Abstract

The impact of missing data on quantitative research can be serious. It could lead to biased estimates of parameters, loss of information, increased standard errors or decreased statistical power and weakened results of findings. The aim of the paper is to discuss three missing data methods: regression, imputation and multiple imputation; and their impact on the CCCTB determination and based on the results to identify the most suitable method which will lead to the least distortion. The results gained with the application of those methods are compared with those obtained from the complete data set.

Keywords: Non-profit organization; tax assignation; gift; income tax; non-profit sector

1. Introduction

The efforts to coordinate the corporate taxation systems within the EU or at least setting of the minimum tax rates has been unsuccessful for decades due to the fact that the member states refused to give up their national sovereignty in the area of direct taxation. As already stated in Bolkestein Report although it is clear that the concept of common tax base for companies operating in the EU is defined as long-term goal of EU tax policy, some parties in both business and tax administrations remain totally opposed to it. However, despite these attitudes European Commission have been developing a strategy, repeatedly addressing specific problems connected with the existence of 27 different taxation systems, resulting into the four drafts of possible harmonization schemes of corporate taxation within the EU.
First of them represented Home State Taxation System, under which the European activities of the companies would all have been taxed in the country, where the company would have its seat. European Union Company Tax represented the second drafted model, under which unified system of corporate taxation would be applied on MNEs and would have been administrated by the body on the EU level. Common Compulsory Harmonized Tax represented another model, which would implement common corporate tax base for all EU companies. None of the above mentioned models was further developed. The only model, comprising certain level of optionability represented Common Consolidated Corporate Tax Base (hereinafter as CCCTB), under which, companies opting for the system would construe the tax base according to the unified set of rules.

In September 2004 ECOFIN supported the establishment of CCCTB Working group, whose task was to start investigate the concepts for introduction of CCCTB. This effort was completed on 16th March 2011, when the European Commission released the draft of the Directive on a Common Consolidated Corporate Tax Base. The major benefits connected with the establishment of the CCCTB can be seen in the elimination of transfer pricing problems (due to the possibility of consolidation), the establishment of fair tax competition, avoiding the double taxation arising from conflicting tax claims, the possibility of cross-border loss compensation and reduction of compliance costs of taxation.

The consolidation regime incorporated in CCCTB system is of course connected with the problem of the allocation of the consolidated tax base between the EU Member States. The directive proposals in that context comprise the allocation formula with three equally weighted factors as the mechanism for sharing the tax base. It is necessary to mention that the discussion in academic literature on the topic of economic and revenue impact resulting from the consolidation and sharing mechanism is very limited. As is showing the review of the literature, all the research is based on using the data from AMADEUS or ORBIX database, which provide the information on the greatest number of private and public companies. Unfortunately, the financial statements of the companies in the databases are very much missing the data, which are needed for the allocation formula (information on payroll, assets and sales). The impact of missing data on quantitative research can be serious. It could lead to biased estimates of parameters, loss of information, increased standard errors or decreased statistical power and weakened results of findings. Moreover this fact is leading to the distortions of the results in the form of revenue impacts on the national budgets resulting from the changes in the sharing of the consolidated tax bases between the EU Member States.

Therefore the aim of the paper is to research the most suitable method of missing data imputation into the sample of 1,190 companies from Amadeus Database which would not significantly distort the allocation of the group tax base between the EU Member States. The paper is the result of research under the project GACR no. 13-21683S “Quantification of the impact of CCCTB system introduction on the budget revenues of the Czech Republic”.

2. Theoretical background

European Commission estimates that CCCTB can save to European business EUR 700 million every year in the form of reduced compliance costs of taxation, and EUR 1.3 billion due to the possibility of consolidation (mainly due to the possibility of cross-border loss compensation). Moreover, remarkable amount could be saved by business looking to expand cross-border, as states Erasmus-Koen (2011). In the proposal, there can be identified five main principles on which the CCCTB system is based – optionability, common tax base, possibility of consolidation, formulary apportionment of consolidated tax bases among member states and the principle of one tax administrator.

Looking on the current treatment of the tax bases of MNEs within the European Union, there can be identified two main approaches. Some of the states, not allowing consolidation and any form of group taxation schemes treat every company as separate entity. As mentions Oestreicher (2008), each subsidiary of MNEs is dealing at arm’s length with its parent for it is treated as separate entity. Countries with separate entity approach are using separate accounting for the division of the group tax base amongst the individual jurisdictions. As further add by Oestreicher (2000) this splits MNEs on the unrelated enterprises expecting to act as independents – i.e. all transactions between the members of the group have to be at arm’s length. This is caused by the fact that in situation when all the intra-group transactions are taken into account for taxation purposes, they might be used as a tool for tax planning. As mentioned by Solilova and Nerudova (2013) arm's length transfer prices have significant impact on both the taxable income (and therefore on the tax burden) of entities and tax revenues of participating states. Vann (1991) mentions that even though the MNEs can use intra-group transactions as a toll for tax planning to benefit from tax differentials, they are most frequently accused by fiscal administrators of manipulating their transfer prices. The same was concluded by Picciotto (1992) who mentioned that construed transfer prices based on the arm’s length principle can be adjusted by
local tax authorities to reflect profit or loss which would have been reached in situation when all internal transfers
would be realized under open-market conditions.

On the contrary to the separate entity approach, countries allowing full tax consolidation or group taxation
schemes are usually applying unitary approach. This means that they look on parents, subsidiaries and other
associated companies as on one single unit for income tax purposes. That system requires also the implementation of
the mechanism which would distribute the group tax base between the individual companies (states). According to
the Weiner and Mintz (2002) the technique under which the cross-country consolidated tax base of MNE is split
between the jurisdictions represents the formulae apportionment. At present, formulary apportionment is applied in
practice in Canada and U.S.A. Under that system, apportionment of the tax base is done according to the allocation
formula, taking into account the different factors as for example assets, labor force or sales of the enterprise. On the
contrary to the separate entity approach, in the system of unitary apportionment substance-over-form principle is
applied.

As was already mentioned above, CCCTB system comprise the possibility of the consolidation and group taxation
scheme, therefore is accompanied by the system according to which the group tax base is distributed between the
Member States. European Commission decided to apply the formulary apportionment similar to the system applied in
Canada and U.S.A. with slight differences in the factors.

First scientific research focusing on the topic of formula apportionment by Musgrave (1972) highlighted the fact
that formulary apportionment could eliminate the problem with transfer pricing. Miller (1984) mentions that the
formula should reflect the elements measuring the processes involved in the earning of net income and that formula
should be easy to administer. They were followed by Gordon and Wilson (1986) and McLure (1981) which were
discussing mainly the distortions arising from the formulary apportionment. The factors comprised in formula
apportionment were discussed by McLure (1980) and later by Golsbee and Maydew (2000) who proved that factors
as property, payroll and sales, corporate income tax transforms into a tax on property, payroll and sales. Wellish
(2000) pointed out the reduction of labor demand in the states caused by using labor as a factor of formula, for
according to him it results into the higher labor costs.

Later Gordon and Wilson (1986) examined how corporate taxation of multinational firms using formula
apportionment affects the incentives faced by individual firms and individual states. Musgrave (1984) has proved that
when a formula consists of the factors as property of the company, payroll and sales, corporate income tax transforms
into a tax on property, payroll and sales. This has also been proved by Goolsbee, and Maydew (2000). Also Wellish
(2000) shows, that when a labor is used as the factor, then the costs of labor are exceeding the local wage rate, which
reduces the demand for labor in each state. The suggestion of a theoretical model of tax competition in apportionment
rules amongst states in the U.S.A. can be find in the research by Anand and Sansing (2000).

The publication of the intention of the European Commission to employ formulary apportionment as the tax base
sharing mechanism within the CCCTB system led to several researches on that topic. Firstly, Devereux (2004)
pointed out that formulary apportionment can be regarded as a system of source taxation. Further, Sorensen (2004),
Mintz and Weiner (2003) or Agúndez-García (2006) discussed the problems stemming from the sharing mechanism
within the EU. The impacts of CCCTB introduction in EU15 have been researched by Fuest, Hemmelgarn, Ramb
(2006), who based on data from Deutsche Bank concluded that Ireland, Belgium and Netherlands will lose a
significant part of their tax base, while larger countries will lose less tax base. Devereux and Loretiz (2007) in their
static model based on the data from ORBIS database deduce that overall EU tax revenues would decline by 1 % and
that the new EU countries as Slovak Republic, Hungary and to a lesser extent also Czech Republic would gained
increased tax revenue.

3. Methodology

The empirical analysis is based on the company-level data from the Amadeus database which is provided by
Bureau van Dijk. These data were taken from update 227 (August 2013) of the database including standardized
financial information of more than 18 million public and private companies in 43 European countries.

In the paper the authors use the similar assumption as Devereux and Loretiz (2007) that corporations do not change
their behavior in response to the tax reform, which provides a useful benchmark for the analysis. Further, the paper
follows the approach of Devereux and Loretiz (2007, Fuest, Hemmelgarn and Ramb (2006) or Clien, Neubig, Phillips,
Sanger and Walsh (2010) which is based on the data from the Amadeus database due to the fact that it contains data
on more than 18 millions companies.
With respect to the fact that the aim of the paper is to research the most suitable method of missing data, which would not significantly distort the allocation of the group tax base between the EU Member States, it was needed to gain the group of EU companies, which would under CCCTB system qualify for consolidation regime and group treatment. Therefore only European companies (EU28) fulfilling the two-tier test confirming the eligibility for consolidation (group membership) were further analyzed. This test consists of two layers – control, which is assumed if the controlling company holds at least 50.01 % in the controlled company and ownership, which is assumed if the ownership rights amount to more than 75 % of the company’s capital.

In the next step the gained sample of companies was researched in order to identify the Czech parent company and its subsidiaries in the EU Member States. Secondly, needed data from the proposed apportionment under CCCTB were researched from the accounting reports of the sample of companies from the Amadeus Database.

The proposed formulary apportionment under CCCTB comprises three factor formula equally weighted according the factors of sales, labor and assets:

\[
ShareA = \left( \frac{1}{3} S^A + \frac{1}{3} \left( \frac{1}{2} P^A + \frac{1}{2} E^A \right) + \frac{1}{3} A^A \right) \times CCCTB
\]  

(1)

where S represents sales, which are based on the sales of goods and services. P represents payroll, which includes the costs of salaries, wages, bonuses and all other employee compensation, including related pension and social security costs borne by the employer. E represents the number of employees, which are considered part of the group that pays the remuneration, unless they are under the control of a different group member, in which case they are considered part of that group. Employees are included if they are employed for at least three uninterrupted months. And finally, A represents assets, which include all fixed tangible assets, including buildings, airplanes and machinery, owned, rented, or leased by a group member.

However, the required information (operating revenue, number of employees, cost of employees, tangible fixed assets) from accounting reports of individual companies was often missing in Amadeus database. In order to eliminate the negative impacts of missing data on the research, three methods, namely the method of regression, imputation (hereinafter Im) and multiple imputation (hereinafter MI), were further researched in details with aim to identify the most suitable method, which would not significantly distort the allocation of the group tax base between the EU Member States.

As was mentioned above, data set includes 1190 entities and was distinguished on EU-West and EU-East, altogether it covers 58 NACE codes in EU-East and 9 NACE codes in EU-West. The largest NACE code was identified NACE 46 – Wholesale trade, except of motor vehicles and motorcycles, which covers 189 entities (15.88 %), and was research further in details.

There are used standard scientific research methods in the paper. The paper is dived into theoretical and empirical part. In the first theoretical part the method of description and analysis was applied in order to research the current state of research with respect to the formula apportionment and tax consolidation. The empirical part of the paper presents the empirical analysis of the individual missing data methods. And further, in order to decide about the most suitable method which will lead to the least distortion of the allocation of the group tax base between the EU Member States, the method of comparative analysis was applied. Finally in the conclusions the method of induction, deduction and synthesis was applied.

3.1. Regression and multiple imputation

The regression method is a basic method for estimation of missing data. Formulas 2–4 based on the linear regression model were used for estimation of individual missing data, namely number of employees, operating revenue and payroll, as follow:

\[
No.Employees \_imputed = koeficient \beta_0 + TFA * koeficinnet \beta_1
\]  

(2)
where as independent variables were used tangible fixed assets (hereinafter TFA) for estimation of number of employees and operating revenue, and number of employees for estimation of payroll, see table 1 below

Table 1. Determination of independent variables for individual formulas

<table>
<thead>
<tr>
<th>Formula</th>
<th>Regression</th>
<th>MI</th>
<th>TMI*</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. Employees (2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Operating_revenue (3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Payroll (4)</td>
<td></td>
<td></td>
<td>Number of employees</td>
</tr>
</tbody>
</table>

* TMI presents multiple imputation with independent variables in the form of tangible fixed assets.

Source: own processing

The same formulas were used for estimation of missing data through multiple imputation, specifically multiple imputation without independent variables (hereinafter MI), and multiple imputation with independent variables (hereinafter TMI). The multiple imputation estimated the missing data on the base of Markov Chain Monte Carlo method, which assumed arbitrary missing data. The number of iterations of 20 was used for both methods to avoid correlations and to remove noises from estimations, as state Schafer and Graham (2002). Then the Multivariate normal regression was made, where in case of MI, number of employees, operating revenue and payroll were identified as imputed variables. Contrary with TMI, where was also added independent variable in the form of TFA.

3.2. Imputation

Another method that can be used as long as the data set provides partial information, is a single imputation. This method replaces the missing items with plausible values and allows to proceed with the analysis rather than to discard the unit (in our case company) entirely.

Imputation has several desirable features. It is potentially more efficient than case deletion, which discards units whose information is incomplete and is used by default in many statistical programs. Further, this method retains the full sample and if the observed data contain useful information for predicting the missing values, an imputation procedure can make use of this information and maintain high precision.

The method of imputation is based on the observed data of 1190 companies in the same industry for EU-East and EU-West. Furthermore, missing operating revenue, number of employees, and payroll were imputed using ratios of the factor to assets for companies in which both variables are observed, with respect of methodology used by Clien, Neubig, Philips, Sanger and Walsh (2010). Companies reporting fixed assets as not available were excluded from the imputation.

Missing operating revenue amounts were imputed using reported tangible fixed asset data and the ratio of observed average operating revenue to the tangible fixed assets for other companies in the same industry:

\[
\text{Operating revenue} = (A\text{OperR} \div \text{ATFA}) \times \text{TFA}_\text{reported}
\]  

Missing employee data were imputed based on the reported tangible fixed assets of the company and the ratio of observed average numbers of employees to tangible fixed assets for the other companies in the same industry:

\[
\text{No. Employees}_\text{imputed} = (A\text{NoE} \div \text{ATFA}) \times \text{TFA}_\text{reported}
\]
Missing cost of employment data were imputed based on the imputed employee headcount and the ratio of observed average payroll cost to employee headcount for other companies in the same industry:

\[
Payroll = \left(\frac{APayr}{ANoE}\right) \times \text{No. Employees}_{\text{imputed}}
\]  

(7)

4. Results

The decision, which method is the most suitable for estimation of missing data and currently is not significantly distorting the allocation of the group tax base between the EU Member States, was based on the comparison of their average deviations from real data.

Firstly the real data set of 1190 companies was deleted and new values were estimated through the missing data methods, namely regression, imputation, multiple imputation, as was described above. Then the deviations from reality were determined for each variables (operation revenue, number of employees, payroll). Finally, the average deviations were calculated.

As can be seen in the table 2 below, the method of imputation presents better results of average deviations as well as multiple imputation methods. The best estimates, i.e. the lowest average deviations are reached in case of number of employees, contrast with the operating revenue, where the worst estimates are reached. However, the worst estimated values were determined through regression method, due to the fact this method was excluded from further research.

<table>
<thead>
<tr>
<th>Missing data methodology</th>
<th>Operating Revenue (th. EUR)</th>
<th>No. Employees</th>
<th>Payroll (th. EUR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average deviation from reality</td>
<td>-18,742.37</td>
<td>-934,531.66</td>
<td>-17,920,098.50</td>
</tr>
<tr>
<td>Imputation (Im)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average deviation from reality</td>
<td>170.51</td>
<td>-19.49</td>
<td>-873.52</td>
</tr>
<tr>
<td>Multiple imputation (MI)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average deviation from reality</td>
<td>-2,504.90</td>
<td>-13.02</td>
<td>-2,630.81</td>
</tr>
<tr>
<td>Multiple imputation (TMI)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average deviation from reality</td>
<td>-3,874.12</td>
<td>-16.86</td>
<td>-994.17</td>
</tr>
</tbody>
</table>

Source: own processing, data from Amadeus database.

As regards to the research of deviations in details, it was needed to compare all individual values of variables with real data. For this purpose was used the largest NACE code 46 – Wholesale trade, except of motor vehicles and motorcycles, which covers 189 entities, almost of 16% of data set. Reached average deviations are presented in table 3 below.

<table>
<thead>
<tr>
<th>Missing data methodology</th>
<th>Operating Revenue (th. EUR)</th>
<th>No. Employees</th>
<th>Payroll (th. EUR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Imputation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average deviation from reality</td>
<td>833.16</td>
<td>-7.31</td>
<td>-740,982</td>
</tr>
<tr>
<td>Multiple imputation (MI)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average deviation from reality</td>
<td>-2,457.35</td>
<td>-13.37</td>
<td>-2,676.12</td>
</tr>
<tr>
<td>Multiple imputation (TMI)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average deviation from reality</td>
<td>-3,816.55</td>
<td>-16.93</td>
<td>-1,028.43</td>
</tr>
</tbody>
</table>

Source: own processing, data from Amadeus database.
As can be seen, the best estimations are also reached by imputation method. The lowest average deviations are presented in case of number of employees, the highest one in case of operating revenue. However, if individual values are compared, the results are different.

Firstly, the comparison of imputed operating revenue data with real data is presented in the following figure 1. As can be seen MI and TMI methods reached similar results, however, they did not copy of the real data trend. Unlike the Im method is doing this in almost cases.

Secondly, the comparison of imputed number of employees data with real data is presented in the figure 2 below. As can be seen all imputation methods reached similar results and copies the real data trend in almost cases. However, this result is not repeated in the last comparison of imputed payroll data with real data that is presented in the figure 3 below. The Im method reached better results in almost cases than others.

Fig. 1. Comparison of imputed operating revenue data with real data in NACE 46.
Source: own processing, Amadeus database.

Secondly, the comparison of imputed number of employees data with real data is presented in the figure 2 below. As can be seen all imputation methods reached similar results and copies the real data trend in almost cases. However, this result is not repeated in the last comparison of imputed payroll data with real data that is presented in the figure 3 below. The Im method reached better results in almost cases than others.
Fig. 2. Comparison of imputed No. Employees data with real data in NACE 46.
Source: own processing, Amadeus database.

Fig. 3. Comparison of imputed payroll data with real data in NACE 46.
Source: own processing, Amadeus database.
Given these facts and based on the overall evaluation of methods (see Table 4 below) it can be concluded, that the most suitable methods of missing data imputation, which would not significantly distort the allocation of the group tax base between the EU Member States, can be considered the imputation method. This method used the ratios of the factor to assets for imputation of missing values.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Operating revenue</th>
<th>No. Employees</th>
<th>Payroll</th>
<th>Operating revenue</th>
<th>No. Employees</th>
<th>Payroll</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Imputation (Im)</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Multiple imputation (MI)</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Multiple imputation (TMI)</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>3</td>
</tr>
</tbody>
</table>

*4 the worst, 1 the best, **based on NACE 46 only, figure 1–3.
Source: own processing.

5. Conclusion

The paper aimed to research which missing data method is the most suitable for CCCTB determination – i.e. the research the method which would not lead to the significantly distortion of the allocation of the group tax base between the EU Member States. The research of three basic missing data methods, namely regression, imputation, multiple imputation with and without independent variable, was performed on the sample of 1,190 companies from Amadeus Database. The performed analysis clearly shows the best results are reached through imputation method. Its average deviations for operating revenue and payroll estimations are the lowest than at other methods as well as the individual deviations. The average deviations reached the value 170.51 th. EUR for operating revenue and −873.52 th. EUR for payroll costs. Only the average deviation for the number of employees reached worse result through this method, specifically −19.49, unlike MI (−13.02) and TMI (−16.86). However, based on the comparison of individual deviations the imputation method is much closer to the real data, therefore the imputation method should not lead to significantly distortion of the CCCTB determination.

It can be concluded that the most suitable method from the three researched missing data methods, namely regression, imputation and multiple imputation, is the imputation method using the ratios of the factor to assets for imputation of missing values.

Acknowledgements

The paper is the result of the GA ČR no. 13-21683S „The quantification of the impact of the introduction of Common Consolidated Corporate Tax Base on the budget revenues of the Czech Republic“.

References

Amadeus Database, Bureau van Dijk. Update 227, August 2013.


