EUROMOD

WORKING PAPER SERIES

EUROMOD Working Paper No. EM1/01

IMPUTATION OF GROSS AMOUNTS FROM NET INCOMES IN HOUSEHOLD SURVEYS. AN APPLICATION USING EUROMOD

Herwig Immervoll and Cathal O'Donoghue

June 2001

Imputation of Gross Amounts from Net Incomes in Household Surveys. An Application using EUROMOD[1]

Herwig Immervoll* and Cathal O'Donoghue

EUROMOD Working Paper No. EM1/01
June 2001

Abstract
Household micro-datasets often do not contain information on gross incomes. We present an algorithm which exploits the tax- and contribution rules built into tax-benefit models to convert net income information into gross amounts. Using EUROMOD, a multi-country tax-benefit model covering all fifteen countries of the European Union, net-to-gross conversions can be performed for a large number of countries utilising existing models of relevant fiscal rules. The algorithm takes into account all relevant complexities of tax- and contribution rules and can, thus, produce much more accurate results than statistical models which estimate net-to-gross ratios using only a few explanatory variables. Among the features of the algorithm is the ability to distinguish between different individuals in the same household. Even if individuals' incomes are taxed jointly, the algorithm is able to approximate separate net-to-gross factors for individuals in the same fiscal unit. This is possible since EUROMOD can accurately assign people to appropriate fiscal units. In addition, it is in certain cases possible to produce different net-to-gross ratios for different income components. We undertake a case study to illustrate the importance of deriving separate net-to-gross factors for different individuals within a household/fiscal unit and for different income sources of the same individual.

JEL Classification: C81; C63

Keywords: Microsimulation; Imputation; Income; Net/Gross

*Author for Correspondence.

---

[1]Immervoll is Research Associate at the Microsimulation Unit, Department of Applied Economics at the University of Cambridge and Research Fellow at the European Centre for Social Welfare Policy and Research. O'Donoghue is Research Associate at the Microsimulation Unit, Department of Applied Economics at the University of Cambridge. This paper was written as part of the EUROMOD project, financed by *Targeted Socio-Economic Research* programme of the European Commission (CT97-3060). We are grateful for access to micro-data from the Socio-Economic Panel Living in Luxembourg (PSELL 2) made available by Centre d'Études de Populations, de Pauvreté et de Politiques Socio-Économiques / International Networks for Studies in Technology, Environment, Alternatives, Development (CEPS/INSTEAD), Luxembourg. We are also grateful to Frédéric Berger, Carlos Farinha Rodrigues, Holly Sutherland and Géza Tarcali for helpful comments and suggestions. Any errors are the authors'.

Imputation of Gross Amounts from Net Incomes in Household Surveys. An Application using EUROMOD

Herwig Immervoll and Cathal O'Donoghue

1. Introduction

Income variables in household survey data are frequently recorded net of income taxes and other levies on income, notably social insurance contributions. Indeed, for some of the main purposes of these household surveys, such as analysing the economic well-being of households, net income is the type of information which is ultimately required. However, for many research tasks gross income information is essential. Important examples are the calculation of tax wedges and effective tax rates, issues related to the distribution or determinants of 'primary' income or studies trying to derive aggregate statistics from micro-data (or reconcile the two).

Another application of household micro-data where the lack of gross incomes can be a major problem is tax-benefit microsimulation. These computer models contain detailed social and fiscal policy rules as they apply to households and are widely used in governments and by academic researchers. By using them in conjunction with representative household micro-data, it is possible to perform detailed analyses of the distributional, incentive and revenue consequences of existing or prospective policy measures. Apart from their main use as tools for analysing the effects of fiscal and social policy measures, these models are frequently used to impute tax- and benefit variables that are not elicited in the survey questionnaire. Weinberg (1999), for example, finds that, of seventeen data sources surveyed, taxes and/or social insurance contributions are imputed in five. Since gross incomes are one of the main determinants of the amount of taxes and transfer payments that households pay/receive, tax-benefit models require detailed information on gross income as one of their principal inputs. Providing this information for all fifteen country datasets used in a novel European tax-benefit model (EUROMOD) has been the main motivation for the present paper (for more information on EUROMOD see Immervoll et al, 1999). For nine of these countries, some or all income components are not available 'gross' in the underlying original dataset.[2] As a result, methods needed to be devised to convert net incomes to gross amounts.

In what follows, we briefly discuss various approaches that can and have been used to convert income information from net to gross. In section 3, we describe an algorithm which can be used in conjunction with existing tax-benefit models and is thus able to exploit the detailed representation of fiscal rules contained in them. Several complications which, depending on the specific fiscal system at hand, can arise in using this approach are also discussed here. Section 4 then illustrates the use of the method by explaining how it can be applied to a Luxembourg household panel. The results of this exercise are presented in section 5 where we also attempt to test the various features of our approach by contrasting the results with those of an alternative ('statistical') method. Lastly, we provide an indication of the level of

[2] Austria, Belgium, Denmark, France, Greece, Italy, Luxembourg, Portugal, Spain.

accuracy that can be obtained using our algorithm by comparing model results from a tax-benefit microsimulation exercise to external reference figures.

2. Net-to-gross conversion: approaches

Reflecting the structure of existing tax and contribution rules, the conversion of net incomes to gross amounts is a complex problem. For some purposes, it may be sufficient to reduce the complexity by not taking into account all tax- and contribution relevant characteristics of the individuals represented in the data. Such an approach is, for example, taken by Eurostat in deriving a 'net/gross ratio' which is supplied in the European Community Household Panel User Database (ECHP UDB) along with the income variables (which are all recorded as 'net').[3] Their method is an example of what we would like to term a 'statistical' approach. Based on information on both net *and* gross income (which are *both* available for a subset of individuals), a statistical model is formulated. Fitting this model to the data then yields estimates of net/gross ratios for those cases (the majority) where survey respondents failed to provide information on both net and gross incomes.

The 'statistical' approach raises several issues. First, the variation which can be reproduced by such a model is necessarily rather limited. One crucial input into the statistical model, the number of respondents who supply *both* net and gross incomes, can vary depending on the data source one is concerned with. However, even in cases where the fraction of people with recorded values of both net and gross incomes is relatively large, the number of observations on which the model is estimated is much too small to support the incorporation of all or even most determinants of income tax and contribution payments.[4] This is because the complexity of tax and benefit systems in general precludes any precise statistical modelling of taxes and transfer payments. For example, tax-benefit systems are highly non-linear which makes it very difficult to produce an adequate functional form, even with a large number of data points. Instead, 'deterministic' tax-benefit models are usually required to provide reliable results for all types of households, including the more 'unusual' cases (large family sizes, unstable work patterns, etc.), which, in terms of social and fiscal policy, are often the more interesting ones.

A second important problem with the 'statistical' approach as used for deriving the ECHP net-to-gross factor is the non-differential treatment of different recipients and different types of income *within* the household. Clearly, one single household-wide conversion factor for all persons and income components within the household ignores the potentially very large differences in tax- and contribution burdens that different individuals and income types in one single household can be subject to. In many cases, for instance, individuals who are entirely exempt from taxation will live together with high-income earners. Since, in the majority of cases, taxes and contributions depend on the incomes of units smaller than the household, these variations of net-to-gross ratios are highly relevant. In our case study below, we attempt to shed further light on the problems associated with the 'statistical' approach.

---

[3] See Eurostat (1996).

[4] For example, in the Austrian ECHP User Database (1996 wave), employment incomes are available in both net and gross for about 2322 employees out of a total 3403.

It has been recognised that "[...] the estimated net/gross ratios [resulting from the 'statistical' approach] are a rather simplistic solution to a complex problem" and that "some data users may wish to do their own conversion, using more sophisticated approaches based on country-specific modelling."[5] While an exact computation of a net-to-gross factor is conceptually preferable to the 'statistical' approach, the implementation of such an alternative in practice is resource intensive. The most direct solution to the problem would be an analytic inversion of all relevant tax and contribution rules prevailing in the year to which net income data refer to. While there are some computational problems, this method has successfully been implemented in practice.[6] The main disadvantage relates to the effort required to build such a model. International experiences with tax-benefit model building demonstrate that an accurate and reliable representation of complex fiscal rules is a very time consuming task (McCrae 1999, Mot 1992). Furthermore, unlike tax-benefit models, which have multiple uses, the inversion of such a model is only useful for imputation purposes. Since tax and contribution rules are subject to frequent adaptations, the development and maintenance of a dedicated net-to-gross model, which would need to be changed almost every year, is often not feasible. In cases where, as with the ECHP, one institution is ultimately responsible for the provision of data from a range of countries, this task would be even more demanding since each tax/contribution system would need to be represented in a separate model.

Given these difficulties, an attractive route would be to exploit models of tax and contribution rules that already exist. This is true for tax-benefit models. Rather than analytically inverting the tax and contribution system, it is therefore useful to explore whether the structure of *existing* tax-benefit models could be adapted to utilise the built-in fiscal algorithms for performing net-to-gross conversions (contrary to their usual purpose which is computing taxes, contributions and other instruments based on gross income information). Because EUROMOD incorporates the tax-benefit rules of fifteen countries, the pay-off of adding this capability to this model seemed especially attractive. But in principle, the net-to-gross functionality can be added to any existing tax-benefit model.

The basic idea is to adopt an iterative approach. For each household in the data where net incomes are recorded, this approach would 'try' different levels of gross incomes. For each 'provisional' gross income which is being 'tried', one would in a next step apply all relevant tax and contribution rules and subtract simulated taxes and contributions from (provisional) gross income to arrive at a simulated net income. It is then possible to compare whether this simulated net income is a good approximation of net income as recorded in the original data. If it is, then the provisional gross income can be seen as a good approximation of actual gross income. The following section explains the approach in more detail.

[5] Eurostat (1999), p.6.

[6] Since taxes and contributions are not necessarily monotonuously increasing functions of gross income (i.e., parts of the 'budget line' may be flat or downward sloping), expressing gross income as a function of net income may not result in a unique solution. Depending on the tax and contribution rules one has to resort to more or less restrictive assumptions to make the inverted system solvable. Berliri et al (1999) demonstrate how this can be done in the case of Italy.

3. The iterative method

EUROMOD is designed to incorporate fifteen very different tax-benefit systems in a framework which is flexible enough to accommodate a multitude of country specific concepts and rules (Immervoll and O'Donoghue 2001). Similarly, in developing the generalised net-to-gross algorithm presented here we have tried to take into account the large variation of different countries' tax and contribution rules. We therefore had to consider extensions to existing methodologies already in use for national purposes in some EUROMOD partner institutions.[7] As a result, the algorithm presented here may be somewhat more complicated than would be required for any specific fiscal system. In return, it should be possible to use this algorithm for many different countries. The objective of the method is to find gross income, which, after applying the tax and contribution rules, results in the correct net income. The basic algorithm consists of the following steps:

a. As a first estimate for imputed gross income ($grossY_0$), the algorithm simply takes the net income supplied in the data (the original net income, *ori_netY*), i.e., in iteration $x = 0$ the estimate of the net-to-gross factor k_0 is simply 1[8]:

$$k_x = 1 \qquad x = 0 \tag{1}$$

$$grossY_x = k_x * ori\_netY \tag{2}$$

b. Applying the tax-benefit rules (as implemented in the tax-benefit model[9]) to this gross income, we produce a new value for net income (the simulated net income, $netY_0$), which will, as long as taxes and contributions are positive, be smaller than *ori_netY*.[10]

$$netY_x = (1-t_x) * grossY_x \tag{3}$$

where t_x is the effective average tax rate.

c. Test if the exit condition (4) is met. Given potential discontinuities in the *effective* tax- and contribution schedules, it is theoretically possible that the algorithm does not converge. For cases where no solution is found after a certain number of iterations (which can, of course, be freely chosen), the algorithm automatically starts over with a (randomly) different starting value $grossY_0$.

$$\left|(ori\_netY - netY_x) / ori\_netY\right| < |\delta| \qquad \text{exit condition} \tag{4}$$

If the value of $netY_0$ is not sufficiently 'close' to *ori_netY* (i.e., as long as the deviation is not less than δ), a new estimate of gross income, $grossY_1$, is produced. $grossY_1$ is generated (equation 2) using a revised estimate of the net-to-gross factor, k_1, as follows

[7] See Berliri et al (1999) for the method applied to Italian data; Levy and Mercader-Prats (1999) for Spain and Bourguignon and Sastre (1999) for France. A similar method is in use in Portugal (Farinha-Rodrigues, 2000).

[8] Or any other number which is seen to be a good starting point (e.g., the expected average net-to-gross factor).

[9] Depending on the comprehensiveness of the underlying tax-benefit model, it is possible to take into account alternative choices which may affect individuals' tax burdens. For countries where it exists, optional joint taxation is, for example, accounted for in EUROMOD by assuming that individuals chose the alternative which minimises the family's overall tax burden.

[10] Of course, the tax and contributions rules used here need to be those, which prevailed during the period to which the net income values in the data relate.

$$k_{x+1} = k_x\ (ori\_netY\ /\ netY_x) \qquad (5)$$

$grossY_1$ is again subjected to taxes and contributions (equation 3) producing a new value $netY_1$, etc.. This process is repeated until the value of $netY_x$ converges to the net income as recorded in the data (*ori_netY*).

In applying this approach in practice, however, several complications arise. First, it may, in cases where a fiscal unit consists of more than one person, be difficult to produce separate net-to-gross factors for each member of the unit. Also, since different income components may be taxed at different rates, a separate net-to-gross factor should be computed for each of these components. Finally, because the algorithm depends on convergence towards a solution, it is important to analyse whether the specific process of convergence is a source of systematic errors. We now discuss these three issues in turn.

3.1. Different net-to-gross factors for different individuals

Fiscal rules relate to certain fiscal units. For generating appropriate net-to-gross factors, it is important to correctly specify the required fiscal unit to solve for (e.g., 'married couple', 'family', individual', 'household', etc.). However, if different rules *within one* country are based on different fiscal units then different types of income may be aggregated across different persons before being subject to the tax or contribution. It could, for example, be the case that employment income is assessed at the individual level for the purpose of computing social insurance contributions while the base of an income tax system based on the joint income of married couples would include the employment income of the spouse as well. In cases where different tax/contribution rules are based on different fiscal units it is, thus, no longer clear which unit of assessment to solve the net-to-gross algorithm for. Primarily, the algorithm should produce results, which will be correct for the chosen unit as a whole (i.e., the sum of all gross incomes across the unit will represent the "true" picture). In addition, we have extended the algorithm to also approximate the correct distribution of gross incomes *within* the unit. This is possible by (a) making assumptions on how, in a joint system, individuals share the tax/contribution burdens; and (b) simultaneously varying the net-to-gross factors for all individuals in the unit. Regarding (a), we suggest as the default assumption (which can of course be adapted if needed) that in a multi-person fiscal unit, the shares in the overall tax burden are equal to the shares in the overall tax base. In order to make (b) feasible, we not only compare original and simulated net incomes for the unit as a whole but also at the individual level. The algorithm compares, for each individual *i*, original net income ${}_i ori\_netY$ with simulated net income ${}_i netY$ and forms a ratio of the two. The exit condition is that the difference between both the unit-wide *and* the individual original and simulated net incomes fall within the specified error range. In our implementation of the algorithm using EUROMOD, one can simply specify the largest fiscal unit relevant for the tax/contribution system at hand (in the above example this would be 'married couple'). The following equations show the extensions necessary for deriving individual net-to-gross factors in tax/contribution systems with multi-person fiscal units:

$${}_ik_x = 1 \qquad\qquad x = 0 \qquad\qquad (1.1)$$

$$_i grossY_x = {}_i k_x * {}_i ori\_netY \qquad (2.1)$$
$$_i netY_x = (1 - {}_i t_x) * {}_i grossY_x \qquad (3.1)$$
$$\left| ({}_i ori\_netY - {}_i netY_x) / ori\_netY \right| < \left| \delta \right| \qquad \text{exit condition} \qquad (4.1)$$
$$_i k_{x+1} = {}_i k_x \, ({}_i ori\_netY / {}_i netY_x) \qquad (5.1)$$

3.2. Different net-to-gross factors for different income components

For any individual, all income components, which are subject to taxes or contributions are increased by the same percentage (equation 2). This, however, assumes that all incomes are taxed at the same rate. Clearly, this does not always correspond to the true picture. For example, employment income will typically face social insurance contributions and income taxes, while investment or property income will only be subject to income tax. From the data, we know the net value of both. So if the different components are taxed in completely separate ways (e.g., by being subject to separate schedules, etc.) then there is no problem because one can simply convert property income using one factor and employment income using another. If, however, both components are assessed together for tax purposes and, therefore, enter the *same* tax base then separate net-to-gross conversion is not feasible. The reason for this is that we do not know how taxes or contributions are distributed between the different income components. This can be shown as follows. Using the example with employment income and property income the iterative procedure solves the following equation for a net-to-gross factor k:

$$net\_empY + net\_propY = k * (net\_empY * (1 - t_{IT,SIC}) + net\_propY * (1 - t_{IT})) \qquad (6)$$

Where *net_empY* and *net_propY* denote net employment income and net property income (as recorded in the original 'net' data) and t_{IT} and $t_{IT,SIC}$ are the effective taxes rates for income tax alone and income tax plus social insurance contributions, respectively. If we want to allow for separate net-to-gross factors k_{empY}, k_{invY} then this becomes

$$net\_empY + net\_propY = k_{empY} * net\_empY * (1 - t_{IT,SIC}) + k_{propY} * net\_propY * (1 - t_{IT}) \quad (7)$$

Clearly, due to two unknown variables k_{empY} and k_{propY}, (7) is not solvable. In general, therefore, the algorithm will produce one common k for all income components. However, in many cases it is still possible to improve on this result and approximate estimates of component specific net-to-gross factors. The first way to do this is to specify a fixed correction ratio of the different ks. For example, if one knows that, in most cases, the effective average tax burden on property income is about 80% of the burden on employment income (because property income is not subject to contribution payments), then one would specify a correction factor of 0.8 for property income. This ensures that in solving for k in the iterative procedure, the net-to-gross factor applied to property income is always 20% smaller than the net-to-gross factor applied to employment income. Clearly this type of 'correction' is a fairly rough approximation since the ratio between k_{empY} and k_{propY} will usually vary between individuals. However, depending on how the income information in the data has been collected, it may often be possible to get a better approximation. Often, income tax payable on employment income and replacement incomes is withheld at source. As a result, employment income is taxed separately during the year even if it is assessed together with all other taxable income components for the purpose of computing the 'final' end-of-year income tax. If net income from employment as reported in the data is net of withholding tax then there is no need to tax employment income together with other incomes (such as property income) for the purpose of net-to-gross conversion. The problem with more than one unknown k in equation (7) can thus be avoided. Instead, one can simply apply the contribution and withholding tax rules to convert any incomes subject to withholding tax from net to gross. Then in a second step, one can use the end-of-year tax rules to find estimates of net-to-gross factors for those income components which are taxable but not subject to the withholding tax. The case study below provides an example of this approach.

3.3. Estimation bias and non-unique solutions

Another potential problem is that the successive approximation of the net-to-gross factors ${}_ik_x$ can result in an estimation bias which depends on the progressivity of the tax and contribution rules. Combining (2), (3) and (5), one can show that

$$netY_x = (1\text{-}t_x) * ori\_netY \,/\, (1\text{-}\ t_{x\text{-}1}) \qquad (8)$$

In general, the average tax rate is a function of gross income. In the special case where all taxes and contributions are proportional, the average tax rate does not vary with income so that $t_x = t_{x\text{-}1}$. For example, if the tax rate was 20% regardless of gross income then the algorithm would produce the following:

1. $k_0 = 1$
2. $grossY_0 = k_0 * ori\_netY = ori\_netY$
3. $netY_0 = (1\text{-}t_0) * grossY_0 = 0.8 * ori\_netY$
4. $k_1 = k_0\ (ori\_netY \,/\, netY_0) = 1 / 0.8$
5. $grossY_1 = k_1 * ori\_netY = ori\_netY / 0.8$
6. $netY_1 = (1\text{-}t_1) * grossY_1 = ori\_netY$

Here, the algorithm would stop since with $netY_1 = ori\_netY$, the exit condition is met. Thus, an exact approximation of the net-to-gross factor (1/0.8) has been found. For progressive parts of the schedule, however, $t_{x\text{-}1} < t_x$, so that $netY_x$ will always remain smaller than the 'true' original net income. The algorithm will thus always converge from below. The reverse is true for regressive parts of the schedule. However, in the algorithm, one can control any such bias by specifying a convergence interval δ that is sufficiently small.

A further problem arises in the case of flat rate instruments such as self-employed social insurance contributions, which, in some countries are constant in value regardless of the income of the individual. In the net-to-gross procedure, small net incomes may thus become zero or negative if one takes net income (ori_netY) reported in the survey as the starting point for estimating gross income ($grossY_0$). So for example in equation (9), if the self-employed social contribution ($sesic$) is equal to or greater than the starting gross income ($grossY_0$), then the new net income ($netY_1$) will be zero or negative and thus the net-to-gross factor computed in equation (5) will be undefined or negative. However, it is relatively easy to deal with this problem. In cases where the sum of taxes and contributions is greater than our estimate of *grossY*, the value of the net-to-gross factor *must* be greater than 2. Thus, if a simulated net income goes to or below zero while original net income in the data is non-negative, then we multiply k by 2+e (equation 10).

$$netY_1 = grossY_0 - sesic \leq 0 \qquad sesic \geq grossY_0 \tag{9}$$

$$k_x = k_{x\text{-}1} * (2+e) \qquad ori\_netY > 0\ ;\ netY_{x\text{-}1} \leq 0 \tag{10}$$

A final point relates to the fact that there may not necessarily be a unique solution to the optimisation problem for all cases. Because of the complexity and non-linearities in tax systems, a number of different gross incomes may result in the same net output. In these cases, the resulting net-to-gross ratio simply reflects the first solution found by the algorithm.

4. A case study

This section provides an illustration of the above net-to-gross conversion method by applying it to one of the fifteen countries which are part of EUROMOD. We have chosen Luxembourg because the tax and contribution system in this country seemed best suited to test the algorithm's ability to deal with the various complications discussed in section 3. The household micro-data available for Luxembourg is the 1999 wave (containing 1998 incomes) of the Socio-Economic Panel Living in Luxembourg (PSELL 2).

The Luxembourg income tax system is a joint system where all tax bases in the family are taxed together. Married couples enjoy a tax advantage in that the family income base is divided by two before being subjected to the tax rate schedule. The resulting tax is then multiplied by two ("split taxation"). In a progressive tax system this method results in lower overall tax burdens in cases where the income bases of the spouses are different. The social insurance contributions (SICs) are, as usual, individual based. For details on both income tax and contributions the reader is referred to Berger and Borsenberger (2001).

In converting incomes from net to gross, a complication arises because the employment and replacement incomes are elicited net of withholding or 'pay-as-you-earn' (PAYE) tax. The rules for this tax are rather different than those for the final 'end-of-year' income tax. Most importantly, the tax base is assessed individually. Different deductions, tax schedules and tax credits apply depending on the individual's characteristics (income, employment status, age, family circumstance, etc). However, not all income components are subject to PAYE tax. Several types of income are only taxed under the 'end-of-year' tax. This is, for example, true for self-employment income, income from capital (property, savings, etc.), private pensions but also social assistance. At the end of the (tax) year, these incomes are assessed together with those subject to PAYE tax. The resulting 'final' tax is compared with the taxes already paid under the PAYE system with any shortfall being met by the taxpayer (any overpayments are refunded). We are thus faced with a situation where different net-to-gross factors should apply to different income components. However, as shown above (equation 7) the algorithm is not able to differentiate between different income components.

But, as also noted above, it may, in certain cases, be possible to approximate correct net-to-gross factors for more than one income component. Since PAYE and 'end-of-year' tax are not levied simultaneously but consecutively, we do not have to solve for both net-to-gross factors at once. Instead, it is possible to adopt a two-stage approach. First, we run the net-to-gross algorithm only for those incomes subject to PAYE tax (y_a). Applying, for each individual, the PAYE tax rules and the relevant SICs, we find, in an iterative fashion, the gross values for employment income, unemployment benefits and (public) pensions, $y_{ag} = k_a\, y_{an}$. Due to the individual nature of the withholding tax, this step is relatively straightforward since we do not have to differentiate between different people in the tax unit (as done in equations 1.1 to 5.1). Then, in a second step, we take these gross incomes from step 1 as given and apply the 'end-of-year' tax rules and the relevant SICs to find the net-to-gross factors for the remaining (net) income components y_{bn}. Clearly, the income tax paid on y_b is not independent of y_a since, for the purpose of the 'final' income tax, the tax base is $y_{ag} + y_{bg}$. In converting net incomes to gross, the iterative algorithm now only multiplies net incomes y_{bn} by the factor k_b. In other words, since the incomes y_a are already gross, we are now looking for the k_b which approximates the following

$$(y_{ag} + k_b\, y_{bn})\,(1 - t) = y_{ag} / k_b + y_{bn} \qquad (11)$$

where

$$t = t(y_{ag}, y_{bg}) = t(y_{ag}, y_{bn}, k_b)$$

Equation (11) is a slightly modified version of (4) where δ is zero and *ori_netY* has been substituted by $y_{ag} / k_b + y_{bn}$ because y_{ag} is already a gross amount.

In addition, to the separate treatment of y_a and y_b, the second step is complicated by the fact that the 'final' income tax is characterised by a joint tax base. As a result, there is, potentially, more than one person in each tax unit (family). We use the approach as outlined in section 3.1 to identify separate net-to-gross factors for each individual in the unit. Since we have two separate *k*s (relating to the two 'steps') for each individual, we get 2*n* net-to-gross factors for a tax unit (family) with *n* individuals.

5. Results

Since all representative household surveys in Luxembourg (and in many other countries) elicit net incomes, there is no direct way to assess the quality of the net-to-gross procedure presented here. Nevertheless, there seem to be two useful ways in which to present the results. One possibility is a descriptive analysis of the resulting net-to-gross factors and a comparison of these descriptives to those of alternative net-to-gross methods. The second alternative would be to examine the results of the tax-benefit model which uses the generated gross incomes as an input. For example, validating aggregates of the different tax or transfer instruments against external reference information would provide a useful picture of the quality of the net-to-gross approach. This is, of course, especially true if (as in the case of EUROMOD) the original purpose of the generation of gross incomes is the use of these data as input for tax-benefit models.

5.1. Descriptive analysis

A first look at the net-to-gross factors resulting from the iterative algorithm reveals the following (table 1). There are obvious differences between the net-to-gross factors computed for the two different income groups (those incomes that are (y_a) and are not (y_b) subject to PAYE withholding tax). The net-to-gross factors k_a and k_b are *never* the same for individuals with non-zero amounts of both y_a and y_b (999 of 6566 individuals fall into this category). This is, of course, due to the different tax rules under PAYE and 'final' income tax and the fact that all income components of y_a are also subject to social insurance contributions while most components of y_b are not (the exception is self-employment income).[11] In more than 12% of the cases, k_a and k_b differ by more than 10%. We take this as evidence that using the same net-to-gross factor for all income types can produce a misleading picture and that it is, indeed, worth the effort to investigate possibilities to differentiate between different income sources.

The other feature of the algorithm we wanted to test is its ability to produce different net-to-gross factors for different people. If the fiscal unit is the individual (as in the case of the PAYE tax or the social insurance contributions) this is straightforward since other people's incomes have no relevance for an individual's tax burden. Not surprisingly, we find that in more than 73% of households with more than one earner of y_a (1171 of 2539 households fall into this category), the lowest and highest k_a are different by more than 10%. In 21.5% of these households, the gap is larger than 20%. However, in section 3 we have claimed that it is possible to also differentiate between individuals even if they are in the same fiscal unit. This is relevant if the tax base is joint as is the case for the 'final' income tax, which y_b is subject to. It is, therefore, interesting to see whether the algorithm produces a meaningful spread of k_b values between individuals in households with two or more income earners. We can confirm that this is the case: While the k_bs of different individuals are within 10% of each other in only about 30% of the 501 households with more than one earner of y_b, we find that in 43% of these households they differ by more than 20%. In this context, it should be noted that the average household size in Luxembourg (approximately 2.6) is smaller than in some other European countries. Countries with larger household sizes will be more likely to exhibit even larger variations of intra-household net-to-gross factors.

To put these results into perspective, it is useful to compare them to an alternative net-to-gross method. As a representative of what we have termed the 'statistical' approach above, we have chosen the net-to-gross factors included in the ECHP user database. It should be emphasised that the explicit purpose of this factor is not to convert net incomes to gross but to convert those income components, which are only collected 'gross' (self-employment income and

[11] Interestingly, the net-to-gross factors k_b are mostly higher than k_a indicating a lower effective tax burden on employment income and replacement incomes. In most countries, one might expect the reverse because of the higher social insurance contribution burden on y_a. However, in Luxembourg social insurance contributions are relatively low for employees, pensioners and unemployed (about 10.5% for a white collar employee in 1998) while those for self-employed people are more than twice as high (since they have to pay the 'employer' part themselves). In addition, a minimum contribution basis applies for self-employed people. Another important reason is the difference between the PAYE and 'final' income tax rules. In particular, people with both y_a and y_b will normally pay a higher marginal tax rate under the 'final' income tax rules than under the PAYE scheme. On one hand, the PAYE income tax schedule is often flatter. In addition, y_a and y_b are assessed together under the 'final' income tax system whereas only y_a is subject to PAYE tax. In a progressive tax system, a higher tax base will, of course, result in higher tax rates.

income from capital) into net amounts. Clearly, however, the quality of this factor is important even if it is used in the latter way. In addition, given its name ("net/gross ratio") and in the absence of other solutions, users of the ECHP may be tempted to use this variable for a 'grossification' of income information (see also footnote 16). Below, we take the net-to-gross factors derived using the exact tax and contribution rules in the 'iterative' algorithm as a baseline reference and assess how well a statistical model is able to reproduce the results.

Comparing the ECHP net/gross ratios to our computations is not entirely straightforward. First, since we use 1998 policy rules in our 'iterative' approach, we would also need to produce the ECHP net/gross ratios from 1998 incomes. However, we do not currently have access to the 1999 wave of the ECHP (with 1998 incomes). More importantly, the samples of the 1999 waves of the ECHP and PSELL 2 (which was used for the case study) are, in any case, not congruent. To get around this, we chose to generate an ECHP-style net-to-gross ratio based on the 1999 wave of PSELL using the documentation supplied by Eurostat (Eurostat 2000, pp. 85-86). Essentially, their method uses a regression model estimated on households where survey respondents supply both net and gross information. The model is household based, i.e., all incomes are summed over household members and the net/gross ratio is modelled as being the same for all household members and all income components. Households are split into five different groups (according to size and family structure). The only explanatory variable in the regression model is total household income (relative to the average of this household group). For each household group, a separate model is estimated which is then used to predict the net/gross ratios for those households in the respective group where net and gross incomes are not both known. The regression model is a simple (least square) quadratic model as follows.

$$Y = \log (GWAGE / NWAGE) \qquad (12.1)$$
$$X = \log (HHINC / AVG\_HHINC) \qquad (12.2)$$
$$Y = c_0 + c_1X + c_2X^2 \qquad (12.3)$$

Where *NWAGE* and *GWAGE* are the sum of all household members' net and gross wages, *HHINC* is total (net) household income and *AVG_HHINC* is the average total (net) household income in the respective group.

For the present case study, we have estimated a net/gross regression model in a similar way using as independent variables the net incomes from the original PSELL data and the gross incomes as computed by our algorithm. The explanatory power of the resulting statistical model depends, of course, on the data on which it is estimated and the extent to which the model structure resembles the actual dependencies between the dependent and independent variables. As discussed in section 2, the number of cases supplying both net and gross incomes will be a crucial determinant of the quality of the model. However, since we are using the gross information as computed by the 'iterative' method, we can eliminate this data restriction by estimating the statistical model on *all* individuals in the data and, thus, focus on the structure of the regression model. On the other hand, we do, of course introduce another source of error since the gross incomes imputed by the 'iterative' method may themselves be inaccurate.

We are confident, however, that due to extensive validation of the fiscal rules coded into the EUROMOD tax-benefit model, results from the 'iterative' method do provide a good benchmark. The difficulty is that in most countries, there are no ideal benchmarks for net-to-gross conversions since net and gross information is rarely available in the same data source. Nevertheless, we realise that estimating the regression on imputed gross incomes is of limited value. We would therefore stress that the main purpose of the results below is to highlight the different extents to which the two different methods are able to capture the *variability* of net-to-gross factors in the population.

Figure 1 illustrates the fit of the regression models. All regression lines have the expected slopes. However, supporting our points about the difficulties of finding an appropriate functional form for statistical modelling of complex fiscal rules in section 2, the simple regression model is not able to capture the complexities of the tax and contribution systems. This is documented by the large spread of data points around the regression line. Consequently, the R^2 values in table 2 are, for the purpose of imputing missing variables, quite low. The poor performance of the regression model in capturing the variability of net-to-gross factors is also illustrated by the much lower standard deviations vis-à-vis the iterative approach (tables 2a and 2b).

Turning to the comparison of net-to-gross factors between the 'iterative' algorithm and the statistical model, we find the following. For 23% of individuals with incomes subject to PAYE withholding tax (3871 out of 6566 individuals have non-zero y_a), the statistical model produces net/gross ratios which differ from the reference net-to-gross factors by more than 10%. For 3% the difference exceeds 20%. Since only employment incomes are used to derive the net/gross ratios which the model is estimated on (*NWAGE* / *GWAGE* in equation 12.1 above) the results are considerably worse for incomes y_b. Only 65% of individuals with non-zero y_b (1141 individuals are in this category) are predicted within 10% of the reference net-to-gross factors while more than 9% are at least 20% off. We should stress that, when used for practical purposes, results are likely to be worse than those reported here since, in addition to a problematic specification, the limited number of cases where both gross and net incomes are observed would further impact on the quality of the estimation.

5.2. Validation of output from a tax-benefit model based on generated gross data

The Luxembourg country module of EUROMOD uses as its input gross income data, which have been derived according to the 'iterative' approach described in the case study above. One test of the quality of the results of the net-to-gross conversion is, therefore, to what extent the Luxembourg results of EUROMOD match external reference figures. It is important to note that any deviations may be due to a number of reasons. Among them are problems related to the original data which have been used as the basis for the net-to-gross conversion (sample selection, under reporting, etc.) as well as issues related to the way social and fiscal policy rules have been translated into the computer code underlying the tax-benefit model (e.g., simplifications of actual rules in the simulations, omissions due to data limitations, etc.).[12] Ideally, one would therefore want to separate these issues arising with any tax-benefit model

[12] An appendix in Immervoll and O'Donoghue (2001) provides a detailed list of issues encountered in validating results of tax-benefit microsimulation models.

from those, which are due to the net-to-gross conversion. Clearly, this is not possible with any great accuracy. However, obvious limitations on the tax-benefit model side need to be taken into account to avoid erroneously attributing matches/mismatches of model results to the net-to-gross procedure. In spite of these caveats we nevertheless believe that comparing model output with external reference figures is a very useful way to form an opinion about the quality of the net-to-gross conversion. If the errors are considered 'acceptable' and in a similar range as those found for tax-benefit models which use *actual* gross income data as their input then the net-to-gross approach is likely to be sufficiently accurate – at least for the purpose of tax-benefit modelling.

As a simple benchmark, we look at aggregate amounts simulated by the model and compare these to external reference numbers. Since we want to test the quality of the income information resulting from our net-to-gross conversion, it is sufficient to look at those social- and fiscal policy instruments, which depend on gross income. These are Income Tax, Social Insurance Contributions and, various means-tested benefits, most importantly Social Assistance ('RMG') and Housing Benefit. However, there are two reasons why the amounts simulated for these means tested instruments are not a good benchmark to be used for the net-to-gross conversion.[13] First, the simulation of these instruments raises numerous major issues that are not related to the net-to-gross conversion.[14] Any mismatches between simulated and reference amounts would therefore be hard to interpret. Secondly, the incomes of families who benefit from these transfers are, in any case, not usually subject to income taxes or social insurance contributions.[15] As a result, net-to-gross conversions are of little relevance for these low-income groups.

We see from table 3 that EUROMOD is able to match the reference figures of the major instruments quite well. The overestimation of income taxes can be explained by the fact that EUROMOD does not take into account most itemised deductions (charity, special expenses, etc.) on which no information is available in the microdata. This explanation is consistent with the very reasonable match of the main social insurance contributions, which are mainly levied on the same base as income tax and for which such deductions are not relevant. While more work remains to be done to track down in more detail the various reasons for discrepancies between simulated and actual numbers, these deviations are clearly within a range that can be regarded good quality and sufficient precision for the purpose of tax-benefit modelling.

6. Conclusion

We have presented an algorithm that exploits the tax- and contribution rules built into tax-benefit models to convert net income information into gross amounts. Using EUROMOD, a multi-country tax-benefit model covering all fifteen countries of the European Union, net-to-

[13] We have also omitted the results for instruments which are mainly relevant for self-employed people. The reason is that, as is common in survey-based micro-data, selfemployment incomes are seriously underrepresented in the micro-data underlying the simulations.
[14] Non-take-up of benefits which people would be entitled to, levels of discretion on the part of authorities, etc. See Sutherland (2001).
[15] The exception is health insurance contributions which are payable by all social assistance recipients in Luxembourg.

gross conversions can be performed for a large number of countries utilising existing models of relevant fiscal rules. The algorithm presented here takes into account all relevant complexities of tax- and contribution rules and can, thus, produce much more accurate results than statistical models which estimate net-to-gross ratios using only a few explanatory variables. Among the features of the algorithm is the ability to distinguish between different individuals in the same household. Even if individuals' incomes are taxed jointly, the algorithm is able to approximate separate net-to-gross factors for individuals in the same fiscal unit. This is possible since EUROMOD can accurately assign people to appropriate fiscal units. In addition, it is in certain cases possible to produce different net-to-gross ratios for different income components which may be subject to different tax/contribution rates.

In a case study of Luxembourg, we show that these features are important since (a) individuals in the same household/fiscal unit and (b) income components of a given individual can be subject to very different tax and contribution burdens. We also contrast our results with net/gross ratios derived from an alternative approach which is used for producing the 'net/gross ratio' included in the User Database of the European Community Household Panel (ECHP). Instead of computing an exact net-to-gross factor for each individual, that method is an example of a 'statistical' approach where fiscal rules are not modelled explicitly. Based on our comparison, we would caution against using such simplified methods for research purposes where accurate gross income information is essential.[16]

[16] It should be noted that Eurostat does in no way endorse such uses of the 'net/gross ratio' which is provide in the ECHP. It is, of course, not the purpose of the ECHP to provide accurate gross income information. Instead, the ratio is used for converting some income components which are only elicited gross into net amounts. As we understand it, the 'net/gross ratio' is included in the data to enable users of the ECHP to reconstruct the original gross incomes so that they can use their own – possibly more sophisticated – method for generating net incomes.

References

Berger, F. and M. Borsenberger, 2001, "EUROMOD Country Report Luxembourg", Department of Applied Economics, University of Cambridge.

Berliri, C., R. De Carli and V. Atella, 1999, "Moving from net to gross income in EUROMOD. An analysis of two alternative procedures", Paper presented at a EUROMOD project meeting, Copenhagen, February.

Bourguignon, F. and J. Sastre, 1999, "Euromod data bases: from Net to Gross incomes", Paper presented at a EUROMOD project meeting, Copenhagen, February.

Eurostat, 1996, "ECHP Documentation", Doc PAN 15.2/96, Luxembourg.

Eurostat, 1999, "ECHP UDB manual. European Community Household Panel Longitudinal Users' Database. Waves 1, 2 and 3", Luxembourg, November.

Eurostat, 2000, "Imputation of Income in the ECHP", Doc PAN 15.2/96, Luxembourg, November.

Farinha-Rodrigues, C., 2000, Personal Communication.

Immervoll, H. and C. O'Donoghue, 2001, "Towards a Multi-Purpose Framework for Tax-Benefit Microsimulation: A Discussion by Reference to EUROMOD, a European Tax-Benefit Model', EUROMOD Working Paper No. 2/01. Department of Applied Economics, University of Cambridge.

Immervoll, H., C. O'Donoghue and H. Sutherland, 1999, "An Introduction to EUROMOD", EUROMOD Working Paper No. 0/99, Department of Applied Economics, University of Cambridge. Available through http://www.econ.cam.ac.uk/dae/mu/emod.htm.

Levy, H. and M. Mercader-Prats, 1999, "A Note on the Net to Gross Procedure: The case of Spain", Paper presented at a EUROMOD project meeting, Rome, September.

McCrae, J., 1999. "The Development and Uses of Tax and Benefit Simulation Models", *Brazilian Electronic Journal of Economics*, Vol 2 No 1.

Mot, E. S. , 1992, *Survey of Microsimulation Models*, Social Security Research Committee, The Hague: VUGA.

Sutherland, H. (ed.), 2001, "Final Report: EUROMOD: An Integrated European Benefit-Tax Model", Department of Applied Economics, University of Cambridge.

Weinberg, D.H., 1999, "Income Data Collection in International Household Surveys", Paper presented at the 3rd Meeting of the Canberra Group in Ottawa, Canada, June 7-9.

Table 1a. k(a) versus k(b) for individuals with y(a)>0 and y(b)>0

	Frequency	%
Difference > 10%	121	12.1
Difference < 10%	878	87.9
Total	999	100.0

Table 1b. Spread of k(a) in multi-earner households

	Frequency	%
Difference between k(a)'s < 10%	311	26.6
Difference between k(a)'s > 20%	252	21.5
(Total)	(1171)	

Table 1c. Spread of k(b) in multi-earner households

	Frequency	%
Difference between k(b)'s < 10%	148	29.5
Difference between k(b)'s > 20%	217	43.3
(Total)	(501)	

Table 2a. 'Iterative' algorithm vs. 'statistical' model: Descriptive statistics for k(a) (individuals where y(a)>0).

	N	Min	Max	Mean	Std.
'iterative' method	3871	1.000	1.778	1.183	.122
'statistical' model	3871	1.000	1.894	1.176	.080

Table 2b. 'Iterative' algorithm vs. 'statistical' model: Descriptive statistics for k(b) (individuals where y(b)>0).

	N	Min	Max	Mean	Std. Dev.
'iterative' method	1141	1.000	2.295	1.27	.202
'statistical' model	1141	1.004	1.894	1.19	.095

Table 2c. 'Iterative' algorithm vs. 'statistical' model: Deviation of k(a) (individuals where y(a)>0).

	Frequency	%
Difference < 10%	2980	77.0
Difference > 20%	124	3.2
(Total)	(3871)	

Table 2d. 'Iterative' algorithm vs. 'statistical' model: Deviation of k(b) (individuals where y(b)>0).

	Frequency	%
Difference < 10%	745	65.3
Difference > 20%	104	9.1
(Total)	(1141)	

Note: descriptives in all tables are for *unweighted* data.

Table 3. Comparisons of Aggregates.

Instrument	EUROMOD aggregate amount[1] 1998, million LUF	Actual aggregate amount[1] 1998, million LUF	Deviation of simulated result
Income Tax[2]	45,530.7	41,271.4	+10.3%
Social Insurance Contributions – Health[3]	16,932.4	17,913.3	-5.5%
Social Insurance Contributions - Pensions[3]	33,737.0	33,999.3	-0.8%

Source: EUROMOD and Berger and Borsenberger (2001).
Note 1: Both simulated and reference figures exclude so called 'border workers' who live abroad but work in Luxembourg.
Note 2: Includes 'Solidarity Tax' which is used to finance the unemployment fund.
Note 3: Employee Contrib. + Employer Contrib. + Self Employed Contrib. + Benefit Recipients' Contrib.

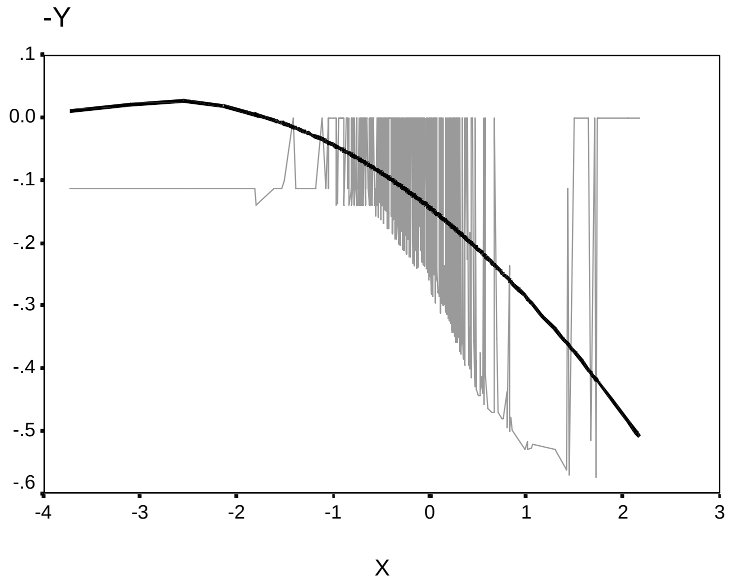
Figure 1. Fit of 'statistical' net-to-gross model.

$Y = \log\,(GWAGE\ /\ NWAGE)$
$X = \log\,(HHINC\ /\ AVG\_HHINC)$
$Y = c_0 + c_1X + c_2X^2$

'Single'

Independent: X

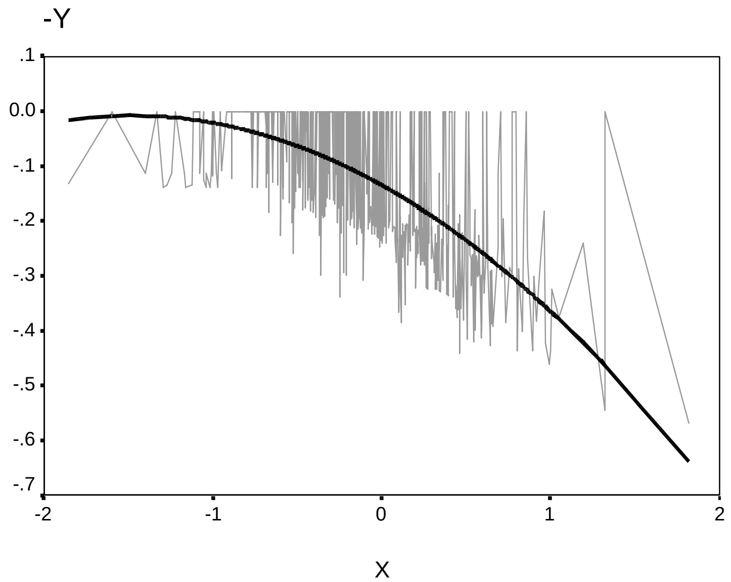
Dependent	Rsq	c0	c1	c2
-Y	.201	-.1444	-.1218	-.0215



'Couple; No Children'

Independent: X

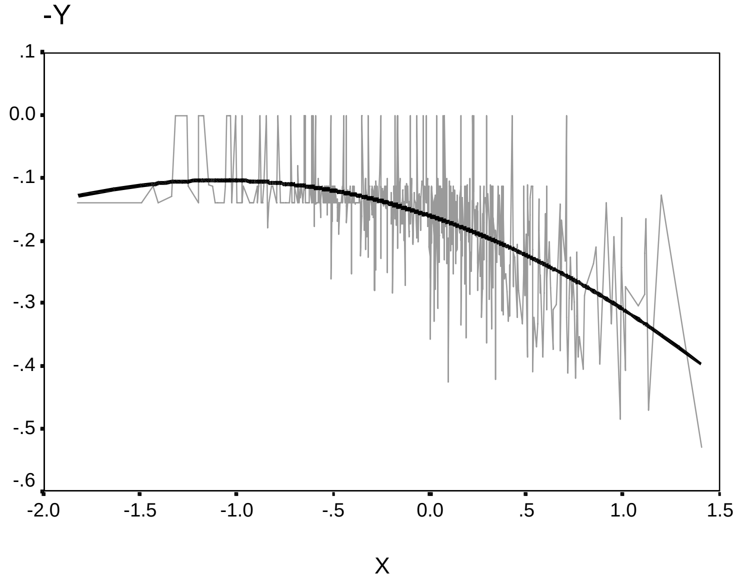
Dependent	Rsq	c0	c1	c2
-Y	.410	-.1351	-.1719	-.0580



’Single or Couple with Children Aged 16-’

Independent: X

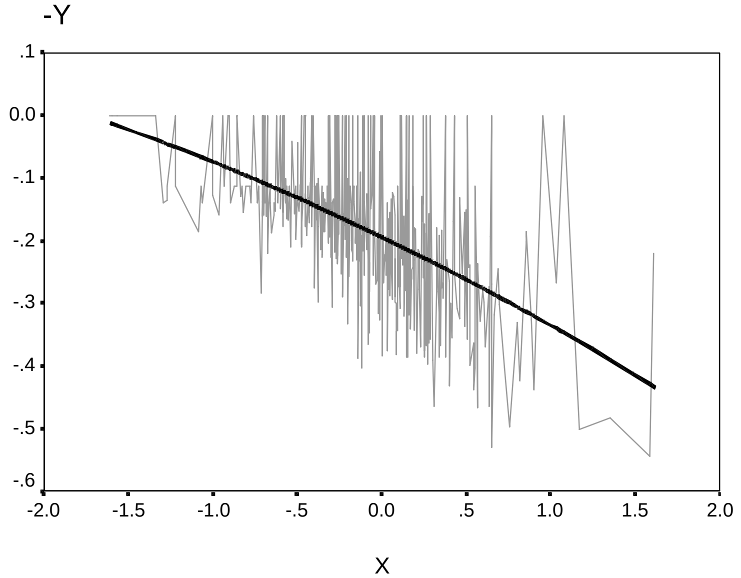
Dependent	Rsq	c0	c1	c2
-Y	.362	-.1603	-.1027	-.0471



’Other Households with Max. 4 Persons’

Independent: X

Dependent	Rsq	c0	c1	c2
-Y	.283	-.1943	-.1305	-.0107



’Other Households with Min. 5 Persons’

```
Independent:  X

  Dependent    Rsq     c0       c1       c2

   -Y         .430   -.1722   -.1471   -.1311
```