low Q (Jovanovic and Rousseau 2002, 2008) and similar mergers with similar (Rhodes-Kropf and Robinson 2008) among others. The impact of technology and innovations is also investigated (Bena and Li 2014). However, due to data limitations, especially on the measures of innovation and technology adoption, the puzzle has not yet been solved.

In this section, I present the results of the following linear probability model

$$\begin{aligned}\text{M\&A Event}_{fgt} =& \alpha \mid \Delta\text{PTF} \mid_{fgt-1} + \beta\text{PTF Similarity}_{fgt-1} + \eta \mid \Delta\text{PTF Breadth} \mid_{fgt-1} \\ &+ \gamma\text{HP Competition}_{fgt-1} + \delta \mid \Delta\text{PTF} \mid_{fgt-1} \times \mid \Delta\text{Firm Characteristics} \mid_{fgt-1} \\ &+ \mid \Delta\text{Firm Characteristics} \mid_{fgt-1} + \text{Pairwise Industry-Year FE} + \epsilon_{fgt}.\end{aligned}$$

M&A event is an indicator ($= 1$ if merger deal announces, $= 0$ otherwise). $\mid \Delta \cdot \mid$ is the absolute value of the difference of the respective variables of the acquirer and the target. PTF similarity is as described in section 1.4.1. HP competition is the product competition measure drawn from Hoberg and Phillips (2016). Firm characteristics include log total asset, cash holding, log sales growth, log Tobin's Q, book leverage, and ROA. All firm characteristic variables are demeaned. Acquirer industry-target industry-year fixed effects are included to absorb any invariant trends within the matched group.

Standard error is clustered at acquirer industry-target industry-year level.

The results are shown in Table 1.8. Panel (a) presents the baseline model without interaction terms. In column (1), I find that firms with similar technology but larger distance of proximity to technology frontier merge with a higher likelihood. In column (2), when control variables are included, the pattern does not change. Furthermore, firms with closer total asset, cash holding, sales growth and Tobin's Q tend to merge, which indicates a similar merges with similar pattern. Firms with larger distance in PTF breadth tend to merge, which suggests a conglomerate merges with specific pattern. In columns (3) and (4), product level competition is controlled for (with a cut on the size of the sample). The pattern stays the same while firms more likely to compete with each other (high HP competition) and firms with similar cash holding, investment opportunities, and profitability tend to merge.

In panel (b) of Table 1.8, I investigate the effect of $\mid \Delta \text{PTF} \mid$ interacting with $\mid \Delta \log(\text{total asset}) \mid$ and find that firms with a larger difference in proximity to technology frontier tend to merge but the effect will be mitigated when the difference in firms size (total asset) between the potential merger firms is large. When product competition is controlled for, the larger the difference in firms' size, the more likely the firms will merge, which suggests a large merges with small pattern in the presence of product competition. This is consistent with most existing findings. In panels (c), (d), and (e), I interact $\mid \Delta \text{PTF} \mid$ with $\mid \Delta \log(\text{sales growth}) \mid$, $\mid \Delta \log(\text{Tobin's Q}) \mid$, and $\mid \Delta \text{ROA} \mid$ and find a similar pattern among the three. The positive impact of $\mid \Delta \text{PTF} \mid$ on merger likelihood is mitigated by larger differences in firms sales growth, investment opportunity (Tobin's Q) and profitability (ROA) while the closer the distance in sales growth, Tobin's Q, and ROA, the more likely the firms merge. All the results are robust when control variables are adjusted by industry average as provided in columns (5)-(8) in all the panels in Table 1.8.

These results shed light on some of the inconclusiveness in the literature. The results on the effects of $\mid \Delta\text{PTF} \mid$ on merger probability seem to support Jovanovic and Braguinsky (2004). They build a model to explain bidder discounts and target premia in takeovers and argue that firms with low tech but poor management buy firms with high tech but good management. However, this results does not necessarily imply high Q firm buys low Q firm since Tobin's Q combines firms' investment opportunities which should be a combination of, at least, both technology and management. The empirical results on firm level characteristics, especially firm performance, suggest a similar buys similar pattern. The results remain significant when controlling for industry effects and product market competition and firm level technology similarity. Therefore, empirical evidence suggests that the determinants of mergers is more subtle than currently known, and both high buys low and similar buys similar patterns exist.

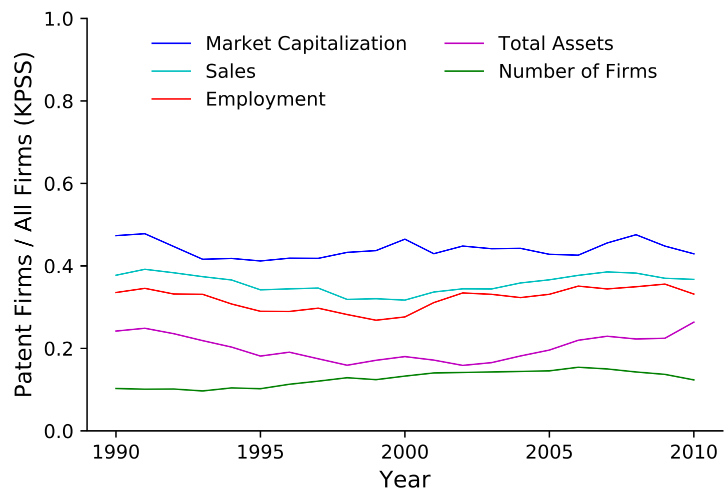
1.5 Conclusion

In this paper, I construct a novel measure of technological innovation and technology adoption, proximity to technology frontier (PTF), through a natural language processing approach. The PTF extends the coverage of traditional innovation measures such as patenting and R&D expenditures to the space of all publicly trade firms. It tracks the rise and fall of technologies and matches these technologies to firms. The validity of PTF is asserted as it predicts better firm level performance in investment rate, employment, output, sales and TFP growth. It also correlates with corporate patenting activities and R&D inputs. The measure also has implications on firm merger and acquisition behavior. It provides more evidence on the pattern of M&A deals in addition to the existing arguments (high buys low and similar buys similar). It points to future research for more consistently measuring technological innovation in a large coverage, and its impact on corporate behavior, labor markets, industry structure, and the macroeconomy.

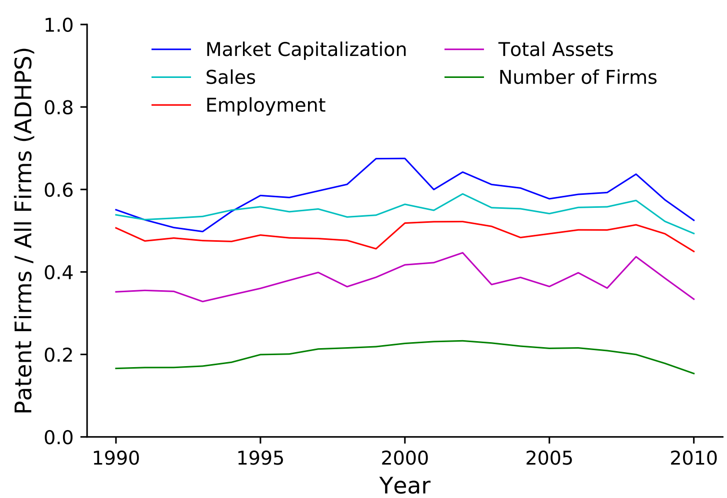
Figures

Figure 1.1: Coverage of Patent as Innovation Measure

(a) Grant Year



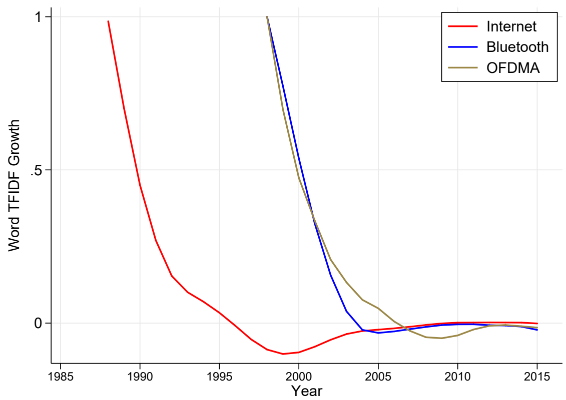
(b) Application Year



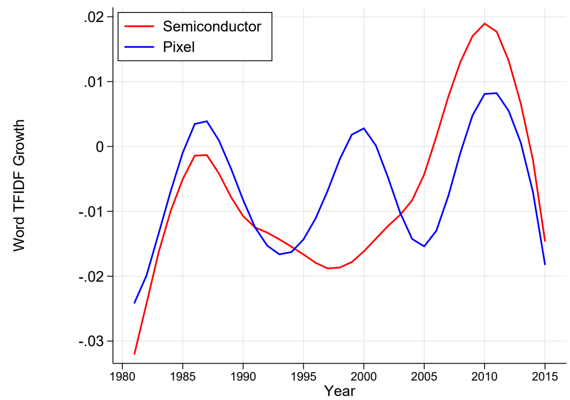
Notes: The sample is created by matching patent data drawn from Kogan, Papanikolaou, Seru and Stoffman (2017) and Autor, Dorn, Hanson, Pisano and Shu (2019) with COMPUSTAT.

Figure 1.2: Words in Patents

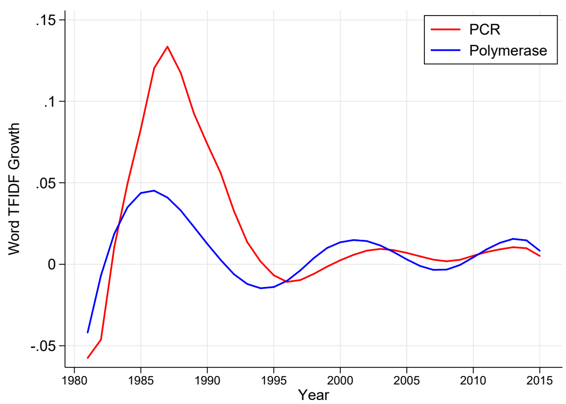
(a) Cutting Edge Technologies - IT



(b) Technology and Industry - Electronics



(c) Mature Technologies - Biomedical



(d) Mature and New Technologies - Biomedical

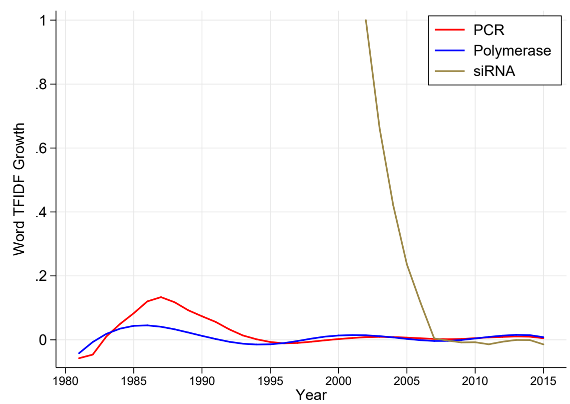
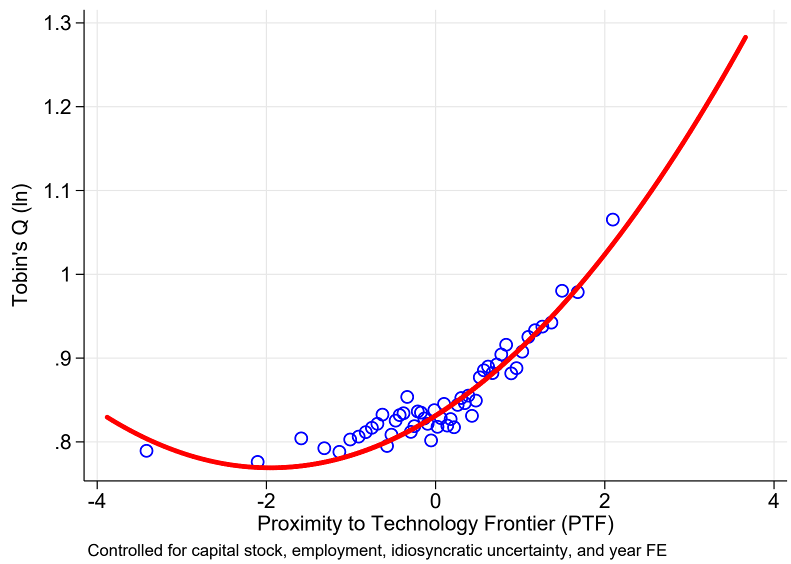
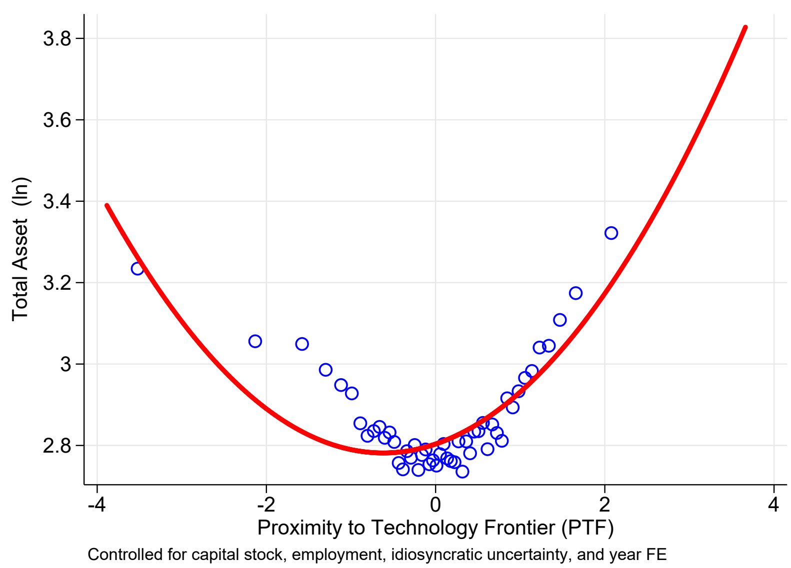


Figure 1.3: Binscatter: PTF vs Growth Opportunities and Size

(a) PTF vs Tobin's Q



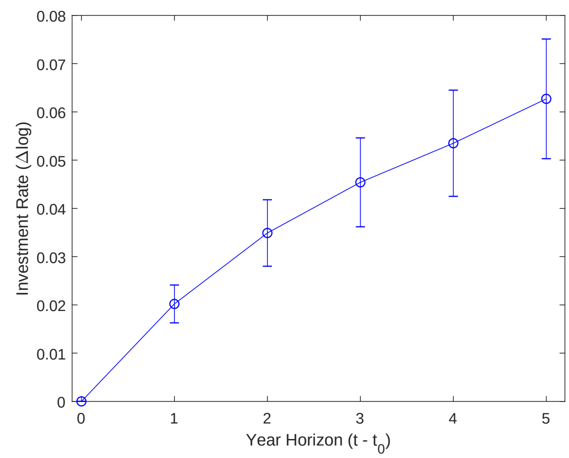
(b) PTF vs Total Asset



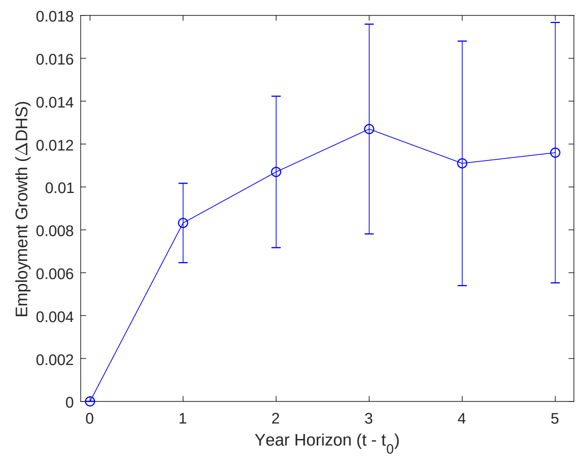
Notes: This figure plots binscatter regressions (Cattaneo, Crump, Farrell and Feng 2019) of PTF on Tobin's Q or total asset controlling for total capital stock, employment, idiosyncratic uncertainty and year fixed effects.

Figure 1.4: PTF and Firm Performance

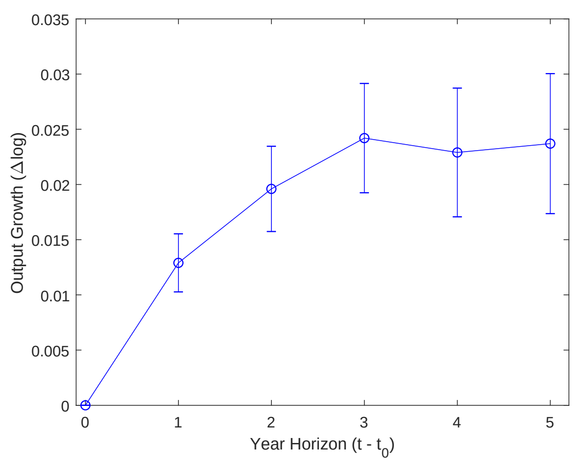
(a) Investment Rate



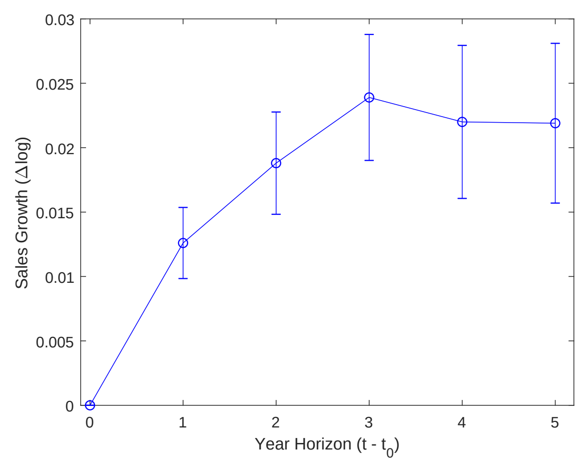
(b) Employment Growth



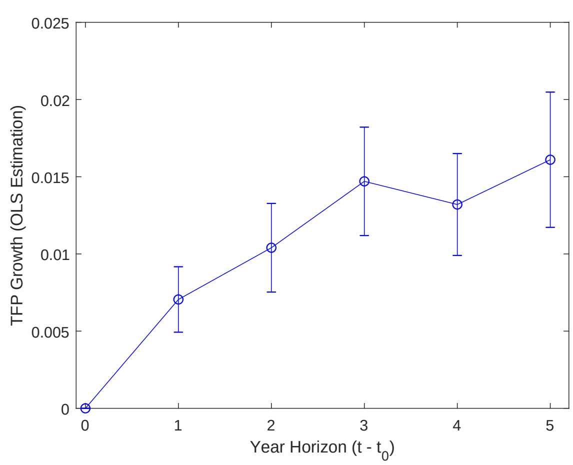
(c) Output Growth



(d) Sales Growth



(e) TFP Growth (OLS)



(f) TFP Growth (OP 1996)

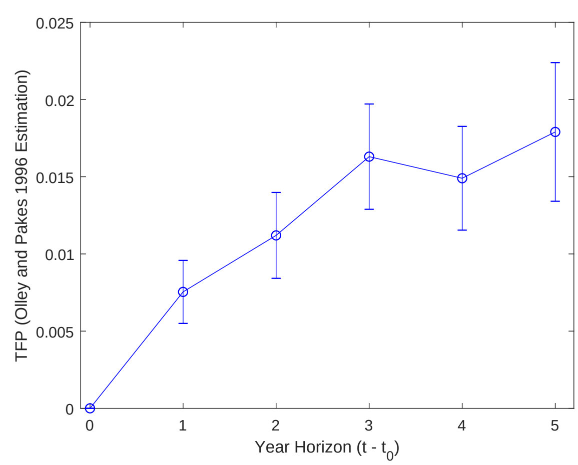
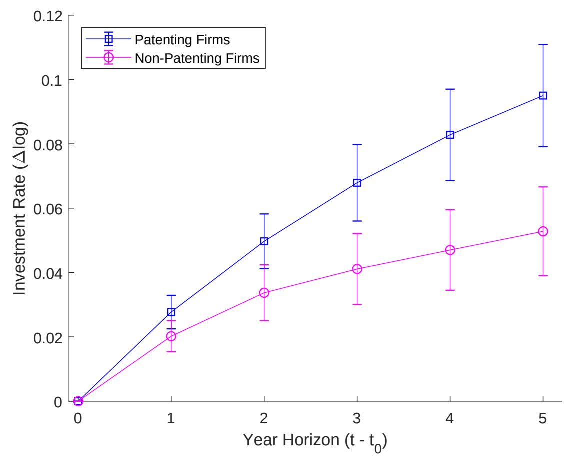
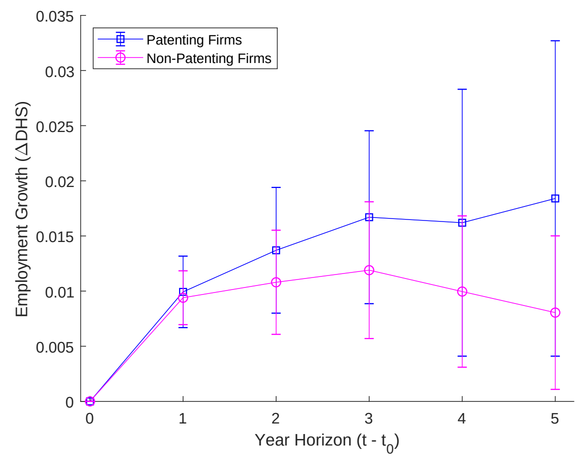


Figure 1.5: PTF and Firm Performance - Patenting vs Non-Patenting Firms

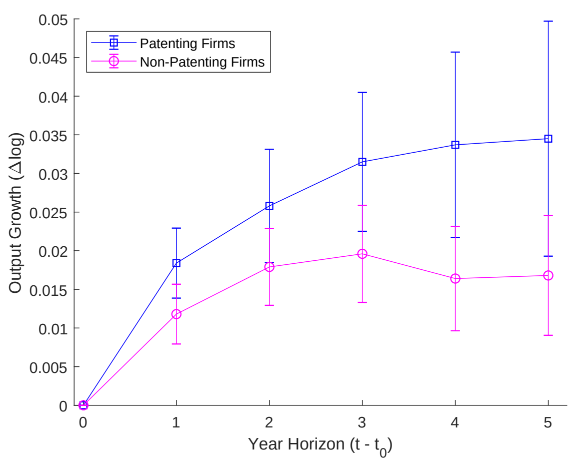
(a) Investment Rate



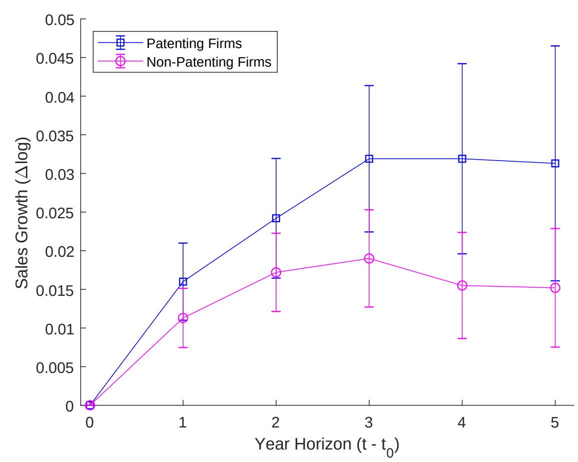
(b) Employment Growth



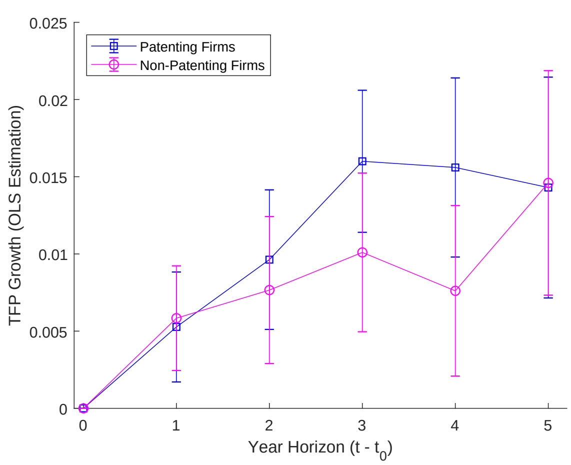
(c) Output Growth



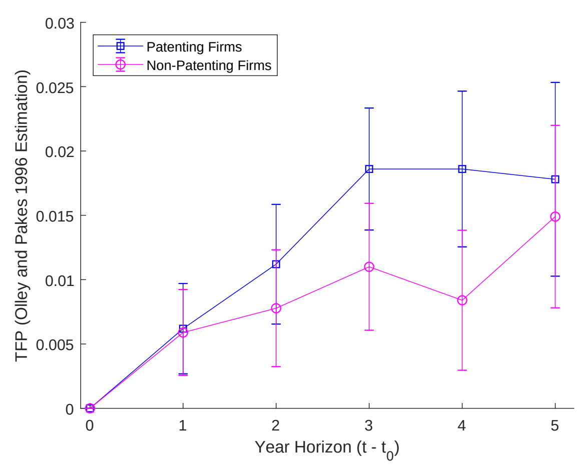
(d) Sales Growth



(e) TFP Growth (OLS)



(f) TFP Growth (OP 1996)



Tables

Table 1.1: PTF and Firm Performance Summary Statistics

VARIABLES	(1) N	(2) mean	(3) sd	(4) min	(5) max
PTF (t)	101,269	0.0284	0.993	-4.255	4.085
CRUX (t)	101,269	0.0420	0.0341	0	0.540
Log Capital Stock (t)	101,269	4.012	2.858	-4.510	10.69
Log Employment (t)	101,269	-0.449	2.385	-6.908	4.934
	Time Horizon τ = 1				
Capital Stock Growth (Δln, t+1→t)	79,894	0.0929	0.209	-0.110	1.560
Employment Growth (Δln, t+1→t)	93,250	0.0268	0.326	-2.048	1.609
Output Growth (Δln, t+1→t)	90,326	0.0561	0.423	-2.181	2.251
Sales Growth (Δln, t+1→t)	91,025	0.0578	0.430	-2.139	2.276
TFP Growth (Δln, OLS, t+1→t)	34,820	0.00848	0.241	-1.011	1.208
TFP Growth (Δln, Olley Pakes 1996, t+1→t)	34,820	0.164	0.247	-0.866	1.418
	Time Horizon τ = 2				
Capital Stock Growth (Δln, t+2→t)	74,821	0.176	0.356	-0.216	2.442
Employment Growth (Δln, t+2→t)	84,802	0.0482	0.499	-2.597	2.307
Output Growth (Δln, t+2→t)	82,295	0.0923	0.615	-2.926	3.187
Sales Growth (Δln, t+2→t)	83,035	0.0945	0.620	-2.752	3.283
TFP Growth (Δln, OLS, t+2→t)	31,277	0.00973	0.303	-1.142	1.468
TFP Growth (Δln, Olley Pakes 1996, t+2→t)	31,277	0.318	0.310	-0.893	1.737
	Time Horizon τ = 3				
Capital Stock Growth (Δln, t+3→t)	67,345	0.250	0.469	-0.321	3.169
Employment Growth (Δln, t+3→t)	74,324	0.0681	0.616	-2.944	2.847
Output Growth (Δln, t+3→t)	72,307	0.125	0.738	-3.305	3.798
Sales Growth (Δln, t+3→t)	73,022	0.127	0.743	-3.282	3.868
TFP Growth (Δln, OLS, t+3→t)	28,306	0.00833	0.337	-1.260	1.599
TFP Growth (Δln, Olley Pakes 1996, t+3→t)	28,306	0.470	0.344	-0.870	2.034
	Time Horizon τ = 4				
Capital Stock Growth (Δln, t+4→t)	60,430	0.317	0.556	-0.417	3.472
Employment Growth (Δln, t+4→t)	65,337	0.0851	0.708	-3.219	3.073
Output Growth (Δln, t+4→t)	63,735	0.155	0.833	-3.627	4.105
Sales Growth (Δln, t+4→t)	64,404	0.159	0.835	-3.605	4.136
TFP Growth (Δln, OLS, t+4→t)	25,803	0.00594	0.359	-1.366	1.642
TFP Growth (Δln, Olley Pakes 1996, t+4→t)	25,803	0.621	0.368	-0.682	2.247
	Time Horizon τ = 5				
Capital Stock Growth (Δln, t+5→t)	53,984	0.380	0.631	-0.520	3.726
Employment Growth (Δln, t+5→t)	57,485	0.101	0.782	-3.714	3.195
Output Growth (Δln, t+5→t)	56,238	0.186	0.902	-3.750	4.188
Sales Growth (Δln, t+5→t)	56,859	0.191	0.900	-3.685	4.205
TFP Growth (Δln, OLS, t+5→t)	22,724	0.00643	0.377	-1.444	1.712
TFP Growth (Δln, Olley Pakes 1996, t+5→t)	22,724	0.777	0.387	-0.629	2.575

Notes: This table provides summary statistics on the effects of PTF on firm growth. PTF is winsorized by 3 standard deviation from mean by year. All other variables are winsorized by 1% and 99% by year except for CRUX.

Table 1.2: PTF and Firm Performance

(a) Investment Rate

	Dependent Variable: Log Change in Capital Stock (Δln, t+τ→t)									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Time Horizon (τ)	1		2		3		4		5	
PTF (t)	0.0192***	0.0202***	0.0344***	0.0349***	0.0456***	0.0454***	0.0545***	0.0535***	0.0643***	0.0627***
	[0.00421]	[0.00392]	[0.00742]	[0.00689]	[0.00993]	[0.00921]	[0.0119]	[0.0110]	[0.0136]	[0.0124]
Log Capital Stock (t)		-0.0114***		-0.0343***		-0.0540***		-0.0698***		-0.0813***
		[0.00197]		[0.00413]		[0.00621]		[0.00780]		[0.00895]
Log Employment (t)		0.0143***		0.0376***		0.0568***		0.0718***		0.0825***
		[0.00200]		[0.00375]		[0.00534]		[0.00669]		[0.00788]
CRUX (t)		-0.203***		-0.357***		-0.452***		-0.504**		-0.550**
		[0.0591]		[0.101]		[0.141]		[0.192]		[0.246]
Constant	0.0927***	0.153***	0.176***	0.347***	0.248***	0.512***	0.314***	0.651***	0.376***	0.767***
	[0.000995]	[0.00799]	[0.00206]	[0.0177]	[0.00318]	[0.0272]	[0.00431]	[0.0341]	[0.00545]	[0.0386]
Industry-Year FE	√	√	√	√	√	√	√	√	√	√
Observations	79,815	79,815	74,742	74,742	67,267	67,267	60,355	60,355	53,916	53,916
R-squared	0.105	0.110	0.112	0.123	0.115	0.130	0.116	0.133	0.117	0.134
Number of Firms	8510	8510	8071	8071	7441	7441	6881	6881	6335	6335

Notes: Standard errors are clustered at firm and year level. Dependent variable is log(K(t+τ))-log(K(t)) where K is firm's total capital stock. PTF (t) measures firm's proximity to technology frontier. Log capital stock and log total employment are included to control for the effects of firm size on dependent variable. Firm idiosyncratic uncertainty (CRUX) is drawn from Handley and Li (2020). Time horizon ranges from 1 to 5 years from base year. All of the regression specification include industry-year fixed effects, where industry is at 3-digit NAICS code level. PTF is winsorized at mean ± 3sd by year, log(capital stock), log(employment) and dependent variable are winsorized at 1% level by year.

Table 1.2: PTF and Firm Performance

(b) Employment Growth

	Dependent Variable: Log Change in Employment (Δln, t+τ→t)									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Time Horizon (τ)	1		2		3		4		5	
PTF (t)	0.00944***	0.00832***	0.0136***	0.0107***	0.0162***	0.0127**	0.0158***	0.0111*	0.0172***	0.0116*
	[0.00190]	[0.00185]	[0.00381]	[0.00353]	[0.00491]	[0.00489]	[0.00540]	[0.00570]	[0.00561]	[0.00607]
Log Capital Stock (t)		0.00834**		0.0209***		0.0298***		0.0389***		0.0498***
		[0.00305]		[0.00643]		[0.00845]		[0.00954]		[0.0111]
Log Employment (t)		-0.0148***		-0.0373***		-0.0563***		-0.0744***		-0.0923***
		[0.00312]		[0.00695]		[0.00945]		[0.0111]		[0.0134]
CRUX (t)		-0.000698		-0.00803		-0.195		-0.258		-0.325
		[0.0648]		[0.0959]		[0.130]		[0.159]		[0.197]
Constant	0.0265***	-0.0133	0.0479***	-0.0511	0.0671***	-0.0661	0.0840***	-0.0884*	0.0996***	-0.120**
	[0.000530]	[0.0147]	[0.00167]	[0.0310]	[0.00284]	[0.0403]	[0.00407]	[0.0454]	[0.00538]	[0.0523]
Industry-Year FE	√	√	√	√	√	√	√	√	√	√
Observations	93,167	93,167	84,711	84,711	74,236	74,236	65,254	65,254	57,412	57,412
R-squared	0.057	0.059	0.061	0.068	0.062	0.072	0.062	0.075	0.062	0.077
Number of Firms	12083	12083	10852	10852	9426	9426	8254	8254	7217	7217

Notes: Standard errors are clustered at firm and year level. Dependent variable is log(emp(t+τ))-log(emp(t)) where emp is firm's total employment. PTF (t) measures firm's proximity to technology frontier. Log capital stock and log total employment are included to control for the effects of firm size on dependent variable. Firm idiosyncratic uncertainty (CRUX) is drawn from Handley and Li (2020). Time horizon ranges from 1 to 5 years from base year. All of the regression specification include industry-year fixed effects, where industry is at 3-digit NAICS code level. PTF is winsorized at mean ± 3sd by year, log(capital stock), log(employment) and dependent variable are winsorized at 1% level by year.

Table 1.2: PTF and Firm Performance

(c) Output Growth

	Dependent Variable: Log Change in Output (Δln, t+τ→t)									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Time Horizon (τ)	1		2		3		4		5	
PTF (t)	0.0158***	0.0129***	0.0246***	0.0196***	0.0297***	0.0242***	0.0300***	0.0229***	0.0325***	0.0237***
	[0.00273]	[0.00263]	[0.00405]	[0.00386]	[0.00528]	[0.00495]	[0.00597]	[0.00583]	[0.00625]	[0.00634]
Log Capital Stock (t)		-0.0211***		-0.0222***		-0.0151**		-0.00937		-0.00307
		[0.00282]		[0.00522]		[0.00686]		[0.00727]		[0.00803]
Log Employment (t)		0.0165***		0.00972*		-0.00459		-0.0183**		-0.0317***
		[0.00290]		[0.00515]		[0.00704]		[0.00830]		[0.00907]
CRUX (t)		0.110		0.111		0.0206		0.0647		0.121
		[0.0884]		[0.148]		[0.203]		[0.230]		[0.277]
Constant	0.0556***	0.144***	0.0917***	0.184***	0.123***	0.187***	0.153***	0.189***	0.184***	0.191***
	[0.000558]	[0.0145]	[0.00187]	[0.0268]	[0.00321]	[0.0348]	[0.00457]	[0.0378]	[0.00597]	[0.0408]
Industry-Year FE	√	√	√	√	√	√	√	√	√	√
Observations	90,245	90,245	82,210	82,210	72,223	72,223	63,659	63,659	56,170	56,170
R-squared	0.057	0.061	0.060	0.063	0.058	0.061	0.057	0.062	0.058	0.064
Number of Firms	11766	11766	10592	10592	9186	9186	8057	8057	7066	7066

Notes: Standard errors are clustered at firm and year level. Dependent variable is log(output(t+τ))-log(output(t)) where output is firm's total output (sales + inventory). PTF (t) measures firm's proximity to technology frontier. Log capital stock and log total employment are included to control for the effects of firm size on dependent variable. Firm idiosyncratic uncertainty (CRUX) is drawn from Handley and Li (2020). Time horizon ranges from 1 to 5 years from base year. All of the regression specification include industry-year fixed effects, where industry is at 3-digit NAICS code level. PTF is winsorized at mean ± 3sd by year, log(capital stock), log(employment) and dependent variable are winsorized at 1% level by year.

Table 1.2: PTF and Firm Performance

(d) Sales Growth

	Dependent Variable: Log Change in Sales (Δln, t+τ→t)									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Time Horizon (τ)	1		2		3		4		5	
PTF (t)	0.0155***	0.0126***	0.0238***	0.0188***	0.0293***	0.0239***	0.0291***	0.0220***	0.0307***	0.0219***
	[0.00284]	[0.00276]	[0.00412]	[0.00397]	[0.00531]	[0.00489]	[0.00608]	[0.00594]	[0.00619]	[0.00620]
Log Capital Stock (t)		-0.0216***		-0.0229***		-0.0149**		-0.00983		-0.00523
		[0.00269]		[0.00521]		[0.00661]		[0.00696]		[0.00801]
Log Employment (t)		0.0166***		0.00968*		-0.00572		-0.0197**		-0.0314***
		[0.00282]		[0.00494]		[0.00642]		[0.00778]		[0.00884]
CRUX (t)		0.0837		0.0721		-0.0188		0.00366		0.0223
		[0.0930]		[0.144]		[0.192]		[0.215]		[0.259]
Constant	0.0573***	0.149***	0.0939***	0.191***	0.126***	0.191***	0.157***	0.198***	0.189***	0.209***
	[0.000514]	[0.0142]	[0.00186]	[0.0269]	[0.00319]	[0.0331]	[0.00453]	[0.0356]	[0.00591]	[0.0395]
Industry-Year FE	√	√	√	√	√	√	√	√	√	√
Observations	90,941	90,941	82,946	82,946	72,935	72,935	64,325	64,325	56,787	56,787
R-squared	0.057	0.060	0.060	0.063	0.058	0.062	0.058	0.063	0.059	0.066
Number of Firms	11808	11808	10642	10642	9231	9231	8107	8107	7109	7109

Notes: Standard errors are clustered at firm and year level. Dependent variable is log(sales(t+τ))-log(sales(t)) where sales is firm's total sales. PTF (t) measures firm's proximity to technology frontier. Log capital stock and log total employment are included to control for the effects of firm size on dependent variable. Firm idiosyncratic uncertainty (CRUX) is drawn from Handley and Li (2020). Time horizon ranges from 1 to 5 years from base year. All of the regression specification include industry-year fixed effects, where industry is at 3-digit NAICS code level. PTF is winsorized at mean ± 3sd by year, log(capital stock), log(employment) and dependent variable are winsorized at 1% level by year.

Table 1.2: PTF and Firm Performance

(e) TFP Growth (OLS)

	Dependent Variable: Log Change in TFP (Δln, OLS, t+τ→t)									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Time Horizon (τ)	1		2		3		4		5	
PTF (t)	0.00797***	0.00705***	0.0117***	0.0104***	0.0167***	0.0147***	0.0159***	0.0132***	0.0191***	0.0161***
	[0.00202]	[0.00212]	[0.00266]	[0.00287]	[0.00346]	[0.00351]	[0.00339]	[0.00330]	[0.00441]	[0.00438]
Log Capital Stock (t)		-0.0150***		-0.0161***		-0.0161**		-0.0140*		-0.0101
		[0.00279]		[0.00463]		[0.00623]		[0.00765]		[0.00856]
Log Employment (t)		0.0147***		0.0144***		0.0132**		0.00988		0.00388
		[0.00275]		[0.00424]		[0.00586]		[0.00776]		[0.00887]
CRUX (t)		0.0804		0.0778		0.131		0.183		0.153
		[0.0604]		[0.0835]		[0.117]		[0.133]		[0.146]
Constant	0.00893***	0.0704***	0.00976***	0.0765***	0.00888***	0.0733**	0.00664**	0.0601*	0.00662**	0.0448
	[1.72e-05]	[0.0114]	[0.000833]	[0.0191]	[0.00167]	[0.0261]	[0.00236]	[0.0322]	[0.00308]	[0.0362]
Industry-Year FE	√	√	√	√	√	√	√	√	√	√
Observations	34,180	34,180	30,709	30,709	27,780	27,780	25,322	25,322	22,286	22,286
R-squared	0.042	0.044	0.047	0.049	0.051	0.053	0.055	0.057	0.056	0.058
Number of Firms	4608	4608	4054	4054	3545	3545	3146	3146	2717	2717

Notes: Standard errors are clustered at firm and year level. Dependent variable is log(TFP(t+τ))-log(TFP(t)) where TFP is firm's total factor productivity eastimated under OLS. PTF (t) measures firm's proximity to technology frontier. Log capital stock and log total employment are included to control for the effects of firm size on dependent variable. Firm idiosyncratic uncertainty (CRUX) is drawn from Handley and Li (2020). Time horizon ranges from 1 to 5 years from base year. All of the regression specification include industry-year fixed effects, where industry is at 3-digit NAICS code level. PTF is winsorized at mean ± 3sd by year, log(capital stock), log(employment) and dependent variable are winsorized at 1% level by year.

Table 1.2: PTF and Firm Performance

(f) TFP Growth (OP 1996)

	Dependent Variable: Log Change in TFP (Δln, Olley Pakes 1996, t+τ→t)									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Time Horizon (τ)	1		2		3		4		5	
PTF (t)	0.00852***	0.00754***	0.0126***	0.0112***	0.0182***	0.0163***	0.0176***	0.0149***	0.0209***	0.0179***
	[0.00197]	[0.00204]	[0.00257]	[0.00278]	[0.00338]	[0.00341]	[0.00344]	[0.00336]	[0.00448]	[0.00449]
Log Capital Stock (t)		-0.0191***		-0.0234***		-0.0260***		-0.0258***		-0.0236**
		[0.00284]		[0.00458]		[0.00615]		[0.00747]		[0.00831]
Log Employment (t)		0.0191***		0.0221***		0.0238***		0.0224***		0.0183**
		[0.00282]		[0.00422]		[0.00584]		[0.00761]		[0.00873]
CRUX (t)		0.0767		0.0738		0.130		0.176		0.146
		[0.0607]		[0.0883]		[0.121]		[0.138]		[0.149]
Constant	0.164***	0.244***	0.318***	0.416***	0.470***	0.578***	0.622***	0.727***	0.777***	0.874***
	[1.49e-05]	[0.0114]	[0.000850]	[0.0188]	[0.00171]	[0.0256]	[0.00240]	[0.0313]	[0.00313]	[0.0353]
Industry-Year FE	√	√	√	√	√	√	√	√	√	√
Observations	34,180	34,180	30,709	30,709	27,780	27,780	25,322	25,322	22,286	22,286
R-squared	0.055	0.059	0.063	0.067	0.069	0.073	0.080	0.084	0.088	0.092
Number of Firms	4608	4608	4054	4054	3545	3545	3146	3146	2717	2717

Notes: Standard errors are clustered at firm and year level. Dependent variable is log(TFP(t+τ))-log(TFP(t)) where TFP is firm's total factor productivity eastimated under Olley and Pakes (1996). PTF (t) measures firm's proximity to technology frontier. Log capital stock and log total employment are included to control for the effects of firm size on dependent variable. Firm idiosyncratic uncertainty (CRUX) is drawn from Handley and Li (2020). Time horizon ranges from 1 to 5 years from base year. All of the regression specification include industry-year fixed effects, where industry is at 3-digit NAICS code level. PTF is winsorized at mean ± 3sd by year, log(capital stock), log(employment) and dependent variable are winsorized at 1% level by year.

Table 1.3: PTF and Firm Performance - Control for Patenting Activity

(a) Investment Rate

	Dependent Variable: Log Change in Capital Stock (Δln, t+τ→t)									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Time Horizon (τ)	1		2		3		4		5	
PTF (t)	0.0204***	0.0210***	0.0354***	0.0354***	0.0447***	0.0437***	0.0524***	0.0501***	0.0592***	0.0563***
	[0.00533]	[0.00494]	[0.00966]	[0.00898]	[0.0124]	[0.0114]	[0.0140]	[0.0129]	[0.0152]	[0.0140]
1(patent) (t)	-0.000318	0.000467	-0.00298	0.00866	-0.00111	0.0211**	0.00336	0.0344***	0.00635	0.0438***
	[0.00490]	[0.00411]	[0.00894]	[0.00708]	[0.0126]	[0.00967]	[0.0143]	[0.0113]	[0.0158]	[0.0134]
1(patent) × PTF (t)	0.00307	0.00379	0.00773	0.00895	0.0140	0.0156	0.0194	0.0216	0.0244	0.0271*
	[0.00474]	[0.00476]	[0.00923]	[0.00934]	[0.0127]	[0.0128]	[0.0142]	[0.0142]	[0.0150]	[0.0151]
Log Capital Stock (t)		-0.0124***		-0.0370***		-0.0572***		-0.0734***		-0.0852***
		[0.00217]		[0.00454]		[0.00672]		[0.00823]		[0.00920]
Log Employment (t)		0.0151***		0.0394***		0.0582***		0.0729***		0.0835***
		[0.00225]		[0.00411]		[0.00574]		[0.00699]		[0.00801]
CRUX (t)		-0.203***		-0.357***		-0.453***		-0.497**		-0.576**
		[0.0693]		[0.115]		[0.153]		[0.199]		[0.251]
Constant	0.0946***	0.158***	0.180***	0.357***	0.253***	0.522***	0.317***	0.659***	0.377***	0.775***
	[0.00152]	[0.00866]	[0.00305]	[0.0190]	[0.00450]	[0.0291]	[0.00565]	[0.0356]	[0.00677]	[0.0394]
Industry-Year FE	√	√	√	√	√	√	√	√	√	√
Observations	67,561	67,561	63,227	63,227	59,041	59,041	55,123	55,123	51,408	51,408
R-squared	0.109	0.114	0.117	0.128	0.119	0.134	0.119	0.136	0.118	0.136
Number of Firms	8066	8066	7661	7661	7178	7178	6716	6716	6250	6250

Notes: Standard errors are clustered at firm and year level. Dependent variable is log(K(t+τ))-log(K(t)) where K is firm's total capital stock. PTF (t) measures firm's proximity to technology frontier. 1(patent) is a indicator function = 1 if the firm have patent granted in the given year, and = 0 otherwise. Log capital stock and log total employment are included to control for the effects of firm size on dependent variable. Firm idiosyncratic uncertainty (CRUX) is drawn from Handley and Li (2020). Time horizon ranges from 1 to 5 years from base year. All of the regression specification include industry-year fixed effects, where industry is at 3-digit NAICS code level. PTF is winsorized at mean ± 3sd by year, log(capital stock), log(employment) and dependent variable are winsorized at 1% level by year.

Table 1.3: PTF and Firm Performance - Control for Patenting Activity

(b) Employment Growth

	Dependent Variable: Log Change in Employment (Δln, t+τ→t)									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Time Horizon (τ)	1		2		3		4		5	
PTF (t)	0.00944***	0.00863***	0.0109**	0.00887*	0.0118*	0.00906	0.0109	0.00680	0.01000	0.00526
	[0.00274]	[0.00273]	[0.00497]	[0.00483]	[0.00617]	[0.00624]	[0.00667]	[0.00678]	[0.00681]	[0.00678]
1(patent) (t)	0.00531	0.0130**	0.00551	0.0252**	0.0119	0.0454***	0.0135	0.0601***	0.0104	0.0653***
	[0.00556]	[0.00459]	[0.0105]	[0.00891]	[0.0141]	[0.0118]	[0.0158]	[0.0141]	[0.0190]	[0.0181]
1(patent) × PTF (t)	0.00400	0.00275	0.0113	0.00836	0.0181*	0.0140	0.0224*	0.0174	0.0267*	0.0207
	[0.00457]	[0.00451]	[0.00789]	[0.00758]	[0.00985]	[0.00948]	[0.0121]	[0.0116]	[0.0147]	[0.0143]
Log Capital Stock (t)		0.00971**		0.0242***		0.0339***		0.0410***		0.0491***
		[0.00333]		[0.00700]		[0.00844]		[0.00922]		[0.0107]
Log Employment (t)		-0.0171***		-0.0429***		-0.0635***		-0.0803***		-0.0955***
		[0.00347]		[0.00765]		[0.00969]		[0.0112]		[0.0135]
CRUX (t)		0.0109		-0.00317		-0.227		-0.304*		-0.382*
		[0.0735]		[0.108]		[0.142]		[0.169]		[0.204]
Constant	0.0223***	-0.0256	0.0419***	-0.0759**	0.0587***	-0.0993**	0.0768***	-0.115**	0.0952***	-0.132**
	[0.00117]	[0.0161]	[0.00286]	[0.0341]	[0.00432]	[0.0412]	[0.00556]	[0.0449]	[0.00710]	[0.0517]
Industry-Year FE	√	√	√	√	√	√	√	√	√	√
Observations	78,342	78,342	71,022	71,022	64,720	64,720	59,365	59,365	54,672	54,672
R-squared	0.058	0.061	0.061	0.069	0.062	0.073	0.062	0.076	0.061	0.078
Number of Firms	10985	10985	9881	9881	8802	8802	7901	7901	7082	7082

Notes: Standard errors are clustered at firm and year level. Dependent variable is log(emp(t+τ))-log(emp(t)) where emp is firm's total employment. PTF (t) measures firm's proximity to technology frontier. 1(patent) is a indicator function = 1 if the firm have patent granted in the given year, and = 0 otherwise. Log capital stock and log total employment are included to control for the effects of firm size on dependent variable. Firm idiosyncratic uncertainty (CRUX) is drawn from Handley and Li (2020). Time horizon ranges from 1 to 5 years from base year. All of the regression specification include industry-year fixed effects, where industry is at 3-digit NAICS code level. PTF is winsorized at mean ± 3sd by year, log(capital stock), log(employment) and dependent variable are winsorized at 1% level by year.

Table 1.3: PTF and Firm Performance - Control for Patenting Activity

(c) Output Growth

	Dependent Variable: Log Change in Output (Δln, t+τ→t)									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Time Horizon (τ)	1		2		3		4		5	
PTF (t)	0.0147***	0.0122***	0.0210***	0.0168***	0.0228***	0.0178***	0.0227***	0.0157**	0.0249***	0.0163**
	[0.00369]	[0.00363]	[0.00492]	[0.00480]	[0.00630]	[0.00600]	[0.00701]	[0.00684]	[0.00742]	[0.00735]
1(patent) (t)	-0.000811	0.0155***	0.00303	0.0331***	0.0172	0.0574***	0.0252*	0.0791***	0.0227	0.0883***
	[0.00429]	[0.00424]	[0.00940]	[0.00810]	[0.0113]	[0.0117]	[0.0133]	[0.0141]	[0.0163]	[0.0175]
1(patent) × PTF (t)	0.00562	0.00495	0.0152*	0.0133	0.0257**	0.0230**	0.0281**	0.0245**	0.0290**	0.0246*
	[0.00471]	[0.00471]	[0.00783]	[0.00782]	[0.00890]	[0.00893]	[0.0102]	[0.0101]	[0.0127]	[0.0126]
Log Capital Stock (t)		-0.0211***		-0.0217***		-0.0148*		-0.0110		-0.00727
		[0.00339]		[0.00587]		[0.00724]		[0.00743]		[0.00823]
Log Employment (t)		0.0165***		0.00822		-0.00693		-0.0205**		-0.0325***
		[0.00340]		[0.00575]		[0.00757]		[0.00864]		[0.00930]
CRUX (t)		0.148		0.148		0.0345		0.0428		0.0581
		[0.0954]		[0.163]		[0.222]		[0.240]		[0.285]
Constant	0.0574***	0.140***	0.0945***	0.175***	0.122***	0.173***	0.152***	0.183***	0.183***	0.194***
	[0.00112]	[0.0165]	[0.00290]	[0.0294]	[0.00426]	[0.0359]	[0.00571]	[0.0379]	[0.00731]	[0.0413]
Industry-Year FE	√	√	√	√	√	√	√	√	√	√
Observations	76,137	76,137	69,160	69,160	63,124	63,124	58,006	58,006	53,523	53,523
R-squared	0.058	0.061	0.058	0.061	0.056	0.060	0.056	0.062	0.057	0.064
Number of Firms	10769	10769	9704	9704	8619	8619	7735	7735	6950	6950

Notes: Standard errors are clustered at firm and year level. Dependent variable is log(output(t+τ))-log(output(t)) where output is firm's total output (sales + inventory). PTF (t) measures firm's proximity to technology frontier. 1(patent) is a indicator function = 1 if the firm have patent granted in the given year, and = 0 otherwise. Log capital stock and log total employment are included to control for the effects of firm size on dependent variable. Firm idiosyncratic uncertainty (CRUX) is drawn from Handley and Li (2020). Time horizon ranges from 1 to 5 years from base year. All of the regression specification include industry-year fixed effects, where industry is at 3-digit NAICS code level. PTF is winsorized at mean ± 3sd by year, log(capital stock), log(employment) and dependent variable are winsorized at 1% level by year.

Table 1.3: PTF and Firm Performance - Control for Patenting Activity

(d) Sales Growth

	Dependent Variable: Log Change in Sales (Δln, t+τ→t)									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Time Horizon (τ)	1		2		3		4		5	
PTF (t)	0.0142***	0.0116***	0.0206***	0.0165***	0.0228***	0.0178***	0.0226***	0.0156**	0.0243***	0.0158**
	[0.00363]	[0.00359]	[0.00507]	[0.00498]	[0.00647]	[0.00608]	[0.00729]	[0.00712]	[0.00743]	[0.00729]
1(patent) (t)	-0.00194	0.0149***	-0.000735	0.0303***	0.0136	0.0549***	0.0185	0.0749***	0.0172	0.0862***
	[0.00482]	[0.00469]	[0.00926]	[0.00814]	[0.0109]	[0.0116]	[0.0126]	[0.0142]	[0.0164]	[0.0180]
1(patent) × PTF (t)	0.00423	0.00361	0.0127	0.0109	0.0235**	0.0208**	0.0242**	0.0205*	0.0243*	0.0199
	[0.00462]	[0.00463]	[0.00803]	[0.00803]	[0.00853]	[0.00862]	[0.0100]	[0.0102]	[0.0124]	[0.0124]
Log Capital Stock (t)		-0.0219***		-0.0222***		-0.0147*		-0.0117		-0.00948
		[0.00325]		[0.00582]		[0.00707]		[0.00735]		[0.00835]
Log Employment (t)		0.0170***		0.00806		-0.00799		-0.0214**		-0.0322***
		[0.00333]		[0.00532]		[0.00682]		[0.00819]		[0.00910]
CRUX (t)		0.132		0.113		-0.00449		-0.0126		-0.0461
		[0.0987]		[0.159]		[0.210]		[0.224]		[0.266]
Constant	0.0598***	0.147***	0.0985***	0.183***	0.126***	0.179***	0.158***	0.194***	0.189***	0.215***
	[0.00118]	[0.0162]	[0.00286]	[0.0290]	[0.00417]	[0.0342]	[0.00557]	[0.0364]	[0.00727]	[0.0404]
Industry-Year FE	√	√	√	√	√	√	√	√	√	√
Observations	76,717	76,717	69,769	69,769	63,722	63,722	58,592	58,592	54,100	54,100
R-squared	0.057	0.061	0.058	0.061	0.057	0.061	0.057	0.063	0.058	0.066
Number of Firms	10822	10822	9763	9763	8666	8666	7786	7786	6993	6993

Notes: Standard errors are clustered at firm and year level. Dependent variable is log(sales(t+τ))-log(sales(t)) where sales is firm's total sales. PTF (t) measures firm's proximity to technology frontier. 1(patent) is a indicator function = 1 if the firm have patent granted in the given year, and = 0 otherwise. Log capital stock and log total employment are included to control for the effects of firm size on dependent variable. Firm idiosyncratic uncertainty (CRUX) is drawn from Handley and Li (2020). Time horizon ranges from 1 to 5 years from base year. All of the regression specification include industry-year fixed effects, where industry is at 3-digit NAICS code level. PTF is winsorized at mean ± 3sd by year, log(capital stock), log(employment) and dependent variable are winsorized at 1% level by year.

Table 1.3: PTF and Firm Performance - Control for Patenting Activity

(e) TFP Growth (OLS)

	Dependent Variable: Log Change in TFP (Δln, OLS, t+τ→t)									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Time Horizon (τ)	1		2		3		4		5	
PTF (t)	0.00589***	0.00498**	0.00599*	0.00471	0.00762*	0.00599	0.00653	0.00430	0.0128**	0.00975
	[0.00178]	[0.00190]	[0.00317]	[0.00331]	[0.00402]	[0.00402]	[0.00543]	[0.00508]	[0.00593]	[0.00568]
1(patent) (t)	0.00554	0.0124***	0.0123	0.0216***	0.0125	0.0218***	0.0158*	0.0260***	0.0161*	0.0310***
	[0.00460]	[0.00388]	[0.00731]	[0.00611]	[0.00862]	[0.00713]	[0.00888]	[0.00788]	[0.00861]	[0.00904]
1(patent) × PTF (t)	0.00258	0.00245	0.0102*	0.00997*	0.0166***	0.0165**	0.0160*	0.0158*	0.0108	0.0104
	[0.00291]	[0.00291]	[0.00490]	[0.00489]	[0.00566]	[0.00566]	[0.00868]	[0.00876]	[0.00762]	[0.00772]
Log Capital Stock (t)		-0.0150***		-0.0145***		-0.0133*		-0.00938		-0.00794
		[0.00307]		[0.00494]		[0.00644]		[0.00808]		[0.00937]
Log Employment (t)		0.0143***		0.0118**		0.0103		0.00464		-0.000509
		[0.00305]		[0.00470]		[0.00629]		[0.00848]		[0.00961]
CRUX (t)		0.0972		0.0843		0.138		0.176		0.135
		[0.0690]		[0.0956]		[0.132]		[0.155]		[0.162]
Constant	0.00582***	0.0634***	0.00371	0.0589***	0.00211	0.0499*	-0.00116	0.0281	0.000506	0.0236
	[0.00159]	[0.0121]	[0.00294]	[0.0188]	[0.00385]	[0.0252]	[0.00437]	[0.0324]	[0.00482]	[0.0381]
Industry-Year FE	√	√	√	√	√	√	√	√	√	√
Observations	28,880	28,879	25,883	25,882	23,372	23,371	21,297	21,297	19,491	19,491
R-squared	0.042	0.044	0.047	0.048	0.048	0.050	0.052	0.054	0.054	0.056
Number of Firms	4256	4256	3747	3747	3283	3283	2916	2916	2575	2575

Notes: Standard errors are clustered at firm and year level. Dependent variable is log(TFP(t+τ))-log(TFP(t)) where TFP is firm's total factor productivity eastimated under OLS. PTF (t) measures firm's proximity to technology frontier. 1(patent) is a indicator function = 1 if the firm have patent granted in the given year, and = 0 otherwise. Log capital stock and log total employment are included to control for the effects of firm size on dependent variable. Firm idiosyncratic uncertainty (CRUX) is drawn from Handley and Li (2020). Time horizon ranges from 1 to 5 years from base year. All of the regression specification include industry-year fixed effects, where industry is at 3-digit NAICS code level. PTF is winsorized at mean ± 3sd by year, log(capital stock), log(employment) and dependent variable are winsorized at 1% level by year.

Table 1.3: PTF and Firm Performance - Control for Patenting Activity

(f) TFP Growth (OP 1996)

	Dependent Variable: Log Change in TFP (Δln, Olley Pakes 1996, t+τ→t)									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Time Horizon (τ)	1		2		3		4		5	
PTF (t)	0.00615***	0.00517**	0.00624**	0.00496	0.00866**	0.00698*	0.00753	0.00525	0.0133**	0.0102*
	[0.00167]	[0.00178]	[0.00294]	[0.00308]	[0.00400]	[0.00396]	[0.00537]	[0.00498]	[0.00595]	[0.00567]
1(patent) (t)	0.00514	0.0135***	0.0118	0.0235***	0.0123	0.0247***	0.0153*	0.0296***	0.0154*	0.0345***
	[0.00459]	[0.00402]	[0.00716]	[0.00600]	[0.00854]	[0.00710]	[0.00873]	[0.00787]	[0.00860]	[0.00909]
1(patent) × PTF (t)	0.00317	0.00304	0.0115**	0.0113**	0.0180***	0.0180***	0.0179*	0.0176*	0.0137*	0.0134
	[0.00290]	[0.00290]	[0.00484]	[0.00481]	[0.00577]	[0.00576]	[0.00856]	[0.00860]	[0.00770]	[0.00775]
Log Capital Stock (t)		-0.0193***		-0.0222***		-0.0238***		-0.0222**		-0.0223**
		[0.00313]		[0.00498]		[0.00649]		[0.00800]		[0.00918]
Log Employment (t)		0.0187***		0.0198***		0.0214***		0.0180**		0.0148
		[0.00309]		[0.00475]		[0.00640]		[0.00844]		[0.00957]
CRUX (t)		0.0927		0.0756		0.134		0.166		0.125
		[0.0691]		[0.102]		[0.139]		[0.160]		[0.164]
Constant	0.166***	0.242***	0.322***	0.409***	0.477***	0.569***	0.630***	0.714***	0.785***	0.869***
	[0.00159]	[0.0120]	[0.00289]	[0.0188]	[0.00384]	[0.0251]	[0.00435]	[0.0319]	[0.00485]	[0.0375]
Industry-Year FE	√	√	√	√	√	√	√	√	√	√
Observations	28,880	28,879	25,883	25,882	23,372	23,371	21,297	21,297	19,491	19,491
R-squared	0.055	0.059	0.060	0.064	0.063	0.066	0.071	0.074	0.078	0.082
Number of Firms	4256	4256	3747	3747	3283	3283	2916	2916	2575	2575

Notes: Standard errors are clustered at firm and year level. Dependent variable is log(TFP(t+τ))-log(TFP(t)) where TFP is firm's total factor productivity eastimated under Olley and Pakes (1996). PTF (t) measures firm's proximity to technology frontier. 1(patent) is a indicator function = 1 if the firm have patent granted in the given year, and = 0 otherwise. Log capital stock and log total employment are included to control for the effects of firm size on dependent variable. Firm idiosyncratic uncertainty (CRUX) is drawn from Handley and Li (2020). Time horizon ranges from 1 to 5 years from base year. All of the regression specification include industry-year fixed effects, where industry is at 3-digit NAICS code level. PTF is winsorized at mean ± 3sd by year, log(capital stock), log(employment) and dependent variable are winsorized at 1% level by year.

Table 1.4: PTF and Firm Performance - Patenting vs Non-Patenting Firms

(a) Investment Rate

	Dependent Variable: Log Change in Capital Stock (Δln, t+τ→t)									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Patent = 0	Patent > 0	Patent = 0	Patent > 0	Patent = 0	Patent > 0	Patent = 0	Patent > 0	Patent = 0	Patent > 0
Time Horizon (τ)	1		2		3		4		5	
PTF (t)	0.0202***	0.0277***	0.0337***	0.0497***	0.0411***	0.0679***	0.0470***	0.0828***	0.0528***	0.0950***
	[0.00482]	[0.00522]	[0.00868]	[0.00850]	[0.0110]	[0.0119]	[0.0125]	[0.0142]	[0.0138]	[0.0159]
Log Capital Stock (t)	-0.0118***	-0.0179***	-0.0359***	-0.0503***	-0.0556***	-0.0794***	-0.0707***	-0.107***	-0.0811***	-0.132***
	[0.00219]	[0.00418]	[0.00457]	[0.00857]	[0.00696]	[0.0111]	[0.00870]	[0.0127]	[0.00977]	[0.0150]
Log Employment (t)	0.0159***	0.0156***	0.0405***	0.0465***	0.0594***	0.0730***	0.0738***	0.0985***	0.0836***	0.121***
	[0.00240]	[0.00416]	[0.00444]	[0.00771]	[0.00632]	[0.00987]	[0.00783]	[0.0120]	[0.00906]	[0.0148]
CRUX (t)	-0.177**	-0.274***	-0.296**	-0.529***	-0.373**	-0.679***	-0.416*	-0.710**	-0.503*	-0.738**
	[0.0742]	[0.0937]	[0.129]	[0.154]	[0.175]	[0.205]	[0.227]	[0.277]	[0.283]	[0.332]
Constant	0.158***	0.180***	0.357***	0.416***	0.523***	0.625***	0.659***	0.818***	0.772***	0.992***
	[0.00876]	[0.0177]	[0.0192]	[0.0376]	[0.0301]	[0.0502]	[0.0376]	[0.0580]	[0.0420]	[0.0681]
Industry-Year FE	√	√	√	√	√	√	√	√	√	√
Observations	52,766	14,527	49,147	13,809	45,689	13,077	42,475	12,370	39,438	11,694
R-squared	0.120	0.143	0.135	0.163	0.142	0.167	0.146	0.169	0.146	0.170
Number of Firms	7457	2437	7045	2336	6580	2233	6140	2092	5718	1949

Notes: Standard errors are clustered at firm and year level. Dependent variable is log(K(t+τ))-log(K(t)) where K is firm's total capital stock. PTF (t) measures firm's proximity to technology frontier. Log capital stock and log total employment are included to control for the effects of firm size on dependent variable. Firm idiosyncratic uncertainty (CRUX) is drawn from Handley and Li (2020). Time horizon ranges from 1 to 5 years from base year. All of the regression specification include industry-year fixed effects, where industry is at 3-digit NAICS code level. PTF is winsorized at mean ± 3sd by year, log(capital stock), log(employment) and dependent variable are winsorized at 1% level by year. Column (1), (3), (5), (7), and (9) represent sample of firms that do not have patent granted in the given year. Column (2), (4), (6), (8), and (10) represent sample of firms that have patent granted in the given year.

Table 1.4: PTF and Firm Performance - Patenting vs Non-Patenting Firms

(b) Employment Growth

	Dependent Variable: Log Change in Employment (Δln, t+τ→t)									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Patent = 0	Patent > 0	Patent = 0	Patent > 0	Patent = 0	Patent > 0	Patent = 0	Patent > 0	Patent = 0	Patent > 0
Time Horizon (τ)	1		2		3		4		5	
PTF (t)	0.00940***	0.00993***	0.0108**	0.0137**	0.0119*	0.0167**	0.00996	0.0162	0.00805	0.0184
	[0.00244]	[0.00324]	[0.00472]	[0.00570]	[0.00620]	[0.00784]	[0.00686]	[0.0121]	[0.00696]	[0.0143]
Log Capital Stock (t)	0.0113***	-0.00404	0.0275***	-0.00677	0.0373***	-0.00219	0.0451***	-0.00149	0.0543***	-0.00323
	[0.00356]	[0.00412]	[0.00743]	[0.00816]	[0.00874]	[0.0114]	[0.00928]	[0.0169]	[0.0109]	[0.0200]
Log Employment (t)	-0.0194***	0.000986	-0.0486***	0.000598	-0.0704***	-0.0109	-0.0884***	-0.0196	-0.105***	-0.0227
	[0.00353]	[0.00493]	[0.00763]	[0.0105]	[0.00950]	[0.0144]	[0.0112]	[0.0205]	[0.0138]	[0.0240]
CRUX (t)	0.0585	-0.144*	0.0582	-0.204	-0.167	-0.444	-0.222	-0.605*	-0.332	-0.577
	[0.0851]	[0.0773]	[0.115]	[0.165]	[0.150]	[0.257]	[0.187]	[0.286]	[0.221]	[0.372]
Constant	-0.0335*	0.0505**	-0.0920**	0.0869**	-0.115**	0.102*	-0.134***	0.133	-0.156***	0.162
	[0.0161]	[0.0189]	[0.0342]	[0.0395]	[0.0404]	[0.0556]	[0.0432]	[0.0793]	[0.0510]	[0.0944]
Industry-Year FE	√	√	√	√	√	√	√	√	√	√
Observations	62,011	16,090	55,784	14,987	50,483	13,981	46,022	13,076	42,119	12,285
R-squared	0.064	0.094	0.075	0.099	0.081	0.096	0.086	0.095	0.089	0.093
Number of Firms	10185	2984	9106	2743	8089	2515	7246	2297	6498	2091

Notes: Standard errors are clustered at firm and year level. Dependent variable is log(emp(t+τ))-log(emp(t)) where emp is firm's total employment. PTF (t) measures firm's proximity to technology frontier. Log capital stock and log total employment are included to control for the effects of firm size on dependent variable. Firm idiosyncratic uncertainty (CRUX) is drawn from Handley and Li (2020). Time horizon ranges from 1 to 5 years from base year. All of the regression specification include industry-year fixed effects, where industry is at 3-digit NAICS code level. PTF is winsorized at mean ± 3sd by year, log(capital stock), log(employment) and dependent variable are winsorized at 1% level by year. Column (1), (3), (5), (7), and (9) represent sample of firms that do not have patent granted in the given year. Column (2), (4), (6), (8), and (10) represent sample of firms that have patent granted in the given year.

Table 1.4: PTF and Firm Performance - Patenting vs Non-Patenting Firms

(c) Output Growth

	Dependent Variable: Log Change in Output (Δln, t+τ→t)									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Patent = 0	Patent > 0	Patent = 0	Patent > 0	Patent = 0	Patent > 0	Patent = 0	Patent > 0	Patent = 0	Patent > 0
Time Horizon (τ)	1		2		3		4		5	
PTF (t)	0.0118***	0.0184***	0.0179***	0.0258***	0.0196***	0.0315***	0.0164**	0.0337**	0.0168**	0.0345**
	[0.00387]	[0.00453]	[0.00496]	[0.00733]	[0.00628]	[0.00898]	[0.00676]	[0.0120]	[0.00774]	[0.0152]
Log Capital Stock (t)	-0.0211***	-0.0258***	-0.0211***	-0.0365***	-0.0131	-0.0451***	-0.00866	-0.0494***	-0.00440	-0.0508**
	[0.00364]	[0.00792]	[0.00649]	[0.0108]	[0.00814]	[0.0140]	[0.00795]	[0.0160]	[0.00893]	[0.0180]
Log Employment (t)	0.0168***	0.0207**	0.00762	0.0252**	-0.00826	0.0250	-0.0227**	0.0208	-0.0352***	0.0145
	[0.00363]	[0.00778]	[0.00642]	[0.0110]	[0.00812]	[0.0158]	[0.00902]	[0.0183]	[0.00983]	[0.0201]
CRUX (t)	0.177	0.0593	0.172	0.118	0.0340	0.111	0.0505	0.0678	0.0725	0.0599
	[0.102]	[0.143]	[0.159]	[0.287]	[0.210]	[0.384]	[0.239]	[0.476]	[0.296]	[0.511]
Constant	0.142***	0.169***	0.176***	0.259***	0.172***	0.345***	0.180***	0.409***	0.190***	0.453***
	[0.0177]	[0.0354]	[0.0316]	[0.0526]	[0.0387]	[0.0692]	[0.0390]	[0.0793]	[0.0438]	[0.0874]
Industry-Year FE	√	√	√	√	√	√	√	√	√	√
Observations	60,010	15,879	54,117	14,791	49,080	13,786	44,833	12,902	41,138	12,108
R-squared	0.065	0.079	0.066	0.084	0.065	0.083	0.068	0.080	0.073	0.079
Number of Firms	9955	2956	8913	2721	7896	2493	7074	2274	6353	2072

Notes: Standard errors are clustered at firm and year level. Dependent variable is log(output(t+τ))-log(output(t)) where output is firm's total output (sales + inventory). PTF (t) measures firm's proximity to technology frontier. Log capital stock and log total employment are included to control for the effects of firm size on dependent variable. Firm idiosyncratic uncertainty (CRUX) is drawn from Handley and Li (2020). Time horizon ranges from 1 to 5 years from base year. All of the regression specification include industry-year fixed effects, where industry is at 3-digit NAICS code level. PTF is winsorized at mean ± 3sd by year, log(capital stock), log(employment) and dependent variable are winsorized at 1% level by year. Column (1), (3), (5), (7), and (9) represent sample of firms that do not have patent granted in the given year. Column (2), (4), (6), (8), and (10) represent sample of firms that have patent granted in the given year.

Table 1.4: PTF and Firm Performance - Patenting vs Non-Patenting Firms

(d) Sales Growth

	Dependent Variable: Log Change in Output (Δln, t+τ→t)									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Patent = 0	Patent > 0	Patent = 0	Patent > 0	Patent = 0	Patent > 0	Patent = 0	Patent > 0	Patent = 0	Patent > 0
Time Horizon (τ)	1		2		3		4		5	
PTF (t)	0.0113***	0.0160***	0.0172***	0.0242***	0.0190***	0.0319***	0.0155**	0.0319**	0.0152*	0.0313*
	[0.00383]	[0.00499]	[0.00506]	[0.00774]	[0.00629]	[0.00947]	[0.00686]	[0.0123]	[0.00767]	[0.0152]
Log Capital Stock (t)	-0.0219***	-0.0259***	-0.0220***	-0.0333***	-0.0134	-0.0417**	-0.00972	-0.0479***	-0.00717	-0.0491**
	[0.00354]	[0.00771]	[0.00653]	[0.0108]	[0.00810]	[0.0143]	[0.00785]	[0.0163]	[0.00889]	[0.0181]
Log Employment (t)	0.0174***	0.0201**	0.00775	0.0218*	-0.00904	0.0208	-0.0235**	0.0181	-0.0344***	0.0105
	[0.00358]	[0.00769]	[0.00612]	[0.0107]	[0.00752]	[0.0157]	[0.00866]	[0.0181]	[0.00965]	[0.0201]
CRUX (t)	0.170	0.0234	0.145	0.0529	0.0280	-0.0375	0.0278	-0.1000	-0.00418	-0.147
	[0.110]	[0.117]	[0.167]	[0.226]	[0.214]	[0.322]	[0.237]	[0.394]	[0.286]	[0.452]
Constant	0.149***	0.173***	0.185***	0.248***	0.178***	0.337***	0.191***	0.411***	0.211***	0.458***
	[0.0176]	[0.0352]	[0.0320]	[0.0520]	[0.0384]	[0.0703]	[0.0383]	[0.0802]	[0.0428]	[0.0876]
Industry-Year FE	√	√	√	√	√	√	√	√	√	√
Observations	60,546	15,929	54,675	14,844	49,631	13,835	45,378	12,943	41,673	12,152
R-squared	0.064	0.078	0.066	0.081	0.066	0.082	0.070	0.082	0.075	0.081
Number of Firms	10008	2967	8968	2736	7944	2504	7125	2284	6397	2080

Notes: Standard errors are clustered at firm and year level. Dependent variable is log(sales(t+τ))-log(sales(t)) where sales is firm's total sales. PTF (t) measures firm's proximity to technology frontier. Log capital stock and log total employment are included to control for the effects of firm size on dependent variable. Firm idiosyncratic uncertainty (CRUX) is drawn from Handley and Li (2020). Time horizon ranges from 1 to 5 years from base year. All of the regression specification include industry-year fixed effects, where industry is at 3-digit NAICS code level. PTF is winsorized at mean ± 3sd by year, log(capital stock), log(employment) and dependent variable are winsorized at 1% level by year. Column (1), (3), (5), (7), and (9) represent sample of firms that do not have patent granted in the given year. Column (2), (4), (6), (8), and (10) represent sample of firms that have patent granted in the given year.

Table 1.4: PTF and Firm Performance - Patenting vs Non-Patenting Firms

(e) TFP Growth (OLS)

	Dependent Variable: Log Change in TFP (Δln, OLS, t+τ→t)									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Patent = 0	Patent > 0	Patent = 0	Patent > 0	Patent = 0	Patent > 0	Patent = 0	Patent > 0	Patent = 0	Patent > 0
Time Horizon (τ)	1		2		3		4		5	
PTF (t)	0.00584	0.00527	0.00766	0.00963**	0.0101*	0.0160***	0.00761	0.0156**	0.0146*	0.0143*
	[0.00339]	[0.00356]	[0.00476]	[0.00452]	[0.00514]	[0.00460]	[0.00552]	[0.00580]	[0.00727]	[0.00715]
Log Capital Stock (t)	-0.0151***	-0.0150**	-0.0156**	-0.0132*	-0.0127*	-0.0160	-0.00703	-0.0172	-0.00417	-0.0173
	[0.00269]	[0.00620]	[0.00540]	[0.00717]	[0.00701]	[0.00992]	[0.00854]	[0.0120]	[0.00968]	[0.0125]
Log Employment (t)	0.0168***	0.00928	0.0158**	0.00428	0.0123	0.00752	0.00482	0.00752	-0.00256	0.00486
	[0.00321]	[0.00643]	[0.00583]	[0.00647]	[0.00770]	[0.00979]	[0.0101]	[0.0117]	[0.0108]	[0.0118]
CRUX (t)	0.0687	0.155	0.0488	0.143	0.0752	0.250	0.101	0.333	0.0708	0.243
	[0.0626]	[0.111]	[0.103]	[0.163]	[0.145]	[0.206]	[0.163]	[0.262]	[0.170]	[0.287]
Constant	0.0674***	0.0746**	0.0681***	0.0720**	0.0539*	0.0783*	0.0263	0.0805	0.0159	0.0874
	[0.0104]	[0.0264]	[0.0213]	[0.0296]	[0.0270]	[0.0426]	[0.0347]	[0.0507]	[0.0397]	[0.0565]
Industry-Year FE	√	√	√	√	√	√	√	√	√	√
Observations	18,429	10,263	16,241	9,461	14,434	8,768	12,953	8,183	11,679	7,659
R-squared	0.056	0.060	0.058	0.072	0.061	0.075	0.067	0.071	0.070	0.069
Number of Firms	3613	1868	3146	1676	2757	1511	2435	1377	2148	1240

Notes: Standard errors are clustered at firm and year level. Dependent variable is log(TFP(t+τ))-log(TFP(t)) where TFP is firm's total factor productivity eastimated under OLS. PTF (t) measures firm's proximity to technology frontier. Log capital stock and log total employment are included to control for the effects of firm size on dependent variable. Firm idiosyncratic uncertainty (CRUX) is drawn from Handley and Li (2020). Time horizon ranges from 1 to 5 years from base year. All of the regression specification include industry-year fixed effects, where industry is at 3-digit NAICS code level. PTF is winsorized at mean ± 3sd by year, log(capital stock), log(employment) and dependent variable are winsorized at 1% level by year. Column (1), (3), (5), (7), and (9) represent sample of firms that do not have patent granted in the given year. Column (2), (4), (6), (8), and (10) represent sample of firms that have patent granted in the given year.

Table 1.4: PTF and Firm Performance - Patenting vs Non-Patenting Firms

(f) TFP Growth (OP 1996)

	Dependent Variable: Log Change in TFP (Δln, Olley Pakes 1996, t+τ→t)									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Patent = 0	Patent > 0	Patent = 0	Patent > 0	Patent = 0	Patent > 0	Patent = 0	Patent > 0	Patent = 0	Patent > 0
Time Horizon (τ)	1		2		3		4		5	
PTF (t)	0.00589*	0.00619*	0.00778	0.0112**	0.0110**	0.0186***	0.00840	0.0186***	0.0149*	0.0178**
	[0.00334]	[0.00351]	[0.00454]	[0.00465]	[0.00493]	[0.00474]	[0.00544]	[0.00605]	[0.00709]	[0.00753]
Log Capital Stock (t)	-0.0199***	-0.0184***	-0.0242***	-0.0189**	-0.0246***	-0.0238**	-0.0213**	-0.0273**	-0.0204*	-0.0285**
	[0.00282]	[0.00602]	[0.00555]	[0.00700]	[0.00724]	[0.00976]	[0.00875]	[0.0118]	[0.00971]	[0.0122]
Log Employment (t)	0.0216***	0.0129*	0.0246***	0.0106	0.0247***	0.0161	0.0194*	0.0186	0.0142	0.0174
	[0.00322]	[0.00629]	[0.00602]	[0.00632]	[0.00808]	[0.00964]	[0.0103]	[0.0115]	[0.0110]	[0.0116]
CRUX (t)	0.0633	0.152	0.0392	0.136	0.0647	0.262	0.0843	0.341	0.0500	0.255
	[0.0634]	[0.110]	[0.111]	[0.165]	[0.153]	[0.209]	[0.170]	[0.268]	[0.176]	[0.286]
Constant	0.247***	0.250***	0.422***	0.416***	0.579***	0.588***	0.718***	0.757***	0.869***	0.922***
	[0.0105]	[0.0256]	[0.0218]	[0.0288]	[0.0277]	[0.0420]	[0.0353]	[0.0498]	[0.0399]	[0.0552]
Industry-Year FE	√	√	√	√	√	√	√	√	√	√
Observations	18,429	10,263	16,241	9,461	14,434	8,768	12,953	8,183	11,679	7,659
R-squared	0.067	0.079	0.070	0.093	0.073	0.099	0.083	0.099	0.089	0.106
Number of Firms	3613	1868	3146	1676	2757	1511	2435	1377	2148	1240

Notes: Standard errors are clustered at firm and year level. Dependent variable is log(TFP(t+τ))-log(TFP(t)) where TFP is firm's total factor productivity eastimated under Olley and Pakes (1996). PTF (t) measures firm's proximity to technology frontier. Log capital stock and log total employment are included to control for the effects of firm size on dependent variable. Firm idiosyncratic uncertainty (CRUX) is drawn from Handley and Li (2020). Time horizon ranges from 1 to 5 years from base year. All of the regression specification include industry-year fixed effects, where industry is at 3-digit NAICS code level. PTF is winsorized at mean ± 3sd by year, log(capital stock), log(employment) and dependent variable are winsorized at 1% level by year. Column (1), (3), (5), (7), and (9) represent sample of firms that do not have patent granted in the given year. Column (2), (4), (6), (8), and (10) represent sample of firms that have patent granted in the given year.

Table 1.5: Innovation Activities and PTF Summary Statistics

VARIABLES	(1) N	(2) mean	(3) sd	(4) min	(5) max
PTF (t)	98,150	-0.0156	0.987	-4.255	4.085
Sales Growth (Δln, t-1)	98,150	0.119	0.435	-1.890	2.645
Leverage (t-1)	98,150	0.807	2.541	-13.51	23.02
Employment (ln, t-1)	98,150	-0.430	2.265	-6.908	5.258
Tobin's Q (ln, t-1)	98,150	0.525	0.705	-0.856	6.049
CRUX (t-1)	98,150	0.0407	0.0324	0	0.540
		Full Sample Innovation Measure			
R&D (t-1)	51,396	0.127	0.267	0	5.429
KPSS Patent Value (t-1)	86,338	0.0162	0.0778	0	1.626
R&D (t-2)	49,729	0.134	0.281	0	3.109
R&D (t-3)	46,681	0.134	0.279	0	2.891
		Positive Innovation Mesure			
R&D (t-1)	39,759	0.166	0.314	0.000402	6.314
KPSS Patent Value (t-1)	17,019	0.107	0.306	8.81e-05	6.066
R&D (t-2)	38,528	0.177	0.332	0.000402	3.989
R&D (t-3)	36,184	0.177	0.334	0.000422	3.821

Notes: This table provides summary statistics on the effects of previous innovation measure on PTF. Control variables are winsorized at 1% and 99% by year, PTF is winsorized by 3 standard deviations from mean by year. Missing R&D reports are dropped, KPSS patent value is replaced by zero for non-patenting firms in full sample innovation measure.

Table 1.6: Innovation Activities and PTF

(a) Full Sample

	Dependent Vairable: Proximity to Technology Frontier (PTF, t)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Measure & Horizon	R&D (t-1)		R&D (t-2)		R&D (t-3)		KPSS Patent Value (t-1)	
Innovation Measure	0.0772***	0.0406**	0.0840***	0.0603***	0.0720***	0.0604***	0.172***	0.145***
	[0.0160]	[0.0165]	[0.0157]	[0.0157]	[0.0155]	[0.0153]	[0.0424]	[0.0423]
Leverage (t-1)		0.000175		0.000619		0.000807		0.000647
		[0.00158]		[0.00160]		[0.00166]		[0.00104]
Employment (ln, t-1)		-0.000622		0.000607		0.00247		-0.00566
		[0.00771]		[0.00770]		[0.00799]		[0.00574]
Tobin's Q (ln, t-1)		0.0439***		0.0436***		0.0428***		0.0333***
		[0.00725]		[0.00741]		[0.00788]		[0.00631]
Sales Growth (Δln, t-1)		0.0306***		0.0279***		0.0233***		0.0244***
		[0.00653]		[0.00666]		[0.00718]		[0.00543]
CRUX (t-1)		1.452***		1.458***		1.460***		1.697***
		[0.201]		[0.207]		[0.211]		[0.158]
Constant	0.0324***	-0.0620***	0.0136***	-0.0809***	-0.0118***	-0.105***	0.0472***	-0.0413***
	[0.00198]	[0.0108]	[0.00208]	[0.0111]	[0.00204]	[0.0114]	[0.000695]	[0.00706]
Firm FE	√	√	√	√	√	√	√	√
Industry-Year FE	√	√	√	√	√	√	√	√
Observations	50,295	50,295	48,715	48,715	45,707	45,707	84,899	84,899
R-squared	0.684	0.685	0.680	0.682	0.676	0.677	0.685	0.686
Number of Firms	5871	5871	5661	5661	5244	5244	10363	10363

Notes: Standard errors are clustered at firm level. Dependent variable is proximity to technology frontier (PTF) at year t. Innovation measure captures 1, 2, or 3 year lagged R&D expenditure or 1 year lagged economic value of patent based on Kogan, Papanikolaou, Seru and Stoffman (2017). Control variables include leverage, log total employment, log Tobin's Q, log sales growth, and idiosyncratic uncertainty (CRUX) drawn from Handley and Li (2020). All control variables are 1 year lagged. All of the regression specification include firm fixed effects and industry-year fixed effects, where industry is at 3-digit NAICS code level. PTF is winsorized at mean ± 3sd by year, other variables (except for CRUX) are winsorized at 1% level by year.

Table 1.6: Innovation Activities and PTF

(b) Positive R&D and Patent

	Dependent Vairable: Proximity to Technology Frontier (PTF, t)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Measure & Horizon	R&D (t-1)		R&D (t-2)		R&D (t-3)		KPSS Patent Value (t-1)	
Innovation Measure	0.0667***	0.0327**	0.0695***	0.0479***	0.0601***	0.0505***	0.0745***	0.0531**
	[0.0146]	[0.0152]	[0.0144]	[0.0144]	[0.0138]	[0.0135]	[0.0237]	[0.0232]
Leverage (t-1)		0.000122		0.000406		0.000464		0.00351
		[0.00179]		[0.00178]		[0.00186]		[0.00310]
Employment (ln, t-1)		0.00439		0.00674		0.0112		0.00514
		[0.00915]		[0.00910]		[0.00949]		[0.0192]
Tobin's Q (ln, t-1)		0.0482***		0.0486***		0.0482***		0.0506***
		[0.00786]		[0.00804]		[0.00857]		[0.0164]
Sales Growth (Δln, t-1)		0.0330***		0.0298***		0.0236***		0.0286*
		[0.00707]		[0.00721]		[0.00772]		[0.0150]
CRUX (t-1)		1.334***		1.332***		1.394***		1.973***
		[0.229]		[0.236]		[0.240]		[0.431]
Constant	0.0560***	-0.0401***	0.0374***	-0.0581***	0.0105***	-0.0858***	0.0374***	-0.0881***
	[0.00238]	[0.0133]	[0.00250]	[0.0136]	[0.00238]	[0.0140]	[0.00264]	[0.0255]
Firm FE	√	√	√	√	√	√	√	√
Industry-Year FE	√	√	√	√	√	√	√	√
Observations	38,780	38,780	37,629	37,629	35,283	35,283	16,039	16,039
R-squared	0.677	0.679	0.674	0.676	0.668	0.670	0.696	0.697
Number of Firms	4577	4577	4430	4430	4084	4084	2348	2348

Notes: Standard errors are clustered at firm level. Dependent variable is proximity to technology frontier (PTF) at year t. Innovation measure captures 1, 2, or 3 year lagged R&D expenditure or 1 year lagged economic value of patent based on Kogan, Papanikolaou, Seru and Stoffman (2017). Control variables include leverage, log total employment, log Tobin's Q, log sales growth, and idiosyncratic uncertainty (CRUX) drawn from Handley and Li (2020). All control variables are 1 year lagged. All of the regression specification include firm fixed effects and industry-year fixed effects, where industry is at 3-digit NAICS code level. PTF is winsorized at mean ± 3sd by year, other variables (except for CRUX) are winsorized at 1% level by year. The sample is restricted to positive innovation measures.

Table 1.7: PTF and Mergers Summary Statistics

VARIABLES	(1) N	(2) mean	(3) sd	(4) min	(5) max
PTF Similarity (t-1)	15,824,898	0.0504	0.0725	2.55e-06	1
Competition (HP, t-1)	11,908,707	0.109	0.0975	0	0.967
Merger Deal Indicator	15,824,898	0.000108	0.0104	0	1
|Δ PTF| (t-1)	15,824,898	0.000990	0.000938	0	0.00908
|Δ PTF Breadth| (t-1)	15,824,898	3.42e-05	3.29e-05	0	0.000224
|Δ log(total asset)| (t-1)	15,824,898	2.264	1.899	0	15.92
|Δ cash| (t-1)	15,824,898	1.276	1.135	0	7.856
|Δ sales growth (ln)| (t-1)	15,824,898	0.378	0.477	0	4.410
|Δ log(Tobin's Q)| (t-1)	15,824,898	0.520	0.678	0	6.885
|Δ book leverage| (t-1)	15,824,898	0.240	0.584	0	15.24
|Δ ROA| (t-1)	15,824,898	0.289	0.933	0	20.93
Adjusted by Industry (4-digit NAICS code) Average					
|Δ PTF| (t-1)	15,824,898	0.000970	0.000921	0	0.00877
|Δ PTF Breadth| (t-1)	15,824,898	3.34e-05	3.20e-05	0	0.000298
|Δ log(total asset)| (t-1)	15,824,898	2.104	1.754	0	18.14
|Δ cash| (t-1)	15,824,898	1.152	1.050	0	9.252
|Δ sales growth (ln)| (t-1)	15,824,898	0.374	0.472	0	4.831
|Δ log(Tobin's Q)| (t-1)	15,824,898	0.485	0.625	0	7.877
|Δ book leverage| (t-1)	15,824,898	0.241	0.573	0	15.24
|Δ ROA| (t-1)	15,824,898	0.299	0.914	0	20.93

Notes: This table provides summary statistics for analysis on merger and acquisitions. The sample is created by taking Cartesian interaction of all possible mergers from industries that involve a merger deal in the given year. There are total 1707 announced deals. Industry is classified at 4-digit NAICS code. PTF is winsorized at 3 standard deviation from mean by year. Total assets, sales growth, Tobin's Q, book leverage, and ROA are winsorized at 1% and 99% by year.

Table 1.8: PTF and Determinants of Mergers

(a) PTF and Mergers

	Dependent Variable: Indicator of Merger Deal							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES					Industry Average Adjusted Measure			
|Δ PTF| (t-1)	0.0164***	0.0134***	0.0161***	0.0132***	0.0183***	0.0151***	0.0180***	0.0148***
	[0.00350]	[0.00346]	[0.00431]	[0.00430]	[0.00360]	[0.00357]	[0.00443]	[0.00444]
PTF Similarity (t-1)	0.00159***	0.00155***	0.00121***	0.00121***	0.00159***	0.00157***	0.00121***	0.00122***
	[0.000311]	[0.000303]	[0.000289]	[0.000287]	[0.000311]	[0.000306]	[0.000290]	[0.000289]
|Δ PTF Breadth| (t-1)		0.383***		0.362***		0.384***		0.347***
		[0.101]		[0.129]		[0.101]		[0.127]
Competition (HP, t-1)			0.00130***	0.00126***			0.00130***	0.00129***
			[0.000206]	[0.000202]			[0.000206]	[0.000205]
|Δ log(total asset)| (t-1)		-0.000689***		0.000309		-0.000218		0.000802***
		[0.000168]		[0.000226]		[0.000175]		[0.000245]
|Δ cash| (t-1)		-0.00213***		-0.00201***		-0.00165***		-0.00124***
		[0.000280]		[0.000397]		[0.000280]		[0.000407]
|Δ sales growth (ln)| (t-1)		-0.00378***		-0.00346***		-0.00384***		-0.00355***
		[0.000565]		[0.000789]		[0.000554]		[0.000781]
|Δ log(Tobin's Q)| (t-1)		-0.00516***		-0.00542***		-0.00537***		-0.00572***
		[0.000536]		[0.000757]		[0.000529]		[0.000771]
|Δ book leverage| (t-1)		-4.78e-06		-0.00157		-0.000682*		-0.00266**
		[0.000360]		[0.00122]		[0.000399]		[0.00129]
|Δ ROA| (t-1)		-0.000131		-0.00348***		0.000221		-0.00190***
		[0.000229]		[0.000697]		[0.000218]		[0.000627]
Acq Ind×Tar Ind×Year FE	√	√	√	√	√	√	√	√
Observations	15,824,898	15,824,898	11,908,707	11,908,707	15,824,898	15,824,898	11,908,707	11,908,707
R-squared	0.004	0.004	0.005	0.005	0.004	0.004	0.005	0.005

Notes: Standard errors are clustered at acquirer industry - target industry - year level. Dependent variable is an indicator = 1 if a merger deal is announced and = 0 otherwise. |Δ PTF| is the absolute value of the difference of PTF of the acquirer and the target. PTF Similarity measures the pairwise similarity of firm-level technology frontier. |Δ PTF Breadth| captures the absolute value of the difference of technology frontier breadth of the acquirer and the target. Competition is drawn from Hoberg and Phillips (2016) to gauge pairwise product competition between firms. Control variables include the absolute value of the difference of total asset, cash holding, sales growth, Tobin's Q, book leverage, and ROA, of the acquirer and the target, adjusted by dividing 100. All independent variables except for PTF Similarity and Competition are demeaned by industry average in columns (5)-(8). Acquirer industry - target industry - year fixed effects are included in all specifications. Industry is defined at 4-digit NAICS level.

Table 1.8: PTF and Determinants of Mergers

(b) PTF, Assets and Mergers

	Dependent Variable: Indicator of Merger Deal							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES					Industry Average Adjusted Measure			
|Δ PTF| (t-1)	0.0174***	0.0141***	0.0154***	0.0124***	0.0189***	0.0155***	0.0174***	0.0144***
	[0.00365]	[0.00360]	[0.00437]	[0.00436]	[0.00371]	[0.00366]	[0.00447]	[0.00449]
PTF Similarity (t-1)	0.00157***	0.00156***	0.00122***	0.00122***	0.00159***	0.00158***	0.00122***	0.00122***
	[0.000309]	[0.000305]	[0.000290]	[0.000288]	[0.000310]	[0.000307]	[0.000291]	[0.000289]
|Δ PTF Breadth| (t-1)		0.387***		0.363***		0.386***		0.348***
		[0.101]		[0.129]		[0.101]		[0.128]
Competition (HP, t-1)			0.00130***	0.00126***			0.00131***	0.00129***
			[0.000205]	[0.000202]			[0.000206]	[0.000205]
|Δ PTF| (t-1) × |Δ log(total asset)| (t-1)	-0.525***	-0.556***	-0.654***	-0.667***	-0.383**	-0.411***	-0.378*	-0.388*
	[0.136]	[0.136]	[0.201]	[0.202]	[0.149]	[0.149]	[0.220]	[0.219]
|Δ log(total asset)| (t-1)	-0.000560***	-0.000110	0.000680**	0.00100***	-0.000165	0.000203	0.00103***	0.00120***
	[0.000205]	[0.000208]	[0.000305]	[0.000310]	[0.000222]	[0.000229]	[0.000329]	[0.000333]
|Δ cash| (t-1)		-0.00213***		-0.00201***		-0.00165***		-0.00124***
		[0.000280]		[0.000397]		[0.000280]		[0.000407]
|Δ sales growth (ln)| (t-1)		-0.00377***		-0.00346***		-0.00383***		-0.00354***
		[0.000566]		[0.000789]		[0.000554]		[0.000781]
|Δ log(Tobin's Q)| (t-1)		-0.00517***		-0.00542***		-0.00538***		-0.00572***
		[0.000536]		[0.000758]		[0.000529]		[0.000771]
|Δ book leverage| (t-1)		-2.38e-05		-0.00161		-0.000694*		-0.00268**
		[0.000360]		[0.00122]		[0.000399]		[0.00129]
|Δ ROA| (t-1)		-0.000149		-0.00349***		0.000207		-0.00191***
		[0.000228]		[0.000700]		[0.000217]		[0.000627]
Acq Ind×Tar Ind×Year FE	√	√	√	√	√	√	√	√
Observations	15,824,898	15,824,898	11,908,707	11,908,707	15,824,898	15,824,898	11,908,707	11,908,707
R-squared	0.004	0.004	0.005	0.005	0.004	0.004	0.005	0.005

Notes: Standard errors are clustered at acquirer industry - target industry - year level. Dependent variable is an indicator = 1 if a merger deal is announced and = 0 otherwise. |Δ PTF| is the absolute value of the difference of PTF of the acquirer and the target. PTF Similarity measures the pairwise similarity of firm-level technology frontier. |Δ PTF Breadth| captures the absolute value of the difference of technology frontier breadth of the acquirer and the target. Competition is drawn from Hoberg and Phillips (2016) to gauge pairwise product competition between firms. Control variables include the absolute value of the difference of total asset, cash holding, sales growth, Tobin's Q, book leverage, and ROA, of the acquirer and the target. All independent variables except for PTF Similarity and Competition are demeaned by industry average in columns (5)-(8). Acquirer industry - target industry - year fixed effects are included in all specifications. Industry is defined at 4-digit NAICS level.

Table 1.8: PTF and Determinants of Mergers

(c) PTF, Sales Growth and Mergers

	Dependent Variable: Indicator of Merger Deal							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES					Industry Average Adjusted Measure			
|Δ PTF| (t-1)	0.0149***	0.0124***	0.0139***	0.0114***	0.0169***	0.0142***	0.0163***	0.0134***
	[0.00354]	[0.00352]	[0.00437]	[0.00436]	[0.00363]	[0.00361]	[0.00446]	[0.00447]
PTF Similarity (t-1)	0.00159***	0.00156***	0.00121***	0.00122***	0.00160***	0.00158***	0.00122***	0.00122***
	[0.000311]	[0.000304]	[0.000290]	[0.000289]	[0.000312]	[0.000307]	[0.000290]	[0.000290]
|Δ PTF Breadth| (t-1)		0.375***		0.352***		0.378***		0.340***
		[0.101]		[0.129]		[0.101]		[0.128]
Competition (HP, t-1)			0.00130***	0.00126***			0.00130***	0.00129***
			[0.000206]	[0.000202]			[0.000206]	[0.000205]
|Δ PTF| (t-1) × |Δ sales growth (ln)| (t-1)	-2.579***	-2.416***	-2.803***	-2.658***	-2.053***	-1.930***	-1.967**	-1.882**
	[0.542]	[0.541]	[0.742]	[0.742]	[0.568]	[0.568]	[0.795]	[0.795]
|Δ log(total asset)| (t-1)		-0.000681***		0.000317		-0.000210		0.000808***
		[0.000167]		[0.000226]		[0.000175]		[0.000245]
|Δ cash| (t-1)		-0.00212***		-0.00200***		-0.00165***		-0.00123***
		[0.000280]		[0.000397]		[0.000280]		[0.000407]
|Δ sales growth (ln)| (t-1)	-0.00262***	-0.00145*	-0.00215*	-0.000885	-0.00299***	-0.00202***	-0.00276**	-0.00178
	[0.000788]	[0.000787]	[0.00110]	[0.00111]	[0.000786]	[0.000775]	[0.00109]	[0.00109]
|Δ log(Tobin's Q)| (t-1)		-0.00516***		-0.00541***		-0.00537***		-0.00572***
		[0.000535]		[0.000756]		[0.000529]		[0.000770]
|Δ book leverage| (t-1)		-1.19e-05		-0.00156		-0.000687*		-0.00265**
		[0.000359]		[0.00122]		[0.000399]		[0.00129]
|Δ ROA| (t-1)		-0.000138		-0.00349***		0.000218		-0.00190***
		[0.000229]		[0.000699]		[0.000218]		[0.000628]
Acq Ind×Tar Ind×Year FE	√	√	√	√	√	√	√	√
Observations	15,824,898	15,824,898	11,908,707	11,908,707	15,824,898	15,824,898	11,908,707	11,908,707
R-squared	0.004	0.004	0.005	0.005	0.004	0.004	0.005	0.005

Notes: Standard errors are clustered at acquirer industry - target industry - year level. Dependent variable is an indicator = 1 if a merger deal is announced and = 0 otherwise. |Δ PTF| is the absolute value of the difference of PTF of the acquirer and the target. PTF Similarity measures the pairwise similarity of firm-level technology frontier. |Δ PTF Breadth| captures the absolute value of the difference of technology frontier breadth of the acquirer and the target. Competition is drawn from Hoberg and Phillips (2016) to gauge pairwise product competition between firms. Control variables include the absolute value of the difference of total asset, cash holding, sales growth, Tobin's Q, book leverage, and ROA, of the acquirer and the target. All independent variables except for PTF Similarity and Competition are demeaned by industry average in columns (5)-(8). Acquirer industry - target industry - year fixed effects are included in all specifications. Industry is defined at 4-digit NAICS level.

Table 1.8: PTF and Determinants of Mergers

(d) PTF, Tobin's Q and Mergers

	Dependent Variable: Indicator of Merger Deal							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES						Industry Average Adjusted Measure		
|Δ PTF| (t-1)	0.0155***	0.0128***	0.0147***	0.0117***	0.0171***	0.0143***	0.0164***	0.0133***
	[0.00357]	[0.00354]	[0.00429]	[0.00428]	[0.00366]	[0.00363]	[0.00438]	[0.00439]
PTF Similarity (t-1)	0.00158***	0.00156***	0.00121***	0.00122***	0.00158***	0.00158***	0.00121***	0.00122***
	[0.000309]	[0.000305]	[0.000290]	[0.000289]	[0.000310]	[0.000308]	[0.000291]	[0.000291]
|Δ PTF Breadth| (t-1)		0.378***		0.356***		0.379***		0.341***
		[0.101]		[0.129]		[0.100]		[0.127]
Competition (HP, t-1)			0.00128***	0.00126***			0.00128***	0.00129***
			[0.000203]	[0.000202]			[0.000204]	[0.000204]
|Δ PTF| (t-1) × |Δ log(Tobin's Q)| (t-1)	-1.459***	-1.386***	-1.484***	-1.461***	-1.508***	-1.427***	-1.355**	-1.342**
	[0.308]	[0.307]	[0.476]	[0.474]	[0.352]	[0.351]	[0.563]	[0.561]
|Δ log(total asset)| (t-1)		-0.000682***		0.000314		-0.000208		0.000809***
		[0.000167]		[0.000226]		[0.000175]		[0.000245]
|Δ cash| (t-1)		-0.00212***		-0.00200***		-0.00164***		-0.00123***
		[0.000280]		[0.000398]		[0.000280]		[0.000407]
|Δ sales growth (ln)| (t-1)		-0.00378***		-0.00347***		-0.00384***		-0.00355***
		[0.000564]		[0.000788]		[0.000554]		[0.000781]
|Δ log(Tobin's Q)| (t-1)	-0.00494***	-0.00379***	-0.00528***	-0.00395***	-0.00478***	-0.00401***	-0.00526***	-0.00443***
	[0.000605]	[0.000617]	[0.000912]	[0.000891]	[0.000613]	[0.000633]	[0.000950]	[0.000946]
|Δ book leverage| (t-1)		-1.06e-05		-0.00157		-0.000685*		-0.00267**
		[0.000359]		[0.00122]		[0.000398]		[0.00129]
|Δ ROA| (t-1)		-0.000170		-0.00352***		0.000191		-0.00192***
		[0.000228]		[0.000699]		[0.000216]		[0.000627]
Acq Ind×Tar Ind×Year FE	√	√	√	√	√	√	√	√
Observations	15,824,898	15,824,898	11,908,707	11,908,707	15,824,898	15,824,898	11,908,707	11,908,707
R-squared	0.004	0.004	0.005	0.005	0.004	0.004	0.005	0.005

Notes: Standard errors are clustered at acquirer industry - target industry - year level. Dependent variable is an indicator = 1 if a merger deal is announced and = 0 otherwise. |Δ PTF| is the absolute value of the difference of PTF of the acquirer and the target. PTF Similarity measures the pairwise similarity of firm-level technology frontier. |Δ PTF Breadth| captures the absolute value of the difference of technology frontier breadth of the acquirer and the target. Competition is drawn from Hoberg and Phillips (2016) to gauge pairwise product competition between firms. Control variables include the absolute value of the difference of total asset, cash holding, sales growth, Tobin's Q, book leverage, and ROA, of the acquirer and the target. All independent variables except for PTF Similarity and Competition are demeaned by industry average in columns (5)-(8). Acquirer industry - target industry - year fixed effects are included in all specifications. Industry is defined at 4-digit NAICS level.

Table 1.8: PTF and Determinants of Mergers

(e) PTF, ROA and Mergers

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dependent Variable: Indicator of Merger Deal							
					Industry Average Adjusted Measure			
|Δ PTF| (t-1)	0.0158***	0.0131***	0.0130***	0.0104**	0.0178***	0.0148***	0.0154***	0.0122***
	[0.00349]	[0.00347]	[0.00414]	[0.00414]	[0.00361]	[0.00358]	[0.00434]	[0.00436]
PTF Similarity (t-1)	0.00159***	0.00155***	0.00121***	0.00122***	0.00159***	0.00157***	0.00122***	0.00122***
	[0.000310]	[0.000304]	[0.000290]	[0.000288]	[0.000311]	[0.000307]	[0.000291]	[0.000290]
|Δ PTF Breadth| (t-1)		0.380***		0.355***		0.381***		0.341***
		[0.101]		[0.129]		[0.101]		[0.127]
Competition (HP, t-1)			0.00129***	0.00126***			0.00129***	0.00129***
			[0.000205]	[0.000202]			[0.000205]	[0.000205]
|Δ PTF| (t-1) × |Δ ROA| (t-1)	-0.774***	-0.716***	-2.265***	-2.237***	-0.740***	-0.709***	-2.057***	-2.102***
	[0.134]	[0.133]	[0.590]	[0.580]	[0.143]	[0.144]	[0.646]	[0.641]
|Δ log(total asset)| (t-1)		-0.000681***		0.000317		-0.000211		0.000810***
		[0.000168]		[0.000226]		[0.000175]		[0.000245]
|Δ cash| (t-1)		-0.00213***		-0.00200***		-0.00165***		-0.00124***
		[0.000280]		[0.000397]		[0.000280]		[0.000407]
|Δ sales growth (ln)| (t-1)		-0.00376***		-0.00346***		-0.00383***		-0.00354***
		[0.000565]		[0.000789]		[0.000554]		[0.000781]
|Δ log(Tobin's Q)| (t-1)		-0.00517***		-0.00543***		-0.00538***		-0.00573***
		[0.000536]		[0.000758]		[0.000529]		[0.000771]
|Δ book leverage| (t-1)		-2.49e-05		-0.00167		-0.000697*		-0.00274**
		[0.000358]		[0.00122]		[0.000397]		[0.00129]
|Δ ROA| (t-1)	-0.00159***	0.000529**	-0.00340***	-0.00131	-0.00119***	0.000863***	-0.00165*	9.92e-05
	[0.000279]	[0.000263]	[0.000945]	[0.000949]	[0.000254]	[0.000263]	[0.000945]	[0.000977]
Acq Ind×Tar Ind×Year FE	√	√	√	√	√	√	√	√
Observations	15,824,898	15,824,898	11,908,707	11,908,707	15,824,898	15,824,898	11,908,707	11,908,707
R-squared	0.004	0.004	0.005	0.005	0.004	0.004	0.005	0.005

Notes: Standard errors are clustered at acquirer industry - target industry - year level. Dependent variable is an indicator = 1 if a merger deal is announced and = 0 otherwise. |Δ PTF| is the absolute value of the difference of PTF of the acquirer and the target. PTF Similarity measures the pairwise similarity of firm-level technology frontier. |Δ PTF Breadth| captures the absolute value of the difference of technology frontier breadth of the acquirer and the target. Competition is drawn from Hoberg and Phillips (2016) to gauge pairwise product competition between firms. Control variables include the absolute value of the difference of total asset, cash holding, sales growth, Tobin's Q, book leverage, and ROA, of the acquirer and the target. All independent variables except for PTF Similarity and Competition are demeaned by industry average in columns (5)-(8). Acquirer industry - target industry - year fixed effects are included in all specifications. Industry is defined at 4-digit NAICS level.

CHAPTER II

Measuring the Effects of Firm Uncertainty on Investment: New Evidence on One Million Documents

2.1 Introduction

Risk and uncertainty can be defined carefully and specifically in theory to encompass time-varying aggregate, industry or firm-level shocks. But constructing empirical proxies is difficult in practice, regardless of aggregation levels. This is because uncertainty, by its very nature, is a summary concept for the degree to which the future is unknown. Our approach to measuring uncertainty at the firm level rests on a simple hypothesis: if firms say they are uncertain, then we take their word for it — literally. We measure the usage of variations of the word "uncertainty" relative to the total number of meaningful words found in required company reports filed with the U.S. Securities and Exchange Commission (SEC). We summarize over one million documents from 1994-2016 and build a new panel database of firm-level, time-varying uncertainty measures. With these measures in hand, we can then ask three important but related questions: (1) should we take firm declarations of "uncertainty" at face value; (2) how important are fluctuations in firm-level uncertainty for investment in aggregate and disaggregated data; and (3) does the firm-level variation in uncertainty explain micro and macro fluctuations independently of other aggregate uncertainty measures?

A growing body of recent work investigates the effects of uncertainty on economic activity. The dynamics of uncertainty and investment have been theoretically understood for some time (cf. Bernanke, 1983; Dixit, 1989; Abel and Eberly, 1993). In the presence some form of partially irreversible costs, firms will delay projects and be less responsive to shocks when uncertainty about future business conditions is high. Recent work links reduced investment to realized or forecasted stock return volatility using firm-level data (Leahy and Whited, 1996; Bloom et al., 2007). It can also affect aggregate export dynamics (Novy and Taylor, 2014) or firm decisions to enter and invest in new export markets (Handley, 2014; Handley and Limão, 2015, 2017).

Understanding and measuring uncertainty at the firm-level is important. Recent evidence suggests that idiosyncratic shocks to particular firms play a substantial role in observed aggregate fluctuations (Gabaix, 2011), but evidence on uncertainty shocks is limited and mixed. Bachmann and Bayer (2013; 2014) suggest that small shocks to firm productivity are important to explaining pro-cyclical productivity dispersion using German firm-level data, but are not a driver of major business cycles. In contrast, Bloom et al. (forth.) find evidence that micro uncertainty rises in recessions and that uncertainty shocks lead to reductions in GDP, productivity and reallocation in a heterogeneous firms DSGE model. To better understand these dynamics, we build firm-level measures of uncertainty from textual analysis of company reports that we can use to explore both the micro and macroeconomic dynamics. This has two important advantages. First, we can explore firm-specific fluctuations in uncertainty in a long panel against performance outcomes that include investment rate at both the firm and establishment level using data from both COMPUSTAT and the U.S. Census firm and establishment microdata. Second, we can aggregate up from the firm-level time variation to a macro index of uncertainty and explore the implications for aggregate investment and GDP growth rate.

Constructing quantitative measures of uncertainty, even at the aggregate level, is fraught with difficulty. As Jurado et al. (2015) show, many common proxies for uncertainty may reflect a mixture of first moment shocks and other sources of variation unrelated to fundamental uncertainty. For example, aggregate proxies such as the CBOE Volatility Index (VIX) measure the volatility of the S&P 500 index implied by equity call options on its underlying components. But the prices of these options and their implied volatility can move substantially even if the underlying fundamentals of the business enterprise are unchanged. Nevertheless, these derived uncertainty proxies operate like uncertainty measures in empirical work. Barerro et al. (2017) and Stein and Stone (2013) find that increases in implied volatility derived from equity options reduces hiring and investment. Unfortunately, there are no traded equity options for large number of public firms, including components of the S&P 500 Index.

Other measures at the firm level are qualitative and derived from surveys of firms or professional forecasters. What these measures lack in quantitative precision is offset, at least partially, by the fact that they are sourced from managers, insiders, and professionals with a detailed knowledge of specific companies or industry sectors. For example, the Federal Reserve Board's Greenbook and other proprietary professional forecasts contain significant narrative components. Sharpe et al. (2017) show the tonality of Fed Board forecasts has predictive power for GDP growth and inflation. Moreover, financial market participants and consumers must value this information given that they continue to pay for the analysis. Similarly, surveys of managers should capture the uncertainty about future business conditions by the decision-making economic agents within the firm. Bachmann et al. (2013) find in German and U.S. data that disagreement in forecasts precede reductions in output driven by "wait-and-see" dynamics that are short-lived in Germany and more persistent in the United States.[1] Guiso and Parigi (1999) use survey data on Italian firms and find that uncertainty about future demand reduces investment. More recently, Bloom et al. (2017) describe evidence from the Management and Organization Practices Survey (MOPS) on the business expectation of managers in the manufacturing sector. They find managers responses to questions about future sales and production expectations are logically consistent and strongly correlated with realized first and second moments in the data.

Our approach measures business uncertainty by analyzing the documents filed with the U.S. Securities and Exchange Commission (SEC). All publicly traded companies are required to file form 10-K annually and quarterly form 10-Q's. These forms report the firm's activities and financial information to investors, shareholders and the public.[2] For each document we measure the frequency of the word "uncertainty" and its variations relative to the total number words to construct a firm-specific **C**ompany **R**eported **U**ncertainty Inde**X**, or **CRUX**. There are several advantages to this measures relative to previous work. First, the basic index methodology is consistent with the other text-based aggregate measures such as the BBD EPU index, but it is available at the firm-level. Second, it relates to work using survey data to extract forecast errors and firms' subjective probability distributions. The filings are not a survey where we ask specific questions of firms, but we build our measure from what is effectively a mandatory census of all publicly traded companies that includes a mixtures of free-form written responses and financial information. Third, our methodology can easily be extended to longer time series and used in subsequent research as new document filings are added to the EDGAR database.

[1]Morikawa (2016) also uses forecast survey data and finds a negative relationship between uncertainty and investment.

[2]The SEC phased-in electronic filing from 1994-1996 and makes the forms available through EDGAR, the Electronic Data Gathering, Analysis, and Retrieval system. We describe nature of forms and our methods in the data section.

In Figure 2.1(a) we plot a quarterly time-series of our baseline CRUX measure from 1998 to 2016. The firm-level measures components are demeaned by firm and then aggregated up by taking a simple mean so it is centered on zero.[3] For comparison, we also plot the VIX. Both measures are standardized for comparison to have mean zero and unit standard deviation. Our CRUX measure rises during well known periods of higher uncertainty: the 1999-2001 period of the tech bubble, September 11 and recession; the Iraq War (2003); the Great Recession (2007-09); and the 2016 election cycle. The comparison with the Baker, Bloom and Davis (2016) index of economic policy uncertainty, hereafter BBD EPU, in Figure 2.1(b) in similar.

Naturally, there is a concern that the index captures a mixture of first and second moment shocks, or other unrelated noise. For example, firms may cite uncertainty when times are bad as an excuse for poor performance. To preview some of our results, we find this is the clearly not the case for several reasons. First, the measure is correlated with other quantitative measures like the VIX and qualitative measures such as the BBD EPU index. Nevertheless, the VIX and EPU indices may also combine first and second moment shocks. Second, in aggregate regressions we find increases in CRUX are negatively correlated with growth in GDP, Gross Private Domestic Investment (GPDI), and the growth in the GPDI/GDP ratio even after we control for other standard predictors. To demonstrate, in Figure 2.2 we plot the semi-parametric fit of the log change in both real GPDI and the gross investment/GDP ratio to the aggregate CRUX measure. This figure already partials out a first moment control, the change in the S&P 500 index, and a second moment control, the VIX. The relationship is clearly negative for real gross investment in the left panel. On the right, we see this relationship is robust even after we normalize by GDP and not driven purely by unobserved shocks to

[3] We describe the construction of this index is more detail in Section 2.2. To aggregate up, we compute a total word weighted firm-level aggregate for the current quarter and 3 lags. We then demeaned each firm's reported uncertainty measure by it's sample mean to remove firm-specific variation. We take an average over all firms by quarter.

aggregate conditions. We will show these findings are robust in subsequent aggregate and firm-level regressions.

Building on this motivating evidence, we then verify the uncertainty mechanism by linking the firm-level version of the CRUX to microdata on investment from two sources: (1) publicly available investment data in COMPUSTAT available at the firm level and (2) the confidential U.S. Census microdata on firm *and* establishment dynamics we construct detailed investment data from the Annual Survey of Manufactures (ASM) and the Census of Manufactures (CM). Using this rich microdata, we need not infer the presence of uncertainty by computing cross-section dispersion in growth rates of investment, productivity, or sales within industries or for the whole economy—measures that may be endogenous to firm decisions.

After controlling for time invariant firm effects and industry-by-year shocks, we can then identify first-order investment delay effects from uncertainty, but more importantly we can also test for second-order caution effects — attenuated responsiveness of investment to demand shocks when uncertainty is high. We leverage the establishment level variation within firm to absorb unobserved firm-level first moment shocks through a complete set of firm-year fixed effects. Similar to across firm variation, within firm reallocation in response to uncertainty is non-trivial in the data and has a strong theoretical basis in the investment under uncertainty literature (cf. Bloom, 2009; Bloom, Bond, and van Reenen, 2007).

A final concern is that the 10-K and other associated company reports do not contain accurate and reliable information disclosures. There is a broader empirical literature on textual analysis, to which we contribute, that suggests this is not the case. First, compliance is mandatory and there are penalties for making false or misleading statements. Second, a number of studies using contextual information for SEC filings surveyed in Loughran and McDonald (2016) and Li (2010) find that language, tone, sentiment and specific words are predictive of firm behavior, identify important firm characteristics, or explain other economic and financial outcomes. For example, Hoberg and Phillips (2016) construct new measures of product characteristics and industry network linkages; Bodnaruk et al. (2015) use 10-K text to measure financial constraints; Buehlmaier and Whited (2018) show textual measures of financial constraints help explain equity returns; Avramov et al. (2016) show contextual measures of downside risk affect numerous corporate policies; and Li et al. (2013) develop measures of market competition consistent with existing measures that contain additional new information. Third, it's possible that managers make persistent forecast errors and under/over-estimate volatility as shown by Ben-David et al. (2013). But even if we concede that firm statements about uncertainty are revealed *ex post* to be misguided or driven by cognitive biases (see Kahneman, 2003) they still reflect thinking that can affect managerial decisions and firm outcomes in important ways. Our task is to understand and empirically quantify that channel.

We contribute to a broader literature on uncertainty that uses text- and news-based measures of uncertainty. The seminal work in this area is the Baker, Bloom and Davis (2016) index of Economic Policy Uncertainty (BBD EPU). They construct their measure from searches of major news sources that contain variations of the words "economic", "policy", and "uncertainty" normalized by the total number of news articles within a given time period. This measure is correlated with major anecdotal episodes of policy uncertainty; increases in the EPU index are associate with lower investment rates, hiring, and downturns in GDP and investment. Similar measures have been constructed for a number of other countries, are widely disseminated on-line, and have been used more follow-up studies than can be usefully cited.[4] More recent approaches include topic modeling analysis of textual information using computational

[4]These measures are available from www.policyuncertainty.com and can also be downloaded from FRED.

linguistic methods (e.g. Huang et al., forth.). Hassan et al. (2017) use pattern-based sequence-classification method to generate measures of firm-level political uncertainty derived from quarterly earnings conference calls. Similarly, Shepotylo and Stuckatz (2017) measure trade policy uncertainty in Ukraine and the firm-level response of FDI and export market entry and exit relative to the likelihood of trade agreement with the European Union.

The rest of the this paper is organized as follows. In Section 2.2, we describe our underlying data source of SEC mandatory filings and the construction of the uncertainty index. Section 2.3 begins with firm and establishment level evidence the investment response to our measure of uncertainty. After validating our measure in a standard investment framework we turn to employment effects. Section 2.4 provides evidence of using aggregate data and Section 2.5 concludes.

2.2 Measurement and Validation

2.2.1 Measuring Uncertainty in Context

To measure uncertainty, we exploit the text of SEC reports from the EDGAR Database. All public domestic firms are required to make reports to U.S. Securities and Exchange Commission (SEC) on a regular basis. For example, the firms must submit annual reports (Form 10-K), quarterly reports (Form 10-Q) and current reports (Form 8-K) which must comply with certain disclosure requirements.

We parse and match SEC EDGAR reports against a dictionary of English words from all 10-K, 10-Q reports and their amendments each year. Our sample includes 1,000,313 documents — all 10-Q and 10-K forms, their variations, and amendments — filed by 41,418 firms from 1994 to 2016. Each filing can be exactly identified by three factors: a Central Index Key (CIK), filing date, and filing type (10-K, 10-Q or amendments). The CIK identifier is used to match with data from COMPUSTAT. When two companies merge or a company changes their name, the CIK of the surviving/new entity can be associated with the new name by updating the company profile in the EDGAR database.[5]

We count the total number of uncertain words each in each document from the list {*uncertain*, *uncertainty*, *uncertainties*, *uncertainly*}. We then aggregate by firm CIK identifiers (indexed by i) and filing period (year and quarter, indexed by t) over all the forms to obtain total Total uncertain words$_{it}$. We normalize this count by the total number of meaningful words to compute the Company Reported Uncertainty Index (CRUX) by firm and filing period

$$\text{CRUX}_{it} = \frac{\text{Total uncertain words}_{it}}{\text{Total number of meaningful words}_{it}} \times 100.$$

The denominator, total number of meaningful words, counts all the words that are present in the filing, but excludes all the stop words, e.g. 'a', 'an', 'the', etc. from the total count.[6] We scale the measure up by 100 so that it measures percentage point frequency.

In Table B1, we report raw summary statistics on over 1 million parsed SEC EDGAR documents. First, we note that the average document contains 4.4 words that are forms of the word "uncertainty." The median count is 3 and there is a high variability across documents ranging from zero to 133 with a standard deviation of 5.7. In part, this is due to commingling quarterly 10-Q and annual 10-K reports along with mixing larger firms that have complex business structures and smaller firms that may have only a few product lines. Because of this variation across form and firm types, we normalize by total word count (excluding stop words). The stop word adjustment

[5]http://www.merrilldirect.com/cps/rde/xchg/merrilldirect/hs.xsl/edgar-getting-started-with-edgar-filing.htm

[6]We provide a list of these words in the appendix. Our results are robust to counting all words.

drops the average word count in a document from over 19,000 to just over 11,000 words. Our results are robust to normalization by all words in a document, but slightly less precise.

2.2.2 Examples in Context

The reason we focus exclusively on "uncertain" words is that we intend to measure subjective expressions of uncertainty. Other words may be of interest for different applications, but do not have the same connotation. For example, words such as "risk" may just indicate lines of business, i.e. "risk management firm" and insurance companies, or describe objective and measurable risks that firm has taken.[7] Moreover, firms may generally have a positive outlook about future sales or profits, an expectation about the future, but still express uncertainty around that outlook. The evidence that we are measuring some degree of uncertainty ultimately lies in the empirical results we discuss in Section 2.3. But to help understand the source of our measure for some well-known companies, we provide the following three contextual examples:

Apple Computer: global and regional shocks

> "The Company's operations and performance depend significantly on global and regional economic conditions. ***Uncertainty*** about global and regional economic conditions poses a risk as consumers and businesses postpone spending in response to tighter credit, higher unemployment, financial market volatility, government austerity programs, negative financial news, declines in income or asset values and/or other factors." (Apple, 2013 10-K, Risk Factors)

General Motors: unspecified generic uncertainty and risk

[7] We show in robustness checks that a measure based on "risk"-related words does not have a robust impact on investment.

> "We face a number of significant risks and ***uncertainties*** in connection with our operations. Our business, results of operations and financial condition could be materially adversely affected by the factors described below."
> (GM, 2010 10-K, Risk Factors)

Apple describes a number of traditional sources of uncertainty that are both regional and global. One could argue these are measurable aggregate or industry shocks. But while these could affect all firms, they may be more important to Apple because it operates in multiple jurisdictions and sells products all over the world. For example, the volatility in the Dollar to Renminbi exchange rate may indirectly affect nearly all U.S. firms, but they are more likely to discuss that source of uncertainty if it is important to their business lines. So even if these reported global uncertainties are simply Apple's report of aggregate shocks, they are being mentioned because they have some specific relevance to Apple's outlook on the future.

In other cases, reports are similar to General Motors. Sources of uncertainty specific to its operations are mentioned. Our index measure will count this mention of uncertainty and if it doesn't contain meaningful information our ultimate empirical results should find little effect.

Wal-Mart: litigation uncertainty

> "However, because of the ***uncertainty*** of the outcome of the appeal from the District Courts certification decision, because of the ***uncertainty*** of the balance of the proceedings contemplated by the District Court, and because the Company's liability, if any, arising from the litigation, including the size of any damages award if plaintiffs are successful in the litigation or any negotiated settlement, could vary widely, the Company cannot reasonably estimate the possible loss or range of loss which may arise from the

litigation." (Walmart, 2006 first 10-Q, Financial Information)

Harley-Davidson: policy uncertainty

> "The European Union has enacted tariffs on various U.S.-manufactured products, including Harley-Davidson motorcycles. [...] risks and ***uncertainties*** include the following, among other factors: (i) ***uncertainties*** regarding the quantity and mix of motorcycles that the company exports from the U.S. during the periods in question; (ii) ***uncertainties*** regarding the import prices of motorcycles; (iii) whether the EU tariffs apply to shipments that had already commenced at the effective time of the tariffs; (iv) ***uncertain*** timing associated with shifting production from the U.S. to international facilities; and (v) ***uncertainties*** regarding the size and duration of EU tariffs." (Harley-Davidson, June 2018, 8-K, Regulation FD Disclosure)

We highlight the Walmart disclosure as it concerns litigation uncertainties about potential financial liabilities that are difficult to objectively quantify. Likewise, the Harley-Davidson excerpt is a specific discussion of the unilateral trade policy actions by the U.S. and the EU that have impacted Harley's costs and overseas pricing. While the additional costs can be accounted for there is uncertainty about shipments timing, product mix, and the duration of tariffs. Moreover, these discussions take place in a quarterly 10-Q (Wal-mart) and 8-K disclosure (Harley) both of which would be missed if we focused only on the 10-K Risk Factors discussions.

2.3 Firm-level Estimation and Quantification

We employ three different data sources at the firm-level to estimate panel regressions of the effect of uncertainty on investment growth. First, we use total corporate investment from COMPUSTAT to estimate the broader effects of firm uncertainty on investment across all sectors of the economy. Second, we turn to detailed establishment level investment data for the manufacturing sector from the Census Bureau's Annual Survey of Manufactures and Census of Manufactures. The latter investment data are closely aligned with canonical models of investment under uncertainty.

2.3.1 Identification and Estimation Approach

To understand our identification strategy and its limits we proceed in two steps. First, we describe the requirements to identify effects of firm-specific exposure to uncertainty and show they hold in our data. Second, we discuss our estimation equation and how we handle other threats to identification.

To fix ideas, we let $\text{CRUX}_{it} = f(\mu_{it}, \sigma_{it}) + \varepsilon_{it}$. This makes explicit that our measure is a function of firm-specific uncertainty σ_{it}, possible first moment shocks μ_{it}, and measurement error ε_{it}. We then take a first order Taylor approximation around firm-level mean deviations in $\bar{\mu}_i$ and $\bar{\sigma}_i$ to write

$$\text{CRUX}_{it} = f(\bar{\mu}_i, \bar{\sigma}_i) + f_\mu(\bar{\mu}_i, \bar{\sigma}_i)(\mu_{it} - \bar{\mu}_i) + f_\sigma(\bar{\mu}_i, \bar{\sigma}_i)(\sigma_{it} - \bar{\sigma}_i) + e_{it} \tag{2.1}$$

where e_{it} is higher-order approximation and measurement error. A key assumption is that $f_\sigma(\cdot)$ is positive so that CRUX_{it} captures a relative ranking of more or less uncertain states across firms. The approximation also makes clear that identification requires within firm, time variation in uncertainty, i.e. $\text{Var}(\sigma_{it} - \bar{\sigma}_i) > 0$. Otherwise, differences across firms may be driven entirely by time-invariant uncertainty (and first moment) differences in a firm's business environment.

Given these requirements, one might worry that corporate reports only reflect aggregate fluctuations and differences in firm-invariant characteristics in most years. To address this concern, we decompose the variance of the CRUX measure into aggregate, industry, firm fixed effects, and firm-year components. We regress CRUX measure on a large set of fixed effects and report our results in Table 2.2.[8] Column (1) reports the R-squared when CRUX is regressed one-by-one on a set of fixed effects for time, industry, industry-time, and firm fixed effects in each row. The largest components are a firm effect, 49% of variation, and time effect, 21% of variation. Industry effects contribute a small amount unless they are interacted with time effects,i.e. industry-year shocks. In Column (2) we additively regress CRUX on time fixed effects, industry fixed effects, industry-year fixed effects, and then firm fixed effects. The remaining residual variation in the CRUX measure is 37% after controlling for firm and industry-time fixed effects. So clearly the CRUX measure captures aggregate fluctuations and firm-invariant characteristics, but still retains significant amount of variation at firm-year level. The latter variation is used for identification in all our firm and establishment level regressions.

Next we turn to our baseline empirical model to estimate the effect of uncertainty on investment. We have panel model where a firm (using COMPUSTAT) or an establishment (ASM/CM) over time. Specifically, our regression function is

$$\Delta y_{it} = \lambda \text{CRUX}_{it} + \eta \text{CRUX}_{it} \cdot \Delta \log(sales)_{it-1} + \beta^{sales} \Delta \log(sales)_{it-1} + \boldsymbol{\beta} \cdot \boldsymbol{X_{it-1}} + \alpha_i + \alpha_t + \epsilon_{it}.$$

The dependent variable is log growth rates of investment (corporate, equipment, structures) computed as the change in the capital stock. We include firm-level effects α_i and time fixed effects α_t to identify idiosyncratic uncertainty, as discussed above, and any other unobservable firm characteristics and aggregate shocks that might influence Δy_{it}.

Our primary interest is on the two RHS terms that contain the CRUX measure.

[8] These results use the matched SEC EDGAR - COMPUSTAT sample we describe below.

First, we include CRUX_{it}, the contextual measure of uncertainty from SEC reports described above.[9] The coefficient on λ measures first order effect of uncertainty on the outcome. In the presence of any non-convex adjustment costs or irreversible sunk costs of decisions, we expect to find a "delay" effect whereby new investment and capital improvements are reduced when uncertainty is high. So we predict that $\lambda < 0$. We also include the interaction term $\text{CRUX}_{it} \cdot \Delta \log(sales)_{it-1}$ to estimate a second order "caution" effect in response to shocks. Specifically, it measures how firms respond to demand shocks under uncertainty. Caution effects will attenuate investment response when uncertainty is high. Hiring and investment typically increase in response to sales shocks so that $\beta^{sales} > 0$. So we predict attenuation from the caution effect such that $\eta < 0$.

The main threat to identification of caution and delay effects is that CRUX_{it} may also capture first moment shocks, rather than uncertainty. Here, we again refer to equation (2.1), where the dependence on first moments is explicit. The first two solutions are econometric and the third is more subtle. Let the first moment parameter be decomposed as follows: $\mu_{it} = \bar{\mu}_i + \bar{\mu}_t + \tilde{\mu}_{it}$.

First, our data are already differenced and we also include firm fixed effects (α_i) that absorb persistent firm shocks to μ_i, firm idiosyncratic growth trends, and the propensity by specific firms to use "uncertainty"-type words more frequently. We also include year fixed effects (α_t) and ultimately industry-year effects that absorb industry and aggregate demand shocks in μ_t.

Second, we include in the the vector of controls $\boldsymbol{X_{it-1}}$ a set of first moment controls that proxy for $\tilde{\mu}_{it}$: average Tobin's Q ($\log(q_{it-1})$) to a proxy of firms' investment opportunities, lagged log sales growth ($\Delta \log(sales)_{it-1}$)to control for firm level demand

[9] We take CRUX_{it} instead of $\text{CRUX}_{i,t-1}$ because our measure is based on date of filing, which reports mixture of firm information in the previous year or quarter and forward looking statements about the future.

shocks, and squared log sales growth to captures nonlinear effects.

Third, finding cautionary effects is an important piece of evidence in favor of the CRUX as a firm-level uncertainty measure. Suppose firms use "uncertainty" as a catch all term, or even an excuse, to describe bad demand or cost shocks when $\mu_t < 0$ or $\tilde{\mu}_{it} < 0$. Alternatively, the same firm may not mention uncertainty all after a positive sales shock, e.g. claiming that "we knew it all along." That would negatively bias estimate toward $\lambda < 0$. But the same would not hold for the predicted, negative cautionary effect of $\eta < 0$. To see this, suppose a firm gets a negative sales shock, $\Delta \log(sales)_{it-1} < 0$, and responds by declaring that the world is a very uncertain place, i.e. driving up CRUX$_{it}$. If we suppose CRUX, in the extreme, contains only first moment shocks, then it is an inverse measure of demand or sales; we would spuriously find $\hat{\lambda} < 0$. But this also means that the sales growth and CRUX interaction term would be the product of a negative sales shock and an inverse demand proxy that is large and positive. If we maintain our predictions that $\lambda < 0$ and $\beta^{sales} > 0$, then this suggests that we should estimate $\eta > 0$, i.e. employment or investment decline more when negative sales shocks are large. The latter only occurs if CRUX$_{it}$ is a proxy, in whole or in part, for first moment shocks. In our results below, we clearly reject $\eta > 0$ and this suggests that on average the CRUX measure is capturing some element of firm-specific uncertainty.

A related concern is that uncertainty is endogenously generated by low investment growth. Because our panel data allows us to control for many of these shocks through fixed effects and we include a number of first moment controls, this issue is not likely to be severe in our application. Moreover, if the feedback from outcomes to uncertainty were strong, then we should expect CRUX to be negatively correlated with across all investment decision margins within the firm.

Given the set of fixed effects we employ and the examples noted in section 2.2.2, it's important to be careful interpreting the resulting estimates. To see this, note that our linear approximation of CRUX in (2.1) applies to each firm i. The coefficient on uncertainty exposure $f_\sigma(\bar{\mu}_i, \bar{\sigma}_i)$ is firm specific. This can matter in the estimation if we want to know whether the effect of uncertainty on an outcome is a firm-specific reaction to its exposure to aggregate uncertainty or a response to its own firm-specific uncertainty. We operationalize this by letting $\sigma_{it} = \sigma_t^A + \tilde{\sigma}_{it}$. We can consider σ_t^A aggregate uncertainty and $\tilde{\sigma}_{it} = \tilde{\sigma}_{it} - \sigma_t^A$ are firm deviations.

In our baseline, we estimate the average partial effect of CRUX on an outcome. The estimated $\hat{\gamma}$ is the average of firm level heterogeneous responses given by $\gamma_i \times \text{CRUX}_{it}$ and therefore $\gamma = E(\gamma_i)$. But firms may respond heterogeneously to aggregate *and* firm specific shocks: $\gamma_i^A \sigma_t^A + \tilde{\gamma}_i \tilde{\sigma}_{it} = (\gamma_i^A - \tilde{\gamma}_i)\sigma_t^A + \tilde{\gamma}_i \sigma_{it}$. If $\tilde{\gamma}_i \approx 0$ and $E(\gamma_i^A) < 0$, then we would still estimate a negative effect of uncertainty because of heterogeneity in how firms respond to aggregate uncertainty. Alternatively, if $\gamma_i^A = \gamma^A$, then it is absorbed by time or industry-time fixed effects out. In that case $\gamma = E(\tilde{\gamma}_i)$ is identified from the firm response to its own uncertainty. Using our rich firm-level data we can rule out our results are driven purely to by heterogeneity in response to aggregate or industry-specific uncertainty. We do so by estimating different slope coefficients directly on aggregated version of CRUX using several different methods and firm data. In the results that follow, we find both channels are important.

2.3.2 Firm and Establishment Level Data

Our firm or establishment level outcomes are draw from publicly available data in the COMPUSTAT database and confidential microdata from the U.S. Census.

2.3.2.1 COMPUSTAT

Firm information is drawn from COMPUSTAT data on firm balance sheets, cash flow and income statements. We match the CRUX measure with COMPUSTAT through firm identifier CIK and year from fiscal years ending from 1994 to 2016. After removing missing dependent and independent variables, we have over 95,000 observations on about 11,800 firms in our investment panel.

We measure the investment rate by taking the log difference of firm level capital stock between two consecutive years, i.e. $\log(K_{it}) - \log(K_{it-1})$. Capital stocks are computed through the perpetual-inventory method. [10]

We capture firms' demand shocks by lagged sales log growth and firms' investment opportunities by lagged Tobin's Q. Specifically, we compute

$$\Delta \log(\text{sales}_{it-1}) = \log(\text{sales}_{it-1}) - \log(\text{sales}_{it-2})$$

We obtain sales directly from COMPUSTAT and find no significant difference when we use reported revenues instead. We also calculate squared sales log growth as demand shock might have convex effects on firms' investment or hiring decisions. Our measure of lagged log Tobin's Q is

$$\text{Tobin's Q} = \frac{\text{Market Capitalization} + \text{Market Value of Liability}}{\text{Total Asset Value}}.$$

We compute market capitalization as common shares outstanding (*csho*) $\times$ price closed

[10] Specifically,

$$K_{it} = \pi_t((1-r)K_{it-1} + I_{it-1}),$$

where π_t is the producer price index ratio between year t and $t-1$, I_{it-1} is the capital investment (*capx*) and the initial capital stock of each firm K_{i0} is measured by total property, plant and equipment (*ppent*). The producer price index by commodity for finished goods (capital equipment) is aggregated from seasonally adjusted monthly data. If the value of capital investment is missing for a single year, it is interpolated with the mean of the preceding and following values.

at fiscal year ($prcc\_f$). Market value of liability is assumed to be approximately equal to the book value of liability and calculated by total asset (at) $-$ total common/ordinary equity (ceq). The summary statistics of COMPUSTAT sample (investment and employment) are reported in Table 2.1.

2.3.2.2 US Census Microdata

To study the effects of uncertainty on firms' investment behavior in further detail, we rely on high quality investment and capital measures from Annual Survey of Manufacturing (ASM), and Census of Manufactures (CM) in data collected by the U.S. Census Bureau. This permits us to link our firm level measure directly to establishment measures of total investment broken out into structures and equipment. These databases also provide information on total value of shipments, manufacturing industry codes (6-digit NAICS), and capital expenditures broken out by equipment and structures. We use these measures to construct log changes in capital stock (investment) and total shipments at the establishment level.[11]

Firm identifiers in ASM and CM are linked to COMPUSTAT through COMPUSTAT-BR (Census Business Register) bridge. This allows us to link financial data and our CRUX measure to the firm and establishment level data in the Census. Specifically, the bridge provides annual link between a Compustat CUSIP and firm identifier in LBD. We then link CRUX measure to ASM/CM data through CIK-CUSIP-ASM/CM identifiers for 1994-2013.[12] Table 2.4 provides summary statistics of matched COMPUSTAT-ASM/CM sample.

[11] We use a Census Bureau provided measure of capital stock computed using the perpetual inventory method. Details on methodology are in Appendix A of Foster et al. (2016).

[12] The COMPUSTAT-BR bridge in our approved project ends in 2011, but we track the extant matched firms in 2011 through 2013.

2.3.3 Firm Level Corporate Investment Results - COMPUSTAT

Before turning to confidential microdata, we focus on corporate investment measured using COMPUSTAT. We show a robust negative effect on investment from delay and caution effects to our text-based measure of uncertainty.

2.3.3.1 Motivating Figures

To Motivate the main regression results, we derive non-parametric and binsreg figures of investment rate on uncertainty. A robust negative first order delay effects can be visualized in Figure 2.3 and Figure 2.4. Figure 2.3 plots the binscatter (Cattaneo et al. 2019) of CRUX against investment rate fitted by second order polynomial without controls. The investment rate is reduced by almost a half when uncertainty (CRUX) rises from low (0) to high (0.15). Our next pass at firm-level investment data is simple non-parametric evidence of reductions in investment growth rate. We divide the sample into high and low uncertainty by the median value of CRUX. We estimate a the kernel density of the investment rate distribution for high CRUX vs low CRUX firms, and plot the result in Figure 2.4. The low uncertainty distribution is shifted right and we reject equality via a Kolmogorov-Smirnov test.

To obtain evidence on second order caution effects, we apply binscatter of sales growth against investment rate under high and low uncertainty. The investment sample is split the same way as in Figure 2.4 and the binscatters are plotted separately. After controlling for year fixed effects, it is clear that sales-investment line under high uncertainty is flatter, which implies that the response of investment to high sales growth is attenuated by high uncertainty.

2.3.3.2 Regression Results

Regression evidence in Table 2.3 confirms the negative impact of uncertainty, as measured by CRUX, on the corporate investment rate. In column (1) we simply regress firms' log investment rate on CRUX with firm and year fixed effects and find a negative and significant coefficient. This is the delay effect we predicted. The estimated effect is robust to including NAICS 3 digit industry-year effects in column (2). Column (3) adds log sales growth (demeaned within sample) and the interaction term between CRUX and log sales growth. We find a positive coefficient on sales growth, as expected, and a negative coefficient on the interaction term. The latter reflects the second-order caution effect we predicted.

We then add other controls for first-moment shocks that are standard in the investment literature. To control for non-linearities in the adjustment of investment to sales, we add the squared log sales growth rate in column (4) and the other coefficients only change slightly. We can also confirm the convex impact of demand shocks on corporate investment. In column (5) we introduce log Tobin's Q to our regression to capture the potential investment opportunities.[13] In column (6) we run the regression with all control variables and our results are robust.

To quantify the effects, we focus on column (3) and compute the effect of a one standard deviation above the mean shocks for delay and caution effects. A one SD above the mean increase in CRUX will result in a 2.3 log point ($= -0.303 \times 0.076 \times 100$) decrease in the corporate investment rate. The coefficient of the interaction term reflects the second order cautionary effect. Firms with high uncertainty would reduce investment rate even when they face high demand growth (sales growth). This caution effect is also reasonably large. A one SD above the mean shock to sales would increase investment by about 4.7 log points ($= 0.0775 \times 0.607 \times 100$). That effect would be

[13] Note that this is average Q rather than marginal Q.

attenuated by 2.3 log points ($= -0.494 \times 0.076 \times 0.607 \times 100$), or 48% of the total effect, if it was accompanied by a one standard deviation shock to CRUX.

2.3.4 Establishment Level Manufacturing Investment

Next we turn to our detailed establishment level manufacturing investment. We show that CRUX has a negative effect on uncertainty and quantify the impact.

The summary statistics for the ASM/CM sample appear in Table 2.4. We have roughly the same mean of 0.0409 for the CRUX measure as in the corporate investment sample. There are more observations relative to the COMPUSTAT sample because these are establishment level data, but fewer firms overall because all non-manufacturing firms are dropped.[14]

Starting in Table 2.5 with total investment growth rates, the sum of equipment and structures investment, we find a strong first order delay effect in columns 1-3 even after we control for sales growth (measured as the change in total value of establishment shipments) and industry-year fixed effects. In column (4) we add a demeaned interaction of sales growth with CRUX. The coefficient on sales growth alone is positive, but it's interaction with CRUX is negative and significant. This is the attenuation through a caution effect that we predicted. We add controls for firm-level Tobin's Q and the square of sales growth in columns (5) and (6). These variables are significant, but don't affect coefficients the CRUX measure even when we include all of them together in column (7).

In Section 3.3.1 we noted caution effects were important evidence that CRUX was not simply of proxy for first moment shocks. We have already shown the caution effect is robust to inclusion of a battery of controls a fixed effects. In columns (8) and (9), we go one step further and take advantage of the *establishment level* variation

[14] Exact breakdowns across the samples cannot be provided due disclosure requirements.

in investment rates by adding a firmid-year fixed effect. This controls for all firm specific shocks to supply, demand, and the firm component of any unobserved first *and* second moment shocks, such as firm's changing tone towards corporate reports and introduction of new mandatory sections to 10-K filings (e.g., the phase-in of Item 1A Risk Factors after 2006). Both the CRUX measure and any firm-specific controls like Tobin's Q are not identified. But because we have establishment level sales growth measures we can identify the investment effect of sales growth and its interaction with the firm-level CRUX measure.

In column (8), we see that the caution effect coefficient remains negative and significant and the coefficient on sales growth is positive. The magnitudes of identified coefficients are nearly unchanged in the saturated regression and robust to including the square of sales growth in column (9). These regression models strongly suggest that CRUX captures meaningful variation in uncertainty on average, even if some firms report uncertainty erroneously or use it as a catch-all term for negative shocks. Moreover, the stability of our caution effect coefficients to the inclusion of firmid-year effects strongly suggests we have adequately controlled for endogenous feedback of firm performance to CRUX in the less saturated baseline specification where we can still identify delay effects.

We repeat these specifications in Table 2.7, breaking out equipment and structures investment separately to better understand the mechanism through which uncertainty impacts investment. The delay effect is primarily driven by equipment investment as seen in the left hand panel. There is a negative effect on structures, but it is not significant. The latter may be due to the slower moving nature of investment in structures. For example, commitments to remodel a plant or repair a roof may respond more to long-run, persistent uncertainty. Whereas machine replacement or retooling that can be more easily delayed. Nevertheless, the caution effects are negative and significant for both types of investment and robust in columns (2) and (4) to the inclusion of firm-year fixed effects.[15]

To visualize the effects of uncertainty, we take the coefficients from column (6) and plot establishment investment rate response to log sales growth under uncertainty in Figure 2.6. Each curve plots the investment response at the 5th, 25th, 50th, 75th and 95th percentile of our CRUX measure.[16] As uncertainty increases, the response curve shifts downward – a delay effect – and the slope of the curve flattens out – a caution effect. We also plot the response curve when CRUX is equal to one standard deviation above the mean in the regression sample, which about about equal to a 75th percentile uncertainty shock.

We quantify the delay and caution effects relative to one SD shocks above mean CRUX within sample, as we did for corporate investment. A one SD shock to CRUX above mean (0.07) reduces the investment rate by 0.5 log points ($= -0.0681 \times 0.07 \times 100$). At more than half the average rate in sample (0.818), this is economically significant. If CRUX is zero, a one SD shock to total shipment growth increases investment by 1.75 log points ($= 0.0437 \times 0.401 \times 100$). But if uncertainty is also high, that effect is attenuated by 0.88 log points ($= -0.314 \times 0.07 \times 0.401 \times 100$), or slightly more than 50%. Figure 2.7 decomposes the effects of high uncertainty (one standard deviation above mean) into delay and caution effects. The delay effect is independent of sales growth and represented by a level shift in investment rates. The magnitude of the caution effect is increasing with respect to sales growth, which we show as a further rotation of the marginal effect in the figure. Specifically, for a one SD shock to shipment growth, the caution effect (0.88 log points) is more than 70% higher than the delay effect (0.5 log points) under high uncertainty.

[15] A full set of results showing this breakout across all controls is available on request

[16] These percentiles are from EDGAR-COMPUSTAT matched sample including only manufacturing firms (2-digit NAICS code ranging from 31 to 33). We do not use Census-based sample percentiles in order to avoid Census disclosure restrictions.

Finally, we want to put the delay effect on investment into sharper contrast. It is well-known that investment is lumpy. New investment can be small to zero at many establishment and might only cover depreciation in many years (e.g. Cooper and Haltiwanger, 2006). To capture the lumpiness, we create a binary indicator for investment spikes that equals 1 if the arithmetic investment rate is higher than 20% ($\frac{I_{it}}{K_{it-1}} \geq 20\%$). We run a simple linear probability model with this binary indicator as the dependent variable and the same set of RHS variables and controls as our continuous baseline regressions. The results, in Table 2.6, show strong evidence of both delay and caution effects that are robust to our full set of baseline controls in column (2) and firm-year effects in column (3). Taking a CRUX shock that is 1 SD above the mean again, we find the probability of an investment spike declines by almost 1 percent (about 20% below the sample mean spike rate). Moreover, the caution effect also reduces the probability of an investment spike following a sales shock from 1.6 percent under no uncertainty to 0.8 percent when CRUX is high.

In sum, when firms use uncertainty related words in public disclosures their investment behavior is fully consistent with models of investment under uncertainty. The evidence suggests a robust and economically significant link between high uncertainty measured by CRUX and the firm and establishment level investment response. A battery of further robustness tests appears in section 2.3.6, but next we investigate the impact on jobs and industry spillovers.

2.3.5 Industry-level Measurement and Applications

This section addresses several related questions about external validity and aggregation. First, we ask whether our results for publicly traded firms are robust and after adjusting for re-weighting for the propensity we observe of publicly traded firm in the set of all private employers. Second, is there is a common industry component to the CRUX measure that can be captured through industry aggregation? Third, if we control for these industry aggregates, does the firm's own CRUX measure still have explanatory power, i.e. are we only picking up common industry shocks or is the idiosyncratic exposure to uncertainty important? We address these questions across several samples and collect the results here.

2.3.5.1 Propensity Score Weighting

Since the CRUX measure is only available for a set of publicly traded firms, results may not generalize to the set of all private sector employers. To handle this issue, we treat the ASM/CM as the population universe of all manufacturing establishments and estimate propensity scores for publicly traded firms in our sample that we use to inverse probability weight our regressions or construct aggregated CRUX measures.

The propensity scores are constructed by fitting logit specifications for each fiscal year

$$\log \frac{p(\boldsymbol{X_{it}})}{1 - p(\boldsymbol{X_{it}})} = \boldsymbol{\theta_t X_{it}},$$

which implies that $\mathbb{P}(I_{it} = 1 \mid X_{it}) = \frac{1}{1+e^{-\boldsymbol{\theta_t X_{it}}}}$ where I_{it} is the indicator equal to 1 if the firm/establishment is selected in the SEC EDGAR - COMPUSTAT - ASM/CM matched sample. To account for the fact that ASM is survey data and non-random, the control variables $\boldsymbol{X_{it}}$ include establishment characteristics: 4-digit NAICS industry code, employment classes (1-9, 10-19, 20-29, 30-49, 50-99, 100-149, 150-249, 250-499, 500-999, 1000 or more), age class (1-5, 6-10, 11-15, 16-20, 21 years or more), payroll class (1 thousand dollars or less, 1-20, 20-200, 200-1000, 1000 thousands dollars or more), and indicator variable equal to 1 if the firm is included in the COMPUSTAT-BR bridge.[17]. The indicator variable equal to 1 if the establishment is both in the

[17]We choose these classes based on Foster et al. (2016) and the propensity score model in Davis et al. (2014).

ASM/CM sample *and* the establishment operates in two consecutive years so we can compute investment rates.

The inverse propensity scores allow us to estimate a weighted linear regression of the baseline model or to compute CRUX measure for industry level aggregates. Our results are largely unchanged and we report them in Table B13 for total manufacturing investment rate and investment spikes, and Table B14 for equipment and structure investment rate. This is likely because publicly traded firms tend to be larger than private firms in terms of sales and investment and they contribute to a substantial share of aggregate output. So studying the uncertainty dynamics within publicly traded firms may be illuminating about firm behavior in the aggregate economy.

2.3.5.2 Peer Effects and Industry Aggregation

We now turn to the spillovers effect from peer uncertainty in the same industry and determine whether within industry aggregation of our measure explains investment in all manufacturing establishments.

We construct a "Peer" CRUX measure at the firm level (LBD sample[18]) to measure within sample spillovers. First we demean our CRUX measure at the firm-level. Second, for each establishment i, the Peer CRUX is the simple average of the firm-demeaned CRUX of all other firms with operations in the same NAICS 4 digit industry.

In Table 2.8, we regress manufacturing investment rate on Peer CRUX and the firm-level CRUX. The sample is the same set of establishment that appear in manufacturing investment baseline sample. Column (1) includes only the peer measure, where we find both caution and delay effects as in our baseline. In column (2), we add the firm-level CRUX and find both firm and peer caution and delay effects are negative and significant. Moreover, adding the own CRUX measure diminishes the effect of peer

[18]Details of construction in Chapter 3.

uncertainty by a sizable amount. In short, the baseline CRUX measure includes firm-specific variation that is robust to including a measure of a common uncertainty shock in the industry. These findings are robust to adding additional controls for in columns (3-4), or when identifying only the caution effect after including firmid-year fixed effects in columns (5-6).[19]

We have ruled out that our firm-level measure is only capturing industry level uncertainty shocks, but the within sample evidence suggests the firm CRUX does contain industry specific variation in uncertainty. So we construct an industry level CRUX measure by taking the average firm demeaned CRUX of all establishments within the same industry (4-digit NAICS). We can then estimate effects of industry level uncertainty on *all* establishments in U.S. manufacturing sector from the Census data.[20]

In Table 2.10, we regress establishment total investment rate on both a simple mean and propensity socre-weighted industry CRUX. We find strong effects on both delay and caution, and the coefficients barely change no matter how we weight the measure. The magnitude of the effects is sizable. Taking the coefficients from column (1), a one SD shock to industry CRUX above mean (0.015) reduces invest rate by 0.28 log points ($= -0.184 \times 0.015 \times 100$), almost 40% of the average rate in sample (0.654). If CRUX is zero, a one SD shock to total shipments growth increases investment by 0.96 log points ($= 0.0230 \times 0.417 \times 100$). But if uncertainty is also high, that effect is readuced by 0.23 log points ($= -0.370 \times 0.015 \times 0.417 \times 100$), or about 25%. We find similar results when breaking out equipment and structures in Table 2.11 and Table B15.

[19] We also compute industry CRUX measure by taking the simple of propensity score weighted average of CRUX of all firms within the same industry (4-digit NAICS). Results are nearly the same and available on request.

[20] Summary statistics are in Table 2.9.

2.3.6 Robustness and Alternative Measures

We perform several robustness checks that focus on the COMPUSTAT-based corporate investment results in Table 2.3. Our identification of both caution and delay effects across a rich set of fixed effects already indicate that our estimates are robust to unobserved industry- and firm-time varying factors that would bias our results. In what follows, we describe a number of robustness checks against alternative measures of uncertainty.

- *Alternative Investment Rate Measures* To address the concern that capital stock created by perpetual inventory method does not account for unexpected change in capital prices, we construct alternative investment rate measures in the following ways: $\frac{I_{it}}{K_{it}}$ and $\log(\frac{I_{it}}{K_{it}})$ where I_{it} is capital expenditure of firm i at year t drawn from *capx* from COMPUSTAT, and total capital K_{it} is measured by *ppent* (total net property, plant and equipment) or *ppegt* (total gross property, plant and equipment). As capital expenditure and total capital stock are measured in the same time frame, the impact of price change vanishes. We report regression results of the impact of CRUX on alternative measures of investment rate in Appendix Tables B2-B5. Both the signs and magnitudes are robust to our baseline.

- *Alternative NAICS Classifications* We use COMPUSTAT Segments database, which reports firm sales by line of business, to adjust firm level NAICS classification with segment sales, according to Bloom et al. 2019. It allows us to account for the fact that the industry classifications of some conglomerates are not perfectly assigned. The results are reported in Appendix Table B6 and the coefficients barely change compared with our baseline.

- *Measurement Error in CRUX.* To address the measurement error in CRUX, we define a binary uncertainty measure by taking above sample median CRUX as high uncertainty and low otherwise. Appendix Table B7 reports regression results of high vs. low CRUX measure on corporate investment with same specification as continuous CRUX measure. The results are robust to our baseline.

- *Using "Risk"-Related Word Index.* We construct an alternative text based measure the same way we construct CRUX by exploiting "risk"-related word in EDGAR filings. The word list includes the words *risk, risked, riskier, riskiest, riskily, riskiness, risking, risks, risky.* In Appendix Table B8, we find such "risk"-related measure has a positive effect first-order effect on the investment rate and negative effect on second order interactions with sales growth. This suggests that "risk"-words may not be used in the context of uncertainty or second moment shocks in SEC filings. One explanation is that the use of the word "risk" does not describe business conditions, forecasts, etc.. For example, an insurance company may describe managing risk as a business line. Another firm may describe taking on risk or the "upside risk" of a project. These usages are either asymmetric or lead to noise in the measure. Another reason is that while economists have assigned distinct but related meaning to risk and uncertainty, usage in business reporting need not respect those definitions.

- *Alternative Market Based Measure: Realized Volatility.* We compute realized volatility by taking standard deviation of firms' monthly stock returns at year $t-1$. In Appendix Table B9, we control for realized volatility in our baseline regression. The CRUX measure coefficient signs and magnitudes are robust, suggesting it captures additional uncertainty factors not present in backward-looking, realized volatility measure. Realized volatility has a negative first order effect on corporate investment rate, but no caution effect.

- *Alternative Market Based Measure: Implied Volatility.* In Appendix Table B10, we control for implied volatility in our baseline CRUX regression on corporate investment.[21] The CRUX measure is robust when controlling for implied volatility and explanatory power of implied volatility is weak. Another limitation of using implied volatility as an uncertainty measure is that it shrinks the sample by almost two thirds since only a small set of publicly traded firms have exchange traded options. That latter highlights another advantage of our measure: it can be computed for all public firms regardless of market capitalization.

- *Heterogeneity from introduction of "Item 1A: Risk Factors" requirement.* The SEC required a risk factors (Item 1A) discussion in all reports from 2006 forward. This could have increased discussion of uncertainty (or risk) in company filings, but our data clearly show firms were mentioning risk and uncertainty regardless of SEC requirements before 2006. The introduction of Item 1A may have simply spurred a reorganization of company reports. We have already shown the caution effect is robust to firmid-year fixed effect controls, which rules out this change driving our results. We also created a Risk Factors indicator variable equal to 1 if the firm reports text under the Item 1A Risk Factors section of their filings during the year and 0 otherwise.[22] We find no evidence of heterogeneity from a risk factors indicator on delay effects and some mild evidence on caution effects (Table B11).[23] We get similar results when replacing Risk Factors indicators as post 2006 indicator as in Table B12.

[21] See Appendix for detailed construction.

[22] Even though SEC made Item 1A mandatory from 2006, not every firm reports it every year and not every firm started reporting it at the same time. This gives our indicator variable firm-year variation.

[23] The introduction of the Risk Factors section precedes the financial crisis and Great Recession, when uncertainty was high, which is another reason a Risk Factors indicator may be generate heterogeneity in the estimates. A more coarse post-2006 indicator variable applied to all firms regardless of the actual text in their reports suggests this is the case.

2.4 Aggregate Validity and Effects

2.4.1 Aggregation

To demonstrate the validity of our measure in aggregate time series, we construct a quarterly, aggregate CRUX measure. Moving from firm to aggregate, we first construct a 4 quarter word-weighted moving average CRUX by firm. We weight within firm and over 4 quarters using the total number of meaningful words, which gives more weight to longer documents, e.g. 10-K reports, within a 4 quarter period. Using all documents available, we then demean these measures by the CIK firm identifier in EDGAR, essentially removing a firm fixed effect. We take a simple average by quarter of the firm level measures in mean deviations, $CRUX_{it}$, to obtain a 76 period quarterly measure $CRUX_t$ from the first quarter of 1998 to the fourth quarter of 2016.

In the aggregate analysis, we drop observations before 1998 because the electronic filing protocol was phased in over 3 years from 1994 to 1996. As such, 1998 is the first year in which an aggregate moving average measure reflects the filings of all publicly traded firms. This is measure we plot in Figure 2.1 against the VIX and the BBD EPU index.

2.4.2 Gross Investment and GDP

Our aggregate measure is constructed as a mean of firm-level residual deviations from their mean CRUX measure in each document. This is the variation identifying the firm-level estimation of employment and investment effects. Increases in the measure are associated with reductions in gross investment, GDP, and the investment/GDP ratio even after conditioning on first moment controls and other measures of aggregate uncertainty.

We use the quarterly aggregate CRUX that we plotted in Figure 2.1 to explore its relationship with aggregate investment and output. We obtain quarterly data on real GDP, real Gross Domestic Private Investment (GPDI), and the GPDI to GDP ratio. We take annual changes in the quarterly outcomes so that $\Delta \ln Z_t = \ln Z_t - \ln Z_{t-4}$. We then regress the change in these aggregate outcomes on the CRUX index and a set of controls for aggregate conditions that we add one by one. Our ultimate regression specification is:

$$\Delta \ln Z_t = \lambda^Z \mathrm{CRUX}_t + \beta_1^Z \Delta \ln Z_{t-1} + \beta_2^Z \Delta \ln \mathrm{SP500}_t + \beta_3^Z \mathrm{VIX}_t^{MA(4)} + \varepsilon_t. \qquad (2.2)$$

We report results in Table 2.12. These are simple, reduced-form regressions that omit some underlying dynamics and other controls. As such, we focus on first on Panel A, the investment/GDP ratio, because the GDP normalization of gross investment is adjusted for unobserved shocks to aggregate output and prices. Because these are time series data, we report Newey and West (1994) heteroskedasticity and autocorrelation robust standard errors for the OLS regressions in the first 3 columns. In column (1) we includes no controls other than a constant. Consistent with the semi-parametric evidence in Figure 2.2 that included other controls, the relationship is clearly negative and significant. Column (2) shows the effect is robust to adding a first moment control, the change in the S&P 500 Index. In column (3) we add the 4 quarter moving average of the VIX as an alternative second moment shock since our CRUX measure is also 4 quarter weighted, moving average. We then add the 4th lag of the dependent variable, i.e. quarterly year-on-year growth in the previous year, and estimate the model using maximum likelihood (columns labeled ML). We then include all the controls in column (5). Adding all the controls reduces the magnitude of the coefficient of our CRUX measure from -4 to -3, but the negative effect remains robust.[24]

[24] We find similar results when we include quarterly dummies for seasonality, use OLS/IV to estimate the regressions with lagged dependent variables, or use the first lag instead of the 4th lag of the dependent variable.

In Panels B and C of Table 2.12, we analyze real gross investment and real GDP separately. Because they are measured in real terms, Panels B and C are not a perfect decomposition of the Panel A effects, but they are close. Taking the difference of coefficients on columns (1) we find $\hat{\lambda}^{rGPDI} - \hat{\lambda}^{rGDP} = -4.1$, which is close to $\hat{\lambda}^{GPDI/GDP} = -4.3$. We find a robust negative relationship between the aggregate CRUX uncertainty measure and gross investment. It is reduced primarily by the inclusion of the change in the S&P 500 index as a first moment control. While the VIX also has a negative effect on investment, these measures are not perfectly correlated and appear to capture different aspects of uncertainty. The effects of the CRUX measure on aggregate real GDP growth is strongly negative in columns 1-5 when adding controls individually or all together.

In sum, the aggregate dynamics of our CRUX uncertainty measure have a strong negative effect on the change in gross investment even after normalizing by GDP. These effects are large in comparison to traditional measures like the VIX. For example, we can compare one standard deviation shock to CRUX and VIX using the estimates in Panel B, column (5). A one standard deviation shock to the VIX would reduce investment growth by 1 log point ($=-0.0016 \times 6.4 \times 100$). In contrast, a one SD shock to the CRUX would reduce investment growth by 2.7 log points ($= -4.37 \times 0.0061 \times 100$). The latter is comparably large given it is the conditional effect after controlling for the VIX, the S&P 500, and lagged investment growth. However, we don't place too much weight on these magnitudes as our regressions omit a number of higher order dynamics and controls that should be included in aggregate time-series regressions for investment and GDP.

2.5 Conclusion

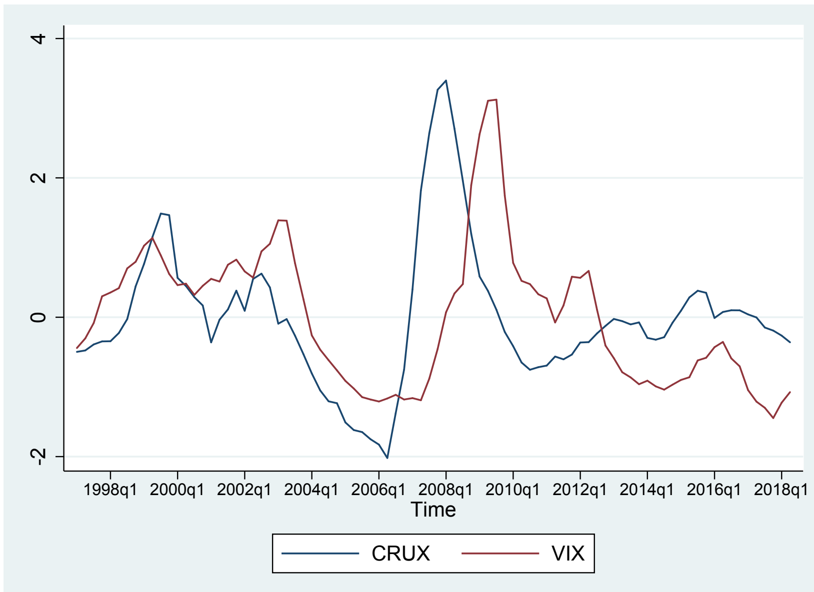
We construct a new time-varying measure of firm-specific uncertainty from analyzing the text of company reports filed with the U.S. Securities and Exchange Commission. We explore the implications of idiosyncratic variation in firm-level uncertainty in the aggregate and with firm-level microdata. We find the new measure is negatively correlated with growth in firm-level investment, aggregate investment, and GDP even after controlling for other measures of first moment shocks and aggregate uncertainty. Using firm-level panel data on investment with a rich set of controls, we find our measure of firm-specific uncertainty has large negative effects on investment rates. Moreover, the response of investment to positive demand shocks is attenuated by 50%.

An implications of our findings is that gross job reallocation, or job churning is reduced by firm-specific uncertainty. To the extent that reallocation facilitates the process of creative destruction and promotes productive growth (cf. Decker et al., 2017), our results suggest high uncertainty during and after recent recessions may have contributed to reductions in economic dynamic and productivity growth. Future research should further investigate these channels.

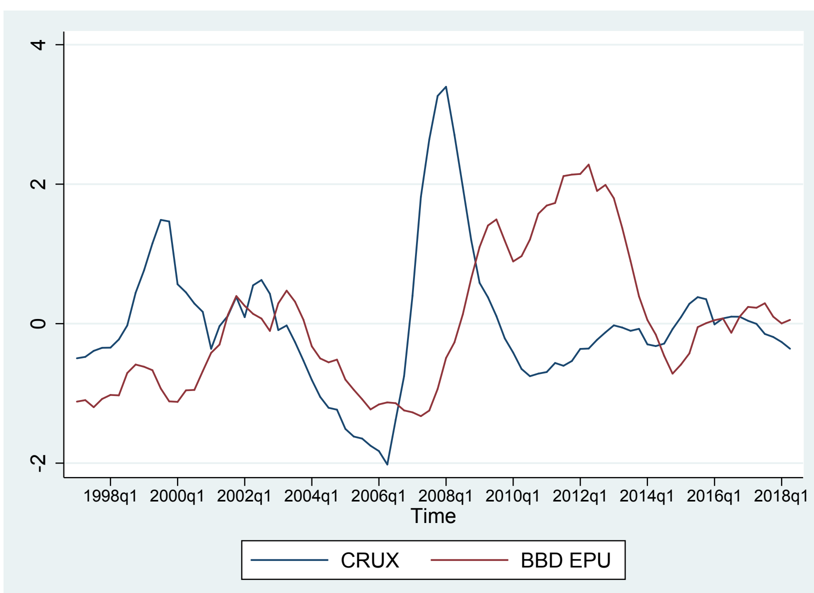
Figures

Figure 2.1: Evolution of Company Reported Uncertainty Index (CRUX) and Alternative Uncertainty Measures (mean/variance standardized)

(a) CRUX vs VIX

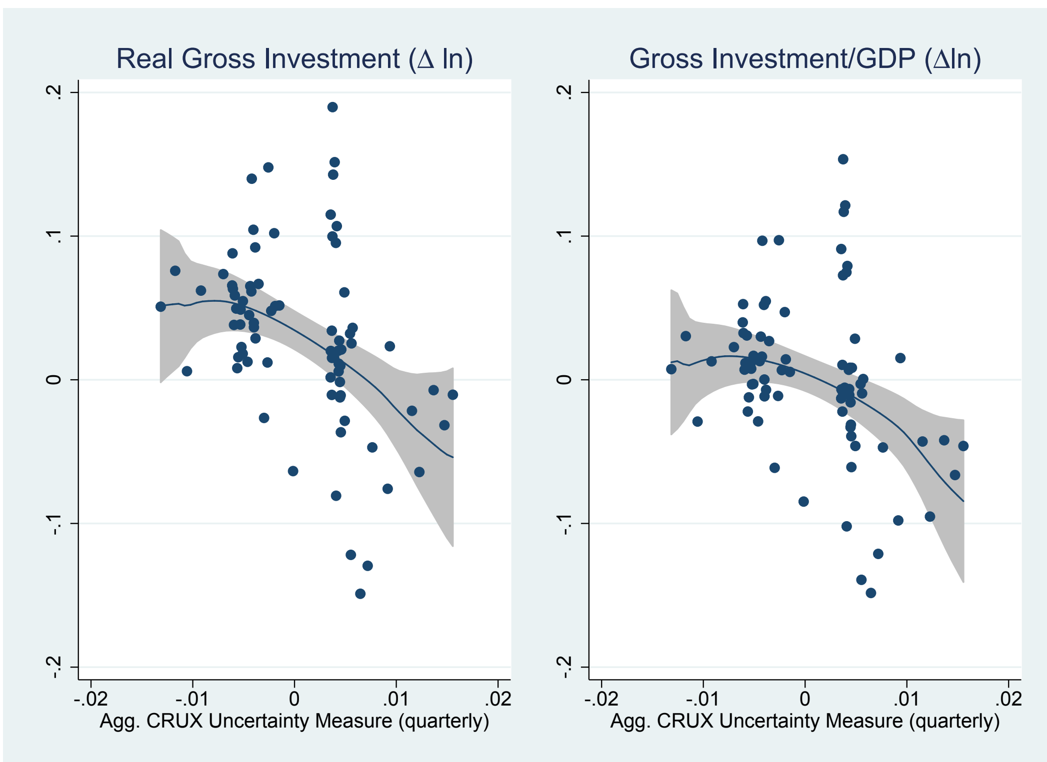


(b) CRUX vs BBD Economic Policy Uncertainty (text based index)



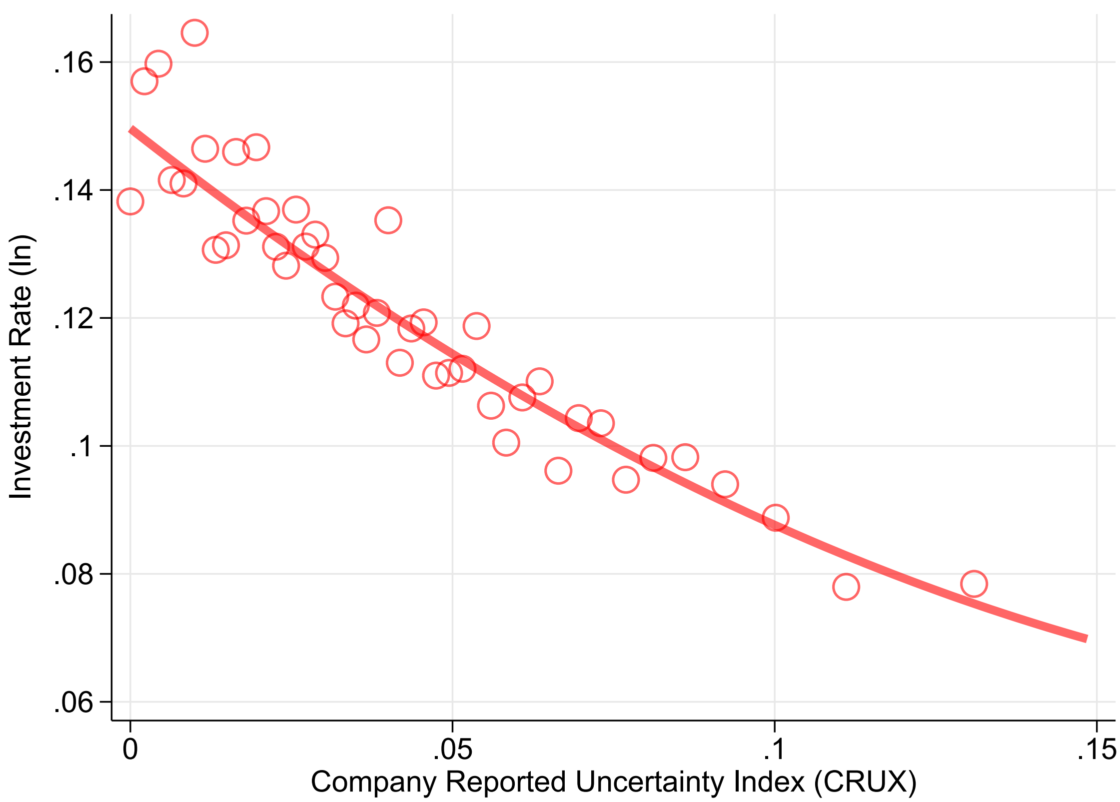
Notes: For comparison only, all measures are standardize. Quarterly aggregated CRUX is constructed as 4 quarter rolling average measure and standardized with zero sample mean and unit sample variance. VIX and BBD EPU are constructed as 4 quarter moving averages and standardized with zero sample mean and unit sample variance.

Figure 2.2: Aggregate Firm Level Uncertainty and Gross Investment–Semiparametric Evidence



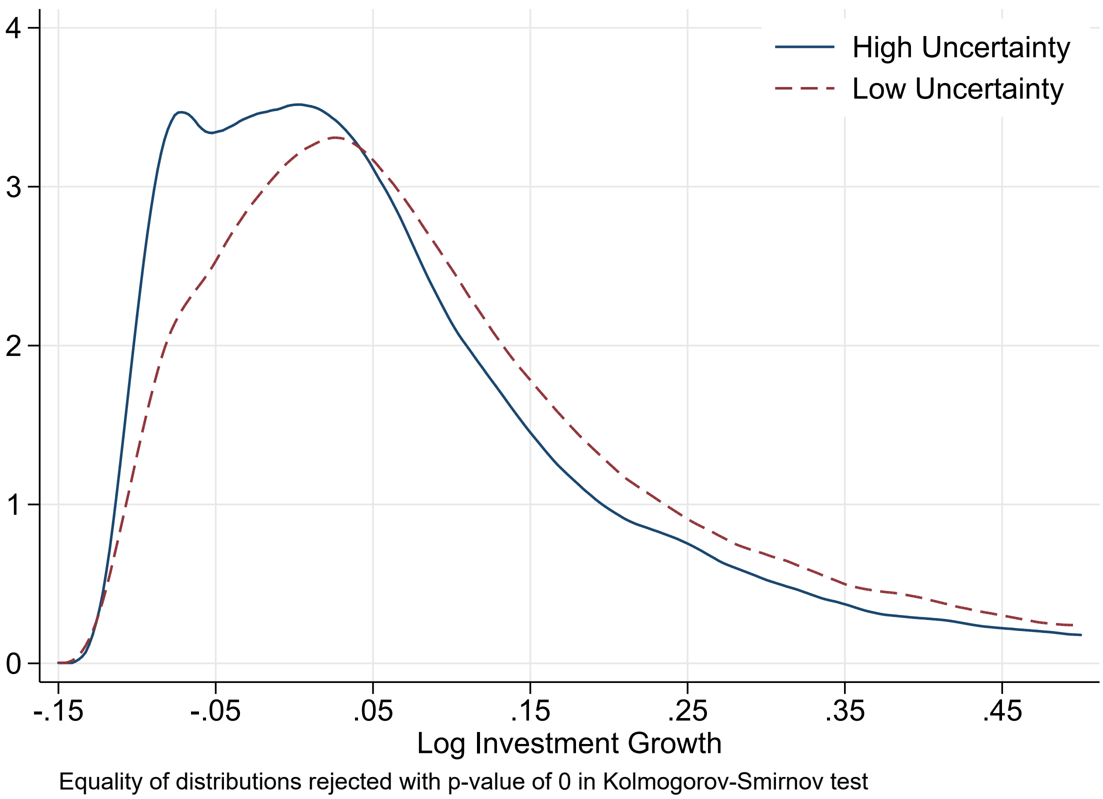
Note: Semiparametric fit of gross investment ($\Delta\,\ln$) and gross investment/GDP ($\Delta\,\ln$) to aggregate CRUX uncertainty measure after partialling out the change in the S&P 500 Index, a first moment control, and the VIX index of equity market volatility, a second moment control.

Figure 2.3: Binscatter: Investment negatively related to uncertainty



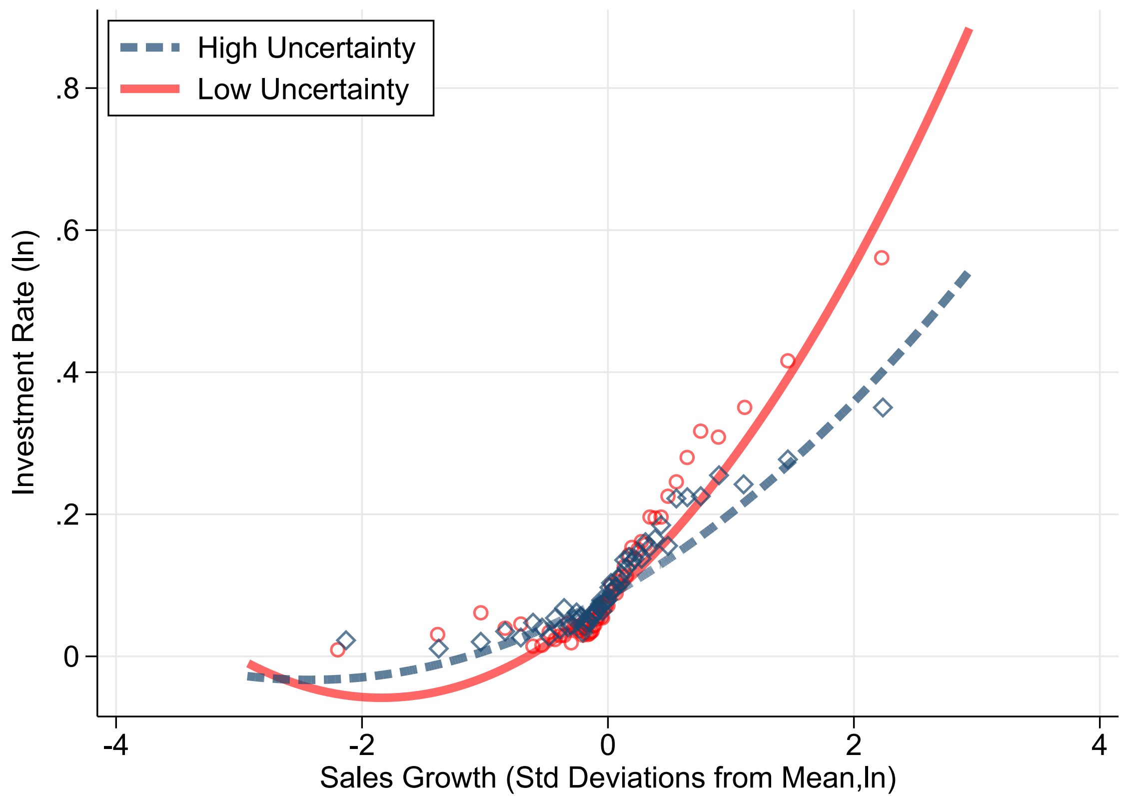
Notes: Binsreg plot of CRUX against investment rate fitted by second order polynomial without controls.

Figure 2.4: Firm Investment Growth under High vs Low Uncertainty



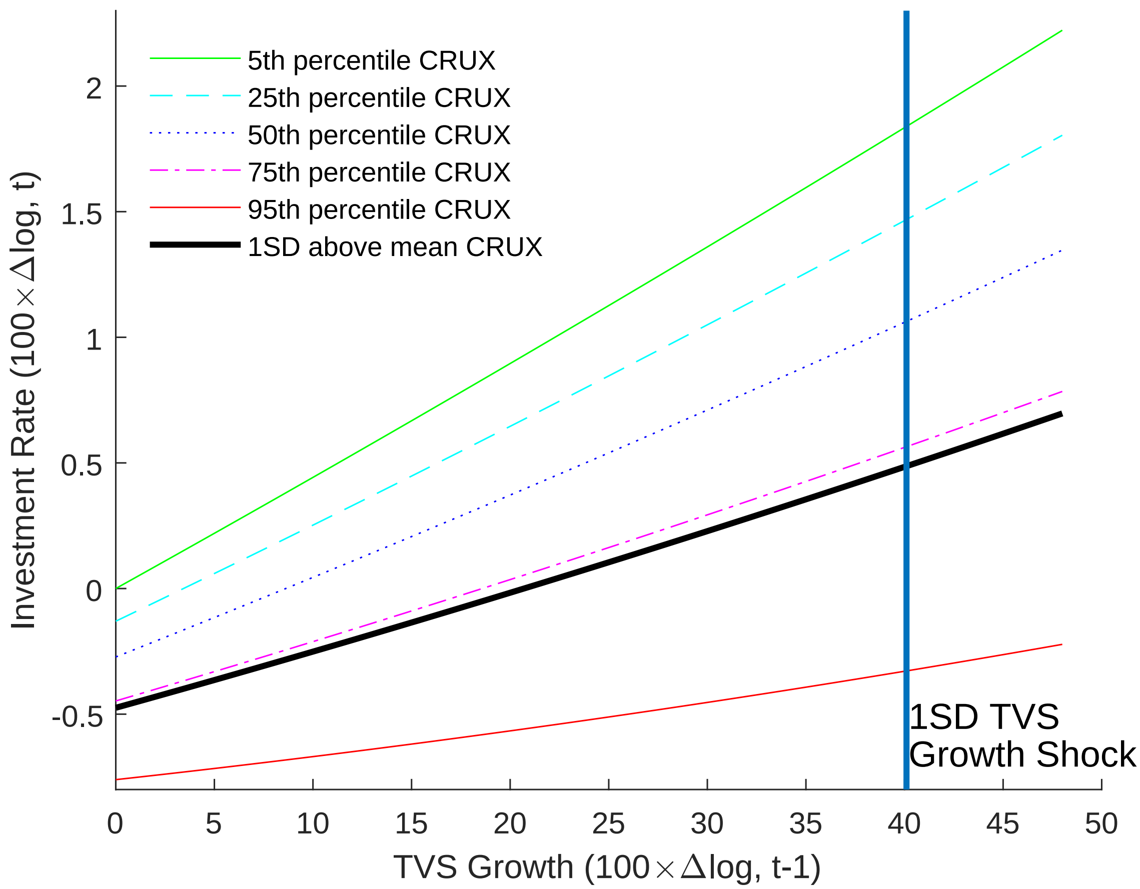
Notes: Epanechnikov kernel density estimates. High uncertainty defined as above median in uncertainty measure in Edgar-Compustat matched sample. Low uncertainty is below the median. Firm investment growth rate is calculated in log differences of total capital stock.

Figure 2.5: Investment Response to Sales Growth Attenuated for High Uncertainty



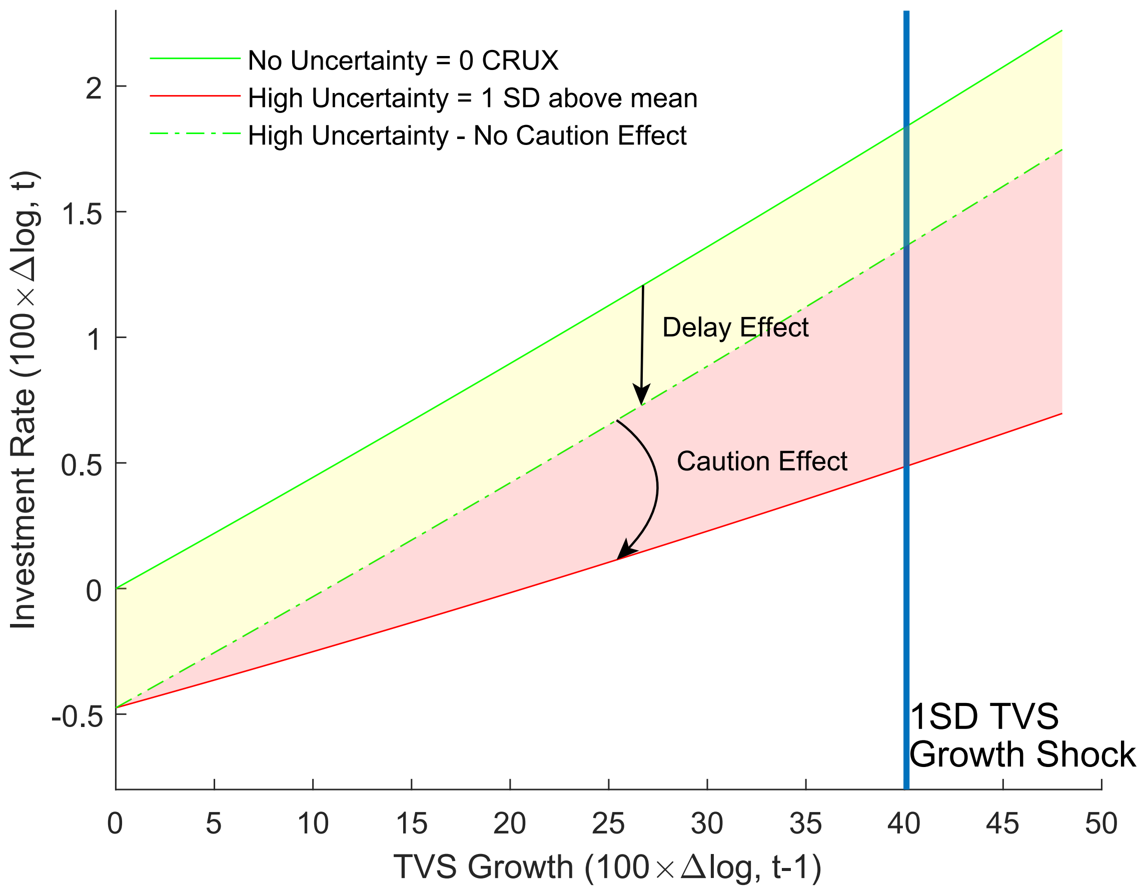
Notes: Binsreg plot of investment response to sales growth under high vs low uncertainty, controlling for year fixed effects.

Figure 2.6: Investment Response to Sales Growth under Uncertainty (Census Establishment Level)



Notes: Plotting investment response to sales growth under different level of uncertainty, taking coefficients from column 6 in Table 6.

Figure 2.7: Delay vs Caution Effect of Investment Response to Sales Growth under Uncertainty



Notes: Comparing the differences between 5th percentile CRUX and one standard deviation above mean CRUX.

Tables

Table 2.1: Summary Statistics - Compustat Firm Level Data

VARIABLES	Panel A Investment Log Growth	Panel B Employment DHS Growth
CRUX	0.0429 [0.0334]	0.0419 [0.0325]
Investment growth	0.123 [0.261]	-
Employment growth	-	0.0307 [0.330]
Lag sales growth (log)	0.119 [0.607]	0.115 [0.564]
Lag Tobin's Q (log)	0.595 [0.766]	0.529 [0.728]
Lag sales growth (log) squared	0.382 [2.272]	0.332 [2.059]
N	95,223	107,031

Panel A reports summary statistics of regression and non-parametric analysis on Compustat investment data. Investment growth is calculated as log growth. Panel B reports summary statistics of regression and non-parametric analysis on Compustat employment data. Employment growth is calculated in DHS form. All missing observations are dropped in both panels.

Table 2.2: Decomposition of Variance of CRUX

	Compustat Corporate Investment Sample	
	(1)	(2)
	R-squared	Incremental R-squared
Time FE	20.75%	20.75%
Industry FE (3-digit NAICS)	4.77%	4.26%
Industry (3-digit NAICS) × Time FE	26.81%	1.80%
Firm FE	48.86%	36.19%
Unexplained Residual	-	37.00%
Number of Industries	95	95
Number of Firms	11,864	11,864
Number of Observations	95,223	95,223

Notes: This table reports variance decomposition of CRUX in Compustat corporate investment sample. Column (1) reports the R-squared values when regressing CRUX on each of the fixed effects alone. Column (2) reports the incremental R-squared values when regressing CRUX on additional fixed effects from previous row. Thus the firm-level variation within industry × time fixed effects accounts for 37% + 36.19 = 73.19%, and the residual variation unexplained by firm fixed effects and industry × time fixed effects is 37%.

Table 2.3: Effects of Uncertainty on Corporate Investment Rate (Δln, Compustat)

	Dependent Variable: Log Change in Capital Stock K (Δln, t)					
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
CRUX (t)	-0.350***	-0.309***	-0.303***	-0.304***	-0.300***	-0.301***
	[0.0430]	[0.0430]	[0.0425]	[0.0425]	[0.0413]	[0.0413]
Sales Growth (Δln t-1)			0.0775***	0.0757***	0.0712***	0.0698***
			[0.00546]	[0.00533]	[0.00521]	[0.00511]
CRUX (t) × Sales Growth (Δln t-1)			-0.494***	-0.480***	-0.456***	-0.445***
			[0.0705]	[0.0708]	[0.0686]	[0.0688]
Sales Growth (Δln t-1) sqaured				0.00299***		0.00229***
				[0.000826]		[0.000765]
Log Tobin's Q (t-1)					0.0893***	0.0890***
					[0.00338]	[0.00338]
Firm Fixed Effects	√	√	√	√	√	√
Year Fixed Effects	√					
Industry × Year Fixed Effects		√	√	√	√	√
Observations	95,222	93,741	93,741	93,741	93,741	93,741
R-squared	0.081	0.391	0.405	0.405	0.427	0.427
Number of Firms	11,864	10,445	10,445	10,445	10,445	10,445

Notes: Standard errors are clustered at firm level. Dependent variable is calculated as log(K(t)) - log(K(t-1)) where K is firm's total capital stock. Firm idiosyncratic uncertainty measure CRUX is at time t, which as mentioned in the text captures firm uncertainty at time t-1. Sales growth is calculated as log(Sales(t-1)) - log(Sales(t-2)) and demeaned by sample mean. Squared sales growth is calculated using demeand sales growth. Tobin's Q is taken as log average Q at time t-1. All of the regression specifications include firm FE. Column (1) includes Year FE, Column (2)-(6) include industry × year FE, where industry is at 3-digit NAICS code level. Column (2)-(6) loses some observations and firms compared with Column (1) due to singleton observations. Time ranges from 1994 to 2016.

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 2.4: Summary Statistics - ASM/CM Establishment Level Data (Matched Sample)

VARIABLES	Mean	Std Dev
CRUX	0.0409	[0.0288]
Peer CRUX	0.0804	[1.25]
Log Total Investment Rate	0.818	[14.7]
Investment Spike	4.7	[21.2]
Log Structure Investment Rate	0.178	[11.7]
Log Equipment Investment Rate	1.13	[17.5]
Lag TVS growth (log)	0.94	[40.1]
Lag TVS Growth (log) squared	0.161	[1.011]
Lag Tobin's Q (log)	0.437	[0.378]
N	133000	

Note: All variables except for lag Tobin's Q, and lag TVS growth sqaured, are in percentage points (value form sample × 100). This is establishment level data.

Table 2.5: Effect of Uncertainty on Manufacturing Establishment Level Total Investment Rate (Δln)

	Dependent Variable: Log Total Investment Rate (Δln, t)								
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
CRUX (t)	-0.0670**	-0.0662*	-0.0685*	-0.0690*	-0.0676*	-0.0681*	-0.0667*		
	[0.0321]	[0.0355]	[0.0353]	[0.0353]	[0.0354]	[0.0354]	[0.0354]		
TVS Growth (Δln t-1)			0.0327***	0.0460***	0.0458***	0.0437***	0.0435***	0.0470***	0.0448***
			[0.00375]	[0.00694]	[0.00697]	[0.00647]	[0.00650]	[0.00690]	[0.00642]
CRUX (t) × TVS Growth (Δln t-1)				-0.335***	-0.337***	-0.314***	-0.316***	-0.373***	-0.354***
				[0.114]	[0.114]	[0.110]	[0.110]	[0.121]	[0.118]
TVS Growth (Δln t-1) squared						0.00540***	0.00541***		0.00492***
						[0.00186]	[0.00186]		[0.00190]
Log Tobin's Q (t-1)					0.0153***		0.0154***		
					[0.00322]		[0.00320]		
Firm Fixed Effects	√	√	√	√	√	√	√		
Firm × Year Fixed Effects								√	√
Year Fixed Effects	√								
Industry × Year Fixed Effects		√	√	√	√	√	√	√	√
R-squared	0.063	0.073	0.08	0.081	0.081	0.082	0.083	0.226	0.227

Notes: Standard errors are clustered at firm level. Dependent variable is calculated as log(K(t)) - log(K(t-1)) where K is establishment's total capital stock. Firm level idiosyncratic uncertainty measure CRUX is at time t, which as mentioned in the text captures firm uncertainty at time t-1. Establishment level TVS (total value of shipment) growth is calculated as log(TVS(t-1)) - log(TVS(t-2)) and demeaned by sample mean. Squared TVS growth is calculated using demeand TVS growth. Firm level Tobin's Q is taken as log average Q at time t-1. Columns (1)-(7) include firm FE, Columns (8)-(9) include firm × year FE, which absorbs CRUX (t) and Log Tobin's Q (t-1). Column (1) includes Year FE, Columns (2)-(9) include industry × year FE, where industry is at 3-digit NAICS code level. Time ranges from 1998 to 2014. Number of observations is 133000 and number of firms is 2000, both rounded to the nearest thousands.

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 2.6: Effect of Uncertainty on Manufacturing Establishment Level Investment Spike ($\frac{I_{it}}{K_{it-1}} \geq 20\%$)

	Dependent Variable: Indicator Arithmetic Investment Rate ≥ 20% (t)		
VARIABLES	(1)	(2)	(3)
CRUX (t)	-0.116**	-0.114**	
	[0.0527]	[0.0526]	
TVS Growth (Δln t-1)	0.0405***	0.0373***	0.0371***
	[0.00489]	[0.00468]	[0.00462]
CRUX (t) × TVS Growth (Δln t-1)	-0.298***	-0.273***	-0.290***
	[0.0875]	[0.0861]	[0.0882]
TVS Growth (Δln t-1) squared		0.00691***	0.00609***
		[0.00145]	[0.00136]
Log Tobin's Q (t-1)		0.0157***	
		[0.00510]	
Firm Fixed Effects	√	√	
Firm × Year Fixed Effects			√
Industry × Year Fixed Effects	√	√	√
R-squared	0.071	0.072	0.214

Notes: Standard errors are clustered at firm level. Dependent variable is an indicator = 1 if (K(t)- K(t-1))/K(t-1) ≥ 20% and = 0 if otherwise, where K is establishment's total capital stock. Firm level idiosyncratic uncertainty measure CRUX is at time t, which as mentioned in the text captures firm uncertainty at time t-1. Establishment level TVS (total value of shipment) growth is calculated as log(TVS(t-1)) - log(TVS(t-2)) and demeaned by sample mean. Squared TVS growth is calculated using demeand TVS growth. Firm level Tobin's Q is taken as log average Q at time t-1. Columns (1) and (2) include firm FE. Column (3) includes firm × year FE, which absorbs CRUX (t) and Log Tobin's Q (t-1). All columns include industry × year FE, where industry is at 3-digit NAICS code level. Time ranges from 1998 to 2014. Number of observations is 133000 and number of firms is 2000, both rounded to the nearest thousands.

*** p<0.01, ** p<0.05, * p<0.1

Table 2.7: Effect of Uncertainty on Equipment vs. Structure Investment Rate (Δln)

	Log Equipment Investment Rate (Δln, t)			Log Structure Investment Rate (Δln, t)		
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
CRUX (t)	-0.0787*	-0.0760*		-0.0284	-0.027	
	[0.0437]	[0.0441]		[0.0252]	[0.0251]	
TVS Growth (Δln t-1)	0.0554***	0.0527***	0.0545***	0.0254***	0.0242***	0.0246***
	[0.00808]	[0.00759]	[0.00737]	[0.00500]	[0.00471]	[0.00498]
CRUX (t) × TVS Growth (Δln t-1)	-0.378***	-0.358***	-0.408***	-0.205**	-0.197**	-0.196**
	[0.136]	[0.132]	[0.138]	[0.0823]	[0.0791]	[0.0903]
TVS Growth (Δln t-1) squared		0.00566***	0.00502**		0.0025	0.00221
		[0.00204]	[0.00205]		[0.00157]	[0.00165]
Log Tobin's Q (t-1)		0.0199***			0.0113***	
		[0.00416]			[0.00219]	
Firm Fixed Effects	√	√		√	√	
Firm × Year Fixed Effects			√			√
Industry × Year Fixed Effects	√	√	√	√	√	√
R-squared	0.094	0.095	0.24	0.048	0.049	0.185

Notes: Standard errors are clustered at firm level. Dependent variable is calculated as log(K(t)) - log(K(t-1)) where K is establishment's total capital stock in equipment or structure. Firm level idiosyncratic uncertainty measure CRUX is at time t, which as mentioned in the text captures firm uncertainty at time t-1. Establishment level TVS (total value of shipment) growth is calculated as log(TVS(t-1)) - log(TVS(t-2)) and demeaned by sample mean. Squared TVS growth is calculated using demeand TVS growth. Firm level Tobin's Q is taken as log average Q at time t-1. Columns (1)-(2) and (4)-(5) include firm FE. Columns (3) and (6) include firm × year FE, which absorbs CRUX (t) and Log Tobin's Q (t-1). All columns include industry × year FE, where industry is at 3-digit NAICS code level. Time ranges from 1998 to 2014. Number of observations is 133000 and number of firms is 2000, both rounded to the nearest thousands.
*** p<0.01, ** p<0.05, * p<0.1

Table 2.8: Effect of Peer vs Idiosyncratic Uncertainty on Manufacturing Establishment Level Total Investment Rate (Δln)

	Dependent Variable: Log Total Investment Rate (Δln, t)					
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
Peer CRUX (t)	-0.237*	-0.268*	-0.219	-0.249*		
	[0.141]	[0.137]	[0.140]	[0.136]		
Peer CRUX (t) × TVS Growth (Δln t-1)	-0.734***	-0.512*	-0.702***	-0.496	-0.814***	-0.594*
	[0.258]	[0.308]	[0.256]	[0.305]	[0.277]	[0.312]
CRUX (t)		-0.0768**		-0.0739**		
		[0.0344]		[0.0346]		
CRUX (t) × TVS Growth (Δln t-1)		-0.248*		-0.232*		-0.248*
		[0.136]		[0.132]		[0.133]
TVS Growth (Δln t-1)	0.0333***	0.0430***	0.0315***	0.0406***	0.0315***	0.0411***
	[0.00368]	[0.00766]	[0.00350]	[0.00724]	[0.00361]	[0.00688]
TVS Growth (Δln t-1) squared			0.00547***	0.00538***	0.00493**	0.00486**
			[0.00189]	[0.00187]	[0.00191]	[0.00190]
Log Tobin's Q (t-1)			0.0152***	0.0152***		
			[0.00317]	[0.00316]		
Firm Fixed Effects	√	√	√	√		
Firm × Year Fixed Effects					√	√
Industry × Year Fixed Effects	√	√	√	√	√	√
R-squared	0.081	0.081	0.083	0.083	0.227	0.227

Notes: Standard errors are clustered at firm level. Dependent variable is calculated as log(K(t)) - log(K(t-1)) where K is establishment's total capital stock. Firm level idiosyncratic uncertainty measure CRUX is at time t, which as mentioned in the text captures firm uncertainty at time t-1. Uncertainty measure of establishment's peers is calculated by taking the average of firm FE demeaned CRUX measure of all establishments from other firms within the same industry (4-digit NAICS code). Establishment level TVS (total value of shipment) growth is calculated as log(TVS(t-1)) - log(TVS(t-2)) and demeaned by sample mean. Squared TVS growth is calculated using demeand TVS growth. Firm level Tobin's Q is taken as log average Q at time t-1. Columns (1)-(4) include firm FE. Columns (5)-(6) include firm × year FE, which absorbs CRUX (t), Peer CRUX (t) and Log Tobin's Q (t-1). All columns include industry × year FE, where industry is at 3-digit NAICS code level. Time ranges from 1998 to 2014. Number of observations is 133000 and number of firms is 2000, both rounded to the nearest thousands.

*** p<0.01, ** p<0.05, * p<0.1

Table 2.9: Summary Statistics - ASM/CM Establishment Level Data (Full Sample)

VARIABLES	Mean	Std Dev
Industry CRUX (Equally-Weighted)	0.219	[1.26]
Industry CRUX (IPS-Weighted)	0.219	[1.26]
Log Total Investment Rate	0.654	[15.2]
Investment Spike	4.75	[21.3]
Log Structure Investment Rate	-0.073	[11.1]
Log Equipment Investment Rate	0.98	[18.4]
Lag TVS growth (log)	0.697	[41.7]
Lag TVS Growth (log) squared	0.174	[1.037]
N	472000	

Note: All variables except for lag TVS growth sqaured, are in percentage points (value form sample × 100). This is establishment level data.

Table 2.10: Effect of Industry Uncertainty on Manufacturing Establishment Level Total Investment Rate (Δln)

	Dependent Variable: Log Total Investment Rate (Δln, t)					
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Equally-Weighted Industry CRUX			IPS-Weighted Industry CRUX		
Industry CRUX (t)	-0.184***	-0.186***		-0.182***	-0.184***	
	[0.0565]	[0.0564]		[0.0565]	[0.0565]	
TVS Growth (Δln t-1)	0.0230***	0.0223***	0.0245***	0.0230***	0.0223***	0.0245***
	[0.00172]	[0.00168]	[0.00239]	[0.00172]	[0.00168]	[0.00239]
Industry CRUX (t) × TVS Growth (Δln t-1)	-0.370***	-0.324***	-0.389***	-0.369***	-0.323***	-0.388***
	[0.102]	[0.0994]	[0.150]	[0.102]	[0.0994]	[0.150]
TVS Growth (Δln t-1) squared		0.00392***	0.00432***		0.00392***	0.00432***
		[0.000875]	[0.00116]		[0.000875]	[0.00116]
Firm Fixed Effects	√	√		√	√	
Firm × Year Fixed Effects			√			√
Industry × Year Fixed Effects	√	√	√	√	√	√
R-squared	0.124	0.125	0.423	0.124	0.125	0.423

Notes: Standard errors are clustered at industry (4-digit NAICS) × year level. Dependent variable is calculated as log(K(t)) - log(K(t-1)) where K is establishment's total capital stock. Industry level uncertainty measure Inudustry CRUX (t) is calculated by taking equally weighted or inverse propensity score weighted average of firm FE demeaned CRUX measure of all establishments within the same industry (4-digit NAICS code). Inudstry CRUX in columns (1)-(3) is equally weighted. Industry CRUX in columns (4)-(6) is weighed by inverse propensity score constructed by fitting logit specifications. Establishment level TVS (total value of shipment) growth is calculated as log(TVS(t-1)) - log(TVS(t-2)) and demeaned by sample mean. Squared TVS growth is calculated using demeand TVS growth. Columns (1)-(2) and (4)-(5) include firm FE, Columns (3) and (6) include firm × year FE, which absorbs Inudstry CRUX (t). All columns include industry × year FE, where industry is at 3-digit NAICS code level. Time ranges from 1998 to 2014. Number of observations is 472000 and number of firms is 21000, both rounded to the nearest thousands.
*** p<0.01, ** p<0.05, * p<0.1

Table 2.11: Effect of Industry Uncertainty on Equipment vs. Structure Investment Rate (Δln)

	Log Equipment Investment Rate (Δln, t)			Log Structure Investment Rate (Δln, t)		
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
Industry CRUX (t)	-0.242***	-0.245***		-0.0342	-0.0351	
	[0.0751]	[0.0750]		[0.0345]	[0.0345]	
TVS Growth (Δln t-1)	0.0300***	0.0292***	0.0316***	0.00974***	0.00950***	0.0117***
	[0.00197]	[0.00192]	[0.00266]	[0.00121]	[0.00121]	[0.00174]
Industry CRUX (t) × TVS Growth (Δln t-1)	-0.508***	-0.458***	-0.547***	-0.158**	-0.142**	-0.205*
	[0.115]	[0.112]	[0.166]	[0.0727]	[0.0697]	[0.108]
TVS Growth (Δln t-1) squared		0.00424***	0.00462***		0.00132**	0.00153*
		[0.000937]	[0.00124]		[0.000667]	[0.000854]
Firm Fixed Effects	√	√		√	√	
Firm × Year Fixed Effects			√			√
Industry × Year Fixed Effects	√	√	√	√	√	√
R-squared	0.132	0.133	0.436	0.092	0.092	0.375

Notes: Standard errors are clustered at industry (4-digit NAICS) × year level. Dependent variable is calculated as log(K(t)) - log(K(t-1)) where K is establishment's total capital stock in equipment or structure. Industry level uncertainty measure Inudustry CRUX (t) is calculated by taking equally weighted average of firm FE demeaned CRUX measure of all establishments within the same industry (4-digit NAICS code). Establishment level TVS (total value of shipment) growth is calculated as log(TVS(t-1)) - log(TVS(t-2)) and demeaned by sample mean. Squared TVS growth is calculated using demeand TVS growth. Columns (1)-(2) and (4)-(5) include firm FE. Columns (3) and (6) include firm × year FE, which absorbs Industry CRUX (t). All columns include industry × year FE, where industry is at 3-digit NAICS code level. Time ranges from 1998 to 2014. Number of observations is 472000 and number of firms is 21000, both rounded to the nearest thousands.
*** p<0.01, ** p<0.05, * p<0.1

Table 2.12: Regressions of Gross Investment and GDP on Aggregate Firm-Level Uncertainty

	OLS			ML	
Panel A: Dependent Variable: Annual Change in Gross Investment to GDP Ratio (ln)					
	(1)	(2)	(3)	(4)	(5)
Controls:	Baseline	S&P 500 (Δln)	VIX	Lagged Dep. Var.	ALL
CRUX(t)	**-4.328***	**-2.815***	**-4.331***	**-4.267***	**-3.151***
	[2.429]	**[1.592]**	**[1.233]**	**[1.134]**	**[0.798]**
Change in S&P 500(ln)		0.283***			0.247***
		[0.0494]			[0.0465]
VIX (4 period MA)			-0.00650***		-0.00277**
			[0.00167]		[0.00136]
Lagged Gross Invest/GDP(Δln)				-0.239	-0.239
				[0.146]	[0.146]
R-Squared	0.114	0.488	0.399	-	-
Newey-West Bandwidth	14	19	19	-	-
Panel B: Dependent Variable: Annual Change in Real Gross Investment (ln)					
	(1)	(2)	(3)	(4)	(5)
Controls:	Baseline	S&P 500 (Δln)	VIX	Lagged Dep. Var.	ALL
CRUX(t)	**-5.811****	**-3.939***	**-5.815***	**-5.812***	**-4.372***
	[2.304]	**[1.145]**	**[1.375]**	**[1.393]**	**[0.797]**
Change in S&P 500(ln)		0.350***			0.336***
		[0.0587]			[0.0435]
VIX (4 period MA)			-0.00684***		-0.00164
			[0.00228]		[0.00136]
Lagged Gross Invest (Δln)				-0.000145	-0.316***
				[0.142]	[0.122]
R-Squared	0.000851	0.000851	0.000851	-	-
Newey-West Bandwidth	19	16	19	-	-
Panel C: Dependent Variable: Annual Change in Real GDP (ln)					
	(1)	(2)	(3)	(4)	(5)
Controls:	Baseline	S&P 500 (Δln)	VIX	Lagged Dep. Var.	ALL
CRUX(t)	**-1.687***	**-1.331***	**-1.688***	**-1.744***	**-1.524***
	[0.331]	**[0.243]**	**[0.374]**	**[0.338]**	**[0.175]**
Change in S&P 500(ln)		0.0665***			0.0559***
		[0.00794]			[0.00736]
VIX (4 period MA)			-0.00130**		-0.000430*
			[0.000570]		[0.000223]
Lagged GDP(Δln)				-0.0854	-0.248**
				[0.117]	[0.106]
R-Squared	0.000851	0.000851	0.000851	-	-
Newey-West Bandwidth	19	19	10	-	-

Notes: All regressions have 76 observations (quarters) of data. Robust standard errors in brackets. OLS regressions employ Newey-West heteroskedasticity and autocorrelation robust standard errors with automatic bandwidth selection columns 1-3.

CHAPTER III

Uncertainty and Firm Labor Reallocation

3.1 Introduction

Recent studies correlate uncertainty with slow economic growth, high unemployment rate, low labor reallocation, and slow economic recovery, especially after the financial crisis. In this chapter, we examine the role of uncertainty on labor market dynamics in both micro and macro level. Easy as it seems in theory, measuring risk and uncertainty (at the firm level) is difficult in practice. Our analysis in this chapter relies on the Company Reported Uncertainty Index (CRUX) measure developed in chapter 2. The CRUX index takes the usage of "uncertain" related words relative to the total number of meaningful words in corporate report filed with U.S. Securities and Exchange Commission (SEC) to create a time-varying aggregate, industry and firm-level measure of uncertainty shocks. Then we ask (1) how important are fluctuations in firm-level uncertainty for employment in aggregate and disaggregated data; (2) does the firm-level variation in uncertainty explain micro and macro labor reallocation independently of other aggregate uncertainty measures?

To better understand these dynamics, we explore the implications of both firm-specific and aggregated industry level fluctuations in uncertainty for employment growth and establishment dynamics using data from both COMPUSTAT and the U.S. Census firm and establishment microdata. An ideal firm-level measure would (1) summarize idiosyncratic implications of all risk factors to firm's outlook on future; (2) disentangle expectations from uncertainty (subjective or not), forecast from forecast error; (3) and provide identifiable variation across firms and time. The Company Reported Uncertainty Index (CRUX) employs the words firms use in SEC filings to measure "uncertainty"-related words in business context. The SEC EDGAR database provides a standard and consistent panel of textual descriptions, which allows the CRUX to capture uncertainty unmeasured by observable volatility and make cross-sectional and time-series comparisons.

We then verify the uncertainty mechanism by linking the firm-level version of the CRUX to microdata employment from two sources: (1) publicly available data in COMPUSTAT available at the firm level and (2) the confidential U.S. Census microdata on firm *and* establishment dynamics we construct from the the Longitudinal Business Database (LBD) and Business Registrar (BR). Using this rich microdata, we can quantify first order delay and second order caution effects on hiring decisions, after controlling for firm and industry-by-year fixed effects and a set of first and higher moment firm-level characteristics. Real options models with irreversible costs yield two predictions: (1) high uncertainty can reduce investment/hiring (delay effects); (2) increased uncertainty attenuates response to demand shocks (caution effects). We leverage establishment level variation within firm to identify labor reallocation pattern and the delay and caution effects of uncertainty through firm's reallocation of activity across multiple margins. Aggregating up CRUX from firm to industry and macro level, we provide border implications on the importance of idiosyncratic vs aggregate uncertainty in macro and firm level dynamics.

The rest of the this chapter is organized as follows. In Section 3.2, we describe our underlying data source of uncertainty measurement and firm and establishment level data drawn from COMPUSTAT and U.S. Census. Section 3.3 provides firm and establishment level evidence the employment response to our measure of uncertainty. Section 3.4 provides evidence of using aggregate data and Section 3.5 concludes.

3.2 Data and Measurement

3.2.1 Measuring Uncertainty

To measure uncertainty, we take the CRUX measure constructed in Chapter 2. The measure takes advantage of the fact that the U.S. Securities and Exchange Commission (SEC) mandates all publicly trade firms to make reports on certain circumstances. Firms must submit an annual report (Form 10-K) and three quarterly reports (Form 10-Q) and their amendments according to their fiscal year schedule. As described in Chapter 2, we parse and analyze reports from SEC EDGAR database to construct the uncertainty measure (CRUX) by normalizing the count of uncertainty words with total number of meaningful words[1] within documents from the entire filing year.

3.2.2 Firm and Establishment Level Data

Our firm or establishment level outcomes are draw from publicly available data in the COMPUSTAT database and confidential microdata from the U.S. Census.

3.2.2.1 COMPUSTAT

Firm information is drawn from COMPUSTAT data on firm balance sheets, cash flow and income statements. We match the CRUX measure with COMPUSTAT through firm identifier CIK and year from fiscal years ending from 1994 to 2016. After removing missing dependent and independent variables, we have over 107,000

[1] We compute CRUX as $\text{CRUX}_{it} = \frac{\text{Total uncertain words}_{it}}{\text{Total number of meaningful words}_{it}} \times 100$. Details in Chapter 2.

observations over more than 13,000 firms in the employment panel.

We measure the employment growth rate by taking the DHS midpoint growth rates $\frac{(emp_{it}-emp_{i,t-1})}{0.5\cdot(emp_{it}+emp_{i,t-1})}$. The DHS growth rate is bounded on $[-2, 2]$, which helps with the somewhat noisy employment data in COMPUSTAT.[2] As in Chapter 2, we obtain sales directly from COMPUSTAT and compute lagged sales log growth and squared sales log growth to capture firms' demand shocks and its convex effects on firm hiring decisions. We use lagged Tobin's Q[3] to proxy firms' investment opportunities. Summary statistics are provided in Table 3.1.

3.2.2.2 US Census Microdata

To study the effects of uncertainty on firms' hiring behavior in further detail, we rely on Longitudinal Business Database (LBD) in data collected by the U.S. Census Bureau.

We rely on the LBD to track employment, firm-identifiers, and establishment dynamics within firms. The LBD is derived from the Census Business Register (BR) and covers the entire non-farm private sector and is compiled from administrative records and survey sources. Firms are defined based on operational control, and all establishments majority owned by a parent firm are included in the parent's activity measures.

Firm identifiers in the LBD are linked to COMPUSTAT and the EDGAR based CRUX measure using the Census COMPUSTAT-BR bridge. This allows us to link financial data and our CRUX measure to the firm and establishment level data in the LBD. Specifically, the bridge provides annual link between a Compustat CUSIP and firm identifier in LBD. We then link CRUX measure to LBD data through CIK-CUSIP-

[2] It is common in the literature that COMPUSTAT employment variables are trimmed or winsorized due to the quality of the data. The DHS growth rate is one method to accomplish that objective. Our results are robust to omitting outliers or winsorizing as well.

[3] As in Chapter 2, we compute Tobin's Q as $\frac{\text{Market Capitalization+Market Value of Liability}}{\text{Total Asset Value}} = \frac{csho \times prcc\_f + at - ceq}{at}$.

LBD identifiers for 1994-2013.[4]

The main variables in the LBD that we use are employment and six-digit industry codes in the North American Industrial Classification (NAICS). We assign consistent NAICS industry codes to establishments using the concordance in Fort and Klimek (2018). For multi-unit firms, we compute employment shares across 3 digit NAICS codes with the firm and assign firm NAICS based on the largest 3 digit industry. The LBD reports total employment in the payroll period containing the week of March 12th. We are interested in how reported uncertainty in CRUX effects employment dynamics, hiring, and reorganization so matching the CRUX measure to this timing is important. Because annual growth in employment between year $t-1$ and t is March to March, we match this timing using a CRUX measure from the second quarter of year $t-1$ to the first quarter of year t.

The firm identifiers in the LBD enable us to compute growth rate measures for establishments and firms and to track their entry, exit and ownership changes. Because we have establishment level detail linked to firm-level uncertainty measure, we can explore within-firm restructuring activity in response to uncertainty relative to other controls. We summarize LBD data in Table 3.3. Total employment growth rate is the sum of job creation rate and job destruction rate. We decompose firm net job creation rates into gross job creation – the sum of contributions from establishment births (job creation in newly established plants), acquisitions (job creation in plants acquired from other companies), and continuer expansions (job creation in existing plants) – and gross job destruction – the sum of contributions from death (job destruction froms plant closure), divestiture (job destruction from selling plants to other firms), and continuer contractions (job destruction in existing plants). We can further compute the organic job growth and destruction rate by removing acquisition and divestiture

[4]The COMPUSTAT-BR bridge in our approved project ends in 2011, but we track the extant matched firms in 2011 through 2013.

from gross job creation and destruction measure. The second panel in Table 3.3 shows the growth rate at the number of establishment level.

3.3 Firm-level Estimation and Quantification

In this section, we discuss our estimation equation, how we handle other threats to identification, and the empirical results and quantification.

3.3.1 Estimation Approach

Our baseline empirical model to estimate the effect of uncertainty on employment is a panel model with a firm (using the LBD and COMPUSTAT) over time. Specifically, our regression function is

$$\Delta y_{it} = \lambda \mathrm{CRUX}_{it} + \eta \mathrm{CRUX}_{it} \cdot \Delta \log(sales)_{it-1} + \beta^{sales} \Delta \log(sales)_{it-1} + \boldsymbol{\beta} \cdot \boldsymbol{X_{it-1}} + \alpha_i + \alpha_t + \epsilon_{it}.$$

The dependent variable is the arithmetic differences over the average of two consecutive years known as DHS growth rates or midpoint growth rates. These are $\Delta y_{it} = (y_{it} - y_{it-1})/(0.5 \times (y_{it} + y_{it-1}))$ where y_{it} is total employment. We include in the vector of controls $\boldsymbol{X_{it-1}}$ a set of first moment controls: average Tobin's Q ($\log(q_{it-1})$) to a proxy of firms' investment opportunities, lagged log sales growth ($\Delta \log(sales)_{it-1}$)to control for firm level demand shocks, and squared log sales growth to captures nonlinear effects. We also include firm-level effects α_i that absorb persistent firm shocks, time fixed effects α_t and ultimately industry-year effects that absorb industry and aggregate demand shocks, and any other unobservable firm characteristics and aggregate shocks that might influence Δy_{it}. Chapter 2 discusses our empirical model and identification strategy in detail.

3.3.2 Firm Level Employment Results - COMPUSTAT

Before turning to confidential microdata, we focus on employment measured using COMPUSTAT. We show a robust negative effect on employment from delay and caution effects to our text-based measure of uncertainty.

We first present a simple non-parametric evidence of reductions in employment growth rates. We divide the sample into high and low uncertainty by the median value of CRUX. We estimate a the kernel density of the employment growth rate distribution for high CRUX vs low CRUX firms, and plot the result in Figure 3.1. The low uncertainty distribution is shifted right and we reject equality via a Kolmogorov-Smirnov test.

Regression evidence in Table 3.2 confirms the negative impact of uncertianty, as measured by CRUX, on the employment growth rate. The employment growth rate is calculated as a DHS midpoint growth rate that is bounded on $[-2, 2]$ and equivalent to log changes up to a second-order Taylor approximation. This growth rate is more robust to outliers, of which there are a fairly large amount in COMPUSTAT employment. As in the investment results in Chapter 2, the employment results are consistent with our predictions. In columns (1) and (2) we find significant negative coefficients when only firm and year or industry-year fixed effects are included in the controls. The results are robust to adding demand shocks, the interaction terms, and Tobin's Q. The impact of CRUX on the employment rate of an average sales growth firm barely moves with or without all the controls. In column (3) a one standard deviation increase in CRUX results in nearly a 0.74 log point ($= -0.227 \times 0.0325 \times 100$) decrease in net job growth. The coefficient on the interaction term confirms that firms are more cautious when they are uncertain and wouldn't respond less to positive demand shocks. These effects are also fairly large. A one standard deviation sales shock would increase employment growth by 1.65 log points ($= 0.0293 \times 0.564 \times 100$) under no uncertainty, i.e.

if $CRUX = 0$. In the face of a one standard deviation shock to CRUX, that would be reduced by more than one half through a countervailing caution effect of 0.94 log points ($= -0.288 \times 0.0325 \times 100$). To account for the concern that the industry classifications of some conglomerates are not perfectly assigned, we adjust NAICS classification using COMPUSTAT Segment database and report the robustness checks in Table C1.

3.3.3 Employment Growth Effects

We now broaden our scope to understand the employment effects of uncertainty. Specific hiring and firing decisions are difficult to estimate without detailed data on labor adjustment costs, search and matching frictions, and regulations. Our approach exploits establishment job growth dynamics within firms across the margins of job growth.

The LBD data allow us to measure job growth at continuing establishments as well as births, deaths, acquisitions, divestitures. Observing the birth or acquisition of new establishment within the firm, for example, is an indicator of firm investment. We will estimate how these margins respond to changes in CRUX. We then compute the contribution of the response of each margin to CRUX for net job creation.

Table 3.3 provides summary stats and breakdown of employment growth across margins. Total employment growth in the top panel and net growth in the number of establishment are slightly negative on average in our sample, which includes the periods of the jobless recovery after the 2000-2001 recession, the financial crisis, Great Recession and its aftermath. In the margins, we see a strong contributions of continuing establishments to job creation (JC) and destruction (JD) of about 50% each. There is also a nontrivial contribution of the extensive margin, particularly for births (26% of total JC) and deaths (29% of JD). For growth in the *number* of establishments, births and deaths contribute to large amount of churning, following by acquisitions and divestitures. Our mean CRUX measure is 0.038 and we report sales growth for the LBD sample using data linked from COMPUSTAT.

We focus initially on net firm employment growth in Table 3.4 and step through the same set of controls we used in the COMPUSTAT results. In the first column we find a negative and precisely estimated effect of CRUX on employment growth when controlling for firmid panel fixed effects and year fixed effects. This holds up when adding additional first moment controls: industry-year effects in column (2) and sales growth in column (3). More importantly, we continue to find delay and caution effects when adding the interaction of CRUX with lagged sales growth. Both of these effects are robust to inclusion of Tobin's Q, squared sales growth, and all controls together in columns 5-7.

The magnitudes of these effects on employment growth are large on average. When CRUX is one SD above its mean (0.07) employment growth falls by 1.4 log points ($= -0.202 \times 0.07 \times 100$). If there is a one SD sales growth shock, employment grows by 5.7 log points when $CRUX = 0$. But when uncertainty is also high, that employment growth is attenuated by 30%, or 1.5 log points ($= -0.560 \times 0.07 \times 37$). To quantify the range of potential outcomes, we take coefficients from Table 3.4, column (4) and plot firm employment growth response to log sales growth under uncertainty in Figure 3.2. The delay effect shifts the entire response curve down and caution effects flatten the response.

The employment margins where firm-level uncertainty operates are important in their own right, but they also provide further evidence that CRUX measures firm exposure to uncertainty. We decompose employment growth into margins and estimate our baseline with CRUX, sales growth, their interaction, and a full set of firm and industry-year fixed effects.[5]

[5] The highlighted results are robust to the full set of controls in Table 3.9, but omitted for brevity.

We start with gross job creation and gross job destruction and their "organic" subcomponents that omit the acquisitions and divestitures. Table 3.6 reports these results and shows when firms are uncertain, the reduction in net employment growth occurs primarily at the job creation margin. Moreover, about 80% of the Gross JC effect is due to reductions in organic JC rather than margins of corporate restructuring. In short, firms create fewer new jobs when uncertainty is high the bulk of the response is not from reductions in acquisitions.

These results across margins are further evidence that CRUX measures of uncertainty conditional on our set of controls. We expand the level of detail in Table 3.5 to births, acquisitions, and continuers, and so on. About 3/4 of the estimated negative effect of uncertainty flows from reductions in job creation in establishment births and acquisitions; the latter are jobs-based measures of foregone investment opportunities. To corroborate this, we also find the divestiture margin has a have positive sign, indicating fewer job losses through this margin. The effect on job destruction through establishment death (shutdowns) is negative, but imprecisely different from zero. Moreover, the response of gross JD is not significantly different from zero (columns 3 and 4). The muted response on the job destruction margins is consistent with "wait and see" dynamics. A firm is less likely to disinvest in an ongoing operations when uncertainty is high.

To visualize these margins, Figure 3.3 plots the decomposition of the effect of uncertainty on all margins of gross job creation. The reduction in employment at establishment births is 51% of the delay effect ($= 0.105/(0.105 + 0.0463 + 0.0566)$) and 34% of the caution effect ($= 0.185/(0.185+0.164+0.202)$). If we focus only on organic gross job creation, establishment birth contributes to 65% of delay effects and more than 65% of the caution effect of reduction in employment. But even continuing establishment that are creating jobs do so at a slower rate when uncertainty is high.

We confirm these same patterns hold when looking only at the extensive margin growth in the number of operating establishments within the firm and their decomposition into births, deaths, acquisition and divestiture. As noted above, establishment openings and closings are measures of investment activity, but since we are more interested in the jobs associated with this extensive margin adjustment, we relegate these regressions to Tables 3.8 and 3.9.

In sum, our employment results across margins are strongly consistent with an (s, S) model investment and the job creation and destruction margins with non-convex adjustment costs. When uncertainty goes up the band of investment inaction widens. The hurdle rate for investment in the next new plant, acquisition, or added job goes up and we see less job creation as a result. Moreover, the decisions to shutdown plants, divest operations or reduce the workforce may be delayed while firms "wait and see" what the future holds.

3.3.4 Industry-level Measurement and Applications

This section addresses aggregation and external validity issues. First, we re-weight our regression and measure for the propensity we observe of publicly traded firm in the set of all private employers and test the robustness of our baseline results. Second, we construct industry-level CRUX measure to see whether idiosyncratic uncertainty can be captured through industry aggregation.

3.3.4.1 Propensity Score Weighting

Since we only observe CRUX for publicly traded firms, the results might not generalize to the private sector. To handle this issue, we treat the LBD as the population universe of all firms and estimate propensity scores for publicly traded firms in our sample that we use to inverse probability weight our regressions or construct aggregated

CRUX measures.

The propensity scores are constructed by fitting logit specifications for each fiscal year

$$\log \frac{p(\boldsymbol{X_{it}})}{1 - p(\boldsymbol{X_{it}})} = \boldsymbol{\theta_t X_{it}},$$

which implies that $\mathbb{P}(I_{it} = 1 \mid X_{it}) = \frac{1}{1+e^{-\boldsymbol{\theta_t X_{it}}}}$ where I_{it} is the indicator equal to 1 if the firm/establishment is selected in the SEC EDGAR - COMPUSTAT- CENSUS matched sample. In the LBD, sample, the control variables $\boldsymbol{X_{it}}$ include firm characteristics: 4-digit NAICS industry code, employment classes (1-249, 250-499, 500-999, 1000 or more), age class (1-5, 6-10, 11-15, 16-20, 21 years or more), payroll class (1 thousand dollars or less, 1-20, 20-200, 200-1000, 1000 thousands dollars or more), and indicator variable equal to 1 if the firm is included in the COMPUSTAT-BR bridge.[6] Our baseline results are largely unchanged when weighted by IPS scores and we report them in the appendix for the employment growth margins (Tables C2 and C3).

3.3.4.2 Peer Effects and Industry Aggregation

We now turn to the effect spillovers from peer uncertainty in the same industry and construct a "Peer" CRUX measure at the firm level (LBD sample) to measure within sample spillovers. In EDGAR-COMPUSTAT-LBD matched sample, we create Peer CRUX by taking the average firm demeaned CRUX of all other firms within the same industry. We find that Peer CRUX has limited effects on firm employment growth as in columns (1) and (3) in Table 3.7. When controlling for idiosyncratic uncertainty, the effects from the peers become insignificant as reported in columns (2) and (4). These patters are similar for gross job creation and destruction margins as reported in Table 3.10. Thus the reduction in employment growth is driven by idiosyncratic firm

[6]We choose these classes based on Foster et al. (2016) and the propensity score model in Davis et al. (2014).

uncertainty and peer uncertainty doesn't contribute much as in manufacturing results. One explanation could be because labor and capital have different adjustment costs or searching frictions. Our results are robust using IPS-weighted Peer CRUX as reported in Appendix Table C4 and C5.

3.4 Aggregate Validity and Effects

We construct a quarterly, aggregate CRUX measure as in Chapter 2. The measure ranges from the first quarter of 1998 to the fourth quarter of 2016.[7] To provide some aggregate evidence on employment dynamics and demonstrate that our measure has explanatory power on publicly available data, we use the BLS Business Employment Dynamics data. These data track changes in employment at the establishment level on a quarterly basis unemployment insurance data that represent job flows from about 90% of private non-farm payrolls. The data includes four types of job flows:

(1) employment gains at opening establishments (including re-openings)

(2) employment gains at expanding establishments

(3) employment losses at closing establishments (including seasonal closings)

(4) employment losses at contracting establishments.

Items (1) and (2) can be summed into total employment gains and (3) and (4) can be summed into total employment losses, while the difference between total gains and losses is the net employment change. The growth rate of type x is calculated by $(x_t - x_{t-1})/(0.5 \times (\text{total emp}_t + \text{total emp}_{t-1}))$ where x represents the level of the type and total emp is the total employment in the data.[8] The statistics are seasonally adjusted by the BLS.

[7] Measure construction details are introduced in Chapter 2.

[8] x can be all four types of job flows or gross employment gains/losses or net employment changes.

Establishment growth rate is calculated from levels using similar method. The four types of establishments associate with employment flow types:

(1) opening establishments (including re-openings)

(2) expanding establishment (continuing establishment with job creation)

(3) closing establishments (including seasonal closings)

(4) contracting establishments (continuing establishments with job destruction).

We first investigate the correlation between CRUX and aggregate employment decompositions. The BLS variables are in percentages, e.g., average net private employment DHS growth rate is around 0.2%. Note that all the BLS variable values are non-negative. Summary statistics are in Table 3.11.

Tables 3.12 and 3.13 report the correlation between CRUX and employment changes in different margins by running simple OLS regressions without controls. In Table 3.12, the dependent variable in column (1) the net change of total employment is computed as the difference between the dependent variables in column (2), gross employment gains, and column (5), gross employment losses. Other gross changes can be computed as discussed in the data section.

The results in Table 3.12 shows a strong negative correlation between CRUX and employment changes in all margins. Specifically, column (1) shows a one standard deviation increase in CRUX is associated with about 0.18% ($= -0.293 \times 0.00610 \times 100$) decrease in employment growth which is roughly equal to the average employment growth in the series. Columns (2)-(4) indicate that uncertainty reduces gross employment job gains and across the margins of establishment opening and expansion. Columns (5)-(7) suggest that uncertainty not only reduces the job creation rate, it reduces job destruction rate as well. This is consistent with "wait and see" effects whereby firms defer both hiring and firing, which entail irreversible costs, when uncertainty is high. The reduction in job creation is more than enough to offset the "wait and see" effects on job destruction such that net effect is negative. These results also imply that the job churning rate goes down when uncertainty is high.

Table 3.13 reports DHS growth rate of employment across establishment margins: job gains at continuers, job gains at new openings, job losses at continuers and job losses from closing. Columns (2)-(4) show CRUX has a negative effect on gross job growth at continuing and newly opened establishments. We find most of the effect comes from reductions in expansions at continuers, likely because new openings are a small share of total employment in the aggregate. Columns (5)-(7) show that CRUX also significantly decreases gross job losses. But it has a significant effect on job losses through establishment closings (column 6). To quantify the net effects using column (1), we find a one standard deviation increase in CRUX is associated with a reduction of almost 1% ($= -1.147 \times 0.00610 \times 100$) in net establishment growth.

3.5 Conclusion

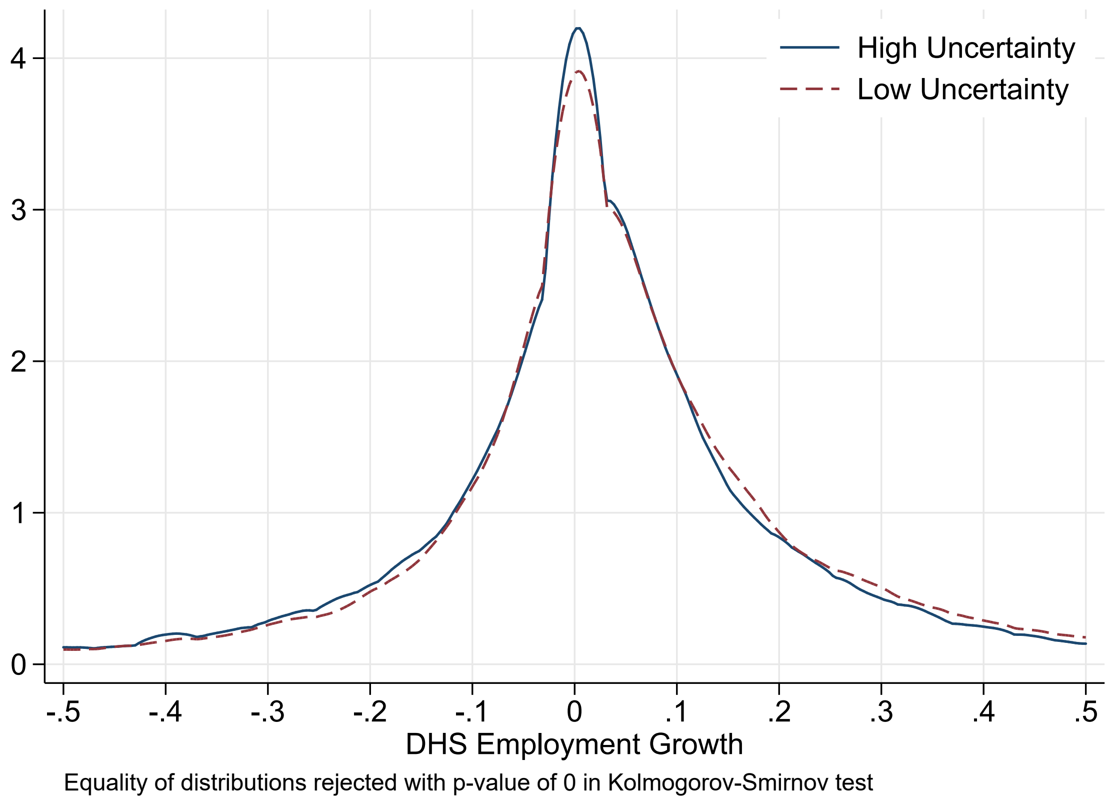
The effects of uncertainty on specific hiring and firing decisions are difficult to estimate without detailed data on regulations, labor adjustment costs, search and matching frictions. In this paper, we construct a new time-varying measure of firm-specific uncertainty from analyzing the text of company reports filed with the SEC and explore its implications in the firm-level hiring and within firm employment and establishment dynamics.

Using detailed establishment-level panel data on employment dynamics with a rich set of controls, we find our measure of firm-specific uncertainty has large negative effects on employment growth rates and firms' response of hiring to positive demand shocks. Most of the response within firms occurs on the gross job creation margin, primarily through reductions employment growth and new establishments, acquisitions, and continuer establishment job creation. The effect on gross job destruction is smaller and less precisely estimated. These two results are consistent with an (s,S) model of investment and hiring where uncertainty increases the hurdle rate for new investment, hiring, or expansions and reduces the threshold for disinvestment, firing, or plant shutdown.

An implications of our findings is that gross job reallocation, or job churning is reduced by firm-specific uncertainty. To the extent that reallocation facilitates the process of creative destruction and promotes productive growth (cf. Decker et al., 2017), our results suggest high uncertainty during and after recent recessions may have contributed to reductions in economic dynamic and productivity growth. Future research should further investigate these channels.

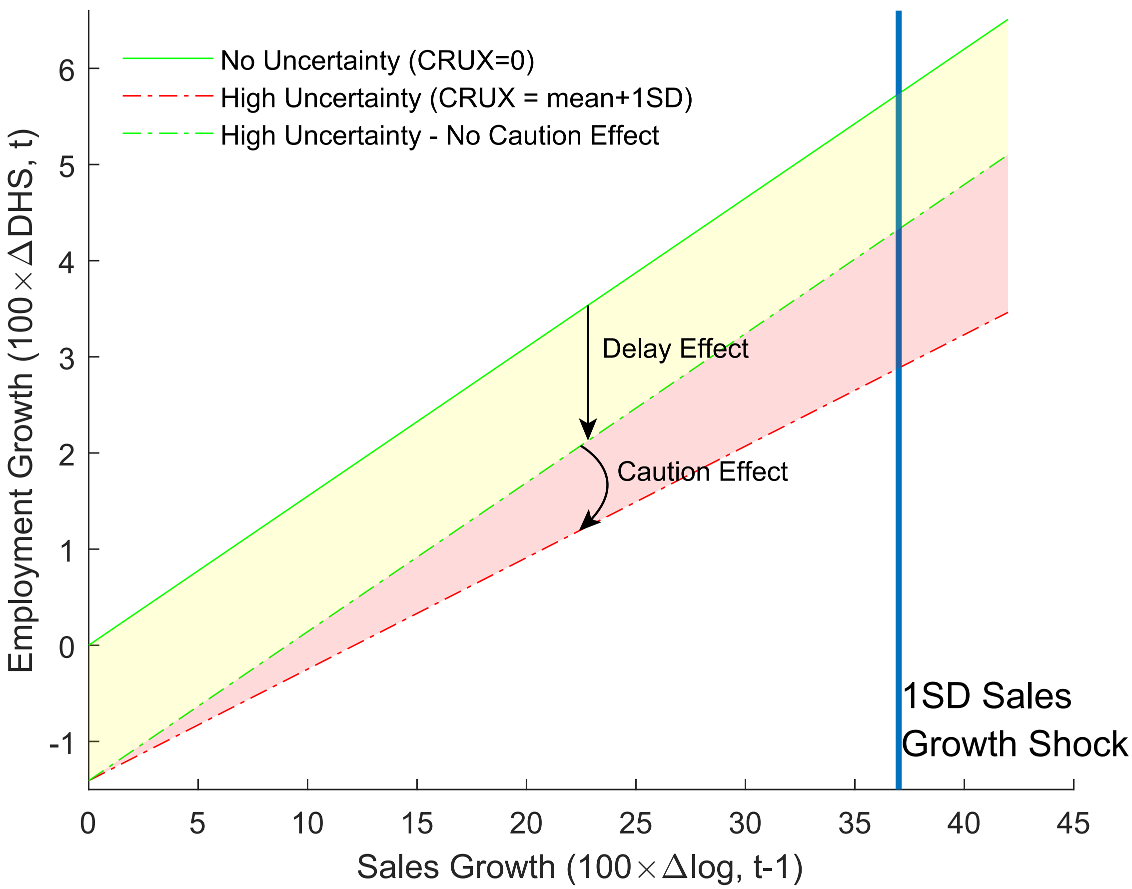
Figures

Figure 3.1: Firm Employment Growth under High vs Low Uncertainty



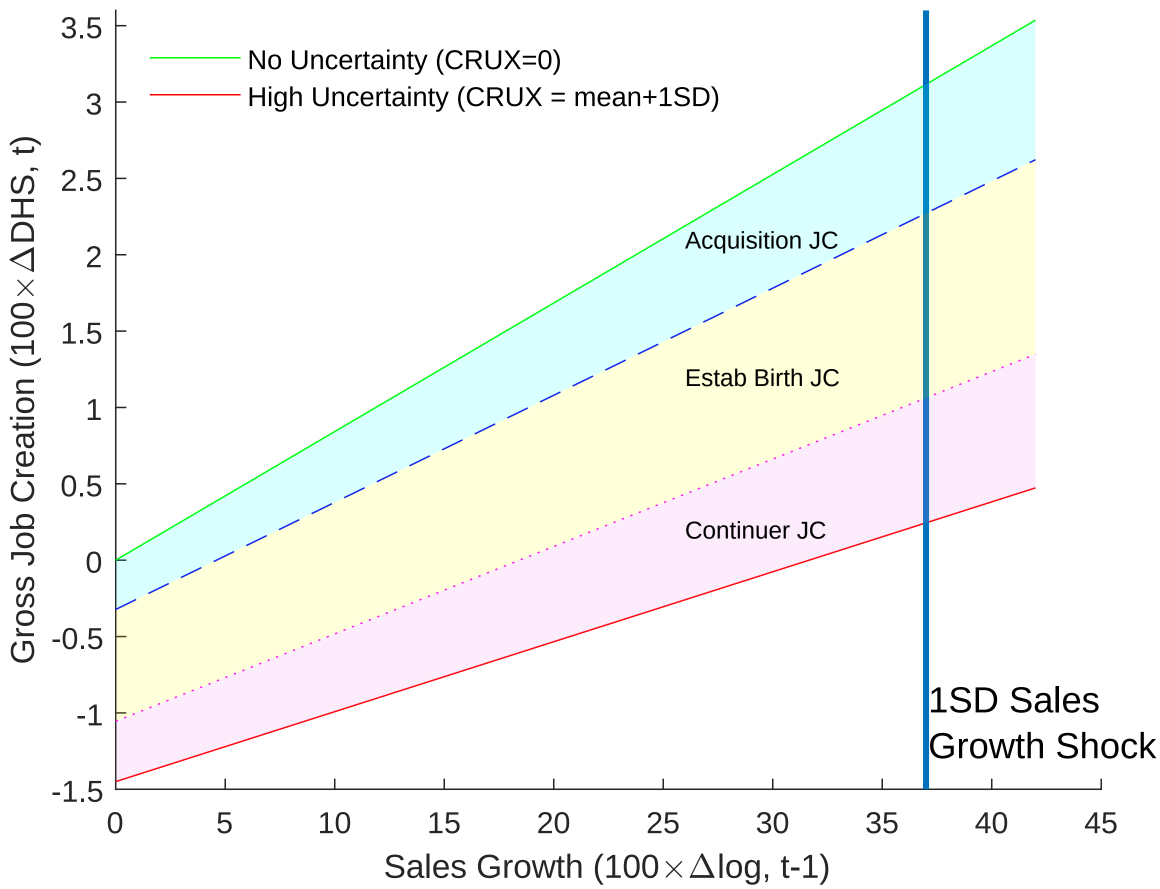
Notes: Epanechnikov kernel density estimates. High uncertainty defined as above median in uncertainty measure in Edgar-Compustat matched sample. Low uncertainty is below the median. Firm employment growth rate is calculated as the midpoint growth rate.

Figure 3.2: Delay vs Caution Effect of Employment Response to Sales Growth under Uncertainty



Notes: Comparing the differences between 5th percentile CRUX and one standard deviation above mean CRUX.

Figure 3.3: Decomposing Employment Response to Sales Growth under Uncertainty



Notes: Decompose job creation into birth, acquisition, and continuer margins in the U.S. Census sample, and plot the effects of high vs zero uncertainty on different margins.

Tables

Table 3.1: Summary Statistics - Compustat Firm Level Data

VARIABLES	Panel A Investment Log Growth	Panel B Employment DHS Growth
CRUX	0.0429 [0.0334]	0.0419 [0.0325]
Investment growth	0.123 [0.261]	-
Employment growth	-	0.0307 [0.330]
Lag sales growth (log)	0.119 [0.607]	0.115 [0.564]
Lag Tobin's Q (log)	0.595 [0.766]	0.529 [0.728]
Lag sales growth (log) squared	0.382 [2.272]	0.332 [2.059]
N	95,223	107,031

Panel A reports summary statistics of regression and non-parametric analysis on Compustat investment data. Investment growth is calculated as log growth. Panel B reports summary statistics of regression and non-parametric analysis on Compustat employment data. Employment growth is calculated in DHS form. All missing observations are dropped in both panels.

Table 3.2: Effect of Uncertainty on Firm Employment Growth Rate (ΔDHS, Compustat)

	Dependent Variable: DHS Change in Employment (Δdhs, t)					
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
CRUX (t)	-0.233***	-0.225***	-0.227***	-0.229***	-0.220***	-0.221***
	[0.0560]	[0.0567]	[0.0568]	[0.0568]	[0.0553]	[0.0553]
Sales Growth (Δln t-1)			0.0293***	0.0279***	0.0204***	0.0194***
			[0.00661]	[0.00666]	[0.00643]	[0.00652]
CRUX (t) × Sales Growth (Δln t-1)			-0.288***	-0.278**	-0.248**	-0.241**
			[0.108]	[0.108]	[0.107]	[0.108]
Sales Growth (Δln t-1) sqaured				0.00270*		0.00185
				[0.00152]		[0.00142]
Log Tobin's Q (t-1)					0.122***	0.121***
					[0.00447]	[0.00448]
Firm Fixed Effects	√	√	√	√	√	√
Year Fixed Effects	√					
Industry × Year Fixed Effects		√	√	√	√	√
Observations	107,030	105,412	105,412	105,412	105,412	105,412
R-squared	0.027	0.226	0.227	0.227	0.249	0.249
Number of Firms	13,130	11,569	11,569	11,569	11,569	11,569

Notes: Standard errors are clustered at firm level. Dependent variable is calculated as (emp(t)-emp(t-1))/(0.5×(emp(t)+emp(t-1))) where emp is firm's total employment. Firm idiosyncratic uncertainty measure CRUX is at time t, which as mentioned in the text captures firm uncertainty at time t-1. Sales growth is calculated as log(Sales(t-1)) - log(Sales(t-2)) and demeaned by sample mean. Squared sales growth is calculated using demeand sales growth. Tobin's Q is taken as log average Q at time t-1. All of the regression specifications include firm FE. Column (1) includes Year FE, Column (2)-(6) include industry × year FE, where industry is at 3-digit NAICS code level. Column (2)-(6) loses some observations and firms compared with Column (1) due to singleton observations. Time ranges from 1994 to 2016.

*** p<0.01, ** p<0.05, * p<0.1

Table 3.3: Summary Statistics - LBD

VARIABLES	Mean	Std Dev	Share
Total Employment Growth(A + B)	-0.630	[39.1]	-
A. Job Creation Rate (a + b + c)	16.830	[23.3]	-
a. Birth (Δdhs, Organic)	4.450	[13.5]	26%
b. Acquisition (Δdhs)	3.280	[13.2]	19%
c. Continuer (Δdhs, Organic)	9.100	[13.4]	54%
B. Job Destruction Rate (d + e + f)	-17.460	[30.4]	-
d. Death (Δdhs, Organic)	-5.080	[18.2]	29%
e. Divestiture (Δdhs)	-3.130	[20.4]	18%
f. Continuer (Δdhs, Organic)	-9.250	[14.6]	53%
Job Churning Rate (a + b + c - d - e - f)	34.290	[37.6]	-
Total Establishment Growth (a + b + c + d)	-0.360	[37.9]	-
a. Birth (Δdhs)	7.550	[17.8]	-
b. Death (Δdhs)	-8.510	[21.1]	-
c. Acquisition (Δdhs)	3.820	[13.9]	-
d. Divestiture (Δdhs)	-3.220	[20]	-
CRUX	0.038	[0.0317]	-
Peer CRUX	-0.058	[1.33]	-
Peer CRUX (weighted)	-0.061	[1.34]	-
Lag Tobin's Q (log)	0.399	[0.529]	-
Lag Sales Growth (log)	0.0774	[0.37]	-
Lag Sales Growth (log) Squared	0.143	[1.258]	-
N	55000		

Note: All variables except for lag Tobin's Q, lag sales growth and lag sales growth sqaured, are in percentage points (value form sample × 100).

Table 3.4: Effect of Uncertainty on Firm Employment Growth Rate (ΔDHS)

	Dependent Variable: DHS Change in Employment (Δdhs, t)						
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)
CRUX (t)	-0.256***	-0.241***	-0.191**	-0.202**	-0.199**	-0.204**	-0.201**
	[0.0839]	[0.0872]	[0.0860]	[0.0849]	[0.0843]	[0.0848]	[0.0842]
Sales Growth (Δln t-1)			0.132***	0.155***	0.148***	0.155***	0.147***
			[0.0113]	[0.0158]	[0.0155]	[0.0158]	[0.0155]
CRUX (t) × Sales Growth (Δln t-1)				-0.560**	-0.541**	-0.591**	-0.584***
				[0.235]	[0.232]	[0.231]	[0.226]
Sales Growth (Δln t-1) squared						-0.003	-0.00421
						[0.00395]	[0.00383]
Log Tobin's Q (t-1)					0.0806***		0.0814***
					[0.00778]		[0.00780]
Firm Fixed Effects	√	√	√	√	√	√	√
Year Fixed Effects	√						
Industry × Year Fixed Effects		√	√	√	√	√	√
R-squared	0.028	0.241	0.253	0.253	0.256	0.253	0.257

Notes: Standard errors are clustered at firm level. Dependent variable is calculated as (emp(t)-emp(t-1))/ (0.5×(emp(t)+emp(t-1))) where emp is firm's total employment. Firm idiosyncratic uncertainty measure CRUX is at time t, which as mentioned in the text captures firm uncertainty at time t-1. Sales growth is calculated as log(Sales(t-1)) - log(Sales(t-2)) and demeaned by sample mean. Squared sales growth is calculated using demeand sales growth. Tobin's Q is taken as log average Q at time t-1. All of the regression specifications include firm FE. Column (1) includes Year FE, Columns (2)-(7) include industry × year FE, where industry is at 3-digit NAICS code level. Time ranges from 1994 to 2014. Number of observations is 55000 and number of firms is 6000, both rounded to the nearest thousands.
*** p<0.01, ** p<0.05, * p<0.1

Table 3.5: Effect of Uncertainty on Firm Employment Growth Rate Decomposition (ΔDHS)

	Dependent Variable: DHS Change in Employment (Δdhs, t)								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
		Extensive Margin					Continuer		
VARIABLES	Total	Birth	Death	Acquisition	Divestiture	Net	Creation	Destruction	Churning
CRUX (t)	-0.202**	-0.105***	-0.0361	-0.0463*	0.0841*	-0.0987**	-0.0566*	-0.0421	-0.214**
	[0.0849]	[0.0305]	[0.0448]	[0.0276]	[0.0462]	[0.0484]	[0.0294]	[0.0356]	[0.0891]
Sales Growth (Δln t-1)	0.155***	0.0219***	0.0331***	0.0247***	0.00464	0.0707***	0.0377***	0.0330***	0.0135
	[0.0158]	[0.00349]	[0.00692]	[0.00375]	[0.00830]	[0.00982]	[0.00473]	[0.00733]	[0.0119]
CRUX (t) × Sales Growth (Δln t-1)	-0.560**	-0.185***	-0.114	-0.202***	0.0495	-0.109	-0.164***	0.0552	-0.542**
	[0.235]	[0.0483]	[0.137]	[0.0484]	[0.142]	[0.166]	[0.0633]	[0.134]	[0.236]
R-squared	0.253	0.236	0.259	0.231	0.249	0.192	0.236	0.236	0.327

Notes: Standard errors are clustered at firm level. Dependent variable is calculated as (emp(t)-emp(t-1))/ (0.5×(emp(t)+emp(t-1))) in different margins. Column (1) represents firm's total employment change. Columns (2)-(5) represent firm's employment change from establishment birth, death, acquisition and divestiture. Column (6) represents gross employment change from firm's continuing establishments. Columns (7)-(8) represent job creation and destruction in firm's continuing establishments. Column (9) represents job churning rate, which is constructed as the sum of birth, acquisition and continuer job creation less the sum of death, divestiture and continuer job destruction. Coefficients in Columns (2)-(6) should add up to coefficients in Column (1). Coefficients in Columns (7)-(8) should add up to coefficients in Column (6). Firm idiosyncratic uncertainty measure CRUX is at time t, which as mentioned in the text captures firm uncertainty at time t-1. Sales growth is calculated as log(Sales(t-1)) - log(Sales(t-2)) and demeaned by sample mean. All of the regression specifications include firm FE and industry × year FE, where industry is at 3-digit NAICS code level. Time ranges from 1994 to 2014. Number of observations is 55000 and number of firms is 6000, both rounded to the nearest thousands.
*** p<0.01, ** p<0.05, * p<0.1

Table 3.6: Effect of Uncertainty on Firm Organic Employment Growth Rate (ΔDHS)

	Dependent Variable: DHS Change in Employment (Δdhs, t)			
	(1)	(2)	(3)	(4)
VARIABLES	Gross Job Creation	Organic Job Creation	Gross Job Destruction	Organic Job Destruction
CRUX (t)	-0.208***	-0.162***	0.00591	-0.0782
	[0.0493]	[0.0420]	[0.0717]	[0.0570]
Sales Growth (Δln t-1)	0.0842***	0.0596***	0.0707***	0.0661***
	[0.00755]	[0.00610]	[0.0118]	[0.00768]
CRUX (t) × Sales Growth (Δln t-1)	-0.551***	-0.349***	-0.00899	-0.0585
	[0.0969]	[0.0786]	[0.215]	[0.149]
R-squared	0.283	0.276	0.291	0.3

Notes: Standard errors are clustered at firm level. Dependent variable is calculated as (emp(t)-emp(t-1))/ (0.5×(emp(t)+emp(t-1))) in different margins. Columns (1) represents firm's gross job creation, which is the sum of organic job creation and job creation from establishment acquisition. Columns (2) represents firm's organic job creation, which is the sum of job creation from establishment birth and continuing establishments. Columns (3) and (4) are similar but represent job destruction margin. Firm idiosyncratic uncertainty measure CRUX is at time t, which as mentioned in the text captures firm uncertainty at time t-1. Sales growth is calculated as log(Sales(t-1)) - log(Sales(t-2)) and demeaned by sample mean. All of the regression specifications include firm FE and industry × year FE, where industry is at 3-digit NAICS code level. Time ranges from 1994 to 2014. Number of observations is 55000 and number of firms is 6000, both rounded to the nearest thousands.
*** p<0.01, ** p<0.05, * p<0.1

Table 3.7: Effect of Peer vs Idiosyncratic Uncertainty on Firm Employment Growth Rate (ΔDHS)

	Dependent Variable: DHS Change in Employment (Δdhs, t)			
VARIABLES	(1)	(2)	(3)	(4)
Peer CRUX (t)	-0.137	-0.0937	-0.121	-0.0771
	[0.314]	[0.316]	[0.314]	[0.315]
Peer CRUX (t) × Sales Growth (Δln t-1)	-1.341*	-0.876	-1.370**	-0.886
	[0.703]	[0.727]	[0.684]	[0.718]
CRUX (t)		-0.202**		-0.201**
		[0.0852]		[0.0844]
CRUX (t) × Sales Growth (Δln t-1)		-0.463*		-0.485**
		[0.250]		[0.243]
Sales Growth (Δln t-1)	0.129***	0.149***	0.120***	0.141***
	[0.0112]	[0.0162]	[0.0118]	[0.0158]
Sales Growth (Δln t-1) squared			-0.00348	-0.00415
			[0.00375]	[0.00383]
Log Tobin's Q (t-1)			0.0815***	0.0814***
			[0.00780]	[0.00780]
R-squared	0.253	0.253	0.256	0.257

Notes: Standard errors are clustered at firm level. Dependent variable is calculated as (emp(t)-emp(t-1))/ (0.5×(emp(t)+emp(t-1))) where emp is firm's total employment. Firm idiosyncratic uncertainty measure CRUX is at time t, which as mentioned in the text captures firm uncertainty at time t-1. Uncertainty measure of firm's peers is calculated by taking equally weighted average of firm FE demeaned CRUX measure of all other firms within the same industry (4-digit NAICS code). Sales growth is calculated as log(Sales(t-1)) - log(Sales(t-2)) and demeaned by sample mean. Squared sales growth is calculated using demeand sales growth. Tobin's Q is taken as log average Q at time t-1. All of the regression specifications include firm FE and industry × year FE, where industry is at 3-digit NAICS code level. Time ranges from 1994 to 2014. Number of observations is 55000 and number of firms is 6000, both rounded to the nearest thousands.
*** p<0.01, ** p<0.05, * p<0.1

Table 3.8: Effect of Uncertainty on Firm Establishment Growth Rate (ΔDHS)

	Dependent Variable: DHS Change in Number of Establishments (Δdhs, t)						
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)
CRUX (t)	-0.132	-0.145*	-0.12	-0.13	-0.129	-0.13	-0.129
	[0.0843]	[0.0874]	[0.0866]	[0.0864]	[0.0864]	[0.0863]	[0.0863]
Sales Growth (Δln t-1)			0.0663***	0.0884***	0.0854***	0.0884***	0.0853***
			[0.00867]	[0.0123]	[0.0123]	[0.0124]	[0.0124]
CRUX (t) × Sales Growth (Δln t-1)				-0.529***	-0.521***	-0.530***	-0.527***
				[0.199]	[0.198]	[0.199]	[0.198]
Sales Growth (Δln t-1) squared						-0.0000843	-0.000584
						[0.00291]	[0.00287]
Log Tobin's Q (t-1)					0.0334***		0.0335***
					[0.00709]		[0.00711]
Firm Fixed Effects	√	√	√	√	√	√	√
Year Fixed Effects	√						
Industry × Year Fixed Effects		√	√	√	√	√	√
R-squared	0.02	0.23	0.233	0.233	0.234	0.233	0.234

Notes: Standard errors are clustered at firm level. Dependent variable is calculated as (est(t)-est(t-1))/ (0.5×(est(t)+est(t-1))) where est is firm's total number of establishments. Firm idiosyncratic uncertainty measure CRUX is at time t, which as mentioned in the text captures firm uncertainty at time t-1. Sales growth is calculated as log(Sales(t-1)) - log(Sales(t-2)) and demeaned by sample mean. Squared sales growth is calculated using demeand sales growth. Tobin's Q is taken as log average Q at time t-1. All of the regression specifications include firm FE. Column (1) includes Year FE, Columns (2)-(7) include industry × year FE, where industry is at 3-digit NAICS code level. Time ranges from 1994 to 2014. Number of observations is 55000 and number of firms is 6000, both rounded to the nearest thousands.
*** p<0.01, ** p<0.05, * p<0.1

Table 3.9: Effect of Uncertainty on Firm Establishment Growth Rate Decomposition (ΔDHS)

	Dependent Variable: DHS Change in Establishment (Δdhs, t)				
	(1)	(2)	(3)	(4)	(5)
VARIABLES	Total	Birth	Death	Acquisition	Divestiture
CRUX (t)	-0.13	-0.144***	-0.00786	-0.0649**	0.0865*
	[0.0864]	[0.0410]	[0.0521]	[0.0297]	[0.0450]
Sales Growth (Δln t-1)	0.0884***	0.0260***	0.0375***	0.0209***	0.00393
	[0.0123]	[0.00405]	[0.00691]	[0.00323]	[0.00733]
CRUX (t) × Sales Growth (Δln t-1)	-0.529***	-0.193***	-0.169	-0.162***	-0.00597
	[0.199]	[0.0576]	[0.140]	[0.0461]	[0.119]
R-squared	0.233	0.231	0.251	0.217	0.244

Notes: Standard errors are clustered at firm level. Dependent variable is calculated as (est(t)-est(t-1))/ (0.5×(est(t)+est(t-1))) in different margins. Column (1) represents firm's total number of establishments change. Columns (2)-(5) represent firm's establishment change from birth, death, acquisition and divestiture. Coefficients in Columns (2)-(5) should add up to coefficients in Column (1). Firm idiosyncratic uncertainty measure CRUX is at time t, which as mentioned in the text captures firm uncertainty at time t-1. Sales growth is calculated as log(Sales(t-1)) - log(Sales(t-2)) and demeaned by sample mean. All of the regression specifications include firm FE and industry × year FE, where industry is at 3-digit NAICS code level. Time ranges from 1994 to 2014. Number of observations is 55000 and number of firms is 6000, both rounded to the nearest thousands.
*** p<0.01, ** p<0.05, * p<0.1

Table 3.10: Effect of Peer vs Idiosyncratic Uncertainty on Firm Employment Growth Rate (ΔDHS) on Different Margins

	Dependent Variable: DHS Change in Employment (Δdhs, t)									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
VARIABLES	Gross Job Creation		Gross Job Destruction		Organic Job Creation		Organic Job Destruction		Job Churning	
Peer CRUX (t)	-0.207	-0.163	0.0699	0.0689	-0.141	-0.107	-0.103	-0.0877	-0.277	-0.231
	[0.167]	[0.169]	[0.243]	[0.243]	[0.145]	[0.146]	[0.188]	[0.188]	[0.274]	[0.274]
Peer CRUX (t) × Sales Growth (Δln t-1)	-1.029***	-0.533	-0.312	-0.343	-0.707**	-0.403	-0.15	-0.107	-0.716	-0.19
	[0.337]	[0.346]	[0.558]	[0.600]	[0.278]	[0.289]	[0.446]	[0.481]	[0.596]	[0.656]
CRUX (t)		-0.206***		0.00437		-0.161***		-0.0767		-0.210**
		[0.0495]		[0.0718]		[0.0422]		[0.0571]		[0.0892]
CRUX (t) × Sales Growth (Δln t-1)		-0.492***		0.0298		-0.305***		-0.0469		-0.522**
		[0.0960]		[0.234]		[0.0803]		[0.162]		[0.255]
Sales Growth (Δln t-1)	0.0591***	0.0805***	0.0696***	0.0683***	0.0435***	0.0568***	0.0634***	0.0654***	-0.0106	0.0122
	[0.00546]	[0.00751]	[0.00808]	[0.0127]	[0.00460]	[0.00618]	[0.00578]	[0.00859]	[0.00802]	[0.0132]
R-squared	0.283	0.284	0.291	0.291	0.276	0.276	0.3	0.3	0.326	0.327

Notes: Standard errors are clustered at firm level. Dependent variable is calculated as (emp(t)-emp(t-1))/ (0.5×(emp(t)+emp(t-1))) in different margins. Columns (1)-(2) represent firm's gross job creation, which is the sum of organic job creation and job creation from establishment acquisition. Columns (5)-(6) represent firm's organic job creation, which is the sum of job creation from establishment birth and continuing establishments. Columns (3)-(4) and (7)-(8) are similar but represent job destruction margin. Columns (9)-(10) represent job churning rate, which is constructed as the sum of birth, acquisition and continuer job creation less the sum of death, divestiture and continuer job destruction. Firm idiosyncratic uncertainty measure CRUX is at time t, which as mentioned in the text captures firm uncertainty at time t-1. Uncertainty measure of firm's peers is calculated by taking equally weighted average of firm FE demeaned CRUX measure of all other firms within the same industry (4-digit NAICS code). Sales growth is calculated as log(Sales(t-1)) - log(Sales(t-2)) and demeaned by sample mean. All of the regression specifications include firm FE and industry × year FE, where industry is at 3-digit NAICS code level. Time ranges from 1994 to 2014. Number of observations is 55000 and number of firms is 6000, both rounded to the nearest thousands.

*** p<0.01, ** p<0.05, * p<0.1

Table 3.11: Summary Statistics - BLS

	Employment	Establishment
CRUX	0.000851	0.000851
	[0.00610]	[0.00610]
Net	0.00203	0.00204
	[0.00622]	[0.0168]
Gross Gain	0.0683	0.272
	[0.00679]	[0.0164]
Gain (opening)	0.0137	0.0562
	[0.00224]	[0.00255]
Gain (expansion)	0.0546	0.216
	[0.00472]	[0.0151]
Gross Loss	0.0663	0.270
	[0.00730]	[0.0156]
Loss (closing)	0.0130	0.0538
	[0.00218]	[0.00346]
Loss (contraction)	0.0533	0.216
	[0.00535]	[0.0136]
N	76	76

This is quarterly data from the 1st quarter of 1998 to the 4th quarter of 2016. Uncertianty measure CRUX is constructed as the average CRUX measure of current quarter and 3 lags weighted by length of SEC Edgar documents demeaned at firm level. BLS data are calculated in DHS form, i.e., difference over average of two consecutive periods.

Table 3.12: Aggregate Employment Effects—BLS Business Employment Dynamics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			Employment Gain			Employment Loss	
VARIABLES	Net Emp Change	Gross	Opening	Expansion	Gross	Closing	Contraction
CRUX	-0.293**	-0.860***	-0.268***	-0.591***	-0.567***	-0.225***	-0.342***
	[0.113]	[0.105]	[0.0393]	[0.0704]	[0.133]	[0.0372]	[0.0992]
Constant	0.00228***	0.0691***	0.0140***	0.0551***	0.0668***	0.0132***	0.0536***
	[0.000650]	[0.000495]	[0.000181]	[0.000347]	[0.000713]	[0.000196]	[0.000539]
Observations	76	76	76	76	76	76	76
R-squared	0.083	0.597	0.534	0.582	0.224	0.395	0.152

Notes: The data ranges from the 1st quarter of 1998 to the 4th quarter of 2016. Uncertianty measure CRUX is constructed as the average CRUX measure of current quarter and 3 lags weighted by length of SEC Edgar documents demeaned at firm level. Employment data is retrieved from BLS Business Dynamics which covers around 98% of total private nonfarm payroll of the US economy. Column (1) represents net employment change rate which can be decomposed into Column (2) and (5). Column (2) represents gross employment gains rate which can be decomposed into Column (3) employment gains rate from opening establishments and Column (4) employment gains rate from establishment expansion. Column (5) represents gross employment losses rate which can be decomposed into Column (6) employment losses rate from closing establishments and Column (7) employment losses rate from establishment contraction. The
*** p<0.01, ** p<0.05, * p<0.1

Table 3.13: Aggregate Establishment Effects—BLS Business Employment Dynamics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			Establishment Gain			Establishment Loss	
VARIABLES	Net Est Change	Gross	Opening	Expansion	Gross	Closing	Contraction
CRUX	-1.147***	-2.007***	-0.150***	-1.857***	-0.860***	0.128**	-0.988***
	[0.298]	[0.274]	[0.0493]	[0.244]	[0.309]	[0.0602]	[0.262]
Constant	0.00302*	0.274***	0.0563***	0.217***	0.271***	0.0537***	0.217***
	[0.00161]	[0.00120]	[0.000267]	[0.00109]	[0.00157]	[0.000358]	[0.00131]
Observations	76	76	76	76	76	76	76
R-squared	0.173	0.555	0.128	0.563	0.113	0.051	0.197

Notes: The data ranges from the 1st quarter of 1998 to the 4th quarter of 2016. Uncertianty measure CRUX is constructed as the average CRUX measure of current quarter and 3 lags weighted by length of SEC Edgar documents demeaned at firm level. Employment data is retrieved from BLS Business Dynamics which covers around 98% of total private nonfarm payroll of the US economy. Column (1) represents net establishment change rate which can be decomposed into Column (2) and (5). Column (2) represents gross establishment gains rate which can be decomposed into Column (3) establishment change rate of opening and Column (4) establishment change rate of expanding establishments. Column (5) represents gross establishment losses rate which can be decomposed into Column (6) establishment change rate of closing establishments and Column (7) establishment change rate from contracting establishments. The data is seasonally

*** p<0.01, ** p<0.05, * p<0.1

APPENDICES

APPENDIX A

Appendix to Chapter 1

Table A.1: PTF and Mergers Summary Statistics

VARIABLES	(1) N	(2) mean	(3) sd	(4) min	(5) max
PTF Similarity (t-1)	8,068,691	0.0619	0.0745	2.55e-06	1
Competition (HP, t-1)	6,066,320	0.133	0.0952	0	0.949
Merger Deal Indicator	8,068,691	0.000213	0.0146	0	1
|Δ PTF| (t-1)	8,068,691	0.000982	0.000942	0	0.00849
|Δ PTF Breadth| (t-1)	8,068,691	3.37e-05	3.25e-05	0	0.000224
|Δ log(total asset)| (t-1)	8,068,691	2.119	1.806	0	15.61
|Δ cash| (t-1)	8,068,691	1.129	1.048	0	7.808
|Δ sales growth (ln)| (t-1)	8,068,691	0.356	0.477	0	4.410
|Δ log(Tobin's Q)| (t-1)	8,068,691	0.444	0.655	0	6.645
|Δ book leverage| (t-1)	8,068,691	0.204	0.505	0	15.24
|Δ ROA| (t-1)	8,068,691	0.241	0.850	0	20.93
Adjusted by Industry (4-digit SIC code) Average					
|Δ PTF| (t-1)	8,068,691	0.000963	0.000926	0	0.00850
|Δ PTF Breadth| (t-1)	8,068,691	3.29e-05	3.15e-05	0	0.000272
|Δ log(total asset)| (t-1)	8,068,691	1.977	1.666	0	15.67
|Δ cash| (t-1)	8,068,691	1.017	0.955	0	8.896
|Δ sales growth (ln)| (t-1)	8,068,691	0.352	0.469	0	4.676
|Δ log(Tobin's Q)| (t-1)	8,068,691	0.412	0.611	0	7.047
|Δ book leverage| (t-1)	8,068,691	0.204	0.497	0	15.24
|Δ ROA| (t-1)	8,068,691	0.246	0.837	0	20.93

Notes: This table provides summary statistics for analysis on merger and acquisitions. The sample is created by taking Cartesian interaction of all possible mergers from industries that involve a merger deal in the given year. There are total 1707 announced deals. Industry is classified at 4-digit SIC code. PTF is winsorized at 3 standard deviation from mean by year. Total assets, sales growth, Tobin's Q, book leverage, and ROA are winsorized at 1% and 99% by year.

Table A.2: PTF and Determinants of Mergers

(a) PTF and Mergers

	Dependent Variable: Indicator of Merger Deal							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES					Industry Average Adjusted Measure			
|Δ PTF| (t-1)	0.0306***	0.0250***	0.0305***	0.0245***	0.0359***	0.0304***	0.0367***	0.0312***
	[0.00652]	[0.00652]	[0.00802]	[0.00803]	[0.00676]	[0.00676]	[0.00829]	[0.00829]
PTF Similarity (t-1)	0.00261***	0.00254***	0.00217***	0.00217***	0.00262***	0.00257***	0.00219***	0.00220***
	[0.000418]	[0.000408]	[0.000381]	[0.000380]	[0.000419]	[0.000412]	[0.000383]	[0.000382]
|Δ PTF Breadth| (t-1)		0.764***		0.770***		0.752***		0.691***
		[0.209]		[0.268]		[0.206]		[0.261]
Competition (HP, t-1)			0.00221***	0.00213***			0.00220***	0.00217***
			[0.000300]	[0.000292]			[0.000300]	[0.000296]
|Δ log(total asset)| (t-1)		-0.00162***		9.59e-05		-0.00137***		6.86e-05
		[0.000369]		[0.000455]		[0.000370]		[0.000456]
|Δ cash| (t-1)		-0.00439***		-0.00438***		-0.00256***		-0.00186**
		[0.000678]		[0.000913]		[0.000662]		[0.000920]
|Δ sales growth (ln)| (t-1)		-0.00615***		-0.00494***		-0.00611***		-0.00526***
		[0.00123]		[0.00163]		[0.00120]		[0.00161]
|Δ log(Tobin's Q)| (t-1)		-0.0101***		-0.0114***		-0.00900***		-0.0102***
		[0.00133]		[0.00185]		[0.00130]		[0.00184]
|Δ book leverage| (t-1)		0.000265		-0.00158		-0.00123		-0.00546*
		[0.000955]		[0.00275]		[0.00109]		[0.00294]
|Δ ROA| (t-1)		-0.000215		-0.00775***		0.000591		-0.00287*
		[0.000624]		[0.00171]		[0.000570]		[0.00149]
Acq Ind×Tar Ind×Year FE	√	√	√	√	√	√	√	√
Observations	8,068,691	8,068,691	6,066,320	6,066,320	8,068,691	8,068,691	6,066,320	6,066,320
R-squared	0.011	0.011	0.014	0.014	0.011	0.011	0.014	0.014

Notes: Standard errors are clustered at acquirer industry - target industry - year level. Dependent variable is an indicator = 1 if a merger deal is announced and = 0 otherwise. |Δ PTF| is the absolute value of the difference of PTF of the acquirer and the target. PTF Similarity measures the pairwise similarity of firm-level technology frontier. |Δ PTF Breadth| captures the absolute value of the difference of technology frontier breadth of the acquirer and the target. Competition is drawn from Hoberg and Phillips (2016) to gauge pairwise product competition between firms. Control variables include the absolute value of the difference of total asset, cash holding, sales growth, Tobin's Q, book leverage, and ROA, of the acquirer and the target, adjusted by dividing 100. All independent variables except for PTF Similarity and Competition are demeaned by industry average in columns (5)-(8). Acquirer industry - target industry - year fixed effects are included in all specifications. Industry is defined at 4-digit SIC level.

Table A.2: PTF and Determinants of Mergers

(b) PTF, Assets and Mergers

	Dependent Variable: Indicator of Merger Deal							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES					Industry Average Adjusted Measure			
|Δ PTF| (t-1)	0.0323***	0.0260***	0.0294***	0.0233***	0.0371***	0.0308***	0.0366***	0.0310***
	[0.00687]	[0.00680]	[0.00823]	[0.00822]	[0.00695]	[0.00688]	[0.00836]	[0.00835]
PTF Similarity (t-1)	0.00257***	0.00255***	0.00219***	0.00219***	0.00258***	0.00258***	0.00219***	0.00220***
	[0.000414]	[0.000410]	[0.000383]	[0.000382]	[0.000416]	[0.000413]	[0.000383]	[0.000383]
|Δ PTF Breadth| (t-1)		0.770***		0.772***		0.755***		0.692***
		[0.209]		[0.268]		[0.206]		[0.261]
Competition (HP, t-1)			0.00219***	0.00213***			0.00220***	0.00217***
			[0.000298]	[0.000291]			[0.000298]	[0.000296]
|Δ PTF| (t-1) × |Δ log(total asset)| (t-1)	-1.053***	-1.090***	-1.346***	-1.369***	-0.419	-0.465	-0.256	-0.294
	[0.290]	[0.291]	[0.426]	[0.427]	[0.298]	[0.298]	[0.429]	[0.429]
|Δ log(total asset)| (t-1)	-0.00130***	-0.000486	0.000898	0.00151**	-0.00149***	-0.000895*	6.95e-05	0.000367
	[0.000438]	[0.000440]	[0.000609]	[0.000618]	[0.000460]	[0.000461]	[0.000614]	[0.000617]
|Δ cash| (t-1)		-0.00439***		-0.00437***		-0.00256***		-0.00186**
		[0.000677]		[0.000912]		[0.000662]		[0.000920]
|Δ sales growth (ln)| (t-1)		-0.00616***		-0.00494***		-0.00611***		-0.00527***
		[0.00123]		[0.00162]		[0.00120]		[0.00161]
|Δ log(Tobin's Q)| (t-1)		-0.0101***		-0.0114***		-0.00901***		-0.0102***
		[0.00133]		[0.00185]		[0.00130]		[0.00184]
|Δ book leverage| (t-1)		0.000258		-0.00166		-0.00123		-0.00548*
		[0.000946]		[0.00275]		[0.00108]		[0.00294]
|Δ ROA| (t-1)		-0.000266		-0.00780***		0.000574		-0.00288*
		[0.000616]		[0.00171]		[0.000566]		[0.00148]
Acq Ind×Tar Ind×Year FE	√	√	√	√	√	√	√	√
Observations	8,068,691	8,068,691	6,066,320	6,066,320	8,068,691	8,068,691	6,066,320	6,066,320
R-squared	0.011	0.011	0.014	0.014	0.011	0.011	0.014	0.014

Notes: Standard errors are clustered at acquirer industry - target industry - year level. Dependent variable is an indicator = 1 if a merger deal is announced and = 0 otherwise. |Δ PTF| is the absolute value of the difference of PTF of the acquirer and the target. PTF Similarity measures the pairwise similarity of firm-level technology frontier. |Δ PTF Breadth| captures the absolute value of the difference of technology frontier breadth of the acquirer and the target. Competition is drawn from Hoberg and Phillips (2016) to gauge pairwise product competition between firms. Control variables include the absolute value of the difference of total asset, cash holding, sales growth, Tobin's Q, book leverage, and ROA, of the acquirer and the target. All independent variables except for PTF Similarity and Competition are demeaned by industry average in columns (5)-(8). Acquirer industry - target industry - year fixed effects are included in all specifications. Industry is defined at 4-digit SIC level.

Table A.2: PTF and Determinants of Mergers

(c) PTF, Sales Growth and Mergers

	Dependent Variable: Indicator of Merger Deal							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES					Industry Average Adjusted Measure			
|Δ PTF| (t-1)	0.0277***	0.0228***	0.0266***	0.0213**	0.0339***	0.0290***	0.0344***	0.0295***
	[0.00671]	[0.00672]	[0.00828]	[0.00829]	[0.00686]	[0.00686]	[0.00845]	[0.00844]
PTF Similarity (t-1)	0.00262***	0.00255***	0.00219***	0.00219***	0.00263***	0.00258***	0.00220***	0.00221***
	[0.000418]	[0.000410]	[0.000383]	[0.000383]	[0.000419]	[0.000413]	[0.000384]	[0.000384]
|Δ PTF Breadth| (t-1)		0.750***		0.752***		0.742***		0.681***
		[0.208]		[0.268]		[0.206]		[0.261]
Competition (HP, t-1)			0.00220***	0.00213***			0.00220***	0.00216***
			[0.000299]	[0.000291]			[0.000299]	[0.000295]
|Δ PTF| (t-1) × |Δ sales growth (ln)| (t-1)	-4.790***	-4.348***	-5.120***	-4.742***	-3.141***	-2.785**	-2.750*	-2.474
	[1.085]	[1.086]	[1.495]	[1.495]	[1.114]	[1.118]	[1.566]	[1.571]
|Δ log(total asset)| (t-1)		-0.00161***		0.000112		-0.00136***		7.81e-05
		[0.000368]		[0.000455]		[0.000370]		[0.000456]
|Δ cash| (t-1)		-0.00438***		-0.00437***		-0.00256***		-0.00185**
		[0.000677]		[0.000912]		[0.000662]		[0.000920]
|Δ sales growth (ln)| (t-1)	-0.00401**	-0.00201	-0.00259	-0.000403	-0.00507***	-0.00351**	-0.00436**	-0.00296
	[0.00164]	[0.00162]	[0.00224]	[0.00224]	[0.00164]	[0.00161]	[0.00221]	[0.00220]
|Δ log(Tobin's Q)| (t-1)		-0.0101***		-0.0114***		-0.00900***		-0.0102***
		[0.00133]		[0.00185]		[0.00130]		[0.00184]
|Δ book leverage| (t-1)		0.000263		-0.00150		-0.00123		-0.00543*
		[0.000951]		[0.00275]		[0.00108]		[0.00294]
|Δ ROA| (t-1)		-0.000233		-0.00777***		0.000588		-0.00287*
		[0.000625]		[0.00171]		[0.000570]		[0.00148]
Acq Ind×Tar Ind×Year FE	√	√	√	√	√	√	√	√
Observations	8,068,691	8,068,691	6,066,320	6,066,320	8,068,691	8,068,691	6,066,320	6,066,320
R-squared	0.011	0.011	0.014	0.014	0.011	0.011	0.014	0.014

Notes: Standard errors are clustered at acquirer industry - target industry - year level. Dependent variable is an indicator = 1 if a merger deal is announced and = 0 otherwise. |Δ PTF| is the absolute value of the difference of PTF of the acquirer and the target. PTF Similarity measures the pairwise similarity of firm-level technology frontier. |Δ PTF Breadth| captures the absolute value of the difference of technology frontier breadth of the acquirer and the target. Competition is drawn from Hoberg and Phillips (2016) to gauge pairwise product competition between firms. Control variables include the absolute value of the difference of total asset, cash holding, sales growth, Tobin's Q, book leverage, and ROA, of the acquirer and the target. All independent variables except for PTF Similarity and Competition are demeaned by industry average in columns (5)-(8). Acquirer industry - target industry - year fixed effects are included in all specifications. Industry is defined at 4-digit SIC level.

Table A.2: PTF and Determinants of Mergers

(d) PTF, Tobin's Q and Mergers

	Dependent Variable: Indicator of Merger Deal							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES					Industry Average Adjusted Measure			
|Δ PTF| (t-1)	0.0290***	0.0236***	0.0276***	0.0217***	0.0344***	0.0293***	0.0351***	0.0297***
	[0.00679]	[0.00677]	[0.00827]	[0.00826]	[0.00695]	[0.00693]	[0.00847]	[0.00846]
PTF Similarity (t-1)	0.00259***	0.00256***	0.00218***	0.00219***	0.00261***	0.00259***	0.00219***	0.00221***
	[0.000416]	[0.000411]	[0.000383]	[0.000384]	[0.000418]	[0.000414]	[0.000384]	[0.000385]
|Δ PTF Breadth| (t-1)		0.753***		0.757***		0.744***		0.685***
		[0.208]		[0.268]		[0.206]		[0.261]
Competition (HP, t-1)			0.00218***	0.00213***			0.00219***	0.00216***
			[0.000297]	[0.000291]			[0.000298]	[0.000295]
|Δ PTF| (t-1) × |Δ log(Tobin's Q)| (t-1)	-2.920***	-2.698***	-3.002***	-2.902***	-2.258***	-2.015***	-1.577	-1.477
	[0.613]	[0.614]	[0.961]	[0.961]	[0.684]	[0.688]	[1.141]	[1.140]
|Δ log(total asset)| (t-1)		-0.00160***		0.000113		-0.00135***		7.88e-05
		[0.000368]		[0.000455]		[0.000370]		[0.000456]
|Δ cash| (t-1)		-0.00437***		-0.00435***		-0.00255***		-0.00185**
		[0.000678]		[0.000913]		[0.000663]		[0.000920]
|Δ sales growth (ln)| (t-1)		-0.00615***		-0.00494***		-0.00610***		-0.00526***
		[0.00123]		[0.00162]		[0.00120]		[0.00161]
|Δ log(Tobin's Q)| (t-1)	-0.00945***	-0.00740***	-0.0106***	-0.00847***	-0.00862***	-0.00707***	-0.00996***	-0.00876***
	[0.00149]	[0.00145]	[0.00217]	[0.00203]	[0.00149]	[0.00149]	[0.00225]	[0.00219]
|Δ book leverage| (t-1)		0.000288		-0.00157		-0.00121		-0.00546*
		[0.000942]		[0.00275]		[0.00108]		[0.00294]
|Δ ROA| (t-1)		-0.000305		-0.00785***		0.000544		-0.00290*
		[0.000614]		[0.00171]		[0.000562]		[0.00148]
Acq Ind×Tar Ind×Year FE	√	√	√	√	√	√	√	√
Observations	8,068,691	8,068,691	6,066,320	6,066,320	8,068,691	8,068,691	6,066,320	6,066,320
R-squared	0.011	0.011	0.014	0.014	0.011	0.011	0.014	0.014

Notes: Standard errors are clustered at acquirer industry - target industry - year level. Dependent variable is an indicator = 1 if a merger deal is announced and = 0 otherwise. |Δ PTF| is the absolute value of the difference of PTF of the acquirer and the target. PTF Similarity measures the pairwise similarity of firm-level technology frontier. |Δ PTF Breadth| captures the absolute value of the difference of technology frontier breadth of the acquirer and the target. Competition is drawn from Hoberg and Phillips (2016) to gauge pairwise product competition between firms. Control variables include the absolute value of the difference of total asset, cash holding, sales growth, Tobin's Q, book leverage, and ROA, of the acquirer and the target. All independent variables except for PTF Similarity and Competition are demeaned by industry average in columns (5)-(8). Acquirer industry - target industry - year fixed effects are included in all specifications. Industry is defined at 4-digit SIC level.

Table A.2: PTF and Determinants of Mergers

(e) PTF, ROA and Mergers

	Dependent Variable: Indicator of Merger Deal							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES					Industry Average Adjusted Measure			
|Δ PTF| (t-1)	0.0296***	0.0245***	0.0245***	0.0193**	0.0352***	0.0299***	0.0316***	0.0266***
	[0.00653]	[0.00655]	[0.00793]	[0.00797]	[0.00680]	[0.00681]	[0.00836]	[0.00839]
PTF Similarity (t-1)	0.00260***	0.00254***	0.00218***	0.00219***	0.00262***	0.00258***	0.00220***	0.00221***
	[0.000417]	[0.000408]	[0.000382]	[0.000382]	[0.000418]	[0.000412]	[0.000385]	[0.000384]
|Δ PTF Breadth| (t-1)		0.759***		0.753***		0.746***		0.676***
		[0.209]		[0.268]		[0.206]		[0.261]
Competition (HP, t-1)			0.00218***	0.00213***			0.00219***	0.00216***
			[0.000298]	[0.000292]			[0.000298]	[0.000295]
|Δ PTF| (t-1) × |Δ ROA| (t-1)	-1.497***	-1.367***	-5.194***	-4.942***	-1.394***	-1.295***	-4.628***	-4.446***
	[0.305]	[0.299]	[1.270]	[1.261]	[0.310]	[0.311]	[1.391]	[1.389]
|Δ log(total asset)| (t-1)		-0.00161***		0.000119		-0.00136***		8.82e-05
		[0.000369]		[0.000455]		[0.000370]		[0.000456]
|Δ cash| (t-1)		-0.00439***		-0.00437***		-0.00256***		-0.00185**
		[0.000678]		[0.000912]		[0.000662]		[0.000920]
|Δ sales growth (ln)| (t-1)		-0.00614***		-0.00491***		-0.00609***		-0.00523***
		[0.00123]		[0.00163]		[0.00120]		[0.00161]
|Δ log(Tobin's Q)| (t-1)		-0.0101***		-0.0114***		-0.00902***		-0.0102***
		[0.00133]		[0.00185]		[0.00130]		[0.00184]
|Δ book leverage| (t-1)		0.000254		-0.00162		-0.00124		-0.00550*
		[0.000943]		[0.00275]		[0.00107]		[0.00294]
|Δ ROA| (t-1)	-0.00301***	0.00105*	-0.00679***	-0.00313	-0.00210***	0.00178***	-0.00225	0.00117
	[0.000682]	[0.000629]	[0.00233]	[0.00221]	[0.000579]	[0.000608]	[0.00221]	[0.00222]
Acq Ind×Tar Ind×Year FE	√	√	√	√	√	√	√	√
Observations	8,068,691	8,068,691	6,066,320	6,066,320	8,068,691	8,068,691	6,066,320	6,066,320
R-squared	0.011	0.011	0.014	0.014	0.011	0.011	0.014	0.014

Notes: Standard errors are clustered at acquirer industry - target industry - year level. Dependent variable is an indicator = 1 if a merger deal is announced and = 0 otherwise. |Δ PTF| is the absolute value of the difference of PTF of the acquirer and the target. PTF Similarity measures the pairwise similarity of firm-level technology frontier. |Δ PTF Breadth| captures the absolute value of the difference of technology frontier breadth of the acquirer and the target. Competition is drawn from Hoberg and Phillips (2016) to gauge pairwise product competition between firms. Control variables include the absolute value of the difference of total asset, cash holding, sales growth, Tobin's Q, book leverage, and ROA, of the acquirer and the target. All independent variables except for PTF Similarity and Competition are demeaned by industry average in columns (5)-(8). Acquirer industry - target industry - year fixed effects are included in all specifications. Industry is defined at 4-digit SIC level.

APPENDIX B

Appendix to Chapter 2

B.1 SEC EDGAR Data Parsing

We download all the raw text of annual and quarterly reports[1] from 1994 to 2016 through links provided by EDGAR that are active as of January 2018. There are 1,000,313 documents and they are cleaned using Python 3.5.4 in the following steps (similar to Loughren and McDonald[2]):

1. Rewrite the entire text into lower case.

2. Remove all built-in graphic, zip, excel, pdf, xml documents indicated by their tags (e.g. all characters between "<type>graphic" and "</document>" are removed).

3. Extract "conformed period of report" for each document.[3]

4. Remove titles of each sub section of the document.

5. Remove the header and footer of the documents (e.g. all characters in front of "</sec-header>").

[1]Those include 10-K, 10-K405, 10-KSB, 10-Q, 10-QSB and their amendments. Some 10-K reports such as transition reports 10-KT are excluded from the sample.

[2]https://www3.nd.edu/ mcdonald/Word_Lists.html

[3]Conformed period of report is usually coded in a standard way in the report. We hand checked the documents whose "conformed period of report" is irregularly coded in the text.

6. Remove phrase “table of contents”. The phrase might show up at the end of each page to link the reader back to the table of contents.

7. Remove all XBRL (eXtensible Business Reporting Language) which is provided by the companies for machine reading.

8. Remove HTML entities, carriages and non ASCII encoded characters.

9. Remove HTML tags.

10. Remove punctuation.

11. Create variable to indicate whether or not the document contains “Item 1A. Risk Factors” section by scanning the text and locate the position of the phrase “item 1a risk factors”.

12. Remove tables that contain a significant amount of digit. The reason we remove tables is because tables are usually non textual such as balance sheet. However, some documents use tables to report everything including descriptive contents. Therefore we decide to keep the tables with certain amount of textual information, specifically the tables whose ratio of digits out of all characters is below that of the entire document.[4]

13. Remove digits.

14. Count the frequency of “uncertain”, and a dictionary of “uncertain” related words obtained from combining synonyms of “uncertain”, “uncertainty”, “risk”, “risky” and their derivatives[5] and word list from Hassen et al. (2017)

15. Count number of words, number of distinct words, number of words when stop words are removed[6].

---

[4]There are other cutoffs used in the literature, but we believe those numbers are too big for our need. We also keep a version that removes all the tables since only a small fraction of companies use table to report textual descriptive contents in some of their reports, and our results are robust to this version of measure.

[5]We use 2007 version of Oxford dictionary and thesaurus as our reference.

[6]List of stop words are from python package NLTK PorterStemmer. The list: 'a', 'about', 'above', 'after', 'again', 'against', 'ain', 'all', 'am', 'an', 'and', 'any', 'are', 'aren', 'as', 'at', 'be', 'because', 'been', 'before', 'being', 'below', 'between', 'both', 'but', 'by', 'can', 'couldn', 'd', 'did', 'didn', 'do', 'does', 'doesn', 'doing', 'don', 'down', 'during', 'each', 'few', 'for', 'from', 'further', 'had', 'hadn', 'has', 'hasn', 'have', 'haven', 'having', 'he', 'her', 'here', 'hers', 'herself', 'him', 'himself', 'his', 'how', 'i', 'if',

16. Stem the whole document and recount and repeat step 15.

With the information collected from the cleaned text we can calculate some useful characteristics of each filing document, such as average frequency of each word, ratio of number of distinct words in stemmed and unstemmed documents, etc. We also pick up some policy changes that have potential impacts on our measure. We count "uncertain tax positions" frequency as it is a proper noun on tax issues that firms start reporting in mid 2000s complying accounting rule FIN 48, which requires publicly traded entities to disclose income tax risks. Another policy change is the requirement of risk factors disclosure. SEC mandate firms to report RISK FACTORS section (usually in item 1A) in their annual reports from fiscal year ending in 2005. Most quarterly reports comply with the requirement as well. We are not decomposing which kind of uncertainty the firms talk about, so that we will not be eliminating any related word. The summary statistics on Edgar parsing results are reported in Table B1.

B.2 Aggregate Time-series Variables

We obtained aggregate data for use as outcomes and controls via the St. Louis Federal Reserve's FRED portal. We describe each outcome and whether it was aggregated to the quarterly level.

Real Gross Private Domestic Investment (GPDI): We refer to this simply as real gross investment. Quarterly and seasonally adjusted.

Real Gross Domestic Product (GDP): Quarterly and seasonally adjusted.

'in', 'into', 'is', 'isn', 'it', 'its', 'itself', 'just', 'll', 'm', 'ma', 'me', 'mightn', 'more', 'most', 'mustn', 'my', 'myself', 'needn', 'no', 'nor', 'not', 'now', 'o', 'of', 'off', 'on', 'once', 'only', 'or', 'other', 'our', 'ours', 'ourselves', 'out', 'over', 'own', 're', 's', 'same', 'shan', 'she', 'should', 'shouldn', 'so', 'some', 'such', 't', 'than', 'that', 'the', 'their', 'theirs', 'them', 'themselves', 'then', 'there', 'these', 'they', 'this', 'those', 'through', 'to', 'too', 'under', 'until', 'up', 've', 'very', 'was', 'wasn', 'we', 'were', 'weren', 'what', 'when', 'where', 'which', 'while', 'who', 'whom', 'why', 'will', 'with', 'won', 'wouldn', 'y', 'you', 'your', 'yours', 'yourself', 'yourselves'.

Ratio of GPDI to GDP: The ratio of nominal gross investment to GDP. Quarterly, not seasonally adjusted.

CBOE Volatility Index (VIX): This is an index of volatility measured by the underlying options, when available, on the S&P 500 index components. We take the quarterly mean of this measure to match it with our CRUX uncertainty index. This index is commonly used as a second moment uncertainty shock (cf. Bloom, 2009).

S&P 500 Index: We use this as a first moment control in our aggregate time series regressions. We aggregate daily closing index values to the quarterly level by taking a simple mean.

B.3 COMPUSTAT, CRSP and Option Metrics

This section provides detailed information on how COMPUSTAT firm level data, realized volatility measure and implied volatility is created. We download COMPUSTAT-Capital IQ North America Fundamentals Annual data from WRDS. In order to match with SEC EDGAR data (CRUX), all missing CIKs are removed. We replace missing fiscal year by COMPUSTAT defined variable datayear if data date is in the second half of the year or by datayear - 1 if data date is in the first half of the year. We take the maximum value of the variables of interest if there are duplicate observations for firms within the same year. Negative sales (*sale*) and capital investment (*capx*) are dropped. Tobin's Q is calculated by $(csho \times prcc\_f + at - ceq)/at$ as described in the main text. Missing capital investment (*capx*) is interpolated by the average of capital investment of the preceding and following years if neither of the two is missing. We start calculating capital stock at the year total property, plant and equipment (*ppent*) is first observed and set the first $ppent = K_{i0}$. Then capital stock for each year is calculated by perpetual inventory formula $K_{it} = PPI_t((1-r)K_{it-1} + capx_{it-1})$. Capital investment is taken at $t-1$ as investment takes time. Construction of sales growth, taking logs, DHS and lags are straightforward. The empirical analysis is on the matched COMPUSTAT-EDGAR data where missings are dropped.

We download firm level stock return data from CRSP to construct realized volatility. The CRSP U.S. Stock database contains end-of-day and end-of-month prices on primary listings from major stock exchange markets, including NYSE, NASDAQ, etc. We calculate the annual volatility of monthly holding period return (RET) as realized volatility of the firm within the year. Then we take the lag and match realized volatility data from CRSP with COMPUSTAT firm level data to evaluate the effect of realized volatility on corporate behavior. Our result is robust if we use daily holding period return to calculate realized volatility.

Implied volatility is downloaded from Option Metrics Standardized Options. The construction of annual implied volatility follows Barrero, Bloom and Wright (2017). First, we take the average of firm-day implied volatility across calls and puts. Then we compute the annual measure of implied volatility by taking the Euclidean mean of daily implied volatility within the year: $\text{impl vol}_{i,year} = \sqrt{\frac{1}{\#\ of\ days} \sum_{\#\ of\ days} \text{impl vol}_{i,day}}$. Similarly, we take the lag and match implied volatility data from Option Metrics with COMPUSTAT firm level data to evaluate the effect of implied volatility on corporate behavior.

B.4 Appendix Tables

Table B.1: Summary Statistics - Parsed SEC Edgar Documents

VARIABLES	N	mean	median	sd	min	max
		Original Document				
Total Word Count	1,000,313	19,247	10,329	35,782	0	9,597,104
Unique Word Count	1,000,313	1,690	1,392	1,850	0	350,324
Total Word Count (without stopwords)	1,000,313	11,357	6,116	23,385	0	7,334,912
Unique Word Count (without stopwords)	1,000,313	1,602	1,301	1,841	0	350,197
Total Stopword Count	1,000,313	7,890	4,196	13,127	0	2,262,192
Unique Stopword Count	1,000,313	88.11	91	18.11	0	137
"Uncertain" count	1,000,313	4.404	3	5.698	0	133
		Stemmed Document				
Unique Word Count	1,000,313	1,220	1,031	1,615	0	337,215
Unique Word Count (without stopwords)	1,000,313	1,137	944	1,610	0	337,139
Unique Stopword Count	1,000,313	83.22	86	15.75	0	122

Table B.2: Effects of Uncertainty on Corporate Investment Rate ($\log(\frac{I_{it}}{K_{it}})$, Compustat - *ppegt*)

	Dependent Variable: Log Investment Rate (I/K, t)					
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
CRUX (t)	-1.697***	-1.475***	-1.429***	-1.430***	-1.411***	-1.410***
	[0.188]	[0.187]	[0.185]	[0.185]	[0.181]	[0.181]
Sales Growth (Δln t-1)			0.213***	0.211***	0.183***	0.183***
			[0.0168]	[0.0174]	[0.0161]	[0.0167]
CRUX (t) × Sales Growth (Δln t-1)			-0.574**	-0.567**	-0.518**	-0.521**
			[0.259]	[0.259]	[0.252]	[0.253]
Sales Growth (Δln t-1) sqaured				0.00178		-0.000592
				[0.00330]		[0.00318]
Log Tobin's Q (t-1)					0.346***	0.346***
					[0.0115]	[0.0115]
Firm Fixed Effects	√	√	√	√	√	√
Year Fixed Effects	√					
Industry × Year Fixed Effects		√	√	√	√	√
Observations	92,718	91,246	91,246	91,246	91,246	91,246
R-squared	0.082	0.491	0.499	0.499	0.517	0.517
Number of Firms	11,533	10,128	10,128	10,128	10,128	10,128

Notes: Standard errors are clustered at firm level. Dependent variable is calculated as log(I(t)/K(t)) where I is capital expenditure (capx) and K is firm's total gross property, plant and equipment (ppegt). Firm idiosyncratic uncertainty measure CRUX is at time t, which as mentioned in the text captures firm uncertainty at time t-1. Sales growth is calculated as log(Sales(t-1)) - log(Sales(t-2)) and demeaned by sample mean. Squared sales growth is calculated using demeand sales growth. Tobin's Q is taken as log average Q at time t-1. All of the regression specifications include firm FE. Column (1) includes Year FE, Column (2)-(6) include industry × year FE, where industry is at 3-digit NAICS code level. Column (2)-(6) loses some observations and firms compared with Column (1) due to singleton observations. Time ranges from 1994 to 2016.

*** p<0.01, ** p<0.05, * p<0.1

Table B.3: Effects of Uncertainty on Corporate Investment Rate ($\frac{I_{it}}{K_{it}}$, Compustat - *ppegt*)

	Dependent Variable: Investment Rate (I/K, t)					
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
CRUX (t)	-0.158***	-0.139***	-0.135***	-0.136***	-0.134***	-0.135***
	[0.0228]	[0.0228]	[0.0227]	[0.0227]	[0.0222]	[0.0222]
Sales Growth (Δln t-1)			0.0271***	0.0261***	0.0237***	0.0229***
			[0.00238]	[0.00238]	[0.00228]	[0.00229]
CRUX (t) × Sales Growth (Δln t-1)			-0.146***	-0.142***	-0.136***	-0.132***
			[0.0368]	[0.0366]	[0.0362]	[0.0360]
Sales Growth (Δln t-1) sqaured				0.00159***		0.00131***
				[0.000437]		[0.000423]
Log Tobin's Q (t-1)					0.0449***	0.0448***
					[0.00159]	[0.00159]
Firm Fixed Effects	√	√	√	√	√	√
Year Fixed Effects	√					
Industry × Year Fixed Effects		√	√	√	√	√
Observations	94,848	93,397	93,397	93,397	93,397	93,397
R-squared	0.065	0.441	0.448	0.448	0.469	0.469
Number of Firms	11,693	10,305	10,305	10,305	10,305	10,305

Notes: Standard errors are clustered at firm level. Dependent variable is calculated as I(t)/K(t) where I is capital expenditure (capx) and K is firm's total gross property, plant and equipment (ppegt). The dependent variable is winsorized at 1% and 99% level by year. Firm idiosyncratic uncertainty measure CRUX is at time t, which as mentioned in the text captures firm uncertainty at time t-1. Sales growth is calculated as log(Sales(t-1)) - log(Sales(t-2)) and demeaned by sample mean. Squared sales growth is calculated using demeand sales growth. Tobin's Q is taken as log average Q at time t-1. All of the regression specifications include firm FE. Column (1) includes Year FE, Column (2)-(6) include industry × year FE, where industry is at 3-digit NAICS code level. Column (2)-(6) loses some observations and firms compared with Column (1) due to singleton observations. Time ranges from 1994 to 2016.

*** p<0.01, ** p<0.05, * p<0.1

Table B.4: Effects of Uncertainty on Corporate Investment Rate ($\log(\frac{I_{it}}{K_{it}})$, Compustat - *ppent*)

	Dependent Variable: Log Investment Rate (I/K, t)					
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
CRUX (t)	-0.832***	-0.812***	-0.786***	-0.785***	-0.768***	-0.766***
	[0.157]	[0.159]	[0.158]	[0.158]	[0.155]	[0.155]
Sales Growth (Δln t-1)			0.138***	0.140***	0.111***	0.115***
			[0.0141]	[0.0147]	[0.0136]	[0.0142]
CRUX (t) × Sales Growth (Δln t-1)			-0.548**	-0.558**	-0.472**	-0.491**
			[0.218]	[0.220]	[0.214]	[0.217]
Sales Growth (Δln t-1) sqaured				-0.00222		-0.00429*
				[0.00265]		[0.00255]
Log Tobin's Q (t-1)					0.299***	0.299***
					[0.00976]	[0.00976]
Firm Fixed Effects	√	√	√	√	√	√
Year Fixed Effects	√					
Industry × Year Fixed Effects		√	√	√	√	√
Observations	99,936	98,361	98,361	98,361	98,361	98,361
R-squared	0.040	0.532	0.535	0.535	0.549	0.550
Number of Firms	12,518	11,008	11,008	11,008	11,008	11,008

Notes: Standard errors are clustered at firm level. Dependent variable is calculated as log(I(t)/K(t)) where I is capital expenditure (capx) and K is firm's total net property, plant and equipment (ppent). Firm idiosyncratic uncertainty measure CRUX is at time t, which as mentioned in the text captures firm uncertainty at time t-1. Sales growth is calculated as log(Sales(t-1)) - log(Sales(t-2)) and demeaned by sample mean. Squared sales growth is calculated using demeand sales growth. Tobin's Q is taken as log average Q at time t-1. All of the regression specifications include firm FE. Column (1) includes Year FE, Column (2)-(6) include industry × year FE, where industry is at 3-digit NAICS code level. Column (2)-(6) loses some observations and firms compared with Column (1) due to singleton observations. Time ranges from 1994 to 2016.
*** p<0.01, ** p<0.05, * p<0.1

Table B.5: Effects of Uncertainty on Corporate Investment Rate ($\frac{I_{it}}{K_{it}}$, Compustat - *ppent*)

	Dependent Variable: Investment Rate (I/K, t)					
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
CRUX (t)	-0.179***	-0.167***	-0.163***	-0.163***	-0.158***	-0.158***
	[0.0350]	[0.0358]	[0.0357]	[0.0356]	[0.0351]	[0.0351]
Sales Growth (Δln t-1)			0.0319***	0.0314***	0.0267***	0.0265***
			[0.00334]	[0.00338]	[0.00322]	[0.00328]
CRUX (t) × Sales Growth (Δln t-1)			-0.190***	-0.187***	-0.172***	-0.170***
			[0.0545]	[0.0545]	[0.0536]	[0.0537]
Sales Growth (Δln t-1) sqaured				0.000876		0.000422
				[0.000607]		[0.000582]
Log Tobin's Q (t-1)					0.0671***	0.0670***
					[0.00227]	[0.00227]
Firm Fixed Effects	√	√	√	√	√	√
Year Fixed Effects	√					
Industry × Year Fixed Effects		√	√	√	√	√
Observations	102,224	100,659	100,659	100,659	100,659	100,659
R-squared	0.033	0.499	0.502	0.503	0.518	0.518
Number of Firms	12,698	11,198	11,198	11,198	11,198	11,198

Notes: Standard errors are clustered at firm level. Dependent variable is calculated as I(t)/K(t) where I is capital expenditure (capx) and K is firm's total net property, plant and equipment (ppent). The dependent variable is winsorized at 1% and 99% level by year. Firm idiosyncratic uncertainty measure CRUX is at time t, which as mentioned in the text captures firm uncertainty at time t-1. Sales growth is calculated as log(Sales(t-1)) - log(Sales(t-2)) and demeaned by sample mean. Squared sales growth is calculated using demeand sales growth. Tobin's Q is taken as log average Q at time t-1. All of the regression specifications include firm FE. Column (1) includes Year FE, Column (2)-(6) include industry × year FE, where industry is at 3-digit NAICS code level. Column (2)-(6) loses some observations and firms compared with Column (1) due to singleton observations. Time ranges from 1994 to 2016.

*** p<0.01, ** p<0.05, * p<0.1

Table B.6: Effects of Uncertainty on Corporate Investment Rate (Δln, Compustat - Segment NAICS)

	Dependent Variable: Log Change in Capital Stock K (Δln, t)					
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
CRUX (t)	-0.350***	-0.300***	-0.295***	-0.296***	-0.290***	-0.291***
	[0.0430]	[0.0425]	[0.0421]	[0.0420]	[0.0409]	[0.0409]
Sales Growth (Δln t-1)			0.0763***	0.0744***	0.0700***	0.0686***
			[0.00544]	[0.00531]	[0.00519]	[0.00509]
CRUX (t) × Sales Growth (Δln t-1)			-0.486***	-0.471***	-0.451***	-0.440***
			[0.0696]	[0.0698]	[0.0679]	[0.0681]
Sales Growth (Δln t-1) sqaured				0.00293***		0.00230***
				[0.000814]		[0.000755]
Log Tobin's Q (t-1)					0.0890***	0.0887***
					[0.00330]	[0.00329]
Firm Fixed Effects	√	√	√	√	√	√
Year Fixed Effects	√					
Industry × Year Fixed Effects		√	√	√	√	√
Observations	95,222	93,717	93,717	93,717	93,717	93,717
R-squared	0.081	0.394	0.407	0.407	0.429	0.429
Number of Firms	11,864	10,447	10,447	10,447	10,447	10,447

Notes: Standard errors are clustered at firm level. Dependent variable is calculated as log(K(t)) - log(K(t-1)) where K is firm's total capital stock. Firm idiosyncratic uncertainty measure CRUX is at time t, which as mentioned in the text captures firm uncertainty at time t-1. Sales growth is calculated as log(Sales(t-1)) - log(Sales(t-2)) and demeaned by sample mean. Squared sales growth is calculated using demeand sales growth. Tobin's Q is taken as log average Q at time t-1. All of the regression specifications include firm FE. Column (1) includes Year FE, Column (2)-(6) include industry × year FE, where industry is at 3-digit NAICS code level adjusted by segment sales. Column (2)-(6) loses some observations and firms compared with Column (1) due to singleton observations. Time ranges from 1994 to 2016.

*** p<0.01, ** p<0.05, * p<0.1

Table B.7: Effects of High vs Low Uncertainty on Corporate Investment Rate (Δln, Compustat)

	Dependent Variable: Log Change in Capital Stock K (Δln, t)					
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
High CRUX (t)	-0.0160***	-0.0140***	-0.0137***	-0.0138***	-0.0127***	-0.0128***
	[0.00241]	[0.00239]	[0.00236]	[0.00236]	[0.00232]	[0.00232]
Sales Growth (Δln t-1)			0.0751***	0.0738***	0.0690***	0.0680***
			[0.00538]	[0.00527]	[0.00510]	[0.00501]
High CRUX (t) × Sales Growth (Δln t-1)			-0.0368***	-0.0367***	-0.0341***	-0.0340***
			[0.00591]	[0.00592]	[0.00571]	[0.00571]
Sales Growth (Δln t-1) sqaured				0.00320***		0.00249***
				[0.000830]		[0.000771]
Log Tobin's Q (t-1)					0.0892***	0.0888***
					[0.00338]	[0.00337]
Firm Fixed Effects	√	√	√	√	√	√
Year Fixed Effects	√					
Industry × Year Fixed Effects		√	√	√	√	√
Observations	95,222	93,741	93,741	93,741	93,741	93,741
R-squared	0.081	0.391	0.404	0.405	0.427	0.427
Number of Firms	11,864	10,445	10,445	10,445	10,445	10,445

Notes: Standard errors are clustered at firm level. Dependent variable is calculated as log(K(t)) - log(K(t-1)) where K is firm's total capital stock. Firm idiosyncratic uncertainty measure CRUX is at time t, which as mentioned in the text captures firm uncertainty at time t-1 and we further define high CRUX as above median in the sample. Sales growth is calculated as log(Sales(t-1)) - log(Sales(t-2)) and demeaned by sample mean. Squared sales growth is calculated using demeand sales growth. Tobin's Q is taken as log average Q at time t-1. All of the regression specifications include firm FE. Column (1) includes Year FE, Column (2)-(6) include industry × year FE, where industry is at 3-digit NAICS code level. Column (2)-(6) loses some observations and firms compared with Column (1) due to singleton observations. Time ranges from 1994 to 2016.

*** p<0.01, ** p<0.05, * p<0.1

Table B.8: Effects of "Risk" on Corporate Investment Rate (Δln, Compustat)

	Dependent Variable: Log Change in Capital Stock K (Δln, t)					
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
RISK (t)	0.0645***	0.0353**	0.0277*	0.0287*	0.0254	0.0262
	[0.0171]	[0.0167]	[0.0166]	[0.0166]	[0.0163]	[0.0164]
Sales Growth (Δln t-1)			0.0594***	0.0587***	0.0569***	0.0563***
			[0.00601]	[0.00602]	[0.00569]	[0.00570]
RISK (t) × Sales Growth (Δln t-1)			-0.0532	-0.0573*	-0.0676**	-0.0708**
			[0.0339]	[0.0343]	[0.0327]	[0.0329]
Sales Growth (Δln t-1) sqaured				0.00327***		0.00256***
				[0.000839]		[0.000775]
Log Tobin's Q (t-1)					0.0898***	0.0894***
					[0.00338]	[0.00338]
Firm Fixed Effects	√	√	√	√	√	√
Year Fixed Effects	√					
Industry × Year Fixed Effects		√	√	√	√	√
Observations	95,222	93,741	93,741	93,741	93,741	93,741
R-squared	0.080	0.391	0.403	0.403	0.425	0.426
Number of Firms	11,864	10,445	10,445	10,445	10,445	10,445

Notes: Standard errors are clustered at firm level. Dependent variable is calculated as log(K(t)) - log(K(t-1)) where K is firm's total capital stock. RISK is constructed by taking word risk and its derivatives following the same method as uncertainty. Sales growth is calculated as log(Sales(t-1)) - log(Sales(t-2)) and demeaned by sample mean. Squared sales growth is calculated using demeand sales growth. Tobin's Q is taken as log average Q at time t-1. All of the regression specifications include firm FE. Column (1) includes Year FE, Column (2)-(6) include industry × year FE, where industry is at 3-digit NAICS code level. Column (2)-(6) loses some observations and firms compared with Column (1) due to singleton observations. Time ranges from 1994 to 2016.

*** p<0.01, ** p<0.05, * p<0.1

Table B.9: Effects of Realized Volatility vs CRUX on Corporate Investment Rate (Δln, Compustat)

	Dependent Variable: Log Change in Capital Stock K (Δln, t)					
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
CRUX (t)	-0.368***	-0.319***	-0.310***	-0.311***	-0.284***	-0.285***
	[0.0443]	[0.0441]	[0.0432]	[0.0431]	[0.0408]	[0.0408]
Sales Growth (Δln t-1)			0.0962***	0.0910***	0.0798***	0.0752***
			[0.00853]	[0.00833]	[0.00784]	[0.00766]
CRUX (t) × Sales Growth (Δln t-1)			-0.587***	-0.567***	-0.555***	-0.537***
			[0.102]	[0.100]	[0.0950]	[0.0934]
Real Vol (t-1)	0.0163	-0.00372	0.00209	0.000662	-0.0413***	-0.0425***
	[0.0132]	[0.0128]	[0.0135]	[0.0135]	[0.0127]	[0.0127]
Real Vol (t-1) × Sales Growth (Δln t-1)			0.00447	0.0143	0.00758	0.0165
			[0.0175]	[0.0172]	[0.0149]	[0.0149]
Sales Growth (Δln t-1) sqaured				0.00517***		0.00468***
				[0.00132]		[0.00118]
Log Tobin's Q (t-1)					0.128***	0.128***
					[0.00356]	[0.00355]
Firm Fixed Effects	√	√	√	√	√	√
Year Fixed Effects	√					
Industry × Year Fixed Effects		√	√	√	√	√
Observations	73,795	72,606	72,606	72,606	72,606	72,606
R-squared	0.097	0.422	0.439	0.440	0.477	0.478
Number of Firms	9,127	8,023	8,023	8,023	8,023	8,023

Notes: Standard errors are clustered at firm level. Dependent variable is calculated as log(K(t)) - log(K(t-1)) where K is firm's total capital stock. Firm idiosyncratic uncertainty measure CRUX is at time t, which as mentioned in the text captures firm uncertainty at time t-1. Realized Volatility is calculated as the standard deviation of firms' monthly stock returns at t-1. Sales growth is calculated as log(Sales(t-1)) - log(Sales(t-2)) and demeaned by sample mean. Squared sales growth is calculated using demeand sales growth. Tobin's Q is taken as log average Q at time t-1. All of the regression specifications include firm FE. Column (1) includes Year FE, Column (2)-(6) include industry × year FE, where industry is at 3-digit NAICS code level. Column (2)-(6) loses some observations and firms compared with Column (1) due to singleton observations. Time ranges from 1994 to 2016.
*** p<0.01, ** p<0.05, * p<0.1

Table B.10: Effects of Implied Volatility vs CRUX on Corporate Investment Rate (Δln, Compustat)

	Dependent Variable: Log Change in Capital Stock K (Δln, t)					
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
CRUX (t)	-0.399***	-0.373***	-0.359***	-0.359***	-0.317***	-0.316***
	[0.0588]	[0.0576]	[0.0563]	[0.0562]	[0.0541]	[0.0541]
Sales Growth (Δln t-1)			0.111***	0.111***	0.0895***	0.0884***
			[0.0113]	[0.0107]	[0.00971]	[0.00905]
CRUX (t) × Sales Growth (Δln t-1)			-0.789***	-0.788***	-0.703***	-0.691***
			[0.136]	[0.133]	[0.118]	[0.115]
Impl Vol (t-1)	-0.00105	-0.00162	-0.00222*	-0.00223*	-0.00181**	-0.00183**
	[0.00153]	[0.00139]	[0.00129]	[0.00129]	[0.000900]	[0.000912]
Impl Vol (t-1) × Sales Growth (Δln t-1)			0.000167	0.000168	0.000584	0.000591
			[0.00133]	[0.00133]	[0.000948]	[0.000951]
Sales Growth (Δln t-1) sqaured				0.000153		0.00113
				[0.00246]		[0.00214]
Log Tobin's Q (t-1)					0.121***	0.121***
					[0.00502]	[0.00501]
Firm Fixed Effects	√	√	√	√	√	√
Year Fixed Effects	√					
Industry × Year Fixed Effects		√	√	√	√	√
Observations	34,971	34,202	34,202	34,202	34,202	34,202
R-squared	0.122	0.498	0.517	0.517	0.558	0.558
Number of Firms	4,841	4,156	4,156	4,156	4,156	4,156

Notes: Standard errors are clustered at firm level. Dependent variable is calculated as log(K(t)) - log(K(t-1)) where K is firm's total capital stock. Firm idiosyncratic uncertainty measure CRUX is at time t, which as mentioned in the text captures firm uncertainty at time t-1. The implied volatility is calculated following Barrero, Bloom and Wright (2017) by year using 91 day duration daily implied volatility. Sales growth is calculated as log(Sales(t-1)) - log(Sales(t-2)) and demeaned by sample mean. Squared sales growth is calculated using demeand sales growth. Tobin's Q is taken as log average Q at time t-1. All of the regression specifications include firm FE. Column (1) includes Year FE, Column (2)-(6) include industry × year FE, where industry is at 3-digit NAICS code level. Column (2)-(6) loses some observations and firms compared with Column (1) due to singleton observations. Time ranges from 1997 to 2016.
*** p<0.01, ** p<0.05, * p<0.1

Table B.11: Effects of Risk Factors vs CRUX on Corporate Investment Rate (Δln, Compustat)

	Dependent Variable: Log Change in Capital Stock K (Δln, t)					
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
CRUX (t)	-0.197***	-0.230***	-0.223***	-0.225***	-0.244***	-0.245***
	[0.0639]	[0.0633]	[0.0632]	[0.0632]	[0.0614]	[0.0614]
Risk Factors (t)	0.0426***	0.0327***	0.0303***	0.0301***	0.0254***	0.0252***
	[0.00446]	[0.00446]	[0.00433]	[0.00433]	[0.00421]	[0.00421]
CRUX (t) × Risk Factors (t)	-0.259***	-0.141*	-0.140*	-0.139*	-0.101	-0.101
	[0.0734]	[0.0738]	[0.0734]	[0.0733]	[0.0715]	[0.0714]
CRUX (t) × Sales Growth (Δln t-1)			-0.499***	-0.466***	-0.473***	-0.448***
			[0.0847]	[0.0852]	[0.0836]	[0.0841]
CRUX (t) × Sales Growth (Δln t-1) × Risk Factors (t)			0.00676	-0.0173	0.0222	0.00378
			[0.0804]	[0.0809]	[0.0801]	[0.0805]
Sales Growth (Δln t-1)			0.0773***	0.0754***	0.0710***	0.0696***
			[0.00545]	[0.00532]	[0.00520]	[0.00510]
Sales Growth (Δln t-1) sqaured				0.00296***		0.00227***
				[0.000822]		[0.000762]
Log Tobin's Q (t-1)					0.0890***	0.0886***
					[0.00338]	[0.00337]
Firm Fixed Effects	√	√	√	√	√	√
Year Fixed Effects	√					
Industry × Year Fixed Effects		√	√	√	√	√
Observations	95,222	93,741	93,741	93,741	93,741	93,741
R-squared	0.083	0.392	0.405	0.406	0.427	0.428
Number of Firms	11,864	10,445	10,445	10,445	10,445	10,445

Notes: Standard errors are clustered at firm level. Dependent variable is calculated as log(K(t)) - log(K(t-1)) where K is firm's total capital stock. Firm idiosyncratic uncertainty measure CRUX is at time t, which as mentioned in the text captures firm uncertainty at time t-1. Risk Factors is an indicator equals to 1 if firm reports item 1A risk factors section during the year, and 0 otherwise. Sales growth is calculated as log(Sales(t-1)) - log(Sales(t-2)) and demeaned by sample mean. Squared sales growth is calculated using demeand sales growth. Tobin's Q is taken as log average Q at time t-1. All of the regression specifications include firm FE. Column (1) includes Year FE, Column (2)-(6) include industry × year FE, where industry is at 3-digit NAICS code level. Column (2)-(6) loses some observations and firms compared with Column (1) due to singleton observations. Time ranges from 1994 to 2016.

*** p<0.01, ** p<0.05, * p<0.1

Table B.12: Effects of Post 2006 Indicator vs CRUX on Corporate Investment Rate (Δln, Compustat)

	Dependent Variable: Log Change in Capital Stock K (Δln, t)					
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
CRUX (t)	-0.300***	-0.305***	-0.278***	-0.282***	-0.277***	-0.280***
	[0.0572]	[0.0573]	[0.0563]	[0.0562]	[0.0550]	[0.0550]
CRUX (t) × Post 2006	-0.110	-0.0100	-0.0641	-0.0581	-0.0568	-0.0523
	[0.0694]	[0.0701]	[0.0686]	[0.0686]	[0.0674]	[0.0674]
CRUX (t) × Sales Growth (Δln t-1)			-0.383***	-0.368***	-0.377***	-0.366***
			[0.0753]	[0.0760]	[0.0742]	[0.0747]
CRUX (t) × Sales Growth (Δln t-1) × Post 2006			-0.214***	-0.214***	-0.152*	-0.153*
			[0.0818]	[0.0829]	[0.0815]	[0.0822]
Sales Growth (Δln t-1)			0.0769***	0.0751***	0.0708***	0.0694***
			[0.00542]	[0.00529]	[0.00518]	[0.00508]
Sales Growth (Δln t-1) sqaured				0.00299***		0.00229***
				[0.000827]		[0.000767]
Log Tobin's Q (t-1)					0.0891***	0.0888***
					[0.00337]	[0.00337]
Firm Fixed Effects	√	√	√	√	√	√
Year Fixed Effects	√					
Industry × Year Fixed Effects		√	√	√	√	√
Observations	95,222	93,741	93,741	93,741	93,741	93,741
R-squared	0.081	0.391	0.405	0.405	0.427	0.427
Number of Firms	11,864	10,445	10,445	10,445	10,445	10,445

Notes: Standard errors are clustered at firm level. Dependent variable is calculated as log(K(t)) - log(K(t-1)) where K is firm's total capital stock. Firm idiosyncratic uncertainty measure CRUX is at time t, which as mentioned in the text captures firm uncertainty at time t-1. Post 2006 is an indicator equals to 1 if fiscal year is after 2006, and 0 otherwise. Sales growth is calculated as log(Sales(t-1)) - log(Sales(t-2)) and demeaned by sample mean. Squared sales growth is calculated using demeand sales growth. Tobin's Q is taken as log average Q at time t-1. All of the regression specifications include firm FE. Column (1) includes Year FE, Column (2)-(6) include industry × year FE, where industry is at 3-digit NAICS code level. Column (2)-(6) loses some observations and firms compared with Column (1) due to singleton observations. Time ranges from 1994 to 2016.

*** p<0.01, ** p<0.05, * p<0.1

Table B.13: Effect of Uncertainty on Manufacturing Establishment Level Investment Rate (Δln) and Investment Spike ($\frac{I_{it}}{K_{it-1}} \geq 20\%$) - IPS Weighted Regression

	Log Total Investment Rate (Δln, t)			Investment Spike (t)		
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
CRUX (t)	-0.0630*	-0.0608*		-0.0997*	-0.0972*	
	[0.0345]	[0.0346]		[0.0530]	[0.0528]	
TVS Growth (Δln t-1)	0.0447***	0.0424***	0.0437***	0.0400***	0.0370***	0.0369***
	[0.00665]	[0.00624]	[0.00615]	[0.00493]	[0.00474]	[0.00475]
CRUX (t) × TVS Growth (Δln t-1)	-0.311***	-0.294***	-0.336***	-0.274***	-0.251***	-0.272***
	[0.107]	[0.104]	[0.111]	[0.0892]	[0.0876]	[0.0918]
TVS Growth (Δln t-1) squared		0.00538***	0.00491***		0.00702***	0.00619***
		[0.00174]	[0.00176]		[0.00143]	[0.00135]
Log Tobin's Q (t-1)		0.0155***			0.0164***	
		[0.00318]			[0.00509]	
Firm Fixed Effects	√	√		√	√	
Firm × Year Fixed Effects			√			√
Industry × Year Fixed Effects	√	√	√	√	√	√
R-squared	0.071	0.073	0.215	0.071	0.073	0.215

Notes: Standard errors are clustered at firm level. Dependent variable in columns (1)-(3) is calculated as log(K(t)) - log(K(t-1)) where K is establishment's total capital stock. Dependent variable in columns (4)-(6) is an indicator = 1 if (K(t)- K(t-1))/K(t-1) ≥ 20% and = 0 if otherwise, where K is establishment's total capital stock. Firm level idiosyncratic uncertainty measure CRUX is at time t, which as mentioned in the text captures firm uncertainty at time t-1. Establishment level TVS (total value of shipment) growth is calculated as log(TVS(t-1)) - log(TVS(t-2)) and demeaned by sample mean. Squared TVS growth is calculated using demeand TVS growth. Firm level Tobin's Q is taken as log average Q at time t-1. Columns (1)-(2) and (4)-(5) include firm FE. Columns (3) and (6) include firm × year FE, which absorbs CRUX (t) and Log Tobin's Q (t-1). All columns include industry × year FE, where industry is at 3-digit NAICS code level. Regressions are weighted by inverse propensity score constructed by fitting logit specifications.Time ranges from 1998 to 2014. Number of observations is 133000 and number of firms is 2000, both rounded to the nearest thousands.
*** p<0.01, ** p<0.05, * p<0.1

Table B.14: Effect of Uncertainty on Equipment vs. Structure Investment Rate (Δln) - IPS Weighted Regression

	Log Equipment Investment Rate (Δln, t)			Log Structure Investment Rate (Δln, t)		
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
CRUX (t)	-0.0703	-0.0678		-0.0257	-0.0244	
	[0.0430]	[0.0432]		[0.0253]	[0.0251]	
TVS Growth (Δln t-1)	0.0539***	0.0514***	0.0534***	0.0249***	0.0237***	0.0241***
	[0.00781]	[0.00734]	[0.00716]	[0.00475]	[0.00447]	[0.00471]
CRUX (t) × TVS Growth (Δln t-1)	-0.348***	-0.330***	-0.387***	-0.201***	-0.193***	-0.194**
	[0.129]	[0.126]	[0.132]	[0.0762]	[0.0734]	[0.0835]
TVS Growth (Δln t-1) squared		0.00566***	0.00504***		0.00264*	0.00237
		[0.00192]	[0.00192]		[0.00143]	[0.00150]
Log Tobin's Q (t-1)		0.0198***			0.0114***	
		[0.00416]			[0.00218]	
Firm Fixed Effects	√	√		√	√	
Firm × Year Fixed Effects			√			√
Industry × Year Fixed Effects	√	√	√	√	√	√
R-squared	0.095	0.097	0.247	0.048	0.049	0.188

Notes: Standard errors are clustered at firm level. Regressions are weighted by inverse propensity score constructed by fitting logit specifications. Dependent variable is calculated as log(K(t)) - log(K(t-1)) where K is establishment's total capital stock in equipment or structure. Firm level idiosyncratic uncertainty measure CRUX is at time t, which as mentioned in the text captures firm uncertainty at time t-1. Establishment level TVS (total value of shipment) growth is calculated as log(TVS(t-1)) - log(TVS(t-2)) and demeaned by sample mean. Squared TVS growth is calculated using demeand TVS growth. Firm level Tobin's Q is taken as log average Q at time t-1. Columns (1)-(2) and (4)-(5) include firm FE. Columns (3) and (6) include firm × year FE, which absorbs CRUX (t) and Log Tobin's Q (t-1). All columns include industry × year FE, where industry is at 3-digit NAICS code level. Time ranges from 1998 to 2014. Number of observations is 133000 and number of firms is 2000, both rounded to the nearest thousands.
*** p<0.01, ** p<0.05, * p<0.1

Table B.15: Effect of Industry Uncertainty on Equipment vs. Structure Investment Rate (Δln) - IPS Weighted Industry CRUX

	Log Equipment Investment Rate (Δln, t)			Log Structure Investment Rate (Δln, t)		
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
Industry CRUX (t)	-0.240***	-0.243***		-0.0321	-0.0331	
	[0.0752]	[0.0751]		[0.0345]	[0.0345]	
TVS Growth (Δln t-1)	0.0300***	0.0292***	0.0316***	0.00973***	0.00949***	0.0117***
	[0.00197]	[0.00192]	[0.00266]	[0.00121]	[0.00121]	[0.00174]
Industry CRUX (t) × TVS Growth (Δln t-1)	-0.507***	-0.458***	-0.546***	-0.157**	-0.141**	-0.204*
	[0.115]	[0.112]	[0.166]	[0.0728]	[0.0698]	[0.108]
TVS Growth (Δln t-1) squared		0.00424***	0.00462***		0.00132**	0.00153*
		[0.000938]	[0.00124]		[0.000667]	[0.000854]
Firm Fixed Effects	√	√		√	√	
Firm × Year Fixed Effects			√			√
Industry × Year Fixed Effects	√	√	√	√	√	√
R-squared	0.132	0.133	0.436	0.092	0.092	0.375

Notes: Standard errors are clustered at industry (4-digit NAICS) × year level. Dependent variable is calculated as log(K(t)) - log(K(t-1)) where K is establishment's total capital stock in equipment or structure. Industry level uncertainty measure Inudustry CRUX (t) is calculated by taking inverse propensity score weighted average of firm FE demeaned CRUX measure of all establishments within the same industry (4-digit NAICS code). The inverse propensity score is constructed by fitting logit specifications. Establishment level TVS (total value of shipment) growth is calculated as log(TVS(t-1)) - log(TVS(t-2)) and demeaned by sample mean. Squared TVS growth is calculated using demeand TVS growth. Columns (1)-(2) and (4)-(5) include firm FE. Columns (3) and (6) include firm × year FE, which absorbs Industry CRUX (t). All columns include industry × year FE, where industry is at 3-digit NAICS code level. Time ranges from 1998 to 2014. Number of observations is 472000 and number of firms is 21000, both rounded to the nearest thousands.

*** p<0.01, ** p<0.05, * p<0.1

APPENDIX C

Appendix to Chapter 3

C.1 Data Appendix

C.1.1 BLS Business Employment Dynamics

We download the entire BLS Business Dynamics data set from BLS website. Then we restrict our sample of analysis to all industries (no industry specific variation), all size classes, and all states. We keep the seasonally adjusted quarterly rate on employment/establishment change on different margins. There are 12 dependent variables:

Percent of gross job gains for the total private sector in the U.S. (as a percentage of total employment) — Gross Employment Gain;

Percent of employment gained from expansions for the total private sector in the U.S. (as a percentage of total employment) — Employment Gain from Expansion;

Percent of employment gained from openings for the total private sector in the U.S. (as a percentage of total employment) — Employment Gain from Opening;

Percent of gross job losses for the total private sector in the U.S. (as a percentage of total employment) — Gross Employment Loss;

Percent of employment lost from contractions for the total private sector in the U.S. (as a percentage of total employment) — Employment Loss from Contraction;

Percent of employment lost from closings for the total private sector in the U.S. (as a percentage of total employment) — Employment Loss from Closing;

Percent of establishments with gross job gains for the total private sector in the U.S. — Gross Establishment Gain;

Percent of establishments with employment gained from expansions for the total private sector in the U.S. — Gain from Expanding Establishment;

Percent of establishments with employment gained from openings for the total private sector in the U.S. — Gain from Opening Establishment;

Percent of establishments with gross job losses for the total private sector in the U.S. — Gross Establishment Loss;

Percent of establishments with employment lost from contractions for the total private sector in the U.S. — Loss from Contracting Establishment;

Percent of establishments with employment lost from closings for the total private sector in the U.S — Loss from Closing Establishment.

Note that establishment gain does not necessarily mean new establishment, it also includes expanding establishments. Similar for establishment loss. We then calculate the Net changes in employment and establishment by taking the difference between Gross Employment Gain and Gross Employment Loss, and Gross Establishment Gain and Gross Establishment Loss. It's also worthwhile to point out that Gross Employment Gain should be the sum of Employment Gain from Expansion and Employment Gain from Opening. Other categories are constructed similarly.

C.2 Appendix Tables

Table C.1: Effect of Uncertainty on Firm Employment Growth Rate (ΔDHS, Compustat - Segment NAICS)

	Dependent Variable: DHS Change in Employment (Δdhs, t)					
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
CRUX (t)	-0.233***	-0.217***	-0.221***	-0.222***	-0.211***	-0.212***
	[0.0560]	[0.0557]	[0.0558]	[0.0558]	[0.0543]	[0.0543]
Sales Growth (Δln t-1)			0.0266***	0.0251***	0.0177***	0.0165**
			[0.00664]	[0.00671]	[0.00646]	[0.00657]
CRUX (t) × Sales Growth (Δln t-1)			-0.278**	-0.268**	-0.240**	-0.232**
			[0.109]	[0.109]	[0.108]	[0.109]
Sales Growth (Δln t-1) sqaured				0.00294*		0.00216
				[0.00157]		[0.00147]
Log Tobin's Q (t-1)					0.123***	0.122***
					[0.00437]	[0.00439]
Firm Fixed Effects	√	√	√	√	√	√
Year Fixed Effects	√					
Industry × Year Fixed Effects		√	√	√	√	√
Observations	107,030	105,374	105,374	105,374	105,374	105,374
R-squared	0.027	0.239	0.239	0.240	0.261	0.261
Number of Firms	13,130	11,569	11,569	11,569	11,569	11,569

Notes: Standard errors are clustered at firm level. Dependent variable is calculated as (emp(t)-emp(t-1))/(0.5×(emp(t)+emp(t-1))) where emp is firm's total employment. Firm idiosyncratic uncertainty measure CRUX is at time t, which as mentioned in the text captures firm uncertainty at time t-1. Sales growth is calculated as log(Sales(t-1)) - log(Sales(t-2)) and demeaned by sample mean. Squared sales growth is calculated using demeand sales growth. Tobin's Q is taken as log average Q at time t-1. All of the regression specifications include firm FE. Column (1) includes Year FE, Column (2)-(6) include industry × year FE, where industry is at 3-digit NAICS code level adjusted by segment sales. Column (2)-(6) loses some observations and firms compared with Column (1) due to singleton observations. Time ranges from 1994 to 2016.
*** p<0.01, ** p<0.05, * p<0.1

Table C.2: Effect of Uncertainty on Firm Employment Growth Rate Decomposition (ΔDHS) - IPS Weighted Regression

	Dependent Variable: DHS Change in Employment (Δdhs, t)								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
		Extensive Margin					Continuer		
VARIABLES	Total	Birth	Death	Acquisition	Divestiture	Net	Creation	Destruction	Churning
CRUX (t)	-0.225**	-0.113***	-0.0507	-0.0405	0.0882*	-0.108**	-0.0630**	-0.0451	-0.209**
	[0.0932]	[0.0337]	[0.0530]	[0.0303]	[0.0474]	[0.0517]	[0.0316]	[0.0378]	[0.0963]
Sales Growth (Δln t-1)	0.148***	0.0193***	0.0358***	0.0232***	0.00442	0.0652***	0.0350***	0.0301***	0.00721
	[0.0161]	[0.00363]	[0.00821]	[0.00368]	[0.00881]	[0.0107]	[0.00452]	[0.00819]	[0.0119]
CRUX (t) × Sales Growth (Δln t-1)	-0.543**	-0.156***	-0.152	-0.189***	0.0439	-0.0907	-0.152**	0.0614	-0.451*
	[0.245]	[0.0469]	[0.164]	[0.0469]	[0.138]	[0.170]	[0.0600]	[0.139]	[0.230]
R-squared	0.268	0.253	0.278	0.257	0.257	0.203	0.248	0.241	0.339

Notes: Standard errors are clustered at firm level. Regressions are weighted by inverse propensity score constructed by fitting logit specifications. Dependent variable is calculated as (emp(t)-emp(t-1))/ (0.5×(emp(t)+emp(t-1))) in different margins. Column (1) represents firm's total employment change. Columns (2)-(5) represent firm's employment change from establishment birth, death, acquisition and divestiture. Column (6) represents gross employment change from firm's continuing establishments. Columns (7)-(8) represent job creation and destruction in firm's continuing establishments. Column (9) represents job churning rate, which is constructed as the sum of birth, acquisition and continuer job creation less the sum of death, divestiture and continuer job destruction. Coefficients in Columns (2)-(6) should add up to coefficients in Column (1). Coefficients in Columns (7)-(8) should add up to coefficients in Column (6). Firm idiosyncratic uncertainty measure CRUX is at time t, which as mentioned in the text captures firm uncertainty at time t-1. Sales growth is calculated as log(Sales(t-1)) - log(Sales(t-2)) and demeaned by sample mean. All of the regression specifications include firm FE and industry × year FE, where industry is at 3-digit NAICS code level. Time ranges from 1994 to 2014. Number of observations is 55000 and number of firms is 6000, both rounded to the nearest thousands.

*** p<0.01, ** p<0.05, * p<0.1

Table C.3: Effect of Uncertainty on Firm Organic Employment Growth Rate (ΔDHS) - IPS Weighted Regression

	Dependent Variable: DHS Change in Employment (Δdhs, t)			
VARIABLES	(1) Gross Job Creation	(2) Organic Job Creation	(3) Gross Job Destruction	(4) Organic Job Destruction
CRUX (t)	-0.217***	-0.176***	-0.00761	-0.0958
	[0.0540]	[0.0461]	[0.0779]	[0.0641]
Sales Growth (Δln t-1)	0.0775***	0.0543***	0.0703***	0.0659***
	[0.00758]	[0.00612]	[0.0120]	[0.00796]
CRUX (t) × Sales Growth (Δln t-1)	-0.497***	-0.308***	-0.0462	-0.0901
	[0.0956]	[0.0769]	[0.218]	[0.161]
R-squared	0.296	0.286	0.305	0.317

Notes: Standard errors are clustered at firm level. Regressions are weighted by inverse propensity score constructed by fitting logit specifications. Dependent variable is calculated as (emp(t)-emp(t-1))/ (0.5×(emp(t)+emp(t-1))) in different margins. Columns (1) represents firm's gross job creation, which is the sum of organic job creation and job creation from establishment acquisition. Columns (2) represents firm's organic job creation, which is the sum of job creation from establishment birth and continuing establishments. Columns (3) and (4) are similar but represent job destruction margin. Firm idiosyncratic uncertainty measure CRUX is at time t, which as mentioned in the text captures firm uncertainty at time t-1. Sales growth is calculated as log(Sales(t-1)) - log(Sales(t-2)) and demeaned by sample mean. All of the regression specifications include firm FE and industry × year FE, where industry is at 3-digit NAICS code level. Time ranges from 1994 to 2014. Number of observations is 55000 and number of firms is 6000, both rounded to the nearest thousands.
*** p<0.01, ** p<0.05, * p<0.1

Table C.4: Effect of Peer vs Idiosyncratic Uncertainty on Firm Employment Growth Rate (ΔDHS) - IPS Weighted Peer CRUX

	Dependent Variable: DHS Change in Employment (Δdhs, t)			
VARIABLES	(1)	(2)	(3)	(4)
Peer CRUX (t)	-0.146	-0.103	-0.13	-0.0872
	[0.312]	[0.313]	[0.311]	[0.313]
Peer CRUX (t) × Sales Growth (Δln t-1)	-1.278*	-0.803	-1.314*	-0.822
	[0.704]	[0.728]	[0.687]	[0.719]
CRUX (t)		-0.201**		-0.201**
		[0.0852]		[0.0844]
CRUX (t) × Sales Growth (Δln t-1)		-0.471*		-0.492**
		[0.249]		[0.242]
Sales Growth (Δln t-1)	0.129***	0.149***	0.120***	0.141***
	[0.0112]	[0.0162]	[0.0118]	[0.0158]
Sales Growth (Δln t-1) squared			-0.00349	-0.00416
			[0.00375]	[0.00383]
Log Tobin's Q (t-1)			0.0816***	0.0814***
			[0.00780]	[0.00780]
R-squared	0.253	0.253	0.256	0.257

Notes: Standard errors are clustered at firm level. Dependent variable is calculated as (emp(t)-emp(t-1))/ (0.5×(emp(t)+emp(t-1))) where emp is firm's total employment. Firm idiosyncratic uncertainty measure CRUX is at time t, which as mentioned in the text captures firm uncertainty at time t-1. Uncertainty measure of firm's peers is calculated by taking inverse propensity score weighted average of firm FE demeaned CRUX measure of all other firms within the same industry (4-digit NAICS code). Peer CRUX is weighed by inverse propensity score constructed by fitting logit specifications. Sales growth is calculated as log(Sales(t-1)) - log(Sales(t-2)) and demeaned by sample mean. Squared sales growth is calculated using demeand sales growth. Tobin's Q is taken as log average Q at time t-1. All of the regression specifications include firm FE and industry × year FE, where industry is at 3-digit NAICS code level. Time ranges from 1994 to 2014. Number of observations is 55000 and number of firms is 6000, both rounded to the nearest thousands.

*** p<0.01, ** p<0.05, * p<0.1

Table C.5: Effect of Peer vs Idiosyncratic Uncertainty on Firm Organic Employment Growth Rate (ΔDHS) - IPS Weighted Peer CRUX

	Dependent Variable: DHS Change in Employment (Δdhs, t)									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
VARIABLES	Gross Job Creation		Gross Job Destruction		Organic Job Creation		Organic Job Destruction		Job Churning	
Peer CRUX (t)	-0.217	-0.173	0.0712	0.0702	-0.143	-0.109	-0.107	-0.092	-0.288	-0.244
	[0.167]	[0.168]	[0.240]	[0.240]	[0.144]	[0.146]	[0.187]	[0.187]	[0.271]	[0.271]
Peer CRUX (t) × Sales Growth (Δln t-1)	-1.021***	-0.524	-0.256	-0.28	-0.701**	-0.395	-0.0929	-0.0415	-0.765	-0.244
	[0.335]	[0.344]	[0.560]	[0.600]	[0.277]	[0.288]	[0.446]	[0.481]	[0.596]	[0.654]
CRUX (t)		-0.206***		0.00444		-0.160***		-0.0766		-0.210**
		[0.0495]		[0.0718]		[0.0422]		[0.0571]		[0.0892]
CRUX (t) × Sales Growth (Δln t-1)		-0.493***		0.0228		-0.305***		-0.0542		-0.516**
		[0.0961]		[0.232]		[0.0804]		[0.161]		[0.254]
Sales Growth (Δln t-1)	0.0591***	0.0806***	0.0698***	0.0688***	0.0436***	0.0568***	0.0635***	0.0658***	-0.0107	0.0118
	[0.00546]	[0.00751]	[0.00808]	[0.0127]	[0.00459]	[0.00619]	[0.00579]	[0.00858]	[0.00801]	[0.0132]
R-squared	0.283	0.284	0.291	0.291	0.276	0.276	0.3	0.3	0.326	0.327

Notes: Standard errors are clustered at firm level. Dependent variable is calculated as (emp(t)-emp(t-1))/ (0.5×(emp(t)+emp(t-1))) in different margins. Columns (1)-(2) represent firm's gross job creation, which is the sum of organic job creation and job creation from establishment acquisition. Columns (5)-(6) represent firm's organic job creation, which is the sum of job creation from establishment birth and continuing establishments. Columns (3)-(4) and (7)-(8) are similar but represent job destruction margin. Columns (9)-(10) represent job churning rate, which is constructed as the sum of birth, acquisition and continuer job creation less the sum of death, divestiture and continuer job destruction. Firm idiosyncratic uncertainty measure CRUX is at time t, which as mentioned in the text captures firm uncertainty at time t-1. Uncertainty measure of firm's peers is calculated by taking inverse propensity score weighted average of firm FE demeaned CRUX measure of all other firms within the same industry (4-digit NAICS code). The inverse propensity score is constructed by fitting logit specifications. Sales growth is calculated as log(Sales(t-1)) - log(Sales(t-2)) and demeaned by sample mean. All of the regression specifications include firm FE and industry × year FE, where industry is at 3-digit NAICS code level. Time ranges from 1994 to 2014. Number of observations is 55000 and number of firms is 6000, both rounded to the nearest thousands.
*** p<0.01, ** p<0.05, * p<0.1

BIBLIOGRAPHY

Andrew B. Abel and Janice C. Eberly. An Exact Soultion for the Investment and Market Value of a Firm Facing Uncertainty, Adjustment Costs, and Irreversibility. NBER Working Papers 4412, National Bureau of Economic Research, Inc, July 1993. URL `https://ideas.repec.org/p/nbr/nberwo/4412.html`.

Daron Acemoglu, Ufuk Akcigit, Harun Alp, Nicholas Bloom, and William Kerr. Innovation, Reallocation, and Growth. *American Economic Review*, 108(11): 3450–3491, November 2018. URL `https://ideas.repec.org/a/aea/aecrev/v108y2018i11p3450-91.html`.

David Autor, David Dorn, Gordon H. Hanson, Gary Pisano, and Pian Shu. Foreign Competition and Domestic Innovation: Evidence from U.S. Patents. IZA Discussion Papers 12638, Institute of Labor Economics (IZA), September 2019. URL `https://ideas.repec.org/p/iza/izadps/dp12638.html`.

Doron Avramov, Minwen Li, and Hao Wang. Risk and corporate policies: A text-based analysis. *Available at SSRN 2732564*, 2016.

Rüdiger Bachmann and Christian Bayer. 'Wait-and-See' business cycles? *Journal of Monetary Economics*, 60(6):704–719, 2013. doi: 10.1016/j.jmoneco.2013.05.

Rüdiger Bachmann and Christian Bayer. Investment Dispersion and the Business Cycle. *American Economic Review*, 104(4):1392–1416, April 2014.

Rüdiger Bachmann, Steffen Elstner, and Eric R. Sims. Uncertainty and Economic Activity: Evidence from Business Survey Data. *American Economic Journal: Macroeconomics*, 5(2):217–249, April 2013.

Scott R. Baker, Nicholas Bloom, and Steven J. Davis. Measuring economic policy uncertainty. *The Quarterly Journal of Economics*, 131(4):1593–1636, 2016. doi: 10.1093/qje/qjw024.

Natarajan Balasubramanian and Jagadeesh Sivadasan. What Happens When Firms Patent? New Evidence from U.S. Economic Census Data. *The Review of Economics and Statistics*, 93(1):126–146, February 2011. URL `https://ideas.repec.org/a/tpr/restat/v93y2011i1p126-146.html`.

Jose Maria Barrero, Nicholas Bloom, and Ian Wright. Short and Long Run Uncertainty. NBER Working Papers 23676, National Bureau of Economic Research, Inc, August 2017.

Gustaf Bellstam, Sanjai Bhagat, and J. Anthony Cookson. A Text-Based Analysis of Corporate Innovation. *SSRN*, 2019.

Itzhak Ben-David and John R. Graham. Managerial Miscalibration. *The Quarterly Journal of Economics*, 128(4):1547–1584, 2013.

Jan Bena and Kai Li. Corporate Innovations and Mergers and Acquisitions. *Journal of Finance*, 69(5):1923–1960, October 2014. URL `https://ideas.repec.org/a/bla/jfinan/v69y2014i5p1923-1960.html`.

Ben S Bernanke. Irreversibility, Uncertainty, and Cyclical Investment. *The Quarterly Journal of Economics*, 98(1):85–106, February 1983.

Nicholas Bloom. The Impact of Uncertainty Shocks. *Econometrica*, 77(3):623–685, 05 2009.

Nicholas Bloom, Steven J. Davis, Lucia Foster, Brian Lucking, Scott Ohlmacher, and Itay Saporta Eksten. Business-level expectations and uncertainty. *Available at SSRN 3085377*, 2017a.

Nicholas Bloom, Charles I Jones, John Van Reenen, and Michael Webb. Are Ideas Getting Harder to Find? CEP Discussion Papers dp1496, Centre for Economic Performance, LSE, September 2017b. URL `https://ideas.repec.org/p/cep/cepdps/dp1496.html`.

Nicholas Bloom, Max Floetotto, Nir Jaimovich, Itay Saporta-Eksten, and Stephen J. Terry. Really Uncertain Business Cycles. *Econometrica*, forthcoming.

Nick Bloom, Stephen Bond, and John Van Reenen. Uncertainty and Investment Dynamics. *Review of Economic Studies*, 74(2):391–415, 2007.

Andriy Bodnaruk, Tim Loughran, and Bill McDonald. Using 10-K Text to Gauge Financial Constraints. *Journal of Financial and Quantitative Analysis*, 50(4):623–646, August 2015. URL `https://ideas.repec.org/a/cup/jfinqa/v50y2015i04p623-646_00.html`.

Matthias M. M. Buehlmaier and Toni M. Whited. Are financial constraints priced? evidence from textual analysis. *The Review of Financial Studies*, page hhy007, 2018a.

Matthias M M Buehlmaier and Toni M Whited. Are Financial Constraints Priced? Evidence from Textual Analysis. *Review of Financial Studies*, 31(7):2693–2728, 2018b. URL `https://ideas.repec.org/a/oup/rfinst/v31y2018i7p2693-2728..html`.

Matias D. Cattaneo, Richard K. Crump, Max H. Farrell, and Yingjie Feng. Binscatter Regressions. Papers 1902.09615, arXiv.org, February 2019. URL `https://ideas.repec.org/p/arx/papers/1902.09615.html`.

Russell W. Cooper and John C. Haltiwanger. On the Nature of Capital Adjustment Costs. *The Review of Economic Studies*, 73(3):611–633, 07 2006.

Steven J. Davis, John Haltiwanger, Ron Jarmin, and Javier Miranda. Volatility and Dispersion in Business Growth Rates: Publicly Traded versus Privately Held Firms. In *NBER Macroeconomics Annual 2006, Volume 21*, NBER Chapters, pages 107–180. National Bureau of Economic Research, Inc, 2007.

Steven J. Davis, John Haltiwanger, Kyle Handley, Ron Jarmin, Josh Lerner, and Javier Miranda. Private Equity, Jobs, and Productivity. *American Economic Review*, 104(12):3956–3990, December 2014.

Avinash K Dixit. Entry and Exit Decisions under Uncertainty. *Journal of Political Economy*, 97(3):620–638, June 1989. doi: 10.1086/261619. URL `https://ideas.repec.org/a/ucp/jpolec/v97y1989i3p620-38.html`.

Avinash K. Dixit and Robert S. Pindyck. *Investment under Uncertainty*. Number 5474 in Economics Books. Princeton University Press, 1994. ISBN ARRAY(0x61d23878).

Michael Ewens, Ryan H. Peters, and Sean Wang. Measuring Intangible Capital with Market Prices. NBER Working Papers 25960, National Bureau of Economic Research, Inc, June 2019. URL `https://ideas.repec.org/p/nbr/nberwo/25960.html`.

Teresa C. Fort and Shawn D. Klimek. The Effects of Industry Classification Changes on US Employment Composition. Working Papers 18-28, Center for Economic Studies, U.S. Census Bureau, June 2018. URL `https://ideas.repec.org/p/cen/wpaper/18-28.html`.

Lucia Foster, Cheryl Grim, and John Haltiwanger. Reallocation in the Great Recession: Cleansing or Not? *Journal of Labor Economics*, 34(S1):293–331, 2016.

Lucia S. Foster, Cheryl A. Grim, John Haltiwanger, and Zoltan Wolf. Macro and Micro Dynamics of Productivity: From Devilish Details to Insights. NBER Working Papers 23666, National Bureau of Economic Research, Inc, August 2017. URL `https://ideas.repec.org/p/nbr/nberwo/23666.html`.

Xavier Gabaix. The Granular Origins of Aggregate Fluctuations. *Econometrica*, 79(3):733–772, May 2011.

Zvi Griliches. Patent Statistics as Economic Indicators: A Survey. *Journal of Economic Literature*, 28(4):1661–1707, December 1990. URL `https://ideas.repec.org/a/aea/jeclit/v28y1990i4p1661-1707.html`.

Luigi Guiso and Giuseppe Parigi. Investment and Demand Uncertainty. *The Quarterly Journal of Economics*, 114(1):185–227, 1999.

Bronwyn H. Hall, Adam B. Jaffe, and Manuel Trajtenberg. The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools. NBER Working Papers 8498, National Bureau of Economic Research, Inc, October 2001. URL `https://ideas.repec.org/p/nbr/nberwo/8498.html`.

Bronwyn H. Hall, Adam Jaffe, and Manuel Trajtenberg. Market Value and Patent Citations. *RAND Journal of Economics*, 36(1):16–38, Spring 2005. URL `https://ideas.repec.org/a/rje/randje/v36y20051p16-38.html`.

Kyle Handley. Exporting under trade policy uncertainty: Theory and evidence. *Journal of International Economics*, 94(1):50–66, 2014.

Kyle Handley and J. Frank Li. Measuring the Effects of Firm Uncertainty on Economic Activity: New Evidence from One Million Documents. Technical report, Mimeo, 2019.

Kyle Handley and Nuno Limão. Trade and Investment under Policy Uncertainty: Theory and Firm Evidence. *American Economic Journal: Economic Policy*, 7(4), 2015.

Kyle Handley and Nuno Limão. Policy uncertainty, trade, and welfare: Theory and evidence for china and the united states. *American Economic Review*, 107 (9):2731–83, September 2017. doi: 10.1257/aer.20141419.

Stephen Hansen, Michael McMahon, and Andrea Prat. Transparency and Deliberation Within the FOMC: A Computational Linguistics Approach. *The Quarterly Journal of Economics*, 133(2):801–870, 2018. URL `https://ideas.repec.org/a/oup/qjecon/v133y2018i2p801-870..html`.

Tarek A. Hassan, Stephan Hollander, Laurence van Lent, and Ahmed Tahoun. Firm-level political risk: Measurement and effects. Working Paper 24029, National Bureau of Economic Research, November 2017.

Gerard Hoberg and Gordon Phillips. Real and Financial Industry Booms and Busts. *Journal of Finance*, 65(1):45–86, February 2010a. doi: 10.1111/j.1540-6261.2009. URL `https://ideas.repec.org/a/bla/jfinan/v65y2010i1p45-86.html`.

Gerard Hoberg and Gordon Phillips. Product Market Synergies and Competition in Mergers and Acquisitions: A Text-Based Analysis. *Review of Financial Studies*, 23(10):3773–3811, October 2010b. URL `https://ideas.repec.org/a/oup/rfinst/v23y2010i10p3773-3811.html`.

Gerard Hoberg and Gordon Phillips. Text-Based Network Industries and Endogenous Product Differentiation. *Journal of Political Economy*, 124(5):1423–1465, 2016.

Allen H. Huang, Reuven Lehavy, Amy Y. Zang, and Rong Zheng. Analyst information discovery and interpretation roles: A topic modeling approach. *Management Science*, forthcoming.

Boyan Jovanovic and Serguey Braguinsky. Bidder Discounts and Target Premia in Takeovers. *American Economic Review*, 94(1):46–56, March 2004. URL `https://ideas.repec.org/a/aea/aecrev/v94y2004i1p46-56.html`.

Boyan Jovanovic and Peter L. Rousseau. The Q-Theory of Mergers. *American Economic Review*, 92(2):198–204, May 2002. URL `https://ideas.repec.org/a/aea/aecrev/v92y2002i2p198-204.html`.

Boyan Jovanovic and Peter L. Rousseau. Mergers as Reallocation. *The Review of Economics and Statistics*, 90(4):765–776, November 2008. URL `https://ideas.repec.org/a/tpr/restat/v90y2008i4p765-776.html`.

Kyle Jurado, Sydney C. Ludvigson, and Serena Ng. Measuring Uncertainty. *American Economic Review*, 105(3):1177–1216, March 2015. URL `https://ideas.repec.org/a/aea/aecrev/v105y2015i3p1177-1216.html`.

Daniel Kahneman. Maps of Bounded Rationality: Psychology for Behavioral Economics. *American Economic Review*, 93(5):1449–1475, December 2003.

Tor Jakob Klette and Samuel Kortum. Innovating Firms and Aggregate Innovation. *Journal of Political Economy*, 112(5):986–1018, October 2004. doi: 10.1086/422563. URL `https://ideas.repec.org/a/ucp/jpolec/v112y2004i5p986-1018.html`.

Leonid Kogan, Dimitris Papanikolaou, Amit Seru, and Noah Stoffman. Technological Innovation, Resource Allocation, and Growth. *The Quarterly Journal of Economics*, 132(2):665–712, 2017. URL `https://ideas.repec.org/a/oup/qjecon/v132y2017i2p665-712..html`.

John V Leahy and Toni M Whited. The Effect of Uncertainty on Investment: Some Stylized Facts. *Journal of Money, Credit and Banking*, 28(1):64–83, February 1996.

Feng Li. Textual Analysis of Corporate Disclosures: A Survey of the Literature. *Journal of Accounting Literature*, 29:143–165, 2010.

Feng Li, Russell Lundholm, and Michael Minnis. A Measure of Competition Based on 10-K Filings. *Journal of Accounting Research*, 51(2):399–436, May 2013.

Tim Loughran and Bill Mcdonald. When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *Journal of Finance*, 66(1):35–65, February 2011. doi: j.1540-6261.2010.01625.x. URL `https://ideas.repec.org/a/bla/jfinan/v66y2011i1p35-65.html`.

Tim Loughran and Bill Mcdonald. Textual Analysis in Accounting and Finance: A Survey. *Journal of Accounting Research*, 54(4):1187–1230, September 2016. doi: 10.1111/1475-679X.12123. URL `https://ideas.repec.org/a/bla/joares/v54y2016i4p1187-1230.html`.

Masayuki Morikawa. Business uncertainty and investment: Evidence from Japanese companies. *Journal of Macroeconomics*, 49(C):224–236, 2016.

Whitney K. Newey and Kenneth D. West. Automatic Lag Selection in Covariance Matrix Estimation. *Review of Economic Studies*, 61(4):631–653, 1994.

Dennis Novy and Alan M. Taylor. Trade and Uncertainty. NBER Working Papers 19941, National Bureau of Economic Research, Inc, February 2014. URL `https://ideas.repec.org/p/nbr/nberwo/19941.html`.

G Steven Olley and Ariel Pakes. The Dynamics of Productivity in the Telecommunications Equipment Industry. *Econometrica*, 64(6):1263–1297, November 1996. URL `https://ideas.repec.org/a/ecm/emetrp/v64y1996i6p1263-97.html`.

Gordon M. Phillips and Alexei Zhdanov. R&D and the Incentives from Merger and Acquisition Activity. *Review of Financial Studies*, 26(1):34–78, 2013. URL `https://ideas.repec.org/a/oup/rfinst/v26y2013i1p34-78.html`.

Matthew Rhodes-Kropf and David T. Robinson. The Market for Mergers and the Boundaries of the Firm. *Journal of Finance*, 63(3):1169–1211, June 2008. doi: 10.1111/j.1540-6261.2008. URL `https://ideas.repec.org/a/bla/jfinan/v63y2008i3p1169-1211.html`.

Paul M Romer. Endogenous Technological Change. *Journal of Political Economy*, 98(5):71–102, October 1990. URL `https://ideas.repec.org/a/ucp/jpolec/v98y1990i5ps71-102.html`.

Edouard Schaal. Uncertainty and Unemployment. *Econometrica*, 85(6):1675–1721, Nov 2017.

Joseph A. Schumpeter. *Capitalism, Socialism and Democracy*. 1942.

Steven A. Sharpe, Nitish R. Sinha, and Christopher A. Hollrah. What's the Story? A New Perspective on the Value of Economic Forecasts. Finance and Economics Discussion Series 2017-107, Board of Governors of the Federal Reserve System (U.S.), November 2017.

Oleksandr Shepotylo and Jan Stuckatz. Political uncertainty, fdi, and trade in intermediate goods: Evidence from ukrainian firms. *Available at SSRN 2983695*, 2018.

Luke C.D. Stein and Elizabeth Stone. The effect of uncertainty on investment, hiring, and r&d: Causal evidence from equity options. *Available at SSRN 1649108*, 2013.

Fabrizio Zilibotti, Philippe Aghion, and Daron Acemoglu. Distance to Frontier, Selection, and Economic Growth. Scholarly Articles 4554122, Harvard University Department of Economics, 2006. URL `https://ideas.repec.org/p/hrv/faseco/4554122.html`.