7 Modelling climate impacts

7.1 Introduction

Although econometric theory (Section 7.1.1 and 7.1.2) suggests that the statistical modelling methodologies (OSR and ANN) discussed previously (Chapter 4.2) may be appropriate for socio-economic upscaling analysis (Chapter 1.3), econometrics is not the only method by which socio-economic impacts can be assessed in a quantative fashion. Other approaches, examples of which are discussed below, offer distinct advantages and disadvantages.

The Ricardian approach is among the leading methods for measuring sensitivity to climate (Mendelsohn et al., 2000). It assumes that climate variables may affect profitability of a given activity, e.g. agriculture (Mendelsohn et al., 1994), forestry (Mendelsohn and Sohngen, 1996), or energy production (Morrison and Mendelsohn, 1998), and therefore the value of the land upon which such activity takes place (Mendelsohn et al., 1999). The Ricardian method uses land prices as a cipher for costing socio-economic impacts. Adaptation measures such as the substitution of crops or other inputs are therefore implicitly incorporated in the method, which can include a wide range of potential forcing factors and measures sectoral impact in a form that can be easily understood by decision makers (Mendelsohn et al., 1994; Mendelsohn et al., 2000). However, the Ricardian approach may not thoroughly isolate climate influence from other important factors (Mendelsohn et al., 2000), cannot be specifically utilised for impacts upon types of crop (e.g. wheat, or maize), or energy production, or forestry (Mendelsohn et al., 1994), and it is not useful where a given form of activity is not directly tied to land price (e.g. energy consumption, or mortality).

The macro-economic method relies upon an extended input-output matrix of supply and delivery concerning economic goods (Aaheim and Schjolden, 2004). Due to the interlocking nature of both economic systems and the input-output matrix, the method is highly effective at placing downscaling studies into a broader perspective (particularly in terms of productivity) and analysing impacts that may have knock-on effects from one sector to another (Aaheim and Schjolden, 2004). In addition, extreme climate events can be represented within an input-output model as shocks to the economy (in terms of effects upon productivity). However, macro-economics requires the aggregation of climate conditions to the same scale as socio-economic data, must assign economic value to impacts for cross-sector analysis, and requires many assumptions based upon detailed knowledge of national economic systems at the micro-level (Aaheim and Schjolden, 2004).

Although econometric predictand indices must be carefully chosen to avoid or incorporate issues specific to the sector under consideration (Chapter 6.2), the econometric models used within this study (Section 7.2) measure impacts upon mortality in lives, and upon agriculture in yield per hectare. These impacts are not easily compared between sectors, and interactions between both sectors and countries are left for future work (Chapter 8.4). However, advantages of the econometric approach used here include:

- an empirical basis,
- a methodology that can be applied consistently between sectors, while applying the same (linear and non-linear) function fitting methods used for climatological downscaling in previous chapters.
- the avoidance of substantial difficulties associated with the valuation of human life (a problem for any cost-benefit based analysis concerning mortality),
- and cross-scale analysis, without the need for aggregation of climate extremes.

Examples of the use of econometric modelling techniques, and a review of their application for this chapter, are therefore given in Section 7.2. The most common form of econometric modelling (OLS regression) is also summarised (Section 7.2.1), and applied so that results may be compared to earlier work. The performance of OLS regression, OSR, and ANN models is discussed in Chapter 7.3 for the socio-economic indices defined in Chapter 6.2.2. The performance of a given method may, however, vary from region to region. The regional nature of sensitivity is discussed in Chapter

7.4.1. Where econometric models successfully replicate socio-economic indices, a strong statistical link exists (Chapter 6.4) and the previous discussion suggests a physical process for the link (Chapter 3.4 and Chapter 5.3), weights of predictors can then be utilised to quantatively estimate the sensitivity of socio-economic activity to climate (Chapter 7.4.2).

7.1.1 Econometrics

Any given econometric equation consists of a dependent socio-economic variable, a number of independent explanatory variables, and an error term (Gujurati, 2003). For example, *economically*, the demand for any household commodity (D):

$$D = f(\text{the cost of available commodities, household income}) \qquad \textbf{(7.1)}$$

Where the given commodity may be, for instance, electricity (Anderson, 1973). However, if, for the sake of simplicity, we assume a linear relationship, *econometric* demand for electricity might be written as:

$$D_e=a_0+a_1X_1+a_2X_2+u \qquad \textbf{(7.2)}$$

Where X_1 represents the cost of electricity, X_2 represents household income, a values are coefficients (including a constant/intercept value a_0), and u the stochastic disturbance (error) term, which includes all the factors not taken into account (which in this oversimplification are likely to be numerous) (Anderson, 1973; Gujurati, 2003). For the above example, variables might include factors such as the costs of competing commodities (e.g. oil and gas), or of related equipment (e.g. heating systems), or of climatic variables such as temperature (Anderson, 1973). Where climate variables are utilised in econometric models they are usually accompanied by socio-economic factors also related to the activity under consideration (Anderson, 1973; Galeotti et al., 2004; Bigano et al., 2004).

The definition of the variable under investigation can substantially affect the explanatory variables that may be relevant in, for instance, Equation 7.2. Aggregated electricity demand per household (AD_e) for a given region reflects average characteristics of households within that region, and therefore only variables that are likely to vary significantly from region to region are relevant within a given model (Anderson, 1973; Gujurati, 2003). A model for comparing the energy demand of individual states within the U.S. might, therefore, be described as:

$$AD_e = a_0 + a_1 H + a_2 R + a_3 TX + a_4 TN + u \qquad \textbf{(7.3)}$$

Where H is the average size of households for a given region, R is the ratio of urban to non-urban population for the region, TX is a measure of heat, and TN a measure of cold (Anderson, 1973). Similarly, electricity consumption may vary as in equation 7.3, but with a measure of wealth emplaced as goods in the place of household size (Chapter 5.3.4). The models used for the rest of this chapter are all variations upon this form (Equation 7.3), using the socio-economic indices (e.g. excess mortality, electricity consumption, yield anomaly) defined in the previous chapter (Chapter 6.2.2) as dependent variables, and the climate indices (e.g. consecutive dry days, rainfall intensity) defined in Chapter 3.4.1 and discussed in Chapter 6.4 as independent variables. Where relevant, additional socio-economic independent variables (e.g. proportion of land irrigated), as discussed in Chapter 6.4, have been included within a given model.

7.1.2 Terminology and assumptions

Econometric models are subject to many of the same constraints that apply to models for climate analysis (Table 3.9). However, terminology differs between the two fields. Many of the terms specific to econometric modelling are related to the choice of an appropriate functional form (Chapter 1.3), and the strict need to meet assumptions relevant to that form. Table 7.1 presents the issues related to functional form in econometric language, but model requirements are generally identical to those given in Table 3.9.

Complex temporal lags and serial effects may influence excess mortality (Chapter 5.3.5). One of the assumptions of econometric theory, as applied to classical linear regression models (e.g. Equation 7.3), prohibits the use of lagged values of the dependent variable as an independent variable (Table 7.1). However, the inclusion of lagged dependent variables in econometric models is not uncommon (Peirson and Henley, 1994; Bigano et al., 2004), as a range of techniques exist for safely bypassing this assumption through mathematical manipulation (e.g. partial adjustment models, adaptive expectations models, the general method of moments approach) (Johnston, 1972; Anderson and Hsiao, 1982). Unless applied with great care, these techniques may lead to further problems, including invalid test procedures, and bias and inconsistency in the estimation of model coefficients (particularly problematic for small sample sizes) (Gujurati, 2003; Bigano et al., 2004). In this study, the use of adjustment and expectation techniques has been left for further work, and instead lagged independent climate variables are included in the model to compensate for effects that may persist from one season to the next (Agnew and Palutikof, 1997).

7.2 Econometric application of models

Both of the models discussed in Chapter 4 are utilised for econometric analysis in this Chapter. However, the Basin wide OSR approach (BOSR) has not been applied as it has been shown that statistical links to climate vary substantially from one country to the next (Chapter 6.4). Additionally, models have been applied with respect to the sensitivity testing discussed in Chapter 4.2.3. Additional testing using socio-economic data has found that the same starting conditions apply, but that the Thin Plate Spline basis function neural network consistently provided lower skill than neural networks constructed with other basis functions. No other form of basis function (Chapter 4.2.2) could be rejected. All the models discussed in this chapter have been constructed with a pre-selected climate index predictor set (Table 6.2) restricted by the results of correlation testing (Chapter 6.4).

In this part of the study, the (Regional) OSR model described in Chapter 4.2.1 is applied only to mortality and agricultural yield data, as there is not enough data available to apply OSR to electricity consumption. For both an ordinary least-squares regression method (Section 7.2.1) and the neural network approaches, electricity consumption models that are unlikely to provide complete calibration and verification periods (Chapter 4.2.1) for the entirety of the bootstrapping process through missing climate data have also been removed from the analysis (listed in Table 7.2). A brief review of the models utilised in this chapter, as applied econometrically, is given below.

7.2.1 Linear regression

Equations 7.2 and 7.3 are both examples of linear econometric models. The coefficients of these models are generally solved via the use of ordinary least-squares (OLS) regression (Gujurati, 2003; Kennnedy, 2003; Barreto and Howland, 2006). When applying econometrics to climate variables OLS regression is a common form of analysis (e.g. Anderson, 1973, Galeotti et al., 2004; Bigano et al., 2005). OLS regression is described in detail by Wilks (1995), and Von Storch and Zwiers (1999). Linear regression models demand that equations are linear in the parameters, but not necessarily) in the variables considered (Barreto and Howland, 2006). In most of the OLS regression literature referenced above, squared variables are entered as equation parameters in order to take the potentially u-shaped distribution of temperature based impacts (e.g. Fig 5.3) into account within the functional form. Although associated errors become a function of the manipulated variable (the parameter), rather than the original variable, manipulation of variables in this way, prior to their entry as econometric parameters, does not violate any of the conditions required for analysis. As an example:

$$EM = a_0+a_1X_1+a_2X_2+a_3X_3+a_4X_4+u \qquad \textbf{(7.4)}$$

Where Expected Mortality (*EM*) is a function of time (X_1), influenza deaths (X_2), temperature (X_3) and temperature squared (X_4) and the function shown is a valid polynomial econometric equation that can be solved by linear regression (Bentham, 1997).

Although trends due to time have been absorbed into the dependent variable (Chapter 6.2.2) this study otherwise follows the methodology used by Galeotti et al. (2004). Using the results of Chapter 6 to determine potential predictors (Table 6.2) a first estimation OLS regression is performed. The estimates of each coefficient (a_0, a_1, a_2, ...) are then checked for statistical significance (at the 0.05 level) and the residuals (observed minus simulated time series) are tested for stationarity and the need for non-linearity. Where a u-shaped distribution is apparent in the residuals, for instance, squared variables are introduced (with respect to relationships evident within the literature presented in Chapter 5.3) before the regression is estimated again. Where parameters are insignificant they are removed and the OLS regression is re-estimated (Figure 7.1).

For this study a substantial volume of station-level data is available. Classical OLS regression cannot, however, be used for simultaneous analysis of multiple climate sites due to collinearity issues (Gujurati, 2003). For each national-scale socio-economic (dependent) index (e.g. Greek winter Excess Mortality), separate models have therefore been created for each climate station, drawing predictors from the independent climate indices at that site (e.g. minimum temperatures at Athens). In some cases only one significant model can be found (at one station) for each country, in others, many similar models (at multiple stations) can be found across the region. In the latter case, the variation of coefficients, skill, and error between models is generally small (i.e. less than 5% of mean values).

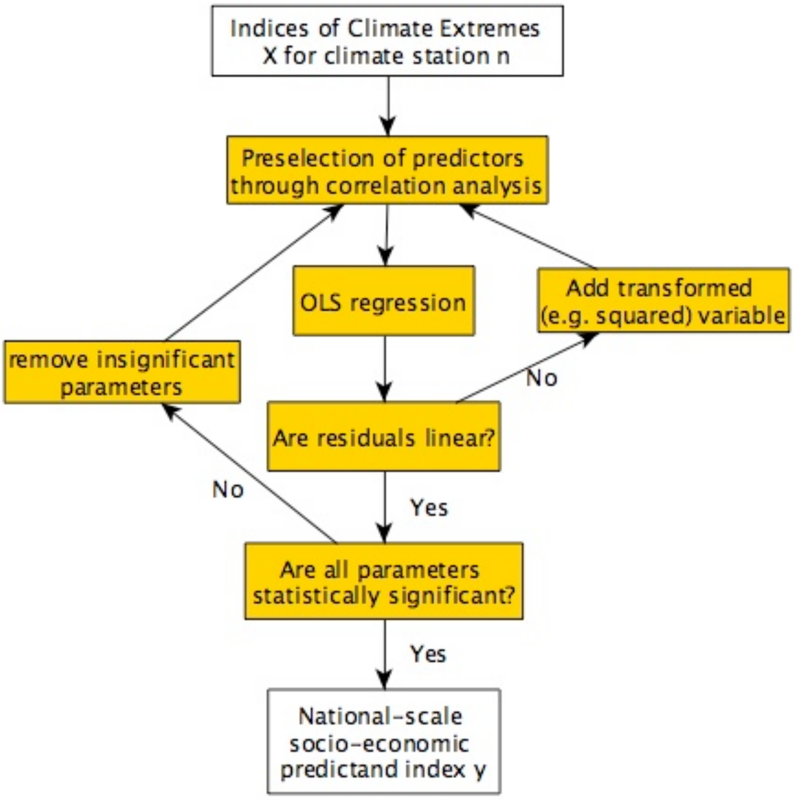

Figure 7.1: Flow diagram of OLS regression process. X refers to all temperature and precipitation climate indices. This process is repeated for all *N* climate stations (*n*), and all *Y* socio-economic predictands (*y*). *N* refers to all stations relevant to a given national socio-economic predictand (e.g. all Portuguese stations for a Portuguese Electricity Consumption model).

7.2.2 OSR

OSR has been utilised to solve issues of multicollinearity in both climatological analysis (Jones et al., 1987) and econometrics (Johnston, 1972; Fomby et al., 1984; Barreto and Howland, 2006). Econometrically, PC regression can be mathematically reduced to a restricted least squares estimator that satisfies all the conditions shown in Table 7.1: for a full proof see Fomby et al. (1984). Although the PC regression method is relatively common in economic applications to trade, market change (Alexander, 2001), monetary policy (Favero et al., 2005), and other financial sectors, recent examples of PC regression in Mediterranean applied econometrics are scarce.

Oliveira (2006) uses a methodology analogous to OSR to counter multicollinearity when constructing predictors influencing Portuguese fertility. Principal components regression has also been utilised to model the electricity load and consumption (for one month) of France, Greece, and Italy using temperature-based parameters as predictors. (Manera and Marzullo, 2003). Results of the Manera and Marzullo (2003) study suggest that for electricity consumption modelling, PC regression may outperform other forms of analysis (i.e. fourier analysis and smoothing spline estimation).

Unlike OLS regression, the OSR method can be applied to multiple climate sites simultaneously. In this study one model is developed for each socio-economic dependent variable (e.g. Wheat Yield) for each country (e.g. Portuguese Maize Yield), drawing predictors from all independent climate indices at all stations for the relevant country (e.g. all temperature and precipitation indices, lagged by up to a year, for all Portuguese stations). The number of extreme climate index predictors is restricted based upon both the results of correlation testing (Chapter 6.3) and the variance-based deletion of components (Chapter 4.2.1), as previous analysis (Chapter 4.3) has shown that prior selection and restriction of candidate predictors may result in greater OSR performance, and rarely reduces levels of skill (Chapter 4.4.1). The inclusion of the OSR approach (Fig 7.2) at this stage allows for an assessment of whether or not a single model that constructs spatially aggregated predictors (with respect to variance) is more appropriate than a single model at one predictor site, or multiple models over multiple sites.

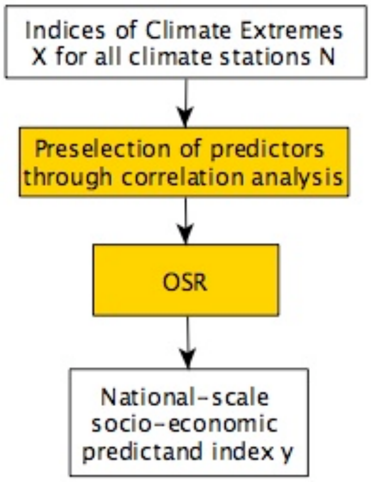

Figure 7.2: Flow diagram of econometric OSR process. This process is repeated for all *Y* socio-economic predictands (*y*). *N* refers to all stations relevant to a given national socio-economic predictand (e.g. all Portuguese stations for a Portuguese Electricity Consumption model). Chapter 4.2.1 also discusses the OSR model .

7.2.3 ANN

The RBF network discussed in Chapter 4 can be described, in econometric terms, as a nonlinear least squares estimation approach with a highly flexible functional form (Kennedy, 2003), fixed via a large number of iterations. Neural networks have found uses in econometrics, much as they have in climatology, particularly in the field of power generation and forecasting (Feinberg and Genthliou, 2005). Several ANN methods have been used for the forecasting of Greek energy load (Papalexopoulos *et al.*, 1994). Beccali *et al.* (2004) have successfully utilised neural networks to forecast urban and suburban electricity load in Palermo (Italy). However, although neural networks have proven useful in the analysis of satellite derived land cover and the potential for Mediterranean crop development (Moriondo and Bindi, 2006), there is a general lack of studies that focus upon the econometric application of neural networks to Mediterranean agriculture, or mortality. Further afield, econometrically applied nonlinear ANN techniques have been favourably compared with regression approaches by Pao (2006). Pao (2006) found that an ANN approach outperformed a selection of regression methods (including multiple log-linear regression, and response surface regression) when attempting to forecast Taiwan's consumption of electricity. Liu et al. (1991) also compare econometric methods and apply a neural network approach to energy consumption for Singapore. However, their study shows that neural networks may not always outperform regression-based forecasts of future consumption. Crowley and Joutz (2005) used ANN results to verify an OLS regression focussed on energy demand, finding that the two gave similar results for the Pennsylvania-New Jersey-Maryland region of the U.S. Electricity consumption models developed for one country may not, however, be appropriate if applied to another (Pao, 2006).

A popular approach to the econometric application of neural networks, where an appropriate functional form is unknown (Pelaez, 2006), is the use of a non-linear 'logit' basis function:

$$y = \alpha + \sum_{i=1}^{k} \beta_i \frac{e^{\theta_i}}{1 + e^{\theta_i}} + \varepsilon \qquad \textbf{(7.5)}$$

However, with enough logits (*k*), these highly flexible basis functions are capable of fitting any functional form, can be particularly prone to overfitting (Kennedy, 2003), and tell us little about the appropriate function for a particular model. In this study the radial basis functions given in Chapter 4.2.2 are used so that, where non-linear performance is greater than linear performance, the improvement can be assigned to a particular mathematical function. However, where nonlinear function neural networks provide greater performance than linear networks, uncertainty over the activation of the basis function may lead to further uncertainty over the internal weighting structure of a given network (Chapter 4.2.2). As with the climatologically applied neural networks in Chapter 4, rigorous calibration and validation (Chapter 4.2) is essential in order to avoid overfitting (Kennedy, 2003). Neural networks are limited in application by multicollinearity issues between predictors (Pao, 2006; Chapter 4.2.2). They are therefore applied here in a similar fashion to the OLS models detailed above (Section 7.2.1), to produce an array of models across each country (one for every climate station site) (Fig 7.3). As ANNs weight irrelevant predictors to zero internally, all models produced by this method possess only statistically significant predictors.

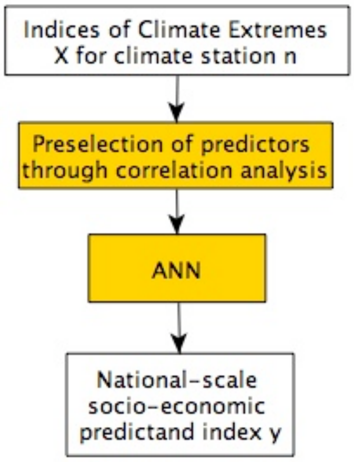

Figure 7.3: Flow diagram of the econometric ANN process. This process is repeated for all *N* relevant climate stations (*n*), and all *Y* socio-economic predictands (*y*). *X* refers to all temperature and precipitation climate indices. *N* refers to all stations relevant to a given national socio-economic predictand (e.g. all Portuguese stations for a Portuguese Electricity Consumption model). Chapter 4.2.2 also discusses the ANN model.

7.2.4 Summary of models

In this section the most commonly utilised form of econometric model (OLS regression) has been introduced so that results can be compared between novel applications of econometric upscaling models and a more traditional method. In addition, the models developed in previous chapters have been placed within an econometric context. The methods used throughout the rest of this chapter are summarised below and in Table 7.3:

- An OLS regression applied for one socio-economic predictand (Chapter 6.2) at a time, for each country, utilising a limited number of extreme climate index predictors (Table 6.2) from one relevant station (Table 3.1) at a time.
- A regional OSR model (ROSR) applied for one socio-economic predictand (Chapter 6.2) at a time, for each country, utilising a limited number of extreme climate index predictors (Table 6.2) from all relevant stations (Table 3.1) simultaneously.
- A Linear RBF neural network (LBF) applied for one socio-economic predictand (Chapter 6.2) at a time, for each country, utilising all

extreme climate index predictors from one relevant station (Table 3.1) at a time.

- Gaussian, Thin Plate Spline, Multiquadratic, Inverse Multiquadratic, and cubic RBF neural networks applied in an identical fashion to the Linear RBF.

7.3 Performance of upscaling models

7.3.1 Definitions of model performance

This section discusses the performance of each modelling method (Section 7.2) when applied socio-economically. In trying to assess the skill in simulation of socio-economic predictands (Chapter 6.2) as part of a socio-economic sensitivity study, measures of skill and significance that are useful in climatology are just as useful when applied to econometrics (if they are applied with care given to the assumptions required of a given model). Where skill is referred to in the following sections (7.3.2-7.3.5) it has been assessed using r (Gujuratti, 2003) and described using the definitions given in Section 3.3.5 and Table 3.2. Taylor diagrams have also been constructed for the socio-economic models discussed above, as discussed in Chapter 4.3.1 (Figures 7.4-7.19). The RMS error of any OSR or ANN model can be estimated from the distance between any point shown in these diagrams, and the 'perfect model' (r=1, v=1) point (Chapter 4.3.1).

7.3.2 Agricultural Yield Anomaly

This section concerns Table 7.3 (OLS regression results), Figure 7.4 (OSR results), and Figures 7.6 to 7.10 (ANN results).

Citrus

For Citrus Yield Anomalies, the greatest level of skill can be found when utilising the OSR method to model Spanish values (r=0.67) (Figure 7.4). Neural network models applied to the same predictand (Spanish Citrus Yield Anomaly) result in less skill (Figure 7.6), but in the case of the linear basis function neural network, much better simulation of Citrus Yield variance.

For the Portuguese Citrus Yield Anomaly, the OSR model produces a high level of skill (r=0.51). OSR based estimates of Portuguese Citrus Yield Anomaly (variance underestimated) show much lower levels of error than those for Spain (variance overestimated), and show performance comparable to the upper limits of skill found with the Gaussian, Multiquadratic and Inverse Multiquadratic basis function neural network models.

Both Greek and French OSR performance is poor, but French ANN performance is high. The upper limits of French Gaussian neural network show both relatively high levels of skill (r=0.55) and variance estimation (v=0.81). When modelling Italian Citrus Yield Anomaly, the upper limits of neural network skill are very high ($r > 0.90$) using the Multiquadratic, Inverse Multiquadratic and Cubic basis functions. However, errors are also very high, with a very poor estimation of variance. Of these three neural networks, the Inverse Multiquadratic basis function neural network produces the lowest error, although differences between the performance of the Multiquadratic and Inverse Multiquadratic basis function neural networks are minimal. For these models, the estimation of variance improves substantially if choosing a model with slightly lower skill.

It can be seen that nonlinear ANN methods generally produce better performance than linear ANN methods for Citrus Yield (although estimation of variance tends to decline at very high levels of skill). Each of the non-linear functions are capable of modelling relationships with at least one shift in behaviour (i.e. one or many ‘change-points’), unlike linear methods, they are well suited to modelling complex interactions that may vary over a range of values. As an example, the yield of a given permanent crop may increase with temperature up to a given point, then decrease up to a second threshold, and then decrease rapidly. Such a process may reflect an increase in growing period, followed by wilting, followed by scorching (Chapter 5.3.3), and is likely to be more accurately represented by a cubic or quadratic function, than a linear one.

Grape

For French and Spanish Grape Yield Anomalies the greatest levels of skill are provided by ANN methods (Figure 7.7). Upper limits of skill are r=0.69 for France (using a Gaussian basis function), and r=0.61 for Spain (using either Multiquadratic or Inverse Multiquadratic basis functions). However, for both regions, OSR models replicate observed data more skillfully (with r=0.43 in both cases) than the majority of French and Spanish ANN models, with a much better estimation of variance (Figure 7.4), which tends to be very poor for the ANN models. The upper limits of the ANN skill distribution provide very high levels of skill for Italian Grape Yield Anomalies, particularly when using cubic basis function models (r=0.95), but again the estimation of variance is very poor.

Although, in the majority of cases, Greek and Italian, Multiquadratic, Inverse Multiquadratic, and Gaussian ANN models do not perform as well as OSR models, they show upper limit results that are better in terms of both skill (r=0.30 and 0.49), and variance estimation (v=0.95 and 0.98). As for Citrus Yield, the most successful ANN models are generally those that are capable of non-linearities. However, for both France and Spain, the OSR method clearly shows better performance.

Maize

As for Grape Yield performance, ANN Maize models tend to underestimate Yield variance (Figure 7.8). However, for Maize, this overdispersion problem (Chapter 4.3.2) is also evident for the relevant OSR models (Figure 7.4). For all Maize models, as skill increases, variance estimates decline.

The best performance (in terms of both skill and variance) is found with a Cubic neural network for France, a Multiquadratic neural network for Portugal, and a Linear neural network for Spain (Figure 7.8). For Greece, OLS regression skill exceeds that gained when using ANN models (which is generally less than r=0.30) resulting in a value of r=0.40 and comparatively low error (Table 7.4). It is clear that to successfully model Maize Yield across the Mediterranean a range of functional forms is required.

Potato

The highest levels of skill that can be found when modelling French and Italian Potato Yield Anomaly models (0.65 r for France, and 0.41 r for Italy) are both gained through OSR (Figure 7.4), suggesting that aggregate conditions are useful for prediction, rather than the variance of climate at a given site (discussed in further detail in Section 7.4.1). For Greece the OLS method (Table 7.4) gives the highest levels of skill (0.60 r). Although Cubic function neural networks demonstrate relatively high levels of skill for Greece (Figure 7.9), they are not as high as the OLS results, and ANN performance is otherwise poor. Variance is substantially underestimated for all basis function neural networks.

French and Greek Potato Yield Anomalies are more successfully modelled with OSR than French and Greek yield anomalies for any other crop, although the Mediterranean climate varies substantially between the two regions (Chapter 3.4.2). The one quality that both regions share is a relatively cold winter regime, which has been shown to possess a statistical link with anomalous Potato Yield (Chapter 3.4.2 and Chapter 6.3.2).

The best results for Potato Yield Anomaly are all gained through the use of linear methods.

Wheat

Wheat Yield Anomalies are modelled with comparable levels of skill for OLS regression models (Table 7.4) for Italy (r=0.42), Portugal (r=0.47), and Spain (r=0.47). The highest levels of skill are also found when using OLS regression to model Wheat Yield for Greece (r=0.58). For Greece the greatest performance is produced with OLS regression models for all Yield Anomalies (i.e., for all crops). There may be a linear quality within Greek agriculture that persists between crops.

ANN (Figure 7.8) and OSR models are generally not as effective when modelling Wheat Yield Anomalies. OSR performance is generally very poor, and the majority of ANN performance is comparable. However, Linear models applied to

French and Spanish Yield may produce performance (for single sites) comparable to that for OLS regression. It is clear that in some instances sophisticated models may not offer much improvement in performance over relatively simple linear regression methods (Liu et al., 1991), as Wheat Yields are generally simulated better by linear OLS regression than through the other methods considered here.

7.3.3 Electricity Consumption

Issues concerning the availability of data and the nature of OSR methodology have meant that meaningful OSR models cannot be constructed for Electricity Consumption (Section 7.2.3), and that when applying OLS and ANN models, certain stations cannot be used to provide extreme climate predictors (Table 7.2). This section concerns Table 7.5 (OLS regression results), and Figures 7.11 to 7.14 (ANN results).

Winter Commercial Electricity Consumption

For Spanish, Portuguese, and Italian Winter Commercial Energy Consumption, OLS regression models provide the greatest skill (r=0.93, 0.55, and 0.44 respectively) with very low errors (Table 7.5). Greek Winter Commercial Energy Consumption is simulated most successfully by Cubic and Linear (r=0.70) radial basis function neural network models (Figure 7.11). For French Winter Commercial Energy Consumption, Multiquadratic, Inverse Multiquadratic and Gaussian models produce the greatest levels of skill (r=0.77).

The requirement for different functional forms for different countries may reflect differences in energy consumption dependent on climate regime. As seen in Chapter 5.3.4, winter energy use may decline in a near-linear fashion with decreasing temperature to a balance point, and then increase non-linearly, as large deviations from the balance point may require large increases in heating. Where winters are relatively mild (Italy, Portugal, and Spain) any decrease in winter temperature may require a linear increase in heating. Where winter regimes are colder (France and Greece) and lower temperature values are reached (Chapter 3.4.2), non-linearities may appear in the relationship between electricity consumption and temperature.

Winter Residential Electricity Consumption

Winter Residential Energy Consumption is the most skilfully modelled index of those considered in this study. For all regions, skill for Winter Residential Energy Consumption models is substantially greater than for Winter Commercial Energy Consumption. This contrast between relationships is suggested by discussion in Chapter 5.3.4. In addition, where both linear and non-linear model functions are successful for Commercial Electricity Consumption, the most successful models for Residential Consumption are all linear. Linear basis function neural networks produce the most skilful Winter Residential Energy Consumption models for all countries, with an average (not upper limit) skill of r=0.93 r for France, r=0.96 for Greece, r=0.81 for Italy, and r=0.96 for Portugal and Spain. Although OLS regression models produce high levels of skill and low errors (Table 7.5), Linear basis function neural network performance (in terms of both skill and error) is superior (Figure 7.12).

Summer Commercial Electricity Consumption

Summer Commercial Energy Consumption is successfully modelled only for France. The OLS regression method (Table 7.5) outperforms all ANN basis function models r=0.78), with the greatest level of ANN skill (Figure 7.13) acquired with Linear basis function neural network models (r=0.51, but poor estimation of variance). Errors associated with the OLS models applied to summer values are substantially higher than those for winter, however. Compared to winter values, Summer Commercial Electricity Consumption is generally poorly simulated.

Summer Residential Electricity Consumption

As for Summer Commercial Energy Consumption, application of the OLS regression method to French Summer Residential Energy Consumption results in the greatest levels of skill (r=0.85). OLS regression also demonstrates the greatest skill for Portugal (r=0.66), although errors are higher than for French models (0.67, compared to 0.52 for France). ANN models perform very poorly for all countries except France, which displays a very poor estimation of variance (Figure 7.13). Again, the best results are all gained using linear methods, and OLS outperforms the more sophisticated methods deployed in this study (Liu *et al.*, 1991).

7.3.4 Excess Mortality

This section concerns Table 7.6 (OLS regression results), Figure 7.5 (OSR results), and Figures 7.15 to 7.19 (ANN results).

Winter

When modelling Excess Winter Mortality the greatest levels of skill are found when modelling Italian values with Multi-quadratic, Inverse Multi-quadratic, Gaussian, and Cubic basis function neural network models (Figure 7.15), all of which produce comparable skill values of up to r=0.64. Average values of skill are low however ($r < 0.32$), and, for the majority (over 75%) of neural network models, the Italian Excess Winter Mortality OSR model (Figure 7.5) offers greater skill (r=0.45). Spanish Excess Winter Mortality values are also modelled well by Multi-quadratic, Inverse Multi-quadratic, Gaussian, and Cubic basis function neural network models at the upper end of the skill distribution (with skill up to r=0.58), but are again outperformed by the OSR approach when considering average performance. It is implied that Winter Excess Mortality for both Italy and Spain is a non-linear process with a single shift in behaviour (as Gaussian basis function neural networks perform as well as Quadratic basis function neural networks), possibly as suggested by discussion in Chapter 5.3.5 (i.e., a u-shaped distribution).

For Portuguese Excess Winter Mortality, the ANN approach performs poorly (for all models) and OLS regression methods are the most successful (r=0.44).

Spring

Results for Spring Excess Mortality illustrate the fact that for some indices different methods are required to produce skilful and significant models for different regions. The greatest performance for Spring Excess Mortality models is found in Greece (r=0.56) using an OSR model (Figure 7.5), for Spain with a Gaussian basis function neural network (r=0.42) (Figure 7.16), for Italy with a Cubic function neural network (r=0.49) and for France with an OLS regression model (r=0.46). However, the neural networks tend to underestimate variance substantially, and the best results are evident only for one or two models in the array, as also seen when modelling other socio-economic indices.

The physical processes that lead to mortality during spring and autumn are less clearly defined than for winter and summer (Chapter 5.3.5), and may vary from one end of the Mediterranean to the other with variations in the dominant climate regime. The spatial variation of the Mediterranean spring climate has been shown to exceed that for autumn (Chapter 3.4.2). It is interesting to note, however, that the two coolest spring climates in the Mediterranean display linear processes.

Summer

The highest levels of skill can be found for for Greek Summer Excess Mortality with an OLS regression model (r=0.44), and for Spanish and Italian values with a Cubic function neural network (up to r=0.85, and r=0.39 respectively) (Figure 7.17). Although, the highest levels of skill are associated with reduced variance (as shown for other socio-economic indices), the Cubic function ANN is less prone to overdispersion (when applied to Summer Mortality) than other forms of network applied to Excess Summer Mortality. A cubic function may be more useful when modelling Summer Excess Deaths where the number of deaths with temperature increases, increases more beyond a given threshold, and then begins to flatten out as the limit provided by the pool of susceptible population is approached (Chapter 5.3.5).

The very high level of skill that can be gained for Spain implies a greater Spanish link between climate and mortality during summer than other seasons, as suggested by discussion in Chapter 5.3.5, but also a greater link than for other regions, which has not been suggested by the literature. There is, however, a general lack of work quantitatively comparing relationships between mortality and climate between countries (Chapter 8.4).

Autumn

As for Spring results, the most successful method of modelling Autumn values of Excess Mortality largely varies from country to country. The greatest skill can be found Spain and Greece with OSR models (up to r=0.61 and r=0.56 respectively, although error is very high in the latter case) (Figure 7.5), for Portugal with an OLS regression model (r=0.56) (Table 7.6), and for Italy with a Linear basis function neural network (r=0.65) (Figure 7.18).

Although the neural network approaches are competitive (compared to the OLS and OSR methodologies also used in this chapter) only when considering the upper end of the skill distribution, the highest levels of skill may be associated with poor replication of variance, as with the other socio-economic indices discussed in this chapter. Skill in replicating autumn excess mortality, in all cases, exceeds that for spring. It should be noted that Greek mortality is skilfully modelled with a linear functional form for all seasons but winter.

Elderly Mortality

Again, skill for ANN methods applied to Elderly Mortality (Figure 7.19) can display very high values (r >0.90), but variance is poorly estimated, ANN models with lower errors, however, can be seen for both Italian (r=0.39, v=1.22), and Greek (r=0.35, v=0.91) applications of the Cubic basis function neural network. For Greece, comparable levels of performance (r=0.39) can be gained through the use of an OLS regression (Table 7.6). For other models, performance is not as good. Skill for a Spanish application of the OSR model is comparable to the Greek value (r=0.36), but errors are substantially higher.

7.3.5 Summary

Given the results discussed above it is difficult to recommend any one modelling approach over any other for a given application, although certain points can be summarised from the previous Sections:

- In some cases it can be seen that socio-economic predictands can be skilfully replicated using a set of purely or largely climate-derived predictors (e.g. Winter Residential Electricity Consumption)

- The most skilfully modelled Yield Anomalies are for Italian permanent crops (Citrus and Grape), with nonlinear basis function (Cubic, Multi-quadratic, Inverse Multi-quadratic) neural networks. However, at the upper limits of skill, estimation of crop variance is poor.
- The most successful models for Grape are also generally nonlinear basis function neural networks.
- Models applied to Maize Yield are particularly prone to the overdispersion problem (i.e. an underestimation of variance).
- The 'best' Wheat Yield models are those utilising an OLS regression approach.
- Potato Yield Anomaly models are also all linear (either OLS or OSR).
- Winter Electricity Consumption is more successfully modelled than other socio-economic predictands, particularly in terms of Residential Consumption.
- Winter Electricity Consumption is more successfully modelled than Summer Consumption, and Residential Consumption more than Commercial Consumption.
- The Electricity Consumption models that demonstrate the greatest performance are linear, except those for Winter Commercial Consumption in France and Greece.
- The best Greek models for all socio-economic predictands, except Winter Commercial Energy Use, are linear.
- The functions that provide the highest level of skill for Excess Mortality vary highly from one region to another, and may be linear in one season and non-linear in the next.
- The highest skill acquired when modelling Excess Mortality is for Spanish Summer Excess Mortality, with a cubic function non-linear network.
- Spanish Excess Mortality models generally perform relatively well in comparison to those for other regions. Otherwise there are few consistencies between results for seasons or model functions.

As stated by Pao (2006) different regions generally require different modelling methods to accurately capture the variability of relevant socio-economic factors.

Sophisticated techniques may not always provide the greatest levels of skill (Liu *et al.,* 1991), and the best results can only be gained by utilising a number of different techniques. There are, however, certain methodological advantages and disadvantages from using one model over another:

- Because the development of OLS regression models is often 'manual' rather than automatic (i.e. is conducted via the inspection of residuals and repeated testing) the most parsimonious functional form may be developed. In instances where relationships require only a small number of predictors, OLS regression methods may be the most appropriate approach.
- OLS regression modelling can utilise non-linear transformations of variables, as long as model parameters are linear.
- OLS regression is, however, the most time-consuming of the above techniques in terms of application.
- Although strictly linear, an OSR approach greatly reduces the number of models required, as it is largely immune to multi-collinearity issues and can be applied with the same predictors over a given region, rather than for individual sites.
- OSR also includes a preliminary spatial analysis stage (in that is conducts PCA as a first step), so results should reflect the need for either a large-scale aggregate index, or a smaller-scale aggregate as determined by climate variability.
- OSR generally will not utilise very (spatially or statistically) small components of variance however, so if climate conditions at a single station are particularly useful as predictors, but all others are not, performance will suffer.
- A collection of individual ANN models, developed for every relevant set of climate predictors (i.e. for every climate site) does not exhibit this problem, and when looked at as a distribution it can be seen whether one or many models/sites are statistically relevant. The issue of a 'one site or many' approach is discussed further below (Section 7.4.2).
- However, the ANN models that produce the highest levels of skill may not be the most appropriate, as such models often greatly underestimate variance, and are thus prone to high errors.

- The ANN approach is capable of implicit non-linearity, and when applied with a selection of basis functions it can be seen whether there is a discrepancy in skill between linear basis functions and non-linear basis functions. In the above discussion it can be seen that non-linear methods often return similar levels of skill, so a relatively simple distinction can be made between linear and non-linear models.
- Cubic basis function neural networks often return high levels of skill, but may also possess the greatest errors. Either Gaussian (less flexible) or Multi-quadratic (more flexible) approaches may offer substantially less error for a small reduction in skill.

7.4 Upscaling model structure

It can be seen that appropriate functional forms are determined largely by the socio-economic activity and region under consideration. In this Section the issue of climate sensitivity is explored further, as the models discussed above are analysed in terms of their predictors (where models exhibit relatively high levels of skill). This study is interested in the forms of climate variability that produce the greatest change in socio-economic activities (i.e. the most appropriate predictors and their weightings) (Section 7.4.2), and the locations that, when climate extremes occur at them, such activities are most sensitive to (Section 7.4.1).

7.4.1 Spatial sensitivities

The division between countries and socio-economic indicators for which local climate at every applicable site can be utilised to analyse national socio-economics, and others where only one or two sites are useful, is explored further through the use of OLS (Tables 7.4-7.6) and ANN results. Where ANN models perform relatively well (Section 7.3.2-7.3.4), ANN skill values are shown as map plots of all models calculated (Figures 7.20-7.22). However, for the regions of highest skill, errors may be high (Section 7.3.2-7.3.4). Spatial issues are less relevant to OSR methods, as they automatically discriminate between the large-scale (country wide) or small-scale (a small number of stations) aggregation of predictors through PCA (Chapter 4.2.1 and Chapter 7.2.2).

Agriculture

This section concerns Table 7.4 (OLS regression results), and Figure 7.20.

Results for Citrus Yield models (Figure 7.20) show that although national aggregates of climate are generally inappropriate, high levels of skill may be gained when using multiple climate sites. Models for central Spain, southern Italy, Paganella (in northern Italy), and north-west Greece all produce relatively high levels of skill. The Linear Spanish neural network model shows regional sensitivity consistent with the agricultural regions most associated with Citrus growing, as do Italian models (World Book, 2006). However, Italian Yields also display sensitivity to climate conditions in the south, where summer and spring temperatures (Table 6.2) are high (Chapter 3.4.2). Greek Citrus Yields are most sensitive to conditions in the north-east, consistent with both growing region (World Book, 2006), and low spring temperatures (Chapter 3.4.2), which may negatively affect Citrus Yield (Chapter 6.4.2).

Grape Yield results (Figure 7.20) illustrate a wide range of regionality. For Spain, the majority of stations show strong links with Grape Yield, for Italy, only one (Paganella, in northern Italy), and for Greece, a number of stations, largely in the south-east (consistent with growing region) (World Book, 2006). Paganella is the most northern, and highest, station within the climate data set (Table 3.1). It has been shown that climate model skill for Paganella is high for multiple seasons, indices, and models (Chapter 4.3.3). It may be that due to its position, the climate at Paganella is particularly representative of large-scale mean climate throughout northern Italy. This would explain the high levels of skill found when modelling Grape yield with Paganella climate data, and also the relatively poor estimation of variance, which may be determined by regional variations in climate. Greek Grape Yield has been shown (Table 6.2) to be (positively) sensitive to autumn dry periods (PCDD) and spring high temperatures (TX90), both of which are higher in the south-east than the north-west. For Spain and Italy, Maize Yield performance displays similar regionality. As the regions generally associated with Maize are not associated with Grape, it seems likely that large-scale (national) climate is more important to Yield sensitivity for both Spain and Italy.

For both Potato and Wheat Yields, the OLS approach outperforms ANN methods (Section 7.3.2), and it can be seen (Table 7.4) that although skill is high throughout Greece (both Wheat and Potato) and Portugal (Wheat), predictors are otherwise more useful when drawn from single sites. For Wheat Yields, climate conditions at Alicante (eastern Spain) and Forli (north eastern Italy) are useful as predictors. Both of these regions of sensitivity are consistent with regions of Wheat farming. Frequent summer rainfall may benefit Wheat Yield (Chapter 6.4.2), and for Spain such conditions are generally a result of easterly flow into the Iberian interior (Chapter 3.4.2). The economy of Alicante is largely agricultural (Britannica, 2006), wheat is a major local crop (World Book, 2006), and the region is positioned such that it is likely to receive easterly rainfall when it occurs. Cold but dry conditions during winter and spring are conducive to good Italian wheat yields (Chapter 6.4.2), and Forli is northern, largely sheltered from westerly flow, and within the largely agricultural Emilia-Romagna region of Italy (Britannica, 2006).

Electricity Consumption

This section concerns Table 7.5 (OLS regression results), and Figure 7.21.

For Winter Commercial Consumption, Neural networks produce high levels of performance for both Greece and France (Figure 7.21). It is clear that, in the former case, Electricity Consumption influences regional skill, rather than regional climate variability. For Greece, the greatest levels of skill are gained using stations close to (or located within) large cities (i.e., Athens and Thessaloniki). However, the greatest levels of skill for French Winter Commercial Consumption are not found using predictors representative of conditions within large cities (Bordeaux and Marseille), but those nearby (Agen and Montelimar). Winter temperatures for Agen and Montelimar are consistently lower than those for coastal (and highly urbanised) Bordeaux and Marseille (Chapter 3.4.2), and it may be that regional performance illustrates sensitivity to inland (more rural) conditions, in a similar fashion to the displacement of sensitivity seen above for Italy and crop yield.

OLS methods show greater Winter Commercial Consumption skill in Italy, Portugal, and Spain (Table 7.5), and although skill is high throughout both Portugal and Spain, results for Italy are good only when using climate predictors from Porretta Terme (north-eastern Italy). Porretta Terme is close to Bologna, the capital city of the northern Emilia-Romagna region. Winter temperatures for Rome tend to be mild, but winters in the region surrounding Bologna can be very cold (Chapter 3.4.2).

For Winter Residential Consumption, and both Commercial and Residential Summer Consumption, it is clear that regional sensitivities are much less important, as both ANN (Figure 7.21) and OLS methods (Table 7.5) show high skill across very large areas for countries where they perform well.

Excess Mortality

This section concerns Table 7.6 (OLS regression results), and Figure 7.21.

Italian Excess Winter Mortality shows regional skill in northern Italy (Figure 7.21), and specifically for the coldest Italian regions (Chapter 3.4.2) considered in this study (i.e. Paganella, Turin, and Lazzaro), rather than regions consistent with a large population. Spanish Excess mortality appears sensitive to westerly circulation, which is an important factor in Spanish winter climate (Chapter 4.4.2). However, climate data is not available for the Madrid region of Spain (Table 3.1), and results are also consistent with a regional sensitivity that increases toward the Spanish capital. Portuguese Excess Winter Mortality performance is high throughout the country (Table 7.6). Without more data concerning the central region of Spain, it is difficult to assess whether Excess Winter Mortality is more sensitive to climate variability, or regional demographics.

For spring, the sensitivity of Excess Mortality is clearly related to the distribution of population (Figure 7.21), as model performance is greatest when using predictors consistent with large cities (i.e., Barcelona and Rome).

Summer Excess Mortality model performance is high throughout Greece (Table 7.6), as for all other seasons but winter. It is evident that Greek Excess Mortality is sensitive to climate conditions across the country in a linear fashion. Successful Spanish and Italian models are non-linear, however (Section 7.3.4), and in the former case sensitive to conditions in the south-east (the warmest region in Iberia) (Chapter 3.4.2), and near Barcelona (Figure 7.21). Italian skill is (again) relatively high for Rome.

Predictors representative of Rome's climate also produce high levels of skill for an Autumn Excess Mortality model (Figure 7.21). However, levels of skill associated with Turin (another highly populated region of Italy) (Britannica, 2006), are higher, possibly due to the link between Autumn mortality and rainfall (Chapter 6.4.4), which is generally heavier in the north of Italy than the south (Chapter 3.4.2). For Autumn, skill is also high throughout Portugal (Table 7.6). Although it may be a function of the close proximity of the Portuguese climate stations considered within this study (Table 3.1), OLS results generally show levels of skill consistent between Portuguese regions.

The skill shown by Elderly Excess Mortality models seems more consistent with regional climate variability than population density (Figure 7.21). This study does not, however, consider the regional distribution of population by age. Sensitivities are high for southern Greece (the warmest region under consideration) and Paganella (possibly representative of northern Italian climate variability).

7.4.2 Model predictors

As for the statistical downscaling models developed in Chapter 4, the weights/coefficients associated with the predictors used by the most successful econometric models (Section 7.3) are of interest to this study (summarised in Table 7.7). However, although coefficients may offer insight into the socio-economic sensitivity to given climate predictors (Table 6.2) only Electricity Consumption models have been shown to produce particularly high levels of performance in terms of both skill and variance. In some cases Electricity Consumption models show high levels of skill for both linear and non-linear approaches (Section 7.3.2). The following discussion is limited to linear model coefficients for ease of comparison.

It can be seen that for both winter and summer, negative relationships exist between temperature indices and Electricity Consumption. As temperatures increase (decrease) in the western (eastern) basin (Chapter 3.4.4), winter Electricity Consumption is likely to decline (increase). In winter, this effect is significant for average temperatures (across the target region), and the occurrence of frosts (France and Spain). Even after considering that the range of frost day values may be up to 3 times the range of average winter temperature for a given location (Chapter 3.4.2), it is clear that Electricity Consumption is more sensitive to the latter. For summer, and French Consumption, indices of extreme cold are more significant than either average temperature (TAVG), or frost occurrence. These results, when considered along-side discussion in Section 7.3.3, imply that links between winter (and summer) low temperatures and the need for heating, are much more successfully modelled in this study than links between summer high temperatures and the need for cooling (Table 6.2, Chapter 5.3.4). Further, regional sensitivities in the previous chapter are more likely to reflect the need for heating, than cooling. However, as (during summer) French Residential Consumption is sensitive to low temperatures (TMIN) and Commercial Consumption is sensitive to very low temperatures (TN10), it is also evident that in some circumstances indices of extremes may be as, or more useful, than mean conditions when considering Electricity Consumption.

Further, it is clear that there is variation in sensitivity between sectors. For Italy, the average winter Residential Consumption coefficient (-8.44) shown in Table 7.7 is consistent with similar work conducted by Galeotti *et al.* (2004), but Italian winter Commercial Consumption shows greater sensitivity (for both averages, and values apparent for Porretta Terme alone). This sectoral variation also exists for France, Italy, and Spain, but the coefficients shown in Table 7.7 reflect variations in sensitivity to temperature, regardless of sectoral variations in consumption. On average, Residential Electricity Consumption is much smaller than Commercial Electricity Consumption for France, Italy, and Spain, (proportionately- 0.36, 0.27, and 0.17, respectively) (Eurostat, 2006). Although, for a degree increase in temperature the response for Commercial Electricity Consumption is likely to be greater than that for Residential Consumption, the sensitivity of the latter is therefore likely to represent a much larger proportion of total sectoral Electricity Consumption than for the former.

For winter, temperature coefficients for Commercial Consumption are greater for France, Italy, and Spain, than Portugal, or Greece. This regional variation in results may reflect either acclimatisation to temperature, or differences in wealth (Chapter 5.3.4), but a similar variation is also apparent in comparative averages of electricity consumption, and GDP (Eurostat, 2006), while the climate of Portugal is closer to that of Spain, than Greece (Chapter 3.4.2). For both Portugal and Greece, temperature coefficients are also greater for winter Residential Electricity Consumption than Commercial Electricity Consumption. It would appear that the Electricity Consumption behaviour for Portugal and Greece is markedly different from that of the other countries considered here. However, this difference is more likely to reflect socio-economic variation, than regional climate variability.

For Greece, results show that coefficients are greater close to Athens, than for the Thessaloniki region, and the same is true for French values closer to Marseille, than Biarritz. It appears that both model skill, and coefficients, can be seen to reflect a sensitivity to temperature consistent with regional variations in population density, and that this result is common between countries that otherwise differ in their sensitivity to climate.

7.5 Summary

This chapter has shown that through econometric upscaling (Section 7.1) and application of the modelling techniques developed in this thesis (Section 7.2), successful models can be constructed for some aspects of Mediterranean socio-economic activity (Section 7.3) that reflect underlying regional variations in sensitivity (Section 7.4.1). Further, that although overdispersion is generally as important an issue for econometric upscaling as it is for the statistical downscaling of climate (Section 7.3), in some instances (i.e., Electricity Consumption) statistically significant, and distinct, climate sensitivities can be found for both mean climate, and extremes (Section 7.4.2). Further conclusions regarding both variations in extreme climate (with respect to the mean) and their ramifications for the sensitivities found in this chapter, are supplied in the following chapter.

Table 7.1: Model assumptions relevant to functional form, their meaning, and how they have been met in this study (Gujurati, 2003; Kennedy, 2003).

Assumption	Meaning
Specification is correct	Regressors are appropriate for the quantity under consideration, i.e. predictors are physically meaningful and strong statistical relation ships exist between predictors and predictands
	A set of relevant regressors remains constant over the target period, i.e. relationships are stationary
	The relationship between regressors and the dependent variable is linear. This part of the assumption may not apply for non-parametric forms of modelling
Errors are symmetric around zero	Model parameters should have been neither under- or over-estimated, otherwise the intercept/constant value may be biased.
	Least squares regression based methods cannot violate this assumption as the mean of the error term is incorporated in the model constant. Normalisation methods also bypass this assumption.
Errors are homoskedastic	Independent variables and any errors in their measurement are drawn from the same statistical distribution over both the range of values and the target period. Homogeneity checks ensure that this is the case.
	Values are not auto-correlated with each other. The application of models independently by season minimises this problem.
Independent variables alone are responsible for variations in the dependent variable	Independent variables are accurately measured. All indices of climate extremes have been checked for homogeneity
	Lagged values of the dependent variable are not included as an independent variable. Instead, lagged independent variables are included.
	Independent variables are exogenous, and are not functions of the dependent variable in any way.
Independent variables do not display colinearity	Non-linear inter-relations are generally allowed, but otherwise multi-colinearity must be avoided through appropriate application of models, or the use of techniques such as PC-regression.
There are a greater number of observations than independent variables	

Table 7.2: Stations removed from Electricity Consumption analysis due to data availability issues. Sources are credited as ARPA-SMR (Agenzua Regionale Protezione Ambientale – Servizio Meteo Regionale), or FIC/KNMI (Fundación para la Investigación del Clima / Royal Netherlands

	Country	Name	Code	Latitude	Longitude	Elevation	Source
40	Italy	Alghero/Fertillia	16520	4038	817	23	ARPA-SMR
46		Ligonchio	LIGONCH	4432	1035	928	FIC/KNMI
47		Monteombraro	MONTEOM	4438	1100	727	FIC/KNMI
48		Monzuno	MONZUNO	4426	1126	620	FIC/KNMI
49		Porretta Terme	PORRETT	4415	1098	349	FIC/KNMI
50	Portugal	Santarem	1320000	3924	-870	54	FIC/KNMI
51		Pegoes	1670000	3863	-865	64	FIC/KNMI
52		Alvega	2120000	3946	-804	51	FIC/KNMI
53		Mora	2260000	3893	-816	110	FIC/KNMI
54		Penhas Douradas	5680000	4041	-755	1380	FIC/KNMI
55		Portalegre	5710000	3928	-741	597	FIC/KNMI
84	Spain	Sabiñanigo	946000	4252	-35	790	FIC/KNMI

Table 7.3: Model variants selected as a part of sensitivity testing.

Method	Model variant	Notes	
OSR	ROSR: Regional Predictor Set	Regionally selected predictors (Table 6.2), applied to each country (France, Greece, Italy, Portugal, Spain) individually.	
ANN	LBFNN: Linear Neural Network	Neural network with no preconditioning	12 nodes, Basis width =10, Constant starting seed.
	GBFNN: Gaussian Basis Function	Gaussian preconditioning signal	
	CBFNN: Thin Plate Spline Basis Function	Cubic preconditioning signal	
	MQFNN: Gaussian Basis Function	Multi-quadratic preconditioning signal	
	IMFNN: Thin Plate Spline Basis Function	Inverse Multi-quadratic preconditioning signal	