Implementing procedures for spatial panel econometrics in Stata

Gordon Hughes
University of Edinburgh

Andrea Piano Mortari & Federico Belotti
CEIS, Universita Roma Tor Vergata

Spanish Stata User Group Meeting
12th September 2012

Spatial analysis in Stata

- Variety of special purpose routines written by users and available through SSC
  - Manipulation of spatial data
  - Cross-section spatial regressions
- StataCorp-related routines – also through SSC
  - shp2dta converts ESRI shapefiles to dta files – similar to programs converting to csv or xls files
  - spmat, spreg, spivreg, etc for construction & manipulation of spatial weights and for cross-section spatial regressions

Nature of panel data

- Large N and/or large T?
- Balanced or unbalanced panels
- Spatial weights – interactions with missing data
- Examples:
  - State tax and fiscal policies
  - Cross-country models of economic development

Econometric specification

- Fixed or random effects – can we talk about random effects with complete sample of states or countries?
- Lagged dependent variable or within panel serial correlation
- Why are data missing – missing at random assumption

Key models

Spatial auto-regression model (SAR)

$$y_{it} = \rho W y_t + X_{it}\beta + \mu_i + \varepsilon_{it}$$

Spatial Durbin model (SDM)

$$y_{it} = \rho W y_t + X_{it}\beta + W X_t \varphi + \mu_i + \varepsilon_{it}$$

Spatial autocorrelation model (SAC)

$$y_t = \rho W y_t + X_t\beta + \mu + v_t \text{ with } v_t = \lambda M v_t + \varepsilon_t$$

Key models 2

Spatial error model (SEM)

$$y_{it} = X_{it}\beta + \mu_i + \nu_{it} \ with\ \nu_{it} = \lambda W \nu_t + \varepsilon_{it}$$

Generalised spatial random errors (GSPRE)

$$y_t = X_t\beta + \mu + \nu_t \ with\ \mu = \rho_1 W\mu + \eta \ and\ \nu_t = \rho_2 M\nu_t + \varepsilon_t$$

Procedure xsmle - syntax

xsmle varlist [if] [in] [weight], WMATrix(string) [MODel(string) FE RE EMATrix(string) DMATrix DURBin(varlist) ROBust DKRAAY(#) DLAG ERRor(#) NOConstant]

- "varlist" = depvar indvars [required].
- "wmat(WN)", “emat(WE)”, “dmat(WD)” refer to an N x N matrices of spatial weights for spatial lags, spatial errors and Durbin variables [at least one of wmat() or emat() is required].
- “model(string)” specifies the type of model to be estimated. The default is “sar” and alternatives are “sdm”, “sem”, “sac” and “gspre”.
- “fe | re" specifies that a fixed or random effects model should be used – the default varies according to the model specified.

Procedure xsmle – syntax 2

- “durbin(varlist)” specifies a set of spatially-weighted regressors.
- "robust" specifies that cluster robust standard errors should be used.
- "dkraay(#)" specifies that Driscoll-Kraay robust standard errors should be computed using a maximum lag equal to the integer contained in the brackets. If this is zero, a default value for the maximum lag equal to floor(4*((T/100)^(2/9))) will be used. If the integer is negative, the robust option is ignored.
- "dlag" includes the lagged dependent variable in the model. This is only available for model(sar) and model(sdm).
- “err(#)” specifies the error structure for the GSPRE model. The default is the most general version ($\rho 1 \neq \rho 2 \neq 0$).
- "noconstant" specifies that the model should be estimated without adding a constant term.

Illustration – US electricity demand

- State data – continental US, 1990-2010
  - Electricity demand by sector
  - Regressors - prices, weather (heating & cooling days)
- Focus on price elasticities and weather impacts
- Likely to be spatial interactions due to
  - Common factors in unobserved variables
  - Competition between states for industry and/or movement of households

Model estimation 1

```
. xsmle ln_sales_rpop ln_rinc_cap ln_gprice_res ln_hunit_pop
>       ln_degday_cool ln_degday_heat,
>       wmat(WN_rook) model(sdm) durbin(ln_gprice_res) fe robust dlag;
Iteration 0:   LL =  2183.6871
Iteration 1:   LL =  2216.2479
Iteration 2:   LL =   2220.072
Iteration 3:   LL =  2220.0925
Iteration 4:   LL =  2220.0925

FE-SDM - fixed effects + spatially lagged dependent & independent variables

Number of panel units = 49                        Number of time periods =  21

Number of observations used = 1029

Type of fixed effects: Individual

Log-likelihood =   2219.493

R-sq:    within  = 0.8929
         between = 0.9073
         overall = 0.8930
```

Model estimation 2

ln_sales_rpop	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
ln_sales_rpop						
Y[t-1]	.500087	.041702	11.99	0.000	.4162395	.5839344
W_ln_sales_rpop	.3271284	.0394606	8.29	0.000	.2477875	.4064694
ln_rinc_cap	.0446942	.0281321	1.59	0.119	-.0118692	.1012576
ln_gprice_res	-.1748491	.0159389	-10.97	0.000	-.2068964	-.1428019
ln_hunit_pop	.1726343	.0631151	2.74	0.009	.0457329	.2995357
ln_degday_cool	.0706395	.0119532	5.91	0.000	.0466059	.094673
ln_degday_heat	.1333204	.0218373	6.11	0.000	.0894136	.1772273
W_ln_gprice_res	.0968248	.0246288	3.93	0.000	.0473053	.1463443
_anc						
sigma_eps^2	.0006446	.0000516	12.50	0.000	.000541	.0007483

Mean of fixed effects = -1.8463

Error components:
sigma_eps 0.025390

Matrix of direct, indirect & total effects by independent variable

	Direct	Indirect	Total
ln_rinc_cap	0.0460	0.0204	0.0664
ln_gprice_~s	-0.1714	0.0554	-0.1160
ln_hunit_pop	0.1776	0.0790	0.2566
ln_degday_~l	0.0727	0.0323	0.1050
ln_degday_~t	0.1371	0.0610	0.1981

Fixed effects models

Table 1 – Fixed effects models for residential electricity demand

	Dependent variable – ln(residential electricity consumption per person)						
Variables	Non-spatial panel	SAR	SAR + Dlag	SDM	SDM + Dlag	SEM	SAC
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Y[t-1]			0.524***		0.500***		
			(0.042)		(0.042)		
W*Y		0.388***	0.283***	0.456***	0.327***		0.379***
		(0.056)	(0.038)	(0.050)	(0.040)		(0.107)
ln(Real personal income per person)	0.381***	0.179***	0.0326	0.198***	0.0447	0.359***	0.184**
	(0.042)	(0.046)	(0.028)	(0.044)	(0.028)	(0.061)	(0.055)
ln(Real average residential price)	-0.243***	-0.246***	-0.146***	-0.294***	-0.175***	-0.283***	-0.248***
	(0.037)	(0.034)	(0.015)	(0.035)	(0.016)	(0.040)	(0.052)
ln(Housing units per person)	1.039***	0.756***	0.199**	0.658***	0.173**	0.815***	0.757***
	(0.123)	(0.106)	(0.060)	(0.110)	(0.063)	(0.171)	(0.113)
ln(Cooling degree days)	0.0718***	0.0527***	0.0722***	0.0523***	0.0706***	0.0670***	0.0533***
	(0.013)	(0.011)	(0.013)	(0.010)	(0.012)	(0.014)	(0.012)
ln(Heating degree days)	0.189***	0.139***	0.143***	0.126***	0.133***	0.155***	0.140***
	(0.025)	(0.027)	(0.023)	(0.026)	(0.022)	(0.032)	(0.031)
W*ln(Real average residential price)				0.190***	0.0968***		
				(0.044)	(0.025)		
Lambda (spatial error)						0.420***	0.0211
						(0.093)	(0.182)

Calculating elasticities

- Direct effect (spatial Durbin model)

$$M_{dir}(k) = trace([I - \rho W]^{-1}[I_N \beta_k + W\varphi_k])(1/N)$$

  - impact of a unit change in variable X_k in state i on demand in state i averaged over all states i = 1...N

- Total effect

$$M_{tot}(k) = i_N'([I - \rho W]^{-1}[I_N \beta_k + W\varphi_k])i_N(1/N)$$

  - the impact of the same unit change in variable X_k in all states on demand in state i, again averaged over all states

Direct and total price elasticities

Table 4 - Direct and total elasticities of residential electricity demand

Variable / specification	Elasticities in fixed effects models			Elasticities in random effects models		
	Direct	Indirect	Total	Direct	Indirect	Total
	(1)	(2)	(3)	(4)	(5)	(6)
Real income per person						
Non-spatial panel	0.38		0.38	0.36		0.36
SAR	0.19	0.10	0.29	0.19	0.09	0.28
SDM	0.21	0.15	0.36	0.25	0.16	0.41
Real average price						
Non-spatial panel	-0.24		-0.22	-0.22		-0.22
SAR	-0.26	-0.14	-0.40	-0.22	-0.10	-0.32
SDM	-0.29	0.10	-0.19	-0.28	0.19	-0.09
Housing units per person						
Non-spatial panel	1.04			1.03		1.03
SAR	0.79	0.44	1.23	0.81	0.37	1.18
SDM	0.70	0.41	1.21	0.70	0.44	1.15
Heating degree-days						
Non-spatial panel	0.19			0.19		0.19
SAR	0.14	0.12	0.23	0.15	0.07	0.22
SDM	0.13	0.10	0.23	0.14	0.09	0.23
Climate average temperature						
Non-spatial panel				1.87		1.87
SAR				1.60	0.73	2.33
SDM				1.79	-0.54	1.25

Unbalanced panels - options

- Listwise deletion
  - Can mean loss of all or most of sample
- Single imputation
  - Particularly useful for spatial lags
  - See Cameron & Trivedi, Chap 27
- Multiple imputation - mi
  - May be computationally expensive
  - xsmle is set up with property(mi) and has been tested with mi
  - Care is needed in setting up the imputation

Results using mi or impute

Table 5 – SAR fixed effects model for residential electricity demand with missing values

	Full data	X variables 20% missing		X variables 50% missing		Y & X variables - 20% missing	
Imputation method		Multiple	Single	Multiple	Single	Multiple	Single
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
W*Y	0.388***	0.400***	0.392***	0.398***	0.400***	0.352***	0.333***
	(0.0560)	(0.0565)	(0.0607)	(0.0652)	(0.0685)	(0.0579)	(0.0435)
ln(Real personal income per person)	0.179***	0.184***	0.201***	0.232***	0.260***	0.225***	0.283***
	(0.0461)	(0.0461)	(0.0470)	(0.0513)	(0.0490)	(0.0501)	(0.0340)
ln(Real average residential price)	-0.246***	-0.223***	-0.202***	-0.145***	-0.126***	-0.208***	-0.136***
	(0.0341)	(0.0355)	(0.0313)	(0.0395)	(0.0207)	(0.0401)	(0.0289)
ln(Housing units per person)	0.756***	0.737***	0.699***	0.604***	0.554***	0.771***	0.613***
	(0.106)	(0.110)	(0.105)	(0.141)	(0.0934)	(0.134)	(0.116)
ln(Cooling degree days)	0.0527***	0.0527***	0.0559***	0.0564***	0.0569***	0.0409***	0.0402***
	(0.0111)	(0.0113)	(0.0105)	(0.0118)	(0.0102)	(0.0103)	(0.00988)
ln(Heating degree days)	0.139***	0.141***	0.142***	0.140***	0.135***	0.124***	0.121***
	(0.0274)	(0.0265)	(0.0265)	(0.0257)	(0.0251)	(0.0240)	(0.0245)