Modeling the Spatial Distributional Agricultural Incomes

Cathal O'Donoghue*, Eoin Grealis *, Niall Farrell****Teagasc Rural Economy and Development Programme, Athenry, Co. Galway** SEMRU, National University of Ireland Galway*** Economic and Social Research Institute, Dublin**

Paper prepared for presentation at the 150th EAAE Seminar

"The spatial dimension in analysing the linkages between agriculture, rural development and the environment"

Jointly Organised between Scotland's Rural College (SRUC) and Teagasc

Scotland's Rural College, Edinburgh, Scotland
October 22-23, 2015

Modelling the Spatial Distributional Agricultural Incomes

Cathal O'Donoghue[1]*, Eoin Grealis** *, Niall Farrell***

*Teagasc Rural Economy and Development Programme, Athenry, Co. Galway

** SEMRU, National University of Ireland Galway

*** Economic and Social Research Institute, Dublin

Abstract.

Agricultural incomes are quite heterogeneous relying as they are in part on the environmental context which land is farmed. In addition a very significant proportion of agricultural income results from public policy via the Farm Direct Payments within in the Common Agricultural Policy. In this paper we develop and test a methodology to spatially model the distribution of Agricultural Activity and associated income across place utilising a spatial microsimulation model. In particular we build upon a quota sampling method used in the development household based spatial microsimulation models to account for spatial heterogeneity in relation to stocking rate. We utilise this framework to model the spatial distribution of activity, incomes and viability across Ireland.

[1] Corresponding Author. Email Cathal.ODonoghue@teagasc.ie. The authors are grateful for funding provided by the Irish Local Development Network and comments provided at the International Microsimulation Association World Congress and at the Teagasc-AESI CAP reform workshop.

Modelling the Spatial Distributional Effect of Common Agricultural Policy Reform

1. Introduction

There is significant spatial heterogeneity in agriculture in Ireland (See Crowley et al., 2008), with by and large the better land in the South and East and the poorer land in the North and West (See Frawley and Commins, 1996). The most profitable sub-sectors within agriculture, dairy and to some extent tillage farming, are predominantly concentrated in the South and East. The lower margin beef and sheep sectors are to a large extent located in the Midlands, North and West of the country. It is important to understand this spatial heterogeneity so as to be able to better target policy interventions. In particular the spatial distribution of Agricultural income and the consequential impact of policy reform such as CAP reform are important in targeting for example Agricultural extension resources or the development of localised rural development interventions.

The challenge in understanding the spatial distribution of farm incomes is one of data. While it may be possible to simulate the spatial pattern of farm direct payments using administrative data as in the case of Bergmann et al. (2011) at a spatial scale in Scotland and Donnellan et al (2013) at an aspatial scale in Ireland, these datasets often lack contextual information, limiting the depth of analysis possible. Typically Censuses of Agriculture and Administrative data provide spatial information on the structure of agriculture, but have no income or farm structure data. On the other hand Farm Accountancy Data Network (FADN) type data contain excellent farm income and structural data, but have weak spatial dimensions. Data imputation/enhancements methods known as spatial microsimulation (O'Donoghue et al., 2014; Hermes and Poulsen, 2012) however have been developed for to combine the strengths of both types of data.

In terms of agricultural income, Hynes et al. (2009b) developed a model of spatial farm incomes, which has been used to examine the impact of EU Common Agricultural Policy Changes (Shresatha et al., 2007). This methodology has also been applied in other countries such as Van Leeuwen et al. (2008) in the Netherlands. These methods involve resampling or reweighting farm survey data to be consistent with spatial calibration totals. While this method performs satisfactorily for calibration totals, it performs less satisfactorily for variables that cannot be included as a constraint total for example stocking rates. In this paper we improve this methodology to in particular produce better spatial stocking rate estimates. We will also test the sensitivity of results to specification choices. We will then model the spatial distribution of agricultural incomes. This paper is structured as follows. Section 3 describes the methodological framework for modelling the spatial distribution of income. In section 4 we discuss the preparation of the data and some summary statistics. Subsequently we report results in 3 sections in relation to the sensitivity results to alternative modelling assumptions, the spatial distribution of Agriculture.

2. Policy Context

3. Methodology

Given a lack of spatially disaggregated farm survey data, the objective of the methodological exercise in this paper is to create a synthetic spatial farm dataset, combining the best of both farm level survey data and spatially disaggregated Census of Agriculture data.

Small area statistical analysis can be used for this purpose (See Ghosh, 1994). However for our purposes, we are interested not only in inter-spatial variation in incomes but also intra spatial area variation of incomes. Therefore we require a method that maintains both spatial variability and micro-level variability.

Spatial microsimulation (Clarke, 1996) is a potential methodology achieving both of these dimensions within its data enhancement process. There is an extensive literature described in O'Donoghue et al (forthcoming) covering many different policy areas, utilising various methodologies described in Hermes and Poulsen, 2012.

The methodology has been applied in a number of instances within agriculture and rural development. Ballas et al., (2006) utilised iterative proportional fitting to examine CAP reform as part of the Luxembourg agreement. Hynes et al. (2009b) developed a model of spatial farm incomes utilising simulated annealing, which has been used to examine the impact of EU Common Agricultural Policy Changes (Hennessy et al., 2007). This forms part of the Simulation Model of the Irish Local Economy (SMILE) (O'Donoghue et al., 2013). O'Donoghue (2013) extended the farm focused models to include wider household income sources to be able to assess the wider economic sustainability of farm households. Clancy et al. (2013) utilised the model in Ireland to assess the optimal spatial location for the growth of willow and miscanthus for biomass production. Lindgren and Elmquist (2005) linked natural sciences and economics in their Systems AnaLysis for Sustainable Agricultural production (SALSA) model to evaluate the economic and environmental impact of alternative farm management practices on a site specific arable farm in Sweden.

A variant of the agricultural dimension of SMILE (Hynes et al.,2009b), focuses on recreational activity in forests within a single city (Cullinan et al., 2008). Also with a small area focus (a number of municipalities), van Leeuwen et al. (2008) have developed a model exploring the linkages between on and off-farm employment, which is becoming an increasing part of farmer's incomes in the EU. While there have been many examples of aspatial static microsimulation models that have simulated greenhouse gas emissions, the spatial models that have modelled these emissions tend to be those where spatial context is relevant such as agricultural models (Hynes et al., 2009a), land use (Moeckel et al., 2007) or transportation issues (Mavoa, 2007). In terms of environmental and biodiversity related issues, microsimulation models were used to look at a range of issues including wildlife-recreation interaction (Bennett et al., 2009) and the non market value of wild bird conservation (Hynes et al., 2010), landscape services from Agriculture (Pfeiffer et al., 2012) and participation in Rural Environmental Protection Schemes, (Hynes et al., 2008).

In order to undertake a spatial impact analysis of the CAP reform, we need to statistically combine farm level survey data (NFS) with spatial Census of Agriculture data. The most recent Census of Agriculture was collected in 2010 and released for research purposes in 2013 (CSO 2010). We would like to combine this with the 2010 Teagasc National Farm Survey (2008).

O'Donoghue et al. (forthcoming) and Hermes and Poulsen (2012) describes a number of potential methodologies to do this. Potential options include

- Iterative Proportional Fitting
- Deterministic Reweighting
- Combinatorial Optimisation
- Quota Sampling

The deterministic approach to reweighting national sample survey data is an attempt to fit small area statistics tables or benchmarks for each small area without the use of random sampling procedures (Ballas et al., 2005). Iterative Proportional Fitting (Deming and Stephen, 1940) may be used to generate cross-tabulated control totals at the small area. These are compared with similar cross-tabulated totals from the survey data to produce weights. The method typically produces weights where the control totals and the survey data have the same unit of analysis. This allows any number of changes to the data in the model to be made until an optimal reweighting methodology is reached (Smith et al., 2007; 2009).

An alternative mechanism for generating weights for generating spatial micro data is to use a regression based reweighting method. An example is GREGWT, which is a generalised regression reweighting algorithm written by the Australian Bureau of Statistics (ABS) which was developed to reweight their survey data to constraints from other Australian data sources (see Tanton et al, 2011). GREGWT is a constrained distance minimisation function which uses a generalised regression technique to get an initial weight and iterates the regression until an optimal set of household or individual weights for each small area is derived. GREGWT is also a deterministic, in that it generates the same result each time it is run. Optimisation is achieved when the difference between the estimated count and the known census count for each of the constraint variables is minimised or a predefined number of iterations is made at which stage the iteration stops. Once the reweighting process is finished, each household in the survey dataset should have a weight for each census small area that had counts for the constraint variables used. The method pose some problems for areas with small sizes.

The final approach to generating spatially disaggregated microdata is the use of combinatorial optimisation methods which can be used to reweight an existing microdata sample to fit small area population statistics. For example, aspatial microdata sets can be reweighted to estimate the micro population at a local spatial scale (Williamson et al., 1998, Ballas and Clarke, 2000). The method differs from IPF primarily in that it reweights or samples from a micro dataset until a new micro-dataset is generated that reflects the characteristics of the small area. In a geographical context, this method has been applied to examine a number of policy areas, including the SMILE agri-environmental model (Hynes et al., 2009).

Quota Sampling

In determining the methodology to use for the creation of a farm level spatial microsimulation model, we faced a number of issues. While IPF could potentially be used to produce small area weights, it struggles to deal with the issue of heterogeneous stocking rates. As the survey has a greater sample size than the cell size for most districts, resulting in weights of less than 1, it is likely that this approach will smooth the heterogeneity of farm incomes. Similarly given how many districts have small numbers of farms in Ireland, the GREGWT method is potentially challenging and may smooth incomes. Simulated Annealing was used to generate an earlier version of the model (Hynes et al, 2009) but has significant computational costs and also struggles with the heterogeneous stocking rate issue.

Thus we were motivated to develop a methodology that was sample based to avoid the income smoothing concern of the weighting methodology, was computationally efficient and could be adjusted to improve the spatial heterogeneity of stocking rates.

We have thus developed in parallel with Farrell et al. (2013) a method known as Quota Sampling (QS) which is a probabilistic reweighting methodology developed which operates in a similar fashion to Simulated Annealing (SA) (Wu and Wang, 1998), whereby survey data are reweighted according to key constraining totals for each small area, with amendments made in the sampling procedure in order to improve computational efficiency. We call the resulting model SMILE-FARM[2]. The basic sampling procedure, and its implementation in the overall simulation process, is now outlined.

Similar to SA, quota sampling selects observations at random and considers whether they are suitable for selection for a given small area based on conformance with aggregate totals for each small area characteristic. Unlike SA, Quota Sampling only assigns units (in this case farms) that conform to aggregate constraint totals and once a unit is deemed selected, it is not replaced; the main computational improvement.

To accommodate this, small area aggregate totals for each constraining variable are required as the initial values for what we term 'quotas', or running totals for each constrained variable, which are recalculated once a unit is admitted to a small area population. The method randomly sorts the population of farms and allocates one unit at a time, in the presence of a number of constraints. If the unit sum of each constraining characteristic (e.g. a Dairy Specialist Farm) is less than or equal to each small area total (e.g. 10 Dairy Specialist Farms in the small area), the unit is assigned to the small area population. Once a unit is selected for a given small area, quota counts are amended, reduced by the sum of the characteristics of the assigned unit(s). For individual level constraints, we increment the running totals per constraint by the number of units with that particular constraint. This procedure continues until the total number of simulated units is equal to the small area population aggregates (i.e. all quotas have been filled).

The quota sampling process therefore involves the following steps

- Thus for each unit *i*, we draw a random number *v*
- Sort units by *v*.
- Select the unit for spatial sample if $x_{j,s}^{acc} + x_{j,i} \le x_{j,s}^{total} \forall j$, where $x_{j,i}$ is the value of the variable j for the unit i, $x_{j,s}^{total}$ is the target total for district s for variable j and $x_{j,s}^{acc}$ is the running total for variable j for district s.
- If $x_{j,s}^{acc} + x_{j,i} > x_{j,s}^{total}$ for any j, then the we do not sample the unit i.

[2] SMILE-FARM: Simulation Model of the Irish Local Economy, Farm Model

Thus, one can see that the variation of admitted units cumulates in a random sort which is consistent with aggregate constraint totals. This mechanism of sampling without replacement avoids the repeated sampling procedure of SA and is fundamental to the efficiency gains of the quota sampling procedure relative to other methods. One can see that the process is analogous to the type of quota sampling undertaken by market researchers, whereby only individuals considered relevant to concurrent quota counts are admitted to a sample.

This method of improving efficiency does present a number of convergence issues, however. Disparities in population distributions between census and survey totals may create a number of problems for unit-based microsimulation procedures. This is because survey microdata are representative at the national level, whereas small area census data are representative at the district level. This poses little difficulty in simulating small areas that have a population distribution similar to that of the national distribution, but areas that differ from the national distribution may lead to some demographic groups consistently being underrepresented in a given district. These differences may cause some districts to consistently fail in reaching adequate convergence.

Also, the use of sampling without replacement in quota sampling results in quota counts becoming increasingly more restrictive as the simulation progresses. As quota counts reach their target, the search space is continuously refined in accordance with concurrent quotas, whereby all units no longer eligible given updated quota totals are removed from the subset and the procedure is repeated[3]. When each constraint allocation reaches its target quota, all individuals of that characteristic are removed from the candidate search space. These mechanisms cumulate to offer a continuously diminishing search space and may prohibit convergence, whereby no unit is able to satisfy all concurrent quota counts.

Improving the fit of the Spatial Stocking Rate

Hynes et al. (2009) utilised farm size, farm speciality and soil code to generate the spatial distribution of agriculture. This however ignores differences in stocking rate, which given that that Irish Agriculture is largely animal based is likely to be a significant driver of farm income heterogeneity not accounted for by farm system, size and soil type. In addition to economic considerations, it is also likely to be an important driver of the environmental impact of agriculture.

While we know the average stocking rate in each spatial district and we know the stocking rate of each farm, we are unable to utilise this variable within the quota sampling process of Farrell et al. (2013) or the Simulated Annealing process of Hynes et al., (2009), which requires the number of farms with a particular characteristic to be sampled. These methods cannot handle spatial averages.

Thus the objective of this new methodology is to improve the spatial heterogeneity of the stocking rate. In devising a method, consider the following relationship between match variable (soil, system, size) dummies and stocking rate

[3] e.g. with a remaining quota count of *n* individuals of class *k* to be filled, the search space is refined to exclude households containing *n+1* individuals of class *k*.

$$stocking\_rate_s = \sum_j match\,\text{var}\_share_j \beta_j + u_s \tag{1}$$

Where the stocking rate of the district s is a function of the share of farms by system, farms by size and farms by soil type, with unobserved heterogeneity being accounted for by a stochastic term u_s

Consider now the stocking rate for farm i

$$stocking\_rate_i = \sum_j match\,\text{var}\_share_j \beta_j + u_s + \varepsilon_i \tag{2}$$

where the stocking rate of the farm i is a function of the share of farms by system, farms by size and farms by soil type, with spatial unobserved heterogeneity u_s and farm level unobserved heterogeneity being accounted for by a stochastic term ε_i.

Thus if we believe in the consistency of our spatial and survey data, where by the underlying relationship between the stocking rate and match variables are the same, then rather than randomly selecting farms for selection, we would like to select farms where the unobserved heterogeneity is similar.

We can partially identify this by estimating β_j using our spatial data and deriving an area effect u_s, applying the coefficients β_j to the micro data and deriving farm level unobserved heterogeneity $u_s + \varepsilon_i$. A selection of farms that can result in a similar spatial stocking rate from sampling to the actual spatial stocking rate are farms farm level unobserved heterogeneity $u_s + \varepsilon_i$ is closest to the spatial unobserved heterogeneity u_s.

To improve the fit, therefore rather than sorting randomly, we sort on the difference between the two residuals. Thus, before selection commences, farms are ranked by the smallest absolute difference between the stocking rate residual for the current district and the stocking rate residual contribution reported for the sample farms. This step means that farms with residuals which most closely resemble the residual stocking rate of the target district are more likely to be selected first. The SMILE-FARM model then considers each ranked farm in the micro data file for inclusion in target district. The application of this ranking is designed so that each target Districts residual stocking rate, unexplained by the linear regression model, can be somewhat preserved.

This assumption rests on the basis that if spatial unobserved heterogeneity is important then u_s is high as a share of $u_s + \varepsilon_i$ in which the approximation of the absolute difference between the residuals will largely account for the spatial effect. On the other hand if unobserved spatial heterogeneity is small, then the absolute difference will be largely driven by the aspatial stochastic term which is assumed to be random.

Post Sample Adjustment of Variability

One of the consequences of sampling with the absolute residual difference adjustment is that while the mean fit may improve, the intra district variation may be reduced. If this proves to be an issue, one potential alternative is to do an post sampling adjustment to the variability, while maintaining the means.

To do this we estimate a series of fixed effects models for income components y_k (gross output, direct costs, overhead costs, subsidies) such that

$$Y = \sum_k y_k = \sum_k \left(X\beta_k + \sum_r \mathrm{Re}\,gion_r \right) + u_{kr} + v_k \tag{3}$$

Where u_{kr} are regional fixed effects with standard deviation σ_{ukr}, normally distributed.

Post sampling, we can also estimate a fixed effects model:

$$Y = \sum_k y_k = \sum_k \left(X\beta_k^* + \sum_r \mathrm{Re}\,gion_k^* \right) + u_{kr}^* + v_k^* \tag{4}$$

With corresponding standard deviation σ_{ukr}^*.

To improve the spatial variability of the model, we can adjust each income source as follows

$$Y^* = \sum_k y_k^* = \sum_k \left(X\beta_k^* + \sum_r \mathrm{Re}\,gion_k^* \right) + u_{kr}^* . \frac{\sigma_{ukr}}{\sigma_{ukr}^*} + v_k^* \tag{5}$$

So that the resulting regional fixed effects variability is the same as in the raw data

4. Data and operational implementation

In this section we discuss the data required for our analysis and provide some summary statistics. In designing a framework for spatial microsimulation models, the basic goal is to ensure that units from the micro data are simulated to the destination spatial unit by matching the characteristics of the micro units selected to the spatially heterogeneous characteristics of the spatial unit.

Data Description

In the SMILE-FARM model, farms from the Teagasc National Farm Survey (NFS) 2010 are sampled to reflect the structure of an Electoral Division (ED) on the basis of aggregate farm totals reported for that district in the Census of Agriculture (CoA) 2010.

Teagasc's National Farm Survey (NFS) to describe the distributional and incentive implications of the SFP. The NFS is collected as part of the Farm Accountancy Data Network of the European Union (FADN 2005). It determines 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 (Connolly et al. 2010).

A random sample of approximately 1,200 farms is surveyed each year.[4] In the Teagasc National Farm Survey (NFS), the principal measure of the income which arises from the year's farming activities is Family Farm Income per farm (FFI). The FFI is calculated by deducting all farm costs (direct and overhead) from the value of farm gross output and adding farm subsidises. Farm Gross Output (GO) does not include income from non-farming sources and thus may not be equated to household income. Most farms in Ireland contain multiple enterprises (beef cattle, sheep, dairy cows, cereals etc.) and so the National Farm Survey classifies a farm by the dominant enterprise. The dominant enterprise is defined as the system with the highest share of gross margin (output for the enterprise minus direct costs). There are substantial variations in margins across enterprises.

The Census of Agriculture is collected approximately every 10 years. It collects primarily physical data in relation to the number of animals by type, the size of the farm and the land use on the farm as well as some demographic data. 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. Due to the Commission decision 78/463ECC all the farms covered in the 2010 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, five 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, Other.

The CoA contains information on people who have registered with the Dept. of Agriculture to avail of agricultural subsidies, and to comply with the Departments agricultural regulations. It contains information on approximately 139,000 active farms (CSO, 2010).

Hynes et al. (2009) identify limitations associated with the NFS and the Census of Agriculture. The NFS contains a large amount of 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). Therefore, 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 to form a static farm level spatial microsimulation model the resulting dataset would provide policy-makers with detailed synthetic microdata as to inform their decision-making at a spatially disaggregated level.

Operational Implementation

In order to have a basis for the application of any microsimulation methodology, match variables common to both the micro data and the spatial data must first be

[4] Very small farms, and pig farms are excluded.

identified. For the SMILE-FARM model, farms are matched to destination districts by the main basic farm characteristics i.e. farm size, speciality and soil type. The choice of variables is determined by those that overlap between the two datasets and that account for over half the variability of family farm income. The CoA provides the aggregate totals for these match variables for each ED. A part-time rate variable by region and speciality is also simulated and applied to the CoA totals on the basis of information from the NFS.

In order to prepare the data for Quota Sampling Matching Process, we require a number of steps

- Step1. Prepare data
- Step2. Create target totals
- Step 3. Preparation and Selection for each district

Step1. Prepare data: Within the 2010 Teagasc NFS dataset , farms are identified, categorized and dummied by the farm speciality (5 categories), farm size (4 categories), soil code (3 categories) and whether the farm is part-time or full-time. These are the match variables which are used for the spatial microsimulation match. In addition, each farms stocking rate per hectare is calculated based on the total number of livestock units per hectare.

Step2. Create target totals: The number of farms in each of the categories of three match variables are then calculated. Because of data confidentiality issues, occasionally categories are rounded. As a result the sum of the number of farms per category does not always sum to the number of farms in the district. We adjust to ensure that the new totals are integers that sum to the total number of farms per district . This gives an integer total for each category with the sum of all categories equalling the districts target farm total.

Step 3. Preparation and Selection for each district: Separately and sequentially, each individual district from the uprated CoA file is then merged with the micro data file and the matching process begins. The selection sample size is limited to those farms matching the dominant soil type for the target district. Target totals or “quotas” for the match variables and the part-time rate are then created and updated each time a farm is selected. Farms are then selected without replacement for inclusion until any one of the totals or “quotas” for that district is filled. The model then skips all farms with the characteristic of the filled bin and fills the district sequentially with the remaining farms until a second bin is filled. The process then repeats until all quotas are filled or until the remaining farms which can be selected has shrunk to zero, i.e. there is no farm remaining in the micro data that can be added without overfilling one of the already filled quotas. If the target total number of farms for the district has not been reached within two iterations of searching the micro data file, the part-time constraint is relaxed and the model moves to the next iteration. This process repeats until either the total target number of farms for the district has been reached or the number of iterations reaches a predetermined termina.

Summary Statistics

In table 1, we report summary statistics that compare aggregates from the micro data used in this analysis with the census aggregates. As with all spatial microsimulation models, the initial consideration is that of choosing which variables constrain the data fusion (Smith et al., 2009). O'Donoghue et al. (2011) outline the process of choosing constraints in SMILE using bivariate regressions of candidate variables against farm income in the NFS microdata.

Table 1. Summary Statistics

	NFS	Census	Ratio
Speciality			
Tillage	7.6	3.8	2.03
Dairy	19.7	20.6	0.95
Cattle	58.9	65.9	0.89
Sheep	13.8	9.7	1.42
Size			
< 20 Ha	31.7	46.4	0.68
20-30 Ha	22.1	17.9	1.23
30-50 Ha	24.7	20.4	1.21
>50 Ha	21.5	15.3	1.41
Soil			
Best	47.6	44.5	1.07
Middle	40.9	37.6	1.09
Worst	11.3	17.8	0.64

Source: Teagasc National Farm Survey, 2010; Census of Agriculture, 2010
Note: Cattle farms include mixed farms, which may partially account for differences in cattle and sheep.

Table 1 reports the share of farms by the 3 constraint variables used in the quota sampling methodology. What stands out is that farm enterprises that are more "commercial", with higher market incomes are more highly represented in the NFS, reflecting the sampling frame used. In particular, while tillage farms, which are largely commercial on better soils typically account for nearly 8% of the NFS, compared with about 4% in the Census of Agriculture. The share of dairy farms in reasonably similar, while cattle farms, which typically have lower incomes have a higher share in the Census of Agriculture. Special sheep farms have the opposite direction, perhaps reflecting the fact that the Cattle farming category include mixed farms, which may partially account for differences in cattle and sheep.

The most vivid difference is the significantly higher share of small farms and farms on the poorest soil quality in the Census relative to the NFS reflecting the sampling frame.

5. Results 1. Impact of Methodological Choices

There are a number of methodological choices when undertaking the data enhancement methodology in producing the spatial distribution of Agricultural Income. These choice include

- Sample from with farms within the same region or from the national sample in the National Farm Survey
- Sample within Less Favoured Areas or not
- Sample randomly or adjusting for localised stocking rate
- Post-sample Adjust for Differences in Income Variability

The first set of choices relate to the sampling frame that is used in the data enhancement process. One can choose to sample from the entire NFS sample or from a subset such as the 8 NUTS3 regions. Sampling from a relevant subset such as only sampling farms from the Western region when generating data for Western region districts allows for some spatial heterogeneity to be preserved, recognising for example that beef farms that are selected are more likely to be suckler farms in the West and Cattle rearing farms in the East. However sampling from a smaller spatial unit can result in a smaller cell size which may result in greater difficulties in achieving convergence. However this may come at the cost of a smaller cell size.

A slightly more aggregated sampling unit is to sample separately for less favoured areas that comprise over 70% of farms and non less favoured areas. It has some advantage in enhancing heterogeneity without as much sample size constraints.

The standard method involves ranking farms randomly and then selecting until quotas are filled. An alternative is to utilise the alternative ranking method described above, where farms are ranked on the absolute difference in residuals. A fourth choice is the post sampling regional fixed effect adjustment described above.

In this section we test the performance of a number of different options as follows. Table 2 describes the nature of the 12 scenarios used in the paper of the potential 16 options. We do not consider both sampling within Less Favoured Areas and within region in the same scenario.

Table 2. Methodological Scenarios

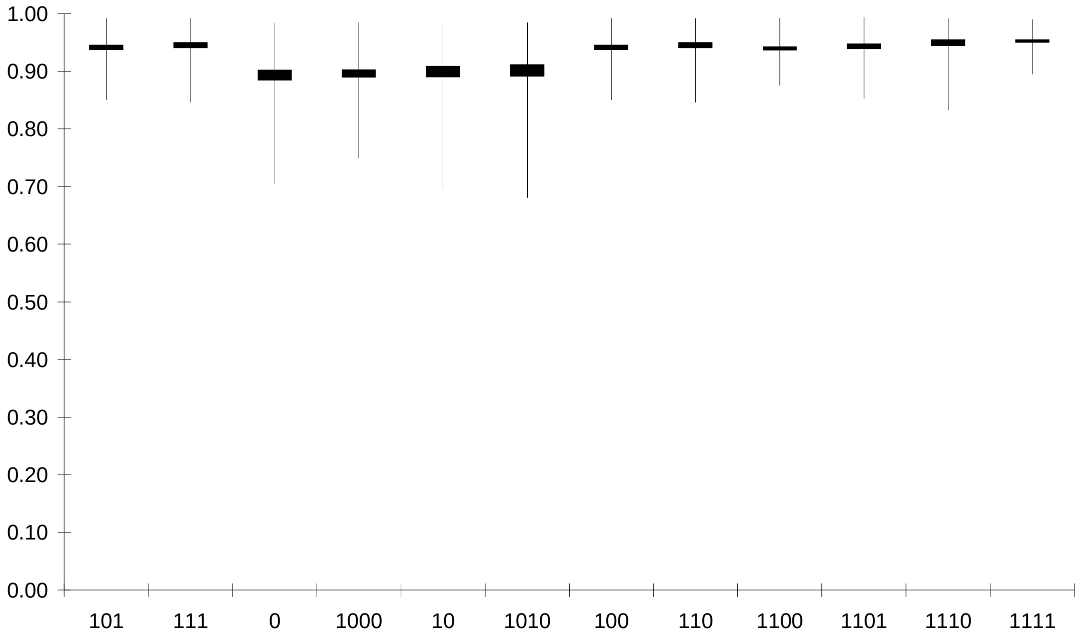
Scenario	101	111	0	1000	10	1010	100	110	1100	1101	1110	1111
LFA sub-group	1	1	0	0	0	0	0	0	0	1	0	1
Stocking Rate Adjustment	0	1	0	0	1	1	0	1	0	0	1	1
National (1)/Regional(0) Sample	1	1	0	0	0	0	1	1	1	1	1	1
Regional Fixed Effect Adjustment	0	0	0	1	0	1	0	0	1	1	1	1

The SMILE-FARM match for 2010 achieves the target total number of farms for all districts. In order to test the effectiveness of each match, we report the correlations between the raw and sampled constraint variables (soil, system, size) and with the non constraint variable stocking rate per hectare in Table 3. We summarise the data as a box plot in figure 1. By and large correlations with constraint variables are good. The Census of Agriculture constraints with the poorest performance (in terms of a correlation of less than 0.85) are specialist sheep and the two smallest size categories, reflecting the nature of the National Farm sample frame.

When we quantify the average correlation by scenario, the four scenarios where the sample was selected from within the same region in the NFS as the Census district (0, 10, 1000, 1010) are the poorest performing, with average correlations of about 0.89.

Nevertheless, given the fact that the survey and the Census have slightly different sampling frames, even the worst performing scenarios have a reasonable match.

Figure 1. Box Plot of Validation Statistics



Note:
1. Model Classification: 1000*Adjust Regional Error + 100*National Sample + 10*Stocking Rate Adjustment + 1*LFA Sample
2. The correlations described in this figure are reported in Table 3 below

Utilising a national sample relative to a regional sample has the biggest improvement in the average correlation, increasing the fit by about 5 percentage points. This is due to the fact that when we use the national sampling frame, the cell sizes are larger, giving the algorithm a wider choice of farms from which to select.

Table 3. Correlation matrix for Target Totals and simulated outcomes for SMILE-FARM

Model	101	111	0	1000	10	1010	100	110	1100	1101	1110	1111
Constraint Variable												
Specialist Tillage	0.970	0.972	0.809	0.802	0.805	0.796	0.859	0.855	0.867	0.862	0.884	0.862
Specialist Dairy	0.968	0.967	0.935	0.937	0.938	0.926	0.970	0.972	0.977	0.965	0.968	0.964
Speciality Beef and Mixed	0.990	0.990	0.941	0.947	0.944	0.950	0.968	0.967	0.971	0.972	0.973	0.976
Speciality Sheep	0.896	0.895	0.960	0.945	0.961	0.963	0.990	0.990	0.980	0.996	0.998	0.992
<=20 Ha	0.895	0.888	0.818	0.812	0.821	0.826	0.896	0.895	0.908	0.889	0.902	0.893
20-30 Ha	0.853	0.848	0.706	0.717	0.698	0.723	0.895	0.888	0.845	0.865	0.891	0.945
30-50 ha	0.932	0.932	0.780	0.781	0.789	0.747	0.853	0.848	0.876	0.857	0.883	0.865
50+ Ha	0.992	0.992	0.869	0.861	0.870	0.872	0.932	0.932	0.948	0.923	0.920	0.915
Good Soil	0.992	0.992	0.984	0.983	0.984	0.984	0.992	0.992	0.992	0.992	0.991	0.992
Medium Soil	0.994	0.994	0.978	0.979	0.979	0.980	0.992	0.992	0.988	0.989	0.989	0.991
Poor Soil	0.884	0.932	0.962	0.961	0.962	0.962	0.994	0.994	0.990	0.993	0.993	0.994
Average	0.937	0.940	0.884	0.889	0.890	0.891	0.937	0.940	0.937	0.939	0.944	*0.950*
Target Variable												
No LU per ha	*0.397*	*0.861*	*0.411*	*0.377*	*0.681*	*0.680*	*0.397*	*0.861*	*0.391*	*0.370*	*0.860*	*0.874*

Note: Model Classification: 1000*Adjust Regional Error + 100*National Sample + 10*Stocking Rate Adjustment + 1*LFA Sample

Amongst the other choices, there is a marginal improvement as the algorithm becomes more sophisticated with the stocking rate adjustment, selection from within relevant Less Favoured Area category and regional error adjustment. However these marginal changes are small relative to the impact of the national/regional sampling choice. When we consider the performance relative to the Census constraints, the best performing is the 1111 choice with an average correlation of 0.95. However the improvement is small relative to the simplest choice, 100, where farms are selected from the national sample, without adjustment.

Given the importance of animal systems in Irish Agriculture, the performance of the selection relative to the actual district stocking rate is important. Here we find a substantial difference in the performance. Selection scenarios that do not make the stocking rate adjustment are poorer with correlations typically 0.4 or lower than those that make the adjustment. The best performing scenarios are those with a stocking rate adjustment, with a national sample at about 0.85 correlation. Again there are marginal improvements with the regional error adjustment for the national sample, but a reduction for the less favoured area selection. The performance of the national sample is slightly better than when we select from within a less favoured area. When we select from within region with a stocking rate adjustment, the correlation is about 0.7

The nature of the match-process is such that there is a trade off b/w methodological complexity and computational efficiency. While it is possible a more accurate match for the match variables may have be obtained using the previous simulated annealing method developed by Hynes et al, (2008) the computational cost of simulated annealing approach is high. The quota sampling method provides a high level of accuracy for the match variables and allows the simulation to be modelled in a number of hours. The inclusion of a ranking mechanism provides the added benefit of preserving much of the spatial heterogeneity of each districts stocking rate.

6. Results 2: Spatial Structure of Agriculture

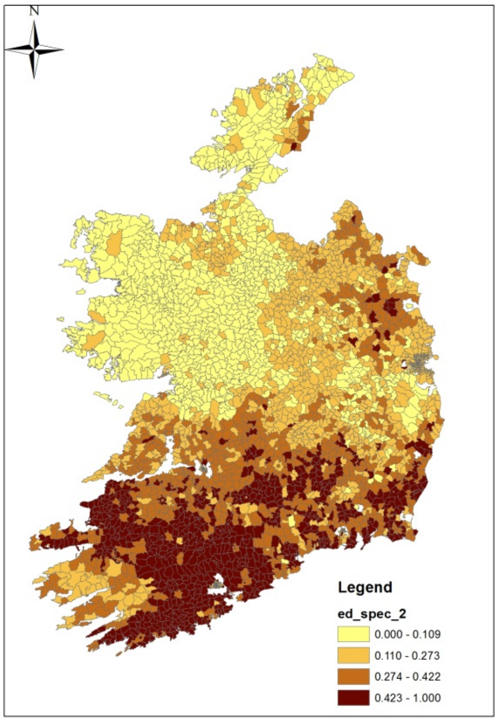
In this section we will describe the spatial structure of Agriculture observed in our synthetic population. Figure 2 reports the distribution of agricultural activity in Ireland utilising the spatially generated data within the SMILE-FARM model. The specialist dairy sector is located primarily in the South-West and close to the Northern border, with the mixed Dairy and Other sector having a similar but more widespread concentration. The next most profitable sector, the Tillage sector is primarily concentrated in the South-East and East, while the less profitable drystock sectors, Cattle and Sheep are located respectively in the Midlands and West and in the more peripheral areas across the Western sea-board. Thus the more profitable sub-sectors are located relatively close to the most populous regions of the country in the East and in the South.

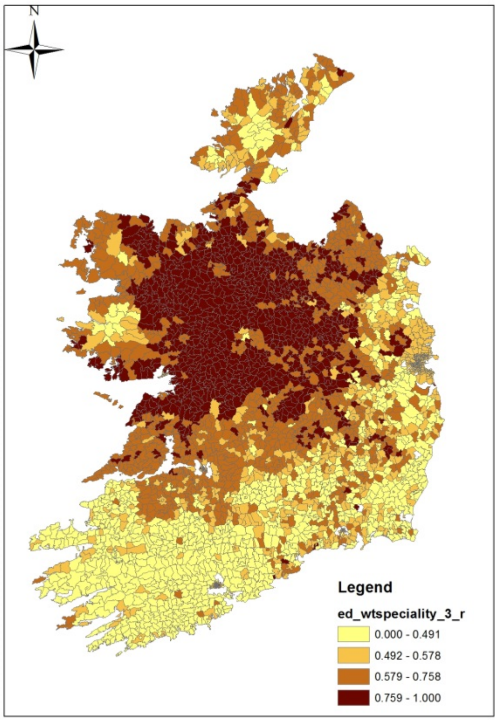
Market Returns and Viability

Underlying the economic sustainability indicators for agriculture in Ireland is the balance of different sectors. The Teagasc National Farm Survey in 2010 (Hennessy et al., 2011) reported that specialist dairy enterprises had an average income from farming (and subsidies in brackets) of €44432 (€21255), compared with €7023 (€13574) from cattle rearing enterprises, €12269 (€16528) from specialist sheep enterprises and €26759 (€24791) from specialist tillage enterprises. Thus as we can see, there is both a significant variability in incomes, but particularly so in net income from the market, defined as income minus direct payments which are dairy (€23177), beef (-€6551), sheep (-€2259) and tillage (€1962). Thus cattle and sheep enterprises are loss making from the market, relying on subsidies for income sustainability, while, dairy and tillage farming is largely profitable. Thus the spatial pattern of income will depend significantly on the predominance of these activities in particular locations.

Figure 2. Structure of Agriculture

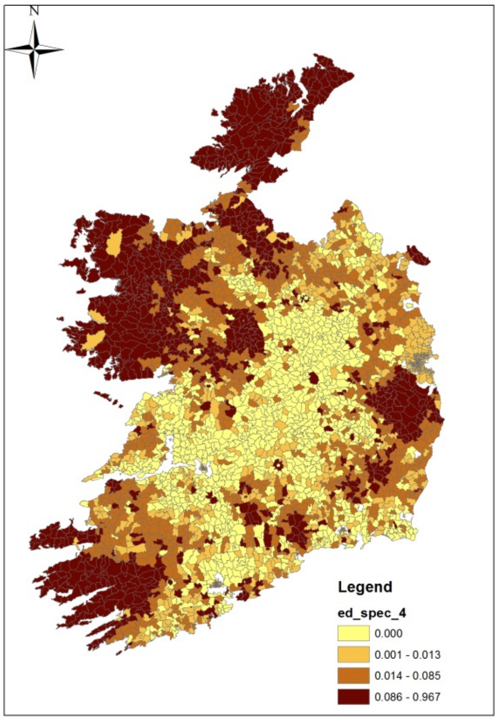
(a) Dairy (b) Cattle

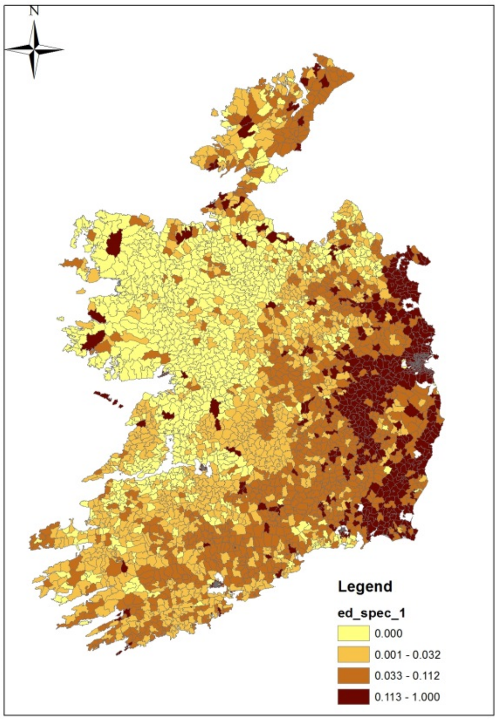




(c) Sheep

(d) Tillage

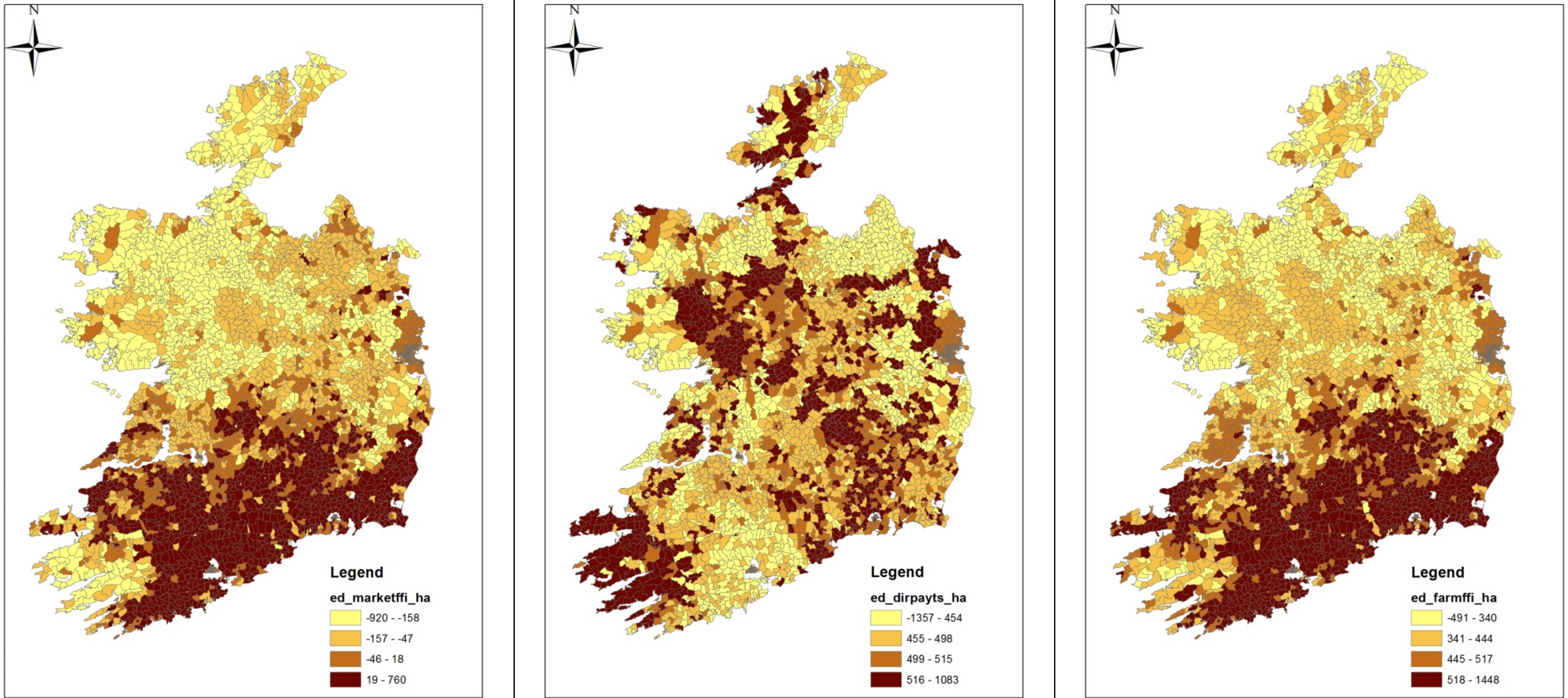




Source: SMILE-FARM 2010

Figure 3. Farm Incomes

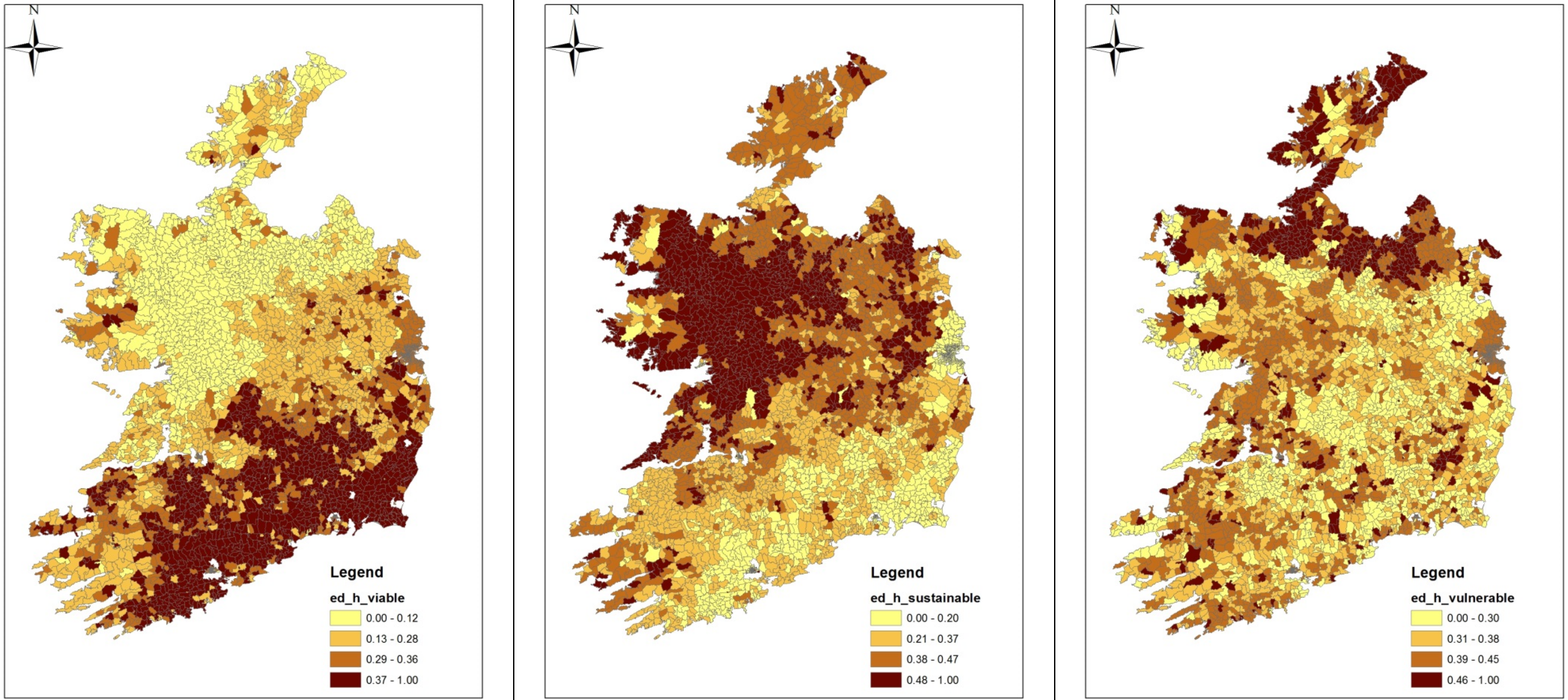
(a) Market Family Farm Income per Ha | (b) Direct Payments per Ha | (c) Family Farm Income per Ha



Source: SMILE-FARM 2010

Figure 4. Farm Viability, Sustainability and Vulnerability

(a) Proportion of Viable Farms | (b) Proportion of Sustainable Farms | (c) Proportion of Vulnerable Farms



Source: SMILE-FARM 2010

In Figure 3 describes the pattern of, market farming income from farming (excluding a return to labour or land) per hectare, direct payments per hectare and their sum family farm incomes per hectare. Market Income from farming reflects the location of dairy and tillage farming in the South and East and corresponds to the better land and consistent with the Commins-Frawley line from Dundalk to Limerick (Frawley and Commins, 1996) which divides what are effectively two agricultural economies. The spatial pattern of direct payments is less clear cut. While the pattern of single farm payments will be a function of particularly the intensity of cattle production and tillage production, largely down the East coast, the prevalence of disadvantaged area payments and agri-environmental payments will typically be more likely to be on the West and in the North in poorer agronomic zones. Combining the two measures, we find that market income driving the overall pattern.

The maps in figure 3 reflect spatial averages. One simple distributional measure is the share of viable farms viable farm is defined as having (a) the capacity to remunerate family labour at the minimum agricultural wage,[5] and (b) the capacity to provide an additional 5 per cent return on non-land assets. Figure r, report the spatial pattern of viability, which tracks the spatial pattern of family farm incomes South and East of the Commins-Frawley line.

Reflecting the balance of sectors, particularly the location of the more profitable dairy and tillage sectors, we see in figure 4, the balance of viable farms across the south and east, with 50% or higher of farms in most districts being viable. Meanwhile the location of the (on average) less profitable sheep and beef enterprises influences the pattern of unsustainable farms (farms without a viable income and without off-farm employment) across the West, Border and North, with 50% or more of farms in most districts being unsustainable. However these viability indicators are based around, the relatively low, minimum agricultural wage paid to farm labourers which is less than half the average wage earned by industrial workers.

Many farm households therefore require other sources of income to have household income sustainability. Sustainable farms are those farms that are not viable, but have off-farm employment, while unsustainable farms are neither viable nor have off-farm employment. In figure 4.b we report the spatial share of sustainable farms, with lower income farms within commuting distance of urban centres in the West and the midlands having a higher share of sustainable farms. The residual category in Figure 4.c are those farms with farm incomes below the viability threshold and without another source of employment income, with concentrations in peripheral areas and the North-West outside of commuting range of urban centres.

7. Conclusions

The objective of this paper was develop a methodology to assess spatial distributional impact of the Common Agricultural Policy Pillar 1 Reforms that will take place from 2015. In Ireland, these reforms will move from a historical based payments system in place since 2005 with a transition towards a flatter system with a combination of a

---

[5] In the absence of an average Irish agricultural wage, the minimum wage for agricultural workers as set by the Labour Court annually is used here.

basic payment and a greening related payment. There are is significant spatial pattern of different farming systems reflecting agronomic and environmental conditions, which historically had different direct payment eligibility and as a result the reforms may have a spatial impact.

The challenge in undertaking such an exercise is that there is no suitable dataset available. While the Irish Farm Accountancy Data Network Data, the Teagasc National Farm Survey contain sufficient data to measure the distributional impact of the reform, they do not have sufficient data to model the spatial distribution.

This paper develops a Quota sampling method which is a simpler but computationally more efficient measure than other methods. We also develop a number of extensions to potentially address a range of issues such as the ability to account for the spatial stocking rate. We validate the model against a range of external totals including Census size, system and soil type, local stocking rate and regional CAP reform winners and losers. The best model is one where farms are selected within region and a stocking rate adjustment is applied.

The paper then models the spatial distribution of Agricultural systems, farm incomes and direct payments. It captures well the pattern of farm activity on either side of the Commins-Frawley line that divides the country. We also modelled the spatial pattern of viable, sustainable and vulnerable farms. This reflects the pattern of farm and off-farm incomes and highlights areas in the country with high proportions of vulnerable farms, requiring other rural development initiatives.

We complete the study by modelling the spatial distribution of CAP reform. Primarily the biggest share of winners are in areas with lower average single farm payments in the West, North and coastal areas. Overall the reform reduced inequality in payments but increased inequality in incomes taking some payments from farms with high direct payments but low market income and giving them to farmers with lower direct payments but higher market income. However there are some areas in the East with a higher share of winners, which is cancelled out by a higher share of losers. This reflects our result that approximately 85% of farm income and direct payment variability is accounted for within area, between farm variability. As a result with heterogeneous farms within a district there can be winners and losers within the same area. Nevertheless the share of between district variability is higher than for other types of income, reflecting spatial heterogeneity in environmental and agronomic conditions. The net impact of the reform was however to reduce spatial variability.

8. References

Allanson, Paul. (2006). "The redistributive effects of agricultural policy on Scottish farm incomes." *Journal of Agricultural Economics* 57.1: 117-128.

Allanson, Paul, and Benedetto Rocchi. "A comparative analysis of the redistributive effects of agricultural policy in Tuscany and Scotland." *Review of agricultural and environmental studies* 86.1 (2008): 6-28.

Alfaro-Navarro, J. L., Mondejar-Jimenez, J., Vargas-Vargas, M., Gazquez-Abad, J. C., & Jimenez-Guerrero, J. F. (2011). The Effects Of The Distribution Of Agricultural Direct Payments. *Review of Business Information Systems (RBIS)*, *15*(5), 25-30.

Ballas, D. and Clarke, G.P., (2000). ‘GIS and microsimulation for local labour market policy analysis’, *Computers, Environment and Urban Systems*, Vol. 24, pp. 305–30.

Ballas, D., Clarke, G.P. and Wiemers, E. (2005). Building a dynamic spatial microsimulation model for Ireland, *Population, Space and Place*, 11, 157-172.

Ballas, D., Clarke, G.P., Wiemers, E., (2006). Spatial microsimulation for rural policy analysis in Ireland: The implications of CAP reforms for the national spatial strategy. *Journal of Rural Studies* Vol. 22 pp 367–378.

Bennett, V.J., M. Beard, P. A. Zollner, E. Fernandez-Juricic, L. Westphal, C.L. LeBlanc (2009). Understanding wildlife responses to human disturbance through simulation modelling: A management tool. Ecological Complexity, Vol. 6 pp. 113 – 134.

Bergmann, Holger, Eva Maria Noack, Kenneth J. Thomson (2011) The Distribution of CAP Payments - Redistributional Injustice or Spatially Adapted Policy? 85rd Annual Conference of the Agricultural Economics Society Warwick University 18 - 20 April

Clancy, Daragh James Breen, Karyn Morrissey, Cathal O’Donoghue and Fiona Thorne (2013) The Location Economics of Biomass Production for Electricity Generation, in Cathal O’Donoghue, Stephen Hynes, Karyn Morrissey, Dimitris Ballas, and Graham Clarke (Eds.), Spatial Microsimulation for Rural Policy Analysis. Springer-Verlag - Advances in Spatial Science.

Clarke G.P., (1996), *Microsimulation for Urban and Regional Policy Analysis*, London: Pion,

Hennessy, T., Moran B., Kinsella A. and Quinlan G (2011) *National Farm Survey* 2010. Teagasc, Agricultural Economics and Farm Surveys Department.

Crowley, C. , Walsh, J. and Meredith, D., (2008). Irish Farming at the Millennium: A Census Atlas. NIRSA, NUI Maynooth

Cullinan J., Hynes, S., and O’Donoghue, C., (2008) “Using Spatial Microsimulation to Analyse Catchment Populations – An Application to a Recreation Forest in Ireland” Irish Geography.

Deming, W.E. and Stephan, F. F.(1940). On a least square adjustment of a sampled frequency table when the expected marginal totals are known. *Annals of Mathematical Statistics*, 11(4):427–444, December.

Donnellan, Trevor, John Finn, Kevin Hanrahan, Thia Hennessy, Catherine Keena, James McDonnell, Pat Murphy, Daire O hUallachain, Fintan Phelan,, Fiona Thorne (2013).Teagasc submission made in response to the Department of Agriculture, Food and the Marine CAP Public Consultation Process. Teagasc Working Group on CAP Reform.

El Benni, Nadja, et al. "The distributional effects of agricultural policy reforms in Switzerland." *Agricultural Economics (Zemědělská Ekonomika)* 58.11 (2012): 497-509.

Farrell Niall, O'Donoghue, Cathal, Karyn Morrissey, John Lennon, Dimitris Ballas, Graham Clarke and, Stephen Hynes (2013) The SMILE model: construction and calibration, in Cathal O'Donoghue, Stephen Hynes, Karyn Morrissey, Dimitris Ballas, and Graham Clarke (Eds.), Spatial Microsimulation for Rural Policy Analysis. Springer-Verlag - Advances in Spatial Science.

Finger, R., & El Benni, N. (2011). Spatial analysis of income inequality in agriculture. *Economics Bulletin*, 31(3), 2138-2150.

Frawley, J. P and P. Commins (1996). The Changing Structure of Irish Farming: Trends and Prospects. Rural Economy Research Series No.1. Dublin: Teagasc.

Frawley, J., P. Commins, S. Scott and F. Trace (2000) Low Income Farm Households: Incidence, Characteristics and Policies Combat Poverty Agency Research Series, Oak Tree Press, Dublin

Ghosh, Malay, and J. N. K. Rao. (1994) "Small area estimation: an appraisal." Statistical science: 55-76.

Hennessy, T and Rehman T (2007). An Investigation into the Factors Affecting the Occupational Choices of Farm Heirs. Journal of Agricultural Economics. Vol. 58. No. 1.

Hennessy, T and Rehman T (2008). Assessing the Impact of the Decoupling Reform on Irish Farmers' off-farm Labour Market Participation Decisions. Journal of Agricultural Economics. Vol. 59.

Hermes, Kerstin, and Michael Poulsen. "A review of current methods to generate synthetic spatial microdata using reweighting and future directions." Computers, Environment and Urban Systems 36.4 (2012): 281-290.

Hynes, S., Farrelly, N., Murphy, E. and O'Donoghue, C. (2008). "Modelling habitat conservation and participation in agri-environmental schemes: A spatial microsimulation approach." Ecological Economics, vol. 66, issue 2-3, pages 258-269

Hynes, S., Morrissey, K., O'Donoghue, C. and Clarke, G.. (2009a). "A Spatial Microsimulation Analysis of Methane Emissions from Irish Agriculture." Journal of Ecological Complexity 6: 135– 146.

Hynes, S., Morrissey, K., O'Donoghue, C. and Clarke, G. (2009b). "Building a Static Farm Level Spatial Microsimulation Model for Rural Development and Agricultural Policy Analysis in Ireland." International Journal of Agricultural Resources, Governance and Ecology, 8 (3): 282-299.

Hynes, S., Hanley, N. and O'Donoghue, C. (2010). A Combinatorial Optimization Approach to Non-market Environmental Benefit Aggregation via Simulated Populations, Land Economics, 86 (2): 345–362

Keeney, Mary. (2000) "The distributional impact of direct payments on Irish farm incomes." *Journal of Agricultural Economics* 51.2: 252-265.

Lindgren, U. and Elmquist, H., (2005). Environmental and Economic Impacts of Decision-Making at an Arable Farm: An Integrative Modeling Approach. *Ambio: A Journal of the Human Environment*. Vol. 34, No. 4.5.

Moeckel, R., Schwarze, B., Spiekermann, K., Wegener, M., (2007). Simulating Interactions between Land Use, Transport and Environment, Paper presented at the 11th World Conference on Transport Research, University of California at Berkeley, 24-28 June 2007

Mavoa, S., (2007). Estimating the Social Impact of Reduced CO2 Emissions from Household Travel Using GIS Modelling. In Morris, Jenny and Rose, Geoff, Eds. Proceedings 30th Australasian Transport Research Forum (ATRF), Melbourne, Australia

O'Donoghue Cathal, Stephen Hynes, Karyn Morrissey, Dimitris Ballas, and Graham Clarke (2013), *Spatial Microsimulation for Rural Policy Analysis*. Springer-Verlag - Advances in Spatial Science.

O'Donoghue, Cathal (2013b). Modelling Farm Viability, in Cathal O'Donoghue, Stephen Hynes, Karyn Morrissey, Dimitris Ballas, and Graham Clarke (Eds.), Spatial Microsimulation for Rural Policy Analysis. Springer-Verlag - Advances in Spatial Science.

O'Donoghue, Cathal, Karyn Morrissey, Philip Hayes, Jason Loughrey, Joanne Banks and Stephen Hynes (2013c) The Spatial Distribution of Household Disposable Income, in Cathal O'Donoghue, Stephen Hynes, Karyn Morrissey, Dimitris Ballas, and Graham Clarke (Eds.), Spatial Microsimulation for Rural Policy Analysis. Springer-Verlag - Advances in Spatial Science.

O'Donoghue Cathal and Hennessy Thia, 2014. Economics of the Irish Agri-Food Sector. Teagasc Working Paper Series.

O'Donoghue, Cathal, John Lennon, Karyn Morrissey (2014). Survey of Spatial Microsimulation Modelling. *International Journal of Microsimulation*. 7 (2)

Oskam, Arie, Gerrit Meester, and Huib Silvis (2011) EU policy for agriculture, food and rural areas. Wageningen Academic Publishers.

Pfeifer, Catherine, Marthijn PW Sonneveld, and Jetse J. Stoorvogel. (2012)"Farmers' contribution to landscape services in the Netherlands under different rural development scenarios." Journal of environmental management 111: 96-105.

Pirzio-Biroli, C. (2008), An Inside Pesrpsective on the Political Economy of the Fischler Reforms in Swinnen, J. (ed.) *The Perfect Storm: The Political Economy of the Fischler Reforms of the Common Agricultural Policy*. Brussels@ Centre for European Policy Studies.

Riordan, Brendan, (2012) The contribution of the 'biosector' to Ireland's net foreign earnings. Dublin: Department of Agriculture, Food and the Marine.

Severini, S., & Tantari, A. (2012). The effect of the EU farm payments policy and its recent reform on farm income inequality. *Journal of Policy Modeling*. Vol. 35: 212-

227Shrestha, S., Hennessy, T., & Hynes, S. (2007). The effect of decoupling on farming in Ireland: A regional analysis. *Irish Journal of Agricultural & Food Research*, *46*(1).

Smith, D.M., Harland, K. and Clarke, G.P., (2007). SimHealth: estimating small area populations using deterministic spatial microsimulation in Leeds and Bradford. University of Leeds School of Geography, Working Paper 07/06.

Smith D M, Clarke G P, Harland K, (2009), "Improving the synthetic data generation process in spatial microsimulation models" *Environment and Planning A* 41(5) 1251 – 1268

Tanton, R., Yogi Vidyattama, Binod Nepal and Justine McNamara, (2011), Small Area Estimation Using A Reweighting Algorithm . *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 174, 4, 931-951.

van Leeuwen, E., & Dekkers, J. (2013). Determinants of off-farm income and its local patterns: a spatial microsimulation of Dutch farmers. *Journal of Rural Studies*, *31*, 55-66.

Williamson, Paul, Mark Birkin, and Phil H. Rees. "The estimation of population microdata by using data from small area statistics and samples of anonymised records." Environment and Planning A 30.5 (1998): 785-816.