Regression Modeling of Spatial Relationships

ECON6027 Spatial Econometrics & Data Analysis, 7a, theory

Introduction to Spatial Econometrics

Outline

Outline

The traditional linear regression model

The traditional linear regression model can be given as,

$$Y_n = X_n\beta + \epsilon_n \tag{1}$$

where, Y_n is a vector of n observations* of the dependent variable, X_n is a matrix of n observations on $k-1$ non **stochastic, exogenous regressors** and a constant term, β is a vector of k unknown parameters to be estimates and ϵ_n is a vector of stochastic disturbances. $\theta = (\beta', \sigma_\epsilon^2)$ is the vector of parameters to be estimated.
*Assume the n observations refer to *areal* units to suit our purposes.

Classic linear regression assumptions

Normality, identically and independence of stochastic disturbances conditional upon the k regressors:

$$\epsilon_n | X_n \sim i.i.d.N(0, \sigma_\epsilon^2 I_n)$$

This means,

- exogeneity: $E(\epsilon_n | X_n) = 0$ (or $E(\epsilon_n' X_n) = 0$ if X_n is non-stochastic)
- spherical disturbances (homoskedasticity): $E(\epsilon\epsilon' | X_n) = \sigma_\epsilon^2 I_n$

Further it is also assumed that,

- Regressors are independent.

Outline

Traditional estimates

The three main estimates for the parameters of the traditional model are

1. Ordinary least squares estimate
2. Maximum likelihood estimate
3. Method of moments estimate

1. OLS estimate and it's properties

The OLS estimate for β in model (1) is,

$$\hat{\beta}_{OLS} = \left(X_n' X_n\right)^{-1} X_n' Y_n$$

Under the classic linear regression assumptions (also known as the Gauss-Markov conditions), the $\hat{\beta}_{OLS}$ is

- the *best linear unbiased* estimator (BLUE).
  - unbiased: $E(\hat{\beta}_{OLS}|X_n) = \beta$
  - best: $Var(\hat{\beta}_{OLS}|X_n) = \left(X_n' X_n\right)^{-1} \sigma_\epsilon^2$ is the minimum variance amongst all possible linear estimators (full efficiency) and tends to 0 when $n \to \infty$ (weak consistency).
- asymptotically normally distributed,

$$\hat{\beta}_{OLS}|X_n \xrightarrow{d} N\left[\beta, \sigma_\epsilon^2 \left(X_n' X_n\right)^{-1}\right]$$

2. The likelihood function for disturbances

Given $\epsilon_n = Y_n - X_n\beta$, and under the assumption of $\epsilon_n | X_n \sim i.i.d.N(0, \sigma_\epsilon^2 I_n)$ a single stochastic disturbance has the pdf,

$$f(\epsilon_i|\theta) = \frac{1}{\sqrt{2\pi\sigma_\epsilon^2}} \exp\left\{-\frac{\epsilon_i^2}{2\sigma_\epsilon^2}\right\}$$

Under the assumption of independence, the n areal units have a joint pdf of,

$$f(\epsilon_{1,\ldots,n}|\theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_\epsilon^2}} \exp\left\{-\frac{\epsilon_i^2}{2\sigma_\epsilon^2}\right\}$$

The *likelihood function* has the same form except that it is a function of the parameters $\theta = (\beta', \sigma_\epsilon^2)$,

$$L(\theta|\epsilon_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_\epsilon^2}} \exp\left\{-\frac{\epsilon_i^2}{2\sigma_\epsilon^2}\right\}$$

The *log*-likelihood function for disturbances

The log-likelihood function, $\ln(L(\theta))$ of the parameters can be written as,

$$
\begin{aligned}
\ell(\theta) &= \ln\left(\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_\epsilon^2}} \exp\left\{-\frac{\epsilon_i^2}{2\sigma_\epsilon^2}\right\}\right) \\
&= \sum_{i=1}^{n} \ln\left(\frac{1}{\sqrt{2\pi\sigma_\epsilon^2}} \exp\left\{-\frac{\epsilon_i^2}{2\sigma_\epsilon^2}\right\}\right) \\
&= \sum_{i=1}^{n} \ln\left(\frac{1}{\sqrt{2\pi\sigma_\epsilon^2}} \exp\left\{-\frac{(Y_i - X_i\beta)^2}{2\sigma_\epsilon^2}\right\} \cdot 1\right) \\
&= -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma_\epsilon^2 - \frac{1}{2\sigma_\epsilon^2}\left(Y_n - X_n\beta\right)'\left(Y_n - X_n\beta\right)
\end{aligned}
$$

Note, $\sum_{i=1}^{n} \epsilon_i^2 = (Y_n - X_n\beta)'(Y_n - X_n\beta)$. The transformation from ϵ_i to Y_i is via $\epsilon_i = Y_i - X_i\beta$, thus, the Jacobian of the transformation for each observation is, $|\partial\epsilon_i/\partial Y_i| = 1$

First order conditions

Taking derivatives of $\ell(\theta)$ with respect to θ gives us the first order conditions (FOCs) that can be solved to give the maximum likelihood estimators. These derivatives are also known as the *score* function,

$$S(\theta) = \begin{cases} -\frac{1}{\sigma_\epsilon^2} X_n' \left(Y_n - X_n \beta \right) \\ -\frac{n}{2\sigma_\epsilon^2} + \frac{1}{2\sigma_\epsilon^4} \left(Y_n - X_n \beta \right)' \left(Y_n - X_n \beta \right) \end{cases}$$

Maximum likelihood estimates

Solving the FOCs, i.e., setting $S(\theta) = 0$, we get the maximum likelihood estimators of θ.

$$\begin{aligned} \hat{\beta}_{MLE} &= \left(X_n' X_n\right)^{-1} X_n' Y_n \\ \hat{\sigma}^2_{\epsilon, MLE} &= \frac{\epsilon_n' \epsilon_n}{n} \end{aligned}$$

- If the assumption of normality of the disturbances does not hold then the MLEs become *quasi* maximum likelihood estimates (QMLEs).
- Under the classic Gauss-Markov assumptions, OLS and (Q)ML estimates for β coincide and thus enjoy the same asymptotic properties: asymptotic normality, consistency, asymptotic unbiasedness, efficiency and scale invariance
- The estimator $\hat{\sigma}^2_{\epsilon, MLE}$ is not unbiased in finite samples as a the unbiased estimator is $\hat{\sigma}^2_{\epsilon,} = \frac{\epsilon_n' \epsilon_n}{n-k}$. However, $\hat{\sigma}^2_{\epsilon, MLE}$ is asymptotically unbiased.

3. Method of moments estimator

Following exogeneity assumption we have the theoretical moment condition $E(X_n'\epsilon) = 0$. The empirical version gives,

$$\frac{1}{n}X_n'\epsilon_n = 0$$

Substituting model in (1) to ϵ_n, gives the MOM estimator:

$$\frac{1}{n}X_n'(Y_n - X_n\beta) = 0 \Longrightarrow$$

$$\hat{\beta}_{MOM} = (X_n'X_n)^{-1} X_n'Y_n = \hat{\beta}_{MLE} = \hat{\beta}_{OLS}$$

Under the classic assumptions, OLS and MOM estimators for β coincide and thus enjoy the same asymptotic properties.

Outline

Goodness of fit

The traditional goodness of fit statistics that can be used to measure the degree of fit for a linear regression model are:

- coefficient of determination: $R^2 = \frac{SSR}{SST}$
- adjusted coefficient of determination: $\bar{R}^2 = 1 - \frac{n-1}{n-k+1}(1 - R^2)$
- Akaike information criterion: $AIC = \ln\left(\frac{\epsilon'\epsilon}{n}\right) + \frac{2k}{n}$
- Schwartz (or Bayesian) information criterion: $BIC = \ln\left(\frac{\epsilon'\epsilon}{n}\right) + \frac{k \ln n}{n}$

Hypothesis test for parameter significance

$$
\begin{aligned}
H_0: &\quad \beta_2 = \ldots = \beta_k = 0 \\
H_1: &\quad \beta_i \neq 0
\end{aligned}
$$

Test statistics:

- F-test: $F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)} \sim F_{n-k}^{k-1}$
- Likelihood ratio test: $LR = -2[\ell(\theta_0) - \ell(\hat{\theta})] \sim \chi^2_{k-1}$
- Wald test: $W = n(\theta_0 - \hat{\theta})^2 I(\theta_0) \sim \chi^2_{k-1}$
- Lagrange multiplier (or Rao's score) test: $LM = S(\theta_0)' I(\theta_0)^{-1} S(\theta_0) \sim \chi^2_{k-1}$

Hypothesis test for parameter significance (contd.)

$S(\theta)$ is the score function representing the first order conditions of the likelihood function and $I(\theta)$ is the Fisher's Information matrix and are given by,

$$
\begin{aligned}
S(\theta) &= \begin{cases} \frac{1}{\sigma_\epsilon^2}\left(X_n' Y_n - X_n' X_n \beta\right) \\ -\frac{n}{2\sigma_\epsilon^2} + \frac{1}{2\sigma_\epsilon^4}\left(Y_n - X_n\beta\right)'\left(Y_n - X_n\beta\right) \end{cases} \\
I(\theta) &= \begin{pmatrix} \frac{1}{\sigma_\epsilon^2} X_n' X_n & 0 \\ 0 & \frac{n}{2\sigma_\epsilon^4} \end{pmatrix}
\end{aligned}
$$

- The LM test in particular has the advantage of only requiring the estimation of the null model.
- In all these test statistics, the score function and the information matrix can be estimated using the plug-in estimate for σ_ϵ^2 as, $\hat{\sigma}_\epsilon^2 = \frac{e'e}{n}$ where e is the vector of the residual under the null model.

Outline

1. Hypothesis test for normality of disturbances

Jarque-Bera test for the normality of disturbances jointly tests the third and fourth moments of the empirical distribution of the residuals are not significantly different to that of a Gaussian distribution

$$JB = \frac{n}{6}\left[\mathcal{S}^2 + \frac{(\kappa+3)^2}{4}\right] \sim \chi^2_2$$

- where $\mathcal{S} = \frac{\frac{1}{n}\sum(e_i - \bar{e})^3}{\left[\frac{1}{n}\sum(e_i-\bar{e})^2\right]^{3/2}}$ is the measure of skewness,
- and $\kappa = \frac{\frac{1}{n}\sum(e_i - \bar{e})^4}{\left[\frac{1}{n}\sum(e_i-\bar{e})^2\right]^{2}}$ is the measure of kurtosis of the residuals. (Do note $\bar{e} = 0$ for OLS).

2. Non-sphericity of disturbances

The fundamental assumption under which all previous results are derived is the *sphericity of errors.*

When disturbances are *non-spherical*, the $E(\epsilon_n \epsilon_n' | X)$ is no longer equal to $\sigma_\epsilon^2 I_n$, but instead is,

$$E(\epsilon_n \epsilon_n' | X) = \sigma_\epsilon^2 \Omega_n$$

When this is the case, OLS estimates are unbiased but no longer efficient as,

$$Var(\hat{\beta}_{OLS} | X_n) = \sigma_\epsilon^2 \left(X_n' X_n \right)^{-1} X_n' \Omega_n X_n \left(X_n' X_n \right)^{-1} > \sigma_\epsilon^2 \left(X_n' X_n \right)^{-1}$$

Two potential sources of non-spericity are:

- Heteroskedasticity: Diagonal elements of Ω_n are non-constant
- Autocorrelation: Off-diagonal elements of Ω_n are non-zero.

Hypothesis test for sphericity of disturbances

1. Breusch-Pagan test for homoskedasticity:

$$BP = \frac{1}{2} g_n' X_n \left(X_n' X_n \right)^{-1} X_n' g_n \sim \chi^2_{k-1}$$

   where, g_n is the vector of transformed disturbances $g_n = \frac{e_i^2}{\frac{e_n' e_n}{n} - 1}$.

2. White test for homoskedasticity: Uses a consistent estimator for $Var(\hat{\beta}_{OLS}|X_n)$ and give the test statistic as

$$WH = nR^2 \sim \chi^2_{k-1}$$

   However, do note that WH test (also PB test) requires the independence of disturbances (no autocorrelation) and hence is generally not applicable for spatial data if spatial data exhibits *spatial* autocorrelation.

3. Durbin-Watson test for no autocorrelation for time series data.
4. Moran's I (Geary's C and Getis-Ord's G) test for no autocorrelation in spatial data. In fact, one can show that the Durbin-Watson test for autocorrelation is a *special case* of the Moran's test for a specific

GLS estimate for a linear regression model with non-spherical disturbances

- Decomposing Ω_n as $\Omega_n = P_n P_n'$ we transform the original model in (1) as;

$$Y_n^* = X_n^* \beta + \epsilon_n^* \tag{2}$$

where $Y_n^* = P_n^{-1} Y_n$, $X_n^* = P_n^{-1} X_n$ and $\epsilon_n^* = P_n^{-1} \epsilon_n$.

- Note that, $E(\epsilon_n^* \epsilon_n^{*\prime}) = \sigma_\epsilon^2 I_n$. The OLS estimate of (2) is called the generalised least squares (GLS) estimator, and is given by,

$$\hat{\beta}_{GLS} = \left(X_n' \Omega_n^{-1} X_n \right)^{-1} X_n' \Omega_n^{-1} Y_n$$

where $E(\hat{\beta}_{GLS}) = \beta$ and $Var(\hat{\beta}_{GLS}) = \sigma_\epsilon^2 \left(X_n' \Omega_n^{-1} X_n \right)^{-1}$.

Exercise A

1. Show that the OLS estimate of (2) is equivalent to $\hat{\beta}_{GLS}$.
2. Show that the GLS estimator can be shown to be a maximum likelihood estimator as well (show this! Hint: define the likelihood on ϵ_n^*).
3. Give an unbiased estimator for σ_ϵ^2.

3. Endogeneity of regressors

- Endogeneity refers to the violation of the condition $E(\epsilon X_n) \neq 0$.
- When the regressors are endogenous, the OLS estimators in general will be biased and inconsistent.
- The usual fix is to use an instrumental variables and apply two stage least squares (2SLS) estimation.
- Instruments H_n should be such that,
  - it is uncorrelated with errors (**exogenous**): $E(\epsilon H_n) = 0$ and
  - it is correlated with regressors (**relevant**): $X_n = H_n\gamma + \eta_n$.

Two Stage Least Squares Estimation (2SLS)

- Stage 1:

$$\begin{aligned}
\hat{\gamma}_{OLS} &= \left(H_n' H_n\right)^{-1} H_n' X_n \\
\hat{X} &= H_n \hat{\gamma} = H_n \left(H_n' H_n\right)^{-1} H_n' X_n
\end{aligned}$$

- Stage 2:

$$\begin{aligned}
Y_n &= \hat{X}_n \beta + \epsilon \\
\hat{\beta}_{2SLS} &= \left(\hat{X}_n' \hat{X}_n\right)^{-1} \hat{X}_n' Y_n \\
&= \left(X_n' H_n \left(H_n' H_n\right)^{-1} H_n' X_n\right)^{-1} X_n' H_n \left(H_n' H_n\right)^{-1} H_n' Y_n
\end{aligned}$$

Properties of 2SLS

$$
\begin{aligned}
E(\hat{\beta}_{2SLS}) &= \beta \\
Var(\hat{\beta}_{2SLS}) &= \sigma_{\epsilon}^{2}\left(X_n' H_n \left(H_n' H_n\right)^{-1} H_n' X_n\right)^{-1}
\end{aligned}
$$

Exercise B (regression revision)

Consider the data given in uk.xls.

1. Estimate the model that explains the GVA as a function of both labour productivity and business birth rate (model 1).
2. Estimate two (nested) models that explain,
   1. GVA as a function of labour productivity (model 2).
   2. GVA as a function of business birth rate (model 3).
   3. compare models 1, 2 and 3 and explain which one would you pick as the most preferred model. Justify your answer.
3. Regress labour productivity on business birth rate (model 4).
4. Compute the residuals of model 1 and
   1. test for normality
   2. test for homoskedasticity

Outline

Spatial autocorrelation

- We have spatial autocorrelation when the non-diagonal elements of the VC matrix, Ω_n are non-zero due to
  - positive spatial autocorrelation: units that are close to one another are more similar than units that are far apart.
  - spatial heterogeneity: when some areas present more variability than others.
- The impact of neighbours are measured using spatially lagged variables.

Outline

Spatially lagged variable

- For given attribute Y_n the spatially lagged value of this attribute can be given as,

$$L(Y_n) = W_n Y_n$$

- $L(Y_n))$ represent the average of variable Y_n observed in all the locations that it is a neighbour to, according to the criterion defined by the spatial weights matrix W_n.[1]

[1] See Chapter 6 notes for more on W_n

Example

Table: Regions

A	B
D	C

Row normalised weights matrix based on rook contiguity criterion.

$$W_n = \begin{pmatrix} 0 & 0.5 & 0 & 0.5 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.5 & 0 & 0.5 \\ 0.5 & 0 & 0.5 & 0 \end{pmatrix}$$

If $Y_n = (1, 2, 3, 5)'$, the spatially lagged variable is:

$$W_n Y_n = \begin{pmatrix} 3.5 \\ 2 \\ 3.5 \\ 2 \end{pmatrix}$$

Exercise C

Consider the labour productivity variable given in uk.xls, (call it Y_n), use a contiguity based weights matrix to obtain the spatially lagged variable.

Outline

Spatial autocorrelation coefficients

1. Join counts
2. **Moran's I** (the standard test that is almost exclusively used in spatial econometrics)
3. Gear's C
4. Getis-Ord G

   All theories and tests discussed in 8c are applicable here in measuring spatial autocorrelation in (i) attributes and/or (ii) regression residuals of areal units.

Autocorrelated regression disturbances

Moran's I statistic applied to measures the correlation between *regression disturbances and it's lagged* values is given by,

$$Corr(\epsilon_n, L\epsilon_n) = \frac{Cov(\epsilon_n, L\epsilon_n)}{\sqrt{Var(\epsilon_n), Var(L\epsilon_n)}}$$

Given that $Var(.)$ is translation invariant, for a **stationary spatial process**, we have, $Var(\epsilon_n) = Var(L\epsilon_n)^2$ and thus,

$$Corr(\epsilon_n, L\epsilon_n) = \frac{Cov(\epsilon_n, L\epsilon_n)}{Var(\epsilon_n)} = \frac{\epsilon_n' W_n \epsilon_n}{\epsilon_n' \epsilon_n} \tag{3}$$

The empirical counterpart of this formula is what we know as the Moran's I statistic.

[2] In stationary time series we know this property as $Var(\epsilon_t) = Var(\epsilon_{t-1})$

Global Moran's I statistic revisited

- The Moran's I statistic considers the biased estimator for the variance in the denominator and a normalising factor equal to the sum of weights. As such the empirical counterpart of (3) can be written as,

$$I = \frac{n}{\sum_i \sum_j w_{ij}} \frac{e_n' W_n e_n}{e_n' e_n}$$

  where e_n is a set of regression residuals. (Note $\bar{e}_n = 0$).

- For a row-normalised W_n matrix, $\sum_i \sum_j w_{ij} = n$ and the Moran's I statistic simplifies to

$$I = \frac{e_n' W_n e_n}{e_n' e_n}$$

- The popular Durbin-Watson test statistic used in time series can shown to be a *special case* of the Moran's I statistic. (Why do you think this is the case?)

Global Moran's I statistic revisited (contd.)

Asymptotic distribution under normality of the residuals:

$$\frac{I - E(I)}{Var(I)} \xrightarrow{D} N(0,1)$$

$$E(I) = \frac{n}{\sum_i \sum_j w_{ij}} \frac{tr(M_X W_n)}{n-k}$$

$$Var(I) = \frac{n^2}{\left(\sum_i \sum_j w_{ij}\right)^2} \frac{tr(M_x W_n M_x W_n') + tr[(M_x W_n)^2] + tr^2(M_x W_n)}{(n-k)(n-k+2)} - E^2(I)$$

where, $M_x = I_n - P_x$ and $P_x = X_n \left(X_n' X_n\right)^{-1} X_n'$

Exercise D

Consider the data given in uk.xls.

1. Compute the lagged values of the variable "gross value added (GVA)" using the weights matrix used in Exercise B.
2. Plot the scatter diagram of GVA against L(GVA). What is this plot known as? What can you infer from the scatterplot?
3. Test for the presence of spatial autocorrelation in the residuals of model 1.

Summary

- A revision of the classic linear regression model
  - model assumptions
  - estimation (OLS, MLE, MOM)
  - goodness of fit
  - violation of regression assumptions
    - non-normality of disturbances
    - non-sphericity of disturbances and GLS
    - endogeneity of regressors and 2SLS
- Spatially autocorrelated areal units
  - lagged variables
  - testing for autocorrelation in regression residuals

References I

G. Arbia.
A primer for spatial econometrics.
Palgrave Macmillan