# Reading China:

## Predicting Policy Change with Machine Learning

Julian TszKin Chan                    Weifeng Zhong

(Bates White)                         (AEI)

November 10, 2018

West Lafayette, IN

# Policy Change Index (PCI) for China

China's industrialization has been a product of gov't direction.

PCI: the first *leading* indicator of China's policy priorities.

- Covers 1951 Q1 – 2018 Q3.
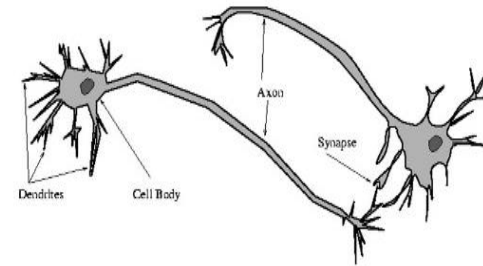
- Can be updated in the future.

# How to predict policy changes?

Build a machine learning algorithm to

- "read" the *People's Daily*;

- detect changes in how it prioritizes policy issues.

Official newspaper, 1946-2018                    (Artificial) neural networks

# 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

**≈**

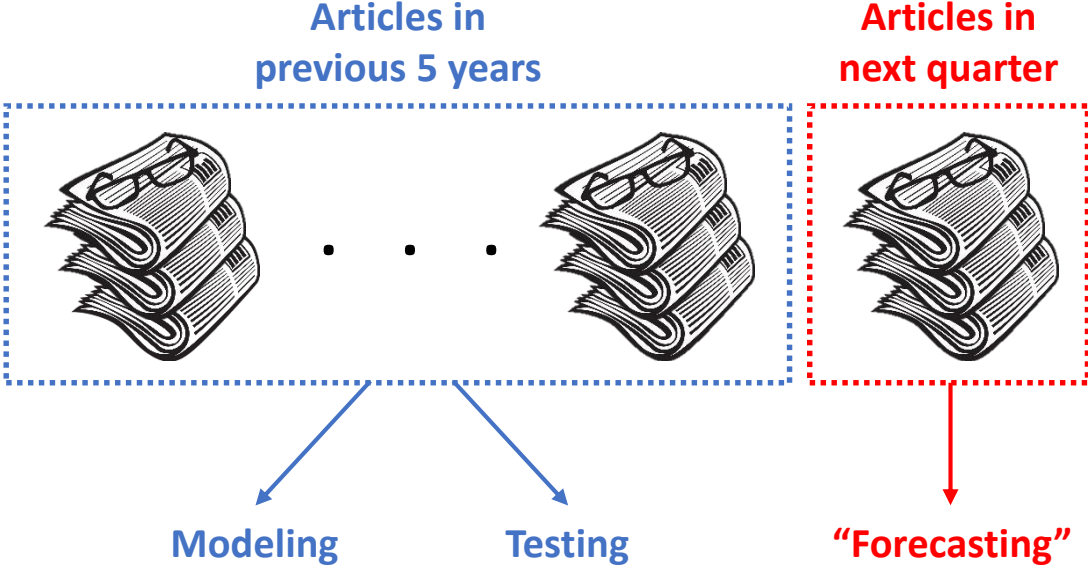Predict changes in gov't policies

Front page?

# Methodology

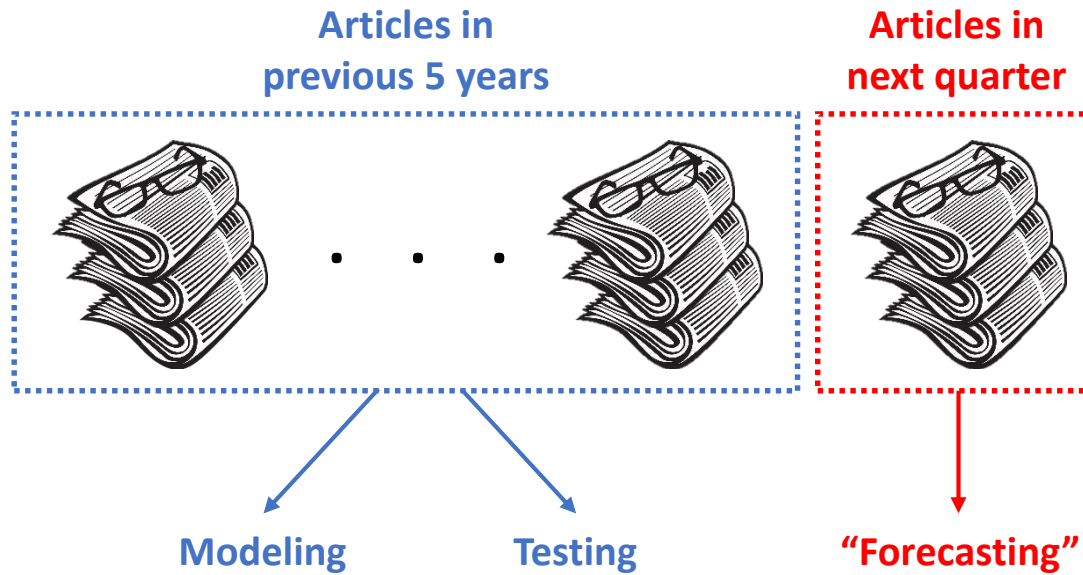Imagine an avid reader of the *People's Daily* who

1. reads recent articles;

2. forms a paradigm about front-page content;

3. tests the paradigm on new articles.

# Methodology

Articles in
previous 5 years

Articles in
next quarter

Modeling      Testing

"Forecasting"

$$\text{PCI} = \left| \text{Testing performance} - \text{"Forecasting" performance} \right|$$

# Methodology

Articles in previous 5 years

Articles in next quarter

Modeling          Testing          "Forecasting"
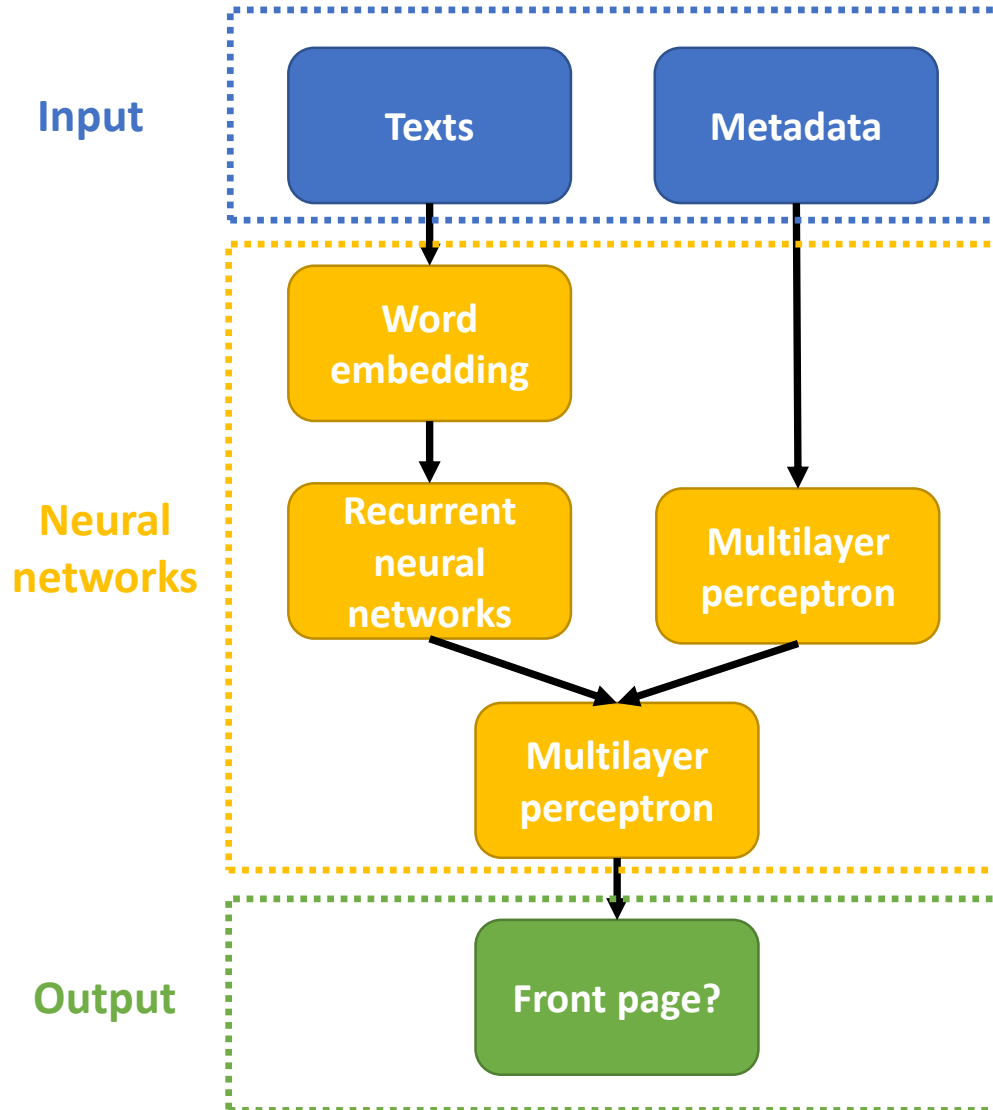
$$PCI = \left|\; \text{Testing performance} - \text{"Forecasting" performance} \;\right|$$

*"Context-free" — it does **not** require an understanding of the Chinese context.*

# Methodology: modelling

**Input**

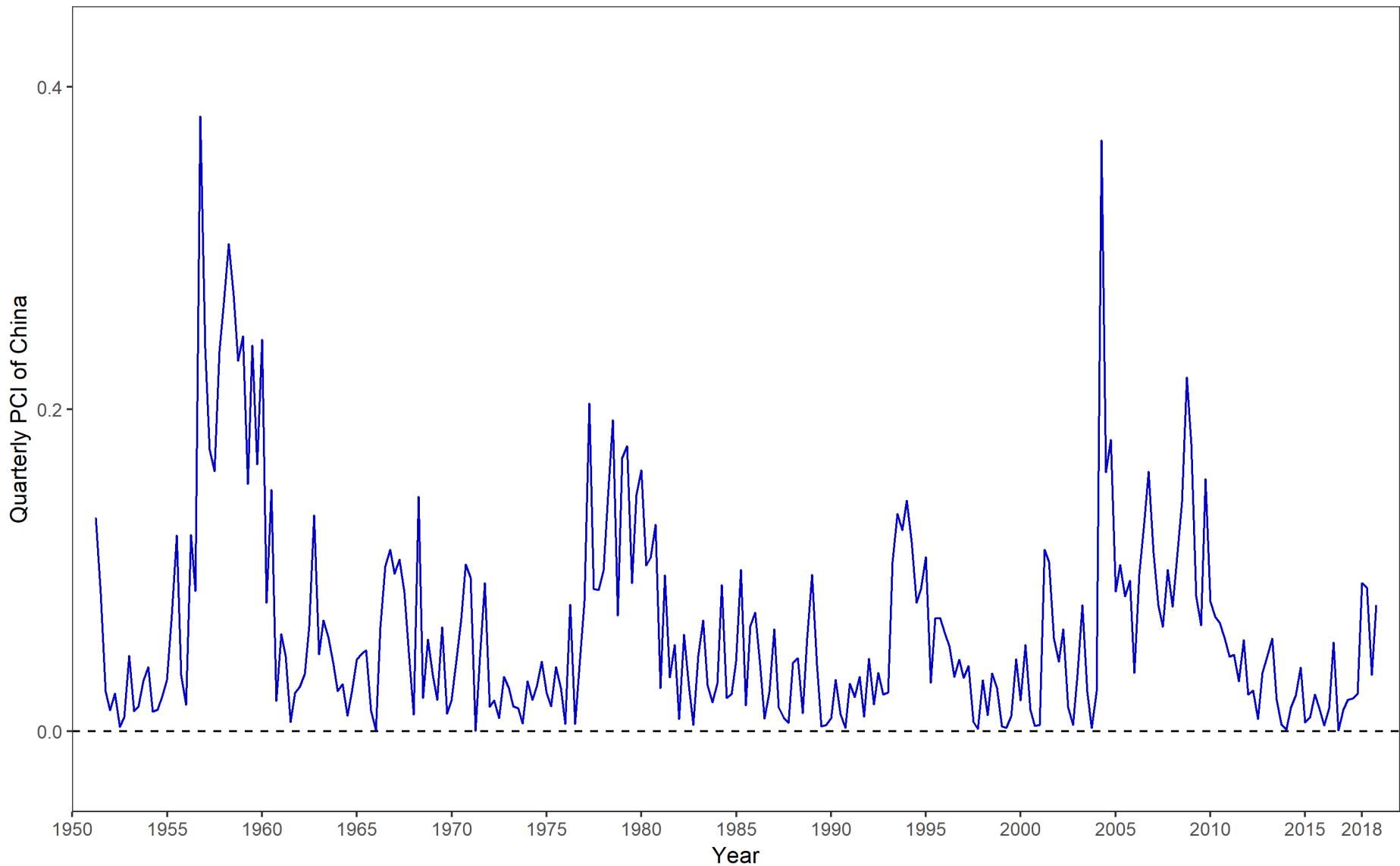Texts    Metadata

$x$ : each article as an observation.

**Neural networks**

Word embedding

Recurrent neural networks    Multilayer perceptron

Multilayer perceptron

$f$ : a complicated function.

**Output**

Front page?
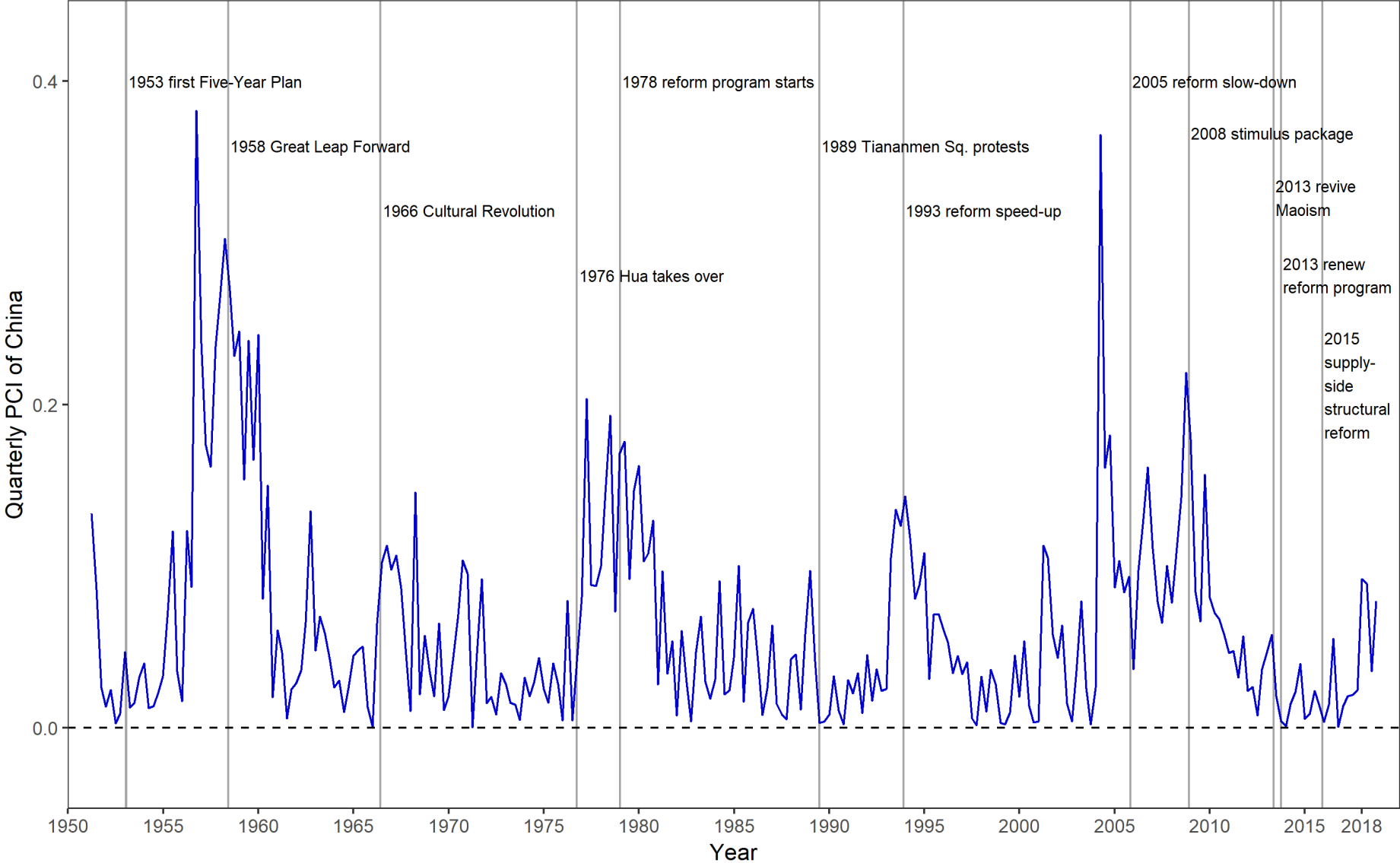
$y = f(x)$

# Results

# Result: PCI

# Result: PCI — with ground truth

# 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.

# Potential applications

# App 1: PCI for other countries

Predicting other (ex-)Communist regimes' policies:

- USSR's *Pravda*

- GDR (East Germany)'s *Neues Deutschland*

- DPRK (North Korea)'s *Rodong Sinmun*

- Cuba's *Granma*

- Vietnam's *Nhân Dân*

# App 1: PCI for other countries

Predicting other (ex-)Communist regimes' policies:

- USSR's *Pravda*

- GDR (East Germany)'s *Neues Deutschland*

- DPRK (North Korea)'s *Rodong Sinmun*

- Cuba's *Granma*

- Vietnam's *Nhân Dân*

PCI Cuba is in the works!

# App 2: measuring media bias

Measuring media bias in the US:

- Replicate the same analysis on US newspapers.

- Compare $"PCI"_{WaPo}$, $"PCI"_{NYT}$, $"PCI"_{WSJ}$, ...

- Divergence among $"PCI"$'s $\dashrightarrow$ Polarization in media

# App 3: predicting vote change

Predicting vote change in legislation:

- *People's Daily* articles ⋯➤ **Legislators' public statements**

- Page numbers of articles ⋯➤ **Legislators' names**

- What if Sen. *A*'s statement is mistaken as Sen. *B*'s?

Source of predictive power:

- Political necessity to justify vote changes by making different statements *in advance*.

# App 4: predicting ideology change

Predicting changes in judges' ideological leaning:

- *People's Daily* articles  ⋯→  **Judges' judicial opinions**

- Page numbers of articles  ⋯→ **Judges' names**

- What if Judge *C*'s opinion is mistaken as Judge *D*'s?

Source of predictive power:

- Changes in ideological leaning tend to be gradual yet long-lasting.

# Next steps

- PCI China will be upgraded to a monthly index.

- PCI Cuba is in the works.

- Other applications…

# Takeaways

- The first index to predict China's "next big things."

- A new way to uncover hidden patterns from trivial labels.

- A wide range of potential applications.

# Interested in DIY?

- Website: www.policychangeindex.com (newsletter sign-up)

- Research paper: www.policychangeindex.com/Reading_China.pdf

- Source code: www.github.com/open-source-economics/PCI

Contact us:

- Julian TszKin Chan: ctszkin@gmail.com

- Weifeng Zhong: weifeng@weifengzhong.com

# Questions?