Introduction to ESDA and Spatial Econometric techniques using STATA

Raul Ramos (AQR-IREA, UB)

http://www.raulramos.cat/stata18

June 2018

AQR

Grup d'Anàlisi Quantitativa Regional
Universitat de Barcelona

UNIVERSITAT DE BARCELONA

Session 2

Spatial econometrics

- Spatial economics is now widely recognised in the economics and econometrics mainstream
- Much insight can be gained by using spatial econometric tools in addition to more standard time series methods
- Time series and spatial econometrics come together in the analysis of spatial panels

What is spatial econometrics?

Time series vs spatial analysis

$$Y_t = \beta_1 + \beta_2 \cdot X_{2t} + u_t \qquad t = 1, 2, \dots, T$$

$$Y_i = \beta_1 + \beta_2 \cdot X_{2i} + u_i \qquad i = 1, 2, \dots, N$$

But extended to include different spatial interaction effects:

- Endogenous interaction effects (spatial lags of Y)
- Exogenous interaction effects (spatial lags of X)
- Spatial autocorrelation (spatial dependence in u)

Spatial interaction effects – W spatial matrix

- Endogenous interaction effects ρWY

  Dependent variable of region i $\leftrightarrow$ Dependent variable of region j

- Exogenous interaction effects $WX\gamma$

  Independent variables of region j $\rightarrow$ Dependent variable of region i

- Interaction effects among error terms θWu

  Error term of region i $\leftrightarrow$ Error term of region j

Linear spatial econometric model - cross-section data

General nesting spatial model

$$Y = \rho WY + X\beta + WX\gamma + u \qquad u = \theta \mathrm{W}u + \varepsilon$$

Spatial Autoregressive Combined Model (SAC) – WY, Wu

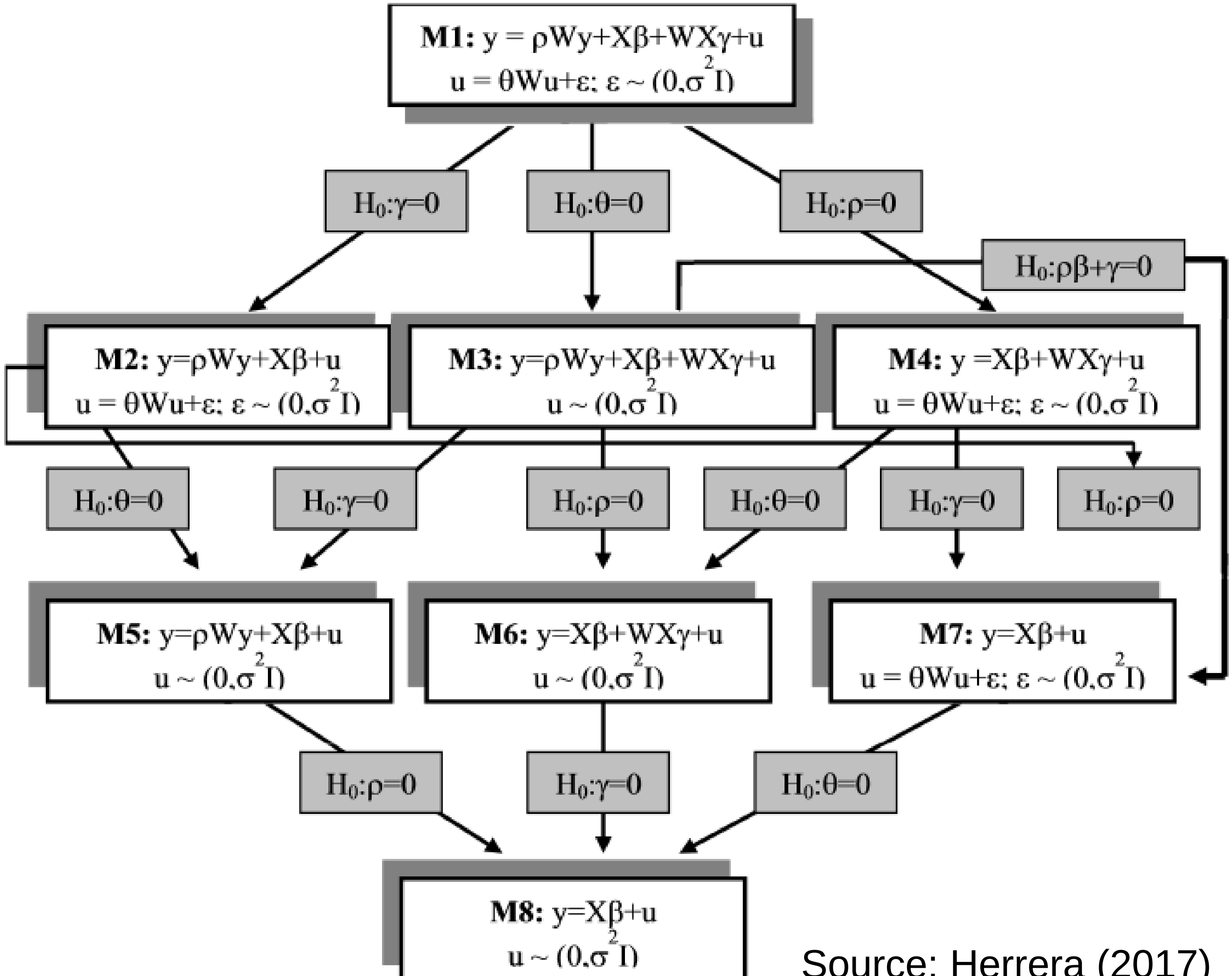
Spatial Durbin Model (SDM) – WY, WX

Spatial Durbin Error Model (SDEM) – WX, Wu

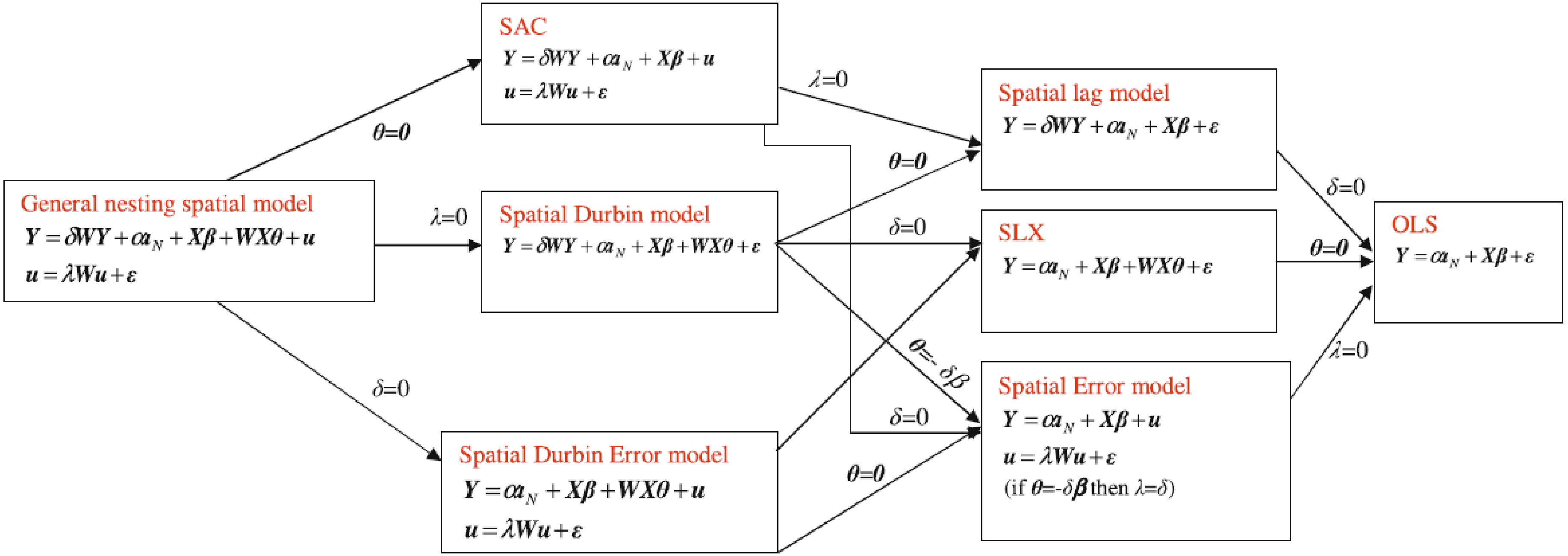
Spatial Autoregressive Model (SAR) – WY

Spatial Lag of X Model (SLX) – WX

Spatial Error Model (SEM) – Wu



Source: Herrera (2017)



Source: Elhorst (2017)

Two basic identification problems:

- How to find the right/best spatial econometric model specification (and estimation method – OLS/ML/GMM)?

  General to specific approach (LM tests), but do spillovers make sense from a theoretical perspective?

- How to find the right/best specification of the spatial weights matrix?

  W must be exogenous – Robustness checks to different specifications

Spatial econometrics

- Modelling spatial dependence with cross-sectional data spatial autoregressive model, spatial error model, spatial durbin model
- Spatial spillover effects (direct and indirect effects, global and local spillovers)

- Spatial panel models
- Fixed effects/Random effects
- Non-dynamic vs dynamic models

STATA 15: https://www.stata.com/new-in-stata/spatial-autoregressive-models/

Spatial econometrics

Spatial autoregressive model (SAR)

Spatial error model (SEM)

Different commands for cross-sectional data:

spatreg	lag / error	ML	W
spreg	lag & error	ML + GS2SLS	W/M
spivreg	*spreg* + endogenous regressor		

spregcs

Detecting spatial dependence

```
. reg CRIME INC HOVAL

      Source |       SS           df       MS      Number of obs   =        49
-------------+----------------------------------   F(2, 46)        =     28.39
       Model |  7423.32674         2  3711.66337   Prob > F        =    0.0000
    Residual |  6014.89274        46  130.758538   R-squared       =    0.5524
-------------+----------------------------------   Adj R-squared   =    0.5329
       Total |  13438.2195        48  279.962906   Root MSE        =    11.435

------------------------------------------------------------------------------
       CRIME |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         INC |  -1.597311   .3341308    -4.78   0.000    -2.269881   -.9247405
       HOVAL |  -.2739315   .1031987    -2.65   0.011    -.4816597   -.0662033
       _cons |   68.61896   4.735486    14.49   0.000     59.08692      78.151
------------------------------------------------------------------------------
```

spatdiag →

```
Diagnostic tests for spatial dependence in OLS regression

Fitted model
---------------------------------------------------------------
CRIME = INC + HOVAL
---------------------------------------------------------------

Weights matrix
---------------------------------------------------------------
Name: w1s
Type: Imported (binary)
Row-standardized: Yes
---------------------------------------------------------------

Diagnostics
---------------------------------------------------------------
Test                           |  Statistic    df   p-value
-------------------------------+-------------------------------
Spatial error:                 |
  Moran's I                    |     2.840      1    0.005
  Lagrange multiplier          |     5.206      1    0.023
  Robust Lagrange multiplier   |     0.044      1    0.834
                               |
Spatial lag:                   |
  Lagrange multiplier          |     8.898      1    0.003
  Robust Lagrange multiplier   |     3.736      1    0.053
---------------------------------------------------------------
```

AQR

```
. spreg ml CRIME INC HOVAL, id(id) dlmat(w1s) elmat(w1s)

Performing a grid search.... finished

Iteration 0:   log likelihood = -182.57222
Iteration 1:   log likelihood = -182.55505
Iteration 2:   log likelihood = -182.55502
Iteration 3:   log likelihood = -182.55502

Optimizing unconcentrated log likelihood

Iteration 0:   log likelihood = -182.55502

Spatial autoregressive model                      Number of obs   =        49
(Maximum likelihood estimates)                    Wald chi2(2)    =   33.0764
                                                  Prob > chi2     =    0.0000

------------------------------------------------------------------------------
       CRIME |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
CRIME        |
         INC |  -1.042749   .3357464    -3.11   0.002      -1.7008   -.3846985
       HOVAL |  -.2798409   .0941962    -2.97   0.003     -.464462   -.0952197
       _cons |   47.91536   9.340775     5.13   0.000     29.60778    66.22294
-------------+----------------------------------------------------------------
lambda       |
       _cons |   .3693742   .1807057     2.04   0.041     .0151976    .7235509
-------------+----------------------------------------------------------------
rho          |
       _cons |    .146417   .3024333     0.48   0.628    -.4463414    .7391754
-------------+----------------------------------------------------------------
sigma2       |
       _cons |   97.04344    19.7707     4.91   0.000     58.29359    135.7933
------------------------------------------------------------------------------
```

2.7 Direct and Indirect (or Spillover) Effects

AQR

Many empirical studies use the point estimates of one or more spatial regression model specifications (δ, θ and/or λ) to draw conclusions as to whether or not spatial spillovers exist. One of the key contributions of LeSage and Pace's book (2009, p. 74) is the observation that this may lead to erroneous conclusions, and that a partial derivative interpretation of the impact from changes to the variables of different model specifications represents a more valid basis for testing this hypothesis. To illustrate this, they give an example of a spatially lagged independent variable WX whose coefficient is negative and insignificant (ibid, Table 3.3), while it's spatial spillover effect is positive and significant (ibid, Table 3.4). The explanation for this can be seen by the derivation below.

By rewriting the general nesting spatial (GNS) model in (2.5a, b) as

$$Y = (\boldsymbol{I} - \delta \boldsymbol{W})^{-1}(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{W}\boldsymbol{X}\boldsymbol{\theta}) + \boldsymbol{R} \tag{2.12}$$

where $\boldsymbol{R}$ is a rest term containing the intercept and the error terms, the matrix of partial derivatives of the expected value of $\boldsymbol{Y}$ with respect to the kth explanatory variable of X in unit 1 up to unit N in time can be seen to be

$$\begin{aligned}
\left[\frac{\partial E(\boldsymbol{Y})}{\partial x_{1k}} \quad . \quad \frac{\partial E(\boldsymbol{Y})}{\partial x_{Nk}}\right] &= \begin{bmatrix} \frac{\partial E(y_1)}{\partial x_{1k}} & . & \frac{\partial E(y_1)}{\partial x_{Nk}} \\ . & . & . \\ \frac{\partial E(y_N)}{\partial x_{1k}} & . & \frac{\partial E(y_N)}{\partial x_{Nk}} \end{bmatrix} \\
&= (\boldsymbol{I} - \delta \boldsymbol{W})^{-1} \begin{bmatrix} \beta_k & w_{12}\theta_k & . & w_{1N}\theta_k \\ w_{21}\theta_k & \beta_k & . & w_{2N}\theta_k \\ . & . & . & . \\ w_{N1}\theta_k & w_{N2}\theta_k & . & \beta_k \end{bmatrix}
\end{aligned} \tag{2.13}$$

Table 2.1 Direct and spillover effects of different model specifications

	Direct effect	Indirect effect
OLS/SEM	β_k	0
SAR/SAC	Diagonal elements of $(I-\delta W)^{-1}\beta_k$	Off-diagonal elements of $(I-\delta W)^{-1}\beta_k$
SLX/SDEM	β_k	θ_k
SDM/GNS	Diagonal elements of $(I-\delta W)^{-1}(\beta_k + W\theta_k)$	Off-diagonal elements of $(I-\delta W)^{-1}(\beta_k + W\theta_k)$

Source Halleck Vega and Elhorst (2012)

AQR

Table 2.2 Model comparison of the estimation results explaining the crime rate

	OLS	SAR	SEM	SLX	SAC	SDM	SDEM	GNS
Intercept	0.686**	0.451**	0.599**	0.750**	0.478**	0.428**	0.735**	0.509
	(14.49)	(6.28)	(11.32)	(11.32)	(4.83)	(3.38)	(8.37)	(0.75)
Income	−1.597**	−1.031**	−0.942**	−1.109**	−1.026**	−0.914**	−1.052**	−0.951**
	(−4.78)	(−3.38)	(−2.85)	(−2.97)	(−3.14)	(−2.76)	(−3.29)	(−2.16)
House value	−0.274**	−0.266**	−0.302**	−0.290**	−0.282**	−0.294**	−0.276**	−0.286**
	(−2.65)	(−3.01)	(−3.34)	(−2.86)	(−3.13)	(−3.29)	(−3.02)	(−2.87)
W * Crime rate		0.431**			0.368*	0.426**		0.315
		(3.66)			(1.87)	(2.73)		(0.33)
W * Income				−1.371**		−0.520	−1.157**	−0.693
				(−2.44)		(−0.92)	(−2.00)	(−0.41)
W * House value				0.192		0.246	0.112	0.208
				(0.96)		(1.37)	(0.56)	(0.73)
W * Error term			0.562**		0.166		0.425**	0.154
			(4.19)		(0.56)		(2.69)	(0.15)
R^2	0.552	0.652	0.651	0.609	0.651	0.665	0.663	0.651
Log-Likelihood	13.776	43.263	42.273	17.075	43.419	44.260	44.069	44.311

**Significant at 5 %; *Significant at 10 %; T-values in parentheses, W = Binary contiguity matrix

AQR

Table 2.3 Model comparison of the marginal effects of the explanatory variables on the crime rate

	OLS	SAR	SEM	SLX	SAC	SDM	SDEM	GNS
Direct effects								
Income	−1.597**	−1.086**	−0.942**	−1.109**	−1.063**	−1.024**	−1.052**	−1.032**
	(−4.78)	(−3.44)	(−2.85)	(−2.97)	(−3.25)	(−3.19)	(−3.29)	(−3.28)
House value	−0.274**	−0.280**	−0.302**	−0.290**	−0.292**	−0.279**	−0.276**	−0.277
	(−2.65)	(−2.96)	(−3.34)	(−2.86)	(−3.10)	(−3.13)	(−3.02)	(0.32)
Indirect or spatial spillover effects								
Income		−0.727*		−1.371**	−0.560	−1.477*	−1.157**	−1.369
		(−1.95)		(−2.44)	(−0.18)	(−1.83)	(−2.00)	(0.02)
House value		−0.188*		0.192	−0.154	0.195	0.112	0.163
		(−1.71)		(0.96)	(−0.39)	(0.66)	(0.56)	(−0.03)

**Significant at 5 %; *Significant at 10 %; T-values in parentheses, W = Binary contiguity matrix

Extension to panel data models

- Large N and/or large T
- Balanced vs unbalanced panels (multiple imputation)
- Fixed effects vs Random effects
- Static and Dynamic panel data models

net install xsmle.pkg

Munnell Productivity Data
48 Continental U.S. States, 17 years,1970–1986

STATE = State name,
ST ABB=State abbreviation,
YR =Year, 1970, . . . ,1986,
PCAP =Public capital,
HWY =Highway capital,
WATER =Water utility capital,
UTIL =Utility capital,
PC =Private capital,
GSP =Gross state product,
EMP =Employment
UNEMP=Unemployment

Checking for cross-sectional dependence

xtcsd, pesaran

```
. xtreg gsp pcap pc emp unemp

Random-effects GLS regression                   Number of obs     =        816
Group variable: state                           Number of groups  =         48

R-sq:                                           Obs per group:
     within  = 0.9587                                         min =         17
     between = 0.9883                                         avg =       17.0
     overall = 0.9874                                         max =         17

                                                Wald chi2(4)      =   23577.54
corr(u_i, X)   = 0 (assumed)                    Prob > chi2       =     0.0000

------------------------------------------------------------------------------
         gsp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        pcap |  -.1416344   .0379326    -3.73   0.000    -.2159808   -.0672879
          pc |   .1709655   .0171498     9.97   0.000     .1373524    .2045786
         emp |    34.1288   .7821401    43.64   0.000     32.59583    35.66177
       unemp |  -274.6708   60.48787    -4.54   0.000    -393.2248   -156.1167
       _cons |  -3200.844   1063.519    -3.01   0.003    -5285.303   -1116.385
-------------+----------------------------------------------------------------
     sigma_u |  5664.4081
     sigma_e |  2668.9728
         rho |  .81832162   (fraction of variance due to u_i)
------------------------------------------------------------------------------

. xtcsd, pesaran


Pesaran's test of cross sectional independence =    31.255, Pr = 0.0000

.
```

Checking for autocorrelation

xtserial

```
. xtserial gsp pcap pc emp unemp

Wooldridge test for autocorrelation in panel data
H0: no first-order autocorrelation
    F(  1,      47) =    344.303
           Prob > F =      0.0000
```

Spatial analysis with STATA

```
SAR with random-effects                              Number of obs =       816

Group variable: state                             Number of groups =        48
Time variable: year                                   Panel length =        17

R-sq:    within  = 0.9588
         between = 0.9870
         overall = 0.9860

Log-likelihood = -7712.2964
```

gsp	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Main						
pcap	-.1834767	.0425254	-4.31	0.000	-.266825	-.1001285
pc	.1620073	.0183108	8.85	0.000	.1261189	.1978957
emp	34.50343	.8179892	42.18	0.000	32.9002	36.10666
unemp	-263.4233	60.69184	-4.34	0.000	-382.3772	-144.4695
_cons	-3625.349	1426.626	-2.54	0.011	-6421.485	-829.2124
Spatial						
rho	.0209289	.0124308	1.68	0.092	-.0034351	.0452928
Variance						
lgt_theta	-2.427323	.1243094	-19.53	0.000	-2.670965	-2.183681
sigma_e	7054569	361942.3	19.49	0.000	6345175	7763963
Direct						
pcap	-.1841726	.0360694	-5.11	0.000	-.2548674	-.1134778
pc	.1634317	.0202483	8.07	0.000	.1237458	.2031176
emp	34.47741	.8365757	41.21	0.000	32.83775	36.11707
unemp	-264.9011	57.75121	-4.59	0.000	-378.0914	-151.7108
Indirect						
pcap	-.0040906	.0026423	-1.55	0.122	-.0092694	.0010881
pc	.0035439	.002055	1.72	0.085	-.0004839	.0075716
emp	.7542546	.4389235	1.72	0.086	-.1060198	1.614529
unemp	-5.857454	3.893037	-1.50	0.132	-13.48767	1.772758
Total						
pcap	-.1882632	.0371924	-5.06	0.000	-.261159	-.1153675
pc	.1669755	.0205064	8.14	0.000	.1267838	.2071673
emp	35.23167	.9610624	36.66	0.000	33.34802	37.11531
unemp	-270.7586	59.45698	-4.55	0.000	-387.2921	-154.225

Spatial analysis with STATA

```
. xsmle gsp pcap pc emp unemp, fe wmat(usaww) dlag(1) mod(sar)
Iteration 0:   Log-likelihood = -7034.0276
Iteration 1:   Log-likelihood = -7028.9233
Iteration 2:   Log-likelihood = -7028.8837
Iteration 3:   Log-likelihood = -7028.8837

Dynamic SAR with spatial fixed-effects              Number of obs =       768

Group variable: state                            Number of groups =        48
Time variable: year                                  Panel length =        16

R-sq:    within  = 0.9630
         between = 0.9931
         overall = 0.9920

Mean of fixed-effects = -7.8e+02

Log-likelihood = -7033.9965
------------------------------------------------------------------------------
         gsp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
Main         |
         gsp |
         L1. |   .5283284   .0372959    14.17   0.000     .4552297    .6014271
             |
        pcap |  -.3137184   .0506232    -6.20   0.000    -.4129381   -.2144987
          pc |   .0212704   .0212018     1.00   0.316    -.0202844    .0628251
         emp |     20.097   1.452237    13.84   0.000     17.25067    22.94333
       unemp |  -268.6159   54.95544    -4.89   0.000    -376.3266   -160.9052
-------------+----------------------------------------------------------------
Spatial      |
         rho |   .0621197   .0123729     5.02   0.000     .0378694    .0863701
-------------+----------------------------------------------------------------
Variance     |
    sigma2_e |    5532924   265761.9    20.82   0.000      5012040    6053808
------------------------------------------------------------------------------
```

Now try what you have learned by answering exercise 2