Essays on the Use of Unstructured Data in Macroeconomics

Inaugural-Dissertation

zur Erlangung des Grades eines Doktors
der Wirtschafts- und Gesellschaftswissenschaften

durch

die Rechts- und Staatswissenschaftliche Fakultät der
Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Lukas Konstantin Püttmann

aus Frankfurt am Main

2018

Dekan:	Prof. Dr. Daniel Zimmer, LL. M.
Erstreferent:	Prof. Dr. Jürgen von Hagen
Zweitreferent:	Prof. Dr. Moritz Schularick
Tag der mündlichen Prüfung:	02.08.2018

Acknowledgements

I am thankful to my advisors for guiding me through my doctoral studies. My first supervisor Jürgen von Hagen provided advice that helped me find a clearer direction for my research. His style of supervision had a good mix of freedom and encouragements. I also enjoyed the cooperative atmosphere at his institute. I am further grateful to my second supervisor Moritz Schularick. Discussions with him sharpened my research ideas and I could always rely on him for help on my projects. I also benefited from the intellectual environment of his macroeconomic history research group in Bonn which offered many seminars and conferences.

My project on patents started out as an idea in Michael Böhm's research course. He accompanied the development of this project all the way from the inception of the idea to the final paper. Hans-Martin von Gaudecker offered courses and seminars with a focus on computing which I found inspiring. In the end, I used most of the methods he introduced me to. I thank Michael Evers for many interesting discussions over coffee. Christian Bayer, Benjamin Born, Narly Dwarkasing, Thomas Hintermaier, Simon Jäger, Keith Kuester, Moritz Kuhn, Gernot Müller and Christopher Roth were generous with their comments and ideas.

Katja Mann has been a diligent coauthor and it has been a pleasure to work with her. Carsten Detken, Anna Kalbhenn and Eric Persson were very welcoming at the European Central Bank and were open to the ideas I contributed to our joint project.

The *Bonn Graduate School of Economics* has provided me with a supportive environment for which I thank Silke Kinzig, Urs Schweizer and Benny Moldovanu. I thank the *Deutsche Forschungsgemeinschaft* and the *Wissenschaftsförderung der Sparkassen-Finanzgruppe e. V.* for financial support. In particular, I would like to thank Klaus Krummrich and Gregor Mauer for their help. I also thank Bénédicte Pillot-Bechtold for the kind interactions.

My friends made the PhD experience a pleasant one. I thank Thomas Graeber, Jens Herold, Fabian Walders, Justus Winkelmann and many other PhD colleagues at the BGSE. I thank Clara for being the better angel of my nature.

I am indebted to my parents Anita and Bernhard and my siblings Sebastian, Frederik and Clara who accompany me through all ups and downs. I dedicate my dissertation to them.

Contents

List of Figures

List of Tables

1

Introduction

Economists depend on data to find patterns and to test their theories. This data often takes the form of official statistics, such as national accounts. With the rise of cheap computer storage, faster computers and better algorithms, it has become feasible to analyze larger and less structured datasets. Insights gained from parsing text corpuses or scraping websites allow to measure concepts that are not found in the traditional data sources. Combining these new indicators with existing data and methods has proven particularly fruitful.

This thesis contains three essays in empirical economics, which leverage the novel abundance of electronic unstructured data. I address two – quite distinct – topics in this dissertation. First, I explore the effect of automation on labor markets. Second, I study financial stress.

Chapter 2: "Benign Effects of Automation: New Evidence from Patent Texts" – in joint work with Katja Mann – studies the effect of automation on labor markets. The question of whether automation will replace workers with machines has been discussed as early as Ricardo (1821) and academic findings resurface periodically in public debates on this issue. Yet, to make progress on this issue, we need good measurements of automation.

This project starts from the observation that existing indicators of automation technology have undesirable properties. Some automation proxies suffer from already imposing what the effects of automation are. Other studies use data on firm activities, but this is only remotely related to the underlying trends in technology. Instead, we make use of patents to track innovations in automation technology. Patents are a rich information source, however researchers usually only use their metadata, such as names of innovators or their citation counts (Griliches, 1990). In contrast, we analyze the actual texts of all 5 million patents granted in the United States between 1976 and 2014. Our text classification algorithm learns to recognize automation patents from a sample of 560 patents and we show that the share of automation patents has risen strongly from 25 to 67 percent of all patents. We identify industries where automation patents are likely to be used and assign patents to regions in the United States.

We rely on an identification strategy that puts three layers of separation between where and why patents originate from how they are applied. First, we disentangle the industries of invention from those of patent use. Second, we study national patenting and local economic effects as it is unlikely that idiosyncratic conditions in small regions will affect national inventive activity. Third, we identify patent owners and show similar results for automation patents from foreigners, universities and governments – groups, whose innovations are less likely to be motivated by US business interests. We document positive effects of automation technology on employment across local labor markets. This effect is driven by an increase in service sector jobs, which more than compensates for a fall in manufacturing employment.

Chapter 3: “Patterns of Panic: Financial Crisis Language in Historical Newspapers" is similar to the first in methodology, but not in content. It addresses a long-standing problem in economic history and macroeconomics: There is no accepted measurement for when financial crises occurred. Researchers depend on financial stress indicators to study the reasons why such events happen and the effects they have on society and the economy. The existing literature relies on a wide array of data, such as narrative accounts by historians or financial market data. In contrast, this paper builds on newspaper reporting.

Newspapers have been around for a long time and are published continuously and on a daily frequency. Other studies, such as Baker, Bloom, and Davis (2016), have also used newspaper data. My approach is different in that I make use of a previously untapped archive of 35 million titles of newspaper articles, published by five major US newspapers since the 19^{th} century. This allows me to go much further than the comparable literature as I am able to analyze the full-texts of titles and am not constrained to using online search masks. I identify articles that report negatively about financial markets and measure the emotional connotation of titles using established sentiment dictionaries with a total of 11,000 words.

The resulting indicator series displays plausible behavior and comoves strongly with other financial stress indicators. I validate the indicator using 23,000 manually coded articles. A time series analysis shows that spikes in the new indicator are followed by lower production, higher unemployment, lower stock market returns and higher credit spreads.

Chapter 4: “Unexpectedly Broke: Expectation Errors and Credit Cycles" – coauthored with Carsten Detken, Anna Kalbhenn and Eric Persson – follows the train of thought from the second chapter, but concentrates on another facet of the issue of financial stress. A robust finding in the literature on financial crises is that these events tend to be preceded by booms in private debt (Schularick and Taylor, 2012). But the question remains, why people accumulate such excessive debts in the first place. A mechanism that has been proposed is that agents in the economy might become overoptimistic about the future course of the economy and their income paths in boom times. This may entice them to take on too much debt.

To investigate this hypothesis, we collect a dataset of 2.6 million macroeconomic forecasts by banks and research institutes with a large international coverage of 32 countries since 1989. A fixed effects panel data analysis shows that expectations were overoptimistic in exactly those periods when credit in the economy expanded. This is the case for households, but not for firms, and thus points to a role of biased expectation formation of the former. This pattern holds for both industrialized and developing countries and is robust to controlling for the state of the business cycle, inflation expectations and current interest rates.

We provide further evidence that participants are overprecise, which means they are unreasonably confident about their predictions. Using a method by Coibion and Gorodnichenko (2012) to analyze information processing, we compare forecast revisions with forecast errors for individual forecasters. In this way, we show that forecasters – in particular in the run-up to the global financial crisis starting in 2007 – overreacted to recent positive news across the world.

References

Baker, Scott R., Nicholas Bloom, and Steven J. Davis (2016): "Measuring Economic Policy Uncertainty." *Quarterly Journal of Economics*, 131 (4), 1593–1636. [2]

Coibion, Olivier and Yuriy Gorodnichenko (2012): "What Can Survey Forecasts Tell Us about Information Rigidities?" *Journal of Political Economy*, 120 (1), 116–159. [3]

Griliches, Zvi (1990): "Patent Statistics as Economic Indicators: A Survey." *Journal of Economic Literature*, 28, 1661–1707. [1]

Ricardo, David (1821): "On Machinery." In *On the Principles of Political Economy and Taxation*. 3rd. London: John Murray. eprint: http://www.econlib.org/library/Ricardo/ricP7.html. [1]

Schularick, Moritz and Alan M. Taylor (2012): "Credit Booms Gone Bust: Monetary Policy, Leverage Cycles and Financial Crises, 1870–2008." *American Economic Review*, 102 (2), 1029–61. [3]

2

Benign Effects of Automation: New Evidence From Patent Texts

Joint with Katja Mann

2.1 Introduction

What is the effect of automation technology on employment? The answer to this question is not obvious: While machines may replace workers, new jobs could also be created. For example, if self-driving vehicles become widely used, taxi and truck drivers might lose their jobs. Other sectors such as retail could, however, experience employment growth through lower transport costs.

To identify the employment effects of automation, this paper introduces a new indicator of automation technology. The large literature addressing this question has so far relied on indirect proxies of automation, such as routine task input (Autor, Katz, and Kearney (2008), Autor, Levy, and Murnane (2003), Goos and A. Manning (2007), Autor and Dorn (2013)), investment in computer capital (Beaudry, Doms, and Lewis, 2010; Michaels, Natraj, and van Reenen, 2014) or investment in robots (Graetz and Michaels, 2015; Acemoglu and Restrepo, 2017). Many of these papers find evidence for job polarization, but the smaller literature on aggregate employment changes reports more ambiguous results. This may be due to difficulties in measuring automation comprehensively.

Our proposed automation indicator relies on patent grant texts. Patents are a natural candidate for measuring technological progress and frequently serve as proxies of innovation. However, few studies examine the consequences of technological progress through patents. Also, while patent meta-data such as citation counts or the identity of innovators is used regularly (Hall, Jaffe, and Trajtenberg, 2001; Acemoglu, Akcigit, and Celik, 2014; Bell, Chetty, Jaravel, Petkova, and Reenen, 2017), the actual patent texts have not been in the focus so far. We classify patents as automation patents if their texts describe physical inventions (such as robots) or immaterial or conceptual inventions (such as software), which carry out a process independently of human interference.

We extract the texts of all 5 million U.S. patents granted between 1976 and 2014 and train a machine learning algorithm on a sample of 560 manually classified patents to sort patents into automation and non-automation innovations. As a result, we document a strong rise in both the absolute and the relative number of automation patents. As a share of total patents, automation patents have increased from 25 percent in 1976 to 67 percent in 2014. Applying a probabilistic matching that is based on Canadian patents, we link patents to the 956 4-digit SIC industries where they are likely to be used. In this way, we quantify trends in newly available technology at the industry level.

Next, we compare the indicator to established measures of automation. The number of automation patents is positively correlated across industries both with investment in computer capital and with robots shipments. More automation patents have been granted in industries with a larger share of employment in routine occupations in 1960, a result that is in line with the literature on routine-biased technological change. Also, industries with more automation patents were characterized by a rise in non-routine cognitive and non-routine interactive task input and a fall in routine cognitive and routine manual task input.

To estimate the labor-market effects of automation, we transfer our industry-level data to U.S. commuting zones through industry-county employment counts. Commuting zones approximate local labor markets as workers tend to look for jobs within commuting distance from where they live. We obtain a panel dataset of new automation technology across 722 commuting zones over 39 years. Up to the late 1980s, there was a higher density of automation in the Great Lakes region, but automation technology has become less geographically concentrated over time.

Our empirical analysis benefits from the fact that we examine *local* economic outcomes which are impacted by, but unlikely to affect, the innovation activity of industries at the *national* level. Our key assumption is that commuting zone-specific developments in the medium-run do not affect automation innovation in industries that operate there. This is plausible for the following reasons: First, we separate the industries where patents originate from where they are used. Second, many patents belong to foreigners and universities who respond to other incentives than local firms. And third, local industries are small in comparison to national aggregate industries. Our approach thus follows Bartik (1991).

Our main econometric analysis is a fixed effects panel regression for five-year periods. Interpreting the automation index as a flow measure of technology, we assess the relationship between the sum of automation and changes in employment. While we find a positive effect of automation on total employment, this is driven by job growth in the service sector, which compensates for a fall in manufacturing employment. This result is robust to adding a variety of other economic and demographic controls and to weighting patents by the number of citations they received. We also consider separately patents belonging to specific groups of assignees: universities and public research institutes, foreigners and governments. All three should be less responsive to US labor market trends than US companies. Our results hold in the regressions for the subgroups of patentees as well as in an instrumental variable regression. Lastly, we find that automation is associated with more job creation in commuting zones where the share of routine occupations is low.

All in all, our study thus shows automation to be more beneficial for employment than some of the previous literature (Autor et al., 2015; Acemoglu and Restrepo, 2017), which might be due to our broader definition of automation. Our results are in line with Gregory, Salomons, and Zierahn (2016), who show that the detrimental substitution effect of automation on routine jobs is more than compensated by a positive labor demand effect due to larger product demand.

In the final part of our paper, we apply our indicator to replicate two central papers (Autor and Dorn (2013) and Autor, Dorn, and Hanson (2015)) that study the influence of automation on labor markets using the routine task share of jobs. First, we show that non-college employment rose in commuting zones where more automation patents could be used and where more people worked in routine occupations. Second, we find that automation leads to rises in employment levels even when controlling for Chinese import competition, which stands in contrast to Autor et al. (2015). We provide further evidence that employment increases were driven to a larger extent by flows into the labor force than by a fall in unemployment.

There are strengths and weaknesses to our approach to quantifying automation technology. Text classification is an inherently imprecise activity and we introduce further inaccuracies through probabilistic matchings of patents to industries and commuting zones. Also, we make assumptions on the usefulness of patents and the way they are implemented. On the upside, we have to impose fewer ex-ante assumption on the nature of advances in automation technology, compared to the literature using routine task shares or computer and robot investment. Our indicator allows us to closely track the technology frontier, translating newly granted patents into a fine-grained industry- or commuting zone-level dataset. With these caveats in mind, we consider our indicator a complement to previous measures of automation.

2.2 New Automation Index

This section introduces the new automation index. We start by arguing why patents are a suitable data source for measuring technological progress and then define automation. We show how we construct the indicator and how the classification algorithm works. Then, we explain how to link patents to industries in which they are likely to be used. The resulting indicator traces the technology frontier across 956 industries and 39 years and displays plausible co-movement with existing indicators of automation such as computer investment, the number of robots used in production and the share of routine tasks across industries.

2.2.1 Patents As Indicators of Technological Progress

The purpose of patents is to encourage innovation and technological progress by offering a temporary monopoly on an invention. Once granted, no one can re-engineer, create or sell the same object or idea. In return, the text of the patent is made publicly available. The language in the patent text is technical and highly standardized. Applicants have an incentive to provide exact and correct information about their innovation to obtain full protection of their ideas. Professional patent examiners judge a patent's claims and make changes where appropriate. In return for disclosing the content of the innovation to the public, an intellectual property right is granted for 20 years. To be patentable, an innovation must be *novel*, *non-obvious* and *useful*. The description must further be exact and detailed enough to allow for replication and it must name the invention's most important application. All these characteristics make patents a valuable data source.

Researchers in economics have made frequent use of patents, often in the form of the database established by Hall et al. (2001). Griliches (1990) provides an extensive survey of various issues related to using patents in economics. However, patents are so far usually interpreted as proxies for innovative activity, not as increments of technological progress whose effects can be studied (for an overview of the more recent literature, see Nagaoka et al. (2010)). This is related to the fact that existing research almost exclusively uses patents' metadata, such as the location or affiliation of a patentee or a patent's importance.[2]

[2]Patent citations, in particular, are widely applied as indicators of the value of an invention, for example by Bell et al. (2017).

Magerman, Looy, and Song (2010) note that there is almost no research which uses the actual texts of the patent document, although this has been recommended as early as Griliches (1990). An exception is Bessen and Hunt (2007), who identify software patents by searching patent texts for keywords. Our approach differs as we do not specify a priori which words to search for, but use a state-of-the-art text classification algorithm. Also, we apply the derived measure to study the effects of technology on the labor market, whereas the goal of Bessen and Hunt (2007) is to characterize firms that file software patents.

In other areas of economics, text search has become common. However, patent texts hold several advantages for researchers over other document collections: The precise technical language with a high degree of standardization, the incentive to deliver correct information, the additional check through the patent examiners' review and the public access to patent grant texts make patents well suited for text search analysis.

Patent text analysis is common in the private sector for prior art and freedom-to-operate searches by firms and lawyers. However, none of these providers – to the best of our knowledge – offers a comparison of technological trends over time, which leads us to develop our own approach.

2.2.2 Patent Data

We obtain all 5 million utility patent documents granted in the United States from 1976 to 2014 from Google.[3] While Europe, Japan and increasingly China are also important patent legislations, of the roughly 10.9 million patents effective ("in force") worldwide in 2014, the largest fraction (about one fourth) had been granted in the United States (WIPO, 2016). In addition, the most important innovations are usually patented in all major patent legislations. These properties make U.S. patents a good proxy for the technological frontier in the United States and beyond. Also, given that this paper studies the effect of automation in the United States, U.S. patents are an obvious candidate for how available technology changes.

We only consider utility patents, which account for around 90 percent of all patents. Utility patents are "issued for the invention of a new and useful process, machine, manufacture, or composition of matter, or a new and useful improvement thereof" (USPTO, 2015). Other patent types are design, plant and reissue patents and do not track technology that we aim to measure. According to the United States Patent and Trademark Office (USPTO), in the period 1976-2014, 83 percent of all patents granted were owned by firms – mostly large multinational corporations. 15 percent of patents were owned by individuals and less

[3] google.com/googlebooks/uspto-patents.html

than 2 percent by the U.S. government. About half of all patents are granted to foreign applicants, a share that has increased over time. During the period of our analysis, IBM, Canon and Samsung were the corporations with the largest number of patents granted (USPTO, 2014).

The patent grant document includes the title, patent number, name of the inventor, date, citations of other patents, legal information, drawings, abstract and a detailed description, as well as information on the technology class of the invention. Every patent is assigned one or more technology classification numbers by the patent examiner which describes technological and functional characteristics of a patent and on which we base our link from patents to industries. We exclude chemical and pharmaceutical patents from our classification.[4] The overwhelming majority of these patents do not meet our definition of an automation patent (14 out of 560 manually classified patents were automation patents from those sectors), but including these patents might distor our classification.

2.2.3 Definition of Automation

We define an automation patent to describe a *device that carries out a process independently*.[5] This broad definition captures technologies such as software, a robot used in a production or the self-driving vehicle mentioned in the introduction. The “device" can be a physical machine, a combination of machines, an algorithm or a computer program. The process it automates may be a production process, but also anything else where an input is altered to generate an output. An important element of the definition is the notion of independence: It works without human intervention, except at the start or for supervision. We require the automation innovation to be a reasonably complete process, product or machine. In addition, we require it to have an at least remotely-recognizable application. This excludes inventions that are minor parts of an automation innovation and highly abstract patents with no obvious application. We make no difference between process and product innovations, so an automation patent could describe either. Table 2.1 displays some examples of automation and non-automation patents.

[4]Excluded USPC technology numbers: 127, 252, 423, 424, 435, 436, 502, 510-585, 800, 930, 987.

[5]This is a standard definition that can be found in encyclopedias. For example, the Encyclopedia Britannica defines automata as “any of various mechanical objects that are relatively self-operating after they have been set in motion" and adds that “the term automaton is also applied to a class of electromechanical devices—either theoretical or real—that transform information from one form into another on the basis of predetermined instructions or procedures" (Encyclopædia Britannica (2015)).

Table 2.1. Examples of Automation and Non-automation Patents

Patent title	Patent number	Automation patent?
"Automatic taco machine"	5531156	Yes
"Color measuring method and device"	6362849	Yes
"Coinfusion apparatus"	8857476	Yes
"Hair dye applicator "	6357449	Yes
"Hand-held scanner having adjustable light path"	5552597	No
"Bicycle frame with device cavity"	7878521	No
"Process for making pyridinethione salts"	4323683	No
"Golf ball"	4173345	No

Note: Authors' classifications according to manual coding guidelines.

2.2.4 Classification of Patents

Based on the definition above, all patents can be classified as either automation or non-automation patents. We use an automated approach. To train a classification algorithm, we need reliable and objective classifications on which we can base the comparison. To this end, we manually classify 560 randomly drawn patents according to rules laid out in manual coding guidelines.[6] We aim to minimize coding mistakes and biases by providing a structured classification process, by classifying patents in random order and by reviewing every classification by a second person.

The language in patent texts might have changed over time. But patents from the 1970s read very similar to those from the 2000s and important technological classes such as computers and robots are developed and patented throughout the sample period. The technical nature of the documents and the fact that legal terms change more slowly than other language also makes it less likely that there are short-lived trends that could pose a problem for a classification based on specific terms.

From our sample of patents, we extract word stems, called *tokens*, with the Porter2 stemming algorithm. This shortens "automation", "automated", "automatically", "automatable" to "automat". Table 2.2 summarizes these tokens. A typical title contains about 5 tokens, a typical abstract about 36 and the rest of the patent (the "body") about 500 to 600.

[6]See: `http://lukaspuettmann.com/assets/pdf/manual_coding_guidelines.pdf`

Table 2.2. Tokens in 560 Manually Classified Patents

Part	All tokens	Unique tokens	Mean	Median
Title	2796	1301	4.99	5
Abstract	20781	3971	37.11	36
Body	339366	31499	606.01	506.5

In principle, one could now record for all 5 million patents whether they contain one of the roughly 32,000 tokens that we can assign probabilities to. But to keep the computation-intensive data collection feasible and to remove noise features, we use the *mutual information criterion* to extract those tokens which are most informative about which class a patent belongs to. This is an established statistic for feature selection which prefers tokens that appear significantly more often in one of the classes and punishes tokens that appear rarely overall (C. D. Manning et al., 2009). We then pick the highest ranked (according to the mutual information criterion) 50 title tokens, 200 abstract tokens and 500 patent body tokens. The final search dictionary consists of 623 tokens.

Figure 2.1 visualizes the 150 tokens with the highest mutual information criterion. The most important token is unsurprisingly “automat”. After that come “output”, “execut”, “inform”, “input” and “detect”. Some tokens are indicative of software, such as “microprocessor”, “database”, “comput”, “program” or “transmiss". Others are more likely to appear in descriptions of physical machines, such as “motor”, “move”, “metal” or “apparatus”. The last discernible group of tokens are action verbs that appear in descriptions of a wide range of independently operating devices, such as “distinguish”, “command”, “respons” or “perform”.

Figure 2.1. Words That Indicate an Automation Patent

Note: Token size is proportional to the value of the mutual information criterion in sample 560 classified patents. We show only the 150 highest ranked tokens excluding chemical and pharmaceutical words.
Source: USPTO, Google and own calculations.

Our algorithm emulates how a human being would have classified each patent. We apply the Naive Bayes algorithm which is a supervised learning method which is easy to interpret and which computationally scales well with large amounts of data. The "naive" assumption the probability of a token to appear in a document is independent from the appearance of other tokens. Despite its simplicity it has been shown to perform quite well (Domingos and Pazzani, 1997).[7] One reason for this that the low number of parameters it estimates make it unlikely to overfit (Murphy, 2012). The assumption of tokens appearing independently of each other also makes this classifier more robust to conceptual drift than other methods such as k-nearest neighbors (C. D. Manning et al., 2009).

C. D. Manning et al. (2009) explain how this algorithm picks the class c for every document d with maximum a posteriori probability $P(c \mid d)$. In our analysis, the documents d correspond to patent grant texts and the two different classes are automation patents and non-automation patents. In the *Bernoulli* Naive Bayes that we use, every document d is represented by a vector e, where entry e_i $(i = 1, \ldots, M)$ is 1 if token i appears at least once in the document and 0 if it does not. Patent texts contain matter-of-fact language, where words are often repeated. So the occurrence of a word is more important than the frequency of its appearance and we therefore ignore how often a word appears in a document.

According to this language model, in any document in class c the token e_i occurs with conditional probability $P(e_i \mid c)$. Therefore, the probability of a document d to show up in class c is

$$P(d \mid c) = \prod_{1 \leq i \leq M} P(e_i \mid c), \tag{2.1}$$

and the conditional probability of document d to belong to class c is according to Bayes' rule[8]

$$P(c \mid d) \propto P(c) \prod_{1 \leq i \leq M} P(e_i \mid c). \tag{2.2}$$

We estimate the prior $\widehat{P}(c)$ as the relative frequency of documents in class c in the training set. This is $\widehat{P}(\text{autom}) = \frac{147}{483} = 0.304$, as about a third of eligible

[7]Gentzkow et al. (2017) also recommend this algorithm if the number of observed features (tokens) is much larger than the size of the training sample, as is the case in our analysis.

[8]$P(c \mid d) = \frac{P(c)P(d \mid c)}{P(d)} \propto P(c)P(d \mid c)$.

patents (i.e., after removing chemical and pharmaceutical patents) were manually labeled as automation patents. We then estimate the conditional probabilities of a certain token to occur in class c, $\widehat{P}(e_i \mid c)$ as

$$\widehat{P}(e_i \mid c) = \widehat{P}(i \mid c)e_i + (1 - \widehat{P}(i \mid c))(1 - e_i), \tag{2.3}$$

where $\widehat{P}(i \mid c)$ is the share of documents with token i in class c. In this way, we calculate posterior probabilities for all 5 million patents to belong to either class and assign each patent to the class with the higher posterior probability.

Table 2.3. Contingency Table

		Computerized: No	Computerized: Yes	
Manual	No	323	88	411
Manual	Yes	25	124	149
		348	212	560

"No": not automation patent

Table 2.3 shows how human examiners and how the computer algorithm classified the set of manually investigated patents. Both the manual coding and the algorithmic classification judged around a quarter of patents to be automation patents. In 80 percent of cases ($= \frac{323+124}{560}$) both approaches agreed. The probability of a false positive (type I error) is 21 percent ($= \frac{88}{411}$). The probability of a false negative (type II error) is 17 percent ($= \frac{25}{149}$).

While some share of misclassified patents remains, as long as there is no underlying bias in the classication this should only add noise to our indicator series as we only aim to approximate trends in technology over time. Any noise should therefore push our empirical results towards zero, making it harder to detect an effect of automation.

A more precise classification might be possible when including patents' other observable characteristics such as their technological class (USPC and IPC numbers), grant years, the origins of inventors or the sector of firms. But we keep the classification into automation and non-automation separate from these observables to allow comparing automation trends across time and industries, without making these associations automatic.

2.2.5 Aggregate Properties of the Indicator

Figure 2.2 is a graphical representation of all 5 million patents granted in the United States between 1976 and 2014. We show patents by when they were

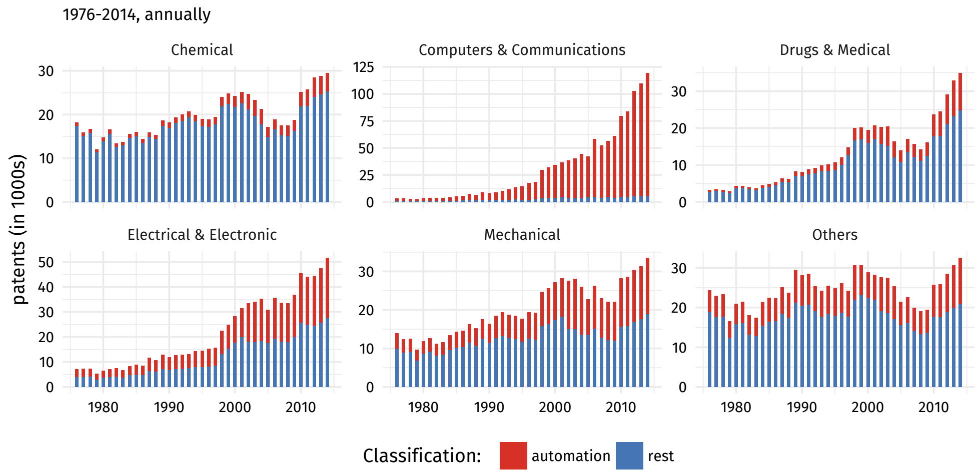


Figure 2.2. Patents, 1976-2014

Note: See text for classification of automation patents and assignment of patents to categories.

Source: USPTO, Google, Hall, Jaffe, and Trajtenberg (2001) and own calculations.

granted, not when applied for, as inventions are unlikely to be shared before they are protected by a patent.

There has been a steady increase from 70,000 granted patents in 1976 to more than 300,000 patents in 2014. Over the whole period, we classify 2.2 million of these as automation patents. The red-shaded parts of the bars show the patents which we classified as automation patents and blue colors signal all other patents. We observe a sharp upward trend in automation patents from 16,000 in 1976 to 180,000 in 2014. The share of patents related to automation also increased, from 25 percent of patents in 1976 to 67 percent of patents in 2014. Table 2.A.1 in the Appendix provides the yearly numbers.

Figure 2.2 further split up into broad categories of patents based on an aggregation method by Hall et al. (2001) which relies on the technological classification (USPC number) of patents.[9] Patents in the sub-category computers and communication have become much more frequent over the sample period and we mostly classify them as automation. Many of these are likely software patents. Electrical, electronic and mechanical patents also contribute significantly to the stock of automation patents. Robots, for example, fall in this category. By design, most

[9] Note that this is a different classification than the one we will employ to match patents to the industries they are likely to be used in. See section 2.2.6.

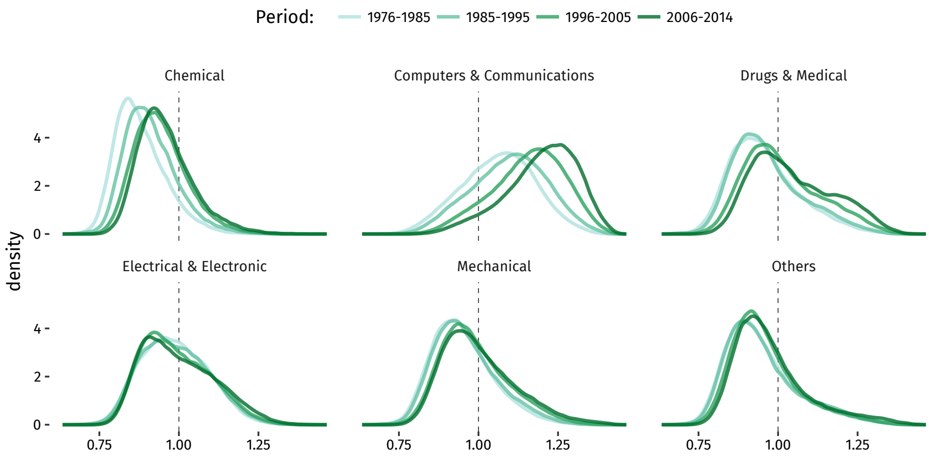


Figure 2.3. Ratio of Posteriors

Note: Every individual patent is classified as *automation* if ratio of posterior is greater than one. The figure shows smoothed densities by Hall et al. (2001) technological category and by decade.

Source: USPTO, Google, Hall, Jaffe, and Trajtenberg (2001) and own calculations.

chemical and pharmaceutical patents are not classified as automation patents, but they make up a large portion of the non-automation patents.

After our applying our algorithm, every patent has a posterior probability of belonging to either class automation or the rest. The ratio of these two posteriors is a sign of how strongly the method recommends putting a patent in either class. We plot the distribution of these ratios in Figure 2.3, separately for the four decades our dataset spans. What becomes obvious that all densities apart from “Computers & Communications" peak to the left of the vertical line, so most of the patents are not classified as automation. The distribution for “Computers & Communications" moves further to the right as time progresses, a sign that more of these patents become automation, but it could also mean that we capture later software patents better than earlier ones. The category “Electrical & Electronic" also has a second peak to the right of the vertical line, but there is no shift over time. The categories “Chemical" and “Drugs & Medical" are shown for completeness, but most of them we will assign to the non-automation category.

The rise in the total number of patents granted is a potential concern for the interpretation of the time-dimension of patent texts. If the nature of patents had changed in parallel with the number, so if the increase in patents is due to something else than an increase in research productivity, the data might not be comparable across time. An increase in the number of automation patents would then not be interpretable as an increase in automation technology. Kortum and Lerner

(1999) evaluate different possible explanations for why the number of patent grants has changed: increased patent protection due to patentee-friendly court rulings, regulatory capture by large firms that patent eagerly, new technology fields producing patentable inventions (e.g., information technology, biotechnology and financial intermediation) and more applied research. The authors refute all hypotheses except for the increase in research productivity. This result is in line with an OECD survey (OECD (2004)) in which 94 percent of surveyed firms responded that an increase in the number of inventions was an important or very important driver of their increased patenting activity (66 percent very important). In contrast, changes in patentability played only a minor role. We therefore conclude that the quality of patents granted has not changed over time and that we do not need to worry about any distortive effects of a change in grant numbers. As an additional check, we compute a deflated version of our indicator, for which we divide the number of automation patents in each industry and year by the total number of patents granted in that specific year relative to the number of patents granted in 1990. The resulting measure is an automation count in units of 1990 patents, which takes higher values for earlier years and lower values for later years than the original measure. Our empirical results in section 2.4 are insensitive to the time deflation.

2.2.6 From Patents to Industries

Various researchers have proposed matchings of patents to industries. Hall et al. (2001) identify firms filing for patents and Lybbert and Zolas (2014) propose an automated approach that compares descriptions of industries with descriptions of patents' technological classes. The OECD (2011) reviews these techniques in more detail and Griliches (1990) describes the difficulties in matching patents to industries.

However, we are interested in how automation technology affects labor markets. Therefore, we aim to find the industries where automation patents are *used*, not where they originate. These two need not be the same, so that the industry of the patentee is not necessarily the industry we want to assign the patent to. As an example, IBM owns many patents that are not used in the computer industry, but by companies in the manufacturing or in the retail sector. These patents are either sold or licensed out. Attributing them to the computer industry would overstate the automation intensity there, while understating it in the other sectors.

Linking patents to the industries of their use is difficult. If we wanted to measure the *actual* usage of a specific patent in a certain industry, we would need data on out-licensing. But this information is not available, as firms and research institutions have incentives to keep their licensing agreements private. Interpreting

Table 2.4. Automation Patents Across Industries of Use

Industries	Manufac-turing	Automation patents (1000s)	Share	SICs (1987)
Computers	✓	499	88%	357
Other electronics	✓	250	46%	36*
Measuring instruments; watches	✓	193	60%	38
Telephones and telegraphs	✓	185	68%	3661
Machines	✓	183	40%	35*
Hospitals		137	46%	8062
Househ. audio and video equip.	✓	104	69%	3651
Other services		118	47%	70-89*
Transportation equipment	✓	115	39%	37
Chemicals, rubber, plastics, oil	✓	101	18%	28, 30, 29
Utilities (transport, gas, sanitary)		57	44%	E
Fabricated metal products	✓	51	33%	34
Medical laboratories		37	64%	8071
Construction		34	24%	C
Printing publishing; paper	✓	34	32%	26, 27
Metal, stone, clay, glass, concrete	✓	29	22%	32, 33
Retail and wholesale trade		26	32%	G, F
Agriculture, forestry and fishing		24	33%	A
R&D, management	✓	23	64%	87
Miscellaneous manufacturing	✓	20	38%	39
Public administration; finance		20	47%	J, H
Food, tobacco	✓	19	24%	20, 21
Mining		16	37%	B
Apparel, wood, furniture	✓	15	17%	22-25, 31
total		2,290	46%	

Note: Sums of patents 1976-2014. Patents are counted if they can be used in an industry, as described in text. An asterisk * indicates that some subindustries are shown separately.
Source: USPTO, Google, Silverman (2002) and own calculations.

patents more indirectly as a proxy for automation technology rather than a direct measure, we can use information about the areas in which patents can *potentially* be applied. There have been attempts by Schmookler (1966) and Scherer (1984) to manually classify patents and link them to industries of use, but this would not be feasible for a large number of patents. Patent offices themselves usually do not provide information on the link of patents to industries. However, we benefit from an exception to this rule by the Canadian patent office. Between 1978 and 1993, Canadian patent officers assigned industries of use for all granted patents. Based on this information, Kortum and Putnam (1997) assembled the "Yale Technology Concordance", a way to link patents through their technological classification to the industries in which they are likely to be used. This is based on the assumption that the pattern linking patents' technological class to industries of use should be similar in Canada and the United States. We use the files provided by Silverman (2002), who calculates empirical frequencies for cross-overs from patent technology classes (IPCs) to 1987 SIC industries using 148,000 patents granted between 1990 and 1993.[10]

This allows for a probabilistic matching. We connect a patent to an industry with the probability of being used in that industry. So if patent A is used in two industries X and Y, then half the patent count is assigned to industry X and half to Y. However, patents are often assigned several IPC technology classifications. In that case, we divide each value for that patent by the number of its IPCs. So if patent A now is assigned another IPC number, then only a quarter of its value will now be attributed to industries X and Y each and the rest to industries in the new IPC. This fractional counting of patents ensures that more general patents that are assigned to several IPCs do not have get more weight than more specialized patents that are assigned to fewer IPCs.[11]

As a result, we obtain an annual dataset of new patents and new automation patents that can be used in 956 industries and over 39 years. Table 2.4 displays all automation patents by industries of use over the whole time period 1976-2014. (The totals differ slightly from Appendix Table 2.A.1 due to rounding errors and the probabilistic conversion to patent equivalents as described before.)

Out of a total of 2.3 million automation patents, 1.8 million (79 percent) are used in the manufacturing sector (division D in SIC 1987). Half a million automation patents could be used in the production of computers (SIC 357) which includes personal computers, mainframes, storage devices, terminals, billing machines, automatic teller machines and peripheral equipment such as printers, scanners, office equipments or typewriters. The production of electronic devices, sensors and communication equipment also received a large number of automation patents. Outside of the manufacturing sector, hospitals, utilities and medical laboratories were assigned a large number of automation patents. In large parts of the economy – such as agriculture, mining, public administration, finance or

[10] `http://www-2.rotman.utoronto.ca/~/silverman/ipcsic/documentation_IPC-SIC_concordance.htm`, accessed 25.10.2015. The fact that we use only data for 1990–1993 means that the matching should be most precise during this period, while becoming less exact the further away we move from this period. It helps that this periodis in the middle of our sample, but the fact that patents grow much more near in the later years is some cause for concern.

[11] This also enables us to interpret the resulting indicator as full patent equivalents which we will still refer to simply as "patents" in the remainder of the paper.

retail – only few automation patents were granted. We also calculate the share of patents used in an industry that we classify as automation. This ratio is high for the computer industry or communication-related industries and is low for the chemical industry or "Apparel, wood, furniture".

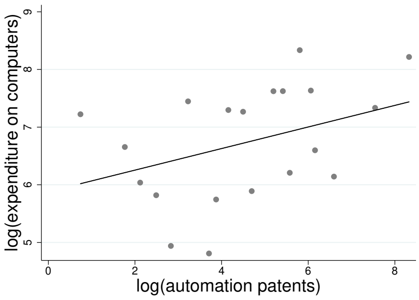
In our following empirical analysis, we interpret these indices as worker intensities by fully assigning all new (automation) patents in an industry to each person employed in that industry and year. This is equivalent to assuming that patents assigned to an industry will potentially be used by everyone working in that industry. If we considered our indicator narrowly as an exact measure of the use of patents in the production process, this would not be a realistic assumption. But to us, a patent is just one part of an innovation process that will produce many types of outputs. Being a measurable outcome of this process, patents serve as a proxy for it.In our regressions we will use the total number of automation patents as our main explanatory variable, but we will also control for the amount of all other patents that can be used in an industry.

2.3 Comparison with Previous Automation Proxies

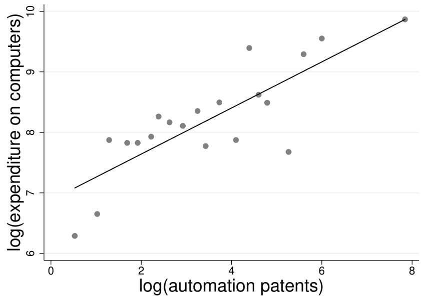
Next, we analyze how our new industry measure of automation technology is related to established automation indicators. Previous proxies of automation differ from ours along two lines. First, they are indicators of realized automation in the production process, not indicators of automation technology. Second, most capture only one specific facet of automation technology, such as computers or robots, while our indicator incorporates both and even allows delineating it from other kinds of technological progress.

As a measure of computerization, studies use survey data of computer use at the workplace (Autor et al. (2003), Beaudry et al. (2010)) or industry-level investment in computer capital (Autor et al. (2003), Michaels et al. (2014)). Frey and Osborne (2017) manually assess the probability of computerization of a number of occupations. Akerman et al. (2015) exploit a natural experiment, the introduction of broadband internet in Norway, to study employment effects of automation.

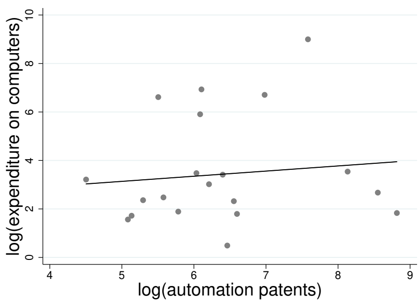
As a proxy for physical automation innovations, Graetz and Michaels (2015), Acemoglu and Restrepo (2017) and Dauth et al. (2017) count the number of robots used in production, a dataset assembled by the International Federation of Robotics. Lewis (2011) applies a more general understanding of automation by looking at adoption rates for new automation technologies, but with limited coverage of industries.



(a) Computer investment (NIPA), 1976-2001



(b) Computer investment (ASM), 2002-2014



(c) Robots (IFR), 2004-2014

Figure 2.4. Comparison with Other Indicators of Automation

Note: NIPA computer investment is the mean of 1976-2001 in millions of 1996 U.S. dollars, ASM computer investment is the mean of 2002-2014 in thousands of 2009 U.S. dollars. Robots is the mean number of robot shipments in the U.S. over 2003-2014 (U.S. data for 2003-2010 are imputed from North America data). Automation patents are counted for the same time period as the respective comparison data. All three figures show binscatters of log values.
Source: USPTO, Google, Silverman (2002), NIPA, ASM and IRF (2014).

To show how our index relates to some of these measures, Figure 2.4 correlates automation patents with investment in computer capital and shipments of robots. We use two different data sources for investment in computer capital: The National Income and Product Accounts (NIPA), which provides annual data until 2001 for 71 2- and 3-digit SIC industries and the Annual Survey of Manufactures (ASM), which is available annually from 2002 onwards and for 465 4-digit SIC industries, the majority of them being manufacturing industries. As a measure of robots, we use the dataset on robot shipments by the International Federation of Robotics, which is provided at an annual frequency for North America starting from 2004 for 24 SIC industries. All correlations are highly positive, which indicates that our automation measure captures both advances in robotics and in software, which are then translated into production and trade of computers and robots.

Another way to contextualize our indicator is to evaluate how it relates to the nature of jobs. A large strand of literature, pioneered by Autor et al. (2003), analyzes the labor market effects of automation based on the assumption that automated machines are good at carrying out repeated tasks and fail at complex intellectual or manual tasks. For each occupation, they calculate what share of a job comprises routine (manual or cognitive) tasks. The resulting routine-task index thus measures the outcome of automation given specific – theory- and data-supported – assumptions. Weighing the index by occupation-specific employment, Autor et al. (2003) further create a routine task intensity measure across 140 industries from 1960 to 1998, based on which they show that changes in routine-task intensity are predicted by investment in computer capital: The share of non-routine tasks increases, whereas that of routine tasks decreases as a result of computer investment.

Figure 2.5 plots the routine task share of industries in 1960 against new automation technology patented between 1976 and 2014.[12] The relationship between automation patents and the routine-task index is positive: The larger the routine task share of an industry in 1960, the more automation technology was subsequently invented, patented and potentially used in that industry in the following decades. Our indicator thus seems to be capturing the same phenomenon as described by the literature on routine-biased technological change.

The correlations between the two variables fell monotonously throughout the four decades.[13] The correlation is strongest in the 1970s to 1980s and declines over time. We interpret this as a sign that the nature of automation technology may have changed: While in the 1970s until 1990s, automation technology mostly replaced routine tasks, it nowadays spreads into other tasks. This could be because many routine jobs have already been replaced by automation, so that additional research in this area is less demanded and less profitable. Another possible explanation is that recent advances in the automation technology frontier affect non-routine workers by being able to replace more complex intellectual or manual tasks. (The self-driving vehicle comes to mind.)

To explore this finding further, we examine the effects of technological change separately for routine manual, routine cognitive, non-routine analytic and non-routine interactive tasks. We regress changes in industry task input within each

[12] Data on routine-task intensities at the industry level is obtained from David Autor's website `economics.mit.edu/faculty/dautor` (accessed 14.07.2015). Their dataset is for U.S. Census industries which we translate into SIC industries using a concordance scheme of the U.S. Census Bureau.

[13] 1976-1985: 0.35; 1986-1995: 0.34; 1996-2005: 0.32; 2006-2014: 0.28

Table 2.5. Automation and Industry Task Input

		Outcome: Within-industry change in task input		
		1970-1980	1980-1990	1990-98
Δ Non-routine analytic	Auto Technology	-0.012	0.033***	0.011
		(0.011)	(0.005)	(0.014)
	Constant	0.068***	0.110***	0.139***
		(0.011)	(0.014)	(0.019)
	R^2	0.004	0.019	0.001
Δ Non-routine interactive	Auto Technology	0.017*	0.062***	0.007
		(0.010)	(0.008)	(0.018)
	Constant	0.131***	0.206***	0.279***
		(0.017)	(0.030)	(0.036)
	R^2	0.004	0.016	0.000
Δ Routine cognitive	Auto Technology	-0.032**	-0.066***	-0.031***
		(0.016)	(0.011)	(0.011)
	Constant	-0.081***	-0.185***	-0.254***
		(0.022)	(0.024)	(0.038)
	R^2	0.008	0.027	0.003
Δ Routine manual	Auto Technology	-0.010***	-0.022***	-0.003
		(0.003)	(0.004)	(0.004)
	Constant	0.002	-0.058***	-0.095***
		(0.007)	(0.009)	(0.011)
	R^2	0.008	0.021	0.000

Note: The table presents separate OLS regressions for the subperiods 1970-1980, 1980-1990 and 1990-1998, always using as explanatory variable the average change of new automation patents between 1976 and 1998 (divided by 1000). The dependent variable is the change in industry-level task input as calculated by Autor et al. (2003). Standard errors in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1.

decade on our measure of new automation technology. This is a replication of a regression analysis by Autor et al. (2003), but we replace investment in computer capital with our index. To stay as close to the analysis of Autor et al. (2003) as possible we calculate the left-hand side variable separately for 1970-1980, 1980-1990 and 1990-1998 whereas on the right-hand side, we use the mean of new automation patents over the whole time period from 1976 to 1998.[14]

[14] Results are very similar when we use the whole period that our indicator covers, 1976-2014. Alternatively, we can count only automation patents of the decade for which the change in task input is calculated. The results stay qualitatively the same. Regression outputs are available from the authors upon request.

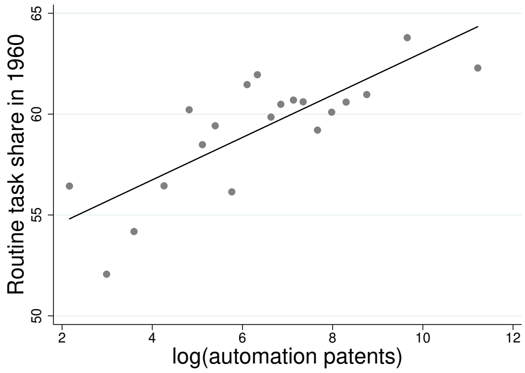


Figure 2.5. Automation Patents and Routine Labor

Note: Binscatter of log of total number of automation patents in industries against routine task input share in 1960 across 258 SIC 3-digit industries, 1976-2014.
Source: Autor et al. (2003) and see text.

Table 2.5 shows that *more* automation patents were granted in industries where routine cognitive and routine manual task inputs *declined* and where the share of non-routine analytic and non-routine interactive tasks *increased*. It is noteworthy that for all four task inputs the effect is strongest in 1980-1990. This differs from Autor et al. (2003) who found that for routine tasks the effect had monotonically increased over time.

2.4 Labor Market Effects of Automation Technology

In this section, we first motivate our unit of analysis, local labor markets, before explaining how we translate our index from industries to U.S. commuting zones. We show graphically how automation across commuting zones changed over time. Then, we apply the derived measure in our econometric analysis of employment effects. In the regression set-up, we rely on fixed effects five-year overlapping time periods, which we explain in detail before discussing the results. We run regressions for the full sample and separately for manufacturing and non-manufacturing employment.

2.4.1 Commuting Zones As Level of Analysis

We study the effects of automation on employment at the level of U.S. commuting zones. Tolbert and Sizer (1996) have grouped all counties of the U.S. mainland into 722 commuting zones which each exhibit strong commuting ties within, but weak commuting ties between one another. These regions are meant to approximate local labor markets. In response to a shock to labor demand, most adjustments in the short- and medium-run will take place within the local labor market (Blanchard and Katz (1992), Moretti (2011)). Workers, when laid off, usually first look for a new job within the same commuting zone. This is particularly true for low-skill workers, who are likely to be affected the most by automation (Notowidigdo (2011)). Therefore, studying the effects of automation on employment on the level of commuting zones gives us a more complete picture of the employment effects of automation than an industry-level analysis, which would neglect worker flows from one industry to another. This is of particular relevance because of the substantial shift of employment from manufacturing to services in the sample period.

We use employment data by the *County Business Patterns* (CBP) to convert patents per industry to worker patent automation intensities on a commuting zone level.[15] To create the commuting zone measure of automation, we first take (one plus) the natural logarithm of industry-level automation patents in order to account for the different levels of patenting across industries: In some industries the pace of technological progress is too fast for patents to be a feasible way to protect innovations, while in others, inventors have strategic reasons not to file for a patent. We then divide the employment-weighted sum of automation patents by total employment in the commuting zone. The resulting measure is

$$\underbrace{\text{autoint}_{c,t}}_{\text{automation intensity}} = \frac{\sum_i \ln(1 + \text{automation patents}_{i,t}) L_{i,c,t}}{L_{c,t}}, \tag{2.4}$$

where L is employment, i stands for industry, c for commuting zone and t for time period.

Figure 2.6 shows the number of automation patents per worker across U.S. commuting zones in four subperiods: 1976-85, 1986-95, 1996-2005 and 2006-14. The colors represent four quartiles of the distribution of automation intensity (in levels) in these subperiods: dark red color signals the 25 percent of commuting zones with the most patents, white color signals the 25 percent with the least patents. The map thus indicates which commuting zones have a high

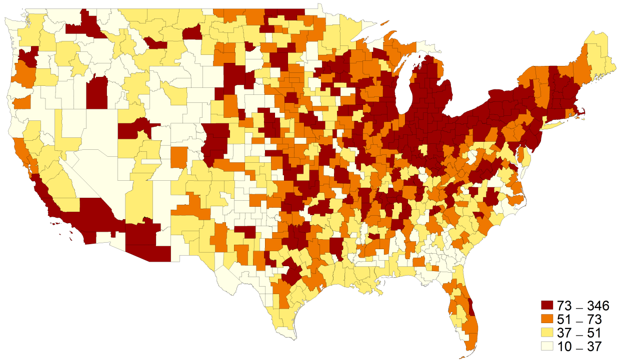
[15] In this dataset, employment numbers are reported by county and 4-digit SIC (6-digit NAICS) industry. In contrast to Census data, which is sometimes used for commuting zone analysis, CBP provides annual data for the whole period of analysis. Agriculture (SIC < 1000) and public administration (SIC > 9000) are excluded from CBP. To avoid imprecision due to SIC-NAICS correspondences and missing CBP employment data for some particular industries, we aggregate employment and the automation index on the 3-digit SIC level before matching.

or low share of patents *relative* to the rest of the United States in the specific sub-period.[16]

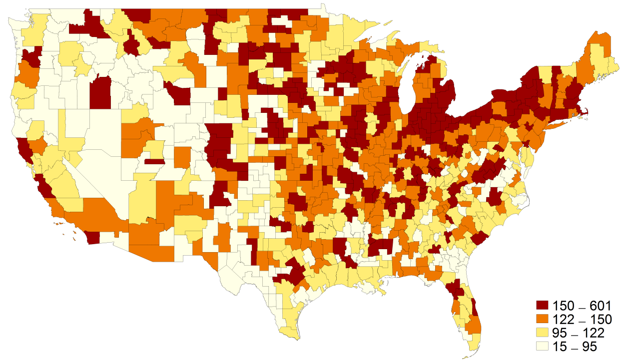
There are pronounced regional patterns in the dispersion of available automation technology. Between 1976 and 1995, the region around the Great Lakes had a large automation patent intensity relative to the rest of the United States. This stems from the conjunction of both a high number of patents in manufacturing industries and a large share of industrial employment in this area. Starting in the mid-1990s, many commuting zones in this region move to a lower quartile as the number of manufacturing employees decreased relative to the number of employees in sectors with fewer patents. But our map of automation density is not simply a reflection of the manufacturing share. In a particular the Southern United States have lower potential automation use than would be expected from their manufacturing density.

The commuting zones with the highest automation intensities are more dispersed in the 1990s and 2000s. Commuting zones in Montana, North and South Dakota and Nebraska attract many automation patents per employee. The Rocky Mountain region has a low share of patents throughout the whole sample period. The map therefore reveals substantial geographic variation over time, which we exploit in the regression analysis.

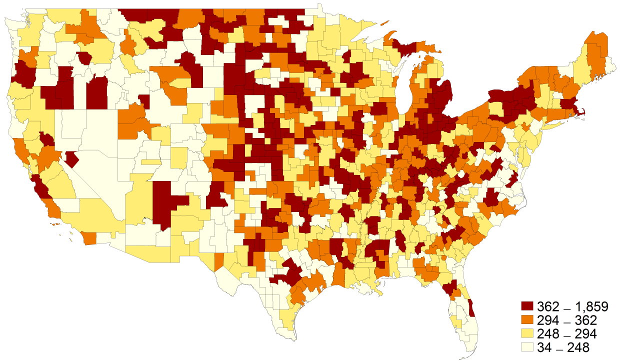
[16] As the legend shows, the absolute number of patents has increased across all quartiles. An individual commuting zone may thus have had its absolute number of patents increase constantly over time, but change from dark red to white because the index increased relatively more slowly than in other commuting zones.



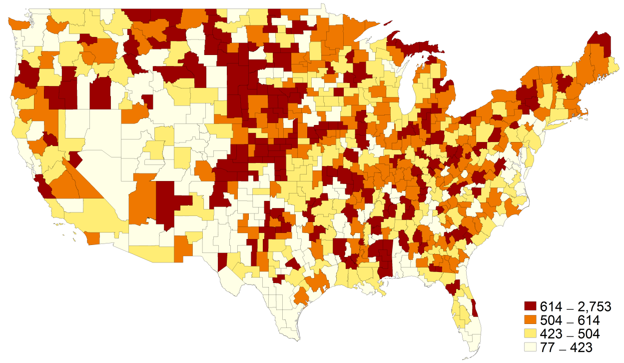
(a) 1976-1985



(b) 1986-1995



(c) 1996-2005



(d) 2006-2014

Figure 2.6. Map of Automation Patent Intensities

Note: Shows averages of the number of national automation patents that can be used by a single worker.
Source: USPTO, Google, Silverman (2002), CBP and own calculations.

2.4.2 Empirical Strategy

Our dependent variable is the five-year change in the employment-to-population ratio L_c/pop_c in commuting zone c:

$$\Delta \frac{L_{c,t}}{\mathrm{pop}_{c,t}} = \frac{L_{c,t}}{\mathrm{pop}_{c,t}} - \frac{L_{c,t-5}}{\mathrm{pop}_{c,t-5}},$$

where in contrast to automation, we observe employment directly at the commuting-zone level. We choose a medium term period as new patents might start to be used by firms only with some lags.[17] This also holds the additional benefit of smoothing out business cycle effects.

The main explanatory variable is the five-year sum of the automation intensity in a commuting zone: $\sum_{s=0}^{4} \mathrm{autoint}_{c,t-s}$. By using sums, we interpret patents as a flow measure of technology and therefore, the five-year sum of new patents is the five-year difference in the stock of patents.

In our econometric analysis we ask the following question: What is the impact of newly available nationwide automation technology on changes in the employment structure at the local level? In order to answer this question causally, we need to argue convincingly that our automation measure is exogenous to employment changes. The main potential source of endogeneity is that in their research activity, firms may be reacting to local developments, for example changes in labor costs, regulations or demand, thus introducing a reverse causality bias. There are several reasons why this is less of a concern for us:

Automation by industry of use: Assigning patents to the industries where they are likely to be used, not filed, weakens the danger of reverse causality: The research effort of a firm in one industry is less directly linked to employment trends in another industry than, for example, data on actual investment in automation technology. Additionally, many patents are granted to universities, research institutes or individuals that might follow other objectives than profit maximization, for example intellectual curiosity or an interest in advancing science. These sources of innovation are of relevance, as in year 2000 about 7000 patent licenses to firms were issued by U.S. universities and U.S. public research institutions (OECD, 2003). Further, around half of the patents granted by the USPTO are filed by foreign applicants. This reduces the potential for a feedback from industry wage structure to innovative activity, as a patent from, for example, a manufacturer in Japan is less likely to respond to employment conditions in the manufacturing industry in the United States.

[17] Results are robust to changing the length of a period.

National innovation, local effects: We measure innovations at the level of national industries, whereas we observe employment changes locally. Our constructed commuting zone automation measure is thus a proxy for unobserved locally applicable innovation in the spirit of Bartik (1991), as recently explained by Goldsmith-Pinkham, Sorkin, and Swift (2017). A national industry is unlikely to react to local employment trends in its research activity unless the following conditions hold: First, the specific commuting zone is of key importance to the industry (by hosting a large share of industry employment) and second, the industry is represented strongly in the commuting zone, so that industry trends will translate directly into commuting zone employment trends. These conditions do not drive our findings: In our sample, only two commuting zones are above the 25 percent double threshold (CZ 35002 in Arizona and CZ 37601 in Nevada, in both of which mining is dominant) and only 34 commuting zones are above the 10 percent double threshold. Excluding these does not significantly change the results.

Fixed industry structure: We fix the employment structure in equation (2.4) to the beginning of each five-year period. This means that in the following five years we assign all patents to a commuting zone according to the initial employment share of relevant industries in this commuting zone. Our indicator thus does not pick up employment changes that happen within the five-year period. A downside of keeping the employment structure fixed is that we potentially do not count all those patents which workers in a commuting zone can use, but might over-represent declining and under-represent growing industries.[18]

Additionally, in Secion 2.4.6 we exploit information on the owners of patents in order to identify innovations that more likely result from research effort that is unrelated to trends in US labor markets. We show that our baseline regression results hold when focusing only on patents held by foreigners, governments or universities and public research institutes, or when using these as instruments for the patents held by US companies.

2.4.3 Regression Set-up

We consider changes in overlapping five-year time periods and the sample therefore comprises 34 consecutive five-year periods across 722 commuting zones.[19]

---

[18] The results are however robust to using an adaptive industry structure.

[19] The overlapping data structure generates serial correlation. We correct the standard errors by using the Driscoll and Kraay (1998) estimator, which corrects both for serial and spacial correlation. An alternative would be to use non-overlapping time periods. But not only would this mean losing a considerable amount of observations (and thus precision), but it would also require us to choose cut-off points for the five-year intervals, which would always be to some extent arbi-

The estimation equation takes the form

$$\begin{aligned}
\Delta \frac{L_{c,t}}{pop_{c,t}} = {} & \alpha_k + \gamma_t + \beta_1 \sum_{s=0}^{4} \text{autoint}_{c,t-s} \\
& + \beta_2 \sum_{s=0}^{4} \text{non-autoint}_{c,t-s} + \beta_3 \text{routine}_{c,t-5} \\
& + \beta_4 \left(\sum_{s=0}^{4} \text{autoint}_{c,t-s} \times \text{routine}_{c,t-5} \right) \\
& + X'_{c,t-5} \beta_5 + \varepsilon_{c,t,t-5}\,,
\end{aligned} \tag{2.5}$$

where γ_t are time fixed effects and α_k are state fixed effects. $X_{c,t-5}$ are additional control variables. The main variable of interest *autoint* is automation intensity, *non-autoint* is the intensity of any non-automation patents and *routine* is the routine task share which we describe below. To construct the left-hand side variable, we take county level population data from the *Census Population and Housing Unit Estimates* and county-level employment data from CBP. Because the CBP omits employment in some SIC industries for certain years, there are a few large jumps in the outcome variable, which we exclude from the analysis by dropping data below the 1^{th} and beyond the 99^{th} percentile in each year.[20]

In addition to commuting zone intensities of automation patents, we include intensities of non-automation patents (*non-autoint*) in the regression, computed analogously to equation (2.4). This variable controls for the effect of technological change other than in automation technology. Given that some industries generally patent more, it is likely that the number of automation patents and non-automation patents granted annually are correlated across industries and commuting zones. At the same time, non-automation inventions may also have an independent effect on employment. In particular, they may be interpreted as an indicator for local growth potential, which we might otherwise suspect to be accountable for correlations between automation and employment: If growing industries increase their workforce as well as invest more in R&D, this should be reflected by the coefficient on *non-autoint*.

As described in Section 2.3, an often-used measure of susceptibility to automation is the routine-task index by Autor et al. (2003). The different construction of this measure from ours creates the opportunity to explore how the effects

---

trary. As shown in the appendix, all main results go through using this more standard estimation procedure instead.

[20] For details, see `census.gov/program-surveys/cbp/technical-documentation`. The number of commuting zones in each year falls to 708.

Table 2.6. Summary Statistics of Main Variables in Baseline Regression

Variable	Mean	Overall Std. Dev.	Between Std. Dev.	Within Std. Dev.	Min	Max
Δ emp/pop	1.19	2.71	0.710	2.62	-9.40	13.2
Δ manu emp/pop	-0.342	1.08	0.457	0.977	-5.35	4.33
Δ non-manu emp/pop	1.53	2.19	0.542	2.12	-8.63	12.9
autoint	16.4	3.02	1.23	2.76	7.63	28.6
non-autoint	18.8	1.89	1.38	1.29	8.88	26.7
routine	34.4	5.32	4.25	3.20	8.51	56.3

of these two are related and to ask the question: How does the effect of automation depend on the routine task share of a commuting zone? We therefore include the initial ($t-5$) routine task share (*routine*) in the regression as well as an interaction term between this measure and the variable for automation intensity.

We further include the initial share of manufacturing employment in total employment (CBP) to capture structural change in the economy. Automation patents occur to a larger extent in the manufacturing sector than in the service sector, so an increase in the automation index may parallel a decline in the manufacturing industry for other reasons, such as the cheap import of manufactured goods from abroad or changes in the demand for goods. If not included as a control, any effect stemming from non-automation-related structural change might be attributed to automation technology.

Similar to Acemoglu and Restrepo (2017), our set-up also includes the log of initial commuting zone population because employment in larger and smaller commuting zones – in particular when interpreting this as a proxy for urban vs. rural areas – might react differently to automation. We also control for the share of non-white citizens in the commuting zone population and for the (log of) per capita level of personal income. Data on the demographic variables are taken from the *Census Population and Housing Unit Estimates*, data on income come from the Bureau of Economic Analysis' *Regional Economic Information System* (REIS), which exploits county-level data from administrative records and censuses.

Table 2.6 summarizes the main variables of interest. Employment per population grew on average over the sample period.[21] Employment changes were negative

[21]This is mainly driven by increases in female labor market participation, which rose from 47 percent in 1976 to 57 percent in 2014, peaking at 60 percent in 1999. (See the BLS

on average for the manufacturing sector and positive for the non-manufacturing sector with more within and across variation for the latter.[22] Our automation intensity measure *autoint* takes the value 16.4 on average across years and commuting zones. This value is equivalent to a commuting zone with a flat industry structure (i.e., all 377 SIC 3-digit industries having the same employment share) where 25 new automation patents are granted every year in all industries. Because patents are skewed across industries, this number will be larger for most industries.

2.4.4 Estimation Results: Total Employment

Table 2.7 presents the baseline results. Throughout almost all specifications, *autoint* has a significantly positive coefficient in the range of around 0.10 to 0.23 percentage points. So new automation technology per worker is significantly related to employment gains in the same commuting zone. This result is robust to controlling for several economic and demographic variables.

Column (1) shows the positive association between automation and employment when no further controls but time and industry fixed effects are included. The relationship becomes more pronounced when we control for other non-automation patents in column (2). Columns (3) shows our preferred regression specification. The coefficient on *autoint* in column (3) can be interpreted such that a one-unit increase in the automation intensity leads to a 0.178 percentage point increase in the employment-to-population ratio. As laid out in Table 2.6, this is about one sixth of the average five-year increase across all observations. The within-year interquartile range of *autoint* lies between 1.23 and 2.15, so a one-unit increase is well within the range of variation of the sample. In terms of the actual number of new patents that this implies, a one-unit increase in *autoint* around its mean is equivalent to the number of new automation patents in a commuting zone with a flat industry structure rising from 23 to 29 per year.

A particularly interesting result is how automation technology interacts with the routine task share. In the setup with both variables in column (4), the coefficients on automation and on routine-intensity become insignificant. This is likely due to the fact that the variables measure overlapping concepts, as argued in Section 2.3. However, both coefficients are significant when we include the interaction between the two variables. The negative coefficient on the interaction shows that the magnitude of the effect of automation on employment varies

series LNS11300000, LNS11300001 and LNS11300002.) Male participation rates fell quite monotonously from 78 percent in 1976 to 69 percent in 2014. We take care of these structural long-run changes in the labor market not related to automation through time fixed effects.

[22] We will use “non-manufacturing” and “services” interchangeably, but “non-manufacturing” also includes mining and construction.

Table 2.7. Labor Market Effects of Automation, Five-year Overlapping Time Periods

	Outcome: Employment-to-population				
	(1)	(2)	(3)	(4)	(5)
autoint	0.105***	0.222***	0.178**	0.144	0.563**
	(0.0363)	(0.0783)	(0.0853)	(0.0886)	(0.214)
non-autoint		-0.120	-0.0245	0.0249	-0.0170
		(0.0997)	(0.0931)	(0.0920)	(0.0989)
manufacturing			-1.782*	-1.211	-1.177
			(1.016)	(1.082)	(1.121)
population			0.0875	0.0745	0.0525
			(0.114)	(0.108)	(0.102)
income			-1.319***	-1.284***	-1.232***
			(0.351)	(0.347)	(0.338)
non-white			-1.222***	-1.256***	-1.383***
			(0.259)	(0.273)	(0.283)
routine				-0.0257	0.143*
				(0.0161)	(0.0787)
autoint × routine					-0.0109**
					(0.00468)
Observations	24,064	24,064	24,064	24,064	24,064
R^2	0.42	0.42	0.43	0.43	0.43

Note: autoint and *non-autoint* are five-year sums of new automation and non-automation technology. *routine* is the initial percentage of routine tasks in commuting zone employment. Includes state and year fixed effects and a constant. Driscoll-Kraay standard errors in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1.

with the level of the routine task share: In commuting zones with more routine labor, automation technology has a less positive effect. The total effect of automation in column (5) turns negative for commuting zones with a routine task share larger or equal to 54.5 percent. The mean of *routine* is 34.4 and in only 0.1 percent of all observations it exceeds 54.5 percent. So, the total effect of automation is positive in the overwhelming majority of commuting zones.

Non-automation patents are not associated with changes in employment. This might be driven by the nature of these innovations. Many non-automation patents are chemical or pharmaceutical and some are patents without any clear applications. In contrast, automation patents are required by our definition to have at least a distantly recognizable application.

The initial manufacturing share has a mildly significant negative coefficient in our baseline setup of column (3), which might capture the part of the secular trend from manufacturing to services that takes place in the five-year periods we study. The population size is not significantly related to employment changes. A higher per capita income negatively predicts employment changes across all specifications. The employment level is generally higher in commuting zones with a higher per capita income. This could be a sign of convergence in employment shares across commuting zones, but could also reflect a reversely causal effect: as personal income is composed to a large extent of labor income, there could be slower employment growth in commuting zones with a higher wage level, because it is more costly to create jobs. A higher share of the non-white population is negatively associated with employment changes.

Our findings thus paint a more positive picture of the net employment effects of automation than Autor et al. (2015), Graetz and Michaels (2015) and Acemoglu and Restrepo (2017), who found negative or insignificant effects of automation on jobs.[23] It is, however, in line with the findings by Gregory et al. (2016), who show that next to a substitution effect on routine-task jobs, automation lower the production costs. Declining goods prices boost product demand, and so new (non-routine) jobs are created. The positive product demand effect trumps the negative substitution effect. Both the positive level effect of automation and the negative coefficient on the interaction term with the routine task share in our regressions support this explanation. By using a broader measure of automation, we can thus extend the knowledge on its employment effects beyond the findings of a literature that focuses on specific types of automation.

2.4.5 Estimation Results: Sectoral Employment

We further study the effect of automation on different types of employment separately. Table 2.8 shows pointedly different effects of automation technology on manufacturing and non-manufacturing employment.

Panel A consistently shows that manufacturing employment falls when the automation intensity increases. The effect is significant in our preferred specification (3) and when adding the routine task share in column (4). In contrast to the total US population, the group of manufacturing workers experiences job losses - even when controlling for the initial manufacturing share, which itself has a significantly negative effect. The negative employment effect of automation is more pronounced in commuting zones with a higher routine task share, as the interaction term shows. It turns positive only for commuting zones with a routine task share below 20.9 percent. This is only the case for 115 out of 24,058

[23] Section 2.4.5 sheds light on why this is the case.

Table 2.8. Labor Market Effects of Automation for Manufacturing and Non-manufacturing Employment, Fixed Employment Structure

	(1)	(2)	(3)	(4)	(5)
A. Outcome: Manufacturing employment-to-population					
autoint	-0.0169	-0.0480	-0.173***	-0.200***	0.144
	(0.0176)	(0.0665)	(0.0300)	(0.0300)	(0.0911)
non-autoint		0.0317	0.235***	0.275***	0.240***
		(0.0747)	(0.0299)	(0.0296)	(0.0218)
manufacturing			-2.581***	-2.142***	-2.127***
			(0.587)	(0.617)	(0.656)
population			-0.0335**	-0.0437***	-0.0608***
			(0.0133)	(0.0128)	(0.0149)
income			-0.739***	-0.712***	-0.668***
			(0.206)	(0.206)	(0.201)
non-white			-0.122	-0.150	-0.259
			(0.238)	(0.232)	(0.214)
routine				-0.0200***	0.119**
				(0.00247)	(0.0437)
autoint × routine					-0.00898***
					(0.00243)
Observations	24,058	24,058	24,058	24,058	24,058
R^2	0.21	0.21	0.25	0.25	0.26
B. Outcome: Non-manufacturing employment-to-population					
autoint	0.113***	0.278***	0.372***	0.370***	0.420***
	(0.0344)	(0.0984)	(0.0768)	(0.0799)	(0.147)
non-autoint		-0.169	-0.293***	-0.290***	-0.296***
		(0.112)	(0.0870)	(0.0840)	(0.0894)
manufacturing			0.852	0.883	0.887
			(0.728)	(0.719)	(0.726)
population			0.118	0.117	0.115
			(0.109)	(0.103)	(0.101)
income			-0.612**	-0.610**	-0.604*
			(0.291)	(0.299)	(0.298)
non-white			-1.105***	-1.107***	-1.122***
			(0.178)	(0.188)	(0.194)
routine				-0.00136	0.0186
				(0.0173)	(0.0384)
autoint × routine					-0.00129
					(0.00256)
Observations	24,067	24,067	24,067	24,067	24,067
R^2	0.38	0.39	0.39	0.39	0.39

Note: All regressions include state and year fixed effects and a constant. Driscoll-Kraay standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

observations. Panel B paints a very different picture. In non-manufacturing industries, automation has a very robust job-creating effect. The coefficients are twice as large as in Table 2.7. Non-manufacturing occupations are clear beneficiaries from automation in terms of employment numbers. In contrast to Panel A, the routine task share in the commuting zone does not play a significant role for the size of the automation effect.

Related to this, the coefficient on the routine task share also reveals strong differences between manufacturing and non-manufacturing employment. Commuting zones with a lot of routine labor lose more manufacturing jobs, but this is not the case for non-manufacturing employment. This is likely due to the larger share of routine tasks in the manufacturing than in the service sector. These findings may explain why Acemoglu and Restrepo (2017), in their analysis of the impact of robot use on employment, found automation to be harmful for employment and why Graetz and Michaels (2015), using the same dataset, found evidence for skill polarizing effects of robots: Robots are mainly used in the manufacturing sector and indeed 19 out of the 24 industries covered by IRF robot data are manufacturing industries. Other types of automation innovations, in particular those that can be used in the non-manufacturing sector, may have a more positive effect on employment than industrial robots. Indeed, Acemoglu and Restrepo (2017) show that the effect of robots is less negative or even positive in non-manufacturing industries. They also find that computer usage tends to increase the demand for labor.

We add to the existing literature by documenting different effects of automation on manufacturing and non-manufacturing employment: Next to a polarization in skills and tasks, automation has lead to a sectoral shift. Manufacturing sector jobs win, while non-manufacturing jobs lose from automation.

The results presented in this and the previous section are robust to weighing patents by how often they have been cited. Patent citations are sometimes used as an indicator of the value of an invention and therefore, giving stronger weight to highly cited patents might paint a more realistic picture of the degree to which a patent is used in the production process. In Tables 2.B.6 and 2.B.7 we replicate the regressions presented in Tables 2.7 and 2.8 using a citations-weighted measure of automation, which we explain further in the Appendix. While our sample is thus shortened by several years, we still find a mildly positive effect of automation for total employment and a pronounced disparity between manufacturing and non-manufacturing.

2.4.6 Effects of Automation by Assignees

Patents contain information on who owns (or "is assigned") a patent. This information is valuable, because it hints on how closely a patentee's research activities are linked to developments in US labor markets. Innovation activity by entities that do not have business interests in US markets is less likely to be influenced by developments on US labor markets. By focusing on new automation technologies that are originating from such groups, we therefore get a cleaner identification.

To classify the patents, we use data by Lai, D'Amour, Yu, Sun, Doolin, and Fleming (2011), who extract the names of assignees from 1976 until 2012 and provide a host of other information about patents and their owners. We focus on patents held by three groups of assignees, who we believe to be less directly responsive to US labor market trends than US companies: foreigners (these can be companies, individuals or public entities), government bodies (US or foreign) and universities and public research institutes.[24]

Research by foreigners can be assumed to respond to developments in their home country rather than in the United States, as long as the following two conditions are met: The company does not operate on a large scale in the United States, and the domestic labor market trends are not linked to US trends. We do not observe if these conditions hold, so the group of foreigners is the most endogenous of the three. Universities and public research institutes conduct more basic research than corporations, so for them, the immediate applicability or profit maximization might only be a distant motivation. Government patents are also unlikely to be motivated by labor market developments, but should rather respond to military buildups, the needs of certain ministries or cycles in budgetary planning.

Table 2.9 shows summary statistics for patents by the different groups of assignees. US firms are the largest group with around 1.9 million patents. The second largest group are foreigners, who hold 1.8 million patents. Based on the classification by Lai et al. (2011), we identify 45 thousand patents that are assigned to governments. The most important assignees in this category are the US Navy with 10,922 patents, the US Army with 6,217 patents, the US Department of Energy with 4,416 patents, the US Air Force with 3448 patents and NASA with 2,823 patents. The largest foreign government institutions owning US patents are French nuclear energy and aviation commissions and the British and Canadian defense ministries. To identify patents assigned to universities, we

[24]These groups are mostly mutually exclusive, but we count foreign governments (a small group) in both the "foreign" and the "governments" category and foreign universities also show up in the foreigners category.

Table 2.9. Assignee Summary Statistics, 1976-2012

Assignee	Patents (1000s)	Automat (1000s)	Share	Cit.	Cit. (weighted)	Excl.	Length
US firm	1875.7	877.9	47%	12.2	1.24	14%	1012.4
foreigners	1827.8	746.3	41%	7.1	0.78	12%	831.5
universities	115.1	39.5	34%	10.4	1.03	41%	1435.9
governments	44.8	16.2	36%	8.6	0.75	17%	701
missing	609.9	169	28%	9.7	0.91	9%	653.7

Note: “Automat" are automation patents as described in text. “Cit." are the average number of citations, “Cit. (weighted)" are the number of citations after removing time-subclassification (HJT) means, where subgroups correspond to those of Table 2.A.2. “Excl." is the share of excluded patents due to being pharmaceutical and chemical patents. “Length" is the average number of lines in a patent document.
Source: Lai et al. (2011) and own calculations.

inspected the 10,000 assignees with the most patents and determined whether they are an university or a public research institute. There are 581 such entities holding a total of 115 thousand patents. The most productive are the University of California (5,400 patents), the Industrial Research Institute of Taiwan (4,289 patents), the Massachusetts Institute of Technology (3897 patents), the Electronics and Telecommunications Research Institute from South Korea (3,606 patents) and the French Institute of Petroleum (2,471 patents). For the remaining 610 thousand patents, we do not know the assignee, as this information is missing in Lai et al. (2011). A casual inspection of these patents suggests that most of these also belong to US firms or individuals.

The automation patents assigned to foreigners, universities or governments may be of a different nature than those held by US firms – not just for their less direct link to economic developments in the United States, but for reasons related to their applicability. We might see different effects of automation on employment if they were not representative of the technology frontier in automation. Table 2.9 shows that patents held by US firms are characterized by a larger share of automation patents and are more widely cited than those held by other patentees. However, automation patents are highly correlated across groups at the industry level, as Table 2.10 shows. Automation innovations by governmental, foreign and university patentees seem to be applicable in similar industries as automation innovations patented by US firms or individuals. This is not the case when considering all patents. So while it is reasonable to assume that patented automation technology is similar across assignee groups, this is not the case for technology in general.

Indeed, the types of patented innovations differ across technology subgroups. As Table 2.A.2 shows, US firms hold a particularly high share of “Communication

Table 2.10. SIC-level Correlation of Patents in Assignee Subcategories with US Companies

Assignee	Patents		Automation	
	year	year & SIC	year	year & SIC
foreigners	0.33	0.33	0.94	0.95
universities	0.35	0.36	0.88	0.88
governments	-0.45	-0.43	0.02	0.04

Note: Shows correlations of subcategories with the categories of US firms and missing assignees. "year" indicates that year trends are taken out, "year & SIC" indicates that year and industry trends are taken out.

& Computer" patents, which contain a large number of automation patents. Foreigners hold fewer pharmaceutical patents, but many mechanical patents and their patents are cited least often. The column "Cit. (weighted)" in Table 2.9 shows that this holds even after controlling for time and subgroup fixed effects. Universities hold many chemical and pharmaceutical patents and few in the "Communication & Computer" category. These patents are also particularly lengthy. In contrast, governments hold many patents on electric and electronic innovations, and the corresponding patent texts are shorter than those from other assignees.

We replicate our empirical analysis from the previous section in two ways. First, we repeat the panel data regressions of Table 2.7 and Table 2.8, but for *autoint* and *non-autoint* we use the intensities computed from either only university patents, foreign patents or government patents. Second, we use all three automation sub-indicators as instrumental variables for possibly more endogenous category of US companies and non-identified assignees. The purpose of this exercise is to extract only the component of automation that is unrelated to US labor market developments. As we only have assignee data until 2012, we limit our analysis to the period 1976 to 2012.

For university patents, we document positive net effects of automation on employment. The same holds when using all three groups of automation patents as instruments in column (4). It is striking that again none of the effects of automation on total employment is negative. The size of the coefficient in Table 2.7 lies in the middle of the new estimates. Table 2.12 reports separate results for manufacturing and non-manufacturing employment. We find negative effects of automation on manufacturing employment for all assignee groups apart from university patents. All types of patented automation technology lead a rise in non-manufacturing employment. The magnitude of the coefficients again frame the previous estimates. The findings strongly support the results from our base-

Table 2.11. Labor Market Effects of Automation, Various Assignee Groups

	Outcome: Employment-to-population			
	(1) university	(2) foreign	(3) gov't	(4) IV
autoint	0.410*	0.153	-0.108	0.128*
	(0.217)	(0.128)	(0.223)	(0.0717)
non-autoint	-0.332	0.0145	0.379	0.0344
	(0.238)	(0.144)	(0.252)	(0.0756)
manufacturing	-0.769	-2.017	-2.061*	-1.961***
	(1.217)	(1.203)	(1.058)	(0.377)
population	0.121	0.110	0.113	0.119***
	(0.114)	(0.112)	(0.116)	(0.0232)
income	-1.225***	-1.393***	-1.358***	-1.358***
	(0.342)	(0.369)	(0.370)	(0.192)
non-white	-1.277***	-1.256***	-1.301***	-1.255***
	(0.233)	(0.256)	(0.250)	(0.255)
Observations	22,648	22,648	22,648	22,648
R^2	0.41	0.42	0.42	0.42

Note: All columns replicate column (3) of Table 2.7. In columns (1) - (3), the full automation measure is replaced by automation by universities, foreigners and governments, respectively. The non-automation measure is constructed accordingly. The last column represents an IV regression, where university, foreign and government (automation) patents are used as instruments for the remaining (automation) patents. Standard errors in parenthesis. *** p<0.01, ** p<0.05, * p<0.1.

line analysis and thus show that the earlier findings were likely not biased by endogeneity of the regressors.

While having roughly the same effects on employment, we can detect slight differences between the patent assignee categories. Automation technology patented by universities and public research institutes has the most strongly positive effects on employment and even the manufacturing sector does not lose from this type of technology. The negative employment effects of automation on the manufacturing sector are strongest when we consider only government patents. Why could this be the case? Universities hold many chemical and pharmaceutical patents, while governments patent many electrical and mechanical patents (Table 2.A.2). But as explained before, we exclude most chemical and pharmaceutical patents and the classification algorithm further extracts only a relevant subset of patents. As Table 2.A.3 shows, the makeup of the final automation patents does not differ much between those two groups of assignees. Pharmaceutical patents make for 4 percent of university automation patents and 1 percent of government university patent. A more likely explanation is that the innovations by universities and governments differ along other dimensions that we do not measure.

Table 2.12. Labor Market Effects of Automation for Manufacturing and Non-manufacturing Employment, Various Assignee Groups

	(1) university	(2) foreign	(3) gov't	(4) IV
A. Outcome: Manufacturing employment-to-population				
autoint	-0.120	-0.208***	-0.435***	-0.216***
	(0.114)	(0.0314)	(0.128)	(0.0329)
non-autoint	0.157	0.286***	0.518***	0.303***
	(0.145)	(0.0375)	(0.169)	(0.0331)
manufacturing	-1.796**	-2.827***	-2.441***	-2.807***
	(0.672)	(0.652)	(0.693)	(0.171)
population	-0.0399***	-0.0429***	-0.0419***	-0.0321***
	(0.0137)	(0.0138)	(0.0140)	(0.00941)
income	-0.807***	-0.793***	-0.862***	-0.724***
	(0.213)	(0.234)	(0.213)	(0.0746)
non-white	-0.287	-0.130	-0.310	-0.0937
	(0.264)	(0.257)	(0.270)	(0.125)
Observations	22,642	22,642	22,642	22,642
R^2	0.24	0.25	0.25	0.25
B. Outcome: Non-manufacturing employment-to-population				
autoint	0.518***	0.380***	0.354***	0.374***
	(0.170)	(0.112)	(0.125)	(0.0598)
non-autoint	-0.479**	-0.304**	-0.175	-0.314***
	(0.204)	(0.135)	(0.148)	(0.0637)
manufacturing	0.912	0.897	0.337	0.963***
	(0.849)	(0.842)	(0.605)	(0.310)
population	0.157	0.150	0.150	0.147***
	(0.108)	(0.108)	(0.110)	(0.0200)
income	-0.431	-0.611*	-0.506	-0.661***
	(0.304)	(0.311)	(0.314)	(0.167)
non-white	-0.989***	-1.143***	-0.978***	-1.188***
	(0.138)	(0.131)	(0.146)	(0.207)
Observations	22,650	22,650	22,650	22,650
R^2	0.37	0.37	0.37	0.37

Note: All columns replicate column (3) of Table 2.8. In columns (1) - (3), the full automation measure is replaced by automation by universities, foreigners and governments, respectively. The non-automation measure is constructed accordingly. The last column represents an IV regression, where university, foreign and government (automation) patents are jointly used as instruments for the remaining (automation) patents. Standard errors in parenthesis. *** p<0.01, ** p<0.05, * p<0.1.

2.5 Reassessing the Literature

With our new dataset we revisit findings from two important papers of the literature on the local labor market effects of automation. We investigate whether our measure of automation predicts different effects for the growth of non-college service sector jobs (Autor and Dorn (2013)) and how the effects of automation compare with those from China import competition (Autor et al. (2015)).[25] Apart from gaining additional insights through our new indicator, this allows comparing our results to the findings from the literature using the established routine-share measure.

2.5.1 Revisiting Autor and Dorn (2013): The Non-college Service Sector and Employment Polarization

Autor and Dorn (2013) address the issue why there has been an increasing polarization in both employment and wages in 1980-2005. They focus on non-college service sector jobs (e.g., cleaners or security guards), which have grown more rapidly than other less-educated and low-paying occupations (such as factory work) and which have experienced wage increases. The authors hypothesize that this is due, among other things, to an increase in automation technology: Automation has reduced the demand for routine manual tasks, while increasing the demand for non-routine manual tasks, thus benefiting non-college service sector jobs at the expense of non-college production jobs.

In their empirical analysis, Autor and Dorn (2013) use the routine-task share as a proxy for automation and show that in commuting zones where initially more people worked in routine occupations, there was a larger increase in non-college service employment. In Table 2.A.4, column (1), we reproduce their finding to the letter.

We then add *autoint*, our new automation intensity measure. The interaction term in column (4) between *autoint* and *routine* is positive and significant: Non-college service jobs rise in commuting zones with a high routine-task share initially *and* where many new automation patents could be used. This is consistent with the model presented by Autor and Dorn (2013) and highlights an important piece of evidence: the presence of those routine jobs that can be easily automated is necessary for the shift of low-skilled employment into the service sector, not the availability of automation technology by itself.

[25] Data and replication files for both papers are from David Dorn's website, `ddorn.net/data` (accessed 10.02.2017).

However, the total effect of automation changes from negative to positive only at a routine-task share of 0.38, a number reached by just 2 out of 2,166 observations and the coefficient on *autoint* in columns (2) and (3) is insignificant. So although we found in Section 2.4.5 that automation creates non-manufacturing jobs, the rise in non-college service jobs depends crucially on the mix between automation and the existence of routine jobs.

2.5.2 Revisiting Autor, Dorn and Hansen (2015): Employment Effects and Relation to Exposure to Chinese Trade Competition

Since the 1990's, there has been a strong rise in trade between the United States and China. A number of papers, such as Autor et al. (2013), Acemoglu et al. (2016) and Pierce and Schott (2016), argue that Chinese import competition is responsible for employment losses in those regions where firms reside that are most exposed to it. Autor et al. (2015) investigate whether this "China shock" or automation has a larger impact on U.S. labor markets. They find that while import competition reduces employment in local labor markets, automation – as measured by the routine task share – is not related to employment changes.

We revisit this finding with our dataset. Table 2.A.5 replicates the baseline analysis of Autor et al. (2015), Table 1, in which the authors regress 10-year equivalent changes in the employment-to-population ratio, unemployment-to-population ratio and non-participation rate among working age adults. The two main variables of interest are the contemporaneous change in Chinese import exposure per worker and the start-of-decade employment share in routine occupations, both of which are being instrumented.[26]

Columns (1) and (4) of Table 2.A.5 are exact replications of columns (1) and (3) of Autor et al. (2015), one containing only the initial routine share, the other one both the routine share and the China shock as explanatory variables. In columns (2) and (5), we replace the routine share by our commuting zone automation intensity. While the coefficient on the routine share is always insignificant, our automation measure has a significantly positive effect on the employment share and a significantly negative effects on both share of unemployed workers and the share of workers that are not in the labor force. This even holds when including both *autoint* and the routine task share. Automation patents have positive effects

[26]The instrument for the trade variable is imports from China to other advanced economies. For the initial routine task share, Autor et al. (2015) use its 1950 value in all states but the one that contains the commuting zone, weighted by 1950 employment shares. They argue that in this way, they can isolate the stable, long-run differences in the production structure across commuting zones.

by reducing the unemployment rate and the number of people outside of the labor force, with a larger effect on the latter group.

An additional finding is that while the effect of the routine task share stays insignificant when including the China shock in column (5), the estimates become even more strongly positive when using our automation indicator. The coefficient on the China shock change little when using *autoint* (column (5)) instead of the *routine* (column (4)). This lends further support to the findings of Autor et al. (2015) on the detrimental effect of Chinese import competition, while automation is playing a more positive role now.

2.6 Conclusion

This paper makes two contributions: First, it provides a new indicator of automation by applying a text classification algorithm to the universe of U.S. patents granted since 1976. Linking patents to their industry of use and, ultimately, to commuting zones, we construct geographical intensities of newly available automation technology. The second contribution is a fresh assessment of the labor market effects of automation. In an econometric analysis, we show that in commuting zones where more newly-invented automation technology becomes available, the employment-to-population ratio increases. At the same time, there is a shift from routine manufacturing jobs towards non-routine service sector jobs. These results hold when we study only patents by universities, governments or foreigners, which are likely less responsive to developments in US labor markets than domestic firms.

While rising employment ratios in response to automation technology are good news, the benefits of automation may be unevenly distributed. We hope that future research will provide more insights in this respect. A more general contribution of this paper is that it pioneers a way of extracting trends in innovation which can also be used to study the effects of other technologies on the economy.

References

Acemoglu, Daron, Ufuk Akcigit, and Murat Alp Celik (2014): "Young, Restless and Creative: Openness to Disruption and Creative Innovations." NBER Working Paper No. 19894. [6]

Acemoglu, Daron, David Autor, David Dorn, Gordon H Hanson, and Brendan Price (2016): "Import Competition and the Great US Employment Sag of the 2000s." *Journal of Labor Economics*, 34 (1), 141–198. [44]

Acemoglu, Daron and Pascual Restrepo (2017): "Robots and Jobs: Evidence from US Labor Markets." NBER Working Paper No. 23285. [5, 7, 20, 31, 34, 36]

Akerman, Anders, Ingvil Gaarder, and Magne Mogstad (2015): "The Skill Complementarity of Broadband Internet." *The Quarterly Journal of Economics*, 130 (4), 1781–1824. [20]

Autor, David H. and David Dorn (2013): "The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market." *American Economic Review*, 103 (5), 1553–1597. [5, 7, 43, 50]

Autor, David H., David Dorn, and Gordon H. Hanson (2013): "The China Syndrome: Local Labor Market Effects of Import Competition in the United States." *American Economic Review*, 103 (6), 2121–2168. [44]

Autor, David H., David Dorn, and Gordon H. Hanson (2015): "Untangling Trade and Technology: Evidence from Local Labor Markets." *The Economic Journal*, 125 (584), 621–646. [7, 34, 43–45, 51]

Autor, David H., Lawrence F. Katz, and Melissa S. Kearney (2008): "Trends in US Wage Inequality: Revising the Revisionists." *The Review of Economics and Statistics*, 90 (2), 300–323. [5]

Autor, David H., Frank Levy, and Richard J. Murnane (2003): "The Skill Content of Recent Technological Change: An Empirical Exploration." *The Quarterly Journal of Economics*, 118 (4), 1279–1333. [5, 20, 22–24, 30]

Bartik, Timothy J (1991): *Who Benefits from State and Local Economic Development Policies?* Kalamazoo, MI: W.E. Upjohn Institute for Employment Research. [6, 29]

Beaudry, Paul, Mark Doms, and Ethan Lewis (2010): "Should the Personal Computer Be Considered a Technological Revolution? Evidence from US metropolitan areas." *Journal of Political Economy*, 118 (5), 988–1036. [5, 20]

Bell, Alex, Raj Chetty, Xavier Jaravel, Neviana Petkova, and John Van Reenen (2017): "Who Becomes an Inventor in America? The Importance of Exposure to Innovation." NBER Working Paper No. 24062. [6, 8]

Bessen, James and Robert Hunt (2007): "An Empirical Look at Software Patents." *Journal of Economics and Management Strategy*, 16 (1), 157–189. [9]

Blanchard, Olivier Jean and Lawrence F. Katz (1992): "Regional Evolutions." *Brookings Papers on Economic Activity*, 1992 (1), 1–75. [25]

Dauth, Wolfgang, Sebastian Findeisen, Jens Südekum, and Nicole Woessner (2017): "German Robots - The Impact of Industrial Robots on Workers." CEPR Discussion Paper 12306. [20]

Domingos, Pedro and Michael Pazzani (1997): "On the Optimality of the Simple Bayesian Classifier Under Zero-One Loss." *Machine Learning*, 29 (2), 103–130. [13]

Driscoll, John C and Aart C Kraay (1998): "Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data." *Review of Economics and Statistics*, 80 (4), 549–560. [29]

Encyclopædia Britannica (2015): *Automaton*. [10]

Frey, Carl Benedikt and Michael A Osborne (2017): "The Future of Employment: How Susceptible Are Jobs to Computerisation?" *Technological Forecasting and Social Change*, 114, 254–280. [20]

Gentzkow, Matthew, Bryan T. Kelly, and Matt Taddy (2017): "Text as Data." *Journal of Economic Literature*. (Forthcoming). [13]

Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift (2017): "Bartik Instruments: What, When, Why and How." Unpublished manuscript. [29]

Goos, Maarten and Alan Manning (2007): "Lousy and Lovely Jobs: The Rising Polarization of Work in Britain." *Review of Economics and Statistics*, 89 (1), 118–133. [5]

Graetz, Georg and Guy Michaels (2015): "Robots at Work." CEPR Discussion Paper, 1335. [5, 20, 34, 36]

Gregory, Terry, Anna Salomons, and Ulrich Zierahn (2016): "Racing With or Against the Machine? Evidence from Europe." [7, 34]

Griliches, Zvi (1990): "Patent Statistics as Economic Indicators: A Survey." *Journal of Economic Literature*, 28, 1661–1707. [8, 9, 17, 52]

Hall, Bronwyn H., Adam B. Jaffe, and Manuel Trajtenberg (2001): "The NBER Patent Citation Data File: Lessons, Insights and Methodologial Tools." NBER Working Paper No. 8498. [5, 8, 15–17, 49, 50, 52–54]

IRF (2014): "World Robotics: Industrial Robots." Tech. rep. International Federation of Robotics. [21]

Kortum, Samuel and Josh Lerner (1999): "What is Behind the Recent Surge in Patenting?" *Research Policy*, 28 (1), 1–22. [16]

Kortum, Samuel and Jonathan Putnam (1997): "Assigning Patents to Industries: Tests of The Yale Technology Concordance." *Economic Systems Research*, 9 (2), 161–176. [19]

Lai, Ronald, Alexander D'Amour, Amy Yu, Ye Sun, David M. Doolin, and Lee Fleming (2011): "Disambiguation and Co-authorship Networks of the U.S. Patent Inventor Database (1975 -2010)." Fung Institute. [37, 38, 49, 50, 52]

Lewis, Ethan (2011): "Immigration, Skill Mix, and Capital Skill Complementarity." *The Quarterly Journal of Economics*, 126 (2), 1029–1069. [20]

Lybbert, Travis J. and Nikolas J. Zolas (2014): "Getting Patents and Economic Data to Speak to Each Other: An 'Algorithmic Links with Probabilities' Approach for Joint Analyses of Patenting and Economic Activity." *Research Policy*, 43 (3), 530–542. [17]

Magerman, Tom, Bart Van Looy, and Xiaoyan Song (2010): "Exploring the Feasibility and Accuracy of Latent Semantic Analysis Based on Text Mining Techniques to Detect Similarity between Patent Documents and Scientific Publication." *Scientometrics*, 82 (2), 289–306. [9]

Manning, Christopher D., Prabhakar Raghavan, and Hinrich Schütze (2009): *An Introduction to Information Retrieval*. Cambridge University Press. [12, 13]

Michaels, Guy, Ashwini Natraj, and John van Reenen (2014): "Has ICT Polarized Skill Demand? Evidence From Eleven Countries Over Twenty-Five Years." *The Review of Economics and Statistics*, 96 (1), 60–77. [5, 20]

Moretti, Enrico (2011): "Local Labor Markets." *Handbook of labor economics*, 4, 1237–1313. [25]

Murphy, Kevin P. (2012): *Machine Learning: A Probabilistic Perspective*. MIT press. [13]

Nagaoka, Sadao, Kazuyuki Motohashi, and Akira Goto (2010): "Patent Statistics as an Innovation Indicator." In *Handbook of Economics of Innovation*. Ed. by Bronwyn H. Hall and Nathan Rosenberg. Amsterdam: North Holland, 1083–1127. [8]

Notowidigdo, Matthew J. (2011): "The Incidence of Local Labor Demand Shocks." NBER Working Paper No. 17167. [25]

OECD (2003): *Turning Science Into Business: Patenting and Licensing at Public Research Organisations*. OECD Publishing. [28]

OECD (2004): "Understanding Business Patenting and Licensing: Results of a Survey." In *Patents, Innovation and Economic Performance. OECD Conference Proceedings*, 89–110. [17]

OECD (2011): *OECD Guide to Measuring the Information Society 2011*. OECD Publishing. [17]

Pierce, Justin R and Peter K Schott (2016): "The Surprisingly Swift Decline of US Manufacturing Employment." *American Economic Review*, 106 (7), 1632–1662. [44]

Scherer, Frederic M. (1984): "Using Linked Patent and R&D Data to Measure Interindustry Technology Flows." In *R&D, Patents, and Productivity*. Ed. by Zvi Griliches. University of Chicago Press. [18]

Scherer, Frederic M. and Dietmar Harhoff (2000): "Technology Policy for a World of Skew-distributed Outcomes." *Research Policy*, 29 (4-5), 559–566. [52]

Schmookler, Jacob (1966): *Invention and Economic Growth*. Harvard University Press. [18]

Silverman, Brian S. (2002): *Technological Resources and the Logic of Corporate Diversification*. Routledge. [18, 19, 21, 27]

Tolbert, Charles M. and Molly Sizer (1996): "U.S. Commuting Zones and Local Labor Areas: A 1990 Update." Economic Research Service Staff Paper No. 9614. [24]

USPTO (2014): "Extended Year Set. All Technologies (Utility Patents) Report." Tech. rep. United States Patent and Trademark Office. [10]

USPTO (2015): "Types of Patents." Tech. rep. United States Patent and Trademark Office. [9]

WIPO (2016): "WIPO IP Facts and Figures 2016." Tech. rep. No. 943E/16. [9]

Appendix 2.A Additional Tables

Table 2.A.1. Yearly automation and non-automation patents

	#A	#P			#A	#P			#A	#P	
1976	16279	70194	(25%)	*1989*	27928	95565	(35%)	*2002*	77267	167400	(54%)
1977	15433	65215	(26%)	*1990*	25925	90421	(34%)	*2003*	82017	169077	(56%)
1978	15412	66087	(26%)	*1991*	28037	96561	(35%)	*2004*	84372	164384	(58%)
1979	11721	48840	(28%)	*1992*	29165	97472	(36%)	*2005*	69602	143891	(54%)
1980	14937	61815	(28%)	*1993*	30439	98385	(38%)	*2006*	91201	173822	(59%)
1981	15885	65770	(28%)	*1994*	33699	101695	(39%)	*2007*	83196	157331	(60%)
1982	15092	57877	(31%)	*1995*	35135	101431	(41%)	*2008*	86705	157788	(62%)
1983	14546	56863	(31%)	*1996*	40411	109654	(44%)	*2009*	92843	167463	(62%)
1984	17665	67212	(31%)	*1997*	40217	112019	(44%)	*2010*	121163	219835	(62%)
1985	19415	71668	(32%)	*1998*	57293	147577	(46%)	*2011*	126328	224871	(63%)
1986	19515	70867	(32%)	*1999*	58464	153591	(45%)	*2012*	147550	253633	(65%)
1987	24359	82963	(34%)	*2000*	61273	157595	(45%)	*2013*	163112	278507	(66%)
1988	22006	77938	(33%)	*2001*	64796	166158	(46%)	*2014*	178422	301643	(67%)
								total	**2158825**	**4971078**	**(43%)**

Note: #A: number of automation patents as classified by own algorithm; the patent totals #P are reported as counted by us in the patent files. The USPTO reports slightly different numbers for total patent counts on its website, but the difference is below 0.5% in all years.
Source: USPTO, Google and own calculations.

Table 2.A.2. Assignee's Patents Across Technological Categories, 1976-2012

Assignee	Patents (1000s)	Chem-ical	Comm., Comput.	Drugs, Med.	Electr., Electron.	Mech-anical	Oth-ers
US firm	1875.7	17%	25%	11%	16%	13%	18%
foreigners	1827.8	16%	23%	7%	20%	18%	16%
universities	115.1	23%	13%	31%	17%	6%	9%
governments	44.8	21%	15%	11%	21%	14%	18%
missing	609.9	11%	8%	11%	9%	21%	39%

Note: Technological classifications are based on USPC numbers and aggregated using the scheme by Hall et al. (2001).
Source: Lai et al. (2011), Hall et al. (2001) and own calculations.

Table 2.A.3. Share of Automation Patents After Excluding Patents

Assignee	Patents (1000s)	Chem-ical	Comm., Comput.	Drugs, Med.	Electr., Electron.	Mech-anical	Oth-ers
US firm	1875.7	2%	24%	2%	8%	5%	6%
foreigners	1827.8	1%	20%	1%	7%	7%	4%
universities	115.1	2%	12%	4%	10%	2%	4%
governments	44.8	2%	11%	1%	11%	4%	7%
missing	609.9	1%	7%	2%	4%	6%	8%

Note: Technological classifications are based on USPC numbers and aggregated using the scheme by Hall et al. (2001). This table excludes all patents based on the selected pharmaceutical and chemical industries as explained in text.
Source: Lai et al. (2011), Hall et al. (2001) and own calculations.

Table 2.A.4. Automation and Non-college Service Employment, 1980-2005

	Outcome: 10 × annual change in share of non-college employment in service occupations			
	(1)	(2)	(3)	(4)
routine	0.105***		0.105***	-0.336
	(0.0320)		(0.0284)	(0.230)
autoint		-0.00100	-0.000990	-0.00533**
		(0.000688)	(0.000645)	(0.00227)
routine × autoint				0.0139*
				(0.00695)
Constant	-0.00632	0.0568***	0.0241	0.161**
	(0.0104)	(0.0210)	(0.0202)	(0.0740)
R^2	0.179	0.171	0.185	0.188

Note: 2,166 observations (3 time periods × 722 commuting zones); robust standard errors in parentheses; all models include state fixed-effects and period fixed effects and are weighted by start of period commuting zone share of national population.
*** p<0.01, ** p<0.05, * p<0.1
Source: Own calculations following Autor and Dorn (2013), Table 5.

Table 2.A.5. Labor Market Effects of Automation Patents, Routine Employment Share and Exposure to Chinese Import Competition, 1990-2007

	(1)	(2)	(3)	(4)	(5)	(6)
	A. Outcome: Share of employed in workage population					
routine	-0.0481		-0.0369	-0.207		-0.185
	(0.224)		(0.233)	(0.254)		(0.260)
autoint		0.215***	0.206***		0.331***	0.297***
		(0.0670)	(0.0748)		(0.0757)	(0.0792)
Δ (Imports from China to US)/Worker				-0.831***	-0.832***	-0.942***
				(0.215)	(0.181)	(0.221)
	B. Outcome: Share of unemployed in workage population					
routine	-0.0144		-0.0247	-0.00513		-0.0104
	(0.0616)		(0.0653)	(0.0702)		(0.0728)
autoint		-0.0579**	-0.0645**		-0.0926***	-0.0914***
		(0.0255)	(0.0282)		(0.0222)	(0.0285)
Δ (Imports from China to US)/Worker				0.186***	0.249***	0.221***
				(0.0527)	(0.0676)	(0.0612)
	C. Outcome: Share of not in labor force in workage population					
routine	0.0624		0.0616	0.213		0.195
	(0.172)		(0.178)	(0.194)		(0.197)
autoint		-0.158***	-0.141**		-0.239***	-0.206***
		(0.0538)	(0.0608)		(0.0667)	(0.0672)
Δ (Imports from China to US)/Worker				0.645***	0.583***	0.721***
				(0.188)	(0.155)	(0.190)

Note: The table is based on Autor et al. (2015), Table 1, juxtaposing the effect of Chinese import competition and routine biased technological change on 10-year equivalent changes in the employment status of the working-age population. N = 1444 (2 time periods 1990-2000, 2000-2007, 722 commuting zones). All regressions control for the start of period levels of share of employment in manufacturing, share of population that is college educated, share of population that is foreign born, employment rate among females and Census division dummies. Robust standard errors in parentheses are clustered on state. Models are weighted by start of period commuting zone share of national population. *** p < 0.01, ** p < 0.05, * p < 0.1.

Appendix 2.B Further Robustness Checks

2.B.1 Patent Citations

Not all patents are of the same importance. Scherer and Harhoff (2000) show that the returns on innovation are highly concentrated, with the 10 percent most valuable patents accounting for around 80 percent of realized value. While Griliches (1990) argues that using a large number of patents partly addresses this concern, we can count how often a patent was cited by other patents as an indicator of its value. We use the patent citations files by Lai et al. (2011) until 2009. The number of citations per patents follow a well-known hump-shape, as newer patents are cited less frequently, but the propensity to cite has risen. Also, some industries (such as pharmaceutical and chemical patents) cite many more patents than others (such as electronics). To control for this, we demean citations across years and the broad technology classes defined by Hall et al. (2001). This is the "fixed effect" method proposed by Hall et al. (2001).

We then weight patents by how often they were cited and replicate our analysis. The analysis shows similar results: Manufacturing employment falls and service employment rises when more (citation-weighted) automation patents become available. The baseline effect on all employment becomes insignificant in this specification, but the interaction between automation and routine task share is still significant.

2.B.2 Non-overlapping Five-year Periods

As an alternative to the five-year overlapping regressions presented in the main part of the paper, we show regression results for non-overlapping periods. These are 1977-1981, 1982-1986, 1987-1991, 1992-1996, 1997-2001, 2002-2006, 2007-2011 and 2012-2014, for which we compute five-year equivalents for the last period that covers only three years. The panel therefore comprises 8 time periods and 708 commuting zones. The results are similar to those presented in the main text. The coefficients in Table 2.B.8 are slightly larger and more significant than those presented in Table 2.7. The effects of automation for the two employment groups of Table 2.B.9 are also each slightly more positive than those of Table 2.8, but the finding of the contrary effect of automation is strongly supported.

Table 2.B.6. Labor Market Effects of Automation of Citations-weighted Patents

	Outcome: Employment-to-population				
	(1)	(2)	(3)	(4)	(5)
autoint	0.0896**	0.177*	0.0337	-0.0226	0.456**
	(0.0339)	(0.0990)	(0.0748)	(0.0834)	(0.212)
non-autoint		-0.0917	0.104	0.182*	0.106
		(0.124)	(0.0885)	(0.0949)	(0.107)
manufacturing			-2.391**	-1.458	-1.264
			(0.887)	(1.080)	(1.171)
population			0.192**	0.171**	0.146*
			(0.0846)	(0.0800)	(0.0771)
income			-1.337***	-1.285***	-1.222***
			(0.389)	(0.380)	(0.378)
non-white			-1.374***	-1.420***	-1.559***
			(0.129)	(0.136)	(0.123)
routine				-0.0417**	0.142*
				(0.0152)	(0.0772)
autoint × routine					-0.0117**
					(0.00424)
Observations	20,524	20,524	20,524	20,524	20,524
R^2	0.32	0.32	0.34	0.34	0.34

Note: Uses citation-weighted patents. Uses only observations until 2009. Citations are adjusted with the Hall et al. (2001) fixed effect method. Driscoll-Kraay standard errors in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1.

Table 2.B.7. Labor Market Effects of Citations-weighted Automation Patents for Manufacturing and Non-manufacturing Employment

	(1)	(2)	(3)	(4)	(5)
A. Outcome: Manufacturing employment-to-population					
autoint	-0.0198	-0.110	-0.262***	-0.292***	0.0443
	(0.0186)	(0.0762)	(0.0373)	(0.0425)	(0.136)
non-autoint		0.0948	0.332***	0.375***	0.323***
		(0.0884)	(0.0418)	(0.0490)	(0.0488)
manufacturing			-2.880***	-2.374***	-2.237***
			(0.536)	(0.590)	(0.613)
population			-0.0310**	-0.0425***	-0.0595***
			(0.0151)	(0.0146)	(0.0169)
income			-0.792***	-0.765***	-0.721***
			(0.169)	(0.162)	(0.169)
non-white			-0.167	-0.192	-0.287
			(0.224)	(0.216)	(0.175)
routine				-0.0228***	0.106*
				(0.00440)	(0.0592)
autoint × routine					-0.00826**
					(0.00320)
Observations	20,520	20,520	20,520	20,520	20,520
R^2	0.19	0.19	0.24	0.24	0.26
B. Outcome: Non-manufacturing employment-to-population					
autoint	0.103***	0.295***	0.315***	0.297***	0.405**
	(0.0313)	(0.0722)	(0.0615)	(0.0705)	(0.157)
non-autoint		-0.202**	-0.258***	-0.233***	-0.250***
		(0.0832)	(0.0741)	(0.0799)	(0.0882)
manufacturing			0.519	0.818	0.861
			(0.673)	(0.776)	(0.829)
population			0.218**	0.211**	0.205**
			(0.0832)	(0.0782)	(0.0757)
income			-0.580*	-0.563	-0.549
			(0.332)	(0.334)	(0.336)
non-white			-1.275***	-1.289***	-1.321***
			(0.110)	(0.105)	(0.0997)
routine				-0.0132	0.0281
				(0.0169)	(0.0413)
autoint × routine					-0.00263
					(0.00296)
Observations	20,529	20,529	20,529	20,529	20,529
R^2	0.26	0.26	0.28	0.28	0.28

Note: Automation and non-automation are citations-weighted. Only uses observations until 2009. Citations are adjusted with the Hall et al. (2001) fixed effect method. All regressions include state and year fixed effects and a constant. Driscoll-Kraay standard errors in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1.

Table 2.B.8. Labor Market Effects of Automation, Five-year Non-overlapping Time Periods

	Outcome: Employment-to-population				
	(1)	(2)	(3)	(4)	(5)
autoint	0.154***	0.324***	0.258**	0.246*	0.611***
	(0.0334)	(0.0892)	(0.126)	(0.134)	(0.162)
non-autoint		-0.173**	-0.0776	-0.0610	-0.0825
		(0.0740)	(0.125)	(0.137)	(0.133)
manufacturing			-1.191*	-1.031*	-0.886
			(0.616)	(0.601)	(0.595)
population			0.107***	0.102***	0.0908***
			(0.0256)	(0.0236)	(0.0241)
income			-0.644***	-0.627***	-0.601***
			(0.228)	(0.224)	(0.223)
non-white			-1.215***	-1.232***	-1.281***
			(0.447)	(0.444)	(0.427)
routine				-0.00751	0.132**
				(0.0136)	(0.0507)
autoint*routine					-0.00969***
					(0.00356)
Observations	5,663	5,663	5,663	5,663	5,663
R^2	0.40	0.40	0.41	0.41	0.41

Note: The table presents fixed effects panel data regressions using non-overlapping five-year equivalent changes in employment as percent of commuting zone population as the dependent variable. *autoint* and *non-autoint* are five-year sums of new automation technology and non-automation technology, as defined in the text. *routine* is the initial percentage of routine tasks in commuting zone employment. Further controls are the initial manufacturing employment share, the log of the initial commuting zone employment, the log of initial per capita income and the initial share of non-white citizens in the population. All regressions include state and year fixed effects and a constant. Standard errors clustered at the state level in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 2.B.9. Labor Market Effects of Automation for Manufacturing and Non-manufacturing Employment

	(1)	(2)	(3)	(4)	(5)
A. Outcome: Manufacturing employment-to-population					
autoint	0.00382	-0.0365*	-0.110***	-0.137***	0.255***
	(0.0102)	(0.0216)	(0.0269)	(0.0297)	(0.0653)
non-autoint		0.0409**	0.164***	0.205***	0.179***
		(0.0187)	(0.0285)	(0.0328)	(0.0390)
manufacturing			-1.588***	-1.209***	-1.055***
			(0.222)	(0.223)	(0.231)
population			-0.00100	-0.0118	-0.0239*
			(0.0117)	(0.0128)	(0.0125)
income			-0.710***	-0.666***	-0.645***
			(0.139)	(0.143)	(0.140)
non-white			-0.149	-0.191	-0.237
			(0.198)	(0.208)	(0.190)
routine				-0.0181***	0.133***
				(0.00402)	(0.0262)
auto*routine					-0.0104***
					(0.00184)
Observations	5,660	5,660	5,660	5,660	5,660
R^2	0.21	0.21	0.23	0.24	0.26
B. Outcome: Non-manufacturing employment-to-population					
autoint	0.137***	0.363***	0.368***	0.382***	0.260
	(0.0317)	(0.0778)	(0.113)	(0.124)	(0.160)
non-autoint		-0.230***	-0.253**	-0.274**	-0.267**
		(0.0612)	(0.104)	(0.121)	(0.123)
manufacturing			0.321	0.115	0.0700
			(0.425)	(0.419)	(0.424)
population			0.105***	0.111***	0.114***
			(0.0246)	(0.0207)	(0.0210)
routine				0.00957	-0.0371
				(0.0138)	(0.0390)
autoint*routine					0.00324
					(0.00239)
income			0.0393	0.0176	0.0101
			(0.172)	(0.178)	(0.176)
non-white			-1.032***	-1.013***	-0.992***
			(0.229)	(0.219)	(0.219)
Observations	5,662	5,662	5,662	5,662	5,662
R^2	0.36	0.36	0.37	0.37	0.37

Note: Uses non-overlapping five-year equivalent changes. All regressions include state and year fixed effects and a constant. Standard errors clustered at the state level in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1.

3

Patterns of Panic: Financial Crisis Language in Historical Newspapers

3.1 Introduction

To analyze the causes and consequences of financial crises, we need to know when they occured. But while the 1873, 1907 or 1929 financial crises are well-documented, many smaller banking panics and liquidity crunches in the 19th and 20th century are less well known and difficult to classify from traditional sources. For many questions which we would like to explore, it is more relevant what people at the time thought was happening on financial markets rather than what we perceive with hindsight.

Newspaper archives are a rich source of narrative history that can help us understand and quantify our recent financial history. This paper uses 35 million titles of five major U.S. newspapers since 1889 which can be used to measure financial sentiment. The new dataset complements existing indicators of financial stress, such as “narrative” accounts (Bordo and Meissner, 2016; Laeven and Valencia, 2012; Schularick and Taylor, 2012), the interplay of financial variables (Baron et al., 2018; von Hagen and Ho, 2007; Gilchrist and Zakrajšek, 2012) or the behavior of credit aggregates (Rancière et al., 2008). Long-run historical text archives offer a new path to measure beliefs and opinions about financial markets.

Text has long been used as quantitative data in macroeconomics and finance. Niederhoffer (1971) and Cutler et al. (1989) investigate the intersection of news stories and financial markets, but these early attempts were limited in the scope of textual data they could handle. Antweiler and Frank (2004) and Tetlock (2007) use better algorithms and faster computers to explore much larger text corpora. However, the time coverage of these studies is short. Instead, this project offers a novel both long-run and high-frequency measure of financial stress. To construct this new dataset, I analyze the universe of published newspaper titles by five major US newspapers (*Boston Globe*, *Chicago Tribune*, *Los Angeles Times*, *Wall Street Journal* and *Washington Post*) from when they were first published. Combining these newspapers provides a rich temporal, spatial, political and topical coverage of newspaper language.

The new indicator has several desirable properties. It is based on a number of newspapers to smooth out particularities of individual media outlets which reduces noise in the final indicator. It is based on text archives and does not use web search masks which limit what other papers in this literature could do. The uninterrupted publication of newspapers mean that such a series can be generated throughout the World Wars, periods when other macroeconomic variables are unreliable or difficult to obtain. Newspaper reporting is available on a daily basis, a high frequency that is rare for historical macroeconomic and even financial series. They also cover weekends and bank holidays and so they offer a deep and high-frequent historical coverage even when financial markets were closed. And whereas previous indicators were often binary, I provide a continuous indicator.

Alexopoulos and Cohen (2009) and Baker et al. (2016) follow a similar route, but they search for “economic policy uncertainty” which is a different, more abstract and more vague idea and so arguably harder to identify in text documents. The financial market is a more precise concept which makes it easier to find set of keywords to describe it. Also, newspapers write regularly and reliably about financial markets, so it is a plausible concept to monitor over time. A similar argument applies to Caldara and Iacoviello (2016) who create an indicator of geopolitical risk from newspaper article counts. Manela and Moreira (2015) and García (2013) use newspaper language as a measure of financial stress, but both are limited to one newspaper. O’Connor et al. (2010) show that sentiment measured on tweets correlates strongly with consumer sentiment and Bollen et al. (2011) provide evidence that moods on Twitter predict stock market performance. Jalil (2015) also builds on historical newspapers to create a new narrative indicator for financial panics for the United States before 1929.

I proceed in three steps. First, I describe the new dataset and explain how to measure financial sentiment. I select all articles that are concerned with financial markets using a broad dictionary of 120 words and show results are robust to altering this dictionary. I study the emotional content of newspaper titles by counting words with positive and negative connotations based on four sentiment lexicons including a total of 11,407 words with positive or negative connotations. In this way, I construct a new daily measurement of how financial markets were covered from 1889 to 2016. I create four different indicators, one for each of the sentiment lexicons. Surprisingly, all four series strongly comove, a result that is not driven by overlap between the dictionaries. To create the main financial stress indicator, I normalize the individual newspaper series and average across them. Results are unchanged when using alternative ways to create the indicator through removing newspaper-specific trends or eliminating newspapers from the sample.

Second, I validate the indicator using a third-party data source by the professional media analysis company Media Tenor AG. Trained experts read and labeled 23,000 articles in the *Wall Street Journal*. A comparison of these articles with the new indicator shows that my approach reliably identifies days with strong coverage of negative financial market reporting. This dataset covers only a subset of years and newspapers, but its large size and high level of detail allow for an in-depth comparison with the new indicator.

Third, I analyze what happened in the US economy during and after times of increased negative financial sentiment. Local projections in quarterly and monthly frequencies provide estimates of how other macroeconomic variables behave after an increase in financial stress. Spikes in my indicator are followed by lower economic activity, rising unemployment, lower stock market returns and increases in corporate bond spreads.

3.2 Data

The dataset contains the titles of all newspaper articles of five U.S. American newspapers which are the *Boston Globe*, *Chicago Tribune*, *Los Angeles Times*, *Wall Street Journal* and *Washington Post*. All five newspapers were founded in the 19th century and have been published continuously since. Table 3.1 shows the first dates of publication for these newspapers, which range from 1853 for the *Chicago Tribune* to 1889 for the *Wall Street Journal*.

Table 3.1. Summary of Newspaper Data

Name	Since	Titles (total)	Financial titles	Titles (/ day)	Words	Unique words
Chicago Tribune	1853	9.0m	4.0%	154	36m	380k
Boston Globe	1872	6.7m	3.5%	129	26m	370k
Washington Post	1877	7.7m	4.3%	153	31m	340k
Los Angeles Times	1881	7.9m	4.8%	164	40m	315k
Wall Street Journal	1889	3.9m	19.5%	107	18m	170k

I only keep articles, editorial articles and front matter. This excludes adverts, obituaries or the weather, as these are unlikely to be of relevance to the analysis in this paper. Also, I only have access to the *titles* of newspaper articles. Titles are written to capture the gist of an article and - most of the time - convey a strong sense of what articles are about. Given the very extensive and comprehensive nature of this dataset, it is therefore likely that I capture trends in reporting well.

These five newspapers have commonly been used in the literature, such as by Baker et al. (2016) and Caldara and Iacoviello (2016). They cover different regions, have other political alignments and emphasize different topics in their reporting. A typical number of articles for a newspaper is about 30,000 to 60,000 per year and between 100 and 200 per day. Figure 3.1 shows how the number of titles per newspaper changed over the years. Trends are quite distinct for each newspaper which is to be expected as the number of titles will vary with stylistic choices, the size of the readership, a newspaper's commercial success, business strategy and changes in the media landscape such as the shift to online publishing. While the *Wall Street Journal* contained about 30,000 articles per year, the *Chicago Tribune* published about 60,000 articles per year. The number of articles in the *Los Angeles Times* grew steadily until the early 1990s and has declined since. For newspaper language to become interpretable, it is therefore necessary to normalize article counts by newspaper length.

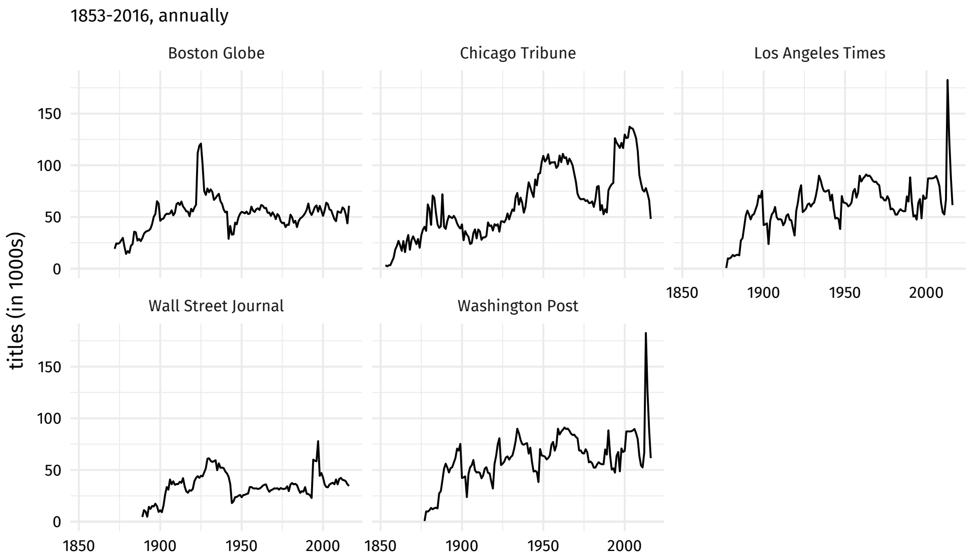


Figure 3.1. Number of Titles per Newspaper

While the *Wall Street Journal* has the lowest total number of titles and the lowest number of titles per day, the ratio of hard news (politics, economics) to soft news (sport, crime) is higher for this newspaper. In the next section, I will explain how I classify *financial* titles. About a fifth of titles in the *Wall Street Journal* are concerned with financial markets, while this percentage is only between 3.5 and 4.8 percent for the other newspapers. The total number of written words (also called tokens) per newspaper is between 18 and 40 million and the size of the total dictionary of unique words that these newspapers used is between 170 and 380 thousand.

Newspaper titles have become longer over time. Figure 3.A.1 plots the average number of words per title. The average length of titles doubled, from about four words to about nine words per title. Newspapers further provide a dense news coverage throughout the week. The five newspapers are published on all days of the week, with the exception of the *Wall Street Journal* which does not appear on Sundays. It is thus another attraction of highly-frequent newspaper data that it is possible to construct indicators covering all days of the week, even when stock markets are closed.

3.3 Measuring Financial Sentiment

In the previous section I described the newspaper data and explained the structure of the data. I now turn to measuring financial sentiment in the titles. This involves two steps: I first find out *what* is being reported and then I measure *how* it is reported.

3.3.1 Sentiment Dictionaries

The main benefit of having the full corpus of titles at hand – as opposed to searching for keywords through online databases – is that it allows to search for much larger lists of words.[2] These compilations are called "dictionaries" or "lexicons" and have the purpose of providing an easy way to measure emotional content in texts. The dictionaries are compiled by a mix of computational linguists and economists and rate unigrams (individual words) as being of positive or negative connotation.

I use of four sentiment dictionaries which are provided by different researchers and have subtly different purposes. The *NRC* (National Research Council Canada) dictionary by S. Mohammad and P. Turney (2010) and S. M. Mohammad and P. D. Turney (2013) is not specific to any domain but is meant to be quite general. The *bing* lexicon by Hu and Liu (2004) and Liu et al. (2005) serves primarily to evaluate customer reviews. Notowidigdo (2011) constructs the *AFINN* (Affective Norms for English Words) with the purpose of measuring language on microblogs and tweets. The dictionary by Loughran and McDonald (2011) is the one most tailored to measuring emotional connotations in business reports, press briefings and has also been previously applied to the financial press (e.g. by García, 2013). Only *AFINN* assigns scores from -5 to +5 to words and the other three dictionaries sort words into simple "positive" and "negative" categories.

While the Loughran and McDonald (2011) dictionary may be the one most suitable for the purpose of this paper, there are good reasons to measure sentiment with different dictionaries. It serves as a strong robustness check to test if the constructed series are similar and I combine the results from the four dictionaries to create one baseline indicator. Of the total of 11,407 unique words that are contained in one of the four dictionaries, 8,229 only show up in one of them. 2,633 words belong to two dictionaries, 542 are contained in three dictionaries and only three words ("abundance", "confess" and "unexpected") show up in all four. As Table 3.2 shows, between 28% and 61% of words are unique to any

[2] Baker et al. (2016) and Caldara and Iacoviello (2016) search for short lists of words through web-based masks which limits the number of phrases one can include. Gentzkow and Shapiro (2010) search for a large number of words and phrases, but do this only for one year of newspaper data. Tetlock (2007) analyzes one column in the *Wall Street Journal* from 1984 to 1999 with dictionary methods. Manela and Moreira (2015) similarly examines titles and first paragraphs of the same newspaper using a machine learning model.

Table 3.2. Lexicon Statistics

	nrc	bing	loughran	AFINN
		positive		
Words	2312	2006	354	878
Unique	61%	51%	33%	34%
Cos. sim. with *loughran*	0.12	0.23	1	0.21
		negative		
Words	3324	4782	2355	1597
Unique	39%	51%	49%	28%
Cos. sim. with *loughran*	0.22	0.25	1	0.24

Note: For AFINN, assigns words to "negative" and "positive" depending on score sign. "Unique" refers to the share of articles that only show up in this one lexicon. The respective third rows report cosine similarity with the Loughran and McDonald (2011) dictionary.

of these dictionaries, so they may capture different moods. Table 3.2 also displays the number of positive and negative words in the dictionaries. The "bing" dictionary is particularly sizable, while the "loughran" dictionary contains few positive words.

Another way to compare documents is through their *cosine similarity*. Cosine similarity is an established indicator of comovement in high-dimensional spaces and is the non-centered normal (Pearson's) correlation coefficient and ranges from 0 to 1 for positive count data. As Table 3.2 shows, the *nrc*, *bing* and *AFINN* dictionary have low similarities with the the *loughran* dictionary. The *loughran* might be different from the others dictionaries due to its specific purpose of measuring *financial* sentiment, but the small overlap also shows the value of using different dictionaries for this analysis.

Simply applying these dictionaries to newspaper titles produces misleading trends. Large sections of contemporary newspapers cover topics such as sport, cooking or health. The topic "crime" is particularly prevalent in newspapers. Consider the words: "dead", "death", "dies", "fight", "fire", "killed", "police" and "shot". These are classified by at least one dictionary as negative and are related to violent crime. The eight terms show up in 3.5% of titles in the *Boston Globe*, 5.5% of titles in the *Chicago Tribune*, 4.0% of titles in the *LA Times*, 0.7% of titles in the *Wall Street Journal* and 4.0% of titles in the *Washington Post*. This shows the importance of first selecting only those article titles that are concerned with financial markets.

3.3.2 Finding Articles about Financial Markets

I select beforehand a range of words about financial markets. It is one of the advantages of searching for a precise concept such as reporting about the financial market that picking words to describe it is relatively easy.

I liberally assemble a large list of 120 words and phrases that signal that a title is about economics, financial markets or business more generally. I base these on words from several content areas (or topics) and usually add their singular and plural forms and some alternative word forms and combinations. Here are some examples of the words in these topics:[3]

- *bonds*: “bonds”, “credit”, “debt”, “loan”, “mortgage”
- *business*: “commerce”, “entrepreneur”, “industry”, “profits”
- *central banks*: “boj”, “bernanke”, “bundesbank”, “central bank”,“currency”, “dollar”, “draghi”, “ecb”, “fed”, “franc”, “money supply”, “reichsbank”, ““volcker”, “yellen”, “yen”
- *economy*: “consumers”, “economy”, “production”, “purchases”
- *general*: “banks”, “financial markets”, “fund”, “interest rates”, “losses”, “wall street”
- *gold, silver*: “bullion”, “coin”, “gold", “silver”
- *inflation*: “consumer prices”, “cpi”, “deflated”, “inflation”, “producer prices”
- *railroads*: “railroad”
- *stocks*: “capital markets”, “dividends”, “equities”, “stocks”
- *trade*: “embargo”, “exports”, “imports”, “tariff”
- *trouble*: “bail out”, “credit crunch”, “market declines”, “stress test”

Table 3.3 shows the number of words per topic, the total number of titles that contain a word from this topic and the total share of financial articles that is due to a topic. The topic “general” is the most important one and accounts for almost a fourth of financial articles.

[3] The full list is in Appendix Section 3.B.

Table 3.3. Importance of Topics

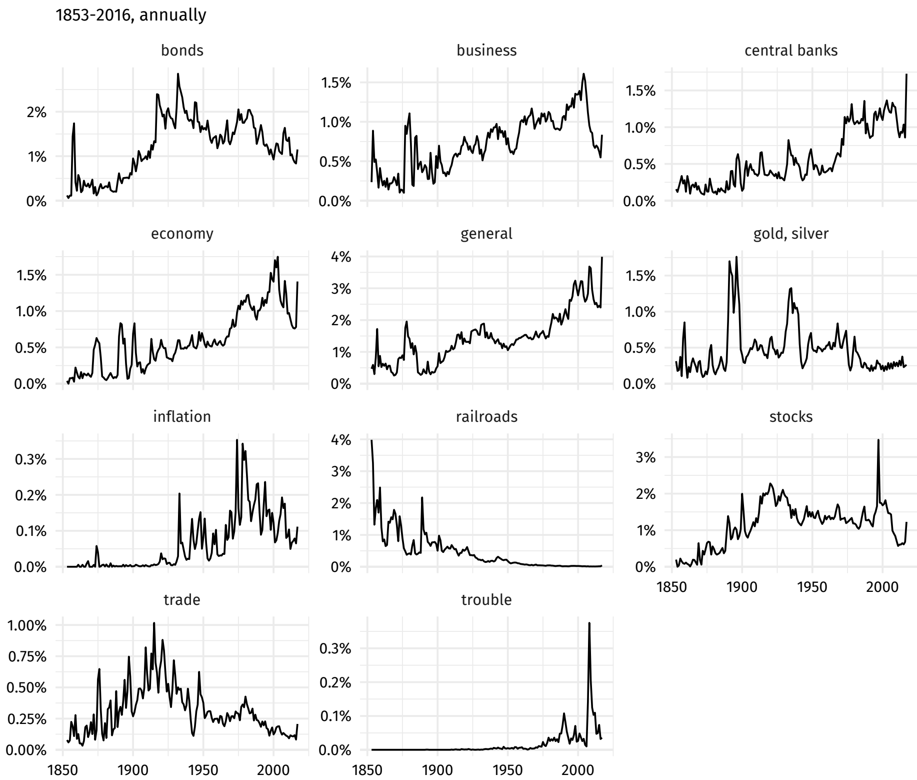
Topic	Words	Titles	Share
general	30	501k	23%
bonds	8	398k	18%
stocks	7	336k	15%
business	9	240k	11%
economy	12	216k	10%
central banks	24	201k	9.8%
gold, silver	5	142k	6.4%
trade	10	92k	4.2%
railroads	1	58k	2.6%
inflation	8	25k	1.1%
trouble	8	7k	0.3%
total	**122**	**2200k**	**100%**

Figure 3.2 plots the occurence of these topics over time. Differences between the topics are not completely sharp. The topics “economy” and “business” tend to move together, for example. Other topics are related concepts, such as “central banks” and “inflation”. In general, however, these topics cover most relevant areas of financial markets and capture articles that we would normally consider to be concerned with a broad notion of finance. Railroad companies and their earnings and dividends were an important asset class before WWI (Homer and Sylla, 2005) and the financial press of those times reflects that. Price movements in gold and silver were more central to financial markets and the global monetary system, but these monetary commodities have – since the end of the Bretton Woods system in the 1970s – lost in importance. Instead, reporting about the actions of central banks has become more prevalent since the 1980s. Trade and in particular the term “tariffs” also used to hold greater prominence in financial reporting.[4]

The point of searching for this large list of words is to be very inclusive in which articles to count as finance. But choosing these words by hand raises the question of whether the new indicator is sensitive to exactly which words are included.

This is not the case, as the occurence of financial terms in titles is highly skewed. Counting all titles that contain at least one of the words in Section 3.B returns about 2.2 million titles. Only ten words account for 41 percent of these financial articles. These words are: “stock(s)”, “bond(s)”, “fund(s)”, “loss”, “gold”,

[4]Baker et al. (2016) also include “tariffs” in their group of words to measure historical economic policy uncertainty.



Note: Figure shows averages across the five newspapers. Values are shares of titles containing a word from a specific topic.

Figure 3.2. Financial Topics in Newspapers

Table 3.4. Changing Financial Dictionary

Number of words	10	20	50
Share finance titles	41%	62%	92%
Correlations with baseline	0.81-0.95	0.90-0.98	>0.99

Note: Shows what happens when restricting the size of the dictionary used to classify titles as being about financial articles to the x most frequent words. See text.

"profit", "industry". Adding the next 10 most important words ("loan", "banks", "railroad", "financial", "silver", "finance", "economy", "investors", "credit" and "dividend") raises this percentage to 62 percent. I also repeat all steps to construct the baseline indicator (of negative financial sentiment) and compare the resulting quarterly time series. The second row in Table 3.4 shows the range of resulting correlations when using the four different sentiment dictionaries. The

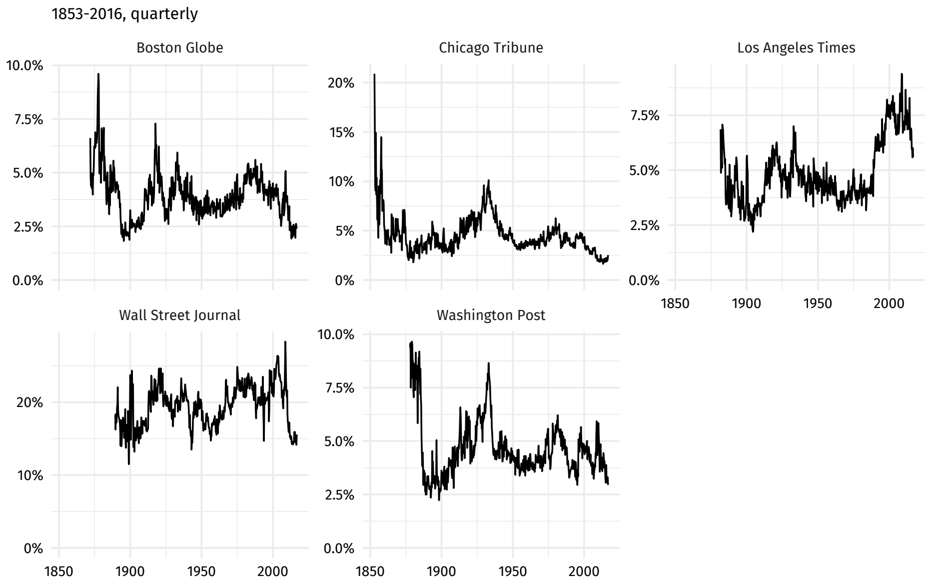


Figure 3.3. Articles about Financial Markets

Note: Financial articles as share of all articles.

series constructed from the dictionary using only 10 words has the smallest correlation of 0.81 (for the *nrc* dictionary). Correlations are very high throughout and the two indicators are indiscernible using the 50 most frequent financial terms.

This means that picking a larger and more inclusive dictionary is unlikely to change results much, but this approach catches some articles that we would plausibly consider to be about financial markets. The term “reichsmark”, for example, only shows up in 11 titles, but such titles are nonetheless a good candidate to be concerned with finance.

Figure 3.3 plots the share of articles about financial markets by newspaper. The data start when a newspaper was founded, which is why there are about forty years more data for the *Chicago Tribune* than for the *Wall Street Journal*. A much larger share of articles (about 20 percent) in the *Wall Street Journal* is about financial markets than in the four other newspapers (where this share is about 7 percent). The *Los Angeles Times* started reporting more on financial markets in the 1990s. The *Boston Globe* has the lowest overall share of reporting on financial markets of about 5 percent.

Reporting on financial markets increased during the Great Depression, the 2001 dotcom bubble and, especially for the *Wall Street Journal*, during the heat of the financial crisis 2008. The earlier series are more volatile. This might be due to the shorter titles which also make finding a good financial language dictionary more difficult. This also shows that the indicator becomes most trustworthy when I start averaging sentiment values across all five newspapers in 1889.

3.3.3 Measuring Emotional Content of Articles

After identifying which articles are concerned with financial markets, the next goal is to measure the tone of those titles. Sometimes, measuring sentiment from newspaper titles is easy. Take for example the following title in the *Wall Street Journal* on February 10th 2017:

> “Equities: Tax-Cut Talk Lifts Stocks"

This would likely be considered good news for financial markets. Consider another article from the *Wall Street Journal*, this one from March 3rd 1938:

> “Sugar Market Declines Following Publication Of [the Secretary of Agriculture Henry A.] Wallace’s Views"

This title conveys the news that the price of sugar is falling and might be considered to have neutral or negative tone. One could theoretically carry out such a classification by hand and measure the tone in articles. But to do such an analysis at scale, I use of the previously-described sentiment dictionaries.

I count the number of negative and positive words for every article on financial markets and sort articles into “positive”, “negative” and “neutral” financial language depending on their net sentiment score. For example, if a title contains two negative and one positive words, I count it as negative.[5] Figure 3.4 shows the share of newspaper articles that are both about financial markets and have net negative sentiment. For all newspapers, the four series are strongly correlated with all pairwise correlations between 0.44 and 0.96 (mean: 0.83, median: 0.87).[6] Correlations are higher for the *Wall Street Journal* than for the *Boston Globe* which might be caused by the much higher share of financial reporting in the former newspaper. Each series was constructed using a different sentiment dictionary and their strong comovement is a powerful robustness check for this method and a further sign of the plausibility of measuring financial stress using newspapers. The previous section also showed that overlap between the dictionaries is low, so the correlation does not arise mechanically.

---

[5] The *AFINN* dictionary uses numericals scores per word, so there I calculate a net sentiment score and then count titles as positive or negative.
[6] See Table 3.A.1.

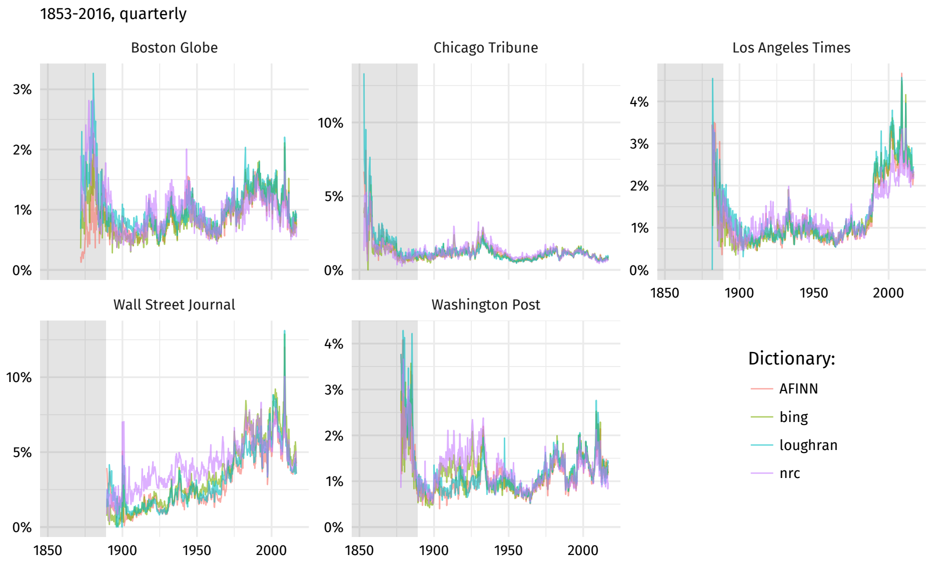


Figure 3.4. Negative Financial Articles

Note: Calculated as all newspaper articles that are both about financial markets and have net negative sentiment divided by all articles in that quarter.

For the final new indicator, I normalize values to mean 100 and unit standard deviation and average across newspapers and sentiment dictionary series. Alternatively, one could weigh newspapers by their circulation numbers (such as Doms and Morin, 2004) or by some other measure of importance or reliability, but this is not feasible due to the long time span of the data considered in this paper. The *Wall Street Journal* is first published in 1889, so from this point onwards we can be most confident about the quality of the new financial sentiment indicator.

Figure 3.5 plots this key new indicator. Spikes occur around known episodes of financial stress, such as the panics of 1893, the land boom and crash in Florida in 1926, the crash of 1929, the spring of 1933, Black Monday in 1987, the Asian Financial Crisis of 1997, the stock market crash and recession of 2001 onwards, the European Sovereign Debt Crisis of 2011 and “shutdown" of the US government in 2015. The indicator takes its highest value during the fall of 2008 when Lehman Brothers went bankrupt and the US and global financial systems were on the verge of collapse. It stayed highly elevated in spring 2009.

Trends in neutral and positive articles about financial markets are also insightful. Figure 3.A.2 shows the share of neutral and positive financial titles across newspapers and Figure 3.A.3 shows the normalized mean across newspapers. Neutral reporting rose initially and then peaked during WWI and the Great Depression

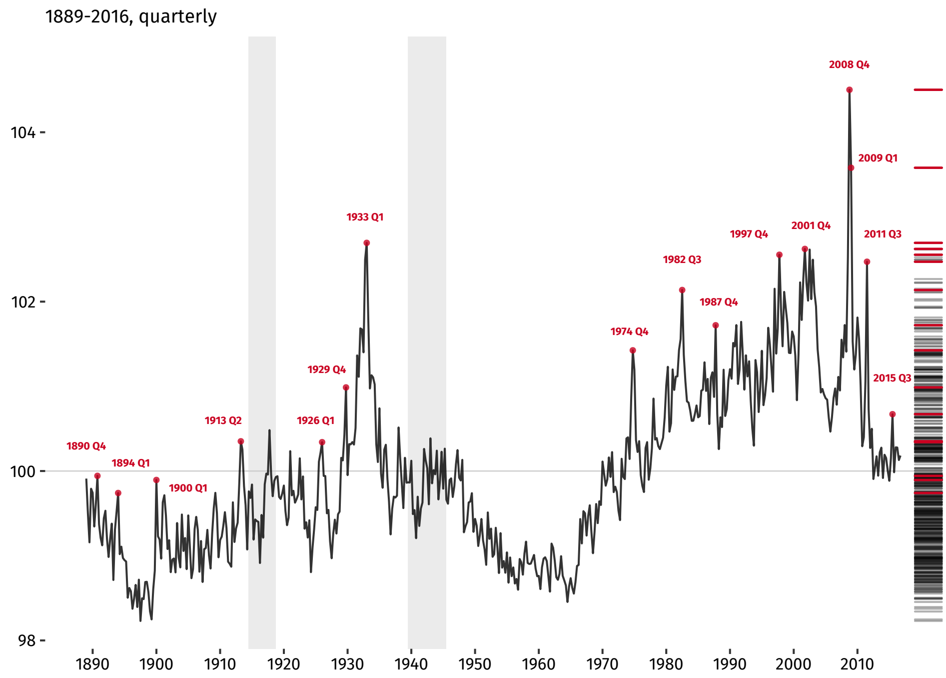


Figure 3.5. New Financial Stress Indicator

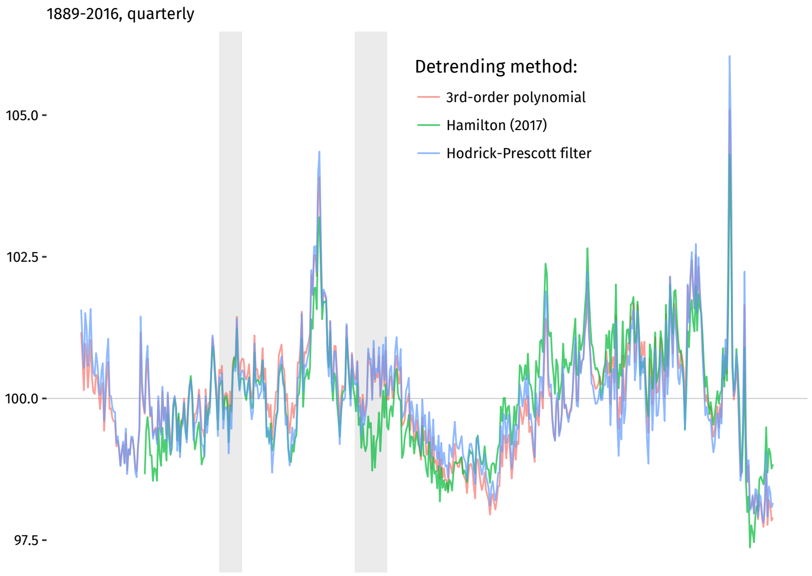
Note: Calculated as normalized share all newspaper articles that are both about financial markets and have net negative sentiment divided by all articles in that quarter. Figure shows averages across the five newspapers and across sentiment dictionaries. Gray bars mark the world wars.

and has trended downwards since. Positive financial reporting was high during the 1920s, the late 1990s and the 2000s before the financial crisis. This may be a sign of the "exuberant" years on financial markets, when the economy was booming and stock prices were rising.

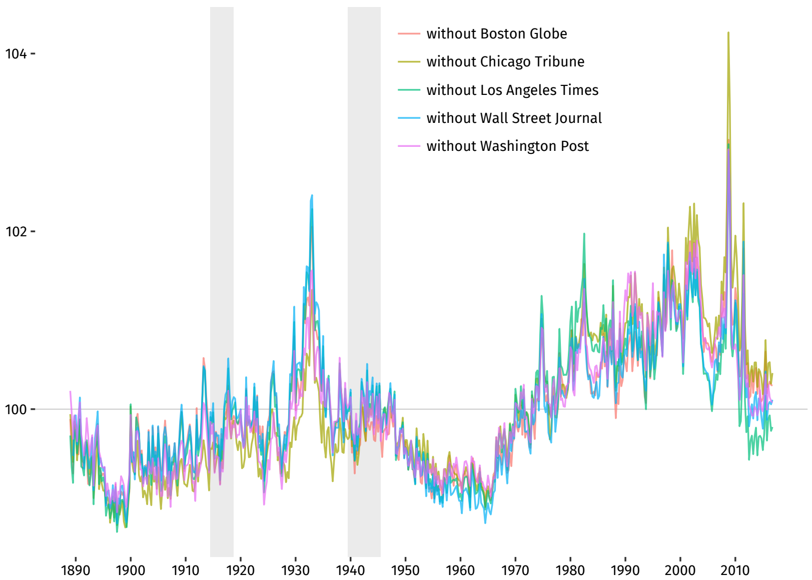
The period between the end of WW2 and the starting financial globalization in the 1970s was a period of calm on financial markets. Strong financial regulations and tight capital controls led to low levels of credit and leverage (Jordà et al., 2016) and almost an elimination of financial crises (Reinhart and Rogoff, 2009).

Newspapers mirror these calm financial conditions. All newspapers in the sample – apart from the *Wall Street Journal* – talked less about financial markets in general (Figure 3.4). And when they reported on finance, they wrote to similar degrees as before in neutral and positive terms (Figure 3.A.3), but used less negative language (Figure 3.5).

It is surprising how much negative financial reporting has fallen since the financial crisis of 2008. After the last strong spike in 2011, the indicator has had an average value of 100, about the level of the 1970s or late 1930s.



(a) Detrending on newspaper level



(b) Eliminating individual newspapers from sample

Figure 3.6. Alternative Indicators

Note: Shows indicator from Figure 3.5, constructed in some alternative ways by detrending or by eliminating one newspaper from the sample. Gray bars mark the world wars.

3.3.4 Alternative Indicators

As Figure 3.4 shows, there are long-run trends in reporting that are idiosyncratic to individual newspapers. Those could be driven by editorial policies or style changes. It is therefore a worry that the new indicator in Figure 3.5 might be dominated by these long-run trends but that the more relevant variation lies in the short run spikes around the trend.

I therefore also construct three alternative indicators that detrend the series in Figure 3.4 on the level of the individual newspapers. The first detrending uses a third-order polynomial regression in time of the form, $\mathrm{fs}_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \varepsilon_t$, where fs_t is the financial sentiment indicator in quarter t. The detrended new indicator then becomes the residual ε_t. The second detrending method is a very slow-moving Hodrick-Prescott filter with a λ parameter of 5×10^6 (5×10^7 for the monthly indicator). Third, I apply the detrending procedure proposed by Hamilton (2017) using a cycle length of 10 years. This last method uses the first 10 years of data to estimate the trend which truncates the beginning of the indicator by a decade.

An additional concern is that the very distinct movements in the newspaper series in Figure 3.4 might mean that some newspapers drive the final indicator much more than others. I therefore exclude one newspaper at a time and and create separate indicators from the remaining four newspapers.

Figure 3.6 visualizes these alternative indicator series. Neither taking out newspaper-specific trends nor eliminating individual newspapers from the sample alters the indicator much. The shape of the indicator is thus robust to alternative ways of creating it. Table 3.A.2 shows that correlations of the detrended indicators are above 0.72 for the quarterly and above 0.63 for the monthly baseline indicator. After taking out one of the newspapers, indicators still have great resemblance with the version using all newspapers and all correlations are above 0.96 for the quarterly and the monthly data. As expected, correlations with the baseline fall slightly when changing both detrending and eliminating some newspapers at the same time: The lowest correlation coefficient is attained when taking out an HP trend and dropping the *Chicago Tribune* from the sample which reduces the correlation to 0.59 for quarterly and 0.55 for monthly values.

I conclude that changes to the construction of the indicator do not affect it much. This is a further sign that the procedure picks up the underlying signal well and is not an artifact of arbitrary choices made in its construction.

3.4 Testing the New Indicator

After constructing the new financial sentiment dataset this section validates the new dataset with a large sample of manually coded newspaper articles and compares the new indicator to established measures of financial stress.

3.4.1 Comparison with Manually Coded Articles

The private media analysis company Media Tenor provides 23,316 manually coded articles from 2007-2016 from the *Wall Street Journal*. Expert human coders read whole newspaper articles and label articles according to a structured set of questions. The main purpose of this dataset is to measure the description of economic conditions in the United States and other countries. The coders also classify the topic of articles, e.g., “financial markets", “politics" or “labor markets". The dataset also indicates the tone of the article (good, neutral, bad) and whether the article is concerned with the past, present or future. Table 3.5 shows some statistics about this dataset. The data is split in three samples with about 7,000 to 8,000 articles in each. The earliest article is from January 2007 and the last from August 2016. Unfortunately, the dataset does not cover the height of the financial crisis starting in October 2008, as the first sample stops in June 2008 and the second only begins in July 2012. Most reporting is about the present (83%) which is a useful validation of the approach taken in this paper and in the literature. A negative tone prevails in economics and finance reporting with a negative tone in 44% of articles and a positive tone in only 15% of articles. An extensive literature reports that negative incidents are more likely to be reported.[7]

The largest topic is “business cycles” and the most relevant category “financial market” makes up only 8 percent of articles. The lack of coverage of sports, health, culture, entertainment and crime and the low volume of reporting on politics shows that Media Tenor selected only articles that were related to economics and business, broadly defined. This should not be an issue for the comparison I carry out here, as these are precisely the topics that my method using sentiment dictionaries is also trying to capture. While I cannot use the Media Tenor data to analyze trends in overall reporting on the economy and on business, I can use it to measure the distribution of more detailed topics within economics and business reporting.

Ideally, I would match the Media Tenor data with my data on the newspaper article level. I could then compare whether mine and Media Tenor’s classifications agree for every article. However, I do not have unified identifiers (such as article IDs) or newspaper titles (in the Media Tenor sample). I can, however, compare the time trends from the new indicator series with those I calculate from the Media Tenor data. I therefore define the number of articles in the Media Tenor sample that are of a specific topic and have a negative connotation, which is analogous to how I constructed the indicator.

[7] See Gieber (1955), Combs and Slovic (1979), Bohle (1986), **eisensee2007news** and Miller and Albert (2015).

Table 3.5. Summary of *Wall Street Journal* Coded Sample

	Articles	Share
Period		
Jan 2007–Jun 2008	7897	33%
Jul 2012–May 2013	7106	30%
Jan 2015–Aug 2016	8313	36%
Tonality		
positive	3574	15%
neutral	9512	41%
bad	10230	44%
Time reference		
past	2031	9%
present	19338	83%
future	1947	8%

Source: Media Tenor

	Articles	Share
Topic		
business cycle	8637	37%
economic growth	3538	15%
economic policy	2630	11%
financial markets	2376	8%
monetary policy	1866	8%
fiscal policy	1735	7%
labor market	1116	5%
global economies	1042	4%
politics	231	1%
other	105	0.5%
business	40	0.2%
Section		
business/economy	13117	56%
politics/news (page 2-)	5157	22%
politics/news (page 1)	3031	13%
finance/markets	1075	5%
other (13 more)	936	4%
total	**23316**	**100%**

I use a regression to investigate the comovement of these two series:

$$m_t = \gamma_w + \delta_p + \beta f_t + \varepsilon_t \,. \tag{3.1}$$

Here, m_t are the Media Tenor topic trends on day t, β is the regression coefficient of interest and f_t is my new indicator series (share of financial articles in the *Wall Street Journal* with a negative sentiment). ε_t is the error term. Newspaper language may vary depending on the day of the week, as weekdays might see more news or the publishing of statistics on special days might drive both markets and reporting. I therefore include a dummy indicating the day of the week γ_w. The three samples cover quite different time periods and trends in reporting and the economic situation might show a correlation where there are just joint trends for other reasons. Equation (3.1) therefore also includes a dummy for the three samples δ_p.

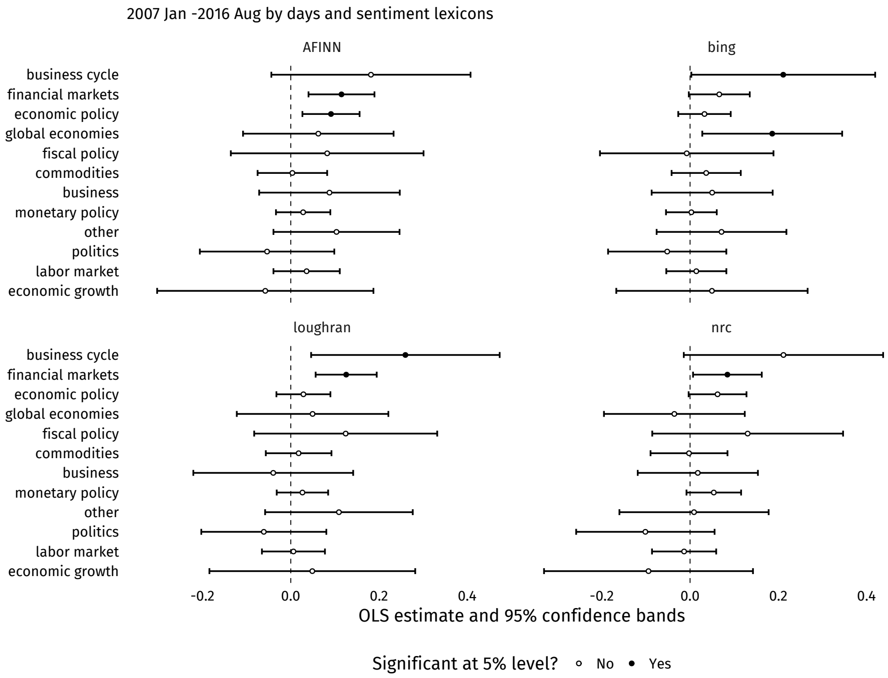


Figure 3.7. Comparison with Media Tenor Sample

Note: Compares the measurement of titles in the *Wall Street Journal* described in text with sample of manually classified articles by Media Tenor. Figure shows the coefficient of a regression of the number of articles with negative connotation per day on the number of negative titles about financial markets identified by one of four sentiment dictionaries. The regression also includes dummies for the day of the week and for the three samples. Table 3.A.3 contains the values.

I estimate this regression model for all Media Tenor topics and using the four indicator series constructed from the different sentiment dictionaries. Figure 3.7 visualizes the regression coefficients and confidence bands (and values are in Table 3.A.3). The topic "financial markets" with a negative tone occurs more often on precisely those days when my derived indicator series of negative financial language is larger. For the *bing* sentiment dictionary, the coefficient is only significant at the 10 percent level. The topic "business cycle" also appears more often on those days that see a high indicator value, but the coefficient is less precisely estimated. The topic "economic policy" is positively associated with the indicator only for the *AFINN* dictionary. The *bing* dictionary stands out in that "global economies" (a topic about worldwide economic trends and foreign economic developments) shows significant comovement.

Overall, this is encouraging. On days where the new indicator was greater, human experts also consider more articles to report negatively on financial markets.

Figure 3.8. Comparing Monthly Financial Stress Measures

Note: Lighter colors signal earlier dates. Monthly data comparing the new indicator to the historical US Economic Policy Uncertainty indicator from Baker et al. (2016), the credit spread index from Gilchrist and Zakrajšek (2012), the risk spread (Fred codes BAA and GS10) and the VIX. The areas surrounding the linear regression line show the 95% confidence bands.

The high quality third-party data source and great level of detail in this dataset allows for a very thorough comparison with my dataset. However, a limitation of this validation exercise is that that it only works with articles from one newspapers and near the end of my sample. We should thus be careful not to infer too much about the reliability with regards to other newspapers and earlier years.

3.4.2 Comparison with Previous Financial Stress Indicators

There are several other indicators that are commonly used to measure financial stress. Figure 3.8 plots the monthly share of articles with negative financial language against four other variables that are indicative of different facets of financial stress. The long time span and high frequency of the new newspaper-based dataset makes comparisons with these other series particularly fruitful. I find strong positive correlations (between 0.59 and 0.74) in all cases.

The "Economic Policy Uncertainty" (EPU) indicator by Baker et al. (2016) shows a firm positive correlation with the new indicator. The colors in the figures are lighter for earlier dates which cluster farther down the figure. This is a sign that the association between the two variables was flatter in the earlier years. A reason might be that the EPU is more volatile even at the start of the joint sample since 1900.

A number of episodes stand out. Both measures take extreme values during the fall of 2008 which was the height of the financial crisis. The two variables also jump in late 2001, when the 9/11 attacks led to a political crisis and losses on financial markets. The new financial stress indicator took a high value in October 1929, but the EPU remained low. The EPU reacts more strongly to wars, such as the beginning and end of WWI.

The other three panels of Figure 3.8 show how the new indicator relates to three other common indicators of financial stress. My indicator jumps earlier than the other series. The indicator reached a new all-time-high already in September 2008. The other series stay elevated for a longer period and remain at high levels into the spring of 2009. This might mirror the nature of media reporting, as facts are reported when they are new.

3.5 Financial Sentiment and the Economy

This section compares the new financial stress measure to events taking place in the economy at the same time. To analyze how the new financial crisis indicator comoves with variables in the aggregate economy, I use local projections – a type of multivariate time series regression. This approach has also been followed by other papers creating new newspaper indices, such as Baker et al. (2016) and Caldara and Iacoviello (2016).

Indicators created from newspaper language are not exogenous to developments in the economy. Instead they measure the media's response to what was happening in the economy and on financial markets. Nevertheless, providing concise summaries of the joint comovement of new indicators and economic variables is insightful. It is likely that newspapers reported more negatively on the economy during and after times of increased financial stress. This does not answer the question of what led to the increase in financial stress or what the economic fallout of an increase in the cost of financial intermediation is. It, however, further validates the new indicator series and allows for a comparison with other financial stress measures.

For this, I estimate local projections as introduced by Jordà (2005). Local projections have become a common tool, as they allow easier modeling of non-linearities, are more robust to misspecified models and can be easily carried out using single-equation estimations. They do not, however, allow to specify cross-equation restrictions which are useful when using vector autoregressions. But given that this is not what I do in this paper, local projections are the right modeling choice.

I present both a quarterly and a monthly analysis, which differ in the periods they cover and in the variables they include.

3.5.1 Quarterly Local Projections

Using quarterly data allows to use data for real GDP and the unemployment rate constructed by Ramey and Zubairy (2017). These authors provide quarterly economic data starting in 1890.

I fit the following regression to the data:

$$
\begin{aligned}
y_{t+k} - y_{t-1} \;=\; & \alpha_k + \beta_k f_t + \sum_{i=1}^{I} \beta_{k,i} f_{t-i} \\
& + \sum_{i=1}^{I} \gamma_{k,i}\, y_{t-i} + \sum_{i=1}^{I} \delta_{k,i}\, z_{t-i} + \varepsilon_{k,t}\,. \qquad (3.2)
\end{aligned}
$$

This is a time series regression that predicts the change in the dependent variable $y_{t+k} - y_{t-1}$ for $k = 0, 1, 2, ...K$ periods ahead. The constant α_k is allowed to differ for every prediction interval and $\varepsilon_{k,t}$ is the error term. In the quarterly set-up, I use (the log of) real GDP and the unemployment rate as dependent variables. The central independent variable is f_t, the new financial sentiment variable developed in this paper (depicted in Figure 3.5). The main parameter of interest is β_k which measures the response of the dependent variable after an increase in f_t.

I also include $i = 1, \ldots, I$ lags of the dependent variables. This will be 4 lags in the quarterly version and 12 lags in the monthly version to account for possible cyclical variations over the year. Apart from lagged values of the dependent

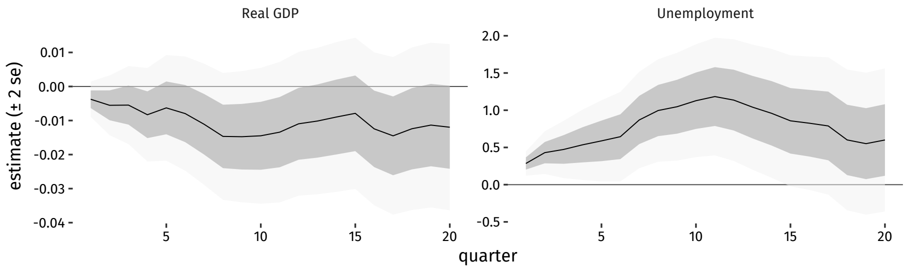


Figure 3.9. Quarterly Local Projections, 1890-2015

Note: Quarterly data. Shows estimates from local projections after a one standard deviation increase in financial stress.

variable (y_t) and financial sentiment (f_t), I also include the respective other dependent variable as z_t. When y_t is real GDP growth, z_t is the unemployment rate (and vice versa). These variables have the purpose of controlling for the state of the business cycle. I then report how a one unit change in the indicators reverberates through the system.

Figure 3.9 shows the impulse response functions. An increase in the new indicator means that more negative language is used, so it is indicative of a rise in financial stress. Real GDP falls slightly, but the response is not precisely estimated. The unemployment rate rises by about three percentage points.

I also estimate the quarterly models for three subperiods: The period of the classical gold standard before 1913, the interwar period from 1919 to 1939 and for the post-war period from 1946 to 2016. I show the impulse responses from these separate estimations in Figure 3.10. The much reduced sample size makes finding significant results more difficult. The markers in Figure 3.10 are filled if a coefficient is estimated to be different from zero at the 5% significance level.

For GDP, results are again imprecise. In the post-war episode, there is a negative response in the first two months. Unemployment rises in the interwar and the post-war period after an increase in the new indicator. This effect is precisely estimated in the post-war period for about 9 months after the jump in financial stress. The response in the interwar period is more pronounced which is likely due to the importance of the Great Depression in the 1930s. The response of unemployment under the gold standard is not significant. This hints at the different functioning of labor markets before 1913. Labor markets were less rigid before WWI (Feinstein et al., 2008), so if wages adjusted faster this might explain the muted response of unemployment.

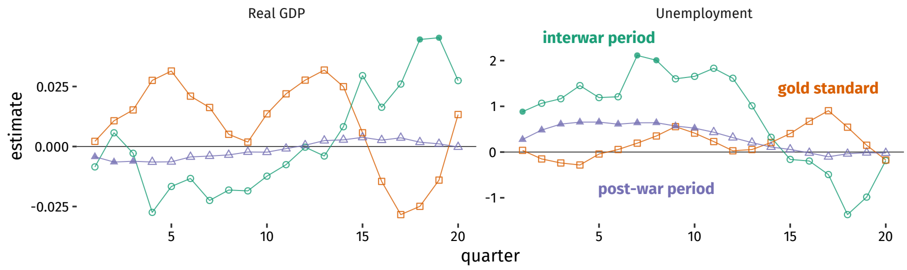


Figure 3.10. Quarterly Local Projections by Time Period, 1890-2015

Note: Quarterly data. Shows estimates from local projections after a one standard deviation increase in financial stress. Uses the *loughran* dictionary for the financial sentiment variable. Markers are filled if estimates are different from zero at the 5% significance level.

3.5.2 Monthly Local Projections

The monthly estimation uses industrial production as the measure of economic activity and I include the growth rate of the S&P stock market indicator. I also use corporate bond spreads, calculated as the difference in the yield on AAA and BAA rated corporate bonds. This is a standard measure of financial stress which is indicative of the time-varying availability ability of the financial market to provide capital to more risky firms. Last, this model version also includes the consumer price index (CPI) to measure inflation.[8] Using monthly data holds the benefit of more than doubling the number of observations from 504 to 1165, but it also truncates the beginning of the sample from 1890 to 1919.

[8] Fred codes: `INDPRO`, `AAA`, `BAA` and `CPIAUCNS`. Stock market data is from Shiller (2000) (`irrationalexuberance.com`).

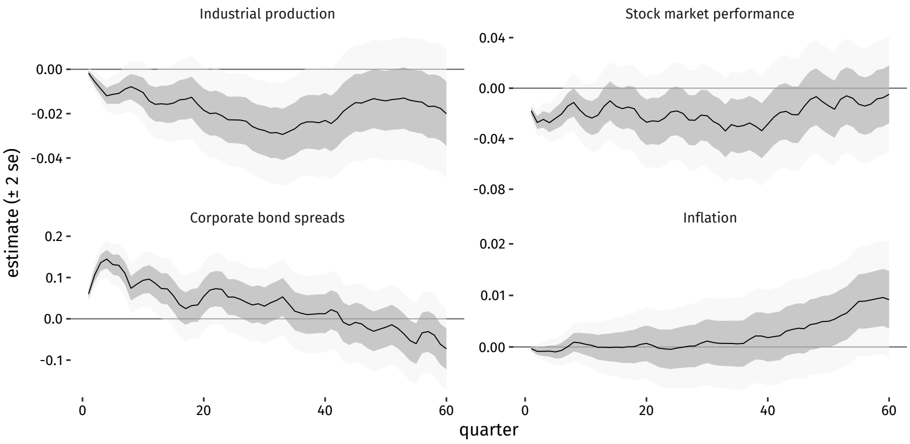


Figure 3.11. Monthly Local Projections, 1919-2016

Note: Monthly data. Shows estimates from local projections after a one standard deviation increase in financial stress.

Figure 3.11 displays the impulse response functions of a one standard deviation shock in financial stress. Industrial production decreases initially and remains subdued for about two years. Stock markets fall, an effect that stays significant for several years after impact. Corporate bond spreads jump markedly. This is a sign that the new financial stress indicator indeed captures periods of financial stress well. Inflation does not respond unambiguously.

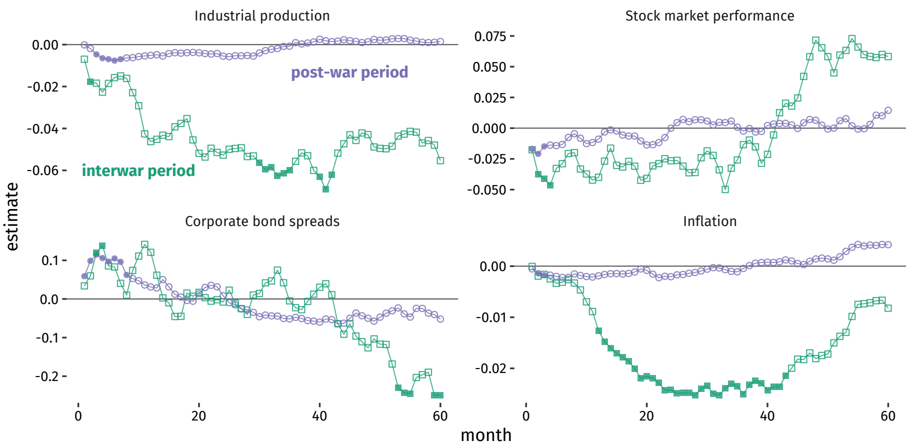


Figure 3.12. Monthly Local Projections by Time Period, 1919-2016

Note: Monthly data. Shows estimates from local projections after a one standard deviation increase in financial stress.

I estimate separate models for the interwar period up to 1945 and the post-war period starting 1946. Industrial production again falls significantly and this effect is larger in the interwar period. Stock market growth is lower in both time periods. Corporate bond spreads also rise for both samples, but in the post-war period there is a reversal after about two years. When estimating models for these separate time periods, the response of the inflation rate also becomes significant. There is a slight fall in inflation after the war, but a strong fall by about seven percentage points in the first time period. This is likely due to the deflation experience of the early 1930s.

3.5.3 Robustness Checks

In Section 3.3.4, I discuss alternative ways to create the new financial stress indicator. Using these series in the local projections does not change the overall conclusions.

Figure 3.A.4 (a) shows the impulse responses obtained from quarterly local projections when using detrended indicator series instead. The impulse responses from all three detrending methods are very similar to those from the baseline indicator. Real GDP falls as in the baseline, but this effect is only significant for the Hamilton (2017) detrending. Unemployment rises significantly across all specifications with estimates that are slightly lower than those of the baseline indicator. Figure 3.A.5 (a) similarly shows how monthly impulse responses change when using detrended indicator series. Most detrended series yield similar results to the non-detrended baseline indicator. Only the estimates from the HP filtered indicator have a slighty different trajectory. The reversal of the stock market returns for the HP filtered indicator after 50 months is not significant at the 95 percent level.

Figure 3.A.4 (b) plots impulse responses from the quarterly model when taking out individual newspapers. Results are again very similar. Effects are somewhat less pronounced when removing the *Chicago Tribune* and are larger when leaving out the *Boston Globe* for real GDP and the *Washington Post* for unemployment. As before, all estimates are insignificant for real GDP and all are effects are significant for unemployment.As Figure 3.A.5 (b) shows, the monthly impulse responses are not affected much when excluding newspapers. While all estimates are nearby, they show some divergence for industrial production and inflation when eliminating the *Boston Globe* from the sample.

3.6 Conclusion

Newspapers have been called “the first draft of history".[9] When looking back at more than a century of financial history, we can learn a lot from reading that first draft again. Events that were salient to observers at the time might be glanced over in the historical narratives. This real-time and subjective view of financial markets serves as an important complement to our usual data sources.

This paper takes the universe of published newspaper titles of five major US newspapers since the 19$^{\text{th}}$ century and provides a novel indicator of financial stress. I identify newspaper titles concerned with financial markets and measure their emotional connotation. From this, I construct a new indicator of negative financial sentiment. A sample of 23,000 hand-coded articles from a media analysis company allows to evaluate the performance of the indicator and I show that the new indicator captures reporting on financial markets well.

I apply the new financial stress indicator in an analysis using local projections to analyze its comovement with key economic variables. I exploit the fact that the new measure is available both on a high frequency and for a long time span and estimate both a quarterly version since 1890 and a monthly version since 1919. I find that an increase in negative financial sentiment is followed by a fall in output, higher unemployment, lower stock market returns and rising corporate bond spreads. This is robust to estimating models using data from different time frequencies, time periods, controlling for several other macroeconomic variables and alternative ways of constructing the new indicator.

The new dataset offers a range of other applications to researchers, as newspaper articles can also be interpreted as the signals that agents received. The literature studying the impact of economic life-time experiences – such as Malmendier and Nagel (2011) – could benefit from employing this new data, as people might not respond to historical measurements of economic variables, but rather to how the situation was perceived at the time. This paper also points to the various other research avenues still open for macroeconomists. Questions of relevance to macroeconomists that could be addressed with newspaper data are measuring inflation expectations or the degree to which government spending shocks were anticipated.

[9] Popik (2009) traces the origin of the quote.

References

Alexopoulos, Michelle and Jon Cohen (2009): "Uncertain Times, Uncertain Measures." Working Paper 352, University of Toronto. [58]

Antweiler, Werner and Murray Z. Frank (2004): "Is All that Talk Just Noise? The Information Content of Internet Stock Message Boards." *The Journal of Finance*, 59 (3), 1259–1294. [58]

Baker, Scott R., Nicholas Bloom, and Steven J. Davis (2016): "Measuring Economic Policy Uncertainty." *Quarterly Journal of Economics*, 131 (4), 1593–1636. [58, 60, 62, 65, 76, 77]

Baron, Matthew, Emil Verner, and Wei Xiong (2018): "Identifying Banking Crises." Unpublished manuscript. [57]

Bohle, Robert H (1986): "Negativism as News Selection Predictor." *Journalism Quarterly*, 63 (4), 789–796. [73]

Bollen, Johan, Huina Mao, and Xiaojun Zeng (2011): "Twitter Mood Predicts the Stock Market." *Journal of Computational Science*, 2 (1), 1–8. [58]

Bordo, Michael D. and Christopher M. Meissner (2016): "Fiscal and Financial Crises." In *Handbook of Macroeconomics*. Vol. 2. [57]

Caldara, Dario and Matteo Iacoviello (2016): "Measuring Geopolitical Risk." Unpublished manuscript. [58, 60, 62, 77]

Combs, Barbara and Paul Slovic (1979): "Newspaper Coverage of Causes of Death." *Journalism Quarterly*, 56 (4), 837–849. [73]

Cutler, David M., James M. Poterba, and Lawrence H. Summers (1989): "What Moves Stock Prices?" *The Journal of Portfolio Management*, 15 (3), 4–12. [57]

Doms, Mark and Norman Morin (2004): "Consumer Sentiment, the Economy, and the News Media." *FRB of San Francisco Working Paper No. 2004-09*. [69]

Feinstein, Charles H., Peter Temin, and Gianni Toniolo (2008): *The World Economy Between the World Wars*. Oxford University Press. [79]

García, Diego (2013): "Sentiment during Recessions." *The Journal of Finance*, 68 (3), 1267–1300. [58, 62]

Gentzkow, Matthew and Jesse M. Shapiro (2010): "What Drives Media Slant? Evidence from US Daily Newspapers." *Econometrica*, 78 (1), 35–71. [62]

Gieber, Walter (1955): "Do Newspapers Overplay 'Negative' News?" *Journalism Quarterly*, 32 (3), 311–318. [73]

Gilchrist, Simon and Egon Zakrajšek (2012): "Credit Spreads and Business Cycle Fluctuations." *American Economic Review*, 102 (4), 1692–1720. [57, 76]

Hamilton, James D. (2017): "Why You Should Never Use the Hodrick-Prescott Filter." *Review of Economics and Statistics*. [72, 82]

Homer, Sidney and Richard Sylla (2005): *A History of Interest Rates*. 4th ed. Wiley. [65]

Hu, Minqing and Bing Liu (2004): "Mining and Summarizing Customer Reviews." *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. [62]

Jalil, Andrew J (2015): "A New History of Banking Panics in the United States, 1825-1929: Construction and implications." *American Economic Journal: Macroeconomics*, 7 (3), 295–330. [58]

Jordà, Òscar (2005): "Estimation and Inference of Impulse Responses by Local Projections." *American Economic Review*, 95 (1), 161–182. [78]

Jordà, Òscar, Moritz Schularick, and Alan M. Taylor (2016): "Macrofinancial History and the New Business Cycle Facts." In *NBER Macroeconomics Annual*. Ed. by Martin Eichenbaum and Jonathan A. Parker. Vol. 31. NBER. University of Chicago Press. [70]

Laeven, Luc and Fabian Valencia (2012): "Systemic Banking Crises Database: An Update." IMF WP/12/163. [57]

Liu, Bing, Minqing Hu, and Junsheng Cheng (2005): "Opinion Observer: Analyzing and Comparing Opinions on the Web." *Proceedings of the 14th International World Wide Web conference*. [62]

Loughran, Tim and Bill McDonald (2011): "When is a Liability not a Liability? Textual Analysis, Dictionaries, and 10-Ks." *Journal of Finance*, 66 (1), 35–65. [62, 63]

Malmendier, Ulrike and Stefan Nagel (2011): "Depression Babies: Do Macroeconomic Experiences Affect Risk Taking?" *The Quarterly Journal of Economics*, 126 (1), 373–416. [83]

Manela, Asaf and Alan Moreira (2015): "News Implied Volatility and Disaster Concerns." Unpublished manuscript. [58, 62]

Miller, Ross A and Karen Albert (2015): "If it Leads, it Bleeds (and if it Bleeds, it Leads): Media Coverage and Fatalities in Militarized Interstate Disputes." *Political Communication*, 32 (1), 61–82. [73]

Mohammad, Saif M. and Peter D. Turney (2013): "Crowdsourcing a Word-Emotion Association Lexicon." *Computational Intelligence*, 29 (3), 436–465. [62]

Mohammad, Saif and Peter Turney (2010): "Emotions Evoked by Common Words and Phrases: Using Mechanical Turk to Create an Emotion Lexicon." *Proceedings of the NAACL-HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*. [62]

Niederhoffer, Victor (1971): "The Analysis of World Events and Stock Prices." *Journal of Business*, 44 (2), 193–219. [57]

Notowidigdo, Matthew J. (2011): "The Incidence of Local Labor Demand Shocks." NBER Working Paper No. 17167. [62]

O'Connor, Brendan, Ramnath Balasubramanyan, Bryan R. Routledge, and Noah A. Smith (2010): "From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series." *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*, 11 (122-129), 1–2. [58]

Popik, Barry (2009): *"First draft of history" (journalism) (23 November 2009)*. https://www.barrypopik.com/index.php/new_york_city/entry/first_draft_of_history_journalism/. Accessed: 2018-04-25. [83]

Ramey, Valerie A. and Sarah Zubairy (2017): "Government Spending Multipliers in Good Times and in Bad: Evidence from U.S. Historical Data." *Journal of Political Economy*. Forthcoming. [78]

Rancière, Romain, Aaron Tornell, and Frank Westermann (2008): "Systemic Crises and Growth." *Quarterly Journal of Economics*, 123 (1), 359–406. [57]

Reinhart, Carmen M. and Kenneth S. Rogoff (2009): *This Time Is Different: Eight Centuries of Financial Folly*. Princeton: Princeton University Press. [70]

Schularick, Moritz and Alan M. Taylor (2012): "Credit Booms Gone Bust: Monetary Policy, Leverage Cycles and Financial Crises, 1870–2008." *American Economic Review*, 102 (2), 1029–61. [57]

Shiller, Robert J. (2000): *Irrational Exuberance*. Princeton University Press. [80]

Tetlock, Paul C. (2007): "Giving Content to Investor Sentiment: The Role of Media in the Stock Market." *Journal of Finance*, 62 (3), 1139–1168. [58, 62]

Von Hagen, Jürgen and Tai-Kuan Ho (2007): "Money Market Pressure and the Determinants of Banking Crises." *Journal of Money, Credit and Banking*, 39 (5), 1037–1066. [57]

Appendix 3.A Additional Figures and Tables

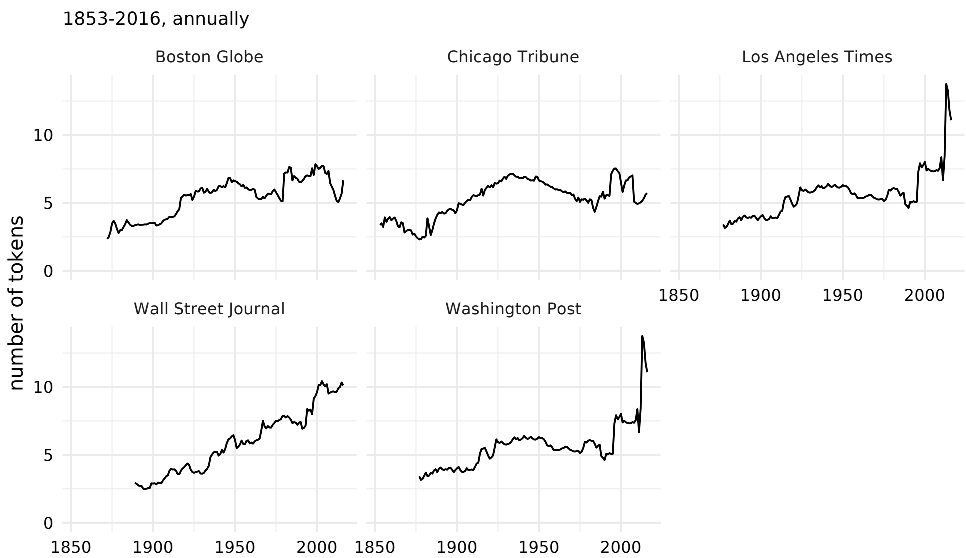
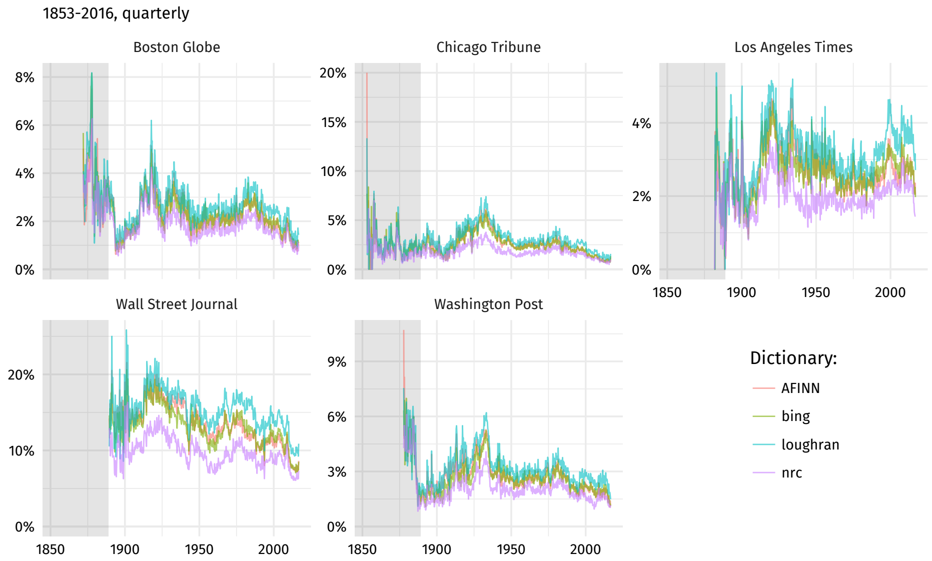
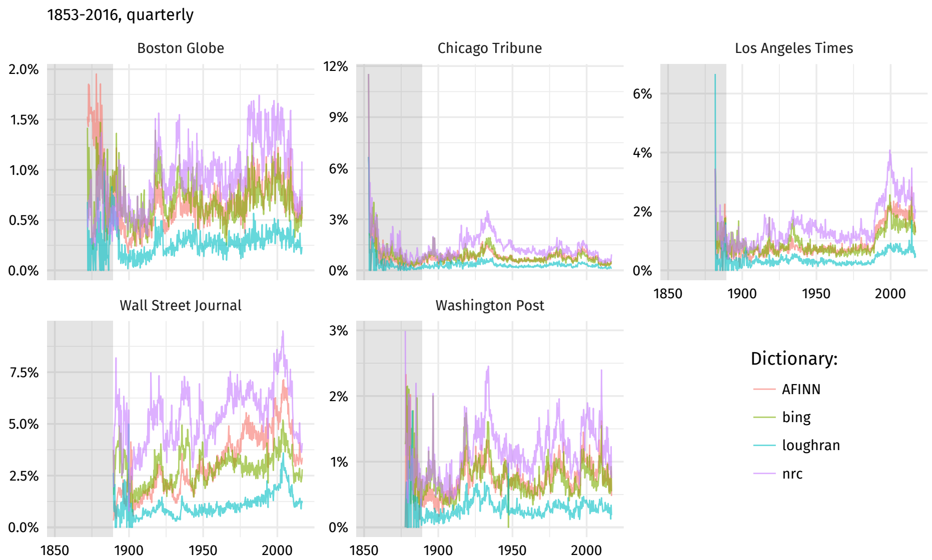


Figure 3.A.1. Length of Newspaper Titles



(a) neutral



(b) positive

Figure 3.A.2. Other Moods in Newspaper Articles

Note: Calculated as all newspaper articles that are both about financial markets and have net neutral/positive sentiment divided by all articles in that quarter.

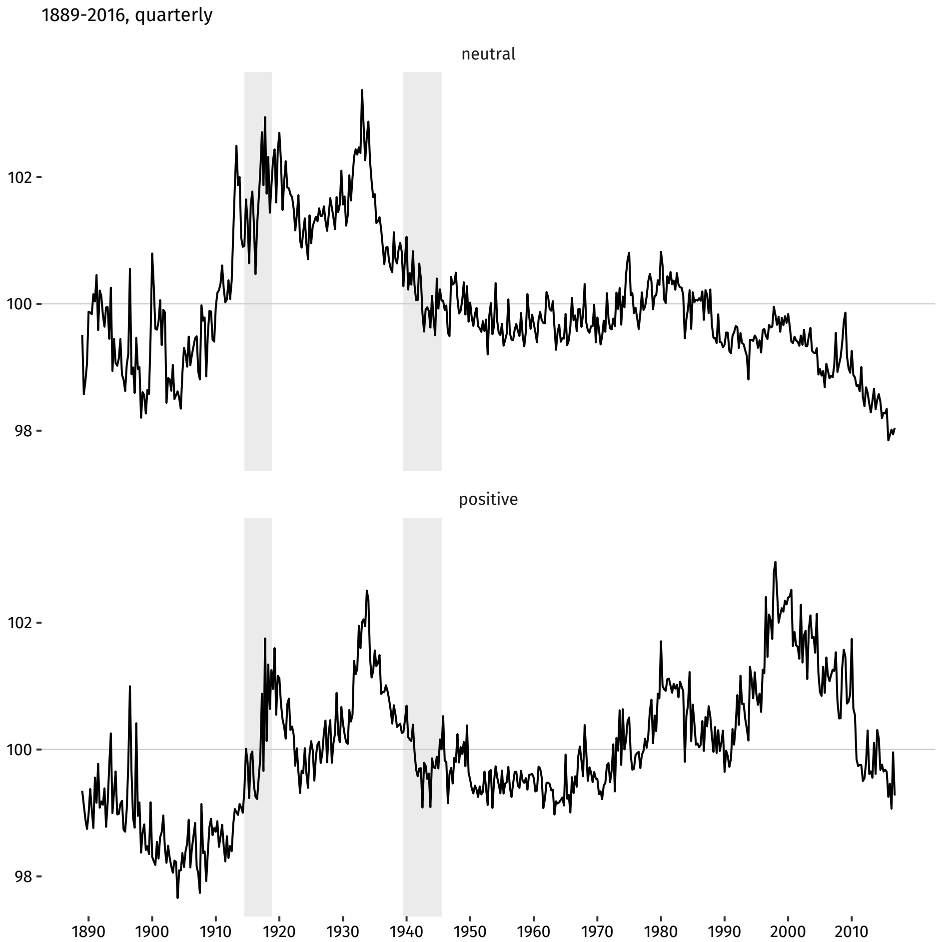
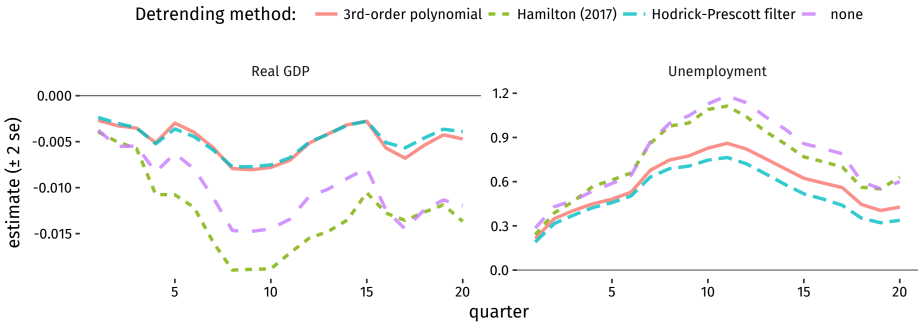
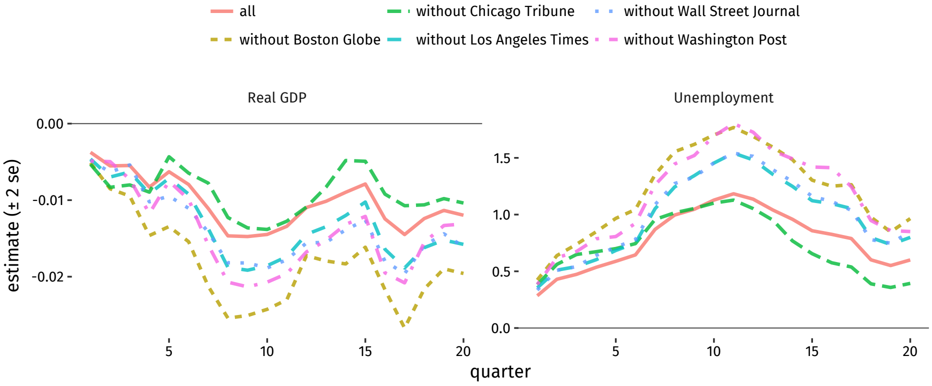


Figure 3.A.3. Other Moods in Financial Sentiment

Note: Financial articles with neutral or net positive sentiments as share of all articles. Shows normalized averages across the five newspapers.



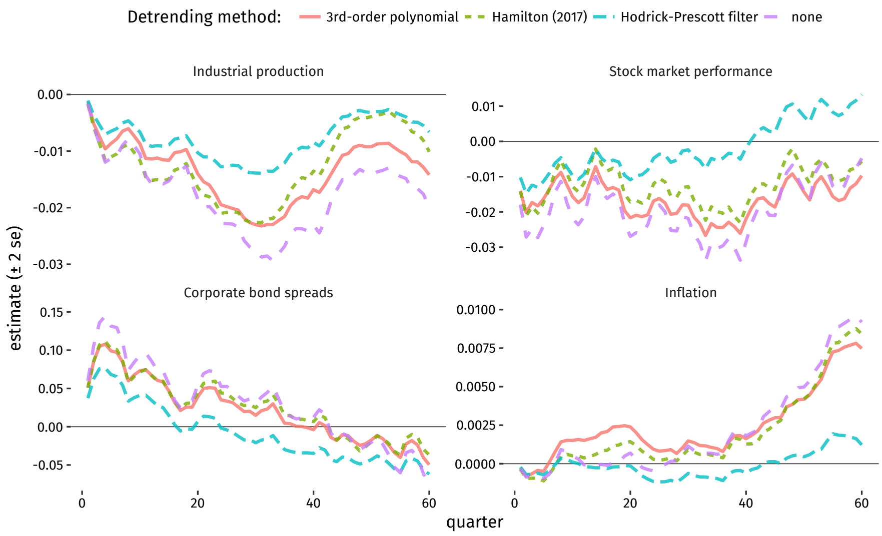
(a) Different detrending methods



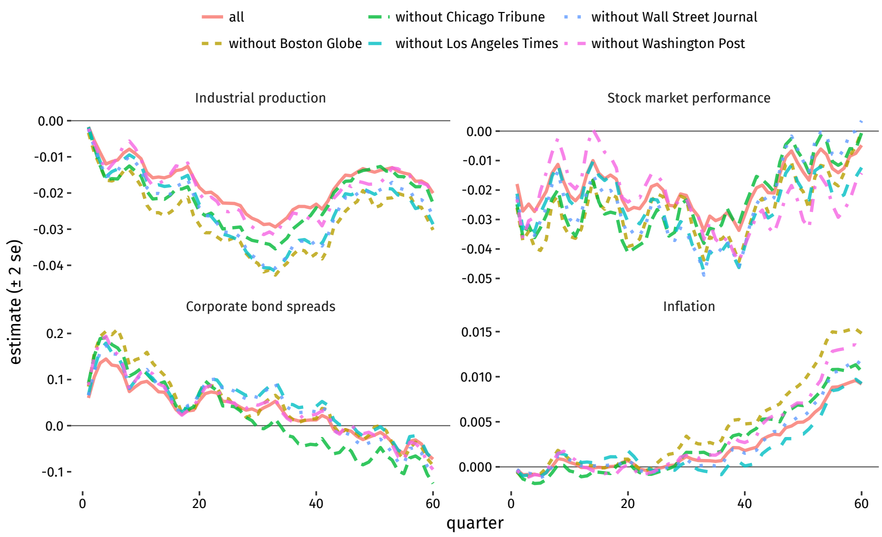
(b) Leaving out newspapers from sample

Figure 3.A.4. Quarterly Local Projections with Alternative Indicators, 1890-2015

Note: Shows estimates from quarterly local projections after a one standard deviation increase in financial stress.



(a) Different detrending methods



(b) Leaving out newspapers from sample

Figure 3.A.5. Monthly Local Projections with Alternative Indicators, 1919-2016

Note: Shows estimates from quarterly local projections after a one standard deviation increase in financial stress by time periods.

Table 3.A.1. Pairwise Correlations of Lexicon Indices

Lexicon	AFINN	bing	loughran	nrc
		Boston Globe		
AFINN	1.0	0.78	0.60	0.44
bing		1.0	0.85	0.72
loughran			1.0	0.73
nrc				1.0
		Chicago Tribune		
AFINN	1.0	0.89	0.94	0.74
bing		1.0	0.82	0.86
loughran			1.0	0.64
nrc				1.0
		Los Angeles Times		
AFINN	1.0	0.96	0.95	0.93
bing		1.0	0.93	0.93
loughran			1.0	0.88
nrc				1.0
		Wall Street Journal		
AFINN	1.0	0.96	0.96	0.87
bing		1.0	0.96	0.93
loughran			1.0	0.87
nrc				1.0
		Washington Post		
AFINN	1.0	0.84	0.92	0.71
bing		1.0	0.77	0.81
loughran			1.0	0.64
nrc				1.0

Note: Shows the pairwise correlations between the time series of the share of negative financial titles by dictionary in Figure 3.4.

Table 3.A.2. Correlations Across Alternative Indicators

	baseline	polynomial	HP filter	Hamilton (2017)
	Quarterly			
all newspapers	1.00	0.75	0.72	0.82
no *Boston Globe*	0.98	0.74	0.71	0.80
no *Chicago Tribune*	0.96	0.60	0.59	0.79
no *Los Angeles Times*	0.97	0.83	0.80	0.87
no *Wall Street Journal*	0.98	0.84	0.81	0.88
no *Washington Post*	0.98	0.72	0.70	0.75
	Monthly			
all newspapers	1.00	0.76	0.63	0.82
no *Boston Globe*	0.98	0.75	0.63	0.81
no *Chicago Tribune*	0.96	0.63	0.55	0.79
no *Los Angeles Times*	0.97	0.84	0.71	0.88
no *Wall Street Journal*	0.98	0.85	0.71	0.88
no *Washington Post*	0.98	0.73	0.61	0.76

Note: Shows correlations when constructing alternative indicators (detrending and/or leaving out individual newspapers from sample). The “baseline" does not detrend and uses all newspapers. Uses quarterly/monthly variables 1889-2016 (detrended indicator for Hamilton 2017 method starts in 1900).

Table 3.A.3. OLS Estimates of Media Tenor Comparison

	Lexicon			
Topic	*AFINN*	*bing*	*loughran*	*nrc*
business cycle	0.182	0.211 **	0.26 **	0.212 *
	(0.115)	(0.106)	(0.109)	(0.115)
financial markets	0.115***	0.066 *	0.126***	0.085 **
	(0.038)	(0.035)	(0.035)	(0.04)
economic policy	0.091***	0.033	0.029	0.062 *
	(0.033)	(0.03)	(0.031)	(0.033)
global economies	0.062	0.186 **	0.049	-0.036
	(0.087)	(0.081)	(0.087)	(0.081)
fiscal policy	0.083	-0.008	0.125	0.131
	(0.111)	(0.1)	(0.106)	(0.11)
commodities	0.004	0.037	0.018	-0.002
	(0.04)	(0.04)	(0.038)	(0.044)
business	0.088	0.05	-0.04	0.018
	(0.078)	(0.067)	(0.088)	(0.066)
monetary policy	0.028	0.003	0.026	0.054 *
	(0.031)	(0.029)	(0.03)	(0.032)
other	0.104	0.071	0.109	0.009
	(0.072)	(0.074)	(0.084)	(0.085)
politics	-0.054	-0.052	-0.061	-0.101
	(0.077)	(0.068)	(0.072)	(0.079)
labor market	0.036	0.014	0.006	-0.014
	(0.038)	(0.035)	(0.036)	(0.037)
economic growth	-0.058	0.05	0.049	-0.094
	(0.125)	(0.11)	(0.119)	(0.121)

Note: Shows OLS estimates and standard errors from a regression of the number of articles from a specific topic on the share of negative financial language, as measured by different lexicons. Also controls for day of the week and sample dummies. *** p<0.01, ** p<0.05, * p<0.1. Shows corresponding estimates to Figure 3.7.

Appendix 3.B Dictionaries

This section includes the dictionaries I use to filter articles of specific topics. I convert all text to lowercase before the matching, so “fed" would match “fed", “Fed", “FED" or any other capitalization.

The following dictionary serves the purpose of selecting articles that are concerned with financial markets:

bail out, bailout, banks, bernanke, boj, bond, bonds, bretton woods, broker, brokerage, brokerages, bullion, bundesbank, capital markets, central bank, central banks, coin, coinage, commerce, commercial, consumer, consumer confidence, consumer price, consumer prices, consumers, cpi, credit, credit crunch, credit crunchs, currency, dax, debt, debts, deflated, deflation, derivative, derivatives, dividend, dividends, dollar, dollars, draghi, ecb, economic, economies, economy, embargo, embargoes, entrepreneur, entrepreneurial, entrepreneurs, equities, equity, export, exported, exports, fed, federal reserve, finance, financial, financial markets, financial system, franc, fund, funds, gold, government bond, government bonds, greenspan, import, imported, imports, income, incomes, industrial, industry, inflated, inflation, interest rate, interest rates, investor, investors, loan, loans, loss, losses, market declines, markets decline, monetary, money market, money supply, mortgage, price movement, price movements, produce, producer prices, producers, profit, profits, purchase, purchases, railroad, reichsbank, reichsmark, silver, sterling, stock, stocks, stress test, stress tests, swap, swaps, tariff, tariffs, treasury bill, treasury bills, tbill, tbills, volcker, wall street, yellen, yen

4

Unexpectedly Broke: Expectation Errors and Credit Cycles

Joint with Carsten Detken, Anna Kalbhenn and Eric Persson

4.1 Introduction

Private debt booms hold risks for the economy. Excessive credit growth has been shown to precede recessions, financial crises and to lower returns to bank capital and bonds.[2] However, the question remains why people accumulate so much debt in the first place. If individuals rationally decide how much to consume and save, why do we observe these ups and downs in credit and economic activity?

Understanding the causes of debt booms is essential for providing policy makers with good advice and the right tools to mitigate their effects. Households might rationally take on debt due to higher anticipated earnings or lower interest rates and an increase in debt might therefore not call for a market intervention. But if savings decisions are based on unreasonable income expectations and this could somehow be found out in advance, then there might be room for policy action.

We ask whether errors in income expectations are a plausible explanation for credit accumulation. For this, we propose a new measure of errors in aggregate income expectations. Some studies — e.g. De Stefani (2017) and Rozsypal and

[2]See Kaminsky and Reinhart (1999), Schularick and Taylor (2012), Greenwood and Hanson (2013), Alessi and Detken (2017), Chen and Rancière (2016), Baron and Xiong (2017), Lopez-Salido et al. (2017) and Mian et al. (2017a).

Schlafmann (2017) — use consumer surveys for measuring income expectations, but such data is only available for few countries. However, credit cycles are slow-moving and financial crises are rare events, so data with sufficient international coverage is necessary to allow making statements with enough statistical power. Good international data coverage also helps eliminate country idiosyncrasies and to find macroeconomic regularities that hold more generally. This paper uses a collection of professional forecasts by the private data provider *Consensus Economic Forecasts* (CEF). The CEF sends a survey every month to financial firms, banks and economic research institutes and asks them to predict macroeconomic variables.

We obtain quarterly GDP forecasts and calculate forecast errors as the difference between mean predictions and realizations. Taking forecasts as proxies for expectations in the economy, we examine what happens with household debt accumulation in periods of overly optimistic expectations. The resulting quarterly dataset covers 32 countries accounting for 79 percent of global output over almost three decades. Our analysis shows that positive forecast errors in 12 month ahead real GDP growth are contemporaneously correlated with booms in household debt growth. This association holds across time periods, within industrialized and developing countries, controlling for time and country fixed effects, excluding banking crises and when controlling for the state of countries' business cycles and ex ante real interest rates. Household debt reacts strongly, but there is no relevant comovement of expectation errors with firm debt growth. This evidence is in line with Mian, Sufi, and Verner (2017a) who show that higher household debt predicts negative GDP forecast errors. We turn their analysis around and ask what drives debt accumulation. Also, we use a different data source which allows us to use quarterly, not annual, observations.

After establishing our baseline result, we provide additional evidence from another dataset. We use the ECB's *Survey of Professional Forecasters* in which forecasters provide subjective confidence bands. On the downside, this dataset only covers the euro area. We show unambiguously that panelists suffer from overprecision in their forecasts. The 95 percent confidence intervals that panelists provide are so narrow that they cover only a third of subsequent realizations. We take this as a further support for our approach of using forecast errors as proxies for expectation errors.

We then dive into how panelists update their forecasts when they receive new information. We use a method from Coibion and Gorodnichenko (2012) and Bordalo, Gennaioli, Ma, and Shleifer (2018) to show strong evidence for overreaction by forecasters, a finding that is robust across countries and two levels of observation. In particular in the run-up to the financial crisis up to 2006 expectation formation showed strong signs of overreaction.

This paper is, to our knowledge, the first to provide comprehensive empirical evidence that misaligned income expectations are a plausible explanation for credit growth across the world. This is relevant for economic policy as it provides further support to the recent efforts to monitor — and maybe regulate — lending to households.

4.2 Theories of Credit Booms

Since the financial markets turmoil of 2007 onwards, economists' interest in understanding the causes and consequences of financial crises has been rekindled. A robust empirical finding is that private credit tends to rise before trouble hits financial markets and the economy. The literature offers several explanations for debt booms and these theories make predictions on how forecast errors and credit cycles should be related in the data.

Cochrane (2017) surveys the field of macrofinance and lays out the unifying framework of a cyclical bias in the representative agent's consumption-savings decision. This wedge leads to a comovement of economic and financial activity in the economy. These biases take different forms, such as neglecting small probability events, extrapolative expectations or habits. But the effects from these distortions tend to be alike: Lenders overestimate the present value of current and future incomes tempting them to hold more debt than they can stomach, while savers overestimate the capacity of households or firms to repay debts.

Whether it is the providers or the receivers of credit who change their behavior is of relevance to how we expect prices to adapt. If banks become more willing to lend, then we would expect interest rates to drop. Mian et al. (2017a) provide evidence that credit booms are driven by fluctuations in credit supply. This idea is in line with Kindleberger (1978), who argues that “in moments of euphoria" (p.57) banks will come up with new ways to lend and create liquidity, thus increasing the supply of credit. Bordalo, Gennaioli, and Shleifer (2017) provide a rationalization for how this might come about. In their model, the agent's savings decision is distorted by the representative heuristic which leads to extrapolative expectations. This means that in good times, agents underestimate the probability for lending firms to default.

Conversely, if households or firms demand more credit, we expect interest rates to rise. In Mian et al. (2017b), this takes the form of a temporarily lower effective interest rate that households face. This, they argue, might be due to financial deregulation or overoptimistic income expectations. The lower interest rate induces households to take on debt to finance higher consumption. Gennaioli, Ma, and Shleifer (2016) find evidence that firm managers have extrapolative expectations. This could similarly lead them to be too optimistic when times are good. Blanchard, L'Huillier, and Lorenzoni (2013) provide a model in which noisy information about future productivity induce agents to consume more than optimal. This, too, might be an explanation for periodical expansions and contractions in household's saving behavior.

The equilibrium outcome in each case with respect to credit is the same: It rises due to mistakes in the expectation formation. In our empirical analysis, we therefore test whether the economy's indebtedness rises when agents are too optimistic.

4.3 Data

4.3.1 Forecasts

Near the middle of every month, forecasters of 32 countries fill out a survey by the private data provider *Consensus Economics* and predict real GDP growth for the current and next year. The CEF dataset is highly fragmented and manually processed as the data providers sends updates as PDF's and Excel sheets to institutions and Central Banks. We go to great lengths to collect and aggregate all available data to create one unified database of macroeconomic expectations.

Participating firms (or "panelists") are a mix of banks, private and public research institutes, market intelligence firms, industrial unions and business organizations. Figure 4.1 shows the panelists with the most forecasts (aggregated to quarterly) which are JP Morgan (5340), UBS (4662) and Goldman Sachs (4356). Forecasters or their subsidiaries reside in the country whose economy they predict. Berger, Ehrmann, and Fratzscher (2011) show that geography proximity increases forecaster accuracy. As participating firms might differ in their access to local information, in how they form forecasts and in the effort they exert, the quality of forecasts might also not be the same across firms. However, with about 17 (st.d.: ±5) forecasters per month and country, the weight of any individual forecasts is low.

Figure 4.2 shows how the number of panelists evolved. The number of panelists was approximately constant for most countries. Over the years, more panelists were surveyed in France and Germany and fewer for Great Britain, for which this number fell from the highest ever in 1994 with 39 forecasts. The Netherlands and Norway had the fewest forecasts with an average of 10 forecasts per month. The approximately constant number of forecasts is a further sign of the quality of the *Consensus Economics* data. This is in contrast to the US *Survey of Professional Forecasters*, where this number fell between 1970 and 1990 (Capistrán and Timmermann, 2009).

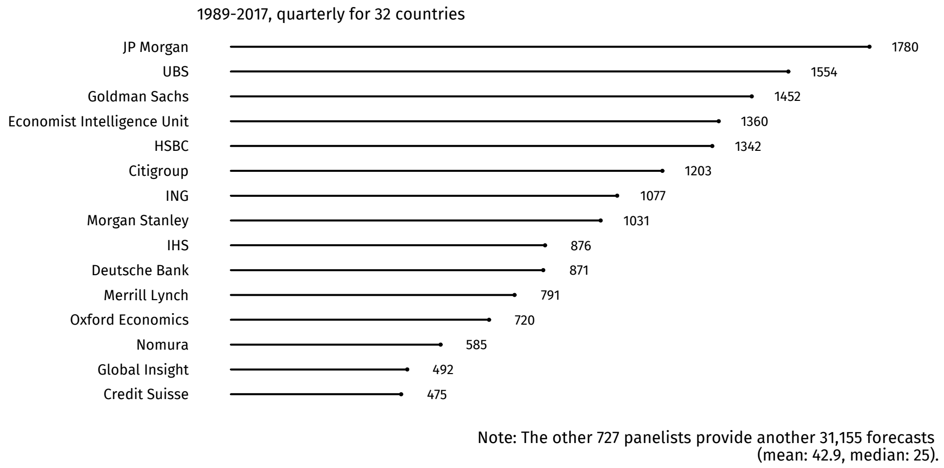


Figure 4.1. Forecasts by Panelist

Panelists in the dataset are not anonymous, so career and reputational concerns might incentivize participating institutions to exert high effort to provide good forecasts. This is in contrast to other surveys such as the Philadelphia Fed's *Survey of Professional Forecasters* or the European Central Bank's *Survey of Professional Forecasters*, where the names of participating firms are not public. However, not being anonymous might also keep panelists from making forecasts that are more unusual for fear of being singled out for large forecast errors.

Some researchers, such as Dovern, Fritsche, and Slacalek (2012) and Rülke, Silgoner, and Wörz (2016), have also worked with the CEF data to tackle different questions, but we are the first to measure cycles of expectation errors and their financial stability implications. We take the errors that participants make in their forecasts as a sign for optimism and pessimism. Ideally we would like to also know participants' individual uncertainty surrounding their point forecasts to assess whether they were really overconfident in their predictions. An advantage of obtaining the microdata of forecasts is our ability to track firms over time and to provide a measure of forecaster dispersion for which we use the standard deviation of point forecasts across panelists at any point in time. We take this as a proxy for forecaster uncertainty, an approach that Bachmann et al. (2013) find support for. The broad coverage of macroeconomic variables also means that we can control for other relevant expectations, such as the expected inflation rate.

Batchelor (2001) and Loungani (2001) analyze the performance of the CEF forecasts and show while they are better than OECD and the IMF forecasts, they are not very good in absolute terms. Breitung and Knüppel (2017) recently provide evidence that the CEF forecasts might not be informative beyond two to four quarters. For the argument in this paper we require that the predictions voiced by professional forecasters are indicative of the opinions held by agents in the economy. People cannot perfectly predict the course of economy and neither can professional forecasters.

Several assumptions are needed for interpreting forecaster errors as expectations of agents in the economy. First, we assume that households hold similar beliefs about the future as do professional forecasters. This might be the case if households and professional forecasters have the same information to construct forecasts. Or it could hold if professional forecasts are published in newspapers and people align their expectations with what they read. A last reason for such a connection between what households expect and what financial firms predict is that both might be driven some third factor such as "optimism", sometimes also called "sentiment" or "exuberance" (Shiller, 2000).

A second assumption becomes necessary when we think about a representative household's savings decision. Typically, households smooth consumption and thus when deciding on how to divide their income into consumption and saving, they not only take next year's income into account, but the discounted sum of all future incomes. So ideally, we would like to measure peoples' lifetime income expectations. However, expectations about GDP growth over the next 12 months are all that we can construct from the CEF data. Several facts mitigate this concern: Households discount incomes more, the further into the future they accrue. Next, if GDP is a random walk, then the one-period (12 month) ahead forecast is the same as the long-run forecast. Neither of these explanations is likely to be exactly true, but both make using the 12 month expected real GDP growth rate a plausible proxy for household's lifetime income expectations.

4.3.2 Macroeconomic Data

For quarterly macroeconomic data we rely on a number of standard sources, such as the OECD, the IMF's International Financial Statistics and Balance, the IMF's Balance of Payments Statistics, national statistical agencies and national central banks.

Data on credit are provided by the BIS. The BIS defines "credit" as loans and debt securities provided to the private non-financial sector which includes non-financial firms, households and non-profit organizations serving households. The lenders can be domestic banks, the rest of the economy and foreigners. We separately use household debt including non-profit organizations serving households (hhd) and non-financial firm debt (fd).

Table 4.1 summarizes the resulting dataset. It provides a good coverage of global economies, as countries in the sample accounted for 79 percent of global

Figure 4.2. Number of Panelists

purchasing power adjusted real GDP in 2015. With 18 out of the largest 20 economies, we cover most major economies. This global coverage also allows us to trace the association between expectation formation and credit accumulation in developing countries which make up about half the countries in the sample. The data for most countries in the sample starts in the 1990s and runs until 2016. For the exact starting dates when countries join our dataset, see Tables 4.A.1 and 4.A.2.

Table 4.1. Data Summary

Number countries	32
(of which developing countries)[1]	(47%)
Coverage starting in 1980s	6
Coverage starting in 1990s	18
Coverage starting in 2000s	8
Share world GDP, PPP (2015)	79%
Forecasts by country-month	17 (±5)

[1]: according to IMF

4.3.3 Forecasts and Realizations

Participating firms provide their forecasts for the annual values of the current and following year. For example, forecasters in June 2016 would provide their guesses for real GDP growth for the year 2016 and 2017. This poses a challenge for the interpretation of this data, as more information becomes available throughout the year. As an extreme case, a forecaster interviewed in mid-December will know with considerable accuracy what happened in the current year. This introduces a seasonality in the forecasts that hinders proper interpretation of this data.

Instead, we would like to use forecasts for the growth rate of real GDP twelve months from survey date. So the challenge is to convert *fixed event* to *fixed horizon* forecasts. Two papers provide methods to overcome this problem, tailored to the data structure of the CEF forecasts. Both use linear weightings of the forecasts for the current and the following year to construct the fixed horizon forecast. The first — Dovern, Fritsche, and Slacalek (2012) — suggests putting progressively less weight on the current year forecast as the year advances. While this holds intuitive appeal, there is no theoretical basis for using this method.

Knüppel and Vladu (2016) instead propose a different weighting which minimizes the expected squared error loss and this method performs better at approximating the fixed horizon forecasts. A key insight from the Knüppel and Vladu (2016) method is somewhat puzzling: For fixed horizon forecasts constructed for the first months in a year, they prescribe to put no weight on the forecasts for the current year. The optimal weights for the current year forecasts even become negative midyear and then positive at the end of the year. Overall, the absolute weight for current year forecasts is very low.[3]

[3]The maximum absolute prescribed weight for the current year under our parameterization ($\rho = 0$) is 8 percent, so at a minimum 92 percent of the constructed forecast comes from next year's forecast. Figure 4.B.1 plots the weights for the two methods.

The reason for this difference between the two approaches is that the hypothetical synthetic forecaster in the ad hoc method by Dovern et al. (2012) puts significant weight on the latest information they received. Acknowledging recent changes is different from making a 12 month forecast, a time span in which recent shocks might have subsided. Due to the better theoretical foundation and the provided empirical evidence we choose the procedure by Knüppel and Vladu (2016), but results remain unchanged when applying the method by Dovern et al. (2012).

We only keep forecasts made in the last month in every quarter to be able to compare them against subsequent real GDP growth realizations, which are only available at quarterly frequency. Figure 4.3 shows forecasts and the subsequent realizations of real GDP growth. The realizations are forward-looking, so the value in the first quarter of 2000 indicates the growth rate of real GDP until the end of the first quarter of 2001. So in the figure, at any point in time we see the four standard error bands (two above and two below) around the consensus forecasts and the ex post true realization. The vertical distance between the two are the forecast errors.

Some of the series (e.g. Argentina) are much more volatile than others (e.g. USA). It is striking how rarely the black line lies within the gray bands, so the GDP growth realizations are more volatile than the forecasts. In particular, forecasts often lag behind the realizations, as if forecasters extrapolated recent realizations. The large recession between 2007 and 2009 surprised forecasters in most countries which led to large positive forecast errors.

4.4 Results

4.4.1 Forecast Errors

Figure 4.4 plots the the forecast errors, calculated as the consensus (mean) forecasts minus realizations. There are periods of positive (red) and negative (blue) forecast errors among all countries. Forecast errors are particularly high when large recessions strike, a result in line with McNees (1992). Magnitudes are much larger in some countries (Argentina, Hong Kong, Malaysia, Russia, Singapore, South Korea, Thailand and Turkey) and this is mostly driven by much higher macroeconomic volatility in developing countries.

To analyze the persistence in the forecast errors, Figure 4.5 shows the autocorrelations of forecast errors. The errors are significantly positively autocorrelated within one year (lag 1 to 4) for most countries. For some countries there is a sig-

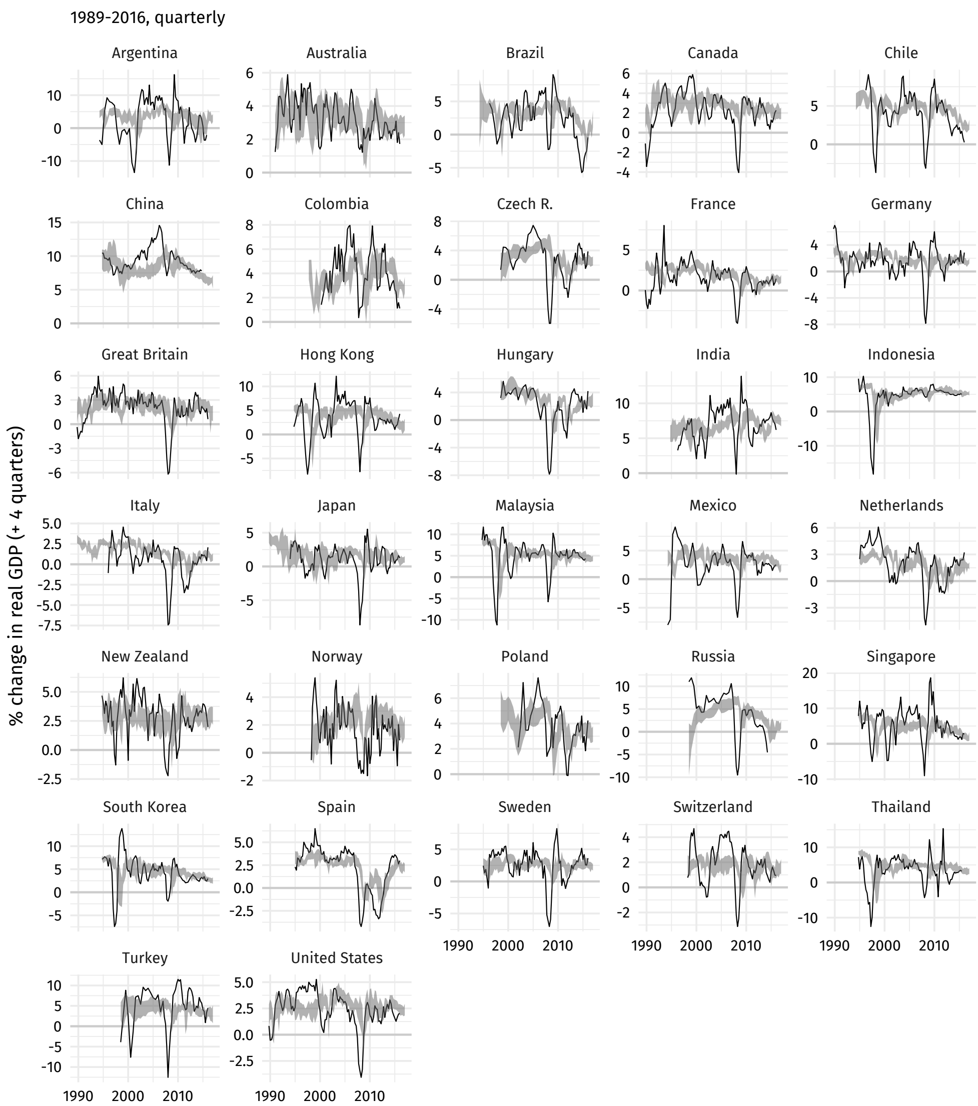


Figure 4.3. GDP Forecasts and Realizations

Note: Shaded areas show 2 standard deviation bands around mean forecasts. Black lines are realizations. Both forecasts and growth rates are forward-looking for 12 months ahead.

nificant reversal towards a negative autocorrelation after about two years (lag 8).[4]

---

[4]As Kučinskas and Peters (2018) explain, the existence of autocorrelation in forecast errors alone is a strong sign of a bias in expectation formation. We will explore this further in Section 4.5.2.

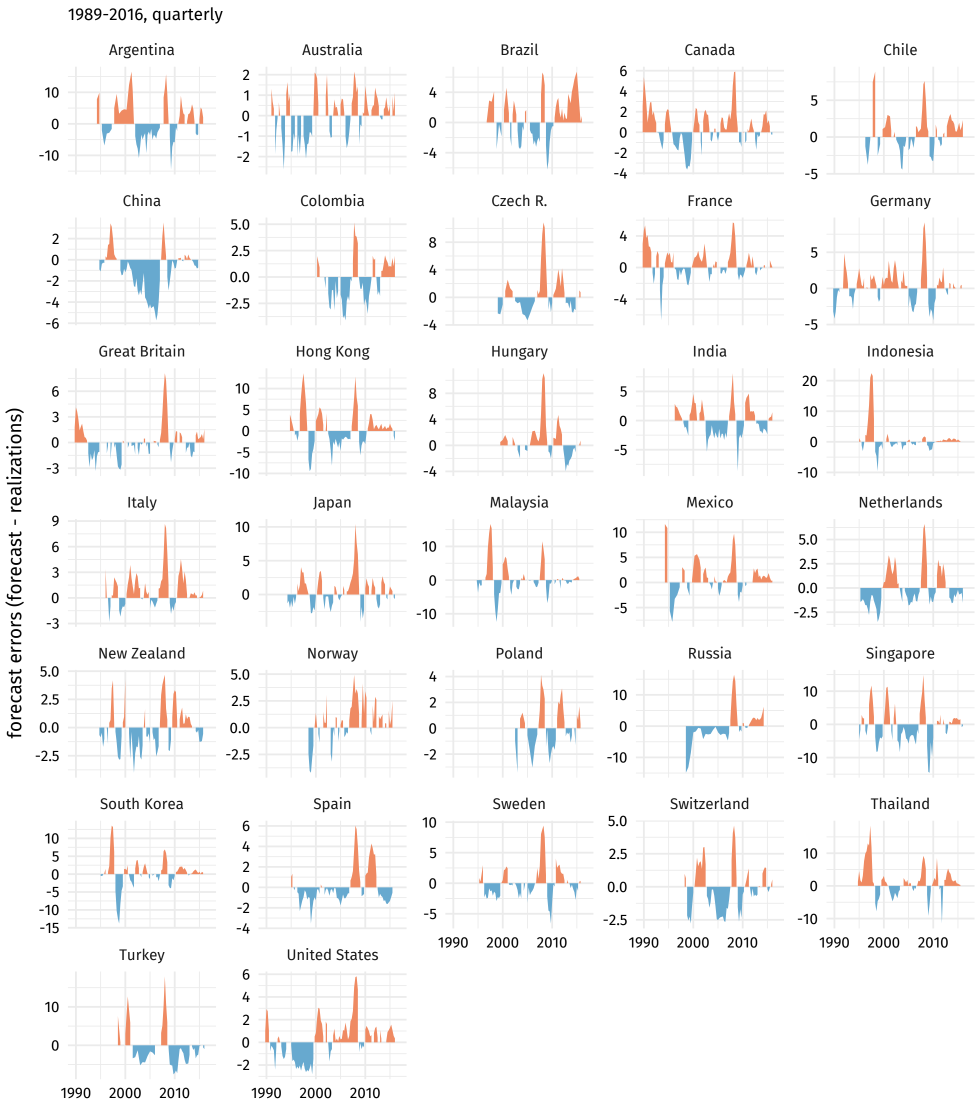


Figure 4.4. Forecast Errors

Note: Positive values (red) show real GDP growth (4 quarters ahead) mean (consensus) forecasts larger than realizations. Vice versa for negative values (blue).

These persistent errors raise the question of how well the hypothetical consensus (mean) forecaster for each country is calibrated.[5] As reported in Table 4.A.1 and 4.A.2, errors are significantly positive at the 95% level for Canada, France, Italy, Japan, Mexico and Thailand, so in these countries forecasters overestimated output growth on average. China is a special case as well: Forecasters underes-

[5]Figure 4.B.2 in the Appendix shows kernel density plots of forecast errors for each of the country in the sample.

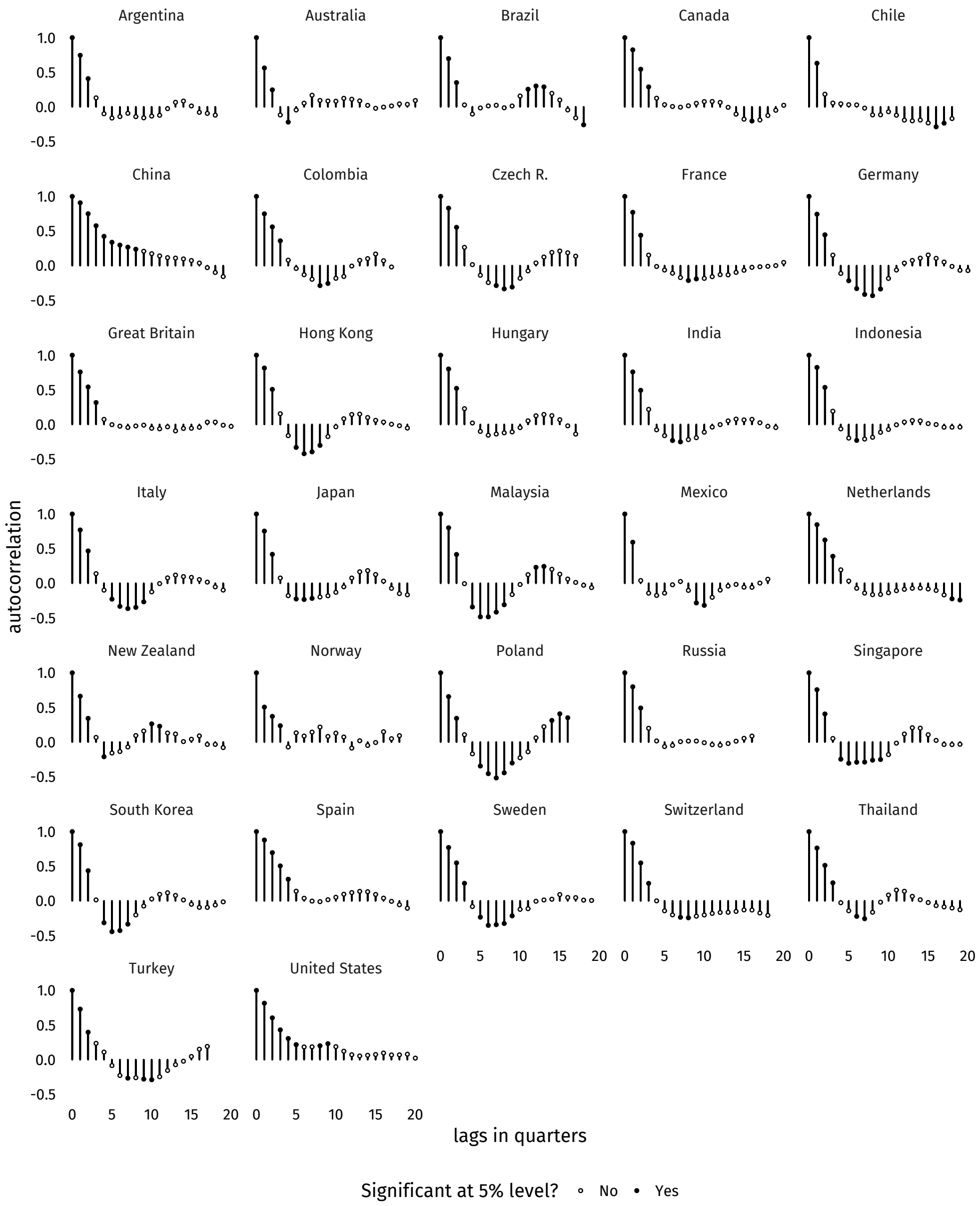


Figure 4.5. Autocorrelations of Forecast Errors

timated China's real GDP growth in every quarter between the first quarter of 1999 and the first quarter of 2007.

Overall, these forecast errors point to extended periods when even professional forecasters were strongly mistaken about aggregate income growth over the coming year. In the following empirical analysis, we examine what else characterized these periods of booms and busts in expectations.

Table 4.2. Baseline Regressions

	Dependent variable: $\text{hhd}_{i,t}$ (household debt growth)					
	(1)	(2)	(3)	(4)	(5)	(6)
Forecast errors	0.91^{**}	0.70^{**}	0.71^{**}	0.72^{**}	0.73^{**}	0.70^{**}
	(0.36)	(0.32)	(0.33)	(0.34)	(0.35)	(0.33)
Real GDP gr.		1.22^{***}	1.35^{***}	1.29^{***}	1.37^{***}	1.38^{***}
		(0.31)	(0.31)	(0.30)	(0.34)	(0.35)
Exp. inflation			0.61^{***}	0.89^{***}	0.89^{***}	0.91^{***}
			(0.20)	(0.27)	(0.26)	(0.25)
Interest rate				-0.24	-0.26	-0.27
				(0.22)	(0.21)	(0.21)
Uncertainty					2.42	2.28
					(2.49)	(2.38)
Banking crises						2.28
						(4.41)
Observations	2348	2301	2301	2295	2295	2295
R^2	0.18	0.24	0.26	0.26	0.26	0.26

Note: All models include country and time fixed effects. Robust standard errors in parentheses. GDP: real GDP y-o-y growth. Exp. inflation: 12-month ahead expected CPI growth. Exp. inflation: 12-month ahead expected CPI growth. Uncertainty: Standard deviation of forecasts. Banking crises are defined by Laeven and Valencia (2012).*** p<0.01, ** p<0.05, * p<0.1.

4.4.2 Comovement with Credit Cycles

We compare periods of positive or negative expectation errors with cyclical expansions and contractions in lending in the economy. Our proxy for the financial cycle are household and firm debt growth. We graph these variables in Figures 4.B.3 and 4.B.4 and we can see distinct financial cycles across countries. Some countries experience only one cyclical swing (in household debt growth) in the sample period (Brazil, Czech Republic, Japan, Spain and Sweden), while for others the measure is trending in this period (Indonesia and the Netherlands). When banking crises hit, GDP in many countries contracts strongly, so we observe sharp drops in their credit growth rates. In the empirical analysis, we carefully exclude the possibility that the association we find is driven by these periods.

Table 4.3. Regressions by Subperiods

	Dependent variable: $\text{hhd}_{i,t}$ (household debt growth)							
	no crisis		2007-2009		-2006		2010-	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Forecast errors	1.12**	0.94**	0.64*	0.13	1.21***	0.93**	0.16	0.077
	(0.45)	(0.41)	(0.35)	(0.55)	(0.41)	(0.37)	(0.26)	(0.26)
Observations	1969	1916	379	379	1188	1144	781	772
R^2	0.18	0.25	0.25	0.32	0.15	0.23	0.12	0.21
Add. controls		✓		✓		✓		✓

Note: All models include country and time fixed effects. Robust standard errors in parentheses. Additional controls are real GDP y-o-y growth, expected inflation, interest rates and forecaster dispersion (uncertainty). *** p<0.01, ** p<0.05, * p<0.1.

We investigate the contemporaneous comovement between forecast errors and credit growth using the following panel regression,

$$c_{i,t} = \gamma_i + \delta_t + \beta_1 f_{i,t} + X_{i,t}\beta_2 + \varepsilon_{i,t}, \tag{4.1}$$

where $c_{i,t}$ are the credit variables, γ_i is a country fixed effect, δ_t is a time fixed effect, β_1 is the regression coefficient of interest, $f_{i,t}$ are forecast errors, $X_{i,t}$ are controls and $\varepsilon_{i,t}$ is the error term. Credit variables, $c_{i,t}$, are known to be autocorrelated (Drehmann et al., 2018) and we have shown the same for the forecast errors $f_{i,t}$ (Figure 4.5). We therefore use robust standard errors. For the covariates $X_{i,t}$, we use real GDP growth (backward-looking, over last 12 months), expected inflation (also from the CEF), interest rates and forecast dispersion (standard deviation across panelists) and banking crises dummies by Laeven and Valencia (2012).

Table 4.2 displays the baseline results. The first column shows the results for the bivariate regression for which the estimated coefficient, $\hat{\beta}_1$, is positive and significant. This means that when professional forecasters were 1 percentage point too optimistic ($f_{it} = 1$), household debt growth was on average 0.91 percentage points higher. This association stays significant when we control for the states of the business cycle, in column (2), the expected ex ante real interest rate in column (3) and (4) and proxies for uncertainty in (5) and dummy variables for banking crises in model version (6).[6]

This establishes the main result: Periods of ex-post excessively optimistic GDP growth expectations also saw expansions in cyclical lending in the economy. In

[6]Adding quarterly dummies does not affect results.

Table 4.4. Regression by Country Groups

	Dependent variable: $hhd_{i,t}$ (household debt growth)								
	Industrialized			Developing			-China		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Forecast errors	0.49**	0.24	0.58***	1.47**	1.26**	1.54**	0.89**	0.71*	0.93**
	(0.20)	(0.18)	(0.19)	(0.49)	(0.47)	(0.64)	(0.36)	(0.35)	(0.37)
Observations	1588	1567	901	760	728	243	2316	2263	1144
R^2	0.28	0.37	0.21	0.31	0.42	0.45	0.17	0.26	0.23
Countries	18	18	18	14	14	14	31	31	31
Add. controls		✓	✓		✓	✓		✓	✓
< 2007			✓			✓			✓

Note: All models include country and time fixed effects. Robust standard errors in parentheses. A minus indicates excluding countries. Additional controls are real GDP y-o-y growth, expected inflation, interest rates and forecaster dispersion (uncertainty). *** p<0.01, ** p<0.05, * p<0.1.

the rest of this section, we explore the heterogeneity in our results by providing separate estimates for different time periods and country subgroups.

We first check whether the observed pattern might be driven by the global financial crisis of 2007-2009. This episode is not classified as a banking crises for all countries, but many countries experienced a strong recession nonetheless. During the crisis, real GDP dropped precipitously for most countries, but forecasters were slow to adapt their expectations (see Figure 4.3). The resulting forecast error is therefore strongly positive, suggesting along our line of argument that during the financial crisis people were far too optimistic about the path of their future incomes. In fact, we find the opposite (see Table 4.3); results hold even when we exclude the financial crisis (column (1)-(2)). They are also robust to estimating the model on the data only before the crisis (columns (5)-(6)), but not on data after the crisis (columns (7)-(8)). During the financial crisis that took place between 2007 to 2009 (columns (3)-(4)) the coefficient becomes insignificant which might be due to the lower number of observations.

In Table 4.4, we report results for country groups. The results for industrialized and developing countries point in the same directions. The greater macroeconomic volatility of developing countries might explain the larger estimates for these countries. When adding additional macroeconomic controls, the results are more pronounced for developing countries. The full model with all covariates is insignificant for the whole sample for industrialized countries, but the association holds before 2007. This is the most relevant period, as it coincides with the build-up of financial imbalances before the financial crisis of 2008. China is a special case as forecasters strongly underestimated its growth performance over several consecutive years. Our results become even stronger when we exclude it in columns (7)-(9).

Strikingly, the relationship between firm debt and forecast errors is mostly insignficant (Tables 4.A.3, 4.A.4 and 4.A.5). We therefore find that expectation errors about aggregate income are contemporaneously related to the growth rates of household debt, but not to non-financial firm debt. This might be a sign that is the expectations by households (or by banks about households), not expectations by firms (or by banks about firms) that are occasionally misaligned.

4.4.3 Robustness Checks

As explained before, there are different ways of converting *fixed event* to *fixed horizon* forecasts. Results are unchanged when we use the alternative Dovern et al. (2012) weights, as we show in Tables 4.A.6 to 4.A.11.

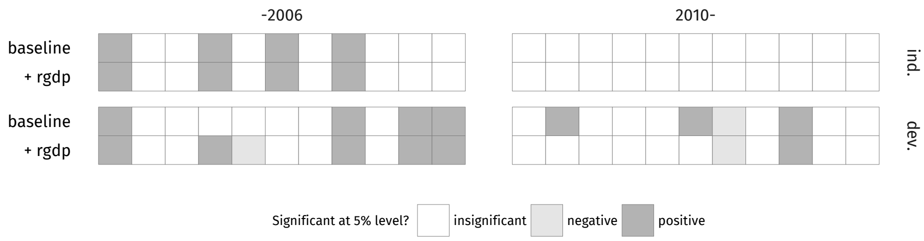


Figure 4.6. Robustness to Alternative Overoptimism Definitions

Note: Squares indicate the sign and significance of $\hat{\beta}$. The baseline is a bivariate of household debt growth on the variable of interest, controlling for time and country fixed effects and using robust standard errors. Columns are estimates using the alternative overoptimism definitions 1, 2, ..., 11. Left panels show results for the sample up to 2006 and right panes for after 2010. Top panels show results for industrialized countries and bottom panes those for developing countries.

Our prefered way of measuring overoptimism is to use forecast errors. One might also define these proxies differently, by taking into account the sign and persistence of errors and comparing them to trend output growth rates. We define eleven alternative ways (see Appendix Section 4.C) to define overoptimism and provide a concise summary of how results change in Figure 4.6. This plot shows the signs and significance of the $\hat{\beta}$ coefficient in Equation 4.1. We show results for the basline bivariate regression without covariates, but including time and country fixed effects and using robust standard errors. Secondly, we also control for the state of the business cycle by including current real GDP growth. We

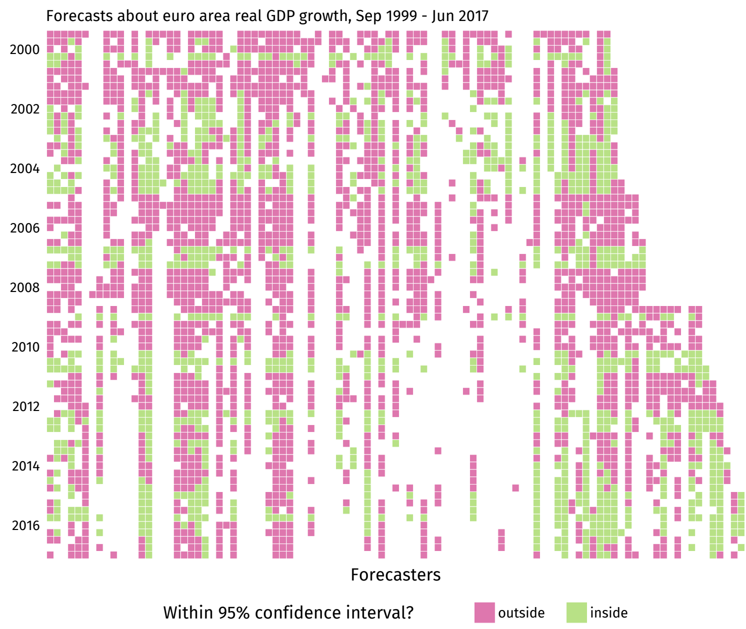


Figure 4.7. Overconfidence in the ECB SPF

Note: Shows whether subsequent realizations of 12 month ahead real euro area GDP growth was within the 95 percent prediction interval of individual forecasters. Columns show the predictions by individual forecasting firms with numbers from 001 to 115. Empty squares show missing data. *Source:* ECB *Survey of Professional Forecasters*

split samples into before 2006 and 2010 and into industrialized and developing country groups.

What becomes apparent is that the estimates are either positively significant or insignificant across most specifications. Especially before 2006, many of the alternative overoptimism definitions also return positive estimates for the comovement with household debt growth. This mirrors our baseline findings which were also only significant for the period before 2006. For developing countries, several of the alternative indicators still return positive estimates after 2010. This is also similar to the findings from Table 4.4.

4.5 Explaining Forecast Errors

4.5.1 Uncertainty of Individual Forecasters

In this paper, we analyze the predictions by professional forecasters. However, every firm only submits point forecasts and they do not report how certain they are about their predictions. We interpret differences between forecasts and subsequent realizations as forecast errors. But how can we do this if we do not know how uncertain forecasters were individually? They might have had very wide distributions of outcomes in mind, but were forced to submit point forecasts.

Our data does not permit us to analyze this further. So we turn to a dataset that does allow us to do so, the *Survey of Professional Forecasters* (SPF) by the European Central Bank.[7] In this survey, participants report a point forecast and subjective probabilities they assign to different bins. This allows us to calculate the width of individual forecasters' confidence bands. Figure 4.B.5 shows the mean of this value across forecasters.

On average, forecasters stated that they were 95 percent certain that one-year-ahead real GDP growth would lie in an interval with a width between 1.1 and 2.6 percentage points. However, Figure 4.7 shows that forecasters were far off with their forecasts. Only 34 percent of realizations lie within the 95 percent confidence bands. This immediately tells us that panelists were overconfident in their forecasts. This is puzzling as firms have no incentive to make such narrow predictions. Forecasters are not identified by name in the SPF and there is no scoring of predictive accuracy that might reward more aggressive predictions. Figure 4.B.5 displays an upward trend in the width of confidence bands, so participants have become more cautious with their predictions.

Overall, we take thees findings as a sign that panelists in the survey are indeed too confident about their forecasts. While we cannot with certainty extrapolate the findings from the ECB SPF to the CEF, the vast amount of overconfidence in the former strongly suggests that a related mechanism might explain the pronounced and persistent forecast errors that we find for many more countries in the CEF.

4.5.2 Forecast Revisions and Information Processing

We have documented that positive forecast errors are associated with debt growth in the economy and that positive forecast errors are likely to be a proxy for overoptimism of forecasters. But this begs the question why forecasters become overoptimistic in the first place. An active literature uses forecast *revisions* to analyze how panelists change their forecast when they receive new information. We use the methodology developed by Coibion and Gorodnichenko (2012) and Bordalo, Gennaioli, Ma, and Shleifer (2018) to test how forecasters in our sample react to new information and discuss how this effect varies over the credit cycle.

[7] A limitation is that this survey only makes forecasts for the euro area. So what we gain in detail, we loose in generalizability.

On the level of individual forecasters, we regress forecast errors on the change in the forecast over the last month:

$$f_{i,k,t} \;=\; \beta\, \mathrm{rev}_{i,k,t} + u_{i,k,t}\,. \tag{4.2}$$

Here, $f_{i,k,t}$ is the forecast error for country i, forecaster k in year t, $\mathrm{rev}_{i,k,t}$ are forecast revisions and $u_{i,k,t}$ is the residual.

Every panelist in the CEF makes two forecasts at the same time, one for the current year and one for the next. We previously aggregated those two fixed horizon forecasts to one fixed event forecasts, but that is not necessary now. Instead, we report separate estimates for the coefficient $\hat{\beta}$ for both forecasts the panelists make.

We also add some indicator variables for the month the forecasts were made in, to control for seasonal trends in information updating. Such patterns are likely, as new information is revealed at fixed points during the year, for example when new GDP forecasts are made public. Also, forecasts mechanically get better as the year progresses. Using microdata on the level of individual panelists shows the strength of the highly detailed dataset we use in this paper. It enables us to also control for the panelist which eliminates panelist idiosyncracies.

As explained by Coibion and Gorodnichenko (2012) and Bordalo et al. (2018), the coefficient $\hat{\beta}$ shows how forecasters react to new infomation. If panelists updated their forecasts rationally, we would not expect a significant relationship with the forecast error $f_{i,k,t}$. But say they received positive news and updated their forecasts upwards. If they overshot and reacted too strongly to new information, the resulting forecast error would be negative. This means that a negative $\hat{\beta}$ coefficient is symptomatic of *overreaction*. If they instead did not adjust their forecast upwards enough, forecast errors would be positive – a sign of *underreaction*.

This method has previously been used to differentiate between classes of macroeconomic models. Coibion and Gorodnichenko (2012) found evidence for underreaction, but Bordalo et al. (2018) argue that this is due to their use of consensus (mean) forecasts. They show that – at least for the United States – overreaction dominates.

Our results are shown in Figure 4.8 for each country, split for forecasts for the current and next year and using consensus and individual data. The figure shows the OLS coefficient and 95 percent confidence bands. There is strong evidence for overreaction. This holds for most countries in the sample using consensus or individual data. The results are more pronounced for the forecasts for the next year than for the current year. This is to be expected, as panelists have much more information on the current year and forecasts errors are much smaller

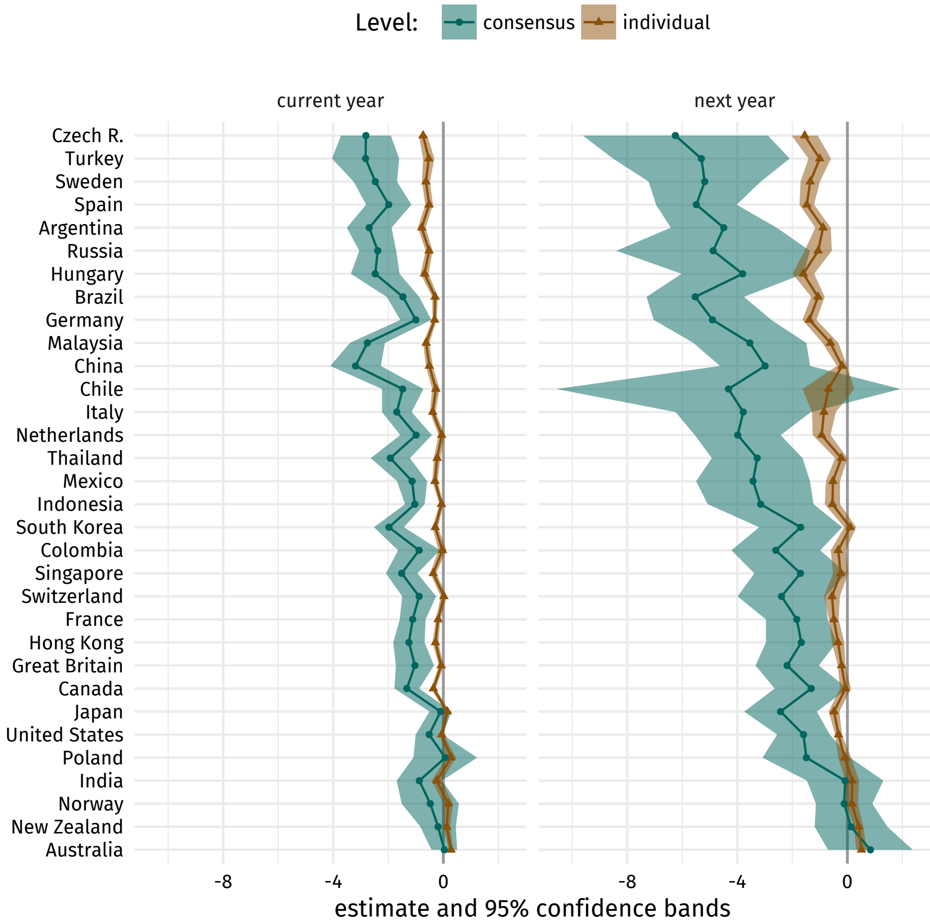


Figure 4.8. Information Processing

Note: Uses one-month forecast revisions for consensus (mean) and for individual forecasts on 1991-2016 monthly forecasts about annual real GDP growth. Shows separate estimates by country of regressing forecast errors on forecast revisions controlling for forecast month. Also controls for the panelists in the case of panelist-level data.

(and plausibly also better calibrated) as a result. In contrast to Bordalo et al. (2018) we also find stronger evidence for overreaction when using aggregate data as opposed to using microdata.

However, this association is strongly driven by the inclusion of the financial crisis period 2007 to 2009 during which period there was pronounced overreaction across countries. We also investigate how information processing changes through the credit cycle. For this, we run the same analysis as before, but pool the data from several countries to get a dense dataset that allows us to display highly detailed results. For this, we run a panel data analysis including country

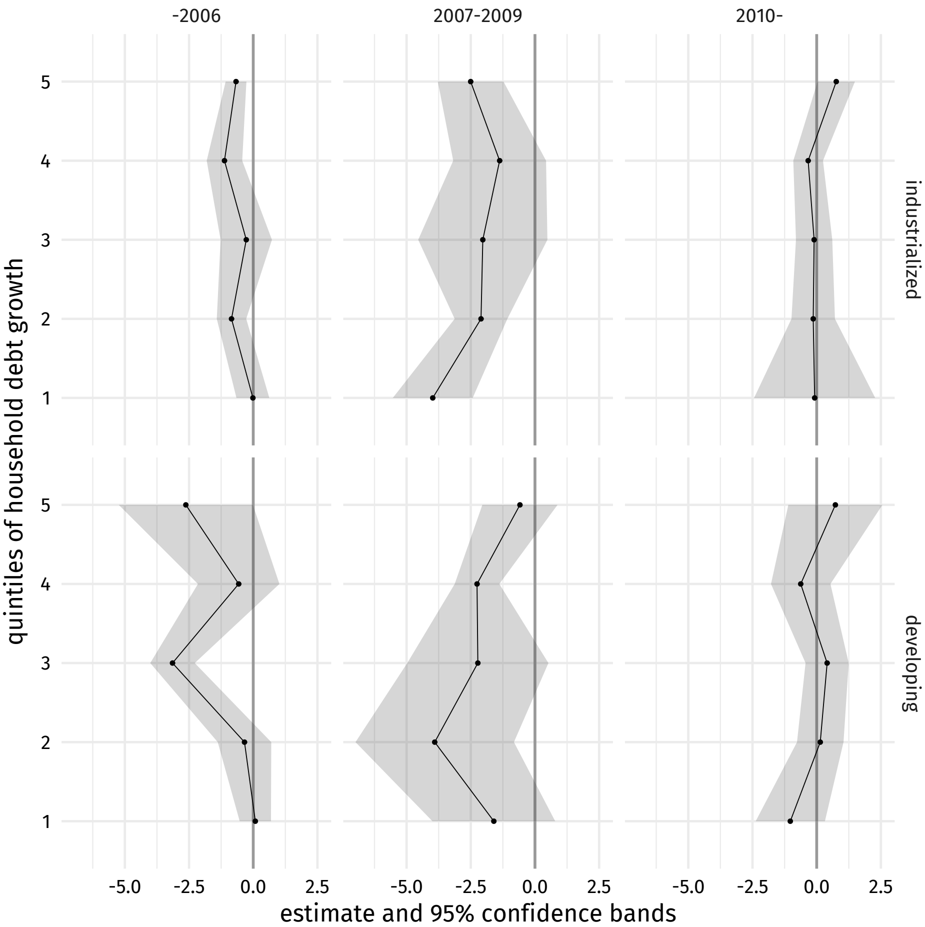


Figure 4.9. Information Processing over the Credit Cycle

Note: Uses one-month forecast revisions for consensus (mean) forecasts on 1991-2016 monthly forecasts about annual real GDP growth. Uses next year forecasts. Shows panel estimates including country fixed effects (but no time fixed effects), controlling for forecast month. Sample is split into five bins depending on country-specific household debt growth quintiles.

fixed effects. In addition, we partition the dataset depending on a country's position in its quintile of household debt growth. Figure 4.9 shows these results, separated by the periods before, during and after the financial crisis and split into industrialized and developing country groups.

Across specifications, we find that overreaction and insignificant results outnumber underreaction. Especially up to 2006 there is very robust evidence of overreaction throughout the household debt growth distribution. As mentioned before, the period from 2007 to 2009 was a period of strong overreaction. Most estimates are insignificant from 2010 onwards. This finding helps explain why forecasters - and other agents in the economy - might have become overoptimistic: During boom times of the Great Moderation before the financial crisis people received positive news and their expectations adjust upwards and overshot. If households form their expectations similarly to professional forecasters, they may have underestimated future risks and therefore took on too much debt in the run-up to the financial crisis.

4.6 Conclusion

This paper seeks to inform the discussion on the buildup of imbalances in the international financial system. We identify periods of positive and negative mistakes in output growth expectations of professional forecasters and show that these periods are characterized by strong credit growth. While household debt rises in such periods of excessive income expectations, firm debt does not respond. These findings are in line with theories in which biased income-savings decisions drive unsustainable debt booms, with harmful consequences for the economy.

We provide more detailed findings that reveal the psychological mechanisms for the formation of expectation errors. First, panelists (at least in the case of the euro area, where we can be sure) display strong signs of overprecision, so they are too confident about their predictions. Second, panelists overshoot when they receive new information. This second insight emerges from an analysis of forecast revisions using established methods from the literature. This overreaction of forecasts was particularly strong before 2006, when leverage in the financial system was growing.

A limitation of this study is the late start of the time series dimension with our first observations starting in 1989. This means that we miss important swings of the national and global financial cycle and are limited to the end of the "financial hockey stick" (Jordà et al., 2016). On the upside, the dataset used in this study has a broad international coverage of 32 countries, allowing us to control for circumstances that might be specific to individual countries and making it possible to report results separately for industrialized and developing countries. The results reported here hold for both subgroups, which is particularly striking considering the differences in macroeconomic volatilities and financial systems of these groups of countries. Access to the microdata of forecasts also allows us to control for the dispersion across predictions which serves as a proxy for uncertainty and to use forecaster fixed effects for some of our analyses.

A downside of our method is that surveys of professional forecasts might only be correlated weakly with expectations of households or firms. A better way forward would be to collect such expectations across groups of agents and for many countries. Rozsypal and Schlafmann (2017), using consumer surveys, are able to measure expectations across household income distribution. Expanding such an approach across countries is a promising direction for future research.

References

Alessi, Lucia and Carsten Detken (2017): "Identifying Excessive Credit Growth and Leverage." *Journal of Financial Stability*. [95]

Bachmann, Rüdiger, Steffen Elstner, and Eric R. Sims (2013): "Uncertainty and Economic Activity: Evidence from Business Survey Data." *American Economic Journal: Macroeconomics*, 5 (2), 217–249. [99]

Baron, Matthew and Wei Xiong (2017): "Credit Expansion and Neglected Crash Risk." *Quarterly Journal of Economics*, 132 (2), 713–764. [95]

Batchelor, Roy (2001): "How Useful are the Forecasts of Intergovernmental Agencies? The IMF and OECD versus the Consensus." *Applied Economics*, 33 (2), 225–235. [99]

Berger, Helge, Michael Ehrmann, and Marcel Fratzscher (2011): "Geography, Skills or Both: What Explains Fed Watchers' Forecast Accuracy of US Monetary Policy?" *Journal of Macroeconomics*, 33 (3), 420–437. [98]

Blanchard, Olivier J., Jean-Paul L'Huillier, and Guido Lorenzoni (2013): "News, Noise, and Fluctuations: An Empirical Exploration." *American Economic Review*, 103 (7), 3045–3070. [98]

Bordalo, Pedro, Nicola Gennaioli, Yueran Ma, and Andrei Shleifer (2018): "Overreaction in Macroeconomic Expectations." Unpublished manuscript. [96, 112–114]

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer (2017): "Diagnostic Expectations and Credit Cycles." *Journal of Finance*, 73 (1), 199–227. [97]

Breitung, Jörg and Malte Knüppel (2017): "How Far Can we Forecast? Statistical Tests of the Predictive Content." Unpublished manuscript. [99]

Capistrán, Carlos and Allan Timmermann (2009): "Disagreement and Biases in Inflation Expectations." *Journal of Money, Credit and Banking*, 41 (2-3), 365–396. [98]

Chen, Sophia and Romain Rancière (2016): "Financial Information and Macroeconomic Forecasts." IMF WP/16/251. [95]

Cochrane, John H. (2017): "Macro-Finance." *Review of Finance*, 2 (3), 945–985. [97]

Coibion, Olivier and Yuriy Gorodnichenko (2012): "What Can Survey Forecasts Tell Us about Information Rigidities?" *Journal of Political Economy*, 120 (1), 116–159. [96, 112, 113]

De Stefani, Alessia (2017): "Waves of Optimism: House Price History, Biased Expectations and Credit Cycles." Unpublished manuscript. [95]

Dovern, Jonas, Ulrich Fritsche, and Jiri Slacalek (2012): "Disagreement among Forecasters in G7 Countries." *Review of Economics and Statistics*, 94 (4), 1081–1096. [99, 102, 103, 110, 122–125]

Drehmann, Mathias, Mikael Juselius, and Anton Korinek (2018): "Going With the Flows: New Borrowing, Debt Service and the Transmission of Credit Booms." NBER Working Paper No. 24549. [108]

Gennaioli, Nicola, Yueran Ma, and Andrei Shleifer (2016): "Expectations and Investment." In *NBER Macroeconomics Annual*. Vol. 30. 1, 379–431. [97]

Greenwood, Robin and Samuel G. Hanson (2013): "Issuer Quality and Corporate Bond Returns." *The Review of Financial Studies*, 26 (6), 1483–1525. [95]

Higgins, Patrick C. and Tao Zha (2015): "China's Macroeconomic Time Series: Method and Implications." Federal Reserve Bank of Atlanta. [120]

Jordà, Òscar, Moritz Schularick, and Alan M. Taylor (2016): "Macrofinancial History and the New Business Cycle Facts." In *NBER Macroeconomics Annual*. Ed. by Martin Eichenbaum and Jonathan A. Parker. Vol. 31. NBER. University of Chicago Press. [116]

Kaminsky, Graciela L. and Carmen M. Reinhart (1999): "The Twin Crises: The Causes of Banking and Balance-of-Payments Problems." *American Economic Review*, 89 (3), 473–500. [95]

Kindleberger, Charles P. (1978): *Manias, Panics, and Crashes: A History of Financial Crises*. Macmillan. [97]

Knüppel, Malte and Andreea L. Vladu (2016): "Approximating Fixed-Hyorizon Forecasts Using Fixed-Event Forecasts." Bundesbank Discussion Paper No 28/2016. [102, 103, 126]

Kučinskas, Simas and Florian Peters (2018): "Measuring Biases in Expectation Formation." Unpublished manuscript. [104]

Laeven, Luc and Fabian Valencia (2012): "Systemic Banking Crises Database: An Update." IMF WP/12/163. [107, 108, 120, 122, 124, 128, 129]

Lopez-Salido, David, Jeremy Stein, and Egon Zakrajšek (2017): "Credit-Market Sentiment and the Business Cycle." *Quarterly Journal of Economics*, 132 (1), 1373–1426. [95]

Loungani, Prakash (2001): "How Accurate are Private Sector Forecasts? Cross-country Evidence from Consensus Forecasts of Output Growth." *International Journal of Forecasting*, 17 (3), 419–432. [99]

McNees, Stephen K. (1992): "The Uses and Abuses of 'Consensus' Forecasts." *Journal of Forecasting*, 11, 703–710. [103]

Mian, Atif, Amir Sufi, and Emil Verner (2017a): "Household Debt and Business Cycles Worldwide." *Quarterly Journal of Economics*, 132 (4), 1755–1817. [95–97]

Mian, Atif, Amir Sufi, and Emil Verner (2017b): "How Do Credit Supply Shocks Affect the Real Economy? Evidence from the United States in the 1980s." Unpublished manuscript. [97]

Rozsypal, Filip and Kathrin Schlafmann (2017): "Overpersistence Bias in Individual Income Expectations and its Aggregate Implications." CEPR Discussion Paper 12028. [95, 117]

Rülke, Jan-Christoph, Maria Silgoner, and Julia Wörz (2016): "Herding Behavior of Business Cycle Forecaster." *International Journal of Forecasting*, 32 (1), 23–33. [99]

Schularick, Moritz and Alan M. Taylor (2012): "Credit Booms Gone Bust: Monetary Policy, Leverage Cycles and Financial Crises, 1870–2008." *American Economic Review*, 102 (2), 1029–61. [95]

Shiller, Robert J. (2000): *Irrational Exuberance*. Princeton University Press. [100]

Appendix 4.A Additional Tables

Table 4.A.1. Country Data Sources and Summary Statistics

	Country	Dev.	Source	First	Last	Obs.	Nr. forec.	CI forec.
1	Argentina	✓	indec	1995 Q2	2015 Q4	71	20 (±5)	[-1.3, 1.9]
2	Australia		oecd	1991 Q1	2016 Q4	104	17 (±2)	[-0.2, 0.2]
3	Brazil	✓	oecd	1996 Q2	2016 Q1	73	18 (±3)	[0.0, 1.3]
4	Canada		oecd	1989 Q4	2016 Q4	109	16 (±2)	**[0.1, 0.7]**
5	Chile	✓	oecd	2003 Q4	2016 Q4	53	18 (±2)	[0.0, 1.3]
6	China	✓	Atlanta Fed[1]	2007 Q1	2014 Q4	32	18 (±3)	[-0.6, 0.3]
7	Colombia	✓	oecd	2000 Q2	2016 Q4	65	13 (±2)	[-0.8, 0.2]
8	Czech R.		oecd	1998 Q3	2016 Q4	57	18 (±2)	[-0.4, 1.2]
9	France		oecd	1989 Q4	2016 Q4	109	19 (±3)	**[0.1, 0.8]**
10	Germany		oecd, destatis	1989 Q4	2016 Q4	109	28 (±2)	[-0.2, 0.7]
11	Great Britain		oecd	1989 Q4	2016 Q4	109	28 (±5)	[-0.3, 0.4]
12	Hong Kong		C&SD[2]	1994 Q4	2016 Q1	86	16 (±2)	[-0.5, 1.4]
13	Hungary	✓	oecd	1998 Q3	2016 Q4	57	17 (±3)	[-0.2, 1.4]
14	India	✓	oecd	2008 Q2	2016 Q4	35	16 (±2)	[-1.0, 0.8]
15	Indonesia	✓	fred, aric[3]	2002 Q4	2015 Q3	52	14 (±2)	[-0.4, 0.2]
16	Italy		oecd	1996 Q1	2016 Q4	84	15 (±2)	**[0.4, 1.3]**
17	Japan		oecd	1989 Q1	2016 Q4	109	20 (±2)	**[0.2, 1.0]**
18	Malaysia	✓	aric[3]	2007 Q1	2015 Q3	35	15 (±2)	[-0.8, 1.5]
19	Mexico	✓	fred	1995 Q4	2016 Q1	71	19 (±3)	**[0.2, 1.4]**
20	Netherlands		oecd	1995 Q1	2016 Q4	88	10 (±2)	[-0.5, 0.3]
21	New Zealand		oecd	1994 Q4	2016 Q4	89	14 (±1)	[-0.6, 0.2]
22	Norway		oecd	1998 Q2	2016 Q4	75	10 (±2)	[0.0, 0.9]
23	Poland	✓	oecd	2002 Q1	2016 Q4	50	19 (±3)	[-0.6, 0.4]
24	Russia	✓	oecd, fred	1999 Q1	2016 Q4	56	17 (±3)	[-0.7, 1.8]
25	Singapore		singstat[4]	1994 Q4	2016 Q1	86	15 (±2)	[-1.6, 0.7]
26	South Korea		oecd	1994 Q4	2016 Q4	89	16 (±2)	[-0.6, 1.1]
27	Spain		oecd	1995 Q1	2016 Q4	88	15 (±2)	[-0.4, 0.3]
28	Sweden		oecd	2000 Q4	2016 Q4	88	14 (±2)	[-0.5, 0.3]
29	Switzerland		oecd	2000 Q4	2016 Q4	65	14 (±2)	[-0.5, 0.3]

Continued.

Table 4.A.2. Country Data Sources and Summary Statistics

	Country	Dev.	Source	First	Last	Obs.	Nr. forec.	CI forec.
30	Thailand	✓	BoT[5]	1994 Q4	2015 Q3	84	14 (±3)	**[0.3, 2.3]**
31	Turkey	✓	oecd	1998 Q3	2016 Q4	57	15 (±3)	[-2.3, 0.5]
32	USA		oecd	1989 Q4	2016 Q4	109	27 (±3)	[-0.1, 0.5]

Note: Developing country classifications ("Dev.") are according to IMF "World Economic Outlook Report" (2017). "Obs." are the number of quarterly observations with complete data for a respective country. "Nr. forec" are the means of the number of forecasts used to calculate an aggregated mean quarterly forecast with standard errors in parentheses. "CI forec." are the 95% confidence intervals of the forecast errors of 4-quarters ahead real GDP growth (boldface shows significance).
[1]: Higgins and Zha (2015), Atlanta Fed; [2]: Census and Statistics Department
[3]: Asia Regional Integration Center; [4]: Statistics Singapore; [5]: Bank of Thailand

Table 4.A.3. Baseline Regressions (Firm Debt)

	Dependent variable: $fd_{i,t}$ (firm debt growth)					
	(1)	(2)	(3)	(4)	(5)	(6)
Forecast errors	0.12	0.20	0.22	0.22	0.21	0.23
	(0.33)	(0.29)	(0.14)	(0.15)	(0.14)	(0.14)
Real GDP gr.		-0.13	0.17	0.18	0.13	0.13
		(0.43)	(0.36)	(0.34)	(0.33)	(0.34)
Exp. inflation			1.48^{**}	1.42^{***}	1.42^{***}	1.41^{***}
			(0.56)	(0.50)	(0.50)	(0.49)
Interest rate				0.057	0.068	0.072
				(0.13)	(0.13)	(0.13)
Uncertainty					-1.46	-1.38
					(1.66)	(1.67)
Banking crises						-1.24
						(1.52)
Observations	2330	2283	2283	2277	2277	2277
R^2	0.13	0.14	0.30	0.30	0.30	0.30

Note: All models include country and time fixed effects. Robust standard errors in parentheses. GDP: real GDP y-o-y growth. Exp. inflation: 12-month ahead expected CPI growth. Exp. inflation: 12-month ahead expected CPI growth. Uncertainty: Standard deviation of forecasts. Banking crises are defined by Laeven and Valencia (2012).*** p<0.01, ** p<0.05, * p<0.1.

Table 4.A.4. Regressions by Subperiods (Firm Debt)

	Dependent variable: $fd_{i,t}$ (firm debt growth)							
	no crisis		2007-2009		-2006		2010-	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Forecast errors	0.085	0.19	0.32	0.090	-0.016	0.23	0.46^{**}	0.46^{*}
	(0.42)	(0.15)	(0.29)	(0.43)	(0.55)	(0.20)	(0.22)	(0.24)
Observations	1951	1898	379	379	1170	1126	781	772
R^2	0.095	0.28	0.34	0.38	0.090	0.34	0.19	0.21
Add. controls		✓		✓		✓		✓

Note: All models include country and time fixed effects. Robust standard errors in parentheses. Additional controls are real GDP y-o-y growth, expected inflation, interest rates and forecaster dispersion (uncertainty). *** p<0.01, ** p<0.05, * p<0.1.

Table 4.A.5. Regression by Country Groups (Firm Debt)

	Dependent variable: $fd_{i,t}$ (firm debt growth)								
	Industrialized			Developing			-China		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Forecast errors	0.41^{**}	0.11	0.17	-0.096	0.25	0.099	0.094	0.18	0.23
	(0.15)	(0.17)	(0.15)	(0.51)	(0.30)	(0.64)	(0.34)	(0.14)	(0.20)
Observations	1570	1549	883	760	728	243	2298	2245	1126
R^2	0.26	0.38	0.35	0.26	0.42	0.53	0.14	0.31	0.34
Countries	18	18	18	14	14	14	31	31	31
Add. controls		✓	✓		✓	✓		✓	✓
< 2007			✓			✓			✓

Note: All models include country and time fixed effects. Robust standard errors in parentheses. A minus indicates excluding countries. Additional controls are real GDP y-o-y growth, expected inflation, interest rates and forecaster dispersion (uncertainty). *** p<0.01, ** p<0.05, * p<0.1.

Table 4.A.6. Baseline Regressions (Alternative Weighting)

	Dependent variable: $hhd_{i,t}$ (household debt growth)					
	(1)	(2)	(3)	(4)	(5)	(6)
Forecast errors	1.05^{***}	0.69^{**}	0.72^{**}	0.73^{**}	0.73^{**}	0.70^{**}
	(0.34)	(0.31)	(0.32)	(0.33)	(0.35)	(0.34)
Real GDP gr.		1.11^{***}	1.23^{***}	1.17^{***}	1.19^{***}	1.21^{***}
		(0.30)	(0.29)	(0.27)	(0.30)	(0.31)
Exp. inflation			0.63^{***}	0.90^{***}	0.91^{***}	0.92^{***}
			(0.19)	(0.27)	(0.26)	(0.25)
Interest rate				-0.23	-0.24	-0.25
				(0.23)	(0.21)	(0.21)
Uncertainty					0.76	0.59
					(3.34)	(3.17)
Banking crises						2.28
						(4.43)
Observations	2348	2301	2301	2295	2295	2295
R^2	0.19	0.24	0.26	0.26	0.26	0.26

Note: Uses alternative Dovern et al. (2012) weighting. All models include country and time fixed effects. Robust standard errors in parentheses. GDP: real GDP y-o-y growth. Exp. inflation: 12-month ahead expected CPI growth. Exp. inflation: 12-month ahead expected CPI growth. Uncertainty: Standard deviation of forecasts. Banking crises are defined by Laeven and Valencia (2012).*** p<0.01, ** p<0.05, * p<0.1.

Table 4.A.7. Regressions by Subperiods (Alternative Weighting)

	Dependent variable: $hhd_{i,t}$ (household debt growth)							
	no crisis		2007-2009		-2006		2010-	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Forecast errors	1.27^{***}	0.96^{**}	0.56^{*}	-0.022	1.34^{***}	0.90^{***}	0.29	0.098
	(0.42)	(0.40)	(0.30)	(0.46)	(0.40)	(0.32)	(0.20)	(0.25)
Observations	1969	1916	379	379	1188	1144	781	772
R^2	0.19	0.25	0.25	0.31	0.17	0.23	0.13	0.21
Add. controls		✓		✓		✓		✓

Note: Uses alternative Dovern et al. (2012) weighting. All models include country and time fixed effects. Robust standard errors in parentheses. Additional controls are real GDP y-o-y growth, expected inflation, interest rates and forecaster dispersion (uncertainty). *** p<0.01, ** p<0.05, * p<0.1.

Table 4.A.8. Regression by Country Groups (Alternative Weighting)

	Dependent variable: $hhd_{i,t}$ (household debt growth)								
	Industrialized			Developing			-China		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Forecast errors	0.52***	0.24	0.57***	1.69***	1.28**	1.45**	1.04***	0.72*	0.90***
	(0.18)	(0.18)	(0.18)	(0.45)	(0.49)	(0.53)	(0.35)	(0.36)	(0.32)
Observations	1588	1567	901	760	728	243	2316	2263	1144
R^2	0.28	0.37	0.22	0.33	0.42	0.44	0.19	0.27	0.23
Countries	18	18	18	14	14	14	31	31	31
Add. controls		✓	✓		✓	✓		✓	✓
< 2007			✓			✓			✓

Note: Uses alternative Dovern et al. (2012) weighting. All models include country and time fixed effects. Robust standard errors in parentheses. A minus indicates excluding countries. Additional controls are real GDP y-o-y growth, expected inflation, interest rates and forecaster dispersion (uncertainty). *** p<0.01, ** p<0.05, * p<0.1.

Table 4.A.9. Baseline Regressions (Firm Debt, Alternative Weighting)

	Dependent variable: $fd_{i,t}$ (firm debt growth)					
	(1)	(2)	(3)	(4)	(5)	(6)
Forecast errors	-0.029	0.051	0.12	0.12	0.11	0.12
	(0.44)	(0.37)	(0.20)	(0.20)	(0.19)	(0.19)
Real GDP gr.		-0.12	0.16	0.18	0.14	0.14
		(0.37)	(0.33)	(0.32)	(0.30)	(0.31)
Exp. inflation			1.49^{**}	1.42^{***}	1.41^{***}	1.41^{***}
			(0.55)	(0.50)	(0.49)	(0.49)
Interest rate				0.062	0.072	0.075
				(0.13)	(0.13)	(0.13)
Uncertainty					-1.28	-1.21
					(1.89)	(1.91)
Banking crises						-0.93
						(1.54)
Observations	2330	2283	2283	2277	2277	2277
R^2	0.13	0.14	0.30	0.30	0.30	0.30

Note: Uses alternative Dovern et al. (2012) weighting. All models include country and time fixed effects. Robust standard errors in parentheses. GDP: real GDP y-o-y growth. Exp. inflation: 12-month ahead expected CPI growth. Exp. inflation: 12-month ahead expected CPI growth. Uncertainty: Standard deviation of forecasts. Banking crises are defined by Laeven and Valencia (2012).*** p<0.01, ** p<0.05, * p<0.1.

Table 4.A.10. Regressions by Subperiods (Firm Debt, Alternative Weighting)

	Dependent variable: $fd_{i,t}$ (firm debt growth)							
	no crisis		2007-2009		-2006		2010-	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Forecast errors	-0.14	0.036	0.34	0.15	-0.35	0.030	0.48^{**}	0.47^{*}
	(0.54)	(0.21)	(0.27)	(0.42)	(0.74)	(0.28)	(0.22)	(0.23)
Observations	1951	1898	379	379	1170	1126	781	772
R^2	0.096	0.28	0.34	0.38	0.098	0.34	0.19	0.21
Add. controls		✓		✓		✓		✓

Note: Uses alternative Dovern et al. (2012) weighting. All models include country and time fixed effects. Robust standard errors in parentheses. Additional controls are real GDP y-o-y growth, expected inflation, interest rates and forecaster dispersion (uncertainty). *** p<0.01, ** p<0.05, * p<0.1.

Table 4.A.11. Regression by Country Groups (Firm Debt, Alternative Weighting)

	Dependent variable: $fd_{i,t}$ (firm debt growth)								
	Industrialized			Developing			-China		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Forecast errors	0.43^{**}	0.094	0.12	-0.40	0.062	-0.31	-0.062	0.071	0.030
	(0.16)	(0.19)	(0.16)	(0.66)	(0.36)	(0.75)	(0.44)	(0.19)	(0.28)
Observations	1570	1549	883	760	728	243	2298	2245	1126
R^2	0.26	0.38	0.36	0.26	0.42	0.53	0.14	0.31	0.34
Countries	18	18	18	14	14	14	31	31	31
Add. controls		✓	✓		✓	✓		✓	✓
< 2007			✓			✓			✓

Note: Uses alternative Dovern et al. (2012) weighting. All models include country and time fixed effects. Robust standard errors in parentheses. A minus indicates excluding countries. Additional controls are real GDP y-o-y growth, expected inflation, interest rates and forecaster dispersion (uncertainty). *** p<0.01, ** p<0.05, * p<0.1.

Appendix 4.B Additional Figures

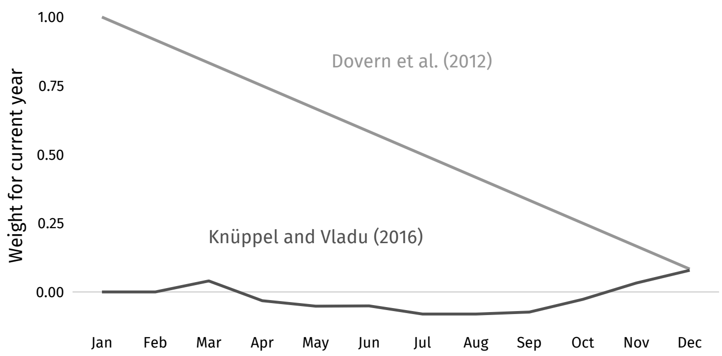


Figure 4.B.1. Weights to Convert Fixed Event to Fixed Horizon Forecasts

Note: The weight put on the forecast for the subsequent year is one minus the weight for the current year. The Knüppel and Vladu (2016) values are shown for $\rho = 0$.

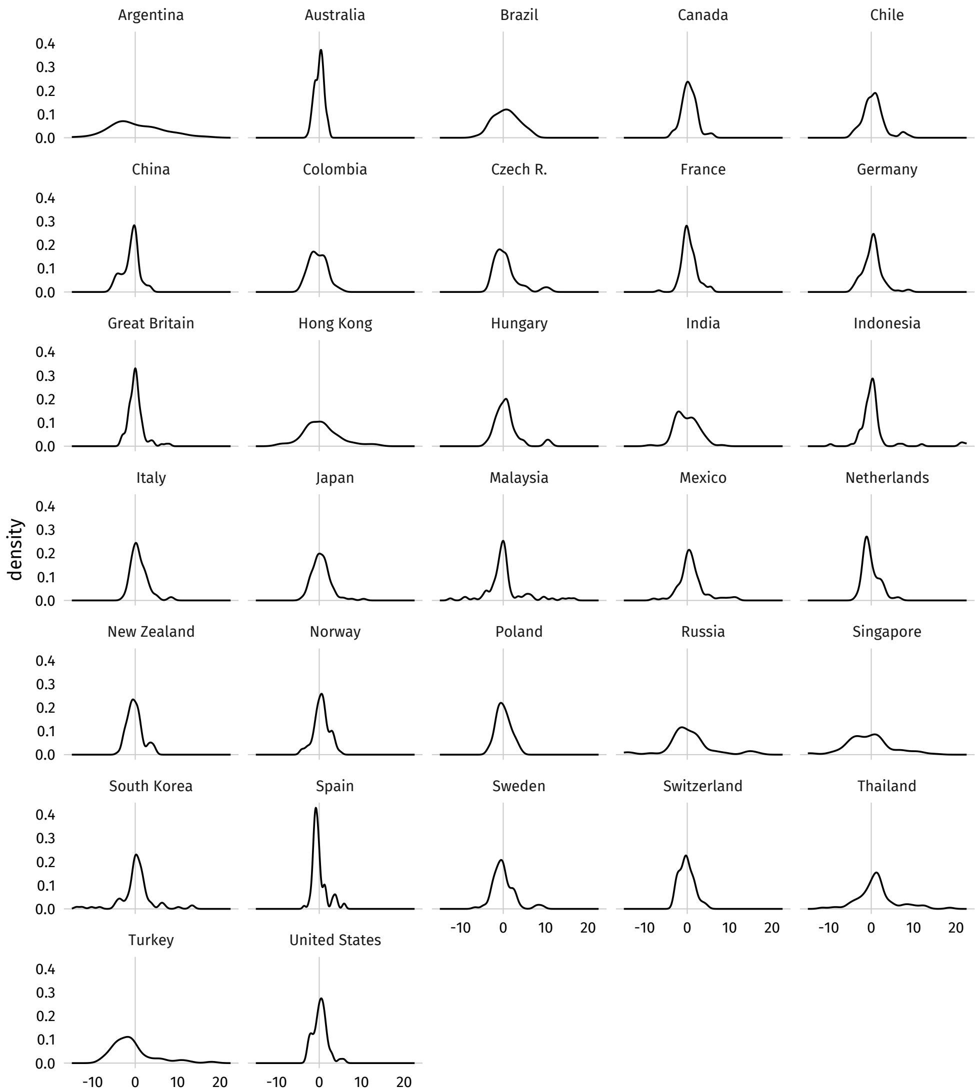


Figure 4.B.2. Distribution of Forecast Errors

Note: Kernel density plots of the distribution of forecast errors across countries. Positive values show real GDP growth (4 quarters ahead) mean (consensus) forecasts larger than realizations.

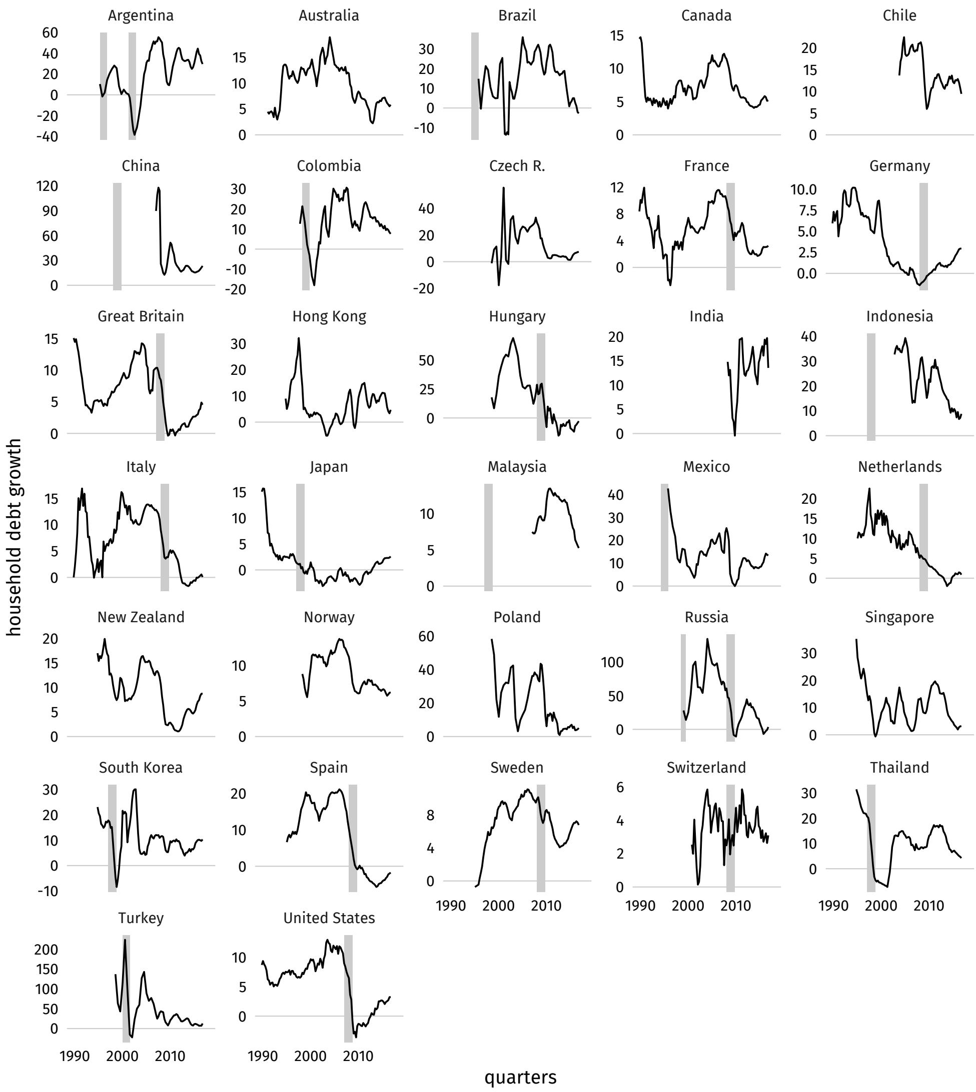


Figure 4.B.3. Household Debt Growth

Note: Backward-looking y-o-y growth rates in credit to households and non-profit organizations serving households. Bars indicate banking crises as classified by Laeven and Valencia (2012). *Source:* BIS and own calculations.

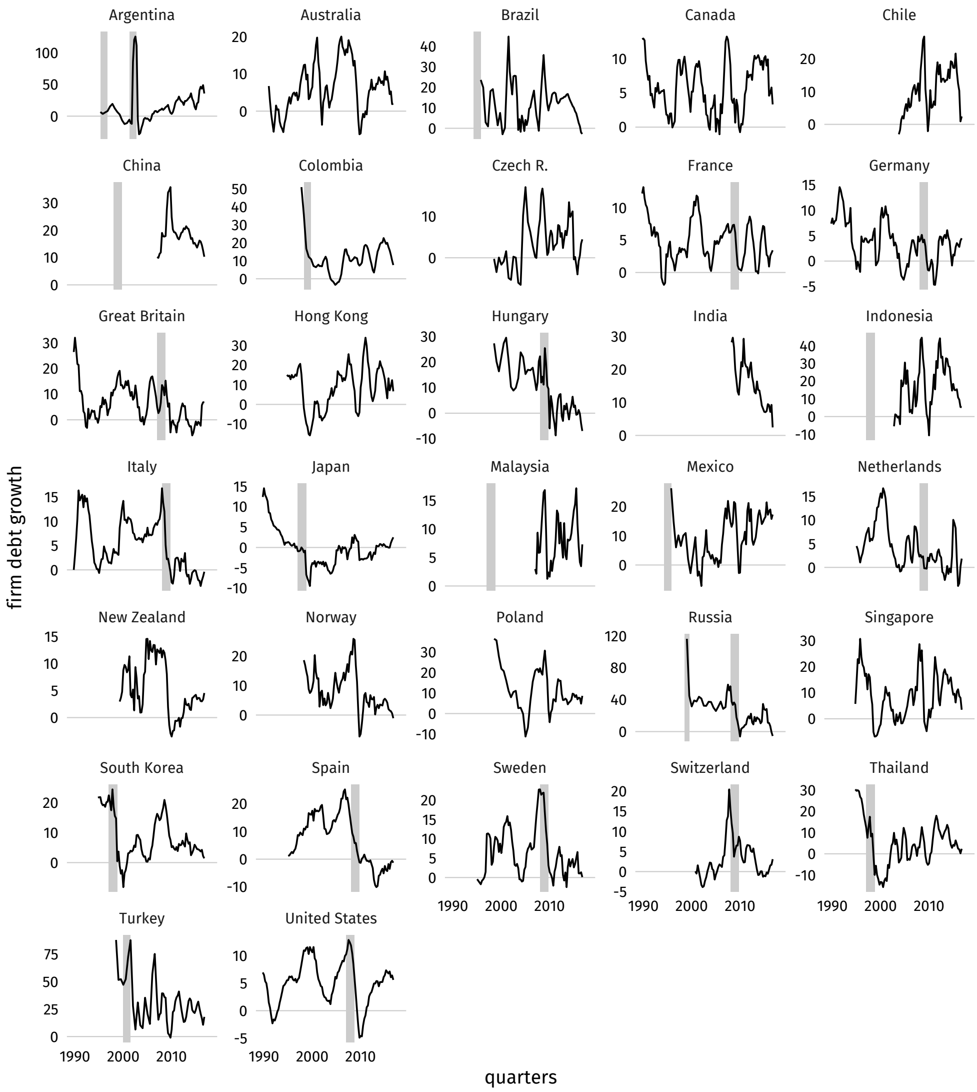


Figure 4.B.4. Firm Debt Growth

Note: Backward-looking y-o-y growth rates in credit to non-financial (private and public) firms. Bars indicate banking crises as classified by Laeven and Valencia (2012). *Source:* BIS and own calculations.

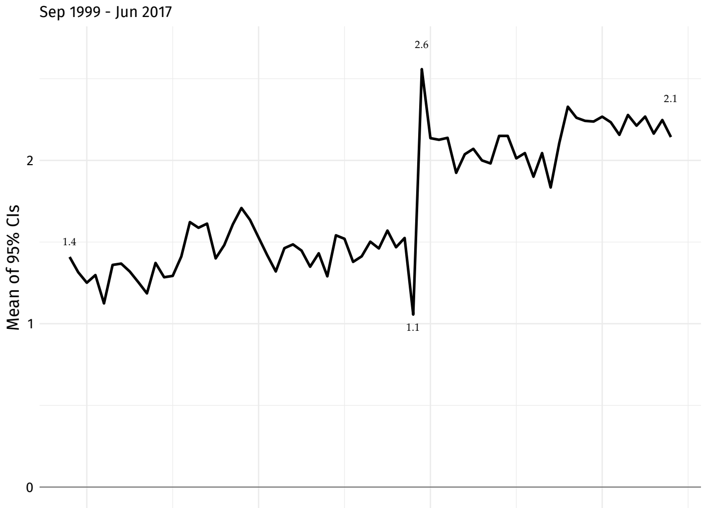


Figure 4.B.5. Individual Uncertainty

Source: ECB SPF

Appendix 4.C Alternative Overoptimism Definitions

We analyze how our results change with different definitions for overoptimism. For this, we also define overoptimism as follows:

1. Forecast errors as raw values. *(Continous variable)*
2. Dummies for periods with positive forecast errors. *(Categorical variable)*
3. Uses only positive forecast errors scaled by magnitude of forecast error. Assigns missing values if forecast error is negative. *(Continous variable)*
4. Same as 3., but assigns zero if forecast error is negative. *(Continous variable)*
5. Dummies if both realizations and forecast errors are positive. *(Categorical variable)*
6. Same as 5., but uses absolute magnitude of forecast error in those periods. Assigns zero instead. *(Continous variable)*
7. Uses forecast errors in periods where forecast errors are positive for at least three continuous periods. Assigns missing values if condition is not met. *(Continous variable)*
8. Same as 7., but assigns zeros if condition is not met. *(Continous variable)*
9. Realizations and forecast errors are positive for at least three continuous periods. *(Categorical variable)*
10. Dummy periods where forecast errors are positive and consensus forecasts are above the long-run real GDP trend. *(Categorical variable)*
11. Scales forecasts by the share of individual forecasts per year that are above the long-run real GDP growth trend. *(Continous variable)*

The long-run growth trends in versions 10. and 11. are taken from the *Penn World Tables 9.0* and are estimated as ten-year moving averages, interpolated to quarterly frequencies.