

A Service of

ZBW

Leibniz-Informationszentrum Wirtschaft Leibniz Information Centre for Economics

Xu, Yonghong; Su, Bingjie; Pan, Wenjie; Zhou, Peng

# Working Paper A high-frequency digital economy index: Text analysis and factor analysis based on big data

Cardiff Economics Working Papers, No. E2024/11

**Provided in Cooperation with:** Cardiff Business School, Cardiff University

*Suggested Citation:* Xu, Yonghong; Su, Bingjie; Pan, Wenjie; Zhou, Peng (2024) : A high-frequency digital economy index: Text analysis and factor analysis based on big data, Cardiff Economics Working Papers, No. E2024/11, Cardiff University, Cardiff Business School, Cardiff

This Version is available at: https://hdl.handle.net/10419/309960

#### Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

#### Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



# WWW.ECONSTOR.EU

# **Cardiff Economics Working Papers**



# Working Paper No. E2024/11

# A High-Frequency Digital Economy Index: Text Analysis and Factor Analysis based on Big Data

# Yonghong Xu, Bingjie Su, Wenjie Pan and Peng Zhou

# April 2024

## ISSN 1749-6010

Cardiff Business School Cardiff University Colum Drive Cardiff CF10 3EU United Kingdom t: +44 (0)29 2087 4000 f: +44 (0)29 2087 4419 business.cardiff.ac.uk

This working paper is produced for discussion purpose. These working papers are expected to be published in due course, in revised form. Cardiff Economics Working Papers are available online from: http://econpapers.repec.org/paper/cdfwpaper/ and https://carbsecon.com/wp/index abstracts.html

Enquiries: EconWP@cardiff.ac.uk

# A High-Frequency Digital Economy Index: Text Analysis and Factor Analysis based on Big Data

Yonghong Xu<sup>a</sup>, Bingjie Su<sup>a</sup>, Wenjie Pan<sup>a</sup>, Peng Zhou<sup>b,\*</sup>

a School of Economics, Xiamen University, China

b Cardiff Business School, Cardiff University, UK

\* Corresponding author. Email: <a href="mailto:zhoup1@cardiff.ac.uk">zhoup1@cardiff.ac.uk</a>

**Abstract:** We propose a high-frequency digital economy index by combining official white papers and big data. It aims to resolve the discrepancy between the new economic reality and old economic indicators used by decision-makers and policymakers. We have demonstrated a significant effect due to keyword rotations on the indices. Further analysis of the Dagum-Gini coefficient shows that spatial heterogeneity and temporal variation of the digital economy indices can be mainly attributed to between-group inequality.

**Keywords:** Digital Economy; High-Frequency Index; Big Data; Text Analysis; Hierarchical Dynamic Factor Model

### JEL codes: O33; O53; C38

**Acknowledgement:** This paper receives support from the National Social Science Fund of China (Major Project No.: 23&ZD124).

### 1 Introduction

The digital economy is not merely a trend, but also a tangible reality. The pervasive and transformative nature of digital technologies has promoted industrial agglomeration on the firm side (Zheng et al., 2024) and income convergence on the worker side (Gao & Li, 2023). Different from the traditional economy, where land, capital, and labor are the primary factors of production, the digital economy places data, information, and knowledge as central inputs. However, these fundamental changes have not yet been fully captured in economic indicators used by decision-makers and policymakers. Mainstream indicators are still those developed under the traditional economy such as aggregate price indices (Charemza & Husssain Shah, 2013).

There are two major problems arising from the discrepancy between the economic reality and economic indicators. On the one hand, the digital economy is characterized by high-frequency changes, which are better captured on a weekly, daily, and even second-to-second basis. In contrast, traditional economic indicators are typically reported on an annual, quarterly, or at best, a monthly basis (Zhou & Dixon, 2019). Low-frequency indicators cannot reflect the high volume, velocity, volatility, and veracity features of the digital economy. As a result, decision-making and policymaking relying on traditional indicators may be subject to a lack of accuracy and timeliness.

On the other hand, the digital economy is constantly evolving in its constituent elements and underlying technologies. For example, the most searched terms related to the digital economy in the 2000s were "internet" and "e-commerce". In the 2010s, terms like "social media", "big data", and "blockchain" became more prominent. Entering the 2020s, "artificial intelligence", "IoT", and "large language model" surged as the most popular search terms. Nevertheless, traditional economic indicators are usually built on relatively stable keyword thesauri identified in literature, annual business reports, and government statements (Zhou et al., 2021). This method often overlooks the dynamic nature of the digital economy.

These drawbacks of traditional indicators can be partly addressed by popular search engine indices like Google Trends and Baidu Index (Fang et al., 2020), but a comprehensive indicator for the entire digital economy is still missing. It is like we can access data on each stock price, but there is no index for the whole stock market. To fill this gap, we propose a high-frequency digital economy index based on big data techniques which are in line with the intrinsic features of the digital economy. The procedure starts with an exploratory text analysis of official white papers, combined with daily search data. The hierarchical dynamic factor model (DHFM) is then applied to extract digital economy indices for any spatial and temporal units. Our first contribution is methodological—to construct a high-frequency index by combining official data and big data.

The second contribution is empirical. We apply various statistical tools to analyze spatial heterogeneities and temporal changes in the digital economy indices. The generalized Gini coefficient (the Dagum-Gini coefficient) is utilized to quantify and decompose the unbalanced distribution of the indicator. We find that between-group inequality is the most important contributor to the regional imbalance.

This paper uses China as the context as it has an important condition for the development of the digital economy—a large population to explore the scale effect. Another advantage of China is that most digital economy terminologies are standardized by official white papers published by the China Academy of Information and Communications Technology (CAICT). It enables us to combine trends identified by the authority ("seed keywords") and reflected by the big data ("hot keywords") in constructing the index, analogous to the Bayesian philosophy of incorporating the "prior" information into the "posterior".

### 2 Method

There are generally two types of methods for index construction in terms of data source. One is based on primary data in the forms of surveys, questionnaires, and interviews. Most traditional indices belong to this type (e.g., price index, consumer confidence index). The other is based on secondary data such as stock quotes, annual reports, and official statements (e.g., stock index, geopolitical risk index). Our method belongs to the latter, but its novelty is to make use of the big data (internet searches) accumulated in the digital economy, so it can cover a higher frequency with lower data collection costs. The high-frequency feature can better capture the temporal variation of the digital economy across different periods.

This type of index is usually constructed by estimating a dynamic factor model, where the extracted factors represent the unobserved trends or cycles in the observed variables. Traditional dynamic factor models, such as Markov-switching dynamic factor models (MS-DFM), factor-augmented VAR models (FAVAR), and mixed-frequency dynamic factor models (MF-DFM), are restricted to single-layer factors. In contrast, DHFM allows for hierarchical transmissions among dynamic factors. This multi-layer feature can better capture the spatial heterogeneity of the digital economy at province, region, and country layers.

Our index is formulated through a five-stage process in Figure 1. Details of each step are elaborated in Supplementary Material. Specifically, in step (v), we apply the DHFM to capture the hierarchical structure and dynamic changes of the index:

$$Z_{ijkt} = \lambda_{ijk}^P P_{jkt} + \epsilon_{ijkt}^Z \tag{1}$$

$$P_{jkt} = \lambda_{jk}^R R_{kt} + \epsilon_{jkt}^P \tag{2}$$

$$R_{kt} = \lambda_k^F F_t + \epsilon_{kt}^R \tag{3}$$

 $Z_{ijkt}$  is the observed Baidu Index of keyword *i* in province *j* of region *k* at time *t*. It is explained by province-layer factors  $P_{jkt}$ , which is then explained by region-layer factors  $R_{kt}$  and further by country-layer factors  $F_t$ . The vector  $\boldsymbol{\lambda} \equiv [\lambda_{ijk}^P, \lambda_{jk}^R, \lambda_k^F]$  contains factor loadings, and the error vector  $\boldsymbol{\epsilon} \equiv [\epsilon_{ijkt}^Z, \epsilon_{jkt}^P, \epsilon_{kt}^R]$  follows a VAR  $\boldsymbol{\epsilon}_t = \boldsymbol{\Psi} \boldsymbol{\epsilon}_{t-1} + \boldsymbol{\xi}_t$ , where  $\boldsymbol{\xi}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma})$ . The DHFM (1)-(3) is estimated using the MCMC algorithm (see Supplementary Material).



Figure 1 Flowchart of the construction of the digital economy index

#### **3** Results

Figure 2 presents the high-frequency (daily) digital economy indices. Panel A compares the indices with and without keyword rotations based on the level criterion, while Panel B follows the growth criterion. It shows that the growth-based index is more sensitive to keyword rotations than the level-based index. If we fix the keywords as in the 2014 list, then the growth-based indices seem to diverge after 2020—the digital economy index with fixed keywords is stagnant, while the index with keyword rotations keeps growing. Without accounting for keyword rotations, we would have neglected the continued growth in the digital economy. It implies that indices with keyword rotations are preferred if we want to capture the dynamics of the digital economy.



Figure 2 High-frequency digital economy indices and time-varying correlation

To confirm the effect of keyword rotations, we estimate a MGARCH model with dynamic conditional correlation (DCC). The implied time-varying correlation between the indices (with and without rotations) is stable for the level-based indices apart from the COVID-19 outlier (Panel C). The growth-based indices, in contrast, show a diminishing correlation over time (Panel D). Structural breaks are also detected by the Chow test and BP test (Supplementary Material).

Focusing on the indices with keyword rotations, we can summarize two patterns. First, there is a rising trend of the digital economy indices. The growth-based index increases at a faster pace because it selects the positive-growth keywords. Second, there is a cyclical feature due to weekly, monthly, and yearly fluctuations in the digital economy. Therefore, it is critical to construct high-frequency indices to understand these high-frequency changes.

#### 3.1 Spatial Heterogeneity and Temporal Variation

If we separately plot the indices by provinces and over time (Figure 3), it demonstrates substantial spatial heterogeneity and significant temporal variation.



Figure 3 Indices by provinces and over time

There is a significant spatial imbalance in the digital economy index. Provinces located in key economic zones or with better infrastructure and access to trade routes may receive more attention due to their strategic importance. Coastal provinces, for example, might initially garner more interest in digital economy development compared to inland provinces.

However, over time, the market attention has shifted from coastal provinces (panels A and C) to inland provinces (panels B and D). It reflects the pattern of knowledge diffusion in China: as the digital economy matures and technology becomes more pervasive, attention may shift from early adopter regions to other areas seeking to catch up or diversify their economic base. This evolution over time contributes to changes in the spatial distribution of the index.

## 3.2 Dagum-Gini Coefficient

Building on the evidence in the previous subsection, we attempt to further quantify the spatial imbalance and temporal variation of the digital economy indices by the Dagum-Gini coefficient

(*D*). It allows for greater flexibility in the underlying distribution, rendering the standard Gini coefficient (*G*) a special case.

$$D \equiv \frac{\pi}{2} \cdot \frac{a}{a+b} \cdot G \tag{4}$$

In the formula, G is the standard Gini coefficient computed from the density function of the following Dagum distribution (a is the scale parameter, b is the location parameter, and p is the shape parameter):

$$f(x; p, a, b) = \frac{ap(a+b)\left(\frac{x}{b}\right)^{a-1}}{\left(1+\left(\frac{x}{b}\right)^{a}\right)^{p+1}}, \text{ where } x \text{ is the underlying random variable.}$$
(5)

The Dagum-Gini coefficient is particularly advantageous when dealing with distributions characterized by heterogeneity. It can be decomposed into three generalized entropy components, which help to understand the sources of the spatial imbalance and temporal variation of the index: (i) the "within-group inequality" component measures the contribution of intra-province inequality to the overall inequality, (ii) the "between-group inequality" component measures the contribution of inter-province inequality, and (iii) the "intensity of transvariation" component measures the cross-impact among provinces (Dagum, 1997).

#### Figure 4 Dagum-Gini coefficients





Figure 4 presents the historical decomposition of the computed Dagum-Gini coefficients over time. It suggests that the spatial imbalance of the digital economy index remains stable (about 0.27) in both level- and growth-based measures. The between-group inequality contributes the most share (about 50%) of the Dagum-Gini coefficients, especially after the COVID-19 pandemic. It suggests that the spatial imbalance of the digital economy index is persistent. More-over, the importance of between-group inequality can be reinforced by exogenous shocks like the pandemic—knowledge diffusion can be hindered by lockdowns. In contrast, the contribution of within-group inequality is less important (about 24%) but remains stable throughout the sample period. The contribution of intensity of transvariation fluctuates and diminishes after the pandemic—a pattern we have seen in the contribution of between-group inequality. This is because both between-group inequality and intensity of transvariation are influenced by barriers of inter-regional knowledge diffusion.

### 4 Conclusion

This paper proposes a high-frequency, multi-layer digital economy index, which integrates official white papers and big data. We have shown that keyword rotation is critical to capture the dynamics of the digital economy. Analysis shows evidence for spatial heterogeneity and temporal variation of the indices. Decomposition of the Dagum-Gini coefficients suggests that the between-group inequality is the key to explaining the regional imbalance in the digital economy index. Decision-makers in businesses can use this high-frequency index to quickly react to market opportunities in the digital economy. Policymakers in governments can use this multi-layer index to monitor spatial imbalances and design developmental strategies for different geographic levels.

Despite its advantages in timeliness and informativeness, the proposed index has a notable limitation in that it only reflects the demand side (web searches) of the digital economy. It does not, however, describe the supply side (provision capability). Future research can collect microdata from firms to identify market gaps for businesses and monitor undeveloped areas for governments.

#### References

- Charemza, W., Husssain Shah, I. 2013. "Stability price index, core inflation and output volatility." *Applied Economics Letters*, 20(8), 737-741.
- Dagum, C. 1997. "A new approach to the decomposition of the Gini income inequality ratio." *Empirical Economics* 22(4): 515-531.
- Fang, J., Gozgor, G., Lau, C.-K. M., Lu, Z. 2020. "The impact of Baidu Index sentiment on the volatility of China's stock markets." *Finance Research Letters* 32: 101099.
- Gao, X., Li, M.-S. 2023. "The impact of the digital economy on the urban-rural income gap: evidence from provincial panel data in China." *Applied Economics Letters*.
- Zheng, Y., Yao, S., Fu, F. 2024. "Digital economy, industrial agglomeration and high-quality development". *Applied Economics Letters*.
- Zhou, B., Zhang, Y., Zhou, P. 2021. "Multilateral political effects on outbound tourism." *Annals of Tourism Research*, 88: 103184.
- Zhou, P., Dixon, H. 2019. "The Determinants of Price Rigidity in the UK: Analysis of the CPI and PPI Microdata and Application to Macrodata Modelling." *The Manchester School*, 87(5): 640-677.

#### **Supplementary Materials**

#### The Method:

Our digital economy index is formulated through a five-stage process as presented in Figure 1.

- (i) **Dictionary compilation.** Perform text analysis of the official white papers published by CAICT (2015-2021), resulting in a crude list of separated words. These separated words are then manually screened to identify an initial list of seed keywords related to digital economy (N = 297).
- (ii) Seed keyword selection. Conduct a further round of text analysis based on word frequency statistics. We exclude keywords appearing only once and those unrelated to digital economy. This results in the final list of seed keywords (N = 60).
- (iii) Data collection and preprocessing. Use Python web crawler to collect daily search data for each province from Baidu Index. We distinguish between the level-based list (search volumes within a period) and the growth-based list (growth rates of search volumes over periods) of searches.
- (iv) Hot keyword identification. For the level-based searches, keep the top 70% searches as the level-based hot keywords. For the growth-based searches, keep the positive searches as the growth-based hot keywords.
- (v) **Index synthesis.** The synthesis starts with the intersection of seed keywords from official white papers in (ii) and hot keywords from social media searches in (iv). This way, the constructed indices account for both reliable "prior" information and the big data. To capture the hierarchical structure and dynamic changes of the index, we apply the DHFM:

$$Z_{ijkt} = \lambda_{ijk}^{P} P_{jkt} + \epsilon_{ijkt}^{Z}$$
(6)

$$P_{jkt} = \lambda_{jk}^{R} R_{kt} + \epsilon_{jkt}^{P} \tag{7}$$

$$R_{kt} = \lambda_k^F F_t + \epsilon_{kt}^R \tag{8}$$

 $Z_{ijkt}$  is the observed Baidu Index of keyword i in province j of region k at time t. It is explained by province-layer factors  $P_{ikt}$ , which is then explained by region-layer factors  $R_{kt}$  and further by country-layer factors  $F_t$ . The vector  $\boldsymbol{\lambda} \equiv [\lambda_{ijk}^P, \lambda_{jk}^R, \lambda_k^F]$  contains factor loadings, and the error vector  $\boldsymbol{\epsilon} \equiv \left[\epsilon_{ijkt}^{Z}, \epsilon_{jkt}^{P}, \epsilon_{kt}^{R}\right]$  follows a VAR  $\boldsymbol{\epsilon}_{t} = \boldsymbol{\Psi}\boldsymbol{\epsilon}_{t-1} + \boldsymbol{\xi}_{t}$ , where  $\boldsymbol{\xi}_{t} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma})$ . The model (1)-(3) is estimated using the following Markov Chain Monte Carlo (MCMC) algorithm:

- 1. Apply Principal Component Analysis to obtain initial values of  $P_{ikt}^{(0)}$ ,  $R_{kt}^{(0)}$ ,  $F_t^{(0)}$  and  $\lambda^{(0)}, \Psi^{(0)}, \Sigma^{(0)}.$
- 2. Simulate  $P_{jkt}^{(1)}$  based on  $R_{kt}^{(0)}, F_t^{(0)}, \lambda^{(0)}, \Psi^{(0)}, \Sigma^{(0)}$ . 3. Simulate  $R_{kt}^{(1)}$  based on  $P_{jkt}^{(1)}, F_t^{(0)}, \lambda^{(0)}, \Psi^{(0)}, \Sigma^{(0)}$ .

- 4. Simulate  $F_t^{(1)}$  based on  $R_{kt}^{(1)}$ ,  $P_{jkt}^{(1)}$ ,  $\lambda^{(0)}$ ,  $\Psi^{(0)}$ ,  $\Sigma^{(0)}$ . 5. Estimate  $\lambda^{(1)}$ ,  $\Psi^{(1)}$ ,  $\Sigma^{(1)}$  based on  $R_{kt}^{(1)}$ ,  $P_{jkt}^{(1)}$ ,  $F_t^{(1)}$ .
- 6. Repeat steps 2-5 until the Markov Chain converges.

We conduct 100,000 simulations and the initial 50,000 are dropped as burnt-in samples. We then keep one result for every 50 simulations, resulting in 1,000 samples. The estimated factors  $[P_{jkt}, R_{kt}, F_t]$  are the digital economy indices of the province, region, and country layers.

Data and codes can be found here: OneDrive Link (200MB).

### Identified Lists of Keywords

Based on the 5-stage process presented in Figure 1, we have identified different lists of keywords in different times. Table S1 shows the evolving keywords of selected years.

Year	Level-based Growth-based			
2014	ICT, information technology, cloud computing,	information technology, data, informatization,		
	Internet of Things (IoT), data, informatization,	IT industry, knowledge economy, cloud com-		
	knowledge economy, intelligent, mobile Inter-	puting, big data, intelligent, Internet finance,		
	net, sharing	industrial robots		
2018	blockchain, artificial intelligence (AI), Internet	digital economy, high-quality development,		
	of Things (IoT), big data, ICT, cloud compu-	blockchain, ICT, data, digital transformation,		
	ting, digital economy, intelligence, information	intelligent, data centers, convergence, Indus-		
	technology, data	trial Internet		
2022	digital economy, big data, data, ICT, artificial	digital economy, digitization, intelligent, shar-		
	intelligence (AI), Internet of Things (IoT),	ing, platforms, regulation, digital transfor-		
	blockchain, cloud computing, digital transfor-	mation, High-quality development, data secu-		
	mation, digitization	rity, digital management		

More on Spatial Heterogeneity and Temporal Variation

Figure S1 and Figure S2 illustrate the level-based and growth-based digital economy indices for the four major regions of China (East, Northeast, Central, and West). They show a gradually increasing trend with less fluctuations from 2014 to 2022. One possible explanation is the introduction of nationwide policies on the digital economy. These policies include action plans for digital economy development, industrial strategies, and subsidies. Government-backed nationwide policies substantially reduced the fluctuations and promoted a stable, balanced growth of digital economy indices.



## Figure S1 Indices by regions and over time

• Tests of Structural Breaks

The following two tables report the detected number of structural breaks of the level-based and growth-based daily digital economy indices. This result supports that rotation of keywords is necessary in constructing a high-frequency digital economy index.

Period		BP Test 1			Chow Test		Model
	Country	Region	Province	Country	Region	Province	
2014-2015	0	1	9	1	4	31	Ι
2015-2016	0	1	3	1	2	15	Ι
2016-2017	1	0	3	1	1	17	Ι
2017-2018	0	1	7	1	3	28	Ι
2018-2019	1	1	3	0	2	9	Ι
2019-2020	0	0	2	1	3	10	Ι
2020-2021	0	0	0	1	2	18	Ι
2021-2022	0	0	0	0	1	0	Ι
2014-2015	0	1	9	1	4	31	I-AR(1)
2015-2016	0	0	2	1	2	15	I-AR(1)
2016-2017	1	1	4	1	1	17	I-AR(1)
2017-2018	0	1	6	1	3	28	I-AR(1)
2018-2019	1	1	3	0	2	9	I-AR(1)
2019-2020	0	0	2	1	3	10	I-AR(1)
2020-2021	0	0	0	1	2	18	I-AR(1)
2021-2022	0	0	0	0	1	0	I-AR(1)
2014-2015	0	0	9	1	4	31	I-AR(p)
2015-2016	0	0	1	0	0	11	I-AR(p)
2016-2017	0	0	1	1	0	20	I-AR(p)
2017-2018	0	1	0	1	3	23	I-AR(p)
2018-2019	0	0	1	0	0	3	I-AR(p)
2019-2020	0	0	2	0	0	3	I-AR(p)
2020-2021	0	0	0	0	0	13	I-AR(p)
2021-2022	0	0	0	0	0	0	I-AR(p)
2014-2015	0	1	9	1	4	31	I-VAR(1)
2015-2016	0	0	3	1	3	17	I-VAR(1)
2016-2017	1	1	4	1	1	15	I-VAR(1)
2017-2018	0	1	7	1	3	29	I-VAR(1)
2018-2019	1	1	3	0	2	10	I-VAR(1)
2019-2020	0	0	1	1	3	12	I-VAR(1)
2020-2021	0	0	0	1	2	18	I-VAR(1)
2021-2022	0	0	0	0	1	0	I-VAR(1)
2014-2015	0	1	6	1	4	31	I-VAR(p)
2015-2016	0	0	2	1	1	15	I-VAR(p)
2016-2017	0	0	4	1	1	25	I-VAR(p)
2017-2018	1	1	2	1	4	27	I-VAR(p)
2018-2019	0	2	2	0	2	4	I-VAR(p)
2019-2020	0	0	3	1	3	11	I-VAR(p)
2020-2021	0	0	0	0	2	19	I-VAR(p)
2021-2022	0	0	0	0	0	0	I-VAR(p)

Table S2 Number of breaks in the level-based daily indices

Period		BP Test 1			Chow Test		Model
	Country	Region	Province	Country	Region	Province	
2014-2015	1	3	21	1	4	30	Ι
2015-2016	0	0	0	1	4	17	Ι
2016-2017	0	0	1	0	0	3	Ι
2017-2018	0	1	0	1	3	17	Ι
2018-2019	0	3	2	1	0	6	Ι
2019-2020	1	1	4	1	3	8	Ι
2020-2021	0	1	2	1	2	7	Ι
2021-2022	0	0	0	1	1	4	Ι
2014-2015	0	1	13	1	4	30	I-AR(1)
2015-2016	0	0	0	1	4	18	I-AR(1)
2016-2017	0	0	2	0	0	2	I-AR(1)
2017-2018	0	2	0	1	3	17	I-AR(1)
2018-2019	0	2	3	1	0	6	I-AR(1)
2019-2020	1	1	3	1	3	8	I-AR(1)
2020-2021	0	2	3	0	1	7	I-AR(1)
2021-2022	0	0	0	1	1	5	I-AR(1)
2014-2015	0	0	11	1	4	31	I-AR(p)
2015-2016	0	0	0	1	1	7	I-AR(p)
2016-2017	0	0	1	0	0	1	I-AR(p)
2017-2018	0	1	0	0	3	14	I-AR(p)
2018-2019	0	1	1	0	0	3	I-AR(p)
2019-2020	1	1	0	0	1	7	I-AR(p)
2020-2021	0	0	1	0	1	1	I-AR(p)
2021-2022	0	0	0	0	0	0	I-AR(p)
2014-2015	1	4	17	1	4	30	I-VAR(1)
2015-2016	0	0	0	1	4	18	I-VAR(1)
2016-2017	0	0	1	0	0	4	I-VAR(1)
2017-2018	0	1	0	1	3	18	I-VAR(1)
2018-2019	0	2	1	1	0	6	I-VAR(1)
2019-2020	1	1	3	1	3	9	I-VAR(1)
2020-2021	0	2	3	0	1	7	I-VAR(1)
2021-2022	0	0	0	0	2	7	I-VAR(1)
2014-2015	1	4	13	1	4	31	I-VAR(p)
2015-2016	0	0	0	1	1	13	I-VAR(p)
2016-2017	0	0	2	0	0	1	I-VAR(p)
2017-2018	0	1	0	0	2	16	I-VAR(p)
2018-2019	0	0	1	0	0	4	I-VAR(p)
2019-2020	1	1	2	1	2	8	I-VAR(p)
2020-2021	0	0	1	0	1	6	I-VAR(p)
2021-2022	0	0	0	0	1	1	I-VAR(p)

Table S2 Number of breaks in the level-based daily indices