

The Spatial Distribution of Labour Force Participation and Market Earnings at the Sub-National Level in Ireland

KARYN MORRISSEY*

National University of Galway

CATHAL O'DONOGHUE

Rural Economic Research Centre, Teagasc, Ireland

The main aim of this paper is to provide a spatial modelling framework for labour force participation and income estimation. The development of a household income distribution for Ireland had previously been hampered by the lack of disaggregated data on individual earnings. Spatial microsimulation through a process of calibration provides a method which allows one to recreate the spatial distribution LFP and household market income at the small area level. Further analysis examines the relationship between LFP, occupational type and market income at the small area level in Co. Galway Ireland.

Keywords: Household Market Income Distribution, Employment, Spatial Microsimulation, Calibration, Mapping

JEL Classifications: J21, J31

1 Introduction

The distribution of individual wealth varies greatly across countries. This is well documented through the use international cross-sectional data and indicators such as Gini coefficients. However, individual wealth also varies greatly within countries. Indeed at the sub-national level, individual wealth can vary across provinces, regions, cities and urban and rural classifications. Ireland, similar to the UK, Portugal and the US, has a relatively high level of income inequality (Wilkinson and Pickett, 2009), which has remained stable over time (Nolan, 2009). When the most widely-used summary measures are calculated from household survey data, Ireland ranks: 10-12th within the EU-15, 17-18th within the EU-27 and 18-22nd within the OECD (Nolan, 2009). However, why should income inequality be a concern for public policy in the first instance? Why is income inequality anymore important than gender inequality, opportunity inequality.

Substantial research has shown that income inequality may be a key factor in producing or exacerbating a wide range of social ills such as educational disadvantage, health inequalities and crime, and undermining social cohesion. As Nolan points out, there is a substantial research literature documenting the extent to which childhood disadvantage underpins poor adult outcomes across various domains, such as educational attainment and adult earnings (Cunha and Heckman, 2007). In terms of health status and income inequality, Wilkinson and Pickett, (2009) demonstrate that countries with lower levels of income inequality are healthier and happier. Given the well documented relationships between health status, educational attainment and general status quo of a population and aggregate economic performance, factors which adversely affect some or all of these factors, such as economic inequality, may have a knock on effect on a country's economic performance.

Bourguignon et al., (2002) note that given such widespread inequality exists and the well-documented effects that inequality has on health, educational attainment, etc, that there ought to be considerable interest in understanding why income distributions vary so much within countries. Are these spatial differences due to for example, the domination of one region economically over another? Or is it due to differences in labour market institutions or household characteristics? And if, as is likely, differences in income distributions reflect all of these factors, in what manner and to what extent does each one contribute?

Against this background, substantial progress has been made in our ability to understand differences in individual level wage distributions (please see Bourguignon et al., (2002) for a concise overview of previous work on the estimation of wage distributions). However, it is only in the very recent past that work has been done in estimating the distribution of household incomes. This is primarily due to the additional complexities involved in modelling household income compared to individual wages. For example, a simple Mincerian earnings equation ($Y = \beta X_i + \epsilon_i$) is sufficient to model the wage distribution for a given sample of employees. These estimates may then be used to compare wage inequality across different sub-samples of employees. In contrast, the distribution of household income also depends on the returns and characteristics of its employed household members and draws on earning models, household income is also determined through a system of inter-related levels.

Bourguignon et al., (2004) separate these levels into three categories: population/endowment effects (such as the age, area of residences, ownership of physical and financial capital, etc), price effects (which may be defined as the returns to factors of production, such as human capital) and occupation effects (the occupation structure of the population). These levels are not independent of each other and a change in one level (for example education level), will interact with another level (for example, the occupational choice decision) to generate a change in household income. A number of recent studies have used such a system-wide approach to estimate the household income distribution

(Bourguignon et al., 2002 and Bourguignon et al., 2004). Using a series of reduced form (non-linear) models which include; employee earnings models, self-employed earnings models, capital earnings models, occupational choice models education levels models and household size models, these studies estimated the distribution of total household income for various East Asian and Latin American countries. These studies than used decomposition techniques to study the nature of the dynamics of the income distribution of each the countries involved in the study.

With regard to spatial income generation models, Kalogirou and Hatzichristos, (2006) point to an extensive literature on spatial econometrics including, modelling income heterogeneity (Jenkins, 2000), and income growth (Azzoni, 2001) across space. Kalogirou and Hatzichristos, (2006) further point to the recent interest in studying income and growth levels in European regions (Le Gallo & Ertur, 2003; Dall'erba, 2005). However, the main aim of such studies is to identify spatial inequalities in terms of income, rather than the underlying determinants of these spatial differentials.

To accurately model household market income, one needs to estimate a system of equations that represents each household's market income generation mechanism. Furthermore to accurately model household market income across space, one must be able to capture spatial differentials in household earning distributions and their underlying drivers. Within an Irish context, the estimation of household market income differentials across space has not been achieved to date. This lack of analysis was previously attributable to the non-availability of data on labour, household, education and health data at lower spatial scales. This paper recreates the entire household market income distribution for County Galway, Ireland at the small area level using spatial microsimulation techniques. In particular, this paper focuses on the relationship between labour force participation (LFP), occupation structure and household market income in County Galway. The principal reason that Galway was chosen as the area of interest for this analysis, is that while Galway is a predominately rural county, it is also home to one of Ireland's top five urban concentrations. This allows us to achieve further spatial diversity with regard to market income differentials.

This paper continues as follows: Section 2 introduces the spatial microsimulation methodology and SMILE (Simulation Model of the Irish Local Economy) a spatial microsimulation model developed by the Rural Economic Research Centre, Teagasc and the School of Geography, University of Leeds. Section 3 introduces the process of model calibration. The rationale for model calibration is first outlined. The alignment process used to calibrate SMILEs labour force participation (LFP) variables and subsequently SMILEs market income variables are outlined. Section 4 presents the results of the alignment process for both LFP and household market income for Co. Galway. The relationship between LFP, occupation type and average household market income at the ED level is examined. Section 5 offers concluding comments.

2. Spatial Microsimulation

Previous Data Limitations

Detailed spatial profiles of household income may be extremely valuable to both government and non-government organisations that want to strengthen the impact that their spending and or policy interventions in areas such as health, social welfare, pension and poverty. The development of a household income distribution for Ireland had previously been hampered by the lack of disaggregated data on individual earnings. Census data, although available at the small area level does not offer any information on household income. Furthermore, although census does include demographic and occupational data at the small area level, this data is not cross-tabulated. As such, it is difficult to produce meaning income distributions using the Irish census. On the other hand, while survey data (such as the Living in Ireland Survey) often contains detailed income data at the individual level, this data is usually aspatial in nature. Spatial microsimulation techniques provide a method of merging spatial and aspatial data from a number of different data sources to provide an attribute rich, geo-referenced dataset. The dataset that is created therefore allows one to investigate individual/household occupational structure and earnings at a very local level of spatial resolution.

Spatial microsimulation is a method used to create spatially disaggregated microdata that previously did not exist. Thus, an important issue associated with spatial microsimulation modelling is the validation of model outputs. Validation techniques examine model outputs in systematic ways to reveal deficiencies/errors in the model outputs. As such, model validation forms an integral part of the overall development and application of any model. Oketch & Carrick (2005) point out that it is only through model validation that the creditability and reliability of simulated data can be assured. There are a number of methods one can use to validate outputs from microsimulation models. These methods include in-sample validation, out-of-sample validation and multiple-module validation. Caldwell (1996) provides a concise overview of these methods.

Spatial microsimulation methods have been developed and utilised to assist researchers to study issues associated with the spatial distribution of income. A frequent and early use of spatial microsimulation has been data enhancement. Spatial Microsimulation was originally developed as a technique to allow policy analyses at a spatial level when spatial data is not available. A widespread example is the poverty mapping methodology largely developed at the World Bank (WB). Although not labelled spatial Microsimulation, the WB methodology uses identical methods for similar purposes as analyses within the spatial microsimulation literature (See, Hentschel et al., 1998; Elbers et al., 2003). The method involves parametric statistical matching of micro household data such as a budget survey to spatial census data to develop poverty maps. The WB methodology is widely used for spatial policy analysis throughout the developing world (Elbers et al., 2003).

The WB methodology, based on parametric statistical matching techniques, has been developed and utilised in parallel in developed countries to consider spatial policy similar issues. For example NATSEM in the University of Canberra, has developed a suite of spatial microsimulation models. Their modelling framework was initially built around a regional income model (Lloyd et al, 2000) and a model Marketinfo focusing on local expenditure and incomes focusing on market clients (King et al., 2002). There have been a number of models built in the UK for spatial poverty and inequality analysis. SimLeeds (Ballas and Clarke, 2001) was developed to examine the labour market in and around the Leeds metropolitan area. SimLeeds adopts a similar approach to that of Williamson (1999). Ballas used this framework to look at changes in poverty and inequality in Leeds and Sheffield between the 1991 and 2001 censuses. Following on from SimLeeds, the Leeds team extended their modelling framework to York and Wales (Ballas et al., 2005) and later to cover the whole country, SimBritain to look at the income and spatial distributional issues. Anderson (2007) has also developed a spatial microsimulation model for studying deprivation in England, while Tiglao (2002) extended the methodology to develop a model for the Phillipines.

Although, developing spatial indicators of income inequality is in itself a useful tool to facilitate spatial planning, the addition of tax-benefit microsimulation modelling techniques allows for the spatial impact of policy in reducing poverty and inequality to be assessed. NATSEM's model SYNAGI (Synthetic Australian Geo-demographic Information) has been extended to simulate taxes and benefits (Chin et al., 2005). SYNAGI was developed to study issues related to differential spatial poverty and inequality rates (Harding et al., 2006). It has also been used to provide support for a social policy agency (King et al., 2002), developing indicators of child social exclusion (Harding et al, 2009b, Tanton et al, 2009a) and urban poverty (Tanton et al, 2009b).

While much of the literature in this area has focused on incidence analysis such as classifications of areas with child social exclusion (Harding et al, 2009b, Tanton et al, 2009a) or spatial income inequality (Banks et al., forthcoming), only a few papers have taken advantage of the capacity of microsimulation models to simulate the impact of policy reform. Harding et al, (2009a) simulated the impact of a national family tax benefit reform. The SimLeeds team used a partial tax-benefit model to simulate a number of tax and pension changes on income (See Ballas et al., 2003). SimBritain was also used to simulate the impact of changes to minimum wage, winter fuel payments, working family's tax credits and new child and working credits (Ballas et al, 2007). Drawing on the previous literature and methodologies, the aim of this paper is to model the spatial distribution of labour market outcomes, including participation and associated incomes. The next section introduces SMILE, Simulation model of the Irish Local Economy and its associated methodology for the creation of geo-referenced, micro-level data for the Irish population.

SMILE - Simulation model of the Irish Local Economy

Spatial microsimulation is a means of synthetically creating large-scale micro-datasets at different geographical scales. The development and application of spatial microsimulation models offers considerable scope and potential to analysis the individual composition of an area so that specific policies may be directed to areas with the highest need for that policy. Although there are a number of datasets containing spatial identifiers for each individual in the dataset, these tend to be at a very aggregate level. The Living in Ireland (LII) dataset for example contains a spatial variable broken down by only 12 possible categories: the five cities in Ireland, a category for Dublin County, an 'open-countryside' category, and five categories for towns of varying sizes. The LII survey is the Irish component of the European Community Household Panel (ECHP) dataset. The LII survey was a seven year longitudinal survey that began in 1994 and ended in 2001. The LII dataset for 2000 contained 13,067 individuals and contains a variety of demographic and socio-economic, income and health information at the micro level.

In contrast, the Irish Small Area Population Statistics (SAPS) contains a rich set of census information at the small area level – in this case for electoral divisions (EDs). EDs are the smallest geographical output area for Ireland. There are 3,340 EDs in Ireland. However, as with most censuses, the data available on individual's health status is limited. If we could merge the data in the LII with the ED Census level data we would have a much richer dataset that would allow us to investigate GP utilisation at a very local level of spatial resolution. We use spatial microsimulation techniques to accomplish this.

SMILE (Simulation model of the Irish Local Economy) is a static spatial microsimulation model (Morrissey et al., 2008). Using a combinational optimisation technique, simulated annealing, to match the 2000 Living in Ireland (LII) Survey to the 2002 Small Area Population Statistics (SAPS), SMILE produces a micro-level synthetic dataset for the whole population of Ireland. Using the simulated annealing algorithm the relevant number of people (i.e. equivalent to the total number of people in each ED) are randomly taken from the LII dataset and the errors between the simulated population and SAPS population (constrained by age, sex, education level, whether a farmer or not and the number of individuals in each household) is calculated. When this error is deemed to be acceptably low enough this configuration of LII records is stored as the simulated population. The particular SA algorithm used is adopted from the one employed by Ballas and Clarke (2001) to construct the base population for the SimLeeds spatial microsimulation model (see Ballas et al. (2005) and Morrissey et al., (2008) for further discussion on SA and the SMILE algorithm).

Once the simulated annealing algorithm has generated the baseline population for each ED in Ireland, the remaining variables from the LII survey are merged with the newly created geo-referenced, baseline dataset. The dataset created by SMILE contains demographic, socio-

economic, health, occupation and income variables for both individuals and household units at the small area, electoral division (ED) level. SMILE also contains an agri-environmental component that may be used as a stand alone model or linked with the population model to address agri-environmental policy issues (Hynes et al., 2009). Given the diversity of micro data contained within the SMILE framework, SMILE has been used to examine a range of policy issues. For example, Morrissey et al., (2008) use the health component of SMILE in conjunction with a spatial interaction model to examine access to GP services at the sub-national level in Ireland. Further work, using the SMILE health component examined the rate of self-reported depression at the small area level in Ireland. This work also combined a spatial interaction model to examine access to both acute and community psychiatric facilities for individuals who reported suffering from depression (Morrissey et al., 2010). Using the agri-environmental component of the SMILE model, Hynes et al., (2009) examined the distribution of family farm income at the small are level in Ireland. Further uses of the agri-environmental component also examined the levels of methane emissions from farms at the ED level in Ireland. Combining the family farm income data within this paper, the redistributive effects of a carbon tax on farms with high methane emissions to low emissions farms was further examined (Hynes et al., 2009). Thus, although the development and further calibration of SMILE (as will be outlined below) is both computationally and labour intensive, the data created by SMILE is applicable to a broad range of policy areas. The applicability of the data created by SMILE is further widened within this paper through the development of the income generation component of SMILE and the incorporation of a Tax-Benefit System for Ireland within the framework.

Validation

Common to all data generation techniques that rely on a random process to generate data, validation of the newly created data is an important component of model development. In the case of spatial microsimulation modelling, one samples from a micro-dataset to make it representative at a spatial scale lower than what was originally collected within the survey. Validation is particularly important with regard to variables that have not been used as part of the weighting, calibration or matching process. In this sub-section, we evaluate the quality of the SMILE matching process.

Validation techniques examine model outputs in systematic ways to reveal deficiencies/errors in the model outputs. With regard to spatial microsimulation, the most commonly used approach for validation is to aggregate the simulated data to a geographic level which has known values for the constrained and unconstrained variables (Ballas and Clarke, 2001). However, limitations associated with this technique centre on the difficulty in finding spatial data at varying levels of aggregation to compare the newly created data to. Edwards and Clarke (2009) highlight a new method of validating the results from a spatial

microsimulation model using regression analysis to compare a percentage of the population in each category in the newly simulated to actual data. The regression analysis visually provides information regarding how well the data fit the ideal simulation but the coefficient of determination describes how well the data fit the best fit regression line, rather than the ideal one. However, as Edwards and Clarke point out this considers the precision of the simulated data rather than accuracy. Thus, to examine the accuracy of the simulated data compared to the actual data, an equal variance t-test may be used to determine whether any differences between the two were statistically significant. However, this approach only works for the constrained variables where actual data exists; unconstrained variables cannot be validated using this method.

This paper uses the out-of-sample validation technique defined by Caldwell (1996). Table 1 provides the results of the validation of a number of the key labour force participation variables simulated by SMILE. Out-of-sample validation involves comparing the synthetically created microdata with new, external data. In Table 1, SMILE's simulated data is compared to its 2002 SAPS counterpart. Column four represents the ratio of the count of SMILE's simulated in-work variable for a number of EDs in Co. Galway. As one can see, the margin of error is between + 0.05 and +0.35. That is, SMILE over-estimates the number of individual's in-work within each ED. The average error for this selection of EDs is 1.22. Column seven represents the count of SMILE's simulated employee variable. As one can see, the margin of error is between -0.26 and +1.37. The average error for this selection of EDs is

Table 1. Validation of SMILEs Labour Force Participation at the ED level

ED	SAPS In Work (Count)	SMILE In Work (Count)	Ratio In Work SMILE/ SAPS	SAPS Employee (Count)	SMILE Employee (Count)	Ratio Employee SMILE/ SAPS
3201002	1089	1433	1.32	840	1030	1.23
3201003	182	236	1.30	127	139	1.09
3201004	170	232	1.36	112	153	1.37
3201005	126	149	1.18	90	84	0.93
3201006	147	154	1.05	114	96	0.84
3201007	189	214	1.13	99	119	1.20
3201008	855	1155	1.35	683	929	1.36
3201009	156	198	1.27	103	128	1.24
3201010	430	581	1.35	328	430	1.31
3201011	371	383	1.03	287	300	1.05
3201012	165	171	1.04	103	95	0.92
Average Error			1.22			1.14

Data Sources: SAPS & SMILE

1.14. One can therefore see that there are differences between SMILE's non-matched results and the SAPS results at the ED level of these two variables. Thus, the next question that must be addressed is given the variability between the exogenous data and simulated variables both at the national and spatial level; does this mean that the model is inaccurate?

3. Model Calibration

As highlighted in the previous section, significant spatial heterogeneity remains unexplained by the variables used in our match process. One alternative is to increase the number of variables used in the spatial microsimulation exercise. However, this comes with significant computation cost. As a result, in this section an alternative method for validation is drawn from the dynamic microsimulation literature to correct for this unexplained spatial heterogeneity. This process is known as calibration through alignment.

Baekgaard (2002) points out that the output from a microsimulation model is only as reliable as the original datasets that were used to create the synthetic dataset. For example, datasets, such as the LII survey are prone to a number of errors which arise due to sampling error, data collection error and data processing error. Indeed, Stuggard (1996) estimates that micro-level datasets have approximately a 10% error margin due to survey bias. Therefore, given the (expected) inaccuracies of the initial SMILE match, another method of ensuring that SMILE's unmatched variables replicate the 'real' characteristics of the Irish population must be used.

Such a solution is offered through alignment, also known as model calibration. The objective of calibrating a spatial microsimulation model is to ensure that the simulated output matches exogenous totals at varying levels of spatial disaggregation (Baekgaard, 2002). Similar to the CORSIM (Caldwell et al., 1996) and DYNACAN (Morrison, 2006) models, SMILE incorporates an array of alignment processes. There are a number of different alignment processes one may use and the choice of process depends on the type of data outputted from the microsimulation model and the data type of the exogenous 'target' data.

The data outputted from the microsimulation model may take three broad types – binary data, continuous data and count data. The actual process used to align these three data types is very different. This paper concentrates on the alignment of two variables – labour force participation (LFP) (whether an individual is in-work or not), a binary variable and employee market income, a continuous variable.

Models of binary events such as the presence LFP (yes, in-work, no, not in-work) may be modelled using either a logistic regression or a probit model. This allows one to estimate the probability of the event occurring. On the other hand, when the explanatory variable, such as the market income, is in the form of continuous data, one must use a model suitable to continuous data, such as OLS regression models. The distinction between data types is only one of the properties that define an alignment technique (Baekgaard, 2002). An important

difference between different alignment processes is the method by which the simulated output is matched to the exogenous targets. These methods include:

- Aggregate total alignment enforces an exact match with the exogenous target data. For example, exogenous census data may reveal that 54 males in one ED are in employment. The alignment process ensures that this total is met for that ED.
- A Percentage/Rate alignment: The simulated data is matched to exogenous rates. For example, 80% of the 19-25 year old population in one ED may be an employee. The simulated data is aligned to match this rate.
- An Average Value Alignment: For a monetary value, such as market income, one wants to ensure that the simulated output produces the same aggregate contributions at the county or national level.

As outlined in the introduction, to accurately estimate household income distribution at the small area level, a wide range of household, labour force participation and occupational data is required. The next section describes the alignment process used to calibrate the labour force participation variables and market income variables produced by SMILE.

Calibrating a Binary Choice Model: The Labour Force Participation Model

A logistic regression model may be used to determine the factors that influence LFP. The logistic regression model used to model LFP (being in-work or not) may be expressed as follows. The choice of individual i between LFP or not may be determined by a vector of individual characteristics. The decision to participate in the labour force can be expressed via a logit model as:

$$y_i^* = \text{logit}(p_i) = \ln \frac{P_i}{(1-P_i)} = B_o + \sum_k \beta X_i^k + \varepsilon_i \quad (1)$$

such that

$$y = 1 \text{ if } y_i^* > 0 \quad (2)$$

In order to create the stochastic term, ε_i we use the following relationship:

$$y = 1 \text{ if } u_i < \log \text{it}^{-1} \left(B_o + \sum_k \beta X_i^k \right) = p_i \quad (3)$$

A value of μ_i that satisfies this is:

$$u_i = (Y = 1) * (r * p_i) + (Y = 0) * (p_i + (r * (1 - p_i))) \quad (4)$$

where r is a uniform random number. Lastly ε_i can be defined as:

$$\varepsilon_i = \ln\left(\frac{u_i}{1-u_i}\right) \quad (5)$$

However, due to inherent errors in the model (Demombynes et al., 2002; Elders et al., 2001), the predictive power of these models is poor. The further the probability of an event occurring is from 0.5, the less probable the event is to occur. Therefore, used alone, a logistic regression model may under or over predict the number of events (Duncan and Weeks, 1998). Given these issues, it was decided to use an ‘alignment procedure’ to ensure that SMILE’s LFP variables match the ‘true’ spatial distribution of LFP.

Our calibration routine operates where N cases are required by ranking y_i such that;

$$y = 1 \text{ for highest } N y_i^* \quad (6)$$

This method is undertaken for each of the simulated processes, so that the aggregate number of cases of each variable y is consistent with the control total from the Census small area statistic.

In order to select the ‘true’ number of individual within the Irish labour force, one may use the exogenously specified totals from the 2002 SAPS dataset. Firstly, an initial alignment is performed at the national level to ensure that SMILE’s LFP variables match their counterparts in the SAPS dataset. The second alignment is spatially disaggregated at the ED level to ensure the best data fit across space. To reduce errors in status assignment, the alignment process is also constrained by seven age categories, which are further sub-divided by gender. As such, these groups may be expressed as a $A_{3440} \times A_{2*7}$ table, where 3440 is the number of EDs in Ireland and A_{7*2} are the seven by two age/sex categories.

The generated equation including stochastic term

$$y_i^* = \text{logit}(p_i) = \ln\left(\frac{P_i}{1-P_i}\right) = B_0 + \sum_k \beta X_i^k + \varepsilon_i \quad (7)$$

is ranked from lowest to highest and the exogenously specified number of individuals in-work for each age/sex and ED category are selected from the individuals where the condition $\varepsilon_i < \text{logit}^{-1}(\alpha_i + \beta x_i)$ is satisfied. Once the correct numbers of individuals by age/sex and ED have been aligned to match the SAPS totals, the other LFP variables such as, unemployment, retirement, occupation type for those in the labour force, etc may be aligned. Thus in effect it is a dynamic microsimulation model as, the X’s are simulated endogenously with the model via calibration. Once our dataset contains the representative labour profile for the whole of Ireland, we can then simulate the labour market earnings for each individual within the labour market.

Income Generation Model

Once all of the categorical variables, X^* , have been calibrated, we simulate the continuous, largely income variables as follows:

$$Y_i = \exp\left(X^* B + \varepsilon\right) \quad (8)$$

where Y_i is labour market earnings for individual i . X_i is a row-vector of individual characteristics that influence income level. α is the deterministic component which made up of a vector of covariates for each individual. ε_i (an estimate of the model-based error) and μ_i (an estimate of potential wage) provide a stochastic component for each individual i . Once each individual who previously had a wage under the original statistical match receives a ε_i component, error terms must also be estimated separately for those individuals who previously were not in-work, but were assigned in-work status on-alignment.

While we observe the error term for individuals with incomes in the original data, for those who are simulated to have incomes, we need to generate the error term via Monte Carlo simulation. To do this, we generate a random variable and take the inverse standard normal cumulative distribution ($F^{-1}(y)$) of this variable (Juhn et al., 1993). Next, the standard deviation of the previously estimated ε component is calculated. The final step involves multiplying the inverse standard normal cumulative distribution of the random variable by the standard deviation of ε . This process gives a ε term for each of the newly assigned individuals in work. Once we have created ε , the stochastic term for each individual in the dataset, we may move to the second part of our calibration process - aligning the estimated wages for each worker to the exogenously specified income totals at the county level.

External forecasts from the Irish National accounts give average annual earnings for compensation of employees and self-employed individuals for each county. The National Accounts (NA) is published annually by the CSO. The NA forms a comprehensive framework within which economic data may be presented for each county in Ireland. There are three approaches one may use to measure national income: Output (valued added by producers), income (all income generated) and expenditure (spending on final demand). In Ireland, the income and expenditure approaches are used. For the purpose of this paper, we are interested in the income estimates. With regard to the income estimates the main components include: profits from companies and the self-employed, remuneration of employees including wages, salaries and employers contributions such as social insurance and pension contributions and the rent of dwellings (inputted if owner occupied). Thus, the accounts include a variety of income information at the county level. However, although the NA accounts contain detailed estimates of income, its use in poverty analysis is limited due to its highly aggregated nature.

The object therefore is for the sum of simulated earnings for each county to equal the National Accounts total for each county, $- E_c$:

$$\sum_i Y_i = E_c \quad (9)$$

This is done by creating an alignment co-efficient γ . The alignment co-efficient may be defined as follows:

$$\gamma = \left(\frac{E_c}{\sum_i Y} \right) \quad (10)$$

The labour market earnings for each income recipient are then adjusted by multiplying with the alignment co-efficient γ . This method assumes that earnings growth is constant across all income and person types and thus the earnings distribution remains constant over time. Once each individual that is in-work has received an income, we next calibrate our capital income variables.

One limitation of this methodology arises when calculating the stochastic component of the model for newly assigned workers. To calculate both of the stochastic terms we use the standard deviation of the entire population. Given that the standard deviation for each error term will be different for each age group, each occupation category, etc, by taking the standard deviation of the entire population we may introduce heteroscedasticity into our model estimates.

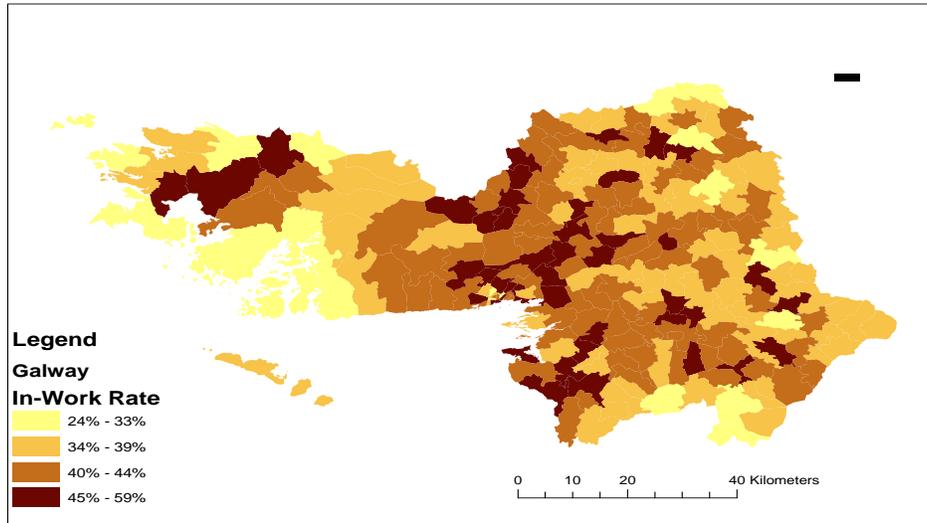
However, as outlined above, this paper is concerned with the generation of household market income, rather than individual market. As such, a final step is required. During the original SMILE matching process each individual is assigned a household. Therefore, the final step involves aggregating the calibrated individual income within each household. This aggregation of individual earnings to household earnings provides us with a representative market income distribution for each household in Ireland. Once we are satisfied that the calibrated data is both reliable and accurate, one can map the spatial distribution of the calibrated variables at the small area level. The next section provides the distribution of occupation, LFP and household income at the ED level for Co. Galway.

4. Results

As outlined above, the aim of this paper is to provide a methodology that may be used to produce a robust spatial distribution of market income and its determining variables at the small area level. This section provides an overview of the initial results obtained from the model. Figure 1 presents the spatial distribution of LFP in County Galway.¹

¹ The source of all figures is SMILE.

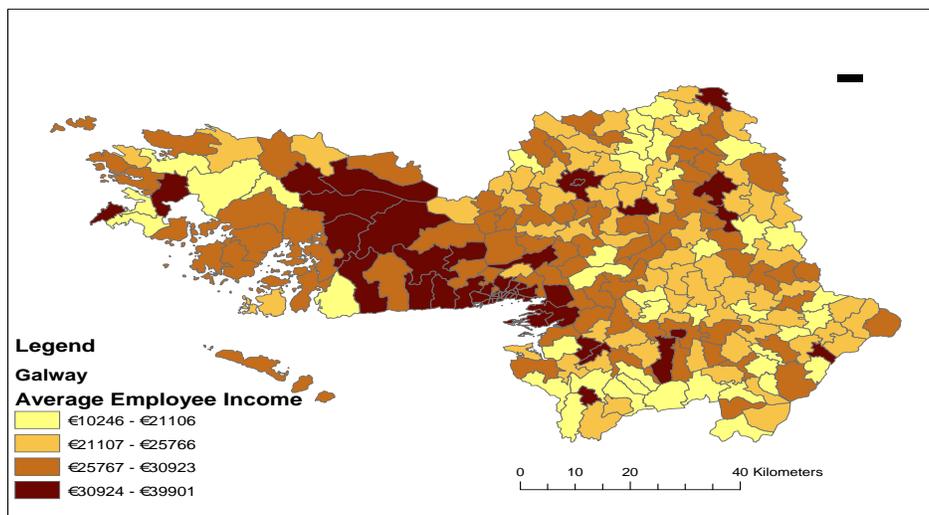
Figure 1. Average Percentage Rate of In-Work in County Galway.



As one can see LFP is highest in and around the hinterland of Galway city. LFP is also high in the East of the County, compared to the West.

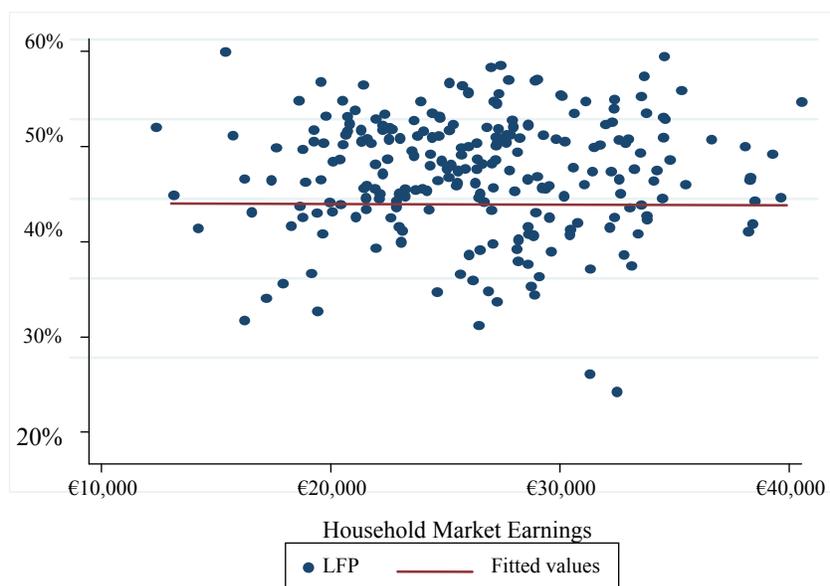
Figure 2 continues by presenting average household income for Co. Galway. As one can see, income levels are highest in and around the hinterland of Galway city, particularly to the West of the city. This is an interesting result. From Figure 1, one can see that LFP is higher in the East of the County, yet Figure 2 indicates the highest household incomes are in the West of the County.

Figure 2. Average Household Incomes in County Galway .



Further analysis, using a scatter graph (Figure 3) and plotting LFP and household income for each ED, shows that average household incomes at the ED level household are constant at approximately 43% LFP in an ED. That is, household income levels for each ED do not increase with higher LFP. This is somewhat surprising and therefore one must ask if LFP (above a certain level) is not driving average household income, what is?

Figure 3. The Relationship between LFP and Household Market Income at the ED Level .



As outlined above, as part of the calibration of market earnings, all LFP variables, such as occupation type and industry type were calibrated to ensure a representative distribution of market income. As demonstrated above LFP rates alone were unable to fully explain the spatial distribution of market incomes in Co. Galway. Thus, Figures 4 to 6 examine the relationship between occupation type and household income levels in Co. Galway.

Figure 4 provides a scattergraph of the relationship between household income levels at the ED level and the percentage of individuals employed within professional occupations. As one can see from Figure 4, there is a clear positive relationship between a higher rate of professionals in an ED and household market income. As the rate of professionals in an ED increase to the right of the scattergraph, so to do household incomes.

Figure 5 continues by providing a scattergraph of the relationship between household income levels at the ED level and the percentage of individuals employed within the construction industry. As one can see from Figure 5, highest household incomes are found in EDs with lower rates of individuals involved in the construction industry. There is a clear

Figure 4. The Relationship between the Percentage of Professionals and Household Market Income at the ED Level.

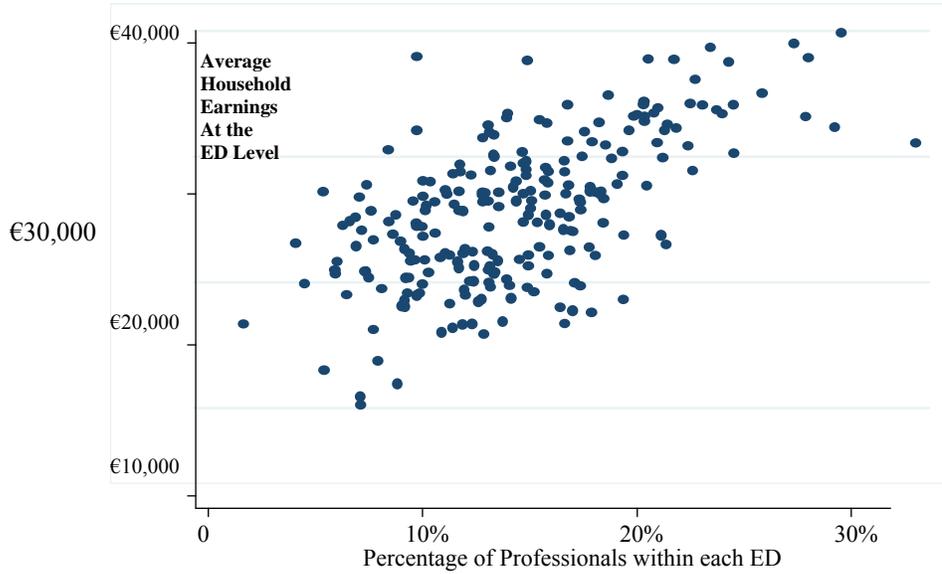
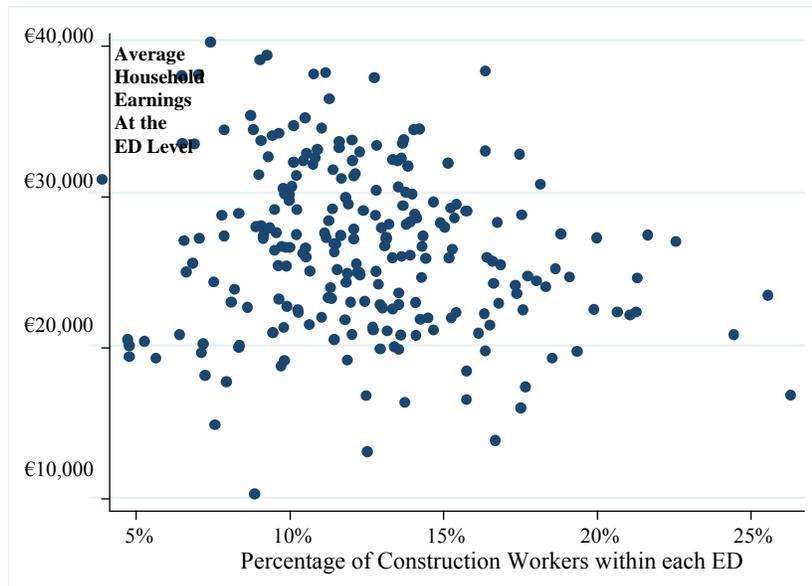


Figure 5 The Relationship between the Percentage of Construction Employees and Household Market Income at the ED Level

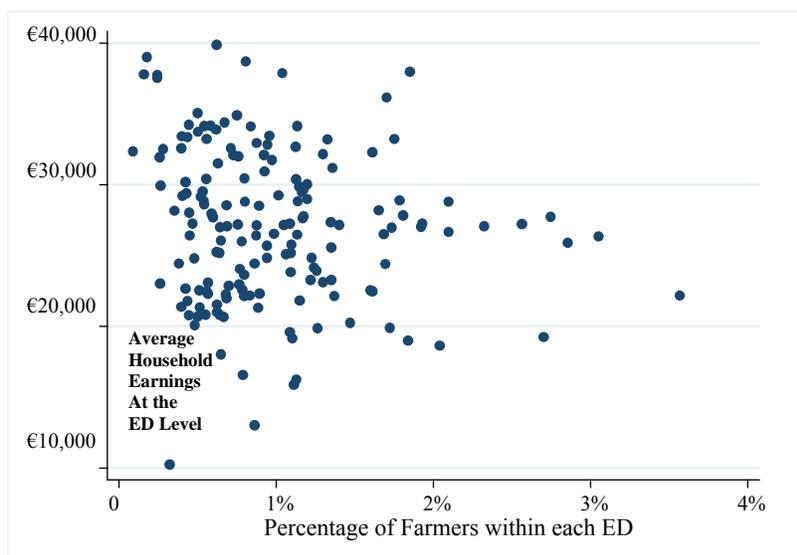


clustering of households with the highest market incomes to the left of the scattergraph. This is in direct contrast to the relationship between average household income and the rate of professionals in each ED presented in Figure 4.

Figure 6 provides a scattergraph of the relationship between household income levels at the ED level and the percentage of farmers. As one can see from Figure 6, highest household incomes are found in EDs with the lowest rates of farmers. The highest average household incomes are clustered in EDs where less than 1% of the population are farmers. This relationship is even more marked than the average number of individuals employed in construction and household market income presented in Figure 5.

As outlined in the introduction, the aim of this paper was to produce the spatial distribution of LFP and market income at the ED level for Co. Galway. Given the spatial differences found to exist with regard to distribution of market incomes, further analysis involved examining the determining which factors influence these differentials. Examining the spatial relationship between different LFP, occupation types and average household market income, it was found that LFP alone does not explain average household income levels. However, when one examines the spatial distribution of different occupation types, in this case, professional, construction workers and farmers, one can see a clear pattern arising. Rather than LFP alone, the spatial distribution of market income is driven by occupation type in Co. Galway.

Figure 6 The Relationship between the Percentage of Farmers and Household Market Income at the ED Level



5 Discussion

The aim of this paper was to generate representative household market income data at the small area level for Co. Galway. Using an alignment technique to calibrate the output from a spatial microsimulation model, the generation of small area data involved a two-step process. First, accurate spatial distributions of the relevant LFP variables such as the in-work, employee, farmer, occupation, etc variables were generated. This ensured that the underlying variables that influence household market income levels were accurate. The second step involved aligning the actual market income totals determined by SMILE in its initial match to the National Accounts totals for each county. On completion of the alignment process, SMILE offers a fully representative profile of LFP and market incomes at both the household and small area level.

Spatial Microsimulation models have long encompassed income components. Indeed Kalogirou and Hatzichristos (2006) state that the most advanced spatial income models are spatial Microsimulation models. For example, Birkin and Clarke's SYNTHESIS spatial Microsimulation model generated small area income data for Leeds (Birkin and Clarke, 1988) and the SimLeeds model (Ballas and Clarke, 2001). However, the unique aspect of the SMILE income generation model is that it provides a fully calibrated market income distribution based on individual differentials in the underlying determinants of market income across space. As such, using the data produced by SMILE, it was found that at the sub-national level, average LFP alone did not fully explain average household market incomes within an ED. Using the occupation-type data generated by SMILE, it was found that the higher the percentage rate of professionals in an ED, the higher average household market income for that ED. Conversely, it was further found that higher rates of both farmers and construction workers within an ED were correlated with lower average household market income. Thus, this analysis found that market income is determined by a number of different dimensions. However, it is important to note that it is only with the use of the data provided by SMILE and the calibration process outlined above that such an analysis and subsequent finding was possible.

The aim of this paper was to use microsimulation and alignment techniques to provide representative data on LFP and income distributions at the small area level for Co. Galway. Using these techniques to examine distributional differences in LFP and income levels, it was found that LFP alone is not a determining factor in wealth generation at the ED level. Instead, occupation and industry type is more indicative of income levels rather than LFP alone. These findings have important implications for a wide range of public policies, including health, education and social welfare in the long term.

For example, in terms of health policy, providing a cross-tabulation on the level of health against income inequality in twenty of the world's richest nations and each of the fifty US

States, Wilkinson and Picketts (2009) examine the effect of income inequality on health outcomes. Their analysis found that in states and nations with high rates of income inequality, mental illness, drug and alcohol and obesity rates are much higher and life expectancy is much lower. The Scandinavian countries and Japan, with their low levels of income inequality, consistently exhibited higher levels of psych-social health outcomes. In contrast, the UK, the USA and Portugal, with their higher levels of income inequality, consistently exhibited lower levels of psych-social health outcomes. The findings of Wilkinson and Pickett (2009) therefore indicate that government and policy-makers need to address income equality levels, rather than economic growth alone, to increase the health and happiness of its population.

Following on from Wilkinson and Pickett's findings, this paper provides empirical evidence that in an effort to target income inequality the Irish government needs to focus on not just employment levels, but also on occupation and industry type. It was found that EDs with a highest percentage of in natural resource based and manual industries, i.e. agriculture and construction had lower incomes on average than EDs with a higher percentage of professionals. This leads policy in the direction of providing more attractive educational opportunities to those initially entering the labour market. For those already involved in the labour market, spatially targeting additional up-skilling opportunities may be appropriate by state agencies. For example, the use of these additional skills may be through advancement into managerial positions for individuals that are employed, or an ability to provide more diversified services or products for those that are employed.

To date, unlike in the UK, USA and Australia, microsimulation models have not been used to guide public policy in Ireland. However, this paper demonstrates the potential uses of microsimulation, particularly spatial microsimulation, in providing a holistic analysis of a particular policy issue for policy-makers and national and regional government.

Future work with regard to model development will include linking SMILE to a national level tax-benefit model, so that non-market incomes and transfers may be simulated. An accurate measure of household income at any spatial level must take into account social welfare transfers and the taxes within a country. This means that the SMILE dataset will then contain a representative net income distribution for the whole of Ireland at the small area level.

Bibliography

- Anderson, B. (2007), Creating Small Area Income Estimates for England: Spatial Microsimulation Modelling, University of Essex Chimera WP: 2007-07.
- Azzoni, C. R. (2001), Economic Growth and Regional Income Inequality in Brazil, *Annals of Regional Science*, 35, 133-152.

- Bekkering, H. (2002), Micro-macro Linkages & the Alignment of Transition Processes, Technical Paper no. 25, National Centre for Social and Economic Modelling, University of Canberra.
- Ballas, D. and G.P. Clarke, (2001), Modelling the Local Impacts of National Social Policies: A Spatial Microsimulation Approach, *Environment and Planning C: Government and Policy*, 19, 587 – 606.
- Ballas D, G.P. Clarke and I. Turton, (2003), A Spatial Microsimulation Model for Social Policy Evaluation. In Boots B, Thomas R (eds), *Modelling Geographical Systems*, The Netherlands: Kluwer, pp. 143–168.
- Ballas, D., G. P. Clarke, D. Dorling, and D. Rossiter, (2007), Using SimBritain to Model the Geographical Impact of National Government Policies, *Geographical Analysis*, 39, 44-77.
- Ballas, D., D. Rossiter, B. Thomas, G.P. Clarke, and D. Dorling (2005), *Geography matters: Simulating the Local Impacts of National Social Policies*, London: Joseph Rowntree Foundation.
- Birkin, M. and M. Clarke, (1988), SYNTHESIS – A Synthetic Spatial Information System for Urban and Regional Analysis: Methods and Examples, *Environment and Planning A*, 20, 1645-1671.
- Bourguignon, F., F. Ferreira, and P. Leite (2002), Beyond Oaxaca-Blinder: Accounting for Differences in Household Income Distributions across Countries, *Texto Para Discussao*, No. 452, Departamento De Economia, Puc-Rio.
- Bourguignon, F., F. Ferreira, and N. Lustig, (2004), *The Microeconomics of Income Distribution Dynamics in East Asia and Latin America*, The International Bank of Reconstruction and Development/The World Bank, Oxford University Press.
- Caldwell, S. (1996), Health, Wealth, Pensions and Lifepaths: The CORSIM Dynamic Microsimulation Model, in Harding, A. (eds.), *Microsimulation and Public Policy*, Amsterdam: North-Holland pp. 505-521.
- Chin, S., A. Harding, R. Lloyd, J. McNamara, B. Phillips, Q. Vu, (2005), Spatial Microsimulation Using Synthetic Small-area Estimates of Income, Tax and Social Security Benefits, *The Australasian Journal of Regional Studies*, 11, 64-75.
- Cunfha, F. and J. Heckman (2007), The Technology of Skill Formation, *American Economic Review*, 97, 31-47.
- Dall'erba, D. (2005), Distribution of Regional Income and Regional Funds in Europe 1989-1999: An Exploratory Spatial Data Analysis, *Annals of Regional Science*, 39, 121-148.
- Demombynes, G., C. Elbers, J. Lanjouw, P. Lanjouw, J. Mistiaen, and B. Özler, (2002), Producing an Improved Geographic Profile of Poverty: Methodology and Evidence from Three Developing Countries, Discussion Paper no. 2002/39, United Nations University.
- Duncan, A. and M. Weeks, (1998), Simulating Transitions using Discrete Choice Models, *Papers and Proceedings of the American Statistical Association*, 106, 151-156.

- Elbers, C., J. Olsen Lanjouw, and P. Lanjouw, (2003), Micro-Level Estimation of Poverty and Inequality, *Econometrica*, 71, 355-364.
- Juhn, C., K. Murphy, and B. Pierce, (1993), Wage Inequality and the Rise in Returns to Skills, *The Journal of Political Economy*, 101, 410-442.
- Harding, A, R. Lloyd, A. Bill, and A. King, (2006), 'Assessing Poverty and Inequality at a Detailed Regional Level: New advances in Spatial Microsimulation' in M. McGillivray and M. Clarke, (eds), *Understanding Human Well-being*, Helsinki, United Nations University Press, pp. 239 - 261.
- Harding, A., N.Q. Vu, R. Tanton, V. Vidyattama (2009), 'Improving Work Incentives and Incomes for Parents: The National and Geographic Impact of Liberalising the Family Tax Benefit Income Test', *Economic Record*, 85, 48-58.
- Harding, A., J. McNamara, A. Daly, and R. Tanton, (2009), Child social exclusion: An Updated Index from the 2006 Census, *Australian Journal of Labour Economics*, 12, 41-64.
- Hentschel, J., J. Olson-Lanjouw, and P. Lanjouw, (1998), Combining Census and Survey Data to Study Spatial Dimensions of Poverty, World Bank Policy Research Working Paper Series No. 1928.
- Jenkins, S. P. (2000). Modelling Household Income Dynamics, *Journal of Population Economics*, 13, 529-567.
- Kalogirou, S. and T. Hatzichristos, (2006), A Spatial Modelling Framework for Income Estimation, *Spatial Economic Analysis*, 2, 1742-178.
- King, A., J. McLellan, and R. Lloyd, (2002), Regional Microsimulation for Improved Service Delivery in Australia: Centrelink's CUSP Model. Paper Prepared for the 27th General Conference of The International Association for Research in Income and Wealth Stockholm, Sweden. August 18 – 24, 2002.
- Le Gallo, J. and C. Ertur, (2003), Exploratory Spatial Data Analysis of the Distribution of Regional Per Capita GDP in Europe, 1980-1995, *Papers in Regional Science*, 82, 175-201.
- Long, J. and J. Freese, (2003), *Regression Models for Categorical Dependent Variables using Stata*, College Station, TX: Stata Press.
- Morrissey K., G.P Clarke, D. Ballas, S. Hynes, C. O'Donoghue, (2008), Analysing Access to GP Services in Rural Ireland using Micro-level Analysis, *Area*, 40, 354-364.
- Morrison, R. (2006). 'Make it so: Event Alignment in Dynamic Microsimulation', DYNACAN Team, Ottawa, Canada.
- Nolan, B. (2009), Income Inequality and Public Policy, *The Economic and Social Review*, 40, 489-510.
- Oketch, T., and M. Carrick, (2005), Calibration and Validation of a Micro-simulation Model in Network Analysis, TRB Annual Meeting, January, 2005, Washington D.C.

- Stuggard, N. (1996). Reconciliation of UK Household Income Statistics with the National Accounts, Paper presented at the Expert Group on Household Income Statistics, Canberra, Australia.
- Tanton, R., A. Harding, A. Daly, J. McNamara and M. Yap, (2009), Australian Children at Risk of Social Exclusion: A Spatial Index for Gauging Relative Disadvantage, *Population, Space and Place*, 10, 531-540.
- Tiglao, N.C. (2002), Small Area Estimation and Spatial Microsimulation of Household Characteristics in Developing Countries with a Focus on Informal Settlements in Metro Manila., unpublished PhD thesis, University of Tokyo.
- Williamson, P., M. Birkin, and P. Rees, (1998). The Estimation of Population Microdata Using Data from Small Area Statistics and Samples of Anonymised Records, *Environment and Planning A*, 30, 785-816.
- Wilkinson, R. and K. Pickett, (2008), *The Spirit Level: Why More Equal Societies Almost Always Do Better*, London: Allen Lane.