

Fezzi, Carlo; Bateman, Ian

Working Paper

Non-linear effects and aggregation bias in Ricardian models of climate change

CSERGE Working Paper, No. 2012-02

Provided in Cooperation with:

The Centre for Social and Economic Research on the Global Environment (CSERGE), University of East Anglia

Suggested Citation: Fezzi, Carlo; Bateman, Ian (2012) : Non-linear effects and aggregation bias in Ricardian models of climate change, CSERGE Working Paper, No. 2012-02, University of East Anglia, The Centre for Social and Economic Research on the Global Environment (CSERGE), Norwich

This Version is available at:

<https://hdl.handle.net/10419/121953>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

2012
02



Non-linear effects and aggregation bias in Ricardian models of climate change

Carlo Fezzi and Ian Bateman

CSERGE Working Paper 2012-02



CSERGE

Centre for Social and Economic Research on the Global Environment

NON-LINEAR EFFECTS AND AGGREGATION BIAS IN RICARDIAN MODELS OF CLIMATE CHANGE

Carlo Fezzi and Ian Bateman

**School of Environmental Sciences,
University of East Anglia,
Norwich Research Park
Norwich NR4 7TJ, UK**

c.fezzi@uea.ac.uk

CSERGE Working Papers

ISSN 0967-8875

CSERGE

(Centre for Social and Economic Research on the Global Environment),

**School of Environmental Sciences,
University of East Anglia,
Norwich Research Park
Norwich NR4 7TJ,
UK**

www.cserge.ac.uk

Abstract

Ricardian models predicting the impact of climate change on agriculture are typically estimated on data aggregated across counties and assuming additively separable effects of temperature and precipitation. We investigate the potential bias induced by such assumptions by using a large panel of farm-level data and estimating a semi-parametric specification. Consistent with the agronomic literature, we observe significant non-linear interaction effects, with more abundant precipitation being a mitigating factor for heat stress. This interaction disappears when the same data is aggregated in the conventional manner, leading to predictions of climate change impacts which are severely distorted.

Key words: Climate Change, Agriculture, Ricardian Analysis, Aggregation Bias, Semi-parametric models.

Acknowledgments

We are grateful to Silvia Ferrini, Luisa Menapace, Marije Schaafsma, Wolfram Schlenker and the participants of the 4th WCERE and the 12th Occasional Workshop on Environmental and Resource Economics at UCSD for their helpful comments. Many thanks to David Hadley for his assistance on the farm survey data. This research was supported by the SEER project, which is funded by the ESRC (reference RES-060-25-0063).

1. INTRODUCTION

There is a growing literature on the potential influence of changing climatic conditions on agriculture (see, for example, Intergovernmental Panel on Climate Change, IPCC, 2007) as variation in temperature and precipitation significantly affect crop and livestock production. Different approaches have been used to model the impact of climate change on agriculture (see Mendelsohn and Dinar, 2009, chap. 3, for a review). Early work focused on crop-growth simulation models (e.g. Adams et al., 1990; Kaufmann and Snell, 1997). Such approaches only included limited behavioral responses in farmers' adaptation to climate change and, therefore, risked over-estimating negative impacts. More recent studies (e.g. Deschênes and Greenstone, 2007; Schlenker and Roberts, 2009; Lobell, Schlenker and Costa-Roberts, 2011) estimated the effect of climate on crop yield or farm profits by fitting statistical or econometric models to time series, cross-section, or panel data. There are strengths and weaknesses in each of these approaches: for example, while time series or panel models can include location-specific fixed effects to absorb possible time-invariant omitted variables, the identification rests on random year-to-year weather shocks, which are different from permanent shifts in climate (Fisher et al., 2011).

Among the econometric approaches, the *Ricardian* (or *hedonic*) *method*, introduced by Mendelsohn et al. (1994), has gained considerable prominence. In recent years, this approach has been applied to various countries across the globe, including the United States (e.g. Schlenker et al., 2005; Schlenker, Hanemann and Fisher, 2006, SHF hereafter), Europe (Lang, 2007), Africa (Seo et al., 2009), Brazil (Timmins, 2006), and Latin America (Seo and Mendelsohn, 2008). The Ricardian method is based on the notion that, in a competitive market, the price of farmland reflects the discounted value of all the expected future profits that can be derived from it (Ricardo, 1817). By regressing land values on climatic determinants and a set of exogenous control variables, this technique estimates the impact of climate on farmers' expected incomes by relying on the cross-sectional variation observed in current climate. This model is most commonly estimated using cross-sectional data as climate has not changed enough over time to allow the identification of its effect in any given location. The major advantage of the Ricardian approach is that it automatically captures adaptation, since farmers are adjusting inputs and outputs to match local conditions. Two major drawbacks are (a) potential omitted variables (although this issue is common to all cross-sectional analyses), and (b) the implicit assumption of fixed prices. While there is no direct test for omitted variable bias, the robustness of the observed relationship across years and settings makes it less likely to be a serious concern. Fixed prices are a common assumption in most statistical studies that use a partial-equilibrium setting where agricultural commodities' prices are set at the global level.

Ricardian models have typically been implemented using aggregate data averaged over counties or large regions (Mendelsohn et al., 1994; Schlenker et al., 2005; SHF; Lang 2008; Seo and Mendelsohn, 2008; Seo et al., 2009). To our knowledge there are only two notable exceptions: the recent studies on farm revenues by Mendelsohn and Dinar (2009) and the analysis by Schlenker et al. (2007) on California, both based on farm-level data. Unfortunately, using aggregated data may conceal individual heterogeneity and non-linear effects and, ultimately, result in biased estimates and conclusions.

Aggregation bias is a long-standing issue in econometrics, recognized since the seminal works by Theil (1954), Grunfeld and Griliches (1960), and Feige and Watts (1972). More recent papers show that aggregation bias can explain the purchasing power parity puzzle (Imbs et al., 2005), anomalies in the factor content of trade (Feenstra and Hanson, 2000) and the excess of consumption in the life-cycle permanent-income model (Goodfriend, 1992). This bias is particularly severe for the estimation of non-linear relations, which normally are not robust to the aggregation process (Stocker, 1984, 1986; Garderen, Lee and Pesaran, 2000). Nevertheless, previous Ricardian analyses have found strong non-linear climatic effects using aggregated information. It is therefore crucial to test whether these estimates are robust and generate un-biased climate change impacts predictions.

In addition, previous Ricardian studies have assumed the effects of precipitation and temperature to be additively separable and to follow simple parametric forms, most commonly quadratic specifications. However, additive effects are at odds with plant physiology since increased heat generates higher demand of water for crop development (e.g. Monteith, 1977; Morison, 1996). Changes in production directly translate into changes in sales and profits. Since farmland values are the discounted value of future profits, any factor that impacts crop yields should also impact farmland values. It is unclear why previous Ricardian studies did not document any interaction effect. A related issue is that restricting the climatic effects to have a specific functional form, such as the quadratic, can be difficult to justify on theoretical grounds. In fact, the lack of theoretical underpinning for parametric specifications is a recognized issue in most hedonic models (e.g. Cropper et al., 1987; Ekeland et al., 2002, 2004) and the advantages offered by semi- and non-parametric alternatives have been demonstrated in several empirical analyses (Anglin and Gencay, 1996; Parmeter et al., 2007; Bontemps et al., 2008).

In light of these issues, this paper makes three methodological contributions to the literature. First, we investigate whether aggregation bias is present in Ricardian models estimated on county-level data. We employ a unique farm-level panel of land values that allows us to examine environmental and climatic determinants at a much more refined spatial resolution than any previous work. This allows us to contrast a farm-level analysis with the results obtained by aggregating the same data up to the county-level, in order to replicate the approaches implemented by earlier studies. Such comparison reveals a strong aggregation bias with severe implications for predictions: on average, climate change impacts estimated on aggregated data differ by a factor of four compared with those derived on farm-level information.

Second, our farm-level analysis reconciles the Ricardian approach with the agronomic literature by confirming a significant interaction effect: precipitation is more valuable when temperatures are high. Similarly, temperature has a positive effect on land value only if there is enough precipitation to prevent possible droughts. We only observe the interaction effect in our farm-level data set, again highlighting the importance of using micro-level data.

Third, we test for functional form miss-specification arising from omitted non-linear effects by relaxing the parametric assumptions implemented in previous Ricardian

studies and estimating a semi-parametric model. We specify the hedonic farmland value function as an Additive Mixed Model (AMM) with penalized splines. Compared to parametric models, the AMM provides a superior fit and reveals an even stronger interaction between rainfall and temperature. However, in the present application the overall climate change impact predictions do not differ significantly from those resulting from a simpler, parametric regression.¹

Our micro-level data set covers farms in Great Britain (GB). While GB is smaller than the spatial extent of other Ricardian analyses, its geographic position surrounded in the south by the Gulf Stream and in the north by sub-Arctic waters generates a diversity of micro-climates yielding a wide range of variation in temperature and precipitation. Obviously, variation in climate is necessary to obtain precise estimates. Focusing on a narrow spatial scale that has significant variation in climate is even preferable as other, potentially confounding, variables are more homogenous. In the extreme, having two identical farms that only differ in climate would be the best-case scenario for identifying the effect of climate on land values. We therefore see the limited spatial scale as an advantage rather than a burden.

The rest of the paper is organized as follows. Section 2 introduces the micro-level data set of individual farm values in GB. Section 3 briefly reviews the Ricardian approach and outlines possible specifications for linking climate with land values. It also describes our strategy to test for aggregation bias and omitted non-linear relations. Section 4 contrasts the empirical estimates of different Ricardian models, including the traditional, parametric county-level model and the parametric and semi-parametric models estimated on farm data. Section 5 compares the same models in terms of climate change impact predictions, using the recently published UK Climate Impacts Programme (UKCIP09, 2009) global warming scenarios. Section 6 concludes.

2. THE DATA

This analysis employs a unique database which covers the whole of Great Britain and integrates multiple sources of information expressed at different spatial resolutions. These are detailed throughout the remainder of this section.

Land value data. Data on land value are derived from the Farm Business Survey and the Scottish Farm Accounts Survey panels which, sampled annually, include information on the physical characteristics and economic performance of farm businesses throughout Great Britain. Farms are retained in the sample for several years, with only 10% of them being replaced in each survey. The two Farm Surveys (FS) include a specific figure for

¹ A common criticism of semi-parametric approaches is that, while they require fewer assumptions than their parametric counterparts, they too easily display idiosyncratic local variations which can lead to overfitting. The penalized splines we implement in this paper address this issue by augmenting the likelihood of the model with penalties proportional to the non-linearity of the estimated functional forms. This resolves the problem of determining model flexibility *a priori* by incorporating this task within the estimation process. The penalty automatically suppresses non-linearities which are not supported by the data (Ruppert, Wand and Carroll, 2003; and Wood, 2006a,b).

land value, which excludes buildings and other improvements used for agriculture (it includes, however, the value of buildings and dwellings older than 30 years) and reflects the expected sale value. This measure is based on information on agricultural land sales taking place within the area where the farm is located and on an assessment of the property provided by professional land agents. Since these estimates are not revised each year, we discard from the analysis all the records in which the farmland value stays constant from one year to the next, interpreting this as a year in which the value has not been updated.²

The database also contains the location of the farm on a 10km grid square basis, which we use to link farm value to environmental and climatic characteristics. In this analysis we consider 10 years of FS data, from 1999 to 2008, consisting of approximately 2500 farm records each year. After eliminating outliers, farms with no owned land and farms for which the land value or the location are missing, about 9500 observations remain for the analysis.³

Climatic variables. Temperature and precipitation are represented by the 5km grid cell data available from the UK Meteorological Office archive (<http://www.metoffice.gov.uk/>). As per SHF, for each observation climatic variables are calculated as averages over the 30 preceding years (e.g. the climate record for 2004 is calculated from the weather during the period 1974-2003). Temperature has been included in Ricardian models as monthly or seasonal averages (Mendelsohn et al., 1994; Seo et al., 2009) or as the number of degree days in the growing season (SHF). Degree days are defined as the sum of degrees between two temperature thresholds. This concept is derived from the agronomic literature and reflects the fact that plant growth is linear in temperature only within a certain range, with temperatures below this interval being irrelevant for crops development, while temperatures above that threshold being potentially harmful.⁴ SHF show how this strategy is superior to including monthly averages, mainly because temperatures in different months can be highly collinear. As in previous contributions, we consider only the degree days during the main growing season, defined for GB as the months from April to September. To derive degree days, we use the common assumption that, during the day, temperature ($temp$) follows a sinusoidal function (Schlenker and Roberts, 2006):

$$temp = 0.5[temp_{max} - temp_{min}] \sin(\chi) + temp_{min} + 0.5[temp_{max} - temp_{min}],$$

Where χ is defined between $-1/2\pi$ and $3/2\pi$ and $temp_{min}$ ($temp_{max}$) is the minimum (maximum) temperature within the day. Since data on the minimum and maximum temperature in each day of the year are not available, we use average monthly minimum

² As an example, consider a farm which is surveyed in the years 2000-2005. The farmland value is estimated to be £1000 per hectare in year 2000, £1500 in 2001, 2002 and 2003 and £1800 in 2004 and 2005. In this case we include in the analysis only the records relative to years 2000, 2001 and 2004.

³ We define as outliers all farms with agricultural area smaller than 30ha, some very large upland farms with extremely low value per hectare and a few farms with a high variance inflation factor which introduce instability in the parameter estimates.

⁴ This is, not surprisingly, just an approximation. Recent literature (e.g. Schlenker and Roberts, 2006) shows how the effect of temperature on yield can present non-linearities even within the two thresholds. However, since the objective of a Ricardian analysis is not to analyze crop growth but to understand the effect of climate on land value, the linearity assumption to compute degree days still constitutes a reasonable approximation.

and maximum temperature to compute the number of degree days. This approximation works well when the variability of the minimum and maximum temperature in each month is not very high, as it is the case in GB, which is characterized by a relatively mild climate. Taking into account the characteristics of GB agriculture, we define the lower and upper threshold as 5.5°C and 32°C respectively. In our sample, the average monthly maximum temperature never surpasses 30°C (the highest average maximum temperature recorded in the sample is 28.24°C) and, therefore, this latter threshold is not relevant for our study. Finally, considering rainfall, we include the total precipitation in the growing season. We then aggregate degree days and precipitation data from 5km to 10km grid squares to match the resolution of the farm location data.

Environmental variables and other determinants. Besides climate, several other variables can significantly influence farmland values. Considering soil characteristics, we include soil texture as the share of fine (clay share between 35% and 60%), medium fine (clay < 35% and sand < 15%), medium (clay between 18% and 35% and sand >15% or clay between 18% and 35% and sand < 65%), coarse (clay < 18% and sand > 65%) and peaty soils, and the depth to rock. These are derived from the 1km grid square data in the European Soil Database (ESDB) maintained by the European soil data centre (<http://eusoils.jrc.ec.europa.eu/>). We also include average slope (derived from the Ordnance Survey, Digital Terrain Model, available at: <http://www.ordnancesurvey.co.uk/oswebsite/>), representing the suitability of land for machinery operations, and population density (computed from 1990 and 2000 census data, <http://casweb.mimas.ac.uk/>), to capture the opportunity value of converting land to residential use, distance to markets and the availability of amenities or off-farm work for the members of the farmer's family. Finally, to capture the impact of location-specific policies we include the share of each 10km grid square classified as National Park, Nitrate Vulnerable Zone (NVZ) or Environmentally Sensitive Areas (ESA) in each year. NVZs, established in 1996 and extended in 2003 and 2008 to now cover more than 70% of English farmland, are designed to reduce surface and groundwater nitrate contamination, and impose some restrictions on the agricultural activities of the farms within their boundaries (e.g. limiting the amount of fertilizer to be used on fields, regulating the storage of organic manure, etc.). However, they do not go beyond good farming practices and, therefore, while being mandatory and uncompensated, might not significantly affect land values. ESAs, introduced in 1987 and extended in subsequent years, are intended to safeguard and enhance areas of high landscape, wildlife or historic value. Unlike NVZs, participation in ESA schemes is voluntary and farmers receive monetary compensation for engaging in environmentally friendly farming practices, such as converting arable land to permanent grassland, establishing hedgerows, etc.

TABLE 1: Descriptive statistics

	symbol	units	\bar{x}	$\hat{s}(x)$	min	max
<i>land value</i>	V	1000£/ha	7.18	5.21	0.02	192.00
<i>degree days</i>	dd	°C	1339.00	184.48	735.50	1673.00
<i>precipitation</i>	$prec$	mm	385.20	98.22	244.40	927.30
<i>soil class</i>						
- <i>fine</i>	s_f	%	15.00	28.04	0.00	100.00
- <i>medium fine</i>	s_{mf}	%	8.66	21.56	0.00	100.00
- <i>medium</i>	s_m	%	58.72	38.77	0.00	100.00
- <i>coarse</i>	s_c	%	12.20	21.52	0.00	100.00
- <i>peat</i>	s_p	%	5.41	15.97	0.00	100.00
<i>depth to rock</i>	dtr	dm	7.42	3.26	0.00	14.00
<i>slope</i>	slo	°	3.19	2.13	0.00	16.20
<i>pop. density</i>	$popd$	pop/km ²	204.60	258.89	7.87	2896.0
<i>income</i>	inc	1000£/head	14.07	1.60	10.09	20.96
<i>park share</i>	s_{park}	%	7.03	22.31	0.00	100.00
<i>ESA share</i>	s_{esa}	%	11.83	26.21	0.00	100.00
<i>NVZ share</i>	s_{nvz}	%	35.34	43.20	0.00	100.00

Notes: \bar{x} indicates the sample mean, $\hat{s}(x)$ the sample standard deviation. Statistics refers to the observations in our FS sample, not to Great Britain as a whole. Land value and income deflated using the GDP deflator with 2008 as the baseline year (source: HM Treasury, http://www.hmtreasury.gov.uk/Economic_Data_and_Tools/GDP_Deflators/data_gdp_index.cfm, accessed on the 16th June 2010). The total number of observations is 9438 (3266 farms in located in 1350 different 10 km grid cells).

Descriptive statistics for the variables included in the analysis are reported in Table 1. As expected, the distribution of farmland values appears to be highly skewed with a long right tail, which could support a log-normal distribution. Considering rainfall levels, although GB covers a relatively modest area compared to those analyzed in other Ricardian studies (e.g. US, Brazil), its location between the warm waters of the Gulf Stream to the West and the cold climates of Scandinavia to the East means that it exhibits a wide range of precipitations across the growing season (from 244 to 947 mm), this range actually exceeding that reported for the rainfed US counties (from 332 to 982mm) analyzed by SHF. Indeed, accounting for irrigation (Schlenker et al., 2005) is necessary in the UK, since most farmland is rainfed (less than 1% of the farmland in our sample is actually irrigated). Furthermore, temperatures rarely reach particularly high values, with the maximum number of degree days per annum being well below 2400, which coincidentally is the optimal value for crop growth indentified by SHF's estimates.

Aggregation. To investigate the impact of aggregation on parameter estimates and climate change predictions, we analyze the same data both at the farm- and at the county-level. If results obtained from the two datasets differ significantly, then there is a clear indication of aggregation bias. In order to obtain units of aggregation comparable with those examined in previous Ricardian studies, we aggregate our farm-level values using the official Nomenclature of Territorial Units for Statistics (NUTS) of third level, as defined by the European Union, which roughly correspond to GB counties. Their size varies considerably, ranging from about 40 km² to more than 10000 km², with an average of 1814 km² (for comparison, the average area of US counties analyzed in SHF is about 3000 km²). We assign each farm to a county based on its spatial location and compute the aggregated values as simple averages. This leads to just over 10 farm

records for each county in each year, and a total of 857 aggregated observations. The next section compares the parameter estimates obtained from the farm-level and county-level models and tests for aggregation bias. Section 5 contrasts their climate change impact's forecasts.

3. METHODOLOGY

3.1 The Ricardian model

The Ricardian approach assumes that each farmer allocates land among different activities in order to maximize net revenues. Consequently, in a competitive market, farmland price equals to the expected present value of the future stream of income derived from land. Therefore, land value per hectare $V_{i,t}$ of farm i in year t can be written as:

$$(1) V_{i,t} = \int_t^{\infty} R_{i,\tau} e^{-\phi\tau} d\tau = \int_t^{\infty} \left[\sum_k p_{k,\tau} q_{k,i,\tau}(\mathbf{z}_{i,\tau}, \mathbf{i}_{i,\tau}, \mathbf{g}_{i,\tau}) - \sum_j r_{j,\tau} i_{j,\tau} \right] e^{-\phi\tau} d\tau,$$

$R_{i,\tau}$ denotes the stream of *expected* net revenues for farm i at time τ which is given by the difference between expected revenues and costs. Revenues are given by the product of expected market prices $p_{k,\tau}$ for each agricultural output k and expected output $q_{k,i,\tau}$. This output is a function of expected weather (climate) $\mathbf{z}_{i,\tau}$, expected input choices $\mathbf{i}_{i,\tau}$ and other environmental and policy determinants $\mathbf{g}_{i,\tau}$ (e.g. soil quality, subsidies, etc.). Costs are the product of expected input prices $r_{j,\tau}$ and input quantities. Finally, the term ϕ indicates the constant discount rate.

We assume that farms are atomistic, and input demand is small enough to not influence input prices. By the same token, idiosyncratic weather shocks do not influence the exogenous output prices since the quantity produced in the UK is not large enough to affect the global market. The Ricardian approach does not model farmers' land allocations, input and output choices explicitly, but rather estimates the overall value of each land characteristic by specifying the hedonic, reduced form model:

$$(2) V_i = f(\mathbf{p}, \mathbf{r}, \mathbf{z}, \mathbf{g}),$$

where $f(\cdot)$ is a functional form unknown *a priori*. As in most hedonic models, economic theory provides little guidance on the shape of this relation, which, while arguably non-linear, remains an open empirical question.

3.2 Testing for aggregation bias and omitted non-linearities

Virtually all Ricardian analyses have translated equation (2) into an empirically tractable model by assuming a linear or semi-log specification with a quadratic formulation for the climatic variables (here degree days, dd , and precipitation, $prec$) and a linear function for all other determinants. Findings reported by SHF suggest that a log-

transformation of the dependent variable outperforms a linear specification, since the distribution of land values is non-negative and typically highly skewed. Estimation is normally implemented on data aggregated over counties or larger regions. As a starting point, we open our analysis by replicating such a model, which has implemented in the majority of Ricardian studies so far. The hedonic equation (*Model A*) is specified as:

$$(3) \ln V_{c,t} = \beta_0 + \beta_1 \text{prec}_{c,t} + \beta_2 \text{dd}_{c,t} + \beta_3 \text{dd}_{c,t}^2 + \beta_4 \text{prec}_{c,t}^2 + \boldsymbol{\gamma}' \mathbf{g}_{c,t} + u_{c,t},$$

where c indicates the county, t indicates the time at which expectations are taken, β_0, \dots, β_4 and $\boldsymbol{\gamma}$ are the county-level parameters to be estimated and $u_{c,t}$ is a residual component which we define as being the sum of a county-specific random effect and a residual, both normally distributed and uncorrelated ($u_{c,t} = \alpha_c + \varepsilon_{c,t}$). The vector $\mathbf{g}_{c,t}$ includes population density ($dpop$ and $dpop^2$, see SFH), depth to rock (dtr), slope, soil texture shares, National Park (s_{park}), ESA (s_{esa}) and NVZ (s_{nvz}) shares, regional fixed effects (for England, Wales and Scotland) and yearly fixed effects.

This quadratic approximation with additively separable climatic effect (on the logarithmic scale) has been implemented in most applications (see Mendelsohn and Dinar, 2009, for a review) because it allows the identification of “optimal” crop growing conditions while maintaining simplicity in estimation. In this specification, climatic effects are multiplicative. For example, the marginal effect of precipitation is:

$$(4) \frac{\partial V_c}{\partial \text{prec}_c} = V_c (\beta_2 + 2\beta_4 \text{prec}_c).$$

This effect, therefore, depends on all the variables which determine the land value V_c . However, this formulation does not encompass all interactions among climatic variables. In fact, the sign of the marginal effect (4) depends solely on the term $\beta_2 + 2\beta_4 \text{prec}_c$, which contains only the parameters of precipitation itself, and none of those relating to other variables. This means that the optimal amount of rainfall will not depend on the level of temperature, and *vice versa*. This constraint might not necessarily be valid. For instance, agronomic experiments have shown that warmer conditions typically lead to an increase in crop requirements for water (e.g. Morison, 1996).

The simplest approach to relax the assumption of additively separable climatic effects is to include an interaction term to equation (3) to allow the effect of precipitation and temperature to be mutually dependent. Therefore, we estimate our second specification (*Model B*) as:

$$(5) \ln V_{c,t} = \beta_0 + \beta_1 \text{prec}_{c,t} + \beta_2 \text{dd}_{c,t} + \beta_3 \text{prec}_{c,t}^2 + \beta_4 \text{dd}_{c,t}^2 + \beta_5 \text{prec}_{c,t} \text{dd}_{c,t} + \boldsymbol{\gamma}' \mathbf{g}_{c,t} + u_{c,t},$$

Investigating the null hypothesis of $\beta_5 = 0$ allows to test whether county-level data show significant non-linear climatic interactions.

A further issue is how the use of aggregated data affects parameter estimates and climate change impacts predictions. There are good theoretical reasons to believe that even simple non-linear relations, such as those represented by equations (3) and (5),

are not robust to the aggregation process. Stocker (1984, 1986), for instance, shows that even the parameters of the simple quadratic or logarithmic models, when estimated on aggregated data, will be a non-linear combination of the micro-level coefficients and of the parameters of the distributions of the exogenous variables. Therefore, even under very stringent conditions, recovering the farm-level parameters using a county-level regression would require the inclusion of squares and cross-products of the explanatory variables and very complex functional forms (Van Garderen, Lee and Pesaran, 2000, provide a few examples). Given that most Ricardian analysis have been implemented on aggregated data, it is important to investigate the size of any bias inherent in such approaches and its implications climate change impacts predictions.

The simplest strategy to test for aggregation bias it to re-estimate model (5) using the same data, but disaggregated at the farm-level, testing whether parameters and climate change impacts predictions are significantly different from those obtained with county-level data. The resulting specification (*Model C*) can be written as:

$$(6) \ln V_{i,j,t} = \alpha_0 + \alpha_1 prec_{i,j,t} + \alpha_2 dd_{i,j,t} + \alpha_3 prec_{i,j,t}^2 + \alpha_4 dd_{i,j,t}^2 + \alpha_5 prec_{i,j,t} dd_{i,j,t} + \xi' \mathbf{g}_{i,j,t} + u_{i,j,t},$$

where i indicates the farm, j the 10km grid square, $\alpha_0, \dots, \alpha_5$ and ξ are the farm-level parameters to be estimated and all other symbols are defined as previously. We specify the residual component to include both a farm- and a cell-specific random effect ($u_{i,j,t} = w_j + \alpha_{ij} + \varepsilon_{i,j,t}$), to take into account that farms located within the same area may share common un-modelled factors which may significantly affect their land value. This is also a simple approach to account for spatial auto-correlation by allowing the residuals of the farms located within the same cell to be correlated with each other.⁵

A drawback of this last approach (and of any other strict parameterization) is that it constrains the effects to assume very specific functional forms. In such model, climatic impacts are forced to be quadratic and their interaction is assumed to be linear. While these could be reasonable approximations, there is no theoretical justification underpinning such a rigid structure, which is mainly adopted for ease of interpretation and estimation. When these restrictions are invalid, however, they could significantly bias welfare estimates and predictions (Cropper et al., 1987, and Anglin and Gencay, 1996, among others, illustrate this issue in standard hedonic models). Therefore, it is worth investigating possible functional form miss-specification by using a more flexible model and comparing the results with those obtained employing model (6). Here we choose to represent the relationships of interest via smooth functions, deriving an Additive Mixed Model (AMM) which is our most general and last specification (*Model D*):

$$(7) \ln V_{i,t} = f(z_{1,t}, z_{2,t}) + s_1(g_{1i,t}) + \dots + s_h(g_{hi,t}) + u_{i,t}.$$

⁵ Including a more general form of residual spatial autocorrelation (e.g. $\varepsilon_{i,t} = \rho W_\varepsilon + \varepsilon_{i,t}$, with ρ = spatial correlation parameter, W_ε = spatial weight matrix, $\varepsilon_{i,t}$ = i.i.d. error term, see SFH and Maddison, 2009) does not change significantly the parameter estimates of Model C. However, the size of our dataset makes this approach computationally too demanding for the estimation of the semi-parametric specification (Model D). Therefore, for ease of comparison we present the simplest specification with farm- and cell-specific random effects for both models.

In this model the joint effects of temperature and precipitation are encompassed by a multidimensional smooth function $f(\cdot)$, which allows the estimation of flexible non-linear relationships and interaction effects. The other exogenous drivers are also included via smooth functions, $s_1(\cdot), \dots, s_h(\cdot)$, to capture possible non-linear relations. However, to maintain simplicity in the interpretation and avoid the well-known curse of dimensionality, their effects are assumed to be additively separable. The marginal effects of rainfall can be derived as:

$$(8) \quad \frac{\partial V_i}{\partial prec_i} = V_i \frac{\partial f(prec_i, dd_i)}{\partial prec_i}.$$

Note that the sign of this marginal effect is a function of both precipitation and temperature. As a result, the model encompasses, in a flexible form, the interaction effects amongst all climatic factors. Furthermore, as in the parametric specification, the size of the marginal effects depends on all the exogenous variables. This is a very general specification and encompasses Model C as a special case. Comparing the estimates from the two models allow us to test for omitted non-linear relations.

4. EMPIRICAL ESTIMATION AND RESULTS

We estimate models (3), (5), (6) and (7) via Maximum Likelihood (ML) using the *R* software (R development core team, 2008). Model A, B and C are standard linear regression with random effects, and are estimated via the package *nlme* (Pinheiro and Bates, 2000). Model D is implemented by representing the smooth functions as natural cubic splines, which fit third degree polynomial functions between a set of knots located between the range of values of each explanatory variable. The number and the location of the knots effectively determine the flexibility of each smooth function. Given a fixed number of knots, the model can be estimated as a standard regression, i.e. by ordinary least squares or ML. In practice, however, there is a trade-off between sufficient knots to accurately represent any unknown, non-linear relation and, at the same time, avoid the risk of overfitting. This is a common problem in semi-parametric approaches. A practical solution is penalized estimation (Ruppert, Wand and Carroll, 2003; Wood, 2006b). The idea here is to augment the likelihood by including a penalty for the excessive roughness (typically indicated with the term ‘wiggleness’) of the smooth functions, which can be expressed as a function of the integral of the square of its second derivative. Various techniques have been proposed to estimate the optimal amount of smoothing directly from the data (see Wood, 2006b, and Keele, 2009). In this paper we follow the approach illustrated by Ruppert, Wand and Carroll (2003), who suggest representing the smoothing splines as random effects. We then estimate the model via ML using the package *mgcv* (Wood, 2006b). The appealing feature of this technique is that it automatically reduces the smooth functions of the variables for which the optimal fit does not include any non-linearity to standard linear forms. In the extreme, a model in which none of the non-linear relationships are supported by the data will be reduced directly to a standard linear regression during estimation. The optimal level of non-linearity of each smooth function is indicated by the ‘Effective Degrees of Freedom’ (EDF). The higher the EDF, the more non-linear is the estimated function. An EDF equal

to 1 suggests that the best smooth function representation is linear. More details on the splines representation and the estimation technique are given in the Appendix I of this paper.

The parameter estimates and diagnostics of the four models are presented in Table 2. The first column reports the coefficients of Model A: the standard Ricardian regression based on county-averaged data. The estimated effect of degree days is positive and the effect of precipitation is negative, with both quadratic terms being insignificantly different from zero. This is not entirely surprising, given the relatively wet and cold conditions which characterize GB. The coefficients of the control variables have intuitive signs. The effect of population density is positive and highly significant, revealing an increase in the opportunity value of switching to non-agricultural uses, and the proximity of off-farm work, shops and other services. Similarly, better terrain (lower slope and deeper soils) translates into higher land values. The coefficient corresponding to the share of the county within a national park is also positive and significant, reflecting that farms located within these areas have access to specific subsidies aimed at conserving wildlife and promoting environmentally friendly practices. Finally, while not reported in the table to preserve space, the yearly fixed effects are also significant, highlighting the presence of important differences among the (deflated) land values among different years, probably reflecting evolutions in market conditions, policy and technology.

Model B, presented in the second column, also employs aggregated county-level data to test whether the interaction effect between degree days and precipitation is significant. The approximate t-test cannot reject the null hypothesis (t-value: -1.01, approximate p-value: 0.32) at any standard significance level and both the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) supports the model with no interaction. In general, our county-level estimates indicate strong and approximately linear climatic effect with no evidence of interactions between precipitation and temperature. Overall, taking into account the wet and cold English climate, these results are in line with those reported by previous Ricardian analyses (see Mendelsohn and Dinar, 2009, chap. 7, for a useful review).

TABLE 2: Parameter estimates and diagnostics

	Model A county-level, no climatic interaction		Model B county-level, climatic interaction		Model C farm-level, climatic interaction		Model D farm-level, semi-parametric	
	Coef.	Sign.	Coef.	Sign.	Coef.	Sign.	Edf	Sign.
Climate variables								
precipitation (prec)	-2.33 (0.98)	*	-2.65 (1.05)	**	-4.15 (1.93)	*	--	
prec ²	-0.30 (0.47)		-0.57 (0.53)		-4.07 (1.21)	***	--	
degree days (dd)	5.38 (1.21)	***	5.54 (1.20)	***	15.73 (1.93)	***	--	
dd ²	-1.05 (0.61)		-1.32 (0.64)	*	-2.48 (1.06)	*	--	
dd & prec	--		-23.32 (23.20)		676.64 (153.09)	***	11.21	***
Control variables								
slope	-0.06 (0.02)	**	-0.06 (0.02)	**	-0.03 (0.01)	***	1.00	***
depth to rock	0.00 (0.00)	*	-0.02 (0.01)	*	0.00 (0.00)		1.00	
pop. density (dpop)	1.29 (0.56)	*	1.29 (0.56)	*	0.40 (0.10)	***	3.17	***
dpop ²	-0.22 (0.45)		-0.21 (0.44)		-0.21 (0.06)	***	Coef.	Sign.
share park	0.33 (0.15)	*	0.32 (0.15)	*	0.21 (0.06)	***	0.16 (0.06)	**
share NVZ	-0.04 (0.04)		-0.04 (0.04)		0.02 (0.04)		0.01 (0.04)	
share ESA	-0.14 (0.10)		-0.15 (0.10)		-0.01 (0.01)		0.01 (0.01)	
Fixed effects	Yes	***	Yes	***	Yes	***	Yes	***
Soils shares	Yes	***	Yes	***	Yes	***	Yes	***
Random effects								
County	0.217		0.212		--		--	
Cell	--		--		0.250		0.240	
Farm	--		--		0.399		0.397	
Residuals	0.211		0.211		0.181		0.180	
Model fit								
LogLik	4.33		4.28		-1776.4		-1754.2	
AIC	49.3		50.4		3614.8		3570.5	
BIC	187.2		192.9		3836.5		3792.2	

Notes: Models A and B estimated on county averages, 857 observations for a total of 94 counties. Models C and D estimated on farm data, 9438 observations for a total of 3266 farms in 1350 cells of 10km. Yearly fixed effect and Scotland and Wales dummy variables strongly significant but not reported in the table to preserve space. In Models A, B, C and D, “Coef.” = coefficient estimate (with standard error in parenthesis) and “Sign.” = parameter significance calculated with an approximate t-value conditional on the random effects (details in Pinheiro and Bates, 2000). In model D, “Edf” = “effective degree of freedom”, “Sign” = parameter significance calculated with an approximate F-test (details in Wood, 2006b). In models A, B and C precipitation, degree days and population density transformed into orthogonal polynomials to reduce multicollinearity. LogLik = log-likelihood, AIC = Akaike Information Criterion, BIC = Bayesian Information Criterion, * = significant at the 0.05 level, ** = significant at the 0.01 level, *** = significant at the 0.001 level.

These conclusions, obtained on aggregated data, are strongly rejected when individual farms are analyzed. As the estimates of Model C (third column) shown, rainfall and degree days present strong non-linear effects, with both the quadratic terms and the interaction being highly statistically significant. This interaction is positive: consistent with the agronomic literature, the optimal amount of rainfall increases with

temperature. These findings suggest a significant aggregation bias which we examine more in detail later on in this section. The estimates of the semi-parametric Model D, reported in last column of Table 2, confirm these results. The ‘Effective Degrees of Freedom’ of the bivariate smooth function of temperature and precipitation is equal to 11.68, indicating very strong non-linear effects and interactions. For comparison, the bivariate smooth function corresponding to the climatic effects estimated in Model C (i.e. two quadratic effects with a linear interaction term) would have an EDF equal to 5. On the other hand, the effect of slope and depth to rock are estimated to be linear (EDF = 1).

We cannot compare Model C and D via a likelihood ratio test, since the two models are not nested. However, the log-likelihoods, the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), reported at the end of Table 2, all support the semi-parametric specification. We also carry-out a J-test for non-nested models (Davidson and MacKinnon, 1981). The t -value for rejecting the quadratic specification in favour of the AMM is 310.49, which strongly favors the more flexible model.⁶ Ignoring this non-linearity also comes at the cost of increasing the un-modelled spatial autocorrelation, which is reflected in the higher value of the cell-specific random component of Model C compared to that of Model D. This is consistent with Kostov (2009)’s findings, which shows that flexible functional forms can significantly reduce residual spatial auto-correlation in farmland price modelling.

Returning to the climatic effects, Figure 1 provides a graphical comparison of the estimates produced by our models. The contour graphs shown in all but the top-left panel represent with iso-value lines the joint effects of rainfall and degree days on land price. We only draw such contours for values of degree days and temperature observed within the estimation sample, as represented in the scatter plot in the top-left panel. For example, in GB there are no areas with both relatively high temperature and high precipitations and, therefore, the top-right corner of each plot is blank. Furthermore, the distribution of the climatic variables appear to be highly skewed, with most farms located in areas characterized by relatively high temperatures ($dd > 1300$) and low precipitations ($prec < 500$). Therefore, the upper-left parts of the contour graphs are the most relevant for climate change predictions in GB.⁷

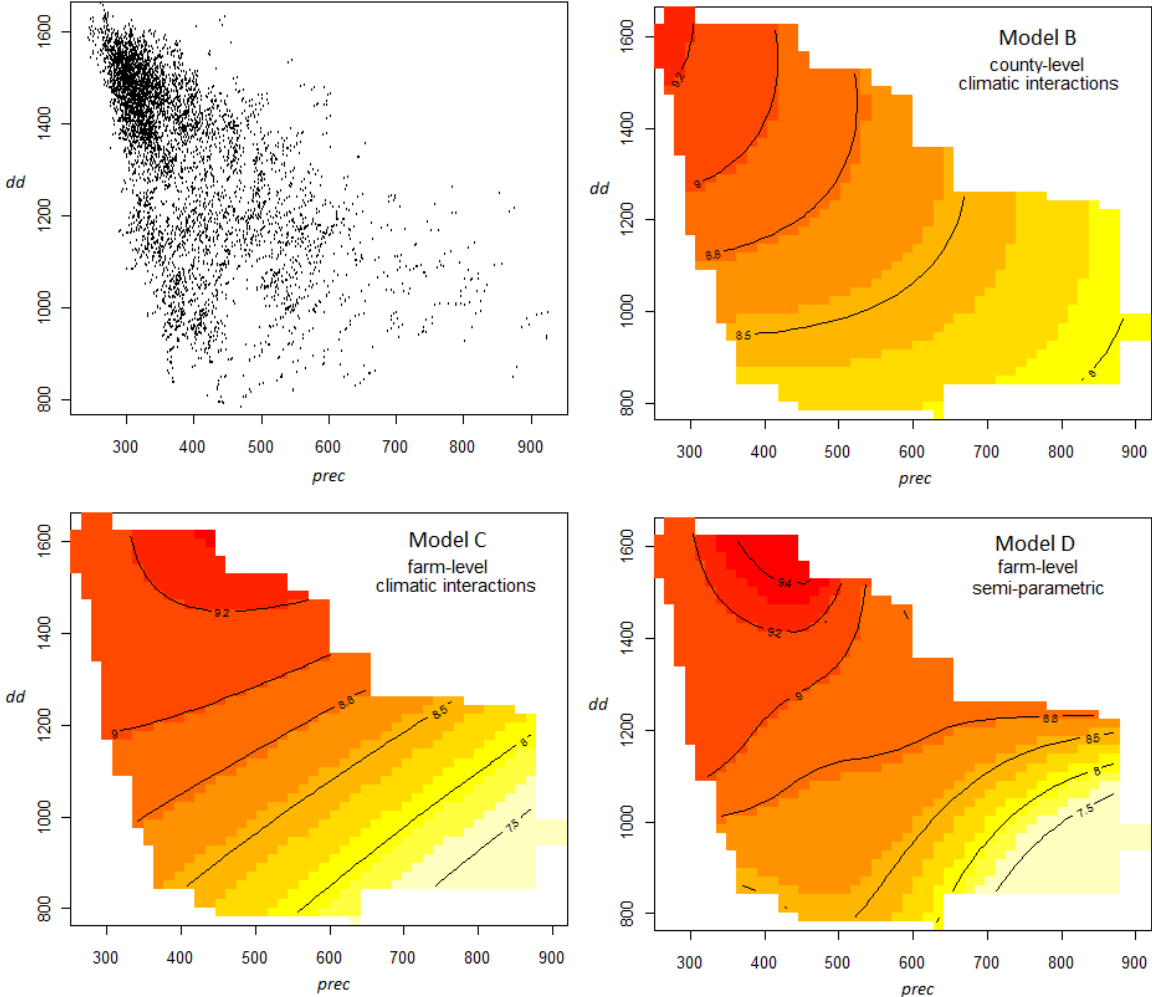
The graphs in the top-right panel of the picture show the effects estimated by Model B: the county-level model with interaction effects (the plot for Model A is almost identical and not reported to preserve space). The highest farm values are found in the driest and warmest climate (top-left corner), whereas the lowest values correspond to the wettest and coldest conditions (bottom-right corner). The shape of the contour graphs changes substantially when using farm-level estimates, as highlighted in the bottom-half panels. The plot corresponding to Model C (left hand side), for example, shows a clear interaction effect. In the colder areas ($dd < 1200$), abundant precipitation reduces

⁶ Ideally we should run a bootstrap simulation to approximate the finite sample distribution of the J-test (Davidson and MacKinnon, 2002). However, this would entail estimating, for each simulation, a new AMM, which requires about 30 minutes. Running the large amount of repetitions required for a bootstrap approximation is, therefore, unfeasible in our case. However, both the large number of observations and the strength of the t -test allow us to confidently reject the null in our specific case.

⁷ The next section shows how our sample of farm is representative of the overall conditions in GB.

agricultural values, since the excess of moisture within the soil can considerably delay machinery operations until late in the growing season. On the other hand, in the warmer areas ($dd > 1400$), the crop requirement for water increases and the effect of rainfall becomes positive. Therefore, beyond a certain level, higher temperatures increase land values only if there is enough precipitation to prevent the risk of drought. This interaction is even more evident in the right-hand lower panel derived from Model D which, by using flexible smooth functions, can better adapt to fit non-linear relations. This contour graph shows an even stronger interaction effect, with a steeper increase in values in the relatively warmer and wetter areas. However, the direction of the interaction and the overall shape of the relation are very similar to those estimated by the simpler Model C.

Figure 1: The effect of temperature and precipitation on the logarithm of land price: contour plots with iso-value lines



Notes: In top-left corner scatter plot of degree days and precipitation in the estimation sample, in the top-right: Model B = county-level model with interaction effect between precipitation and temperature, bottom-left: Model C = farm-level model with interaction effect, bottom right: Model D = farm-level, semi-parametric (AMM) model.

Figure 2 analyzes the interaction effect of the two climatic variables in greater detail. For each level of temperature (varied down the three rows of graphs) we show the impact of rainfall on land value as predicted by Models B, C and D (the estimates from Model A are,

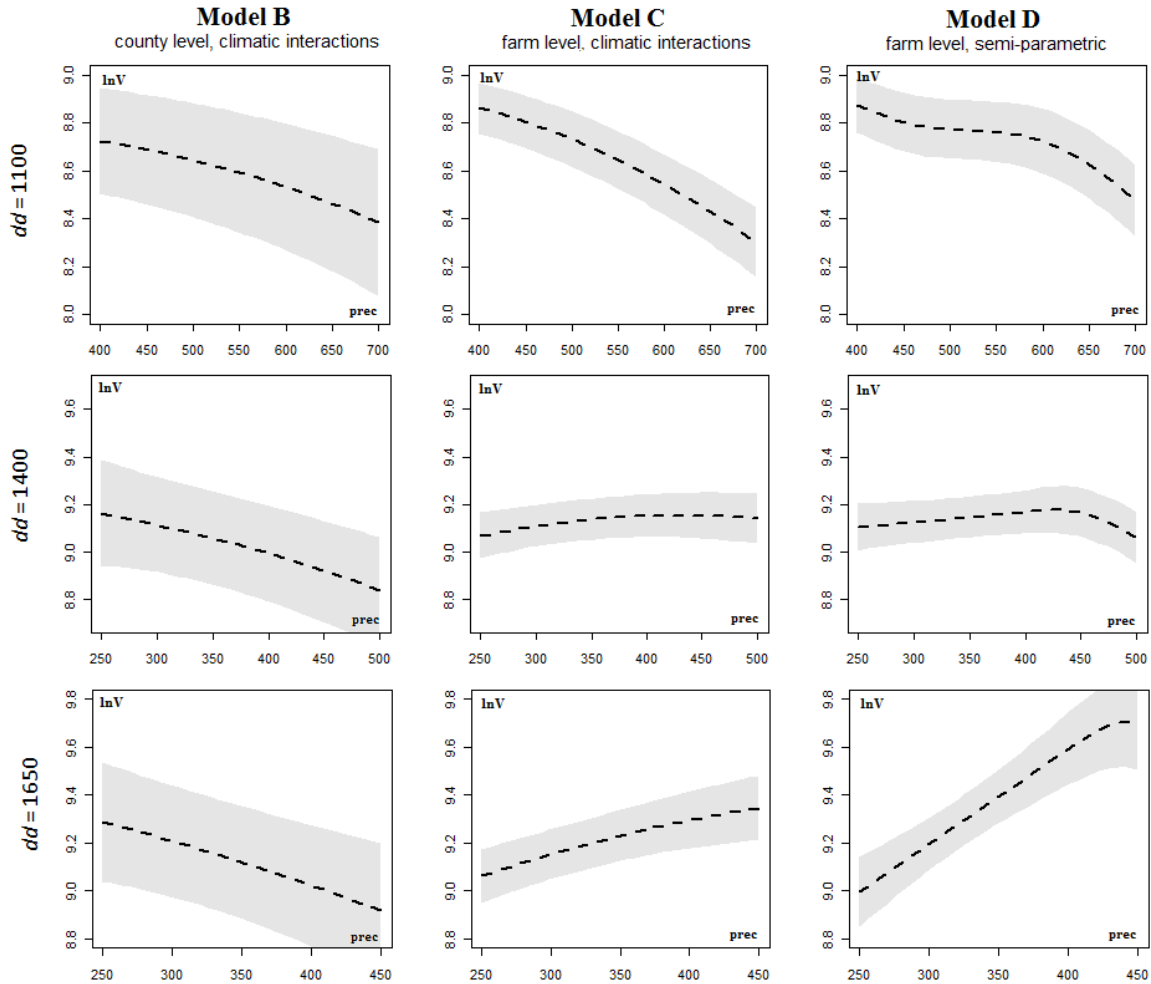
again, virtually identical to those of Model B and therefore we omit them to save space). As in Figure 1, the spans of the four plots differ because the effects are represented only within the width of the data used for estimation. Comparison between the first (Model B) and the second column (Model C) provides an assessment of the aggregation-bias, while comparison of the second and the third column (Model D) highlights the difference between parametric and semi-parametric estimates, providing a visual test of functional form mis-specification. According to the county-level Model B, the effect of rainfall on land value is strong and negative, and virtually the same for all levels of temperature. The optimal amount of rainfall within the plotted range is always the lowest possible one: 400mm for cold temperatures ($dd = 1100$) and 250mm for average ($dd = 1400$) and warm ($dd = 1650$) temperatures. In contrast, both models estimated on farm data show that the effect of rainfall varies considerably with temperature. For cold temperatures the response functions reveal a strong negative effect, very similar to the one estimated on county data. However, as temperature increases the relationship becomes more moderate ($dd = 1400$) and, in the warmest areas ($dd = 1650$), rainfall has a significant and strong positive effect within the entire range of the observed data. These results reconcile Ricardian analyses with the agronomic literature, which shows that in relatively warmer climates crops require more water for development (e.g. Monteith 1977; Morison, 1996). As one may expect, this effect is particularly strong in the semi-parametric Model D, which is the most flexible specification considered here.

Taken together, these findings provide strong evidence of aggregation bias. In particular, the interaction effect between temperature and rainfall, which is very strong and significant in the farm-level analyses, completely disappears in county-level estimates. For this reason, aggregated models are not able to capture the positive effect of rainfall on land values in warmer areas of GB. Since both temperature and rainfall patterns are expected to transform as a result of climate change, this bias may have significant implications for climate change predictions. This hypothesis is tested in the next Section, which compares the projected climate change impacts according to our four Ricardian regressions. On the other hand, the differences in estimates between the parametric and semi-parametric models appear to be fairly minor, indicating no major deviation from quadratic functional forms.

We conclude this Section by undertaking some robustness analyses on our farm-level estimates, testing for omitted variable bias and parameter stability. Concerns about potential omitted cross-sectional variables in Ricardian analyses have been pointed out by Deschênes and Greenstone (2007), among others. A panel fixed-effect estimator provides little help in addressing this issue, since it would eliminate not only the potential bias but also all the cross-sectional variation on which the parameters of the Ricardian model hinge upon. In Ricardian models, in fact, in order to represent climate (as opposed to weather), temperature and rainfall are constructed as long-term averages and, even in long panels, present a time-variability which is almost negligible compared to the cross-sectional (spatial) variability. This means that fixed-effect estimators, intended to eliminate potential time-invariant omitted variables, cannot be implemented in this context, because they would also remove the crucial spatial variation of climate. A second, related, issue, concerns the distortions in the land market which various agricultural policies tend to create (e.g. Barnard et al., 1997). For instance,

crop and environmental payments could be capitalized in the land price and, if correlated with climatic variables, introduce a bias in their coefficient estimates.

Figure 2: The of effect precipitation on the logarithm of land price for different levels of temperature



Notes: Dotted line = estimated logarithm of land value, gray area = 95% approximate confidence interval intervals, (for model D based on a Bayesian representation of the smooth functions (Wood, 2006, for details). All variables except precipitation ($prec$) and temperature (dd) held constant at their average levels in the sample. Model A is very similar to Model B and therefore not reported to preserve space.

We address both issues by testing whether the parameters of degree days and precipitation are stable over time. If omitted variable bias is a concern, in fact, these coefficients should exhibit significant temporal variation, reflecting changes in agricultural prices and policies. Our sample provides a particularly hard benchmark for this test, since, during the 10 years covered by our data, agricultural policies in GB have changed markedly, following various reforms of the EU Common Agricultural Policy. These policy developments culminated with the introduction of decoupled single farm payments in 2005, replacing the system of area-based crop-specific subsidies in place since 1992. Input and output prices have also changed dramatically during this period, with, for instance, cereal prices more than doubling during the period from 2007 to

2008. This variation is reflected in the yearly fixed-effects, which are strongly significant in all specifications. If there are omitted variables correlated with climate, therefore, one would expect the smooth functions of temperature and precipitation to present significant variation over time in such unstable market conditions. To investigate this hypothesis, we choose the first year in our sample (1999) as the baseline and test for parameter stability comparing one year at a time via pairwise F-tests (Pineiro and Bates, 2000). We test our final and more flexible specification, Model D, and in order to use the standard inference for random-effect models, we define the spline bases of the model *a priori* and implement un-penalized estimation via ML. To attain a level of flexibility comparable to the optimal one selected by the penalized likelihood, we choose natural cubic splines with 4 knots for population density and 16 knots for the joint function of precipitation and rainfall, while all other variables are modelled as linear terms. Table 3 reports the results: none of the pairwise instability tests is significant at the 5% level, and only one is significant at the 10%. This is consistent with the null hypothesis of parameter invariance which, therefore, we find no evidence to reject. Overall, these results reassure us about the robustness of our climate impact estimates to omitted variable bias such as changes in prices and agro-environmental policies.

TABLE 3: *Pairwise stability tests*

<i>years</i>	<i>F-test</i>	<i>p-value</i>
<i>1999 and 2000</i>	0.94	0.478
<i>1999 and 2001</i>	0.53	0.930
<i>1999 and 2002</i>	1.34	0.266
<i>1999 and 2003</i>	1.54	0.112
<i>1999 and 2004</i>	1.82	0.062
<i>1999 and 2005</i>	1.07	0.348
<i>1999 and 2006</i>	1.25	0.292
<i>1999 and 2007</i>	1.01	0.478
<i>1999 and 2008</i>	1.58	0.110

Notes: test on Model D estimated with 4 knots for population density and 16 knots for the joint effect of rainfall and temperature. The F-statistics test the stability of the climate parameters from one year to the other and are conditional on the random-effect estimates, as suggested in Pineiro and Bates (2000). Approximated p-values are calculated with 500 bootstrap repetitions.

5. CLIMATE CHANGE IMPACT PREDICTIONS

We combine the estimates presented in the previous section with the recently released UK Climatic Projections 2009 (UKCIP09, source: ukclimateprojections.defra.gov.uk) to forecast the impact of climate change on agriculture in Great Britain and to test whether aggregation bias has any implication for predictions. Specifically, we use the UKCIP09 projected changes in monthly average minimum temperature, maximum temperature and precipitation in the medium level emission scenario for years 2015-2045 (corresponding to the SRES A1B in the IPCC Special Report on Emissions Scenarios, Nakicenovic et al., 2000). This data is available on 25 km grid squares covering the entire UK. We derive the corresponding values of degree days and precipitation in the

growing season by applying these changes to the 10 km Met Office historical averages for the years 1960-1990, which are the baseline climatic conditions identified by the UKCIP09. We also take this climate as our baseline.

To derive the impact of climate change, we predict log-agricultural land price under both the baseline and the climate change scenario. The only difference between the two scenarios is climate: all other determinants (soil, population density, etc.) are kept constant at their 2008 values, the last year of our farm data. Consequently, we do not consider other factors that are likely to change in the future, such as technology, prices, land use and population. Therefore, our results are intended to estimate how climate will affect agriculture *ceteris paribus*, and should not be interpreted as predictions of the future. Table 4 reports descriptive statistics for the climatic and environmental determinants in the baseline and climate change scenarios. The ranges of the exogenous variables are similar to those of the data used for estimation, (reported in Table 1) indicating that our FS sample is representative of the overall environmental conditions in Great Britain. Possible exceptions are the highest upland areas in the North of Scotland, characterized by much lower degree days and higher precipitations than the rest of the country. These are just a few 10 km cells with extremely low land values and, therefore, this is not likely to be significant issue for our analysis.

TABLE 4: *Descriptive statistics of the climatic and environmental variables, baseline climate (1960-1990) and UKCIP 2015-2045 projections (medium emission scenario)*

	units	\bar{x}	$\hat{s}(x)$	min	max
Baseline (1960-1990)					
degree days	°C	1164.0	251.8	367.4	1645.0
precipitation	mm	450.7	169.0	250.8	1504.0
Climate change projections (UKCIP 2015-2045)					
degree days	°C	1424.0	276.1	571.6	1948.0
precipitation	mm	395.8	188.3	158.6	1444.0
Control variables					
depth to rock	dm	6.4	3.5	0.0	14.0
slope	°	4.4	3.6	0.0	24.7
pop. density	pop/Km ²	226.4	434.6	8.0	4924.0

Notes: \bar{x} indicates the sample mean, $\hat{s}(x)$ the sample standard deviation. Data refer to Great Britain 10 km grid square cells, only including only cells in which there is some agricultural land. Control variables are assumed to remain constant between the two scenarios.

Compared to the baseline, the UKCIP 2015-2045 medium emission climate change scenario is characterized by both higher temperatures and lower precipitation during the growing season. The highest value of degree days is 1948, still considerably below 2400: the threshold identified by SHF as the level at which temperature begins to have a negative effect on land values. Therefore, we can extrapolate our model estimates with sufficient confidence. However, to test the robustness to these out-of-sample

projections, we also compute climate change impacts using a “limited” scenario, in which all combinations of rainfall and temperature are restricted to be within the range used for estimation. For example, in this “limited” scenario we set to 1700 the maximum value for degree days and to 240 mm the minimum quantity of precipitation. While we calculate predictions from the parametric specifications (Models A, B and C) using both the “original” and “limited” climate change scenarios, we only use the latter for the semi-parametric regression (Model D) since the bi-dimensional smooth function of degree days and precipitation tends to be very erratic outside the range of values used for estimation, generating unreliable results when used out-of-sample.

TABLE 5: *Climate change impact on agriculture in the 2015-2045 UKCIP medium emission scenario*

Model	sample	mean (%)	Q 10 (%)	Q 90 (%)	Total (M£)	std.err (M£)
Model A (county, no climatic interactions)	original	23.24	13.44	37.13	1167	401
	limited	22.83	11.34	37.13	1150	346
Model B (county, climatic interactions)	original	23.41	17.64	30.39	1175	391
	limited	22.38	12.65	30.39	1123	336
Model C (farm, climatic interactions)	original	7.01	-17.09	39.06	352	194
	limited	8.62	-10.34	38.58	432	177
Model D (farm, semi-parametric)	limited	4.29	-19.69	35.57	215	240

Notes: Weighted statistics (weights = agricultural land area * land value in the baseline). “Q10” indicates the 10% quantile, “Q90” the 90% quantile, “Total” refers to the sum of annual GB farm net revenues assuming a discount rate of 5%, and “std.err” is the standard deviation of the total, calculated via 5000 bootstrap repetitions.

The predicted impacts of climate change on farmland values derived from our four models are reported in Table 5. To provide a meaningful summary of grid squares or counties with different agricultural areas and land values, the percentage changes are weighted by the total agricultural land value in each cell or county in the baseline.⁸ The two county-level regressions (Model A and B) predict strong and positive impacts on the rural sector, with an overall 23% increase in GB farm values compared to baseline levels. By applying a 5% discount rate (as per Mendelsohn et al., 1994), this translates into an increase in farm net revenues of roughly £1 billion per annum. Furthermore, the first decile is +13%, indicating that almost all counties will be significantly better-off. However, the estimates provided by our farm-level models (Model C and D) are

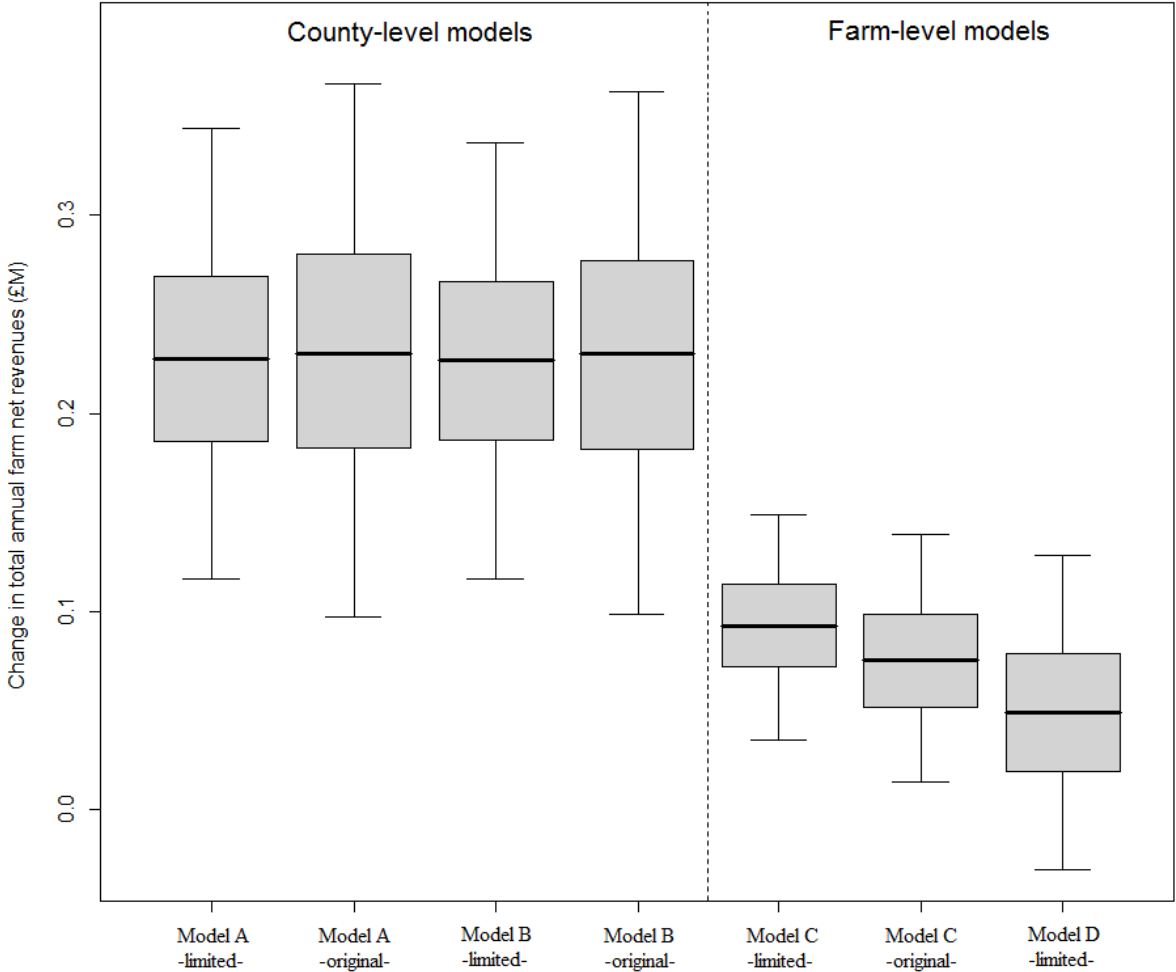
⁸ We compute the total agricultural area within each square or county using 1 km grid data from the Land Cover Map 2000 (LCM2000, <http://www.ceh.ac.uk/LandCoverMap2000.html>), produced by the Centre for Ecology and Hydrology. LCM2000 is a parcel-based classification of satellite image data showing land cover for the entire United Kingdom. We include in agriculture the land classified as: (a) arable and horticultural, (b) improved grassland, (c) semi-natural, rough grass and bracken and (d) mountain, heath and bog. This definition slightly overestimates the amount of agriculture in GB, resulting in a total of about 19 million hectares rather than the 17 million presented in the official statistics (e.g. <http://www.defra.gov.uk/statistics/>). Therefore, we rescaled the agricultural area in each cell or county by multiplying it by a factor of 0.9 to match with the official GB total.

considerably less optimistic. Model C predicts a mean increase in farm value ranging between 7% and 9% (which is lower than the first decile predicted by both county-level models) and large areas experiencing losses, with a first decile of -17%. Even these results overestimate the findings of our most flexible regression, Model D, which shows that climate change will induce a diverse set of impacts on agricultural land values ranging from a lower decile of -20% to an upper decile of more than +35% with a mean value around +5%, comparable to an annual rise in net revenues of about £200 million.

Not surprisingly, the motivation for this difference in predictions can be found in the interaction effect between temperature and rainfall. Recalling Figures 1 and 2, precipitation in farm-level models can either have a negative or a positive effect on land values, depending on the level of temperature. While the warmer growing season projected in this climate change scenario can boost yields, it can also increase crops' water demand and reduce drought-tolerance. The scenario is also characterized by a decrease in precipitation which, in the driest areas of GB, can lead to water deficiency and less favourable farming conditions, ultimately lowering land values. This effect is particularly strong according to Model D, which, in fact, predicts the lowest overall increase in farmland values. In contrast, the failure of the aggregated, county-level models to capture the interaction effect between temperature and precipitation leads to lower rainfall and higher temperatures being erroneously predicted as having consistently positive impacts on land values.

This difference in climate change impact predictions is highlighted in Figure 3, which presents box-plots of the distribution of the change in total GB farm net revenues according to the different models, computed via 5000 bootstrap repetitions. The bias introduced by aggregation is evident, with the county-level models' projections, represented in the first four plots, being significantly different from the farm-level results, shown in the last three plots. In comparison, using a semi-parametric approach (Model D) rather than simpler quadratic regressions (Model C) introduces only minor differences, which are not statistically significant. Similarly, restricting the climate change conditions within the boundaries of the estimation sample ("limited" climate scenario) as opposed to using the original UKCIP temperature and precipitation values ("original climate scenario") does not significantly change the results.

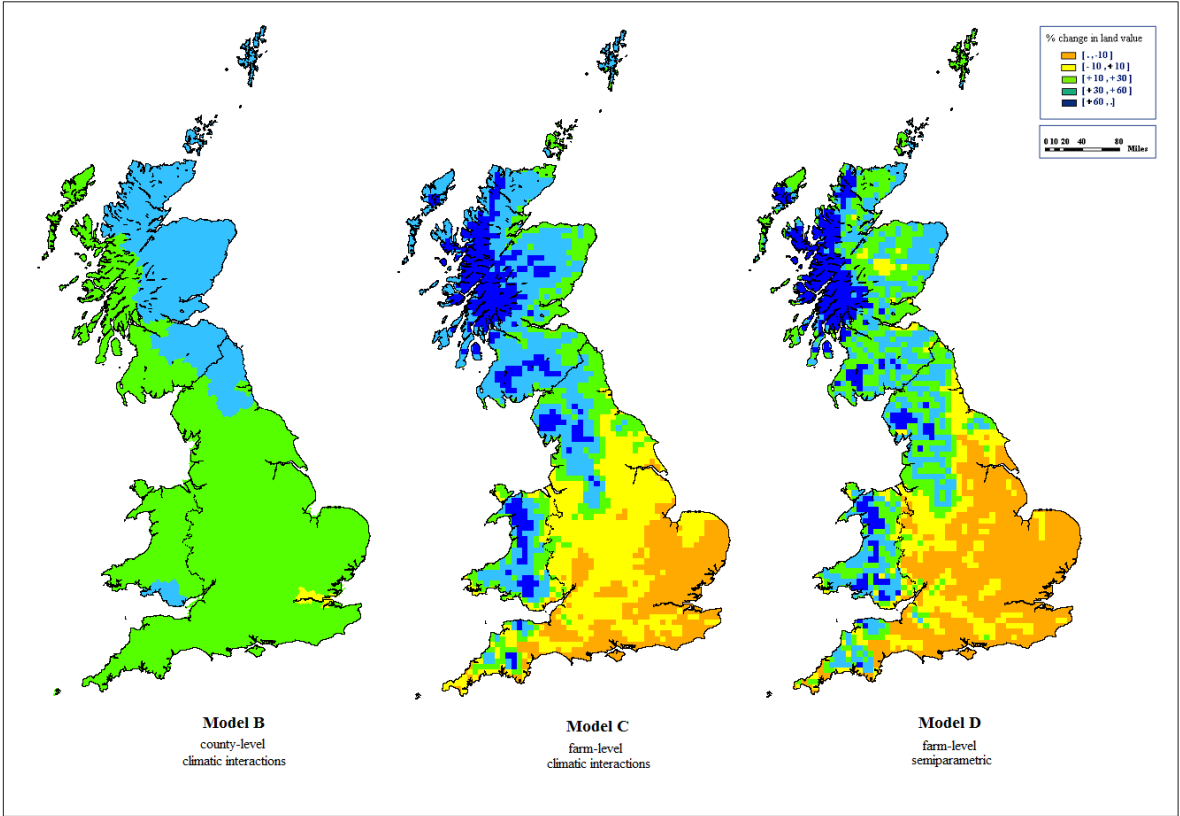
Figure 3: Estimated total impact of climate change on GB agriculture in the 2015-2045 UKCIP medium emission scenario



Notes: The boxplots represent the confidence intervals for the change in total GB annual farm net revenues, calculated with 5000 bootstrap repetitions. The gray box indicates the 1st and 3rd quartile, the whiskers the 95% confidence interval. Models A and B are estimated on county-level data, whether Models C and D are based on farm-specific data, “original” = original climate change scenario, “limited” = “limited” climate change scenario.

As a final analysis, Figure 4 represents the spatial distribution of farmland value changes derived from models B, C and D (Model A is again virtually indistinguishable from Model B and, therefore, is not reported). Consistent with the statistics reported in Table 5, the aggregated, county-level Model B predicts a uniform increase in land values through-out the country driven by the warmer and drier climate. In contrast, both farm-level models predict spatially heterogeneous impacts with the wetter areas of Scotland, Wales and the North of England being significantly better-off and the dry lowland of the South-East of England being the main loser, as a consequence of higher temperatures and reduced precipitations increasing the risk of heat-related stress for crops. This spatial heterogeneity is irredeemably lost by using county-level data. In our case-study the aggregation process has not only affected the parameter estimates, as showed in the previous Section, but has also led to climate change impact projections which are significantly biased.

Figure 4: Predicted percentage change farmland value under the 2015-2045 UKCIP09 climate change scenario (average emission level)



Notes: Model B and C use the original climate scenario, Model D the “limited” climate scenario. Predictions from Model A are virtually indistinguishable to those from Model B and are, therefore, not reported.

6. CONCLUDING REMARKS

This paper presented a Ricardian land value regression using a unique, 10 year panel of more than 3000 farms. Comparing this analysis with the standard models estimated on data averaged across counties, we demonstrate that a strong bias afflicts climatic coefficients based on aggregated data. While county-level regressions confirm the assumption of additive climatic effects implemented in previous Ricardian studies, our farm-level analysis reveals important interactions between precipitation and temperature in determining land values. Consistent with the literature on plant physiology, which shows that the crop requirement for water increases with temperature, we find that higher precipitation is more valuable when temperatures are high. Accordingly, higher temperatures increase land values only if there is enough precipitation to prevent the risk of drought. This interaction effect becomes statistically insignificant when we analyze the same data aggregated over counties. This aggregation bias has important implication for climate change impact projections. Predictions based on county-level data are significantly distorted and, in many areas, even present the incorrect sign.

We also test for functional form miss-specification by estimating a semi-parametric model based on penalized splines. The results do not appear to be significantly different from those obtained with the simpler, quadratic regression. Although farmland values have changed considerably in the 10 years included in the analysis, our estimates remain remarkably robust. As per SHF, we find that the fundamental dependence of agricultural incomes on climatic conditions is independent of government policies and crop prices.

The hedonic equations of farmland values are then used to predict the impact of climate change on agriculture in GB, using the recent UKCIP09 2015-2045 medium emissions scenario. Results from our farm-level models indicate that climate change will be, on average, moderately beneficial for agriculture, with an aggregated increase in farmland value between 4% and 9% and the largest gains arising in the upland areas located in Scotland, Wales and the North of England, where agriculture is currently substantially constrained by cold weather and excess rainfall. Here the warmer, drier climates projected by UKCIP09 will generally improve farm productivity and the accompanying agricultural land value. However, there will be some localized losses, particularly in the East of England, where the decrease in precipitation and the warmer climate may increase the risk of droughts. Aggregation to county-level data produces a significantly distorted picture, with gains predicted for virtually all areas and an average increase in farm land values in excess of 20%.

These results are relevant also to other countries. Since both precipitation and temperature are expected to be altered by the process of climate change, our results casts some doubts on the validity of the projections based on previous county-level Ricardian models, which failed to account for interaction effects. For example, considering the US, the projected increase in average rainfall could substantially mitigate the negative impact of increased temperature in some climate change scenarios (e.g. Nakicenovic et al., 2000). However, given that precipitation is also expected to decrease in some areas, further research is necessary to assess the overall magnitude of the bias introduced by the omission of interactions effect. Given this, a welcomed contribution would be the extension this analysis to farm-level data covering a larger and even more heterogeneous region, such as North America. This would allow the estimation of climate change impacts on land value for temperatures outside those captured within our temperate study area. Finally, the usual caveats of Ricardian analyses apply also here. Prices, population and other drivers are assumed to remain constant between the scenarios. Possible beneficial effects of increased CO₂ fertilization on crop growth are also not taken into account, though some recent research suggests that those may be much smaller than previously believed (Long et al., 2006).

7. REFERENCES

Adams R.M., Rosenzweig C., Pear R.M., Ritchie J.T., McCarl B.A., Glycer J.D., Curry R.B., Jones J.W., Boote K.J, Hartwell A.L. (1990) Global climate change and US agriculture, *Nature*, vol. 344, pp. 219 – 224.

- Anglin P., Gencay R. (1996) Semiparametric estimation of a hedonic price function, *Journal of Applied Econometrics*, vol. 11, pp. 633-648.
- Barnard C.H., Whittaker G., Westenbarger D., Ahearn M. (1997) Evidence of capitalization of direct government payments into U.S. cropland values, *American Journal of Agricultural Economics*, vol. 79, pp. 1642-1650.
- Bontemps C., Simioni M., Surry Y. (2008) Semiparametric hedonic price models: assessing the effects of agricultural nonpoint source pollution, *Journal of Applied Econometrics*, vol. 23, pp. 825-842.
- Cropper M.L, Deck L.B, McConnell K.E. (1988) On the choice of functional form for hedonic price functions, *The Review of Economics and Statistics*, vol. 70, pp. 668-675.
- Davidson R., MacKinnon J.G. (1981) Several tests for model specification in the presence of alternative hypotheses, *Econometrica*, vol. 49, pp. 781-793.
- Davidson R., MacKinnon J.G. (2002) Bootstrap J tests for nonnested linear regression models, *Journal of Econometrics*, vol. 109, pp. 167-193.
- Deschênes O., Greenstone M. (2007) The economic impacts of climate change: evidence from agricultural output and random fluctuations in weather, *American Economic Review*, vol. 97, pp. 354-385.
- Ekeland I., Heckman J., Nesheim L. (2002) Identifying hedonic models, *American Economic Review*, vol. 92, pp. 304-309.
- Ekeland I., Heckman J., Nesheim L. (2004) Identification and estimation of hedonic models, *Journal of Political Economy*, vol. 112, pp. S60-S109.
- Feenstra R.C., Hanson G.H. (2000) Aggregation bias in the factor content of trade: evidence from U.S. manufacturing, *American Economic Review*, vol. 90, pp. 155-160.
- Feige E., Watts H. (1972) An investigation of the consequences of partial aggregation of micro-economic data, *Econometrica*, vol. 40, pp. 343-360.
- Fisher A., Hanemann M., Roberts M., Schlenker, W. (2012) The economic impacts of climate change: evidence from agricultural output and random fluctuations in weather: comment, *American Economic Review* (forthcoming).
- Goodfriend M. (1992) Information-aggregation bias, *American Economic Review*, vol. 82, pp. 508-519.
- Grunfeld Y., Griliches Z. (1960) Is aggregation necessarily bad? *Review of Economics and Statistics*, vol. 42, pp. 1-13.
- Imbs J., Mumtaz H., Ravn M.O., Rey H. (2005) PPP strikes back: aggregation and the real exchange rate, *Quarterly Journal of Economics*, vol. 120, pp. 1-43.

Intergovernmental Panel on Climate Change (2007) *Climate change 2007: synthesis report*, (Eds: A. Allali, R. Bojariu, S. Diaz, I. Elgizouli, D. Griggs, D. Hawkins, O. Hohmeyer, B.P. Jallow, L. Kajfez-Bogataj, N. Leary, H. Lee, D. Wratt), published by the Intergovernmental Panel on Climate Change, Geneva.

Kaufmann R.K., Snell S.E. (1997) A biophysical model of corn yield: integrating climatic and social determinants, *American Journal of Agricultural Economics*, vol. 79, pp. 178-190

Keele L. (2009) *Semiparametric regression for the social sciences*, Wiley and Sons, Chichester.

Kostov P. (2009) Spatial dependence in agricultural land prices: does it exist?, *Agricultural Economics*, vol. 40, pp. 347-353.

Laird N.M., Ware J.H. (1982) Random-effects models for longitudinal data, *Biometrics*, vol. 38, pp. 963-974.

Lang G. (2007) Where are Germany's gains from Kyoto? Estimating the effects of global warming on agriculture, *Climatic Change*, vol. 84, pp. 423-439

Lobell D.B, Schlenker W., Costa-Roberts J. (2011) Climate trends and global crop production since 1980, *Science*, vol. 333, pp. 616-620.

Long S. P., Ainsworth E.A, Leakey A.D.B., Nosberger J., Ort D.R. (2006) Food for thought: lower-than-expected crop yield stimulation with rising CO₂ concentrations, *Science*, vol. 312, pp. 1918-1921.

Maddison D. (2009) A spatio-temporal model of farmland values, *Journal of Agricultural Economics*, vol. 60, pp. 171-189.

Mendelsohn R., Dinar A., (2009) *Climate change and agriculture: an economic analysis of global impacts, adaptation and distributional effects*, Northampton, Edward Elgar.

Mendelsohn R., Nordhaus W.D., Shaw D. (1994) The impact of global warming on agriculture: a Ricardian analysis, *American Economic Review*, vol. 84, pp.753-771.

Morison J.I.L. (1996) Climate change and crop growth, *Environmental Management and Health*, vol. 7, pp. 24-27.

Monteith, J.L. (1977) Climate and the efficiency of crop production in Britain, *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 281, pp. 277-294.

Parmeter C.F., Henderson D.J., Kumbhakar S.C. (2007) Nonparametric estimation of a hedonic price function, *Journal of Applied Econometrics*, vol. 22, pp. 695-699.

Pinheiro J.C., Bates D.M. (2000) *Mixed effects models in S and S-plus*, Springer, New York.

R Development Core Team (2008) *R: a language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, <http://www.r-project.org>.

Ricardo D. (1817) *On the principles of political economy and taxation*, London, John Murray.

Ruppert D., Wand M.P., Carroll R.J. (2003) *Semiparametric regression*, Cambridge University Press, New York.

Schlenker W., Hanemann M., Fisher A.C. (2005) Will U.S. agriculture really benefit from global warming? Accounting for irrigation in the hedonic approach, *American Economic Review*, vol. 95, pp. 395-406.

Schlenker W., Hanemann M., Fisher A.C. (2006) The impact of global warming on U.S. agriculture: an econometric analysis of optimal growing conditions, *Review of Economics and Statistics*, vol. 88, pp. 113-125.

Schlenker W., Hanemann M., Fisher A.C. (2007) Water availability, degree days and the potential impact of climate change on irrigated agriculture in California, *Climatic Change*, vol. 81, pp. 19-38.

Schlenker W., Roberts M. (2006) Nonlinear effects of weather on corn yields, *Review of Agricultural Economics*, vol. 28, pp. 391-398.

Schlenker W., Roberts M. (2009) Nonlinear Temperature Effects indicate Severe Damages to U.S. Crop Yields under Climate Change, *Proceedings of the National Academy of Sciences*, vol. 106, 2009, p. 15594-15598.

Seo S., Mendelsohn R. (2008) A Ricardian analysis of the impact of climate change on Latin American farms, *Chilean Journal of Agricultural Research*, vol. 68, pp. 69-79.

Seo S., Mendelsohn R., Dinar A., Hassan R., Kurukulasuriya P. (2009) A Ricardian analysis of the distribution of climate change impacts on agriculture across agro-ecological zones in Africa, *Environmental and Resource Economics*, vol. 43, pp. 313-332.

Speed T.P. (1991) Comment on: that BLUP is a good thing: the estimation of random effects (by G.K. Robinson), *Statistical Science*, vol. 6, pp. 42-44.

Stocker T. (1984) Completeness, distribution restriction and the form of aggregate functions, *Econometrica*, vol. 52, pp. 887-907.

Stocker T. (1986) Simple tests of distributional effects on macroeconomic equations, *Journal of Political Economy*, vol. 94, pp. 763-795.

Stram D. O., Lee, J. W. (1994) Variance components testing in the longitudinal mixed-effects models, *Biometrics*, vol. 50, pp. 1171–1177.

Theil H. (1954) *Linear Aggregation of Economic Relations*, Amsterdam, North-Holland.

Timmins, C. (2006) Endogenous land use and the Ricardian valuation of climate change, *Environmental and Resource Economics*, vol. 33, pp 119-142.

UK Climate Impacts Programme (UKCIP, 2009) *UK climate projection: briefing report*, Met Office Hadley Centre, Exeter, UK.

Van Garderen K.J, Lee K., Pesaran M.H. (2000) Cross-sectional aggregation of non-linear models, *Journal of Econometrics*, vol. 95, pp. 285-331.

Welham S.J., Cullis B.R., Kenward M.G., Thompson R. (2007) A comparison of mixed model splines for curve fitting, *Australian and New Zealand Journal of Statistics*, vol. 49, pp. 1-23.

Wood, S.N. (2003) Thin plate regression splines, *Journal of the Royal Statistical Society, Series B*, vol. 65, pp. 95-114.

Wood, S.N. (2006a) Low rank scale invariant tensor product smooths for generalized additive mixed models, *Biometrics*, vol. 62, pp. 1025-1036.

Wood S.N. (2006b) *Generalized additive models: an introduction with R*, Chapman and Hall, New York.

Appendix I: the semi-parametric AMM estimation

This Appendix provides the details regarding the estimation of the model in equation (7). We represent the smooth functions as splines, i.e. linear combination of basis functions of the regressors (e.g. Ruppert, Wand and Carroll, 2003; Wood, 2006b; Keele, 2009). For example, considering one of the non-climatic factors g_x and indicating the basis functions with $b_{x,j}(g_x)$ ($j=1,\dots,l_x$), the corresponding smooth function can be written as:

$$(7) \quad s_x(g_x) = \sum_{j=1}^{J_x} \delta_{x,j} b_{x,j}(s_x) = \boldsymbol{\delta}_x' \mathbf{b}_x$$

where $\delta_{x,j}$ ($j = 1,\dots,J_x$) are the parameters to be estimated and J_x is the number of basis functions which determines the maximum possible flexibility of the relation between g_x and V (the higher the value, the more non-linear or 'wiggly' is the estimated effect). Among the simplest basis functions are those corresponding to linear regression splines, which fit a piecewise linear function between a set of knots located between the range of values of the regressor. The number of knots determines the flexibility of the splines and the number of parameters to be estimated. In the linear case with r knots ($\kappa_{1x},\dots, \kappa_{rx}$) and suppressing the subscript x for simplicity, equation (7) becomes:

$$(8) \quad s(g) = \delta_0 + \delta_1 g + \sum_{j=1}^r \delta_{j+1} (g - \kappa_j)_+ ,$$

where $(g - \kappa_j)_+ = \max(0, g - \kappa_j)$ (note that, if there are multiple smooth functions, the constant is only identifiable by imposing a sum-to-zero constraint). Whiles this type of spline is conceptually very intuitive, it can present sharp corners at the knots and, therefore, is too restrictive for many applications. In this paper we use natural cubic splines, which offer computational advantages when applied to large datasets (see Wood, 2006b). These splines fit third degree polynomial functions between each set of knot, with first and second derivatives constrained to be continuous in the entire range of $g(.)$. Furthermore, in order to avoid erratic behavior at the extremes, the fit before the first knot and after the last one is constrained to be linear (i.e. first and second derivatives are set to zero). This results in the number of basis functions J_x being equal to the number of knots r .⁹

The number of knots effectively determines the flexibility of the smooth function. Given a fixed number of knots, the model can be estimated as a standard regression, i.e. by Ordinary Least Squares (OLS) or Maximum Likelihood (ML). However, there is a trade-off between sufficient knots to accurately represent any unknown, non-linear relation and, at the same time, avoid the risk of overfitting. This is a common problem in semi-parametric approaches. A practical solution to this long-standing issue is penalized estimation (Ruppert, Wand and Carroll, 2003; Wood, 2006b). The idea here is to

⁹ Several other types of bases have been proposed in the literature. Ruppert et al. (2003) and Wood (2006b) provide a comprehensive illustration. Welham et al. (2007) demonstrate the links existing among the most commonly used bases and undertake a simulation study from which no clear winner emerges.

augment the likelihood by including a penalty for the excessive roughness (typically indicated with the term ‘wigginess’) of the smooth functions, which can be expressed as a function of the integral of the square of its second derivative. The penalized likelihood corresponding to the smooth function in equation (7) can then be written as:

$$(9) \quad l_p(\delta) = l(\delta) + \lambda \int s''(g)^2 dg ,$$

where $l(\cdot)$ is the model likelihood, $l_p(\cdot)$ is the model penalized likelihood, δ is the vector of parameter to be estimated and λ is the smoothing parameter, which controls the weight given to the ‘wigginess’ penalty. As λ increases so the function becomes smoother, with $\lambda \rightarrow \infty$ corresponding to a straight line fit. In this framework, therefore, the flexibility of the smooth function is regulated by the value of the smoothing parameter λ rather than by the number and placement of the knots, which actually make little difference (see Keele, 2009, for some examples). Ruppert, Wand and Carroll (2003), for example, show that the degrees of freedom of a smoothing spline are just a mathematical transformation of λ .

Various techniques have been proposed to estimate the optimal amount of smoothing (i.e. the parameter λ) directly from the data (see Wood, 2006b, and Keele, 2009). In this paper we use ML estimation techniques representing the smoothing splines as random effects (Ruppert, Wand and Carroll, 2003). The random effect representation of the natural cubic spline corresponding to equation (7) can be written as:

$$(10) \quad s(g) = \delta_0 + \delta_1 g + \sum_{j=1}^{r-2} \phi_j b_j(g) ,$$

where the $b_j(g)$ are non-linear basis functions (whose definition is somewhat lengthy and given, for example, in Welham et al., 2007), δ_0 and δ_1 are the fixed effect (un-penalized) parameters and the ϕ_j are elements of a vector of random effects drawn from a $N(0, \sigma_\phi^2 H)$, where H depends on the penalties (9). This approach models non-linearity as a form of heterogeneity across groups. The data within each set of knots form each group. The intuition behind this representation is that a linear fit ($\phi_1 = \phi_2 = \dots = \phi_{r-2} = 0$) would ignore these differences and capture the relationship with only two parameters, whereas an un-penalized likelihood ($\phi_1, \phi_2, \dots, \phi_{r-2}$ estimated as fixed effect) would provide highly variable and "wiggly" estimates (in the extreme case, with a knot at each data point, it would perfectly interpolate the data). Between these two extremes, the random effect representation provides the optimal (i.e. best linear un-biased predictor, Speed, 1991) trade-off between excessive smoothing and overfitting of the non-linear function.

This specification can be estimated as a standard random effect model, i.e. by ML or Restricted Maximum Likelihood, (REML). By estimating each smoothing parameter λ as $\sigma_u^2 / \sigma_\phi^2$, these techniques resolve the subtle task of determining the model flexibility *a priori*, by incorporating this choice into the actual estimation process. Another important feature of this method is that, if a non-linear relationship is not supported by the data, the corresponding smoothing parameter will automatically be estimated to have a high value, the resulting random effect will be close to zero and the smooth function will reduce to a standard linear form. Moreover, this approach can also be

extended to bivariate functions, in order to flexibly capture any joint non-linear effects of two explanatory variables. In this paper we model the impact of rainfall and temperature on land value by using tensor products (Wood, 2006a), which have the important properties of being invariant to changes in the scale of the regressors and can, therefore, be used to smooth variables expressed in different units. Finally, since this estimation technique expresses smoothing splines as random effect terms, inference can be implemented within the standard framework for this class of models (Pineiro and Bates, 2000; Ruppert, Wand and Carroll, 2003). For instance, model reduction can be implemented with likelihood ratio tests for hypotheses on the random effects and with F-tests for hypotheses on the fixed effects. However, as in standard random effect models, testing for the random effects will be only approximate since it involves setting the variance of certain components of the random effects to zero, which is on the boundary of the parameter region (Stram and Lee, 1994).