Building a Static Farm Level Spatial Microsimulation Model: Statistically Matching the Irish National Farm Survey to the Irish Census of Agriculture

Stephen Hynes

Teagasc and National University of Ireland, Galway

Karyn Morrissey

National University of Ireland, Galway and Teagasc

Cathal O'Donoghue

Teagasc, National University of Ireland, Galway, ICER, CHILD and IZA

Abstract

This paper reviews a statistical matching technique used to match the Irish Census of Agriculture to the Irish National Farm Survey (NFS) to produce a farm level spatial microsimulation model for Ireland. Using statistical matching techniques, economists can now create attribute rich datasets by matching across the common variables in two or more datasets. Static spatial microsimulation then uses theses synthetic datasets to analyse the relationships among regions and localities and to project the spatial implications of economic development and policy changes. The farm level spatial microsimulation model developed in this paper uses one of many combinational optimatisation techniques - simulated annealing - to match the Census and the NFS. We then use this matched NFS and Census information to produce small area farm population microdata estimates for the year 2002. Using the newly constructed farm level spatial microsimulation model and the associated spatially disaggregated farm population microdata set this paper then briefly analyses the spatial distribution of family farm income in Ireland.

Keywords: Static Farm Level Spatial Microsimulation Model, Simulated Annealing, Census of Agriculture, National Farm Survey, Family farm Income.

JEL Classification: Q12, R12

1. Introduction

Over the last decade, there has been an increase in interest in rural policy at the EU level. Indeed, the European Union Agenda 2000 agreement on agriculture contains an increased focus on rural development, acknowledging it as the second pillar of the Common Agricultural Policy. This increased focus on rural development at the EU level means that member states have also increased their emphasis on rural development along a range of spectrums, such as reducing socio-economic disparities between regions, enhancing employment and competitiveness in rural areas, and re-orienting agricultural production. However, while other EU member states try to redirect their rural development policies, rural Ireland is facing uncertainty and concern over its future viability. The main sources of concern for rural Ireland include its declining population, the disadvantages in attracting new jobs and retaining existing employment, persistent relative poverty and the decreasing number of farm and farm related jobs (Ballas et al., 2005). According to the Irish White Paper on Rural Development, one of the main goals of rural development policy in Ireland is to maintain "the maximum number of rural farms and especially family farms" (White Paper, 2000).

It is against this background that Irish rural development policy has developed over the last decade. As a result, a programme of collaboration between the Rural Economy Research Centre, Teagasc and the University of Leeds was initiated to develop a microsimulation model capable of analysing the impacts of different rural development policies. The Simulation Model for the Irish Local Economy (SMILE) was designed to analyse the relationship between regions and to project the spatial implications of economic development and policy change in rural areas (Ballas et al, 2001). The static SMILE model has spatial information on farmers derived from the European Community Farm Panel (ECHP) dataset. However, the information concerning these farmers is limited to their location (Electoral Division, (ED)) and their income level. By statistically matching farms in the Census of Agriculture to the National Farm Survey (NFS) we will have a much deeper understanding of farm activity at the local level than is available in the SMILE model or any other farm dataset in Ireland thus providing policy-makers with a much better spatial map of farming activity in the country. The Static Farm Level Spatial Microsimulation Model developed in this paper is an object-orientated model, built in Java, using the same framework as the SMILE model. Object-orientated modelling has, amongst other benefits, the added advantage of platform-independence. Object-orientated models have especially high potential for modelling in Regional Science fields[1].

As already mentioned the 2 datasets used in this analysis are the Census of Agriculture to the National Farm Survey (NFS). While neither the Census nor the NFS alone provides policy-makers with a complete overview of all of the important farming activities and attributes at the local level, if combined the two datasets would provide policy-makers with detailed synthetic microdata as to inform their decision-making a spatially disaggregated level. It is for this reason that we use a statistical matching technique to combine the Census of Agriculture to the NFS. By doing so, agricultural attributes can be analysed at the electric division (ED) level in Ireland. The rest of this paper is structured as follows. In section 2 we present the datasets used in the matching process. Section 3 then looks at the microsimulation methodology required to statistically match the NFS and the Census of Agriculture. The validation of the matching process takes place in section 4, where we look at the matched tables, z-scores, and z^2-scores. Section 5 presents an application of the static farm level microsimulation model by looking at the spatial distribution of family farm

[1] Spatial microsimulation exploits the benefits of object-orientated programming both as a tool and a concept. Spatial microsimulation frameworks use a list-based approach to microdata representation: a household or an individual has a list of attributes that are stored as lists rather than as occupancy matrices (Williamson et al., 1996). From a computer programming perspective, the list-based approach uses the tools of object-orientated programming because the individuals and households can be seen as objects with their attributes as associated instance variables. For a technical discussion of the java based framework used in the development of the SMILE model and adapted to run the static farm level model discussed in this paper see Kelly (2004).

income in Ireland using GIS mapping techniques. Finally, section 6 concludes with a discussion of the policy implications of our results and some recommendations for further research.

2. Data

In this section we briefly describe the data used in this paper. *The National Farm Survey (NFS)* was set up in 1972 and has been published on annual basis since. The NFS is collected as part of the Farm Accountancy Data Network of the European Union (FADN). The aim of this network is to gather accountancy data from farms in all member states of the EU for the determination of incomes and business analysis of agricultural holdings (FADN, 2005). In line with FADN, the main objectives of the NFS are firstly, to determine the financial situation on Irish farms by measuring the level of gross output, costs, income, investment and indebtedness across the spectrum of farming systems and sizes and secondly, to provide data on Irish farm income to the EU Commission in Brussels and to provide a database for economic and rural development research and policy analysis (The NFS, 2002).

To achieve these objectives, a farm accounts book is recorded for each year on a random sample of farms throughout the country. The data in the NFS is collected on an ongoing basis (3-4 times a year) by NFS 'recorders'. The recorders collect the data in-person from each farm. The information for the sample of farms comes from the Central Statistics Office of Ireland (CSO). In general there are 1,200 farms in the survey each year[2]. The matching process described in the next section uses the 2002 NFS and contains 1,177 farms. Table 1 contains some summary statistics on the sample for 2002.

[2] The weights used to make the NFS representative of the Irish farming population are based on the sample number of farms and the population number of farms (from the Census of Agriculture) in each farm system and farm size category. The sample number of observations by size/system is simply divided by the population number of observations by size/system to get the weights that make the sample representative of the actual farming population.

Table 1. Summary Statistics for Farm Variables from the National Farm Survey.

Variable	Mean	Std. Dev.	Min	Max
Farm Size	33.82	34.12	2.01	319.56
Total labour Hours Employed on Farm per Year	2552.94	2196.44	1.00	23928.00
Gross Output	72304.66	72962.96	1456.00	760150.30
Total Costs	49337.27	55009.03	1215.58	533558.10
Gross Margins	46653.72	46811.25	-12459.89	566768.30
Family Farm Income	22967.40	24330.05	-47039.41	281226.30

The method of classifying farms into farming systems, used in the NFS is based on the EU FADN typology set out in the Commission Decision 78/463. The system titles refer to the dominant enterprise in each group based on Standard Gross Margins (SGMs). Within the NFS, the farm system variable is broken down into six different categories as follows: Dairying, Dairying and Other, Cattle rearing, Cattle Other, Mainly Sheep and Tillage Systems.

The other dataset used in this paper is the *Census of Agriculture.* The Census of Agriculture began in 1847 and was last conducted in June 2000. The Census in 2000 was the first full census to be conducted since 1991, thus keeping in line with the general practice of conducting a full census approximately every 10 years. The 2000 Census of Agriculture was conducted entirely by post. The objective of the Census was to identify every operational farm in the country and collect data on agricultural activities undertaken on them (CSO, 2000). The scope of the census was all farms, where the agricultural area used for farming was at least 1 hectare. The census classifies farms by physical size, economic size, economic type and geographical location.

To obtain the full population of farms in Ireland, a register was drawn up based on the main client file belonging to the Dept of Agriculture. This register comprises people who have registered with the Dept of Agriculture to avail of agricultural subsidies, and to comply with the Departments agricultural regulations, and is though to include the majority of active farms. However, while the register contains about 190,000 farms, it was expected that there would be only about 140,000 active farms (CSO, 2000).

Due to the Commission decision 78/463ECC all the farms covered in the 2000 Census of Agriculture are classified down to the most detailed farm system classification (Projet de Decision de la Commission, 1992). However, as many of the farm system types present in the Commission decision 78/463/EEC are not used in Ireland, seven summary farm type classes of general interest to Irish agriculture were selected from the EU typology as follows (Census of Agriculture, 2000): Specialist Tillage, Specialist Dairying, Specialist Beef Production, Specialist sheep, Mixed grazing livestock, Mixed crops and livestock. In the EU FADN system the main method in which farm systems are determined is by Standard Gross Margins (SGMs). This 'typology' classifies farm systems according to their main source of income, using standard gross margins (Connelly and Kinsella, 2005)[3].

With regard to the matching of the Census and the NFS a problem encountered was that in the Census, for a small proportion of EDs, some details were not made available due to confidentiality or non-response. Furthermore, it was found that the two variables, farm size and farm system, were rounded to the nearest decile in a further effort to increase the confidentiality of the census. To overcome this rounding up problem, a one-stage iteration method was applied. This one-stage iteration involved generating a new variable for the six farm size categories, that is, rounded up to the nearest one rather than ten. The iteration process was also carried out in the same manner for the farm system variable in the Census of Agriculture.

While both the NFS and the Census of Agriculture provide a comprehensive coverage of Irish farm farms they separately have several major limitations. The NFS contains a large amount of

[3] Each type of agricultural production, whether crop or livestock, is assigned a standard gross margin (SGM). SGM is a concept similar to value added. SGM is defined as the difference between the standard value of the gross agricultural product and specific costs that are allocated (Connelly and Kinsella, 2005). Data from a number of sources are used to compile SGMs for enterprises in Ireland. However, the predominant source is the NFS.

information on farming activity but is only nationally representative and cannot be used for analysis at the local level. On the other hand, the Census of Agriculture has limited individual farm information and some information is unavailable due to confidentiality issues. It does however have information on a small number of key farm variables at a very local level (ED). Thus, to enrich our knowledge of farming activity at a more regional level in Ireland, we combine these two datasets in this paper to create a more attribute rich synthetic farm dataset.

3. Methodology

A microsimulation model uses microdata on individuals; farms and firms to build large-scale data sets based on the real-life attributes of individuals, farms or firms and then simulates the effect of changes in policy on each of these units. By permitting analysis at the individual level, microsimulation methods allow one to assess variations in the distributional effects of different policies (Wiemers et al. 2002, Holm, 1996). Within the microsimulation framework, the differences before and after the policy change can be analysed at the micro-level or aggregated to show the overall national effect of the change. It is the dependence on individual information from the micro-data at every stage of the analysis that distinguished microsimulation models from other sorts of economic, statistical and descriptive models. Modern policy problems require analysts to capture the interactions between policy and the complexities of economic and social life as well as between policies of different types (Mitton et al, 2000). With the development of increased computing power and analytical techniques, microsimulation is becoming the chosen technique to analyse these policy problems.

Traditionally, microsimulation models are divided into two types: static and dynamic microsimulation (O'Donoghue, 2001). However, a third type of microsimulation model, spatial microsimulation models are becoming increasingly useful and common. Static models examine micro units at one point in time, and have been used extensively to examine the differential impact of budget changes. Dynamic microsimulation projects the population in the base year forward through time by simulating transitions such as fertility and mortality at the individual level. Finally spatial microsimulation contains geographic information that links micro-units with location and therefore allow for a regional or local approach to policy analysis.

Static spatial microsimulation is designed to analyse the relationships among regions and localities and to project the spatial implications of economic development and policy changes in at a more disaggregated level (Holm, 1996). Spatial microsimulation has four main advantages over more traditional micro models. First, it allows data from various sources to be linked if datasets contain at least one attribute in common (for example the variable farm size in the NFS dataset and the Census of Agriculture dataset). Secondly, the models are flexible in terms of spatial scale that is data can be re-aggregated or disaggregated. For example, the results of the Static Farm Level Spatial Microsimulation Model developed in this paper can be aggregated to counties (by ED), regions (by province) or the country as a whole. Third, spatial microsimulation models store data efficiently as lists. Finally, the models allow for updating and projecting.

The microdata used in static microsimulations generally consists of a list of unidentifiable individuals or farms with associated characteristics obtained as mentioned above, from a survey or census. This data set can then help to fill the deficiency in the information available to policy makers. Melhuish et al, (2002), outlines 3 main benefits of creating synthetic microdata. Firstly it allows the creation of spatially disaggregated data from aggregated such as national surveys (e.g. the NFS). Secondly, the many simulated characteristics of each individual or farm can be used for multivariate analysis, thereby providing a method of identifying and analysing specific socio-demographic groups at the ED level. Finally, creating synthetic spatial microdata gives the researcher the potential to estimate the spatial impact of policy change on particular groups within the population. However, one disadvantage of microdata is that validation of the results is difficult.

This is an obvious fact given that one of the objectives of creating synthetic microdata is to create data that does not currently exist for small geographic areas.

The statistical matching technique used in this paper uses farm survey microdata to generate a synthetic farm population that fits known small area characteristics. While iterative proportional fitting (IPF) methods are commonly used to generate synthetic microdata, our static farm level spatial microsimulation model uses combinational optimisation methods where existing microdata sets are reweighted to fit Small Area Population Statistics (SAPS) (Kelly, 2004)[4].

We use a combinational optimatisation process called simulated annealing. Simulated annealing estimates small area farm microdata by selecting a series of SAPS tables to describe the small areas of interest and determines the records from the microdata survey that best match these tables (Kelly, 2004). The combinational optimisation method of producing small area population microdata requires two types of data as input: population microdata and geographically referenced SAPS. In our farm level model we use the NFS data as the population microdata and the SAPS tables from the Irish Census of Agriculture as the geographically referenced small area population data.

The origin of the simulated annealing method is in thermodynamics and dates back to the 1950s, when Metropolis et al. (1953) suggested an algorithm for the efficient simulation of the evolution of a solid material to thermal equilibrium. Annealing is a physical process in which a solid material is first melted in a heat bath and then it is cooled down slowly until it crystallises (Laarhovern and Aarts, 1987, Dowsland, 1993, Pham and Karaboga, 2000). The solid material is heated up by increasing the temperature of the heat bath to a maximum value at which all particles of the solid have high energies and the freedom to randomly arrange themselves in the liquid phase. This phase is then followed by a cooling phase, in which the temperature of the heat bath is slowly lowered.

[4] For further discussion of the iterative proportional fitting methodology see Norman.1999 and Wong, 1992.

The particles of the material attempt to arrange themselves in a low energy state during the cooling phase. When the maximum temperature is sufficiently high and the cooling is carried out sufficiently slowly then all the particles of the material eventually arrange themselves in a state of high density and minimum energy. According to the laws of thermodynamics at each temperature with value T where the solid is allowed to reach thermal equilibrium, the probability of change in energy of magnitude δE is given by the Boltzmann distribution:

$$p(\delta E) = \exp(-\delta \mathrm{E} / kT)$$

where k is a physical constant known as Boltzmann's constant (Laarhoven and Aarts, 1987, Dowsland, 1993).

Metropolis et al. (1953) first realised that the annealing process can be simulated for a fixed temperature by Monte Carlo methods to generate sequences of energy states. The use of simulated annealing as an optimisation method was suggested in the early 1980s when Kirkpatrick et al. (1983) discovered an analogy between minimising the cost function of a combinatorial optimisation problem and the slow cooling of a solid until it reaches its low energy ground state. Since then, simulated annealing has been employed as an optimisation technique to solve a variety of combinatorial optimisation problems (Dowsland, 1993, Laarhoven and Aarts, 1987).

In this paper we use the simulated annealing approach to estimate spatially disaggregated microdata for Irish rural areas using data from the 2002 NFS and SAPS tables from the 2000 Census of Agriculture. Simulated annealing is used to select a set of farms from the 1177 records of the NFS that best fits the Census small area constraints. These small area constraints are the following SAPS tables: Table 1—Farm Size in hectares; Table 2—Farm System and Table 3—Soil Class. Tables 1 and 2 were adapted so that category definitions from the NFS matched those used in the Census of Agriculture. Broadly, the simulated annealing process works by first reading in these 3 SAPS tables and the NFS data. It then selects NFS farms at random to population the SAPS tables, applies a simulated annealing algorithm to find the best fitting set of farms and saves the set of NFS farms that best fits the SAPS tables. The remainder of this section describes this process in more detail.

An initial random sample of records from the NFS is selected until sufficient farms are represented. These records area then used to create tables that match the selected SAPS tables. Each pair of tables is then compared to calculate the relative error between the two tables. A number of records in the set are then selected at random and replaced with ones chosen at random from the universe of records. The error is then recalculated and the change in error (Δe) is calculated. If Δe is less than zero then there has been an improvement and the changes are accepted. Simulated annealing also allows sub optimal changes to occur.

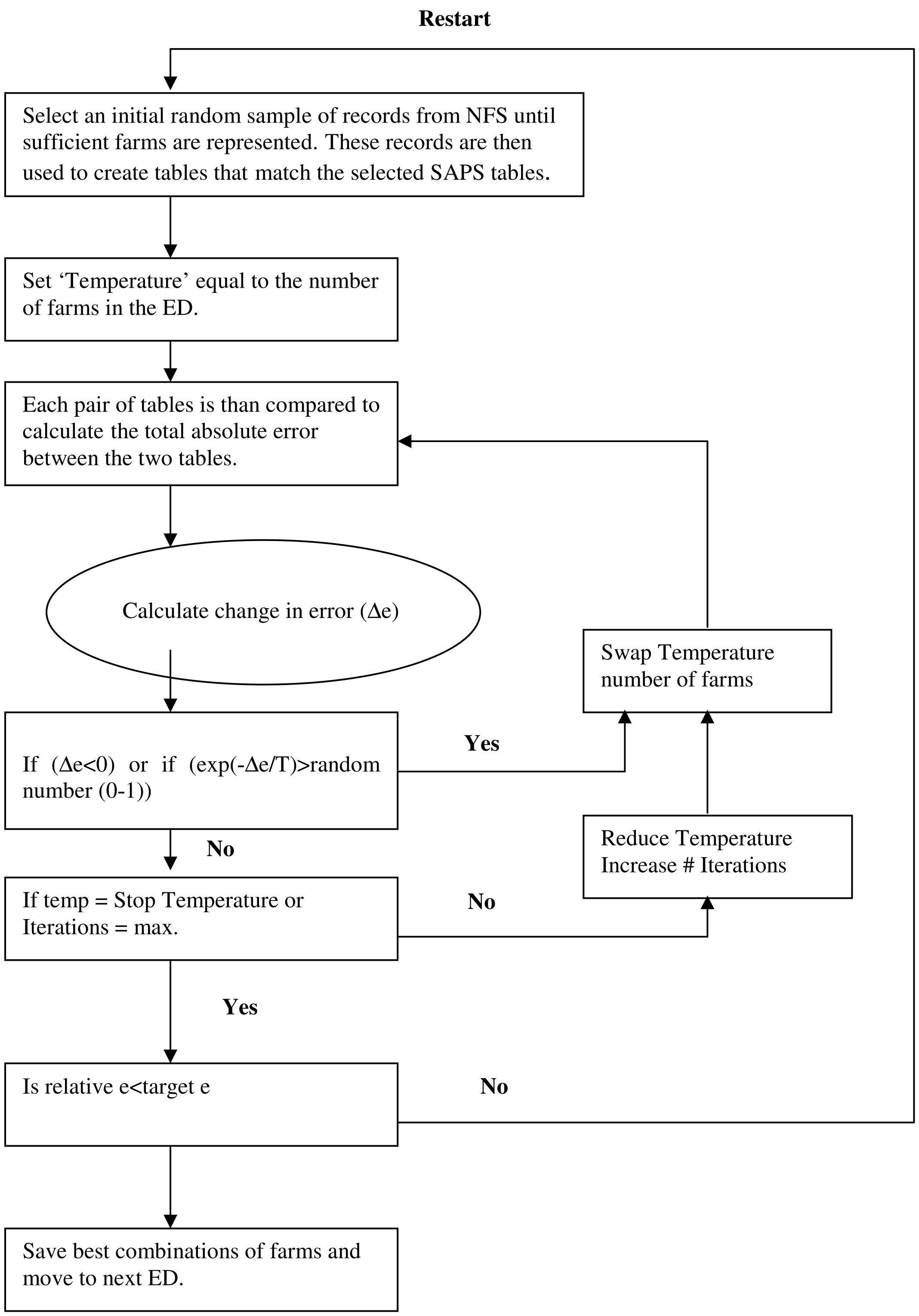
If Δe is positive, $\exp(-\Delta e/T)$ is compared to a random number between 0 and 1. If it is greater than the random numbers then the changes are accepted, otherwise the changes are rejected and reversed. In this implementation if Δe is zero the change is accepted to allow the exploration of a greater part of the solution space. If the new error is the best seen so far the set of farms used is stored. As the simulation progresses, the number of records selected at one time decreases. This process allows a faster rate of improvement in the error term. The static model also employs a restart method. When a restart occurs the simulated annealing process begins again with a new sample of records. The restart is used so that more farm combinations can be explored. The simulation is complete when the total relative error is less than a specified target, in our case 0.05.

Matching the NFS and the SAPS data creates synthetic demographic, socio-economic and farm level variables, such as martial status, age, fertiliser usage, livestock units per farm, etc. When the two datasets are matched, a synthetic population dataset is created, whereby the individuals in the new synthetic dataset have the attributes of the two original datasets combined. As one can imagine, there are a vast number of possible sets that could be drawn from each of the datasets, thus combinational optimisation techniques are used to find a set that fits the target SAPS tables as well as possible.

As shown in Figure 1, the matching process described above works as an analogy of the physical process of annealing, as described by Metropolis et al (1953). The ‘temperature’ (or number of farms in the EDs) is a control parameter and is initially set high and then slowly lowered after a set of iterations has taken place. In practice, the ‘temperature’ selected is equal to half the total number of farms in the ED. This means for an ED containing 100 farms, 50 farms are swapped in the first iteration. As the error decreases, the number of farms per swap is reduced to 1. The number of iterations is inversely proportional to the ‘temperature’, so that as the number of farms per swap is reduced the number of iterations is increased.

There are advantages and disadvantages to using simulated annealing over other methods such as IPF. Its main advantage is that it uses real microdata to generate small area population data rather than using synthetically created microdata. Furthermore, because IPF uses probabilities to create synthetic farms rather than using actual farms from survey data it can produce some unrealistic farms. Simulated annealing only produces farms that exist in the survey dataset. The main disadvantage of simulated annealing however is its computational intensity. The Static Farm Level Spatial Microsimulation Model produces almost 140,000 individual farm records and takes approximately 2 days to run on a DELL workstation.

Figure 1. Static Farm Level Spatial Microsimulation Flowchart



4. Results I: Validation

In this section, we examine ways in which synthetic microdata can be validated. We look at the 4 ways in which the Static Farm Level Spatial Microsimulation Model validates the microdata that it produces – z-scores, z^2-scores, relative error measures and re-aggregating output in order to compare results to other relevant datasets at the re-aggregated level. The first three of these validation methods are hard-coded within the model and are produced every time a simulation is run.

As Ballas et al. (2001) point out; one of the biggest drawbacks of microsimulation models is the difficulty in validating the model outputs. This is due to the fact that microsimulation models estimate distributions of variables which were previously unknown. However, one way of validating microsimulation model outputs is to re-aggregate estimated data sets to levels at which observed data sets exist and compare the estimated distributions with the observed.

The static model developed in this paper uses three different statistics to assess (internally) the models goodness-of-fit: total absolute error, relative error and z-scores (Kelly, 2004). Farms are added or removed in the simulation process based on the total absolute error of all the target tables. The simulation process is ended based on the total relative error of the target tables. The relative error result for each table is calculated by dividing the total deviations of the estimated table from the actual table by the sum of the cells in the actual table. The relative error is chi-squared distributed at the 95% level. Finally, as a further validation exercise Z-scores for cellular fit and Z^2 –scores for tabular fit are calculated and outputted along with the results.

The Z-score is based on the difference between the relative size of the category in the synthetic and actual populations, although an adjustment is made to the formula when dealing with zero counts. A Z-score can be summed and squared to provide a measure of tabular fir similar to a chi-squared statistic. If a cell's Z-score exceeds the critical value, the cell is deemed not to fit, while if a Z^2 –

score exceeds the critical value, then the dataset is deemed not to fit (i.e. |Z|>1.96). The Z score calculation is given by:

$$Z = \frac{\frac{T_{ij} - O_{ij}}{\sum_{ij} O_{ij}} \pm \frac{1}{2 \times \sum_{ij} O_{ij}}}{\sqrt{\frac{\left(\frac{O_{ij}}{\sum_{ij} O_{ij}}\right)\left(1 - \frac{O_{ij}}{\sum_{ij} O_{ij}}\right)}{\sum_{ij} O_{ij}}}}$$

Where: T_{ij} is the estimated data, column i, row j.

O_{ij} is the census data, column i, row j.

$\sum_{ij} O_{ij}$ is the sum of all the elements in the table.

The $\frac{1}{2 \times \sum_{ij} O_{ij}}$ stochastic component is added or subtracted because in some large tables it is possible to have 0 values, and then we would have division by Zero. Add the stochastic component if $T_{ij} < O_{ij}$ and subtract if $T_{ij} > O_{ij}$. Of course if the observed and the expected are the same then Z is 0. We use the above formula to calculate the Z score. It is easy to see from the sample of Z squared results presented in Table 2 which tables and which EDs fit the best.

Table 2. Microsimulation Validation for a Sample of EDs

Farm System	Specialist Tillage	Specialist Dairying	Specialist Beef Production	Specialist sheep	Mixed grazing livestock	Mixed crops & livestock	Other	Total No. of Farms	Z^2 – score
1. Census Table									
DED: 24	10	0	10	20	20	10	0	70	
DED: 25	14	0	0	0	0	14	0	28	
DED: 26	10.33333	0	10.33333	0	0	10.33333	0	31	
DED: 27	0	0	9.666667	0	9.666667	9.666667	0	29	
DED: 28	10	0	10	0	10	0	0	30	
DED: 29	0	0	10.75	10.75	21.5	0	0	43	
DED: 30	11	0	0	0	11	0	0	22	
DED: 31	0	12.25	12.25	0	12.25	12.25	0	49	
2. Estimated Table									
DED: 24	10	0	11	20	20	9	0	70	
DED: 25	14	0	0	0	0	13	0	27	
DED: 26	11	0	10	0	0	10	0	31	
DED: 27	0	0	10	0	10	9	0	29	
DED: 28	10	0	10	0	10	0	0	30	
DED: 29	0	0	11	11	21	0	0	43	
DED: 30	11	0	0	0	11	0	0	22	
DED: 31	0	13	12	0	12	12	0	49	
3. Z-Scores	X Squared Critical Value:		2.16		Degrees of Freedom:		7		
DED: 24	0	0	0.1641	0	0	-0.1641	0		0.0538
DED: 25	0	0	0	0	0	-0.1543	0		0.0238
DED: 26	0	0	0	0	0	0	0		0
DED: 27	0	0	0	0	0	0	0		0
DED: 28	0	0	0	0	0	0	0		0
DED: 29	0	0	0	0	0	0	0		0
DED: 30	0	0	0	0	0	0	0		0
DED: 31	0	0	0	0	0	0	0		0

Information on the relative error and the z-scores are outputted automatically in the static simulation. As shown in table 2, the first line in section 3 of the table shows the degrees of freedom and associated 95% critical value for the Z^2 –score. The degrees of freedom are the number of columns in the table that represent a farm system. As there are seven such columns, the associated degrees of freedom for specialist are 2.16. Taking ED 26 as an example, the z^2 –score of zero indicates that the estimated tables fit the actual tables. Also for this ED, the Z-score is zero across all cells, indicating that the estimated cells fit the actual cells from the Census perfectly. On the other hand in ED 24, cell 3 is 0.1641 as is cell 6. This is above zero but still does not exceed the critical value, i.e. these cells still fit the actual cells at the 95% confidence level and its Z^2 – score is also below the critical value (0.0538), thus indicating that the estimated table still fits the actual table.

Examining the actual and estimated *System* variables in Table 2 will verify these statistics. The census and estimated tables for ED 24 to ED 31 for the variable specialist are shown in the first and second sections of Table 2 respectively. On examining the estimated and actual farm numbers per ED, the two tables do correspond for ED 24, as was indicated by our Z-scores. However, we can see by comparing the estimated and actual tables, that cell 5 for ED 24 tells us that there are 10 specialist beef farmers and 10 mixed grazing and livestock farmers in ED 24, while the information from the Census indicates that are 11 and 9 such farms respectively in ED 24. There are corresponding Z-score results produced by the model for each of the other three SAPS tables – farm numbers, farm size and soil code[5].

As well as these internal validation measures we can also validate the synthetic microdata estimates produced by Static Farm Level Spatial Microsimulation Model by re-aggregating the model results up to the county and national level and then comparing the estimates against Irish Central Statistics Office figures for average farm size at the county level and a cross-tabulation of farm size and system at the national level. This analysis at the national and county level of farm size and system is a further validation of our synthetic microdata and in turn it validates the z-score and z^2-score results discussed above.

[5] To save space these results are not presented here but are available from the authors upon request.

Table 3. Microsimulated Estimates of Average Farm Size at the County Level, compared to Actual Average Farm Size from CSO Statistics

County	Microsimulation Model	Census of Agriculture	% Error
Carlow	37.73	38.30	-1.49
Cavan	25.94	25.20	2.94
Clare	30.68	31.30	-1.98
Cork	36.54	37.50	-2.56
Donegal	27.16	26.20	3.66
Dublin	62.92	42.20	49.10
Galway	23.69	24.60	-3.70
Kerry	32.38	32.70	-0.98
Kildare	44.34	41.80	6.08
Kilkenny	43.46	42.60	2.02
Laois	34.38	35.30	-2.61
Leitrim	23.50	24.60	-4.47
Limerick	33.33	32.60	2.24
Longford	26.76	26.90	-0.52
Louth	40.13	35.10	14.33
Mayo	23.21	21.90	5.98
Meath	41.07	40.20	2.16
Monaghan	22.74	21.80	4.31
Offaly	32.55	34.50	-5.65
Roscommon	25.11	24.80	1.25
Sligo	23.67	24.50	-3.39
Tipperary North	37.66	38.80	-2.94
Tipperary South	39.63	40.70	-2.63
Waterford	41.89	44.60	-6.08
Wexford	39.64	40.10	-1.15
Wicklow	45.18	42.20	7.06
Westmeath	33.69	34.90	-3.47

Table 3 demonstrates that the estimates for average farm size at the county level derived from the synthetic microdata are approximately the same as the average farm sizes from the Census of Agriculture. Only for counties Dublin and Louth is there are a greater than 10% difference between the estimated and actual average farm size. This comparison further validates the z-scores and z^2 –scores taken from the Static Farm Level Spatial Microsimulation Model.

The next comparison between the synthetic microdata estimates and the actual Census of Agriculture results is a cross tabulation of farm size and system at the national level. As one can see from table 4 the majority of results are quite accurate. With regard to farms of less than ten hectares, at the national level, the Static Farm Level Spatial Microsimulation Model estimated that there were 26,746 while according to the Census there are 28,419[6]. With regard to farms between twenty and thirty hectares the Static Farm Level Spatial Microsimulation Model estimated that there were 25,230, while according to the Census there are 25,045.

Table 4. A Cross Tabulation Comparison of the Synthetic SMILE Microdata and CSO Statistics for Farm Size and System at the National Level*

Census	Farm Size											
	10-20		20-30		30-50		50-100		>100		Total	
System												
Tillage	919	**1,107**	633	**599**	877	**955**	890	**568**	671	**474**	4736	**3,703**
Dairying	3159	**2,801**	4992	**5,185**	9038	**9,576**	6975	**6,852**	938	**304**	26292	**26,486**
Beef	21890	**22464**	13637	**13,138**	12236	**10926**	5728	**5,006**	1042	**364**	70141	**75,408**
Sheep	3042	**2,849**	1917	**2,096**	2097	**3,303**	1270	**2,081**	624	**350**	12233	**10,679**
Mixed Grazed Livestock	4541	**4,926**	3206	**4,212**	4215	**5,262**	3487	**4,075**	880	**747**	20729	**20,690**
Mixed Crop Livestock	500	**0**	488	**0**	924	**0**	965	**2,118**	400	**0**	3644	**2,118**
Other	239	**0**	172	**0**	240	**0**	220	**0**	56	**198**	1752	**198**
Total	34290	**34,147**	25045	**25,230**	29627	**30,022**	19,535	**20700**	4611	**2,437**	141527	**139,282**

* Figures in bold represent the national size-system aggregates of the Static Farm Level Spatial Microsimulation Model. Figures not in bold are the corresponding totals according to the Census of Agriculture 2002.

For farms between fifty and a hundred hectares the Static Farm Level Spatial Microsimulation Model estimated that there were 20,700 farms in this category, while the actual Census records 19,535. Thus one can see that while the Static Farm Level Spatial Microsimulation Model estimates are not exact at national level for farm size they do closely approximate the actual Census data. A similar story can be told for the totals of each farm system. The only category of farm system that does compare well at the national level is "other". This may be explained by the fact that within the NFS, data on horticulturalists, pig and poultry producers are not collected. Since these farm types should make up the vast majority of this "other" category it is not surprising

[6] To save space in the table the column for farm size < 10 hectares is not shown but the results are available from the authors upon request.

that the estimate is so poor (1752 being the actual census total of this category and 198 being the estimate from our model).

5. Results II: Spatial analysis using the Static Farm Level Spatial Microsimulation Model

There has been extensive research done on the distribution of Irish farm earnings at particular points in time and at the national level (Nolan.2005, Matthews, 2000, Keeney et al. 1997, Fitzpatrick and associates, 1997, Hynes and O'Donoghue, 2004). On the other hand, surprisingly little attention has been given to the distribution of Irish farm earnings across geographical space. One recent report by Watson et al. (2005) did attempt to conduct an analysis of farms at a more disaggregated level. Their study looked at the distribution of farmers by farm size and age of operator at the county level using Census of Agriculture figures. As with previous studies however the Watson report was constrained to only looking at the *numbers* of farmers in each county by age and size as they did not have the spatial farm level micro dataset that has been constructed for the first time in this paper.

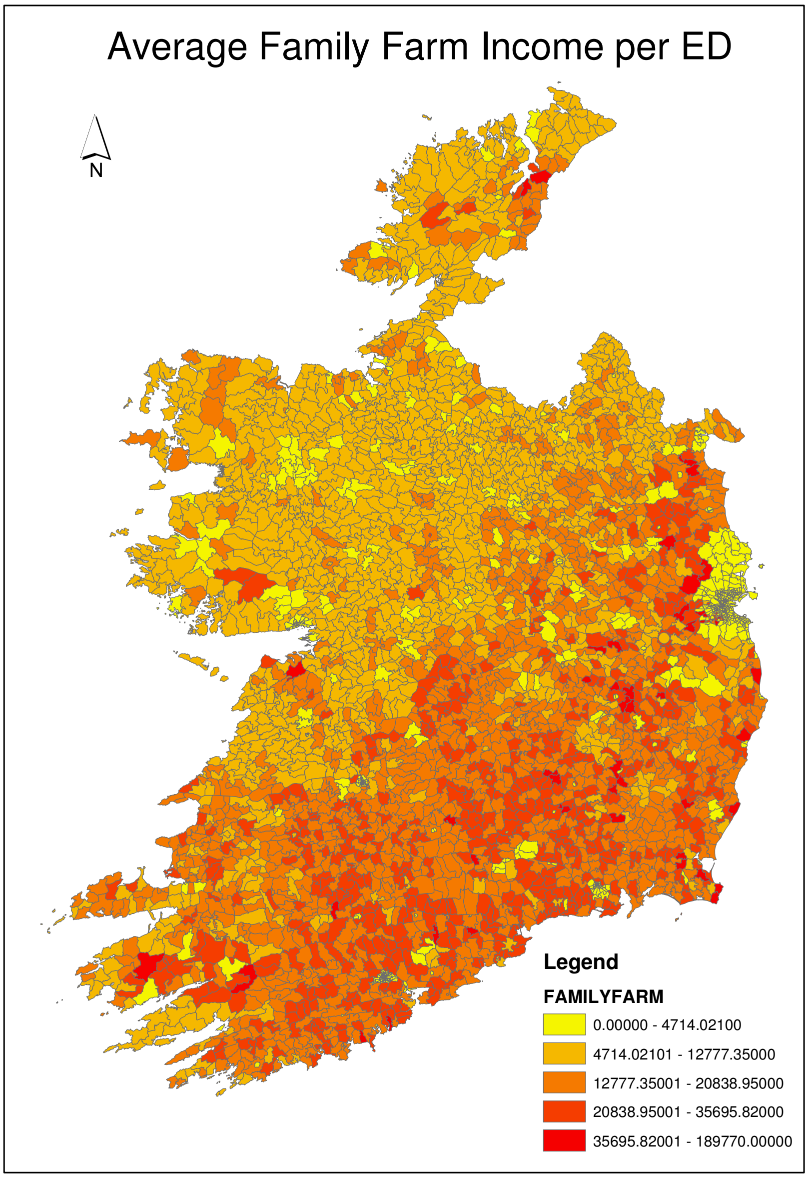
The Teagasc National Farm Survey for 2002 showed a decline of 5.8% in farm income, bringing average income per farm to €14,925. Average income in 2001 was €15,840. According to the NFS report (2003) the income decline resulted from a drop of over 2% in the value of output. Farm production costs were also found to have increased by 3.5%. However, the level of direct payments received by farmers in 2002 increased by 17%, which partially compensated for the decline in returns from the marketplace. The survey also showed an enormous variation in incomes between the larger dynamic full-time farmers and the smaller part-time group, who are highly dependent on direct payments and off-farm employment. This survey however (similar to any other farm dataset in the country for that matter), was unable to analyse the variation in family farm income at anything below the national level as it would not be representative. This paper fills this gap in the literature on Irish farm income by using our Static Farm Level Spatial Microsimulation Model to analyse the distribution of farm earnings at the ED level.

Most farmers' standard of living depends on their own labour and perhaps that of one or more other members in their household, working with them on the family farm. However, in the last decade in particular, off-farm employment has become an increasingly more important source of earnings for farming households. According to the National Farm Survey Report 2002 (Teagasc, 2003), on 35% of farms the main farm operator held an off-farm job compared to 33% in 2001. It is still the case that for the majority of farm families in Ireland, the money earned (and received in the form of farm grants and subsidies) from on-farm employment largely determines how well off its members are, and thus the extent of observed inequality of living standards between farm families.

Therefore, the unit of income that is used in this paper to analyse the distribution of farm earnings is Family Farm Income per farm (FFI). Family Farm Income as defined in the National Farm Survey is calculated by deducting all the farming costs from the value of farming gross output. Family Farm Income represents the financial reward to all members of the family, who work on the farm, for their labour, management and investment. It is important to note however that FFI does not include income from non-farm sources and therefore may not be equal to household income.

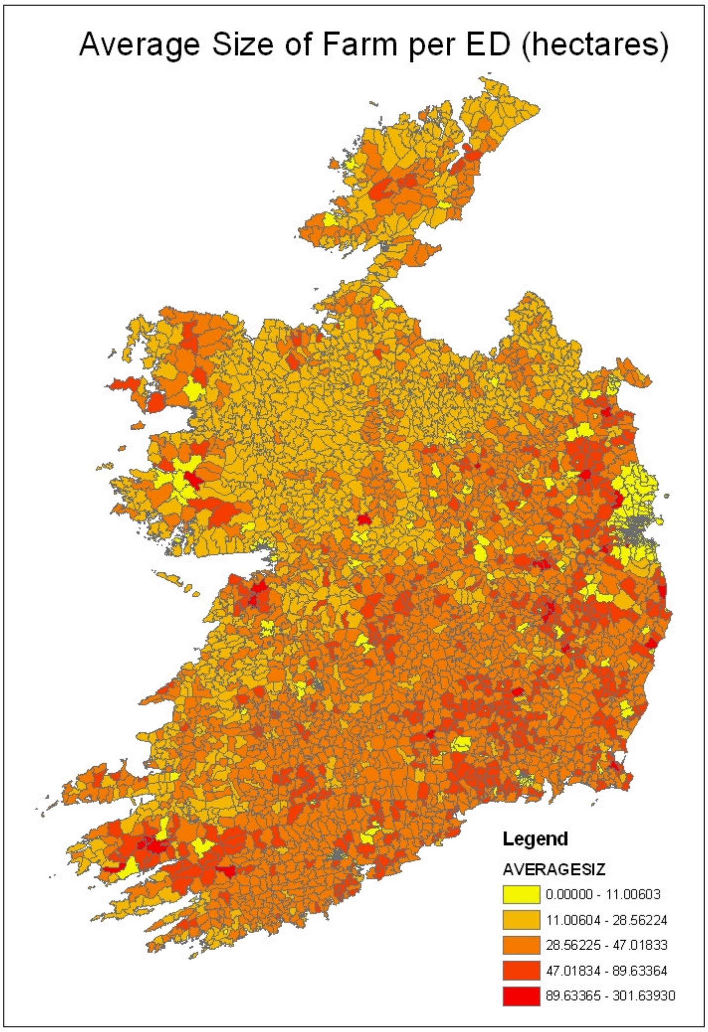
Using the synthetic microdata we were able to produce a spatial analysis of average family farm income across each ED. Figure 2 demonstrates that the majority of family farm incomes is between €12,777 and €35,695. Indeed, according to our Static Farm Level Spatial Microsimulation microdata, in 2002 average family farm income across Ireland was approximately £13,872 while average family farm income by ED was £15,218. It is clear from figure 2 that there are clear regional and local differences in terms of the average income earned on the farm. Although farm earnings has previously been analysed in Ireland (Honohan, 1997) these studies have tended to mask a substantial degree of county and sub-county variation in family farm earnings.

Figure 2. Average Family Farm Income per ED



The results of our Static Farm Level Spatial Microsimulation Model provide clear evidence of the substantial regional variation in family farm income. It is clear that the Border and West region of the country contain the lowest levels of family farm earnings while the provinces of Munster and Leinster in the South and South East of the country enjoy the highest. This however is strongly correlated with the average size of farm holdings in these areas as can be seen from Figure 3.

Figure 3. Average Size if Farm per ED (hectares)



This spatial analysis of family farm income can also be used to validate our Static Farm Level Spatial Microsimulation Model results. The average family farm income estimate from our model can be compared to the weighted NFS average income findings in 2002. According to the NFS weighted average family farm income in 2002 was €14,925 compared to £13,872 according to our Static Farm Level Spatial Microsimulation Model. Thus, as with our comparisons between the simulated microdata estimates and the CSO figures at county and national level, the national estimate from the Static Farm Level Spatial Microsimulation Model is a very close approximation of the actual NFS results at the national level.

5. Discussions and Conclusions

This paper has reviewed the development of a spatial microsimulation farm level model for Ireland. This is the first such static microsimulation model developed for the farming sector. It is envisaged that the models principle contribution will be its ability to analyse policy change in the agricultural sector at a disaggregated spatial level that was not possible previously in Ireland. This is all the more relevant given that the government's new territorial focus of rural development requires modelling economic policy below county level and preferably at the ED level.

With the matched NFS/Census of Agriculture microdata, we will be able to produce spatially disaggregated agricultural data, so that policy-makers can simulate the effect of new policy proposals on farming behaviour down to the ED and individual farm level. For example, the Static Farm Level Spatial Microsimulation Model would allow us to analyse the spatial implications of adhering to the Nitrates Directive for Irish farmers or the spatial implications of further CAP reform or the spatial impact of a new capital tax being placed on land owner. The synthetic microdata can also be used in multivariate analyses where ED location can now be used as an explanatory variable.

Findings from this paper provide useful information for government policy-making purposes. Given the fact that farm income distribution has been shown here, to display high variability across EDs and the very distinctive Northwest/ Southeast divide in terms of average family farm income, government agricultural policy should be aimed at improving the productive capacity of Irish farming in the Northwest of the country rather than simply maintaining the statuesque through direct payment income support. Hynes and O'Donoghue (2002) have shown previously that the low mobility within the farm income distribution means that policies aimed at improving the income of farming, by improving their productive capacity, should have long lasting implications for the level of welfare of Irish farming. Re-orientating agricultural policy in this direction could give rise to a more sustainable basis for the continuation of an economically healthy and productive agricultural sector across Ireland and not just in the Southeast of the country.

Where alleviation of need is the aim, more integration of the farming sector into the Social Welfare system might be worth considering as an option for policy. This should automatically redistribute addition funds to the lower farm income groups in the North West of the country. It needs to be kept in mind however that this area is also associated with smaller farm holdings and the NFS has demonstrated that these types of holdings tend to be associated with a higher level of off-farm employment. Given this fact, overall farm household income (as apposed to on-the-farm family income that was analysed here) may be more evenly distributed across the country than has been shown here. This would mean that further integration of the farming sector into the Social Welfare system might not be necessary. Nevertheless, our analyse demonstrates the main advantage of constructed a spatial microsimulation model, that is the ability to analyse the population (in our case the farm population) across geographical space at a level that was not previously possible due to data constraints.

References

Ballas, D, Clarke GP, Wiemers E (2005), Building a dynamic spatial microsimulation model for Ireland, *Population, Space and Place* (in press)

Ballas, D. and Clarke, G.P. (2001). Modelling the local impacts of national social policies: a spatial microsimulation approach. *Environment and Planning C: Government and Policy*, 19, pp.587-606.

Ballas, D., Clarke, G., and Commins P. (2001) *'Building a dynamic spatial microsimulation model for the Irish rural economy.'* The 12th European colloquium on Theoretical and Quantitative Geography, Rouen, France.

CSO (2000) Census of Agriculture 2000, www.cso.ie*/releasespublications*

Commission Decision (1985), A Community typology for agricultural holdings, *Journal official des Communautes europeennes*, NO. 220/27, P. 0001-0032

Connelly, L and Kinsella, A (2005) Standard Gross Margin (SGM) – Calculation of costs for livestock Grazing in Ireland, Presented on the 15 March 2005 at the Preceding of the Agricultural Research Forum

Dowsland K., (1993), Simulated Annealing. In Reeves C (ed) *Modern Heuristic Techniques for Combinatorial Problems*, Blackwell, Oxford.

FADN (2005), Concept of FADN http://europa.eu.int/comm/agriculture/rica/concept_en.cfm

Fitzpatrick Associates (1997): Operational Programme for Agriculture, Rural Development and Forestry 1994-1999, Mid-Term Review Report. Report to the Community Support Framework Evaluation Unit.

Honohan, P. (1997): EU Structural Funds in Ireland - A Mid-Term Evaluation of the CSF 1994-99, ESRI, PRS No. 31, Dublin.

Holm, E., Lindgren, U., Makila, K., Malmberg, G. (1996). Simulating an entire nation. In G.P. Clarke (Ed.), *Microsimulation for Urban and Regional Policy Analysis*. Pion, London.

Hynes, S. and O'Donoghue, C. (2004), Irish Family Farm Income Mobility and Inequality 1994-2001, National University of Ireland Working paper no. 78.

Keeney, M., A. Matthews and J. Frawley (1997), The Distribution of Direct Payments in Irish Agriculture", paper read to the Economic Policy conference, Dublin Economic Workshop, at Kenmare, October 17-19.

Kelly, D. (2004), SMILE Static Simulator Software User Manual Teagasc, Teagasc Athenry Publication, October 2004

Kirkpatrick, S., Gelatt, C. and Vecchi, M., (1983), Optimization by simulated annealing, *Science* 220 671-680

Matthews, A., 2000. *Farm Incomes: Myths and Reality*, Cork University Press, Cork.

Melhuish, T, Blake, M, and Daly, S. (2002), An Evaluation of Synthetic Household Populations for Census Collection Districts Created Using Spatial Microsimulation Techniques 26th Australia & New Zealand Regional Science Association International Annual Conference, Gold Coast, Queensland, Australia

Mitton, L, Sutherland H, and Weeks M. (2000) *Microsimulation Modelling for Policy Analysis: Challenges and Innovation* Cambridge university Press, P.1-23

Metropolis N, Rosenbluth A W, Rosenbluth M N, Teller A H, Teller E, (1953), Equation of state calculations by fast computing machines. *Journal of Chemical Physics* 21: 1087-1092

NFS, (2002) National Farm Survey 2001, Teagasc Publication, Athenry, Galway

NFS, (2003) National Farm Survey 2002, Teagasc Publication, Athenry, Galway

Nolan, B. (2005), Ireland's Income Distribution in Comparative Perspective, *Review of Income and Wealth*, Series 51, No 4

O'Donoghue C. (2001), Dynamic Microsimulation: A Survey, *Brazilian Electronic Journal of Economics*.

Openshaw S, Rao L (1995) Algorithms for reengineering 1991 Census geography. *Environment and Planning* 27: 425-446

Openshaw S, Schmidt J, (1996) Parallel simulated annealing and genetic algorithms for re-engineering zoning systems, *Geographical Systems* 3: 201-220

Pham D T, Karaboga D (2000), Intelligent Optimisation Techniques: Genetic Algorithms, Tabu Search, Simulated Annealing and Neural Networks, Springer, LondonNorman, P. (1999). Putting Iterative Proportional Fitting (IPF) on the researcher's desk. Working Paper 99/03. School of Geography, University of Leeds, Leeds.

Projet de Decision de la Commission (1992) *Typologie des Exploitations Agricoles* 37eme Reunion du Groupe d'Experts.

Williamson, P., Clarke, G.P., McDonald, A.T. (1996). Estimating small area demands for water with the use of microsimulation. In G.P. Clarke (Ed.) *Microsimulation for Urban and Regional Policy Analysis*. Pion, London.

White Paper on Rural Development. (2000). Department of Agriculture, Dublin, Ireland.

Wiemers, E., D. Ballas and G. Clarke, (2002) A Spatial Microsimulation Model for Rural Ireland: Evidence from the 2002 Irish Census of Population, Teagasc working paper series

Wong, D.W.S. (1992). The Reliability of Using Iterative Proportional Fitting Procedure. Professional Geographer. Vol. 44, 349-348.