Reading China: Predicting Policy Change with Machine Learning

Weifeng Zhong
American Enterprise Institute
Julian Tszkin Chan
Bates White Economic Consulting

AEI Economics Working Paper 2018-11
Updated November 2018

Reading China: Predicting Policy Change with Machine Learning*

Julian TszKin Chan† Weifeng Zhong‡

November 15, 2018

Latest version available here.

Abstract

For the first time in the literature, we develop a quantitative indicator of the Chinese government's policy priorities over a long period of time, which we call the Policy Change Index (PCI) for China. The PCI is a leading indicator of policy changes that covers the period from 1951 to the third quarter of 2018, and it can be updated in the future. It is designed with two building blocks: the full text of the *People's Daily* — the official newspaper of the Communist Party of China — as input data and a set of machine learning techniques to detect changes in how this newspaper prioritizes policy issues. Due to the unique role of the *People's Daily* in China's propaganda system, detecting changes in this newspaper allows us to predict changes in China's policies. The construction of the PCI does not require the understanding of the Chinese text, which suggests a wide range of applications in other settings, such as predicting changes in other (ex-)Communist regimes' policies, measuring decentralization in central-local government relations, quantifying media bias in democratic countries, and predicting changes in lawmakers' voting behavior and in judges' ideological leaning.

Keywords: policy change, machine learning, China, *People's Daily*, propaganda

*We are grateful to Robert Barro, Kevin Hassett, Matt Jensen, Nicola Persico, Si Wu, Ruiqing Zhong, and conference and seminar participants at AACS, ACPS, ICCMS, ISA-South, ISA-ISSS, GMU, GWU, and TEXATA Summit for helpful comments. Laura Hou, Cody Kallen, Claire Li, and Erin Melly provided excellent research assistance. We thank NVIDIA Co. for the generous donation of a Titan Xp used for this research. The views expressed here are solely those of the authors and do not represent the views of the American Enterprise Institute, Bates White Economic Consulting, or their other employees. All errors are our own.

†Bates White Economic Consulting, ctszkin@gmail.com.

‡American Enterprise Institute, weifeng@weifengzhong.com.

1 Introduction

China's impressive economic growth in the past four decades has brought renewed attention to the model of government-led capitalism and its impact on the world. But the visible hand of government has had a prominent role throughout the industrialization process in post-imperial China,[1] be it coercive central planning or ambitious industrial policy. The Belt and Road Initiative and the "supply-side structural reform" under Xi Jinping's administration are just the latest signs of the persistence of China's developmental state.

Despite the pervasive role of government in the Chinese economy, until now there have been no *quantitative* indicators of the Chinese government's policy priorities over a long period of time. In this paper, we fill that gap by devising the first of such measures, the Policy Change Index (PCI) for China.[2] The PCI is a leading indicator — often moving before policy changes occur — that covers the period from 1951 to the third quarter of 2018, and it can be updated in the future.[3] In other words, this index not only helps us understand China's industrialization process in the past but also allows us to make short-term predictions about its future directions.[4]

The design of the PCI has two building blocks: (1) it takes as input data the full text of the *People's Daily* — the official newspaper of the Communist Party of China — since it was founded in 1946; (2) it employs a set of machine learning techniques — such as word embedding, multilayer perceptrons, and recurrent neural networks — to detect changes in how this newspaper prioritizes policy issues.

As we elaborate in Section 2, the source of the PCI's power to predict policy changes rests on the nature of the *People's Daily* — it is at the nerve center of China's propaganda system. Modeled closely after the Soviet counterpart, China's propaganda system is built on the premises that there is a set of goals toward which society should strive and that all social institutions, including the media, should be tools for moving public opinion, mobilizing economic resources, and, ultimately, attaining those goals. Unlike most newspapers in countries with a free press, which typically report current events after the fact, official media like the *People's Daily* are tasked with promoting the social process the regime sets out — often *in advance*. Therefore, by detecting (real-time) changes in propaganda, the PCI is, effectively,

1. For a detailed review, see Brandt, Ma, and Rawski (2014) and the references therein.

2. For simplicity, we will refer to "the PCI for China" simply as "the PCI" when there is no ambiguity.

3. The most up-to-date PCI data are available at `https://policychangeindex.com`. The source code of the project can be found at `https://github.com/open-source-economics/PCI`. The PCI is a quarterly series at the time of writing, but it will be updated to a monthly index.

4. The Economic Policy Uncertainty Index for China based on the method in Baker, Bloom, and Davis (2016) is a quantitative indicator that can potentially cover a long period of time. But it captures how uncertain the public thinks the policies are, not how the government will change them.

predicting (future) changes in policy.[5]

To detect changes in how the *People's Daily* prioritizes policy issues, we develop a machine learning algorithm that "reads" its articles and classifies whether they are published on the front page — a proxy for high priorities.[6] The learning is possible because our data set, described in Section 3, already has the articles' page numbers.

While we formally describe the model in Section 4, the intuition behind it can be explained with the following example. Imagine an avid reader of the *People's Daily*, whose mind our algorithm tries to mimic. If the reader had read, remembered, and thought through all the articles published in recent times, they would have acquired a fairly good sense of what kind of articles "should" or "should not" appear on the front page. But if the reader then woke up to a surprising paper the next morning — that is, their educated guess about the new paper turned out to work either particularly well or exceptionally poorly — it might constitute a signal of change from the reader's perspective. While a small surprise may well be taken as noise, a strong signal would convince the reader that their existing understanding of the page arrangement is no longer valid and that the priorities of the *People's Daily* have fundamentally changed. The design of our model mimics this reasoning in the reader's mind and builds the PCI on the "surprises" to the algorithm's existing understanding of the newspaper.[7]

As a preview of the main result, Figure 1 plots the quarterly PCI for China from 1951 to the third quarter of 2018. When the index hovers near zero, the new articles are largely confirming the "paradigm" the algorithm has acquired, suggesting policy stability. But if the index increases drastically, it would mean a big "surprise" to the algorithm's existing understanding, which, in turn, would indicate a major policy change in the near future.

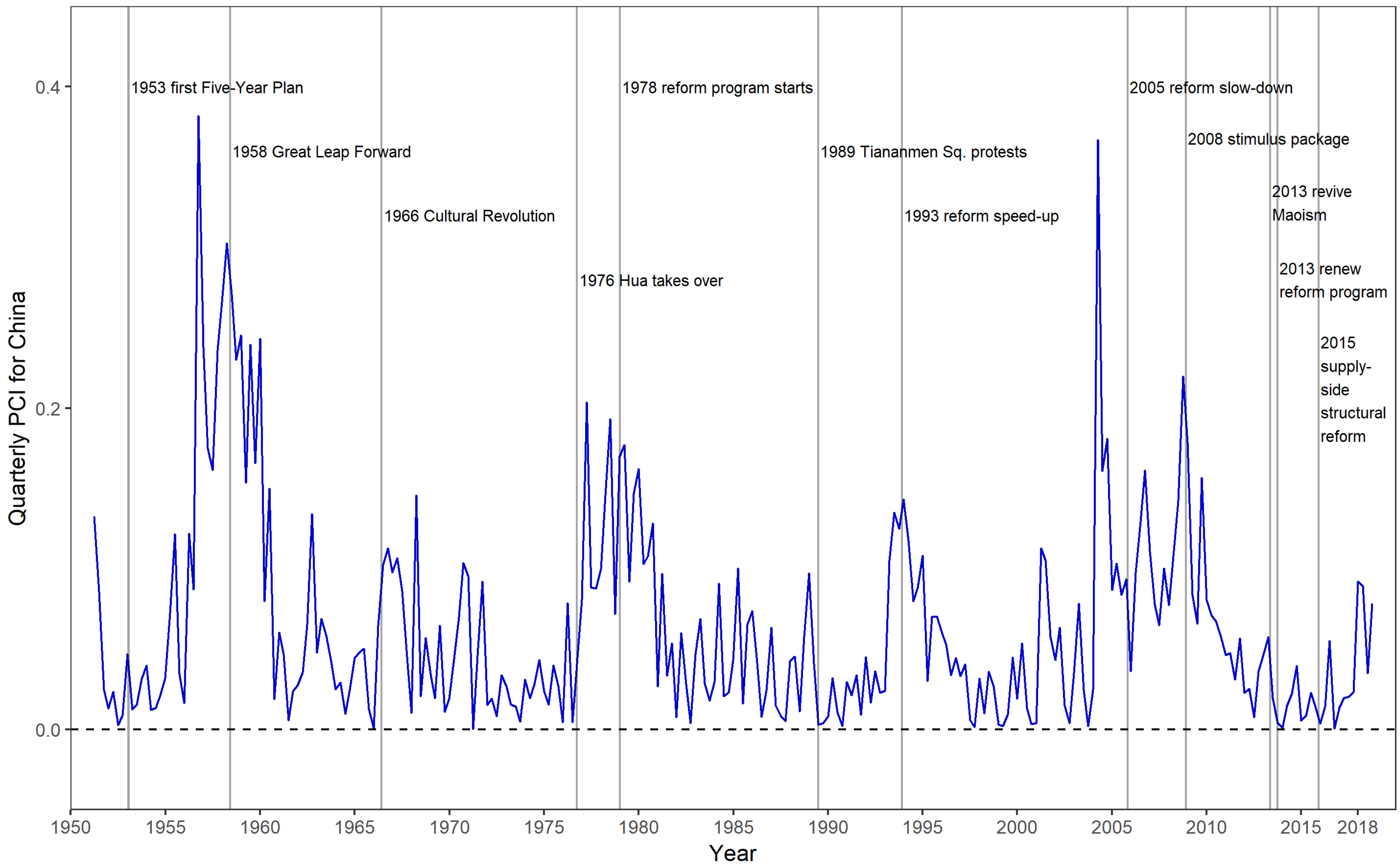
The validity of the PCI, which we address in Section 5, rests on whether it could have predicted important policy changes in China that have been identified in the literature — or, the ground truth. These major historical events are plotted in Figure 1 against the PCI time series for comparison. As shown in the figure, the PCI picks up the beginning of the Great Leap Forward in 1958, that of the Cultural Revolution in 1966, that of the economic reform

5. One may wonder if the commercialization of mass media in the post-reform China has affected the predictive power of our index. While the trend of commercialization has influenced most media outlets that are under less party control, the *People's Daily* has remained largely intact. See Section 2 for more discussion on this point.

6. Unlike newspapers in the West, Chinese newspapers do not have editorial pages. Instead, both news and commentaries are published on different pages — including the front page — depending on how important editors consider they are. Focusing on the front page, therefore, does *not* mean omitting editorials or commentaries.

7. As a consequent of this design, the PCI is agnostic about what *type* of changes it predicts. As long as a policy issue is high-priority enough to affect the *People's Daily*'s page arrangement, it will be reflected in the PCI. The index may pick up non-policy changes. It may also fail to predict policy changes that the government does *not* want to publicize in advance, or at all. See Section 4.6 for more details about the limitations of this design.

Figure 1: Policy Change Index and major events in China



Note: The PCI series is a predictor of policy changes. A spike in the PCI signals an upcoming policy change, while a vertical bar marks the occurrence of an actual policy change labeled by the respective event.

program in 1978, and, more recently, a reform speed-up in 1993 and a reform slow-down in 2005, among others. Furthermore, the PCI often leads these events by months, providing short-term predictions of the latter's occurrences.

Since the (quantitative) PCI does not contain substantive information about policies, one may wonder whether, based on what is known at the time of a PCI spike (but before the respective policy change), an observer would be able to tell what the substance of the upcoming policy change is. In Section 5, we examine the *People's Daily* articles that drive the "surprises" to the PCI, and we show that they are consistent with the policy changes that follow. In other words, the substance of policy change is also knowable to the observer.

The design of the PCI has an added feature that should be clear by now: It is "language-free." That is, predicting the occurrences of policy change, in itself, does *not* require the researcher to read and process the Chinese text.[8] The key ingredient of the algorithm is merely the page on which each article appears. This feature, therefore, suggests that the method could be applicable to a wide range of other settings, such as the possibility of predicting policy change in other (ex-)Communist regimes (such as Cuba and North Korea), measuring decentralization in central-local government relations (such as in China), measuring media bias in democracies (such as the US), and predicting changes in lawmakers' voting behavior and in judges' judicial or ideological leaning. We discuss these potential applications in Section 6.

1.1 Related Literature

This paper draws from the literature in at least three fields: political communication, media economics, and the applications of text analysis and machine learning.

First, at least since the publication of the classic *Four Theories of the Press* (Siebert, Peterson, and Schramm 1956), scholars in political communication have understood that a country's press system reflects the political institutions in which it functions.[9] The most well-studied case of this thesis is perhaps the Soviet model of mass communication — propaganda — which was not only created by the Soviet Union itself but also followed by China and other (ex-)communist states. The way the Soviet press operated, write Siebert, Peterson, and Schramm (1956, p. 5), was "[g]rounded in Marxist determinism and in the harsh political necessity of maintaining the political ascendancy of a party which represents less than ten percent of the country's people." Other well-known work on the Soviet propaganda system (e.g., Inkeles 1950; Kenez 1985) and its Chinese counterpart (e.g., Liu 1971; Schurmann 1966) are also based on the investigation of the nature of the respective regimes. Our study follows

8. Characterizing the substance of an upcoming policy change, however, would involve human reading of at least some of the articles.

9. See Hallin and Mancini (2004) for more recent developments in this literature.

this tradition, but we emphasize the reverse channel: Instead of how the media are influenced by the political system, we ask what inferences about politics and policies can be made from studying the press.

Second, in the field of media economics, substantial progress has been made in recent years in understanding whether and why the media are effective in changing people's preferences, beliefs, or behavior. While communication thinkers have long argued that propaganda "works" (e.g., Lasswell 1927; Lippmann 1922), recent studies show that the media can have a large effect on political outcomes in countries with weak democratic institutions, such as Mexico in the late 1990s (Lawson and McCann 2005) and Russia in the 1990s (Enikolopov, Petrova, and Zhuravskaya 2011). The effect is even more significant in authoritarian or totalitarian regimes. Gentzkow and Shapiro (2004) show that a broader range of information sources reduces hostility to America in Muslim countries. Yanagizawa-Drott (2014) shows that radio propaganda fueled participation in killings in the 1994 Rwandan genocide. Adena et al. (2015) find that radio propaganda helped the Nazis enroll new members and incited anti-Semitic acts.[10] These findings lend support to our analysis for an obvious reason: Had propaganda been ineffective, the Chinese government would not have invested massive resources on it.[11]

Other studies in this empirical literature provide a deeper understanding of why propaganda is effective, given that citizens may be aware that their government is propagandizing. Drawing from the exposure-acceptance model of persuasion (see McGuire 1968), Geddes and Zaller (1989) show that the support for the military regime in Brazil was the strongest among citizens in the broad middle ranges of political awareness — those who are informed enough to be exposed to government indoctrination but who are not so educated to be able to resist it. Kennedy (2009) finds a similar pattern using data in China, where the highest support for the regime is found in the large population of rural residents who have completed just junior high school — but no more — making them susceptible to political persuasion.

Third, this paper joins a booming literature on the applications of text analysis and machine learning to policy problems in economics and political science.[12]. Some studies gauge US monetary policy by examining the deliberation of policymakers on the Federal Open Market Committee (FOMC), such as the forming of opinion groups in FOMC discussions (Zirn, Meusel, and Stuckenschmidt 2015), the influence of FOMC members on one another (Guo et al. 2015; Schonhardt-Bailey 2013), and how the (internal) deliberation is affected by

10. Media effects have also been established in cross-border settings, such as the effect of West German television on East Germans (Kern and Hainmueller 2009), that of Serbian radio on Croats (DellaVigna et al. 2014), and that of Russian television on Ukrainian elections (Peisakhin and Rozenas 2017). But also see Crabtree, Darmofal, and Kern (2015) for a finding of no cross-border effect.

11. In the fiscal year 2015, for example, the budget for the Publicity Department of the Communist Party of China was 2.54 billion yuan — or, $400 million (Yang 2015).

12. This literature is too large for us to summarize all studies in this section. The interested reader can find more details in Athey (2017), Mullainathan and Spiess (2017), and Varian (2014)

external communication (Hansen, McMahon, and Prat 2018). Other scholars apply similar techniques to the prediction of lawmaking. Yano, Smith, and Wilkerson (2012) develop a model to predict whether a US Congressional bill will survive the committee process, while other algorithms are built to predict whether a bill will be voted and enacted into law (Gerrish and Blei 2011; Kraft, Jain, and Rush 2016; Nay 2017). Text analysis and machine learning are also applied to predict court rulings, such as the Supreme Court of the United States (Agrawal et al. 2017; Katz, Bommarito, and Blackman 2017; Sim, Routledge, and Smith 2016), the German Fiscal Courts (Waltl et al. 2017), and the European Court of Human Rights (Aletras et al. 2016). In urban economics, similar methods are used to measure and predict poverty by leveraging novel data, such as mobile phone metadata (Blumenstock, Cadamuro, and On 2015), street view photos (Naik, Raskar, and Hidalgo 2016), and satellite images (Engstrom, Hersh, and Newhouse 2016).

Our study differs from these applications in a critical way: While their common goal is to maximize an algorithm's predictive power on out-of-sample data, this paper focuses on making inferences from when the performance of our algorithm on those data is surprisingly good or surprisingly poor, compared to the performance in training.

2 Significance of the *People's Daily*

Throughout the history of the People's Republic, the *People's Daily* has been the most important print medium for China's party-state to propagate official viewpoints, communicate government directives, and signal future policies to the country and the rest of the world.[13]

The (political) significance of the *People's Daily* is apparent from its organizational connection to the Communist Party of China. The newspaper is under the direct control of the Publicity Department, which is a major division of the Party's Central Committee and the nerve center of China's propaganda system.[14] The Editor-in-Chief of the *People's Daily* is a minister-rank post, whose holder is often concurrently a deputy director of the Publicity Department.[15]

The Chinese leadership has been directly involved in the publication process of the *People's Daily*. Guoguang Wu, who served as a commentary editor of the *People's Daily* from 1985 to 1989, described in penetrating detail how the editorial process of the newspaper works (see G. Wu 1994). According to G. Wu, top officials of the Party are involved in two ways. First, they

13. When first established in 1946, the *People's Daily* was the official newspaper of a regional division of the Communist Party of China. It became the official newspaper of the Party in 1948.

14. The English name of the division used to be the "Propaganda Department," reflecting the fact that the word "propaganda" does not necessarily have a pejorative connotation in China's socialist context.

15. This is the case with Tuo Zhen, the Editor-in-Chief at the time of writing.

command subjects for the editorial team and specify the constraints by which it must abide. Some party leaders even write commentaries themselves. Mao, for example, frequently wrote for the newspaper in the 1940s. Second, editorials and commentators' articles are subject to tight censorship throughout the production process. Drafts and proofs are sent to party leaders for instructions on revision and final approval. The most important editorials have to be approved by the general secretary of the Party; that is, the leader of the country.

The consequence of disobeying the party control of the *People's Daily* is undoubtedly grave. Deng Tuo, the Editor-in-Chief from 1949 to 1957, is a telling example. In 1956, with the support of some reform-leaning party officials, Deng Tuo ran an editorial implicitly rebuking Mao's policies, much to the latter's dismay. The friction intensified in 1957 when Deng Tuo opted not to promote some of Mao's speeches in the newspaper, which resulted in Mao calling him "a dead man running the paper" and removing him from the Editor-in-Chief post.[16]

Given the highly official status of the *People's Daily*, it is not surprising that the change of propaganda messages in the newspaper is closely tied to — and often comes ahead of — the change of major government policies. Before the economic reform program took off in late 1978, for example, the *People's Daily* had seen a drastic shift in content and tone. On February 7, 1977, it published the now-notorious "Two Whatevers" statement — "We will resolutely uphold whatever decisions Chairman Mao made and unswervingly follow whatever instructions Chairman Mao gave" — embraced by Mao's designated successor upon his death. Deng Xiaoping and other reformists, who were increasingly taking hold of power at the time, contested the statement from within the Central Committee. Their efforts eventually succeeded and led to the publication of an almost negating article, "Practice is the Sole Criterion for Testing Truth," on May 12, 1978.[17] This change in the *People's Daily* had effectively foretold the economic reform program led by Deng Xiaoping — months in advance.

Since the economic reform program took off in 1978, China's mass media has undergone a notable change, which one might argue raises a question about the predictive power of our index. As documented by G. Wu (2000) and Y. Zhao (1998), there has been a significant increase in the number of commercialized media outlets in post-reform China, many of which are semi-official, less scrutinized by the Party, and more profit-driven. This change in the media industry even prompted the more optimistic of China scholars to project the possibility of a liberal, democratic press for China's future (e.g., Lynch 1999; Y. Zhao 1998). The reality,

16. See Wang (2008). After leaving the *People's Daily*, Deng Tuo kept writing articles critical of Mao's policies. He was later dubbed a member of the "anti-party group" and committed suicide at the dawn of the Cultural Revolution.

17. The subtext of this article is that, since practice was the *sole* criterion, Mao's decisions or instructions would cease to be a criterion, along with the Soviet-style planning the economic reforms were to overturn.

however, suggests otherwise. While the Chinese government has different policy goals in the reform era, its methodology of propaganda persists (Brady 2008), and the media commercialization has not produced more diverse political voices (Stockmann 2013).[18] In particular, the *People's Daily* remains the least commercialized and the most tightly controlled newspaper, upholding the party line in its reporting.[19] Therefore, the evidence suggests the persistence of the PCI's predictive power in the reform era.

3 Data

Our data set consists of 1,917,930 articles published in the *People's Daily* between May 15, 1946 (the date it was founded) and September 30, 2018. This data set is exhaustive except for 41 dates between 1946 and 1951 when the texts are not available.[20]

For each article, we have data for the date it was published, the page it appeared on, and the full text of the title and body of the article. The total number of front-page articles in our data set is 225,323 — or, 11.7%.

We also derive from the raw data the following metadata variables for each article: whether the date is a weekday or over a weekend, the length (in characters) of the title, the length of the body, the number of articles published that day, and the number of front-page articles published that day.

The fraction of front-page articles varies over time, which is mostly driven by the different number of articles the newspaper publishes each day. Figure 2 plots these two summary statistics in annual terms. As the *People's Daily* publishes more and more articles — from thousands to tens of thousands each year — the fraction of front-page article declines from 48% in 1946 to as low as 6% in 2018.

4 Methodology

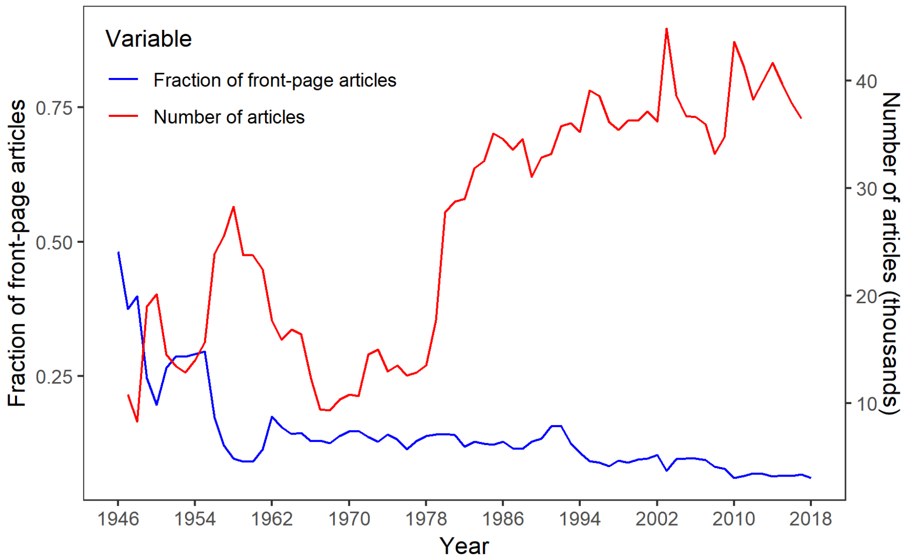
In this section, we describe how the PCI is constructed from the *People's Daily* data set and discuss this methodology in the context of supervised learning. The validity of its predictive

18. By popular measures of press freedom, such as the Reporters Without Borders' Press Freedom Index and the Freedom House's Freedom of the Press ranking, China's media sphere remains one of the most oppressed in the world.

19. In the commercialization process, if the *People's Daily* upholds the party line while other outlets present consumers with relatively more diverse choices, market competition would drive readership away from the *People's Daily*, which is exactly what happened: Its hard-copy circulation plummeted from five million copies per day in the 1980s to 1.8 million in 2004 (see Esarey 2006).

20. On those dates, either the newspaper did not publish or it did but the texts are not available to us. Either way, the fraction of missing dates (0.15%) is negligible.

Figure 2: Fraction of front-page articles and number of articles in the data set



power is discussed in the next section.

4.1 Overview

We model the PCI after the aforementioned avid reader of the *People's Daily* in the following steps: (1) We start with a quarterly rolling window with the length of five years, which we call the training window, and we divide the data within a training window into (i) training and validation data and (ii) testing data. We call the data in the quarter following the training window the "forecasting" data. (2) We train a machine learning algorithm, using the training and validation data, to classify whether each article appears on the front page (as if the reader was forming a paradigm about how the newspaper prioritizes issues). (3) We then capture the "surprise" to the algorithm by taking the difference (in absolute value) between its performance on the testing data (in the training window) and its performance on the "forecasting" data (in the following quarter). The PCI is defined simply as the magnitude of this "surprise" as the training window rolls forward.

A note on terminology: To avoid confusion, we use the word *classification* to refer to the problem of predicting whether an article appears on the front page. We reserve the word *prediction* for the problem of predicting policy changes using the constructed PCI. Furthermore, the data in the quarter following a training window are referred to as the *"forecasting" data*,[21] in contrast to the *testing data* within the window.

21. The word "forecasting" is in quotation marks because our data set already has the answer.

4.2 Data structure

An article in year-quarter t is defined as

$$A_{i,t} \equiv \{Y_{i,t}, X_{i,t}, Z_{i,t}\}, \tag{1}$$

where the index i denotes the i-th article in year-quarter t,[22] $Y_{i,t}$ is an indicator for whether the article is on the front page, $X_{i,t}$ is the texts of the article[23], and $Z_{i,t}$ represents the set of metadata variables mentioned in Section 3.

For each year-quarter s of interest, we define a training window T_s as the period of 20 year-quarters (five years) immediately preceding s. That is,

$$T_s \equiv \{t : s - 20 \leq t < s\}. \tag{2}$$

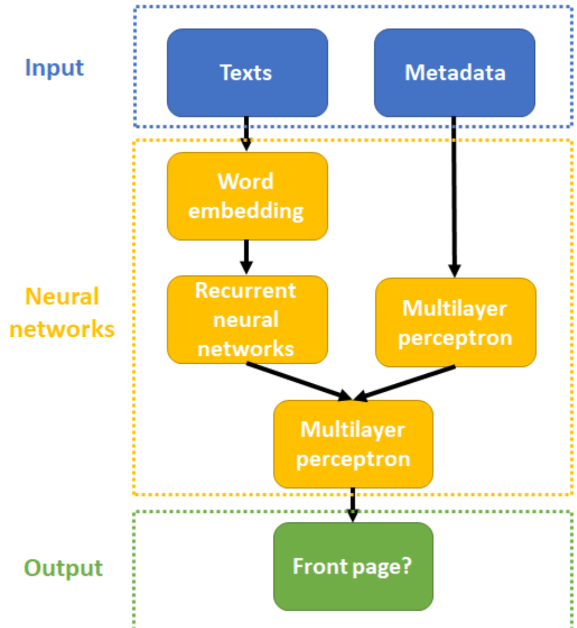
This structure is analogous to the aforementioned avid reader who lives in year-quarter s and remembers the articles that were published during the training window T_s.

Since the data in T_s will be used to train an algorithm, choosing the length of T_s involves a trade-off. On the one hand, a longer training window would contain more data for the algorithm to learn from, which potentially can improve its classification performance. On the other hand, more historical policy changes and noises would be covered by a longer training window, which may create confusion and render the learning process more challenging.[24] We have chosen five years as the length for our model after experimenting with different values. Appendix A.2, for example, investigates the alternative of using a ten-year rolling window, which produces a similar result.

For each training window T_s, we use a stratified sampling method to create training and validation data, which consist of 80% of the articles,[25] and testing data, which consist of the remaining 20%. The training data are used to compute a specified model. The validation data serve the purpose of searching for the optimal model specifications without over-fitting the training data. The purpose of the testing data is to compute how well an optimized model fits the training window without using the training and validation data, which the model has already "seen."

22. Formally, the index i should depend on the year-quarter t; that is, an article should be identified with $A_{i_t,t}$. We omit the subscript t of index i to simplify notations.
23. We concatenate the title and body of each article into one string of texts.
24. Another drawback of having a longer training window is a higher computation cost, of course.
25. We use 60% for training and 20% for validation.

Figure 3: Model overview



4.3 Model

For each year-quarter s, we train a model f_s such that, for each article in the training window T_s, the model classifies whether it is on the front page ($\hat{Y}_{i,t}$, as an estimate of the true $Y_{i,t}$) using the text of the article ($X_{i,t}$) and the respective metadata variables ($Z_{i,t}$), i.e.,

$$\hat{Y}_{i,t} = f_s(X_{i,t}, Z_{i,t}), \ \forall s, \ \forall t \in T_s. \tag{3}$$

Though the purpose of the model f_s is straightforward, its structure consists of a sequence of various types of neural network models, including a word embedding layer, a recurrent neural networks layer, and two multilayer perceptron layers. Figure 3 outlines the structure of the model f_s. The blue, yellow, and green nodes represent the inputs, neural network models, and output, respectively. An arrow means the (interim) output of the origin node is used as the (interim) input of the destination node. The model takes texts and metadata as its two inputs and, ultimately, produces a front-page classification as its output.

The text, as a sequence of words, is first fed to a word embedding layer, which maps each word into a numeric vector. In doing so, it reduces the dimensionality of texts while preserving the semantic relations between words. We then feed the outputs of the word embedding layer to a recurrent neural network, which is specialized in processing sequential data (such as sentences and articles) into a vector of hidden variables. Meanwhile, the metadata are fed to a multilayer perceptron, which creates hidden variables that summarize the features in the metadata. Finally, the hidden variables of the text and metadata are combined using another multilayer perceptron to generate a front-page classification.

Modeling the above layers requires the researcher to choose a set of model specifications (or, hyper-parameters), which control the complexity of each layer. The trade-off in this process lies in the fact that a greater complexity may improve the fit on the training data, but it may also worsen the performance on the validation data by over-fitting. To find the appropriate hyper-parameters, we implement a simulated annealing algorithm (à la Kirkpatrick, Gelatt, and Vecchi 1983) to search for the hyper-parameters that optimize the classification performance on the validation data.

The rest of this subsection briefly discusses each of the model's components. The reader more interested in how the model's output is used in subsequent steps may skip to Section 4.4.

4.3.1 Word Embedding

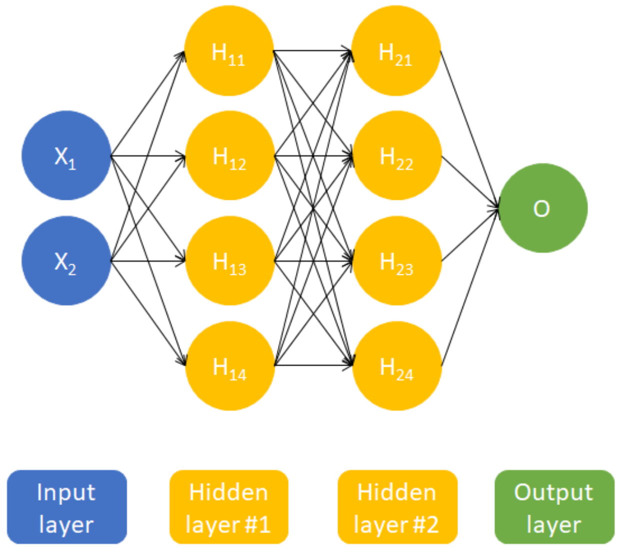
The curse of dimensionality kicks in quickly when texts are to be modeled as data. In the Chinese language, for example, there are tens of thousands of distinct Chinese characters, thousands of which are commonly used. For an article that contains a sequence of a thousand characters, the number of possible character combinations is enormous. Dimensionality reduction, therefore, is almost a necessity in textual analysis.

We adopt word embedding — a widely used technique in natural language processing pioneered by Mikolov et al. (2013) — to reduce the dimensionality of our data. It maps words or phrases to vectors of real numbers in a vector space. The method brings the dimensionality of texts down to that of the vector space while largely preserving the semantic relations between words or phrases. In this paper, we apply the Chinese-language word embedding developed by Li et al. (2018), which is a neural network model trained on all Chinese words and phrases that have appeared on the *People's Daily* — the same data source as ours — between 1946 and 2017.[26]

One of the most important features of word embedding is that it retains the linguistic meaning of words through their relations to other words. For example, the cosine distance between (the embedding of) two words measures the similarity between them. Thus, the cosine distance between "the United States" and "China" is smaller than, say, the cosine distance between "banana" and "umbrella." Moreover, the direction of a vector has linguistic

26. A caveat on using this word embedding, which is trained on the entire horizon, uniformly on all training windows: From the perspective of any training window, the embedding necessarily contains (future) information that has not emerged yet. For example, the meaning of a word in 2017 may be fed to a training window from 1960 to 1969, in which the same word meant differently back then. An alternative would be to train an embedding with data only up to, say, 1969. The drawback of doing this, however, is that the quality of the embedding would be lower due to the lack of data.

Figure 4: An example of multilayer perceptron



meaning as well. For example, the difference between "the United States" and "Washington, DC" — which is itself a vector in the same space — is similar to the difference between "China" and "Beijing."

4.3.2 Multilayer Perceptron

Multilayer perceptron (MLP) is a basic form of artificial neural network models. It contains three types of layers: the input layer, hidden layer(s), and the output layer. Figure 4 is a graphic illustration of an MLP with two hidden layers. Each node represents a variable, and each directional link represents a relationship between the nodes it connects. For example, variable H_{12} — the second note in the first hidden layer — is a function of a linear transformation of inputs X_1 and X_2. The variables in the first hidden layer are used as the inputs of the second hidden layer. Finally, the output layer is a function of a linear transformation of the second hidden layer's outputs.

Formally, the model in Figure 4 can be written as:

$$H_{(1)} = \sigma(X\beta_1 + b_1), \tag{4}$$

$$H_{(2)} = \sigma(H_{(1)}\beta_2 + b_2), \tag{5}$$

$$O = \sigma(H_{(2)}\beta_O + b_O), \tag{6}$$

where σ is a non-linear function — called the activation function in the literature — that maps a real number to a number between zero and one.[27]

The number of hidden layers and the number of hidden variables in each hidden layer (also known as neurons) are important hyper-parameters of the model specification. Optimizing these hyper-parameters and those in the next model component is part of the search algorithm we conduct in this paper.

4.3.3 Recurrent Neural Network

In the MLP model, the input layer is only used (as input) by the first hidden layer, and all inputs are treated equally by the model. This structure limits the ability of an MLP to model text data, which are a sequence of words. Recurrent neural networks (RNN) are a class of neural network models that specialize in handling texts and other sequential data.

Figure 5 gives an example of RNN. The input layer contains three sequential input variables, (X_1, X_2, X_3), which enter the model sequentially. In the context of text data, this input layer will be a sequence of word embeddings — think "She likes apples." The first input variable X_1 and a default variable H_0 are used as inputs to generate the first hidden variable H_1 and the first output variable O_1. The model then uses the second input variable X_2 and the hidden variable H_1 from the previous step to generate the second hidden variable H_2 and the second output variable O_2. Finally, the final output variable O_3 is generated similarly using variables X_3 and H_3.

Formally, the model in Figure 5 can be described as:

$$H_{(\ell)} = \sigma(X_{(\ell)}\beta_\ell + H_{(\ell-1)}\gamma_\ell + b_\ell), \tag{7}$$

$$O_{(\ell)} = \sigma(H_{(\ell)}\beta_O + b_O), \tag{8}$$

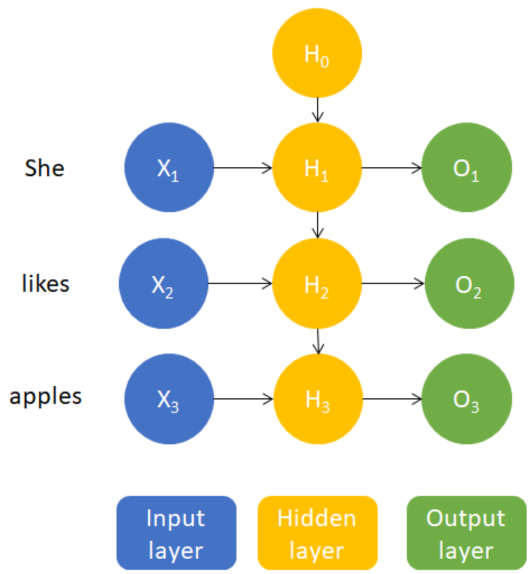
where ℓ is the index for the sequence of input, hidden, and output variables.

Since the model processes the input variables sequentially, there is an output variable at every step. For example, the output variable O_2 would be the output of the model if only the first two input variables were considered. In the context of text data, these interim output variables can be interpreted as the predictions of the model if only a truncated segment of the texts were "read" by the algorithm.

In this paper, we implement the gated recurrent units (GRU) model developed by Cho et al. (2014), which is a variant of RNN. A GRU model includes additional control units, which decide what information should be gated and what should be passed down to the next step to generate the hidden variables and outputs. It improves upon RNN methods by aiming

27. For example, a commonly used activation in the literature is the logistic function: $\sigma(x) = 1/(1 + e^x)$.

Figure 5: An example of recurrent neural network



to solve the vanishing gradient problem recognized in the literature.[28]

4.4 Calculation of the Policy Change Index

The framework described in the previous subsection generates a model f_s, which can be used to classify new articles in both the testing and "forecasting" data. For each article in year-quarter s, for example, the classification of whether it appears on the front page is given by

$$\hat{Y}_{i,s} = f_s(X_{i,s}, Z_{i,s}), \tag{9}$$

where $\hat{Y}_{i,s}$ is the estimate of $Y_{i,s}$ — the ground truth. Similarly, for an article in the testing data in the training window T_s, the classification would be

$$\hat{Y}_{i,t} = f_t(X_{i,t}, Z_{i,t}), \ \forall t \in T_s. \tag{10}$$

We use the F_1 score of the outcome variable to measure the performance of a model. The F_1 score is defined as the harmonic mean of the precision and recall of the model. That is,

$$F_1 = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}, \tag{11}$$

where the precision is the fraction of true positives in all predicted positives, and the recall is

28. See Cho et al. (2014) for more details.

the fraction of true positives in all actual positives.

The F_1 score, as an alternative to the accuracy rate, is meant to account for the issue of imbalanced samples; that is, having much fewer front-page articles than non-front-page articles. Imagine that the performance measure was the accuracy rate, which is the fraction of correct classifications in the whole sample. By this standard, a naïve algorithm that almost always makes negative classification would fare quite well because most articles are not on the front page, to begin with. However, the F_1 score of this naïve algorithm would be nearly zero because the recall would be nearly zero.

Finally, using the classifications in equations (9) and (10), we define the PCI at year-quarter s as the absolute value of the difference in classification performance between the training window T_s and year-quarter s. Or, formally,

$$PCI(s) \equiv |F_1(T_s) - F_1(s)|. \tag{12}$$

The PCI allows us to make policy predictions for the following reason. If $PCI(s)$ is near zero, it would mean that $F_1(T_s)$ and $F_1(s)$ are close to each other, which suggests that the way the *People's Daily* prioritizes articles in s has not changed much compared to T_s and that China's policy priorities have not changed much either. However, if $PCI(s)$ is much larger than zero, it would mean that $F_1(s)$ is either much higher or much lower than $F_1(T_s)$. Either way, such a "surprise" would suggest that the way the *People's Daily* prioritizes articles in s is fundamentally different from that in T_s, which, in turn, would suggest a fundamental shift in China's policy priorities.

4.5 Relationship with Supervised Learning

Although we employed supervised learning in constructing the PCI, it was not directly used to predict changes in policy priorities. In this subsection, we discuss the role of supervised learning in our methodology and how it differs from its conventional, off-the-shelf applications.

A supervised learning algorithm seeks a function θ that maps a set X of features onto a set Y of labels, which is the variable of our interest; that is,

$$\theta : X \to Y. \tag{13}$$

In a data-rich environment, θ could be precisely estimated and accurately predict the labels for "unseen" data.

An off-the-shelf application of supervised learning to our policy change prediction problem would go as follows. The features of each observation would be the text of an article and its metadata variables, such as whether it appears on the front page. The labels of that observation would be the policy priorities implied by the article. Formally, the mapping to be learned is given by

$$\theta : \{(\text{Article}, \text{FrontPage})\} \to \{(\text{Policy}, \text{Priority})\}. \tag{14}$$

For example, a front-page article that talks about how well private companies are doing would suggest privatizing the state sector — as a policy issue — is likely high-priority. In contrast, if a front-page article points out self-dealing and corrupt practices during privatization, it would suggest the opposite.

Once the function θ is estimated, it would be able to assign a Policy-Priority pair to each Article-FrontPage pair. It can, therefore, predict the policy priorities using the latest news articles.

However, this off-the-shelf approach is prohibitively challenging for two reasons. (i) Estimating θ requires a large number of labels (i.e., Policy-Priority pairs) for the training data, and the labeling process could be very costly. In fact, such a cost is one of the primary challenges in machine learning (Oliver et al. 2018). (ii) The labeling process could be subjective and biased because it relies on the researcher's understanding of the Chinese economy and politics, a complex and complicated subject.

The methodology of this paper looks at the relationship between the variables differently. Instead of building a model to predict the variable of interest directly, we consider it as a latent variable in a different supervised learning model:

$$\phi_{\{(\text{Policy},\text{Priority})\}} : \{\text{Article}\} \to \{\text{FrontPage}\}. \tag{15}$$

This model takes the policy priorities as given and simply indicates whether each article is on the front page. Since the policy priorities are removed from the model, we do not face any labeling challenges — page numbers are already available in the data set.

Although we did not model policy priorities explicitly, they are already embedded in the estimated function $\phi_{\{(\text{Policy},\text{Priority})\}}$. Consequently, the task of detecting changes in policy priorities has now been transformed to one of detecting changes in the function ϕ that embeds them. A change in policy priorities now means that a function ϕ that is trained to fit the data from a certain period (i.e., the training window) would perform very differently on the "unseen" data that come from a different period (i.e., the next quarter). The calculation of the PCI in Section 4.4 is based on this reasoning of the transformed problem.

Because we do not apply supervised learning off-the-shelf, the predictive power of our approach does *not* come from the accuracy of any single estimated function. Instead, it comes from the possibility that the estimated function changes because the latent variable — where our interest lies — has changed.

4.6 Advantages and Limitations

Because the *People's Daily* is the only data source for constructing the PCI, this approach has its advantages and limitations tied to the unique input.

While the page numbers of articles are easily available and seemingly trivial labels, they are in fact associated with underlying patterns (i.e., the newspaper's priorities) in which we are interested. Because we have transformed the policy change prediction problem in such a way that the front-page indicator is the label, the fact that the articles are in Chinese is no longer relevant — hence the "language-free" feature. This property suggests that the same methodology can be applied to settings that have a structure similar to ours. We devote Section 6 to exploring various potential applications along this line.

Another advantage of the PCI is that it minimizes the bias a researcher may have when evaluating results against the complicated reality. For example, an alternative design, which may seem more natural, would be to conduct a topic analysis[29] of all front-page articles and then investigate how the distribution over topics changes over time. But this approach would require the researcher to not only read the representative articles in all (estimated) topics and across the entire time series but also characterize their substance at an abstract level. The researcher's judgment in this step would become part of the input to the model and inevitably be carried over when the predictive power of the analysis is evaluated against the ground truth. Our design, in contrast, does not suffer from this subjectivity because the page numbers are objective labels, to begin with.

Our method also has several limitations. First, the PCI may pick up changes in the *People's Daily* that are not inherently about policies. For example, changes in culture, semantics, or terminologies could affect the PCI. However, these changes tend to take place gradually and, hence, are unlikely to cause abrupt fluctuations in the PCI. Editorial changes in the *People's Daily* may also be reflected in the PCI. For example, there have been periods of time when the newspaper published daily weather forecasts or selected quotations of Mao and his designated successor on the front page.

Second, for issues that are important but that the government does not publicize *in advance*, the PCI may not be able to predict them (ex ante) even if it can detect them as they occur. The Cultural Revolution, for example, started as a covert operation of Mao to purge his opponents (MacFarquhar and Schoenhals 2006) and, hence, was not promoted in advance. Also not publicized in advance was the four trillion yuan stimulus package of 2008, which was a policy *in response to* the global financial crisis.

Third, for issues that are important but that the government does not publicize *at all*, the

29. Such as the models built upon Latent Dirichlet Allocation (Blei, Ng, and Jordan 2003; Roberts et al. 2013) or Latent Semantic Analysis (Griffiths and Steyvers 2002; Hofmann 1999).

PCI is, by design, not meant to detect them even as they take place. The Tiananmen Square protests of 1989, as well as the three-year reform hiatus that followed, is a case in point. On June 4, the *People's Daily* published only one article about the suppression, which barely exceeds a hundred words. Yet, both the Director and the Editor-in-Chief of the *People's Daily* were removed from office thereafter (Tan 1990).

Finally, while the PCI can predict the occurrences of policy change, it does not tell an observer how long the new policies will last. This is because policy changes of different eventual durations may be preceded by the same publicity campaign in the *People's Daily*, rendering them indistinguishable ex ante. But if a policy lasts for a long time before it ends — long enough for the algorithm to learn the new patterns — the PCI can, in principle, pick up the end of the policy, but as another policy change.

5 Results

The ultimate test of the PCI's predictive power is the one against the history of China's policy changes. We approach this test by surveying the literature on China's economic history and reforms since the People's Republic was established.[30] We compile a chronology of major events from 1949 to 2018 that the literature has largely agreed on — see Appendix A.1 for details. These events are plotted against the quarterly PCI in Figure 1. A successful prediction would be a spike in the PCI that occurs in advance of a policy change.

5.1 Predicting the Timing of Policy Change

Comparing the events and the spikes in the PCI shows the extent to which the index is predictive. For the following policy changes, the PCI moves drastically, months before they occur: the beginning of the Great Leap Forward in 1958, that of the economic reform program in 1978, a reform speed-up known by the slogan "Socialist Market Economy" in 1993, and a reform slow-down known by the slogan "Socialist Harmonious Society" in 2004. All of these events have been widely recognized in the literature as among the most critical junctures in the economic history of the People's Republic.

While the PCI has not made major false alarms so far, there are important policy changes that the PCI did not pick up in advance: the beginning of the Cultural Revolution in 1966 and the four trillion yuan stimulus package rolled out in 2008. As explained in Section 4.6, the failure to predict these two changes is due to the nature of these events; the former was

30. Included in our survey are the following well-known work: Coase and Wang (2012), Howe, Kueh, and Ash (2002), Lardy (2014), Naughton (2014, 2018), Saich (2015), J. Wu (2014), and S. Zhao (2016).

Mao's covert political operation that he did not want to publicize, while the latter was a policy response to an unforeseen external shock.

There are other policy changes that the PCI did not pick up at all: the Tiananmen Square protests of 1989 and the reform hiatus that followed, as well as the three policy initiatives under Xi's administration — reviving Maoism, renewing reform, and the "supply-side structural reform." The failure to predict the former is, as expected, due to the political sensitivity of the issue. The failure to predict the latter occurred because the three initiatives, when viewed as a whole, largely amount to a continuation of the previous administration's agenda — or, no changes. We return to the discussion of Xi's policies in Section 5.2.3.

5.2 Understanding the Substance of Policy Change

From an observer's perspective, a sudden increase in the PCI time series signals only the occurrence but not the substance of an upcoming policy change. In this subsection, we discuss how to infer the policy substance from a PCI spike.

Recall that the PCI is the difference (in absolute value) between the algorithm's performance on the testing data (i.e., $F_1(T_s)$) and on the "forecasting" data (i.e., $F_1(s)$). Figure 6 plots these two performance measures on the same chart, with the orange curve representing the performance in testing and the purple curve in "forecasting." Because of the rolling-window design, the orange curve lags behind the purple curve by a few years and is smoother than the latter. Moreover, the level of the orange curve is important to understanding the status quo of policies. High performance in testing would mean it is relatively easy for the machine learning model to learn the priorities of the newspapers, which suggests those priorities are coherent and distinct, to begin with. In contrast, low performance in testing would mean it is difficult for the model to learn the newspaper's priorities very well, which suggests those priorities are ambiguous and vague throughout the training window.

By construction, a spike in the PCI in Figure 1 corresponds to a widening of the gap between the orange and purple curves in Figure 6. We show below that the policy implication of a PCI spike depends on whether the purple curve is much higher or much lower than the orange curve.

When a trained algorithm is applied to new data — either the testing data or the "forecasting" data — two types of classification errors occur (see Table 1). False positives are those articles that, according to the algorithm's understanding, "deserve" the front-page status but appear on other pages. False negatives, in contrast, are the articles that appear on the front page despite the algorithm suggesting otherwise.

When the purple curve in Figure 6 is much lower than the orange curve, the algorithm performs substantially more poorly on the "forecasting" data than on the testing data. This

Figure 6: Classification performance: testing versus “forecasting”

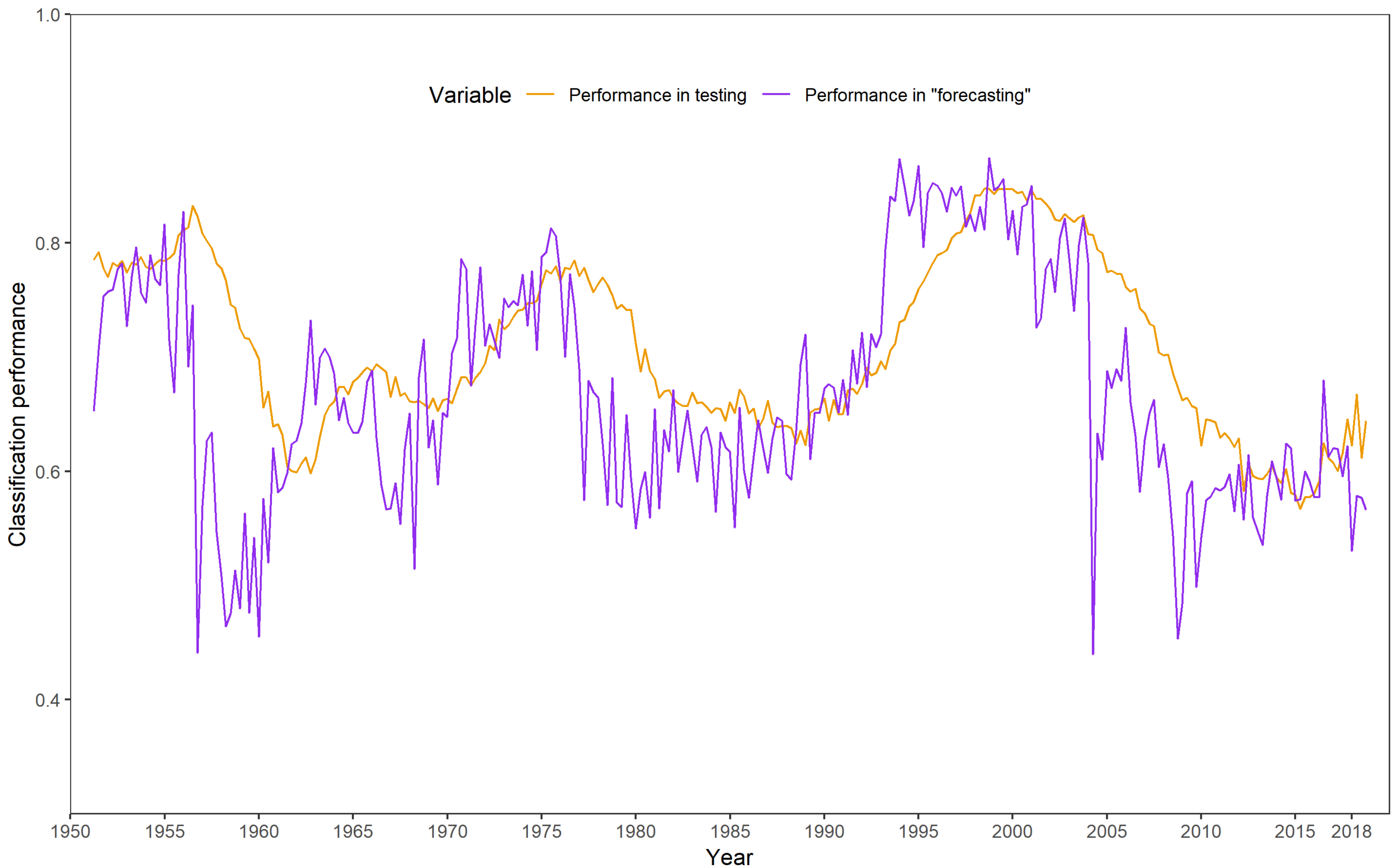


Table 1: Table of confusion for front-page classification

		Classified on front page?	
		No	Yes
On front page?	No	✓	False positives
	Yes	False negatives	✓

typically happens when the model is trained well in the rolling window. We argue that, in this case, *the false negatives in the "forecasting" data* are the articles that are most indicative of the new policy priorities. Imagine a policy issue — say, poverty — that used to be an unambiguously low priority in the training window. The algorithm would likely form a "paradigm" that assigns most poverty articles outside of the front page, and it would be mostly correct. But then in the new quarter, an unusually large number of poverty articles may come on the front page, conflicting with the algorithm's existing understanding. This would imply that poverty has become a high-priority issue in the new quarter.

When the purple curve in Figure 6 is much higher than the orange curve, the algorithm performs substantially better on the "forecasting" data than on the testing data. This typically happens when the priorities in the rolling window are difficult for the model to learn well. In this case, we argue that *the false positives in the testing data* are the articles that best indicate the policy priorities to come. Imagine two factions inside the party leadership, reformists and conservatives, who in the training window disagreed on privatization. The "paradigm" the algorithm forms — say, pro-privatization — would inevitably work poorly. This is because pro-privatization articles may go on the front page on one day but may not on another (false positives), and whether they do is driven by factors that cannot be detected from the newspaper itself (such as power struggle inside the Party). But then in the new quarter, most pro-privatization articles may come on the front page, which would confirm the "paradigm" in the new quarter even better than it used to in the training window. This would imply that privatization has become a high-priority issue in the new quarter.

In what follows, we apply the above analysis to two recent episodes of policy change, corresponding to the two cases, respectively. We end this section with a discussion of Xi's policies on which, at the time of writing, the PCI has not predicted any major changes.

5.2.1 1993: A Reform Speed-up

Since the PCI measures *change*, the policy implication of a PCI spike depends on what has been the status quo up to the point of change. Since 1978, the Chinese economy had experienced over 15 years of "dual-track reform," in which the traditional command economy and a newly established market mechanism coexist as methods to allocate resources. The state sector had remained largely protected by the government and reliant on planning, while the private sector had been allowed to grow concurrently through market competition. This unique strategy has prompted scholars of the Chinese economy to label it as "growing out of the plan" (Naughton 1995), "reform without losers" (Lau, Qian, and Roland 2000), and "incremental reform" (J. Wu 2005). The duality in policy also found its mirror image in the party leadership, where reformists and conservatives debated about the path forward. The antagonism between the two factions grew even more explicit after the Tiananmen Square protests of 1989.

An algorithm trained on this period of duality would work poorly in the training window because, while it may deem reform as a high priority, conservative policies at the same time would be perceived as noise and affect the algorithm's performance. This can be verified in Figure 6, where the orange curve is relatively low around 1992 compared to the rest of the series.[31]

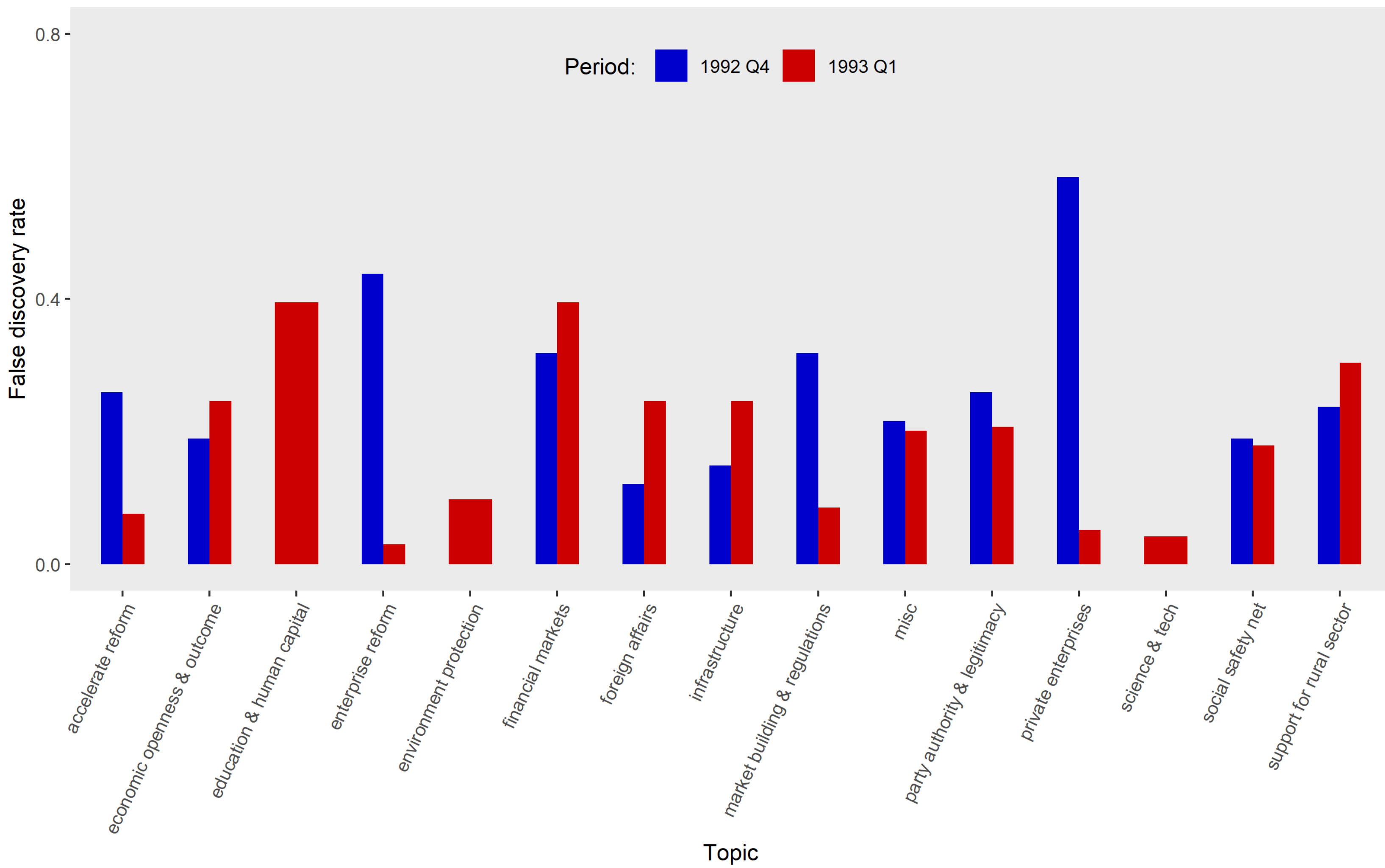
When the purple curve in Figure 6 suddenly increases in the first quarter of 1993, causing a spike in the PCI, we investigate the false positives in the training window to understand what issues are predicted to be the new priorities. We start with the testing data in the last quarter of 1992 — that is, the last quarter in the training window in question. We sample 100 articles from each of two categories — false positives and true positives — and classify them into a set of topics based on their content and our judgment. For each topic, we use the (estimated) fraction of false positives in all predicted positives, i.e., the *false discovery rate*, to measure how heavily the topic is featured in the false-positive category. Finally, we compare the false discovery rate in the last quarter of 1992 with that in the first quarter of 1993. If the latter is much smaller than the former for a certain topic, that topic would be among the predicted new priorities after the PCI spike.

Figure 7 shows the false discovery rate for all topics in the last quarter of 1992 (blue bars) against the first quarter of 1993 (red bars). Topics where the red bar is much lower than the blue bar would be considered, by the above analysis, new policy priorities. As the figure shows, these topics include reforming the state-owned enterprise system, promoting private enterprises, building and regulating markets, and, overall, accelerating the reform agenda.

This prediction is consistent with China's reform speed-up after 1993. Beginning in

31. In the early 1990s, the orange curve represents a low F_1 score around 0.67.

Figure 7: False discovery rate by topic, 1992 Q4 versus 1993 Q1



Note: False discovery rate is the fraction of false positives in all predicted positives.

November 1993, the Chinese government rolled out a broad reform agenda under the slogan of "Socialist Market Economy," which includes restructuring and, eventually, privatizing state-owned enterprises, unifying national markets and abandoning central planning, further liberalizing international trade and investment, enhancing macroeconomic controls, etc.[32]

5.2.2 2004: A Reform Slow-down

The status quo of the Chinese economy from the 1993 reform speed-up to the 2004 policy change had been a comprehensive, pro-market reform program affecting every sector of the economy. Furthermore, unlike the "reform without losers" before 1993, the post-1993 economic system had created significant losses in various social groups and driven up income and wealth inequalities among them. Most severely affected were workers laid off from the state sector and farmers who had not benefited from the urbanization process. An algorithm trained on this period of accelerated reform would easily associate pro-market policy issues as high priorities, and it would work well in the training window. Consistently, the orange curve in Figure 6 is relatively high around 2003 compared to the rest of the series.[33]

When the purple curve in Figure 6 suddenly declines in the first quarter of 2004, causing a PCI spike, we investigate the false negatives in the new quarter to understand what issues are predicted to be the new priorities. Similar to the analysis in Section 5.2.1, we sample and classify articles in the last quarter of 2003 and the first quarter of 2004. But different from before, we now focus on false-negative and true-negative articles. We calculate the fraction of false negatives in all predicted negatives, i.e., the *false omission rate*, as a measure of how heavily each topic is featured in the false-negative category. If the false omission rate is much higher in the first quarter of 2004 than in the last quarter of 2003 for a certain topic, that topic would be among the predicted new priorities after the PCI spike.

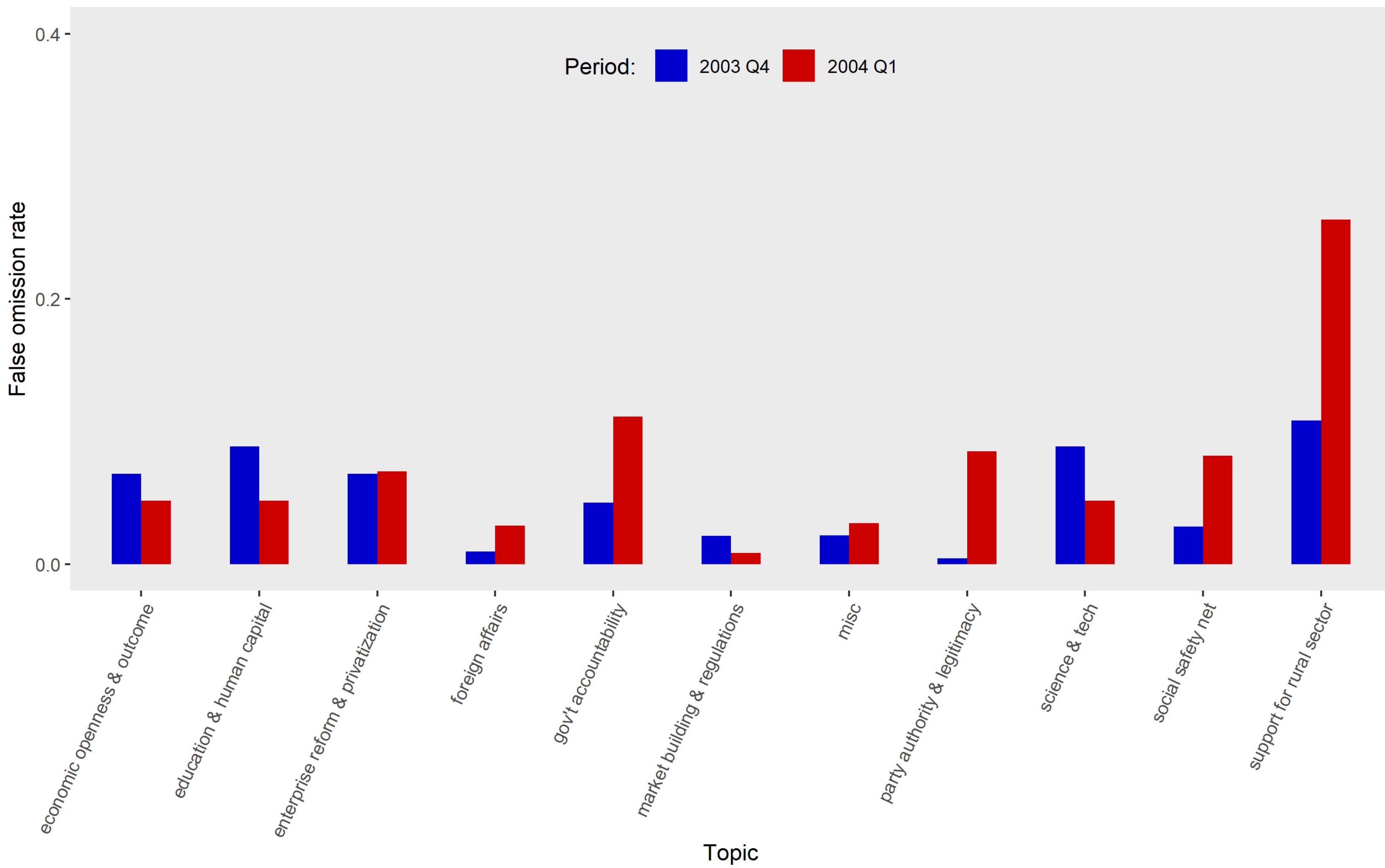
Figure 8 shows the false omission rate for all topics in the last quarter of 2003 (blue bars) against the first quarter of 2004 (red bars). Topics for which the red bar is much higher than the blue bar would be deemed new policy priorities. As the figure shows, these topics include increasing support for the rural sector, strengthening social safety nets (regarding poverty, social security, public health, etc.), enhancing government accountability to the general public, and asserting the Party's authority and legitimacy.

This prediction is also consistent with China's reform path after 2004. Driven by various social issues, which had intensified during the comprehensive reform, the Chinese government launched a set of populist social policies under the slogan of "Socialist Harmonious Society" in October 2005. In the meantime, the pace of pro-market reforms slowed down significantly. The

32. See Appendix A.1 for more details.

33. In the early 2000s, the orange curve represents a high F_1 score around 0.83.

Figure 8: False omission rate by topic, 2003 Q4 versus 2004 Q1



Note: False omission rate is the fraction of false negatives in all predicted negatives.

privatization of the state sector stabilized, and the government support for large, remaining state-owned enterprises was emphasized again.[34]

5.2.3 Xi Jinping's Administration, 2012–present

As shown in Figure 1, despite the several initiatives identified in the literature, the PCI has so far predicted no major policy changes during Xi's administration. We argue below that "no major changes" is indeed an accurate interpretation of the policy priorities under Xi's administration and that, based on past data, a spike in the PCI might be expected.

Between the launch of Hu's "Socialist Harmonious Society" in 2005 and the ascendancy of Xi in 2012, the status quo of policies in China had been the coexistence of maintaining the economic reform program — though at a slower pace — and tackling social issues with populist policies. The two conflicting initiatives rolled out by Xi in 2013, reviving Maoist ideology and calling for renewed reform, are essentially a reinforcement of the duo-agenda. The former initiative ordered government officials to fight the spread of subversive currents, including "ardently market-friendly neo-liberalism and 'nihilistic' critiques of [the Party's] traumatic past" (S. Zhao 2016, p. 85). The latter initiative, in stark contrast, called for "allowing the market to play the decisive role in resource allocation" (Naughton 2014, p. 4). The policy agenda under Xi, therefore, can be understood as having the same duality that had existed since Hu's administration.[35]

However, our "no major changes" prediction does *not* mean the policies under Xi and those under Hu are the same. Because the PCI design uses the front page as the proxy for high priorities, it ignores priority differences that are less important than that. Xi's and Hu's administrations, therefore, may well have policy differences that are present but not salient enough to affect the *People's Daily*'s front-page arrangement.

Consistent with the policy duality, our algorithm trained on this period performs relatively poorly in the training window. As shown in Figure 6, the performance in testing (the orange curve) during Xi's administration hovers around 0.6, substantially lower than other years.

While we are agnostic about how to predict the future PCI using the current PCI,[36] one may argue that having a bundle of incoherent policies is not sustainable. The last time this happened in the PCI time series was the "dual-track reform" leading up to 1993. As discussed in Section 5.2.1, the duality was not stable, and it eventually led to a major policy change, moving the Chinese economy toward an era of comprehensive market reforms. Therefore, if

34. See Appendix A.1 for more details.

35. The "supply-side structural reform," launched in 2015, was a government-led effort to reduce excess capacity in heavy industries and, hence, does not fit the definition of market-oriented reform either (Naughton 2018, p. 123).

36. This would be a different problem from predicting future policies using the current PCI.

history is any guide, a major policy change later under Xi's administration might be expected as well.

6 Other Applications

The backbone of our approach to the PCI for China is taking readily available — and seemingly trivial — labels (i.e., the page numbers of *People's Daily* articles) and inferring underlying changes that are associated with those labels (i.e., policies). In this section, we discuss how the same reasoning can be transformed and applied to a variety of other contexts.

The *People's Daily* is by no means a unique case in (ex-)Communist regimes. In fact, the *People's Daily* is to China as the *Pravda* is to the USSR, the *Neues Deutschland* is to the GDR (East Germany), the *Granma* is to Cuba, the *Rodong Sinmun* is to the DPRK (North Korea), the *Nhân Dân* is to Vietnam, etc. The same design, therefore, can be directly applied to these countries to produce the respective indices of policy change — the PCI for Cuba, the PCI for North Korea, etc.[37]

Even within an (ex-)Communist regime, there are regional official newspapers, which can be used to produce the respective regional indices of policy change. In China, for example, each provincial committee of the Communist Party has its own official newspaper. These regional papers are supposed to toe the party line, but they may not in reality. The less correlated a provincial index is to the PCI for China, the more independent that province likely is from China's central government. Moreover, how divergent provincial indices are from the PCI for China can be interpreted as a measure of decentralization in China's central-local government relations.

Although newspapers in a free press are not official media, applying the same techniques to them may lead to interesting discoveries. Consider US newspapers with different ideological leanings — e.g., *The Washington Post*, *The New York Times*, and *The Wall Street Journal*. Applying the PCI design to each of these papers would produce an index of its own. While these indices would not predict US policy change, their positions relative to each other may provide a (relative) measure of media bias. If the three "PCI curves" diverged, for example, it would signal a more polarized media landscape even though the "PCI for the US" — which probably lies in between them — remains unknown.

Names of people are also readily available labels for studying underlying changes. Consider a large volume of lawmakers' public statements, such as floor speeches and press releases. An algorithm similar to ours can be trained to link lawmakers' statements to their identities, which

37. Moreover, in the cases of North Korea, Cuba, and Vietnam, since the regimes still exist and the newspapers are still in print, the constructed indices would help researchers and policymakers make short-term predictions of these countries' future policies.

potentially could predict changes in the lawmakers' future voting behavior. For example, a speech by Legislator A might be mistaken by the algorithm as coming from Legislator B. This would mean that A is now taking a stance close to where B used to be. To the extent that legislators find it necessary to justify their votes to their constituents, the algorithm's misclassification would suggest that, in the near future, A might vote in a way that is consistent with B's past records, rather than with A's own.

The same reasoning can be applied to predicting changes in judges' judicial opinions as well. An algorithm trained on past judicial opinions might mistake a new opinion written by Judge C as coming from Judge D. This would mean that C is now ruling on cases in a way similar to how D used to. As to making predictions, C's future rulings might resemble D's more and more, considering that when judges' stances shift — e.g., from conservative to liberal ones — such changes can be gradual yet long-lasting. Furthermore, applying this method to supreme courts would be particularly fruitful because the judges (or, justices) have long careers (more data for the algorithm to learn) and they decide on the highest-profile cases that profoundly affect their countries.

We end the discussion on potential applications with a cautionary note. While our "language-free" method to predict changes can be easily replicated on a wide range of applications, the ultimate test of the predictive power should be the one against the ground truth — i.e., the *actual* changes in other (ex-)Communist countries' policies, degree of decentralization, degree of media bias, lawmakers' voting behavior, or judges' judicial or ideological leaning. The researcher's careful analysis and judicious treatment of the respective contexts is crucial to ensuring the validity of the predictive models.

7 Conclusion

We have developed the PCI for China, which covers the period from 1951 to the third quarter of 2018 and can make short-term predictions about China's future policy agendas. We have achieved this by using a unique data set — the full text of the *People's Daily* — and applying recent advances in machine learning to analyze it. The constructed PCI allows us to not only predict the timing of policy changes but also understand the substance of these changes before they are realized.

One may wonder how long the predictive power of the PCI will last; we are agnostic on this question. The fundamental assumptions of our approach — that China is an authoritarian regime and that the *People's Daily* is its mouthpiece — may or may not hold in the future. Though it is not very likely, the Chinese government may designate another official newspaper in the *People's Daily*'s stead. Or, China may cease to be an authoritarian regime altogether.[38] But we conjecture that, before either scenario becomes a reality, the index will likely remain indicative of China's policy agenda.

The design of the PCI has a "language-free" feature, which opens the door to a wide range of applications in settings that are structurally similar to ours. Applying a similar analysis to other (ex-)Communist newspapers, newspapers in a free press, lawmakers' public statements, and judges' judicial opinions are among the examples to which our approach can relate. In two ongoing studies, we explore the applications using Cuba's official newspaper, *Granma*, and *The New York Times*. We leave these and the other potential applications for future research.

38. Obviously, should either scenario occur, there will be no reason to keep updating this index.

References

Adena, Maja, Ruben Enikolopov, Maria Petrova, Veronica Santarosa, and Ekaterina Zhuravskaya. 2015. "Radio and the Rise of the Nazis in Prewar Germany." *The Quarterly Journal of Economics* 130 (4): 1885–1939.

Agrawal, Sharan, Elliott Ash, Daniel Chen, Simranjyot S. Gill, Amanpreet Singh, and Karthik Venkatesan. 2017. *Affirm or Reverse? Using Machine Learning to Help Judges Write Opinions.* Working Paper.

Aletras, Nikolaos, Dimitrios Tsarapatsanis, Daniel Preoţiuc-Pietro, and Vasileios Lampos. 2016. "Predicting Judicial Decisions of the European Court of Human Rights: A Natural Language Processing Perspective." *PeerJ Computer Science* 2:e93.

Athey, Susan. 2017. "Beyond Prediction: Using Big Data for Policy Problems." *Science* 355 (6324): 483–485.

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

Blei, David M., Andrew Y. Ng, and Michael I. Jordan. 2003. "Latent Dirichlet Allocation." *Journal of Machine Learning Research* 3:993–1022.

Blumenstock, Joshua, Gabriel Cadamuro, and Robert On. 2015. "Predicting Poverty and Wealth from Mobile Phone Metadata." *Science* 350 (6264): 1073–1076.

Brady, Anne-Marie. 2008. *Marketing Dictatorship: Propaganda and Thought Work in Contemporary China.* Lanham, MD: Rowman & Littlefield.

Brandt, Loren, Debin Ma, and Thomas G. Rawski. 2014. "From Divergence to Convergence: Reevaluating the History behind China's Economic Boom." *Journal of Economic Literature* 52 (1): 45–123.

Cho, Kyunghyun, Bart van Merriënboer, Çağlar Gülçehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. "Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation." In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing.*

Coase, Ronald, and Ning Wang. 2012. *How China Became Capitalist.* New York, NY: Palgrave Macmillan.

Crabtree, Charles, David Darmofal, and Holger L. Kern. 2015. "A Spatial Analysis of the Impact of West German Television on Protest Mobilization during the East German Revolution." *Journal of Peace Research* 52 (3): 269–284.

DellaVigna, Stefano, Ruben Enikolopov, Vera Mironova, Maria Petrova, and Ekaterina Zhuravskaya. 2014. "Cross-Border Media and Nationalism: Evidence from Serbian Radio in Croatia." *American Economic Journal: Applied Economics* 6 (3): 103–132.

Engstrom, Ryan, Jonathan Hersh, and David Newhouse. 2016. *Poverty in HD: What Does High Resolution Satellite Imagery Reveal about Economic Welfare?* Working Paper.

Enikolopov, Ruben, Maria Petrova, and Ekaterina Zhuravskaya. 2011. "Media and Political Persuasion: Evidence from Russia." *American Economic Review* 101 (7): 3253–3285.

Esarey, Ashley. 2006. *Caught between State and Society: The Commercial News Media in China.* New York, NY: Columbia University Press.

Geddes, Barbara, and John Zaller. 1989. "Sources of Popular Support for Authoritarian Regimes." *American Journal of Political Science* 33 (2): 319–347.

Gentzkow, Matthew A., and Jesse M. Shapiro. 2004. "Media, Education and Anti-Americanism in the Muslim World." *Journal of Economic Perspectives* 18 (3): 117–133.

Gerrish, Sean, and David M. Blei. 2011. "Predicting Legislative Roll Calls from Text." In *Proceedings of the 28th International Conference on Machine Learning,* 489–496.

Griffiths, Thomas L., and Mark Steyvers. 2002. "A Probabilistic Approach to Semantic Representation." In *Proceedings of the Twenty-Fourth Annual Conference of the Cognitive Science Society,* 381–386.

Guo, Fangjian, Charles Blundell, Hanna Wallach, and Katherine Heller. 2015. "The Bayesian Echo Chamber: Modeling Social Influence via Linguistic Accommodation." In *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics,* edited by Guy Lebanon and S. V. N. Vishwanathan, 315–323.

Hallin, Daniel C., and Paolo Mancini. 2004. *Comparing Media Systems: Three Models of Media and Politics.* New York, NY: Cambridge University Press.

Hansen, Stephen, Michael McMahon, and Andrea Prat. 2018. "Transparency and Deliberation within the FOMC: A Computational Linguistics Approach." *The Quarterly Journal of Economics* 133 (2): 801–870.

Hofmann, Thomas. 1999. "Probabilistic Latent Semantic Analysis." In *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence,* 289–296.

Howe, Christopher, Y. Y. Kueh, and Robert Ash. 2002. *China's Economic Reform: A Study with Documents.* London, UK: RoutledgeCurzon.

Inkeles, Alex. 1950. *Public Opinion in Soviet Russia: A Study in Mass Persuasion.* Cambridge, MA: Harvard University Press.

Katz, Daniel M., Michael J. Bommarito II, and Josh Blackman. 2017. "A General Approach for Predicting the Behavior of the Supreme Court of the United States." *PLOS ONE* 12 (4): 1–18.

Kenez, Peter. 1985. *The Birth of the Propaganda State: Soviet Methods of Mass Mobilization, 1917-1929.* New York, NY: Cambridge University Press.

Kennedy, John James. 2009. "Maintaining Popular Support for the Chinese Communist Party: The Influence of Education and the State-Controlled Media." *Political Studies* 57 (3): 517–536.

Kern, Holger Lutz, and Jens Hainmueller. 2009. "Opium for the Masses: How Foreign Media Can Stabilize Authoritarian Regimes." *Political Analysis* 17 (4): 377–399.

Kirkpatrick, S., C. D. Gelatt, and M. P. Vecchi. 1983. "Optimization by Simulated Annealing." *Science* 220 (4598): 671–680.

Kraft, Peter, Hirsh Jain, and Alexander M. Rush. 2016. "An Embedding Model for Predicting Roll-Call Votes." In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing,* 2066–2070.

Lardy, Nicholas R. 2014. *Markets Over Mao: The Rise of Private Business in China.* Washington, DC: Peterson Institute for International Economics.

Lasswell, Harold D. 1927. *Propaganda Technique in the World War.* Oxford, England: Knopf.

Lau, Lawrence J., Yingyi Qian, and Gerard Roland. 2000. "Reform without Losers: An Interpretation of China's Dual-Track Approach to Transition." *Journal of Political Economy* 108 (1): 120–143.

Lawson, Chappell, and James A. McCann. 2005. "Television News, Mexico's 2000 Elections and Media Effects in Emerging Democracies." *British Journal of Political Science* 35 (1): 1–30.

Li, Shen, Zhe Zhao, Renfen Hu, Wensi Li, Tao Liu, and Xiaoyong Du. 2018. "Analogical Reasoning on Chinese Morphological and Semantic Relations." In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics,* 138–143.

Lippmann, Walter. 1922. *Public Opinion.* New York, NY: MacMillan Co.

Liu, Alan P. L. 1971. *Communications and National Integration in Communist China.* Berkeley, CA: University of California Press.

Lynch, Daniel C. 1999. *After the Propaganda State: Media, Politics, and "Thought Work" in Reformed China.* Stanford, CA: Stanford University Press.

MacFarquhar, Roderick, and Michael Schoenhals. 2006. *Mao's Last Revolution.* Cambridge, MA: Harvard University Press.

McGuire, William J. 1968. "Personality and Susceptibility to Social Influence." In *Handbook of Personality: Theory and Research,* edited by Edgar F. Borgatta and William Wilson Lambert, 1130–1187. Chicago, IL: Rand McNally.

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. "Efficient Estimation of Word Representations in Vector Space." In *Proceedings of the International Conference on Learning Representations.*

Mullainathan, Sendhil, and Jann Spiess. 2017. "Machine Learning: An Applied Econometric Approach." *Journal of Economic Perspectives* 31 (2): 87–106.

Naik, Nikhil, Ramesh Raskar, and César A. Hidalgo. 2016. "Cities Are Physical Too: Using Computer Vision to Measure the Quality and Impact of Urban Appearance." *American Economic Review* 106 (5): 128–32.

Naughton, Barry J. 1995. *Growing Out of the Plan: Chinese Economic Reform, 1978–1993.* Cambridge University Press.

———. 2014. "After the Third Plenum: Economic Reform Revival Moves toward Implementation." *China Leadership Monitor* 43.

———. 2018. *The Chinese Economy: Adaptation and Growth.* 2nd ed. Cambridge, MA: MIT Press.

Nay, John J. 2017. "Predicting and Understanding Law-Making with Machine Learning." *PLOS ONE* 12 (5): 1–14.

Oliver, Avital, Augustus Odena, Colin Raffel, Ekin D. Cubuk, and Ian J. Goodfellow. 2018. "Realistic Evaluation of Deep Semi-Supervised Learning Algorithms." *arXiv preprint arXiv:1804.09170.*

Peisakhin, Leonid, and Arturas Rozenas. 2017. "Electoral Effects of Biased Media: Russian Television in Ukraine." *American Journal of Political Science* 62 (3): 535–550.

Roberts, Margaret E., Brandon M. Stewart, Dustin Tingley, and Edoardo M. Airold. 2013. "The Structural Topic Model and Applied Social Science." In *Advances in Neural Information Processing Systems Workshop on Topic Models: Computation, Application, and Evaluation.*

Saich, Tony. 2015. *Governance and Politics of China.* 4th ed. Houndmills, UK: Palgrave Macmillan.

Schonhardt-Bailey, Cheryl. 2013. *Deliberating American Monetary Policy: A Textual Analysis.* Cambridge, MA: MIT Press.

Schurmann, Franz. 1966. *Ideology and Organization in Communist China.* Berkeley, CA: University of California Press.

Siebert, Fred S., Theodore Peterson, and Wilbur Schramm. 1956. *Four Theories of the Press: The Authoritarian, Libertarian, Social Responsibility, and Soviet Communist Concepts of What the Press Should Be and Do.* Urbana, IL: University of Illinois Press.

Sim, Yanchuan, Bryan Routledge, and Noah A. Smith. 2016. "Friends with Motives: Using Text to Infer Influence on SCOTUS." In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing,* 1724–1733.

Stockmann, Daniela. 2013. *Media Commercialization and Authoritarian Rule in China.* New York, NY: Cambridge University Press.

Tan, Frank. 1990. “The People’s Daily: Politics and Popular Will — Journalistic Defiance in China During the Spring of 1989.” *Pacific Affairs* 63 (2): 151–169.

Varian, Hal R. 2014. “Big Data: New Tricks for Econometrics.” *Journal of Economic Perspectives* 28 (2): 3–28.

Waltl, Bernhard, Georg Bonczek, Elena Scepankova, Jörg Landthaler, and Florian Matthes. 2017. “Predicting the Outcome of Appeal Decisions in Germany’s Tax Law.” In *Electronic Participation,* edited by Peter Parycek, Yannis Charalabidis, Andrei V. Chugunov, Panos Panagiotopoulos, Theresa A. Pardo, Øystein Sæbø, and Efthimios Tambouris, 89–99.

Wang, Binbin. 2008. “Deng Tuo yu diyi bu Mao Ze Dong xuanji de chuban [Deng Tuo and the publication of the first *Selected Works of Mao Zedong*].” *Nanfang zhoumo [Southern Weekly]* (April 16).

Wu, Guoguang. 1994. “Command Communication: The Politics of Editorial Formulation in the People’s Daily.” *The China Quarterly* 137:194–211.

———. 2000. “One Head, Many Mouths: Diversifying Press Structures in Reform China.” Chap. 2 in *Power, Money, and Media: Communication Patterns and Bureaucratic Control in Cultural China,* edited by Chin-Chuan Lee, 45–67. Evanston, IL: Northwestern University Press.

Wu, Jinglian. 2005. *Understanding and Interpreting Chinese Economic Reform.* 1st ed. Cincinnati, OH: Thomson/South-Western.

———. 2014. *Understanding and Interpreting Chinese Economic Reform.* 2nd ed. Singapore: Gala Asia.

Yanagizawa-Drott, David. 2014. “Propaganda and Conflict: Evidence from the Rwandan Genocide.” *The Quarterly Journal of Economics* 129 (4): 1947–1994.

Yang, Zhijin. 2015. “Xuexi dangbu yusuan: wu bumen gonggong yusuan zhichu 37.68 yi [Learn about the budgets for Party departments: total budget for five departments reaches 3.768 billion yuan].” *21 shiji jingji baodao [21st Century Business Herald]* (March 18).

Yano, Tae, Noah A. Smith, and John D. Wilkerson. 2012. “Textual Predictors of Bill Survival in Congressional Committees.” In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies,* 793–802.

Zhao, Suisheng. 2016. “Xi Jinping’s Maoist Revival.” *Journal of Democracy* 27 (3): 83–97.

Zhao, Yuezhi. 1998. *Media, Market, and Democracy in China: Between the Party Line and the Bottom Line.* Urbana, IL: University of Illinois Press.

Zirn, Cäcilia, Robert Meusel, and Heiner Stuckenschmidt. 2015. “Lost in Discussion? — Tracking Opinion Groups in Complex Political Discussions by the Example of the FOMC Meeting Transcriptions.” In *Proceedings of Recent Advances in Natural Language Processing,* edited by Galia Angelova, Kalina Bontcheva, and Ruslan Mitkov, 747–753.

A Appendix

A.1 Chronology of Major Events in China

In this appendix, we compile a chronology of major events in China from 1949 (when the regime was established) to 2018. The chronology, Table A.1, is based on our survey of studies on the Chinese economy and reforms that are widely cited in the literature: Coase and Wang (2012), Howe, Kueh, and Ash (2002), Lardy (2014), Naughton (2014, 2018), Saich (2015), J. Wu (2014), and S. Zhao (2016). The chronology serves as the ground truth against which the PCI's predictive power is validated in Section 5.

A.2 Robustness Check with Ten-Year Window

We have used a five-year rolling window as the baseline to train the algorithm. In this appendix, as a robustness check, we explore the alternative of using a ten-year rolling window.

As discussed in the main text, having a longer training window is not necessarily better, regardless of computation cost. However, since the mechanism of predicting policy change remains the same whether we use a five- or ten-year window, we expect that the resulting PCI with a ten-year window will be similar to our main finding.

Figures A.1 and A.2 confirm our conjecture. The classification performance in testing for both the five- and ten-year windows are plotted in Figures A.1, in which the solid orange curve is the same as that in Figure 6. The dashed black curve (i.e., the ten-year window) has a similar pattern, but, due to the fact that it covers data farther in the past then the five-year window, lags behind the solid orange curve by several years.

Despite the differences in classification performance, the resulting PCIs — the index of ultimate interest — are fairly similar. In Figure A.2, the PCI with a ten-year window is represented by the dashed red curve, against the baseline PCI already shown in Figure 1. While differences exist, the pattern of spikes near important policy changes is largely the same in both cases.

Table A.1: Chronology of major events in China, 1949-2018

Date	**Event**	**Note**
10/1/49	Regime established.	Economic recovery starts.
1/1/53	First Five-Year Plan announced.	• Soviet-style institution put in place; investment surges (esp. in heavy and defense-oriented industries. • “High Tide of Socialism” in 1955-56, an abrupt transformation to public ownership.
5/16/58	Great Leap Forward starts.	• Massive transformation of resources from agriculture to industry, with a rapid expansion of steal production. • “People’s Communes” in rural areas to mobilize labor for construction projects.
5/16/66	Cultural Revolution starts.	• Originated as a political campaign to overthrow Mao’s opponents. • Investment and industrial construction surge; consumption restrained. • Brief retrenchment in 1972, after Lin Biao’s death and Nixon’s visit to China.
9/9/76	Hua Guofeng takes over.	• Mao’s death, followed by the purge of the Gang of Four and the resumption of massive investment push under Hua.

Continued on next page

Table A.1 – *Continued from previous page*

Date	Event	Note
12/18/78	Economic reform program starts under Deng Xiaoping.	• Wide-ranging reassessment of the command economy; rural reforms, to reduce the burden on farmers; decentralization of powers in industry; markets allowed; trade surges. • Reforms continues to intensify after the roll-out of "Socialist Planned Commodity Economy" in October 1984.
6/4/89	Tiananmen Square protests	• Protests followed by a conservative ascendancy and campaigns against market-oriented reform.
11/11/93	Reform speed-up under Jiang Zemin	• Prelude: Deng Xiaoping's southern tour in early 1992; the Party's endorsement of the "Socialist Market Economy" slogan in October 1992. • Broad reform program kicks off; restrictive macroeconomic policy; price stabilization; fiscal reform; market unification; further economic openness; SOE restructuring. • Another round of reform deepening in September 1997: overhaul of banking system, housing reform, health care reform, restructuring government bureaucracy, SOE privatization.

Continued on next page

Table A.1 – *Continued from previous page*

Date	**Event**	**Note**
10/8/05	Reform slow-down under Hu Jintao	• Prelude: Hu raised the slogan of "Socialist Harmonious Society" for the first time in September 2004. • Enterprise reforms slow down; support for large SOEs emphasized again. • Cushions the impact of reforms on the "losers" — reducing income inequality, improving access to health care and education, extending social security, moderating environment impact of growth.
11/5/08	Stimulus package	• Four trillion yuan economic stimulus package as a policy response to the global financial crisis.
4/22/13	Initiative to revive Maoism under Xi	• Orders officials to fight the spread of subversive currents, such as Western constitutional democracy, universal values of human rights, Western-inspired notions of media and civil society independence, ardently market-friendly neo-liberalism, and "nihilistic" critiques of the Party's traumatic past.

Continued on next page

Table A.1 – *Continued from previous page*

Date	**Event**	**Note**
11/9/13	Initiative to renew reform under Xi	• Calls for re-commitment to market reforms while retaining emphasis on support for the state sector; Belt and Road Initiative. • Many of the initiatives stall out in less than two years.
11/10/15	"Supply-side structural reform" under Xi	• Calls for reducing excess industrial capacity, esp. steel and coal. • Relies on direct government action, not a market-oriented reform.

Figure A.1: Classification performance in testing: five- versus ten-year windows

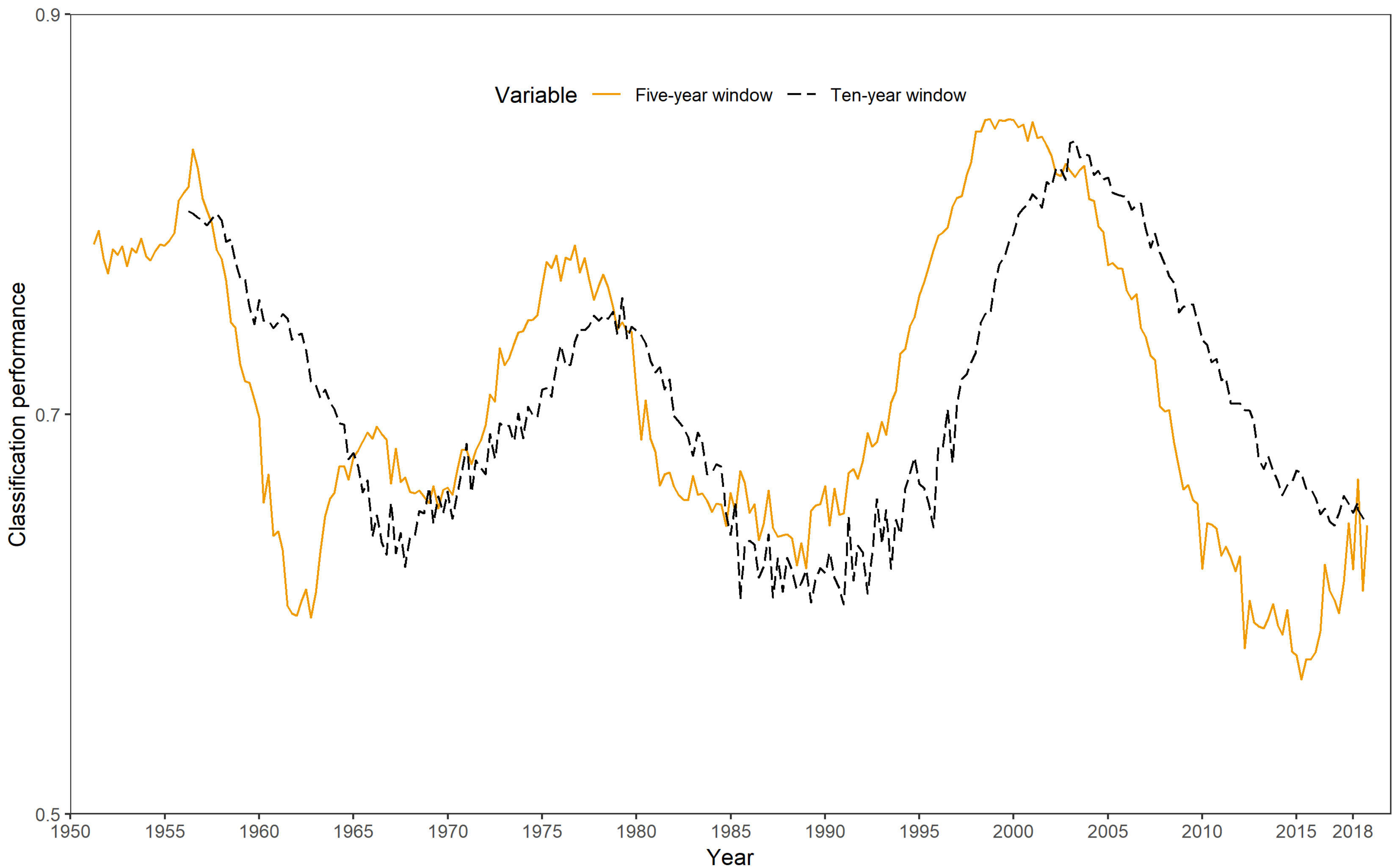


Figure A.2: Policy Change Index: five- versus ten-year windows