

# Reading China:

## Predicting Policy Change with Machine Learning

Julian TszKin Chan  
(Bates White)

Weifeng Zhong  
(American Enterprise Institute)

March 27, 2019  
Analytics Frontiers

*The views expressed here are those of the authors and do not represent the views of AEI, Bates White, or their other employees.*

# Predicting policy change: why?



# Predicting policy change: why?

- China's industrialization: a product of gov't direction.
- Opaque system makes prediction prohibitively difficult.

# Predicting policy change: why?

- China's industrialization: a product of gov't direction.
- Opaque system makes prediction prohibitively difficult...

... until now.

- We construct the first predictive algorithm for China's policy shifts using machine learning:

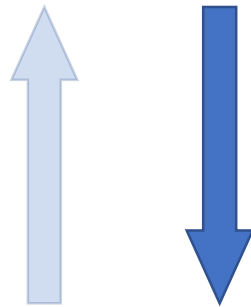
**The Policy Change Index (PCI) for China**

# Value of economics in data science

- Economics: explaining the past, lots of theories.
- Data science: predicting the future, “theory-free.”

# Value of economics in data science

- Economics: explaining the past, lots of theories.



- Data science: predicting the future, “theory-free.”

**1. Design of the PCI**

2. Results

3. Other applications

# Predicting policy change: how?

Build a machine learning algorithm to

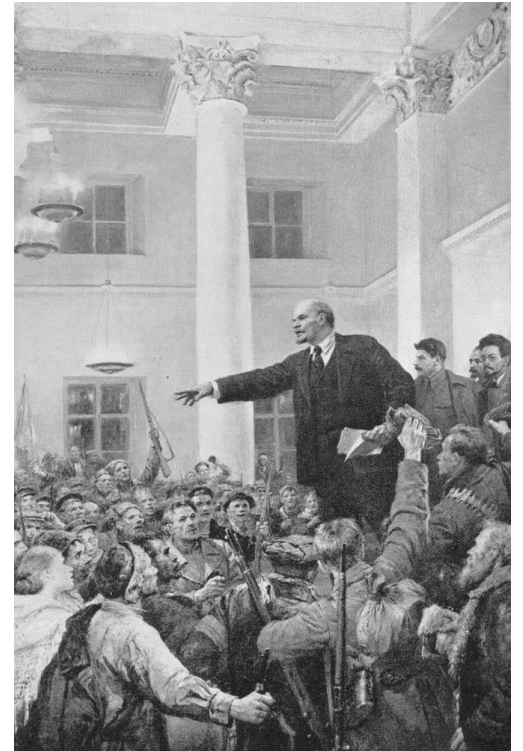
- “read” the *People's Daily*;
- detect structural changes in its priorities.



Official newspaper, 1946-present



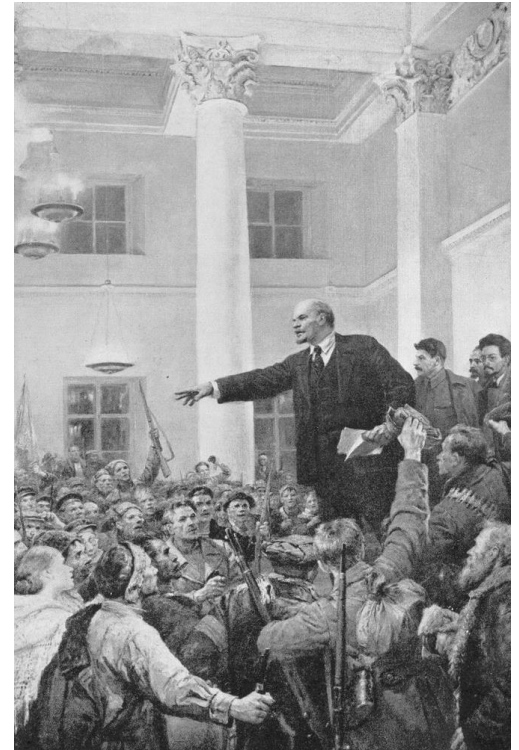
# Source of predictive power



# Source of predictive power

## The Leninist tradition:

- “[T]he whole task of the Communists is to be able to **convince** the backward elements.” (1920)
- It is a fundamental necessity “to transform the press... into a serious organ for the economic **education** of the mass of the population.” (in early USSR years)



# Source of predictive power

The Leninist tradition, in Xi's own words:

- The media should “uphold the **correct orientation** of public opinion in all aspects and at all stages.” (2016)



# Source of predictive power

*People's Daily:*  
nerve center of China's propaganda system

+

Propaganda often precedes policies.

↓

Detect changes in  
newspaper's priorities

≈

Predict changes in  
gov't policies

# Source of predictive power

*People's Daily:*  
nerve center of China's propaganda system

+

Propaganda often precedes policies.

↓

Detect changes in  
newspaper's priorities

≈

Predict changes in  
gov't policies

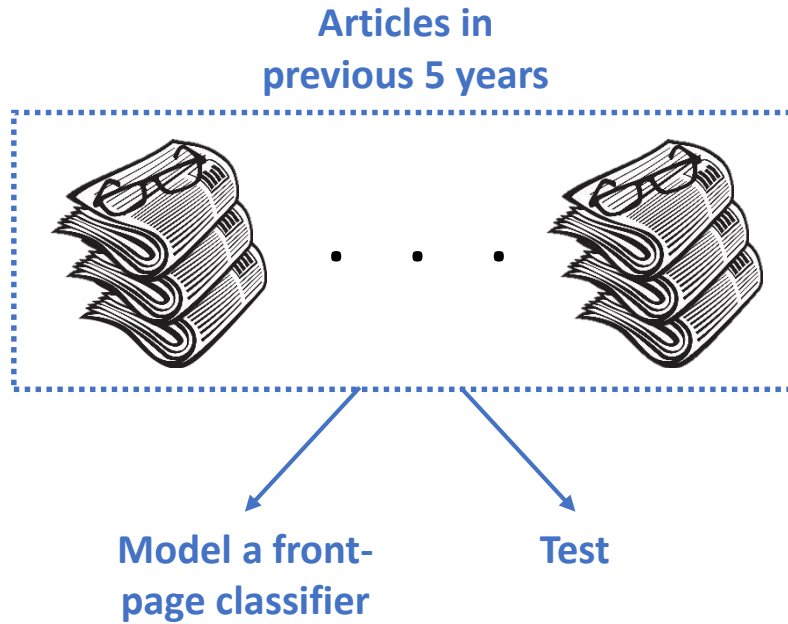
Machine  
learning



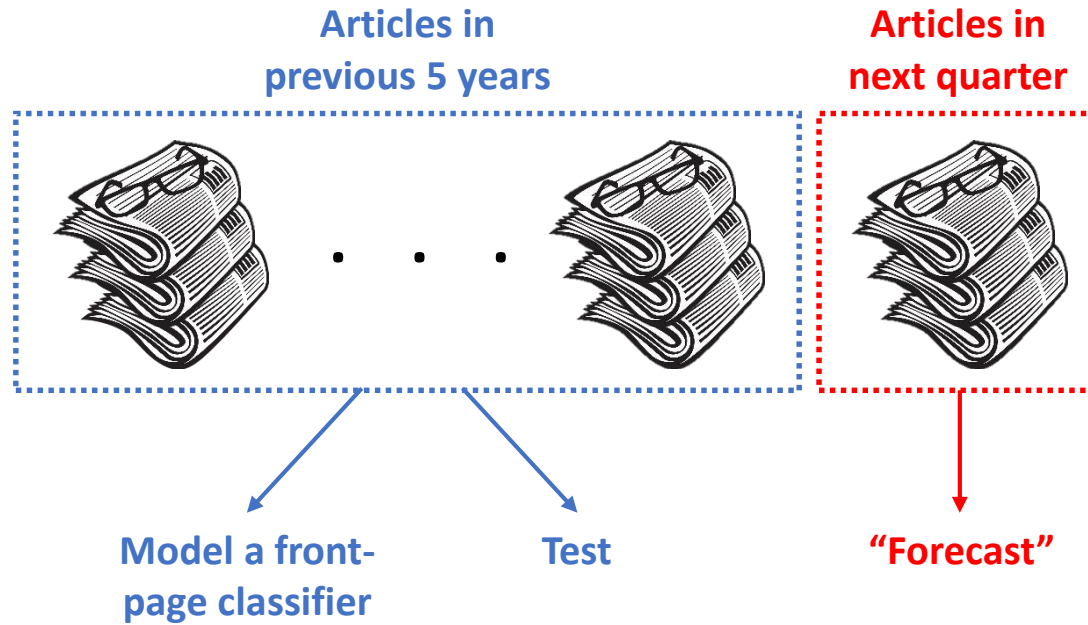
# Method

An algorithm that classifies **front-page** articles  
by “reverse-engineering” the editor’s mind.

# Method

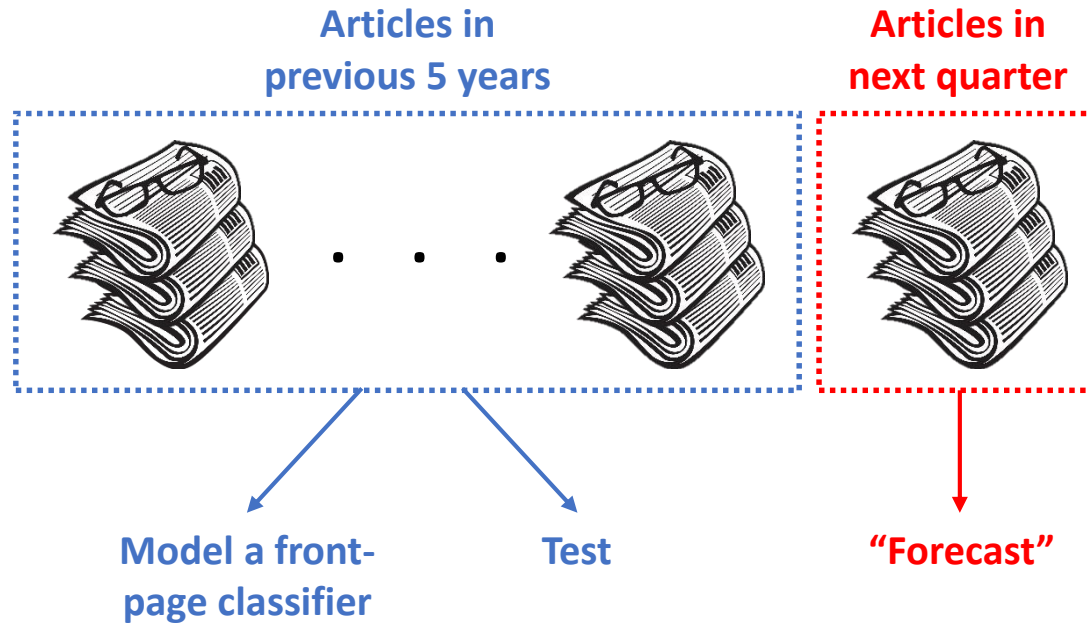


# Method





# Method



$$\text{Policy Change Index} = \left| \text{Test performance} - \text{"Forecast" performance} \right|$$

**PCI**  $\gg 0 \Rightarrow$  Change in the newspaper's priorities

# Method: data

sample\_df

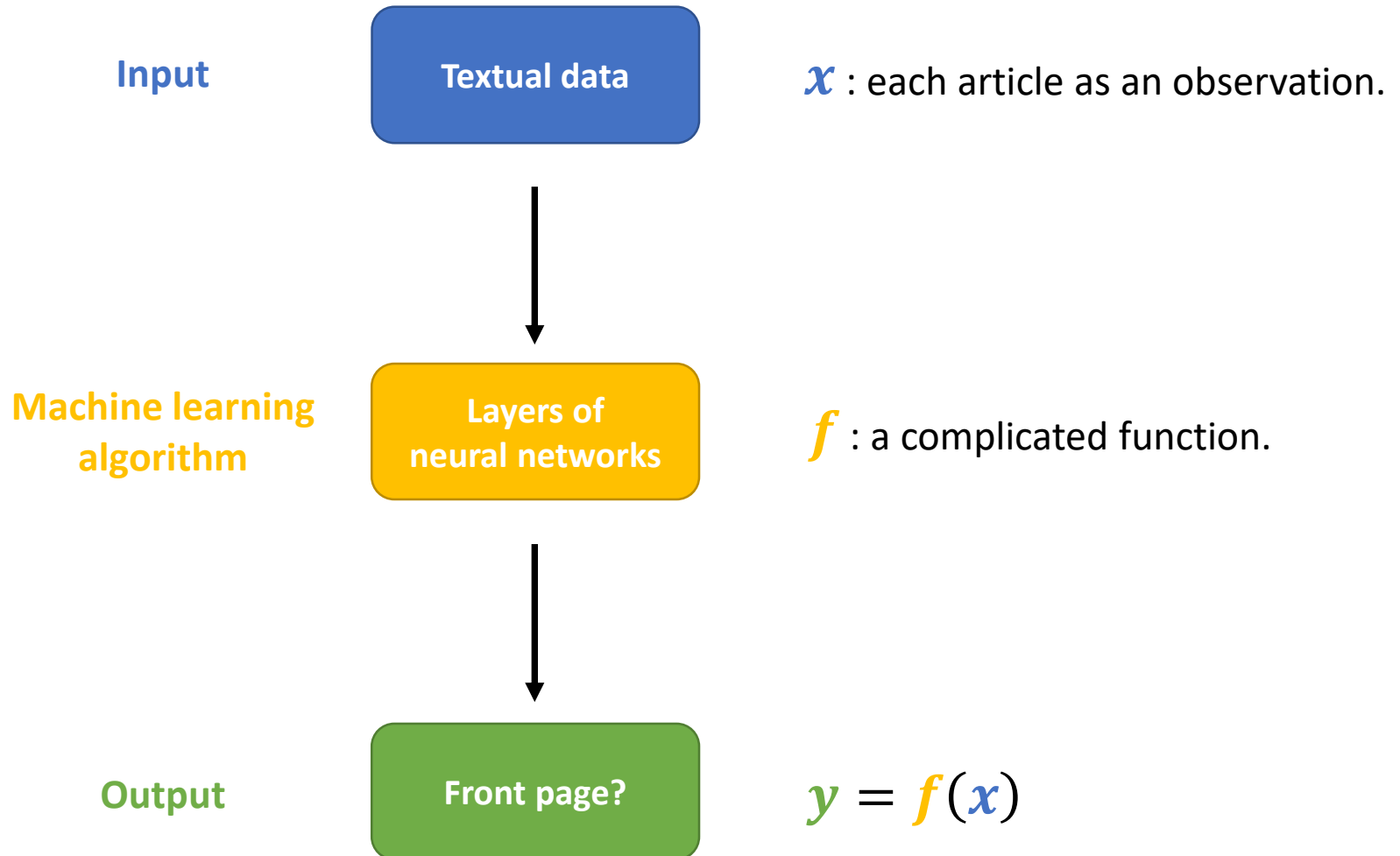
id	date	page	title	body
20181202525	2018-12-31 00:00:00	1.0	今年粮食总产量13158亿斤 农业农村发展取得新成绩，乡村振兴开局良好	本报北京12月30日电 （记者高云才、郁静娴）记者从30日召开的全国农业农村厅局长会议获...
20181202526	2018-12-31 00:00:00	1.0	新时代改革再出发的重要里程碑 ——写在党的十八届三中全会召开五周年之际	本报评论员 改革开放40年来，我们党带领人民绘制了一幅波澜壮阔、气势恢宏的历史画卷。其...
20181202527	2018-12-31 00:00:00	1.0	改革引领中国经济新航向（在习近平新时代中国特色社会主义思想指引下——新时代新作为新篇章） ...	本报记者 许志峰 吴秋余 王珂 林丽鹏 巨轮向前，离不开航标指引方向。 1978年...
20181202528	2018-12-31 00:00:00	1.0	产量连续六年超1200亿斤，实现“十五连丰” 黑龙江 粮食再丰收，粮仓更稳固	本报哈尔滨12月30日电 （记者方圆）此刻的黑土地，正在冬眠蓄力。今年黑龙江粮食又是一...
20181202529	2018-12-31 00:00:00	1.0	伟大的变革——庆祝改革开放40周年大型展览 现场观众达223万人次 展览将持续到2019年...	据新华社北京12月30日电 “伟大的变革——庆祝改革开放40周年大型展览”自11月13日...
20181202530	2018-12-31 00:00:00	1.0	航线不停、景区不休、生意不闲、游客不断 新疆 旅游无淡季，今冬更红火	本报乌鲁木齐12月30日电 （记者杨明方、阿尔达克）一年之中半年是降雪期，冬季雪深超过3...
20181202531	2018-12-31 00:00:00	1.0	图片报道	徜徉知识海洋，感受阅读乐趣。十二月三十日是元旦小长假的第一天，全国各地的特色文化活动丰富...
20181202532	2018-12-31 00:00:00	2.0	创造新的更大奇迹 ——党的十八届三中全会召开五周年述评	“全面发力、多点突破、蹄疾步稳、纵深推进”——庆祝改革开放40周年大会上，习近平总书记用...
20181202534	2018-12-31 00:00:00	3.0	开启中国特色大国外交新征程（2018年度国际特别报道） ——国务委员兼外交部长王毅回顾20...	本报记者 赵嘉鸣 管克江 杜尚泽 焦翔 王海林 国务委员兼外交部长王毅12月29日接...
20181202535	2018-12-31 00:00:00	3.0	二〇一七年岁末，莫斯科大学的中国留学生们欣喜地收到了习近平主席的回信；一年来，牢记习主席的...	本报驻俄罗斯记者 吴焰 “一年前，习近平主席给留学生回信，鼓励我们胸怀大志，刻苦学习...
20181202536	2018-12-31 00:00:00	3.0	外交部发言人就中美建交40周年发表谈话	本报北京12月30日电 外交部发言人陆慷30日就中美建交40周年发表谈话。 陆慷说，2...

# Method: data

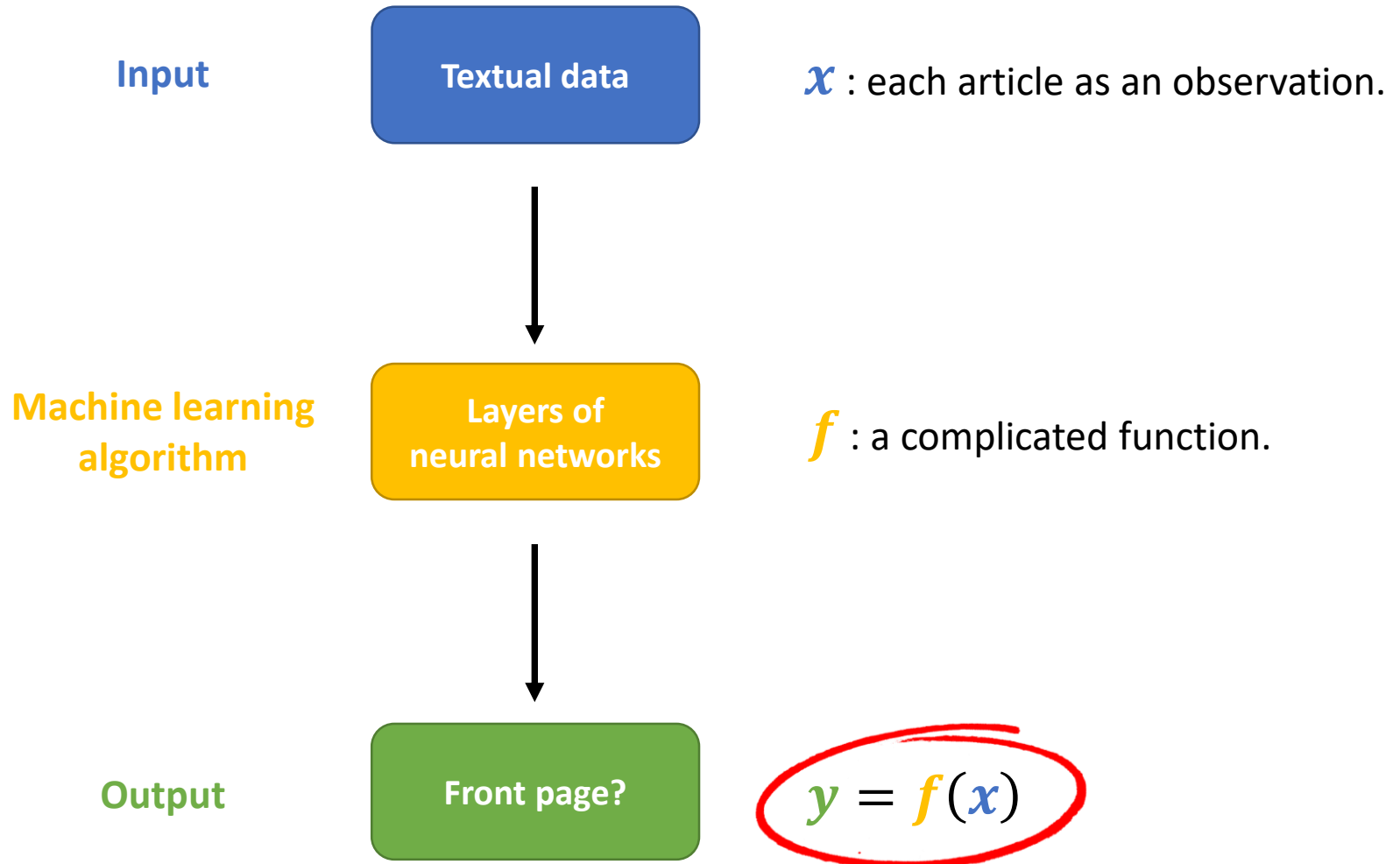
sample\_df

id	date	page	title	body
20181202525	2018-12-31 00:00:00	1.0	今年粮食总产量13158亿斤 农业农村发展取得新成绩，乡村振兴开局良好	本报北京12月30日电 （记者高云才、郁静娴）记者从30日召开的全国农业农村厅局长会议获...
20181202526	2018-12-31 00:00:00	1.0	新时代改革再出发的重要里程碑 ——写在党的十八届三中全会召开五周年之际	本报评论员 改革开放40年来，我们党带领人民绘制了一幅波澜壮阔、气势恢宏的历史画卷。其...
20181202527	2018-12-31 00:00:00	1.0	改革引领中国经济新航向（在习近平新时代中国特色社会主义思想指引下——新时代新作为新篇章） ...	本报记者 许志峰 吴秋余 王珂 林丽鹏 巨轮向前，离不开航标指引方向。 1978年...
20181202528	2018-12-31 00:00:00	1.0	产量连续六年超1200亿斤，实现“十五连丰” 黑龙江 粮食再丰收，粮仓更稳固	本报哈尔滨12月30日电 （记者方圆）此刻的黑土地，正在冬眠蓄力。今年黑龙江粮食又是一...
20181202529	2018-12-31 00:00:00	1.0	伟大的变革——庆祝改革开放40周年大型展览 现场观众达223万人次 展览将持续到2019年...	据新华社北京12月30日电 “伟大的变革——庆祝改革开放40周年大型展览”自11月13日...
20181202530	2018-12-31 00:00:00	1.0	航线不停、景区不休、生意不闲、游客不断 新疆 旅游无淡季，今冬更红火	本报乌鲁木齐12月30日电 （记者杨明方、阿尔达克）一年之中半年是降雪期，冬季雪深超过3...
20181202531	2018-12-31 00:00:00	1.0	图片报道	徜徉知识海洋，感受阅读乐趣。十二月三十日是元旦小长假的第一天，全国各地的特色文化活动丰富...
20181202532	2018-12-31 00:00:00	2.0	创造新的更大奇迹 ——党的十八届三中全会召开五周年述评	“全面发力、多点突破、蹄疾步稳、纵深推进”——庆祝改革开放40周年大会上，习近平总书记用...
20181202534	2018-12-31 00:00:00	3.0	开启中国特色大国外交新征程（2018年度国际特别报道） ——国务委员兼外交部长王毅回顾20...	本报记者 赵嘉鸣 管克江 杜尚泽 焦翔 王海林 国务委员兼外交部长王毅12月29日接...
20181202535	2018-12-31 00:00:00	3.0	二〇一七年岁末，莫斯科大学的中国留学生们欣喜地收到了习近平主席的回信；一年来，牢记习主席的...	本报驻俄罗斯记者 吴焰 “一年前，习近平主席给留学生回信，鼓励我们胸怀大志，刻苦学习...
20181202536	2018-12-31 00:00:00	3.0	外交部发言人就中美建交40周年发表谈话	本报北京12月30日电 外交部发言人陆慷30日就中美建交40周年发表谈话。 陆慷说，2...

# Method: modeling



# Method: modeling

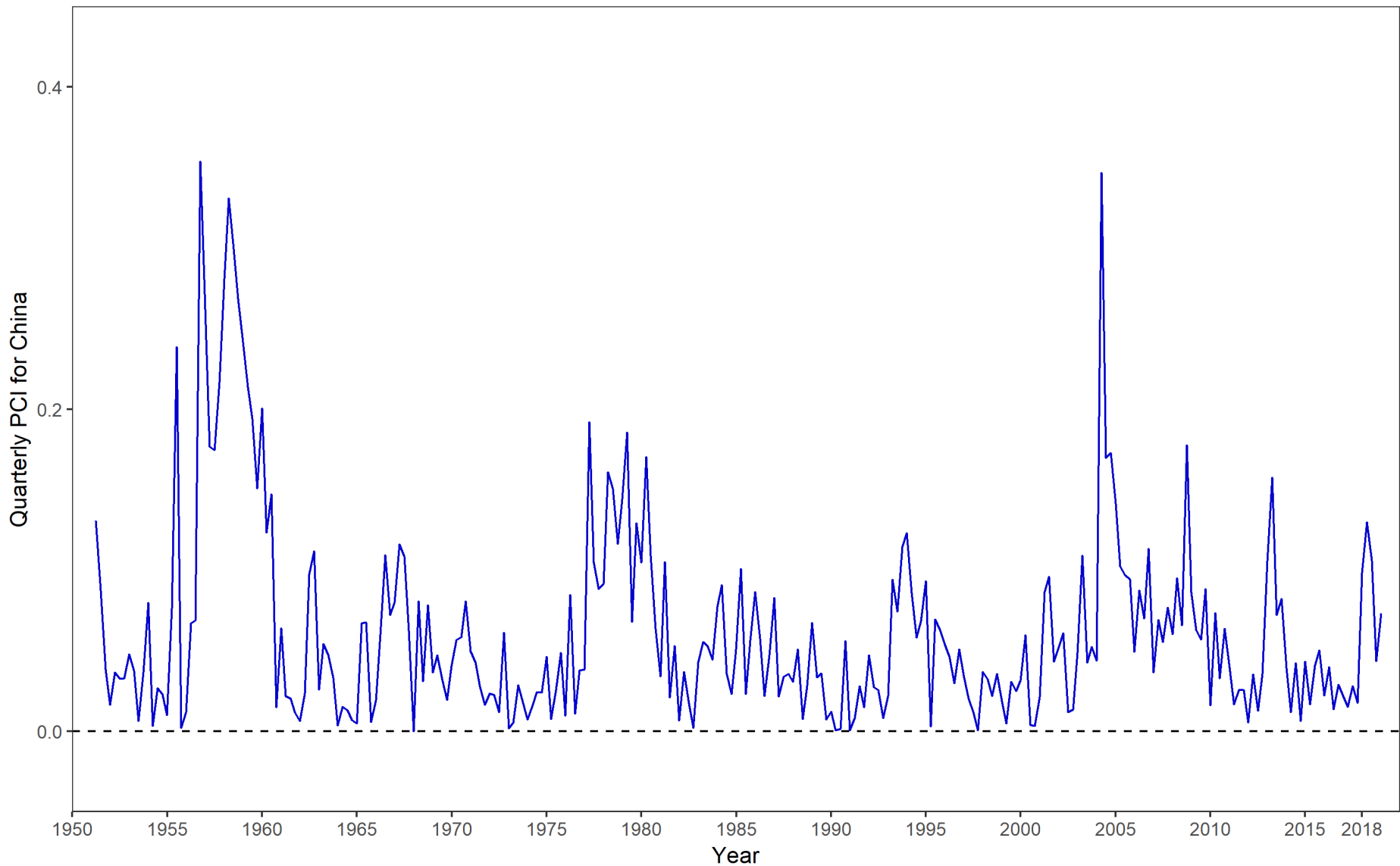


1. Design of the PCI

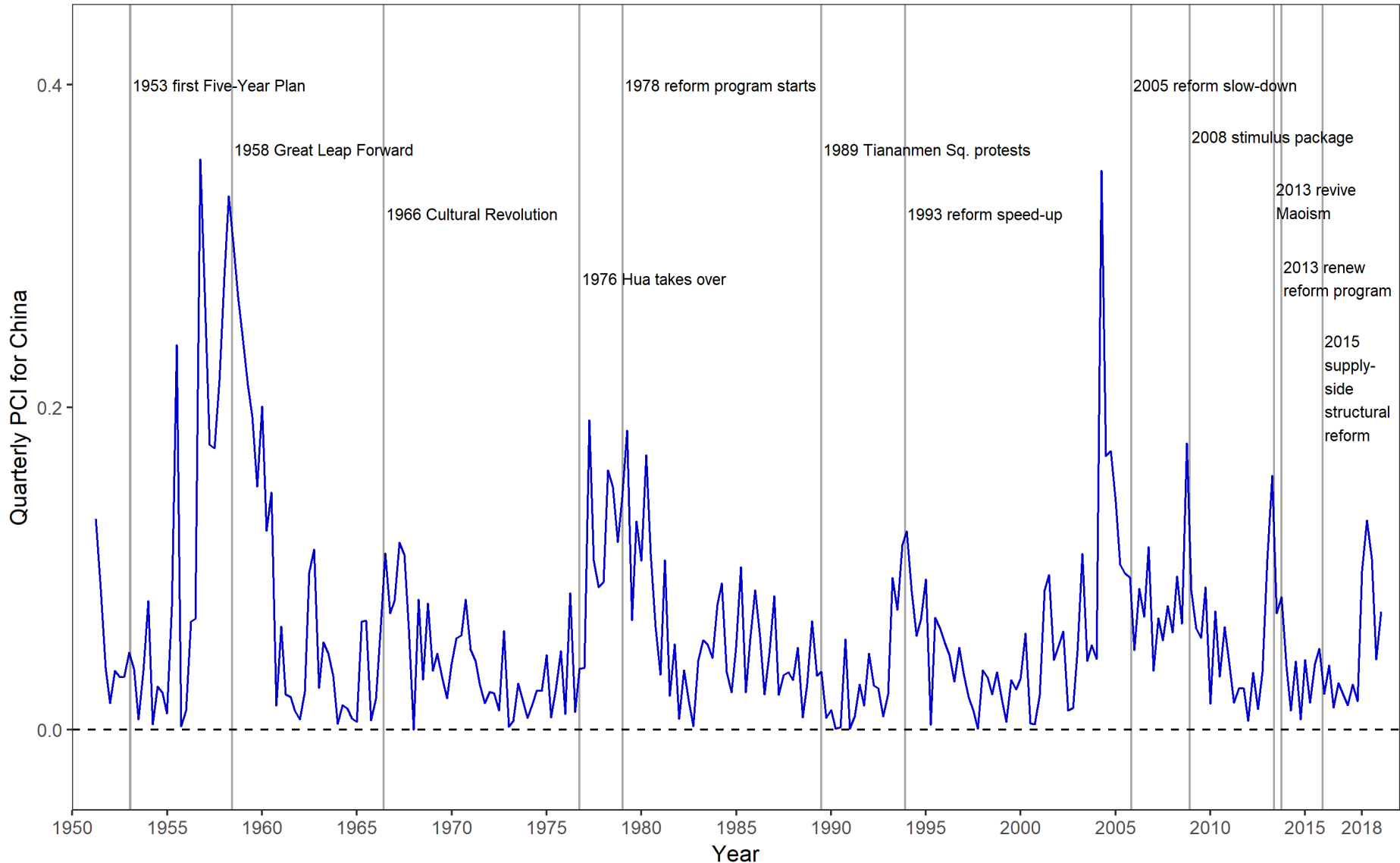
**2. Results**

3. Other applications

# PCI-China

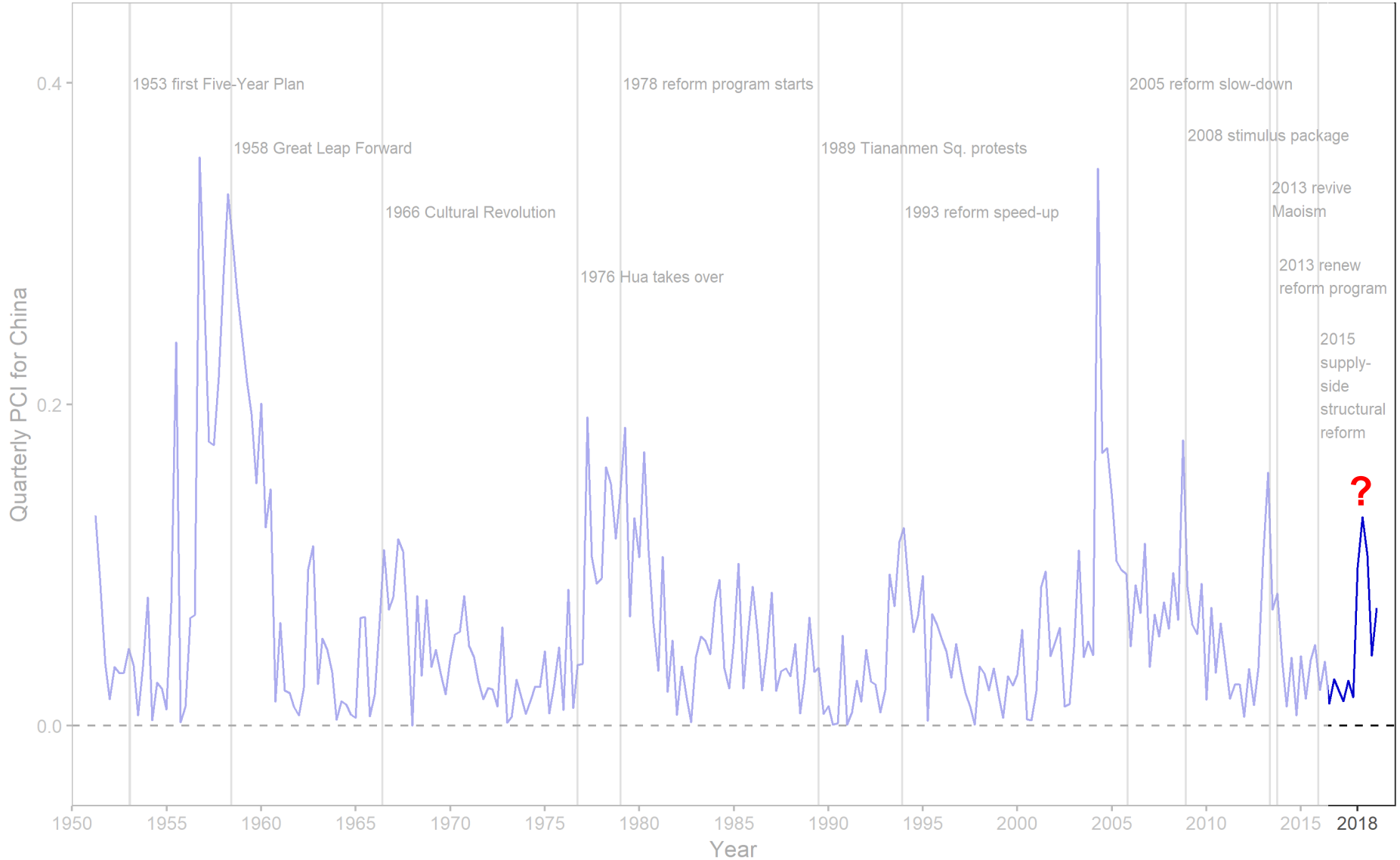


# PCI-China — with ground truth





# PCI-China — going forward



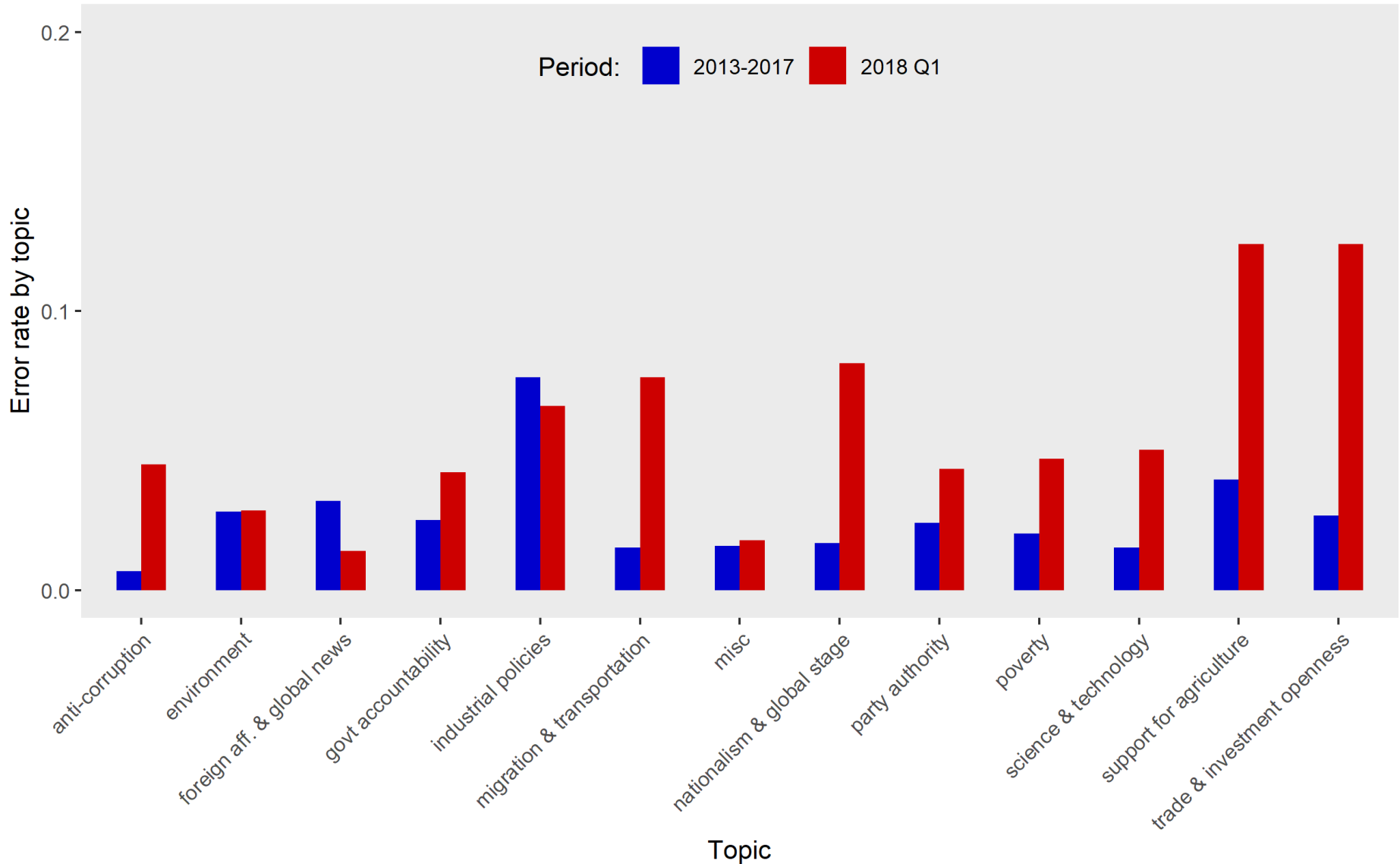
# Understanding substance of change

		Classified on front page?	
		No	Yes
Front page?	No	√	false positives
	Yes	false negatives	√

- Content of *mis*-classified articles has policy substance.

# The 2018 Q1 uptick

# The 2018 Q1 uptick



# The 2018 Q1 uptick

The 2018 Q1 uptick represents:

- internally: strengthening party authority;
- externally: nationalism and global leadership;
- populist policies to boost political support.

# The 2018 Q1 uptick

The 2018 Q1 uptick represents:

- internally: strengthening party authority;
- externally: nationalism and global leadership;
- populist policies to boost political support.

⇒ Curb your enthusiasm for the trade talks!

1. Design of the PCI

2. Results

**3. Other applications**

# Generally

PCIs for other (ex-)Communist regimes, using:

- Soviet Union's *Pravda*
- East Germany's *Neues Deutschland*
- Cuba's *Granma*
- North Korea's *Rodong Sinmun*
- Vietnam's *Nhân Dân*



# Generally

PCIs for other (ex-)Communist regimes, using:

- Soviet Union's *Pravda*
- East Germany's *Neues Deutschland*
- Cuba's *Granma*
- North Korea's *Rodong Sinmun*
- Vietnam's *Nhân Dân*

} Work in progress

# More generally

PCIs for other countries:

- The state may not control the media.
- But politicians/policymakers still talk!

# Even more generally

Applicability well beyond predicting policies.

Omitted here. See our [research paper](#).

# Interested in DIY?

- Website: [policychangeindex.com](http://policychangeindex.com) (newsletter sign-up)
- Paper: [policychangeindex.com/pdf/Reading\\_China.pdf](http://policychangeindex.com/pdf/Reading_China.pdf)
- Source code: [github.com/PSLmodels/PCI](https://github.com/PSLmodels/PCI)
- A [simulated example](#) to show how the PCI works.

**Questions?**