Do Market-oriented Reforms Explain China's Economic Growth?

Amit Khandelwal, Columbia
Suresh Naidu, Columbia
Heiwai Tang, Johns Hopkins
Qing Zhang, Columbia

Aug 2018 (Extremely preliminary. Please do not cite.)

Abstract

The Chinese economy has been growing at a rate of 10% on average since the early 1980s. Many researchers attribute China's stellar economic performance to its government's active pro-market reforms. In this paper, we apply state-of-the-art textual language processing techniques to establish several stylized facts about the evolution of China's institutional landscape using 1.4 million legal documents issued by its government over the past four decades. We find that (1) the share of of laws and regulations issued by local governments in our corpus has been rising; (2) the share of "informal" policies (those that lack a formal legal status) has been increasing; (3) the share of laws on economic issues has been gradually declining, and (4) the Chinese government has been constantly nullifying existing laws, especially after 2000. We then quantify the degree of market orientation of China's policies within and between legal documents using word embedding techniques, and relate them to provincial economic growth and other macroeconomic outcomes. Preliminary empirical results show that only a small part of provincial economic and foreign direct investment growth can be explained by the focus on neoliberal ideas and reforms in the country's laws and regulations.

1 Introduction

The Chinese economy has been growing at a spectacular rate of 10% on average since the beginning of market reforms in the early 1980s. Paradoxically, the country is still widely considered to have poor property rights protection, high corruption, substantial costs of doing business, and a legal system that is heavily influenced by the will of the Communist Party. For instance, according to the World Bank's Doing Business indicators, China was ranked 78 in 2017, below Colombia, El Salvador, Indonesia, Ukraine, and many much less developed economies.

All these *de jure* indicators of weak institutions in China are at odds with the conventional wisdom that good domestic institutions are the prerequisites for growth (La Porta et al. 2008).[1] That said, it is well-known that the *de jure* institutions are not binding in many countries, due to selective implementation of laws and regulations (Hallward-Driemeier and Pritchett, 2015). China is a particularly special case given its entrenched state-private relationships, which gave rise to an economy that was often characterized as crony capitalism (Bai, Hsieh and Song, 2018). An extensive literature has studied a host of reasons for China's economic miracle,[2] proposing determinants ranging from the country's political system's emphasis on meritocracy[3] and regional decentralization of economic policies[4], to factional competition in the central government.[5]

In this paper, we seek to study the evolution of China's *de facto* legal landscape in the past four decades by applying the state-of-the-art natural language processing (NLP) techniques to analyze the corpus of 1.4 million legal documents issued by the Chinese government. Our corpus includes the universe of formal laws, and close to the entirety of policies and regulations ever issued by various departments of both local and central governments in China since 1978. The first goal of the paper is to establish several stylized facts about China's changing laws and regulations. We then gauge the degree of market orientation in the legal documents across time and space using word embedding techniques to find similarity between the textual content of the legal documents and three sets of anchor words representing Marxist and neoliberal languages. The final part of the paper examines whether our constructed measures of market orientation of legal documents are correlated with provincial macroeconomic outcomes.

We find the following stylized facts in our corpus of laws and regulations: (1) the share of laws issued by local governments in our corpus of laws and regulations has been rising; (2)

[1] Sala-i-Martin (1997), Acemoglu, Johnson and Robinson (2005), Acemoglu et al. (2014).
[2] See comprehensive reviews by Malesky and London (2014) and Xu (2011).
[3] Li and Zhou (2005), Masina (2006), Nee and Opper (2012).
[4] Huang (1998), Roland (2000), Blanchard and Shleifer (2001), Jin, Qian, and Weingast (2005).
[5] Cai and Treisman (2006), Shih, Adolph, and Liu (2012).

the share of "informal" policies (those that lack a formal legal status) has been increasing; (3) the share of laws on economic issues has been gradually declining, and (4) the Chinese government has been constantly nullifying existing laws, especially after 2000.

To be more systematic in searching for the content in legal documents that is related to market orientation versus central planning, we use word embedding, a recent technique used in many recent NLP studies. The ultimate goal of word embedding is to identify words in a document that convey ideas closely related to market orientation. We then compute the intensity of those market-oriented words in each document. For example, we want to find words that are used in similar contexts as "privatization". To perform such task, we need to first construct a set of anchor words from external sources that suggest market orientation (or the inverse of it). To this end, we choose to obtain the anchors from the following three sources:

- The top 50 "Marxist" words in Karl Marx's famous class *Das Kapital*.
- The full list of "neoclassical economics" words as appeared in the glossary of Mankiw's best-selling introductory economic textbooks for college students.
- The 10 "neoliberal" words: the 10 economic policy prescriptions for economic growth, proposed by John Williams in 1990 as the so-called "Washington Consensus".

Our NLP exercise shows that the Chinese government has been actively building a market-oriented legal infrastructure by introducing pro-market laws and policies from the mid 1980s up to the beginning of 2000s. For instance, laws on the books in 2002 contain over two more standard deviations of "neoliberal" language than in 1984. The same period also saw an over one standard deviation decline in "Marxist" language. Many of these changes coincide with several important events during the period, such as the 3rd plenary session of the 12th Party Congress in 1984 when the term "commodity economy" was introduced, and the 16th Party Congress in 2002 when the "preliminary completion of the socialist market economy" was announced.

The slowdown in pro-market development in China after 2000, as suggested by the corpus, is surprising, given the country's continuous economic growth. We find in fact that within documents, the intensity of economic-related language (Mankiw words) continued to increase, especially in the "informal" documents. In other words, the decline in the neoliberal and economic content in both formal and informal documents after 2000 was largely driven by the introduction of non-economic related laws and regulations, consistent with one of our stylized facts. A possible interpretation of these findings is that the Chinese governments has shifted from market-orienting reforms to something else.

The last part of the paper relates the measured market orientation in the legal documents to provincial macroeconomic outcomes. We find that while there has been a rise in the frequency of neoliberal language in laws and regulations up to 2000, provinces' market reforms do not seem to contribute much to their GDP (per capita) and FDI growth. Specifically, year and province fixed effects can already explain over 70% of the R-squared of provincial growth regressions. Including market-oriented word shares as additional regressors only improve the R-squared slightly. Using a more general approach to summarize the legal documents and a more flexible regression framework (i.e., LASSO), we also find that the documents have insignificant power to predict provincial macroeconomic outcomes. These findings suggest the importance of studying the informal arrangements between market participants and government officials in more detail, along the lines of Hallward-Driemeier and Pritchett (2015) and Bai, Hsieh and Song (2018).

Our paper is related to several strands of literature. It relates to the studies on the divergence between de jure and the de facto legal environment (Acemoglu and Robinson, 2008; Fisman and Wei, 2004; Carrillo et al., 2014). It is also related to the extensive literature on China's policy reforms, including studies on the dynamic central-local government relationship (Lin, Cai, and Li, 1997; Xu, 2011), the work on meritocracy (Li and Zhou 2005, Masina 2006, Nee and Opper 2012); the body of research on the decentralization of economic policies (Blanchard and Shleifer 2001, Jin, Qian, and Weingast 2005), and factional competition in the central government (Cai and Treisman 2006). Give the methodology of our research, it naturally borrows state-of-the-art techniques from the work on natural language processing, such Hansen, McMahon and Prat's (forthcoming) work on central bankers' communication, Hoberg and Phillips' (2016) study on product differentiation, and Kelly et al.'s (2018) recent paper on the quality of technological innovations and patent citation.

The paper is organized as follows. Section 2 describes our data source. Section 3 establishes several stylized facts using the corpus of legal documents. Section 4 describes the natural language processing approach to gauge the degree of market orientation in China's changing legal landscape. Section 5 examines whether our constructed measures of market orientation can explain provincial economic outcomes. The final section concludes.

2 Data

We obtain the entire corpus of over 1.4 million Chinese legal documents collected by Chinalawinfo, a hi-tech legal information company established by Peking University.[6] The document repository, called PKULaw, contains many daily working documents produced by various ministries and departments of the Chinese government, and nearly the entirety of laws issued by China's central, provincial and prefectural governments since 1949. Specifically, the corpus contains the universe of 64,802 formal laws issued by various departments and levels of the Chinese political system since 1978. Formal laws include those written by the national and local People's Congresses, the State Council and its administrative agencies, and local governments.

A substantial part of our repository (i.e., over 1.3 million or 99%) is composed of documents that lack a formal legal status. These documents were issued without formal congressional process. Thus, we do not call them laws but policies and regulations, or simply "informal" documents. Despite its lack of formal legal status, the "informal documents" includes stipulated concrete policies by central and local Development Reform Commissions (DRC) (e.g., the minimum wage policies enacted by different provincial governments).

Figure 1 provides a screenshot of the PKULaw Interface. It reveals the standardized template PKULaw uses to organize and report each policy document. Each document reported must include the following information: the government department that issued the document, the date of the issuance, the date on which the policy has been nullified and thus became ineffective.

PKULaw also assigned to each document a unique document ID and a policy category. **Figure 2** shows the top 10 categories in terms of the frequency shares in the formal part of our corpus, out of the 104 categories. The top 3 categories are transportation, real estate, and public safety. We identify those that are related to economic policies and use them to guide our analysis. For instance, out of the top 10 categories, half of them (transportation, real estates, labor unions, construction, and agriculture) are related to economic matters. See Table A1 in the appendix for the full list of categories and those that we flag as economic-related documents.

[6] Chinalawinfo collected these documents from open sources. No classified documents are included. A few recent papers have used documents from Chinalawinfo (or its major product PKULaw) to study specific Chinese policies (Fan, 2018; Chari et al., 2018; Tian, 2018). Chari et al. (2018) examines rural land contracting laws in China, while Tian (2018) examines the rollout of migration reforms. We are not aware of any project that uses the entirety of raw PKULaw documents.

Figure 1: Screenshot of the PKULaw Interface

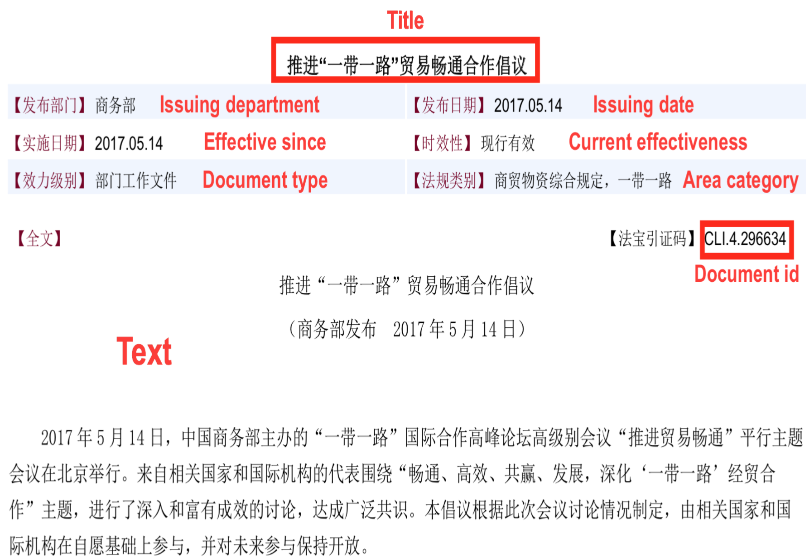
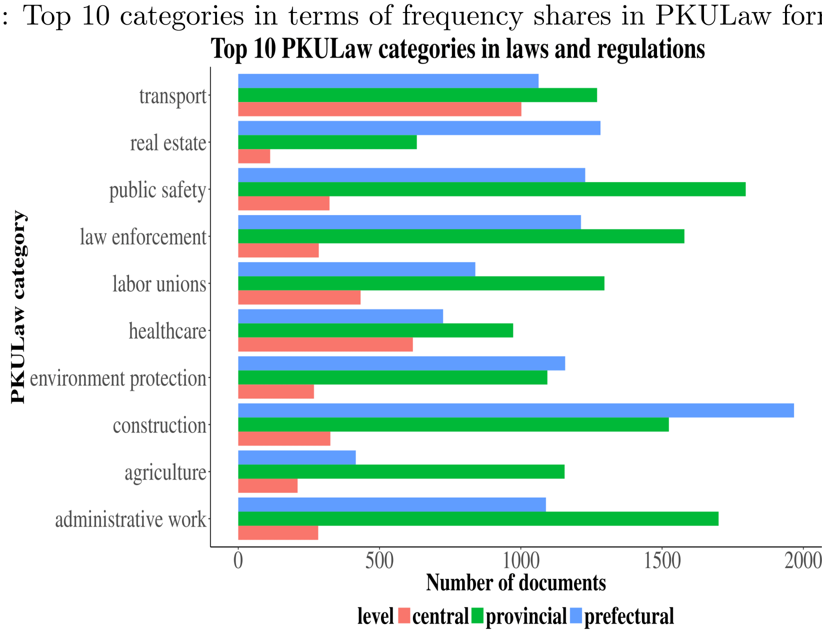


Figure 2: Top 10 categories in terms of frequency shares in PKULaw formal laws



3 Descriptive Statistics

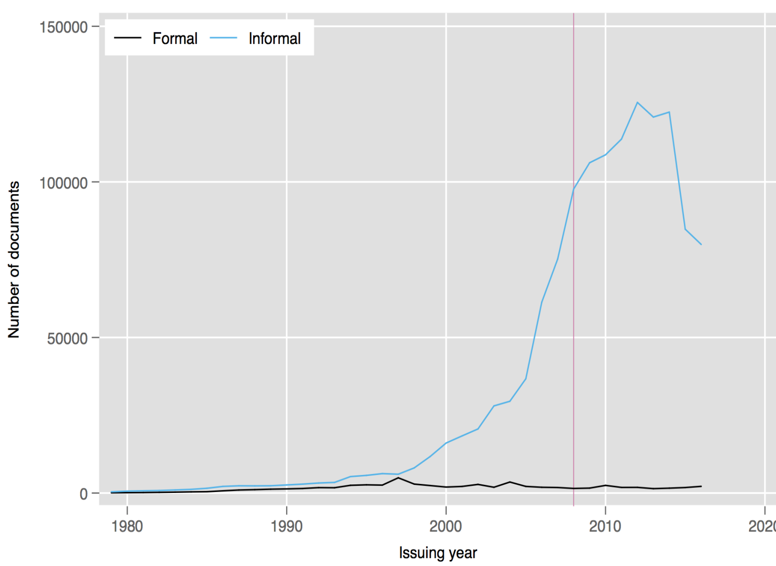
3.1 Coverage over Time

Before we discuss our textual analysis and regression results, let us establish several stylized facts about China's reforms and policy changes according to the corpus.

Fact 1: The share of "informal" laws issued by various levels of the governments has been increasing.

Figure 3 report the numbers of formal laws and "informal" policies and regulations from 1980 to 2016 in our corpus. As is shown, the numbers of informal regulations and policies have been rising significantly faster than that of formal laws, especially since 2005. One reason for the increased coverage for informal documents in recent years is related to the advances of internet technology and the Chinese government's passing of the policy disclosure requirement in 2008, called "Regulation of the People's Republic of China on the Disclosure of Government Information" (Yale Law School, 2009). Since the shares of reported formal and informal documents in each year may be affected by these changes, in the regression analysis below, we will separate formal and informal laws separately.

Figure 3: Formal laws and "informal" regulations over time



Fact 2: The share of formal laws issued by local governments has been increasing.

Figure 4 illustrates the number and share of (formal) laws issued local governments, including provincial and prefecture-level governments, relative to those issued by the central governments and its ministries. As revealed by the red line, the share of local laws has been increasing rapidly from 1980 all the way to the mid 90s, and continue to increase gradually until the recent years. The black line shows that the number of laws has been increasing rapidly from 1980 up to 1997, after which it dropped gradually.

Figure 4: Shares of documents issued by local and central governments, respectively.

Fact 3: The share of formal laws on economic issues has been gradually declining.

Figure 5 shows instead the share of (formal) laws on economic matters, issued by various levels of the Chinese governments and ministries. Economics-related laws include all formal documents that belong to 70 (out of total) 3-digit categories that appear to be obviously related to economic activities, ranging from contract laws, intellectual property rights, to labor laws and the governance of E-commerce (see Table A1 in the appendix for details). As revealed by the red line, the share of economic-related laws has been gradually declining over time.

Figure 5: Shares of economic-related documents

Fact 4: The Chinese government has been constantly nullifying existing laws, especially after 2000.

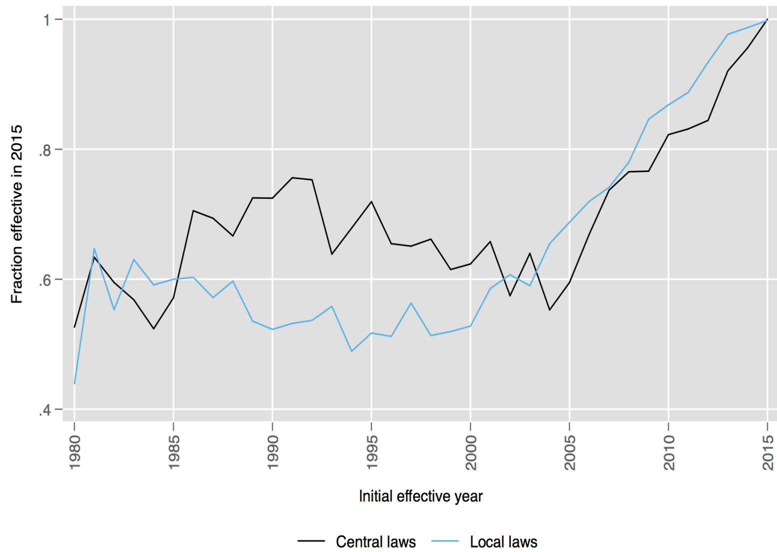
The next fact we want to establish is about how long a formal law will remain effective. **Figure 6** shows the fraction of laws issued in year t (where $t < 2015$), that was still effective in 2015. The black line shows that less than 70% of the central laws issued before 2005 were still effective in 2015. Moreover, less than 25% of laws enacted by local People's Congresses in the 80s are still effective today. Laws made by executive branches have more staying power.

Among the local laws issued before 2005, their survival rates in 2015 were even lower. The blue line shows that less than 60% of the local laws enacted before 2005 were still effective in 2015. Laws that were enacted after 2005 had a substantially higher "survival" rate in 2015. A possible explanation for the low survival rate of laws issued before 2005 could be related to the Chinese government's announced commitment to complete the "socialist legal system with Chinese characteristics" by 2010.[7] Many of the laws that were no longer relevant for such commitment were simply nullified or modified to fit the central government's objective.

Another way to show that is to examine the average effectiveness of laws n year since their issuance. **Figure 7** shows that over 90% of laws still remained effective 2 years after their issuance. The fraction naturally dropped when we expand the time horizon. In particular,

[7] Source: http://www.scio.gov.cn/zfbps/ndhf/2011/Document/1036756/1036756_1.htm

Figure 6: Fractions of laws issued in year t that were still effective in 2015



when we examine a law's effectiveness 10 years since its issuance, the survival rate is about 80% before 2000 but dropped substantially to 70% for those that were issued between 2000 and 2007.

Yet another way to illustrate this nature of frequent turnover of Chinese laws is to examine the average effective rate across all documents n years after their issuance, up to 10 years. By construction, the slope of the inverse hazard rate needs to be negative. Two facts stand out from **Figure 8** shows that 10 years from the date of issuance, about 20% of the laws had be nullified before that. The average nullifying rate year by year appears to be higher for economic-related laws, suggesting that the learning and turnover dynamics appear to be faster for economic-related policies.

4 Measuring Market Orientation in Policies

We now describe our approach of using Natural Language Processing (NLP) techniques to measure market orientation of each document in our corpus. Experts studying the Chinese economy have agreed that the spectacular growth of the Chinese economy is attributed to the government's market-oriented liberalization policies (e.g., Lardy, 2014; Naughton, 2018). While there are many anecdotes describing the country's economic reforms, we are not aware of a systematic analysis of the forty years of the reform processes based on the government's

Figure 7: Laws that are still effective n years since issuance.

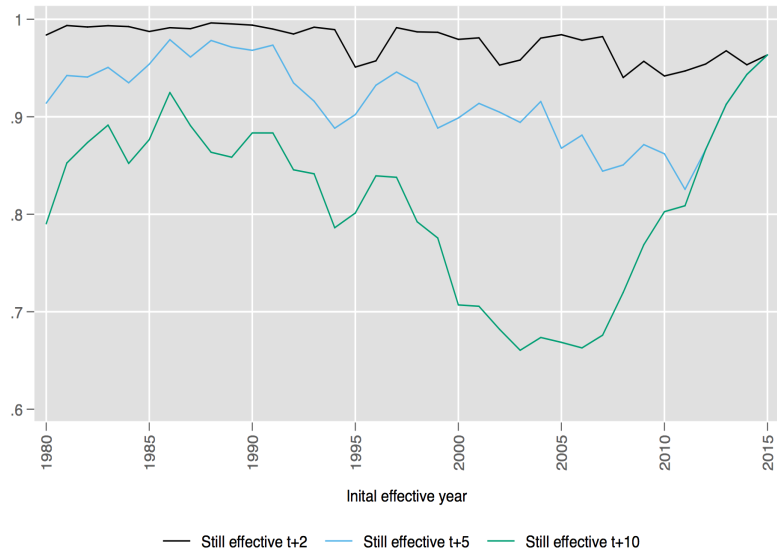
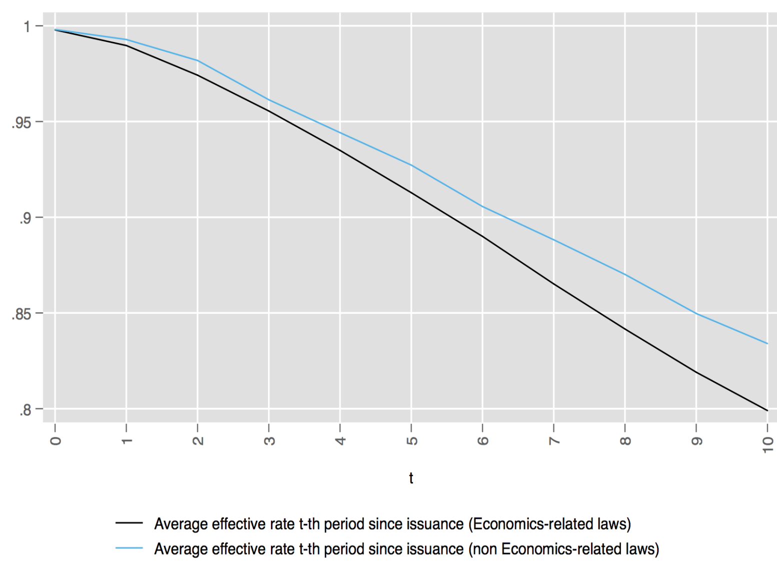


Figure 8: Hazard rates of laws

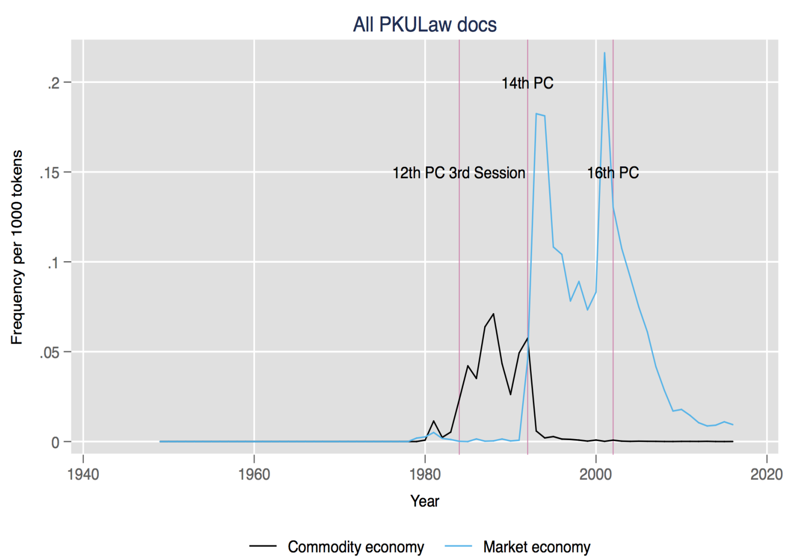


legal documents. That is the goal of this section.

4.1 A Motivating Example

Before discussing our approach of textual analysis, let us conduct a supervised textual analysis for searching for the two obvious keywords that indicate the switch from a Marxist regime to a market economy: "commodity economy" and "market economy". F**igure 9** shows the frequency of these two key words mentioned per 1000 token (roughly a meaning word) in the corpus. The frequency is essentially zero in all documents before 1980, and exhibit interesting "peak and trough" patterns across time. Specifically, the frequency of the word "commodity economy" shot up right after the 1984, and remained high for the rest of the 1980s. In October, 1984, leaders of the Chinese government introduced the first time the idea of "commodity economy" into its economic lexicon, paving the way for further liberalization and market-oriented reforms. Specifically, at the Third Plenary Session of the country's Communist Party (CPC) Central Committee, the idea that a centrally planned economy cannot go with a commodity economy was dropped.

Figure 9: China's Ideological Shift in Two Words



The frequency of the word "market economy" also exhibited interesting fluctuations. The frequency shot up right after 1992, when the CPC announced in the Third Plenum in Oct to completely embrace the private sector as an key component of the economy, right after the famous Southern Tour by the architect of China's economic reforms Deng Xiaoping, earlier that year. The frequency of "market economy" shot up again around 2002, when in Oct that year that the CPC announced the "preliminary completion of the market economy".

4.2 Pre-processing the Corpus

Before conducting any textual analysis, several steps need to be implemented. First, we need to transform the corpus into a document-term matrix, which has an element $Entry\,(i, j)$ as the number of occurrences of term j in document i. Before the construction of such matrix, some pre-processing of documents is needed. Specifically, we pre-process each document through the following steps

1. Remove all non-Chinese characters.
2. Remove tokens containing just one character based on a standard list of stop words (e.g, "the", "and").
3. Segment sentences into tokens using an existing Python module called "Jieba" (Chinese for "to stutter") to partial out phrases as frequently cooccurring words. For instance, many tokens are meaningful phrases rather than simple words (e.g., Three Represents).
4. We obtain over one million unique tokens, the majority of which provides little information for us to categorize documents. We thus filter out tokens whose frequency is particularly low, filtering out words that have a cross-document sum of tf-idf lower than 100.[8]

The remaining tokens are sufficiently frequent, and do not show up in too many documents. Our ultimate filtered corpus is represented as a document-term matrix with dimension equal to 1414736 documents and 12693 tokens.

With the document-term matrix constructed, we are ready to gauge the extent to which each Chinese policy document has moved towards markets from Marx, or not. The methodology we adopt is very similar to the ones used by Groseclose and Milyo (2005), who study Republican and Democrat politicians citations of different think tanks with clear political positions, and Gentzkow, Shapiro and Taddy's (2018), who quantify the evolution of political polarization using data on the text of politicians' speeches in the US Congress from 1873 to 2016. These existing machine-learning innovations are not immediately applicable to our Chinese corpus. The commonly used topic models – an unsupervised technique that automatically finds meaningful groups of words ("topics") is not what we want. While it is relatively

[8] See https://en.wikipedia.org/wiki/Tf%E2%80%93idf for the definition of tf-idf.

easy to find a handful of words that have clear ideological implications (e.g., privatization, competition), like the one we showed in the motivating example in Section 4.1, handpicking keywords does not appear to be an objective approach.

4.3 Identifying market orientation in policy documents

Let us know be more systematic in searching for the keywords that are related to market orientation versus central planning. To this end, we use the technique called word embedding, which is a recent technique that has proven useful in many NLP tasks. The ultimate goal of word embedding is to find a substantial number of words that represent the market orientation. We also want to let the data tell us what these words should be according to ways language is actually used in the policy documents. In other words, we want to find words that are used in similar contexts as for example, "privatization". To this end, we construct the a set of anchor words obtained from other texts that are external to our corpus. We obtain the anchors from the following three sources:

1. The top 50 "Marxist" words in Karl Marx's famous class *Das Kapital*.
2. The full list of "neoclassical economics" words as appeared in the glossary of Mankiw's best-selling introductory economic textbooks for college students.
3. The 10 "neoliberal" words: the 10 economic policy prescriptions for economic growth, proposed by John Williams in 1990 as the so-called "Washington Consensus".

4.3.1 Initial Words

Let us know discuss how we construct our anchors. The approach is to start with small sets of words having relatively clear orientations and expand the sets using word embedding. To construct our database of Marxist words: we first extract keywords from the text of Karl Marx's *Das Kapital*. We use techniques for keyword extraction, *TextRank* (Mihalcea and Tarau 2004), similar to Google's PageRank algorithm for ranking websites, but applied to text. The basic idea is to find the most central words in a graph of text. Two words are connected in the graph if they appear within a window of each other. The importance of a word is determined recursively by examining the importance of words connected to it. We take the 50 most central words from *Das Capital*.

To give an example, our methodology produces the following connected words to the word "privatization": Privatization, privatization (alternative Chinese word), demutualization, joint stock system, reorganize, transform the system, corporation system, shareholding cooperative system, restructuring, transform the mechanism and build the system, asset restructuring, property rights system, state-owned enterprise, cooperative system, corporatization, change system, mixed system, shareholding cooperation, merger, debt-for-equity, state-owned. This example highlights the fact that a single concept is discussed in varied and context-specific ways, hence the importance of letting data speak for what words we should trace.

We construct “Neoclassical economics" words by taking the entire glossary from Gregory Mankiw’s *Principles of Economics*. Alternatively, we adopt the list of “Neoliberal words” from the ten principles for economic reforms laid out in John Williamson’s 1989 article that later became known as the Washington Consensus.

Table 1 shows the top and bottom 10 central words in Marx’s *Das Kapital* and their corresponding shares of occurrence in our corpus. For instance, the word "labor force" showed up 0.25/1000 times in our corpus. **Table 2** instead show the top and bottom 10 central words in Mankiw’s principle economics textbook and their corresponding shares of occurrence in the corpus. **Table 3** shows the 9 Washington Consensus keywords and the corresponding shares in the corpus.

Table 1: Relative shares of top Marxist words in the corpus

Word	Share	Word	Share
labor force	0.250	working class	0.005
worker	0.238	means of subsistence	0.003
ownership	0.150	industrial capital	0.002
productive forces	0.066	farm owner	0.002
means of production	0.057	equivalent	0.002
producer	0.056	capitalism	0.002
owner	0.054	amount of value	0.002
commodity circulation	0.019	usury	0.002
profit rate	0.017	bourgeoisie	0.002
mode of production	0.016	money capital	0.002

With these three sources of words identified, the next step is to find a way to quantify the similarity between each policy document in our corpus and each of the three set of anchor words. In a pioneering paper, Mikolov et al. (2013) propose two models – Continuous Bag-of-Words and Skip-gram, to explicitly target the relationship between words and their contexts (i.e., proximate words) in model training. Each word is affiliated with a vector. The cosine similarity between these vectors then measures the extent to which words are used in similar

Table 2: Relative shares of Mankiw words in the corpus

Word	Share	Word	Share
investment	0.236	currency	0.011
market	0.205	stock	0.010
social security	0.061	welfare	0.010
export	0.059	equilibrium	0.009
cost	0.039	market economy	0.009
consumption	0.039	diversification	0.009
import	0.036	equality	0.009
capital	0.035	principal	0.007
bond	0.033	screening	0.006
property rights	0.031	reserve	0.005

Table 3: Relative shares of Washington Consensus keywords in the corpus

Word	Share
taxation	0.354
trade	0.269
property rights	0.233
interest rate	0.105
deregulation	0.025
exchange rate	0.012
fiscal deficit	0.001
fiscal expenditure	0.001
privatization	0.0001

contexts.

Specifically, we train the Continuous Bag-of-Words model on the entire PKULaw corpus. The model is a simple neural network that seeks to predict each word in the text from its surrounding context. Let $(w_1, w_2, ..., w_T)$ denote our corpus. The objective of the model is to maximize the log likelihood

$$\frac{1}{T}\sum_{t=1}^{T}\log\ p(w_t|w_{t-c}, w_{t-c+1}, ..., w_{t+c})$$

where $(w_{t-c}, w_{t-c+1}, ..., w_{t-1}, w_{t+1}, ..., w_{t+c})$ are the words surrounding w_t.

Abstracting from details of the neural network, the likelihood of a single word can be expressed as

$$p(w_t|w_{t-c}, w_{t-c+1}, ..., w_{t+c}) = \frac{\exp(v_{w_t}'^{T}\frac{1}{2c}\sum_j v_{w_{t+j}})}{\sum_{i=1}^{V}\exp(v_{w_i}'^{T}\frac{1}{2c}\sum_j v_{w_{t+j}})}$$

where v_w and v'_w are two vectors describing how word w is used in different contexts.

Intuitively, we see that if two words are often used in similar contexts (i.e., in proximity to a similar set of words), the algorithm is going to assign high cosine similarity to these two words. In training the model, we represent words as 400-dimensional vectors, and use a window size $c = 5$ (i.e., we find 10 words that are closest to the token of interests in terms of cosine similarity). As an example, **Table 4** shows the top 10 words that are considered to have a similar meaning as "Hong Kong". Intuitively, "Macau", "Hong Kong SAR" and "textile" are words that are often mentioned in the same topic of discussion when "Hong Kong" is mentioned.

Table 4: Most similar words to "Hong Kong"

Word	Similarity
Macau	0.782
Hong Kong SAR	0.658
textile making	0.650
Singapore	0.641
hot pot	0.621
Taiwan	0.620
Macau SAR government	0.614
HK and Macau	0.612
trade promotion bureau	0.603
The Venetian	0.601

4.4 Market orientation of the stock of regulations

Equipped with a set of anchor words, we now examine how market oriented the effective "stock" of laws evolved over time.[9] **Figure 10** shows the frequency of tokens in the formal laws that are similar to the identified "Marx", "Mankiw" and Washington Consensus (WC) words, with the cosine similarity cutoff equal to 0.4 The red line shows a continuous decline in the Marx-similar tokens (words) in the formal documents in our corpus, while a continuous increase in the frequency of Mankiw-similar and WC-similar words. The dashed vertical lines in the graph mark the year when a Third Plenum, in which the CPC would discuss the next 5-year plan, took place. It is interesting to see that the Third Plenum often served as a turning point of one of these keyword trends. It is also interesting to observe that basically by the end of the 1990s, the frequency of both "Marx" and neoliberal related words both started to drop. These trends are consistent with one of the stylized facts that the shares of economic-related documents have been declining since 2000. One possibility is that by the end of the 90s, the CPC already felt that much of the overdue economic reforms have been mostly done, and it was the time to focus on other social and political developments of the country.

Figure 10: New formal laws issued, similarity cutoff 0.4, 5-year moving averages

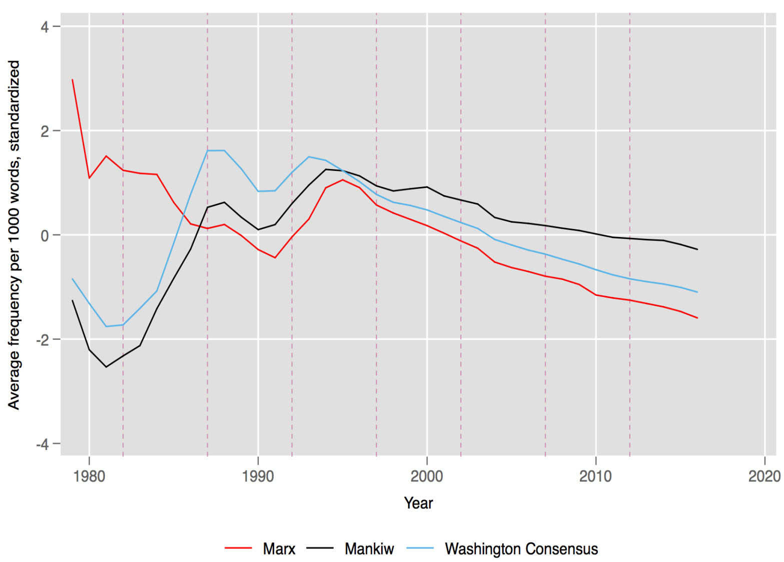


Figure 10 illustrates the trends of the frequency of tokens in the informal document that are similar to the three anchors. While there is a persistent downward trend in the frequency

[9]In this and the following graphs, the three time series are each standardized, as their levels are not really meaningful, but mechanically correlated with the number of words within each set.

of tokens similar to "Marx" words, consistent with the conventional wisdom that China has been moving away from its Marxist root and towards a market-economy system, the trends of the frequency of token similar to "Mankiw" and WC words exhibit an surprising hump shape. As a mirror image to the declining frequency of the Marx-similar words, the frequency of market-oriented words increased constantly from 1980 up to 2000. Since 2000, the frequency of the market-oriented tokens has been gradually declining up to the most recent period in the corpus. One should not interpret these findings as evidence for de-marketization or the return of state capitalism in the Chinese economy. One possibility is that the government's policy documents became richer and more diverse since 2000. After years of institutional building for market reforms, the country's government may feel that the necessary institutions for economic development were largely in place, and started building institutions to foster other aspects of the country's development, such as political reforms and social issues such as food safety. Regardless of the reasons, the downward trend in the frequency of market-oriented words in the informal documents is consistent with the same downward trends in the formal laws and that of economic-related laws and regulations. **Figures 12** and **13** illustrate the same downward trends in terms of flows, rather than stocks, for both formal and informal documents, respectively.

Figure 11: Stock of informal laws, similarity cutoff 0.4

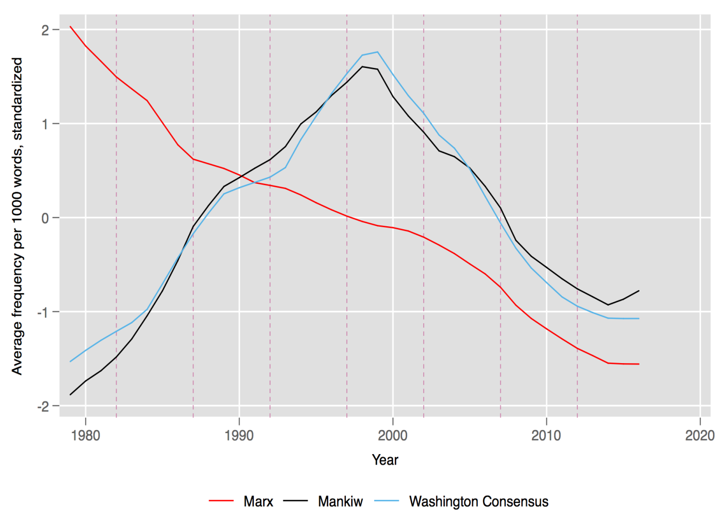


Figure 14 illustrates the frequency of WC-similar tokens for coastal and inland provinces, respectively. For both regions, the downward trend in WC-similar words are observed for both coastal and inland provinces, with the former always having a higher frequency of WC-similar

Figure 12: New formal laws issued, similarity cutoff 0.4, 5-year moving averages

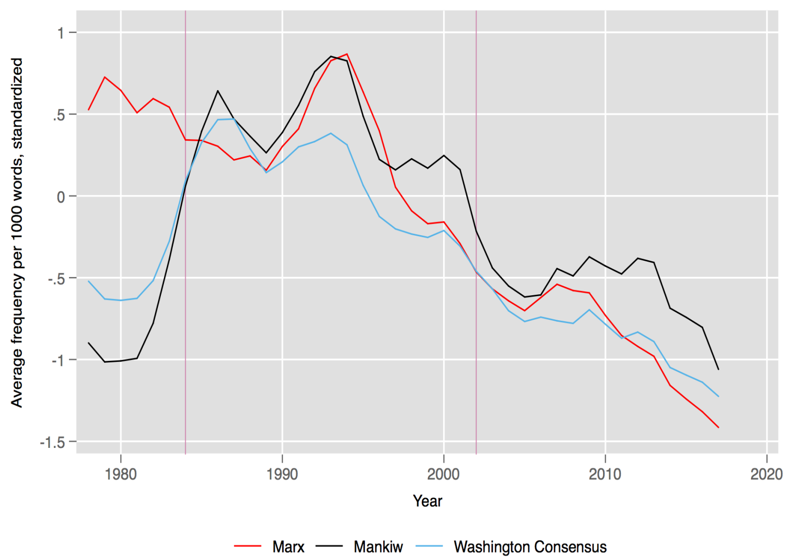


Figure 13: New informal laws issued, similarity cutoff 0.4, 5-year moving averages

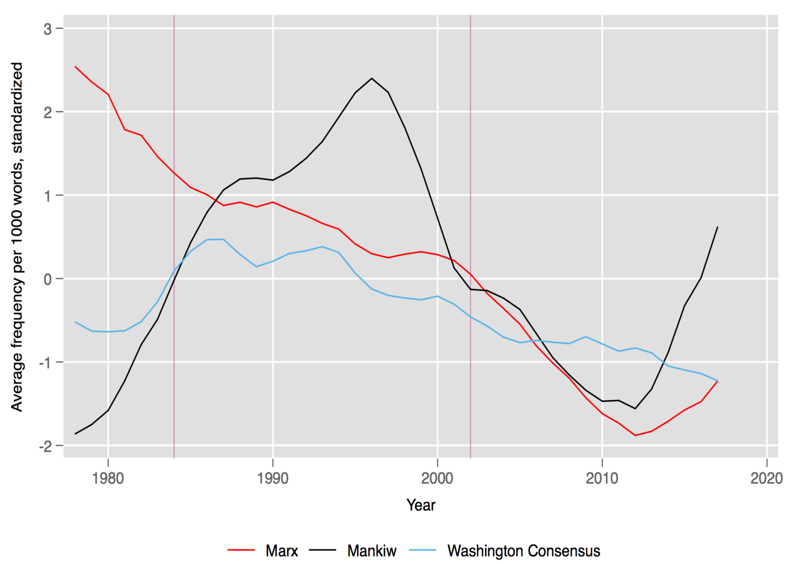
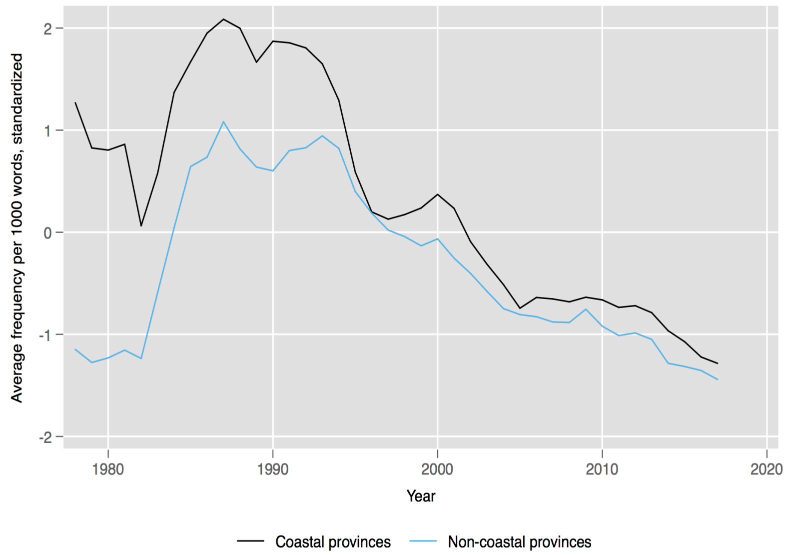


Figure 14: Washington Consensus in new formal laws issued



words, consistent with the common perception that coastal provinces are more economic developed due to their more liberal and outward-oriented policies. **Figure 15** shows similar patterns between coastal and inland provinces. The hump-shaped time-series patterns were observed for both regions. In the appendix, we show similar trends in both formal and informal document for separate components in the WC principles.

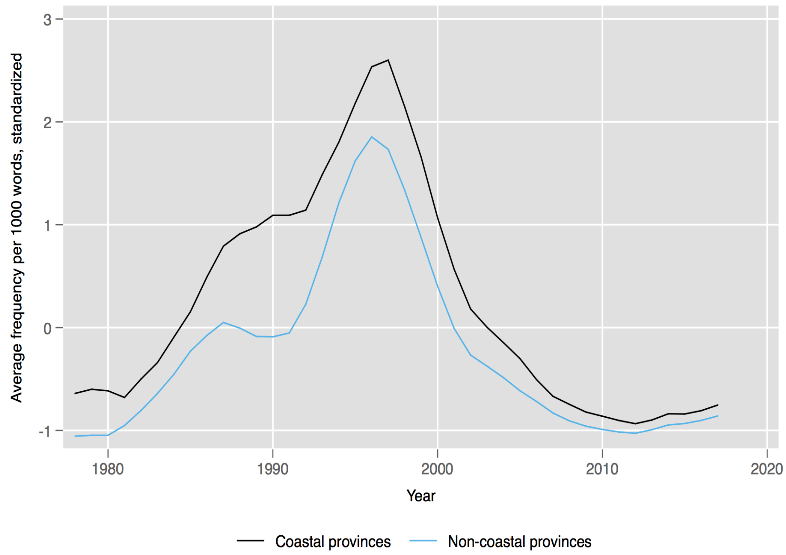
Now let us run some regressions to shed light to the hump-shaped in documents' market orientation. In particular, we want to understand whether the documented trends were simply driven by the compositional changes in the laws and regulations. To this end, we estimate the following specification

$$Z(M_{ict}) = \alpha_{\text{prov}(i)} + \lambda_c + F_t + \epsilon_{ict}, \tag{1}$$

where $Z(M_{ipc})$ is the z-score of the share of the words similar to one of the three anchors in document i that belonged to category c and issued in period t. α is province fixed effect. λ is subcategory fixed effect. These are very fine-grained (1159 subcategories, e.g., land use of foreign enterprises). F is 5-year period fixed effect, corresponding to the 5-year plan time frames. Standard errors clustered at the category level.

Figure 16 plots the estimates of F for every 5-year periods from 1980 to 2015 based on (**1**) using the z-score of the Washington Consensus words as the dependent variable. For both

Figure 15: Washington Consensus in new informal laws issued



formal and informal documents, there appears to be a downward trend in the Washington Consensus language *within* document types between the third 5-year period (1990-1995) and the fifth 5-year period (2000-2005).

Figure 17 plots the estimates of F for each 5-year period from 1980 to 2015 based on (**1**) using the z-score of the Marx-related words as the dependent variable. For both formal and informal documents, there is a downward trend in the Marxist language *within* document types over the entire sample period. The downward trend is significantly more pronounced for the sample of informal documents.

Figure 18 plots the estimates of F for every 5-year periods from 1980 to 2015 for the "Mankiw" regressions. For both formal and informal documents, there appear to be a slight upward trend in the Mankiw-type language *within* document types over the entire sample period. The positive slope is particularly steep for the sample of informal documents.

5 Do market-oriented regulations explain economic growth?

After establishing stylized facts about the evolution of Chinese policy documents, we now relate the extent of market orientation in laws and regulations to economic outcomes. The goal is to study whether market-oriented policies explain China's economic growth. We run a host of province-level regressions to correlate a province p's average GDP per capita growth

Figure 16: Washington Concensus language regressions

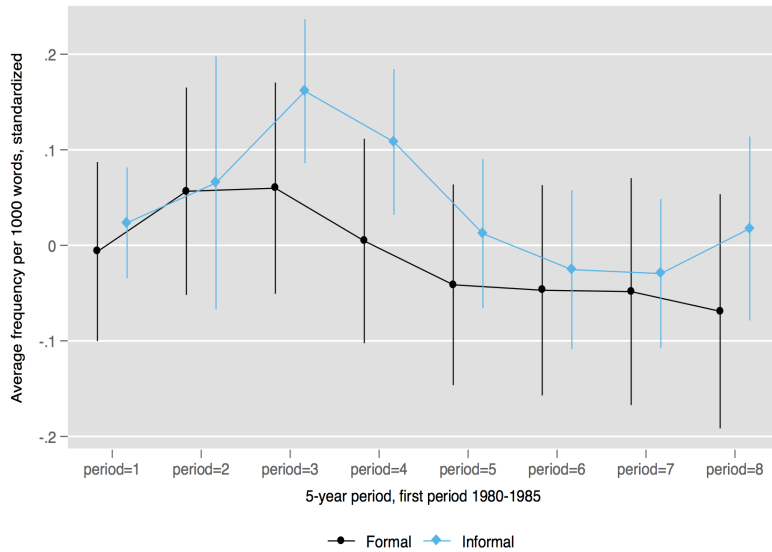


Figure 17: Marxist language regressions

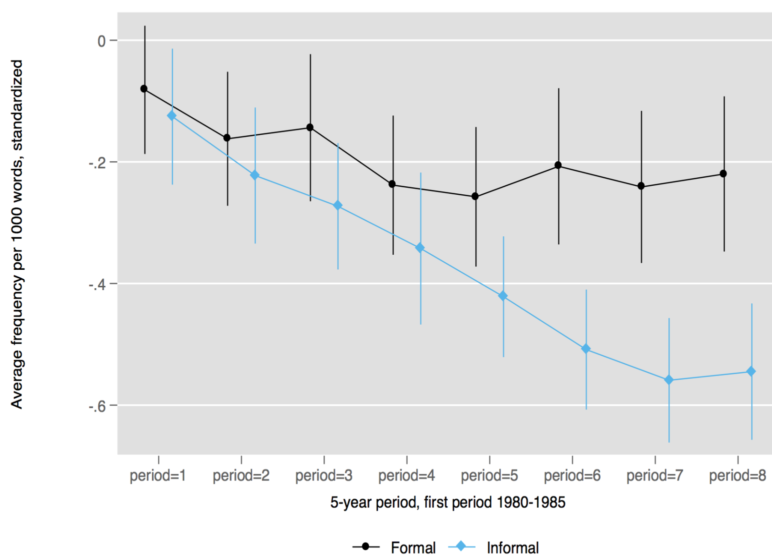
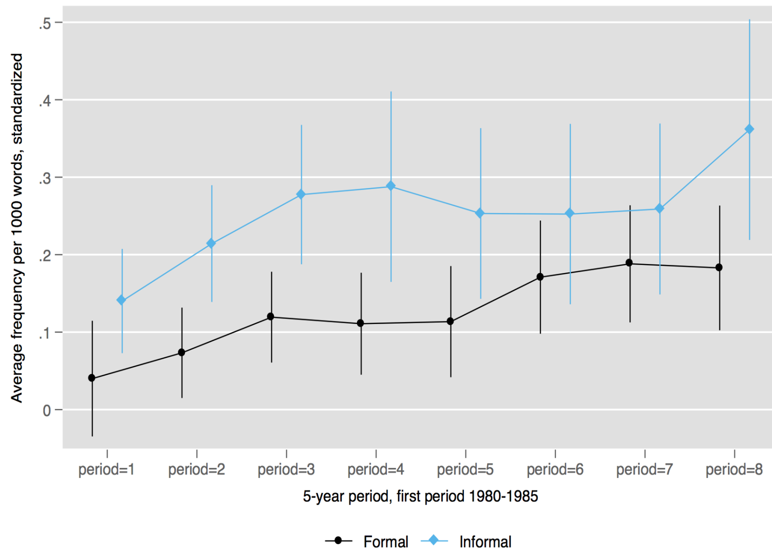


Figure 18: Mankiw language regressions



over a 5-year period with the share of one of the three anchor-related words in each province p's documents, both formal and informal, in the same period.

To this end, we estimate the following specification

$$\Delta y_{pt} = \alpha_p + \lambda_t + \beta \text{word share}_{\text{pt}} + \gamma' X_{pt} + \epsilon_{pt}, \tag{2}$$

where Δy_{pt} is the average annual GDP per capital growth over the 5-year period p. α is the province fixed effect. λ is the 5-year period fixed effect. word share_{pt} is the average frequency share of one of the three sets of anchor-related words in the "stock" of regulations ever issued by the province that still remained effective by period t. In addition to the fixed effects, we also include controls variables such as log number of regulations and log average length of regulations.

Table **5** shows the results of estimating (**1**). As column (1) shows, simply adding province fixed effects can only account for about 4.2% of the variation in gdp per capita growth across provinces and time. When year fixed effects are added in column (2), the R-squared shoots up to 78%. Notice that year fixed effects capture all variation in the central government's national laws across time. The significant increase in R-squared suggests that national laws and other macroeconomic factors still play an important role, compared to local regulations, in driving provincial economic growth. In column (3), we add the (log) share of Mankiw-related words in formal and informal documents separately as regressors, in addition to the year and province fixed effects. We find that the intensity of the Marxist language in both sets of documents does not explain provincial GDP outcomes. In columns (4) and (5), we repeat the exercise of columns (3) by replacing the two regressors of Marxist word shares by Mankiw words shares and WC word shares, respectively. None of them appear to predict provincial GDP per capita. When all of the six regressors of anchor-related word shares are included as regressors, none besides the WC-related word shares in informal documents of the province are positively correlated with provincial economic growth.

Table 5: Relative shares of anchor-similar words in the corpus

	DV: provincial GDP per capita growth over 5-year period					
	(1)	(2)	(3)	(4)	(5)	(6)
log Marx, formal			0.007			0.014
			(0.028)			(0.062)
log Marx, informal			0.041			0.083
			(0.042)			(0.051)
log Mankiw, formal				0.025		0.073
				(0.055)		(0.088)
log Mankiw, informal				-0.019		-0.114
				(0.059)		(0.072)
log WC, formal					0.017	-0.023
					(0.029)	(0.037)
log WC, informal					0.028	0.054**
					(0.022)	(0.025)
Province FE	Yes	Yes	Yes	Yes	Yes	Yes
Period FE	No	Yes	Yes	Yes	Yes	Yes
Observations	217	217	211	211	206	206
R-Squared	0.042	0.784	0.792	0.790	0.796	0.803
Adjusted R-Squared	-0.113	0.741	0.740	0.738	0.744	0.746

One may think that there are many channels through which market-oriented policies affect the structure rather than the level of economic outcomes. Given the drastic increase in trade flows and foreign direct investment (FDI) into China since its government's trade and FDI liberalization policies since the mid 90s, we may expect to see a stronger correlation between the word count shares of market-orienting words in policy documents and performance in provincial trade or FDI activities. We thus replacing average GDP per capita growth in (**1**) by the 5-year average FDI/GDP growth. Table **6** shows the estimation results. Column (1) shows that province fixed effects explain significant variation in FDI outcomes, compared with economic growth as reported in Table **5**. Specifically, the R-squared is 54% just with province fixed effects included as a regressor, suggesting that province time-invariant characteristics (e.g., coastal provinces) explain much of the observed variation in FDI growth across provinces. In column (2), when year fixed effects are included as regressors additionally, the R-squared increases to 66%. In column (3), when Marx-related word share in formal and informal documents are separately added as regressors, the R-squared increases to 70%, with the (log) share of Marx-related words in informal documents issued by the province being negatively and significantly correlated with the 5-year average FDI/GDP growth.

We repeat the same estimations of column (3), but with the two Marx-related word shares replaced by the corresponding Mankiw-related shares in column (4), and by the corresponding WC-related shares in column (5). We find that while the intensity of Mankiw language does not seem to matter for FDI growth, the WC word shares in informal documents of the province are positively correlated with provincial FDI growth.

Table 6: Relative shares of top Marxist words in the corpus

	DV: mean FDI-to-GDP ratio over 5-year period					
	(1)	(2)	(3)	(4)	(5)	(6)
log Marx, formal			-0.000			-0.002
			(0.001)			(0.002)
log Marx, informal			-0.002**			-0.001
			(0.001)			(0.001)
log Mankiw, formal				-0.001		0.002
				(0.001)		(0.004)
log Mankiw, informal				0.000		-0.000
				(0.001)		(0.001)
log WC, formal					-0.002**	-0.001
					(0.001)	(0.001)
log WC, informal					0.001**	0.001*
					(0.001)	(0.001)
Province FE	Yes	Yes	Yes	Yes	Yes	Yes
Period FE	No	Yes	Yes	Yes	Yes	Yes
Observations	202	202	197	197	192	192
R-Squared	0.535	0.660	0.700	0.689	0.710	0.717
Adjusted R-Squared	0.457	0.588	0.621	0.607	0.631	0.630

In sum, based on the small improvement in R-squared when the market-oriented word shares are included as regressors for both sets of regressions, we find no evidence that market oriented policies can explain province-level macroeconomic outcomes.

5.1 Does Any Regulation Predict Growth?

- If the market orientation of policies alone does not explain much growth or FDI, can a richer representation of regulations explain growth?
- We answer this question by resorting to a variable selection technique (LASSO).
- Challenge is how to represent the policies: each document is issued just once for a certain province in a certain year.
- To obtain repeated observations, we use DBSCAN (Ester et al. 1996), a widely-used clustering algorithm, to group similar documents into a cluster.
- We calculate the cosine similarity between each pair of documents, and then cluster on the basis of these similarities.

5.2 Variable Selection

- A richer representation of the text through document clustering and variable selection.
- We estimate the following relationship with LASSO.

$$y_{pt} = \gamma_p + \lambda_t + \sum_{c \in C} \beta_c f_{pc} + \epsilon_{pt}$$

- γ_p and λ_t are province and period fixed effects. C is the set of "stock" policy clusters (c) that are effective throughout period t in province p.
- We force the fixed effects to be in the equation by setting their penalties at 0.
- Select the LASSO penalty that minimizes out-of-sample MSE in a cross-validation.
- Evaluate the predictive power of regulations out of sample (i.e., in a test set), as in-sample R^2 is rendered less meaningful by potential overfitting.

5.3 Does Any Regulation Predict Growth?

- Province and period fixed effects reduce the variance in GDP per capita growth by 53.7%.
- Policy clusters further reduce variance by just 2.33%.

Figure 19: Do Changing Regulations Predict Provincial Economic Growth?

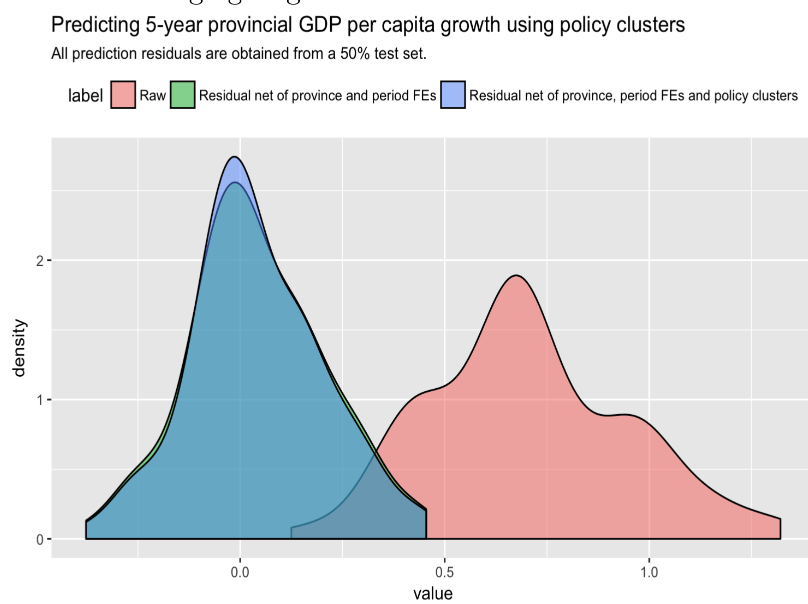
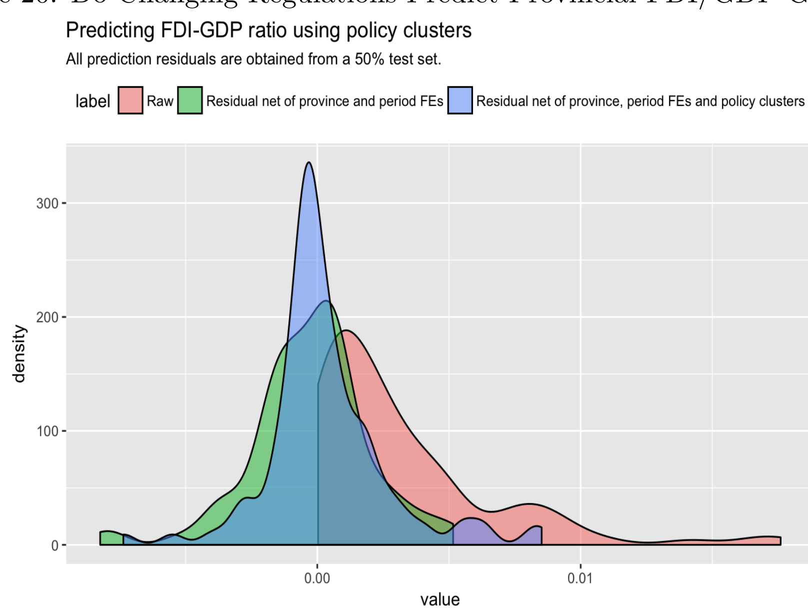


Figure 20: Do Changing Regulations Predict Provincial FDI/GDP Growth?



5.4 Does Any Regulation Predict FDI?

- Province and period fixed effects reduce the variance in FDI-to-GDP ratio by 56.7%.
- Adding policy clusters increases variance by 2.78% due to overfitting.

6 Conclusions

- Active introduction of pro-market institutions from the mid 1980s to around 2000, which slowed down after 2000.
- Within documents, the intensity of economic-related language has been increasing, especially in "informal" documents.
- Thus, the decline in the neoliberal and economic content in both formal and informal documents after 2000 was largely driven by the introduction of non-economic related laws and regulations.
- One interpretation: the Chinese governments has shifted from market orientation to something else.
- That said, the market orientation of regulations only explains a small fraction of the provincial variation in growth and FDI.
- A richer representation of the documents also exhibits small predictive power.
- This suggests the importance of studying the informal arrangements between market participants and government officials in more detail, along the lines of Hallward-Driemeier and Pritchett (2015) and Bai, Hsieh and Song (2018).

References

[1] Acemoglu, Daron, Simon Johnson amd James A. Robinson (2005) "Institutions as a fundamental cause of long-run growth", *Handbook of Economic Growth* 1, 385-472.

[2] Acemoglu, Daron, Suresh Naidu, Pascual Restrepo, and James A. Robinson (2014) "Democracy Does Cause Growth", NBER Working Paper No. 20004.

[3] Bai Chong-en, Chang-tai Hsieh and M. Song (2014) Crony Capitalism with Chinese Characteristics", University of Chicago Manuscript.

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

[5] Blanchard, Olivier and Andrei Shleifer (2001) "Federalism with and without political centralization: China versus Russia", IMF Staff Paper, 48(Spec. Issue):171–79.

[6] Blei, David, Andrew Y. Ng and Michael I. Jordan (2003), "Latent Dirichlet Allocation", *Journal of Machine Learning Research*, 3, 993-1022.

[7] Blei, David and John D. Lafferty (2006), "Dynamic Topic Models", *Proceedings of the 23rd International Conference on Machine Learning*, Pittsburgh, PA.

[8] Cai, Hongbin and Daniel Treisman (2006) "Did Government Decentralization Cause China's Economic Miracle?", *World Politics*, Vol. 58, pp. 505-35.

[9] Fan, Jingting (2018) "Internal Geography, Labor Mobility, and the Distributional Impacts of Trade", forthcoming *American Economic Journal: Macroeconomics*.

[10] Fisman, Raymond and Shang-Jin Wei (2004) Tax Rates and Tax Evasion: Evidence from Missing Imports in China, *Journal of Political Economy*, 26.

[11] Gentzkow, Matthew, Jesse Shapiro, and Matt Taddy (2018) "Measuring Group Differences in High-Dimensional Choices: Method and Application to Congressional Speech", Stanford University Working Paper.

[12] Grimmer, Justin and Gary King (2011) "General purpose computer-assisted clustering and conceptualization", Proceedings of the National Academy of Sciences.

[13] Grimmer, Justin (2015) "We Are All Social Scientists Now: How Big Data, Machine Learning, and Causal Inference Work Together", *PS: Political Science & Politics*, 2015. 48(1), 80-83.

[14] Hallward-Driemeier, Mary and Lant Pritchett (2015) "How Business Is Done in the Developing World: Deals versus Rules", *Journal of Economic Perspectives* 29 (2015), no. 3, 12.

[15] Hansen, Stephen, Michael McMahon, and Andrea Prat (2018) "Transparency and deliberation within the FOMC: a computational linguistics approach", *Quarterly Journal of Economics*.

[16] Heilmann, Sebastian (2008a) "From local experiments to national policy: The origins of China's distinctive policy process", *The China Journal*, (59), 1-30.

[17] Heilmann, Sebastian (2008b) "Policy experimentation in China's economic rise", *Studies in Comparative International Development*, 43(1).

[18] Hsieh, Chang-Tai and Zheng Song (2015) "Grasp the Large, Let Go of the Small: The Transformation of the State Sector in China", *Brookings Papers in Economic Activity*, Spring 2015.

[19] Hoberg, Gerard, and Gordon Phillips (2016) "Text-based network industries and endogenous product differentiation." *Journal of Political Economy* 124, no. 5 (2016): 1423-1465.

[20] Huang, Yasheng (1996) "Central-local relations in China during the reform era: the economic and institutional dimensions", *World Devevelopment*, 24(4):655–72.

[21] Jin, Hehui, Yingyi Qian, and Barry R. Weingast (2005) "Regional decentralization and fiscal incentives: Federalism, Chinese style". *Journal of Public Economics*, 89(9), 1719-1742.

[22] Lardy, Nicholas (2014) *Markets over Mao: The Rise of Private Business in China*, Institute for International Economics publication.

[23] Li, Hongbin and Li-An Zhou (2005) "Political turnover and economic performance: the incentive role of personnel control in China", *Journal of Public Economics*, 89 (9), 1743-1762.

[24] Lin, Justin Yifu, Fang Cai, and Zhou Li (1997) "The China Miracle: Development Strategy and Economic Reform", Beijing: Chinese University Press.

[25] Malesky, Edmund and Jonathan London (2014) "The political economy of development in China and Vietnam", *Annual Review of Political Science,* 17 (2014): 395-419.

[26] Mihalcea, Rada and Paul Tarau (2004) "TextRank: Bringing Order into Texts".

[27] Milokov, Tomas, Ilya Sutskever, Kai Chen, Greg Corrado and Jeffrey Dean (2013), "Distributed Representations of Words and Phrases and their Compositionality", *Proceedings of Workshop at ICLR*.

[28] Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean (2013) "Effcient estimation of word representations in vector space", arXiv preprint arXiv:1301.3781.

[29] Masina, Pietro (2006) *Vietnam's Development Strategies*. New York/London: Routledge.

[30] Naughton, Barry (2018) *The Chinese Economy: Adaptation and Growth*, The MIT Press.

[31] Nee, Victor and Sonja Opper (2012) *Capitalism from below: Markets and institutional change in China*. Harvard University Press.

[32] Rawski, Thomas (1995) "Implications of China's reform experience", *China Quarterly*, (144):1150–73.

[33] Roland, Gerard (2000) *Transition and Economics: Politics, Markets, and Firms.* MIT press.

[34] Sala-I-Martin, Xavier X. (1997) "I Just Ran Two Million Regressions", *American Economic Review*, Vol. 87, No. 2, pp. 178-183.

[35] Shih, Victor, Christopher Adolph, and Mingxing Liu (2012) "Getting ahead in the communist party: explaining the advancement of central committee members in China", *American Political Science Review*, 106(01), 166-187.

[36] Stewart, Brandon (2013) "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts", *Political Analysis*, 21 (3), 267–297.

[37] Tian, Yuan (2018) "International Trade Liberalization and Domestic Institutional Reform: Effects of WTO Accession on Chinese Internal Migration Policy" UCLA Manuscript.

[38] Williamson, John (1990) "The Washington Consensus as Policy Prescription for Development", Institute for International Economics publication.

[39] Xiabo, Wu (2012) Storming 30 Years: 1978–2008 Chinese Enterprises (Chinese Edition), CITIC press.

[40] Xu, Chenggang (2011) “The Fundamental Institutions of China’s Reforms and Development”, *Journal of Economic Perspectives*, 49:4, 1076–1151.

[41] Yale Law School (2009) "Regulation of the People’s Republic of China on the Disclosure of Government Information".

6.1 Trends: Washington Consensus Components

Figure 21: Taxation in new regulations issued

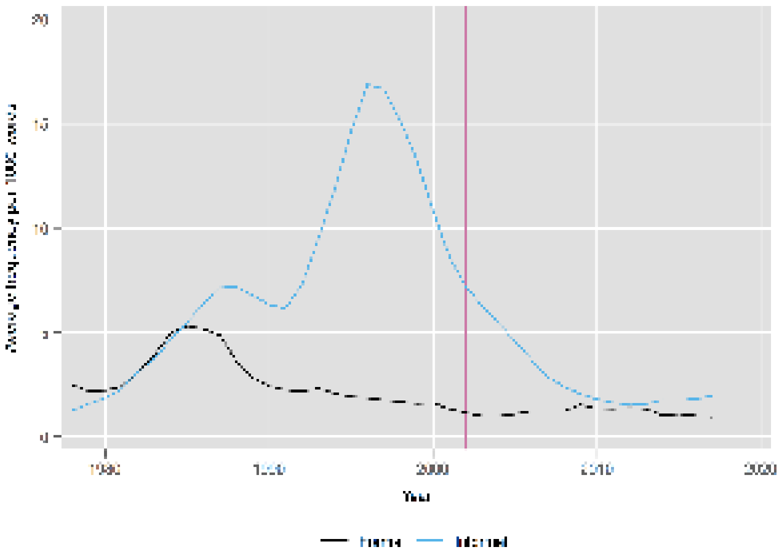


Figure 22: Trade in new regulations issued

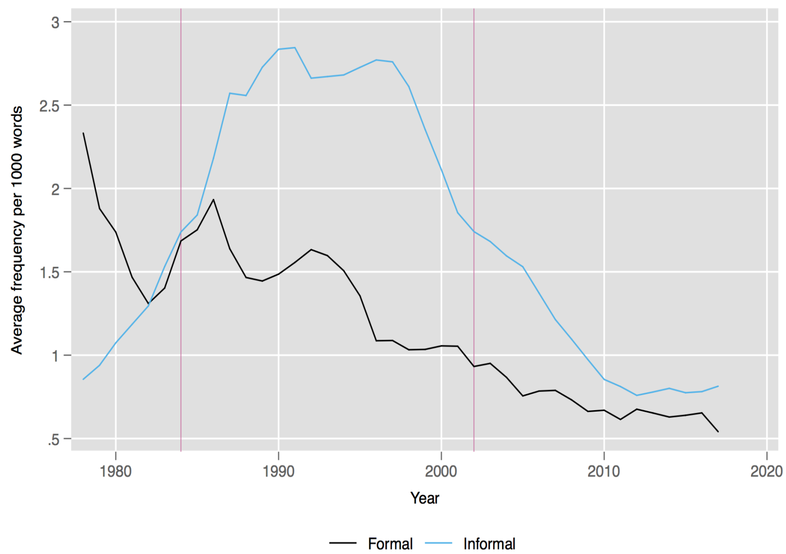


Figure 23: Property rights in new regulations issued

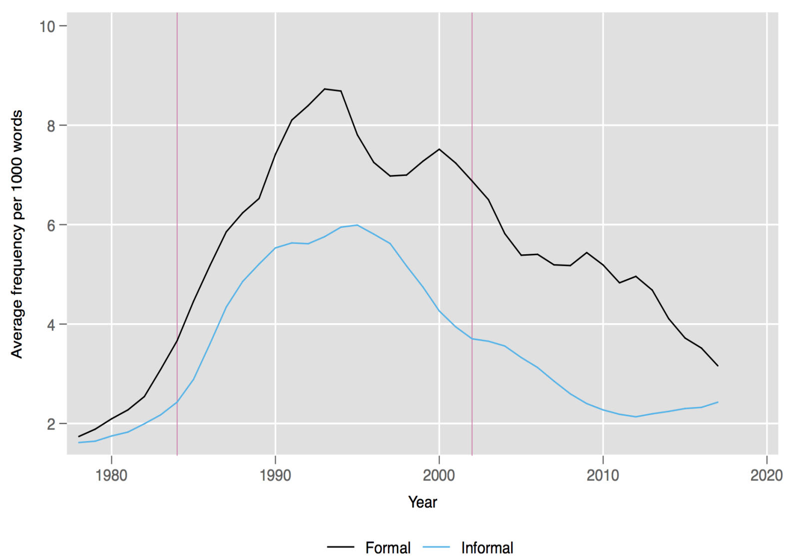


Figure 24: Privatization in new regulations issued

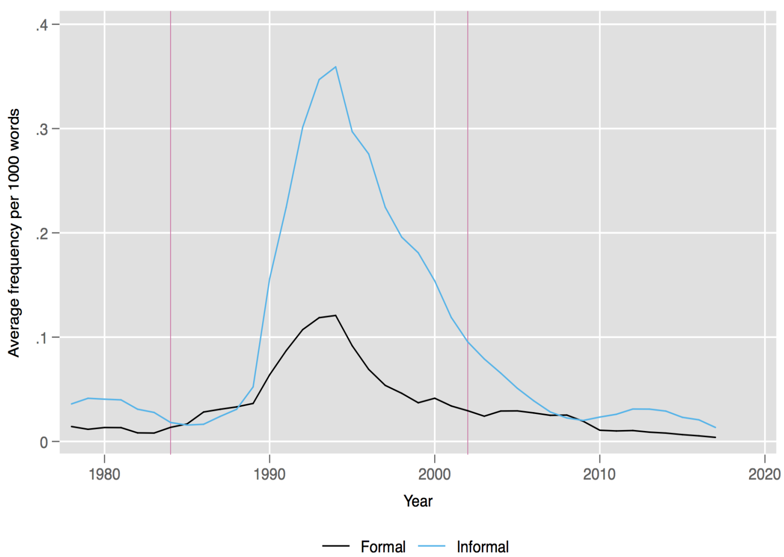


Table A1: 3-digit Area Categories

3-digit Category	PKULaw Category	Econ Related?
1	constitution	0
2	Government offices	0
3	Organs	0
4	Legal Work	0
5	lawyer	0
6	notarization	0
7	National Affairs	0
8	Overseas Chinese Affairs	0
9	Hong Kong and Macao Affairs	0
10	Taiwan Affairs	0
11	Religious Affairs	0
12	National Security	0
13	Foreign Affairs	0
14	public security	0
15	Civil administration	0
16	Civil Law	0
17	contract	1
18	Intellectual property	1
21	Anti-Unfair Competition	1
22	Marriage and adoption inheritance maintenance	1
23	Reform and Opening	1
24	plan	1
25	statistics	1
26	land	1
27	Map	1
28	Resources	1
29	energy	1
30	State-owned Assets	1
31	financial	1
32	tax	1
33	Financial affairs	1
35	accounting	1
36	audit	1
37	bank	1
38	Exchange	1
39	bill	1
40	Securities	1
41	lease	1
42	futures	1
43	price	1
44	Insurance	1
45	enterprise	1
46	the company	1
47	Foreign-invested enterprises	1

48	Individual economy	1
49	Construction industry	1
50	Industrial Management	1
61	agriculture	1
62	forestry	1
63	Animal husbandry	1
64	Fisheries	1
65	Water conservation	1
66	meteorological	1
67	Geology and Mineral Resources	1
68	Transportation	1
69	Warehousing	1
70	Posts and Telecommunications	1
71	Food Service	1
72	tourism	1
73	advisory	1
74	real estate	1
75	Business supplies	1
76	Foreign trade	1
77	Commodity inspection and quarantine of animals and plants	1
78	Customs	1
79	SAR	1
80	Zone	1
81	advertising	1
82	Business management	1
83	Standardization and Certification and Accreditation Administration	1
84	Measure	1
85	Quality management and supervision	1
86	Mediation and Arbitration	1
87	Labor unions	1
88	personnel	1
89	Environmental Protection	1
90	earthquake	1
91	education	1
92	Technology	1
93	Languages	0
94	Cultural relics	0
95	culture	0
96	health	0
97	Population and family planning	0
98	physical education	0
99	military	0
100	criminal law	0
101	Civil litigation	0
102	Economic trial	0
103	Maritime litigation	0
104	Administrative Litigation	0

105	Criminal proceedings	0
106	State compensation	0
107	Labor reform labor education prison prison	0
108	Judicial assistance	0
109	Prosecution business	0
110	Free trade pilot area	1
111	E - commerce	1
112	Judicial case release	0
113	Network car	1
114	Along the way	1
115	Big Data	1
116	the Internet	1
117	artificial intelligence	1