

# Predicting Policy Change with Text Data

Julian TszKin Chan (*Bates White Economic Consulting*)

Weifeng Zhong (*Mercatus Center at George Mason University*)

O'Reilly AI Conference (10/17/2019)

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

# Our use case: China

The Policy Change Index (PCI): a machine learning model that predicts **policy changes** by “reading” this:

习近平：新时代要有新气象 更要有新作为

中国青年报

人民日报  
RENMIN RIBAO

习近平应约同美国总统特朗普通电话  
特电祝贺中共十九大胜利召开，祝贺习近平再次当选中共中央总书记



习近平同志主持会议并作重要讲话  
习近平李克强栗战书汪洋王沪宁赵乐际韩正任中央政治局常委  
党的十九届一中全会产生中央领导机构



会议并作重要讲话  
中央政治局常委  
中央军委主席



南方周末

证券时报

中国共产党第十九

# How can it possibly work?!

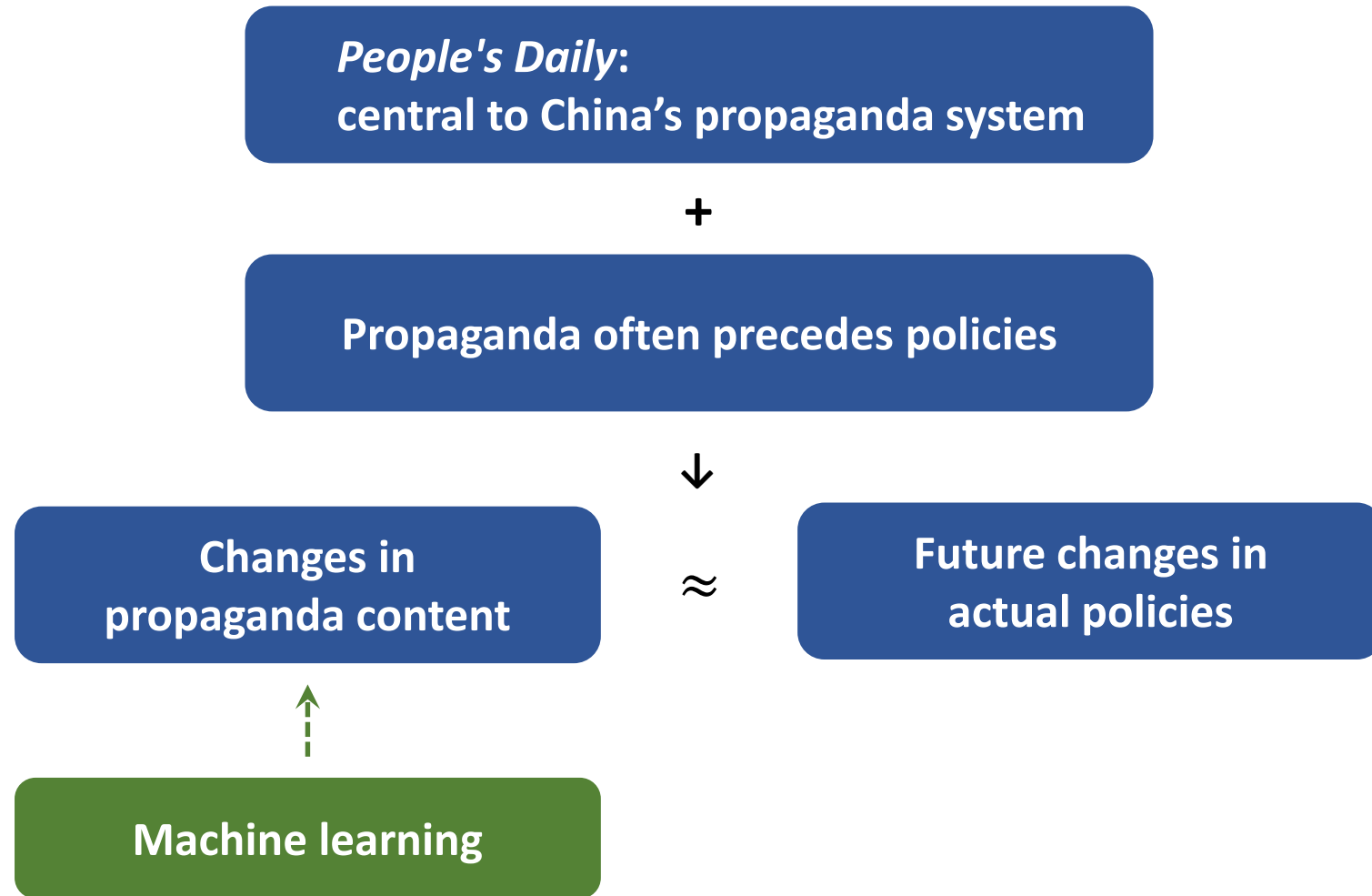
Because propaganda often precedes policies.

# How can it possibly work?!

Because propaganda often precedes policies.



# The “value” of propaganda



# Challenge: lack of labels

- Propaganda is abstract and subjective.
- We narrow it down to editorial priorities.
- Still, no labels—editorial priorities are in the editor's mind!

# What's in the editor's mind??

Transform estimation problems to anomaly detection problems.



# What's in the editor's mind??

Transform estimation problems to anomaly detection problems.



# Estimation: *infeasible*

- Newspaper text comes with page number.
- The infeasible estimation problem:

$$g : \{(text, FrontPage)\} \rightarrow \{\text{"editorial priorities"}\}$$

- $g(\text{"pvt sector is great", on front page}) = (\text{reform is high-priority})$ ;
- $g(\text{"central planning is great", on front page}) = (\text{reform is low-priority})$ .
- Only if  $g$  can be learned...

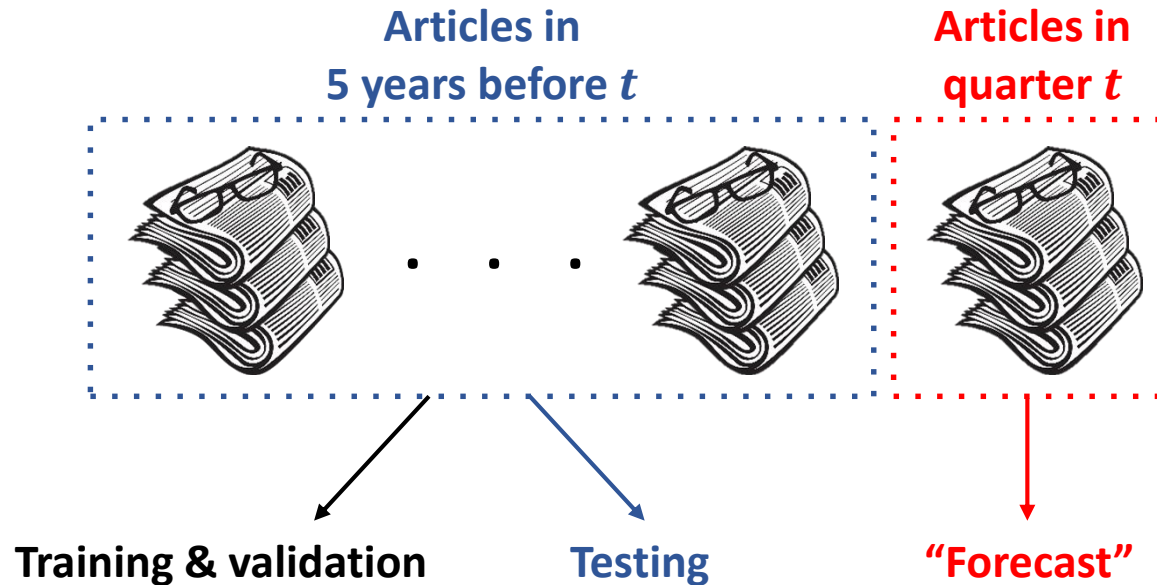
# Anomaly detection: feasible

- Treat “editorial priorities” as a latent variable:

$$f_{\{\text{"editorial priorities"}\}} : \{\textit{text}\} \rightarrow \{\textit{FrontPage}\}$$

- Lots of labels for  $f$ .
- Detecting changes in  $f$ : a feasible anomaly detection problem.
- Change in  $f \Rightarrow$  change in propaganda priorities  $\Rightarrow$  change in policy!

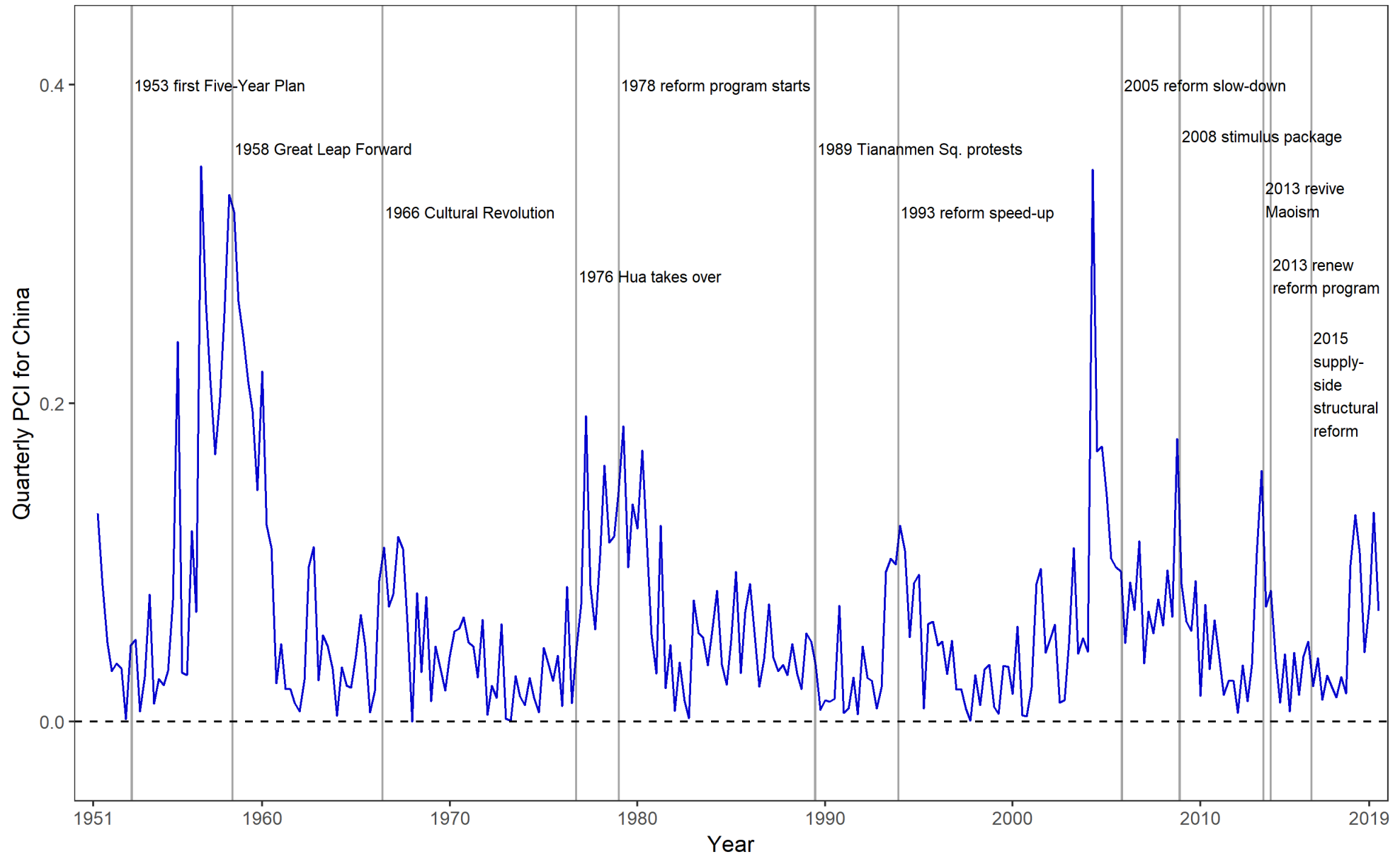
# Policy Change Index for China (PCI-China)



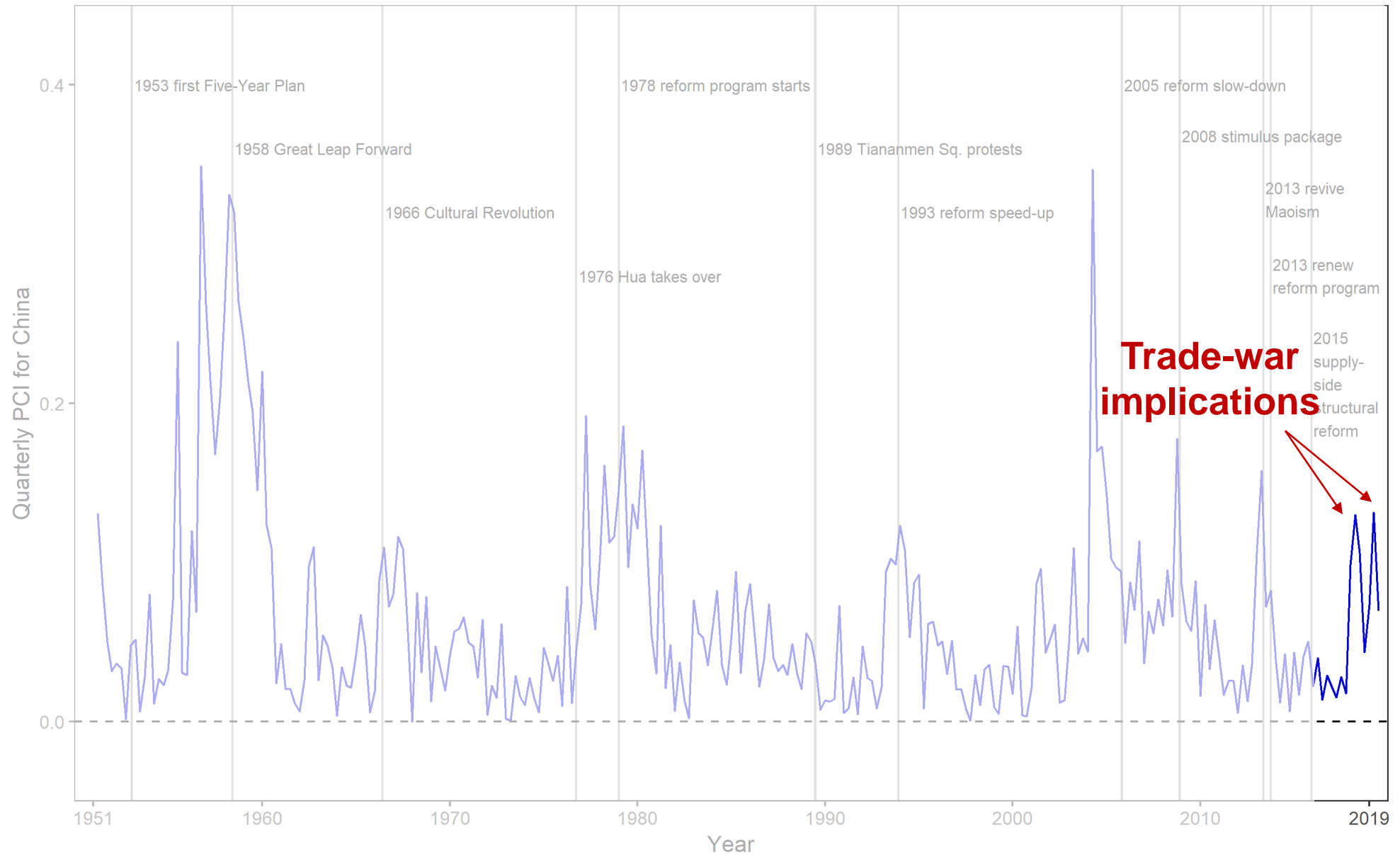
$$\text{PCI-China in quarter } t = \left| \text{Test performance} - \text{“Forecast” performance} \right|$$

Large **PCI-China**  $\Rightarrow$  Structural change in editorial priorities

# PCI-China: result v ground truth



# PCI-China: result v ground truth



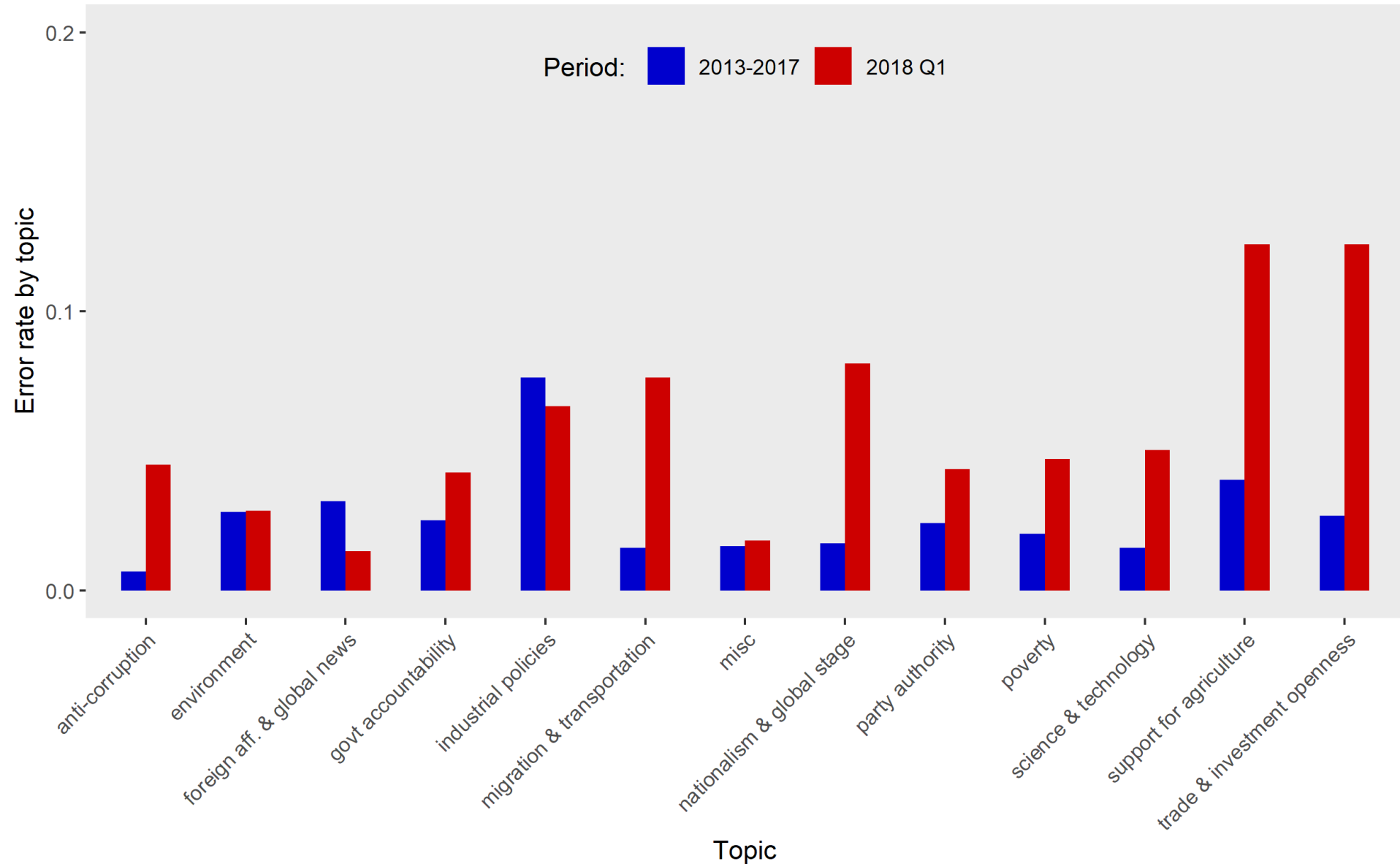
# Understanding changes

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

Content of *misclassified* articles has policy substance:

- False negatives suggest new policies;
- False positives suggest phasing-out policies.

# Understanding changes: 2018 Q1 uptick





# Trade-war implications

Two upticks in 2018 Q1 and 2019 Q1:

- Internally: strengthening party authority;
- Externally: nationalism and global leadership.

A trade deal with structural reforms? Curb your enthusiasm...

How can these be useful to you?

# Remember Trick #2?

When there's no label for the target variable, transform the estimation problem to an anomaly detection problem.

**Many other use cases share the same structure!**

# Future PCI projects

Stay tuned for more in the PCI pipeline:

- PCIs for other countries: North Korea, Russia, Cuba,...
- PCI sub-indices, by subject: trade policy, fiscal policy,...
- And more!

# Back to Trick #1

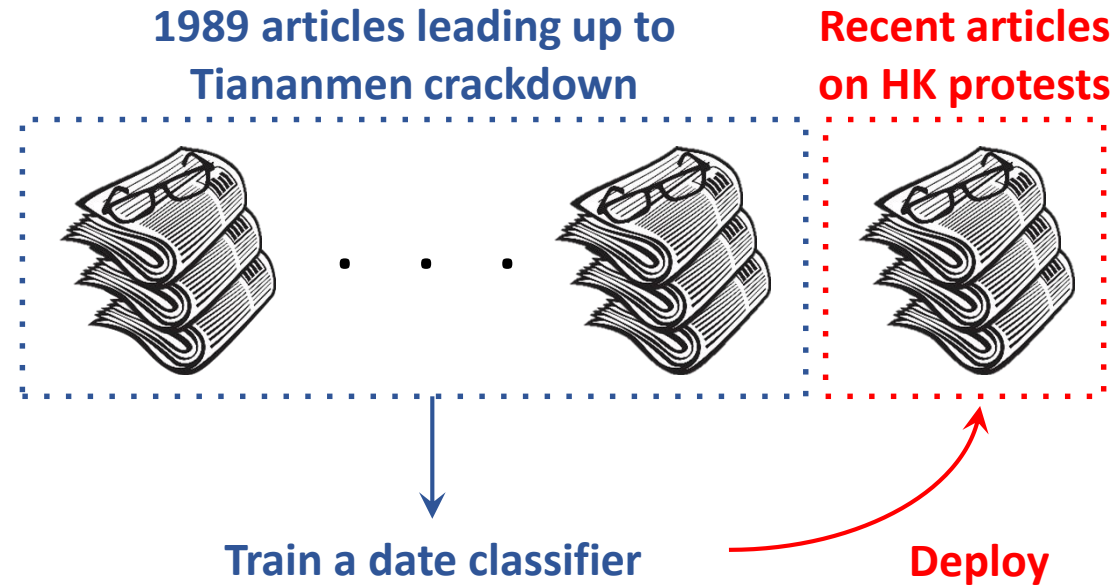
Propaganda often precedes policies.

**More generally, words often foreshadow actions.**

# Consider the Hong Kong protests

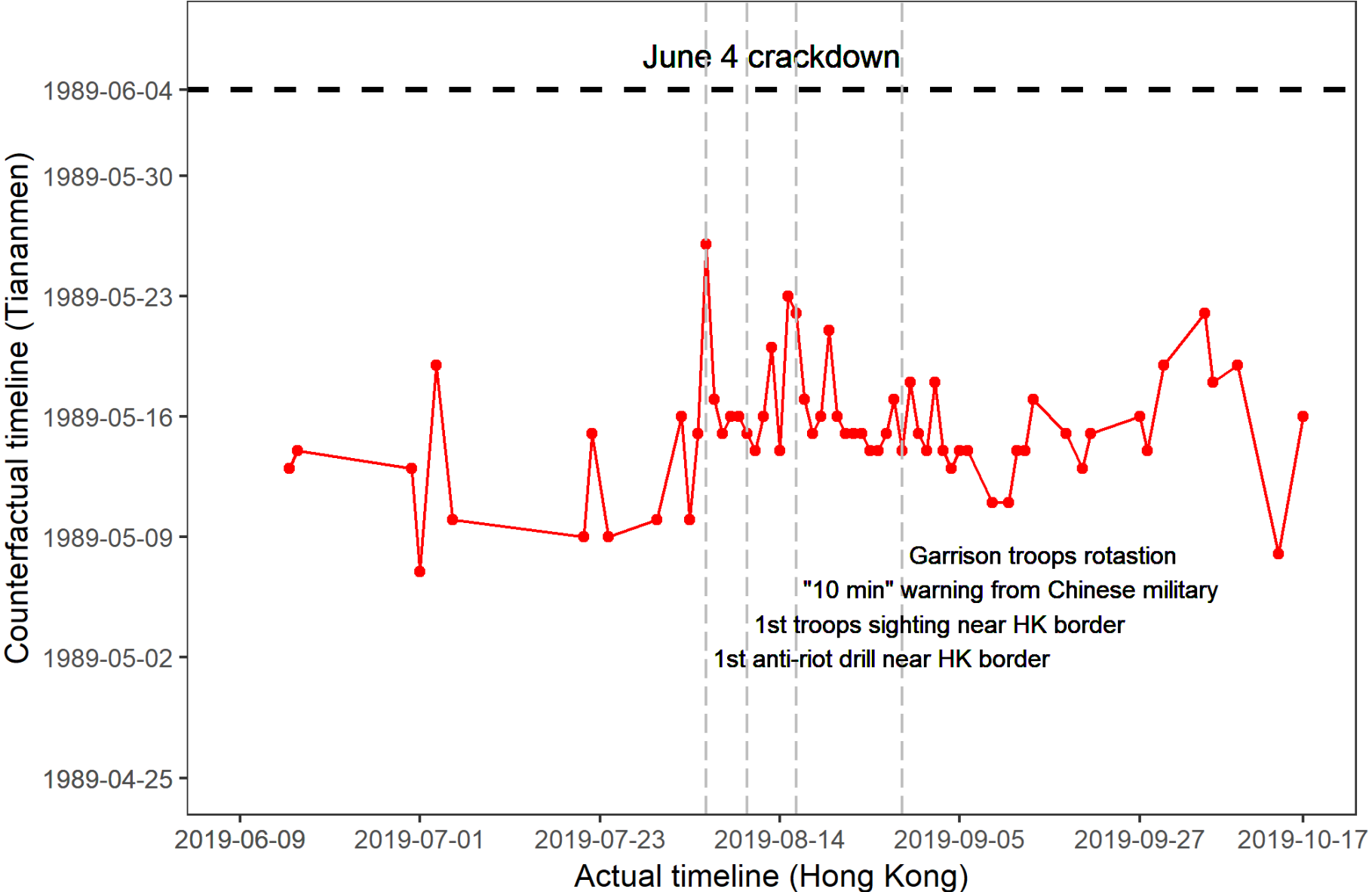
- Will the protests be met with a military crackdown by China?
  - ↳ How hard is the *People's Daily* slamming Hong Kong?
- There's a unique, well-formed precedent: Tiananmen crackdown.

# Policy Change Index for Crackdown (PCI-Crackdown)



**PCI-Crackdown:** mapping each current date to an “as-if” date

# PCI-Crackdown: Hong Kong protests





# Takeaways

1. **Words often foreshadow actions.**
2. **No labels? Try transforming estimation problems to anomaly detection problems.**

Website: [policychangeindex.org](http://policychangeindex.org)

(papers, code, newsletter,...)