Project number: 5562
Funding source: European Commission
Date: February 2012
Project dates: Feb 2006 - Jan 2009

Accurate income measurement for the assessment of public policies

Key external stakeholders:

European Commission, Euromod, Institute for Social and Economic Research (Essex), European government agencies

Practical implications for stakeholders:

This project supports the capacity of researchers to overcome data limitations and generate richer datasets for the analysis of key economic and policy changes.

This project enhances the ability of researchers to evaluate integrated tax and welfare policy at the European level and assess methodological issues relating to the most appropriate methods of analysis across a range of countries, policy and data contexts.

The methods can be used to match spatially disaggregated data such as the Small Area Population Statistics to household and farm level datasets. As a consequence, there will be more scope for Teagasc researchers to spatially illustrate social and economic data within Ireland.

Main results:

We evaluated different methods to link different datasets, so as to increase the available information for analysis. A relatively simple method based upon parametric regression was found to be the best of five methods.

Opportunity / Benefit:

REDP researchers will be able to pose a wider set of research questions as fewer constraints will emerge due to a lack of data availability. The codes used for the policy simulations can be used elsewhere e.g. identify the welfare effects of changes to environmental taxes and shocks to farm incomes.

Collaborating Institutions:

See page two of the full Technology Update

Contact Cathal O Donoghue Email: Cathal.odonoghue@teagasc.ie.

Teagasc project team: Cathal O'Donoghue
Philip Hayes
Darragh Flannery
Jason Loughrey
Stephen Hynes

External collaborators: Prof. Andre DeCoster
Dr Kris DeSwerdt
Dr. Bram DeRock
Dirk Verwerft (Catholic University of Leuven)
Institute for Social and Economic Research
Centre for Economic Research and Environmental Strategy
Economic & Social Research Institute
DIW Berlin
European centre for Social Welfare Policy and Research
CentERdata, University of Tilburg
Autonomous University of Barcelona
Luxembourg Income Study
University of Antwerp

1. Project background:

Research questions in quantitative economic research are typically designed with data availability in mind. Since the mid 1990s, there has been a significant improvement in the availability of household and individual level data on incomes, employment and other socio-economic variables. This has allowed economists and other social scientists to conduct research on topics such as the earnings distribution and the impact of inflation and taxation on household welfare. The Quarterly National Household Survey (QNHS), the Survey of Income and Living Conditions (SILC) and the familiar National Farm Survey (NFS) are examples of such micro level datasets.

All of these datasets are constructed with particular research questions in mind. For example, the calculation of the national unemployment rate is largely based upon the results of the QNHS while the NFS is concerned mainly with agricultural incomes, production costs and closely related issues. For the purposes of examining issues connected with Rural Development, there is a need to analyze changes to the social and economic wellbeing of households not just within agriculture but within wider rural communities. A sole dependence upon either the NFS or the SILC datasets does not provide us with the necessary infrastructure to carry out such research. This is not an unusual constraint faced by researchers. In many cases, there is no single data source available that includes all of the variables that the researcher requires in order to address a particular research question.

The absence of one single data source does not however, preclude researchers from achieving ambitious research objectives. This is due to the availability of statistical matching methods that can be used to impute information from one dataset to another. There exists a good deal of variation in the complexity of these methods. Some of these methods involve the simple ranking of records in each dataset according to the responses to questions that are common to both datasets. The more complex methods involve the use of regression-based econometric techniques. The aim of this AIMAP project was to provide an assessment of five different matching methods. The testing of these five methods involved the imputation of household expenditures from the Household Budget Survey into the Living in Ireland Survey of 2000, a household income survey. These methods can however, be used to match farm household data such as the NFS to a household income survey such as the SILC and therefore contribute to a better data infrastructure within Teagasc.

2. Questions addressed by the project:

- Identify the best generic method of matching a household expenditure survey to a household income survey for a selected set of EU-countries. The Household Budget Survey and the Living in Ireland Survey were the expenditure and income surveys in the case of Ireland.
- Estimate the incidence and distributional outcomes of a policy relevant simulation involving a shift from personal income taxes to indirect taxes or environmental taxes using the matched datasets. Identify who pays and gets how much in different forms of taxes and benefits under the new and old tax-benefit policies.

Contact Cathal O Donoghue Email: Cathal.odonoghue@teagasc.ie.

3. The experimental studies:

Five alternative methods were used to impute expenditures from the household expenditure dataset to the household income dataset. In all of these methods, the expenditure survey is considered to be the source dataset as it contains the original expenditure information. The household income survey is considered to be the target dataset as the expenditure values must be imputed into the household income dataset.

Two of the five methods involved the use of regression based techniques. The other three techniques were based upon the matching of each household in the income survey to a household in the budget survey using ranking methods. These rankings were based upon the responses from households to questions common to both the expenditure survey and the income survey. These included variables such as education level of the household head and the number of adults and children in the household.

The relative matching quality was evaluated by means of two criteria: a goodness of fit measure, and tests of the equality of the distributions of the imputed and the observed budget shares. The goodness of fit tests summarized the discrepancy between the expected and observed values of the imputed variables in terms of the average household in the income distribution. The tests on the equality of distributions were used to analyze the ability of the matching process to make accurate imputations for households at different points of the income distribution. The best matching method was chosen on the basis of these tests.

This was followed by a two-step simulation process involving a combined set of policy instruments: direct taxes and indirect taxes. The first step in the simulation involved a 25% decrease in the social security contributions of the employees in the matched dataset and the second step involved an increase in the standard VAT rate in order that revenue neutrality can be achieved.

4. Main results:

Overall, the parametric and non-parametric methods generated the best fit of the imputed values with respect to the observed values. The parametric match outperformed the non-parametric match in terms of imputing the distribution of expenditures. The parametric method is much less computational intensive to implement relative to the non-parametric method. The authors therefore selected the parametric regression method as the best method.

The results of the matching process show that indirect taxation is an influential, regressive component in the total tax system. It is therefore important to include it in microsimulation studies. The regressivity is due to the regressivity of the savings rate. Households earning more save more, and this compensates for the fact that they buy goods with higher indirect tax rates.

Decreasing the social security contributions of the employees and financing this by an increase in the standard VAT rate is a very regressive measure in all countries studied. This is the consequence of replacing a progressive taxation scheme by a regressive one.

5. Opportunity/Benefit:

REDP researchers will be able to pose a wider set of research questions as fewer constraints will emerge due to a lack of data availability. The codes used for the policy simulations can be used elsewhere e.g. identify the welfare effects of changes to environmental taxes and shocks to farm incomes.

6. Dissemination:

Main publications:

DeCoster, A., J. Loughrey, C. O'Donoghue and D. Verwerft, 2011, Microsimulation of Indirect Taxes, *International Journal of Microsimulation*

Decoster, A., Loughrey, J., O'Donoghue, C. and Verwerft, D. "How Regressive Are Indirect Taxes? A Microsimulation Analysis for Five European Countries" Journal of Policy Analysis and Management, March 2010, 29(2), 326-350

7. Compiled by: Jason Loughrey

Contact Cathal O Donoghue Email: Cathal.odonoghue@teagasc.ie.