Projecting population and labour market trends in rural areas

Project No: 4847

Authors: Teagasc—Emily Wiemers, Patrick Commins, Eamonn Pitts
University of Leeds—Dimitris Ballas, Graham Clarke

Summary

This purpose of this project is to develop a spatial model to project population and labour market variables at the small area level in Ireland. The model is called SMILE (Simulation Model for the Irish Local Economy) and is a static and dynamic spatial microsimulation model. Microsimulation attempts to describe economic and social events by modelling the behaviour of individual agents such as persons or firms. Microsimulation models have proved useful in evaluating the impact of policy changes at the micro level. Spatial microsimulation models contain information on geographic units and allow for a regional or local approach to policy analysis. SMILE is based on modelling work on urban systems and employs similar techniques for analysing rural areas.

The static model creates a spatially referenced synthetic population of Ireland. Each individual enumerated in the 1991 Census of Population is synthetically constructed and is assigned 11 census characteristics including a District Electoral Division (DED) location. The dynamic element incorporated in SMILE ages the synthetic population by modelling demographic processes including fertility, mortality and internal migration. The dynamic process is used to project population in the medium term; it ages the synthetic 1991 population to 1996. For validation purposes, these 1996 projections are then compared to the 1996 Census of Population. The same process was used to project between the 1996 and the 2002 Census of Population. The results indicate that the accuracy at DED and county level is within acceptable limits.

The model will be extended in the next three years, beginning in 2003, with additions including validating individual attributes such as employment status and social class and also including households in the model. This project has created a basic model that can be expanded and developed in the future.

1. Introduction

The primary objective of this project was to develop a spatial, small area population and labour force model. The first step was to identify a suitable methodology. Input/output models are the most common tools used to examine regional economies. However, because this project was concerned with analysing policy change for geographic areas smaller than counties, input/output techniques were not appropriate. Spatial microsimulation is an alternative to input/output methodologies for examining regional economies. The School of Geography at the University of Leeds used spatial microsimulation to examine the implications of a closure of a large factory in the city of Leeds at ward level (Ballas and Clarke, 1999, 2001). The type of analysis carried out in Leeds fit the aims of the project and a programme of collaboration between the Rural Economy Research Centre and the University of Leeds was established to develop a small area population model capable of analysing the differential impacts of policy change in rural areas. Teagasc and the School of Geography at the University of Leeds developed the Simulation Model for the Irish Local Economy (SMILE). SMILE is a microsimulation model that creates a synthetic, spatially referenced population for Ireland. This project focused on developing and validating the small area population model. Outputs are listed at the end of this report.

The model will be extended to analyse the relationships among regions and localities and to project the spatial implications of economic development and policy change in rural areas. The model has been designed to analyse both sectoral and spatial strategies for rural development. The sectoral policies that will be examined include CAP reform and policies to promote economic development. Spatial policies include the National Spatial Strategy and “rural proofing” as outlined by the White Paper on Rural Development.

2. Methodology

SMILE is a static and dynamic, spatial microsimulation model for projecting population variables. It builds a synthetic, spatially referenced population for Ireland using the 1996 Census of Population Small Area Population Statistics (SAPS). It contains two processes. First, the static process creates the ‘base population’ and assigns census attributes to individuals. Second, the dynamic process ages the population by evaluating individuals for fertility, mortality and migration.

Static Process

The static spatial microsimulation procedure constructs a micro-level population at the small area level. The first step of the static process uses a static spatial microsimulation approach based on Iterative Proportional Fitting (IPF) to create small area microdata. SMILE applies IPF to data from the Irish Census SAPS to compute conditional probabilities.

SMILE creates the base population using two SAPS tables. One contains the number of individuals in each District Electoral Division (DED) by five-year age group, sex and marital status, and the other contains the number of individuals in each DED by sex, marital status and employment status. IPF is used to combine these tables and to estimate the probability of an individual (*i*) in each DED (*x*) having a given age (*a*), sex (*s*), marital status (*m*) and employment status (*e*). That is, we want to estimate $p_i^x(a,s,m,e)$ given the known probabilities $p_i^x(a,s,m)$ and $p_i^x(s,m,e)$. Once the joint probability is estimated using the IPF procedure, Monte Carlo sampling is used to assign age, sex, marital status and employment status attributes to each individual in a DED using a process similar to that described in Table 1. The procedure results in a synthetic population for each DED in Ireland where the simulated individuals have age, sex, marital status and employment status attributes.

The second step in the static SMILE model is to estimate further attributes. Among these are the official census variables of occupation, industry and social class. These attributes are estimated using known conditional probabilities. For example, occupation is reported by sex and employment status for each DED. Therefore, for each sex and employment status combination there exists a set of probabilities for occupation. Individuals are assigned an occupation by using Monte Carlo sampling on the set of probabilities for their sex, employment status and DED location. Table 1 outlines the process for assigning occupation.

The first synthetic household in Table 1 has the following characteristics: male, aged 25, single, at work and living in the first DED of County Leitrim. As shown in Table 1, the estimated probability that an individual with these characteristics would have an occupation of a farmer with less than 30 acres is 60 percent. The next step in the procedure is to generate a random number to see if the synthetic individual is estimated to be a farmer with less than 30 acres. The random number in this example is 0.4 and falls within the 0.001 - 0.600 range needed to qualify as a farmer with less than 30 acres.

Steps	**1st**	**2nd**	**…**	**Last**
Age, sex and marital status, employment status and location (DED level) (given)	Age: 25 Sex: Male Marital Status: Single Employment Status: At work GeoCode: Leitrim Co., DED 001 Ballinamore	Age: 76 Sex: Female Marital Status: Married Employment Status: Other (e.g. Retired) GeoCode: Leitrim Co., DED 002 Cloverhill	…	Age: 30 Sex: Male Marital Status: Married Employment Status: At work GeoCode: Leitrim Co., DED 078 Rowan
Probability (conditional on sex and employment status) of individual being an farmer with less than 30 acres	0.6		…	0.2
Random number	0.4		…	0.6
Occupation category assigned on the basis of random sampling	Farmer with less than 30 acres		…	Not a farmer with less than 30 acres

Table 1. A simple example of the microsimulation procedure for the allocation of occupation

The same procedure is followed to assign occupation and other attributes to all of the synthetic households included in the spatial microsimulation database.

Dynamic Process

The dynamic process ages the synthetic micro-level population estimated in the static process. Each individual created in the static model is assessed for mortality, fertility and migration. Mortality and fertility rates are given by the 1991 Report on Vital Statistics. Mortality is reported by age and sex while fertility is reported by age and marital status. Migration is calculated using statistics from the 1996 Census of Population. It is reported by age, sex and location.

This model assumes that mortality is a function of age, sex and county location and that fertility is a function of age, marital status and county location. Counties were clustered to mitigate problems of sample size. For example, there were only two married women under age 20 who gave birth in County Leitrim in 1991. Grouping counties together helps to solve problems of small numbers while preserving spatial differentiation.

Each individual in the database is evaluated every year in the simulation period for mortality and fertility on the basis of random sampling from the respective probabilities. The process is similar to that outlined in Table 1. If an individual is deemed to die he/she is removed from the database. If an individual gives birth, the model creates a new individual. The new individual's attributes are set as follows: age is zero, sex is determined probabilistically (50 percent probability of each sex), marital status is single, social class and location are that of the mother and all other attributes are left blank. In the next simulation period, the new individual is simulated along with the other individuals in the location.

Internal migration is modelled on the basis of random sampling from calculated migration probabilities derived from 1996 Census of Population data at county level. Probabilities of migrating from one county to another are calculated by age, sex and county location. Every individual in the database is assessed for migration using Monte Carlo sampling. The individuals that are assigned migrant status are allocated to a DED within the new county on the basis of its population size. Areas with the biggest populations have the highest probability of attracting migrants. This assumption is a simplification of the migration process but is necessary due to lack of data.

Immigration is not modelled because of the lack of data. EU and non-EU immigrants would need to be included if immigration were to be modelled. EU immigrants are difficult to capture because there is freedom of movement among citizens of EU countries and because Ireland does not have an identity card system where individuals are required to have a state-issued ID. Non-EU immigrants may be easier to capture because they must register each year with the local Gardai. Until recently, these records were not computerised but in the future they may provide a method to account for non-EU immigrants at the small area level. However, even if the total number of immigrants could be found, accurately estimating attributes for these individuals and projecting the likely immigration patterns at the small area level in the future would be difficult.

3. Results

The SMILE model results can be divided into static and dynamic categories. The results presented here are only examples of the type of results generated by the model.

Static Results

The 1991 static model is used to create a synthetic population for 1991. The individuals have 11 census characteristics. The results from the model are linked to a Geographic Information System (GIS) to map the spatial distribution of individuals. The following figures illustrate the type of outputs generated by the model. Figure 1 shows the distribution of married, unemployed males in each DED according to the model. This distribution is an average of five runs of the model. The five-run average reduces the observed error. In Figure 1, data are grouped in quintiles.

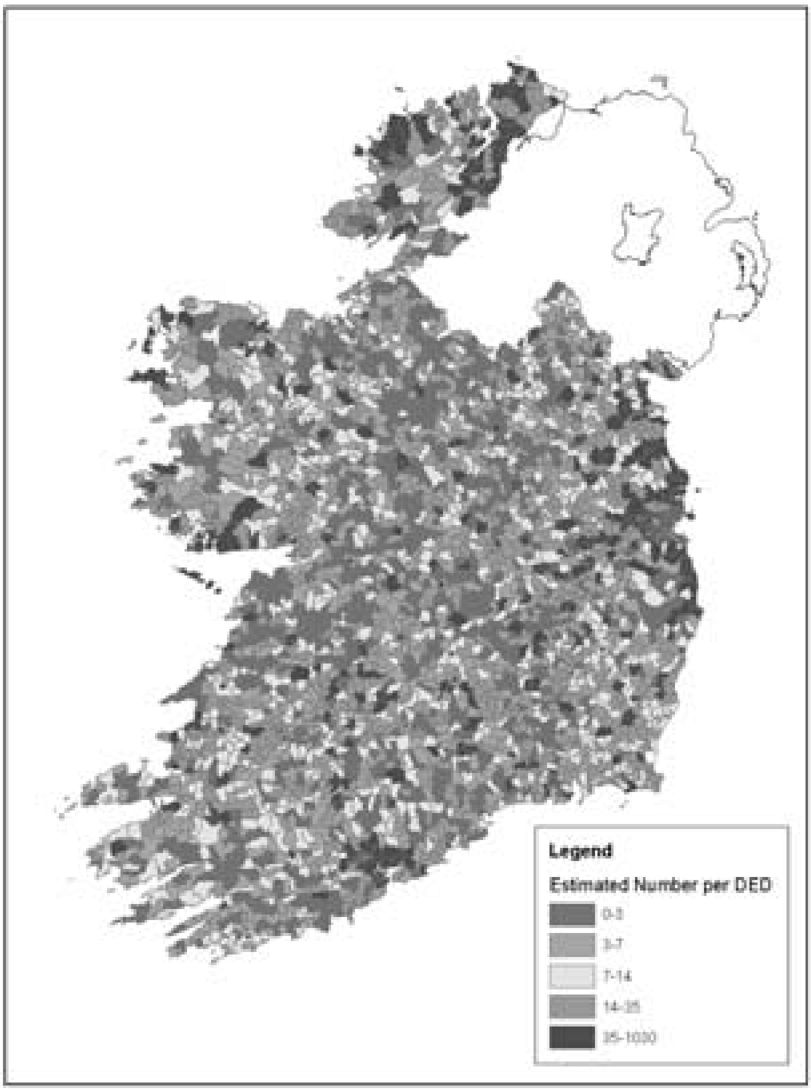

Figure 1. Estimated number of unemployed, married men in each DED, 1991.
Source: SMILE 1991 static model.

Figure 2 shows the actual distribution of these individuals in quintiles from the 1991 Census of Population. Figure 3 maps the absolute percent error between the 1991 Census of Population and the SMILE estimates in quintiles.

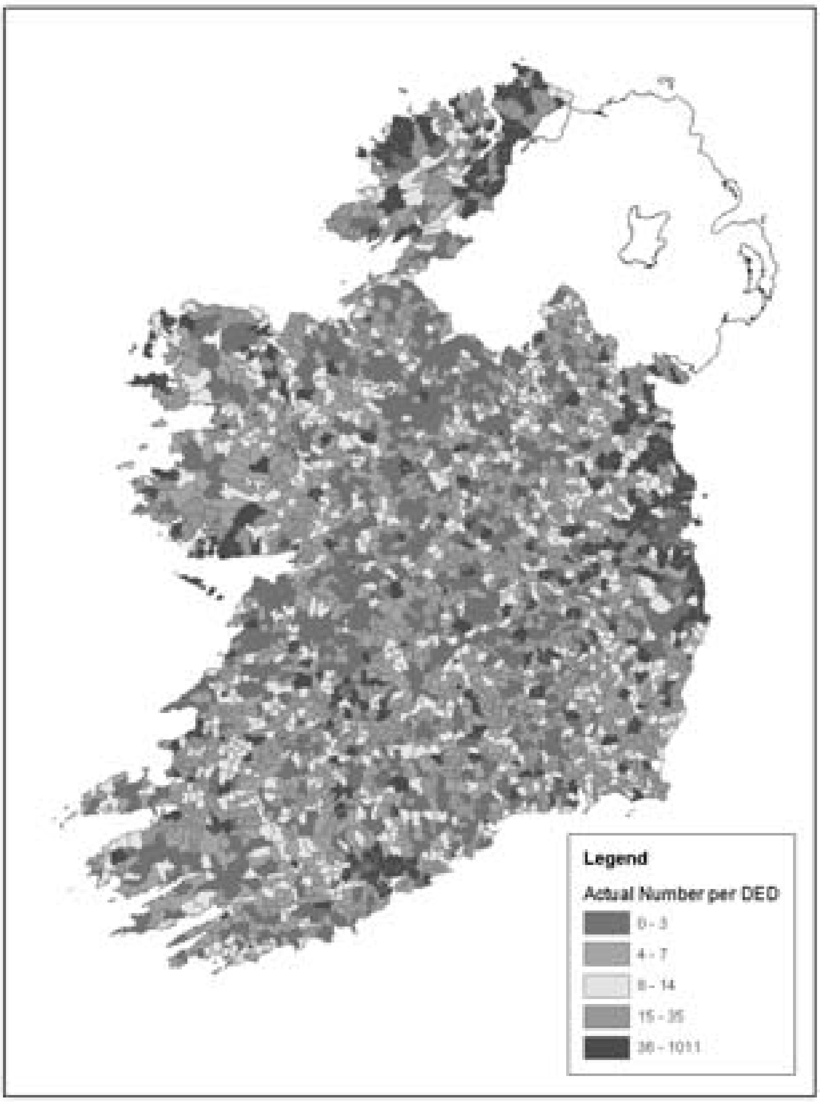

Figure 2. Actual number of unemployed, married men in each DED, 1991.
Source: 1991 Census of Population.

The results of the static model show that 80 percent of DEDs fall below 20 percent absolute error. Most DEDs with over 20 percent absolute error are those with very small numbers of unemployed, married men. An error of 100 percent is generally observed when the model estimates two unemployed, married men in a DED that only contains one according to the census.

Figure 3. Absolute percent error of unemployed, married men in each DED, 1991.
Source: SMILE 1991 static model and 1991 Census of Population.

The static model has the capability to produce the estimated spatial distribution of individuals with any combination of census characteristics. The example was chosen because the actual figures are available from the census data for validation. The static model also uses 1996 Census of Population to produce a static synthetic population for 1996. The results are similar to that from 1991.

Dynamic Results

The dynamic model is used to age the 1991 (1996) static population to 1996 (2002). The results are linked to a GIS for analysis. Two measures were used to validate the model. First, the total estimated population at DED level in 1996 was compared to the actual population at DED level. Figure 4 shows the absolute percent error of total population at DED level in quintiles. The results show that 80 percent of DEDs fall below 10 percent error. As many DEDs have an absolute error under two percent as have an absolute error over 10 percent. Only 128 DEDs out of approximately 3400 contain an absolute error over 20 percent. These estimates are within acceptable limits. Population at DED level is very difficult to project because changes in planning policies and the building of housing estates within towns can radically change the population of DEDs.

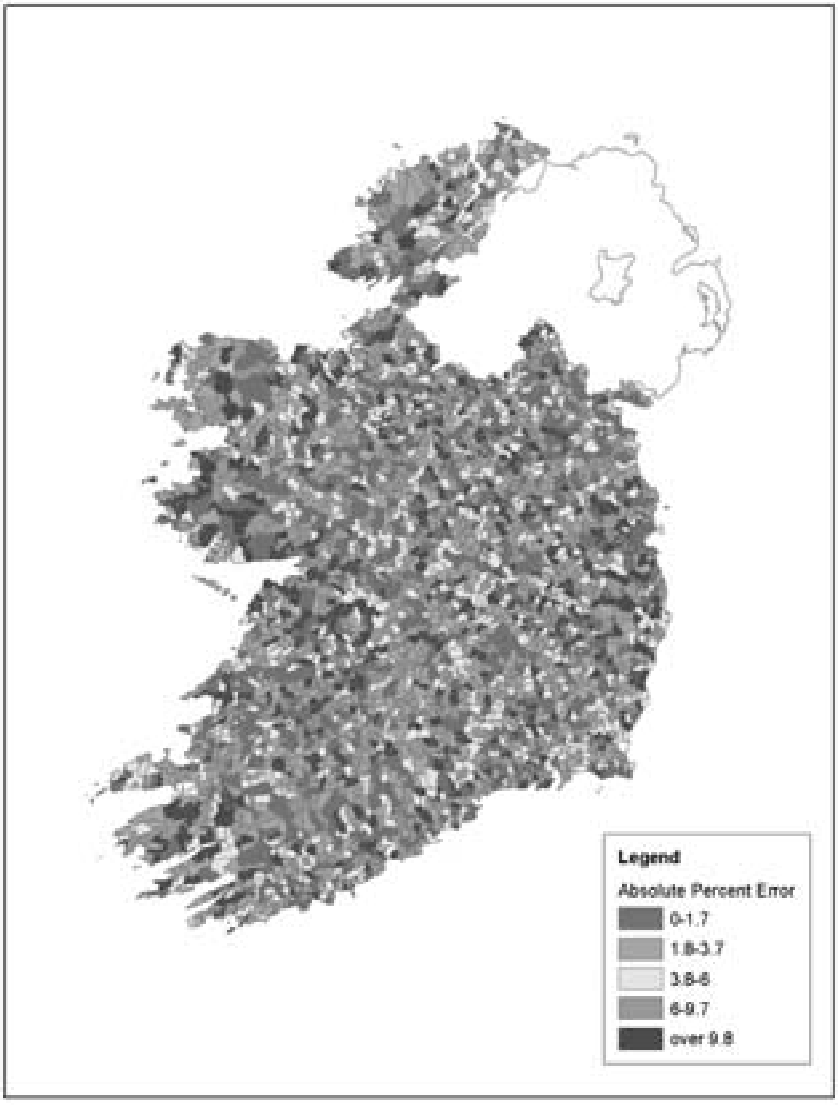

Figure 4. Absolute percent error of total population by DED, 1996.
Source: SMILE 1991 dynamic model and 1996 Census of Population.

The second measure used to validate the dynamic model was county level population estimates. These are easier to project. Table 2 shows the projected population at county level for 2002 taken from the 1996 dynamic model, the actual county population taken from the preliminary results of the 2002 Census of Population, and the error at county level of the 2002 county population projections.

The highest rate of error is observed in County Westmeath where the population is underestimated by 11.32 percent. The rate of error in 23 of 27 counties is below ten percent and in 12 counties it is below five percent. The mean absolute percent error of all counties is 6.22 percent.

Province or County	Without Net Immigration				
	2002 Actual	2002 Estimate	Estimate – Actual	Relative Error	Absolute Error
State	3,917,336	3,779,568	-137,768	-3.52%	3.52%
Leinster	2,105,449	2,048,005	-57,444	-2.73%	2.73%
Carlow	45,845	44,412	-1,433	-3.13%	3.13%
Dublin	1,122,600	1,137,995	15,395	1.37%	1.37%
Kildare	163,995	154,617	-9,378	-5.72%	5.72%
Kilkenny	80,421	76,257	-4,164	-5.18%	5.18%
Laois	58,732	53,269	-5,463	-9.30%	9.30%
Longford	31,127	29,683	-1,444	-4.64%	4.64%
Louth	101,802	96,873	-4,929	-4.84%	4.84%
Meath	133,936	119,903	-14,033	-10.48%	10.48%
Offaly	63,702	57,964	-5,738	-9.01%	9.01%
Westmeath	72,027	63,872	-8,155	-11.32%	11.32%
Wexford	116,543	106,080	-10,463	-8.98%	8.98%
Wicklow	114,719	107,080	-7,639	-6.66%	6.66%
Munster	1,101,266	1,061,281	-39,985	-3.63%	3.63%
Clare	103,333	91,774	-11,559	-11.19%	11.19%
Cork	448,181	432,541	-15,640	-3.49%	3.49%
Kerry	132,424	123,506	-8,918	-6.73%	6.73%
Limerick	175,529	180,395	4,866	2.77%	2.77%
Tipperary North	61,068	54,546	-6,522	-10.68%	10.68%
Tipperary South	79,213	77,933	-1,280	-1.62%	1.62%
Waterford	101,518	100,586	-932	-0.92%	0.92%
Connacht	464,050	435,075	-28,975	-6.24%	6.24%
Galway	208,826	197,765	-11,061	-5.30%	5.30%
Leitrim	25,815	23,177	-2,638	-10.22%	10.22%
Mayo	117,428	107,780	-9,648	-8.22%	8.22%
Roscommon	53,803	49,597	-4,206	-7.82%	7.82%
Sligo	58,178	56,756	-1,422	-2.44%	2.44%
Ulster (part of)	246,571	235,207	-11,364	-4.61%	4.61%
Cavan	56,416	50,969	-5,447	-9.66%	9.66%
Donegal	137,383	133,149	-4,234	-3.08%	3.08%
Monaghan	52,772	51,089	-1,683	-3.19%	3.19%
MAPE					6.22%

Table 2. Actual and projected population by county, 2002.
Source: SMILE 1996 dynamic model and Preliminary Report of 2002 Census of Population.

Figure 5 maps the error in quintiles. Dublin and Limerick are the only counties in which the projected population is greater than the actual population in 2002. Dublin and Limerick may have increased relatively slowly compared to 1996 as negative agglomeration factors such as congestion and high house prices caused people to move to the suburbs and to other counties. Considering the rapid pace of change in Ireland since 1996, the concentration of employment in Dublin and Shannon regions, and the large increase in net immigration, these results are satisfactory.

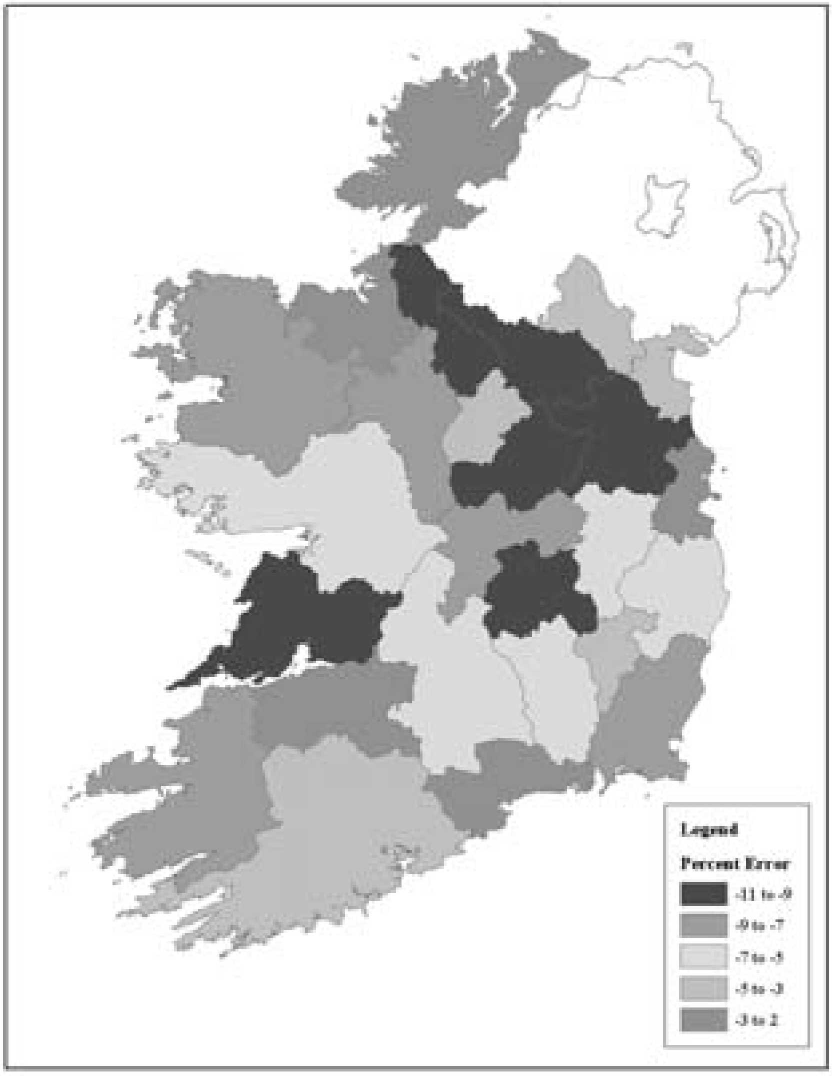

Figure 5. Percent error of total population by county, 2002.
Source: SMILE 1996 dynamic model and 2002 Census of Population preliminary results.

The 1996 dynamic results reported in Table 2 do not include immigration or emigration. However, immigration contributed significantly to population change between 1996 and 2002. Ireland experienced net immigration of 153,067 during the

1996 to 2002 census period. It is the first census period in which average annual net immigration exceeded average annual natural increase, and one of only three census periods where net migration was positive (Preliminary Census Report, 2002). The counties with the largest number of immigrants are Cork, Galway and Dublin along with its surrounding counties. The border counties of Monaghan and Longford received the fewest number of immigrants but every county had positive net immigration. The annual rates of net migration per thousand were greatest in Meath, Kildare, Westmeath and Laois. These counties are part of the wider Dublin commuter belt. If the 1996 dynamic results are adjusted to account for net immigration, only six counties have an absolute error above five percent and the mean absolute percent error for all counties is 3.2 percent.

4. Summary and Implications of Key Findings

The application of spatial microsimulation to rural development analysis is new and still undergoing development. However, the territorial focus of rural development strategy and the government's pledge to "rural proof" all new policy initiatives requires modelling economic policy below county level.

SMILE has been tested against the 1991, 1996 and 2002 Census of Population results with acceptable levels of error. The model already has the capability to examine policies that affect variables currently included in the model. For example, if the IDA were advising a computer chip manufacturer on where to establish a new factory, the agency could use SMILE to identify areas that have high levels of unemployment among skilled manual workers.

The expansion of the SMILE model over the next three years to include labour market transitions, commuting data and agricultural characteristics promises to provide a powerful tool for policy analysis at the small area level. The new model will provide the ability to project labour force characteristics for the medium term below county level, and to apply macro agricultural sectoral model projections to the small area level. The policies that could be analysed using the expanded SMILE model are extensive. The model could be used to identify structurally weak areas, such as those highly dependent on small family farms or on manual labour industries. Areas with high levels of social exclusion and spatial isolation that are likely to be differentially favoured behind by the National Spatial Strategy and other rural development schemes can also be identified. The ability to project and understand the spatial and sectoral impact of policy change is becoming increasingly important. Over the next three years, our goal is to develop SMILE to meet this challenge.

References

Ballas, D. and Clarke, G.P. (1999). *Regional versus Local Multipliers of Economic Change? A microsimulation approach*. Paper presented at the 39th European Regional Science Congress. Dublin, Ireland. August 23-27.

Ballas, D. and Clarke, G.P. (2001). Towards local implications of major job transformations in the city: a spatial microsimulation approach. *Geographical Analysis*, 33, pp.291-311.

Preliminary Census Report. (2002). Central Statistics Office, Dublin, Ireland.

Outputs

Ballas, D., Clarke, G.P. and Commins, P. (2001a). *Spatial microsimulation for rural policy analysis*. Paper presented at the 41st European Regional Science Association (ERSA) congress. Zagreb, Croatia. August 29 – September 1.

Ballas, D., Clarke, G.P. and Commins, P. (2001b). *Building a dynamic spatial microsimulation for the Irish rural economy.* Paper presented at the 12th European colloquium on Theoretical and Quantitative Geography. Rouen, France. September 7-11.

Pitts, E., Ballas, D. (2002). *Projecting population and labour force in rural areas*. Paper presented at the Teagasc Rural Development Conference. Tullamore, Ireland. February 12.

Wiemers, E. (2002). *Using spatial microsimulation to model farming system and income in Ireland*. Paper presented at the 32nd Regional Science Association International Conference, British and Irish Section. Brighton, UK. August 20-23.

Wiemers, E. (2002). *Using spatial microsimulation to model farming system and off-farm employment in Ireland*. Paper presented at the 10th European Association of Agricultural Economists Congress. Zaragoza, Spain. August 28-31.