11th Economics Summer Seminars

Pamukkale University

Denizli

Turkey

Applied Spatial Econometrics

Bernard Fingleton
University of Cambridge
UK

Spatial panels

- Panel data models allow the researcher to control for heterogeneity across individuals (people, firms, countries etc)
  - differences in ability, race or productivity
- Spatial models allow for modelling externalities and spill-over effects
  - spend more money on police in one neighbourhood, you may increase the crime in an adjacent neighbourhood
- Spatial panel models control for *both* heterogeneity and spatial interaction

Spatial panels

- With the increasing availability of micro as well as macro level panel data, spatial panel data models are becoming increasingly attractive in empirical economic research
- spatial panel data applications
  - Does policing expenditures reduce crime across counties? see Kelejian and Robinson (1992)
    - Extra policing in one county may increase crime in neighbouring county
  - Is productivity across US states increased by public capital investment in roads and highways? see Holtz-Eakin (1994)
    - Road improvement in one state may benefit producers in nearby states

Two forms of error components

- The recent literature on spatial panel data models with error components adopts two error processes
- only the remainder/transient error term is spatially correlated but the individual effects are not (Anselin 1988)
  - Individual/permanent effects can be random (RE) or fixed (FE)
- both the individual/permanent and remainder/transient error components follow the same random spatial error process (Kapoor, Kelejian &Prucha, 2007)
  - Individual/permanent effects random (RE)

Two forms of error process

- Spatially autoregressive error process (Anselin, KKP)
  - SAR
- Moving average error process (Fingleton, Baltagi & Pirotte, Pirotte)
  - SMA

Six models

- Models with no spatial autocorrelation in individual component, only in remainder
  - RE-SAR
  - RE-SMA
  - FE-SAR
  - FE-SMA
- Models with spatial autocorrelation in both the individual and remainder component
  - SAR-RE
  - SMA-RE

Revision

Kronecker product

$$\begin{bmatrix}1 & 2\\3 & 4\end{bmatrix} \otimes \begin{bmatrix}0 & 5\\6 & 7\end{bmatrix} = \begin{bmatrix}1\cdot 0 & 1\cdot 5 & 2\cdot 0 & 2\cdot 5\\1\cdot 6 & 1\cdot 7 & 2\cdot 6 & 2\cdot 7\\3\cdot 0 & 3\cdot 5 & 4\cdot 0 & 4\cdot 5\\3\cdot 6 & 3\cdot 7 & 4\cdot 6 & 4\cdot 7\end{bmatrix} = \begin{bmatrix}0 & 5 & 0 & 10\\6 & 7 & 12 & 14\\0 & 15 & 0 & 20\\18 & 21 & 24 & 28\end{bmatrix}$$

$$e_T = \begin{bmatrix}1\\1\\1\\1\end{bmatrix} I_N = \begin{bmatrix}1 & 0 & 0\\0 & 1 & 0\\0 & 0 & 1\end{bmatrix} \qquad \alpha = \begin{bmatrix}1\\2\\3\end{bmatrix}$$

$$e_T \otimes I_N = \begin{bmatrix}1 & 0 & 0\\0 & 1 & 0\\0 & 0 & 1\\1 & 0 & 0\\0 & 1 & 0\\0 & 0 & 1\\1 & 0 & 0\\0 & 1 & 0\\0 & 0 & 1\\1 & 0 & 0\\0 & 1 & 0\\0 & 0 & 1\end{bmatrix} \qquad (e_T \otimes I_N)\alpha = \begin{bmatrix}1\\2\\3\\1\\2\\3\\1\\2\\3\\1\\2\\3\end{bmatrix}$$

Background to the equations

λ is the scalar spatial autoregressive coefficient with $|\lambda| < 1$. W is a known $N \times N$ spatial weight matrix whose diagonal elements are zero

Often W is binary, with $w_{ij} = 1$ when i and j are neighbours and $w_{ij} = 0$ when they are not.

Alternatively, W could be based on physical distances such as port to port or capital to capital, commuting distances, some measure of economic distance or social distance, or distance on a network connecting points.

The weights are commonly standardized so that the elements of each row sum to 1.

W also satisfies the condition that $(I - \lambda W)$ is nonsingular

Anselin (1988) provides more details on the properties of W

Important difference from ordinary panel data

- The data are ordered such that $i = 1, \ldots, N$ is the fast index and $t = 1, \ldots, T$ is the slow one.

RE-SAR

$$y_{it} = x_{it}\beta + \varepsilon_{it} \qquad i = 1,\dots,N; t = 1,\dots,T$$

$$\varepsilon_t = \alpha + u_t$$

Permanent and transient error components. P. component introduces time dependency in the data, transient component differs across individuals (space) and times

$$\varepsilon_t = \left(\varepsilon_{1t},\dots,\varepsilon_{Nt}\right)'$$

$$\alpha = \left(\alpha_1,\dots,\alpha_N\right)' \qquad \alpha \sim iid(0,\sigma_\alpha^2)$$

Permanent component is random process

$$u_t = \left(u_{1t},\dots,u_{Nt}\right)'$$

$$u_t = \lambda W u_t + v_t = (I - \lambda W)^{-1} v_t = B_N^{-1} v_t \qquad v_t \sim iid(0,\sigma_v^2)$$

I,*W* and *B* of dimension *N*

$$\varepsilon = (e_T \otimes I_N)\alpha + (I_T \otimes B_N^{-1})v \qquad e_T = (1,\dots,1)'$$

SAR applies only to transient disturbances

RE-SMA

$$u_t = \lambda W v_t + v_t = (I + \lambda W) v_t = D_N v_t$$

$$\varepsilon = (e_T \otimes I_N)\alpha + (I_T \otimes D_N)v$$

I,*W* and D of dimension *N*

SMA applies only to transient disturbances

FE-SAR,FE-SMA

$$y = X\beta + (e_T \otimes I_N)\alpha + (I_T \otimes B_N^{-1})v$$

$$y = X\beta + (e_T \otimes I_N)\alpha + (I_T \otimes D_N)v$$

Permanent, time constant Individual effects deterministic not random

SAR-RE

$$y_{it} = x_{it}\beta + \varepsilon_{it} \qquad i = 1,\ldots,N; t = 1,\ldots,T$$

$$\varepsilon_t = \lambda W \varepsilon_t + u_t$$

$$u_t = \alpha + v_t \qquad \alpha \sim iid(0,\sigma_\alpha^2) \qquad v_t \sim iid(0,\sigma_v^2)$$

$$\varepsilon = (e_t \otimes B_N^{-1})\alpha + (I_T \otimes B_N^{-1})v$$

SAME spatial interaction applies to unobservable Individual heterogeneity (permanent component)
AND to transient component

SMA-RE

$$\varepsilon = (e_t \otimes D_N)\alpha + (I_T \otimes D_N)v$$

estimation

- Maximum likelihood
  - RE-SAR and RE-SMA typically estimated by ML(Anselin, 1988)
  - FE-SAR, FE-SMA see also Anselin(1988)
  - Many routines available on LeSage website
  - www.spatial-econometrics.com
  - And on Elhorst website
  - http://www.regroningen.nl/elhorst/software.shtml
  - Many other sources such as
  - http://certur.free.fr/vgb/MatlabCodes.html
- GMM
  - For SAR-RE suggested by Kapoor, Keleijian and Prucha(2007)….KKP
  - For SAR-SME suggested by Fingleton(2008)

More ML

- Lee and Yu (2010) considered estimation of a spatial panel data model with individual-specific fixed effects, and proposed a ‘‘transformation approach’’ to eliminate the fixed effects and then apply quasi-ML to the transformed model
- Yu et al. (2008) and Yu and Lee (2010) focused on the properties of the quasi-ML estimator in the case of dynamic, possibly nonstationary, panels with fixed effects and spatial error correlation, assuming both *N* and *T* large

Focus on GMM

- Why GMM?
- ML estimation, even in its simplest form entails substantial computational problems when the number of cross-sectional units N is large
- ML techniques requires the serial correlation processes of the error terms, if any, to be fully specified. In panels where N is relatively large this could be quite demanding
- GMM allows multiple endogeneity
- GMM method is less demanding but still requires moment conditions that correctly take account of specific spatial and serial correlation patterns of the errors
  - Kelejian and Prucha (1999) suggested a generalized moments (GM)estimation method which is computationally feasible even when N is large
  - Kapoor, Kelejian, and Prucha (2007) generalized this GM procedure from cross-section to panel data and derived its large sample properties when T is fixed and $N \to \infty$
  - Fingleton(2008) extends this to MA errors

Fingleton B (2008) 'A Generalized Method of Moments estimator for a spatial panel model with an endogenous spatial lag and spatial moving average errors' *Spatial Economic Analysis*, 3 27-44

Focus on GMM

- GMM is concerned with the estimation of only some of the overall model parameters, namely those relating to the errors
- The starting point is therefore the residuals from an initial model (step 1)
- These provide (consistent) estimates, via GMM, (stage 2) of

$$\lambda, \sigma_v^2, \sigma_1^2 = T\sigma_\alpha^2 + \sigma_v^2$$

- Given these, we find estimates of the other model parameters in step 3
- The process for step 2 is very like that for GMM estimation of cross-sectional data, but is more elaborate for panels because of the different moments equations

Step 2- GMM : preliminary tools

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

$$Q_1 = \frac{J_T}{T} \otimes I_N$$

I_T is a TxT diagonal matrix with 1s on the main diagonal
and zeros elsewhere

I_N is a similar NxN matrix

J_T is a TxT matrix of 1s

Given TNx1 vector θ,

$Q_0\theta$ is a TNx1 vector of deviations from the time mean,
where the mean is mean of θ averaging over time

$Q_1\theta$ is a TNx1 vector comprising N means, averaging across time, and stacked for each T

θ	time	place	timemean	$Q_0\theta$	$Q_1\theta$
1.0000	1.0000	1.0000	2.5000	-1.5000	2.5000
2.0000	1.0000	2.0000	3.5000	-1.5000	3.5000
3.0000	1.0000	3.0000	4.5000	-1.5000	4.5000
4.0000	2.0000	1.0000	2.5000	1.5000	2.5000
5.0000	2.0000	2.0000	3.5000	1.5000	3.5000
6.0000	2.0000	3.0000	4.5000	1.5000	4.5000

GMM for SAR-RE

$$y_{it} = x_{it}\beta + \varepsilon_{it} \qquad i = 1,\dots,N; t = 1,\dots,T$$

$$\varepsilon_t = \lambda W \varepsilon_t + u_t$$

$$u_t = \alpha + v_t \qquad \alpha \sim iid(0,\sigma_\alpha^2) \qquad v_t \sim iid(0,\sigma_v^2)$$

$$\varepsilon = (e_t \otimes B_N^{-1})\alpha + (I_T \otimes B_N^{-1})v$$

$$\varepsilon_t = \lambda W \varepsilon_t + u_t$$

$$\begin{cases} u_t = \varepsilon_t - \lambda W \varepsilon_t = \varepsilon_t - \lambda \bar{\varepsilon}_t \\ \bar{u}_t = W(\varepsilon_t - \lambda W \varepsilon_t) = \bar{\varepsilon}_t - \lambda \bar{\bar{\varepsilon}}_t \end{cases}$$

Moments equations

$$E\left[u'Q_0u / N(T-1)\right] = \sigma_v^2$$

$$E\left[\bar{u}'Q_0\bar{u} / N(T-1)\right] = \sigma_v^2 tr(W'W) / N$$

$$E\left[\bar{u}'Q_0u / N(T-1)\right] = 0$$

$$E\left[u'Q_1u / N)\right] = T\sigma_\alpha^2 + \sigma_v^2 = \sigma_1^2$$

$$E\left[\bar{u}'Q_1\bar{u} / N)\right] = \sigma_1^2 tr(W'W) / N$$

$$E\left[\bar{u}'Q_1u / N)\right] = 0$$

Substituting these expressions we obtain a system of six equations involving the second moments of ε etc. This system involves λ etc and can be expressed as

$$G\begin{bmatrix} \lambda & \lambda^2 & \sigma_v^2 & \sigma_1^2 \end{bmatrix}' - g = 0$$

$$t_1 = tr(W'W) \quad \hat{\bar{\varepsilon}} = (I_T \otimes W)\hat{\varepsilon} \quad \hat{\bar{\bar{\varepsilon}}} = (I_T \otimes W)\hat{\bar{\varepsilon}}$$

Source : KKP 2007
Fingleton 2008

$$G\begin{bmatrix}\lambda & \lambda^2 & \sigma_\nu^2 & \sigma_1^2\end{bmatrix}' - g = \zeta(\lambda \quad \sigma_\nu^2 \quad \sigma_1^2)$$

$$G = \begin{bmatrix} \frac{2}{N(T-1)}\hat{\varepsilon}'Q_0\hat{\bar{\varepsilon}} & \frac{-1}{N(T-1)}\hat{\bar{\varepsilon}}'Q_0\hat{\bar{\varepsilon}} & 1 & 0 \\ \frac{2}{N(T-1)}\hat{\bar{\bar{\varepsilon}}}'Q_0\hat{\bar{\varepsilon}} & \frac{-1}{N(T-1)}\hat{\bar{\bar{\varepsilon}}}'Q_0\hat{e}_{-2}\hat{\bar{\bar{\varepsilon}}} & \frac{1}{N}t_1 & 0 \\ \frac{1}{N(T-1)}(\hat{\varepsilon}'Q_0\hat{\bar{\bar{\varepsilon}}} + \hat{\bar{\varepsilon}}'Q_0\hat{\bar{\varepsilon}}) & \frac{-1}{N(T-1)}\hat{\bar{\varepsilon}}'Q_0\hat{\bar{\bar{\varepsilon}}} & 0 & 0 \end{bmatrix}\begin{bmatrix}\lambda \\ \lambda^2 \\ \sigma_\nu^2 \\ \sigma_1^2\end{bmatrix} \qquad g = \begin{bmatrix}\frac{1}{N(T-1)}\hat{\varepsilon}'Q_0\hat{\varepsilon} \\ \frac{1}{N(T-1)}\hat{\bar{\varepsilon}}'Q_0\hat{\bar{\varepsilon}} \\ \frac{1}{N(T-1)}\hat{\varepsilon}'Q_0\hat{\bar{\varepsilon}}\end{bmatrix}$$

The equations underlying our GM procedures are the sample counterparts to the six equations based on estimated disturbances

$$\tilde{G}\begin{bmatrix}\lambda & \lambda^2 & \sigma_\nu^2 & \sigma_1^2\end{bmatrix}' - \tilde{g} = \tilde{\zeta}(\lambda \quad \sigma_\nu^2 \quad \sigma_1^2)$$

$$\tilde{G} = \begin{bmatrix} \frac{2}{N}\hat{\varepsilon}'Q_1\hat{\bar{\varepsilon}} & \frac{-1}{N}\hat{\bar{\varepsilon}}'Q_1\hat{\bar{\varepsilon}} & 0 & 1 \\ \frac{2}{N}\hat{\bar{\bar{\varepsilon}}}'Q_1\hat{\bar{\varepsilon}} & \frac{-1}{N}\hat{\bar{\bar{\varepsilon}}}'Q_1\hat{\bar{\bar{\varepsilon}}} & 0 & \frac{1}{N}t_1 \\ \frac{1}{N}(\hat{\varepsilon}'Q_1\hat{\bar{\bar{\varepsilon}}} + \hat{\bar{\varepsilon}}'Q_1\hat{\bar{\varepsilon}}) & \frac{-1}{N}\hat{\bar{\varepsilon}}'Q_1\hat{\bar{\bar{\varepsilon}}} & 0 & 0 \end{bmatrix}\begin{bmatrix}\lambda \\ \lambda^2 \\ \sigma_\nu^2 \\ \sigma_1^2\end{bmatrix} \qquad \tilde{g} = \begin{bmatrix}\frac{1}{N}\hat{\varepsilon}'Q_1\hat{\varepsilon} \\ \frac{1}{N}\hat{\bar{\varepsilon}}'Q_1\hat{\bar{\varepsilon}} \\ \frac{1}{N}\hat{\varepsilon}'Q_1\hat{\bar{\varepsilon}}\end{bmatrix}$$

$$(\hat{\lambda}, \hat{\sigma}_\nu^2, \hat{\sigma}_1^2) = \arg\min\{\zeta(\lambda \quad \sigma_\nu^2 \quad \sigma_1^2)'\zeta(\lambda \quad \sigma_\nu^2 \quad \sigma_1^2) + \tilde{\zeta}(\lambda \quad \sigma_\nu^2 \quad \sigma_1^2)'\tilde{\zeta}(\lambda \quad \sigma_\nu^2 \quad \sigma_1^2)\}$$

In general the variances associated with the two separate right hand side terms 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.

3 steps

Step 1 : obtain consistent estimates of ε

$$y_{it} = x_{it}\beta + \varepsilon_{it}$$

$$\hat{\varepsilon}_{it} = y_{it} - x_{it}\hat{\beta}$$

OLS if regressors exogenous, but 2SLS if endogenous

Step 2: estimate parameters of Autoregressive error process

$$\hat{\varepsilon} \rightarrow \hat{\lambda}, \hat{\sigma}_{\nu}^{2}, \hat{\sigma}_{1}^{2}$$

Step 3 : eliminate error dependence then estimate regression coefficients

Estimation

- 3 steps
  - (1) 2sls to obtain 2sls residuals
  - (2) Use OLS/2sls residuals to estimate via GMM
  - (3) robust IV to estimate regression parameters and t ratios
    - First carry out Cochrane-Orcutt transform

$$y^* = (I_T \otimes (I_N - \hat{\lambda} W))y$$

$$X^* = (I_T \otimes (I_N - \hat{\lambda} W))X$$

$$\xi = (I_T \otimes (I_N - \hat{\lambda} W))\hat{\varepsilon}$$

Feasible efficient GMM estimation

$$y^* = X^*\beta + \xi$$

$$\Omega_\xi = \sigma_\nu^2 Q_0 + \sigma_1^2 Q_1$$ error covariance matrix

$$P_H = Q(Q'\hat{\Omega}_\xi Q)^{-1}Q'$$ Q is matrix of instruments

$$\hat{\beta} = \left[(X^{*\prime}Q)(Q'\hat{\Omega}_\xi Q)^{-1}(Q'X^*)\right]^{-1}(X^{*\prime}Q)(Q'\hat{\Omega}_\xi Q)^{-1}(Q'y^*)$$

eqn(9.40)Davidson and Mackinnon

$$\hat{\beta} = (X^{*\prime}P_H X^*)^{-1}X^{*\prime}P_H y^*$$ eqn(A7) Fingleton (2008)

$$\hat{C} = (X^{*\prime}P_H X^*)^{-1}$$

$$s.e.\hat{\beta}^* = \sqrt{diag(\hat{C})}$$

$$t = \frac{\hat{\beta}^*}{s.e.\hat{\beta}^*}$$

This does not produce $s.e.\hat{\lambda}$

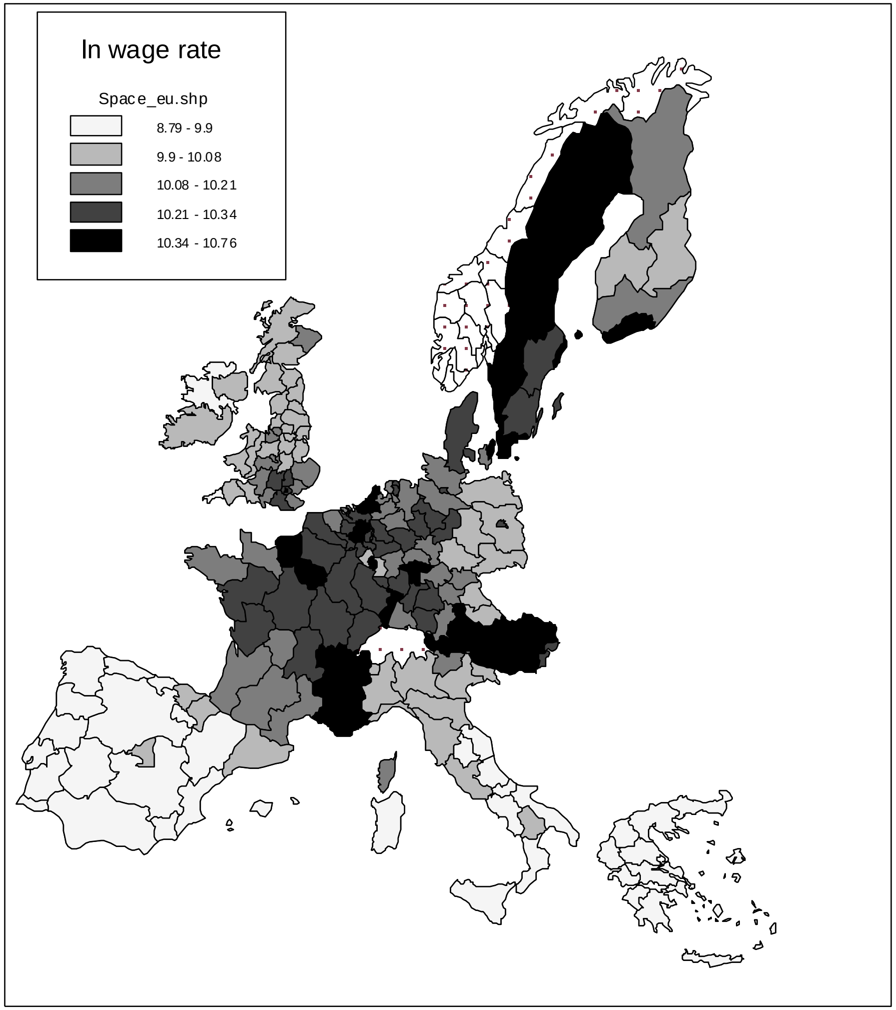
Bootstrap estimation

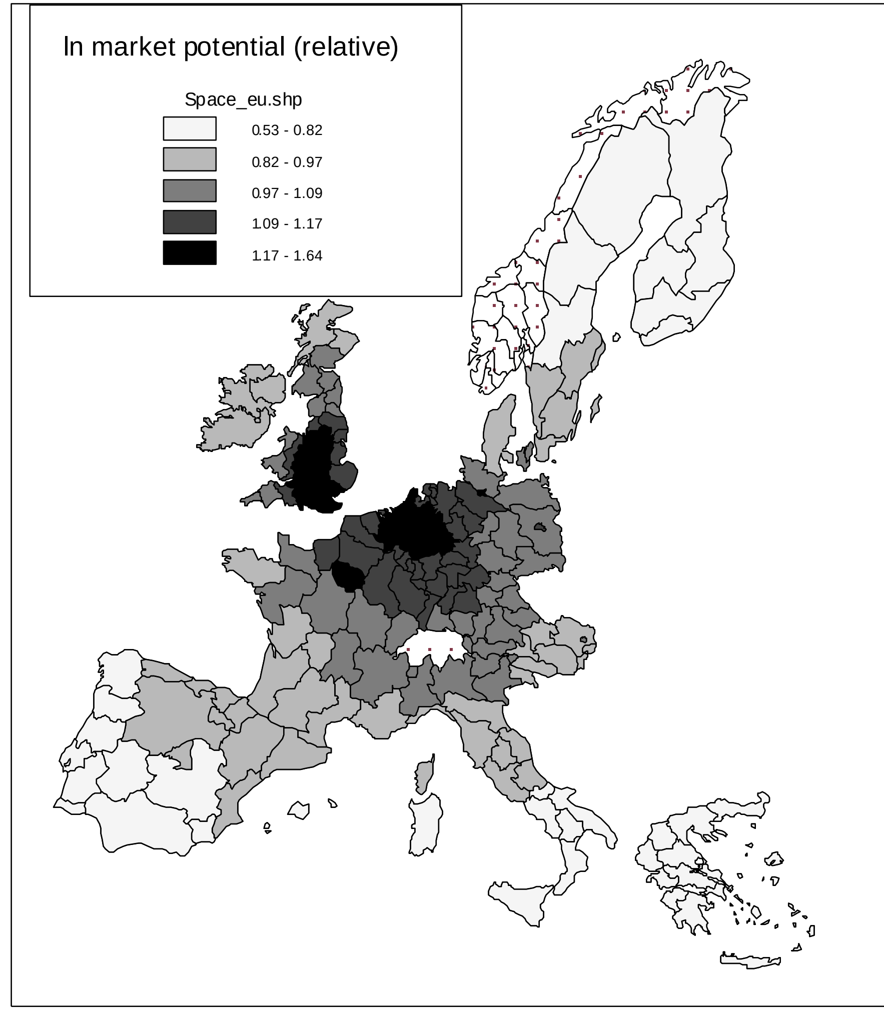
- Need an indication of λ value that would occur when there is no residual autocorrelation
- provided by sampling at random (with replacement) from the residuals
- thus purging residuals of spatial dependence
  - since the order of the sample of residuals is random, so neighbours in residuals will no longer typically be neighbours in the sample of residuals
  - The sample size equal to n and the probability of drawing a specific residual is equal to 1/n
- then calculating λ using the randomly drawn set of residuals

Bootstrap estimation

- Doing this k times gives k λs, each one consistent with null of no residual autocorrelation
- Ranking the k λs gives the empirical cumulative distribution function (or Bootstrap distribution)
- Compare estimated λ with Bootstrap distribution
- Reject the null if estimated λ is extreme with respect to its Bootstrap distribution
  - estimated λ is > 2 standard deviations away from mean of Bootstrap distribution
  - estimated λ may be close to the top ranked λ in the Bootstrap distribution
- indicating that estimated λ will rarely occur when the residuals are random, suggesting that they are not random, i.e. spatially autocorrelated

One year of panel data : 2003





n = 255 t = 9 (1995-2003)

Estimates SAR-RE model

Created by demo_5.m

λ estimate (AR error process) = 0.731368 $\sigma2\_v$ =0.0024103 $\sigma2\_1$ = 0.199964

constant=6.84201 ln MP = 0.386 new entrants =-1.32422

s.e. constant=0.3982 s.e.ln MP = 0.0399597 new entrants =0.0724361

t constant=17.1825 t ln MP = 9.65972 t new entrants =-18.2812

sum of squared residuals = 148.556
correlation between fitted and actual wages = 0.912692

<u>approximate t ratio for lambda</u>

estimated lambda = 0.7314

mean of Bootstrap distribution = -0.0057

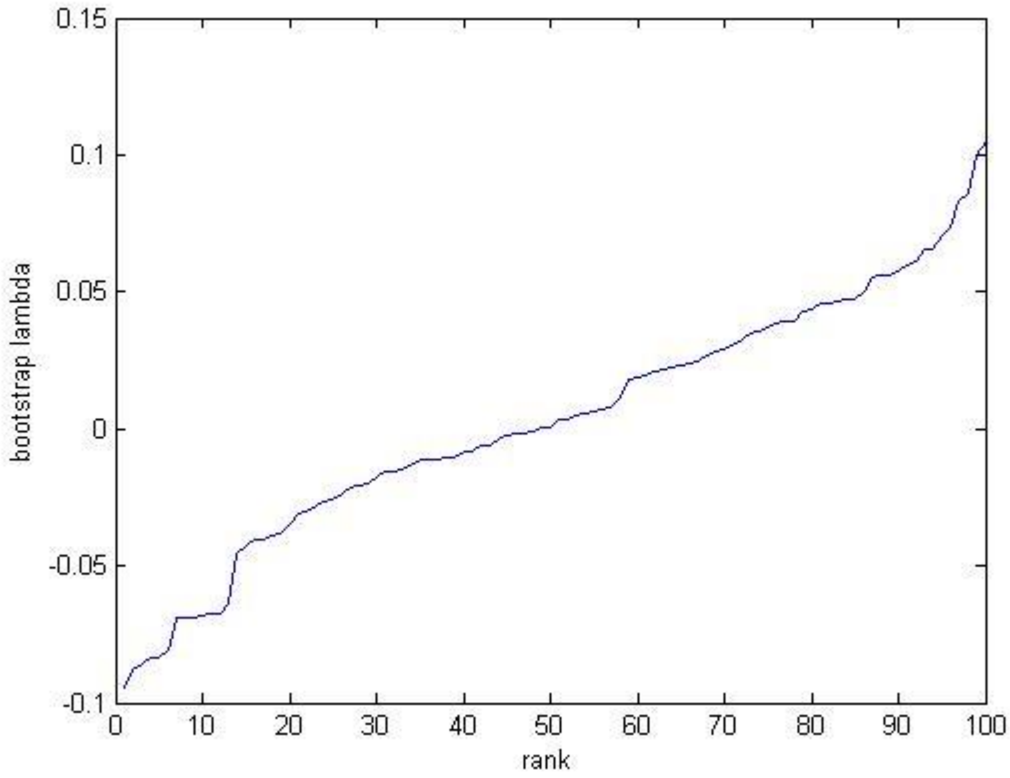
standard deviation of Bootstrap distribution = 0.0453

t_ratio_lambda = 16.0605

n.b. results differ each time because of different random draws

Bootstrap lambda distribution

Created by demo_5.m



n.b. results differ each time because of different random draws

Introducing endogenous variables

- Consistent 2sls residuals used as input into GMM
- Problem of finding instruments
  - lnMP is endogenous by definition
  - Empirical example uses region of area (sq km) and employment per sq. km as instruments

Estimates SAR-RE model

Created by demo_5.m

λ estimate (AR error process) = 0.73128 $\sigma2\_v$ =0.00242336 $\sigma2\_1$ = 0.200278

constant=6.35297 ln MP = 0.435119 new entrants = -1.30697

s.e. constant=0.791254 s.e.ln MP = 0.079462 new entrants =0.076216

t constant= 8.02899 t ln MP = 5.47584 t new entrants =-17.1483

sum of squared residuals = 149.741

correlation between fitted and actual wages = 0.91165

<u>approximate t ratio for lambda</u>

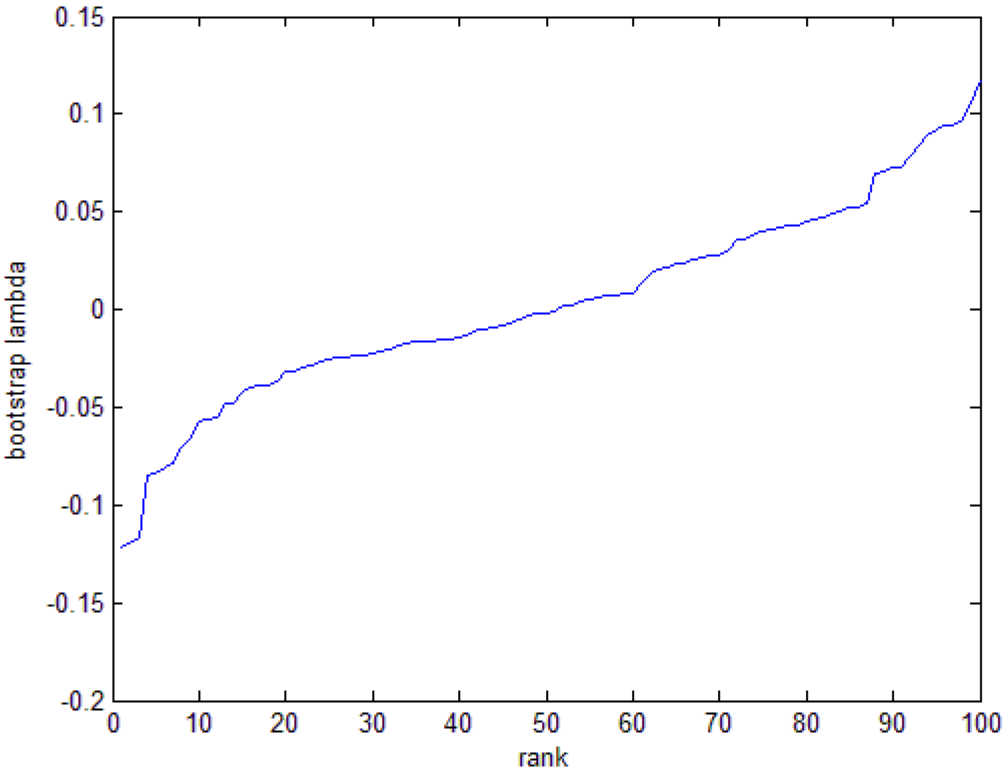
estimated lambda = 0.73128

mean of Bootstrap distribution = -0.0031
standard deviation of Bootstrap distribution = 0.0437
t_ratio_lambda = 16.8011

n.b. results differ each time because of different random draws

Bootstrap lambda distribution

Created by demo_5.m



n.b. results differ each time because of different random draws

Introducing endogenous variables

- Consistent 2sls residuals used as input into GMM
- Problem of finding instruments
- Simplest when instrumenting an endogenous spatial lag Wy
  - Wages (especially for similar work) tend not to be high and low in two nearby places, in-commuting will lower the high wages, and out-commuting will raise the low wages
  - Commuting can also cause efficient workers living in one place to affect labour efficiency in another place where they work, so that labour efficiency spills over
- Use WX as instruments

$$y_{it} = x_{it}\beta + \varepsilon_{it} \rightarrow y_{it} = \rho\sum_{j=1}^{N} W_{ij} y_{jt} + x_{it}\beta + \varepsilon_{it} \qquad i = 1,\ldots,N; t = 1,\ldots,T$$

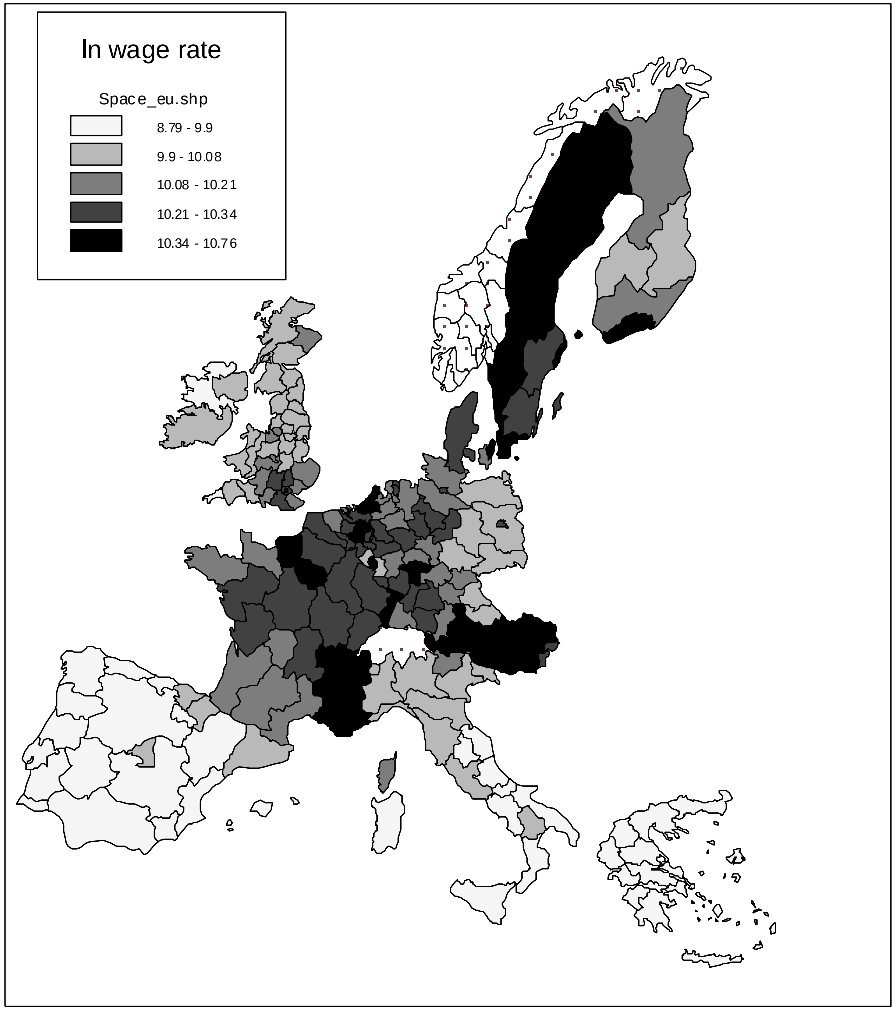
$$\varepsilon_t = \lambda W \varepsilon_t + u_t$$

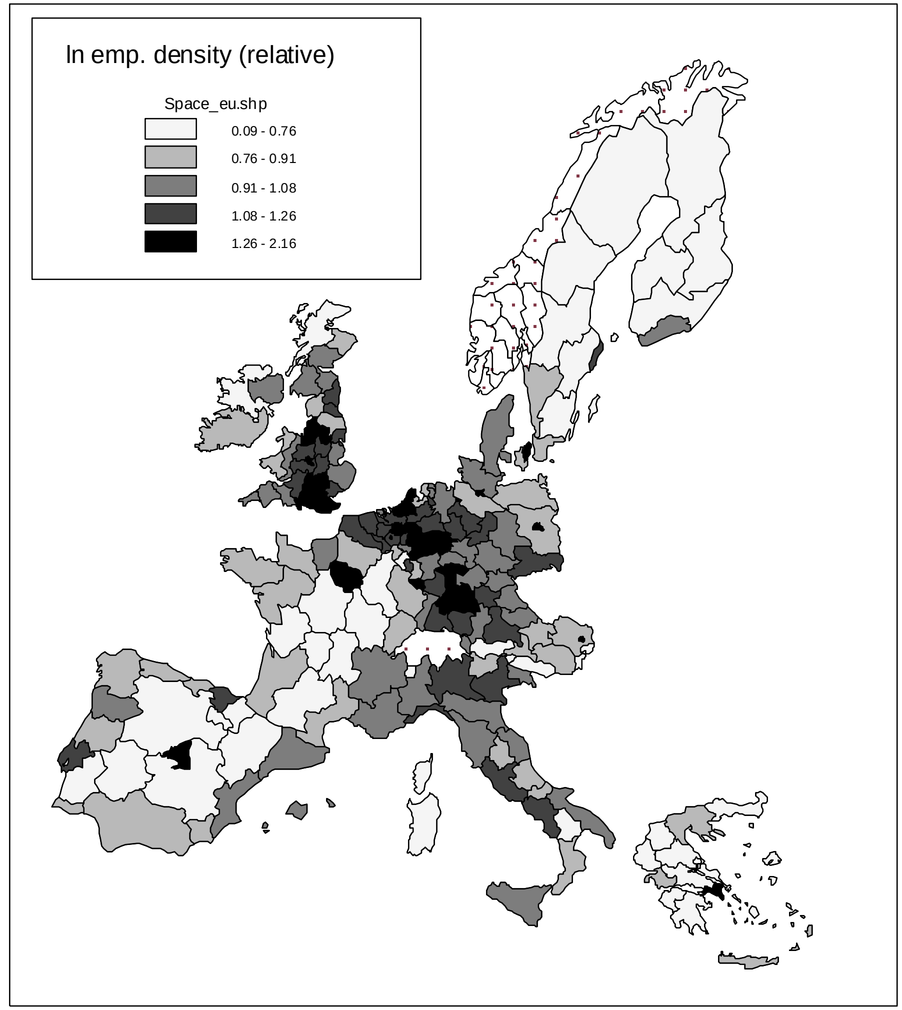
$$u_t = \alpha + v_t \qquad \alpha \sim iid(0,\sigma_\alpha^2) \qquad v_t \sim iid(0,\sigma_v^2)$$

$$\varepsilon = (e_t \otimes B_N^{-1})\alpha + (I_T \otimes B_N^{-1})v$$

Here for simplicity we assume that the matrix W applies to both
The spatial lag and to the spatial error process.
Alternatively, we could use another matrix M (with similar properties
To W) to capture one of the spatial processes.

One year of panel data : 2003





SAR-RE model : with Wy

Created by demo_6.m

λ estimate (AR error process) = 0.659634 $\sigma2\_v$ =0.0038395 $\sigma2\_1$ = 0.195991

constant= 8.13783	Wy =0.21753	ln ed = 0.054	n. entr. = -1.191
s.e. cons. = 0.6892	s.e. Wy = 0.0656	s.e.ln ed = 0.0092	s.e. n entr. =0.075
t cons. = 11.81	t Wy =3.31	t ln ed = 5.79	t n entr. = -15.89

sum of squared residuals = 126.0
correlation between fitted and actual wages = 0.926562

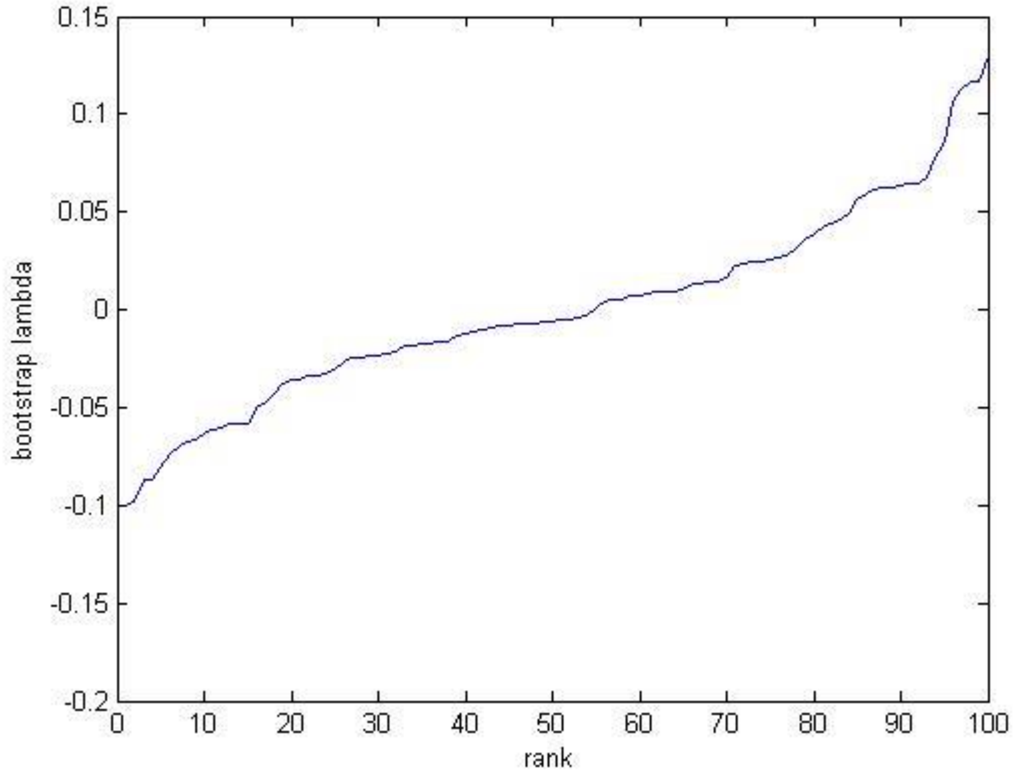
<u>approximate t ratio for lambda</u>

estimated lambda = 0.659634

mean of Bootstrap distribution = 5.6378e-004
standard deviation of Bootstrap distribution = 0.0492
t_ratio_lambda = 13.3956

n.b. results differ each time because of different random draws

Bootstrap lambda distribution



Created by demo_6.m

n.b. results differ each time because of different random draws

prediction

- 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

best linear unbiased predictor (BLUP)

- Goldberger (1962)

$$\hat{y}_{i,T+s} = x_{i,T+s}\hat{\beta} + \omega_i'\Omega^{-1}\hat{\varepsilon}$$

$\hat{y}_{i,T+s}$ = the scalar predicted value for location i in period $T+s$

$x_{i,T+s}$ = 1 by k vector of regressor values at i at $T+s$

ω_i = E[$\varepsilon_{i,T+s}\varepsilon$], an NT by 1 vector of covariances of the future disturbance at location i with the NT by 1 vector of residuals ε

Ω = NT by NT error variance – covariance matrix

$\hat{\varepsilon}$ = estimated residuals

Goldberger, A.S., 1962, Best linear unbiased prediction in the generalized linear regression model, Journal of the American Statistical Association 57, 369.375

BLUP:RE model

- Consider first the RE error components model, that is without any error autocorrelation, so that $\lambda = 0$
- It can be shown that $\hat{y}_{i,T+s} = x_{i,T+s}\hat{\beta} + \omega_i'\Omega^{-1}\hat{\varepsilon}$

reduces to

This equation reappears later

$$\hat{y}_{i,T+s} = x_{i,T+s}\hat{\beta} + \left(\frac{\sigma_\alpha^2}{\sigma_1^2}\right)\left(e_T' \otimes l_i'\right)\hat{\varepsilon}$$

$$e_T = (1,\ldots,1)'$$

$$\sigma_1^2 = T\sigma_\alpha^2 + \sigma_v^2$$

l_i is the i'th column of I_N

$$\hat{y}_{i,T+s} = x_{i,T+s}\hat{\beta} + (T\frac{\sigma_\alpha^2}{\sigma_1^2})\bar{\hat{\varepsilon}}$$

Adds a fraction of the mean residual for Individual i

$$\bar{\hat{\varepsilon}} = \sum_{t=1}^{T}\hat{\varepsilon}_t / T$$

RE-SAR, RE-SMA

- Baltagi & Li(2004,2006)
- The spatial dependence is restricted to the transient disturbances

$$\hat{y}_{i,T+s} = x_{i,T+s}\hat{\beta} + \omega_i'\Omega^{-1}\hat{\varepsilon}$$

reduces to

$$\hat{y}_{i,T+s} = x_{i,T+s}\hat{\beta} + \theta\left(e_T' \otimes l_i' C_1^{-1}\right)\hat{\varepsilon}$$

$$e_T = (1,\ldots,1)'$$

$$C_1 = \left[T\theta I_N + \left(B_N' B_N\right)^{-1}\right]$$

l_i is the i'th column of I_N

SAR-RE,SMA-RE

- The spatial dependence (SAR or MA) occurs in both transient and permanent error components
- The BLUP is identical to BLUP for the basic model with no spatial effects (BBP)

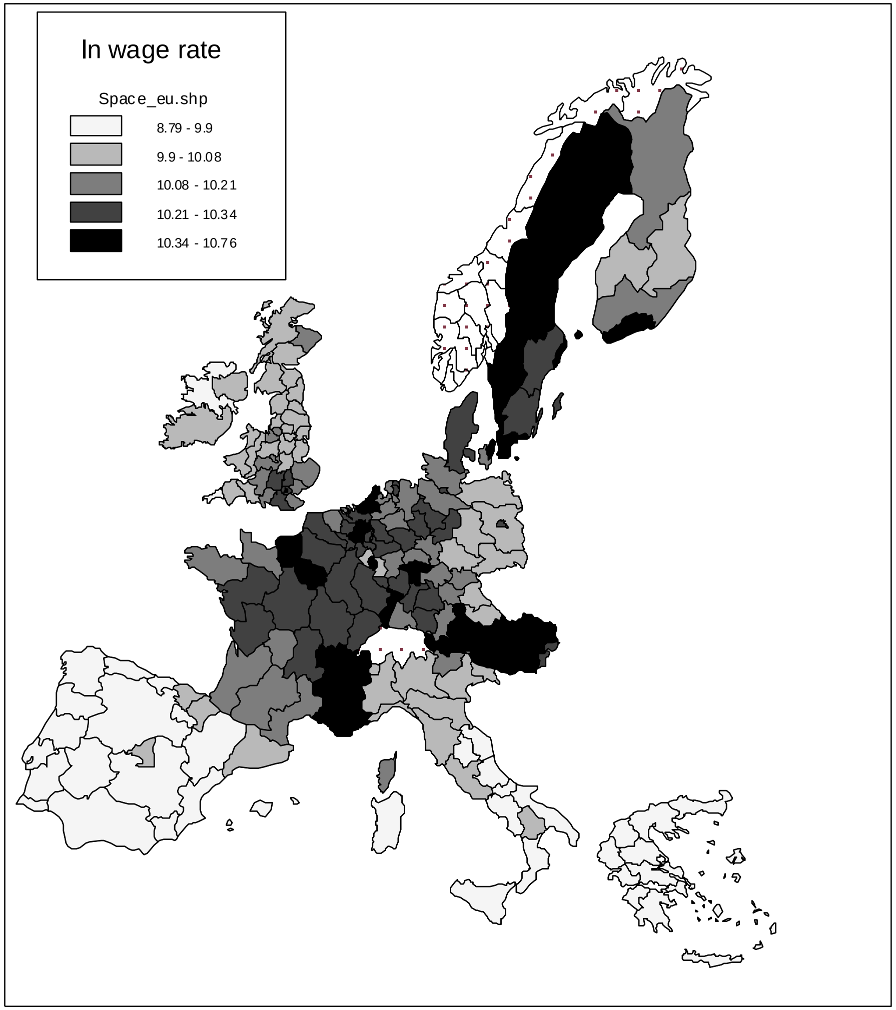
$$\hat{y}_{i,T+s} = x_{i,T+s}\hat{\beta} + \left(\frac{T\sigma_{\alpha}^{2}}{T\sigma_{\alpha}^{2} + \sigma_{\nu}^{2}} \right) \hat{\bar{\varepsilon}}$$

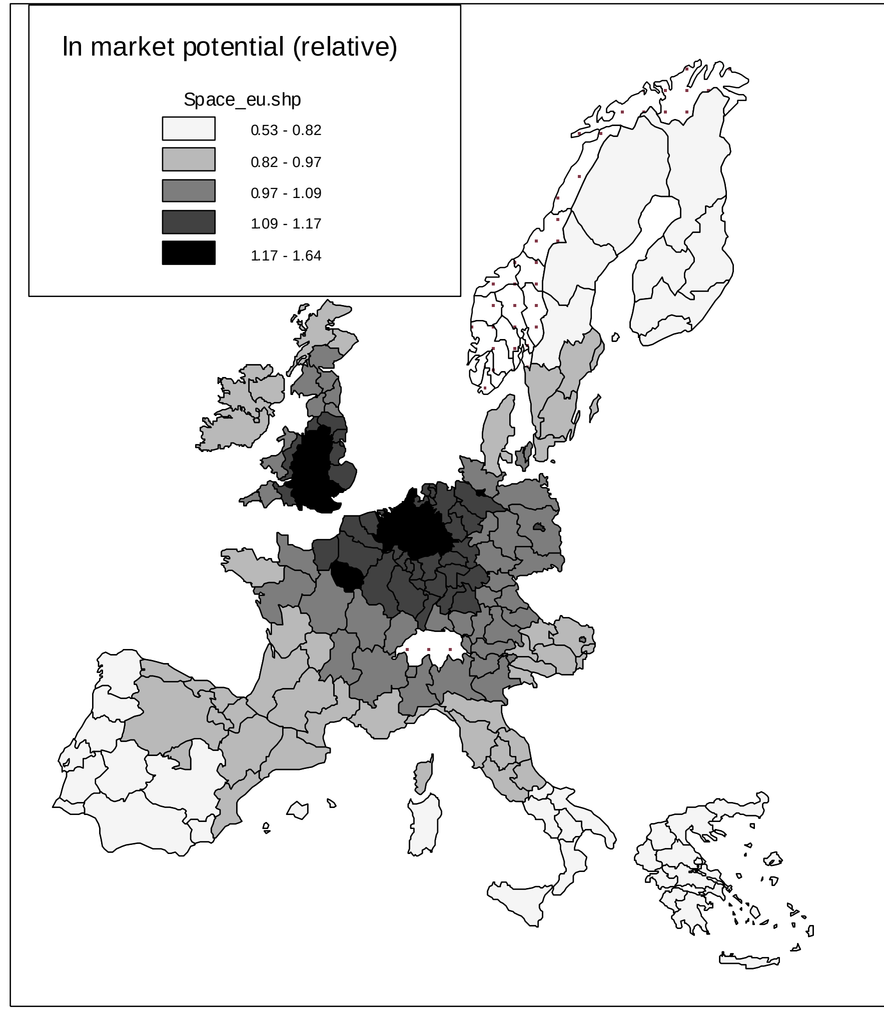
$$\hat{\bar{\varepsilon}}_{j} = \frac{1}{T} \sum_{t=1}^{T} \hat{\varepsilon}_{jt}$$

- But the forecasts are different because the estimators are different (ML versus GLS residuals)

Example: prediction based on SAR-RE model

One year of panel data : 2003





n = 255   t = 7 (1995-2001) n.b. holding 2 years data to check ex-post forecast

Estimates SAR-RE model

Created by demo_7.m

λ estimate (AR error process) = 0.73642 $\sigma2\_v$ =0.00206186 $\sigma2\_1$ = 0.156656

constant=6.34586 ln MP = 0.436234 new entrants = -1.3573

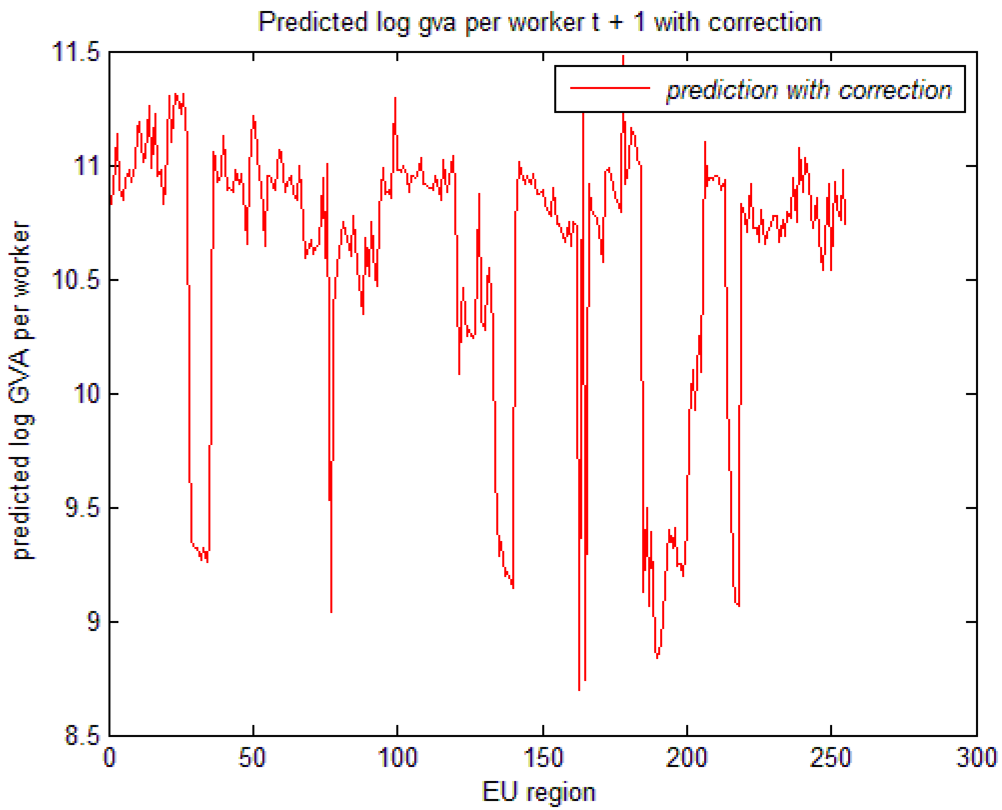
s.e. constant=0.775531 s.e.ln MP = 0.0786159 new entrants = 0.0772075

t constant=8.1826 t ln MP =5.54893 t new entrants =-17.5799

sum of squared residuals = 120.12
correlation between fitted and actual wages = 0.912934

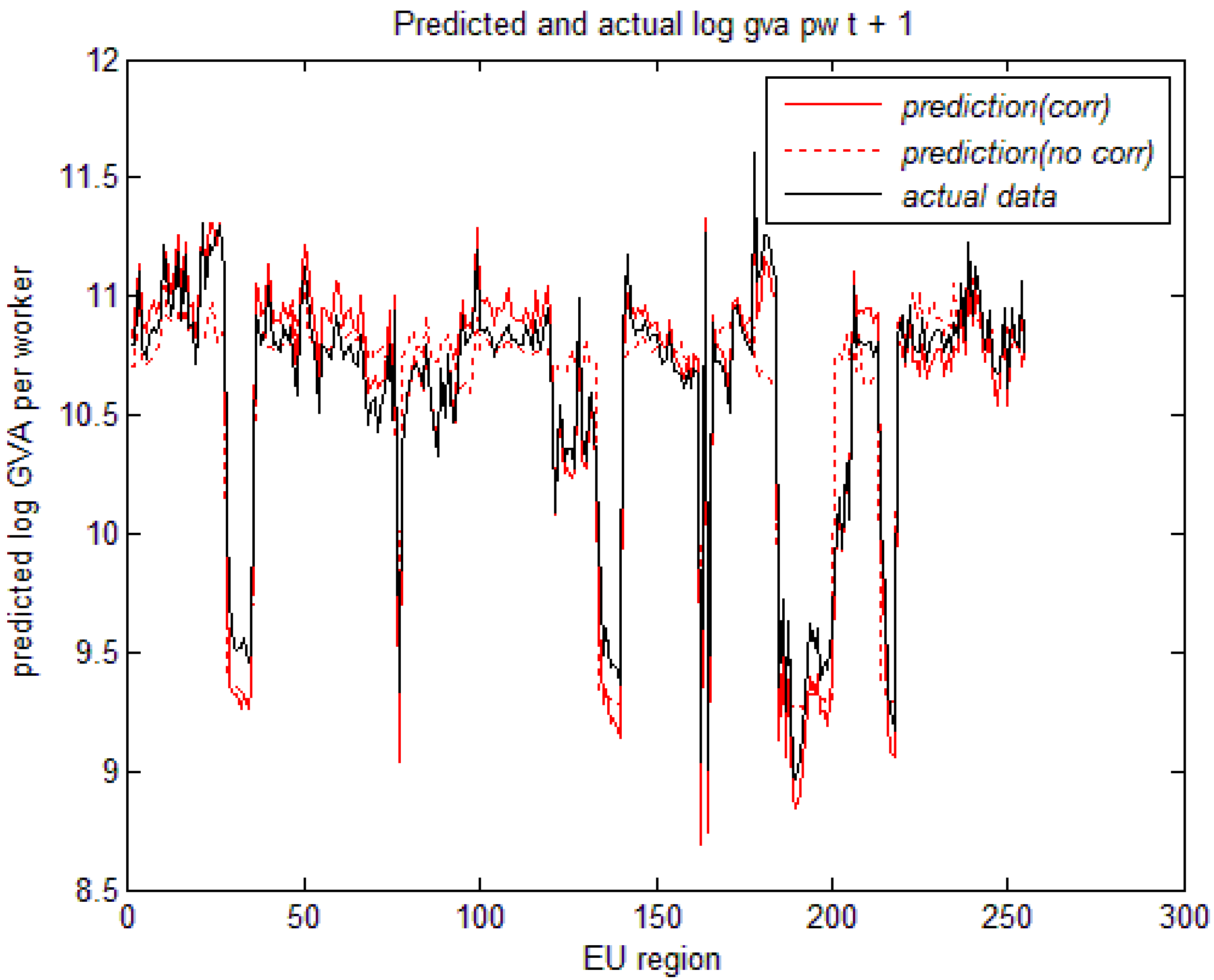
prediction based on SAR-RE model

Created by demo_7.m



prediction based on SAR-RE model

Created by demo_7.m



prediction based on SAR-RE model

Created by demo_7.m

prediction for t + 1

FIT_with_correction =   3.6150

FIT_without_correction =   14.2162

prediction for t + 2

FIT_with_correction =   3.5328

FIT_without_correction = 14.0978

Prediction for SAR-RE model : with spatial lag

- Wy is endogenous, this leads to a specification

$$y_{it} = \rho \sum_{j=1}^{N} W_{ij} y_{jt} + x_{it}\beta + \varepsilon_{it} \qquad i = 1,\ldots,N; t = 1,\ldots,T$$

$$\varepsilon_t = \lambda W \varepsilon_t + u_t$$

$$u_t = \alpha + v_t \qquad \alpha \sim iid(0,\sigma_\alpha^2) \qquad v_t \sim iid(0,\sigma_v^2)$$

$$\varepsilon = (e_t \otimes B_N^{-1})\alpha + (I_T \otimes B_N^{-1})v$$

$$y_t = \rho W y_t + x_t\beta + \varepsilon_t$$

$$(I - \rho W) y_t = x_t\beta + \varepsilon_t \qquad (I - \hat{\rho} W)\hat{y}_{T+s} = x_{T+s}\hat{\beta} + \omega' \Omega^{-1}\hat{\varepsilon}$$

$$y_t = (I - \rho W)^{-1}(x_t\beta + \varepsilon_t) \qquad \hat{y}_{T+s} = (I - \hat{\rho} W)^{-1}(x_{T+s}\hat{\beta} + \omega' \Omega^{-1}\hat{\varepsilon})$$

Prediction for SAR-RE model : with spatial lag

Both equations (1) and (2) are mathematically equivalent

(1) $\hat{y}_{T+s} = (I - \rho W)^{-1}(x_{T+s}\hat{\beta} + \omega'\Omega^{-1}\hat{\varepsilon})$ Fingleton(2008)

(2) $\hat{y}_{i,T+s} = \sum_{k=1}^{K} \hat{\beta}_k \sum_{j=1}^{N} h_{ij} x_{kj,T+s} + \frac{T\sigma_{\alpha}^2}{\sigma_1^2} \sum_{j=1}^{N} h_{ij} \hat{\varepsilon}_j$ Baltagi,Fingleton, Pirotte(2011)

h_{ij} is the i, jth element of $(I - \rho W)^{-1}$

$$\hat{\varepsilon}_j = \frac{1}{T}\sum_{t=1}^{T} \hat{\varepsilon}_{jt}$$

A proof that this is BLUP is available at

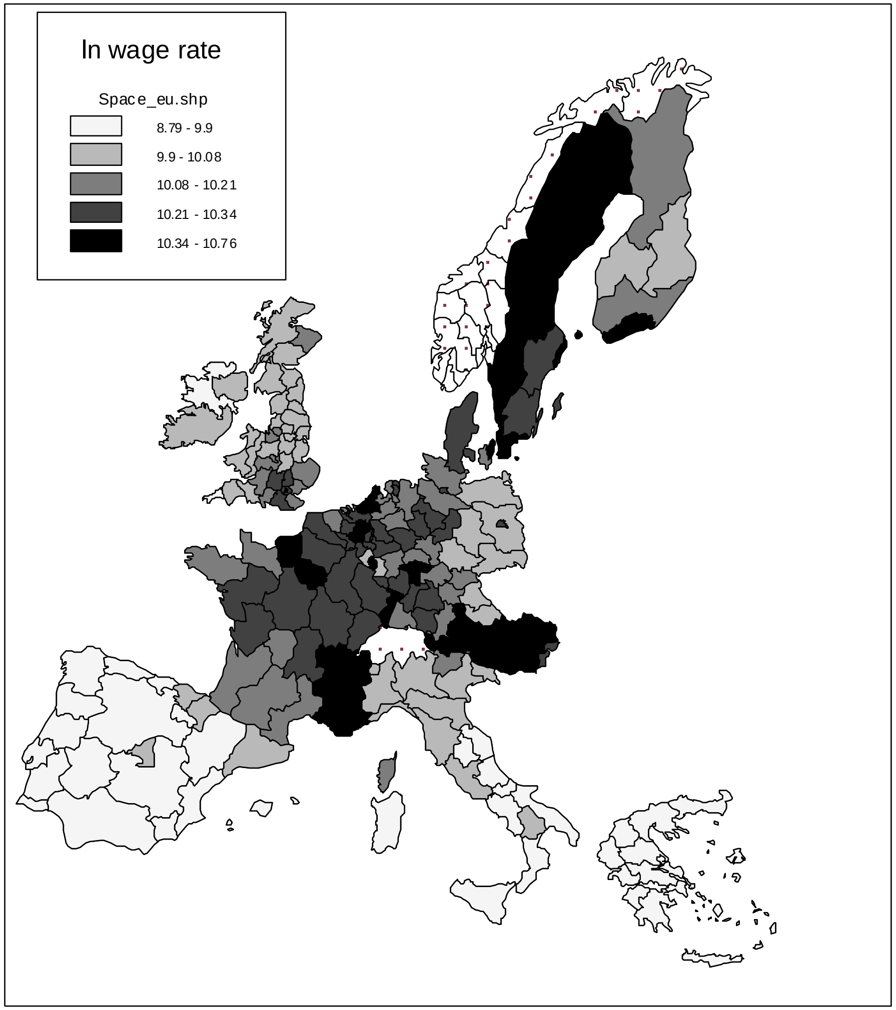
www.spatialeconomics.ac.uk/textonly/SERC/publications/download/sercdp0095.pdf.

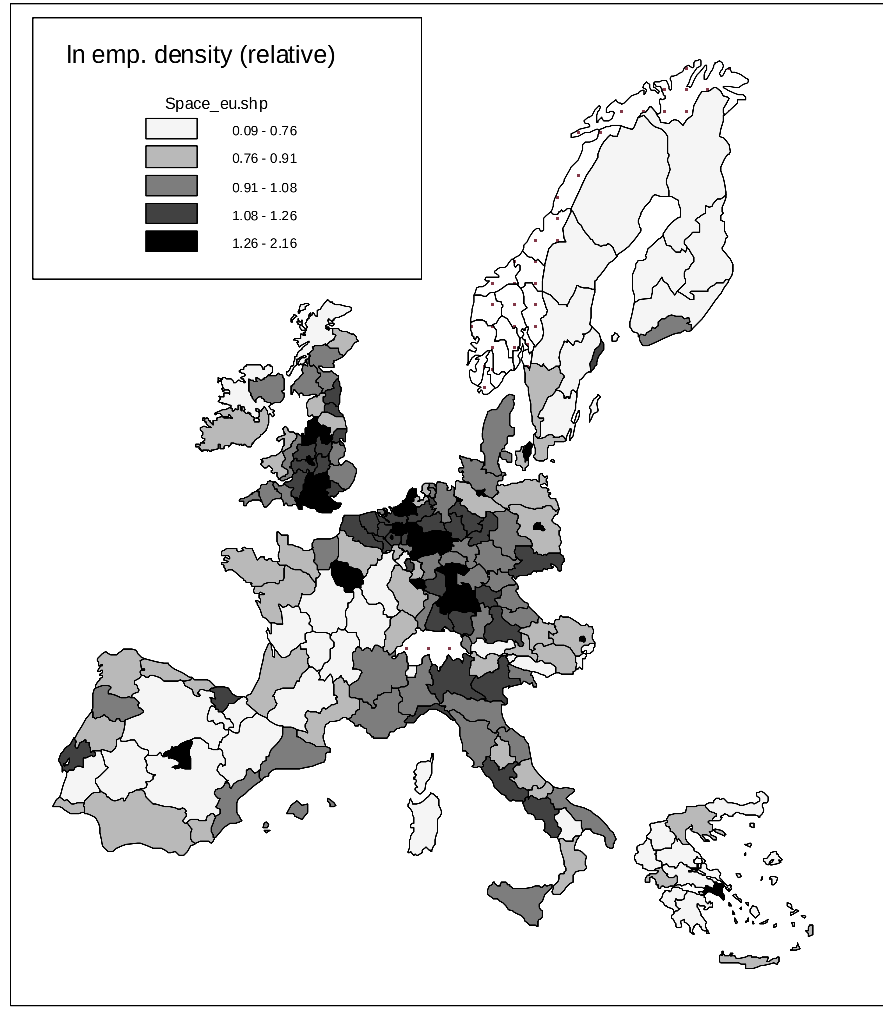
The relative performance of this compared with other prediction equations is given in

Fingleton B(2009) 'Prediction Using Panel Data Regression

with Spatial Random Effects' International Regional Science Review 32 195-220

One year of panel data : 2003





n = 255 t = 7 (1995-2001) n.b. holding 2 years data to check ex-post forecast

SAR-RE model : with Wy

Created by demo_8.m

λ estimate (AR error process) = 0.673976 $\sigma 2\_v$ =0.00301837 $\sigma 2\_1$ = 0.1507

constant= 8.19683 Wy =0.208964 ln ed = 0.0545 n. entr. = -1.245

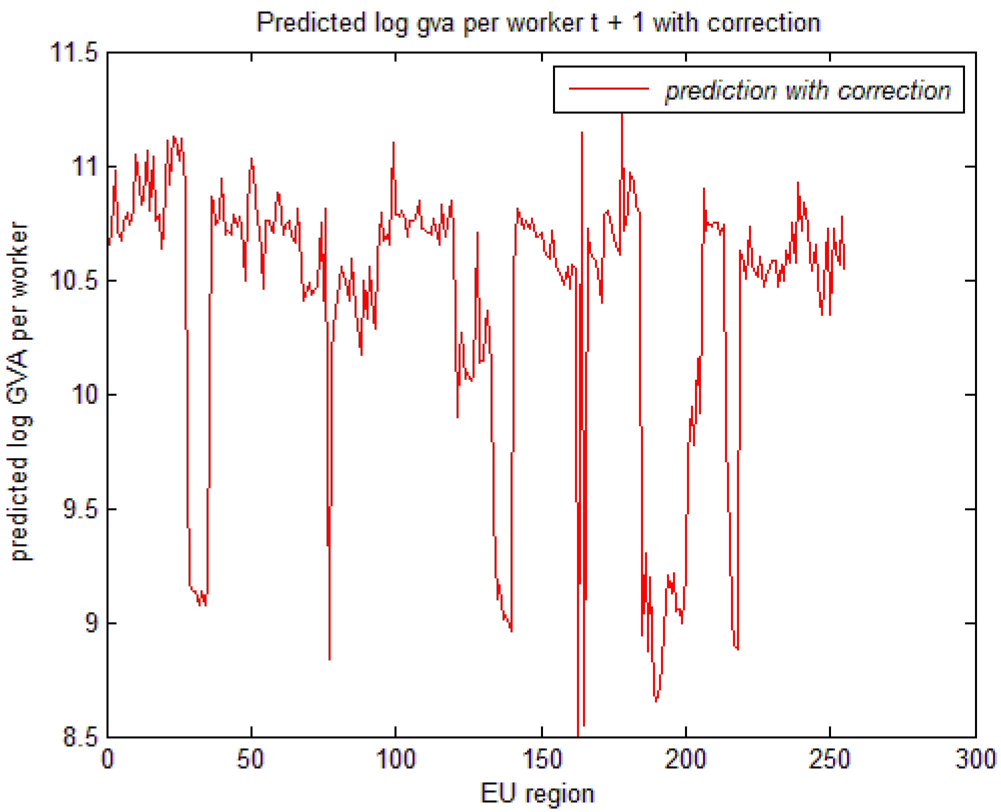
s.e. cons. = 0.67086 s.e. Wy =0.064 s.e.ln ed = 0.0092 s.e. n entr. =0.075

t cons. = 12.22 t Wy =3.26 t ln ed = 5.88 t n entr. = -16.70

sum of squared residuals = 95.0976
correlation between fitted and actual wages = 0.932038

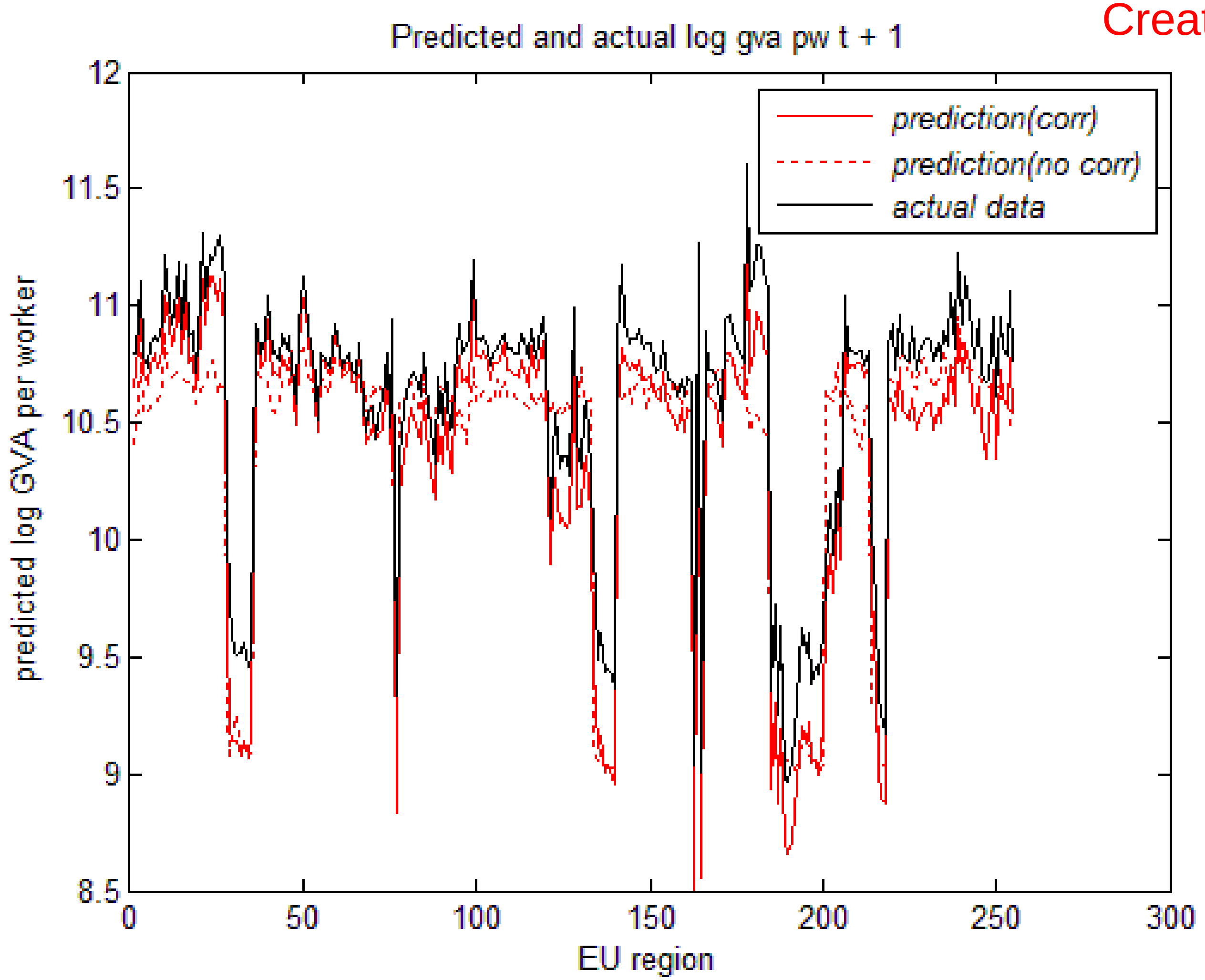
Prediction for SAR-RE model : with spatial lag

Created by demo_8.m



Prediction for SAR-RE model : with spatial lag

Created by demo_8.m



Prediction for SAR-RE model : with spatial lag

Created by demo_8.m

prediction for t + 1

FIT_with_correction = 12.7422

FIT_without_correction = 24.2068

prediction for t + 2

FIT_with_correction = 13.9213

FIT_without_correction = 25.1271

Note worse fit, but this model has an extra parameter estimated compared with the lnMP model

extensions

- Dynamic

a first order spatial autoregressive panel data model

$$y_{it} = a + \gamma y_{it-1} + \rho_1 \sum_{j=1}^{N} w_{ij} y_{jt} + \beta x_{it} + \varepsilon_{it} \qquad i = 1, \ldots, N; \; t = 1, \ldots, T, \quad (44)$$

where the disturbance ε_{it} follows a SAR process:

$$\varepsilon_{it} = \rho_2 \sum_{j=1}^{N} w_{ij} \varepsilon_{jt} + u_{it} \quad (45)$$

where w_{ij} is the (i, j) element of the spatial matrix W_N, and u_{it} has an error component structure

$$u_{it} = \mu_i + v_{it} \quad (46)$$

Baltagi BH, Fingleton B and A Pirotte (2011) ‘Estimating and Forecasting with a Spatial Dynamic Panel Model’ SERCDP0095, Spatial Economics Research centre, London School of Economics

Prediction with dynamic spatial panel

When $\gamma \neq 0$ (i.e. a dynamic model), and $\rho_1 \neq 0$ (i.e. including a spatial lag on the dependent variable) and $\rho_2 \neq 0$ (i.e. including a SAR process on the disturbances ε), the derivation of the predictor is more complicated mainly because the lagged endogenous variable is correlated with the individual effects. Following Chamberlain (1984) and Sevestre and Trognon (1996), we derive the linear predictor of y_{it} conditional upon (y_{10},...,y_{N0},x_{11},...,x_{N1},..., x_{1T},...,x_{NT}) which is given[3] by

$$\begin{aligned} & E^*\left[y_{it} | y_{10}, \ldots, y_{N0}, x_{11}, \ldots, x_{N1}, \ldots, x_{1T}, \ldots, x_{NT}\right] \\ = \; & \gamma^t \sum_{j=1}^{N} h_{ij}^{(t)} y_{j0} + \sum_{l=1}^{t} \gamma^{l-1} \sum_{j=1}^{N} h_{ij}^{(l)} x_{jt-l+1} \beta \\ & + \sum_{l=1}^{t} \gamma^{l-1} \sum_{j=1}^{N} p_{ij}^{(l)} E^*\left[\mu_j | y_{10}, \ldots, y_{N0}\right], \end{aligned} \tag{36}$$

where $h_{ij}^{(l)}$ is the (i,j) element of the matrix $\left(G_N^{-1}\right)^l$, $p_{ij}^{(l)}$ is the (i,j) element of the matrix $\left(\left(\left(G_N\right)^{-1}\right)^l B_N^{-1}\right)$. μ_j and y_{j0} are assumed to be uncorrelated

$$B_N = (I_N - \rho_2 W_N) \qquad G_N^{-1} = (I_N - \rho_1 W_N)^{-1}$$

Conclusion

- ‘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’ Hashem Pesaran(1990)