



THE EFFECTS OF MISSING IMPUTATION ON A COMPOSITE BUSINESS CLIMATE INDEX OF INSTITUTE FOR ECONOMIC AND ENTERPRISE RESEARCH

Miklós Hajdu – István János Tóth
Institute for Economic and Enterprise Research

Abstract

In this paper we attempt to examine whether it is feasible to impute missing data values or not regarding the Quarterly Business Climate Survey of Institute for Economic and Enterprise Research. This subject implies the assessment of the missing data mechanism of the business climate questions and thereby the discussion of the possible imputation methods. But as we do not have sufficient information for analysing the missing data mechanism yet and the aim of the survey is to indicate the changes in some macroeconomic tendencies, we attempt to examine the situation with the assumption that the missing data mechanism is random. This paper presents an evaluation of the effects of multiple imputation from the aspect of the correlation between the GDP growth and the Quarterly Business Climate Index.

Keywords: small and medium enterprises, Hungary, item nonresponse, missing data, business climate

JEL Classification: C18



Introduction

Nonresponse became a key challenge of organizations dealing with surveys¹. The reasons of this phenomenon are more or less obvious, but its effects on the survey results are less clear and thereby the decision between the possible techniques for handling missing data should be based on the consideration of several factors. In any case nonresponse should not be ignored during the analysis of survey data.

Survey nonresponse can be occurred by refusing the complete interview (*unit nonresponse*) or by leaving some questions unanswered in the questionnaire (*item nonresponse*) – this analysis focuses on some possibilities for assessing and treating item nonresponse. Three different types of missing data can be identified:

- 1) missing completely at random (MCAR) – occurs when the factors leading to nonresponse are independent both of observable and unobservable parameters;
- 2) missing at random (MAR) – occurs when nonresponse is not a completely random process, and can be modelled by observed variables;
- 3) missing not at random (MNAR) – occurs when nonresponse depends on unobserved factors.

Some papers suggest that organizational surveys are more affected by nonresponse than individual ones (Tomaskovic-Devey et al., 1994). There are also some analyses about the effects of nonresponse on business tendency surveys concluding that impact of nonresponse on such surveys does not endanger their forecasting performance (Kowalczyk – Tomczyk, 2011; Seiler 2013). The previous studies of Institute for Economic and Enterprise Research (IEER) proved that the extent nonresponse for business tendency questions is both correlated with official macroeconomic indicators (Hajdu, 2014) and with the uncertainty indices of IEER.

The Quarterly Business Climate Survey

This study is based on the Quarterly Business Climate Survey (QBCS) of IEER. A total of 400 companies are surveyed in every quarter since 2010 that are representative of the economic performance and sector distribution of firms

¹ See the special issue (Volume 645.) of *The ANNALS of the American Academy of Political and Social Science*



operating in Hungary. The questionnaire inquires changes regarding ten different business climate indicators related to the business situation and prospects of firms that are analyzed and summarized in the IEER Quarterly Business Climate Index (QBCI)². QBCS is not a panel survey as new, independent samples of firms are taken for every waves of the research, but the quota parameters do not change.

As the survey has been repeated only for 24 waves between 2010 and 2015, we can only make preliminary calculations for the correlations between the QBCI and macroeconomic indicators. Also as the small number of observations did not enable us to calculate the optimal formula for QBCI so far, its current composition is based on theoretical considerations. However, the growing number of observations will allow us soon to work out the possibilities of its development and an interesting opportunity is to deal with missing values.

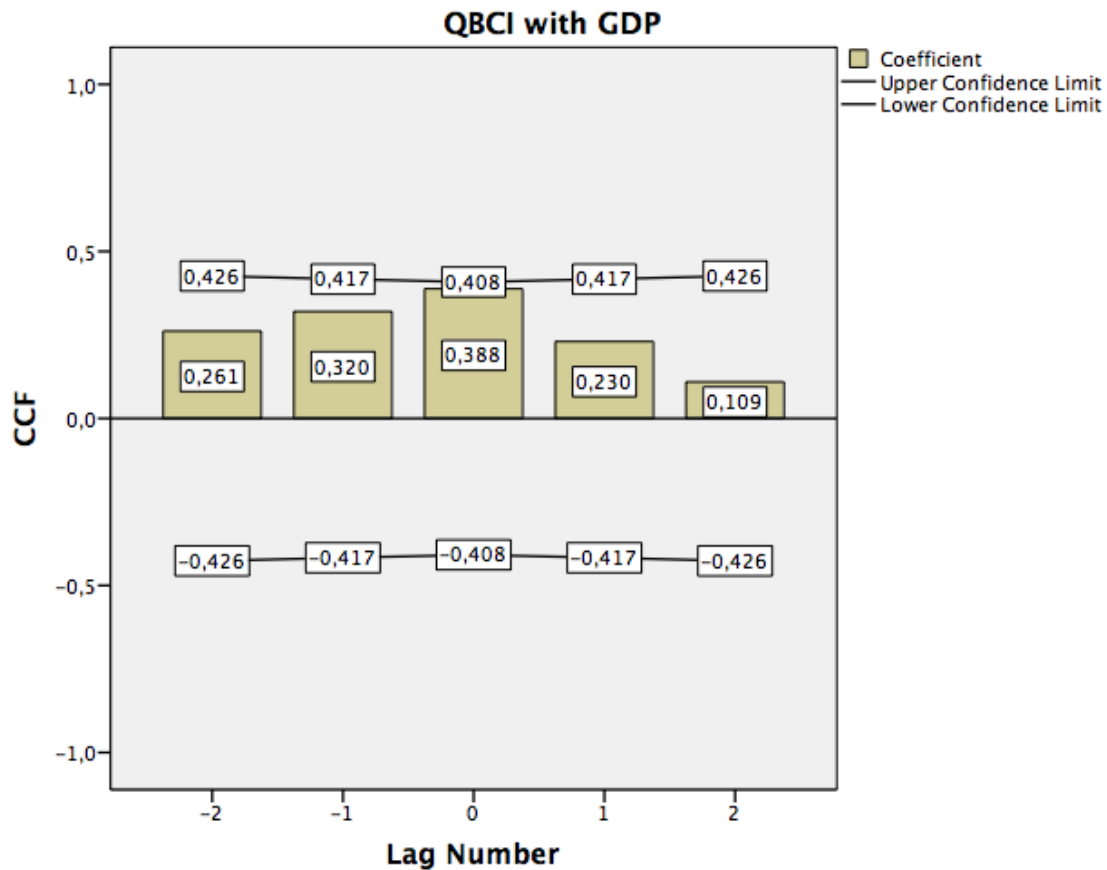
The cross-correlations suggest that a positive correlation without lags is emerging between QBCI and the GDP growth³ (see Figure 1.). The question is whether missing data imputation can contribute to the nowcasting ability of QBCI.

² See the Appendix for the composition of QBCI.

³ Seasonally and calendar effects adjusted and reconciled quarter-by-quarter data is taken into account.



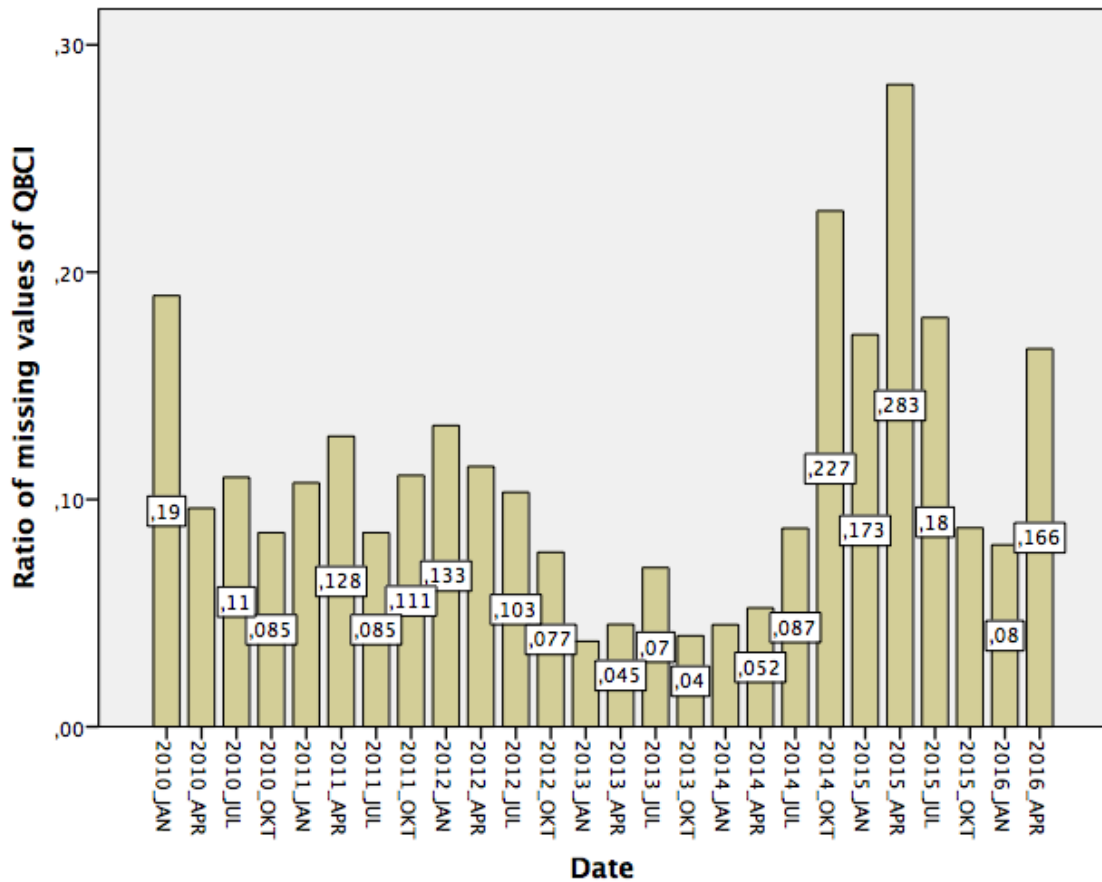
Figure 1. The cross-correlations between the QBCI before missing data imputation and the GDP growth, 2010-2015



As the business climate indicator is based on the ten questions described in the Appendix, even if one them is left unanswered the composite indicator should not be computed because of insufficient information. So even only one item nonresponse leads to missing value of the QBCI. As it is shown on Figure 2, the proportion of the firms with missing data for calculating the QBCI was moving between 4% and 28% with the mean of 11% – these ratios suggest that missing imputation is much more considerable than listwise deletion, the basic option for treating missing data.



Figure 2. The ratio of missing values of QBCI, 2010-2015



Missing imputation

As there are no definitive statistical methods for deciding which missing data mechanism occurred during the surveys, both theoretical and empirical aspects should be considered. The literature mentioned in the introduction about the missing data analysis of business tendency surveys suggests that MNAR is the mechanism how nonresponse occurs during these researches. Additionally, a binary logistic regression model (see Appendix 2.) evinces that the odds ratios for nonresponse are different between the sectors and for smaller and larger firms. Also, there were significant differences between the waves of the research. However, the feedbacks from the interviewers and the distribution of QBCI implies the possibility of MAR.

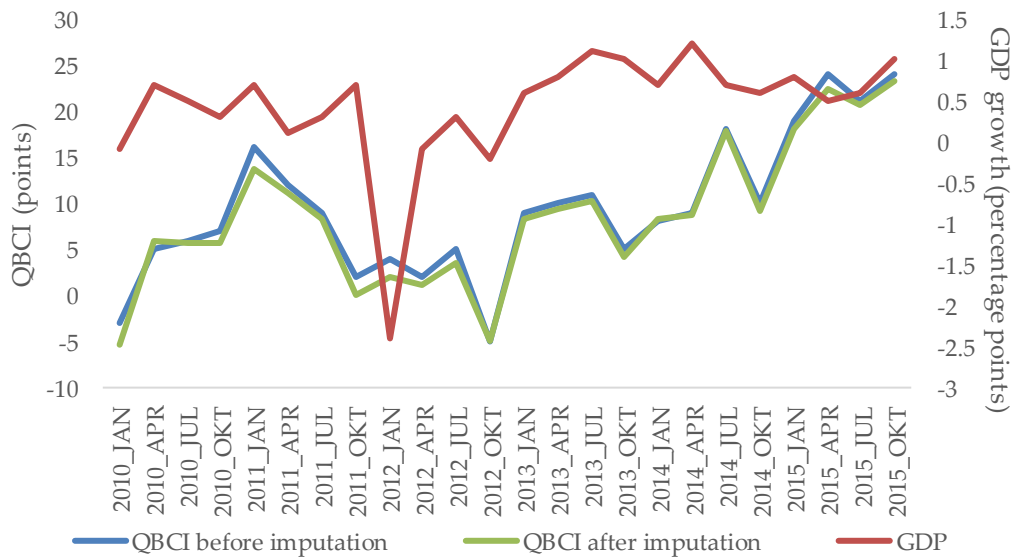


All in all, these factors indicate that missing data mechanism for the business tendency questions of IEER's QBCS may be regarded as MAR⁴.

We chose the fully conditional specification method based on Markov Chain Monte Carlo for missing imputation with five rounds. Our imputation model includes the sector, the size, the region and the wave as complete predictor variables and the ten business tendency questions underlying the QBCI as predictors and imputed variables. As these business tendency questions may have categorical answers, logistic regression models are used during the imputation.

After the imputation, we recalculated the QBCI based on the complete variables of the business tendency questions. Figure 3 presents how did the imputation change the QBCI and the GDP growth is also charted. The imputation decreased the value of QBCI in the case of 21 waves and increased its value only for three waves, suggesting that pessimistic answers remained hidden during the interviews.

Figure 3. The QBCI before and after missing data imputation and the GDP growth, 2010-2015



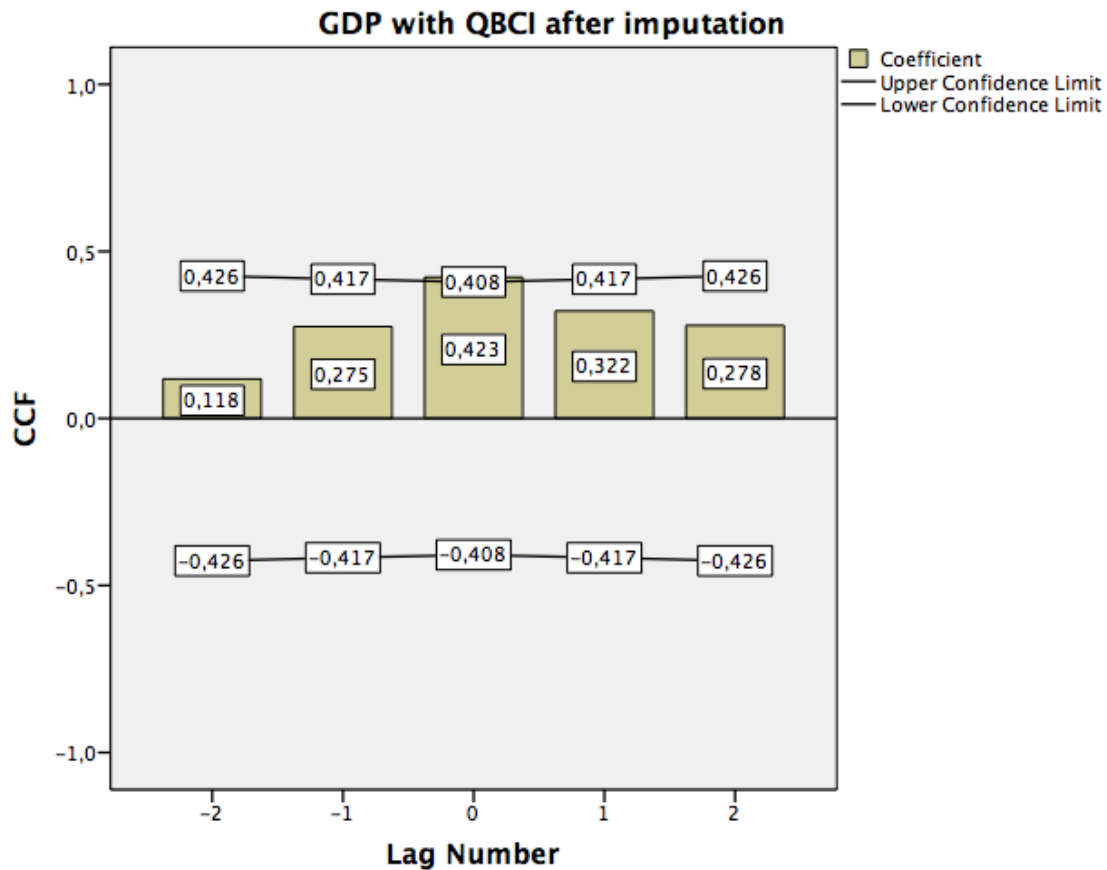
These changes could improve the correlation between the GDP growth and the QBCI (see Figure 4.) that can be regarded as significant with no lags after the imputation. However, it should be kept in mind that these results have to be treated very

⁴ The IEER is currently preparing a follow-up survey that may help in the decision of this question.



cautiously as only 24 observations were analysed, the missing data mechanism has to be explored more thoroughly and also the composition of the QBCI has to be developed.

Figure 4. The cross-correlations between the QBCI after missing data imputation and the GDP growth, 2010-2015



Conclusions

The results draw attention to the fact that it worth to consider missing imputation during the development of composite business situation indicators. In order to do this, an important task is a more detailed exploration of the missing data mechanism of business tendency questions by follow-up surveys and additional data collection. Then the imputation method can be refined. However, our results suggest that multiple imputation can contribute to the nowcasting ability of QBCI, even if the missing data mechanism is not clear.



References

- Hajdu, Miklós (2014) *Item Nonresponse in the SME Outlook Business Tendency Survey*. IEER. Budapest. https://www.ciret.org/media/ciret_papers/hangzhou-2014/0055_ciret_2014_140327.pdf (acc: 2016.03.19.)
- Kowalczyk, Barbara – Tomczyk, Emilia (2011): *Non-response and Weighting Systems in Business Tendency Surveys: Are Expectations Influenced?* Institute of Econometrics. Warsaw School of Economics. http://194.145.100.18/repec/files/wsd/irgpim/wsd_irgpim_v86_2011_i2_n5_2.pdf (acc: 2016.03.18.)
- Seiler, Christian (2013) *Nonresponse in Business Tendency Surveys: Theoretical Discourse and Empirical Evidence*. In: ifo Beiträge zur Wirtschaftsforschung Ifo Institute for Economic Research at the University of Munich. https://edoc.ub.uni-muenchen.de/16071/1/Seiler_Christian.pdf (acc: 2016.03.18.)
- Tomaskovic-Devey, Donald – Leiter, Jeffrey – Thompson, Shealy (1994) *Organizational Survey Nonresponse*. In: *Administrative Science Quarterly*, Vol. 39, No. 3., pp. 439-457.



Appendix

A1. The formula of the IEER Quarterly Business Climate Index (QBCI)

$BCI = (raj_ + uhj_ + vjj_ + tszm_ + lvv_ + uhv_ + vjv_ + cue_ + tszv_ + bv_)/10$, where:

- RAJ_: current stock orders
- UHJ_: current business situation
- VJJ_: current profitability
- TSZM_: production levels – previous quarter
- LVV_: expected change in staff members
- UHV_: expected business situation
- VJV_: expected profitability
- CUE_: expected capacity utilization
- TSZV_: expected production levels
- BVV_: expected investment activity

The answers for the questions relating to the indicators worth the following scores:

Answer	Score
Good / Better / Growing / It will be better / It will grow	100
Satisfactory / Does not change / Will not change	0
Bad / Worse / Decreasing / It will be worse / It will decrease	-100



A2. Detailed results of the binary logistic model explaining the missingness of the business climate indicator (N=10,480)

	B	S.E.	Wald	df	Sig.	Exp(B)
SECTOR			96.71	3	0	
Manufacturing	-0.561	0.099	32.064	1	0	0.57
Construction	-0.026	0.118	0.049	1	0.826	0.974
Commerce	0.157	0.1	2.472	1	0.116	1.17
SIZE			21.737	3	0	
20-49	0.317	0.09	12.551	1	0	1.373
50-99	0.12	0.095	1.585	1	0.208	1.127
100-249	-0.052	0.092	0.316	1	0.574	0.949
REGION			1.865	2	0.394	
Central	0.083	0.076	1.189	1	0.276	1.087
Hungary						
Transdanubia	-0.015	0.083	0.032	1	0.857	0.985
WAVE			321.925	25	0	
2010_JAN	0.124	0.188	0.439	1	0.508	1.132
2010_APR	-0.664	0.222	8.983	1	0.003	0.515
2010_JUL	-0.482	0.211	5.196	1	0.023	0.618
2010_OCT	-0.771	0.224	11.801	1	0.001	0.462
2011_JAN	-0.536	0.211	6.434	1	0.011	0.585
2011_APR	-0.333	0.203	2.698	1	0.1	0.717
2011_JUL	-0.766	0.225	11.647	1	0.001	0.465
2011_OCT	-0.452	0.21	4.622	1	0.032	0.636
2012_JAN	-0.278	0.202	1.891	1	0.169	0.757
2012_APR	-0.449	0.205	4.773	1	0.029	0.638
2012_JUL	-0.547	0.212	6.646	1	0.01	0.578
2012_OCT	-0.911	0.233	15.328	1	0	0.402
2013_JAN	-1.662	0.297	31.219	1	0	0.19
2013_APR	-1.474	0.278	28.106	1	0	0.229
2013_JUL	-1.005	0.24	17.524	1	0	0.366
2013_OCT	-1.589	0.29	29.98	1	0	0.204
2014_JAN	-1.472	0.278	28.016	1	0	0.229
2014_APR	-1.313	0.263	24.85	1	0	0.269
2014_JUL	-0.742	0.225	10.919	1	0.001	0.476
2014_OCT	0.364	0.183	3.972	1	0.046	1.439



A2 (continued)

	B	S.E.	Wald	df	Sig.	Exp(B)
2015_JAN	0.046	0.192	0.057	1	0.812	1.047
2015_APR	0.669	0.177	14.218	1	0	1.953
2015_JUL	0.074	0.19	0.151	1	0.698	1.076
2015_OCT	-0.768	0.225	11.7	1	0.001	0.464
2016_JAN	-0.837	0.231	13.19	1	0	0.433
Constant	-1.525	0.178	73.38	1	0	0.218

Notes: the reference categories were the following:

Sector: business services

Size: 250 employees or more

Region: Great Plain and North

Wave: 2016_APR