Doctorado en Economía Aplicada con Mención en Estudios Regionales

Examen de Calificación: Econometría/Econometría Espacial

Departamento de Economía, UCN

Marzo, 2019

Instruciones:

- Ud. tiene 15 minutos para revisar las preguntas de su examen y realizar preguntas al profesor. Luego de los 15 minutos, el profesor se retirará de la sala.
- Luego de los 15 minutos para revisar las preguntas. Ud. cuenta con 2.5 hrs (150 minutos) para responder las preguntas.
- **De las 4 preguntas, Ud. sólo debe responder 3**.
- Si responde las 4 preguntas, sólo se revisarán las 3 primeras.

1. Preguntas

1. Considere el siguiente modelo:

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \epsilon_i,$$
$$\mathbb{E}(\epsilon_i|\boldsymbol{x}_i) \neq \boldsymbol{0}$$

donde $i = 1, ..., n$, $\boldsymbol{x}_i$ es un vector de $K \times 1$. Suponga que existe un vector de variables aleatorias $\boldsymbol{z}_i$ tal que $\mathbb{E}(\epsilon_i|\boldsymbol{z}_i) = \boldsymbol{0}$, donde $\boldsymbol{z}_i$ es un vector $L \times 1$ con $L > K$.

Se le pide contestar las siguientes preguntas. Si requiere de algún supuesto extra puede incluirlo.

a) (3 pts) Demuestre que $\widehat{\boldsymbol{\beta}}_{OLS}$ de y_i sobre $\boldsymbol{x}_i$ es inconsistente.

b) Asuma que $\boldsymbol{W} = \boldsymbol{Q}_{zz}^{-1}$, donde $\boldsymbol{Q}_{zz} = n^{-1}\boldsymbol{z}_i\boldsymbol{z}_i'$. Encuentre el estimador GMM.

c) Asuma que $\mathbb{E}\|\boldsymbol{x}_i\boldsymbol{z}_i'\| < \infty$ y $\mathbb{E}\|\boldsymbol{z}_i\|^2 < \infty$. Encuentre la distribución asintótica del estimador el la pregunta b).

d) Asuma que $\text{Var}(\epsilon_i|\boldsymbol{z}_i) = \sigma^2$. Encuentre la distribución asintótica.

2. (3 pts) Considere el siguiente modelo:

$$
\begin{aligned}
y_i &= \alpha + \beta S_i + \delta A_i + \epsilon_i \\
S_i &= \gamma A_i + u_i
\end{aligned}
$$

donde y_i es el logaritmo natural del salario, $\ln(w_i)$; A es el background familiar; S es la escolaridad en años; y ϵ no está correlacionado con S y A, y u no está correlacionado con ϵ. Es decir, u y $\log(w_i)$ están correlacionado sólo a través de su dependencia conjunta de A. Sin embargo, A no es observable.

a) Suponga que usted tiene una base de datos de salarios y educación para gemélos idénticos. Suponga que el salario del gemelo i en la familia f está determinado por:

$$\log(w_{if}) = \alpha + \beta S_{if} + \delta A_{if} + \epsilon_{if},$$

donde $S_{if}(i = 1, 2)$ es la escolaridad, A_{if} es habilidad. Ahora considere el siguiente modelo en primeras diferencias:

$$
\begin{aligned}
\ln(w_{1f}) - \ln(w_{2f}) &= \beta^w(S_{1f} - S_{2f}) + \delta(A_{1f} - A_{2f}) + (\epsilon_{1f} - \epsilon_{2f}), \\
\Delta \ln w_f &= \beta^w \Delta S_f + \Delta A_f + \Delta\epsilon_f
\end{aligned}
$$

β^w es el retorno dentro de dos pares de gemelos. Asuma que A es un componente familiar no observado que representa una combinación no especificada de habilidad innata, ambiente familiar, o habilidades generales no observadas que es común para los gemelos. Es decir, $A_{1f} = A_{2f}$. Escriba el estimador OLS para β^w. ¿ Es consistente?

b) Suponga que la escolaridad tiene error de medición debido al autoreporte, tal que $\widetilde{S} = S + v$, donde Ud. observa $\widetilde{S}$ en vez de S, y v no está correlacionado con S_i y ϵ. Demuestre que el estimador de primeras diferencias es inconsistente. (Pista: Muestre que $\text{plim}\, \widehat{\beta}^w = \beta^w\left(1 - \frac{\sigma^2_{\Delta v}}{\sigma^2_{\Delta S} + \sigma^2_{\Delta v}}\right)$).

3. (3 pts) Considere el siguiente Spatial Lag Model:

$$\boldsymbol{y} = \rho\boldsymbol{W}\boldsymbol{y} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

a) Demuestre que los estimadores OLS de este modelo son inconsistentes. Indique todos los supuestos que utilizó para justificar su respuesta.

b) Demuestre que los estimadores OLS son sesgados si son estimados vía OLS.

c) Encuentre los efectos marginales de este modelo.

d) Ahora asuma que $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \boldsymbol{I}_n)$. Encuentre la función log-likelihood y los estimadores ML.

4. (3 pts) The model presented below is based on the dataset used by Harrinson and Rubinfield (1978). In their well-known paper, they were interested in the housing demand and clean air issues by computing a hedonic model with information for the Boston SMSA with 506 observations. In this exercise, the dependent variable is the median housing prices for each 506 census tracts and the explanatory variables are listed in table 1. In addition, a log transformation is applied on the dependent variable and then, this new variable and explanatory variables are standardized by subtracting the means and divided by the standards deviations. Table 2 shows the results of the estimation for this model using OLS, SAR, SEM and SAC.[1] Table 3 reports the estimates of tests for testing for spatial dependence in linear models. Finally, table 4 shows the results of Moran's I for residuals of each model estimated.

   Using this information, response the following questions:

   a) Using the Anselin's criteria, which model is the best? Use all elements that you consider important to justify the answer.

   b) Which are the problems that could arise when the previous criteria is used for choosing a spatial model? Why do we need to be careful about the conclusions that can emerge from this analysis? Use all elements that you consider important to justify the answer.

[1] The W matrix for spatial models is a Queen contiguity of order 1.

Variable	Description
MEDV	A numeric vector of median values of owner-occupied housing in USD 1000
CRIM	A numeric vector of per capita crime
ZN	A numeric vector of proportions of residential land zoned for lots over 25000 sq. ft per town (constant for all Boston tracts)
INDUS	A numeric vector of proportions of non-retail business acres per town (constant for all Boston tracts)
CHAS	A factor with levels 1 if tract borders Charles River; 0 otherwise
NOX	A numeric vector of nitric oxides concentration (parts per 10 million) per town
RM	A numeric vector of average numbers of rooms per dwelling
AGE	A numeric vector of proportions of owner-occupied units built prior to 1940
DIS	A numeric vector of weighted distances to five Boston employment centers
RAD	A numeric vector of an index of accessibility to radial highways per town (constant for all Boston tracts)
TAX	A numeric vector full-value property-tax rate per USD 10,000 per town (constant for all Boston tracts)
PTRATIO	A numeric vector of pupil-teacher ratios per town (constant for all Boston tracts)
B	A numeric vector of 1000*(Bk - 0.63)^2 where Bk is the proportion of blacks
LSTAT	A numeric vector of percentage values of lower status population

Cuadro 1: Description of Explanatory variables

	OLS	SAR	SEM	SAC
CRIM	-0.216***	-0.162***	-0.185***	-0.185***
	(0.028)	(0.024)	(0.022)	(0.023)
ZN	0.067*	0.078**	0.050	0.059
	(0.031)	(0.027)	(0.031)	(0.031)
INDUS	0.041	0.043	0.003	0.019
	(0.041)	(0.035)	(0.048)	(0.046)
CHAS	0.063**	0.017	-0.015	-0.007
	(0.021)	(0.018)	(0.021)	(0.021)
NOX	-0.221***	-0.128***	-0.221***	-0.189***
	(0.043)	(0.038)	(0.060)	(0.055)
RM	0.156***	0.161***	0.200***	0.201***
	(0.029)	(0.025)	(0.024)	(0.024)
AGE	0.015	0.018	-0.062	-0.042
	(0.036)	(0.031)	(0.037)	(0.036)
DIS	-0.253***	-0.216***	-0.208**	-0.248***
	(0.041)	(0.035)	(0.065)	(0.054)
RAD	0.304***	0.272***	0.347***	0.336***
	(0.057)	(0.049)	(0.064)	(0.060)
TAX	-0.258***	-0.222***	-0.257***	-0.258***
	(0.062)	(0.053)	(0.056)	(0.056)
PTRATIO	-0.203***	-0.102***	-0.123***	-0.128***
	(0.028)	(0.025)	(0.032)	(0.030)
B	0.092***	0.078***	0.128***	0.117***
	(0.024)	(0.021)	(0.027)	(0.026)
LSTAT	-0.507***	-0.341***	-0.377***	-0.376***
	(0.035)	(0.033)	(0.035)	(0.035)
ρ		0.449***		0.198***
		(0.035)		(0.059)
λ			0.754***	0.612***
			(0.036)	(0.062)
R^2	0.790			
Adj. R^2	0.784			
Num. obs.	506	506	506	506
RMSE	0.464			
Parameters		15	15	16
Log Likelihood		-260.319	-235.298	-231.427
AIC (Linear model)		674.139	674.139	674.139
AIC (Spatial model)		550.638	500.596	494.854
LR test: statistic		125.500	175.543	183.284
LR test: p-value		0.000	0.000	0.000

$^{***}p < 0{,}001$, $^{**}p < 0{,}01$, $^{*}p < 0{,}05$

Cuadro 2: Results

Model	statistic parameter	D.F	p - value
LMerr	183.25	1	$<$2.2e-16 ***
LMlag	124.824	1	$<$2.2e-16 ***
Robust LMerr	72.312	1	$<$2.2e-16 ***
Robust LMlag	13.885	1	0.0001943 ***
SARMA	197.135	2	$<$2.2e-16 ***

Cuadro 3: Test for spatial dependence in linear models