

# Prediction Using Panel Data Regression with Spatial Random Effects

Bernard Fingleton (SERC, SIRE and Strathclyde University)

September 2008

This work was part of the research programme of the independent UK Spatial Economics Research Centre funded by the Economic and Social Research Council (ESRC), Department for Business, Enterprise and Regulatory Reform (BERR), the Department for Communities and Local Government (CLG), and the Welsh Assembly Government. The support of the funders is acknowledged. The views expressed are those of the authors and do not represent the views of the funders.

© Bernard Fingleton, submitted 2008

# **Prediction Using Panel Data Regression With Spatial Random Effects**

Bernard Fingleton\*

September 2008

\* SERC, SIRE and Strathclyde University

## Acknowledgements

This paper was presented at the 37th Annual Conference, Regional Science Association, International, British and Irish Section.

## Abstract

This paper considers some of the issues and difficulties relating to the use of spatial panel data regression in prediction, illustrated by the effects of mass immigration on wages and income levels in local authority areas of Great Britain. Motivated by contemporary urban economics theory, and using recent advances in spatial econometrics, the panel regression has wages dependent on employment density and the efficiency of the labour force. There are two types of spatial interaction, a spatial lag of wages, and an autoregressive process for error components. The estimates suggest that increased employment densities will increase wage levels, but wages may fall if migrants are under-qualified. This uncertainty highlights the fact that *ex ante* forecasting should be used with great caution as a basis for policy decisions.

Keywords: panel data, spatially correlated error components, economic geography, spatial econometrics

JEL Classifications: C33, O10, F02

## I. Introduction

Prediction is a difficult exercise, but *ex ante* prediction, in which the independent variables themselves have to be forecast, is even more so. *Ex post* prediction, with independent variables known with certainty, is a more feasible and a valuable adjunct to assessing the performance of a model. In this paper we carry out both *ex post* and *ex ante* prediction for panel data regressions using a random effects estimator with spatially autocorrelated disturbances, following, among others, Anselin (1988), Kapoor, Kelejian and Prucha (2007), and Baltagi and Li (2006). The aim of the paper is to highlight issues and assumptions associated with spatial panel prediction, using as an illustration the question of the impact of mass migration into the UK.

The principal projection by the UK Government's actuary department assumes net migration of 4.1M between 2004 and 2031. Assuming mainly economic migration, this raises the interesting question, what will be the effect on wages, and tax revenues, of this net increase in labour supply? Also how will the extra spending be distributed across the cities and regions of Great Britain? It is possible that the projected expansion in the supply of labour will hold down wage levels below what they otherwise would be. On the other hand increasing returns due to additional labour may create positive pecuniary and technological externalities.

The paper tests the increasing returns hypothesis, and also uses the resulting estimated model for *ex ante* and *ex post* prediction. The dependent variable is the average wage rate<sup>1</sup> across all occupations in each local area<sup>2</sup> and the explanatory variables are the density of employment in each area, together with the efficiency levels of employees.

The reduced form of the model also includes a spatial lag, equal to the weighted average of wage rates in 'neighbouring' areas. As argued below, a spatial lag occurs as an outcome of spatial interaction between areas due to commuting. The simplicity of the model means that there will undoubtedly be some unmodelled heterogeneity across areas. We attempt to control for this by means of a spatial autoregressive process for the disturbances, and hence the error components, of the panel data regression. Following

---

<sup>1</sup> In constant 1998 levels.

<sup>2</sup> The paper analyses 408 unitary authority and local authority districts (UALADs) in Great Britain.

Baltagi and Li (2006), the out-of sample prediction is based on Goldberger (1962) who derives best linear unbiased predictors.

## II. Theory

The theoretical model behind the reduced form is well known, and therefore only a sketch of the model is provided here. We follow Rivera-Batiz, (1988), Abdel-Rahman and Fujita (1990), Quigley (1998), Fujita, Krugman and Venables (1999) and Fujita and Thisse (2002) and assume that the city/region economy can be divided into two sectors with different market structures. There is a competitive sector comprising industries and services (hereafter ‘industry’) that operate under constant returns to scale and trade globally, and an immobile ‘producer services’ sector (hereafter ‘services’) providing inputs to the competitive sector which operates under monopolistic competition as described by Dixit-Stiglitz monopolistic competition theory. This seems reasonable for service firms, which are assumed to be typically numerous, small, independent and heterogeneous. In the services sector there are internal increasing returns to scale and a CES subproduction function which defines the level of composite services deployed in the Cobb-Douglas production function of the competitive sector, as defined in equation (1). The net effect could be returns that increase in city employment density, although these externalities may be offset by the effects of congestion (Ciccone and Hall, 1996). Therefore we have the possibility of increasing returns while at the same time each actor’s decision is explicitly stated as one of profit or utility maximization. As cities become denser, increasing diversity or variety in producer inputs can yield external scale economies, even though firms only break even (earn normal profits). There are increasing returns to the city economy as a whole in the context of competitive producers.

Monopolistic competition theory tells us why an increase in service labour increases service variety, rather than creating more of the same variety<sup>3</sup>. Assume that the market structure comprises numerous service firms each producing their own differentiated services with free exit and entry. Profits will be driven to zero (i.e. each

---

<sup>3</sup> Under monopolistic competition theory, equilibrium creates firms with size dependent on exogenous parameters, so an increase in service labour does not affect the size of service firms, just the number, and since each firm produces a different variety, that increase the number of varieties.

firm earns normal profits) which is a stable equilibrium point. Positive profits induce entry and drive profits towards zero, negative profits cause exits, raising profits towards the stable equilibrium. Fixed costs give internal increasing returns. This means that each firm produces its own differentiated service variety rather than varieties being replicated across firms, so as to maximise the internal returns to scale. Service differentiation creates monopoly power and prices are a mark up on marginal costs. The equilibrium levels of output and labour requirement depend on exogenous parameters and are constant across firms.

Monopoly power is determined by the exogenous parameter  $\mu$  which takes a value greater than 1. As  $\mu$  increases, so does monopoly power, and a concomitant of this is falling elasticity of substitution. Increasing monopoly power, ceteris paribus, increases the level of composite services, which is employed to give industry output. Reducing  $\mu$  towards 1 has a reverse effect.

More explicitly<sup>4</sup>, assume that the level of composite services  $I$  is an argument in the production ( $Q$ ) of competitive industry, in other words

$$Q = ((E^C \cdot A)^\beta I^{1-\beta})^\alpha L^{1-\alpha} \quad (1)$$

in which  $E^C \cdot A$  is the number of labour efficiency units in the competitive industry sector and  $L$  is land area used for production. The existence of  $\alpha$  creates the possibility (with  $\alpha < 1$ ) of diminishing returns due to congestion effects, since the factor of production land ( $L$ ) is set equal to 1 and the variables are measured per unit of land. Since  $I$  depends only on the labour efficiency units in the monopolistically competitive sector ( $E^M \cdot A$ ) and the total labour efficiency units is  $N^E = A(E^C + E^M)$ , it follows (see for example Fujita and Thisse (2002)) that

$$Q = ((E^C \cdot A)^\beta I^{1-\beta})^\alpha = \phi (N^E)^\gamma \quad (2)$$

In equation (2)  $\phi$  is a constant and following Ciccone and Hall (1996)  $\gamma = \alpha [1 + (1 - \beta)(\mu - 1)]$ , with the elasticity of substitution between services equal to

---

<sup>4</sup> We develop the model initially for a single cross-section here omitting for simplicity the subscript  $t$ , prior to estimating a panel version.

$\frac{\mu}{\mu - 1}$ . If  $\gamma > 1$  there are increasing returns to employment density. Also, using standard theory, the proportion ( $\alpha$ ) of final production  $Q$  spent on labour is

$$\frac{w^o N^E}{Q} = \alpha \quad (3)$$

in which  $w^o$  is the wage rate. Therefore

$$\ln w^o = \ln Q + \ln \alpha - \ln N^E \quad (4)$$

and on substituting for  $Q$  and for labour efficiency units  $N^E = E \cdot A$  in which  $E$  is the total employment level per square km and  $A$  is each area's level of efficiency, gives

$$\ln w^o = \ln \phi + \gamma \ln(A \cdot E) + \ln \alpha - \ln(A \cdot E) \quad (5)$$

It then follows that

$$\ln w^o = k_1 + (\gamma - 1) \ln E + (\gamma - 1) \ln A \quad (6)$$

in which  $k_1$  denotes a constant.

Consider next what determines the level of efficiency ( $A$ ) in each local area. We assume that worker efficiency is depended on the extent of formal qualification ( $S$ ) in the workforce, skills attributable mainly to schooling and college. We also assume that worker efficiency depends on knowledge and physical capital ( $K$ ) available in the workplace. Details are given in the data section below on the measurement of these variables.

Since workers commute to earn wages, the efficiency level of workers at any one place will also be partly determined by the efficiency levels in other places from which they might commute. Commuting frequency declines with distance increases, and this is reflected in the structure of the so-called  $N$  by  $N$  matrix  $W$ , where  $N$  is the number of areas. Details of the construction of  $W$  are given in the data section below.

Embodying these various factors, and writing in matrix notation, the vector of (log) efficiency levels is given by

$$\begin{aligned} \ln A &= \ln Bb + \rho W \ln A + \zeta \\ \zeta &\sim N(0, \tilde{\Omega}^2) \end{aligned} \quad (7)$$



where  $\ln B$  is a matrix of exogenous variables with  $N$  rows and  $k-1$  columns (equal to 1,  $\ln S$  and  $\ln K$ ),  $b$  is a  $k-1$  by 1 coefficients vector, and the matrix product  $W \ln A$  gives an  $N$  by 1 vector in which the value for area  $i$  is the sum of the log efficiency levels in other areas weighted by commuting distance from area  $i$ . Since this leads to a somewhat distinct endogenous variable, we signify this by the scalar coefficient  $\rho$ . Finally, the vector  $\zeta$  represents excluded variables which behave as random shocks.

Written with only exogenous variables on the right hand side gives

$$\ln A = (I - \rho W)^{-1} (\ln B b + \zeta)$$

This can then be substituted into equation (6) to give

$$\ln w^o = k_1 + (\gamma - 1) \ln E + (\gamma - 1) (I - \rho W)^{-1} (\ln B b + \zeta)$$

which is equal to

$$(I - \rho W) \ln w^o = (I - \rho W) k_1 + (\gamma - 1) (I - \rho W) \ln E + (\gamma - 1) (\ln B b + \zeta)$$

and therefore for time  $t$

$$\begin{aligned} \ln w_t^o &= \rho W \ln w_t^o + (I - \rho W) k_1 + (\gamma - 1) (\ln E_t - \rho W \ln E_t) + \ln B_t c + e_t \\ e_t &\sim N(0, (\gamma - 1)^2 \tilde{\Omega}^2) \end{aligned} \quad (8)$$

### III. Data

The dependent variable at time  $t$  is the natural log of mean pay<sup>5</sup> in each of ( $N =$ ) 408 areas, which have denoted by  $\ln w_t^o$ . The data cover the period 1999-2003 inclusive ( $T =$  5 years), so that  $\ln w^o$  is an  $NT \times 1$  vector. The number of employees is taken from the UK's annual business enquiry employee analysis, for both males and females and across all industries. This gives the employees per square km for the period 1999-2003, denoted by the  $NT \times 1$  vector  $\ln E$ .

---

<sup>5</sup> gross pay for full time workers (male and female), across all occupations. These data have been deflated to 1998 levels.

Educational attainment ( $S$ ), the percentage of residents with no qualifications, as given<sup>6</sup> by the UK's 2001 Census. These census data are constant over the period of the panel, with the same  $N$  values are replicated  $T = 5$  times in the  $NT \times 1$  vector  $S$ . The focus is a lack of any qualifications so as to avoid the complexities associated with the comparability of different educational attainment levels across cultures and time.

The level of workplace provided knowledge and physical capital ( $K$ ) is assumed to depend on the local intensity of knowledge/capital intensive activities, as indicated by the location quotient<sup>7</sup>. This measures each area's relative specialisation in computing and related activities (1992 SIC 72) and in research and development<sup>8</sup> (1992 SIC 73), as reflected in the relative concentration of employees in software and hardware consultancy and supply, data base activities, data processing, office machinery and computer maintenance and repair, other related computer-oriented activities, and engineering, natural, social sciences and humanities research.

The  $W$  matrix is an  $N \times N$  matrix of non-stochastic time constant weights which defines the spatial (commuting) interdependence of areas. For origin  $i$  and destination  $r$ ,  $W$  is defined as

$$\begin{aligned} W_{ir} &= \exp(-\hat{\delta}_i D_{ir}) & i \neq r \\ W_{ir} &= 0 & i = r \\ W_{ir} &= 0 & D_{ir} > 100km \end{aligned} \quad (9)$$

In which  $D_{ir}$  is the (straight line) distance between areas and  $\hat{\delta}_i$  provides an area-specific distance decay. Given observed travel to work patterns<sup>9</sup> for area  $i$ , the estimated value

$\hat{\delta}_i = \arg \min_{\delta_i} f(x)$  in which  $x = \sum_k (p_k - \tilde{p}_k)^2$  where  $p_k$  is the proportion travelling distances ( $d$ ) within commuting distance class  $k$ , so that  $k_l \leq d < k_u$ , and

$\tilde{p}_k = e^{-\delta_i D_{ik_u}} / \sum_k e^{-\delta_i D_{ik_u}}$ . In practice distance class bounds are  $k_l = 0, 2, 5, 10, 20, 30$ , and

<sup>6</sup> Available from the website Casweb, which is a web interface to statistics and related information from the United Kingdom Census of Population.

<sup>7</sup> The location quotient is the share of local employment in these sectors divided by national share.

<sup>8</sup> Calculated from data taken from the annual business enquiry employee, available from the UK's Office of National Statistics online via NOMIS.

<sup>9</sup> 1991 Census of Population - Special Workplace Statistics, available from NOMIS.

40km and  $k_u = 2, 5, 10, 20, 30, 40$  and 60km , as dictated by data availability. This is done for each of the 408 areas, giving  $\hat{\delta}_i$  ( $i = 1 \dots 408$ ), so that  $W$  embodies the differential propensity to commute by job location.

Spatial interdependence is also a feature of the error process assumed for the model, which is used to capture the effects of across-area heterogeneity and omitted spatially autocorrelated variables. The basis of the error process is an  $N$  by  $N$  contiguity matrix. To obtain this, for area  $j$ , contiguous areas, say  $k, l, m$  etc, are coded 1 and non-contiguous areas coded 0. Then each non-zero cell is divided by its respective row total to give the normalised contiguity matrix<sup>10</sup>  $W_c$  .

The short time period in this study and the presence of spatial dependence would seem to preclude testing for unit roots (O’Connell, 1998, Banerjee, 1999, Banerjee et. al., 2004), although recently several panel unit root tests have been proposed that are suitable for the analysis of cross-correlated panels, the properties of which are discussed by Gutierrez (2006). However these generally assume that spatial dependence can be adequately captured by an error component following a factor model with a common factor. This is somewhat different from the error process proposed below.

#### **IV. The panel model with spatially autocorrelated error components**

In the reduced form (8),  $W \ln w_i^o$  is endogenous, with instantaneous interaction across areas in wage levels. A shock to the wage a location  $i$  will affect areas  $k, l, m$  and so on, rebounding back to  $i$ , and vice versa. Assume also that  $\ln E$  is endogenous. While equation (8) tells us that high density will cause high wages, causation will no doubt be two-way, with  $\ln E$  responding to wage rate variations. However assume that in the short run wage levels do not affect the other variables, namely  $S$  and  $K$ . For the purposes of estimation, it is assumed that  $S$  and  $K$  are in effect exogenous variables. Only in the longer-run might they respond to variations in to  $\ln w^o$  .

---

<sup>10</sup> Note that remote island groups (Eilean Saar, Shetland, Orkney and Scilly) are assumed to be contiguous to the nearest mainland UALAD (Highland, Orkney, Highland and Penwith respectively ).

Writing (8) with  $S$  and  $K$  as explicit variables and taking just one cross-section at year  $t$ , the model is

$$\ln w_t^o = \rho W \ln w_t^o + (I - \rho W) \ln(k_1) + (\gamma - 1)(\ln E_t - \rho W \ln E_t) + c_0 + c_1 \ln S_t + c_2 \ln K_t + e_t \quad (10)$$

In which  $e_t$  is a stochastic disturbance. Unfortunately, we do not know  $k_1 = \ln(\alpha \phi)$  so the variable  $(I - \rho W)k_1$  is of necessity omitted from the estimating equation, which is therefore

$$\ln w_t^o = \rho W \ln w_t^o + (\gamma - 1)(\ln E_t - \rho W \ln E_t) + c_0 + c_1 \ln S_t + c_2 \ln K_t + e_t \quad (11)$$

This entails a constraint equalizing the coefficient  $\rho$  for  $W \ln w_t^o$  and  $W \ln E_t$ . In order to satisfy this constraint, an iterative procedure is used, as described subsequently.

Simplifying, we rewrite equation (11) as

$$Y_t = \rho W Y_t + H_t \beta + e_t \quad (12)$$

in which  $Y_t = \ln w_t^o$ ,  $W Y_t = W \ln w_t^o$ ,  $H_t$  is the  $N$  by  $k$  matrix of regressors  $(\ln E_t - \rho W \ln E_t, 1, \ln S_t, \ln K_t)$ ,  $\beta$  is a  $k \times 1$  vector of parameters  $(\gamma - 1, c_0, c_1, c_2)$ .

For the error or disturbance process, assume that in each period  $e_t = \lambda W_c e_t + \xi_t$ , in which  $\lambda$  is an unknown parameter, and  $\xi_t$  is an  $N$  by 1 vector of time  $t$  innovations.

Since we are considering a panel with  $T$  periods rather than purely cross sectional data, we omit  $t$ , hence

$$\begin{aligned} Y &= \rho (I_T \otimes W) Y + H \beta + e = X b + e \\ X &= ((I_T \otimes W) Y, H) \\ b' &= (\rho, \beta') \end{aligned} \quad (13)$$

in which  $Y$  is a  $TN \times 1$  vector of observations obtained by stacking  $Y_t = \ln w_t^o$  for  $t = 1 \dots 5$ ,  $X$  is a  $TN \times (1 + k)$  matrix of regressors, comprising the  $TN \times 1$  vector  $(I_T \otimes W) Y$ , and  $H$  which is a  $TN \times k$  matrix of regressors. Also  $b$  is the  $k+1 \times 1$  vector of parameters  $(\rho, \gamma - 1, c_0, c_1, c_2)$ . In addition, given that  $I_T$  is a  $T \times T$  diagonal matrix with 1s on the

main diagonal and zeros elsewhere, and  $I_N$  is a similar  $N \times N$  diagonal matrix, then  $I_{TN} = I_T \otimes I_N$  is a  $TN \times TN$  diagonal matrix with 1s on the main diagonal and zeros elsewhere. These create the  $TN \times 1$  vector  $e$

$$e = (I_{TN} - \lambda I_T \otimes W_c)^{-1} \xi \quad (14)$$

In which  $\xi$  is an  $TN \times 1$  vector of innovations.

Regarding the error components in space-time, time dependency is introduced into the innovations  $\xi$  via permanent error component  $\mu$ , thus

$$\begin{aligned} \mu &\sim iid(0, \sigma_\mu^2) \\ v &\sim iid(0, \sigma_v^2) \end{aligned} \quad (15)$$

$$\xi = (I_T \otimes I_N) \mu + v \quad (16)$$

so that  $\mu$  is an  $N \times 1$  vector of errors specific to each area,  $v$  is the transient error component comprising an  $TN \times 1$  vector of errors specific to each area and time,  $I_T$  is a  $T \times 1$  matrix with 1s, and  $I_T \otimes I_N$  is a  $TN \times N$  matrix equal to  $T$  stacked  $I_N$  matrices. The result is that the  $TN \times TN$  innovations variance-covariance matrix  $\Omega_\xi$  is nonspherical.

Also  $\sigma_1^2 = \sigma_v^2 + T\sigma_\mu^2$ .

Also each of the two error components  $\mu$  and  $v$  is subject to the same autoregressive process (c.f. Baltagi and Li, 2006), since

$$e = (I_{TN} - \lambda I_T \otimes W_c)^{-1} \xi = (I_{TN} - \lambda I_T \otimes W_c)^{-1} ((I_T \otimes I_N) \mu) + (I_{TN} - \lambda I_T \otimes W_c)^{-1} v \quad (17)$$

Estimating the reduced form involves iterations to ensure the equality of the  $\rho$ s evident in equation (11) for instance, within which is nested GMM, nonlinear least squares and spatial FGLS. In estimating  $\rho, \beta, \sigma_\mu^2, \sigma_v^2$  and  $\lambda$ , the method of Kapoor et al. (2007) is adapted to allow for endogenous variables in the matrix of regressors  $X$ . Regressor matrix  $X$  includes vector  $(I_T \otimes W)Y$  which is endogenous by definition, and the endogenous

variable  $\ln E_t - \rho W \ln E_t$ , stacked for  $t = 1 \dots 5$ . Hence we proceed using instrumental variables, but also take account of the non-sphericity of variance-covariance matrix  $\Omega_\xi$ .

## V. Results

TABLE 1  
Panel regression estimates

		Ungrouped instruments			Grouped instruments		
		Estimate	St. error	t ratio	Estimate	St. error	t ratio
constant	$c_0$	6.10921	0.0684865	89.20	6.11018	0.0688914	88.69
$W \ln w_t^o$	$\rho$	0.00107829	7.456e-005	14.46	0.0010334	9.411e-005	10.98
$\ln E_t - \rho W \ln E_t$	$\gamma - 1$	0.0140515	0.00276572	5.08	0.0190839	0.00600049	3.18
$\ln S_t$	$c_1$	-0.0913607	0.0210325	-4.34	-0.099407	0.0229029	-4.34
$\ln K_t$	$c_2$	0.0514069	0.00532627	9.65	0.049414	0.00558258	8.85
Error process							
	$\lambda$	0.305396			0.317614		
	$\sigma_v^2$	0.00218021			0.00216799		
	$\sigma_1^2$	0.0224078			0.0224084		
RSS		13.5854			13.7098		
R <sup>2*</sup>		0.845178			0.843562		
Instruments		$\ln S, \ln K,$ $\ln E_{1998}$			$\ln S, \ln K, G_{-}$ $\ln E_{1998}$		
		$W \ln E_{1998}$			$G_{-} W \ln E_{1998}$		
		$ZZZ_{-}$			$ZZZ_{-}$		
Sargan p-value		0.3836			0.5878		

RSS =sum of squared residuals

\* squared correlation between observed and fitted values.

Table 1 summarises the outcome of the estimation procedure using a panel of data for the period 1999-2003. Given that variables  $W \ln w_t^o$  and  $\ln E_t - \rho W \ln E_t$  are endogenous, variables uncorrelated with the errors but correlated with the endogenous variables are required as instruments. Two sets of instruments are employed, giving the

two sets of results in Table 1. On the left hand side we have the estimates using log employment density<sup>11</sup> for 1998 ( $\ln E_{1998}$ ),  $W \ln E_{1998}$  and  $ZZZ_{-}$ . The instrument  $ZZZ_{-}$  is, for area  $j$  ( $j = 1, \dots, N$ ), equal to the sum of straight line distances to all other areas ( $k, l, m, \dots$ ). The three right hand side columns of Table 1 are based on  $G_{-} \ln E_{1998}$ , which is a variable with values equal to -1, 0, 1 depending on whether or not  $\ln E_{1998}$  is in the upper, middle or lower third of values when placed in rank order. In addition, we use  $G_{-} W \ln E_{1998}$  which is equal to the matrix product of  $W$  and  $G_{-} \ln E_{1998}$ . Using groups rather than actual values is meant to provide additional insurance against endogeneity, although in either case the Sargan tests indicates that there is no significant correlation between the instruments and the residuals<sup>12</sup>. Each of these instruments is constant across time, and so the values are replicated to create  $NT$  by 1 vectors. The other instruments are the exogenous variables  $\ln S_t$  and  $\ln K_t$ . All further analysis is based on the estimates using ungrouped instruments.

The estimate of  $\gamma - 1$ , which is significantly greater than 0, suggest increasing returns to employment density as a result of pecuniary externalities raising wages in denser cities. Doubling employment density produces an increase in wages of  $\ln(2^{0.014}) \approx 0.01$  or about 1%. The implication is that increasing labour supply as a consequence of positive net migration will raise wage levels, but not by much. There are significant effects on wage levels due to labour efficiency variations as reflected by variations in educational attainment ( $\ln S_t$ ) and availability of knowledge and physical capital ( $\ln K_t$ ), the coefficients of which are significant and appropriately signed. If we halve the percentage without qualifications, wages rise by a factor of about  $\ln(2^{0.091}) \approx 0.06$  or about 6%, while doubling the percentage without qualifications causes wages to fall by about 6% since  $\ln(2^{-0.091}) \approx -0.06$ . Doubling  $K$  causes wages to rise by approximately 3.6%. The implication of this for positive net migration are unclear, since, focusing on educational attainment, it is unclear whether future migrants will be more

---

<sup>11</sup> The assumption being that, since they predate the data period, they will not be contemporaneously correlated with the errors.

<sup>12</sup> Using the (iterated) 2SLS residuals.

qualified than the current population, or less. Certainly the past record of employment suggests that many migrants have been poorly qualified. As indicated by Weale and Riley (2006), according to Labour Force Survey ‘27.6 per cent of post-1997 immigrants work in elementary occupations as compared to 18.9 per cent of the population as a whole’, and increase by a factor of about 1.5.

With regard to the two spatial effects in the model, the evidence that  $\rho > 0$  indicates a significant autoregressive spatial lag involving  $\ln w_t^o$ . The positive sign on the estimated  $\lambda$  reflects positive spatial disturbance correlation and hence positive dependence among the permanent and transient error components, reflecting spatially autocorrelated heterogeneity and omitted variables. The relatively high value of  $R^{2*}$  suggests that the model fits the data quite well.

## VI. *Ex post* Prediction

The model is tested against out-of-sample data for the year 2004. In this case the known 2004 values of the regressors, denoted by  $\ln S_{T+1}$ ,  $\ln K_{T+1}$  and  $\ln E_{T+1}$  are combined with the parameter estimates  $\hat{b}$  from Table 1, but in addition there is a correction allowing for the error process.

Goldberger (1962) shows that the best linear unbiased predictor (BLUP)<sup>13</sup> for location  $i$  in period  $T + s$  is given by

$$\ln \hat{w}_{i,T+s} = x'_{i,T+s} \hat{b} + \omega'_i \Omega^{-1} \hat{e} \quad (18)$$

In which  $\ln \hat{w}_{i,T+s}$  is the scalar predicted value for location  $i$  in period  $T + s$ ,  $x_{i,T+s}$  is a  $k$  by 1 vector of regressor values at  $i$  at  $T + s$ , and the BLUP estimator of  $b$  is the  $k$  by 1 vector  $\hat{b}$ . The scalar term  $\omega'_i \Omega^{-1} \hat{e}$  is an estimate of the prediction disturbance for location  $i$  at  $T + s$ , in which  $\omega_i$  is an  $NT$  by 1 vector of covariances of the prediction disturbance at location  $i$  with the estimate of the  $NT$  by 1 vector of residuals  $e$ , and  $\Omega$  is the  $NT$  by  $NT$

<sup>13</sup> An early purely spatial application of this is Dubin, Kelley Pace and Thibodeau (1999).



error variance-covariance matrix. Given the innovations covariance matrix  $\Omega_\xi$  (see Appendix), the error covariance matrix  $\Omega$  is

$$\Omega = \Omega_\xi [(I_{TN} - \lambda I_T \otimes W_c)'(I_{TN} - \lambda I_T \otimes W_c)]^{-1} \quad (19)$$

Which is estimated using  $\hat{\lambda}$ ,  $\hat{\sigma}_v^2$  and  $\hat{\sigma}_1^2$ . This is similar to Anselin (1988) and Baltagi and Li (2006), although in their cases the spatial error process applies only to the transient error component, hence, rather than equation (17), they have

$$e = ((I_T \otimes I_N)\mu) + (I_{TN} - \lambda I_T \otimes W_c)^{-1}v \quad (20)$$

The term  $\omega_i' \Omega^{-1} \hat{e}$  is equal to a weighted average of the residuals  $\hat{e}$  for the  $N$  locations averaged over the  $T$  time periods,

$$\bar{e}_j = \sum_{t=1}^T \hat{e}_{tj} / T \quad (21)$$

with the weight depending on  $\overline{W_c}$  and  $\hat{\lambda}$ . Hence

$$\omega_i' \Omega^{-1} \hat{e} = T \frac{\sigma_\mu^2}{\sigma_v^2} \sum_{j=1}^N \delta_j \bar{e}_j \quad (22)$$

In which  $\delta_j$  is the  $j$ 'th element in the  $i$ 'th row of estimated  $V^{-1}$  with

$$V = T \frac{\sigma_\mu^2}{\sigma_v^2} I_N + [(I_N - \lambda W_c)'(I_N - \lambda W_c)]^{-1} \quad (23)$$

With the error process applying to both error components, it follows that

$$V = T \frac{\sigma_\mu^2}{\sigma_v^2} I_N [(I_N - \lambda W_c)'(I_N - \lambda W_c)]^{-1} + [(I_N - \lambda W_c)'(I_N - \lambda W_c)]^{-1} \quad (24)$$

To obtain the predictions, we calculate

$$\ln \hat{w}_{T+1} = (I - \hat{\rho} W)^{-1} [\hat{c}_0 1 + (\hat{\gamma} - 1)(\ln E_{T+1} - \hat{\rho} W \ln E_{T+1}) + \hat{c}_1 \ln S_{T+1} + \hat{c}_2 \ln K_{T+1} + \omega' \Omega^{-1} \hat{e}] \quad (25)$$

in which  $(I - \hat{\rho} W)^{-1}$  is an  $N$  by  $N$  matrix, and the square brackets give an  $N$  by 1 vector, and  $\omega' \Omega^{-1} \hat{e}$  is the  $N$  by 1 vector of corrections. Equation (25) is equivalent to (A21) in the Appendix, which is shown by Monte-Carlo simulation to perform better than several competing prediction equations.

It is apparent that the out-of-sample test produces predictions that are quite close to the data for 2004, as is evident from Figures 1,2 and 3. The lower panel of Figure 1 shows the relationship between the 2004 wage rate and the model predictions  $\ln \hat{w}_{i,T+1}$  without applying the correction (in other words in equation (25)  $\omega' \Omega^{-1} \hat{e}$  is assumed to be a vector of zeros). The upper panel is the relationship between the 2004 wage rate and  $\ln \hat{w}_{i,T+1}$  calculated according to equation (25) including the Goldberger-type correction. Also the improvement due to the correction is shown by the respective values of the simple prediction error given by

$$\sum_{i=1}^N (\ln \hat{w}_{i,T+1} - \ln w_{2004}^o)^2 \quad (26)$$

which is equal to 2.3897 without the correction, and equal to 1.3275 with the correction. Figure 2 shows the wage data for 2004, and Figure 3 gives the *ex post* prediction using equation (25).

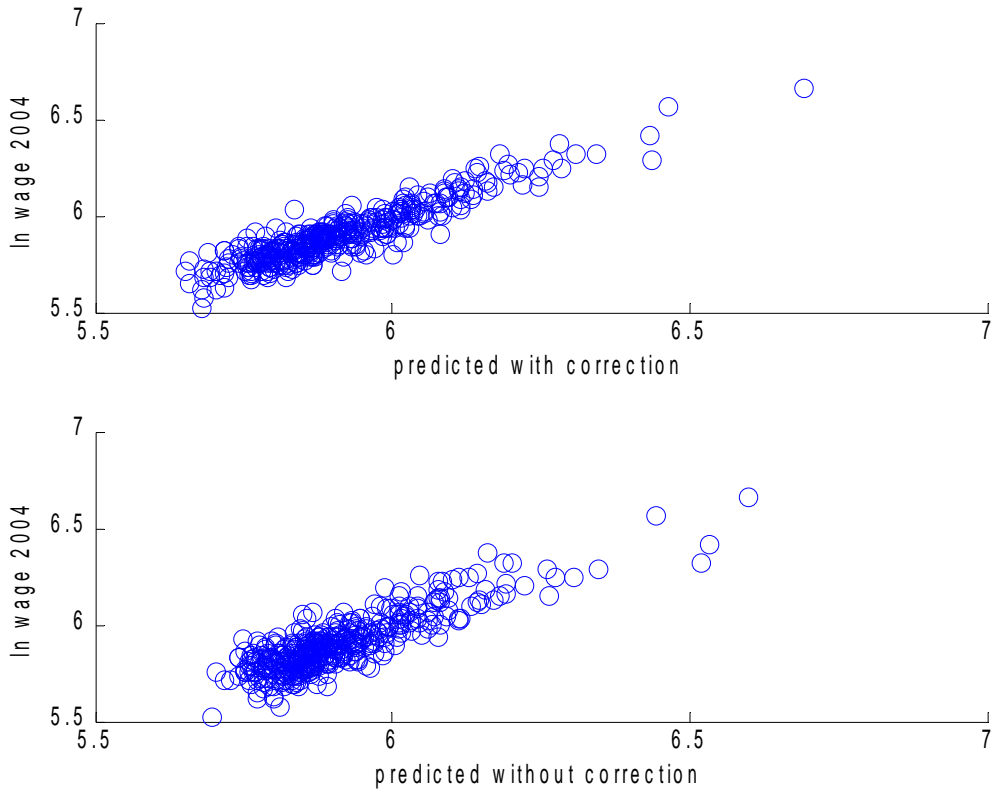


Figure 1. Fit of model to out-of-sample data with and without Goldberger-type correction

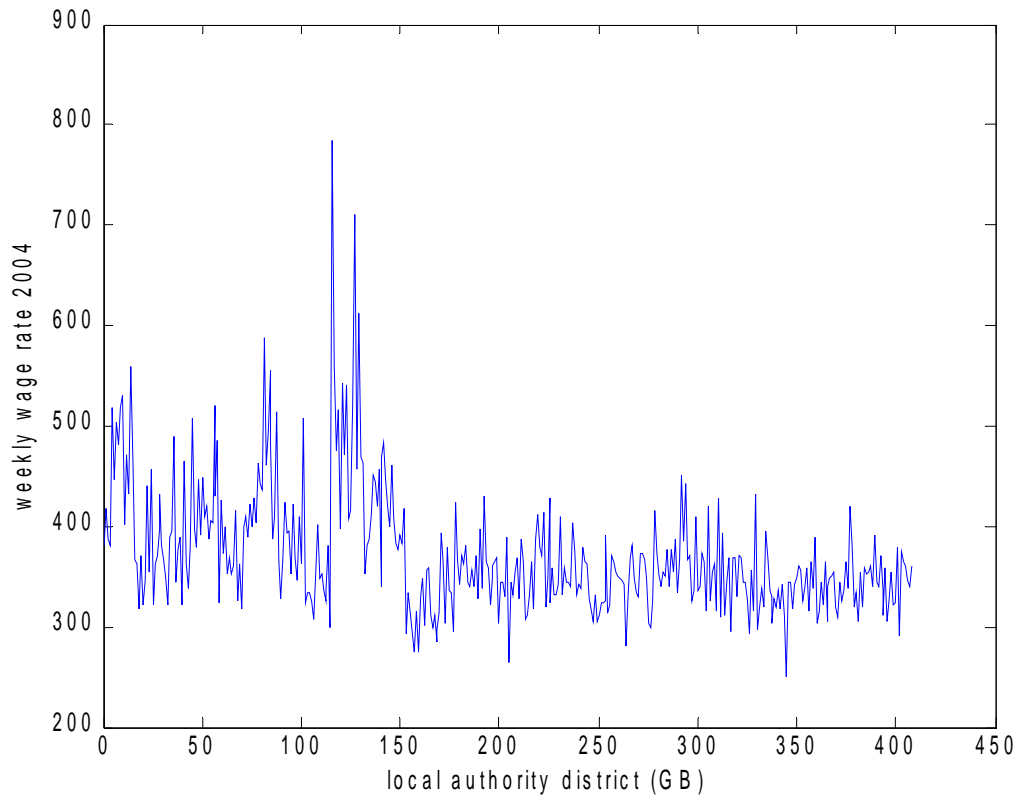


Figure 2. Weekly wage by local authority district

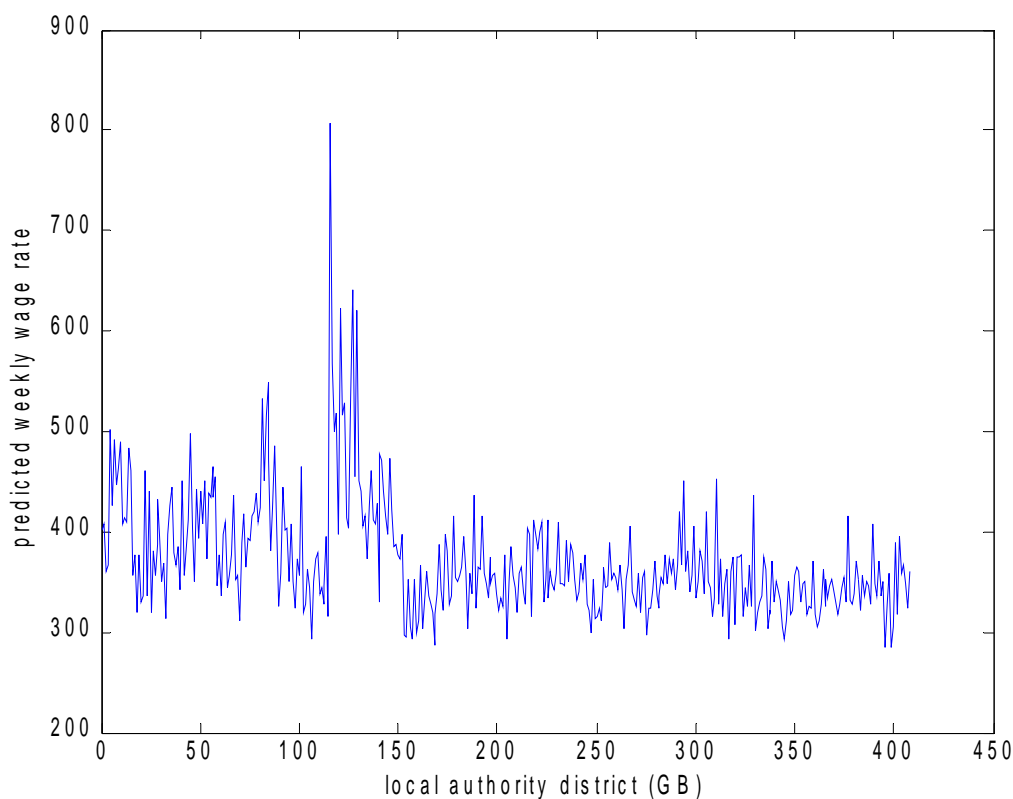


Figure 3. Ex post predicted wage by local authority district.

## VII. *Ex Ante* Prediction

In this section *ex ante* forecasts are produced using the assumption made by the UK Government actuary service that by 2031 there will be 4.1M additional migrants in the UK<sup>14</sup>. It is assumed that these are all economic migrants and that they are all employed. Predicting wages and aggregate income by area as a result of the additional employees calls for a number of additional assumptions and extreme caution, notably because *ex ante* prediction calls for predicted values of the independent regressors. Because of the additional uncertainty and assumptions, *ex ante* predictions are dangerous weapons, and are at best is only a quasi-scientific basis for implementing policy.

<sup>14</sup> We assume 4M for Great Britain.

$$\ln \hat{w}_{T+k} = (I - \hat{\rho}W)^{-1}[\hat{c}_0 1 + (\hat{\gamma} - 1)(\ln E_{T+k} - \hat{\rho}W \ln E_{T+k}) + \hat{c}_1 \ln S_{T+k} + \hat{c}_2 \ln K_{T+k} + \omega' \Omega^{-1} \hat{e}] \quad (27)$$

Equation (27), which is equation (25) with  $T+k$  (2031) in place of  $T+I$ , is the basis of the *ex ante* predictions. To implement this, it is necessary to predict employment density ( $E$ ), educational attainment ( $S$ ) and knowledge and physical capital ( $K$ ).

Additionally, we assume parameter homogeneity, so that the estimated parameters  $\hat{\rho}, \hat{\gamma}, \hat{c}_0, \hat{c}_1, \hat{c}_2$  are applicable to  $T+k$ . It is also assumed that the pattern of residual dependence will remain the same, so that  $\omega' \Omega^{-1} \hat{e}$  applies, and the same commuting patterns is assumed for the future, so that  $W$  remains appropriate. These assumptions really do not have a strong factual basis, so it is important to test the sensitivity of the predictions to alternative assumptions.

Let us make some first pass, naïve, assumptions and then explore the sensitivity of the outcomes to alternative assumptions about  $S$ . In order to allocate migrants to areas, ideally we should have sub-models predicting where this new employment will occur, hence obtain  $E_{T+k}$ , and predict where the migrants will live, and be able to show the implications of migrant settlement for the levels of educational attainment ( $S_{T+k}$ ) in each area. In the absence of such models, we simply allocate migrants' employment proportional to existing employment in each area, and allocate migrants to places of residence in proportion to existing population. Following Weale and Riley (2006), assume first that these migrants are 50% less qualified than the existing population. Averaging across all areas, the mean proportion of residents who are unqualified is 0.28, while averaging across all areas gives 0.42 for the additional migrants. Under this scenario,  $S_{i,T+k} = 100(D_{2003,i} + 1.5D_{2003,i}M_{i,T+k} / P_{2003,i}) / (P_{2003,i} + M_{i,T+k})$ , in which  $D$  is the total number of non-migrant residents without qualifications,  $P$  is the total number of non-migrant residents and  $M$  is the number of migrants, so the unqualified migrant share is 50% greater than the resident unqualified share in each area, which increases the percentage of the overall population without qualifications. The outcome is Figure 4. Secondly, assume that migrants have the same lack-of-qualification rate (28%) as the existing residents, so that there is no effect on  $S_{T+k}$ . Thirdly assume that the additional

workforce is completely qualified, so there are no additional unqualified residents, but a larger overall resident population following the arrival of the immigrants. This reduces the percentage of the resident population without qualifications. The percentage without qualifications ( $S_{T+k}$ ) in area  $i$  is then estimated as  $S_{i,T+k} = 100(D_{2003,i}) / (P_{2003,i} + M_{i,T+k})$ .

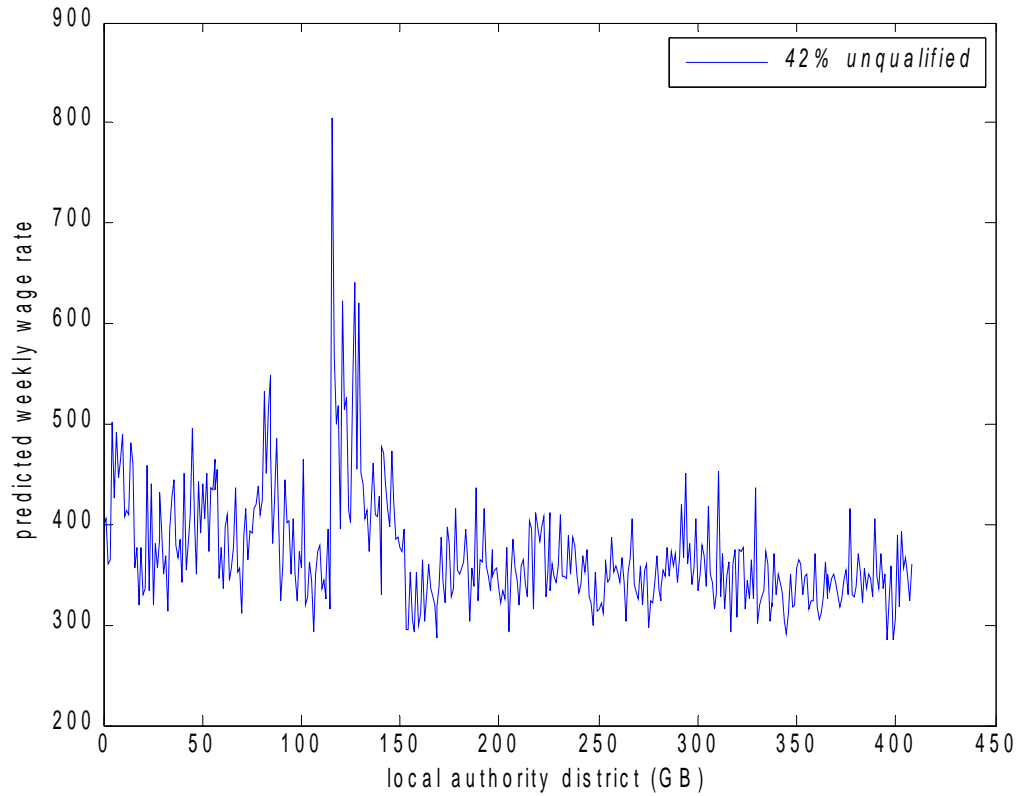


Figure 4. Ex ante predicted wage by local authority district

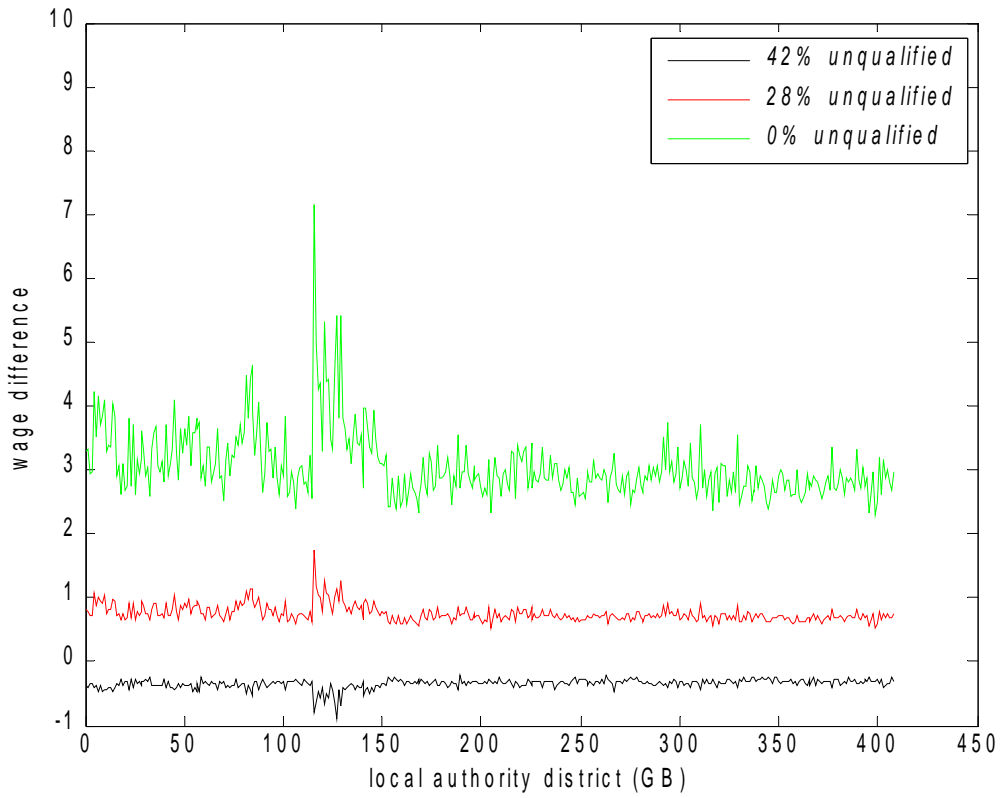


Figure 5. Ex ante predicted weekly wage differences by local authority district

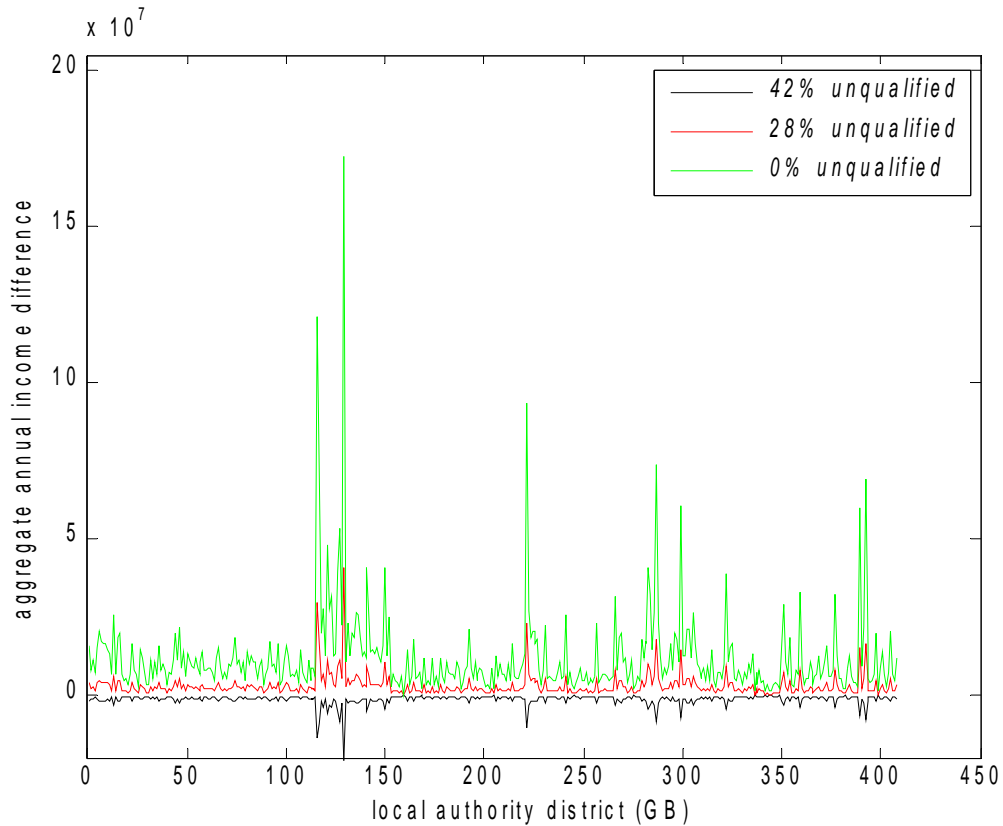


Figure 6. Ex ante predicted annual aggregate income differences by local authority district

Figure 5 shows the wage rate differences comparing the wages under the three assumptions about migrant qualification rates with the predicted wage rates assuming no additional migrants. The lower line shows that when 42% of migrants are unqualified, wages are marginally reduced compared to the no migrants scenario. This outcome is a combination of two effects, first the additional 4M migrants will undoubtedly increase employee densities, but we know from the Table 1 estimates that these would have to double for wage rates to increase by 1%. Any marginal rise in wage rates due to higher densities is more than offset by the reduction due to a fall in the percentage of the overall population in each area that is qualified. If 28% of migrants are unqualified, like the initial resident population, then the outcome is a wage increase of about 1£ per week. Assuming that 0% of the additional workforce is unqualified increases wages by about £3. These differences may seem very small, but when translated into the aggregate annual increase in income, significant numbers occur.



Figure 6 shows the effect on aggregate income, comparing aggregate annual income under the three assumptions about migrant qualifications. In each case the aggregate annual<sup>15</sup> income per UALAD ( $i$ ) is

$$I_{i,B} = 50 \exp(\ln \hat{w}_{i,T+k}) E_{i,T+k} L_i \quad (28)$$

where  $E_{i,T+k} L_i$  is the aggregate level of employment<sup>16</sup> at time  $T+k$  and  $\hat{w}_{i,T+k}$  is the appropriate weekly wage rate.

Figure 6 shows prominent peaks for the city of London, Westminster, and for the other major employment concentrations. For Westminster, the presence of fully qualified migrant workers increases annual aggregate income by about £170m. Outside the main central cities, the typical aggregate income is about £10m more than it would be without the presence of qualified migrants. Summing across all areas gives the effect for the whole of Great Britain. If 0% of migrants are unqualified, the gain is about £4.9bn. If 28% of migrants are unqualified, there is an annual aggregate increase of about £1.2bn. If 42% of migrants are unqualified, there is a loss of aggregate income compared to the level one would obtain with no migrants of about £0.5bn.

## VIII. Conclusions

This paper applies panel regression models with spatially correlated error components, to show that they have become viable tools for the spatial economist and spatial econometrician. In this application, the estimation procedure set out by Kapoor, Kelejian and Prucha (2007) is adapted to allow for the presence of an endogenous lag, and also allowing for a parameter restriction, both of which are necessary because of the theory underpinning the reduced form. The theoretical basis is a model in which increasing returns to employment density are an outcome of assumptions about the existence of internal increasing returns to service firms with a monopolistic competition market structure, modified by variation between areas in worker efficiency. The paper shows that increasing employment density increases wage rates, and wages increase with higher

---

<sup>15</sup> 50 weeks

<sup>16</sup> Density times land area.

educational attainment and the availability of more knowledge and physical capital. The paper examines the outcomes and assumptions of *ex post* and *ex ante* prediction, based on the BLUP correction<sup>17</sup> given by Goldberger (1962). The model error process assumed in the paper is that the entire disturbance structure subject to the same spatial autoregressive process, and this thus entails a modification of the correction used by Baltagi and Li (2006) which considers spatial dependence in the transient error component only (c.f. Anselin, 1988, Baltagi and Li, 2006). The *ex ante* predictions use as an example the impact of an increase in employment attributable to positive net migration. The problem of predicting the independent variable, which unlike *ex post* prediction are not known with certainty, is illustrated by showing the implications of different assumptions about the percentage of migrant workers possessing some form of qualification. Although the implications of individual wage rates appear negligible, when treated as annual income differences by area, quite significant magnitudes ensue. The conclusion is that while it is important and useful to do out-of-sample *ex post* model testing, *ex ante* prediction is more dangerous and difficult, and should be accompanied by a very big health warning! However, as Pesaran (1990) observes, ‘Econometric models are important tools for forecasting and policy analysis, and it is unlikely that they will be discarded in the future. The challenge is to recognise their limitations and to work towards turning them into more reliable and effective tools. There seem to be no viable alternatives’.

## APPENDIX

### Innovations covariance matrix $\Omega_{\xi}$

The vector  $\xi$  is equal to the sum of the errors  $\mu$ , which differ for each area but are constant for the same area at different times, and the errors  $\nu$ , which differ for each area and for each time, with no covariance across area or time.

For areas  $i, j$  and times  $t, s$

$$E(\xi' \xi) = \begin{bmatrix} \sigma_{\nu}^2 + \sigma_{\mu}^2 \\ \sigma_{\mu}^2 \end{bmatrix} \text{ if } i = j; t = s$$

$$E[\xi' \xi] = \begin{bmatrix} \sigma_{\mu}^2 \end{bmatrix} \text{ if } i = j; t \neq s \quad (\text{A1})$$

<sup>17</sup> It is not claimed that the correction used in the present paper is itself BLUP.

$$E[\xi^t \xi^s] = [0] \text{ if } i \neq j; t \neq s$$

We represent the  $TN \times TN$  innovation variance-covariance matrix  $\Omega_\xi$  using the matrices  $Q_0$  and  $Q_1$  defined as follows

$$Q_0 = (I_T - \frac{J_T}{T}) \otimes I_N \quad (\text{A2})$$

in which  $J_T$  is a  $T \times T$  matrix of 1s, and

$$Q_1 = \frac{J_T}{T} \otimes I_N \quad (\text{A3})$$

It follows that  $Q_0 + Q_1 = I_{TN}$  and

$$\begin{aligned} \Omega_\xi &= \sigma_\mu^2 (J_T \otimes I_N) + \sigma_v^2 I_{TN} \\ \Omega_\xi &= \sigma_v^2 Q_0 + \sigma_1^2 Q_1 \end{aligned} \quad (\text{A4})$$

in which

$$\sigma_1^2 = \sigma_v^2 + T\sigma_\mu^2 \quad (\text{A5})$$

## Estimation

Estimation proceeds in three stages. Stages 1 and 3 are almost identical and both give estimates of  $b = (\rho, \gamma - 1, c_0, c_1, c_2)$ , however they differ by the way values are obtained for  $\sigma_v^2$ ,  $\sigma_1^2$  and  $\lambda$ . In both stages 1 and 3 we first filter out the error dependence, using a Cochrane-Orcutt (C-O) transformation, premultiplying by  $I_{TN} - \lambda I_T \otimes W_c$  since  $e = (I_{TN} - \lambda I_T \otimes W_c)^{-1} \xi$ , therefore

$$\begin{aligned} Y^* &= (I_T \otimes (I_N - \hat{\lambda} W_c))Y \\ X^* &= (I_T \otimes (I_N - \hat{\lambda} W_c))X \\ \xi &= (I_T \otimes (I_N - \hat{\lambda} W_c))e \end{aligned} \quad (\text{A6})$$

A linearly independent subset of the exogenous variables is used to give the  $TN \times f \geq (k+1)$  matrix of instruments  $Z$ , and we assume matrices  $X$  and  $Z$  are full column

rank with  $f \geq (k+1)$ . Since the error covariance matrix  $\Omega_\xi = \sigma_v^2 Q_0 + \sigma_1^2 Q_1$  means that there are nonspherical disturbances, following Bowden and Turkington (1984), calculate  $P_z = Z(Z'\hat{\Omega}_\xi Z)^{-1}Z'$ , which is a symmetric matrix ( $P_z \Omega_\xi$  is idempotent) and hence

$$\hat{b} = \left[ (X^{*'}Z)(Z'\hat{\Omega}_\xi Z)^{-1}(Z'X^*) \right]^{-1} (X^{*'}Z)(Z'\hat{\Omega}_\xi Z)^{-1}(Z'Y^*) = (X^{*'}P_z X^*)^{-1} X^{*'}P_z Y^* \quad (A7)$$

The estimated variance-covariance matrix of the parameters is given by

$$\hat{C} = \left[ (X^{*'}Z)(Z'\hat{\Omega}_\xi Z)^{-1}(Z'X^*) \right]^{-1} = (X^{*'}P_z X^*)^{-1} \quad (A8)$$

Greene (2003) also gives the equivalent of (A7) and (A8) as generalized methods of moments (instrumental variables) estimators with nonspherical disturbances.

Also

$$\hat{e} = Y - X\hat{b}$$

Stage 1 uses arbitrary values of 1, 1 and 0 respectively for  $\sigma_v^2$ ,  $\sigma_1^2$  and  $\lambda$ , but estimates are available from the data for Stage 3.

Additionally, within stages 1 and 3 there are iterations in order to satisfy the constraint involving  $\rho$ . The iterations proceed with an arbitrary value for  $\rho$  ( $\tilde{\rho}_1 = 1$ ) in the first iteration to allow calculation of  $\ln E_t - \tilde{\rho}_1 W \ln E_t$ , which then leads to an initial value  $\hat{\rho}_1$  given by (say)  $\hat{b}_1$  ( $\rho$  is an element of vector  $b$ ). The second iteration uses  $\ln E_t - \hat{\rho}_1 W \ln E_t$  to obtain  $\hat{\rho}_2$ , and the third iteration uses  $\ln E_t - \hat{\rho}_2 W \ln E_t$  to obtain  $\hat{\rho}_3$ , and so on, with iterations terminating when for iteration  $r$ ,  $|\hat{\rho}_r - \hat{\rho}_{r-1}| < 0.000005$ , at which point  $\hat{\rho} = \hat{\rho}_r$  and  $\hat{b} = \hat{b}_r$ .

The estimated variance-covariance matrix of the  $b$  parameters is given by

$$\hat{C} = (X^{*'}P_z X^*)^{-1} \quad (A9)$$

and the standard errors of the  $\hat{b}$  are given by the squares roots of the values on the main diagonal of  $\hat{C}$ , which allows 't-ratios' to be calculated for purposes of inference.

For the GMM estimation at Stage 2, following Kapoor et al (2007),

$$\Gamma \phi' - \eta = 0 \quad (\text{A10})$$

and

$$\tilde{\Gamma} \tilde{\phi}' - \tilde{\eta} = 0 \quad (\text{A11})$$

where  $\Gamma$  and  $\tilde{\Gamma}$  are 3 by 4 matrices,  $\eta$  and  $\tilde{\eta}$  are 3 by 1 vectors and  $\phi =$

$[\lambda \quad \lambda^2 \quad \sigma_v^2 \quad \sigma_1^2]$  is a vector of parameters. Using the estimated disturbances  $\hat{e}$ , one

obtains sample counterparts  $g$  and  $\tilde{g}$  of vectors  $\eta$  and  $\tilde{\eta}$ , and sample counterparts

$G$  and  $\tilde{G}$  of matrices  $\Gamma$  and  $\tilde{\Gamma}$ , defined as follows

$$G = \begin{bmatrix} \frac{2}{N(T-1)} \tilde{e}' Q_0 \hat{e}_{-1} & \frac{-1}{N(T-1)} \tilde{e}'_{-1} Q_0 \hat{e}_{-1} & 1 & 0 \\ \frac{2}{N(T-1)} \tilde{e}'_{-2} Q_0 \hat{e}_{-1} & \frac{-1}{N(T-1)} \tilde{e}'_{-2} Q_0 \hat{e}_{-2} & \frac{1}{N} t_1 & 0 \\ \frac{1}{N(T-1)} (\tilde{e}' Q_0 \hat{e}_{-2} + \tilde{e}'_{-1} Q_0 \hat{e}_{-1}) & \frac{-1}{N(T-1)} \tilde{e}'_{-1} Q_0 \hat{e}_{-2} & 0 & 0 \end{bmatrix} \quad (\text{A12})$$

$$g = \begin{bmatrix} \frac{1}{N(T-1)} \tilde{e}' Q_0 \hat{e} \\ \frac{1}{N(T-1)} \tilde{e}'_{-1} Q_0 \hat{e}_{-1} \\ \frac{1}{N(T-1)} \tilde{e}' Q_0 \hat{e}_{-1} \end{bmatrix} \quad (\text{A13})$$

$$G[\lambda \quad \lambda^2 \quad \sigma_v^2 \quad \sigma_1^2]' - g = \zeta(\lambda \quad \sigma_v^2 \quad \sigma_1^2) \quad (\text{A14})$$

$$\tilde{G} = \begin{bmatrix} \frac{2}{N} \tilde{e}' Q_1 \hat{e}_{-1} & \frac{-1}{N} \tilde{e}'_{-1} Q_1 \hat{e}_{-1} & 0 & 1 \\ \frac{2}{N} \tilde{e}'_{-2} Q_1 \hat{e}_{-1} & \frac{-1}{N} \tilde{e}'_{-2} Q_1 \hat{e}_{-2} & 0 & \frac{1}{N} t_1 \\ \frac{1}{N} (\tilde{e}' Q_1 \hat{e}_{-2} + \tilde{e}'_{-1} Q_1 \hat{e}_{-1}) & \frac{-1}{N} \tilde{e}'_{-1} Q_1 \hat{e}_{-2} & 0 & 0 \end{bmatrix} \quad (\text{A15})$$

$$\tilde{g} = \begin{bmatrix} \frac{1}{N} \tilde{e}' Q_1 \hat{e} \\ \frac{1}{N} \tilde{e}'_{-1} Q_1 \hat{e}_{-1} \\ \frac{1}{N} \tilde{e}' Q_1 \hat{e}_{-1} \end{bmatrix} \quad (\text{A16})$$

$$\tilde{G}[\lambda \quad \lambda^2 \quad \sigma_v^2 \quad \sigma_1^2]' - \tilde{g} = \zeta(\lambda \quad \sigma_v^2 \quad \sigma_1^2) \quad (\text{A17})$$

In which  $t_1 = tr(W_c' W_c)$ ,  $\hat{e}_{-1} = (I_T \otimes W_c) \hat{e}$  and  $\hat{e}_{-2} = (I_T \otimes W_c) \hat{e}_{-1}$ , and  $\zeta(\lambda \quad \sigma_v^2 \quad \sigma_1^2)$ ,  $\tilde{\zeta}(\lambda \quad \sigma_v^2 \quad \sigma_1^2)$  are vectors of residuals, and the nonlinear least squares estimators are given by

$$(\hat{\lambda}, \hat{\sigma}_v^2, \hat{\sigma}_1^2) = \arg \min \{ \zeta(\lambda \quad \sigma_v^2 \quad \sigma_1^2)' \zeta(\lambda \quad \sigma_v^2 \quad \sigma_1^2) + \tilde{\zeta}(\lambda \quad \sigma_v^2 \quad \sigma_1^2)' \tilde{\zeta}(\lambda \quad \sigma_v^2 \quad \sigma_1^2) \} \quad (\text{A18})$$

This can be done in various ways. One is to use unconstrained non-linear least squares estimation using a modified Newton-Raphson method which is suitable for minimising any non-linear function and this depends on numerical differences, so there is no need to specify derivatives.

In general the variances associated with the two separate right hand side terms of (A18) differ, and Kapoor et al (2007) suggest weighting to allow for this. However for simplicity we have not introduced differential weighting. Kapoor et al (2007) note that giving equal weight to all six moments equations does give consistent estimates.

### **The comparative performance of alternative predictors**

In this section we compare the predictive performance of some alternative prediction equations, including equation (A21) which is a simplified analogue of equation (25) and which is shown to perform better than competitors across several indicators. The comparisons are based on a Monte-Carlo set-up with the following data generating process

$$\begin{aligned}
Y &= \lambda I_T \otimes W^E Y + H\gamma + e \\
X &= (I_T \otimes W^E Y, H) \\
b' &= (\lambda, \gamma') \\
Y &= Xb + e \\
e &= (I_{TN} - \rho (I_T \otimes W))^{-1} \xi
\end{aligned} \tag{A19}$$

In (1A),  $W$  is a standardized contiguity matrix<sup>18</sup> on a  $\sqrt{N} \times \sqrt{N}$  square. Contiguity matrix  $W^*$  is standardized by dividing each row cell by its row total. Also in practice, for simplicity we assume that  $W^E = W$ .

The matrix  $H$  had columns equal to the  $N(T+1) \times 1$  vectors  $1_{TN}$ ,  $H_1$ , and  $H_2$ , in which  $1_{N(T+1)}$  is a  $N(T+1) \times 1$  vector with 1s. To obtain each  $H$ , first generate time  $t = 0$   $N \times 1$  vectors  $H_1(0), H_2(0)$  by sampling at random from an  $N(0,1)$  distribution. Then for  $t$  equal to  $1 \dots T+1$ , where  $1 \dots T$  is the estimation period,  $H_1(t) = H_1(t-1) + \pi_1$ , in which  $\pi_1 \sim N(0,1)$ , and likewise for  $H_2(t)$  using  $\pi_2 \sim N(0,1)$ . Then stacking these  $N \times 1$  vectors we obtain  $H_1$  and  $H_2$ . In his way the exogenous variables in  $H$  have some time dependency as seems reasonable for panel data. Once generated, the variables  $H_1$  and  $H_2$  remain fixed.

Given the exogenous variables, we next obtain the innovations vector  $N(T+1) \times 1$  vector  $\xi$ . The innovations vector depends on the  $N \times 1$  vector  $\mu$  obtained by sampling from an  $N(0, \sigma_\mu^2 I)$  distribution and on the  $N(T+1) \times 1$  vector  $v$  obtained by sampling from an  $N(0, \sigma_v^2 I)$  distribution, so that  $\xi = (1_{T+1} \otimes I_N)\mu + v$ . This is then used to obtain

$$Y = (I_{N(T+1)} - \lambda (I_{T+1} \otimes W))^{-1} H\gamma + (I_{N(T+1)} - \lambda (I_{T+1} \otimes W))^{-1} (I_{N(T+1)} - \rho (I_{T+1} \otimes W))\xi \tag{A20}$$

We use part of the data to estimate model, and use the information from this estimation to predict the vector  $Y_{T+1}$ . Given  $Y$ ,  $W$  and  $H$ , estimates are obtained of the known parameters  $\rho, \lambda, \gamma_0, \gamma_1, \gamma_2, \sigma_v^2$  and  $\sigma_\mu^2$  using the three stage method outlined above.

---

<sup>18</sup>  $W^*$  is a Rook's case contiguity matrix, comprising 1s and 0s with 0s on the main diagonal.

Since there is an endogenous lag  $(I_T \otimes W)Y$ , estimation uses instruments  $Z$  comprising the exogenous variables,  $H$ , and their spatial lags.

It is evident that there are a number of different forecasting equations with varying bias and precision (see for instance Kelejian and Prucha, 2007 and Le Sage and Pace, 2004 for different spatial prediction equations involving a spatial lag). Using the values of the exogenous variables at time  $T + 1$ , and the estimated  $\rho, \lambda, \gamma_0, \gamma_1, \gamma_2, \sigma_v^2, \sigma_\mu^2$  and hence  $\omega' \Omega^{-1} \hat{e}$  we obtain predicted  $Y$  at  $T + 1$  ( $\tilde{Y}$ ) using several different forecasting

equations and measure the forecast error  $\sum_{i=1}^N (\tilde{Y} - Y_{T+1})^2$ ; this leads us to equation (A21)

and by analogy to (25) as the preferred forecasting equation. The alternative forecasting equations that are considered are

$$\tilde{Y} = (I - \hat{\lambda} W^E)^{-1} (X \hat{b} + \omega' \Omega^{-1} \hat{e}) \quad (\text{A21})$$

$$\tilde{Y} = (I - \hat{\lambda} W^E)^{-1} X \hat{b} + \omega' \Omega^{-1} \hat{e} \quad (\text{A22})$$

$$\tilde{Y} = (I - \hat{\lambda} W^E)^{-1} X \hat{b} \quad (\text{A23})$$

$$\tilde{Y} = X \hat{b} + \omega' \Omega^{-1} \hat{e} \quad (\text{A24})$$

$$\tilde{Y} = X \hat{b} \quad (\text{A25})$$

For simplicity we eliminate subscript  $T + 1$  from (A21) to (A25). This was replicated 1000 times and we summarise the resulting forecast error distribution via the median forecast error, the standard deviation, and the interquartile range (IQR)<sup>19</sup>. We retain  $\sigma_\mu^2 = 1$  and  $\sigma_v^2 = 1$  throughout. The outcomes are presented as ratios of the summarizing statistics, with in each case equation (A25) providing the denominator.

Table 1A The comparative performance of prediction equations A21 to A25 (1)

Simulation parameters	ratio	Median	Standard deviation	IQR
$N = 81 \ T = 2$	A21/A25	0.6696	0.7643	0.7317
$\lambda = 0.5$	A22/A25	0.7006	0.7828	0.7397
$\rho = 0.75$	A23/A25	0.7547	0.8016	0.7639

<sup>19</sup> The median and IQR are chosen for robustness to extreme values.



$\gamma_0 = 1$	A24/A25	0.9460	0.9779	0.9971
$\gamma_1 = 2$	A25/A25	1.0000	1.0000	1.0000
$\gamma_2 = 3$				

Table 2A The comparative performance of prediction equations A21 to A25 (2)

Simulation parameters	ratio	Median	Standard deviation	IQR
$N = 81$ $T = 2$	A21/A25	0.1735	0.8896	0.1954
$\lambda = 0.75$	A22/A25	0.1941	1.0214	0.2195
$\rho = 0.25$	A23/A25	0.2276	1.0340	0.2340
$\gamma_0 = 1$	A24/A25	0.9654	0.9914	1.0001
$\gamma_1 = 1$	A25/A25	1.0000	1.0000	1.0000
$\gamma_2 = 1$				

Table 3A The comparative performance of prediction equations A21 to A25 (3)

Simulation parameters	ratio	Median	Standard deviation	IQR
$N = 49$ $T = 2$	A21/A25	0.7210	0.8209	0.8053
$\lambda = 0.25$	A22/A25	0.7462	0.8439	0.8109
$\rho = 0.5$	A23/A25	0.9181	0.9327	0.8974
$\gamma_0 = 1$	A24/A25	0.8424	0.9034	0.9048
$\gamma_1 = 0.5$	A25/A25	1.0000	1.0000	1.0000
$\gamma_2 = 2$				

Table 4A The comparative performance of prediction equations A21 to A25 (4)

Simulation parameters	ratio	Median	Standard deviation	IQR
$N = 121$ $T = 2$	A21/A25	0.6811	1.7686	0.7922
$\lambda = 0.5$	A22/A25	0.6907	1.7705	0.8099
$\rho = 0.9$	A23/A25	0.7088	1.7719	0.8114

$\gamma_0 = 1$	A24/A25	0.9810	0.9951	1.0021
$\gamma_1 = 5$	A25/A25	1.0000	1.0000	1.0000
$\gamma_2 = 1$				

Table 5A The comparative performance of prediction equations A21 to A25 (5)

Simulation parameters	ratio	Median	Standard deviation	IQR
$N = 81, T = 4$	A21/A25	0.0326	0.1442	0.1235
$\lambda = 0.8$	A22/A25	0.0386	0.1639	0.1434
$\rho = 0.3$	A23/A25	0.0450	0.1731	0.1532
$\gamma_0 = 1$	A24/A25	0.9940	0.9571	0.9656
$\gamma_1 = 5$	A25/A25	1.0000	1.0000	1.0000
$\gamma_2 = 5$				

## References

- Abdel-Rahman, H. and Fujita, M. (1990). 'Product variety, Marshallian externalities and city size', *Journal of Regional Science*, Vol. 30, pp. 165-83.
- Baltagi, B. H. and Li, D. (2006). 'Prediction in the panel data model with spatial correlation: the case of liquor', *Spatial Economic Analysis*, Vol. 1, pp. 175-185.
- Banerjee, A. (1999). 'Panel data unit roots and cointegration : an overview', *Oxford Bulletin of Economics and Statistics*, Special Issue, Vol. 61, pp. 607-629.
- Banerjee, A., Marcellino, M. and Osbat C. (2004). 'Testing for PPP: should we use panel methods?', *Empirical Economics*, Vol. 30, pp. 77-91.
- Bowden, R. J. and Turkington, D. A. (1984). *Instrumental Variables*, Cambridge University Press, Cambridge .
- Ciccone, A. and Hall, R. E. (1996). 'Productivity and the density of economic activity', *American Economic Review*, Vol. 86, pp. 54-70.
- Dubin, R., Kelley Pace, R. and Thibodeau, T. G. (1999). 'Spatial autoregression techniques for real estate data', *Journal of Real Estate Literature*, Vol. 7, pp. 79-95.

- Fujita, M., Krugman, P. R. and Venables, A. (1999). *The Spatial Economy : Cities, Regions, and International Trade*, MIT press, Cambridge Massachusetts.
- Fujita, M. and Thisse, J-F. (2002). *Economics of Agglomeration*, Cambridge University Press, Cambridge.
- Goldberger, A. S. (1962). 'Best linear unbiased prediction in the generalized linear regression model', *Journal of the American Statistical Association*, Vol. 57, pp. 369-375.
- Gutierrez, L. (2006). 'Panel unit-root tests for cross-sectionally correlated panels: a Monte Carlo comparison', *Oxford Bulletin of Economics and Statistics*, Vol. 68, pp. 519-540.
- Greene, W. H. (2003). *Econometric Analysis*, 5<sup>th</sup> Edition, Prentice Hall ,New Jersey.
- Kelejian, H. H. and Prucha, I. (2007) 'The relative efficiencies of various predictors in spatial econometric models containing spatial lags', *Regional Science and Urban Economics*, Vol. 37, pp. 363-374.
- Kapoor, M, Kelejian, H. H. and Prucha, I. (2007) 'Panel data models with spatially correlated error components', *Journal of Econometrics*, Vol. 140, pp. 97-130.
- LeSage, J. P. and Pace, R. Kelley (2004) 'Models for spatially dependent missing data', *Journal of Real Estate Finance and Economics*, Vol. 29, pp. 233-254.
- O'Connell, P. (1998). 'The overvaluation of purchasing power parity', *Journal of International Economics*, Vol. 44, pp. 1-19.
- Quigley, J. M. (1998). 'Urban diversity and economic growth', *Journal of Economic Perspectives* , Vol. 12, pp. 127-38.
- Rivera-Batiz, F. (1988). 'Increasing returns, monopolistic competition, and agglomeration economies in consumption and production', *Regional Science and Urban Economics* , Vol. 18, pp. 125-53.
- Riley, R. and Weale, M. (2006). 'Commentary – Immigration and its effects', *National Institute Economic Review*, Vol. 198, pp. 4.

**BERR**

Department for Business  
Enterprise & Regulatory Reform



Llywodraeth Cynulliad Cymru  
Welsh Assembly Government

**Spatial Economics Research Centre (SERC)**

London School of Economics  
Houghton Street  
London WC2A 2AE

**Tel:** 020 7852 3565

**Fax:** 020 7955 6848

**Web:** [www.spatial-economics.ac.uk](http://www.spatial-economics.ac.uk)

SERC is an independent research centre funded by the Economic and Social Research Council (ESRC), Department for Business, Enterprise and Regulatory Reform (BERR), the Department for Communities and Local Government (CLG) and the Welsh Assembly Government.