Universidad de San Andrés

Departamento de Economía

Maestría en Economía

Estimación del modelo de rezago espacial con pérdida de datos aleatoria en la variable dependiente. Estimadores bietápicos con imputación.

Autor: Alejandro Izaguirre

Legajo: 29.945.269

Mentor: Marcos Herrera Gómez

Fecha: Buenos Aires, 7 de Julio de 2015

Estimación del modelo de rezago espacial con pérdida de datos aleatoria en la variable dependiente. Estimadores bietápicos con imputación.

Alejandro Izaguirre*

Resumen

El siguiente trabajo se enfoca en el desarrollo de un conjunto de estimadores dentro de un contexto de pérdida de observaciones en un modelo de econometría espacial. Específicamente, el modelo sobre el cual trabajaremos es un modelo de rezago espacial con pérdida de datos aleatoria en la variable dependiente. El primer estimador desarrollado es una modificación del estimador IBG2SLS presentado por Wang & Lee (2013). Un segundo estimador es presentado en base a una aproximación de la matriz de instrumentos óptima. Finalmente, se desarrolla un tercer estimador que es $\sqrt{n}-$equivalente al primero. Mediante un experimento Monte Carlo se evalúan las propiedades en muestras finitas de los diferentes estimadores, mostrando las ventajas relativas de nuestra propuesta. Para finalizar, se presenta una aplicación empírica en donde se muestra la utilidad de los estimadores ante diferentes niveles de pérdida de datos bajo un contexto real.

Palabras claves: *Pérdida de datos aleatoria, Estimadores bietápicos, Imputación, Modelo de rezago espacial.*
Clasificación JEL: *C13, C21, R5.*

*E-mail de contacto: izaguirre.ale@gmail.com.

1. Introducción

En las últimas décadas ha habido un importante y creciente desarrollo tanto a nivel teórico como aplicado de las técnicas de econometría espacial. Una de las causas de este progreso ha sido el creciente interés, a nivel teórico, por el impacto de la localización espacial de determinados fenómenos económicos (Krugman, 1991; Fujita et al., 1999). Otro motivo ha sido el surgimiento de diferentes tecnologías como los sistemas de posicionamiento globales (GPS) o los sistemas de información geográfica (SIG) que facilitan la utilización de la información georeferenciada. La combinación de estos elementos ha permitido que diversas áreas científicas relacionadas a la agricultura, medio ambiente, recursos naturales, así como tópicos de economía urbana y regional incorporen rápidamente estos progresos (Florax & Van der Vlist, 2003).

En econometría espacial, los principales desarrollos econométricos se centraron en la detección e incorporación de efectos espaciales en los modelos de regresión (Anselin, 1988). Las alternativas de especificación permitieron desarrollar diferentes modelos espaciales tales como el modelo de rezago espacial (Spatial Lag Model, SLM), el modelo de error espacial (SEM) y la combinación de ambos modelos sintetizada bajo el nombre del modelo SARAR, entre otros. Además, debido a las particularidades que conlleva la dependencia espacial, se han desarrollados métodos de estimación específicos (máxima verosimilitud, método de momentos generalizados, cuasi-máxima verosimilitud) aunque principalmente bajo un entorno de observación completa de todas las variables involucradas.

Los mencionados avances econométricos no han cubierto totalmente algunos tópicos de relevancia metodológica y aplicada. En particular, ha sido escaso el desarrollo de técnicas de estimación ante la problemática de datos faltantes, tópico bastante común en el análisis de datos observacionales. La estadística espacial ha desarrollado diferentes técnicas de predicción que pueden ser de utilidad para el problema de datos faltantes. Por ejemplo, Cressie (1993, cap. 3) detalla diferentes herramientas de predicción óptima, bajo estructuras espaciales, basadas en la técnica de Kriging. En Kelejian & Prucha (2007) se puede encontrar un análisis sobre la eficiencia de cinco tipos diferentes de predictores para un modelo SARAR según diferentes conjuntos de información. Si bien estos trabajos se vinculan con la temática de pérdida de datos, dado que brindan opciones de imputación que podrían servirnos como herramientas para sortear el problema, no hacen referencia alguna a la estimación de los modelos con datos faltantes.

Nuestro interés se focaliza en la búsqueda de una metodología de estimación espacial que nos permita obtener estimadores consistentes ante la pérdida (aleatoria) de observaciones en la variable dependiente en un modelo de rezago espacial (SLM). En la literatura, pueden encontrarse algunos trabajos que se aproximan al problema aquí analizado. Por ejemplo, Kelejian & Prucha (2010) desarrollan tres estimadores bietápicos considerando diferentes configuraciones de pérdida de datos para un modelo espacial (incluyendo pérdidas en la variable dependiente) y muestran que en determinados casos la falta de observaciones puede ser asintóticamente ignorada, no así en otros. LeSage & Pace (2004) aplican el algoritmo EM (basado en el método de máxima verosimilitud) para estimar las y´s faltantes en un modelo de rezago espacial con errores autocorrelacionados, aplicándolo a un estudio sobre precios de vivienda. Recientemente, Wang & Lee (2013) proponen para un modelo SLM, ya sea con error homocedástico o heterocedástico con estructura desconocida,

tres estimadores diferentes: uno basado en mínimos cuadrados no lineales, otro en el método de momentos generalizados y un tercero basado en mínimos cuadrados bietápicos con imputación (I2SLS).

El eje principal de nuestra propuesta, gira en torno al tercer estimador de Wang & Lee (2013), I2SLS (y sus variantes). La diferencia básica entre dicho estimador y el que desarrollaremos se centra en la imputación de los datos faltantes. Puntualmente, en procura de imputar la menor cantidad de datos posibles, nuestra propuesta restringe la imputación al rezago espacial de la variable dependiente, a diferencia de Wang & Lee que imputan, también, la variable explicada. La eficiencia del estimador dependerá de la matriz de instrumentos utilizada, por lo cual, también se propondrá una matriz óptima que minimiza la varianza del estimador, y siguiendo a Kelejian et al. (2004), se propone una aproximación de dicha matriz.

La estructura del trabajo es la siguiente. En la segunda sección, se desarrollará el modelo con datos faltantes haciendo explícita la imputación y su impacto sobre la estructura del error del modelo SLM. En la tercera sección, se presentarán los estimadores propuestos con sus características asintóticas. En la cuarta sección, se desarrollará un ejercicio de simulación de Monte Carlo para conocer el comportamiento relativo de los diferentes estimadores respecto al propuesto por Wang & Lee (2013). En la quinta sección se presenta un ejemplo empírico utilizando los datos de Baller et al. (2001) al que le impondremos diferentes niveles de pérdida de datos para comparar los ajustes. Por último, en la sexta sección, se presentan las principales conclusiones de la investigación. Para facilitar la comprensión del texto, las demostraciones se presentan en un apéndice.

QUAERERE VERUM
Universidad de
SanAndrés

2. Desarrollo del modelo con datos faltantes

2.1. Modelo

Sea el modelo SLM:

$$y_n = \lambda_0 W_n y_n + X_n \beta_0 + e_n, \tag{2.1}$$

donde y_n es un vector de variables dependientes de orden $(n \times 1)$, X_n es una matriz de variables exógenas de orden $(n \times k)$, β_0 es una vector de parámetros de orden $(k \times 1)$, e_n es un vector de errores aleatorios de orden $(n \times 1)$ tal que $e_n \sim \left(0, \sigma_e^2 I_n\right)$, W_n es una matriz de pesos espaciales de orden $(n \times n)$ y λ_0 es un parámetro de autocorrelación espacial.

La matriz W_n permite introducir dependencia espacial entre las unidades determinando qué observaciones son vecinas entre sí. Formalmente, la matriz de pesos espaciales es representada por:

$$W_n = \begin{bmatrix} 0 & w_{12} & \cdots & w_{1n} \\ w_{21} & 0 & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \cdots & 0 \end{bmatrix},$$

donde las filas y columnas identifican a las observaciones de corte transversal, siendo n el tamaño del conjunto de datos y w_{ij} $(i, j = 1, 2, \ldots, n)$ las ponderaciones o pesos que aproximan la relación entre dos unidades i (filas) y j (columnas). La diagonal principal esta formada por ceros, estableciendo que ninguna observación puede ser vecina de si misma. La matriz está conformada por pesos exógenos al modelo establecidos por algún criterio (por ejemplo, contigüidad o distancia geográfica).

Para introducir y analizar la falta de datos en la variable dependiente, expresamos al modelo (2.1) de la siguiente manera:

$$\begin{pmatrix} y_n^o \\ y_n^u \end{pmatrix} = \lambda_0 W_n \begin{pmatrix} y_n^o \\ y_n^u \end{pmatrix} + X_n \beta_0 + e_n,$$

donde y_n^o es un subvector de y_n de orden $(n^o \times 1)$ con n^o igual al número de y´s observadas, de la misma forma, y_n^u es un subvector de y_n de orden $(n^u \times 1)$ con n^u igual al número de y´s no observadas, con $n = n^o + n^u$.

Definamos a J_n^o como una matriz de selección de orden $(n^o \times n)$ que surge de eliminar de una matriz identidad $(n \times n)$ las filas que se corresponden con las y´s no observadas. De la misma manera, sea J_n^u una matriz de selección de orden $(n^u \times n)$ que surge de eliminar de una matriz identidad $(n \times n)$ las filas que se corresponden con las y´s observadas, tal que $y_n^o = J_n^o y_n$ es un subvector de y_n, de orden $(n^o \times 1)$, que contiene a las y´s observadas; y $y_n^u = J_n^u y_n$ es un subvector de y_n, de orden $(n^u \times 1)$, que contiene a las y´s no observadas.

Nuestra intención es trabajar con el modelo de variables endógenas observadas tal que utilizando las matrices de selección, el modelo inicial será:

$$\begin{aligned} J_n^o \begin{pmatrix} y_n^o \\ y_n^u \end{pmatrix} &= J_n^o \lambda_0 W_n \begin{pmatrix} y_n^o \\ y_n^u \end{pmatrix} + J_n^o X_n \beta_0 + J_n^o e_n, \\ y_n^o &= J_n^o \lambda_0 W_n \begin{pmatrix} y_n^o \\ y_n^u \end{pmatrix} + J_n^o X_n \beta_0 + J_n^o e_n. \end{aligned} \tag{2.2}$$

En el caso de las observaciones faltantes imputaremos las mismas a través de una estimación de su esperanza. Little (1992)[1] afirma que, habiendo información completa sobre las x´s y siendo la pérdida de valores de y aleatoria, imputar los valores de las y´s faltantes no agrega información a la regresión de y sobre X, en un modelo sin variables endógenas en el lado derecho de la ecuación. Basándonos en este punto, y procurando imputar la menor cantidad de dato posibles, nuestra propuesta es no imputar los valores de y_n^u que representan a las variables explicadas, sino solo hacerlo sobre el rezago espacial, con lo cual, la imputación solo la haremos sobre $W_n \begin{pmatrix} y_n^o \\ y_n^u \end{pmatrix}$.

De la ecuación (2.1) puede deducirse que $y_n = (I_n - \lambda_0 W_n)^{-1} X_n \beta_0 + (I_n - \lambda_0 W_n)^{-1} e_n$, tal que al aplicar esperanza se obtiene: $E(y_n^u) = J_n^u E(y_n) = J_n^u (I_n - \lambda_0 W_n)^{-1} X_n \beta_0 = F_n(\theta_0)$, con $\theta_0 = \left(\lambda_0, \beta_0'\right)'$.

[1] Véase también Von Hippel (2007, pp. 9-11).

El vector de variables observadas y no observadas, $\begin{pmatrix} y_n^{o'} & , & y_n^{u'} \end{pmatrix}'$ puede ser expresado como:

$$\begin{pmatrix} y_n^o \\ y_n^u \end{pmatrix} = \left[\begin{pmatrix} y_n^o \\ F(\theta_0) \end{pmatrix} + \begin{pmatrix} 0 \\ y_n^u - F_n(\theta_0) \end{pmatrix} \right],$$

y reemplezando en (2.2) obtenemos:

$$\begin{aligned} y_n^o &= J_n^o \lambda_0 W_n \begin{pmatrix} y_n^o \\ F_n(\theta_0) \end{pmatrix} + J_n^o X_n \beta_0 + J_n^o \lambda_0 W_n \begin{pmatrix} 0 \\ y_n^u - F_n(\theta_0) \end{pmatrix} + J_n^o e_n, \\ y_n^o &= J_n^o \lambda_0 W_n \begin{pmatrix} y_n^o \\ F_n(\theta_0) \end{pmatrix} + J_n^o X_n \beta_0 + J_n^o u_n, \end{aligned} \tag{2.3}$$

dado que $(y_n^u - F_n(\theta_0)) = J_n^u S_n^{-1}(\lambda_0) e_n$, con $S_n(\lambda_0) = (I_n - \lambda_0 W_n)$, y siendo $\begin{pmatrix} 0 \\ y_n^u - F_n(\theta_0) \end{pmatrix} = J_n^{u'} J_n^u S_n^{-1}(\lambda_0) e_n$ tenemos que $u_n = \left[\lambda_0 W_n J_n^{u'} J_n^u S_n^{-1}(\lambda_0) + I_n\right] e_n$.

La ecuación (2.3) incorpora la esperanza con el objetivo de salvar los datos faltantes del lado derecho de la ecuación. Esta esperanza involucra parámetros desconocidos que deberán ser estimados previamente. Para ello, asumamos que existe una estimación consistente de estos parámetros, $\hat{\theta}$, tal que:

$$\begin{pmatrix} y_n^o \\ F_n(\theta_0) \end{pmatrix} = \left[\begin{pmatrix} y_n^o \\ F\left(\hat{\theta}\right) \end{pmatrix} - \begin{pmatrix} 0 \\ F_n\left(\hat{\theta}\right) - F_n(\theta_0) \end{pmatrix} \right],$$

y reemplazando en (2.3) se obtiene:

$$\begin{aligned} y_n^o &= J_n^o \lambda_0 W_n \begin{pmatrix} y_n^o \\ F_n\left(\hat{\theta}\right) \end{pmatrix} + J_n^o X_n \beta_0 - J_n^o \lambda_0 W_n \begin{pmatrix} 0 \\ F_n\left(\hat{\theta}\right) - F_n(\theta_0) \end{pmatrix} + J_n^o u_n, \\ y_n^o &= J_n^o \lambda_0 W_n \begin{pmatrix} y_n^o \\ F_n\left(\hat{\theta}\right) \end{pmatrix} + J_n^o X_n \beta_0 + J_n^o \tilde{u}_n, \end{aligned} \tag{2.4}$$

dado que $\begin{pmatrix} 0 \\ F_n\left(\hat{\theta}\right) - F_n(\theta_0) \end{pmatrix} = J_n^{u'} J_n^u \left[\left(S_n^{-1}\left(\hat{\lambda}\right) X_n \hat{\beta}\right) - \left(S_n^{-1}(\lambda_0) X_n \beta_0\right)\right]$, siendo

$$\tilde{u}_n = u_n - \lambda_0 W_n J_n^{u'} J_n^u \left[\left(S_n^{-1}\left(\hat{\lambda}\right) X_n \hat{\beta}\right) - \left(S_n^{-1}(\lambda_0) X_n \beta_0\right)\right]. \tag{2.5}$$

El modelo sobre el cual trabajaremos es el establecido en la ecuación (2.4), cuyo error responde a la forma funcional (2.5).

2.2. Desarrollo de $\tilde{u}_n$

En esta sección nos enfocaremos en el desarrollo de $\tilde{u}_n$ prestando especial atención al término de la ecuación (2.5), $\left[\left(S_n^{-1}\left(\hat{\lambda}\right)X_n\hat{\beta}\right)-\left(S_n^{-1}\left(\lambda_0\right)X_n\beta_0\right)\right]$, el cual captura el error en la estimación de $E\left(y_n^u\right)^2$.

Sean $\left[\left(S_n^{-1}\left(\hat{\lambda}\right)X_n\hat{\beta}\right)-\left(S_n^{-1}\left(\lambda_0\right)X_n\beta_0\right)\right] = \left[S_n^{-1}\left(\hat{\lambda}\right)-S_n^{-1}\left(\lambda_0\right)\right]X_n\hat{\beta} + S_n^{-1}\left(\lambda_0\right)X_n\left(\hat{\beta}-\beta_0\right)$, y $\left(S_n^{-1}\left(\hat{\lambda}\right)-S_n^{-1}\left(\lambda_0\right)\right) = \left(\hat{\lambda}-\lambda_0\right)S_n^{-1}\left(\hat{\lambda}\right)G_n$, con $G_n = W_nS_n^{-1}\left(\lambda_0\right)$. Utilizando estas igualdades obtenemos:

$$\begin{aligned}\left[\left(S_n^{-1}\left(\hat{\lambda}\right)X_n\hat{\beta}\right)-\left(S_n^{-1}\left(\lambda_0\right)X_n\beta_0\right)\right] &= \left[S_n^{-1}\left(\hat{\lambda}\right)G_nX_n\hat{\beta} \quad S_n^{-1}\left(\lambda_0\right)X_n\right]\begin{bmatrix}\left(\hat{\lambda}-\lambda_0\right)\\ \left(\hat{\beta}-\beta_0\right)\end{bmatrix},\\ &= S_n^{-1}\left(\lambda_0\right)C_n\left(\hat{\theta}-\theta_0\right)+R_n\left(\hat{\lambda}-\lambda_0\right),\end{aligned}$$

con $R_n = \left[S_n^{-1}\left(\hat{\lambda}\right)G_nX_n\hat{\beta}-S_n^{-1}\left(\lambda_0\right)G_nX_n\beta_0\right]$ y $C_n = \left(G_nX_n\beta_0 \quad X_n\right)$.

Utilizando los anteriores resultados podemos reexpresar (2.5) como

$$\tilde{u}_n = u_n-\lambda_0W_nJ_n^{u'}J_n^uS_n^{-1}\left(\lambda_0\right)C_n\left(\hat{\theta}-\theta_0\right)+\lambda_0W_nJ_n^{u'}J_n^uR_n\left(\hat{\lambda}-\lambda_0\right). \tag{2.6}$$

Siguiendo la propuesta de Wang & Lee, utilizaremos una estimación no lineal de $E\left(y_n^o\right) = h_n\left(X_n \quad \theta_0\right)+v_n$, donde $h_n = J_n^oS_n^{-1}\left(\lambda_0\right)X_n\beta_0$ y $v_n = B_ne_n$, con $B_n = J_n^oS_n^{-1}\left(\lambda_0\right)$, dado que $E\left(v_n\right)=0$. El método de estimación elegido es el de Mínimos Cuadrados no Lineales (NLS).

Bajo un conjunto de supuestos típicos[3] puede obtenerse la distribución asintótica de los estimadores:

$$\sqrt{n}\left(\hat{\theta}_{nls,n}-\theta\right) = \left[\frac{1}{n}C_n^{'}B_n^{'}B_nC_n\right]^{-1}\frac{1}{\sqrt{n}}C^{'}B_n^{'}B_ne_n+o_p(1)\xrightarrow{d}N\left(0,\Sigma_{nls}\right). \tag{2.7}$$

Utilizando el resultado de la ecuación (2.7), podemos utilizar $\hat{\theta}_{nls,n}$ como una aproximación consistente tal que, al reemplazar en (2.6) tenemos:

$$\begin{aligned}\tilde{u}_n &= u_n-\lambda_0W_nJ_n^{u'}J_n^uS_n^{-1}\left(\lambda_0\right)C_n\left[\left[C_n^{'}B_n^{'}B_nC_n\right]^{-1}C^{'}B_n^{'}B_ne_n+o_p(\frac{1}{\sqrt{n}})\right]\\ &\quad +\lambda_0W_nJ_n^{u'}J_n^uR_n\left(\hat{\lambda}-\lambda_0\right),\end{aligned}$$

reemplazando u_n por su expresión, reordenando términos y distribuyendo nos queda:

[2] Un mayor desarrollo del presentado aquí puede encontrarse en el Apéndice, sección A.

[3] Los supuestos son los mismos que utilizaremos en la sección 3 del presente trabajo, más ciertas condiciones de identificación y existencia de limites.

$$\tilde{u}_n = \left\{\lambda_0 W_n J_n^{u'} J_n^u S_n^{-1} + I_n - \lambda_0 W_n J_n^{u'} J_n^u S_n^{-1}(\lambda_0) C_n \left[C_n' B_n' B_n C_n\right]^{-1} C' B_n' B_n\right\} e_n + R^*,$$

con $R^* = \lambda_0 W_n J_n^{u'} J_n^u \left[R_n\left(\hat{\lambda} - \lambda_0\right) - S_n^{-1}(\lambda_0) C_n l_k o_p(\frac{1}{\sqrt{n}})\right]$ donde l_k es un vector de unos de dimensión $k+1$.

Con todo lo anterior podemos expresar a nuestro modelo con datos faltantes (2.4) de la siguiente manera

$$y_n^o = J_n^o \tilde{Z}_n \theta_0 + J_n^o \left(H_n(\lambda_0) e_n + R^*\right), \tag{2.8}$$

donde: $\tilde{Z}_n = (W_n \tilde{y} \quad X_n)$, con $\tilde{y} = \begin{pmatrix} y_n^o \\ F_n\left(\hat{\theta}_{nls}\right) \end{pmatrix}$; $\theta_0 = \begin{pmatrix} \lambda_0 & \beta_0' \end{pmatrix}'$ y $H_n(\lambda_0) = \left\{\lambda_0 W_n J_n^{u'} J_n^u S_n^{-1}(\lambda_0) + I_n - \lambda_0 W_n J_n^{u'} J_n^u S_n^{-1}(\lambda_0) C_n \left[C_n' B_n' B_n C_n\right]^{-1} C_n' B_n' B_n\right\}$.

3. Estimadores

En esta sección desarrollaremos tres estimadores para los parámetros del modelo (2.8). El primero es una variante del estimador de Wang & Lee (2013) denominado "Best Generalized Two Stage Least Squares with Imputation" (IBG2SLS). El segundo estimador es una alternativa del anterior utilizando una aproximación a la matriz de instrumentos óptima, presentada por Kelejian et al. (2004) en el desarrollo de su estimador "A Series-Type Efficient IV Estimator". Por último, presentaremos un estimador basado en el cálculo del límite del IBG2SLS, el cual es $\sqrt{n}$-equivalente al mismo.

3.1. Supuestos

Supuesto 1: $e_i \sim i.i.d.\left(0, \sigma_e^2\right)$, $\forall i = 1, \ldots, n$, con $0 < \sigma_e^2 < \infty$.

Supuesto 2: Los elementos de X_n se encuentran uniformemente acotados y son no estocásticos, X_n tiene rango completo en columnas y el $\lim_{n\to\infty} \frac{1}{n} X_n' X_n$ existe y es no singular.

Supuesto 3: La suma de las filas y columnas de W_n y de $S_n^{-1}(\lambda)$ están acotadas uniformemente en valor absoluto para $|\lambda| < 1$.

Este supuesto garantiza que $S_n(\lambda)$ sea no singular, y además, permite expresar $S_n^{-1}(\lambda) = \sum_{k=0}^{\infty} \lambda^k W_n^k$.

Supuesto 4: La proporción de observaciones $\frac{n^0}{n}$ tiende a c cuando n va a infinito, donde c es una constante finita tal que $0 < c < 1$.

Este supuesto permite que la convergencia de los estimadores sea expresada en términos de n. La condición establece, básicamente, que la cantidad de datos observados crezca a medida que crece n.

Supuesto 5: Q_n^o es una matriz de instrumentos en función de $J_n^o X_n$ y $J_n^o W_n$. Los elementos de Q_n^o deben ser uniformemente acotados. Además, $\lim_{n\to\infty} \frac{1}{n} Q_n^{o'} Q_n^o$ es no singular y $\lim_{n\to\infty} \frac{1}{n} Q_n^{o'} \left[J_n^o G_n X_n \beta_0 \quad J_n^o X_n \right]$ tiene rango completo en columnas.

3.1.1. Estimador de Mínimos Cuadrados en dos Etapas con Imputación (I2SLS)

Siguiendo a Wang & Lee, comenzaremos por desarrollar un estimador de 2SLS imputando ciertos valores de $J_n^o W_n y_n$. El estimador propuesto bajo el nombre I2SLS es:

$$\hat{\theta}_{i2sls,n} = \left[\tilde{Z}_n^{o'} Q_n^o \left(Q_n^{o'} Q_n^o \right)^{-1} Q_n^{o'} \tilde{Z}_n^o \right]^{-1} \tilde{Z}_n^{o'} Q_n^o \left(Q_n^{o'} Q_n^o \right)^{-1} Q_n^{o'} y_n^o, \tag{3.1}$$

siendo Q_n^o una matriz de instrumentos con rango mayor o igual a $k+1$ construida como función de $J_n^o W_n$ y $J_n^o X_n$, y sea $\tilde{Z}_n^o = J_n^o \tilde{Z}_n$, con $\tilde{Z}_n$ definida en 2.2.

Proposición 1. *Bajo los supuestos anteriormente establecidos, $\hat{\theta}_{i2sls,n}$ es un estimador consistente de θ_0 y $\sqrt{n}\left(\hat{\theta}_{i2sls,n} - \theta_0\right) \xrightarrow{d} N\left(0, \Sigma_{i2sls}\right)$ donde*

$$\begin{aligned} \Sigma_{i2sls} = \; & \sigma_e^2 \lim_{n\to\infty} \left[\frac{1}{n} C_n^{o'} Q_n^o \left(Q_n^{o'} Q_n^o \right)^{-1} Q_n^{o'} C_n^o \right]^{-1} \\ & \left[\frac{1}{n} C_n^{o'} Q_n^o \left(Q_n^{o'} Q_n^o \right)^{-1} Q_n^{o'} H_n^o\left(\lambda_0\right) H_n^{o'}\left(\lambda_0\right) Q_n^o \left(Q_n^{o'} Q_n^o \right)^{-1} Q_n^{o'} C_n^o \right] \\ & \left[\frac{1}{n} C_n^{o'} Q_n^o \left(Q_n^{o'} Q_n^o \right)^{-1} Q_n^{o'} C_n^o \right]^{-1}. \end{aligned}$$

3.1.2. Estimador de Mínimos Cuadrados en dos Etapas Generalizado con Imputación (IG2SLS)

La versión generalizada del estimador I2SLS utiliza la estructura del error para ponderar a las variables. Para ello, utilizaremos la varianza asintótica del error de nuestro modelo (2.4): $Avar\left(\tilde{u}_n^o\right) = Avar\left(H_n^o\left(\lambda_0\right) e_n + J_n^o R^*\right) = \left(\sigma_e^2 H_n^o\left(\lambda_0\right) H_n^{o'}\left(\lambda_0\right)\right)$, con $H_n^o\left(\lambda_0\right) = J_n^o H_n\left(\lambda_0\right)$ y siendo $J_n^o R^* \xrightarrow{p} 0$.

Definiendo $\Omega\left(\lambda_0\right) = H_n^o\left(\lambda_0\right) H_n^{o'}\left(\lambda_0\right)$, entonces $Avar\left(\tilde{u}_n^o\right) = \sigma_e^2 \Omega\left(\lambda_0\right)$. Esta estructura asintótica del error nos permite determinar un nuevo estimador al que llamaremos estimador de Mínimos Cuadrados Generalizados en dos Etapas con Imputación (IG2SLS).

Supuesto 6: Sea $\Omega_n\left(\lambda_0\right)$ uniformemente definida positiva, simétrica de orden $\left(n^o \times n^o\right)$, con $\lim_{n\to\infty} \frac{1}{n} \Omega_n\left(\lambda_0\right)$ definida positiva, simétrica. Además, el $\lim_{n\to\infty} \frac{1}{n} Q_n^{o'} \Omega_n^{-1}\left(\lambda_0\right) Q_n^o$ finito y no singular, y $\lim_{n\to\infty} \frac{1}{n} Q_n^{o'} \Omega_n^{-1}\left(\lambda_0\right) \left[J_n^o G_n X_n \beta_0 \quad J_n^o X_n \right]$ tiene rango completo en columnas.

El estimador IG2SLS es

$$\hat{\theta}_{ig2sls,n} = \left[\tilde{Z}_n^{o'}\Omega_n^{-1}(\lambda_0)Q_n^o\left(Q_n^{o'}\Omega_n^{-1}(\lambda_0)Q_n^o\right)^{-1}Q_n^{o'}\Omega_n^{-1}(\lambda_0)\tilde{Z}_n^o\right]^{-1}$$
$$\tilde{Z}_n^{o'}\Omega_n^{-1}(\lambda_0)Q_n^o\left(Q_n^{o'}\Omega_n^{-1}(\lambda_0)Q_n^o\right)^{-1}Q_n^{o'}\Omega_n^{-1}(\lambda_0)y_n^o.$$

Proposición 2. *Bajo los supuestos 1-4 y 6, $\hat{\theta}_{ig2sls,n}$ es un estimador consistente de θ_0, y $\sqrt{n}\left(\hat{\theta}_{ig2sls,n}-\theta_0\right)\xrightarrow{d}N(0,\Sigma_{ig2sls})$ donde*

$$\Sigma_{ig2sls} = \sigma_e^2\lim_{n\to\infty}\left[\frac{1}{n}C_n^{o'}\Omega_n^{-1}(\lambda_0)Q_n^o\left(Q_n^{o'}\Omega_n^{-1}(\lambda_0)Q_n^o\right)^{-1}Q_n^{o'}\Omega_n^{-1}(\lambda_0)C_n^o\right]^{-1}.$$

3.1.3. Mejor Estimador de Mínimos Cuadrados en dos Etapas Generalizado con Imputación (IBG2SLS)

Hasta acá hemos derivado los estimadores (I2SLS y IG2SLS) basándonos en una matriz de instrumentos Q_n^o sobre la cual solo establecimos que debía ser función de $J_n^oW_n$ y $J_n^oX_n$. Sin embargo, es posible derivar una mejor matriz de instrumentos, es decir, aquella que minimice la varianza del estimador.

Para que la varianza del estimador de IG2SLS sea mínima, debe ser máximo el término entre corchetes de Σ_{ig2sls}. Si reemplazamos en dicho término, $\Omega_n^{-1}(\lambda_0)$, por su igualdad $K_n^{'}K_n$[4] nos queda $C_n^{o'}K_n^{'}K_nQ_n^o\left(Q_n^{o'}K_n^{'}K_nQ_n^o\right)^{-1}Q_n^{o'}K_n^{'}K_nC_n^o$, además, definiendo $A_n=(K_nC_n^o)^{'}$, $X_n=K_nQ_n^o$ y $V_n=I_n$ y aplicando la desigualdad generalizada de Cauchy-Schwarz[5], tenemos que:

$$\begin{aligned}(K_nC_n^o)^{'}I_n(K_nC_n^o) &\geq (K_nC_n^o)^{'}(K_nQ_n^o)\left[(K_nQ_n^o)^{'}I_n(K_nQ_n^o)\right]^{-1}(K_nQ_n^o)^{'}(K_nC_n^o),\\ &\geq C_n^{o'}K_n^{'}K_nQ_n^o\left(Q_n^{o'}K_n^{'}K_nQ_n^o\right)^{-1}Q_n^{o'}K_n^{'}K_nC_n^o,\end{aligned}$$

donde $(K_nC_n^o)^{'}(K_nC_n^o)$ es la cota superior del término que queremos maximizar. La cota máxima será alcanzada cuando $Q_n^o=C_n^o$. Esto implica que la matriz de instrumentos que minimiza la varianza del estimador IG2SLS es $Q_n^{o*}=C_n^o$.

Con todo lo anterior, el estimador IBG2SLS es

$$\begin{aligned}\hat{\theta}_{ibg2sls,n} &= \left[\tilde{Z}_n^{o'}\Omega_n^{-1}(\lambda_0)Q_n^{o*}\left(Q_n^{o*'}\Omega_n^{-1}(\lambda_0)Q_n^{o*}\right)^{-1}Q_n^{o*'}\Omega_n^{-1}(\lambda_0)\tilde{Z}_n^o\right]^{-1},\\ &\quad\tilde{Z}_n^{o'}\Omega_n^{-1}(\lambda_0)Q_n^{o*}\left(Q_n^{o*'}\Omega_n^{-1}(\lambda_0)Q_n^{o*}\right)^{-1}Q_n^{o*'}\Omega_n^{-1}(\lambda_0)y_n^o. \qquad (3.2)\end{aligned}$$

[4] Ver Apéndice, sección C.

[5] La desigualdad generalizada de Cauchy-Schwarz establece que: Siendo V_n una matriz definida, positiva y simétrica de orden $(n\times n)$ y A_n y $X_n^{'}$ matrices de orden $(n\times k)$, con X_n de rango k, entonces se cumple que $A_nV_nA_n^{'}\geq A_nX_n\left(X_n^{'}V_n^{-1}X_n\right)^{-1}X_n^{'}A_n$.

La distribución asintótica de IBG2SLS es la misma que la de IG2SLS, solo que hemos definido como matriz de instrumentos a $Q_n^{o*} = C_n^o$, y esto hace que algunos términos de la varianza asintótica se cancelen.

Proposición 3. *Bajo los supuestos 1-4 y 6,* $\hat{\theta}_{ibg2sls,n}$ *es un estimador consistente de* θ_0*, y* $\sqrt{n}\left(\hat{\theta}_{ibg2sls,n} - \theta_0\right) \xrightarrow{d} N\left(0, \Sigma_{ibg2sls}\right)$ *donde*

$$\Sigma_{ibg2sls} \quad = \sigma_e^2 \lim_{n\to\infty} \quad \left[\frac{1}{n} C_n^{o'} \Omega_n^{-1}\left(\lambda_0\right) C_n^o\right]^{-1}.$$

3.1.4. Estimador Series-Type Efficient IV con imputación (ISTE)

Si nos olvidamos por un momento de la falta de observaciones, y pensamos en un modelo SLM clásico, la matriz de instrumentos ideal, establecida por Kelejian & Prucha (1998) es $Q_n^{kp} = E\left(Z_n\right)$, con $Z_n = [W_n y_n \quad X_n]$. Dado que, la $E\left(y_n\right) = \left(I_n - \lambda_0 W_n\right)^{-1} X_n \beta_0$, podemos expresar dicha matriz como, $E\left(Z_n\right) = [W_n E\left(y_n\right) \quad X_n] = \left[W_n S_n^{-1} X_n \beta_0 \quad X_n\right] = C_n$, como vemos, la matriz óptima sugerida por Kelejian & Prucha, es la misma a la que arribamos en el desarrollo del estimador IBG2SLS.

Si bien Kelejian & Prucha establecen la matriz de instrumentos óptima, inicialmente no la utilizan en su metodología de estimación, sino que solo usan algunas columnas linealmente independientes de $\left\{X_n \quad W_n X_n \quad W_n^2 X_n \ldots\right\}$ que arroja estimadores consistentes y asintóticamente normales, pero que no son asintóticamente óptimos. Luego, en un trabajo posterior (Kelejian et al., 2004), desarrollan un estimador basado en una aproximación a la matriz de instrumentos óptima, al que llaman "Series-Type Efficient IV", y muestran que dicho estimador es consistente, asintóticamente normal y eficiente (dentro de su clase)[6].

Dado que la matriz de instrumentos óptima del estimador IBG2SLS es la misma propuesta por Kelejian & Prucha (1998), y que a través de una aproximación de dicha matriz se logran estimadores asintóticamente eficientes, lo que proponemos en este apartado es utilizar dicha aproximación para arribar a un nuevo estimador, al que llamaremos Estimador Series-Type Efficient IV con imputación (ISTE).

Aproximación a la matriz de instrumentos Dado el supuesto 3, podemos expresar a la esperanza de y_n de la siguiente manera, $E\left(y_n\right) = \sum_{k=0}^{\infty} \lambda_0^k W_n^k X_n \beta_0$. Utilizando esta fórmula en la expresión de la matriz de instrumentos óptima, obtenemos $E\left(Z_n\right) = [W_n E\left(y_n\right) \quad X_n] = \left[\sum_{k=0}^{\infty} \lambda_0^k W_n^{k+1} X_n \beta_0 \quad X_n\right]$.

Kelejian et al. (2004) establecen que, siendo r_n una secuencia de números naturales, tal que $r_n \uparrow \infty$, la matriz de instrumentos óptima puede aproximarse a través de $\hat{Q}_n^{kp} =$

[6] Ver Kelejian et al. (2004). El modelo desarrollado en el trabajo es un SARAR, pero la extensión a un SLM es directa, suponiendo que el parámetro de correlación espacial del error es cero ($\rho = 0$).

$\left[\sum_{k=0}^{r_n} \hat{\lambda}_0^k W_n^{k+1} X_n \hat{\beta}_0 \quad X_n\right]$, para ello es necesario que $\hat{\theta} = \left(\hat{\lambda} \quad \hat{\beta}'\right)'$ sean estimadores $\sqrt{n}$-consistentes de θ[7].

Considerando lo anterior, el estimador que proponemos es el siguiente:

$$\begin{aligned}\hat{\theta}_{iste,n} &= \left[\tilde{Z}_n^{o'} \Omega_n^{-1}(\lambda_0) \hat{Q}_n^{okp} \left(\hat{Q}_n^{okp'} \Omega_n^{-1}(\lambda_0) \hat{Q}_n^{okp}\right)^{-1} \hat{Q}_n^{okp'} \Omega_n^{-1}(\lambda_0) \tilde{Z}_n^{o}\right]^{-1} \\ &\quad \tilde{Z}_n^{o'} \Omega_n^{-1}(\lambda_0) \hat{Q}_n^{okp} \left(\hat{Q}_n^{okp'} \Omega_n^{-1}(\lambda_0) \hat{Q}_n^{okp}\right)^{-1} \hat{Q}_n^{okp'} \Omega_n^{-1}(\lambda_0) y_n^{o},\end{aligned}$$

donde $\hat{Q}_n^{okp} = J_n^o \hat{Q}_n^{kp} = \left[J_n^o \sum_{k=0}^{r_n} \hat{\lambda}_{nls}^k W_n^{k+1} X_n \hat{\beta}_{nls} \quad J_n^o X_n\right]$.

La distribución asintótica de ISTE es la misma que la de IG2SLS, solo que hemos definido como matriz de instrumentos a $Q_n^o = \hat{Q}_n^{okp}$, y dado que (como demostramos en el apéndice) $\frac{1}{n}\hat{Q}_n^{okp} \xrightarrow{p} \frac{1}{n} C_n^o$, tenemos la siguiente proposición.

Proposición 4. *Bajo los supuestos 1-4 y 6, y siendo* $\sqrt{n}\left(\hat{\lambda}_{nls} - \lambda_0\right) = O_p(1)$[8], *con* $|\lambda_0| < 1$, *y* $\left(\hat{\beta}_{nls} - \beta_0\right) = o_p(1)$. *Sea* r_n *una secuencia de números naturales con* $0 \leqslant r_n \leqslant n$, $r_n \uparrow \infty$ *y* $r_n = o(\sqrt{n})$, $\hat{\theta}_{iste,n}$ *es un estimador consistente de* θ_0, *y* $\sqrt{n}\left(\hat{\theta}_{iste,n} - \theta_0\right) \xrightarrow{d} N(0, \Sigma_{iste})$, *donde*

$$\Sigma_{iste} = \sigma_e^2 \lim_{n\to\infty} \left[\frac{1}{n} C_n^{o'} \Omega_n^{-1}(\lambda_0) C_n^o\right]^{-1}.$$

3.1.5. Mejor Estimador de Mínimos Cuadrados en dos Etapas Generalizado con Imputación Asintótico (IBG2SLSA)

El estimador que presentamos en esta sección surge de una propiedad conocida como equivalencia asintótica.

Sea el estimador IBG2SLSA

$$\hat{\theta}_{ibg2slsa,n} = \left(C_n^{o'} \Omega_n^{-1}(\lambda_0) C_n^o\right)^{-1} C_n^{o'} \Omega_n^{-1}(\lambda_0) y_n^o.$$

Dado que, como se demuestra en el apéndice, $\sqrt{n}\left(\hat{\theta}_{ibg2sls,n} - \hat{\theta}_{ibg2slsa,n}\right) = o_p(1)$, ambos estimadores son $\sqrt{n}-$equivalentes, con lo cual tienen la misma distribución asintótica.

Proposición 5. *Bajo los supuestos 1-4 y 6,* $\hat{\theta}_{ibg2slsa,n}$ *es un estimador consistente de* θ_0, *y* $\sqrt{n}\left(\hat{\theta}_{ibg2slsa,n} - \theta_0\right) \xrightarrow{d} N(0, \Sigma_{ibg2slsa})$ *donde*

$$\Sigma_{ibg2slsa} = \sigma_e^2 \lim_{n\to\infty} \left[\frac{1}{n} C_n^o \Omega_n^{-1}(\lambda_0) C_n^o\right]^{-1}.$$

[7] Para el desarrollo formal, que establece las propiedades asintóticas del estimador, es necesario suponer que $0 \leq r_n \leq n$ y además, $r_n = o_p\left(\sqrt{n}\right)$. Esta imposición no tiene implicancias prácticas, ya que, como detallaremos en la simulación, la aproximación se alcanza para valores muy pequeños de r_n.

[8] La proposición se cumple para $\sqrt{n}\left(\hat{\lambda}_n - \lambda_0\right) = O_p(1)$, en este caso hemos reemplazado $\hat{\lambda}_n$ por $\hat{\lambda}_{nls}$ porque es el que utilizamos en el trabajo para imputar los datos faltantes, por ende, es el que nos interesa.

3.2. Estimador de la varianza del error

Como vimos a lo largo del trabajo, la varianza de los estimadores depende de σ_e^2, el cual es un parámetro desconocido que debemos estimar. En este apartado propondremos un estimador del mismo basado en el error del modelo utilizado para imputar los datos faltantes.

En la sección 2.2 establecimos que una alternativa de imputación, para los datos faltantes, era a través de la estimación de

$$J_n^o y_n = J_n^o S_n^{-1}(\lambda_0) X_n \beta_0 + J_n^o S_n^{-1}(\lambda_0) e_n, \tag{3.3}$$

utilizando el método de mínimos cuadrados no lineales.

La varianza de dicho modelo es $\sigma_e^2 \left[J_n^o S_n^{-1}(\lambda_0) S_n^{-1'}(\lambda_0) J_n^{o'} \right]$, siendo $\Sigma_n^{-1} = \left[J_n^o S_n^{-1}(\lambda_0) S_n^{-1'}(\lambda_0) J_n^{o'} \right]^{-1}$, utilizando la descomposición de Cholesky, la misma puede expresarse como $\Sigma_n^{-1}(\lambda_0) = T_n^{'}(\lambda_0) T_n(\lambda_0)$, donde $T_n(\lambda_0)$ es una matriz triangular inferior no singular de orden $(n^o \times n^o)$.

El error de (3.3) es $J_n^o S_n^{-1}(\lambda_0) e_n$, si lo multiplicamos por $T_n(\lambda_0)$, tenemos que, $E\left[T_n(\lambda_0) J_n^o S_n^{-1}(\lambda_0) e_n\right] = 0$, $Var\left[T_n(\lambda_0) J_n^o S_n^{-1}(\lambda_0) e_n\right] = \sigma_e^2 I_n$, y $cov\left[T_n(\lambda_0) J_n^o S_n^{-1}(\lambda_0) e_n\right] = 0$.

Siendo v_i el elemento i-ésimo del vector $v_n = T_n(\lambda_0) J_n^o S_n^{-1}(\lambda_0) e_n$, por ley débil de los grandes números tenemos que, $\frac{1}{n}\sum_{i=1}^n v_i^2 \xrightarrow{p} \frac{1}{n}\sum_{i=1}^n E\left(v_i^2\right) = \frac{1}{n} n \sigma_e^2 = \sigma_e^2$.

Dado que, como mostramos en el apéndice, $\hat{v}_n = T_n\left(\hat{\lambda}\right) J_n^o S_n^{-1}\left(\hat{\lambda}\right) \hat{e}_n$, es un estimador consistente de v_n, el estimador que proponemos para la varianza es,

$$\hat{\sigma}_e^2 = \frac{1}{n}\sum_{i=1}^n \hat{v}_i^2,$$

el cual, es un estimador consistente de σ_e^2.

Cabe mencionar que, una vez estimado σ_e^2 y λ_0 podemos arribar a una estimación de $\sigma_e^2 \left[J_n^o S_n^{-1}(\lambda_0) S_n^{-1'}(\lambda_0) J_n^{o'} \right]$ y utilizarla para obtener una re-estimación de los parámetros del modelo (3.3) a través de Mínimos Cuadrados no Lineales Generalizados (GNLS) minimizando, en λ y β, la siguiente expresión, $\left[\left(J_n^o y_n - J_n^o S_n^{-1}(\lambda) X_n \beta \right)' \left(\sigma_e^2 \Sigma_n \right)^{-1} \left(J_n^o y_n - J_n^o S_n^{-1}(\lambda) X_n \beta \right) \right]$. Dicho estimador es una alternativa eficiente, cuya distribución asintótica está desarrollada en el trabajo de Wang & Lee.

4. Comportamiento de los estimadores ante muestras finitas

El comportamiento de los estimadores ante muestras finitas lo evaluaremos mediante un ejercicio de simulación Monte Carlo. Para poder realizar este ejercicio se generó un código[9] de cada uno de los estimadores bajo el entorno de R, software de uso libre y que puede obtenerse gratuitamente desde http://www.r-project.org/.

[9] El código está a disposición solicitándolo al autor.

4.1. Diseño del experimento

La especificación del modelo inicial es:

$$y = \lambda W y + \beta_0 + \beta_1 X_1 + e,$$

con $\beta_0 = \beta_1 = 1$, $\lambda = 0{,}4$, y X_1 y $e \sim \mathcal{N}(0, 1I_n)$.

Se han considerado diferentes tamaños muestrales:

- Pequeño: $n = 67$.
- Medio: $n = 217$.
- Grande: $n = 417$.

Para cada uno de los tamaños muestrales se han asumido distintas fracciones de pérdida de aproximadamente 10 %, 25 %, y 50 %. El número de réplicas ha sido establecido en 500.

La matriz de contactos W ha sido construida siguiendo el criterio de k vecinos más cercanos. Para ello, las coordenadas de cada uno de los datos espaciales han sido generadas de manera aleatoria asumiendo n pares de datos provenientes de dos distribuciones uniformes, $U(0, 1)$, que simulan latitud y longitud. Para cada réplica se ha obtenido una realización de estas coordenadas, estableciendo así, una determinada estructura para los n puntos de la réplica. Posteriormente, se ha creado la matriz W utilizando el criterio de k vecinos más cercanos (4 u 8, según el caso).

La idea detrás de la elección de dos niveles de vecindad reside en evaluar el comportamiento de los estimadores ante matrices de pesos con diferentes densidades[10]. Definiendo como densidad al número de entradas diferentes de cero sobre el total de posibles entradas, tenemos que, para matrices de 4 vecinos más cercanos con $n = 67$, la densidad es 0.060, con $n = 217$ es 0.018, y con $n = 417$ es 0.009, mientras que, para 8 vecinos más cercanos, con $n = 67$, $n = 217$ y $n = 417$, la misma es, 0.121, 0.037 y 0.019 respectivamente. Estos porcentajes son habituales en trabajos aplicados y no generan importantes problemas en la estimación del modelo.

El primer estimador considerado es el IBG2SLS de Wang & Lee, IBG2SLS (W-L), el cual imputa la totalidad de los datos faltantes (tanto $J^u y$, como los valores no observados de Wy), el segundo es nuestra versión de dicho estimador, identificado como IBG2SLS, que solo imputa los valores no observados de $J^o W y$, el tercero es la versión asintótica del IBG2SLS al que llamamos IBG2SLSA, y el último es el IG2SLS que utiliza la aproximación a la matriz óptima desarrollada por Kelejian et al. (2004), denominado como ISTE.

En lo que respecta al estimador ISTE, la matriz de instrumentos es $\hat{Q}_n^{okp} = \left[J_n^o \sum_{k=0}^{r_n} \hat{\lambda}_{nls}^k W_n^{k+1} X_n \hat{\beta}_{nls} \quad J_n^o X_n \right]$ la cual involucra una suma de términos de orden r_n, si bien en teoría $r_n \uparrow \infty$, en la práctica, utilizamos solo algunos sumandos.

Kelejian et al. (2004) presentan simulaciones donde prueban diferentes valores de r_n, y concluyen que, siendo $r_n = n^\alpha$, con $\alpha = 0{,}25$, los resultados son razonablemente buenos, con lo cual nosotros

[10] Existen diferentes trabajos donde queda en evidencia que cuando la densidad de las matrices es muy elevada, esto afecta la calidad de los test de hipótesis y genera cierto sesgo en los estimadores. Ver Smith, 2009; Farber et al., 2010, entre otros.

seguiremos ese criterio en esta simulación, lo que nos lleva a que (redondeando cifras), para $n = 67$ tenemos un $r_n = 3$, para $n = 217$ tenemos un $r_n = 4$, y para $n = 417$ tenemos un $r_n = 5$.

Por otro lado, si bien se ha demostrado que

$$\begin{aligned}
\Sigma_{iste} &= \sigma_e^2 \lim_{n\to\infty} \left[\frac{1}{n} C_n^{o'} \Omega_n^{-1}(\lambda_0) \hat{Q}_n^{okp} \left(\hat{Q}_n^{okp'} \Omega_n^{-1}(\lambda_0) \hat{Q}_n^{okp} \right)^{-1} \hat{Q}_n^{okp'} \Omega_n^{-1}(\lambda_0) C_n^{o} \right]^{-1}, \\
&= \sigma_e^2 \lim_{n\to\infty} \left(\frac{1}{n} C_n^{o'}, \Omega_n^{-1}(\lambda_0) C_n^{o} \right)^{-1},
\end{aligned}$$

es decir, que la varianza de la distribución asintótica del ITSE es igual a la de los otros dos estimadores, utilizaremos la primer ecuación para calcular el desvío (tanto el estimado como el verdadero) del ITSE para, de esta manera, corroborar dicha igualdad (ya que la segunda ecuación la tenemos presente para el desvío, estimado y verdadero, de los otros dos estimadores).

4.2. Resultados experimentales

En las tablas se presenta el sesgo de los estimadores de (λ, β_0, β_1, σ_e), la raíz del error cuadrático medio (RECM) y, excepto para el IBG2SLS (W-L)[11], el desvío estándar, tanto el verdadero como el estimado.

Los primeros resultados presentados de la simulación son los obtenidos para matrices de 4 vecinos más cercanos (Cuadros 1 y 2). En lo que respecta al comportamiento de los estimadores en relación al tamaño muestral (tanto el observado como el total), vemos el comportamiento que esperaríamos ver, vale decir, a medida que el tamaño muestral aumenta, tanto el sesgo como la varianza y el RCME se reducen, si bien esto no sucede en la totalidad de los casos, está presente en la mayoría de ellos, casi sistemáticamente.

Para $n = 67$, y todas sus fracciones de pérdida, el promedio del desvío estimado para el estimador de λ_0, supera el valor de 0.214, esto hace que, si tomamos dicho promedio, y pensamos en un intervalo de confianza del 95 % basado en su distribución asintótica, tenemos un intervalo cuya amplitud es $2 \times 0{,}214 \times 1{,}96 = 0{,}83$, la cual nos brinda muy poca precisión acerca del verdadero valor del parámetro de correlación espacial, claro está que, por lo dicho en el primer párrafo, dicha amplitud se va reduciendo a medida que el tamaño muestral aumenta.

En lo que respecta a la comparación entre estimadores, el IBG2SLS (W-L), el IBG2SLS y el ITSE, excepto para $n = 67$, donde presentan algunas diferencias con respecto a RECM y sesgo (aunque entre el IBG2SLS (W-L) y el IBG2SLS las diferencias de sesgo son menores), en el resto de los tamaños muestrales se comportan idénticamente.

[11] No presentamos el desvío del IBG2SLS (W-L) porque nuestro interés se centra en comparar el sesgo y la RECM de dicho estimador con respecto al IBG2SLS desarrollado en este trabajo.

Cuadro 1: Simulación con matriz de 4 vecinos más cercanos

n	n^o	Parámetros	IBG2SLS (W-L)		IBG2SLS		
			Sesgo	RECM	Sesgo	RECM	Desv.
67	33	λ_0	0.073	0.357	0.072	0.364	0.276 (0.263)
		β_0	-0.126	0.614	-0.124	0.626	0.483 (0.461)
		β_1	0.012	0.205	0.013	0.206	0.185 (0.190)
		σ_e	-	-	0.055	0.267	-
	51	λ_0	0.048	0.295	0.047	0.299	0.237 (0.228)
		β_0	-0.090	0.534	-0.088	0.542	0.424 (0.404)
		β_1	0.017	0.149	0.017	0.150	0.146 (0.148)
		σ_e	-	-	0.020	0.217	-
	61	λ_0	0.024	0.260	0.023	0.263	0.215 (0.212)
		β_0	-0.029	0.457	-0.028	0.462	0.381 (0.377)
		β_1	0.018	0.137	0.019	0.137	0.133 (0.133)
		σ_e	-	-	0.004	0.198	-
217	109	λ_0	0.009	0.141	0.008	0.141	0.137 (0.136)
		β_0	-0.014	0.248	-0.012	0.248	0.242 (0.242)
		β_1	0.004	0.100	0.005	0.100	0.100 (0.101)
		σ_e	-	-	0.015	0.146	-
	163	λ_0	0.005	0.120	0.004	0.120	0.118 (0.117)
		β_0	-0.009	0.214	-0.008	0.214	0.210 (0.208)
		β_1	0.007	0.083	0.007	0.083	0.080 (0.080)
		σ_e	-	-	0.001	0.121	-
	195	λ_0	-0.003	0.109	-0.003	0.109	0.111 (0.110)
		β_0	0.007	0.195	0.007	0.195	0.198 (0.196)
		β_1	0.010	0.075	0.010	0.075	0.073 (0.073)
		σ_e	-	-	-0.019	0.108	-
417	209	λ_0	0.004	0.098	0.003	0.098	0.097 (0.097)
		β_0	-0.008	0.165	-0.007	0.165	0.171 (0.170)
		β_1	0.000	0.075	0.001	0.075	0.072 (0.073)
		σ_e	-	-	0.007	0.103	-
	313	λ_0	0.005	0.089	0.005	0.089	0.083 (0.083)
		β_0	-0.007	0.159	-0.006	0.159	0.149 (0.148)
		β_1	-0.000	0.059	-0.000	0.059	0.058 (0.058)
		σ_e	-	-	-0.009	0.087	-
	376	λ_0	0.002	0.078	0.002	0.079	0.078 (0.078)
		β_0	-0.006	0.139	-0.005	0.139	0.140 (0.139)
		β_1	0.000	0.056	0.000	0.056	0.053 (0.052)
		σ_e	-	-	-0.017	0.078	-

Nota: entre paréntesis se presenta el desvío verdadero.

Para $n = 67$ el estimador que mejor se comporta en términos de RECM es el IBG2SLSA, el cual presenta los menores valores de dicha medida para todas las fracciones de pérdida, siendo que, para valores muestrales mayores, en la mayoría de los casos, se ubica ligeramente por sobre el resto.

En lo que respecta a sesgo, no es claro que algún estimador sea mejor para un tamaño muestral de $n = 67$ y todas sus fracciones de pérdida.

Universidad de San Andrés

Cuadro 2: Simulación con matriz de 4 vecinos más cercanos (cont.)

n	n^o	Parámetros	IBG2SLSA			ISTE		
			Sesgo	RECM	Desv.	Sesgo	RECM	Desv.
67	33	λ_0	0.072	0.338	0.276 (0.263)	0.069	0.352	0.277 (0.263)
		β_0	-0.124	0.582	0.483 (0.461)	-0.119	0.607	0.484 (0.461)
		β_1	0.008	0.201	0.185 (0.190)	0.012	0.204	0.185 (0.190)
		σ_e	0.055	0.267	-	0.055	0.267	-
	51	λ_0	0.047	0.274	0.237 (0.228)	0.048	0.301	0.238 (0.228)
		β_0	-0.088	0.499	0.424 (0.404)	0.090	0.546	0.425 (0.404)
		β_1	0.015	0.148	0.146 (0.148)	0.017	0.150	0.146 (0.148)
		σ_e	0.020	0.217	-	0.020	0.217	-
	61	λ_0	0.023	0.241	0.215 (0.212)	0.025	0.264	0.215 (0.212)
		β_0	-0.031	0.425	0.381 (0.377)	-0.031	0.464	0.382 (0.377)
		β_1	0.018	0.134	0.133 (0.133)	0.018	0.137	0.133 (0.133)
		σ_e	0.004	0.198	-	0.004	0.198	-
217	109	λ_0	0.008	0.140	0.137 (0.136)	0.009	0.141	0.137 (0.136)
		β_0	-0.012	0.246	0.242 (0.242)	-0.013	0.248	0.243 (0.242)
		β_1	0.005	0.101	0.100 (0.101)	0.004	0.100	0.100 (0.101)
		σ_e	0.015	0.146	-	0.015	0.146	-
	163	λ_0	0.004	0.122	0.118 (0.117)	0.004	0.120	0.118 (0.117)
		β_0	-0.008	0.217	0.210 (0.208)	-0.008	0.214	0.210 (0.208)
		β_1	0.008	0.085	0.080 (0.080)	0.007	0.083	0.080 (0.080)
		σ_e	0.001	0.121	-	0.001	0.121	-
	195	λ_0	-0.001	0.110	0.111 (0.110)	-0.003	0.109	0.111 (0.110)
		β_0	0.005	0.199	0.198 (0.196)	0.007	0.195	0.198 (0.196)
		β_1	0.010	0.075	0.073 (0.073)	0.010	0.075	0.073 (0.073)
		σ_e	-0.019	0.108	-	-0.019	0.108	-
417	209	λ_0	0.005	0.098	0.097 (0.097)	0.004	0.098	0.097 (0.097)
		β_0	-0.009	0.167	0.171 (0.170)	-0.007	0.165	0.171 (0.170)
		β_1	0.001	0.076	0.072 (0.073)	0.001	0.075	0.072 (0.073)
		σ_e	0.007	0.103	-	0.007	0.103	-
	313	λ_0	0.005	0.090	0.083 (0.083)	0.005	0.089	0.083 (0.083)
		β_0	-0.008	0.161	0.149 (0.148)	-0.006	0.159	0.149 (0.148)
		β_1	-0.000	0.060	0.058 (0.058)	-0.000	0.059	0.058 (0.058)
		σ_e	-0.009	0.087	-	-0.009	0.087	-
	376	λ_0	0.003	0.079	0.078 (0.078)	0.002	0.078	0.078 (0.078)
		β_0	-0.006	0.140	0.140 (0.139)	-0.005	0.139	0.140 (0.139)
		β_1	0.000	0.056	0.053 (0.052)	0.000	0.056	0.053 (0.052)
		σ_e	-0.017	0.078	-	-0.017	0.078	-

Nota: entre paréntesis se presenta el desvío verdadero.

En los Cuadros 3 y 4 se presentan los resultados de la simulación con matrices de 8 vecinos más cercanos. En este caso, al igual que en el anterior, vemos que al aumentar el tamaño muestral, tanto el sesgo como la RECM tienden a reducirse, decimos tienden porque esto no se ve para todos los tamaños muestrales, por ejemplo, para $n^o = 109$ los sesgos de λ_0 y β_0 son menores que para $n^o = 163$ para todos los estimadores.

Universidad de San Andrés

Los promedios de los desvíos estimados para $n = 67$, al igual que en la simulación anterior, conducen a intervalos de confianza excesivamente amplios.

Cuadro 3: Simulación con matriz de 8 vecinos más cercanos.

n	n^o	Parámetros	IBG2SLS (W-L)		IBG2SLS		
			Sesgo	RECM	Sesgo	RECM	Desv.
67	33	λ_0	0.162	0.627	0.177	0.722	0.427 (0.402)
		β_0	-0.285	1.084	-0.304	1.170	0.738 (0.691)
		β_1	0.026	0.204	0.028	0.210	0.178 (0.185)
		σ_e	-	-	0.090	0.259	-
	51	λ_0	0.114	0.491	0.109	0.476	0.359 (0.339)
		β_0	-0.203	0.843	-0.197	0.829	0.623 (0.584)
		β_1	0.022	0.151	0.022	0.150	0.143 (0.146)
		σ_e	-	-	0.049	0.206	-
	61	λ_0	0.067	0.417	0.067	0.420	0.320 (0.314)
		β_0	-0.102	0.685	-0.102	0.687	0.548 (0.540)
		β_1	0.003	0.144	0.003	0.144	0.131 (0.133)
		σ_e	-	-	0.025	0.194	-
217	109	λ_0	0.029	0.220	0.028	0.221	0.195 (0.190)
		β_0	-0.047	0.380	-0.048	0.381	0.338 (0.328)
		β_1	0.003	0.097	0.004	0.097	0.098 (0.099)
		σ_e	-	-	0.018	0.145	-
	163	λ_0	0.036	0.189	0.035	0.189	0.168 (0.163)
		β_0	-0.066	0.331	-0.065	0.332	0.291 (0.282)
		β_1	-0.000	0.078	-0.000	0.078	0.078 (0.078)
		σ_e	-	-	-0.002	0.116	-
	195	λ_0	0.014	0.168	0.014	0.168	0.157 (0.154)
		β_0	-0.017	0.281	-0.016	0.281	0.270 (0.267)
		β_1	0.005	0.073	0.005	0.073	0.072 (0.072)
		σ_e	-	-	-0.004	0.104	-
417	209	λ_0	0.014	0.143	0.013	0.143	0.135 (0.133)
		β_0	-0.024	0.245	-0.023	0.245	0.232 (0.230)
		β_1	-0.001	0.073	-0.002	0.073	0.070 (0.071)
		σ_e	-	-	0.014	0.100	-
	313	λ_0	0.003	0.118	0.002	0.118	0.115 (0.115)
		β_0	-0.007	0.207	-0.007	0.207	0.200 (0.200)
		β_1	0.001	0.060	0.001	0.060	0.057 (0.057)
		σ_e	-	-	0.006	0.082	-
	376	λ_0	0.006	0.121	0.006	0.121	0.109 (0.109)
		β_0	-0.009	0.209	-0.008	0.209	0.189 (0.188)
		β_1	0.003	0.052	0.003	0.052	0.052 (0.052)
		σ_e	-	-	-0.000	0.071	-

Nota: entre paréntesis se presenta el desvío verdadero

Universidad de San Andrés

Cuadro 4: Simulación con matriz de 8 vecinos más cercanos (cont.)

n	n^o	Parámetros	IBG2SLSA			ISTE		
			Sesgo	RECM	Desv.	Sesgo	RECM	Desv.
67	33	λ_0	0.143	0.549	0.427 (0.402)	0.176	0.966	0.438 (0.402)
		β_0	-0.255	0.973	0.738 (0.691)	-0.298	1.468	0.756 (0.691)
		β_1	0.026	0.199	0.178 (0.185)	0.032	0.231	0.179 (0.185)
		σ_e	0.090	0.259	-	0.090	0.259	-
	51	λ_0	0.100	0.417	0.359 (0.339)	0.129	0.704	0.360 (0.339)
		β_0	-0.185	0.755	0.623 (0.584)	-0.223	1.060	0.626 (0.584)
		β_1	0.020	0.147	0.143 (0.146)	0.024	0.165	0.143 (0.146)
		σ_e	0.049	0.206	-	0.049	0.206	-
	61	λ_0	0.056	0.355	0.320 (0.314)	0.070	0.436	0.321 (0.315)
		β_0	-0.090	0.590	0.548 (0.540)	-0.108	0.716	0.550 (0.540)
		β_1	0.002	0.142	0.131 (0.133)	0.003	0.144	0.131 (0.133)
		σ_e	0.025	0.194	-	0.025	0.194	-
217	109	λ_0	0.030	0.218	0.195 (0.190)	0.029	0.220	0.195 (0.190)
		β_0	-0.051	0.377	0.338 (0.328)	-0.048	0.380	0.338 (0.328)
		β_1	0.003	0.097	0.098 (0.099)	0.003	0.097	0.098 (0.099)
		σ_e	0.018	0.145	-	0.018	0.145	-
	163	λ_0	0.035	0.188	0.168 (0.163)	0.035	0.189	0.168 (0.163)
		β_0	-0.064	0.330	0.291 (0.282)	-0.065	0.332	0.291 (0.282)
		β_1	-0.001	0.078	0.078 (0.078)	-0.000	0.078	0.078 (0.078)
		σ_e	-0.002	0.116	-	-0.002	0.116	-
	195	λ_0	0.016	0.168	0.157 (0.154)	0.014	0.168	0.157 (0.154)
		β_0	-0.021	0.282	0.270 (0.267)	-0.017	0.281	0.270 (0.267)
		β_1	0.005	0.074	0.072 (0.072)	0.005	0.073	0.072 (0.072)
		σ_e	-0.004	0.104	-	-0.004	0.104	-
417	209	λ_0	0.014	0.142	0.135 (0.133)	0.013	0.143	0.135 (0.133)
		β_0	-0.024	0.245	0.232 (0.230)	-0.023	0.245	0.232 (0.230)
		β_1	-0.001	0.074	0.070 (0.071)	-0.002	0.073	0.070 (0.071)
		σ_e	0.014	0.100	-	0.014	0.100	-
	313	λ_0	0.004	0.118	0.115 (0.115)	0.003	0.118	0.115 (0.115)
		β_0	-0.009	0.208	0.200 (0.200)	-0.007	0.207	0.200 (0.200)
		β_1	0.000	0.060	0.057 (0.057)	0.001	0.060	0.057 (0.057)
		σ_e	0.006	0.082	-	0.006	0.082	-
	376	λ_0	0.009	0.122	0.109 (0.109)	0.006	0.121	0.109 (0.109)
		β_0	-0.013	0.210	0.189 (0.188)	-0.009	0.209	0.189 (0.188)
		β_1	0.003	0.052	0.052 (0.052)	0.003	0.052	0.052 (0.052)
		σ_e	-0.000	0.071	-	-0.000	0.071	-

Nota: entre paréntesis se presenta el desvío verdadero

Comparando ambas simulaciones, lo que vemos es que, con una matriz mas densa, el sesgo y la RECM aumentan fuertemente, aunque los sesgos de ambas simulaciones parecieran ir acercándose a medida que el tamaño muestral aumenta, no así la RECM, la cual mantiene sus diferencias para todos los tamaños muestrales simulados.

Un elemento a considerar, viendo lo que ambas simulaciones nos arrojan, es que, pareciera ser más importante, para la precisión de los estimadores, el tamaño muestral observado que el total, ya que, por ejemplo, vemos que para $n = 417$ con $n^o = 209$, su desvío promedio (para todos los

estimadores) es menor que para $n = 217$ con $n^o = 195$, si bien en el primer caso tenemos 14 datos más que para el segundo, también tenemos una pérdida del 50 % contra el 10 % del segundo.

En lo que respecta a los estimadores, para el menor tamaño muestral y todas sus fracciones de pérdida, el IBG2SLSA se comporta mejor que el resto, tanto en términos de sesgo, como de RECM, para el resto de los tamaños muestrales no hay ningún estimador que sistemáticamente sea mejor que los demás, como así tampoco, no hay ninguno se encuentre claramente por debajo del resto.

Para $n = 217$ y $n = 417$, en todas su fracciones de pérdida, los estimadores IBG2SLS (W-L), IBG2SLS y el ITSE presentan prácticamente los mismos valores de sesgo y RECM, y en los pocos casos en que se ven diferencias, la misma no supera el 0.001 en valor absoluto, también se ve que sus discrepancias con el IBG2SLSA, en términos de RECM, son menores.

Al igual que en la simulación anterior, a medida que aumenta el tamaño muestral el desvío estimado de todos los estimadores se aproxima a su valor verdadero, y también, a medida que se reduce el sesgo se aproximan a la RECM, como es de esperar.

Los desvíos del ITSE, tanto los estimados como los verdaderos, para ambas simulaciones y para tamaños muestrales de moderados a grandes, se aproximan fuertemente al de los otros dos estimadores, corroborando así su igualdad asintótica, y por lo tanto, su eficiencia.

5. Aplicación empírica: Tasa de homicidios en Estados Unidos

En esta sección utilizaremos el herramental desarrollado en las secciones previas para replicar parcialmente los resultados de Baller et al. (2001), "Structural Covariates of U.S. County Homicide Rates: Incorporating Spatial Effects", a fin de evaluar empíricamente los estimadores propuestos. Los datos originales pueden descargarse gratuitamente de http://geodacenter.asu.edu/sdata.

Baller et al. (2001) examinan el impacto de ciertas variables estructurales sobre la tasa de homicidio en Estados Unidos para cuatro momentos diferentes del tiempo (1960, 1970, 1980, 1990). Las unidades observacionales son los condados norteamericanos (3085 regiones), clasificados entre la zona sur (1412) y el resto (1673) debido a razones de heterogeneidad. Entre los modelos propuestos se encuentra el SLM aplicado únicamente a los condados de la región sur de Estados Unidos para cada uno de los años mencionados.

En nuestro caso, replicaremos la estimación SLM para los condados del sur de U.S. considerando únicamente el periodo 1960. En el trabajo original, se utilizó la metodología de 2SLS instrumentalizando Wy mediante WX, por lo que el SLM no fue estimado con la matriz de instrumentos óptima. Debido a que nuestros estimadores para datos faltantes utilizan dicha matriz, se ha re-estimado el modelo con datos completos utilizándola, como así también su aproximación. Por otra parte, cabe aclarar que la intención de este ejercicio es comparar las estimaciones con datos completos y con datos faltante sin analizar en profundidad los resultados finales del trabajo.

Siguiendo a Baller et al., 2001, el modelo a estimar es el siguiente:

$$y = \lambda W y + X\beta + e,$$

con $X = (X_0\, X_1\, X_2\, X_3\, X_4\, X_5)$, $\beta' = (\beta_0\, \beta_1\, \beta_2\, \beta_3\, \beta_3\, \beta_5)$, λ parámetro de correlación espacial, W matriz de 10 vecinos más cercanos, y $e \sim \left(0,\, \sigma_e^2\right)$.

Las variables implicadas en el modelo son:

y= tasa de homicidio.

X_0= vector de unos.

X_1= es una variable representada por un índice de componentes principales que contiene, porcentaje de personas de color, el logaritmo de la media del ingreso familiar, el índice de Gini para la desigualdad del ingreso familiar, porcentaje de familias por debajo de la línea de pobreza, porcentaje de familias donde el jefe de hogar es mujer.

X_2= al igual que la anterior, es una variable representada por un índice de componentes principales que incluye, el logaritmo del tamaño poblacional, el logaritmo de la densidad poblacional.

X_3= tasa de desempleo.

X_4= porcentaje de mujeres mayores de 14 años que están divorciadas.

X_5= promedio de edad.

Para replicar la estimación del SLM, año 1960, haremos uso de tres matrices de instrumentos diferentes, por un lado $(WX\ \ X_0\ \ X_1 \ldots)$, que es el que se utiliza en el trabajo original, por otro, la matriz óptima,$\left(W(I-\lambda)^{-1}X\beta\ \ X_0\ \ X_1 \ldots\right)$, y en tercer lugar la aproximación a la matriz óptima, $\left(\sum_{k=0}^{r_n} \lambda^k W_n^{k+1} X_n \beta\ \ X_0\ \ X_1 \ldots\right)$. El uso de estas matrices de instrumentos nos conduce a tres estimadores diferentes, el primero es el que conocemos como 2SLS, el segundo, que utiliza la matriz óptima, siguiendo a Lee (2003), lo llamaremos "Best Feasible Spatial 2SLS" (BFS2SLS), y al tercero, que utiliza la aproximación de dicha matriz, siguiendo a Kelejian et al. (2004), lo llamaremos "Best Series Feasible Spatial 2SLS" (BSFS2SLS)[12].

Como vemos, la matriz de instrumentos óptima, así como su aproximación, requieren de la estimación de λ y β para ser factibles, dichos parámetros los hemos estimado aplicando la metodología de mínimos cuadrados no lineales. Para la aproximación de la matriz de instrumentos óptima se ha considerado $r_n = 6 \cong 1412^{0,25}$. La pérdida de datos en la variable explicada fue realizada al azar tomando en consideración los diferentes porcentajes de pérdida.

Para tener una medida de la bondad de ajuste, al igual que en el trabajo original, se ha calculado la correlación al cuadrado de las y´s estimadas con las y´s observadas. En el caso de los modelos con datos faltantes, se calculó la correlación al cuadrado para el total de y´s (observadas y no observadas) contra las estimaciones de todas las y´s. Los resultados obtenidos se presentan en el Cuadro 5.

Lo primero que podemos observar es que no hay grandes diferencias entre los estimadores estimados con datos completos (2SLS, BFS2SLS, y BSFS2SLS). Las variables significativas son las mismas bajo las tres metodologías de estimación y estos resultados son similares a los obtenidos por Baller et al..

Ante una pérdida aleatoria del 10 % a los datos, el valor de los estimadores no varía significativamente, aunque la ordenada al origen dejó de ser significativa al 5 %, siéndolo al 10 %, de la misma forma, cuando la pérdida pasó a ser del 25 % tampoco se ven grandes diferencias, las mayores están presentes en los estimadores que nunca fueron significativos, pero, a diferencia del

[12]Para más referencia sobre los estimadores ver Kelejian & Prucha (1998, 1999); Lee (2003) y Kelejian et al. (2004).

caso anterior, la ordenada al origen resultó ser significativa al 5 % para el estimador IBG2SLSA, y al 10 % para el IBG2SLS.

Cuadro 5: SLM con datos completos y con pérdida del 10 %

Parámetros	$n = 1412$			$n = 1271$		
	2SLS	BFS2SLS	BSFS2SLS	IBG2SLS	IBG2SLSA	ITSE
λ	0.742***	0.768***	0.758***	0.769***	0.770***	0.767***
	(6.124)	(7.191)	(7.102)	(6.656)	(6.656)	(6.527)
β_0	3.781**	3.457**	3.578**	3.375	3.365	3.407
	(2.008)	(1.968)	(2.037)	(1.783)	(1.777)	(1.781)
β_1	0.823***	0.790***	0.802***	0.803***	0.773***	0.806***
	(3.358)	(3.395)	(3.447)	(3.248)	(3.130)	(3.241)
β_2	-0.072	-0.084	-0.080	-0.080	0.023	-0.079
	(-0.335)	(-0.402)	(-0.379)	(-0.366)	(0.105)	(-0.360)
β_3	-0.055	-0.047	-0.050	-0.068	-0.100	-0.069
	(-0.706)	(-0.625)	(-0.665)	(-0.833)	(-1.217)	(-0.841)
β_4	0.799***	0.786***	0.791***	0.713***	0.711***	0.714***
	(3.297)	(3.247)	(3.267)	(2.767)	(2.760)	(2.770)
β_5	-0.128***	-0.123***	-0.125***	-0.112**	-0.103**	-0.112**
	(-2.927)	(-2.857)	(-2.901)	(-2.467)	(-2.278)	(-2.466)
$Corr^2$	0.169	0.168	0.168	0.162	0.161	0.162
$Corr^{2*}$	-	-	-	0.168	0.167	0.168

Nota: $Corr^2$es la correlación al cuadrado entre las y´s estimadas y las y´s observadas. $Corr^{2*}$ es la correlación entre las y´s estimadas para las 1412 observaciones y todas las y´s (observadas y no observadas). ** significativo al 5 %; *** significativo al 1 %.

Cuadro 6: SLM con pérdida del 25 y 50 %

Parámetros	$n = 1059$			$n = 705$		
	IBG2SLS	IBG2SLSA	ITSE	IBG2SLS	IBG2SLSA	ITSE
λ	0.775***	0.744***	0.792***	0.940***	0.916***	0.879***
	(7.720)	(7.409)	(7.684)	(11.878)	(11.585)	(6.651)
β_0	2.738	3.250**	2.530	0.901	0.581	1.714
	(1.700)	(2.018)	(1.547)	(0.565)	(0.364)	(0.786)
β_1	0.880***	0.943***	0.850***	0.509**	0.379	0.577**
	(3.468)	(3.718)	(3.310)	(2.287)	(1.703)	(2.213)
β_2	0.228	0.364	0.213	0.307	0.015	0.322
	(1.044)	(1.663)	(0.972)	(1.276)	(0.064)	(1.299)
β_3	-0.066	-0.115	-0.059	-0.066	0.009	-0.084
	(-0.876)	(-1.537)	(-0.778)	(-0.815)	(0.112)	(-0.928)
β_4	0.835***	0.715***	0.825***	0.435	0.412	0.446
	(3.278)	(2.806)	(3.228)	(1.634)	(1.548)	(1.670)
β_5	-0.102**	-0.096**	-0.098**	-0.044	-0.035	-0.057
	(-2.465)	(-2.338)	(-2.373)	(-1.044)	(-0.832)	(-1.474)
$Corr^2$	0.197	0.198	0.196	0.188	0.182	0.191
$Corr^{2*}$	0.167	0.166	0.166	0.155	0.154	0.158

Nota: $Corr^2$ y $Corr^{2*}$ han sido definidas en el Cuadro 5. ** significativo al 5 %; *** significativo al 1 %.

Universidad de San Andrés

Ahora bien, cuando pasamos a una pérdida de alrededor del 50 %, los estimadores presentan considerables diferencias y muchos parámetros pasan a ser no significativos o reducen su nivel de significancia, el parámetro de correlación espacial mantiene su nivel al 1 % (aunque bastante sobrestimado), β_1 es significativo al 5 % para el IBG2SLS y para el ITSE, y al 10 % para el IBG2SLSA, siendo que para los anteriores tamaños muestrales lo era al 1 % para todos ellos, de igual manera β_4 era significativo al 1 % para todos los tamaños muestrales y ahora lo es al 10 % y solo para el IBG2SLS y el ITSE.

En lo que respecta al grado de correlación entre las observaciones y las estimaciones ($Corr^2$) vemos que a medida que estimamos con menos datos dicha correlación mejora, esto puede deberse a la presencia de valores atípicos en la muestra total que hayan quedado afuera en el proceso de pérdida. Respecto a la correlación de las estimaciones para los 1412 datos versus los valores verdaderos, vemos que dicha correlación se mantiene relativamente parecida entre todos los escenarios y esto refuerza la hipótesis anterior.

6. Conclusión

El presente trabajo muestra un conjunto de estimadores consistentes ante la pérdida (aleatoria) de observaciones en la variable dependiente en un modelo de rezago espacial (SLM). El eje principal de nuestra propuesta, gira en torno al estimador de Wang & Lee (2013), I2SLS (y sus variantes). La diferencia básica entre dicho estimador y el que se ha desarrollado se centra en la imputación de los datos faltantes. Puntualmente, nuestra versión, solo imputa ciertos rezagos espaciales no observados de las y´s, mientras que, el de Wang & Lee, imputa todas las y´s no observadas.

Si bien no hemos probado analíticamente la equivalencia asintótica de ambos estimadores, y siendo conscientes de lo limitado de las simulaciones, hemos visto que, para muestras de moderadas a grandes, tanto en sesgo como en RECM, ambos estimadores presentan idéntico comportamiento, y para muestras pequeñas, no se ven diferencias sistemáticas que permitan concluir que uno sea mejor que otro. Además, dichas conclusiones se extienden también a los otros estimadores (ITSE y IBG2SLSA). También hemos podido corroborar que, como se demostró analíticamente, a medida que el tamaño muestral aumenta la varianza del ITSE y del IBG2SLS se igualan, quedando en evidencia la eficiencia asintótica del estimador basado en la aproximación de la matriz de instrumentos óptima.

Por otro lado, en términos aplicados, es importante considerar que la estimación inicial de $\theta = \begin{pmatrix} \lambda_0 & \beta_0' \end{pmatrix}$ a través de NLS, necesaria para llevar adelante la imputación, es altamente demandante en términos computacionales, con lo cual, la posibilidad de obtener una estimación inicial (consistente) que sea computacionalmente más sencilla, mejoraría considerablemente la aplicabilidad de los estimadores propuestos.

Referencias

Anselin, L. (1988). *Spatial econometrics: Methods and models*, volume 4. Boston: Kluwer. Academic Publishers.

Anselin, L. (2005). Spatial regression analysis in R. A workbook. *Urbana*, 51, 61801.

Anselin, L. & Florax, R. (1995). *New directions in spatial econometrics.* Springer.

Baller, R., Anselin, L., Messner, S., Deane, G., & Hawkins, D. (2001). Structural covariates of us county homicide rates: incorporating spatial effects. *Criminology*, 39(3), 561–588.

Baltagi, B., Ed. (2001). *A companion to theoretical econometrics.* Blackwell, Oxford (UK).

Baraldi, A. & Enders, C. (2010). An introduction to modern missing data analyses. *Journal of School Psychology*, 48(1), 5–37.

Billingsley, P. (2008). *Probability and measure.* New York: John Wiley & Sons.

Conniffe, D. (1983). Comments on the weighted regression approach to missing values. *Economic and Social Research Institute*, 14(4), 259–272.

Cressie, N. (1993). *Statistics for Spatial Data.* New York: John Wiley & Sons.

Farber, S., Páez, A., & Volz, E. (2010). Topology, dependency tests and estimation bias in network autoregressive models. In *Progress in Spatial Analysis.* Springer.

Florax, R. & Van der Vlist, A. (2003). Spatial econometric data analysis: Moving beyond traditional models. *International Regional Science Review*, 26(3), 223–243.

Fujita, M., Krugman, P., & Venables, A. (1999). *The Spatial Economics: Cities, Regions and International Trade.* Cambridge, MA: MIT Press.

Kelejian, H. & Prucha, I. (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *The Journal of Real Estate Finance and Economics*, 17(1), 99–121.

Kelejian, H. & Prucha, I. (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. *International economic review*, 40(2), 509–533.

Kelejian, H. & Prucha, I. (2007). The relative efficiencies of various predictors in spatial econometric models containing spatial lags. *Regional Science and Urban Economics*, 37(3), 363–374.

Kelejian, H. & Prucha, I. (2010). Spatial models with spatially lagged dependent variables and incomplete data. *Journal of geographical systems*, 12(3), 241–257.

Kelejian, H., Prucha, I., & Yuzefovich, Y. (2004). Instrumental variable estimation of a spatial autoregressive model with autoregressive disturbances: Large and small sample results. *Advances in econometrics*, 18, 163–198.

Krugman, P. (1991). *Geography and trade.* Cambridge, MIT Press.

Lee, L.-F. (2003). Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances. *Econometric Reviews*, 22(4), 307–335.

Universidad de San Andrés

LeSage, J. & Pace, R. (2004). Models for spatially dependent missing data. *The Journal of Real Estate Finance and Economics*, 29(2), 233–254.

Little, R. (1992). Regression with missing x's: A review. *Journal of the American Statistical Association*, 87(420), 1227–1237.

Little, R. & Rubin, D. (1987). *Statistical Analysis with Missing Data.* New York: John Wiley & Sons.

Pigott, T. (2001). A review of methods for missing data. *Educational research and evaluation*, 7(4), 353–383.

Shao, J. (2013). Estimation and imputation in linear regression with missing values in both response and covariate. *Statistics and its interface*, 6(3), 361–368.

Smith, T. (2009). Estimation bias in spatial models with strongly connected weight matrices. *Geographical Analysis*, 41(3), 307–332.

Von Hippel, P. (2007). Regression with missing ys: An improved strategy for analyzing multiply imputed data. *Sociological Methodology*, 37(1), 83–117.

Wang, W. & Lee, L. F. (2013). Estimation of spatial autoregressive models with randomly missing data in the dependent variable. *The Econometrics Journal*, 16(1), 73–102.

White, H. (1984). *Asymptotic Theory for Econometricians.* New York: Academic Press.

Woolridge, J. (2002). *Econometric analysis of cross section and panel data.* Cambridge: MIT Press.

Universidad de
SanAndrés

7. Apéndice

A. Desarrollo del modelo con datos faltantes

En este apartado desarrollaremos las deducciones de algunas de las igualdades utilizadas en la subsección 2.2.

Sea la siguiente igualdad

$$\begin{aligned} S_n(\lambda_0) - S_n\left(\hat{\lambda}\right) &= (I_n - \lambda_0 W_n) - \left(I_n - \hat{\lambda} W_n\right), \\ &= \left(\hat{\lambda} - \lambda_0\right) W_n. \end{aligned}$$

Utilizando la igualdad anterior podemos expresar lo siguiente

$$\begin{aligned} \left[S_n^{-1}\left(\hat{\lambda}\right) - S_n^{-1}(\lambda_0)\right] &= S_n^{-1}\left(\hat{\lambda}\right)\left[S_n(\lambda_0) - S_n\left(\hat{\lambda}\right)\right] S_n^{-1}(\lambda_0), \\ &= S_n^{-1}\left(\hat{\lambda}\right)\left(\hat{\lambda} - \lambda_0\right) W_n S_n^{-1}(\lambda_0). \end{aligned}$$

Dado que $W_n S_n^{-1}(\lambda_0) = G_n$ y considerando que $\left(\hat{\lambda} - \lambda_0\right)$ es un escalar, tenemos

$$\left[S_n^{-1}\left(\hat{\lambda}\right) - S_n^{-1}(\lambda_0)\right] = \left(\hat{\lambda} - \lambda_0\right) S_n^{-1}\left(\hat{\lambda}\right) G_n.$$

Con todo lo anterior podemos arribar a lo siguiente

$$\begin{aligned} \left[S_n^{-1}\left(\hat{\lambda}\right) X_n\hat{\beta} - S_n^{-1}(\lambda_0) X_n\beta_0\right] &= \left[\left(\hat{\lambda} - \lambda_0\right) S_n^{-1}\left(\hat{\lambda}\right) G_n\right] X_n\hat{\beta} + S_n^{-1}(\lambda_0) X_n\left(\hat{\beta} - \beta_0\right) \quad (7.1) \\ &= \left[S_n^{-1}\left(\hat{\lambda}\right) G_n X_n\hat{\beta} \quad S_n^{-1}(\lambda_0) X_n\right] \begin{bmatrix} \left(\hat{\lambda} - \lambda\right) \\ \left(\hat{\beta} - \beta\right) \end{bmatrix}, \quad (7.2) \\ &= S_n^{-1}(\lambda_0) C_n\left(\hat{\theta} - \theta\right) + R\left(\hat{\lambda} - \lambda\right), \quad (7.3) \end{aligned}$$

con $\left(\hat{\theta} - \theta\right) = \begin{bmatrix} \left(\hat{\lambda} - \lambda\right) \\ \left(\hat{\beta} - \beta\right) \end{bmatrix}$; $C_n = (G_n X_n\beta_0 \quad X_n)$; $R = \left[S_n^{-1}\left(\hat{\lambda}\right) G_n X_n\hat{\beta} - S_n^{-1}(\lambda_0) G_n X_n\beta_0\right]$.

Si desarrollamos la ecuación (7.3) tenemos que

$$\begin{aligned} S_n^{-1}(\lambda_0) C_n\left(\hat{\theta} - \theta\right) + R\left(\hat{\lambda} - \lambda\right) &= S_n^{-1}(\lambda_0)\left[G_n X_n\beta_0\left(\hat{\lambda} - \lambda\right) + X_n\left(\hat{\beta} - \beta\right)\right] \\ &\quad + S_n^{-1}\left(\hat{\lambda}\right) G_n X_n\hat{\beta}\left(\hat{\lambda} - \lambda\right) - S_n^{-1}(\lambda_0) G_n X_n\beta_0\left(\hat{\lambda} - \lambda\right), \\ &= S_n^{-1}\left(\hat{\lambda}\right) G_n X_n\hat{\beta}\left(\hat{\lambda} - \lambda\right) + S_n^{-1}(\lambda_0) X_n\left(\hat{\beta} - \beta\right). \quad (7.4) \end{aligned}$$

Es fácil ver que (7.4) (que es el desarrollo de (7.3)) es igual a (7.2), y de ahí a (7.1).

B. Estimador I2SLS

B.1. Prueba proposición 1: Consistencia

En la siguiente igualdad introduciremos el término $\frac{1}{n}$ el cual no afectan la igualdad, ya que se cancela mutuamente, pero es útil para facilitar la comprensión de los requerimientos sobre el comportamiento de los términos en el límite.

Sea

$$\hat{\theta}_{2sls,n} - \theta_0 \quad = \quad \left[\frac{1}{n}\tilde{Z}_n^{o'} Q_n^o \left(\frac{1}{n}Q_n^{o'} Q_n^o\right)^{-1} \frac{1}{n}Q_n^{o'} \tilde{Z}_n^o\right]^{-1} \frac{1}{n}\tilde{Z}_n^{o'} Q_n^o \left(\frac{1}{n}Q_n^{o'} Q_n^o\right)^{-1} \frac{1}{n}Q_n^{o'} y_n^o - \theta_0,$$

si reemplazamos $y_n^o = \tilde{Z}_n^o \theta_0 + \tilde{u}_n^o$, con $\tilde{u}_n^o = J_n^o \tilde{u}_n$, tenemos que $\hat{\theta}_{2sls,n} - \theta_0 = \left[\frac{1}{n}\tilde{Z}_n^{o'} Q_n^o \left(\frac{1}{n}Q_n^{o'} Q_n^o\right)^{-1} \frac{1}{n}Q_n^{o'} \tilde{Z}_n^o\right]^{-1} \frac{1}{n}\tilde{Z}_n^{o'} Q_n^o \left(\frac{1}{n}Q_n^{o'} Q_n^o\right)^{-1} \frac{1}{n}Q_n^{o'} \tilde{u}_n^o$.

Para probar consistencia, debemos demostrar que $\left[\frac{1}{n}\tilde{Z}_n^{o'} Q_n^o \left(\frac{1}{n}Q_n^{o'} Q_n^o\right)^{-1} \frac{1}{n}Q_n^{o'} \tilde{Z}_n^o\right]^{-1}$ converge a un límite bien definido y de rango completo.

Sea

$$lim_{n\to\infty} \left[\frac{1}{n}\tilde{Z}_n^{o'} Q_n^o \left(\frac{1}{n}Q_n^{o'} Q_n^o\right)^{-1} \frac{1}{n}Q_n^{o'} \tilde{Z}_n^o\right]^{-1} = \left[\left(lim_{n\to\infty}\frac{1}{n}\tilde{Z}_n^{o'} Q_n^o\right) \left(lim_{n\to\infty}\frac{1}{n}Q_n^{o'} Q_n^o\right)^{-1} \left(lim_{n\to\infty}\frac{1}{n}Q_n^{o'} \tilde{Z}_n^o\right)\right]^{-1}.$$

Dado el supuesto 5, $lim_{n\to\infty}\frac{1}{n}Q_n^{o'} Q_n^o$ esta bien definido, con lo cual solo necesitamos probar que $\frac{1}{n}Q_n^{o'} \tilde{Z}_n^o$ converge a un límite bien definido y de rango completo.

Cabe aclarar que, si bien nosotros utilizamos $F_n\left(\hat{\theta}_{nls}\right)$ para imputar los datos faltantes, la consistencia del estimador I2SLS solo requiere que $\hat{\theta}$ sea consistente, con lo cual haremos la demostración pensando en $F_n\left(\hat{\theta}\right)$con $\hat{\theta}$ consistente, como caso especial, tenemos a $F_n\left(\hat{\theta}_{nls}\right)$.

Sea $\tilde{Z}_n^o = J_n^o \tilde{Z}_n$, con $\tilde{Z} = (W\tilde{y}_n \quad X_n)$, y $\tilde{y} = \begin{pmatrix} y_n^o \\ F_n\left(\hat{\theta}\right) \end{pmatrix}$,

$$\tilde{y}_n = \begin{pmatrix} y_n^o \\ F_n\left(\hat{\theta}\right) \end{pmatrix} = J_n^{o'} J_n^o y_n + J_n^{u'} J_n^u S_n^{-1}\left(\hat{\lambda}\right) X_n\hat{\beta} = J_n^{o'} J_n^o S_n^{-1}\left(\lambda_0\right)\left(X_n\beta_0 + e_n\right) + J_n^{u'} J_n^u S_n^{-1}\left(\hat{\lambda}\right) X_n\hat{\beta}.$$

Siendo $\frac{1}{n}Q_n^{o'} \tilde{Z}_n^o = \left[\frac{1}{n}Q_n^{o'} J_n^o W_n \tilde{y}_n \quad \frac{1}{n}Q_n^{o'} J_n^o X_n\right]$, si desarrollamos $Q_n^{o'} J_n^o W_n \tilde{y}_n$ tenemos

$$
\begin{aligned}
Q_n^{o'} J_n^o W_n \tilde{y}_n &= Q_n^{o'} J_n^o W_n J_n^{o'} J_n^o S_n^{-1}(\lambda_0) X_n \beta_0 \\
&\quad + Q_n^{o'} J_n^o W_n J_n^{o'} J_n^o S_n^{-1}(\lambda_0) e_n + Q_n^{o'} J_n^o W_n J_n^{u'} J_n^u S_n^{-1}\left(\hat{\lambda}\right) X_n \hat{\beta},
\end{aligned}
$$

considerando que $J_n^{o'} J_n^o = I_n - J_n^{u'} J_n^u$, tenemos que $Q_n^{o'} J_n^o W_n J_n^{o'} J_n^o S_n^{-1}(\lambda_0) X_n \beta_0 = Q_n^{o'} J_n^o W_n S_n^{-1}(\lambda_0) X_n \beta_0 - Q_n^{o'} J_n^o W_n J_n^{u'} J_n^u S_n^{-1}(\lambda_0) X_n \beta_0$, si reemplazamos esta igualdad en la ecuación anterior nos queda

$$
\begin{aligned}
Q_n^{o'} J_n^o W_n \tilde{y}_n &= Q_n^{o'} J_n^o W_n S_n^{-1}(\lambda_0) X_n \beta_0 + Q_n^{o'} J_n^o W_n J_n^{o'} J_n^o S_n^{-1}(\lambda_0) e_n \\
&\quad + Q_n^{o'} J_n^o W_n J_n^{u'} J_n^u A_n,
\end{aligned}
$$

con $A_n = \left[S_n^{-1}\left(\hat{\lambda}\right) X_n \hat{\beta} - S_n^{-1}(\lambda_0) X_n \beta_0\right]$, cabe recordar que $\left(\hat{\lambda}, \hat{\beta}\right)$ son estimadores consistentes de (λ_0, β_0) por lo tanto, $S_n^{-1}\left(\hat{\lambda}\right) X_n \hat{\beta} \xrightarrow{p} S_n^{-1}(\lambda_0) X_n \beta_0$, y $A_n = o_p(1)$, de esta manera tenemos que

$$
\begin{aligned}
\frac{1}{n} Q_n^{o'} J_n^o W_n \tilde{y}_n &= \frac{1}{n} Q_n^{o'} J_n^o W_n S_n^{-1}(\lambda_0) X_n \beta_0 + \frac{1}{n} Q_n^{o'} J_n^o W_n J_n^{o'} J_n^o S_n^{-1}(\lambda_0) e_n \\
&\quad + \frac{1}{n} Q_n^{o'} J_n^o W_n J_n^{u'} J_n^u A_n, \\
&= \frac{1}{n} Q_n^{o'} J_n^o W_n S_n^{-1}(\lambda_0) X_n \beta_0 + o_p(1),
\end{aligned}
$$

dado que $\frac{1}{n} Q_n^{o'} J_n^o W_n J_n^{o'} J_n^o S_n^{-1}(\lambda_0) e_n = o_p(1)$.

Con lo anterior tenemos que $\frac{1}{n} Q_n^{o'} \tilde{Z}_n^o = \left[\frac{1}{n} Q_n^{o'} J_n^o W_n S_n^{-1}(\lambda_0) X_n \beta_0 + o_p(1) \quad \frac{1}{n} Q_n^{o'} J_n^o X_n\right] = \frac{1}{n} Q_n^{o'} \left[J_n^o G_n X_n \beta_0 + o_p(1) \quad J_n^o X_n\right]$, esto implica que

$$
\lim_{n\to\infty} \frac{1}{n} Q_n^{o'} \tilde{Z}_n^o = \lim_{n\to\infty} \frac{1}{n} Q_n^{o'} \left[J_n^o G_n X_n \beta_0 + o_p(1) \quad J_n^o X_n\right],
$$

y dado que

$$
\lim_{n\to\infty} \frac{1}{n} Q_n^{o'} \left[J_n^o G_n X_n \beta_0 + o_p(1) \quad J_n^o X_n\right] \xrightarrow{p} \lim_{n\to\infty} \frac{1}{n} Q_n^{o'} \left[J_n^o G_n X_n \beta_0 \quad J_n^o X_n\right],
$$

entonces

$$
\lim_{n\to\infty} \frac{1}{n} Q_n^{o'} \tilde{Z}_n^o \xrightarrow{p} \lim_{n\to\infty} \frac{1}{n} Q_n^{o'} \left[J_n^o G_n X_n \beta_0 \quad J_n^o X_n\right], \tag{7.5}
$$

bajo el supuesto 5, este límite esta bien definido.

De esta manera nos encontramos con que $\left[\frac{1}{n}\tilde{Z}_n^{o'}Q_n^o\left(\frac{1}{n}Q_n^{o'}Q_n^o\right)^{-1}\frac{1}{n}Q_n^{o'}\tilde{Z}_n^o\right]^{-1}$ converge, en probabilidad, a una matriz no singular.

Por otro lado tenemos que, con $\tilde{u}_n$ definido en (2.6)

$$\begin{aligned}\frac{1}{n}Q_n^{o'}\tilde{u}_n^o &= \frac{1}{n}Q_n^{o'}\left\{\left[\lambda_0 W_n J_n^{u'}J_n^u S_n^{-1}+I_n\right]e_n-\lambda_0 W_n J_n^{u'}J_n^u\left[\left(S_n^{-1}\left(\hat{\lambda}\right)X_n\hat{\beta}\right)-\left(S_n^{-1}\left(\lambda_0\right)X_n\beta_0\right)\right]\right\},\\ &= \frac{1}{n}Q_n^{o'}\left[\lambda_0 W_n J_n^{u'}J_n^u S_n^{-1}+I_n\right]e_n-\frac{1}{n}Q_n^{o'}\lambda_0 W_n J_n^{u'}J_n^u A_n,\\ &= o_p\left(1\right). \end{aligned} \tag{7.6}$$

Dado (7.5), y (7.6), tenemos que $\left(\hat{\theta}_{2sls,n}-\theta_0\right)\xrightarrow{p}0$.

B.2. Prueba proposición 1: Distribución asintótica

Sea

$$\sqrt{n}\left(\hat{\theta}_{2sls,n}-\theta_0\right)=\left[\frac{1}{n}\tilde{Z}_n^{o'}Q_n^o\left(Q_n^{o'}Q_n^o\right)^{-1}Q_n^{o'}\tilde{Z}_n^o\right]^{-1}\tilde{Z}_n^{o'}Q_n^o\left(Q_n^{o'}Q_n^o\right)^{-1}\frac{1}{\sqrt{n}}Q_n^{o'}\tilde{u}_n^o. \tag{7.7}$$

Para determinar la distribución asintótica de $\sqrt{n}\left(\hat{\theta}_{2sls,n}-\theta_0\right)$debemos determinar primero la distribución asintótica de $\frac{1}{\sqrt{n}}Q_n^{o'}\tilde{u}_n^o$.

$$\frac{1}{\sqrt{n}}Q_n^{o'}\tilde{u}_n^o=\frac{1}{\sqrt{n}}Q_n^{o'}J_n^o\left(H_n\left(\lambda_0\right)e_n+R^*\right)=\frac{1}{\sqrt{n}}Q_n^{o'}J_n^o H_n\left(\lambda_0\right)e_n+\frac{1}{\sqrt{n}}Q_n^{o'}J_n^o R^*,$$

si desarrollamos el segundo término de la última igualdad tenemos

$$\begin{aligned}\frac{1}{\sqrt{n}}Q_n^{o'}J_n^o R^* &= \frac{1}{\sqrt{n}}Q_n^{o'}J_n^o\left\{\lambda_0 W_n J_n^{u'}J_n^u\left[R_n\left(\hat{\lambda}-\lambda_0\right)-S_n^{-1}\left(\lambda_0\right)C_n l_k o_p(\frac{1}{\sqrt{n}})\right]\right\},\\ &= \frac{1}{n}Q_n^{o'}J_n^o\lambda_0 W_n J_n^{u'}J_n^u R_n\sqrt{n}\left(\hat{\lambda}-\lambda_0\right)-\frac{1}{n}Q_n^{o'}J_n^o S_n^{-1}\left(\lambda_0\right)C_n l_k o_p(1),\\ &= o_p\left(1\right).\end{aligned}$$

Ya que, con R_n, definido en la sección 2.2, dada la consistencia de $\hat{\theta}$, y considerando que los términos, $\frac{1}{n}Q_n^{o'}J_n^o\lambda_0 W_n J_n^{u'}J_n^u$ y $\sqrt{n}\left(\hat{\lambda}-\lambda_0\right)$ son $O_p\left(1\right)$, tenemos que $\frac{1}{n}Q_n^{o'}J_n^o\lambda_0 W_n J_n^{u'}J_n^u R_n = o_p\left(\frac{1}{n}\right)$. Por otro lado, $\frac{1}{n}Q_n^{o'}J_n^o S_n^{-1}\left(\lambda_0\right)C_n l_k = O_p\left(1\right)$, con lo cual, $\frac{1}{n}Q_n^{o'}J_n^o S_n^{-1}\left(\lambda_0\right)C_n l_k o_p(1) = o_p\left(1\right)$.

Considerando lo anterior, tenemos que $\frac{1}{\sqrt{n}}Q_n^{o'}\tilde{u}_n^o = \frac{1}{\sqrt{n}}Q_n^{o'}H_n^o\left(\lambda_0\right)e_n + o_p\left(1\right)$, con $H_n^o\left(\lambda_0\right) = J_n^o H_n\left(\lambda_0\right)$, por lo tanto, $\frac{1}{\sqrt{n}}Q_n^{o'}\tilde{u}_n^o - \frac{1}{\sqrt{n}}Q_n^{o'}H_n^o\left(\lambda_0\right)e_n \xrightarrow{p} 0$, de esta manera, dado el lema de equivalencia asintótica, la distribución asintótica de $\frac{1}{\sqrt{n}}Q_n^{o'}H_n^o\left(\lambda_0\right)e_n$ es la misma que la de $\frac{1}{\sqrt{n}}Q_n^{o'}\tilde{u}_n^o$.

A continuación derivaremos la distribución asintótica de $\frac{1}{\sqrt{n}}Q_n^{o'}H_n^o\left(\lambda_0\right)e_n$.

Distribución asintótica de $\frac{1}{\sqrt{n}}Q_n^{o'}H_n^o(\lambda_0)\,e_n$

Para demostrar la distribución asintótica de $\frac{1}{\sqrt{n}}Q_n^{o'}H_n^o(\lambda_0)\,e_n$ debemos basarnos en una versión del TCL para arreglos triangulares. El teorema que presentamos a continuación es un corolario del TCL de Lindeberg y Feller para arreglos triangulares, usando la regla de Cramer-Wold. El mismo puede encontrarse en Kelejian & Prucha (1999) o en Lee (2003), entre otros.

Teorema. *Sea $\{e_{i,N},\ 1\leq i\leq N,\ N\geq 1\}$ un arreglo triangular de variables aleatorias idénticamente distribuidas, y para cada N conjuntamente independientes, con $E(e_{i,N})=0$, y $Var(e_{i,N})=\sigma^2$ con $0<\sigma^2<\infty$. Sea $\{d_{ij,N},\ 1\leq i\leq N,\ N\geq 1\}$, $j=1,...,K$ un arreglo triangular de números reales que son acotados en valor absoluto, es decir, $sup_N sup_{i\leq N,\,j\leq K}|d_{ij,N}|<\infty$. Ademas, sea $\{V_N:n\geq 1\}$ y $\{D_N:n\geq 1\}$, con $V_N=(e_{iN})_{i=1,...,N}$ y $D_N=(d_{ij,N})_{i=1,...,N;\,j=1,...,K}$, una secuencia de vectores aleatorios de orden $(N\times 1)$ y de matrices reales de orden $(N\times K)$ respectivamente, y siendo $\lim_{n\to\infty}N^{-1}D_N^{'}D_N$, finito y positivo definido. Entonces $N^{-1/2}D_N^{'}V_N\xrightarrow{d}N\left(0,\sigma^2\lim_{n\to\infty}N^{-1}D_N^{'}D_N\right)$.*

Los errores del modelo, establecidos en el supuesto 1, cumplen con las especificaciones para $e_{i,N}$ del teorema anterior, ya que para cada sucesión son $i.i.d.$ con $E(e_{i,n})=0$. Por otro lado, asumiendo que $H_n(\lambda_0)$ es uniformemente acotada en la suma de sus filas y sus columnas (absolutamente sumable), tenemos que $Q_n^{o'}H_n^o(\lambda_0)$ es acotado, ya que una matriz acotada multiplicada por una matriz absolutamente sumable da como resultado una matriz acotada.

Si definimos a $Q_n^{o'}H_n^o(\lambda_0)=D_n^o$, con D_n^o matriz de orden $(k+1,\ n^o)$, por lo anterior sabemos que, al ser D_n^o acotado, $sup_N sup_{i\leq N,\,j\leq K}|d_{ij,N}|<\infty$, y asumiendo que $\lim_{n\to\infty}\frac{1}{n}D_n^oD_n^{o'}$ es finito y positivo definido[13], entonces

$$\frac{1}{\sqrt{n}}Q_n^{o'}H_n^o(\lambda_0)\,e_n\xrightarrow{d}N\left(0,\sigma_e^2\lim_{n\to\infty}\frac{1}{n}Q_n^{o'}H_n^o(\lambda_0)\,H_n^{o'}(\lambda_0)\,Q_n^o\right).$$

Basándonos en la derivación anterior, y considerando (7.7), tenemos que, $\sqrt{n}\left(\hat{\theta}_{2sls,n}-\theta_0\right)\xrightarrow{d}N(0,\Sigma_{i2sls})$, con

$$\begin{aligned}\Sigma_{i2sls} &= \sigma_e^2\lim_{n\to\infty}\left[\frac{1}{n}C_n^{o'}Q_n^o\left(Q_n^{o'}Q_n^o\right)^{-1}Q_n^{o'}C_n^o\right]^{-1}\\ &\quad\left[\frac{1}{n}C_n^{o'}Q_n^o\left(Q_n^{o'}Q_n^o\right)^{-1}Q_n^{o'}H_n^o(\lambda_0)\,H_n^{o'}(\lambda_0)\,Q_n^o\left(Q_n^{o'}Q_n^o\right)^{-1}Q_n^{o'}C_n^o\right]\\ &\quad\left[\frac{1}{n}C_n^{o'}Q_n^o\left(Q_n^{o'}Q_n^o\right)^{-1}Q_n^{o'}C_n^o\right]^{-1}.\end{aligned}$$

En la ecuación de la varianza asintótica hemos hecho uso de que, $\lim_{n\to\infty}\frac{1}{n}Q_n^{o'}\tilde{Z}_n^o\xrightarrow{p}\lim_{n\to\infty}\frac{1}{n}Q_n^{o'}C_n^o$, como lo refiere la ecuación (7.5).

[13] Kelejian & Prucha (1999) demuestran que, siendo A y B dos matrices absolutamente sumables, AB también lo es, además, si Z es una matriz acotada, entonces, $Z^{'}AZ=O(n)$, con todo esto, $D_n^oD_n^{o'}=O(n)$.

C. Estimador IG2SLS

Bajo el supuesto 6 tenemos que $\Omega_n(\lambda_0) = H_n^o(\lambda_0) H_n^{o'}(\lambda_0)$, de orden $(n^o \times n^o)$, tiene inversa, la cual también es definida positiva, y así podemos aplicar la descomposición de Cholesky y expresarla de la siguiente manera $\Omega_n^{-1}(\lambda_0) = K_n^{'} K_n$, donde K_n es una matriz triangular inferior no singular de orden $(n^o \times n^o)$, si multiplicamos ambos lados de nuestro modelo por K_n tenemos

$$K_n y_n^o = K_n \tilde{Z}_n^o \theta_0 + K_n J_n^o (H_n(\lambda_0) e_n + R^*) = K_n \tilde{Z}_n^o \theta_0 + K_n \tilde{u}_n^o.$$

Dado que nuestro modelo transformado es función de $y_n^{o*} = K_n y_n^o$, de $\tilde{Z}_n^{o*} = K_n \tilde{Z}_n^o$, y de $\tilde{u}_n^{o*} = K_n \tilde{u}_n^o$, la matriz de instrumentos no va a ser la misma, como vemos a continuación, también hay que aplicarle la transformación

$$\begin{aligned}
y_n^o &= J_n^o (I_n - \lambda_0 W_n)^{-1} X_n \beta_o + J_n^o (I_n - \lambda_0 W_n)^{-1} e_n, \\
K_n y_n^o &= K_n J_n^o (I_n - \lambda_0 W_n)^{-1} X_n \beta_o + K_n J_n^o (I_n - \lambda_0 W_n)^{-1} e_n, \\
K_n y_n^o &= K_n J_n^o \left(I_n + \lambda_0 W_n + \lambda_0^2 W_n^2 + \cdots\right) X_n \beta_o + K_n J_n^o (I_n - \lambda_0 W_n)^{-1} e_n, \\
K_n y_n^o &= K_n \left(J_n^o X_n \beta_0 + \lambda_0 J_n^o W_n X_n \beta_o + \lambda_0^2 J_n^o W_n^2 X_n \beta_o + \cdots\right) + K_n J_n^o (I_n - \lambda_0 W_n)^{-1} e_n. \quad (7.8)
\end{aligned}$$

Bajo el formato tradicional del SLM, basados en un desarrollo similar a (7.8), solo que sin transformar las variables y sin datos faltantes, se llega a que los instrumentos recomendados, podrían ser las columnas linealmente independientes de $\left\{X_n \quad W_n X_n \quad W_n^2 X_n \ldots\right\}$ a la cual llamamos Q_n, siendo que, al introducirle la falta de observaciones, la multiplicamos por la matriz selectora y la llamamos Q_n^o.

En este caso, donde pensamos en un modelo "generalizado", como se muestra en (7.8), debemos multiplicarla por K_n. En otras palabras, la matriz de instrumentos del modelo generalizado no va a ser Q_n^o función de $J_n^o X_n$ y $J_n^o W_n$, sino que en su lugar, va a considerar la transformación, con lo cual sera $K_n Q_n^o$.

Ahora bien, considerando que la matriz de variables explicativas es $\tilde{Z}_n^{o*} = K_n \tilde{Z}_n^o$, que la variable explicada es $y_n^{o*} = K_n y_n^o$, y que la matriz de instrumentos es $Q_n^{o*} = K_n Q_n^o$, el estimador IG2SLS es

$$\begin{aligned}
\hat{\theta}_{ig2sls,n} &= \left[\tilde{Z}_n^{o*'} Q_n^{o*} \left(Q_n^{o*'} Q_n^{o*}\right)^{-1} Q_n^{o*'} \tilde{Z}_n^{o*}\right]^{-1} \tilde{Z}_n^{o*'} Q_n^{o*} \left(Q_n^{o*'} Q_n^{o*}\right)^{-1} Q_n^{o*'} \tilde{Z}_n^{o*} y_n^{o*}, \\
&= \left[\tilde{Z}_n^{o'} K_n^{'} K_n Q_n^o \left(Q_n^{o'} K_n^{'} K_n Q_n^o\right)^{-1} Q_n^{o'} K_n^{'} K_n \tilde{Z}_n^o\right]^{-1} \tilde{Z}_n^{o'} K_n^{'} K_n Q_n^o \left(Q_n^{o'} K_n^{'} K_n Q_n^o\right)^{-1} Q_n^{o'} K_n^{'} K_n y_n^o, \\
&= \left[\tilde{Z}_n^{o'} \Omega_n^{-1}(\lambda_0) Q_n^o \left(Q_n^{o'} \Omega_n^{-1}(\lambda_0) Q_n^o\right)^{-1} Q_n^{o'} \Omega_n^{-1}(\lambda_0) \tilde{Z}_n^o\right]^{-1} \\
&\quad \tilde{Z}_n^{o'} \Omega_n^{-1}(\lambda_0) Q_n^o \left(Q_n^{o'} \Omega_n^{-1}(\lambda_0) Q_n^o\right)^{-1} Q_n^{o'} \Omega_n^{-1}(\lambda_0) y_n^o.
\end{aligned}$$

Prueba proposición 2: Consistencia y Distribución asintótica

La derivación de la consistencia y la distribución asintótica del estimador IG2SLS, es similar a la realizada para I2SLS, con lo cual evitaremos ciertos pasos que se pueden derivar del apéndice B.

Sea

$$\begin{aligned}\left(\hat{\theta}_{ig2sls,n}-\theta_0\right) &= \left[\frac{1}{n}\tilde{Z}_n^{o'}\Omega_n^{-1}\left(\lambda_0\right)Q_n^o\left(Q_n^{o'}\Omega_n^{-1}\left(\lambda_0\right)Q_n^o\right)^{-1}Q_n^{o'}\Omega_n^{-1}\left(\lambda_0\right)\tilde{Z}_n^o\right]^{-1} \\ &\quad \frac{1}{n}\tilde{Z}_n^{o'}\Omega_n^{-1}\left(\lambda_0\right)Q_n^o\left(Q_n^{o'}\Omega_n^{-1}\left(\lambda_0\right)Q_n^o\right)^{-1}Q_n^{o'}\Omega_n^{-1}\left(\lambda_0\right)\tilde{u}_n^o.\end{aligned}$$

La diferencia entre este estimador y el I2SLS, es la introducción de una matriz de ponderaciones, $\Omega_n^{-1}\left(\lambda_0\right)$, por lo tanto, al igual que en I2SLS, debemos mostrar que $\frac{1}{n}Q_n^{o'}\Omega_n^{-1}\left(\lambda_0\right)\tilde{Z}_n^o$ converge a un límite bien definido y de rango completo. Introduciendo $\Omega_n^{-1}\left(\lambda_0\right)$ en la derivación de (7.5) se llega a que $\frac{1}{n}Q_n^{o'}\Omega_n^{-1}\left(\lambda_0\right)\tilde{Z}_n^o = \frac{1}{n}Q_n^{o'}\Omega_n^{-1}\left(\lambda_0\right)\left[J_n^oG_nX_n\beta_0+o_p\left(1\right) \quad J_n^oX_n\right]$, y de esta igualdad tenemos que

$$\frac{1}{n}Q_n^{o'}\Omega_n^{-1}\left(\lambda_0\right)\tilde{Z}_n^o \xrightarrow{p} \lim_{n\to\infty}\frac{1}{n}Q_n^{o'}\Omega_n^{-1}\left(\lambda_0\right)\left[J_n^oG_nX_n\beta_0 \quad J_n^oX_n\right], \tag{7.9}$$

bajo el supuesto 6, este límite esta bien definido.

Siguiendo el mismo desarrollo de B.2, solo que considerando a $\Omega_n^{-1}\left(\lambda_0\right)$, y asumiendo que es una matriz absolutamente sumable, se puede derivar fácilmente que $\frac{1}{\sqrt{n}}Q_n^{o'}\Omega_n^{-1}\left(\lambda_0\right)\tilde{u}_n^o \xrightarrow{d} N\left(0,\sigma_e^2\lim_{n\to\infty}\frac{1}{n}Q_n^{o'}\Omega_n^{-1}\left(\lambda_0\right)Q_n^o\right)$, esto implica que $\sqrt{n}\left(\hat{\theta}_{ig2sls,n}-\theta_0\right) \xrightarrow{d} N\left(0,\Sigma_{ig2sls}\right)$ donde

$$\begin{aligned}\Sigma_{ig2sls} &= \sigma_e^2\lim_{n\to\infty}\frac{1}{n}\left[\frac{1}{n}C_n^{o'}\Omega_n^{-1}\left(\lambda_0\right)Q_n^o\left(Q_n^{o'}\Omega_n^{-1}\left(\lambda_0\right)Q_n^o\right)^{-1}Q_n^{o'}\Omega_n^{-1}\left(\lambda_0\right)C_n^o\right]^{-1} \\ &\quad C_n^{o'}\Omega_n^{-1}\left(\lambda_0\right)Q_n^o\left(Q_n^{o'}\Omega_n^{-1}\left(\lambda_0\right)Q_n^o\right)^{-1}Q_n^{o'}\Omega_n^{-1}\left(\lambda_0\right)\Omega_n\left(\lambda_0\right) \\ &\quad \Omega_n^{-1}\left(\lambda_0\right)Q_n^o\left(Q_n^{o'}\Omega_n^{-1}\left(\lambda_0\right)Q_n^o\right)^{-1}Q_n^{o'}\Omega_n^{-1}\left(\lambda_0\right)C_n^o \\ &\quad \left[\frac{1}{n}C_n^{o'}\Omega_n^{-1}\left(\lambda_0\right)Q_n^o\left(Q_n^{o'}\Omega_n^{-1}\left(\lambda_0\right)Q_n^o\right)^{-1}Q_n^{o'}\Omega_n^{-1}\left(\lambda_0\right)C_n^o\right]^{-1},\end{aligned}$$

cancelando términos nos queda

$$\Sigma_{ig2sls} = \sigma_e^2\lim_{n\to\infty}\left[\frac{1}{n}C_n^{o'}\Omega_n^{-1}\left(\lambda_0\right)Q_n^o\left(Q_n^{o'}\Omega_n^{-1}\left(\lambda_0\right)Q_n^o\right)^{-1}Q_n^{o'}\Omega_n^{-1}\left(\lambda_0\right)C_n^o\right]^{-1}.$$

Prueba proposición 4:

Sea el siguiente modelo

$$\begin{aligned}y_n &= X_n\beta+\lambda_0W_ny_n+u_n, \qquad |\lambda_0|\prec 1, \\ u_n &= \rho_0M_nu_n+e_n, \qquad |\rho_0|\prec 1,\end{aligned} \tag{7.10}$$

donde y_n es un vector de variables dependientes de orden $(n \times 1)$, X_n es una matriz de variables exógenas de orden $(n \times k)$, β_0 es una vector de parámetros de orden $(k \times 1)$, u_n es un vector de orden $(n \times 1)$ de perturbaciones de la regresión, e_n es un vector de errores aleatorios de orden $(n \times 1)$, W_n y M_n son matrices de orden $(n \times n)$ de pesos espaciales cuyos valores son conocidos con ceros en la diagonal, λ_0 y ρ_0 son parámetros de autocorrelación espacial.

Kelejian et al. (2004) demuestran el siguiente lema

Lema 6. *Dado el modelo (7.10), bajo los supuestas 1-3, asumiendo que la suma de las filas de M_n es uniformemente acotada en valor absoluto por uno, ademas, la suma de las columnas de M_n y $(I_n - \rho_0 M_n)^{-1}$ es acotada uniformemente en valor absoluto por una constante finita, asumiendo también, que las innovaciones tienen cuarto momento finito, y siendo $\sqrt{n}\left(\hat{\lambda}_n - \lambda_0\right) = O_p(1)$, con $|\lambda_0| < 1$, y $\left(\hat{\beta}_n - \beta\right) = o_p(1)$. Sea r_n una secuencia de números naturales con $0 \leqslant r_n \leqslant n$, $r_n \uparrow \infty$ y $r_n = o(\sqrt{n})$, y sea $a_n = (a_{1,n}, \ldots, a_{n.n})'$ una secuencia de vectores constantes de orden $(n \times 1)$ cuyos elementos son uniformemente acotados en valor absoluto. entonces*

$$n^{-1} a_n' \left(\widetilde{\overline{y}}_n - E\overline{y}_n\right) = o_p(1),$$

donde, $E\overline{y}_n = W_n (I_n - \lambda_0 W_n)^{-1} X_n \beta$, y $\widetilde{\overline{y}}_n = \sum_{k=0}^{r_n} \hat{\lambda}_0^k W_n^{k+1} X_n \hat{\beta}_0$.

El modelo para el cual esta desarrollado este lema es un SARAR, pero esta claro que también se cumple para el caso particular en que $\rho_0 = 0$, y así estamos bajo el modelo SLM, que es el que nos interesa a nosotros.

Lema 7. *Asumiendo que los supuestos establecidos en el lema 1 se cumplen*[14]*, y siendo $\hat{Q}_n^{kp} = \left(\widetilde{\overline{y}}_n \; X_n\right)$, y $C_n = (E\overline{y}_n \; X_n)$, entonces $\frac{1}{n}\hat{Q}_n^{kp} \xrightarrow{p} \frac{1}{n} C_n$.*

Pensemos en $I_n = \begin{pmatrix} a_{1,n}' \\ \vdots \\ a_{n^o,n}' \end{pmatrix}$, donde $a_{i,n}$ es un vector cuyo i-ésimo elemento vale uno y el resto son todos ceros. Sea

$$n^{-1} I_n \left(\widetilde{\overline{y}}_n - E\overline{y}_n\right) = \begin{pmatrix} n^{-1} a_{1,n}' \left(\widetilde{\overline{y}}_n - E\overline{y}_n\right) \\ \vdots \\ n^{-1} a_{n,n}' \left(\widetilde{\overline{y}}_n - E\overline{y}_n\right) \end{pmatrix},$$

dado que $a_{i,n}'$cumple con las especificaciones del lema 1 para a_n', entonces

$$n^{-1} I_n \left(\widetilde{\overline{y}}_n - E\overline{y}_n\right) = \begin{pmatrix} o_p(1) \\ \vdots \\ o_P(1) \end{pmatrix},$$

[14]Para nuestro modelo, dado que $\rho = 0$ y $M_n = [0]$, esto implica que $(I_n - \rho M_n)^{-1} = I_n$, con los cual los supuestos sobre M_n y $(I_n - \rho M_n)^{-1}$ se cumplen trivialmente.

y por otro lado tenemos que , $n^{-1}I_n\left(\widetilde{\overline{y}}_n - E\overline{y}_n\right) = n^{-1}\left(\widetilde{\overline{y}}_n - E\overline{y}_n\right)$, esto implica que

$$n^{-1}\left(\widetilde{\overline{y}}_n - E\overline{y}_n\right) = \begin{pmatrix} o_p(1) \\ \vdots \\ o_P(1) \end{pmatrix}.$$

Con lo anterior tenemos que $n^{-1}\widetilde{\overline{y}}_n \xrightarrow{p} n^{-1}E\overline{y}_n$, así $\frac{1}{n}\hat{Q}_n^{kp} \xrightarrow{p} \frac{1}{n}C_n$.

La convergencia en probabilidad de una matriz implica que converge elemento a elemento, por lo tanto también es cierto que, $\frac{1}{n}J_n^o\hat{Q}_n^{kp} \xrightarrow{p} \frac{1}{n}J_n^oC_n$, ya que J_n^o es una matriz que selecciona elementos.

La varianza de la distribución asintótica del estimador IG2SLS utilizando la matriz de instrumentos $\hat{Q}_n^{kp}$, es equivalente a la siguiente expresión

$$\Sigma_{iste} = \sigma_e^2 \lim_{n\to\infty}\left[\frac{1}{n}C_n^{o'}\Omega_n^{-1}(\lambda_0)\frac{1}{n}\hat{Q}_n^{okp}\left(\frac{1}{n}\hat{Q}_n^{okp'}\Omega_n^{-1}(\lambda_0)\frac{1}{n}\hat{Q}_n^{okp}\right)^{-1}\frac{1}{n}\hat{Q}_n^{okp'}\Omega_n^{-1}(\lambda_0)C_n^o\right]^{-1}.$$

Considerando que, $\frac{1}{n}J_n^o\hat{Q}_n^{kp} \xrightarrow{p} \frac{1}{n}J_n^oC_n$, tenemos

$$\begin{aligned}\Sigma_{iste} &= \sigma_e^2 \lim_{n\to\infty}\left[\frac{1}{n}C_n^{o'}\Omega_n^{-1}(\lambda_0)\frac{1}{n}C_n^o,\left(\frac{1}{n}C_n^{o'},\Omega_n^{-1}(\lambda_0)\frac{1}{n}C_n^o\right)^{-1}\frac{1}{n}C_n^{o'}\Omega_n^{-1}(\lambda_0)C_n^o\right]^{-1}, \\ &= \sigma_e^2 \lim_{n\to\infty}\left(\frac{1}{n}C_n^{o'},\Omega_n^{-1}(\lambda_0)C_n^o\right)^{-1}.\end{aligned}$$

D. Equivalencia asintótica de $\hat{\theta}_{ibg2sls,n}$y $\hat{\theta}_{ibg2slsa,n}$

Para probar $\hat{\theta}_{ibg2sls,n}$ y $\hat{\theta}_{ibg2slsa,n}$ son $\sqrt{n}-$equivalentes, debemos probar que $\sqrt{n}\left(\hat{\theta}_{ibg2sls,n} - \hat{\theta}_{ibg2slsa,n}\right) = o_p(1)$.

Dado que el estimador IBG2SLS utiliza la matriz de instrumentos optima, $Q_n^{o*} = C_n^o$, tenemos

$$\begin{aligned}\sqrt{n}\left(\hat{\theta}_{ibg2sls,n} - \hat{\theta}_{ibg2slsa,n}\right) &= \left[\frac{1}{n}\tilde{Z}_n^{o'}\Omega_n^{-1}(\lambda_0)C_n^o\left(\frac{1}{n}C_n^{o'}\Omega_n^{-1}(\lambda_0)C_n^o\right)^{-1}\frac{1}{n}C_n^{o'}\Omega_n^{-1}(\lambda_0)\tilde{Z}_n^o\right]^{-1} \\ &\quad \frac{1}{n}\tilde{Z}_n^{o'}\Omega_n^{-1}(\lambda_0)C_n^o\left(\frac{1}{n}C_n^{o'}\Omega_n^{-1}(\lambda_0)C_n^o\right)^{-1}\frac{1}{\sqrt{n}}C_n^{o'}\Omega_n^{-1}(\lambda_0)y_n^o \\ &\quad -\left(\frac{1}{n}C_n^{o'}\Omega_n^{-1}(\lambda_0)C_n^o\right)^{-1}\frac{1}{\sqrt{n}}C_n^{o'}\Omega_n^{-1}(\lambda_0)y_n^o. \qquad (7.11)\end{aligned}$$

De (7.9) sabemos que, $\frac{1}{n}Q_n^{o'}\Omega_n^{-1}(\lambda_0)\tilde{Z}_n^o \xrightarrow{p} \frac{1}{n}Q_n^{o'}\Omega_n^{-1}(\lambda_0)C_n^o$. Reemplazando Q_n^o por $Q_n^{o*} = C_n^o$, tenemos que, $\frac{1}{n}C_n^{o'}\Omega_n^{-1}(\lambda_0)\tilde{Z}_n^o \xrightarrow{p} \frac{1}{n}C_n^{o'}\Omega_n^{-1}(\lambda_0)C_n^o$. Con lo cual,

$$\begin{aligned}\sqrt{n}\left(\hat{\theta}_{ibg2sls,n}-\hat{\theta}_{ibg2slsa,n}\right) &\xrightarrow{p} \left[\frac{1}{n}C_n^{o'}\Omega_n^{-1}(\lambda_0)C_n^o\left(\frac{1}{n}C_n^{o'}\Omega_n^{-1}(\lambda_0)C_n^o\right)^{-1}\frac{1}{n}C_n^{o'}\Omega_n^{-1}(\lambda_0)C_n^o\right]^{-1} \\ &\quad \frac{1}{n}C_n^{o'}\Omega_n^{-1}(\lambda_0)C_n^o\left(\frac{1}{n}C_n^{o'}\Omega_n^{-1}(\lambda_0)C_n^o\right)^{-1}\frac{1}{\sqrt{n}}C_n^{o'}\Omega_n^{-1}(\lambda_0)y_n^o \\ &\quad -\left(\frac{1}{n}C_n^{o'}\Omega_n^{-1}(\lambda_0)C_n^o\right)^{-1}\frac{1}{\sqrt{n}}C_n^{o'}\Omega_n^{-1}(\lambda_0)y_n^o.\end{aligned}$$

Cancelando términos tenemos

$$\begin{aligned}\sqrt{n}\left(\hat{\theta}_{ibg2sls,n}-\hat{\theta}_{ibg2slsa,n}\right) &\xrightarrow{p} I_{k+1}\left(\frac{1}{n}C_n^{o'}\Omega_n^{-1}(\lambda_0)C_n^o\right)^{-1}\frac{1}{\sqrt{n}}C_n^{o'}\Omega_n^{-1}(\lambda_0)y_n^o \\ &\quad -\left(\frac{1}{n}C_n^{o'}\Omega_n^{-1}(\lambda_0)C_n^o\right)^{-1}\frac{1}{\sqrt{n}}C_n^{o'}\Omega_n^{-1}(\lambda_0)y_n^o, \\ \sqrt{n}\left(\hat{\theta}_{ibg2sls,n}-\hat{\theta}_{ibg2slsa,n}\right) &\xrightarrow{p} 0.\end{aligned}$$

Con esto, hemos probado que $\sqrt{n}\left(\hat{\theta}_{ibg2sls,n}-\hat{\theta}_{ibg2slsa,n}\right)=o_p(1)$.

E. Estimador de la varianza del modelo

Como dijimos en el apartado correspondiente, el estimador de la varianza propuesto en el trabajo se basa en el error del modelo utilizado para imputar los datos faltantes, el cual, como se deriva de (3.3), es $J_n^o S_n^{-1}(\lambda_0)e_n$.

La varianza del modelo (3.3) es, $Var\left[J_n^o S_n^{-1}(\lambda_0)e_n\right]=J_n^o S_n^{-1}(\lambda_0)Var(e_n)S_n^{-1'}(\lambda_0)J_n^{o'}$, y dado el supuesto 1, $Var(e_n)=\sigma_e^2 I_n$, tenemos que, $Var\left[J_n^o S_n^{-1}(\lambda_0)e_n\right]=\sigma_e^2\left[J_n^o S_n^{-1}(\lambda_0)S_n^{-1'}(\lambda_0)J_n^{o'}\right]$.

Si asumimos que $\Sigma_n(\lambda_0)=\left[J_n^o S_n^{-1}(\lambda_0)S_n^{-1'}(\lambda_0)J_n^{o'}\right]$ es definida positiva, entonces tiene inversa, la cual también es definida positiva, y así podemos aplicar la descomposición de Cholesky y expresarla como, $\Sigma_n^{-1}(\lambda_0)=T_n^{'}(\lambda_0)T_n(\lambda_0)$, donde $T_n(\lambda_0)$ es una matriz triangular inferior no singular de orden $(n^o\times n^o)$.

Con todo lo anterior, si multiplicamos al error por $T_n(\lambda_0)$ tenemos, $v_n=T_n(\lambda_0)J_n^o S_n^{-1}(\lambda_0)e_n$, esta nueva variable v_n tiene las siguientes características,

$$\begin{aligned}E(v_n) &= E\left(T_n(\lambda_0)J_n^o S_n^{-1}(\lambda_0)e_n\right), \\ &= T_n(\lambda_0)J_n^o S_n^{-1}(\lambda_0)E(e_n), \\ &= 0.\end{aligned}$$

$$
\begin{aligned}
Var\left(v_n\right) &= Var\left[T_n\left(\lambda_0\right) J_n^o S_n^{-1}\left(\lambda_0\right) e_n\right], \\
&= \sigma_e^2\left[T_n\left(\lambda_0\right) J_n^o S_n^{-1}\left(\lambda_0\right) S_n^{-1'}\left(\lambda_0\right) J_n^{o'} T_n^{'}\left(\lambda_0\right)\right], \\
&= \sigma_e^2\left[T_n\left(\lambda_0\right) \Sigma_n\left(\lambda_0\right) T_n^{'}\left(\lambda_0\right)\right], \\
&= \sigma_e^2\left[T_n\left(\lambda_0\right)\left(T_n^{'}\left(\lambda_0\right) T_n\left(\lambda_0\right)\right)^{-1} T_n^{'}\left(\lambda_0\right)\right], \\
&= \sigma_e^2 I_n.
\end{aligned}
$$

$$
\begin{aligned}
Cov\left(v_i,\, v_j\right) &= T_n\left(\lambda_0\right) J_n^o S_n^{-1}\left(\lambda_0\right)\left[Cov\left(e_i,\, e_j\right)\right] S_n^{-1'}\left(\lambda_0\right) J_n^{o'} T_n^{'}\left(\lambda_0\right), \\
&= 0,
\end{aligned}
$$

para $i \neq j$.

Considerando las características de v_n, y siendo v_i el elemento i-ésimo del vector v_n, podemos aplicar ley débil de los grandes números, y tenemos que $\frac{1}{n}\sum_{i=1}^{n} v_i^2 \xrightarrow{p} \frac{1}{n}\sum_{i=1}^{n} E\left(v_i^2\right) = \frac{1}{n} n\sigma_e^2 = \sigma_e^2$.

Ahora bien, nuestro problema reside en que v_n es inobservable para nosotros, ya que depende de λ_0 que es un parámetro desconocido.

Lo que hacemos a continuación es mostrar que $J_n^o S_n^{-1}\left(\hat{\lambda}\right)\hat{e}_n = J_n^o y_n - J_n^o S_n^{-1}\left(\hat{\lambda}\right) X_n\hat{\beta}$, es un estimador consistente de $J_n^o S_n^{-1}\left(\lambda_0\right) e_n$.

Sea

$$
\begin{aligned}
J_n^o S_n^{-1}\left(\hat{\lambda}\right)\hat{e}_n &= J_n^o S_n^{-1}\left(\lambda_0\right) X_n\beta_0 + J_n^o S_n^{-1}\left(\lambda_0\right) e_n - J_n^o S_n^{-1}\left(\hat{\lambda}\right) X_n\hat{\beta}, \\
&= \left[J_n^o S_n^{-1}\left(\lambda_0\right) X_n\beta_0 - J_n^o S_n^{-1}\left(\hat{\lambda}\right) X_n\hat{\beta}\right] + J_n^o S_n^{-1}\left(\lambda_0\right) e_n, \\
&= o_p\left(1\right) + J_n^o S_n^{-1}\left(\lambda_0\right) e_n.
\end{aligned}
$$

Las igualdades anteriores se basan en la consistencia de $\hat{\theta} = \left(\hat{\lambda},\, \hat{\beta}'\right)'$ que hacen que el término entre corchetes sea $o_p\left(1\right)$.

Por otro lado tenemos que $\Sigma_n^{-1}\left(\lambda_0\right) = T_n^{'}\left(\lambda_0\right) T_n\left(\lambda_0\right)$ también dependen de λ_0, pero, al igual que en el caso anterior, dada la consistencia de $\hat{\lambda}$ y asumiendo que $\Sigma_n^{-1}\left(.\right)$ es continua en λ_0 , esto nos garantiza que $\Sigma_n^{-1}\left(\hat{\lambda}\right) \xrightarrow{p} \Sigma_n^{-1}\left(\lambda_0\right)$.

Teniendo en cuenta el desarrollo previo, tenemos que $\hat{v}_n = T_n\left(\hat{\lambda}\right) J_n^o S_n^{-1}\left(\hat{\lambda}\right)\hat{e}_n \xrightarrow{p} v_n = T_n\left(\lambda_0\right) J_n^o S_n^{-1}\left(\lambda_0\right) e_n$, y de esta forma,

$$
\hat{\sigma}_e^2 = \frac{1}{n}\sum_{i=1}^{n}\hat{v}_i^2 \xrightarrow{p} \sigma_e^2.
$$