Econometría Espacial

Profesor: Marcos Herrera Gómez

Homework 2

Lima, Matías

8 de diciembre de 2022

De lo particular a lo general

En este caso, se empieza desde el modelo M8 de la especificación vista en clase: MCO. Se comienza con el modelo más básico (sin especificaciones espaciales) y, progresivamente, se van agregando estructuras para ver si estas son significativas. Es decir, se plantea un modelo:

$$y = X\beta + \mu, \quad \mu \sim (0, \sigma^2 I), \tag{1}$$

donde, en el caso del *mystery process* bajo análisis, X es un vector que incluye ambas variables explicativas: x_1 y x_2.

Correr un test I de Moran para esta regresión muestra un valor de 0.394, con un *p-value* que tiende a 0. Es decir, es posible inferir que existe alguna estructura espacial, pero con este test no se puede determinar específicamente donde.

Por esta razón tiene sentido plantear los tests LM para estructuras espaciales en el término del error y en la variable explicada. Al observar el LM robusto para el error se puede determinar que existe estructura espacial a todos los niveles de significancia tradicional (valor de 81.18, *p-value* = 0). También aparece este resultado para el LM robusto del lag espacial, para la y, con un valor de 526.88.

Se puede concluir que el modelo M8, de MCO, no es suficiente. Por esta razón, se debe sumar complejidad.

Al ser los dos LM robustos significativos, uno podría pasar directamente a un SARAR, modelo que incorpora ambas estructuras espaciales. Sin embargo, tal como sugieren Halleck Vega & Elhorst (2015), estos tests no alcanzan a testear por estructuras espaciales en las X: primero tiene sentido probar el modelo M6, SLX.

Se parte del modelo 1 y se agregan los rezagos espaciales de las X:

$$y = X\beta + W_2 X\gamma + \mu, \quad \mu \sim (0, \sigma^2 I) \tag{2}$$

Al correr la regresión SLX, se puede ver que todos los coeficientes de la regresión son significativos, incluidos los rezagos. Sus valores son $x1 = 3.02$, $x2 = 0.99$, $wx1 = 2.02$ y $wx2 = 4.05$.

El test I de Moran, por su parte, sigue mostrando que existe alguna estructura espacial subyacente (no especificada); el valor observado de 0.117 es suficiente para rechazar la hipótesis nula de no correlación espacial.

Sin embargo, al volver a correr los test LM, se encuentra que el LM robust para los lags deja de ser significativo (el p-valor pasa a ser de 0.3429). Esto permite conjeturar que no existía correlación espacial en la y, si no que existía un problema de omisión de variables explicativas (los rezagos). Más allá de esto, el *RLMerr* sigue siendo significativo, por lo que nuestro algoritmo sugiere sumarle complejidad al modelo 2. Este modelo no es competitivo, ya que sigue mostrando dependencia en el error.

Tras probar el modelo SLX, es posible descartar tanto este modelo, como el SLM (porque la relación con las y parecía ser espuria), pero no dijimos nada del SEM:

$$y = X\beta + \lambda W_3 \mu + \varepsilon, \quad \varepsilon \sim (0, \sigma^2 I), \tag{3}$$

donde W_3 es la matriz de pesos espaciales y λ es el parámetro autorregresivo espacial.

Al correr el modelo M7, SEM, el parámetro λ toma un valor de 0.817 y es estadísticamente significativo: esto suma evidencia para pensar que existe una estructura espacial en el error. Bajo esta especificación, la variable $x2$ deja de ser significativa (su *p-valor* asociado es de 0.378), pero las pruebas previas permiten pensar que un modelo más complejo puede ser el que mejor ajusta al proceso generador de los datos: debe probarse un SDEM.

El modelo M4, SDEM, incluye estructuras tanto para las X, como para el término de error:

$$y = X\beta + W_2 X\gamma + \lambda W_3 \mu + \varepsilon, \quad \varepsilon \sim (0, \sigma^2 I) \tag{4}$$

Bajo esta especificación, se puede ver que son significativas todas las variables de la regresión: $x1$, $x2$, y sus rezagos espaciales. Además, el parámetro λ toma un valor de 0.224 y es estadísticamente significativo a los niveles habituales. Finalmente, como el modelo SEM también es competitivo, es posible comparar el ajuste de los modelos mediante el criterio de información de Akaike (AIC): 2810.4 para SDEM, 5069.4 para SEM. De esta manera, se puede concluir que el modelo que mejor ajusta a los datos es el modelo de error espacial de Durbin.

De lo general a lo particular

La idea del método *à la Hendry* es empezar desde el modelo más complejo posible para luego ir testeando si los coeficientes asociados a las estructuras espaciales no son significativos. Lo mejor sería comenzar por el modelo de Cliff-Ord, con rezagos espaciales endógenos, en las variables explicativas y en el error. Sin embargo, este modelo no está identificado empíricamente, por lo que no es un punto de partida válido.

Teniendo en cuenta esto, se comienza analizando un modelo SDM, ya que este es aquel que más modelos anida:

$$y = \rho W_1 y + X\beta + W_2 X\gamma + \mu, \quad \mu \sim (0, \sigma^2 I), \tag{5}$$

donde se incorporan rezagos para las X y para y.

Ahora, el objetivo es, a través de los test LR, ver si es posible reducir el modelo a un M5, M6 o M7.

Si se piensa en el modelo SLX (2), la hipótesis nula será $\rho = 0$: el rezago para y no es significativo. Sin embargo, este test presenta un *p-valor* de 0.007, lo que indica que se debe rechazar H_0. No es posible la reducción a un modelo SLX.

Al probar esto para la especificación SLM (M5), es posible observar lo mismo: el *p-value* es cercano a 0, por lo que el coeficiente γ es significativo en el modelo 5.

El vínculo con M7 se presenta mediante el test de contraste de factores comunes. El resultado, nuevamente, nos indica que no es posible reducir el modelo.

¿Nos quedamos con el SDM? Siguiendo a Elhorst (2014), lo óptimo es comparar M3 con M4, para ver cuál ajusta mejor a los datos. Es decir, comparamos el SDM con el SDEM. Nuevamente, se comparan los AIC: el del SDM es 2830, mientras que el del SDEM alcanza los 2810.4. Este último modelo, M4, ajusta mejor a los datos.

Finalmente, se puede considerar un SARAR: es posible ver que, tanto los coeficientes de las X, como ambos coeficientes espaciales (ρ y λ), son significativos. Sin embargo, el AIC es de 4856.3, por lo que se descarta esta especificación.

De esta manera, ambas estrategias de especificación llegan a la misma conclusión: el modelo SDEM es el más adecuado para los datos del *mystery process*.