PILOT'S ASSESSMENT OF THE VIETNAM PROVINCIAL INNOVATION INDEX

Hoang Minh
Vietnam Institute of Science, Technology and Innovation

Nguyen Thi Phuong Mai[1], Nguyen Vo Hung
National Institute for Science and Technology Policy and Strategy Studies

Summary:

The Global Innovation Index (GII) of the World Intellectual Property Organization (WIPO) has been used by the Viet Nam Government since 2017 as a management toolkit for various ministries and agencies. As this index operates at the national level, thus authorities at provincial level have not yet played a role in improving Viet Nam's GII index. Provincial authorities with different characteristics require tailored solutions and socio-economic development models based on science, technology and innovation (STI). Establishing a provincial-level innovation index will provide a practice and comprehensive overview of the status of the socio-economic development model based on STI in each locality, offering evidence of strengths, weaknesses, potential factors, and necessary conditions to promote socio-economic development based on STI, thereby contributing to the nation's GII improvement. Building this provincial -level innovation index involves several steps, and the pilot assessment is expected to provide insights and foundations for the formal construction and implementation of this toolkit in the coming time. This article presents the results of the pilot's assessment on the local-level innovation index conducted in 2021-2022, covering 18 localities based on the GII index framework and methodology.

Keywords: *Science and technology; Innovation; Innovation index; Measurement of innovation; Socio-economic development.*

Code: *23101701*

1. Background

At the national level, since 2017, the Government of Vietnam has utilized the Global Innovation Index (GII) issued annually by the World Intellectual Property Organization (WIPO) to identify the strengths and weaknesses of the national innovation system. This serves as a basis for proposing suitable solutions and improvements, as well as timely to develop and issue relevant policies.

At the provincial level, differences in the scale of socio-economic development, population, land, economic structure, and development orientation among regions necessitate localities to select and adopt different models for socio-economic development based on their specific context, natural resource, and STI

[1] Author's email address: npmai.vn@gmail.com

development characteristics. Addressing these issues requires local authorities to understand the status of the socio-economic development model based on their STI unique characteristics/indices. Presently, local-level indices (such as Provincial Competitiveness Index, Administrative Reform Index, Digital Transformation, etc.) mainly assess every specific sector, lacking a comprehensive index that evaluates socio-economic development based on local STI development characteristics. Therefore, researching and developing an index describing the current state of the socio-economic development model based on local STI development characteristics is highly necessary. And conducting experimental assessment research on this index is a crucial step towards proposing and formally implementing it.

2. Experimental assessment Method for the Provincial Innovation Index

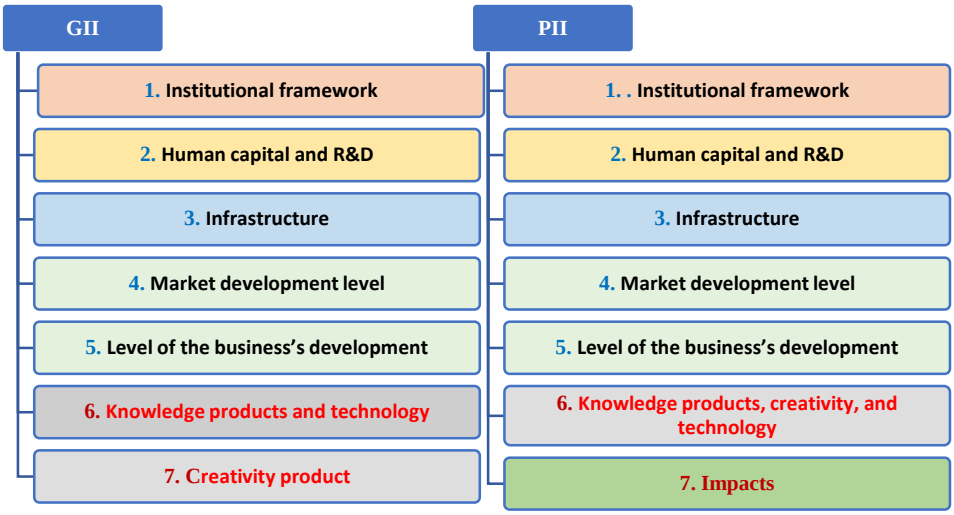
The method of developing the provincial innovation index was implemented by the research group based on the recommendations for constructing composite indices published by OECD in 2005 *(OECD/JRC, 2005).* The theoretical framework for constructing the provincial innovation index is derived from Nguyen Thi Phuong Mai research results of 2020. In the 2020 study, Nguyen Thi Phuong Mai et al.[2] applied the Global Innovation Index (GII) and international experiences to develop a framework for evaluating Provincial Innovation Index.

Internationally, some countries, including China, the United States, have measured local innovation index for several years. And notably, India and Colombia, based on GII methodology also established local IP indices *(Nguyen Thi Phuong Mai, 2020).* The local IP index of Colombia and India clearly illustrates this approach. The Indian index consists of 7 pillars, divided into two main input-output groups: Enablers, comprising 5 pillars related to input conditions, and Outputs, comprising 2 pillars, totaling 33 component indices. India, with its 29 states and 7 territories, has a total of 36 regions. All these regions are evaluated and ranked, categorized into three groups: (i) big major states (17 regions); (ii) northern states and mountainous regions (11 regions); (iii) cities, small states and territories (8 regions). Additionally, regions are grouped by income level for assessment and comparison across each pillar.

However, the mentioned study has some limitations, including: (i) data insufficiency; (ii) inconsistency in data collection and calculation methods for some indices across regions; (iii) trial secondary data collection for three regions; (iv) absence of calculation methods, score conversion, ranking; validity testing, audit methods, examination of the relationship between this index and other

[2] Research project at the Ministry level for 2018-2019. "*A study proposing a framework for the ministry, industry, and localities to improve Vietnam's Global Innovation Index (GII) and the potential application of the GII method to assess innovative capabilities at the provincial level*." National Institute of Science and Technology Policy and Strategy. Hanoi.

relevant provincial indices. Based on these research findings, we developed the framework for the Provincial Innovation Index (PII), consisting of 51 indices organized into 7 pillars based on the principles of the GII. This framework, shown in Figure 1 below, is designed for testing. In comparison to the GII, we merged columns 6 and 7 of the GII into one column in the PII and added column 7, column of "Impact," drawing from the experience of regional IP evaluation indices in the United States and the European Union.



Source: GII 2022 Report (WIPO) and PII Framework proposed by the authors

Figure 1. A comparison between the GII framework and the PII framework

Methods and data sources

Due to limited resources, it is impossible to carry out large-scale testing in all 63 provinces and cities. Within the framework of this study, we conducted testing in 20 localities (accounting for about 1/3 of the total number of localities of Vietnam). Localities are selected according to the following criteria: (i) Geography: distributed in all 6 economic regions; (ii) Income: different levels of per capita income; (iii) Economic structure: different economic structures; (iv) Scale: different sizes of area and population to represent the diversity of all 63 localities nationwide.

Table 1. Pilot's assessment Provinces

No.	Province/Locality	Region	Average Income per Capita/Month in 2021 (thousand VND)	Population Average in 2021 (thousand people)	Area (km^2)
1	Son La	Northern Midlands	1,834.3	1,287.72	14,109.83
2	Bac Giang		3,965.7	1,875.24	3,895.89

3	Thai Nguyen	and Mountains	3,813.7	1,323.15	3.521,96
4	Hai Phong	Red River Delta	5,093.4	2,072.39	1,526.52
5	Quang Ninh		3,991.6	1,350.85	6,207.79
6	Ha Noi		4,372.1	8,330.83	3,359.82
7	Ninh Binh		4,281.7	1,007.57	1,411.78
8	Vinh Phuc		4,511.2	1,191.78	1,236
9	Thanh Hoa	North Central and Central Coast	3,651.6	3,716.43	11,114.71
10	Nghe An		3,094.8	3,409.81	16,486.5
11	Da Nang		5,229.9	1,195.49	1,284.73
12	Quang Nam		3,543.2	1,518.48	10,574.86
13	Phu Yen		3,296.4	875.54	5,025.96
14	Thua Thien Hue				
15	Gia Lai	Central Highlands	2,329.4	1,569.72	15,510.13
16	Dong Thap	Mekong River Delta	4,198.9	1,601.31	3,382.28
17	Ben Tre		3,367.3	1,295.7	2,379.7
18	Can Tho		4,794.2	1,246.99	1,440.4
19	Ho Chi Minh city	Southeast	6,006.6	9,166.84	2,095.39
20	Binh Duong				

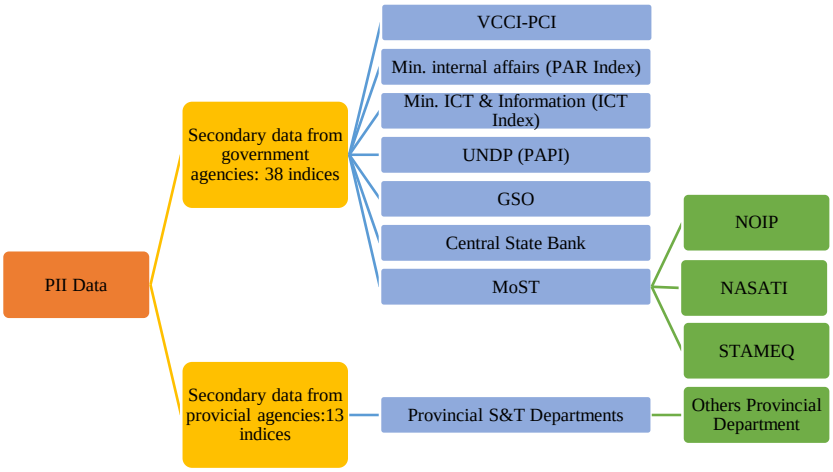
Source: Compiled by the authors from the GSO database Top of Form

The data were collected based on statistics and official management reports from central and local agencies, as well as from various index reports (such as Administrative Reform Index, Provincial Competitiveness Index, Digital Transformation Index, Efficiency of Provincial Governance and Administration Index).

The data source structure is as follows:

- From reports and statistics of central agencies: 35% (18 indices).
- From other index reports: 20% (10 indices).
- From state management data of the Ministry of Science and Technology: 20% (10 indices).
- From locally provided data: 25% (13 indices).

During the experimentation process, two localities did not complete data collection and did not submit within the specified timeframe. Despite providing documentation and training materials for data collection, some localities still made errors, such as using incorrect units and failing to provide sufficient supporting documentation.



Source: Compiled by the authors

Figure 2. The data source for pilot's PII index assessment

Table 2. List of index, sources and data collected year

Index	Year of data
1. Institutional framework	
1.1. Policy environment	
1.1.1. Policies to promote STI to serve socio-economic development [Source: Provided by provinces] 2021	2021
1.1.2. Legal framework and security [Source: PCI]	2021
1.1.3. Policies to support small and medium-sized enterprises [Source: PCI]	2021
1.2. Business environment	
1.2.1. Market entry costs [Source: PCI]	2021
1.2.2. Dynamic features of local government [Source: PCI]	2021
	2021

Index	Year of data
5. Level of the enterprise's development	
5.1. Knowledgeable labor	
5.1.1. Labor training costs, % of total business expenses [Source: PCI]	2021
5.1.2. enterprise's R&D costs/total R&D costs [Source: National Agency for Science and Technology Information]	--
5.1.3. Proportion of enterprises with R&D activities [Source: Government Statistical Office, GSO]	2020
5.2. Creative linkages	
5.2.1. Research cooperation between scientific and technological organizations and enterprises (%) [Source: Provided by provinces]	2021
5.2.2. The proportion of investment projects operating in industrial parks (%) [Source: Provided by provinces]	2021
5.2.3. The proportion of SME, cooperatives businesses in industrial	2020

Index	Year of data
1.2.3. Administrative reform [Source: PAR Index]	
1.2.4. Equal competition [Source: PCI]	2021
2. Human capital and R&D	
2.1. Education	
2.1.1. Education 2.1.1. Education index [Source: GSO]	2020
2.1.2. Ratio of students to teachers in secondary and high schools [Source: GSO]	2021
2.1.3. Ratio of secondary and high schools with STEM/STEAM training [Source: Provided by provinces]	2021
2.1.4. Education budget expenditure/total local budget expenditure [Source: Provided by provinces]	2021
2.2. R&D	
2.2.1. Full-time R&D personnel/10,000 population	2021
2.2.2. R&D expenditure from state budget/total state budget expenditure [Source: Provided by provinces]	2021
2.2.3. R&D expenditure/total state budget expenditure [Source: Provided by NASATI]	2019
3. Infrastructure	
3.1. ICT infrastructure	
3.1.1. ICT infrastructure [Source: ICT]	2020
3.1.2. Online public services [Source: ICT]	2020
3.2. General infrastructure	
3.2.1. General infrastructure [Source: PCI]	2021
3.2.2. Proportion of industrial land with built-in infrastructure (%) [Source: Provided by provinces]	2021

Index	Year of data
clusters on the total number of SME, cooperatives businesses operating in the provinces [Source: Provided by provinces]	
5.3. Knowledge absorption	
5.3.1. Number of research personnel in enterprises per 10,000 population [Source: NASATI]	--
5.3.2. Foreign direct investment/Gross Regional Domestic Product (GRDP) [Source: PCI]	2021
5.3.3. Enterprises with science, technology, and innovation activities [Source: PCI]	2020
5.3.4. Proportion of enterprises with ISO certification [Source: General Department of Standards and Quality]	2021
5.3.5. Digital economy [Source: DTI]	2021
6. Knowledge products, creativity, and technology	
6.1. Intellectual creativity	
6.1.1. Patent applications/10,000 population [Source: National Office of Intellectual Property]	2021
6.1.2. Trademark applications/10,000 population [Source: National Office of Intellectual Property]	2021
6.1.3. Plant variety applications/10,000 population [Source: National Office of Intellectual Property]	--
6.2. Intangible assets	
6.2.1. Trademark registration applications/10,000 population [Source: National Office of Intellectual Property]	2021
6.2.2. Industrial design registration applications/10,000 population [Source: National Office of Intellectual Property]	2021
6.2.3. Number of granted geographical indications protection certificates [Source: National Office of Intellectual Property]	2021
6.3. Knowledge dissemination	
	2021

Index	Year of data
3.2.3. Environmental management [Source: PAPI]	2021
4. Market development level	
4.1. Finance and investment	
4.1.1. Credit in the private sector [Source: Central State Bank]	2021
4.1.2. Microfinance/Gross Regional Domestic Product (%) [Source: Provided by provinces]	2021
4.1.3. Results of local science and technology development fund activities [Source: Provided by provinces]	2021
4.2. Support services	
4.2.1. Number of specialized S&T service enterprises/1,000 enterprises [Source: GSO]	2021
4.2.2. Number of organizations supporting in standards, measurement, and quality/1,000 enterprises [Source: General Department of Standards and Quality]	2021

Index	Year of data
6.3.1. Scientific publications / total number of R&D tasks [Source: Provided by the Provinces]	
6.3.2. Number of newly established enterprises/10,000 population [Source: GSO]	2021
6.3.3. Number of S&T and qualified enterprises as S&T enterprises/1,000 enterprises [Source: Provided by Provinces]	2021
7. Impacts	
7.1. Impact on production and business	
7.1.1. Industrial production index [Source: GSO]	2021
7.1.2. Number of OCOP products/total number of administrative commune level units [Source: Provided by provinces]	2021
7.2. Socio-Economic Impacts	
7.2.1. Poverty rate [Source: GSO]	2021
7.2.2. Ratio of 15 and over 15 aged workers employed /total population [Source: GSO]	2021
7.2.4. Average income per capita [Source: GSO]	2021

Source: Compiled by the authors

Data Processing and Analysis Steps

(i) *Data Cleaning*: Remove indices with no data; Check and adjust locally provided data for unit consistency, year discrepancies, etc.

(ii) *Data Availability Check*: Analyze data availability with a threshold of 66% for each index and 70% for each locality for inclusion into calculations.

(iii) *Data Distribution Check*: Check skewness and kurtosis of data to identify potential outliers using thresholds (Skewness ≤ 2.25 or Kurtosis ≤ 3.5); Assess correlations between indices within and across groups and columns in the PII index using Pearson correlation coefficient.

(iv) *Missing Data Imputation*: the method of imputation of missing data for each index used in PII testing is min-value imputation, which is different from the

method of GII (shadow imputation). Accordingly, the missing value of a locality will be assigned the value of the locality with the lowest result. This is to encourage localities that do not have data to try to improve their data status in the next versions.

(v) *Data Adjustment According to Reference Base*: Divide processed data by their reference base to enhance comparability among provinces and localities.

(vi) *Outlier Treatment*: Employ Winsorization (max 5 points) for indices with skewness > 2.25 or kurtosis > 3.5. If skewness or kurtosis remains problematic after Winsorization, outliers are handled through Box-Cox transformation (Logarithm).

(vii) *Data Normalization and Component Index Ranking*: Utilize Min-Max normalization method on a scale of 0-100 for ranking; It is like the GII approach.

(viii) *Weighting*: Like GII, PII pilot testing also uses the average weight for each level (level of the index). This means that the component indices in an index group will have the same weight. The weight of the index group will be the sum of the weights of the component indices and similarly, the weight of the pillar will be equal to the sum of the weights of the index groups in it, etc. Finally, the total weight of the test PII will be equal to the weight of the innovation input and output index (0.5) and equal to 1.

(ix) *Aggregation Score Calculation and localities Ranking*: Due to that, the PII test uses the average weight between component indices in an index group, between index groups in a pillar... Score The number of index groups will be equal to the average of the scores of its component indexes, and the pillar's score will be equal to the average of the scores of the pillar groups within it.

(x) *Result Verification and Cross-Checking*: Verify correlation levels using Pearson's Correlation coefficient between indices:

- Within sub-columns;
- Between sub-columns within a main column;
- Between main columns;
- Between input and output indices of the PII;
- Check internal consistency using multivariate analyses: Cronbach's Alpha and Principal Component Analysis (PCA).

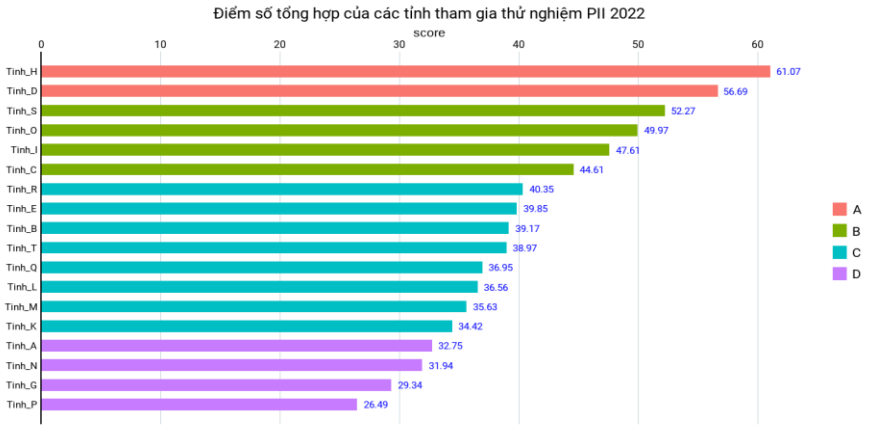
(xi) *Sensitivity and Specificity Testing*: Assess sensitivity and specificity through input factors like assumptions regarding missing data imputation, normalization methods, weight variability, and scoring aggregation methods.

The Provincial Innovation Index (PII) was constructed using appropriate methods and a repeatable analytical process through R software and the COINr

analysis package. This allows for transparent analysis, facilitates replication, and sets the stage for future updates and improvements. The processing and calculations followed a methodical approach at each step, from data availability checks to normalization, outlier handling, and synthesis.

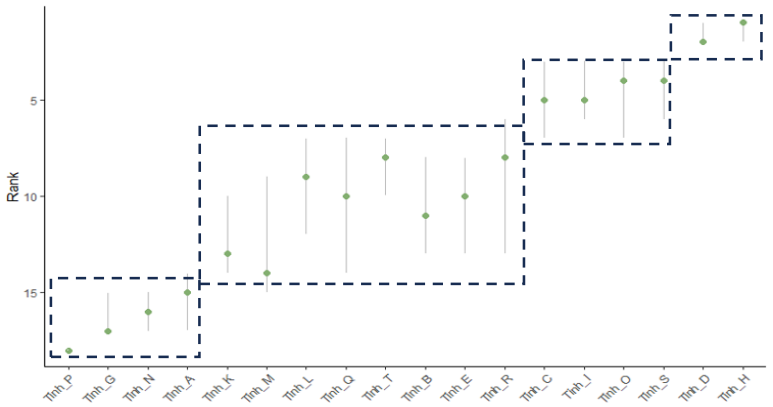
3. Pilot's assessment results

With the data processing, analysis and calculation methods implemented, the pilot's assessment and ranking of localities participating in the test are presented in Figure 3 below. Because this is an experimental assessment, we do not publish the ranking order of localities to avoid misunderstandings.



Source: Compiled by the authors

Fig 3. Score and ranking of the pilot's assessment PII



Source: Compiled by the authors

Fig 4. Sensitivity analysis results of PII 2022

According to the results of sensitivity analysis (range of variation) on the rankings of the tested localities presented in Figure 4, they can be divided into four groups and are shown in different colors in Figure 3. Accordingly, there are 2 localities are in the leading group, the second group has 4 localities, the 3rd group has 8 localities and the remaining group has 4 localities. Comparing with the socio-economic data of the localities participating in the experiment such as GRDP and economic structure, we find the results quite consistent. Localities with high GRDP and developed industries and services tend to have high rankings, on the contrary, localities with low GRDP and undeveloped industries and services tend to have low rankings.

Multivariate analysis indicates a high level of consistency in the dataset. However, there are some minor correlation issues that may warrant further examination in the future, especially when all localities are included in the calculation (detailed in the Discussion section below). A meticulous sensitivity analysis reveals that the PII results are robust, making it a reliable tool for reference and decision-making purposes.

4. Discussion

Regarding the Framework and Component Indices

Most of the indicators reflect the status of the measured object. However, there are still a few indicators that need to be considered for adjustment in the following years, including:

(i) It is crucial to consider adjusting or replacing the index representing the "Number of high school students per teacher" to better reflect the quality of education. Current data shows that in remote, rural, and mountainous areas, the ratio of high school students to teachers is low, but this does not necessarily correlate with high-quality education;

(ii) The index related to the ratio of high schools offering STEM/STEAM education needs reconsideration. According to the guidelines of the Ministry of Education and Training, STEM education activities encompass lectures, extracurricular activities, and scientific and technical research competitions. Consequently, the data reveals that most localities have a very high ratio of high schools offering STEM/STEAM education, with minimal variation between provinces and regions;

(iii) Considering the high correlation between the indices of patent registrations and utility models application registrations, it may be worthwhile to combine these two indices. Similarly, the indices related to the proportion of businesses engaged in R&D and those engaged in Intellectual Property activities show high similarity; hence, a consolidation or selection of a single index may be more appropriate;

(iv) The index reflecting the performance of Provincial Research and Development funds should be reviewed for potential removal. Currently, many local funds are inactive due to objective reasons, including inappropriate regulations and existing policies issued by the central government, making implementation difficult;

(v) The index of published scientific papers per total Research and Development tasks needs re-evaluation or replacement due to unavailable data. Localities would have to aggregate, but the data may not be comprehensive, and there is a lack of consistency in reporting various products and publications among different regions.

Data Source and Data Collection Method:

Indicators with available data account for a large proportion, and data availability is generally very high. Approximately 70% of the indices rely on secondary data obtained from statistical agencies, composite indices, and from central-level organizations, ensuring data consistency and reliability. However, for indices based on data provided by provincial localities, it is crucial to implement comprehensive training and specific guidelines for all regions in subsequent years. A robust process for receiving, verifying, and validating data must be established and rigorously executed to ensure accuracy, objectivity, and transparency.

There are still 03 indices without available data, including the R&D expenditure by businesses, the number of research personnel in businesses, and the number of registered plant varieties. The two indices related to R&D in businesses are essential indicators of investment and human resources in R&D within businesses. The index regarding the registration of plant varieties is a crucial output, particularly considering the significant role of agriculture in Vietnam's economy and various localities. Therefore, there is a need to organize data collection and statistics for these indices in the coming years. While lacking data, alternative indices should be researched.

Data Processing, Analysis, Score Calculation, and Ranking methods:

Consideration can be given to additional methods for handling outliers, normalization based on ranking, and correlation analysis between component indices and other composite indices. Specifically:

- Handling missing data might benefit from considering the GII technique (excluding the index from calculation and ranking). The current approach, using the value of the locality with the lowest result for those without data, might be perceived as unfair and inaccurate;
- Standardized data (used as denominators for assessment) should reflect the most accurate measurement of the index, for example, using the total population per 10,000 people, the total number of businesses per 1,000

businesses, or the total number of businesses in the manufacturing sector, total local budget or GRDP...;

- Applying inverse values for the microfinance index should be considered. The analysis indicates that localities with low microfinance to GRDP ratios are those with many businesses, medium and large business scales (labor, capital), indicating no need (and not being the target) for microfinance. Conversely, economically underdeveloped provinces with many poor households tend to have more access to microfinance. Therefore, this index needs to use inverse calculation. Localities with low microfinance to GRDP ratios should be evaluated more favorably, and vice versa.

5. Conclusion

The framework and component indices of the 2022 PII pilot's assessment have been meticulously selected, aligning with the GII structure. However, it is not an exact replica of GII to suit the local context of Vietnam, aligning both theoretically with the national science and technology development system and statistically. The PII columns closely follow GII, including five input columns, two output columns with index groups, and component indices as per GII's design.

The 2022 PII pilot's assessment adhered to international standards in constructing the composite index. As a result, the test outcomes demonstrate that PII is robust, making it become a suitable tool for leadership at various levels as a reference and basis for decision-making purpose.

Both domestic and international experiences in constructing a composite index and the results of the PII index pilot's test in 2022 indicate the need for a periodic review, examination, and adjustment of the index framework and component indices before nationwide implementation. No index can achieve perfection in its first design. Conversely, annual reviews and adjustments are necessary to adapt to the context and measurement objectives, following a common practice in developing composite indices.

Moreover, data processing techniques and calculation methods need continuous scrutiny to choose the most appropriate ones.

In terms of data sources, a maximum preference should be given to reliable hard data (statistical data) from statistical agencies and state management bodies at the central level. The use of data provided by local authorities should be limited to ensure the objectivity and uniformity of the data, as like the approach in GII./.

REFERENCES

1. Nguyen Thi Phuong Mai (2020). "A study proposing a framework for the ministry, industry, and localities to improve Vietnam's Global Innovation Index (GII) and the potential

application of the GII method to assess innovative capabilities at the provincial level." Research project at the Ministry level for 2018-2019. National Institute of Science and Technology Policy and Strategy. Hanoi.

2. OECD/JRC (2008). Handbook on Constructing Composite Indicators: Methodology and User Guide. OECD. EC.
3. WIPO (2016). The Global Innovation Index 2016. Winning with Global Innovation. WIPO, Geneva.
4. Becker, W.M., Saisana, P. Paruolo. and Vandecasteele (2017). "Weights and Importance in Composite Indicators: Closing the Gap". Ecological Indicators.
5. Carvalho, N, Carvalho, L, Nunes, S (2015). "A Methodology to measure innovation in European Union Through the National Innovation System". International Journal of Innovation and Regional Development. Vol 6. No.2, 2015.
6. Cherchye, L. et al. (2008). "Creating Composite Indicators with DEA and Robustness Analysis: The case of the Technology Achievement Index". Journal of Operational Research Society 59: 239-51.
7. Kotsemir, M. (2013). "Measuring National Innovation Systems Efficiency - A Review of DEA Approach". Working Paper. Basic Research Program. Series: Science, Technology and Innovation WP BRP 16/STI/2013. National Research University. Higher School of Economics.
8. Saisana, M. et al (2005). "Uncertainty and Sensitivity Analysis Techniques as Tools for the Analysis and Validation of Composite Indicators". Journal of the Royal Statistical Society A 168 (2): 307-23.