Liet. mat. rink. LMD darbai, **50**, 2009, 224–229

Composite indicators – methodology and practical aspects

Jurga RUKŠĖNAITĖ
Vilnius Gediminas Technical University
Saulėtekio 11, 10223 Vilnius, Lithuania
e-mail: jurga.ruksenaite@fm.vgtu.lt

Abstract. In this paper methodology of construction Composite Indicators is described. Data standardisation and ranking procedures are proposed. Results of modelling irregular component using different ARIMA models are described.

Keywords: composite indicators, data normalisation, ranking procedure, ARIMA model.

1. Introduction

Composite Indicators (CI) which compare country performance are recognised as a useful tool in setting policy priorities and in benchmarking or monitoring performance and the use of CI around the world is growing year after year. Main features of CI are: CI can summarise complex, multi-dimensional indicators, these factors are easier to interpret than collection of individual indicators. *Composite Indicators* (CI) – given mathematical function-model, composed using a set of individual indicators from complex different fields.

CI provide comparisons of countries that can be used to illustrate complex in some important fields, e.g., environment, economy or technological development. Some examples could be: Technology Achievement Index, Growth Competitiveness Index and other. CI is formed when individual indicators are compiled into a single index on the basis of an underlying model. The CI should ideally measure multidimensional concepts which cannot be captured by a single indicator, e.g., competitiveness, sustainability, knowledge-based society, etc.

The quality of Composite indicators depends not only on the methodology but also on the quality of the framework and the data used. Actually CI are not only economical more mathematical or computational models.

2. Methodology

A brief review (main steps) of construction CI is presented below [1]:

- *theoretical framework* (provides the basis for the selection and combination of variables into meaningful CI);
- *date selection* (based on the analytical importance, measurability, country coverage and relevance of the indicators);

– *pre-adjustment analysis of variables* (analysis of outliers, estimation of missing values);
– *multivariate analysis* (study of structure of the dataset and its suitability);
– *data normalisation* (procedure that enables to compare variables);
– *weighting and aggregation* (selection of appropriate procedures);
– *uncertainty and sensitivity analysis* (analysis of robustness of CI);
– *back to the data* (to reveal the reasons of good or bad performance of CI);
– *links to other indicators* (correlation with existing indicators, identifying linkage through regression);
– *visualisation of the results* (visualisation method/technique can influence interpretability of CI).

In this paper raised mathematical problems are mainly related to *data normalisation* step.

Mathematical problems:

1. Ranking procedure of different countries using some criteria.
2. Composition of economically stable/efficient country (country-standard) with the set of indicators.
3. Formation of statistically appropriate (optimal) model for each of indicators.

2.1. Normalisation of data

Normalisation is one of the steps for constructing CI. The main questions are: selection suitable normalisation procedure(s) that respect both the theoretical framework and the data properties; making scale adjustment; transformation highly skewed indicators. Different normalisation methods will produce different results for the CI. A number of normalisation methods exists [1, 3, 4]:

– ranking

$$I_{qc}(t) = \mathrm{Rank}\big(x_{qc}(t)\big), \tag{1}$$

– standardisation (or z-scores)

$$I_{qc}(t) = \frac{x_{qc}(t) - x_{qc=\overline{c}}(t)}{\sigma_{qc=\overline{c}}(t)}, \tag{2}$$

and other, where $x_{qc}(t)$ is the value of indicator q for country c at time t, $\overline{c}$ is the reference country.

In this paper *modified standardisation procedure* is recommended:

$$I_{qc}(t) = \frac{x_{qc}(t)}{F_{qc^*}(t)}, \tag{3}$$

where $F_{qc^*}(t)$ is a function-model of indicator q for artificial country-standard c^* at time t. $F(t)$ may be constructed using different mathematical techniques therefore using elimination of irregular component.

There is defined a time series $\{\xi(t), t \in T\}$ which may be decomposed into particular function $F(t)$ and irregular component $i(t)$:

$$\xi(t) = F(t) \cdot i(t). \tag{4}$$

For choosing statistically appropriate (optimal) model irregular component $i(t)$ is essential element which consists of additive outliers, transitive changes and random residuals. It indicates if the model is suitable for the analysis. The main characteristics:

$$cov(\varepsilon(t), \varepsilon(s)) = 0 \quad \text{if } t \neq s,\ \varepsilon(t) \sim N(0, \sigma^2). \tag{5}$$

Various criteria tests (criteria of models stability) for residuals, outliers and models' parameters in the analysis have been used:
Residual Autocorrelations (*ACF*), *Residual independence*, *Residual Normality*, *Residual Asymmetry*, *Kurtosis of Residuals*, *Test for the number of outliers* and other. Because of small size of the paper the criteria are introduced briefly, information can be founded in following article [2]. The suitability of ARIMA models is checked by *Akaike* (AIC) and *Bayesian* (BIC) criteria.

2.2. *Procedure of ranking – comparison of criteria*

There are defined a set of countries $N = \{1, 2, \ldots, i, \ldots, n\}$ and a set of indicators $M = \{1, 2, \ldots, j, \ldots, m\}$.

ASSUMPTION 1. *Country's economy is developing stable/effective during period* $T = \{1, 2, \ldots, t\}$ *if the collection of various indicators satisfies particular criteria.*

It depends on the mathematical task and can be chosen a set of criteria but in this paper only one criterion has been chosen.

Brief description on the Fig. 1.

In the first table there are put a set of values of criterion K_{ij} (i – country, j – indicator). For example in the first column there are values of criterion (indicator: 1) of different countries. In the second table the values of criterion (in every column separately) are arranged by increasing order. For example in the first column the value of criterion of first country is biggest and goes to the last position. After the ranking procedure all positions get a rank from 6 (if the value is minimum) to 1 (if the value is in 6th position), other – 0.

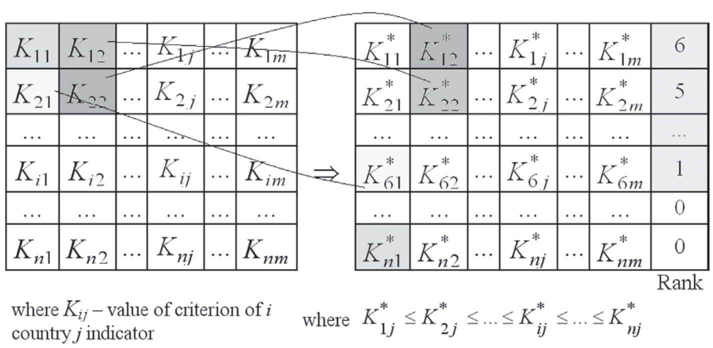

Fig. 1. Ranking procedure.

The rank of individual country i:

$$R_i = \sum_{j=1}^{m} r_{ij}, \tag{6}$$

where r_{ij} – rank of single criterion, indicator j.

3. Normalisation procedure in practise

In this section ranking procedure and analysis of irregular component are presented.

3.1. Composition of country-standard (ranking procedure)

For the selection of statistical data some criteria have been used: relevance and availability, economical importance and statistical characteristics of data and researches of other authors. In the paper economic-social annual statistical indicators of 27 European Union countries have been used: Gross domestic product (GDP), unemployment rate, annual inflation, investment level and labour productivity. The ranking criterion: standard deviation (σ) has been chosen.

Four countries with biggest evaluated individual ranks (6) have been selected: United Kingdom (rank $R_{UK} = 18$), Austria ($R_{AT} = 12$), Spain ($R_{ES} = 10$) and France ($R_{FR} = 10$). After the econometric analysis some similarities and consistency have been noticed, as an example in Fig. 2. (Scale shifting may influence the interpretation of results).

For the next analysis one indicator has been chosen: GDP with quarterly periodicity, 1997 – 2008. The indicator GDP of country-standard (c^*) has been selected as weighted average:

$$x_{GDPc^*}(t) = \sum_i w_{c_i} \cdot x_{GDPc_i}(t), \tag{7}$$

where t – time period, c_i – country i with given weight w_{c_i}.

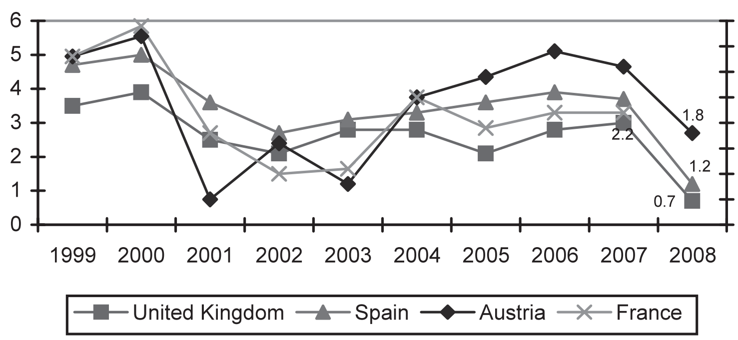

Fig. 2. Real GDP growth rate – percentage change on previous year.

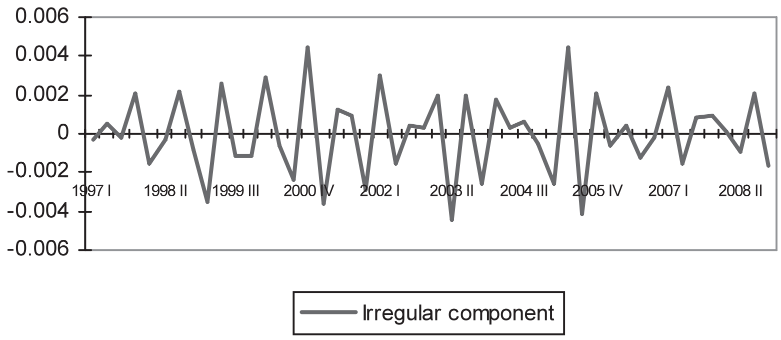

Fig. 3. Separated irregular componet.

Table 1. ARIMA models

(011)(111)	(021)(011)	(010)(011)
(111)(011)	(001)(001)*	(011)(012)
(111)(111)	(011)(001)*	(012)(011)
(211)(111)		(013)(011)
(210)(011)		(013)(012)
(211)(211)*		(023)(012)
(311)(211)*		

3.2. Modelling the irregular component

Procedure of modelling is divided into three steps.

First step. Software module TRAMO/SEATS has been used for the modelling and analysis time series [2]. Parametric model ARIMA$(p, d, q) \times (sp, sd, sq)$ has been applied for time series $x_{GDPc^*}(t)$ to decompose (4) and separate irregular component $i(t) \sim N(0, \sigma^2)$ (see Fig. 3).

Second step. Modelling 200 sequences: $\tilde{i}_k \sim N(0, \hat{\sigma}_k^2)$, $k = 1, 2, \ldots, 200$. Using the same function $F(t)$ from *first step*, time series are constructed:

$$\tilde{\xi}_k(t) = F(t) \cdot \tilde{i}_k(t), \tag{8}$$

Third step. Applying 12 different ARIMA models:

$$\xi_k^h(t) = F_k^h(t) \cdot i_k^h(t), \quad h = 1, 2, \ldots, 12. \tag{9}$$

The modelling has been started using ARIMA(011)(011) model and then changing number of parameters (Table 1). Some models have been rejected (*).

4. Results

After the application different ARIMA models, quality and stability of models (except rejected) have been checked using criteria tests, described in Section 2.1. A set of findings have been identified during the analysis.

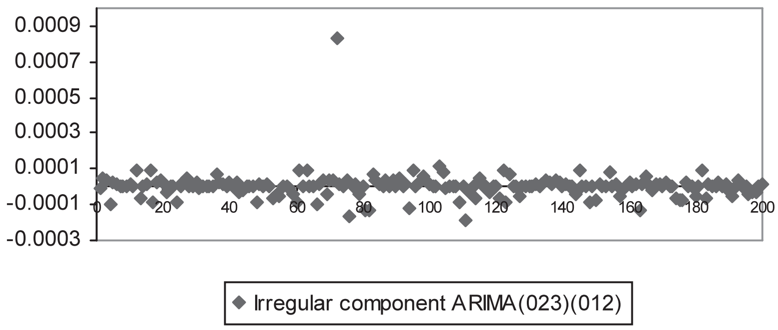

Fig. 4. Average of irregular component of model ARIMA(023)(012).

Gained three statistically appropriate ARIMA models: ARIMA(011)(011), ARIMA(021)(011), ARIMA(010)(011). The average of irregular component of every time series $\zeta_k^h(t)$ are absolute zero: $\mu_k = 0$. In this case functions $F_k(t)$ are stable in time t and can be used in data standardisation procedure (3).

Identified optimal parameter number of ARIMA model is $z \leqslant 5$, otherwise model's quality characteristics are decreasing.

Also identified, that often ARIMA models have insignificant parameters when number is $z > 5$ when finally models are accepted (see Fig. 4).

When $z \geqslant 8$ or integrated part $d \leqslant 1$ – ARIMA models are rejected. When parameter number of autoregressive part is $p > 0$ – quality characteristics of model are not increasing (see Table 1).

References

1. M. Freudenberg. *Composite Indicators of Country Performance*: *A Critical Assessment*. OECD, Paris, 2003.
2. V. Gomez, A. Maravall. *Demetra 2.0. User Manual*. Eurostat, 2002.
3. *Handbook on Constructing Composite Indicators*: *Methodology and User Guide*, 2008.
4. R. Jacobs, P. Smith, M. Goddard. *Measuring Performance*: *An Examination of Composite Performance Indicators*. Centre for Health Economics, Technical Paper Series 29, 2004.

REZIUMĖ

J. Rukšėnaitė. Sudėtiniai rodikliai – metodologijos ir praktiniai aspektai

Šiame straipsnyje yra aprašyta sudėtinių rodiklių konstravimo metodika. Pristatytos duomenų standartizavimo ir rangavimo procedūros. Aprašyti rezultatai, gauti modeliuojant nereguliariąją komponentę naudojat įvairius ARIMA modelius.

Raktiniai žodžiai: sudėtiniai rodikliai, duomenų normalizavimas, rangavimo procedūra, ARIMA modelis.