Designing a Country SME-DI using Firm-level Data: The Case of Thailand

@Conference on the Ecosystem of Vibrant MSMEs for Resilient Growth in Asia and the Pacific

Day-3 / Session 5: MSME Policies and Data Science

Nov 22nd, 2024

Daisuke Miyakawa (UTEcon / Waseda U.)
Shigehiro Shinozaki (ADB)

* This is a result of a joint research project of ADB and U Tokyo Economic Consulting (UTEcon). Usual disclaimers are applied.

Background-1

- MSMEs are showing sizable presence

Table 1: MSMEs in Developing Asia and the Pacific
(percentage share)

	All Countries	Southeast Asia	South Asia	Central and West Asia
MSME development				
• **Number of MSMEs to total enterprises**	96.6%	98.0%	99.6%	99.2%
• **MSME employees to total employees**	55.8%	66.4%	76.6%	51.9%
• **MSME contribution to economic output**	28.0%	41.2%	17.7%	41.5%
• **MSME exports to total export value**	26.3%	13.3%	37.4%	28.3%
Access to finance (bank credit)				
• **MSME loans to national GDP**	10.6%	13.3%	5.2%	11.1%
• **MSME loans to total bank loans**	22.0%	12.3%	12.5%	33.1%
• **MSME NPLs to total MSME loans**	7.2%	5.3%	12.1%	4.3%

GDP = gross domestic product, MSME = micro, small, and medium-sized enterprise, NPL = nonperforming loan.
Notes: Reporting countries only. Data based on latest available data until 2022. Data for all countries cover 25 countries: 10 from Southeast Asia; 5 from South Asia; 7 from Central and West Asia; and 3 from the Pacific.
Source: Asia SME Monitor 2023 database.

Source: Shinozaki, S., D. Miyakawa, and R. Arahan (2024) "Factors Affecting Micro, Small, and Medium-Sized Enterprise Development in Developing Asia: Findings from a Probabilistic Principal Component Analysis," *ADB Economics Working Paper Series* No.715.

Background-2

- There are many extant trials

A) Description of multiple information of MSMEs' activities

- Institutional features, real activities (e.g., sales), import-export, access to finance, innovation etc.

- OECD (SME and Entrepreneurship Outlook), ERIA (ASEAN SME policy Index), ITC (SME Competitiveness Outlook), WIPO (Global Innovation Index), GEDI (Global Entrepreneurship Index)

B) Qualitative Discussions based on detailed surveys

- OECD (Scoreboard), WEF (Global Competitiveness Index)

Research Question

Q. Can we “systematically” measure MSMEs’ activities…?

i. …based on available data

⇔ Able to take care of missing data & sample data

ii. …with little discretion

⇔ A machine learning method

iii. …for flexible targets

⇔ Able to focus on a specific dimension (e.g., industry)

⇒ SME-DI

What We Do

- P-PCA (Probabilistic-Principal Component Analysis)

  - Estimate the PCs driving the variables in firm-level panel data

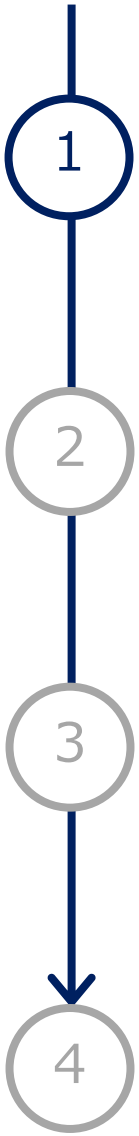
    - Applicable to the data with missing records
    - The choice of inputs ⇔ PCs' interpretation

⇒ Thailand's firm-level granular data (+ASM data)

⇒ Use those factors as SME-DI

≠ C.f. ADB (2022), Shinozaki et al. (2024) for country-year panel

Agenda



1 Methodology

2 Data

3 Results

4 Conclusion

1. Method

- P-PCA

$$x_{i,t} = W z_{k,t} + \mu_i + \epsilon_{i,y}$$

where

$x_{i,t}$: Data (e.g., panel of variables over industry-region-year)

$z_{k,t}$: Extracted factors (e.g, the top 3 PCs, $k \ll i$)

- A selected number of $z_{k,t}$ (SME-DI) to summarize $x_{i,t}$

⇔ "Systematically measure MSMEs' activities"

Agenda

1. Methodology
2. Data
3. Results
4. Conclusion

2. Data

- D&B data (i.e., firm-level “granular” panel data)
  - Firm-year panel data in Thailand
  - Around 400k companies over 2016-2023
  - Variables
    - Sales, #employees, profit (gross & net)
    - Total assets, trade credits
    - Total liabilities, total current liabilities, trade debts
    - Bank borrowing
    - Working capital
  - F/S ending in a year YYYY ⇔ Year = YYYY

2. Data

- D&B data
  - Industry
    - A) Agricultural, forestry, and fishing
    - B) Mining
    - C) Construction
    - D) Manufacturing
    - E) Transportation
    - F) Wholesale
    - G) Retail
    - H) Finance
    - I) Service

2. Data

- D&B data
  - Region
    - 6 regions of Thailand
    - + Bangkok

Source: Muneerat et al. (2022)

- Also, ASM (i.e., country-year) data as a routine work

2. Data

◻ Processing data

(Step-0: Original data)

Firm-year panel data accompanied by multiple variables

(Step-1: Baseline)

Aggregate each variable in industry-region-year (i.e., a cell) and apply P-PCA

↓

We have several options (e.g., mean) to aggregate and use a simple summation (i.e., summing up the value over the firms in a cell) in the current version of our analysis

2. Data

◘ Processing data (cont'd)

(Step-2: Industry SME-DI)

Select an industry, aggregate each variable in region-year, and apply P-PCA

(Step-2': Regional SME-DI)

Select a region, aggregate each variable in industry-year, and apply P-PCA

2. Data

◘ Processing data (cont'd)

(Step-3: Real vs. financial SME-DI)

Filter either real or financial variables, follow (Step-1), and apply P-PCA

(Step-4: Small sample exercise)

Randomly choose 50% or 10% of the data, follow (Step-1), and apply P-PCA

↓

This last exercise aims to see to what extent the dynamics of the extracted PCs depend on the sample size. If the dynamics of the extracted PCs share some important features with that we obtained from the entire dataset, we might be able to use some surveyed (i.e., a limited number of) data to construct SME-DIs

Agenda

1 Methodology

2 Data

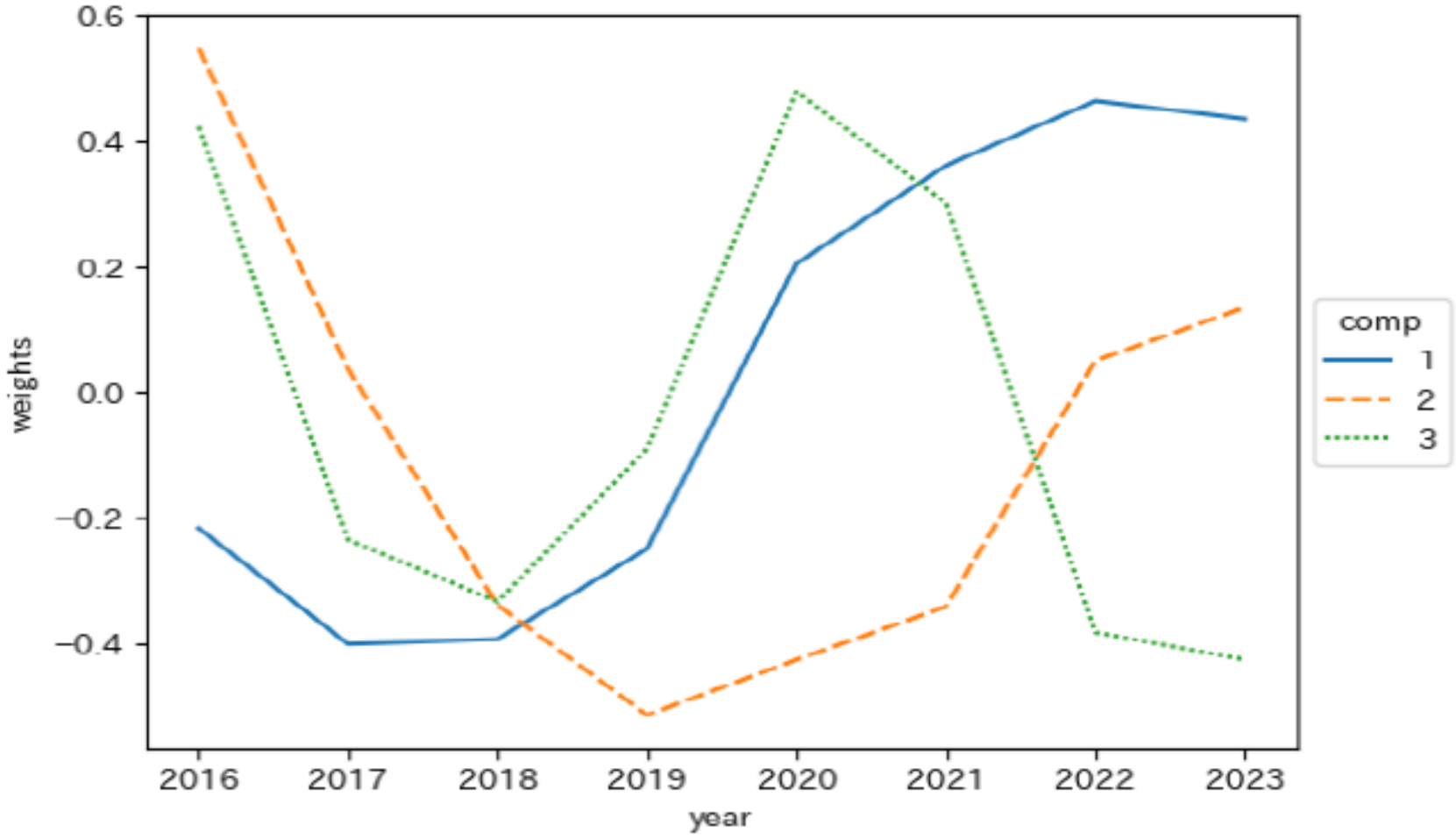
3 Results

4 Conclusion

3. Results

☐ Step-1: Main analysis (note: 87% of data variation ⇔ 3 PCs)

Figure 1: Baseline result on full sample



3. Results

PC1

- Positively correlated with trade debt, working capital, and bank borrowing
- Negatively correlated with a selected variables such as employment
- Largely accounts for the trade credit as well as financial burden carried by the firms in some specific industries and regions
- Given that the first factor started to increase from year 2020 and has stayed in a high level with a slight sign of peak out in 2022, we can infer that the financial side of the Thailand firms have been in a specific situation (i.e., larger borrowing and financial needs) accompanied by smaller employment over the recent COVID-19 pandemic.

3. Results

PC2

- Negatively correlated with gross profit in most of the cases
- Accounts for the “inverse” of firm profitability
- Given that the second factor steadily declined up to 2019 (i.e., firm performance became better) and increased from year 2020, we can infer that the recent pandemic provided a significant negative impact on firm profitability in Thailand.
- One additional information we can extract is that the second factor has not shown any decline in the recent years. This suggest that at least through the lens of P-PCA method, the profitability of individual firms in Thailand are still in its process toward recovery as of 2023.

3. Results

PC3

- Positively correlated with employment and profit

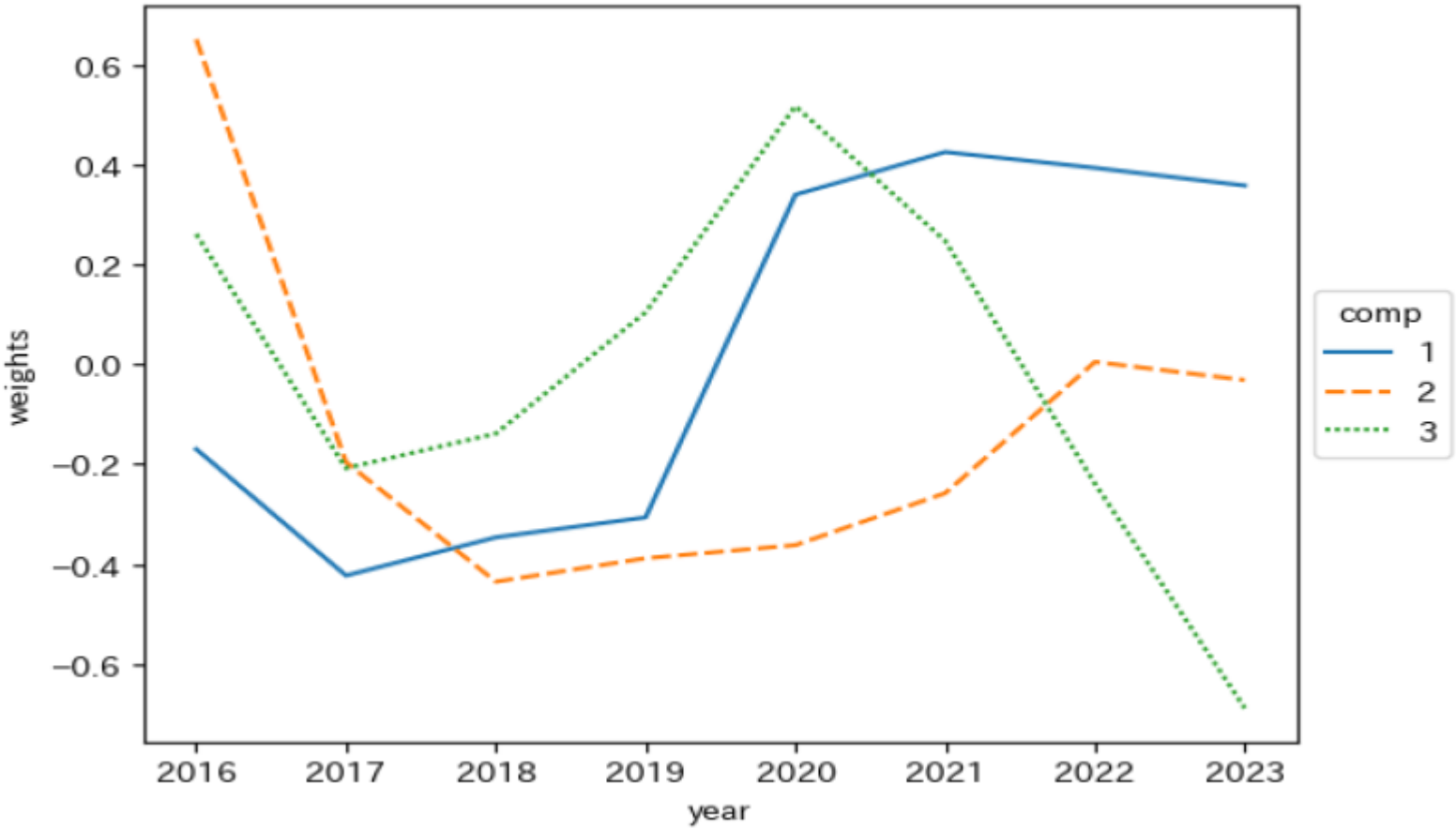
- Complement to the interpretation of the first factor, along with the increasing burden in financial side (i.e., trade credit) captured by PC1, the employment and firm performance faced significant downturn, which is consistent with the implication of the second factor.

- One interesting finding is that the dynamics of the first factor is mainly driven by financial side (i.e., trade credit and borrowing) while the third factor mainly accounts for (the inverse of) firm performance.

3. Results

SME-DI for a specific industry

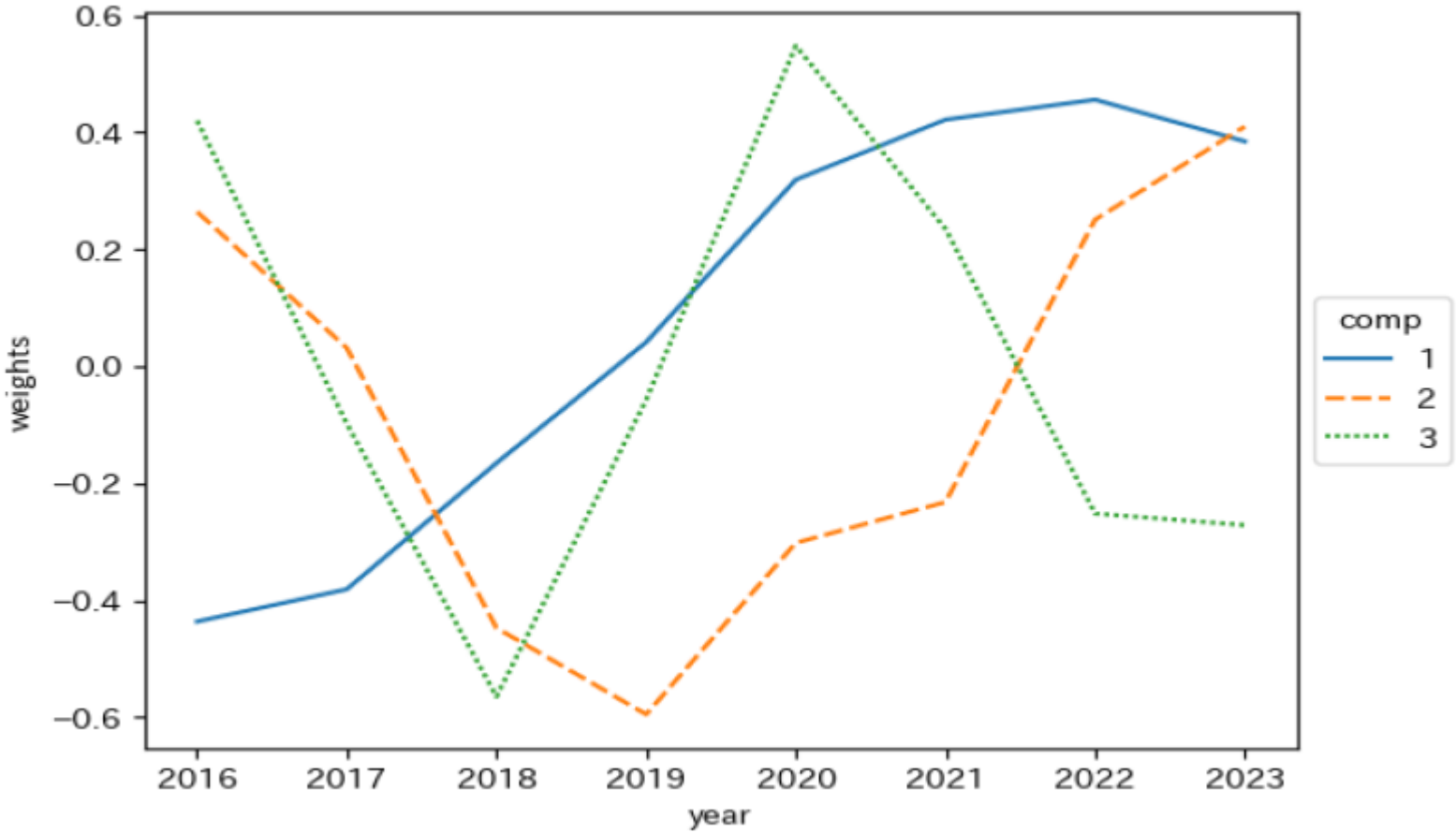
Figure 5: Industry F (wholesale)



3. Results

SME-DI for a specific region

Figure 6: Bangkok



3. Results

◘ Real and Financial SME-Dis

- Associations with each variable need to be carefully examined

Figure 7: Real variables

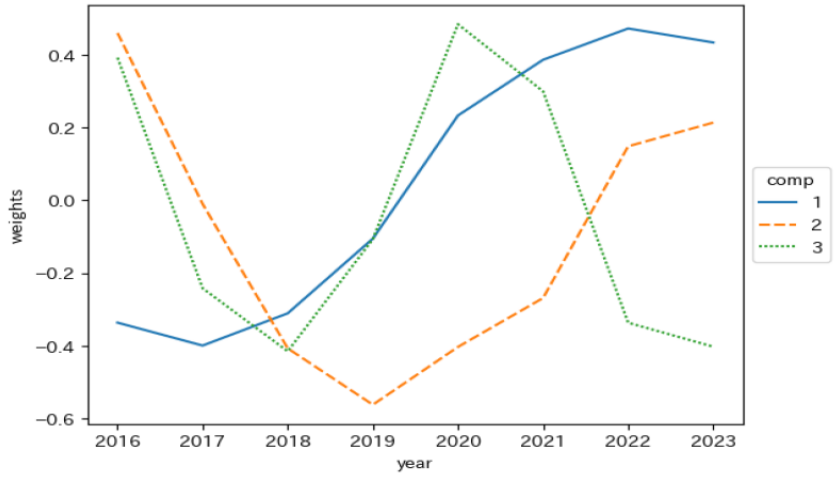
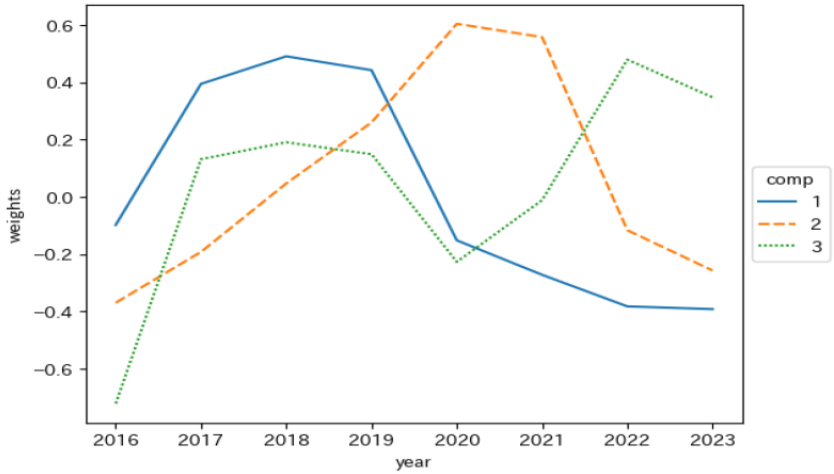


Figure 8: Financial variables



3. Results

◘ Small sample exercises

Figure 1: SME-DI based on a full sample

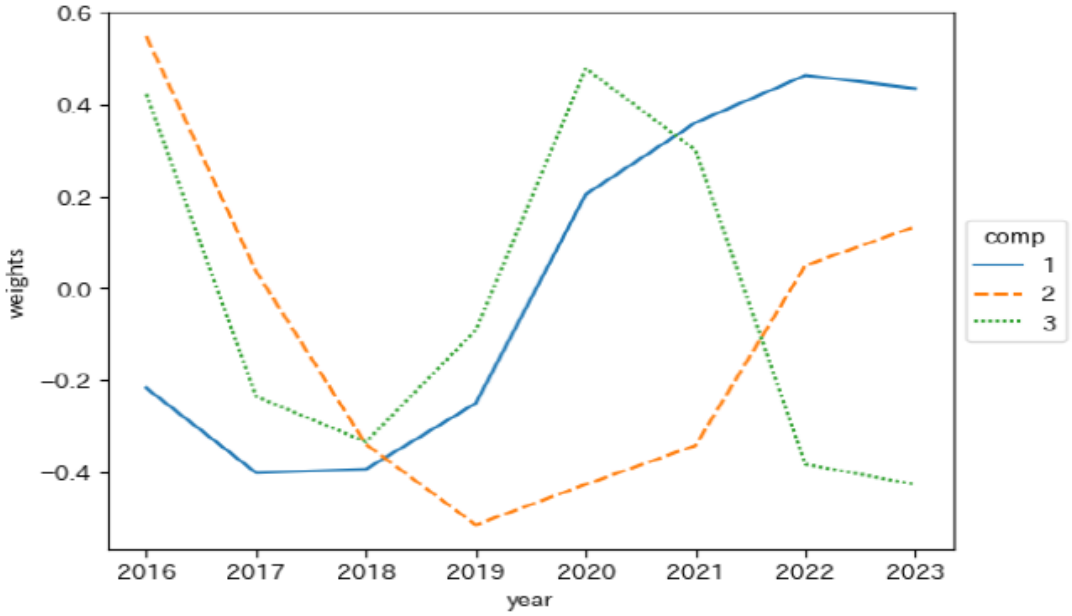
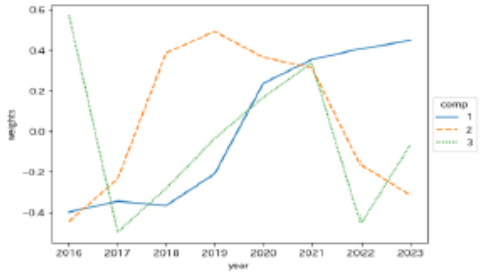
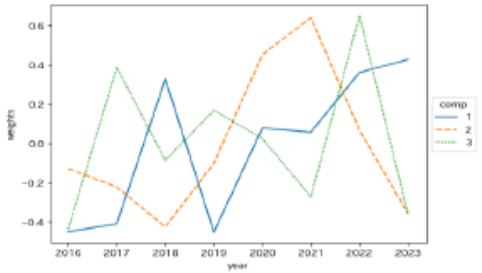


Figure 2: SME-DI based on a small sample

Panel A: 50% case



Panel B: 10% case



3′. Validation of SME-DI

- Compare the following two models (i.e., y_t vs.$\hat{y}_t$):

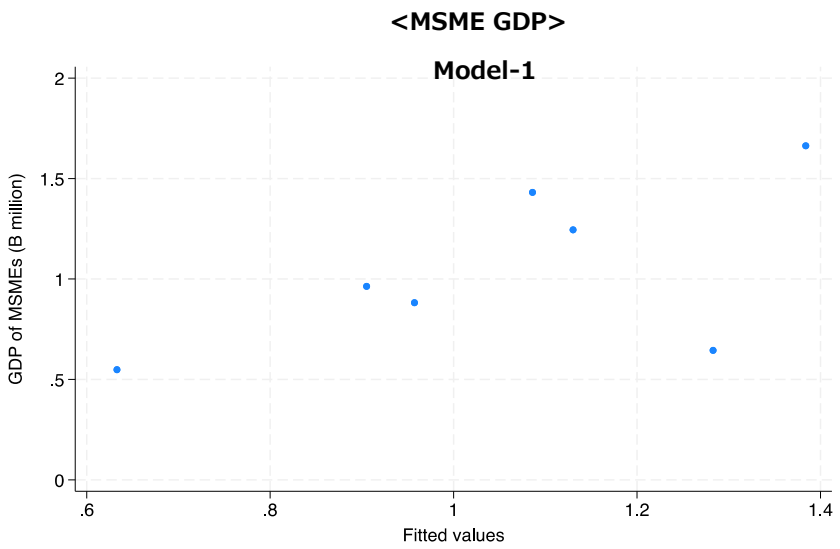
Model-1: $y_t = \alpha + \beta y_{t-1} + \varepsilon_t$

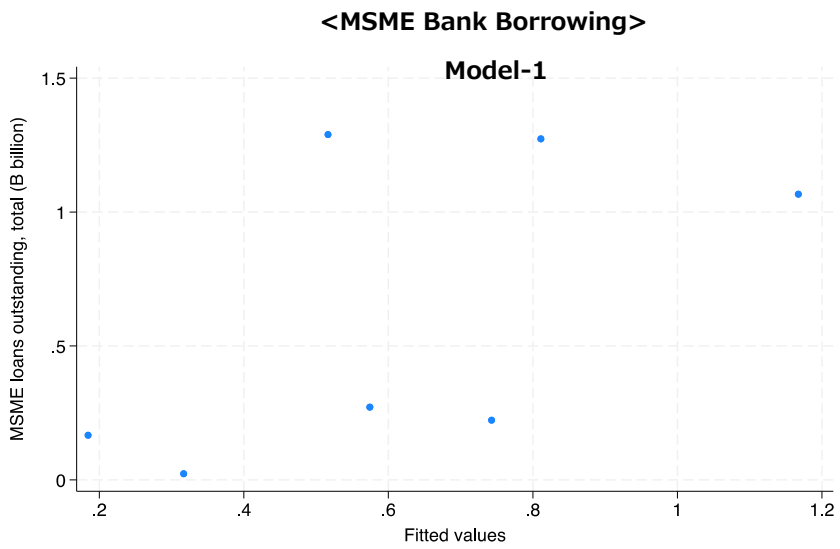
Model-2: $y_t = \alpha + \beta y_{t-1} + \sum_{k=1}^{3} \gamma_k \widehat{PC(k)}_{t-1} + \varepsilon_t$

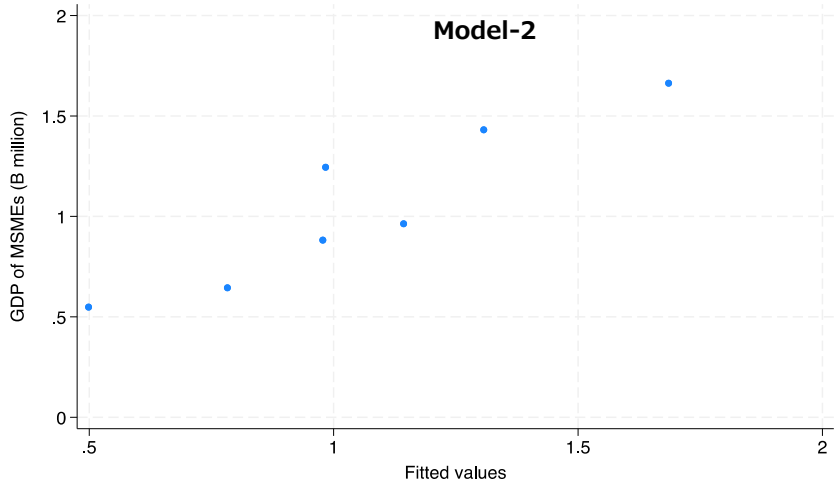
where y_t is either MSEM GDP or MSME borrowing

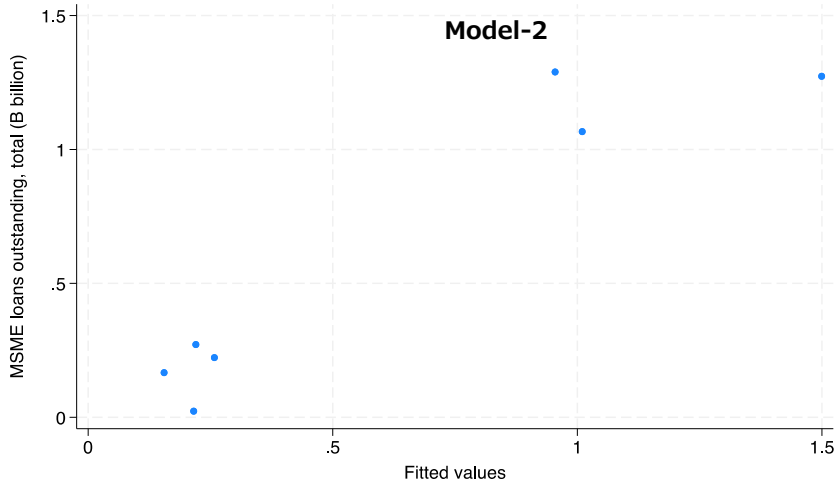
- ✓ **Note**: $\widehat{PC(k)}_{t-1}$ should be estimated sequentially (i.e., using only the information up to $t-1$) while we are now using a serries of $\widehat{PC(k)}_{t-1}$ estimated by using the entire dataset. In this sense, the information in the periods t and onward is leaked to $\widehat{PC(k)}_{t-1}$. Given this limitation, we show the results only for the purpose to show that $\widehat{PC(k)}_{t-1}$ has a valid association with y_t as it should be.

3′. Validation of SME-DI









Agenda

1 Methodology

2 Data

3 Results

4 Conclusion

4. Conclusion

- Using granular data in Thailand, we demonstrate how to systematically measure MSMEs' activities…
  i. …based on available data potentially suffering from missing records,
  ii. …by using a method immune to discretion,
  iii. …accounting for a specific measurement motive
  iv. …applicable to a smaller sample

- Future work?
  - Other countries
  - Systematic interpretation based on the results
  - Incorporate the results into prediction work

Reference

Asian Development Bank (2022) Asia Small and Medium-Sized Enterprise Monitor 2021 Volume IV— Pilot SME Development Index: Applying Probabilistic Principal Component Analysis. http://dx.doi.org/10.22617/TCS220150-2. Manila: ADB.

Muneerat, P., W. Saeheng, T, Khan, I. Kamdar, K. Maliwan, and J. Taweekun (2022) "Topographic Data Analysis of Observed Wind Data Recorded at 15 Different Stations in Thailand," *Journal of Advanced Research in Fluid Mechanics and Thermal Sciences* 92(2):116-137.

Shinozaki, S., D. Miyakawa, and R. Arahan (2024) "Factors Affecting Micro, Small, and Medium-Sized Enterprise Development in Developing Asia: Findings from a Probabilistic Principal Component Analysis," *ADB Economics Working Paper Series* No.715.

<Contact Information>

Daisuke Miyakawa（宮川大介）

Chief Economist
UTokyo Economic Consulting Inc. (UTEcon)
Professor
School of Commerce, Waseda University

Office 1318 Building 11, 1-6-1 Nishiwaseda
Shinjuku-ku, Tokyo 169-8050 Japan
E-mail: damiyak@waseda.jp
https://sites.google.com/site/daisukemiyakawaphd/