

An Exact Analytical Grossing-Up Algorithm for Tax-Benefit Models

Miroslav Verbič, Mitja Čok and Tomaž Turk
University of Ljubljana, Faculty of Economics,
Kardeljeva ploščad 17, 1000 Ljubljana, Slovenia

Keywords: deductive data imputation, household budget survey, microsimulation, tax-benefit models

Received: November 27, 2014

In this paper, we propose a grossing-up algorithm that allows for gross income calculation based on tax rules and observed variables in the sample. The algorithm is applicable in tax-benefit microsimulation models, which are mostly used by taxation policy makers to support government legislative processes. Typically, tax-benefit microsimulation models are based on datasets, where only the net income is known, though the data about gross income is needed to successfully simulate the impact of taxation policies on the economy. The algorithm that we propose allows for an exact reproduction of a missing variable by applying a set of taxation rules that are known to refer to the variable in question and to other variables in the dataset during the data generation process. Researchers and policy makers can adapt the proposed algorithm with respect to the rules and variables in their legislative environment, which allows for complete and exact restoration of the missing variable. The algorithm incorporates an estimation of partial analytical solutions and a trial-and-error approach to find the initial true value. Its validity was proven by a set of tax rule combinations at different levels of income that are used in contemporary tax systems. The algorithm is generally applicable, with some modifications, for data imputation on datasets derived from various tax systems around the world.

Povzetek: Članek predstavlja algoritem obrutenja, ki omogoča izračunavanje bruto dohodkov iz neto dohodkov ob širokem naboru davčnih pravil različnih davčnih sistemov. Algoritem omogoča reproduciranje manjkajočih spremenljivk in je široko uporaben pri mikrosimulacijskem modeliranju.

1 Introduction

There are various techniques for data imputation, which give to the researcher an opportunity to remedy the situation when the dataset is not complete. This does not come without costs, since data imputation can easily introduce biased parameter estimates in statistical applications [1, 2], or in other domains [3, 4]. Imputation techniques rely on deterministic and stochastic approaches, mostly under the assumption that the variable in question is in some way related to other variables under investigation. In this paper, we are exploring a case of deductive approach [5], a possibility to estimate a missing variable by applying a set of rules for which it is known that they refer to the variable in question and to other variables in the dataset. This set of rules might be enforced in various contexts, for instance by legislation, government policy, or other institutional or social constraints. If there is a consistent set of rules, which are enforced in practice, and the rules are comprehensive, the researcher could develop a formal algorithm with respect to the rules and variables in the dataset, which would allow for the data imputation of the missing variable.

Let us consider the case of household budget survey (hereinafter HBS) datasets. HBS surveys are implemented at the national level of EU member states [6], where taxpayers report their net income for different income sources (e.g. wages, rents, pensions), as well as

socio-economic data, which enable estimations of tax allowances and tax credits. HBS datasets are most valuable in many microsimulation tax-benefit models. Such models are standard tools in academia, in financial industry, and for underpinning everyday policy decisions and government legislative processes [7, 8, 9, 10].

Gross income represents a starting point for any tax simulation (including tax-benefit modelling), but HBS datasets are usually reporting only net amounts. As noted in [7], one possibility to generate gross income is the statistical approach based on information on both net and gross income. Using this information, a statistical model can be developed that yields estimates of net/gross ratios. These estimates are then applied to net incomes in order to compute gross amounts.

The second known technique is the iterative algorithm that exploits the tax and contribution rules already built into tax-benefit models to convert gross income into net income [8, 9]. The procedure takes different levels of gross income for each observation, applies them to the tax rules, calculates the net income, and compares it with the actual net income as long as the gross income fits the actual net income within approximation limits.

Both techniques, namely the statistical approach and the iterative algorithm, give gross income values that are estimates and not the actual gross income values. The

task is not trivial, since modern tax systems include various rules for taxation and their combinations, and they usually involve a bracketing system for one or more parameters. Involvement of bracketing systems (and especially their combination) means that the calculation of net income from gross income is analytically non-reversible function.

In this paper, we are presenting a solution to this problem, namely an algorithm that enables a full restoration of the gross income value. The algorithm includes a set of analytical inversions combined with a trial-and-error approach to deal with bracketing system combinations. The proposed algorithm allows for the calculation of gross income from net income for a broad set of taxation possibilities, where only information on net income is available, along with information on tax reliefs. The algorithm is feasible in cases of proportional and progressive tax schedules of personal income tax (hereinafter PIT) and social security contributions. It also covers tax allowances as well as tax credits. It is thereby generally applicable to contemporary tax systems around the world.

The validity and accuracy of the proposed algorithm was tested by its application to a synthetic sample of taxpayers using an artificial system of personal income tax (PIT) and social security contributions. A comparison of gross income, calculated from net income using the proposed technique, with the initial gross income demonstrates the complete accuracy of the algorithm.

The rest of the paper is organized as follows. In Section 2, we analyse taxation rules that are used in contemporary taxation systems. The analysis is a basis for the formalization of the imputation algorithm, which is explained in Section 3, including detailed solutions and proofs for various combinations of tax rules and bracketing systems. A test of the validity and accuracy of the proposed algorithm is presented in Section 4. In the Conclusion, the proposed algorithm is presented in its full form, which can be directly applied in practice.

2 Analysis of taxation rules

Gross income is a starting point for the taxation of personal income, as to which we can distinguish three basic approaches [11, 12]: comprehensive income tax, dual income tax, and flat tax. Under a comprehensive income tax system, all types of labour and capital incomes are taxed uniformly by the same progressive tax schedule. A dual income tax system retains progressive rates on labour income, while introducing a proportional tax rate on capital income, e.g. the Scandinavian dual income tax [13]. The third option, which has been dominating income tax reforms in Eastern Europe [14, 15], is the flat-tax concept, although it is noted that this concept has not been implemented in any country in Western Europe [6].

Hereby we follow the most comprehensive procedure for the taxation of gross income, which is presented in Table 1 and includes a combination of progressive tax schedules and flat rates, with the addition of tax allowances and tax credits.

Gross income
– Social security contributions ^a
▪ determined by social security contributions schedule
▪ set as a proportion of gross income
▪ given in absolute amounts
– Other costs related to the acquisition of net income
▪ determined by a schedule
▪ set as a proportion of gross income
▪ given in absolute amounts
– Tax allowances
= Personal income tax base
× Personal income tax rate ^b
= Initial personal income tax
– Personal income tax credit
= Final personal income tax
Net income = Gross income – Social security contributions
– Final personal income tax

^a Employee social security contributions

^b Either a single (flat) tax rate or set by the tax schedule

Table 1: General procedure for taxing gross income.

Table 1 contains the general procedure for the taxation of gross income. From gross income, employee social security contributions and other costs related to the acquisition of income (e.g. travel allowances or standardized costs set as a proportion of gross income) are deducted. Further, the tax allowances are subtracted and the tax base is obtained, which is subject to a PIT calculation using the tax schedule or a proportional (flat) tax rate. In this way, the initial PIT is calculated, which could be further reduced by a tax credit in order to calculate the final PIT and the net income.

From the taxation point of view, other costs related to the acquisition of income have consequences identical to social security contributions or tax allowances. Therefore, our further development implicitly incorporates these costs into the concepts of social security contributions and tax allowances.

When the schedules are applied, the PIT schedule or social security contributions schedule consists of a number of tax brackets with different marginal tax rates. The amount of PIT is calculated from the tax base according to the PIT schedule. Likewise, the amount of social security contributions is determined by gross income and the social security contributions schedule.

In general, at the annual level tax bases from different income sources are summed up into a single tax base, which is subject to a single-rate schedule, and then the final annual PIT is calculated. An alternative option is a dual-tax system, where the PIT is calculated separately for different income sources (multiple-rate schedule).

The procedure from Table 1 covers the existing tax systems to a great extent. In several OECD countries [16], the employee social security contributions are determined by the schedule (i.e. Austria, France) or set as a proportion of gross income (i.e. Spain, Norway), while social security contributions set by absolute amount are not very common and can be found, e.g., in Slovenia for certain categories of the self-employed.

The algorithm applies the logic of social security contributions to the *Other costs related to the acquisition*

of net income (i.e. cost connected with the real estate maintenance in the case of taxing income from rents) and tax allowances (i.e. for children or interest of housing loan allowances), which are found across the tax systems. The algorithm also covers the case of social security ceiling (i.e. in Austria, Germany). Regarding the calculation of PIT, the algorithm covers the prevailing progressive PIT schedule, as well as flat-tax systems (e.g. in Hungary or Bulgaria).

However, the algorithm (more precisely, equations that cover specific combinations of tax parameters) has to be adapted to certain country specifics, which are not explicitly set out by the procedure from Table 1. For example, if social security contributions are not included in the PIT base, the gross income shall be calculated by the algorithm assuming that social security contributions are zero. Another example refers to above mentioned social security contribution ceiling. In this case, a zero rate tax bracket of social security contribution schedule above the set ceiling should be applied in an appropriate equation that is suitable to the specific combination of tax parameters of the particular country.

The algorithm hereinafter is derived for the case when there is only one PIT instrument and one SSC instrument, thus two instruments in total. However, in actual fiscal systems there are cases when two or more PIT or SSC instruments are applied to a single income source at the same time. In these cases, the net to gross conversion is more complicated, since a “compression” of two (or more) PIT or SSC instruments should be done into a single PIT or SSC instrument.

During the year, when a particular income source is paid out, the advance (in-year) PIT is usually paid at the time of disbursing the income source. This advance PIT is taken into account once the final annual PIT is calculated (i.e. the advance PIT is consolidated with the annual PIT). This procedure is called withholding.

Understanding the mechanism of (in-year) advance PIT is important when we are dealing with survey data, such as HBS datasets. In a typical survey, respondents report their net income from different income sources for a certain period of the year, when their income sources are only subject to (in-year) advance PIT. In order to calculate the overall annual gross income, the reported net income from different income sources should be initially grossed-up using the algorithms that take account of various rules of advance (in-year) PIT and social security contributions for each income source separately. The focus of our paper is the development of these grossing-up algorithms for different income sources. Once the grossed-up amounts from different income sources are calculated, they can be summed up into overall taxpayers’ annual gross income, which is the starting point for building a microsimulation model.

Thus, the calculation of gross income from net income of a single income source (i.e. calculation of gross wages from given net wages) thus depends on different combinations of tax parameters from Table 1, which are described in detail as an algorithm in the paper. For example, in a case when gross wages are subject of: (a) progressive PIT schedule, (b) social

security contributions, which are set as a proportion of gross wage with a ceiling, (c) tax allowances, and (d) without tax credits (e.g. in-year taxation of wages in Croatia), then equation (9) should be applied. Since the ceiling of social security contributions is set, this implies that the applied value of social security contribution rate above the ceiling should be zero.

Table 1 can be transformed into the following expression:

$$N = G - S - PIT, \quad (1)$$

where N and G represent net and gross income, respectively, S is the sum of social security contributions, and PIT is the personal income tax.

Social security contributions S are a function of gross income. Similarly, PIT is a function of the tax base, which is the difference between gross income (reduced by social security contributions) and tax allowances, TA . This can be generalized as follows:

$$N = G - f_s(G) - f_{PIT}(G - f_s(G) - TA). \quad (2)$$

Function $f_s(G)$ can be defined in practice in different ways. A common approach is to use a schedule system, but it can also be defined as a proportion of gross income or as an absolute amount.

In practice, function $f_{PIT}(G - f_s(G) - TA)$ is usually defined by a schedule system (different from the schedule system for social security contributions). As mentioned, function f_{PIT} can also incorporate the concept of a tax credit.

Our task is to estimate gross income G from expression (2) from the known values of N and TA and from a set of constraints that are usually given by social security contributions and PIT schedule systems, or by other legislative rules. The combination of two schedule systems makes solving equation (2) for G particularly challenging. The solution we propose in this paper has a trial-and-error nature. The idea is to prepare a set of all possible (PIT and social security contribution) bracket combinations. Then, we calculate for each taxpayer ‘candidate’ gross income values for each bracket combination, calculate net incomes from these candidate gross incomes, and compare the results to the starting value of net income. The gross income candidate that fits (or equals) the net income is the true gross income value. The fit is exact, i.e. we find the actual gross income in a non-iterative way.

The following section describes the construction and design of the procedures we propose to deal with different income sources taxed by different rules. The general setup of the grossing-up algorithm is explained. Sections 3.2 and 3.3 set out a detailed examination of various taxation rules for social security contributions and tax crediting, together with the proposed grossing-up procedures for specific tax rule combinations.

3 Data imputation algorithm

In this section, we explore a general setup where the tax system involves a combination of the following elements: (1) a social security contributions schedule, (2) a PIT schedule, and (3) tax allowances. This general setup forms the basis for development of the proposed algorithm. In the next steps, we incorporate other tax complexities, i.e. other rules for calculating social security contributions and various rules for determining tax credits.

Function $f_s(G)$ can be expanded by the rules of the social security contributions schedule to:

$$f_s(G) = S = Sr_s(G - L_s) + \sum_{j=1}^{s-1} Sr_j(H_j - L_j), \quad (3)$$

i.e. for each bracket, social security contributions are equal to the social security contributions marginal rate Sr_s , multiplied by the difference between gross income G and the lower bracket margin (s denotes the social security contributions bracket). This amount is added to the social security contributions, which are collected for all 'lower' brackets (i.e. brackets from 1 to $s-1$). H_s and L_s denote the upper and lower social security bracket margins.

Similarly, function f_{PIT} can be expanded by the rules of the schedule system for PIT to:

$$f_{PIT}(G) = PIT = Tr_b(G - f_s(G) - TA - L_b) + \sum_{i=1}^{b-1} Tr_i(H_i - L_i), \quad (4)$$

where Tr_b is the marginal tax rate for PIT bracket b , L_b is the lower margin for bracket b , Tr_i is the marginal rate for bracket i , and H_i and L_i are the upper and lower margins of bracket i , respectively.

By combining (3) and (4), we obtain:

$$\begin{aligned} N_{sb} = G_{sb} & - \left(Sr_s(G_{sb} - L_s) + \sum_{j=1}^{s-1} Sr_j(H_j - L_j) \right) - \\ & - \left(Tr_b \left(G_{sb} - \left(Sr_s(G_{sb} - L_s) + \sum_{j=1}^{s-1} Sr_j(H_j - L_j) \right) - \right. \right. \\ & \left. \left. - TA - L_b \right) + \sum_{i=1}^{b-1} Tr_i(H_i - L_i) \right), \end{aligned} \quad (5)$$

The above equation holds for an individual taxpayer, when PIT was calculated by the tax authorities in such a way that social security bracket s corresponds to gross income G , and PIT bracket b corresponds to $(G - S - TA)$. Since we do not know the actual G and S , we cannot directly establish, which PIT and social security contributions brackets (and corresponding marginal rates) were actually used for each individual taxpayer by the tax authorities.

By reordering expression (5), we can express gross income as follows:

$$G_{sb} = \frac{N_{sb} - Sr_s L_s - Tr_b L_t + Sr_b Tr_b L_s}{(Sr_s - 1)(Tr_b - 1)} - \frac{Tr_b TA + Tr_b \Sigma_s - \Sigma_s - \Sigma_b}{(Sr_s - 1)(Tr_b - 1)}, \quad (6)$$

where

$$\Sigma_b = \sum_{i=1}^{b-1} Tr_i(H_i - L_i)$$

and

$$\Sigma_s = \sum_{j=1}^{s-1} Sr_j(H_s - L_s).$$

Following our general trial-and-error scheme, the grossing-up algorithm is as follows:

1. For each statistical unit, calculate the matrix with $K \cdot B$ candidate gross incomes as its elements, according to equation (6):

$$\begin{Bmatrix} G_{11} & \cdots & G_{K1} \\ \vdots & \ddots & \vdots \\ G_{1B} & \cdots & G_{KB} \end{Bmatrix},$$

where K and B are the number of social security contributions brackets and the number of PIT brackets, both defined by the PIT and social security contributions system, respectively, and where $k = 1, \dots, K$ and $l = 1, \dots, B$.

2. Calculate the net incomes from the matrix of candidate gross incomes according to the tax rules:

$$\begin{Bmatrix} N_{11} & \cdots & N_{K1} \\ \vdots & \ddots & \vdots \\ N_{1B} & \cdots & N_{KB} \end{Bmatrix}.$$

3. In the above matrix, find the net income N_{kl} , which is equal to the starting net income for this individual taxpayer: $N = N_{kl}$.
4. The actual gross income G for this individual taxpayer is then: $G = G_{kl}$.

In the next subsections, we discuss the following extensions to this general setup: (1) social security contributions are not determined by the schedule system,

but as a proportion of gross income or as an absolute amount (Section 3.1), and (2) tax credits are included according to various rules for their determination (Section 3.2).

3.1 Variations of social security contributions

In the following section, we extend the general setup to include cases where social security contributions are not determined by a schedule, but as a proportion of gross income or as an absolute amount.

3.1.1 Social security contributions as a proportion of gross income

When social security contributions are set as a proportion of gross income, equation (2) can be rewritten as:

$$N = (1 - Sr)G - f((1 - Sr)G - TA), \quad (7)$$

where Sr is the rate of social security contributions, expressed as a proportion of the gross income. By simplifying equation (5), we obtain:

$$N_b = (1 - Sr)G_b - \left(Tr_b((1 - Sr)G_b - TA - L_b) + \sum_{i=1}^{b-1} Tr_i(H_i - L_i) \right), \quad (8)$$

which holds for each PIT bracket b . By reordering, we can express the gross income with the equation:

$$G_b = \frac{N_b - Tr_b TA - Tr_b L_b + \sum_b}{(1 - Sr) - Tr_b(1 - Sr)}. \quad (9)$$

From here, we can proceed according to the general setup, outlined above.

3.1.2 Social security contributions as an absolute amount

When social security contributions are set as an absolute amount, we can simplify equation (5):

$$N_b = G_b - S - \left(Tr_b(G_b - S - TA - L_b) + \sum_{i=1}^{b-1} Tr_i(H_i - L_i) \right) \quad (10)$$

and by reordering we obtain:

$$G_b = \frac{N_b + S - Tr_b S - Tr_b TA - Tr_b L_b + \Sigma}{1 - Tr_b}. \quad (11)$$

From here, we can proceed according to the general setup, outlined above.

3.2 Grossing-up procedure when PIT is subject to a tax credit

A tax credit means that PIT is reduced by a certain amount (called a tax credit) and that the gross income source is effectively not taxed with the full PIT (the ‘initial PIT’), but with the PIT reduced by the amount of the tax credit (the ‘final PIT’). In practice, if a tax credit is calculated to be greater than the initial PIT, then net income N equals gross income G , as the net income cannot exceed the gross income (i.e. a tax credit can be as high as the initial PIT).

In various tax systems, a tax credit can be defined in three ways: (1) as a proportion of the initial PIT, (2) as a proportion of the gross income, or (3) as an absolute amount.

3.2.1 Tax credit as a proportion of the initial PIT

In general, we can express a tax credit as a proportion of the initial PIT as:

$$N = G - f_s(G) - f_{PIT}(G - f_s(G) - TA) + c_{PIT} \cdot f_{PIT}(G - f_s(G) - TA), \quad (12)$$

where c_{PIT} is the share of the tax credit in the initial PIT. Following the above, we can write:

$$\begin{aligned} N_{sb} = G_{sb} - & \left(Sr_s(G_{sb} - L_s) + \sum_{j=1}^{s-1} Sr_j(H_j - L_j) \right) - \\ & - \left(Tr_b \left(G_{sb} - \left(Sr_s(G_{sb} - L_s) + \sum_{j=1}^{s-1} Sr_j(H_j - L_j) \right) - \right. \right. \\ & \left. \left. - TA - L_b \right) + \sum_{i=1}^{b-1} Tr_i(H_i - L_i) \right) + \\ & + c_{PIT} \left(Tr_b \left(G_{sb} - \left(Sr_s(G_{sb} - L_s) + \sum_{j=1}^{s-1} Sr_j(H_j - L_j) \right) - \right. \right. \\ & \left. \left. - TA - L_b \right) + \sum_{i=1}^{b-1} Tr_i(H_i - L_i) \right), \end{aligned} \quad (13)$$

which holds for a specific combination of social security contributions and PIT brackets. Solving (13) for G , we obtain:

$$\begin{aligned} G_{sb} = & \frac{N_{sb} - Sr_b L_s - Tr_b L_t + c Tr_b L_t + Tr_b Sr_b L_s}{(1 - Sr_s)(Tr_b(c_{PIT} - 1) + 1)} - \\ & - \frac{c_{PIT} Tr_b Sr_b L_s + Tr_b TA - c_{PIT} Tr_b TA}{(1 - Sr_s)(Tr_b(c_{PIT} - 1) + 1)} - \\ & - \frac{\Sigma_s(Tr_b(1 - c_{PIT}) - 1) + (c_{PIT} - 1)\Sigma_b}{(1 - Sr_s)(Tr_b(c_{PIT} - 1) + 1)}. \end{aligned} \quad (14)$$

When the tax credit is set as a proportion of the initial PIT and social security contributions are defined by a schedule, the above equation should be used instead of expression (6) in the general setup.

Tax credit as a proportion of the initial PIT and social security contributions as a proportion of the gross income

Where social security contributions are set as a proportion of the gross income, the general procedure can be simplified. In this case, the net income can be expressed as:

$$N = (1 - S_r)G - f((1 - S_r)G - TA) + c_{PIT} \cdot f((1 - S_r)G - TA). \quad (15)$$

The following equation holds for a particular tax bracket b :

$$N_b = (1 - S_r)G_b - (Tr_b((1 - S_r)G_b - TA - L_b) + \sum_{i=1}^{b-1} Tr_i(H_i - L_i)) + c_{PIT}(Tr_b((1 - S_r)G_b - TA - L_b) + \sum_{i=1}^{b-1} Tr_i(H_i - L_i)). \quad (16)$$

By reordering we obtain:

$$G_b = \frac{N_b - (1 - c_{PIT})(Tr_b L_b + Tr_b TA - \Sigma_b)}{(1 - S_r)(1 - Tr_b + c_{PIT} Tr_b)}. \quad (17)$$

Thus, when a tax credit is set as a proportion of the initial PIT and social security contributions are set as a proportion of the gross income the above equation should be used instead of expression (6) in the general setup.

Tax credit as a proportion of the initial PIT and social security contributions as an absolute amount

If social security contributions are set as an absolute amount, we can redefine equation (10) to incorporate tax credit as a proportion of the initial PIT:

$$N_b = G_{sb} - S - (Tr_b(G_b - S - TA - L_b) + \sum_{i=1}^{b-1} Tr_i(H_i - L_i)) + c_{PIT}(Tr_b(G_b - S - TA - L_b) + \sum_{i=1}^{b-1} Tr_i(H_i - L_i)) \quad (18)$$

and by reordering we obtain:

$$G_b = \frac{N_b + S + (c_{PIT} - 1)(Tr_b L_b + Tr_b S + Tr_b TA - \Sigma_b)}{(1 - Tr_b + c_{PIT} Tr_b)}. \quad (19)$$

Thus, when a tax credit is set as a proportion of the initial PIT and social security contributions are set in an absolute amount, the above equation should be used instead of expression (6) in the general setup.

3.2.2 Tax credit as a proportion of the gross income

If the amount of a tax credit is defined as a proportion of the gross income, the net income calculation can be formalized as:

$$N = G - f_s(G) - f_{PIT}(G - f_s(G) - TA) + c_G \cdot G, \quad (20)$$

where c_G is the tax credit share of the gross income. For clarity, we can denote the initial PIT as:

$$PIT_I = f_{PIT}(G - f_s(G) - TA) \quad (21)$$

and the final PIT as:

$$PIT_F = f_{PIT}(G - f_s(G) - TA) - c_G \cdot G. \quad (22)$$

Since a tax credit can be as high as the initial PIT, the following rule applies:

$$N = \begin{cases} G - f_s(G) - PIT_F & \text{if } c_G \cdot G < PIT_I; \\ G - f_s(G) & \text{if } c_G \cdot G \geq PIT_I. \end{cases} \quad (23)$$

Due to this rule, the gross income cannot be easily estimated from net income N and tax allowances TA , as PIT_I and $c_G \cdot G$ are not known at this stage. The rule implies that the actual calculation of net income N for each taxpayer was done by the tax authorities either by:

$$N = G - f_s(G) - f_{PIT}(G - f_s(G) - TA) + c_G \cdot G \quad (24)$$

when $c_G \cdot G < f_{PIT}(G - f_s(G) - TA)$, or by:

$$N = G - f_s(G) \quad (25)$$

when $c_G \cdot G \geq f_{PIT}(G - f_s(G) - TA)$.

When we are interested in G , we can use these two approaches in reverse fashion (calculating G and not N), but we do not know which one, (24) or (25), is correct.

Let us consider the case when we calculate G for a particular taxpayer from known values of N , TA and the PIT schedule (as in Table 1), once by using the rule expressed in equation (24) and once by using the rule expressed in (25). We obtain two estimates for the taxpayer's gross income G :

$$G' = N + f_s(G) + f_{PIT}(G - f_s(G) - TA) - c_G \cdot G \quad (26)$$

and

$$G'' = N + f_s(G). \quad (27)$$

If the net income N for this particular taxpayer was actually calculated according to expression (24), this inequality holds true:

$$(N + f_s(G) + f_{PIT}(G - f_s(G) - TA) - c_G \cdot G) > (N + f_s(G)), \quad (28)$$

since $c_G \cdot G < f_{PIT}(G - f_s(G) - TA)$ must hold. By using (26) and (27), we obtain:

$$G' > G''. \quad (29)$$

The proper value of gross income G is G' , since net income N for this particular taxpayer was actually calculated according to expression (24).

Let us consider the opposite case where net income N for our taxpayer was actually calculated (by the tax authorities) according to (25). In this case, we can write:

$$(N + f_s(G) + f_{PIT}(G - f_s(G) - TA) - c_G \cdot G) \leq (N + f_s(G)), \quad (30)$$

since $c_G \cdot G \geq f_{PIT}(G - f_s(G) - TA)$ must hold. By using (26) and (27), we obtain:

$$G' \leq G''. \quad (31)$$

The proper value of gross income G in this case is G'' . Following (29) and (31), we can conclude that in both cases the highest value of G' and G'' is the one that actually holds:

$$G = \max(G', G'') \quad (32)$$

or

$$G = \max((N + f_s(G) + f_{PIT}(G - f_s(G) - TA) - c_G \cdot G), (N + f_s(G))). \quad (33)$$

For the construction of a general setup in the case of tax credits given as a proportion of the gross income, where social security contributions and PIT are calculated according to their schedules, we need to express equation (33) in a more exact way, for a specific combination of social security contribution and PIT brackets. The specific form for equation (26) is then:

$$\begin{aligned} N_{sb} = G_{sb} - & \left(Sr_s(G_{sb} - L_s) + \sum_{j=1}^{s-1} Sr_j(H_j - L_j) \right) - \\ & - \left(Tr_b \left(G_{sb} - \left(Sr_s(G_{sb} - L_s) + \sum_{j=1}^{s-1} Sr_j(H_j - L_j) \right) - \right. \right. \\ & \left. \left. - TA - L_b \right) + \sum_{i=1}^{b-1} Tr_i(H_i - L_i) \right) - c_G G_{sb} \end{aligned} \quad (34)$$

and for equation (27):

$$N_{sb} = G_{sb} - Sr_s(G_{sb} - L_s) + \sum_{j=1}^{s-1} Sr_j(H_j - L_j). \quad (35)$$

From expression (34) we obtain:

$$\begin{aligned} G'_{sb} = & \frac{N_{sb} - Sr_s L_s - Tr_b \Sigma_s - Tr_b L_b}{1 + c_G - Sr_b - Tr_b + Sr_b Tr_b} + \\ & + \frac{Sr_b Tr_b L_s - Tr_b TA + \Sigma_s + \Sigma_b}{1 + c_G - Sr_b - Tr_b + Sr_b Tr_b} \end{aligned} \quad (36)$$

and from expression (35):

$$G''_{sb} = \frac{N_{sb} - Sr_b L_s + \Sigma_s}{1 - Sr_b}. \quad (37)$$

According to expression (32), we can establish the right value for gross income G_{sb} :

$$G_{sb} = \max(G'_{sb}, G''_{sb}). \quad (38)$$

Tax credit as a proportion of the gross income and social security contributions as a proportion of the gross income

Where social security contributions are set as a proportion of the gross income, the calculation of net income N for each taxpayer was done by the tax authorities either by:

$$N = (1 - Sr)G - f_{PIT}((1 - Sr)G - TA) + c_G \cdot G \quad (39)$$

when $c_G \cdot G < f_{PIT}((1 - Sr)G - TA)$, or by:

$$N = (1 - Sr)G \quad (40)$$

when $c_G \cdot G \geq f_{PIT}((1 - Sr)G - TA)$. The reasoning is similar to that above where we constructed equations (34) and (35). These two equations can be simplified since we only have one social security contributions rate Sr , and we obtain:

$$\begin{aligned} N_b = & (1 - Sr)G_b - \left(Tr_b \left((1 - Sr)G_b - TA - L_b \right) + \right. \\ & \left. + \sum_{i=1}^{b-1} Tr_i(H_i - L_i) \right) + c_G G_b \end{aligned} \quad (41)$$

and

$$N_b = (1 - Sr)G_b. \quad (42)$$

From this, we obtain two solutions for G_b :

$$G'_b = \frac{N_b - Tr_b L_b - Tr_b TA + \Sigma_b}{1 + c_G - Sr - Tr_b + Sr Tr_b} \quad (43)$$

and

$$G_b'' = \frac{N_b}{1 - S_r}, \quad (44)$$

which should be used in the general setup instead of (36) and (37), respectively. Again, the matrix of candidate solutions is one-dimensional (a vector for gross income candidates, i.e. one value for each PIT bracket), since there is only one social security contributions rate.

Tax credit as a proportion of the gross income and social security contributions as an absolute amount

In this case, the procedure can follow the same principles we used to construct equations (34) and (35). Since social security contributions are now set as an absolute amount, these two equations can be simplified:

$$N_b = G_b - S - \left(Tr_b (G_b - S - TA - L_b) + \sum_{i=1}^{b-1} Tr_i (H_i - L_i) \right) + c_G G_b \quad (45)$$

and

$$N_b = G_b - S. \quad (46)$$

The gross income for both cases can then be calculated from:

$$G_b' = \frac{N_b + S - Tr_b L_b - Tr_b S - Tr_b TA + \Sigma_b}{1 + c_G - Tr_b} \quad (47)$$

and

$$G_b'' = N_b + S, \quad (48)$$

which should be used in the general setup instead of (36) and (37), respectively.

3.2.3 Tax credit as an absolute amount

If the amount of a tax credit is defined as an absolute amount, the procedure is similar to the one described in Section 2.3.2. The net income can be expressed as:

$$N = G - f_s(G) - f_{PIT}(G - f_s(G) - TA) + C, \quad (49)$$

where C is the amount of the tax credit. The initial PIT is the same as in Section 2.3.2, equation (21), and the final PIT is:

$$PIT_F = f_{PIT}(G - f_s(G) - TA) - C. \quad (50)$$

The following rule applies:

$$N = \begin{cases} G - f_s(G) - PIT_F & \text{if } C < PIT_I; \\ G - f_s(G) & \text{if } C \geq PIT_I. \end{cases} \quad (51)$$

If net income N for a particular taxpayer was actually calculated (by the tax authorities) according to $C < PIT_I$ in (51), this inequality holds true:

$$(N + f_s(G) + f_{PIT}(G - f_s(G) - TA) - C) > (N + f_s(G)), \quad (52)$$

since $C < f_{PIT}(G - f_s(G) - TA)$ must hold. In the opposite case, i.e. if net income N was calculated according to $C \geq PIT_I$, then:

$$(N + f_s(G) + f_{PIT}(G - f_s(G) - TA) - C) \leq (N + f_s(G)), \quad (53)$$

since $C \geq f_{PIT}(G - f_s(G) - TA)$ must hold.

Following a similar reasoning to that in Section 2.3.2, we can conclude that the actual gross income G for a particular taxpayer must be:

$$G = \max((N + f_s(G) + f_{PIT}(G - f_s(G) - TA) - C), (N + f_s(G))) \quad (54)$$

Quantity G' can be estimated from:

$$N_{sb} = G_{sb} - \left(Sr_s (G_{sb} - L_s) + \sum_{j=1}^{s-1} Sr_j (H_j - L_j) \right) - \left(Tr_b \left(G_{sb} - \left(Sr_s (G_{sb} - L_s) + \sum_{j=1}^{s-1} Sr_j (H_j - L_j) \right) - TA - L_b \right) + \sum_{i=1}^{b-1} Tr_i (H_i - L_i) \right) + C \quad (55)$$

and solving for G_{sb} :

$$G_{sb}' = \frac{N_{sb} + C + Sr_s L_s + Tr_b L_b - Sr_s Tr_b L_s}{(Sr_s - 1)(Tr_b - 1)} + \frac{Tr_b TA + Tr_b \Sigma_s - \Sigma_s - \Sigma_b}{(Sr_s - 1)(Tr_b - 1)}, \quad (56)$$

whereas the estimation of G'' is already explained in (36) and (38).

We can conclude that in cases where the amount of a tax credit is defined as an absolute amount, the general setup is the same as that described in Section 2.3.2, except for equation (36), which should be substituted by equation (56).

Tax credit as an absolute amount and social security contributions as a proportion of the gross income

Where when social security contributions are set as a proportion of the gross income, the calculation of net income N for each taxpayer was done by the tax authorities either by:

$$N = (1 - S_r)G - f_{PIT}((1 - S_r)G - TA) + C \quad (57)$$

when $C < f_{PIT}((1 - S_r)G - TA)$, or by:

$$N = (1 - S_r)G \quad (58)$$

when $C \geq f_{PIT}((1 - S_r)G - TA)$. The reasoning is similar to that in Section 2.3.2. Equation (41) can be rewritten in the following form:

$$N_b = (1 - S_r)G_b - \left(Tr_b((1 - S_r)G_b - TA - L_b) + \sum_{i=1}^{b-1} Tr_i(H_i - L_i) \right) + C \quad (59)$$

and from this, we can obtain:

$$G'_b = \frac{N_b - C - Tr_b L_b - Tr_b TA + \Sigma_b}{(S_r - 1)(Tr_b - 1)}, \quad (60)$$

which should be used in the general setup instead of (43), whereas equation (44) also applies in this case for obtaining G''_b . Again, the matrix of candidate solutions is one-dimensional (a vector for gross income candidates, i.e. one value for each PIT bracket).

Tax credit as an absolute amount and social security contributions as an absolute amount

In this case, the procedure can follow the same principle we introduced in Section 2.3.2. Since social security contributions are given as an absolute amount, equation (34) can be written in this way:

$$N_b = G_b - S - \left(Tr_b(G_b - S - TA - L_b) + \sum_{i=1}^{b-1} Tr_i(H_i - L_i) \right) + C, \quad (61)$$

whereas equation (46) also holds in the case social security contributions are set as an absolute amount. The gross income can then be calculated from (61) as:

$$G'_b = \frac{N_b - C + S - Tr_b L_b - Tr_b S - Tr_b TA + \Sigma_b}{1 - Tr_b}, \quad (62)$$

which should be used in the general setup instead of (36), together with (48), which was derived from (46).

3.3 Algorithm in its full form

For clarity, the grossing-up procedure that we developed in the above sub-sections is given below, including all combinations of the taxation rules that we described in the subsections following the basic setup at the beginning of Section 3.

1. For each statistical unit, calculate the matrix of $K \cdot B$ candidate gross incomes according to equation (6):

$$G = \begin{Bmatrix} G_{11} & \cdots & G_{K1} \\ \vdots & \ddots & \vdots \\ G_{1B} & \cdots & G_{KB} \end{Bmatrix},$$

where K and B are the number of social security contribution brackets and the number of tax brackets, both defined by the tax and social security contribution systems, respectively, where $k = 1, \dots, K$ and $l = 1, \dots, B$.

Formulas for specific combinations of taxation rules can be found in Table 2.

In cases where only the tax schedule system is used and social security contributions related to the acquisition of the income are set as one parameter, the above matrix of candidate gross incomes becomes a vector $\{G_1, \dots, G_l, \dots, G_B\}$.

2. Calculate the net incomes from the matrix of candidate gross incomes according to the tax rules:

$$N = \begin{Bmatrix} N_{11} & \cdots & N_{K1} \\ \vdots & \ddots & \vdots \\ N_{1B} & \cdots & N_{KB} \end{Bmatrix}$$

or

$$\{N_1, \dots, N_l, \dots, N_B\}.$$

3. In the above matrix, find net income N_{kl} (or N_l), which is equal to the starting net income:

$$N = N_{kl} \text{ (or } N = N_l \text{)}.$$

4. The actual gross income G is then:

$$G = G_{kl} \text{ (or } G = G_l \text{)}.$$

System without tax credits			Equation
I	Schedule for social security contributions	$G_{sb} = \frac{N_{sb} - Sr_s L_s - Tr_b L_t + Sr_b Tr_b L_s - Tr_b TA - Tr_b \Sigma_s + \Sigma_s + \Sigma_b}{(Sr_s - 1)(Tr_b - 1)}$	(6)
II	Social security contributions as a proportion of gross income	$G_b = \frac{N_b - Tr_b TA - Tr_b L_b + \Sigma_b}{(1 - Sr) - Tr_b (1 - Sr)}$	(9)
III	Social security contributions as an absolute amount	$G_b = \frac{N_b + S - Tr_b S - Tr_b TA - Tr_b L_b + \Sigma}{1 - Tr_b}$	(11)
Tax credit as a proportion of the initial PIT			Equation
IV	Schedule for social security contributions	$G_{sb} = \frac{N_{sb} - Sr_b L_s - Tr_b L_t + c_{PIT} Tr_b L_t + Tr_b Sr_b L_s - c_{PIT} Tr_b Sr_b L_s - Tr_b TA}{(1 - Sr_s)(Tr_b (c_{PIT} - 1) + 1)} + \frac{c_{PIT} Tr_b TA - \Sigma_s (Tr_b (1 - c_{PIT}) - 1) - (c_{PIT} - 1) \Sigma_b}{(1 - Sr_s)(Tr_b (c_{PIT} - 1) + 1)}$	(14)
V	Social security contributions as a proportion of gross income	$G_b = \frac{N_b - (1 - c_{PIT})(Tr_b L_b + Tr_b TA - \Sigma_b)}{(1 - Sr)(1 - Tr_b + c_{PIT} Tr_b)}$	(17)
VI	Social security contributions as an absolute amount	$G_b = \frac{N_b + S + (c_{PIT} - 1)(Tr_b L_b + Tr_b S + Tr_b TA - \Sigma_b)}{(1 - Tr_b + c_{PIT} Tr_b)}$	(19)
Tax credit as a proportion of gross income			Equation
VII	Schedule for social security contributions	$G_{sb} = \max(G'_{sb}, G''_{sb})$	(38)
		$G'_{sb} = \frac{N_{sb} - Sr_s L_s - Tr_b \Sigma_s - Tr_b L_b + Sr_b Tr_b L_s - Tr_b TA + \Sigma_s + \Sigma_b}{1 + c_G - Sr_b - Tr_b + Sr_b Tr_b}$	(36)
		$G''_{sb} = \frac{N_{sb} - Sr_b L_s + \Sigma_s}{1 - Sr_b}$	(37)
VIII	Social security contributions as a proportion of gross income	$G_b = \max(G'_b, G''_b)$	(38)
		$G'_b = \frac{N_b - Tr_b L_b - Tr_b TA + \Sigma_b}{1 + c_G - Sr - Tr_b + Sr Tr_b}$	(43)
		$G''_b = \frac{N_b}{1 - Sr}$	(44)
IX	Social security contributions as an absolute amount	$G_b = \max(G'_b, G''_b)$	(38)
		$G'_b = \frac{N_b + S + Tr_b L_b - Tr_b S - Tr_b TA + \Sigma_b}{1 - c_G - Tr_b}$	(47)
		$G''_b = N_b + S$	(48)
Tax credit as an absolute amount			Equation
X	Schedule for social security contributions	$G_{sb} = \max(G'_{sb}, G''_{sb})$	(38)
		$G'_{sb} = \frac{N_{sb} + C + Sr_s L_s + Tr_b L_b - Sr_s Tr_b L_s + Tr_b TA + Tr_b \Sigma_s - \Sigma_s - \Sigma_b}{(Sr_s - 1)(Tr_b - 1)}$	(56)
		$G''_{sb} = \frac{N_{sb} - Sr_b L_s + \Sigma_s}{1 - Sr_b}$	(37)
XI	Social security contributions as a proportion of gross income	$G_b = \max(G'_b, G''_b)$	(38)
		$G'_b = \frac{N_b - C - Tr_b L_b - Tr_b TA + \Sigma_b}{(Sr - 1)(Tr_b - 1)}$	(60)
		$G''_b = \frac{N_b}{1 - Sr}$	(44)
XII	Social security contributions as an absolute amount	$G_b = \max(G'_b, G''_b)$	(38)
		$G'_b = \frac{N_b - C + S - Tr_b L_b - Tr_b S - Tr_b TA + \Sigma_b}{1 - Tr_b}$	(62)
		$G''_b = N_b + S$	(48)

Table 2: Equations for specific combinations of taxation rules.

4 Results and discussion

Table 3 presents a summary of all possible social security contributions and tax credit combinations explored in Section 3. In reality, for any income source one of these combinations is applicable. Parallel to this, the PIT schedule system and tax allowances in absolute amounts are assumed.

Our approach can also be applied to flat PIT systems (i.e. with a single proportional PIT rate). If this is a case, we apply only one PIT bracket with a positive marginal PIT rate. Where tax allowances are not set as absolute amounts, they can be expressed as an ‘additional layer’ of social security contributions.

To test for the validity and accuracy of the proposed algorithm, we created a synthetic sample of 10,000 taxpayers with a normally distributed gross income, where the mean gross income was 50,000 mu (monetary units) and the standard deviation was 11,500 mu. We assumed the following tax parameters:

1. The PIT schedule includes three brackets:
 - 0 – 20,000 mu, a 15% marginal PIT rate;
 - 20,000 – 50,000 mu, a 25% marginal PIT rate;
 - over 50,000 mu, a 45% marginal PIT rate.
2. The social security schedule includes three brackets:
 - 0 – 10,000 mu, a 17% marginal rate;
 - 10,000 – 40,000 mu, a 20% marginal rate;
 - over 40,000 mu, a 0% marginal rate.
3. Social security contributions as a proportion of the gross income were set at 22%.
4. Social security contributions as an absolute amount were set at 500 mu.
5. Tax allowances were set at an absolute amount of 2,000 mu.
6. The amounts of tax credits were given as follows: 13% of the gross income, 6% of the initial PIT, or 200 mu.

These parameters were applied to the entire population of taxpayers according to the general procedure for taxing gross income (Table 1) and the specific combination of tax rules from Table 3.

In the first step, we generated the amount of gross income for each taxpayer. In the second step, we calculated the net income according to combinations I to XII (from Table 3) of the tax rules, as is done in practice by tax authorities.

In the third step, we applied the proposed grossing-up algorithm to combinations I–XII for each taxpayer. Finally, we compared the grossed-up income with the initial gross income.

	Schedule for social security contributions	Social security contributions as a proportion of the gross income	Social security contributions as an absolute amount
System without tax credits	I	II	III
Tax credit as a proportion of the initial PIT	IV	V	VI
Tax credit as a proportion of the gross income	VII	VIII	IX
Tax credit as an absolute amount	X	XI	XII

Table 3: Summary of tax rules combinations (detailed equations are given in Table 4).

The comparison of the gross income, calculated from the net income using the grossing-up algorithm, with the initial gross income demonstrates the complete accuracy of the algorithm for all income types.

As an example, we can repeat the steps for an individual taxpayer with a gross income equal to 49,433.10 mu. In the second step, we calculated the net income amount for all 12 combinations of the tax rules (see Table 4).

	Schedule for social security contributions	Social security contributions as a proportion of the gross income	Social security contributions as an absolute amount
System without tax credits	33,799.80 (I)	31,418.30 (II)	39,199.80 (III)
Tax credit as a proportion of the initial PIT	34,275.80 (IV)	31,846.70 (V)	39,783.80 (VI)
Tax credit as a proportion of the gross income	40,226.10 (VII)	37,844.60 (VIII)	45,626.10 (IX)
Tax credit as an absolute amount	33,999.80 (X)	31,618.30 (XI)	39,399.80 (XII)

Table 4: Net income for a chosen taxpayer with $G = 49,433.10$ mu.

For each net income from Table 4, we applied the grossing-up algorithm (i.e. equations from Table 2). According to the technique, several gross income candidates were calculated for each of these net incomes.

Due to space limitations, here we (arbitrarily) present the gross income candidates for net income VII:

$$G^{VII} = \begin{Bmatrix} 48,465.20 & 50,134.40 & 48,465.20 \\ 49,907.60 & 51,371.40 & 49,907.60 \\ 47,926.10 & 49,433.10 & 47,926.10 \end{Bmatrix}.$$

To each of these gross income candidates we applied the taxation rules (in this case the combination of taxation rules VII) and calculated the net income:

$$N^{VII} = \begin{Bmatrix} 39,374.40 & 40,843.20 & 39,374.40 \\ 40,643.70 & 41,931.80 & 40,643.70 \\ 38,900.00 & 40,226.10 & 38,900.00 \end{Bmatrix}.$$

By comparing the elements of matrix N^{VII} with the net income for a combination of tax rules VII from Table 4, which equals 40,226.10 (VII), we identified the matching element in the third row and the second column. The corresponding gross income in matrix G^{VII} equals 49,433.10, which is identical to the initial gross income of this particular taxpayer. In other words, for this combination of tax rules (VII), the proposed grossing-up algorithm is accurate.

We repeated such tests for all 12 tax rule combinations and for 10,000 individual cases.

5 Conclusion

In this paper, we presented a detailed construction of deterministic data imputation algorithm. In particular, we described an exact grossing-up algorithm for calculating the pre-tax income from data, which are only available in net (after-tax) form, and proved its successfulness, since it leads to a complete data reconstruction.

Contemporary tax systems are rich in complexity, and some of tax rules combinations might not be covered by our technique. However, we believe that the general architecture of our proposition is sound and flexible enough to incorporate (with some modifications) additional, locally specific tax rules.

In general, if a set of rules that relate to the variables under investigation could be assembled, researchers and policy makers can perform data imputation in deterministic fashion, and construct the algorithm for the exact analytical generation of the missing values.

In future research efforts, a framework for feasibility assessment of such approach could be envisioned, which would employ estimates on rules' consistency and complexity on the one hand, and measures of the quality of replicated data on the other hand.

References

- [1] Rancourt, E. (2007). Assessing and dealing with the impact of imputation through variance estimation. *Statistical Data Editing: Impact on Data Quality*. New York: United Nations.
- [2] Rueda, M. M., Gonzalez, S. & Arcos, A. (2005). Indirect methods of imputation of missing data based on available units. *Applied Mathematics and Computation* 164: 249–261.
- [3] Smirlis, Y. G., Maragos, E. K. & Despotis, D. K. (2006). Data envelopment analysis with missing values: An interval DEA approach. *Applied Mathematics and Computation* 177: 1–10.
- [4] Raghunathan, T. E., Lepkowski, J. M., van Hoewyk, J. & Solenberger, P. (2001). A Multivariate Technique for Multiply Imputing Missing Values Using a Sequence of Regression Models. *Survey Methodology* 27: 85–95.
- [5] Franklin, S. & Walker, C. (2003). *Survey methods and practices*. Ottawa: Statistics Canada.
- [6] Fuest, C., Peichl, A. & Schaefer, T. (2008). Is a flat tax reform feasible in a grown-up democracy of Western Europe? A simulation study for Germany. *International Tax and Public Finance* 15: 620–636.
- [7] Immervoll, H. & O'Donoghue, C. (2001). Imputation of gross amounts from net incomes in household surveys: an application using EUROMOD, *EUROMOD Working Papers* EM1/01. Colchester: ISER-Institute for Social and Economic Research.
- [8] D'Amuri, F. & Fiorio, C. V. (2009). Grossing-Up and Validation Issues in an Italian Tax-Benefit Microsimulation Model. *Econpubblica Working Paper*, 117, Milano: University of Milan.
- [9] Betti, G., Donatiello, G. & Verma, V. (2011). The Siena microsimulation model (SM2) for net-gross conversion of EU-silc income variables. *International Journal of Microsimulation* 4: 35–53.
- [10] ISER – Institute for Social and Economic Research, <https://www.iser.essex.ac.uk/euromod> (April 16th, 2012)
- [11] OECD – Organisation for Economic Co-operation and Development (2006). Reforming Personal Income Tax. *Policy Brief March*. Paris: OECD.
- [12] Zee, H. H. (2005). Personal income tax reform: Concepts, issues, and comparative country developments. *IMF Working Paper* 87. Washington: International Monetary Fund.
- [13] Sorenson, P. B. (2005). Dual income tax: Why and how? *FinanzArchiv* 61: 559–586.
- [14] Ivanova, A., Keen, M. & Klemm, A. (2005). The Russian 'flat tax' reform. *Economic Policy* 20: 397–444.
- [15] Moore, D. (2005). Slovakia's 2004 tax and welfare reforms. *IMF Working Paper* 133, Washington: International Monetary Fund.
- [16] OECD – Organisation for Economic Co-operation and Development (2013). *Taxing Wages 2011–2012*. Paris: OECD.