11th Economics Summer Seminars

Pamukkale University

Denizli

Turkey

Applied Spatial Econometrics

Bernard Fingleton

University of Cambridge

UK

Aim of Course

- To set out the basic approach of spatial econometrics
- To illustrate the practices and usefulness of spatial econometrics for applied economists
- To highlight some of the pitfalls

Structure of course

- Regression and spatial dependence
  - Residual Spatial autocorrelation
- Modelling spatial dependence
  - Spatial lag model, Spatial error model,Spatial Durbin model
- Estimation methods
  - Maximum likelihood estimation (ML)
  - Two stage least squares (2SLS)
  - Generalized Method of Moments (GMM) and Feasible Generalized Spatial 2SLS (FGS2SLS)
- Spatial panel models
  - GMM, FGS2SLS, random effects and spatial dependence
  - Prediction

- Focus on computational aspects and ‘how to do spatial econometrics’ in an applied sense – demonstration programs in MATLAB

Session 1

- The reasons for spatial econometrics :
- Why spatial econometrics?
- What is spatial econometrics?
- Spatial versus time series

Why spatial econometrics?

- Spatial economics now widely recognised in the economics/econometrics mainstream
- Krugman's Nobel prize for work on economic geography
- Importance of network economics (eg Royal Economic Society Easter 2009 School , on 'Auctions and Networks')
- LSE ESRC Centre for Spatial Economics
- Increasing policy relevance : World Bank (2008), *World Development Report 2009,* World Bank, Washington.
- Importantly, much insight can be gained by using spatial econometric tools in addition to more standard time series methods
- Time series methods and spatial econometrics come together in the analysis of spatial panels, which we will look at towards the end of the course

What is spatial econometrics?

- the theory and methodology appropriate to the analysis of spatial series relating to the economy
- spatial series means each variable is distributed not in time as in conventional, mainstream econometrics, but in space.

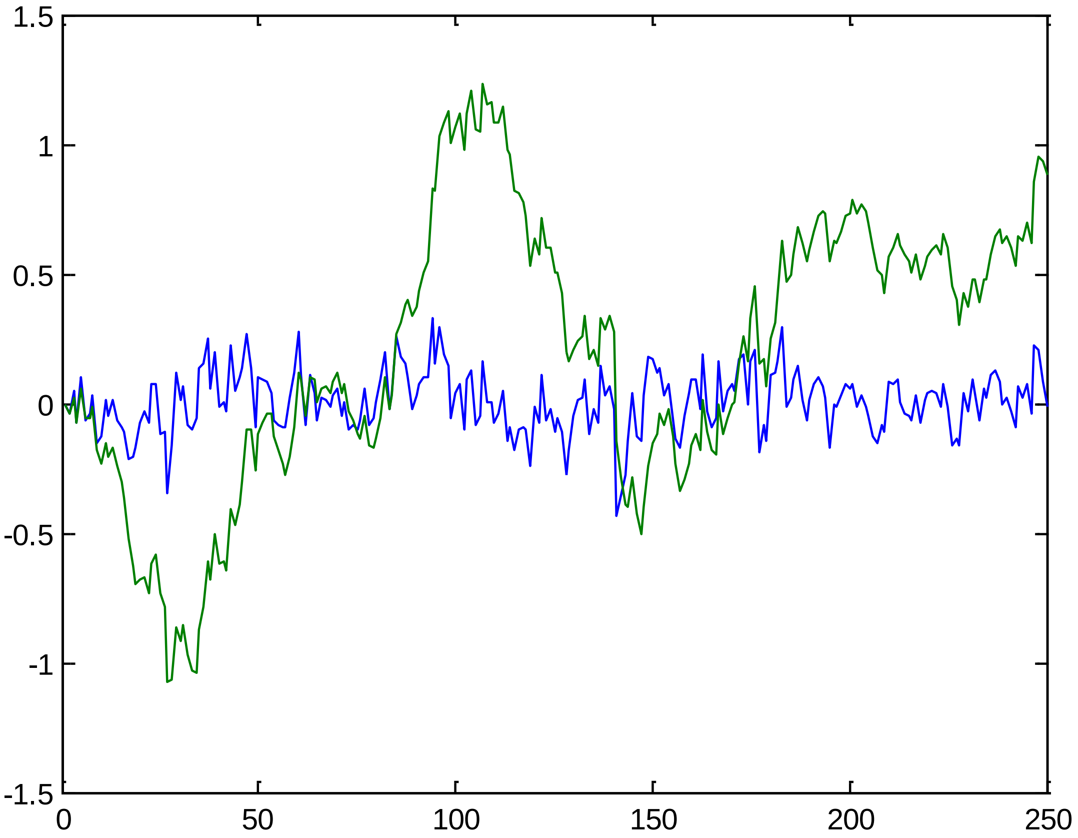
Spatial versus time series

- DGP for time series

$$y(t) = \alpha y(t-1) + \varepsilon(t)$$
$$\varepsilon(1) = y(1) = 0$$
$$\varepsilon \sim iid(0, \sigma^2)$$
$$t = 2...T$$

Spatial versus time series

- DGP for time series



Spatial versus time series

- DGP for time series

$$y = \alpha W y + \varepsilon$$

y is a T x 1 vector

α is a scalar parameter that is estimated

ε is an T x 1 vector of disturbances

DGP for time series

$$y = \alpha W y + \varepsilon$$

W is a TxT matrix with 1s on the minor diagonal, thus for T = 10

$$W = \begin{matrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{matrix}$$

The 1s indicate location pairs that are close to each other in time

DGP for time series

$$y = \alpha W y + \varepsilon$$

Provided Wy and ε are contemporaneously independent we can estimate α by OLS and get consistent estimates, although there is small sample bias.

DGP for spatial series

In spatial econometrics, we have an N x N *W* matrix
N is the number of places.

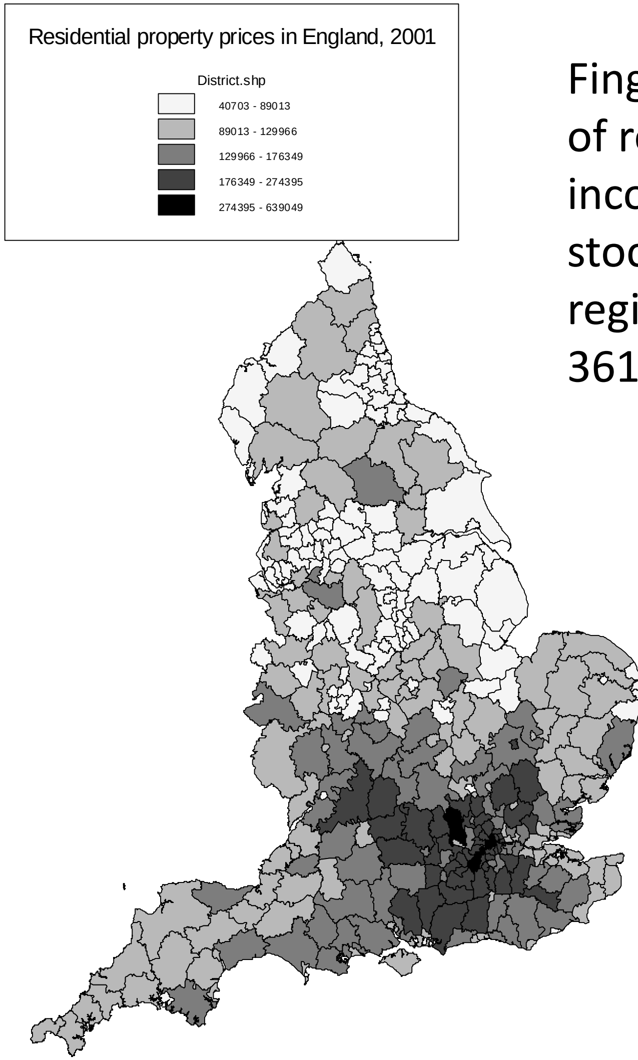
W =

0	0	0	1	0	0	0	0	0	0
0	0	1	1	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0
1	1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	1	1
0	0	0	0	0	0	0	1	0	1
0	0	0	0	0	0	0	0	1	0
0	0	0	0	0	1	0	0	0	1
0	0	0	0	1	0	1	0	0	1
0	0	0	0	1	1	0	1	1	0

N= 353
a portion of the W matrix for Luton(1), Mid Bedfordshire(2), Bedford(3) , South Bedfordshire(4), Bracknell Forest(5), Reading(6), Slough(7), West Berkshire(8), Windsor and Maidenhead(9), Wokingham(10)

The 1s indicate location pairs that are close to each other in space

DGP for spatial series



Fingleton B (2006) 'A cross-sectional analysis of residential property prices: the effects of income, commuting, schooling, the housing stock and spatial interaction in the English regions' *Papers in Regional Science* 85 339-361

N= 353

We refer to these small areas
As UALADs

DGP for spatial series

$$y = \rho W y + \varepsilon$$

y is an N x 1 vector

ρ is a scalar parameter that is estimated

ε is an N x 1 vector of disturbances

DGP for spatial series

$$y = \rho W y + \varepsilon$$

- This is an almost identical set-up to the time series case
  And one might think that it can also be consistently estimated by OLS
- But now there is one big difference
- we cannot estimate the spatial autoregression by OLS
  and obtain consistent estimates of ρ.
- Reason - Wy and ε are not independent.
- Wy determines y but is also determined by y.

But more about this later……

Regression and spatial dependence

- Typically in economics we working with regression models, thus

$$y_t = \sum_k x_{tk}\beta_k + \varepsilon_t$$

- But in spatial economics typically the analysis is cross-sectional, so that

$$y_i = \sum_k x_{ik}\beta_k + \varepsilon_i$$

Regression and spatial dependence

$$y_i = \sum_k x_{ik}\beta_k + \varepsilon_i$$

y_i = Observed value of dependent variable *y* at location i (i = 1,…,N)

x_{ik} = Observation on explanatory variable x_k at location i, with k = 1,…,K

β_k = regression coefficient for variable x_k

ε_i = random error term or disturbance term at location i

Let us assume as in the classic regression model that the errors ε_i simply represent unmodelled effects that appears to be random. We therefore commence by assuming that $E(\varepsilon_i) = 0, Var(\varepsilon_i) = \sigma^2, E(\varepsilon_i, \varepsilon_j) = 0$ for all i,j . The assumption is that the errors are identically and independently distributed. For the purposes of inference we might specify the error as a normal distribution.

Regression and spatial dependence

- Writing our model in matrix terms gives

$$y = X\beta + \varepsilon$$

y is an N x 1 vector

X is an N x k matrix

β is a k x 1 vector

ε is an N x 1 vector

$$E(\varepsilon) = 0, E(\varepsilon\varepsilon') = \sigma^2 I$$

- And spatial dependence manifests itself as spatially autocorrelated residuals

$$\hat{\varepsilon} = y - \hat{y} = y - X\hat{\beta}$$

Residual Spatial autocorrelation

- This term is analogous to autocorrelation in time series, which is when the residuals at points that are close to each other in time/space are not independent.
    - For instance they may be more similar than expected (positive autocorrelation) for some reason.
- suggesting that something is wrong with the model specification that is assuming they are independent.
    - For example the errors/disturbances/residuals may contain the effects of omitted effects that vary systematically across space.

Moran's *I*

- Based on *W* matrix
  - A spatial weights matrix is an N x N with non-zero elements in each row *i* for those columns *j* that are in some way neighbours of location *i*
  - The notion of neighbour is a very general one, it may mean that they are close together in terms of miles or driving time, or it may be distance in some more abstract economic space or social space that is not really connected to geographical distance.
  - The simplest form of distance might be contiguity, with W_{ij}= 1 if locations i and j are contiguous, and W_{ij} = 0 otherwise.
  - Usually (but not necessarily) *W* is standardised so that all the values in row *i* are divided by the sum of the row *i* values.

Calculating Moran's *I*

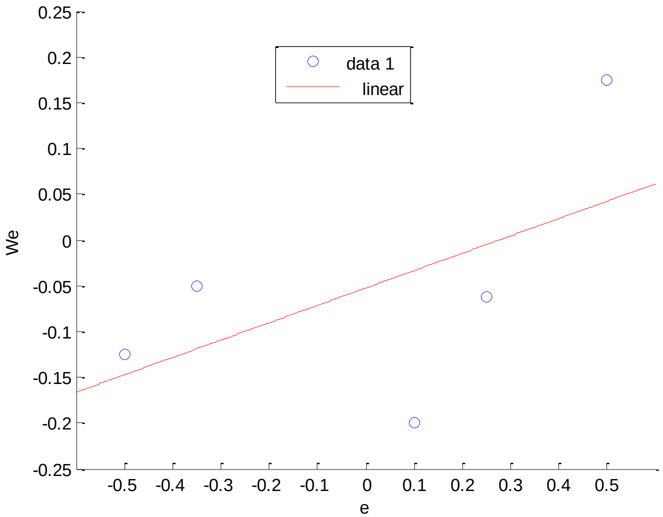
think of Moran's I as approximately the correlation between the two vectors $W\hat{\varepsilon}$ and $\hat{\varepsilon}$. We can show this for a 5 location analysis in graphical form, known as a Moran scatterplot.

$$W = \begin{matrix} 0 & 0.5000 & 0.5000 & 0 & 0 \\ 0.3330 & 0 & 0.3330 & 0.3330 & 0 \\ 0.3330 & 0.3330 & 0 & 0.3330 & 0 \\ 0.2500 & 0.2500 & 0.2500 & 0 & 0.2500 \\ 0 & 0 & 0.5000 & 0.5000 & 0 \end{matrix}$$

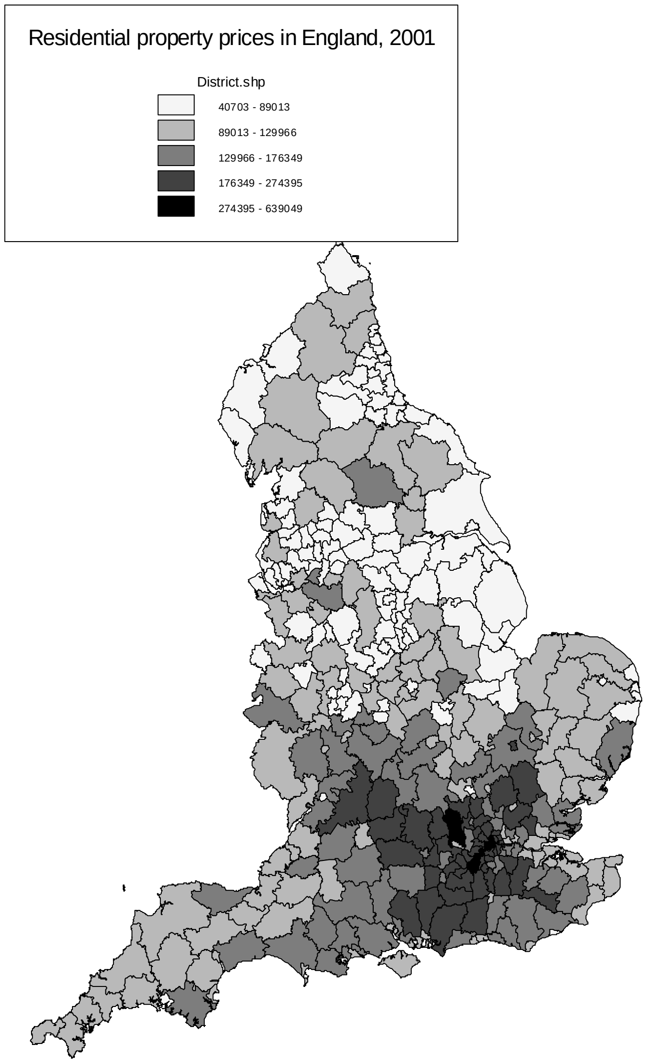
$$\hat{\varepsilon} = \begin{matrix} -0.5000 \\ -0.3500 \\ 0.1000 \\ 0.2500 \\ 0.5000 \end{matrix}$$

Hence -0.1250= 0.5 x -0.35 + 0.5 x 0.1.

$$W\hat{\varepsilon} = \begin{matrix} -0.1250 \\ -0.0500 \\ -0.1998 \\ -0.0625 \\ 0.1750 \end{matrix}$$



Average House prices in local authority areas in England (UALADs)



N= 353

Calculating Moran's I in practice

- Let us look at our map of house prices.
- Can we build a model explaining this variation?
- Do we have spatially autocorrelated residuals?
  - The presence of spatial autocorrelation would suggest there is some specification error,
    - either omitted spatially autocorrelated variable
    - residual heterogeneity
    - or a spatial autoregressive error process

Calculating Moran's I in practice

$$y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \varepsilon$$

y = mean residential property price in each of N local authority areas

X_1 = 1, the constant, an N x 1 vector of 1s

X_2 = total income in each local authority area

X_3 = income earned within commuting distance of each local authority area

X_4 = local schooling quality in each local authority area

X_5 = stock of properties in each local authority area

$$y = X\beta + \varepsilon$$

X is a N x k matrix

β is a k x1 vector

$$\hat{\varepsilon} = y - X\hat{\beta}$$

Created by demo_0.m

Dependent variable *y*		
	estimate	t ratio
Constant *(X_1)*	-571.874	-6.47
Local income *(X_2)*	864.0059	10.02
Within-commuting-distance income *(X_3)*	57.7055	14.08
Schooling quality *(X_4)*	175802.9235	7.74
Number of households *(X_5)*	-0.7112	-6.46
R^2 adjusted	0.567	
Standard Error	42.113	
Moran's I	0.39369	11.29
Degrees of freedom	348	

the value for Moran's I is 11.29 standard errors above expectation. Expectation is the expected value of I under the null hypothesis of no residual autocorrelation. It is clear that there is very significant residual autocorrelation.

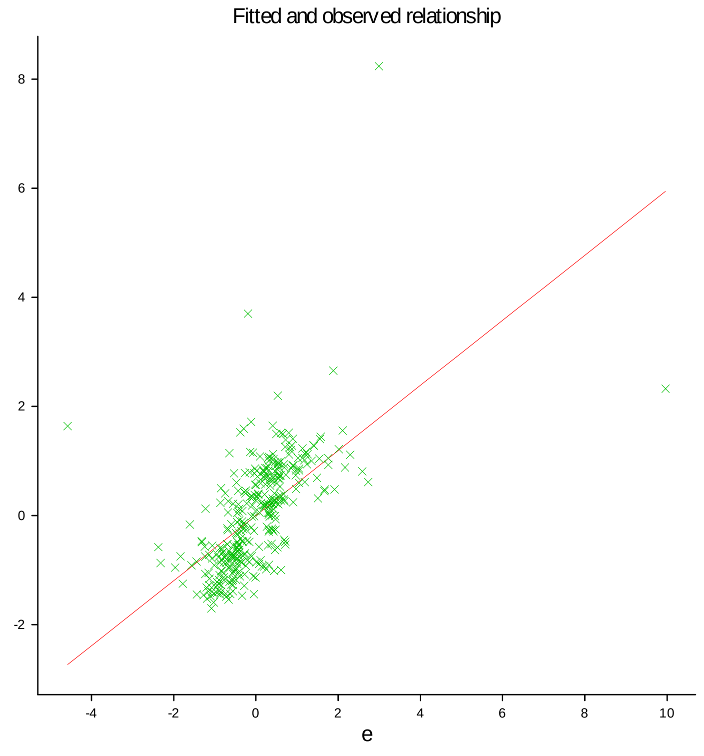
Calculating Moran's I in practice

- What is W?

$$W^* = \frac{1}{d_{ij}^2} \qquad W_{ij} = \frac{W_{ij}^*}{\sum_j W_{ij}^*}$$

Moran scatterplot

$$W\hat{\varepsilon} \text{ versus } \hat{\varepsilon}$$



Calculating Moran's I in practice

- Unfortunately calculating Moran's *I* is not easy without the help of software. I give the formulae below just to show how difficult this is!

$$I = \frac{\hat{\varepsilon}' W \hat{\varepsilon} / S_0}{\hat{\varepsilon}' \hat{\varepsilon} / N}$$

$$S_0 = \sum_i \sum_j W_{ij}$$

- We have seen all of these terms except S_0
- If we row-standardise, so that each row sums to 1 then $S_0 = N$
- So then

$$I = \frac{\hat{\varepsilon}' W \hat{\varepsilon}}{\hat{\varepsilon}' \hat{\varepsilon}}$$

Calculating Moran's I in practice

- Given *I*, we need to compare it with what we would expect under the null hypothesis of no residual autocorrelation

$$E(I) = tr(MW)/(N-K)$$

$$M = I - X(X'X)^{-1}X'$$

$$Var(I) = \frac{tr(MWMW') + tr(MWMW) + [tr(MW)]^2}{(N-K)(N-K+2)} - (E(I))^2$$

- These are the moments we would expect if the residuals were independent draws from a normal distribution

Calculating Moran's I in practice

- The test statistic is Z, which has the following distribution under the null hypothesis

$$Z = \frac{I - E(I)}{\sqrt{Var(I)}} \sim N(0,1)$$

- if *Z* > 1.96 or *Z* < -1.96 then we reject the null of no spatial autocorrelation and infer that there is spatial autocorrelation in the regression residuals. In making this conclusion, we should add the caveat that there is a 5% chance of a Type I error, false rejection of the null
- In the case of our house price data, *I* is 19.96 standard deviations above expectation, a really clear indication that there is positive residual spatial autocorrelation

Calculating Moran’s I in practice

- Positive spatial autocorrelation is when residuals than are close to each other take similar values. Negative spatial autocorrelation would be when the residuals (coded black for negative and white for positive) formed a chequer board pattern if the regions were squares.
- There are several alternatives to Moran’s I, and Moran’s I may also detect things other than spatially autocorrelated residual
  - Moran’s I will also tend to detect heteroscedasticity, that is when the residuals have different variances rather than a common variance.
- However despite this it is the most famous and commonly used method of detecting spatial autocorrelation in regression residuals.

Modelling spatial dependence

- Say we have a significant Moran's I static, what next?
- We need to eliminate the spatial dependence
- one way to do this is to introduce an autoregressive lag (spatial lag model)

$$y = X\beta + \varepsilon$$

X is a N x k matrix

β is a k x1 vector

ε is an N x 1 vector of errors

→

$$y = \rho W y + X\beta + \varepsilon$$

ρ is a scalar parameter

W is an Nx N matrix

Spatial lag model

- Here I list the values of these variables for the first 10 of the UALADs.

district	uaname	y	Wy
1.0	Luton	87464	168313
2.0	Mid_Bedfordshire	138856	151526
3.0	North_Bedfordshire	117530	137574
4.0	South_Bedfordshire	126650	157673
5.0	Bracknell_Forest	167633	200166
6.0	Reading	150094	186756
7.0	Slough	126361	222769
8.0	West_Berkshire	209543	170172
9.0	Windsor_and_Maidenhead	273033	183066
10.0	Wokingham	203059	205737

We can check whether Wy is a significant variable by adding it to our model

$$y = \rho W y + X\beta + \varepsilon$$

Dependent variable *y*	Spatial lag ML	
	estimate	t ratio
Constant *(X_1)*	-541.135534	-8.02
Local income (X_2)	393.33	5.58
Within-commuting-distance income (X_3)	27.45	6.89
Schooling quality (X_4)	149842.21	8.61
Number of households (X_5)	-0.35	-4.10
Spatial lag *(Wy)*	0.6089	14.90
R^2 adjusted	0.6330	
Standard Error	32.13	
Degrees of freedom	347	

Dependent variable *y*	ols	
	estimate	t ratio
Constant *(X_1)*	-571.874	-6.47
Local income (X_2)	864.0059	10.02
Within-commuting-distance income (X_3)	57.7055	14.08
Schooling quality (X_4)	175802.9235	7.74
Number of households (X_5)	-0.7112	-6.46
R^2 adjusted	0.567	
Standard Error	42.113	
Moran's I	0.39369	11.29
Degrees of freedom	348	

Created by demo_0.m

Fingleton B (2006) 'A cross-sectional analysis of residential property prices: the effects of income, commuting, schooling, the housing stock and spatial interaction in the English regions' *Papers in Regional Science* 85, 339-361

Direct, indirect and total effects in spatial lag model

- With *Wy*, the true effect of a variable, which typically is not the same as β, as emphasized by LeSage and Pace (2009)
- the effects on dependent variable of a unit change in an exogenous variable , the derivative $\partial y/\partial X$ is not simply equal to the regression coefficient β
- the true derivative also takes account of the spatial interdependencies and simultaneous feedback embodied in the model, leading to a total effect that differs somewhat (typically) from β
- This derivative is somewhat complicated because it depends on the individual observations but can be represented by a mean
- See also Corrado and Fingleton(2012)

Corrado L. & Fingleton B. (2012) Where is the economics in spatial econometrics? Journal of Regional Science 52 210-239

Direct, indirect and total effects in spatial lag model

- Total effect = direct effect + indirect effect
- So we can partition the average total effect into a direct and an indirect component
- The average direct effect gives the effect of *X* on *y* when the locations of *X* and *y* are the same
  - direct effect is somewhat different from β because at location *r*, a change in *X* affects *y*, which then affects *y* at location *s* (*s n.e. r*) and so on, cascading through all areas and coming back to produce an additional effect on *y* at *r*
- The difference between the total effect and the direct effect is the average indirect effect of a variable
- The average indirect effect gives the effect of *X* on *y* when *X* and *y* are not in the same location

Direct, indirect and total effects given lagged dependent variable
ML estimates : spatial lag model

Variable	Coefficient	Asymptot t-stat	z-probability
const	-541.135534	-8.023904	0.000000
local_income	393.326396	5.577120	0.000000
commuting_income	27.450614	6.894253	0.000000
supply	-0.353572	-4.104351	0.000041
schooling	149842.210059	8.613959	0.000000
rho	0.608979	14.898820	0.000000

Direct	Coefficient	t-stat	t-prob
local_income	439.507710	5.895381	0.000000
commuting_income	30.341573	7.244365	0.000000
supply	-0.394318	-4.179838	0.000037
schooling	167313.290102	8.807929	0.000000

Indirect	Coefficient	t-stat	t-prob
local_income	578.415728	5.010304	0.000001
commuting_income	39.749552	6.981685	0.000000
supply	-0.520811	-3.590056	0.000377
schooling	222208.985696	4.900776	0.000001

Total	Coefficient	t-stat	t-prob
local_income	1017.923438	5.870218	0.000000
commuting_income	70.091125	8.466974	0.000000
supply	-0.915129	-3.999705	0.000077
schooling	389522.275797	6.582510	0.000000

Created by demo_0.m

Spatial error model

- a second way we may model the residual dependence detected by Moran's I
- spatial autocorrelation does not enter as an additional variable. But it is captured by the covariance structure of the errors
- The linear regression with spatially autoregressive errors is the most relevant in many cases, since spatial dependence in the error term is likely to be present in most data sets for contiguous spatial areas

Usually in regression models we assume that $E(\varepsilon) = 0, E(\varepsilon\varepsilon') = \sigma^2 I$, which implies that for each area k the expected value of the error is 0, and the variance of the error distribution is σ^2 . We also assume no covariance across areas, so that $E(\varepsilon_j, \varepsilon_k) = 0$. The error at j is unrelated to the error at k, for all j,k. Now when we model spatial error dependence, we assume that $E(\varepsilon_j, \varepsilon_k) \neq 0$.

an autoregressive error process

$$\varepsilon_i = \lambda \sum_{j=1}^{N} W_{ij} \varepsilon_j + u_i$$

$$E(u) = 0, E(uu') = \sigma^2 I$$

$$E(u_j, u_k) = 0$$

In matrix notation, the spatial error model is

$$y = X\beta + \varepsilon$$

$$\varepsilon = \lambda W \varepsilon + u$$

$$E(\varepsilon\varepsilon') = \sigma^2 (I - \lambda W)^{-1} (I - \lambda W')^{-1}$$

Dependent variable y	Spatial error ML	
	estimate	t ratio
Constant (X_1)	-412.944662	-5.87
Local income (X_2)	291.17	3.15
Within-commuting-distance income (X_3)	49.06	7.83
Schooling quality (X_4)	134152.03	7.29
Number of households (X_5)	-0.29	-3.03
Spatial error lambda	0.740976	19.53
R^2 adjusted	0.7453	
Standard Error	32.07	
Degrees of freedom	347	

Dependent variable y	ols	
	estimate	t ratio
Constant (X_1)	-571.874	-6.47
Local income (X_2)	864.0059	10.02
Within-commuting-distance income (X_3)	57.7055	14.08
Schooling quality (X_4)	175802.9235	7.74
Number of households (X_5)	-0.7112	-6.46
R^2 adjusted	0.567	
Standard Error	42.113	
Moran's I	0.39369	11.29
Degrees of freedom	348	

Created by demo_0.m

Fingleton B (2006) 'A cross-sectional analysis of residential property prices: the effects of income, commuting, schooling, the housing stock and spatial interaction in the English regions' *Papers in Regional Science* 85, 339-361

The spatial Durbin model: a 'catch all' spatial model

This includes a spatial lag *Wy* and a set of spatially lagged exogenous regressors WX

$$y = \rho Wy + X\beta + WX\gamma + \varepsilon$$

y = the dependent variable, an N x 1 vector
Wy = the spatial lag, an N x 1 vector
X = an N x K matrix of regressors, with the first column equal to the constant
β = a K x 1 vector of regression coefficients
ρ = the spatial lag coefficient
ε = an N x1 vector of errors
WX is the N by K matrix of exogenous lags resulting from the matrix product of W and X
γ is the corresponding coefficient vector.

Restricting the parameters of the spatial Durbin leads back to the spatial lag model or to the spatial error model

spatial Durbin model : ML estimates

Created by demo_0.m

Variable	Coefficient	Asymptot t-stat	z-probability
const	-513.835677	-4.146915	0.000034
local_income	-7.730616	-0.083091	0.933780
commuting_income	40.795703	6.257112	0.000000
supply	-0.103221	-1.106877	0.268347
schooling	134249.627896	7.733356	0.000000
Wlocal_income	974.661531	6.096601	0.000000
Wcommuting_income	-25.325850	-3.358633	0.000783
Wsupply	-0.496569	-3.109303	0.001875
Wschooling	8596.323682	0.265708	0.790464
rho	0.621996	13.257551	0.000000

Rbar-squared = 0.6549

Standard Error = 873.1750^0.5 = 29.55

Special cases of the spatial Durbin

spatial lag

$$y = \lambda W y + X\beta + WX\gamma + \varepsilon$$

if $\gamma = 0$

then $y = \lambda W y + X\beta + \varepsilon$

spatial error

$$y = \lambda W y + X\beta + WX\gamma + \varepsilon$$

if $\gamma = -\lambda\beta$

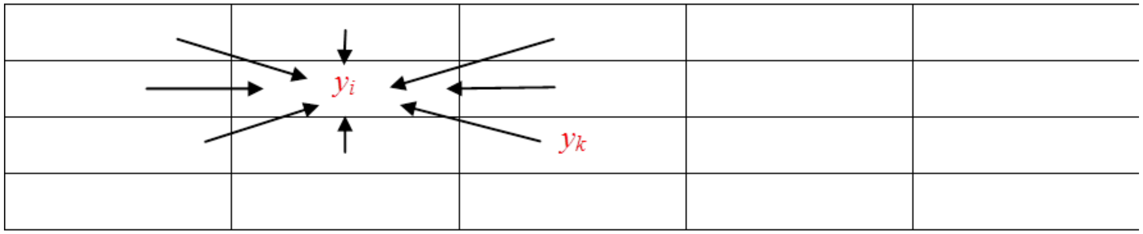
then $y = X\beta + \varepsilon$

and $\varepsilon = \lambda W\varepsilon + u$

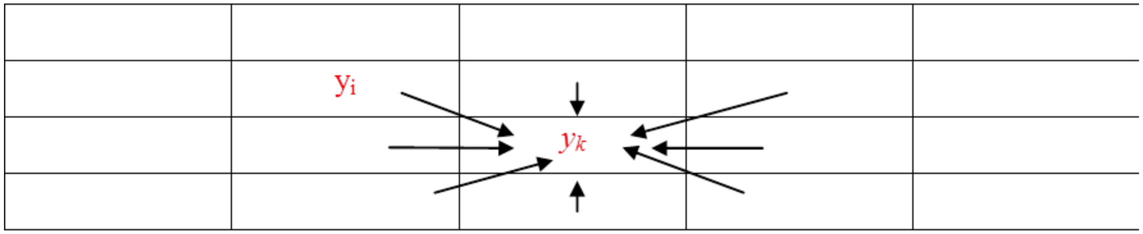
Endogeneity of the spatial lag

$$y = \lambda Wy + X\beta + \varepsilon$$

y_i depends on y_k which is part of Wy



y_k depends on y_i so is a cause of y_i via Wy and a response to y_i hence also ε



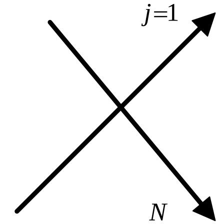
Endogeneity of the spatial lag

$$y = \lambda W y + X\beta + \varepsilon$$

- y_i depends on all other y s, including y_k because they are within Wy.
- But y_k also depends on all other y s, including y_i because they are within Wy.
- So Wy determines y_i and is determined by it.

- So we have to use the appropriate likelihood function or 2sls to obtain consistent estimates.

$$y_i = f(\sum_{j=1}^{N} W_{ij} y_j)$$



$$y_k = f(\sum_{j=1}^{N} W_{kj} y_j)$$

$$y_i = \beta_0 + \beta_1 x_{i1} + \varepsilon_i$$

to estimate by OLS it is convenient to multiply through by *x* and sum over all *i*

$$\sum_i x_{i1} y_i = \beta_1 \sum_i x_{i1}^2 + \sum_i x_{i1} \varepsilon_i$$

Rearranging gives

$$\beta_1 = \frac{\sum_i x_{i1} y_i}{\sum_i x_{i1}^2} - \frac{\sum_i x_{i1} \varepsilon_i}{\sum_i x_{i1}^2}$$

$$\beta_1 = \frac{\sum_i x_{i1} y_i}{\sum_i x_{i1}^2} - \frac{\sum_i x_{i1} \varepsilon_i}{\sum_i x_{i1}^2}$$

OLS estimator is $\hat{\beta}_1 = \frac{\sum_i x_{i1} y_i}{\sum_i x_{i1}^2}$

hence $\beta_1 = \hat{\beta}_1 - \frac{\sum_i x_{i1} \varepsilon_i}{\sum_i x_{i1}^2}$ and rearranging and taking expectations

gives $E(\hat{\beta}_1) = \beta_1 + E\left(\frac{\sum_i x_{i1} \varepsilon_i}{\sum_i x_{i1}^2} \right)$

only if $E\left(\frac{\sum_i x_{i1} \varepsilon_i}{\sum_i x_{i1}^2} \right) = 0$ is $E(\hat{\beta}_1) = \beta$

So far we have been talking about bias, but it might be argued that this will disappear as the sample size gets large.
In other words an estimator can be biased but consistent.

In fact for our OLS estimator this is not the case. Summing over a larger number of cases increases both the numerator and the denominator in

$$\frac{\sum_i x_{i1}\varepsilon_i}{\sum_i x_{i1}^2}$$

So the bias cannot be removed by simply increasing the size of the sample.
In order for an estimator to be consistent, it should tend towards unbiasedness as the sample size goes to infinity.
Using the OLS estimator this will not happen so the OLS estimator is both biased and inconsistent.

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 .... + \hat{\beta}_{k-1} X_{k-1}$$

$\hat{\beta}$ is an unbiased estimator of β if $\mathrm{E}(\hat{\beta}) = \beta$

$\hat{\beta}$ is a consistent estimator of β if $\hat{\beta} \xrightarrow{p} \beta$

this means that as the sample size n increases then

the probability approaches 1 that $\hat{\beta}$ lies

within the range $\beta + c$ to $\beta - c$

where c is a small constant > 0

the small p stands for 'converges in probability' to β

as n goes to infinity

Consistent OLS estimation requires

$\text{plim } n^{-1}(Wy'Wy) = \tilde{Q}$ a finite nonsingular matrix

$\text{plim } n^{-1}(Wy'\varepsilon) = 0$

The first constraint can be satisfied with proper constraints on ρ and W (more about this later)

BUT

the second condition does not hold for spatial data

Instead

$$\text{plim } n^{-1}(Wy'\varepsilon) = \text{plim } n^{-1}\varepsilon' W(I - \rho W)^{-1}\varepsilon$$

The existence of *W* leads to a quadratic form, so that the plim does not equal zero (unless $\rho = 0$)

Inconsistency : simulation

assume that we have 81 regions forming a 9 by 9 lattice (like a chessboard). Assume that with $W_{ij} = 1$ if locations i and j are contiguous, and $W_{ij} = 0$ otherwise

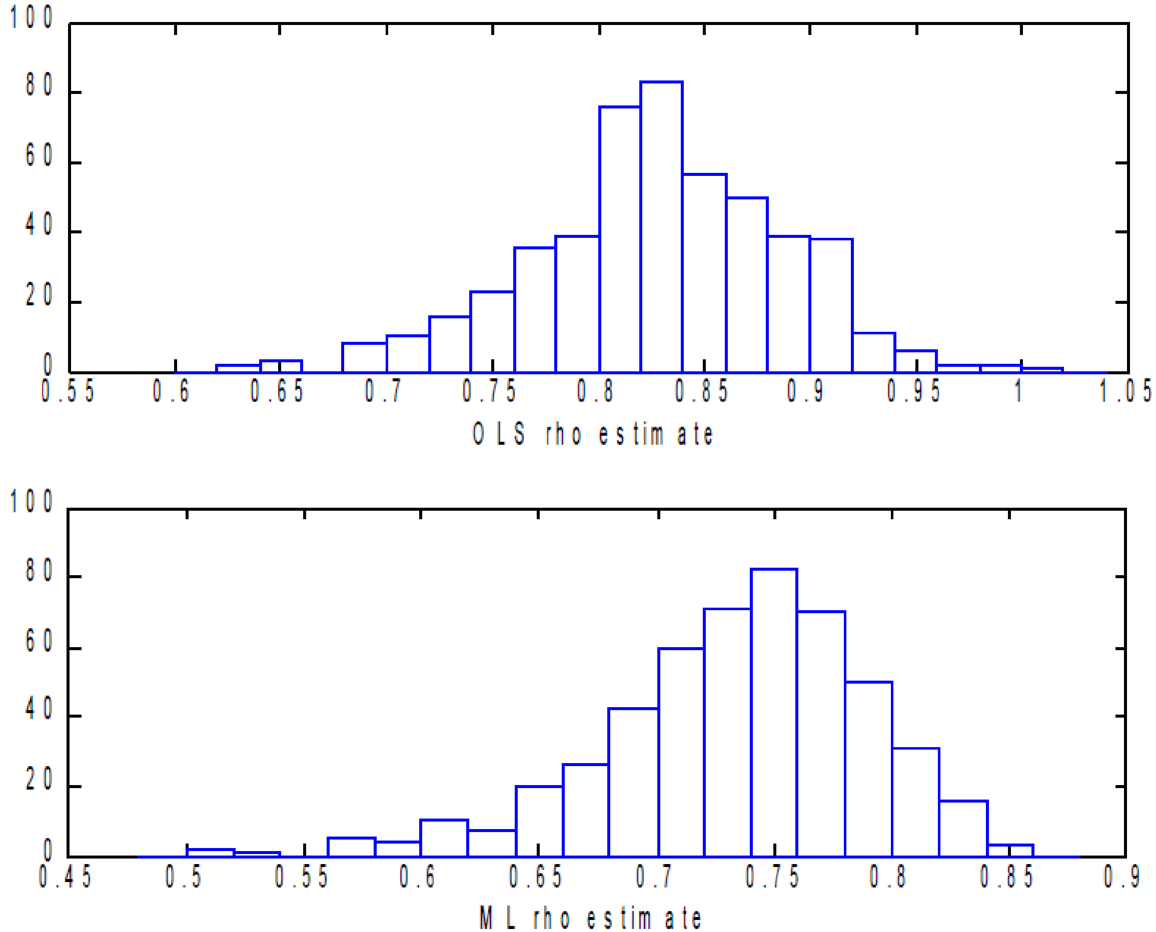
regions are contiguous if they share an edge

W is standardised so that all the values in row i are divided by the sum of the row i values

$$y = \rho Wy + X\beta + \varepsilon$$

y = the dependent variable, an N=81 x 1 vector
Wy = the spatial lag, an N=81 x 1 vector
X = an N x K=3 matrix of regressors, with the first column equal to the constant, other two columns sampled from a uniform distribution
β = a K x 1 vector of regression coefficients, values 1,4,5
ρ = the spatial lag coefficient, value 0.75
ε = an N x1 vector of errors, drawn from an N(0,1) distribution

Outcome of 500 Monte Carlo simulations, in each case drawing
The errors from an N(0,1) distribution to obtain *y*, and then regressing *y*
On *Wy* and *X*



Spatial error model

$$y = X\beta + \varepsilon$$

$$\varepsilon = \lambda W \varepsilon + u$$

$$E(\varepsilon\varepsilon') = \sigma^2 (I - \lambda W)^{-1} (I - \lambda W')^{-1}$$

$$u \sim iid(0, \sigma^2 I)$$

$$\varepsilon = (I - \lambda W)^{-1} u$$

$$y = X\beta + (I - \lambda W)^{-1} u$$

Spatial error model

$$\hat{\beta}_1 \sim N(0, \hat{\sigma}^2 Sxx^{-1})$$

$$Sxx = \sum_i X_i^2 - (\sum_i X_i)^2 / n$$

$$OLS \text{ s.e. } \beta_1 = \sqrt{\sigma^2 Sxx^{-1}}$$

$$\text{true s.e. } \beta_1 = \sqrt{\sigma^2 Sxx^{-1} X'[(I - \lambda W)'(I - \lambda W)]^{-1} X Sxx^{-1}}$$

Typically the true standard error will be greater than the OLS s.e.
and therefore using the OLS s.e. we will often reject the null hypothesis
that $\beta_1 = 0$ when we should not reject.