A C T A U N I V E R S I T A T I S L O D Z I E N S I S
FOLIA OECONOMICA 269, 2012

*Piotr Białowolski**

APPLICABILITY OF THE MULTI-GROUP CONFIRMATORY FACTOR ANALYSIS TO CONSTRUCTION OF BUSINESS SENTIMENT INDICATORS

Abstract. The paper presents arguments that advocate for application of the multi-group confirmatory factor analysis as a tool for constructing sentiment indicators in business surveys. Reliable measurement and comparisons of the sentiment mean between periods require measurement invariance on its three basic levels-configural, metric and scalar invariance. It is hypothesized that only sets of questions that are internally coherent can serve as a group of proxies for business sentiment indicator. An attempt to construct two different sentiment indicators for manufacturing industry is performed. The results show that only for the set of coherent proxies it is possible to estimate model with measurement invariance.
Key words: Confirmatory Factor Analysis, Sentiment Indicators.

I. INTRODUCTION

Standard sentiment measures in business and consumer tendency surveys are constructed as an arithmetic mean of balances of answers to the certain group of question (European Commission 2006). This approach leaves however a vast field for discussion in the problem of measurement, as there is a possibility that the sentiment index constructed in line with this methodology does not reflect any unidimensional phenomenon (sentiment). Even if it is assured that the measured phenomenon is unidimensional in a given period, its unidimensionality might not be maintained in all periods of analysis and, additionally, the understanding of the concept underlying the answers to given set of questions serving as proxies of the sentiment measure might evolve in time. In such situation, comparisons of the values of the sentiment index should be perceived as neither reliable nor valid. As a solution one should provide rules of measurement that would enable answering following questions:

1.Is it possible to measure sentiment with application of a given dataset (given set of questions)?

2. Is the measured phenomenon (sentiment) consistent over time or should it be interpreted differently in different time periods?

---

* Ph.D., Department of Economics I, Warsaw School of Economics.

Only if the positive answer is given to both of these questions, comparisons of the sentiment index can be reliably conducted between periods. As it is rather uncommon to verify consistency of questions in the standard sentiment measurement in business surveys, this paper aims at presenting an approach with multi-group confirmatory factor analysis that facilitates construction of indices with intertemporaly comparable values. Brief description of the method and necessary conditions for conducting time comparisons are shown later in the paper. It is also hypothesized that the set of questions that serve as proxies in the measurement of sentiment should be consistent with respect to the time scope (either forecasting or diagnostic questions) and on the conceptual level it should be possible to explain the answers to each of the questions by one underlying latent concept. Finally, in order to provide an argument in the process of verification of the hypothesis, a case study that is based on the survey in manufacturing industry in Poland is conducted.

II. INTERTEMPORAL COMPARISONS WITH APPLICATION OF MGCFA

In the approach based on the multi-group confirmatory factor analysis, sentiment is treated as a latent phenomenon that is reflected by the answers to a given set of questions – proxies. The formal structure of the model in the case of N proxies (questions), only one latent variable (as only one sentiment measure is assumed) and T time periods can be provided by the formula (see Stenkamp and Baumgartner 1998):

$$\forall_{t \in T} \mathbf{q}^t = \boldsymbol{\tau}^t + \boldsymbol{\gamma}^t CSI^t + \boldsymbol{\varepsilon}^t \text{, where} \tag{1}$$

in all time periods $\mathbf{q}^t$ is $N \times 1$ vector of question answers, $\boldsymbol{\tau}^t$ is $N \times 1$ vector of intercepts, $\boldsymbol{\gamma}^t$ is $N \times 1$ vector of factor loadings and $\boldsymbol{\varepsilon}^t$ is $N \times 1$ vector of measurement errors. In order to ensure identifiability of the model one element of the $\boldsymbol{\gamma}^t$ vector (factor loading) is set to 1[1] and one element (which must correspond to constrained to 1 factor loading) of $\boldsymbol{\tau}^t$ vector (intercept) is set to zero. Additionally, $E\left(\boldsymbol{\varepsilon}^t\right) = \mathbf{0}$ and $\forall_{t \in 1..T, p, q \in 1..N, p \neq q} \operatorname{cov}\left(\boldsymbol{\varepsilon}_p^t, \boldsymbol{\varepsilon}_q^t\right) = 0$.

Nevertheless, it is not sufficient to estimate the model with such assumptions in order to conduct comparisons of the sentiment mean between periods. In order

[1] It is usually the first element of this vector. Instead of constraining one factor loading to 1, the identification of the measurement model can be also ensured by setting the variance of latent variable to 1.

to ensure the reliability of comparisons in time it is essential to provide the measurement invariance (between groups – periods of analysis) that is checked on three levels: 1. Configural invariance, 2. Metric invariance, 3. Scalar invariance (Davidov 2008).

Configural invariance is presented as an identity of conceptualization and operationalization of measured phenomenon. In the case of sentiment measurement it is guaranteed by applying the same group proxies (questions) in all periods of analysis (Kaplan 2009). It is also verified by checking whether in models estimated for each period separately, pattern of signs of factor loadings is the same in all periods. Second step is verification of metric invariance. Metric invariance implies that reaction of an answer to each from the group of questions serving as proxies to unit change in the latent variable is the same in all time periods. It is tested by fixing the factor loadings to be the same for all periods of analysis and checking if the model with such constraint fits the data well. Final step – scalar invariance – means that the observed differences in the values of proxies (answer to questions) between periods are only a consequence of change in the value of the measured sentiment. Scalar invariance implies that the latent variable (sentiment) has in all periods not only the same scale but also the same reference point. It is checked by fixing the intercepts to be equal in all periods and checking the model fit. In practice it is usually very hard to confirm full measurement invariance. However, it was shown (Steenkamp and Baumgartner 1998) that in order to conduct reliable comparisons between groups (periods) it is sufficient to assure partial measurement invariance. It is connected with fixing the factor ladings and intercepts for only two proxies (questions) and letting the remaining ones to vary between periods.

Measurement invariance can be verified stepwise. Nevertheless, in the case of sentiment indicators there is no reason for checking metric invariance separately, so it is sufficient to verify configural invariance and then metric and scalar invariance simultaneously[2] One of the approaches in assessment of the measurement invariance is to verify the model fit with each set of constraints. In the case of business sentiment index, it would imply checking the fit of the model with configural invariance and afterwards, checking the fit of the model with additionally imposed conditions on equality of factor loadings and intercepts (metric and scalar invariance). If full metric invariance cannot be achieved modification indexes can be calculated, which show the possible gain in the model fit after some constraints are relaxed.

In order to assess the model fit at various levels of measurement invariance a set of fit-statistics that are based on chi-square distribution is applied. The most

[2] Metric invariance is necessary to conduct comparisons of correlation between latent variables. However, it is not the point of interest of the analysis in the case of sentiment.

basic is χ^2 statistic itself which is based on the differences in the error term matrix. Yet, it is very sensitive to the sample size and thus very often leads to rejection of true models (Górniak 2000, Bollen 1989). More often statistics that take into account the sample size and the degrees of freedom of the estimated model are taken into account. The most popular ones are: (1) χ^2/df, which should be within the range <1;5>; (2) Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI), which should be in the range <0.9; 1.0>, but also (3) Root Mean Square Error of Approximation (RMSEA) and Standardized Root Mean Square Residuals (SMRM) which should be less than 0.05–0.1 depending on the rigidity of researcher (Górniak 2000, p.134; Hox 2002, p. 239)[3].

III. BUSINESS SENTIMENT INDICATOR IN MANUFACTURING INDUSTRY ESTIMATED WITH MGCFA

This point presents the results of a case study performed in order to verify the applicability of the MGCFA to construct sentiment indicator for manufacturing industry. In the study two different sets of questions from the survey in manufacturing conducted at Research Institute for Economic Development at Warsaw School of Economics were employed. Measurement model was estimated for each of the sets and the procedure proposed for the verification of the set of questions was tested. Additionally, it was checked whether a group of questions comprising information about company's forecasts performs better as a basis for a sentiment indicator than group of questions comprising diagnostic and forecast information.

The first set of questions covered information concerning production forecasts (PROD_F), orders forecasts (ORD_F), financial situation forecasts (FS_F) and general economic situation forecasts (GES_F) for the period 2009Q1 to 2010Q4.[4] It was verified stepwise, whether the model with configural invariance and then full and partial measurement invariance fits the data well.

---

[3] The procedure of model fit assessment in the Multi-group Confirmatory Factor Analysis relies much on the interpretation of the researcher. Different authors apply different thresholds for the given set of fit statistics. For discussion see Brown (2006).

[4] Detailed wording of questions and the set of answers were: 1. What will be the change in production in your company in the forthcoming 3-4 months? (1 „Will go up"; 2 „Will stay at the same level"; 3 „Will go down"); 2.What will be the portfolio of orders in your company in the forthcoming 3-4 months? (1 „Higher"; 2 „The same"; 3 „Lower"); 3. How will the financial position of your company develop in the forthcoming 3-4 months? (1 „Will improve"; 2 „Will stay the same"; 3 „will deteriorate"); 4. How will the general economic situation in Poland develop in the forthcoming 3-4 months? (1 „Will improve"; 2 „Will stay the same"; 3 „will deteriorate").

Table 1. Verification of the model fit with the first set of questions

	Measurement invariance		
	Configural invariance	Full measurement invariance	Partial measurement invariance
χ^2/df	0.925	5.798	1.275
CFI	1.000	0.953	0.998
TLI	1.000	0.955	0.997
RMSEA	0.000	0.106	0.025
SRMR	0.005	0.136	0.033

Source: Own calculations in MPlus.

The configural invariance was assured and the model proved to be very close to full measurement invariance. Nevertheless, it was sufficient to relax the constraint on factor loading and intercept in question concerning the general economic situation to obtain very good fit. The estimated model with partial measurement invariance can be presented by the following system:

$$\forall_{t\in T, i\in N}\begin{cases} PROD\_F_{t,i} = BSI_{t,i} + \varepsilon_{1,t} \\ ORD\_F_{t,i} = 0.020 + 1.049\cdot BSI_{t,i} + \varepsilon_{2,t} \\ FS\_F_{t,i} = 0.720 + 0.691\cdot BSI_{t,i} + \varepsilon_{3,t} \\ GES\_F = \tau_t + \gamma_t \cdot BSI_{t,i} + \varepsilon_{4,t} \end{cases} \tag{2}$$
$$\forall_{t\in T, k\in\{1,2\}, l\in\{1,2,3,4\}, k<l} cor\left(\varepsilon_{k,t},\varepsilon_{l,t}\right) = 0, cor\left(\varepsilon_{3,t},\varepsilon_{4,t}\right) = \lambda_t$$

All of the period specific estimates are presented in table below.

Table 2. Period specific estimates for the model with partial measurement invariance (2).

	2009Q1	2009Q2	2009Q3	2009Q4	2010Q1	2010Q2	2010Q3	2010Q4
$E(BSI_t)$	2.228	2.099	2.093	2.122	2.026	1.840	1.887	1.953
$SD(BSI_t)$	0.616	0.560	0.616	0.556	0.623	0.567	0.599	0.605
τ_t	1.816	1.135	1.440	1.340	1.068	1.183	1.135	1.481
γ_t	0.370	0.598	0.446	0.435	0.523	0.453	0.501	0.365
λ_t	0.061	0.045	0.064	0.081	0.044	0.117	0.058	0.057

Source: Own calculations in MPlus.

In the first two rows of Table 2 the means and standard deviations of sentiment indicators for different periods are presented. As measurement invariance of the model was established, the means can be compared between periods. The lower the value of BSI, the better the business climate is observed. Information concerning standard deviation provides insight concerning the variability of the sentiment on the individual (respondent) level. The higher the standard deviation is, the more diversified the business sentiment among Polish companies is.

Additional remark should be given to coefficients of correlation between error terms (λ_t). In standard specification they are assumed to be zero. Nevertheless, it is possible to relax this constraint, as there might be effects connected with wording that bias answers to given pair of questions in some direction. In the case of estimated model (2), there was imposed a correlation between error terms in answers to question concerning the financial situation of respondent's company (FS_F) and the general economic situation (GES_F). As the correlation in all periods is positive, it implies that if respondents were giving better or worse answer to the question concerning the financial situation of their branch, they were also providing better or worse (respectively) assessment of the general economic situation. It confirms strong relations between branch specific situation and situation in the general economy.

The initial set of proxies proved to have an underlying unidimensional phenomenon with constant meaning in all of the period from the sample. Thus the selected questions can be assumed to reflect some kind of business sentiment and answers to these questions can be modeled by linear functions of business sentiment indicator.

In the second specification a model of business sentiment that comprised information from diagnostic and forecasting questions was estimated. Four questions were selected as proxies of the new sentiment. Two diagnostic questions referring to production and orders were picked, but also two forecasting questions from the same fields were selected[5]. However, the model with such set of proxies proved to be highly unreliable as the tool to assess business sentiment level. It was not possible to establish even configural invariance in the model estimated with MGCFA. The statistics proved very poor (χ^2/df = 34.21; CFI = 0.961; TLI = 0.767; RMSEA = 0.275; SRMR = 0.040). Thus it could not have been stated that in each period there is some latent variable that is able to explain the answers to the set of questions.

[5] Detailed wording of questions and the set of answers were: 1. What was the change in production in the last month in your company (1. "Went up", 2. "Remained the same", 3. "Went down"); 2. What will be the change in production in your company in the forthcoming 3-4 months? (1 "Will go up"; 2 "Will stay at the same level"; 3 "Will go down"); 3. What was the orders' portfolio in the last month in your company (1 „Higher"; 2 „The same"; 3 „Lower"); 4. What will be the portfolio of orders in your company in the forthcoming 3-4 months? (1 „Higher"; 2 „The same"; 3 „Lower").

IV. CONCLUSIONS

Presented approach should start a discussion on measurement issues in business surveys. It should be considered not sufficient to calculate indices that are based on researcher's intuition. It was shown that with application of multi group confirmatory factor analysis it is possible to verify whether set of proposed fields from the business survey questionnaire (represented by answers to questions from these fields) can serve as a proxy of business sentiment i.e. whether answers to these questions can be explained by the value of business sentiment indicator. MGCFA provides also an additional advantage – verifying measurement invariance enables to check not only unidimensionality of business sentiment in a given period, but also consistency of understanding of this phenomenon between periods.

Empirical part of the paper focused on the possibility of construction of a sentiment index in the manufacturing industry. It was shown that with the set of four forecasting questions concerning future production, orders, branch's financial situation and general economic situation, one can provide consistent estimates of business sentiment with one indicator. On the other hand, index that consists of mixture of diagnostic and forecasting indicators, proved to be unreliable.

REFERENCES

Bollen K. (1989) *Structural Equations with Latent Variables*, New York

Brown T. (2006) *Confirmatory Factor Analysis for Applied Research*, The Guilford Press, New York

Davidov E. (2008), A Cross Country and Cross-Time Comparison of the Human Values Measurements with the Second Round of the European Social Survey, *Survey Research Methods*, vol. 2 No.1, pp. 33–46

European Economy (2006): *The Joint Harmonised EU Programme of Business and Consumer Surveys*, European Commission, Special Report no. 5

Górniak J. (2000), *My i nasze pieniądze*, Aureus, Kraków

Hox J. (2002), *Multilevel Analysis: techniques and applications*, Lawrence Erlbaum Associates, Mahwah

Kaplan D. (2009), *Structural equation modeling. Fundation and Extentions*, Sage, Los Angeles.

Steenkamp J., Baumgartner H. (1998) Assessing Measurement Invariance in Cross-National Consumer Research, *The Journal of Consumer Research*, Vol. 25, No. 1 (Jun., 1998), pp. 78–90

Piotr Białowolski

ZASTOSOWANIE KONFIRMACYJNEJ ANALIZY CZYNNIKOWEJ DLA WIELU GRUP DO TWORZENIA WSKAŹNIKÓW KONIUNKTURY

W artykule przedstawiony został zbiór argumentów za korzystaniem z konfirmacyjnej analizy czynnikowej dla wielu grup przy tworzeniu wskaźników nastroju w badaniach koniunktury. Rzetelny pomiar, umożliwiający wiarygodne porównania średnich wartości wskaźnika między okresami wymaga zapewnienia zgodności pomiaru na każdym w trzech podstawowych poziomów – zgodność konfiguracji, zgodność skali i zgodność punktu referencyjnego (skalarnej). W artykule postawiono hipotezę, że jedynie zbiór pytań, które są spójne wewnętrznie, mogą stworzyć podstawę do konstrukcji wskaźnika nastroju. Pierwsza próba zweryfikowania tej hipotezy przeprowadzona została w oparciu o studium przypadku. Podjęto próbę stworzenia wskaźników nastroju w przemyśle przetwórczym na podstawie dwóch zbiorów pytań. Wyniki wskazują, że jedynie dla zmiennych tworzących konceptualną całość i odwołujących się do przyszłości, możliwa była estymacja modelu, w którym spełniony jest warunek częściowej zgodności pomiaru.