ANALIZING WATER FRAMEWORK DIRECTIVE IMPACTS USING A MULTINOMIAL LOGIT LAND USE MODEL[1]

Carlo Fezzi[a] Ian J. Bateman[a] Brett Day[a] Paulette Posen[a] Dan Rigby[b]

[a]CSERGE,
School of Environmental Sciences,
University of East Anglia,
Norwich, NR4 7TJ (UK)

[b]Economics,
School of Social Sciences,
University of Manchester,
Manchester M13 9PL (UK)

Contact: c.fezzi@uea.ac.uk

Paper prepared for presentation at the 107th EAAE Seminar *"Modelling of Agricultural and Rural Development Policies"*. Sevilla, Spain, January 29th -February 1st, 2008.

[1] Funding for this research was provided by the ChREAM project, which is supported by the ESRC and other research councils under the RELU programme (ref: RES-227-25-0024). Ian Bateman is also Adjunct Professor at the Department of Agricultural and Resource Economics at the University of Western Australia, Perth and the in the Department of Agricultural and Resource Economics, Waikato University.

Abstract

We develop a two-stage, multinomial logit model of UK land use to investigate the impact of policy changes upon agriculture. The model utilises a large panel database covering the entirety of England and Wales for 14 years between 1969 and 2004 integrated with the economic and physical environment determinants of all major agricultural land use types. Our model performs well in out-of-sample prediction of current land use and we use it to assess a proposed implementation of the Water Framework Directive via a tax on fertilizer. Results indicate that such policy change would generate substantial switching from arable to grassland systems, reducing significantly the amount of nitrate leaching into UK water-bodies.

Key words: Water Framework Directive; Land use models; Discrete choice models; Multinomial logit

1. Introduction

In this paper we develop an empirical framework to model agricultural land use in the United Kingdom (UK) based on a discrete-choice econometric approach. We consider all major agricultural land use types, comprising both arable (e.g. wheat, barley, root crops, etc.) and grassland and estimate the model using a large panel dataset consisting of $5km^2$ grid square share data. The model is then applied, *inter alia*, to estimate the land uses arising from different policy options related to the implementation of the Water Framework Directive (WFD).

The WFD, in fact, requires European Union (EU) member states to achieve a "good ecological and chemical status" in all water bodies by 2015 (European Commission, 2000). In England and Wales roughly 80% of rivers are currently at risk of not meeting this target. The major source of nutrient pollutnats is from diffuse sources, which are in turn dominated by agriculture (Department for Environment, Food and Rural Affairs, Defra, 2007a). For example, Defra (*ibid*) estimates that the WFD implementation would require halving the amount of phosphates discharged by farms into UK rivers producing a total impact on the agricultural sector of £200 millions. Considering that the total income from agriculture has fallen significantly in the UK during the last decades, providing quantitative estimates of the effects arising from various WFD related measures is of interest not only for those stakeholders directly affected but also for those responsible for the formulation of such policies[2].

Previous studies have identified a range of measures (e.g. cutting fertilizer application rates, reducing livestock stocking rates, switching arable land to grassland) that could be adopted to reduce diffuse water pollution from agriculture in the UK, estimating the economic costs using linear programming (LP) models (Defra, 2007b) or farm account data (Fezzi *et al.*, 2007). Despite being clearly informative, these contributions do not take into account the spatial dimension of the problem, which, given the diversity of environmental characteristics and the agricultural practices that characterize a heterogeneous country such as the UK, is clearly substantial.

Furthermore, in those earlier analyses, the policies options are implemented in a rather mechanical fashion. For example, the livestock stocking rate reduction assumes that every farmer reduces stock levels by a common percentage, regardless the prior livestock density on farm or the livestock type. Therefore, behavioural responses are limited and all farmers are forced to respond in the same way to each policy (even though the disparity in economic impacts is estimated, distinguishing across broad sectors, in Fezzi *et al.*, 2007).

In this paper we tackle both issues developing a spatially explicit econometric model of farmers' land use decision using $5km^2$ grid square data from the June Agricultural Census covering the entire England and Wales (available on line at www.edina.ac.uk). We extend the methodology introduced by Chomitz and Gray (1996) and Chomitz and Thomas (2003), analyzing share data of the major agricultural land uses (including both arable and grassland) and taking into account variation in physical environment characteristics (e.g. rainfall, soil quality, slope), input (e.g. fertilizer, feedings) and output (e.g. wheat, barley, milk) prices. We test the validity of our model developing out-of-sample predictions for 2004 and comparing them with actual land use in that year. This analysis supports the predictive validity of our model which we then apply to simulate the impact of one of the policies proposed for implementation of the WFD: a fertiliser tax. We illustrate predicted response via a series of maps, showing land use pre- and post-policy implementation.

The paper is organized as follows. In section 2 we present our methodology, developing from our economic model our econometric approach. Results are presented in section 3 wherein out-of-sample validation and policy change simulations are provided. Section 4 concludes.

2. Modelling aggregate land use decisions

2.1. The economic model

Our economic model of land use allocation builds upon the original works of Wu and Segerson (1995) and Chomitz and Gray (1996). Assuming that farmers are risk neutral, they will choose to allocate their land to the uses that maximize, every year, their expected profits. Since we are considering agricultural land uses only, choice irreversibility is not an issue here. For simplicity this initial analysis excludes conversion costs and ignores issues of the endogeneity of total agricultural land given reasonable stability in overall area during the time span considered in the analysis.

Given this initial framework we can write the total present profits arising from allocating the various plots managed by farmer i (i = 1,…,N) among the different land uses k (k = 1,…,K) in year y as a function of the exogenous output prices p_{ky}, the exogenous input prices c_{ky} and the physical environment characteristics of the land e_i (which, in turn, determine inputs demand and productivity) as:

$$(1) \qquad \Pi_{iy} = \sum_{k=1}^{K} \pi_{iky} = \sum_{k=1}^{K} \pi_{iky}[p_{ky}, c_{ky}, e_i, a_{iky}(p_{ky}, c_{ky}, e_i)],$$

where a_{iky} indicates the amount of land devoted to land use k by farmer i. When the farmer seeks to maximize profits subject to land constraints, the solution to the optimization problem leads to the optimal land allocation a_{iky}*(p_{ky}, c_{ky}, e_i,). In this specification, the areas devoted to each land use depend upon input and output prices as well as the distribution of physical and environmental characteristics of the farm

Dividing the areas by the total amount of land managed by each farmer i, we can re-write the optimal allocation in terms of shares. Furthermore, the actual shares ϕ_{iky} might differ from the optimal ones s_{iky}*(p_{ky}, c_{ky}, e_i,), due to exogenous shocks (e.g. poor weather, unexpected price changes, etc.) and

[2] Indeed the research reported in this paper arises from direct request from Defra to the Authors for guidance regarding this issue.

unobserved variables (e.g. the experience of the farmer with particular cultivations or livestock) that we encompass in a random error term u_{iky}:

(2) $$\phi_{iky} = s_{iky} * (p_{ky}, c_{ky}, e_i) + u_{iky}.$$

We discuss error term (u_{iky}) assumptions and motivate the empirical econometric specification below.

2.2. The econometric specification

Following Wu and Segerson (1995) we assume a linear, logistic specification for the share functions $s_{iky}*(p_{ky}, c_{ky}, e_i,)$. Therefore, the shares can be written in a logit form as:

(3) $$s_{iky}* = \frac{\exp(\beta_k' X_{iy})}{\sum_{j=1}^{K} \exp(\beta_j' X_{iy})},$$

where X_{iy} includes exogenous input and output prices and environmental characteristics, and the parameters β_k differ across all possible land uses. This model is an equally valid representation even if we do not observe specific farmers decisions but only information on aggregated data, for example on a county or (as in this case) grid square basis.

This specification, implemented, for example, in Wu and Segerson (1995), Chomitz and Gray (1996) and Chomitz and Thomas (2003), implies that the ratio of the probabilities of any two choices is independent of the other alternatives. However, since our empirical model encompasses numerous land uses, including both arable and grassland, differences in substitutability among alternatives may be an issue. For example, when an arable crop (e.g. barley), becomes less profitable, if the farmer's past experience lies in crop cultivation, they may be more inclined to switch to another crop (e.g. wheat) rather than converting some or all of the farm to grassland for livestock rearing. This difference in substitution patterns can be incorporated in the model through a nested logit specification, as in Lubowski *et al.* (2006).

In this paper we opt for a simpler approach, estimating a two-step logit model. In the first step, a binomial logit is used to discriminate between arable land and grassland. In the second step, the share of arable land is allocated to the different crops and the share of grassland to the different types of pasture using a subsequent logit model. Furthermore, since (as we discuss subsequently) the survey points in our dataset are not evenly spaced across time, we assume the residual component to be uncorrelated across time. Therefore, the dynamics are included in the model exclusively through a trend component, representing the evolution of technology and husbandry practices. Even if it is quite simple, this approach provides accurate out-of-sample estimates.

3. The empirical analysis[3]

3.1 The model estimation

We estimate our model using 5km^2 grid squares collected for the whole of England and Wales in the June Agricultural Census conducted for thirteen, unevenly spaced, years between 1969 and 2003. The out-of-sample model performance is assessed by generating predictions for 2004 and comparing these to the actual land use for that year. We consider seven different types of land use, including two pastures (permanent grassland and temporary grassland) and five arable crops categories (wheat, barley, root crops, oilseed rape and other cereals) encompassing more than 80% of the agricultural land in the sample. We do not include set aside land for two reasons: (*a*) its allocation follows rather distinct economic considerations compared to the other land uses and (*b*) with the introduction of the single farm payment (SFP) scheme in 2005, set aside is no longer a requirement to receive decoupled subsides and therefore is expected to disappear. However, we do include the introduction of set aside payments within the model inserting a dummy variable for the year 1993 onwards, representing the implementation of the EU Common Agricultural Policy (CAP) reform in 1992. The only other significant agricultural land use not included in the model is rough grazing, the reason being a lack of data for most of the years. Trends in land use shares, derived from June Agricultural Census data (available on line through www.edina.ac.uk) are summarized in Figure 1.

Figure 1 shows the two main UK crops, barley and wheat, have been affected by a trend which progressively substituted the latter for the former. This land use change cannot be attributed to price, since the price indexes of those two commodities remained fairly close during all the time period showed in Figure 1. Rather, this change seems likely to have been caused by a variation in the husbandry practices. Technology changes such as this one are encompassed in the model by a linear trend.

The environmental and climate variables are also derived on 5km grid squares from the LandIS database, maintained by National Soils Resources Institute[4]. We include: (*a*) temperature as cumulative degree days above 0^{o}C, (*b*) average summer rainfall, (*c*) potential evapotranspiration, (*d*) median number of days of field capacity, (*e*) autumn machinery working days, (*f*) percentages of land within the square belonging to the five land capability classes, distinguishing between land over and below a slope of 6 degrees.

The economic determinants included in the model are crops, livestock (cattle) and milk prices and fertilizer, energy and feeding real price indexes, computed on yearly basis. All have been derived from the records in the Annual Abstract of Statistics (Central Statistical Office, various years) and data available on the Defra web site (http://statistics.defra.gov.uk/esg/). To represent expectations, instead of the actual prices or the price index in year *y* we use the moving average of the price in the three preceding years.

[3] The empirical analysis is undertaken using the statistical software R (R Development Core Team, 2004).

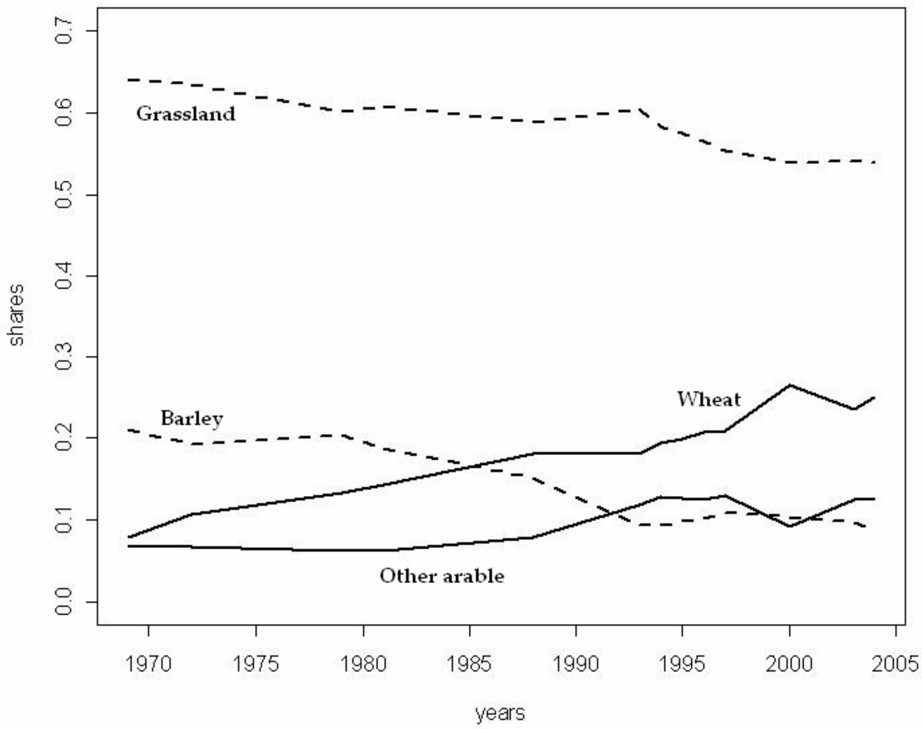


Figure 1: shares of land use in England and Wales

Notes: all shares are referred to the total of land included in the model not to the total available land. Grassland includes both permanent and temporary grassland; other arable includes root crops (potatoes and sugar beet) and other cereals (oats, maize, etc.);

The estimates of the logit model for the first step, i.e. distinguishing between arable and grassland, obtained via Maximum Likelihood (ML), are reported in Table 1. According to the likelihood ratio test, the fit of the model is very satisfactory. Furthermore, all the coefficients have the sign we expect. For example, an increase in the fertilizer and in the energy price indexes decrease the profitability of arable land, which requires more inputs (see, for example, the British Survey of Fertilizer Practice, 2005), whereas an increase in the feed costs decreases the predicted share of grassland. Furthermore, an increase in the prices of arable crops compared to the prices of grassland outputs (represented by the variable prices ratio) increases the share of arable land. Finally, as showed by the negative coefficient of "dummy set", the introduction of set aside reduces the share of arable crops. We point out how the model does not include any lagged dependent variable (e.g. the share of wheat in the each grid square in the previous year). This ensures that the estimated equations are long run, equilibrium relations, and that the ability of the model to predict the land use arising from different policy scenarios is not based on the information embedded in simply observing the past land uses.

Observing the effect of the environmental and climate variables, in the right part of the table, we note how favourable conditions for crop grown (land of better quality and lower slope, more machinery working days, etc.) increase the shares of arable land. On the contrary, an increase in the average summer rainfall has a negative effect on crop land shares. This is a consequence of the high levels of rainfall that characterize the UK.

[4] See http://www.landis.org.uk/gateway/ooi/intro.cfm.

Table 1: First step logit model estimates (reference category: grassland)

Variable	**Parameter**	**z-test**	**Variable**	**Parameter**	**z-test**
Intercept	32.48	26.4	*ln(pot.evo)*	6.10	18.9
			ln(field.cap)	-3.12	-11.9
energy price	-0.53	-2.6	*ln(summer rain)*	-1.38	-8.16
feed price	0.46	6.3	*ln(degrees>0)*	-7.19	-25.8
fertilizer price	-0.22	-1.9	*mach.wk.days*	0.0079	20.0
prices ratio	0.44	4.1	*class 1&2* > 6^0	1.21	1.9
dummy set	-0.13	-2.0	*class 3* > 6^0	3.15	12.7
Trend	0.024	7.7	*class 4* > 6^0	1.45	5.1
% urban	-0.87	-14.8	*class 1&2* < 6^0	3.62	16.2
			class 3 < 6^0	2.97	13.5
Pseudo R^2 (McFadden, 1974): 0.27			*class 4* < 6^0	1.76	7.9
LR test $\chi(19)$: 26055.4 [0.00000]			*class 5* < 6^0	1.21	3.9
Number of observations: 70450					

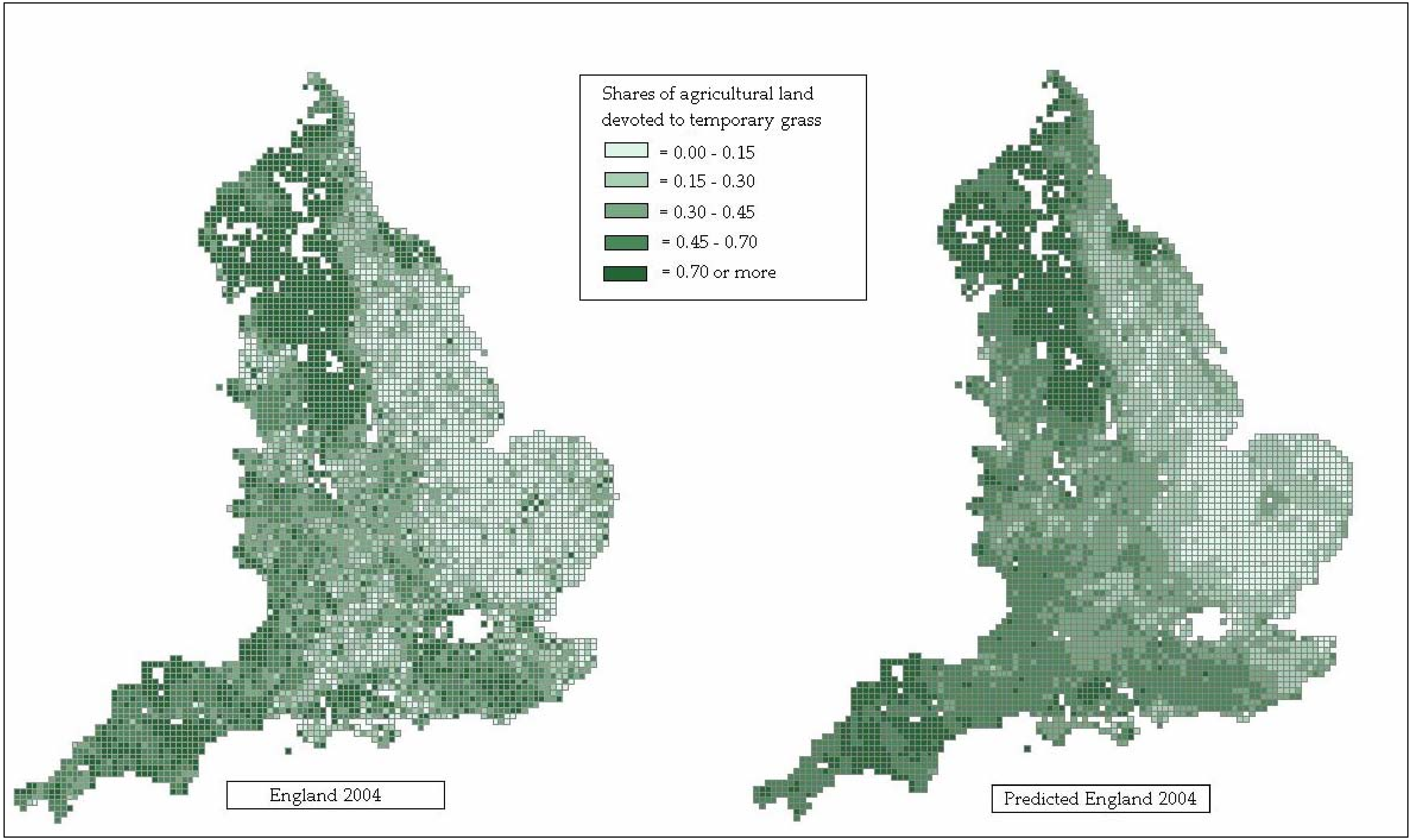
Notes: "energy price", "feed price" and "fertilizer price" indicate the price indexes of these commodities; "prices ratio" = $(p_{wheat} + p_{roots}) / (p_{milk} + p_{cattle})$, an indicator of how profitable is arable cultivation with respect to grassland; dummy set = 1 when year ≥ 1993 and 0 otherwise, "%urban" = percentage of urban area in the square, "pot.evo" = potential evotranspirations, "field.cap" = number of days of field capacity; "summer rain" = average summer rainfall; "degrees>0" = degree days above 0°C; "mach.wk.days"= autumn machinery working days; "class 1&2 >6°" percentage of agricultural land within the square belonging to land capability classes 1&2 and with a slope higher than 6° and the rest of the land classes defined in a congruent way.

The second-stage logit models further subdivide the shares of grassland into permanent and temporary grassland and the shares of arable into wheat, barley, root crops (including sugar beet and potatoes), oilseed rape and other cereals (including oats, maize, rye, etc.). The estimation results are not reported here, but are available under request from the Authors. However, we show the model fitting deriving out-of-sample prediction for England (Wales data are not available for this year) in 2004 and comparing them with the actual land use for the June Agricultural Census for the same year.

Both actual and predicted shares of agricultural land under any given land use can readily be mapped for illustrative purposes. Figures 2 maps the predicted and actual shares of permanent grassland for 2004, while Figure 3 provides the same analysis for wheat. The two-stage logit model performs reasonably well in this forecasting exercise: the predicted shares of agricultural land are fairly close to the actual shares in both figures. However, in particular in the West Midlands, the actual data seem to be much more variable than the predicted, showing grid squares with high shares (of wheat or of grassland) next to grid squares with very low shares. The predictions from our model, instead, appear to be smoother and more uniform. This is not necessary a drawback of the model, but is probably caused by noise present in the rough data, due to the collection and the grid square assignation methodology implemented[5].

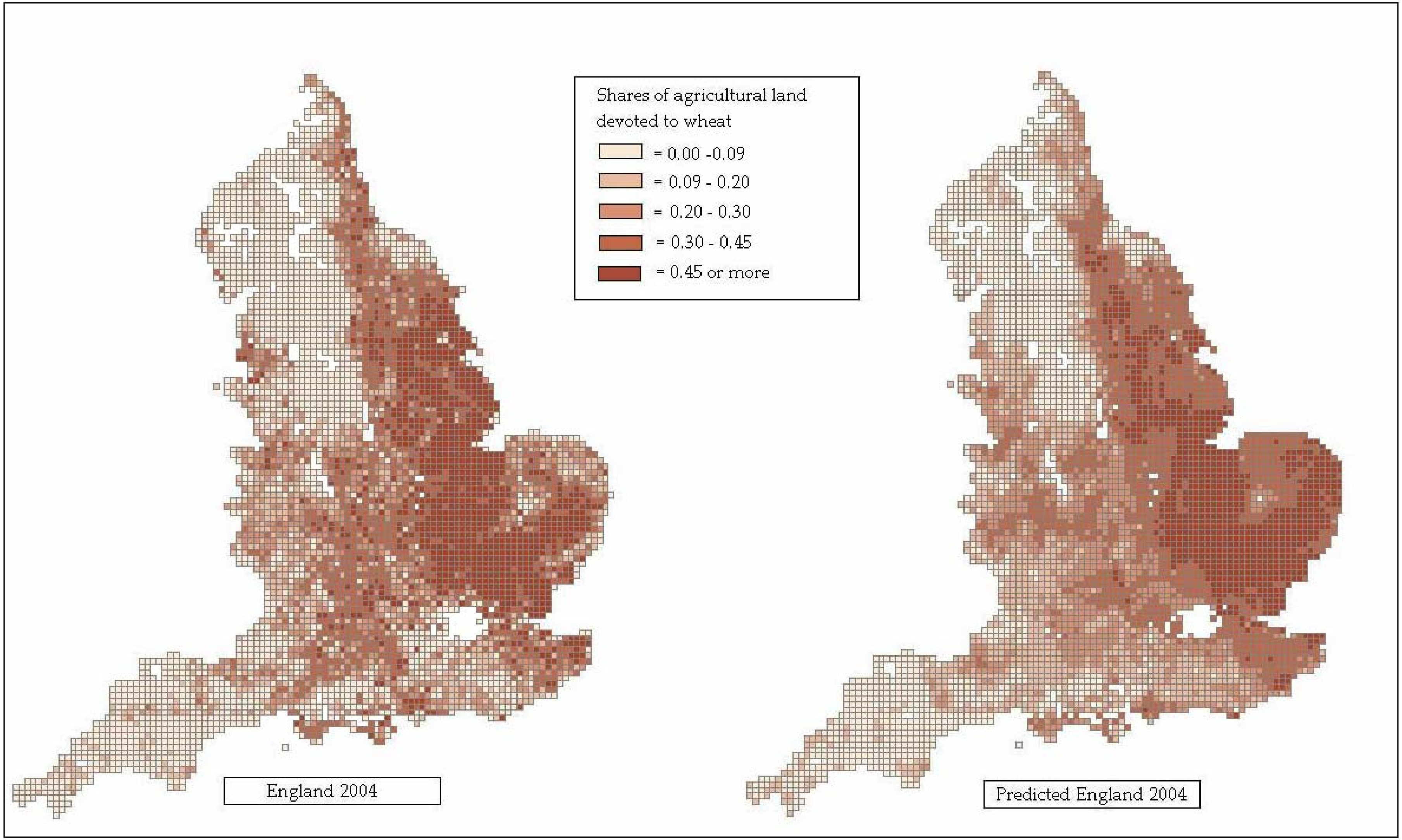
[5] For more information see: http://www.edina.ac.uk/agcensus/description.shtml

Figure 2: Actual shares of permanent grassland (left) and out-of-sample prediction (right); England in year 2004.



source: EDINA (www.edina.ac.uk)

Figure 3: Actual shares of wheat (left) and out-of-sample prediction (right); England in year 2004.



source: EDINA (www.edina.ac.uk)

3.2 Policy effects simulation

The model estimated above provides the basis for evaluating quantitatively the effects of policies targeted to modify agricultural land use. In this analysis, we consider a strategy that could be implemented in order to reduce the amount of nutrients (e.g. nitrates, phosphates, etc.) leaching from farm land into water-bodies, in compliance with the WFD. Therefore, we predict the effect on agricultural land use of a tax that doubles the price of fertilizer compared with the price in England in year 2004. We choose such a major tax increase purely in order to observe and illustrate the significant changes in land use which such an extreme measure would generate. In reality more modest tax increases are still likely to generate substantial responses. Indeed, one of the pollution reduction measures considered in Defra (2007b) is switching arable land to grassland.

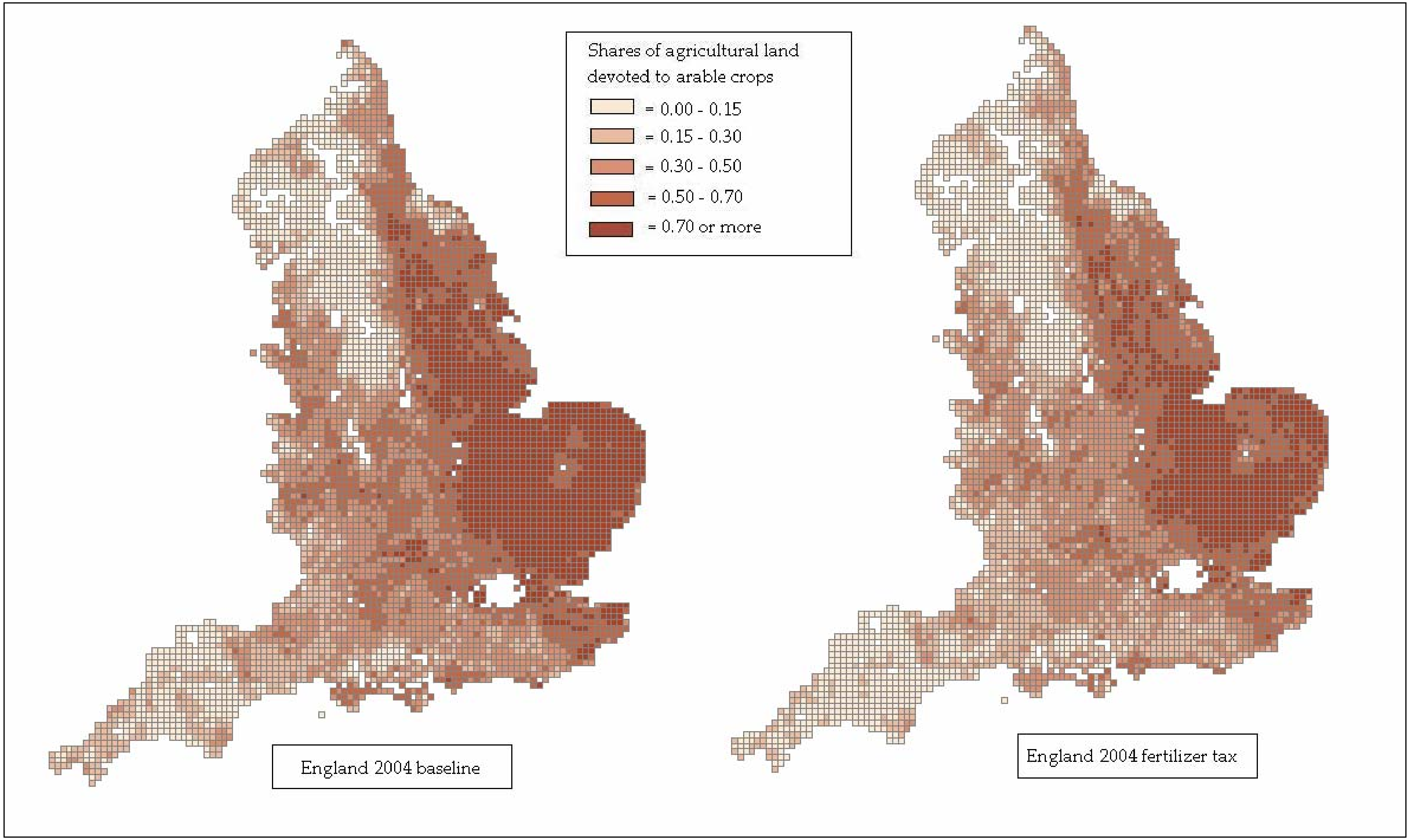
Figure 4 shows the estimated shares of arable land in 2004 before and after the introduction of the fertilizer tax. The resulting reduction of arable land is significant and widespread. Considering that the amount of nitrate and phosphate leaching from grassland is, on average, substantially lower than that leached from arable land these changes should lead to a significant reduction in the in-stream nutrient concentrations.

4. Concluding remarks

In this paper we develop a land use share model for the United Kingdom using an econometric, discrete choice approach estimated on $5km^2$ grid square data. We show that a model including both physical environment and economic factors yields a fairly accurate picture of the land use distribution, even in out-of-sample predictions. We then use our model to simulate the effect of a water pollution reduction policy related to the implementation of the WFD: a substantial tax-driven increase in the price of fertilizer. Our analysis indicates that such a measure would foster major conversion of arable land to grassland across the country, thereby substantially decreasing diffuse nutrient pollution of rivers and waterways.

The limitations of this study have to be acknowledged when assessing the results. Firstly, the empirical framework developed here cannot be used to compute the social welfare impacts of alternative policies, since it cannot translate acreage usage into producer or consumer welfare. Secondly, this analysis is limited to the policy effects at the margin, even though policies will typically affect both land use and land use intensity. For example, the application of a tax on fertilizer would not only encourage land use switching but also reduce the fertilizer application rates. This second aspect cannot be encompassed in the present approach. However, we are currently investigating options to extend our approach and overcome both shortcomings.

Figure 4: Predicted shares of arable land in the baseline (left) and after a 100% increase in the fertilizer price (right); England in year 2004.



References

Central Statistical Office (various years) *Annual Abstract of Statistics*, Her Majesty's Stationery Office, London.

Chomitz K.M., Gray D.A. (1996) Roads, land use, and deforestation: a spatial model applied to Belize, *World Bank Economic Review*, vol. 10, pp. 487-512.

Chomitz K.M., Thomas T. (2003) Determinants of land use in Amazonia: a fine scale spatial analysis, *American Journal of Agricultural Economics,* vol. 85, pp. 1016-1028.

Defra (2005) *British survey of fertiliser practice, fertiliser use on farm crops for crop year 2004*, Defra and Seerad report.

Defra (2007a) *The protection of water against diffuse water pollution from agriculture, consultation on diffuse sources in England*, Defra report.

Defra (2007b) *An inventory of methods to control diffuse water pollution from agriculture*, Report prepared by ADAS and IGER.

European Commission (200) *Directive 2000/60/EC (Water Framework Directive)*, Official Journal, 22 December 2000.

Fezzi C., Rigby D., Bateman I.J., Hadley D., Posen P. (2007) *Estimating the range of impacts arising from nitrate leaching reduction policies using farm accounts*, CSERGE Working Paper, ECM 07-02.

Lubowski R.N., Plantinga A., Stavins R. (2006) Land-use change and carbon sinks: econometric estimation of the carbon sequestration supply function, *Journal of Environmental Economics and Management*, vol. 51, pp. 131-152.

McFadden D. (1974) *Conditional logit analysis of qualitative choice behavior*, in Zarembka P. (ed.), Frontiers in Econometrics, Academic Press.

R Development Core Team (2004) *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, http://www.Rproject.org.

Wu J., Segerson K. (1995) The impact of policies and land characteristics on potential groundwater pollution in Wisconsin, *American Journal of Agricultural Economics*, vol. 77, pp. 1033-1047.