

Electricity consumption in the service sector in Taiwan:

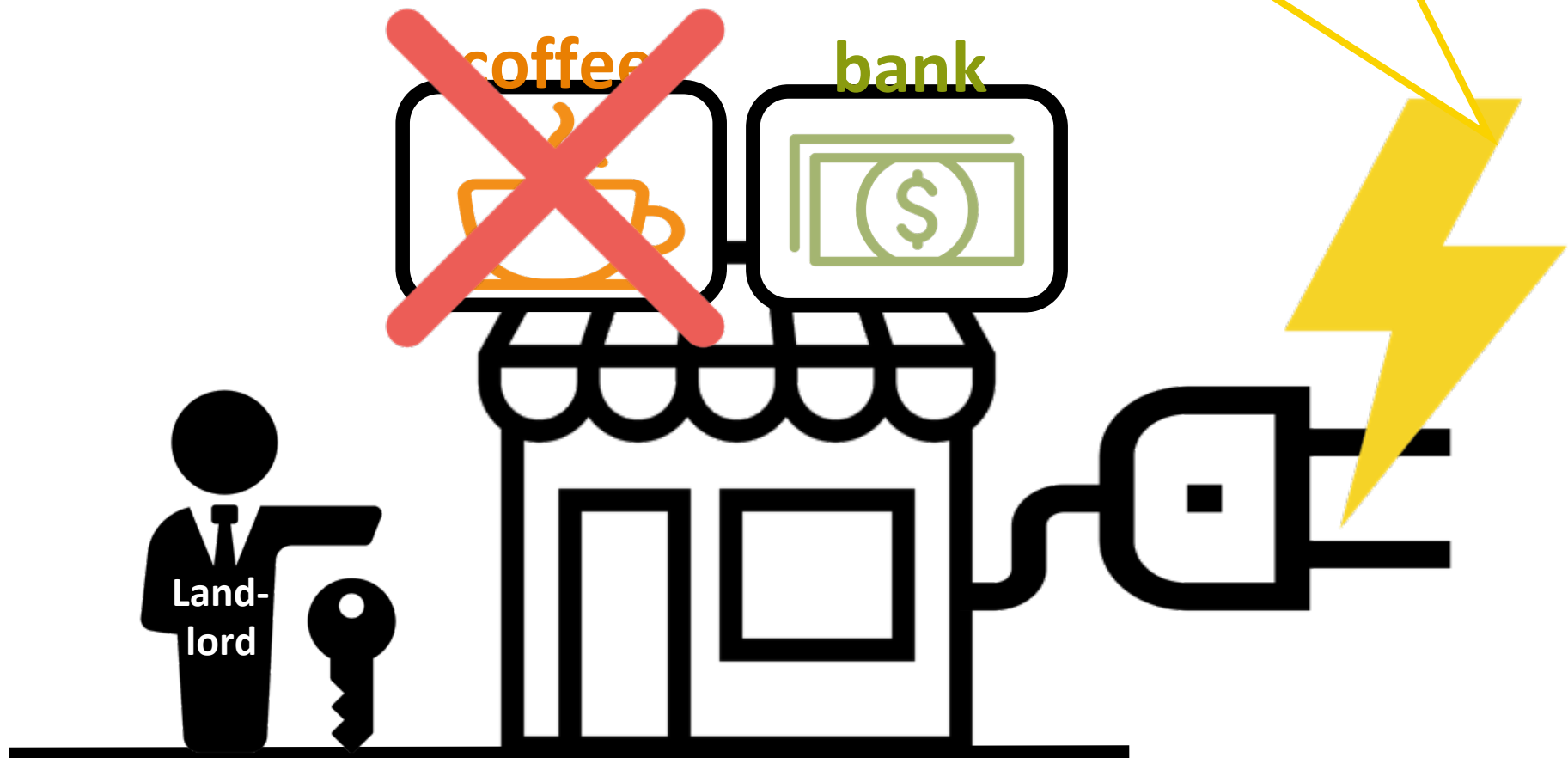
# Cross-dataset calibration

Su, Yu-Wen

Industrial Technology Research Institute (ITRI), Taiwan



~~Food service~~  
~~Financial service?~~  
Other service



# Outline

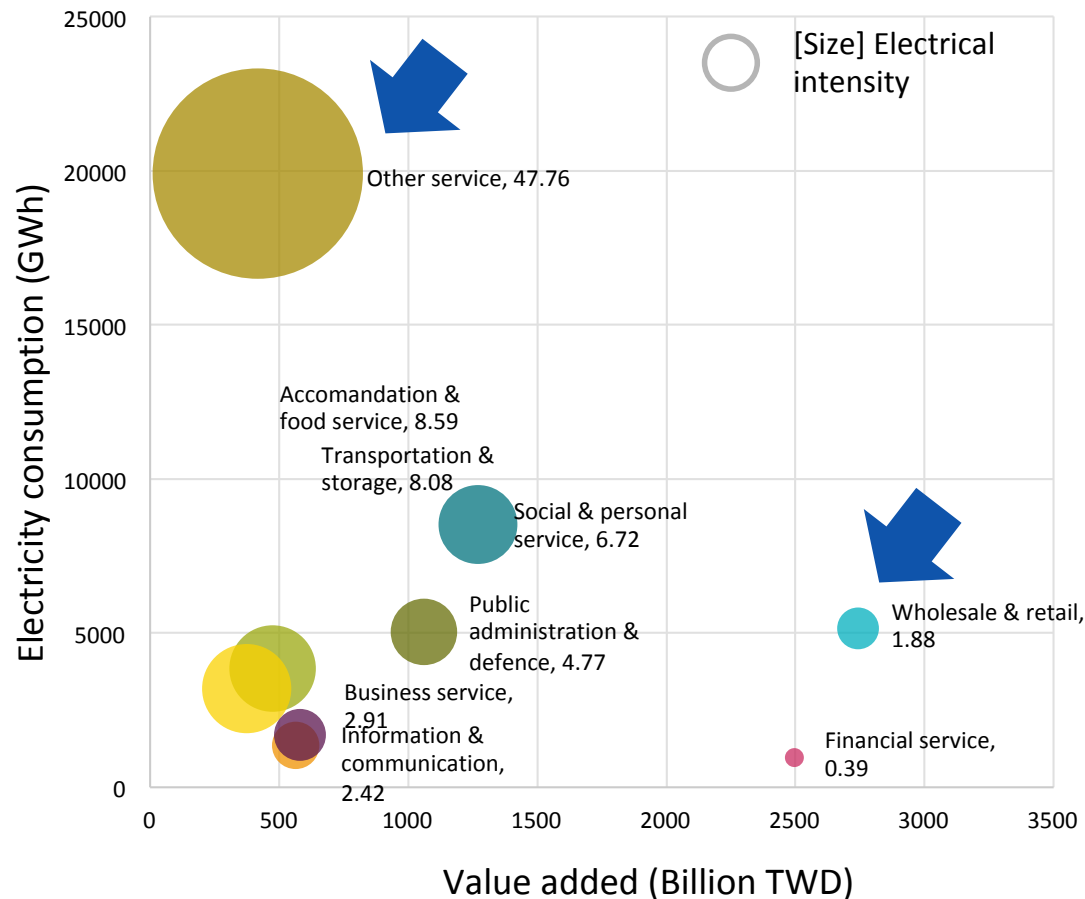
## Research question

Do we have the right data of industrial electricity consumption?

- Motivation: data inconsistency between economic & energy aspects
- Datasets: electricity consumption vs. tax information statements
- Approach: cross-dataset calibration
- Calibrated results
- Conclusions & further research
- Discussions

# Motivation

## Economic & energy data inconsistency



### Note:

1. Electricity consumption (E) come from the Bureau of Energy & Taiwan Power Company
2. Value added (VA) come from the Directorate General of Budget, Accounting and Statistics
3. Electrical intensity =  $E/VA$  (kWh/1000TWD)
4. Business service: Professional, scientific, & technical activities; Administrative & support service
5. Social & personal service: Education; Health & social work activities; Arts, entertainment & recreation

1. Identifying detailed industrial electricity consumption in the service sector is not the priority for electrical supplier
2. Short life cycle of SMEs causes detailed industries difficult to classify (e.g. address transferring or company moving)

# Datasets

Deal with the **population** directly, not just the **sample**

- **Electricity consumption**

- 1,243,004 low-voltage commercial users
- Annual data (2017)
- Source: Taiwan Power Company (Taipower)

- **Tax information statements**

- Company list of registration, non-operation (temporary), defunct (close formally), dissolution (on longer exist), and alternation
- Monthly updated (2017m1 - 2019m1)
- Source: Ministry of Finance in Taiwan

Electricity consumption	... ..	Billing address	Company name	GUI number	... ..	Division/ Section
-		-	-	-		A
-		-	-	-		A
-		-	-	-		B
-		-	-	-		B
-		-	-	-		...

(complete)    (incomplete) (incomplete)

\* GUI: Government Uniform Invoice

# 5 steps

## Step 1

Tax  
information  
statements  
(current ► 2017)

### Date rolling back

Current exist (2019m1)  
– Established after 2018m1  
+ Closed in 2017m1-m12

## Step 2

Electricity  
consumption  
(2017)

Tax  
information  
statements  
(2017)

### Data cleaning

- Full/half characters
- Abbreviation
- Typo correction

## Step 3

Electricity  
consumption  
(2017)

Address  
Company name  
GUI number

Tax  
information  
statements  
(2017)

### Integration

## Step 4

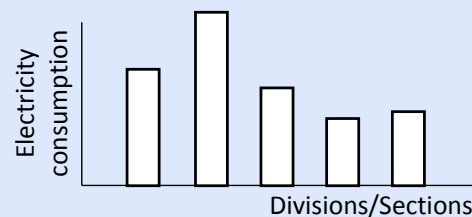
Integrated data  
(2017)

industry 1  
industry 2  
industry 3  
....

96 divisions → 19 sections

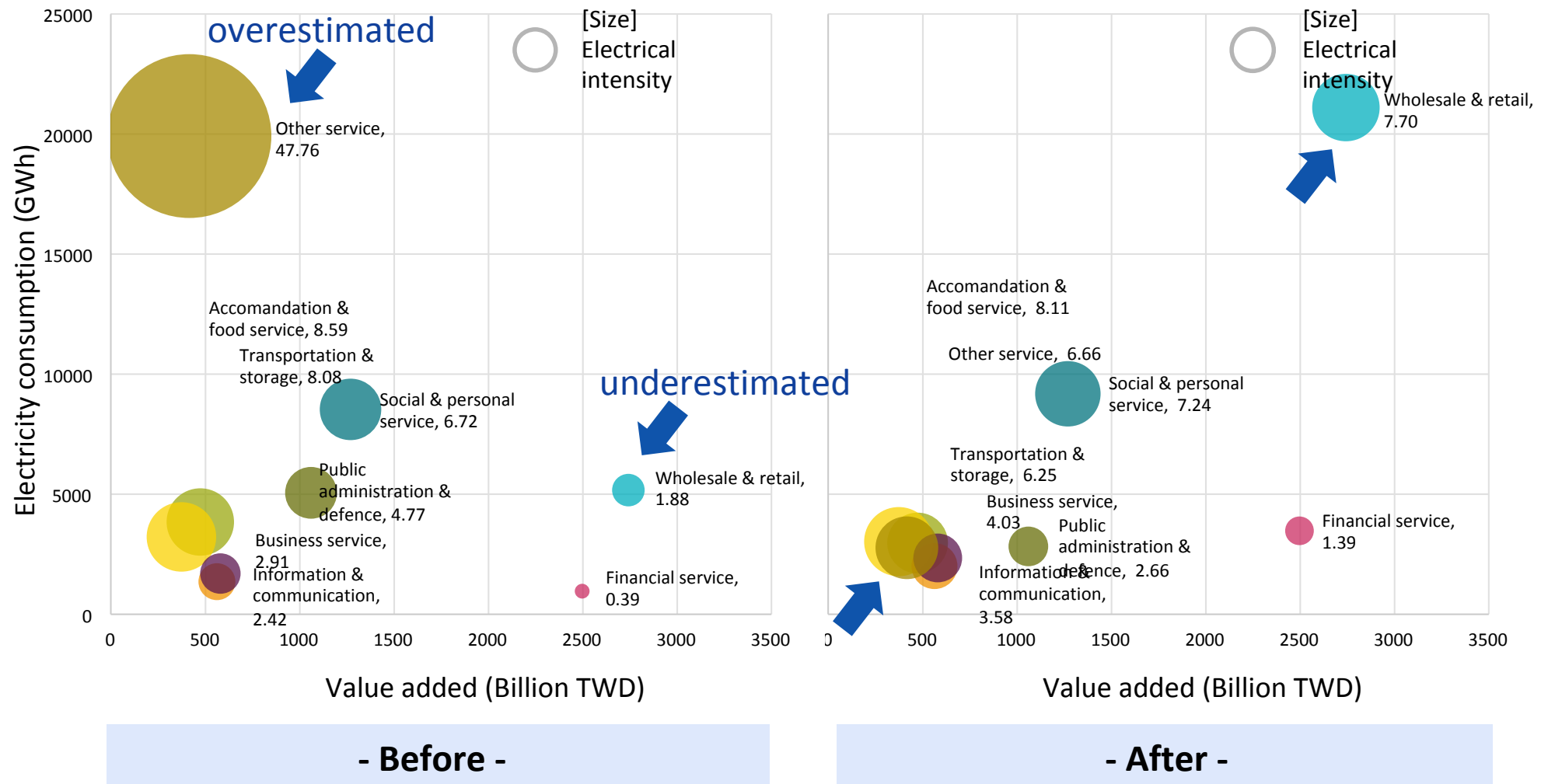
### Industry identification

## Step 5



### Electricity consumption analysis

# Results: calibrated industrial electricity consumption



\* Only the low-voltage commercial electricity consumption was calibrated.

# Results: match conditions

		Company #		Electricity consumption (MWh)	
		Match #	Match rate	Match consumption	Match rate
<p>Best cases</p> <p>Companies move to other places</p> <p>Addresses were transferred to other producers</p>	address + GUI	223,889	18.0%	5,008,681	29.6%
	address + company name	4,397	0.4%	72,452	0.4%
	GUI + company name	99,395	8.0%	2,822,402	16.7%
	GUI	69,887	5.6%	1,470,633	8.7%
	address	367,052	29.5%	3,134,203	18.5%
Total matched		764,620	61.5%	12,508,371	73.9%
Unmatched		478,384	38.5%	4,423,432	26.1%
Total		1,243,004	100.0%	16,931,804	100.0%

Small enterprises, who consume less electricity individually, were the source of unmatched industrial electricity consumption.



# Some notes

- Part of commercial low-voltage electricity (16%) was consumed in the industrial sector (e.g. small factory)
- Multiple divisions → majority, first (main business)
- Address cleaning in step 2 could be easier for alphabet then character language system, e.g.
  - 八, 捌, 8 → 8
  - 臺北市, 北市 → 台北市 (Taipei city)

# Conclusions

- The cross-dataset calibration is an efficient approach to **ensure the correctness of energy statistics** (cost less & deal with population)
- The dataset of electricity consumption and tax information have already existed for years. All we have to do is connect them **regularly**
- Further research
  - Increase the matched rate (with fuzzy match maybe)
  - Include commercial high-voltage and residential electricity consumption

# Discussions

1. How to ensure the correctness of electricity consumption of detailed industries in your cases?
2. Discuss the limitation of cross-dataset calibration
3. Suggest other possible applications of this approach

Thank you 😊

Su, Yu-Wen  
Industrial Technology Research Institute (ITRI), Taiwan  
[yuwensu@itri.org.tw](mailto:yuwensu@itri.org.tw)

