A Local Regression Analysis of Irish Farm Census Data

Caroline Crowley[1], Jim Walsh[2]

[1]Department of Geography, National University of Ireland Maynooth, Kildare, Ireland
Telephone: (+353 21) 4905865
Fax: (+353 1) 7083573
Email: caroline.m.crowley@nuim.ie

[2]Department of Geography, National University of Ireland Maynooth, Kildare, Ireland
Telephone: (+353 1) 7083611
Fax: (+353 1) 7083573
Email: jim.walsh@nuim.ie

1. Introduction

Researchers who use spatial data are increasingly aware of the limitations of global regression techniques, such as Ordinary Least Regression (OLS), which by generating 'global' outputs may mask local variations. Geographically weighted regression (GWR), a local regression technique, has been applied in many instances to successfully reveal local relationships between socio-economic variables to explore phenomena such as regional industrialisation (Huang and Leung 2002), commuting (Lloyd and Shuttleworth 2005), the distribution of food consumption and food poverty (Farrow et al. 2005) and rural poverty (Benson et al. 2005).

In this paper, GWR is applied for the first time to a model that explores the economic scale of Irish farm enterprises throughout the state, and the results are compared with that of an OLS global model comprised of the same variables. The use of both global and local regression analysis provides the twin benefits of the elucidation of broad trends by the global analysis and the revelation of local deviations to those trends by the local analysis, both of which can be seen to characterize Irish farming.

2. Data and Methods

2.1 Data

The models are based on data primarily from the 2000 Census of Agriculture, at the spatial scale of the electoral division (ED). Agricultural activity was enumerated in almost 2,870 of the 3,440 EDs in the state in 2000.

2.2 Selection of the dependent variable

Models were built to assess the policy relevant dependent variable of average farm economic scale measured in European size units (ESU)[1]. Fig. 1 shows the geography of average farm economic scale in the state in 2000. This variable represents farm output as the farm gate value of primary and secondary farm products, inclusive of relevant

[1] The sum of the standard gross margins (SGM) for farm output or the monetary value of its gross production less specific costs, where 1 ESU = €1,200 SGM.

agricultural subsidies and exclusive of value added tax (VAT), minus direct costs related to the production of farm output such as seeds, fertilisers, pesticides, feed and livestock replacement. Direct costs do not include labour, machinery, buildings and fuel.

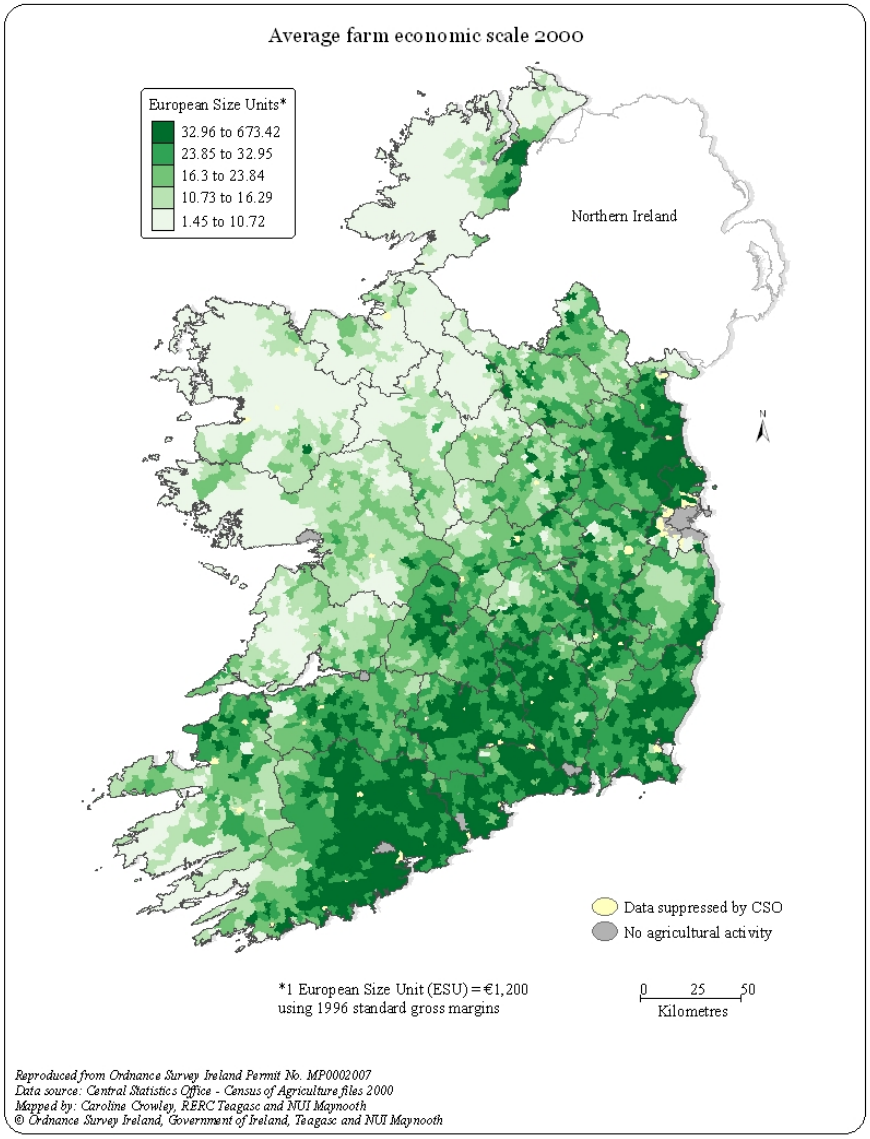

Figure 1. The geography of average farm economic scale in 2000.

Fig. 1 reveals a gradient of increasing values from the north-west where the average farm economic scale is less than 11 ESU to the predominance of high values of 33 ESU and higher in the south and east.

2.3 Selection of independent variables

Following a close examination of the definition of the ESU, independent variables were chosen based on hypotheses developed from assessing the spatial patterns of multiple agricultural variables in 2000 along with existing literature on Irish agricultural geography, and through discussions with Irish agricultural specialists. These steps resulted in the selection of 17 variables that measure land use type and intensity, farm size, farm holder characteristics, mechanisation and urban proximity.

2.4 Analytical Methods

A global model was calibrated first using OLS regression to assess the global statistics of the proposed model and provide a baseline against which to compare the performance of the local model using GWR. OLS regression was conducted in SPSS 12.0, while GWR 3.0 (ncg.nuim.ie/ncg/GWR/) was used to calibrate the local model. The outputs were then combined with the visualization power of the GIS to generate maps for interpretation.

2.5 Geographically weighted regression

GWR recognises that spatial nonstationarity[2] in processes may exist and extends traditional linear regression by allowing the estimation of local parameters, so that the linear regression equation becomes:

$$y_i = \beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i) x_{ik} + \varepsilon_i, \qquad (1)$$

where (u_i, v_i) denotes the coordinates of the ith point and $\beta_k(u_i, v_i)$ represents the continuous function of $\beta_k(u, v)$ at point *i* (Fotheringham et al. 2002) .

A spatial kernel is placed over each calibration point (ED centroid) and the data around that point are weighted, using a weighted least squares approach, according to the distance-decay curve of the kernel. At the regression point *i* (ED centroid), the weight of the data point is unity (equals one); it falls towards zero with increasing distance between the regression point and the data point. Using a Gaussian function, the weighting diminishes at a rate determined by a normal curve, so that the weight w_i, for observation *i* is:

$$w_i = \exp\left[-\frac{1}{2}\left(\frac{d}{a}\right)^2\right], \qquad (2)$$

where *d* is the distance between observation *i* and the kernel centre and *a* is the kernel bandwidth (Lloyd and Shuttleworth 2005). Thus, data observations near to regression point *i* have a greater influence on the estimated parameters of the relationship being measured at point *i* than distant observations. In this way, GWR accounts for the fact that processes and relationships may vary significantly over space related to spatial factors that range from physical, environmental and economic to social, cultural and political.

[2] Whereby the measurement of a relationship is influenced by the location of the measurement.

3. Results

Using stepwise regression, the OLS regression R^2 reached 91 percent with the addition of just three independent variables, namely: land use intensity measured in livestock units per 100 hectares of grassland and commonage, average farm size and the share of total livestock units represented by dairy cows and dairy heifer-in-calf. This means that 91 percent of the variation in average farm economic scale is associated with these variables. But the global model's residuals exhibited significantly positive spatial autocorrelation (Moran's I: 0.02, $p \leq 0.01$), which means that any inferences drawn from the global model are questionable.

When this model was calibrated in GWR using the same variables, the GWR model delivered a significant improvement in goodness-of-fit as confirmed by a statistically significant ANOVA (table 1) and a decline greater than three in the Akaike Information Criterion (AIC_c) (table 2).

Source	Sum of squares	DF	MS	F
OLS residuals	47.5	18		
GWR improvement	31.3	426.29	0.0734	
GWR residuals	16.2	2415.71	0.0067	10.93

Table 1. Local model ANOVA.

Model	AIC_c
Global	*-3565*
GWR	*-5621*

Table 2. Local model AIC_c.

Significant Monte Carlo tests revealed that the intercept and 12 of the 17 independent variables exhibited spatial non-stationarity in their relationships with the dependent variable. Fig. 2 of local R^2 values across the state shows that the model performs better from the north-east through the midlands, and from east Connacht to Clare and south to Cork. The model has lower explanatory power along the north-west coast, the Clare/Galway border, the extreme south-west, around Waterford in the south, and in the Greater Dublin Area of the mid-east.

Local factors that may explain the lower explanatory power of the model in these areas include the negative effects of peripherality in upland western areas reducing access to markets and the particularly strong labour market in the mid-east drawing farmers out of agriculture. Conversely, the growing consumer market related to population growth in the mid-east over the 1990s provided a positive effect. Relatively high percentages of commonage (poor quality land) in 2000 along with increases in farm fragmentation over the 1990s are also characteristic of EDs across most of these areas and may have reduced the explanatory power of the model in those areas.

Overall, the spatial pattern in fig. 2 supports the conclusion that more balanced regional development and thus a more widespread distribution of both the labour market, through job creation, and the consumer market, through population growth, could positively influence average farm economic scale in different locales. On the one hand, by promoting economic development in western peripheral areas, access to input suppliers and consumer markets could grow. On the other hand, by endorsing a more integrated approach to planning and development that takes local agricultural sustainability in eastern areas that are prone to over development and suburbanisation into consideration, the negative influence of rapid economic development on average farm economic scale could be countered. This example highlights that by revealing local variations in the relationships that underpin such models, GWR provides a tool to help in the design of location-specific public policies.

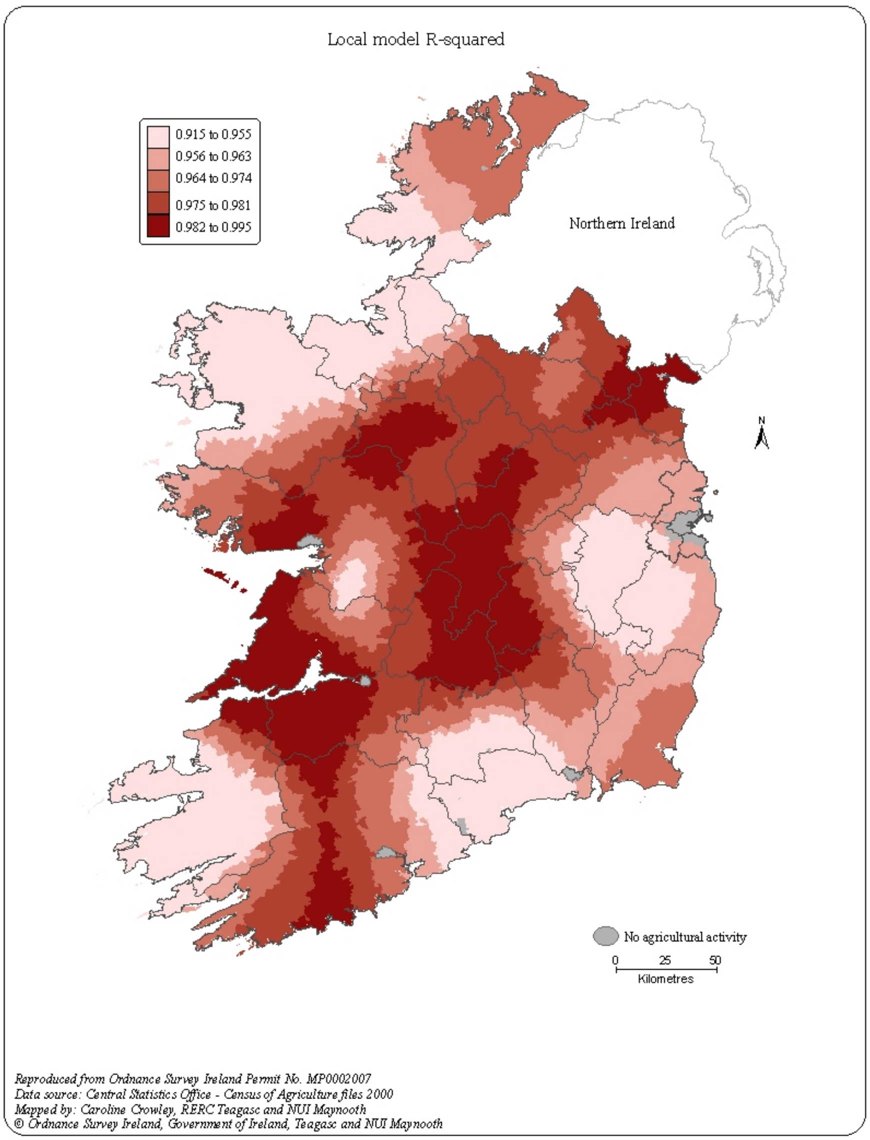

Figure 2. The geography of the local model R^2 values.

4. Acknowledgements

This work is the result of PhD research by the primary author, Caroline Crowley, funded under the Teagasc Walsh Fellowship programme. The authors are grateful to Gregg Patrick at the Central Statistics Office for farm census data provision and assistance with queries, and to Martin Charlton of the National Centre for Geocomputation for generating digital data for input to the regression modelling.

5. References

Benson T, Chamberlin J and Rhinehart I, 2005, An investigation of the spatial determinants of the local prevalence of poverty in rural Malawi. *Food Policy*, 30(5-6): 532-550.

Farrow A, Larrea C, Hyman G and Lema G, 2005, Exploring the spatial variation of food poverty in Ecuador. *Food Policy*, 30(5-6): 510-531.

Fotheringham AS, Brunsdon C and Charlton M, (2002), *Geographically weighted regression: the analysis of spatially varying relationships*. John Wiley & Sons Ltd, Chichester, UK.

Huang Y and Leung Y, 2002, Analysing regional industrialisation in Jiangsu province using geographically weighted regression. *Journal of Geographical Systems*, 4(2): 233-249.

Lloyd C and Shuttleworth I, 2005, Analysing commuting using local regression techniques: scale, sensitivity, and geographical patterning. *Environment and Planning A*, 37(1): 81-103.