



## Estimation of Spatial Lag Model Under Random Missing Data in the Dependent Variable. Two Stage Estimator with Imputation

Alejandro Izaguirre<sup>a,b</sup>

<sup>a</sup>Universidad de San Andrés, Buenos Aires, Argentina

<sup>b</sup>Facultad de Ciencias Económicas y Estadística, Universidad Nacional de Rosario

✉ izaguirre.ale@gmail.com

### Abstract

The main goal of this article is to propose estimators for the Spatial Lag Model (SLM) under missing data context. We present three alternatives estimators for the SLM based on Two Stage Least Squares estimation methodology. The estimators are efficient within their type and consistent under random missing data in the dependent variable. Unlike the IBG2SLS estimator presented in Wang and Lee (2013) which impute all missing data we only impute missing data in the spatial lag. Our first proposal is an alternative version of the IBG2SLS estimator, the second one is based on an approximation to the optimal instruments matrix and the third one is an alternative  $\sqrt{n}$ -equivalent to the first. Thorough a Monte Carlo simulation we assess the estimators performance under finite samples. Results show a good performance for all estimators, moreover, results are quite similar to the IBG2SLS estimator suggesting that a complete imputation (as IBG2SLS does) does not add information.

**Article History:** Received: July 28 2020 / Revised: October 05 2020 / Accepted: December 20 2020

**Keywords:** Random missing data; Two stage estimators; Imputation; Spatial lag model

**JEL Classification:** C13; C21; R5

## 1. Introduction

In the last few decades there has been a renaissance of theoretical and empirical work on the spatial aspects of the economy. This *new economic geography* has emerged as one of the most exciting areas of contemporary economics (Krugman, 1991; Fujita et al., 1999).

One of the reasons for this is the growing interest in the impact of location on multiple economic aspects. Additionally, the availability of georeferenced data has increased considerably over the past two decades, allowing the use of spatial econometric techniques for analyzing data. At the same time, the Internet has greatly facilitated access to such data (Florax and Van der Vlist, 2003).

The main developments in spatial econometrics have focused on identifying and including spatial effects into regression models (Anselin, 1988). There are different model specifications, such as the Spatial Lag Model (SLM); the Spatial Error Model (SEM); and the combination of both, named SARAR, among the best known. At the same time, given the issues involved in spatial dependence, these models require their own estimation methods, such as Maximum Likelihood, the General Method of Moments, and Quasi Maximum Likelihood, among others.

There are some relevant methodological issues which have not been completely covered from a theoretical point of view. In particular, the development of techniques for dealing with missing data, under a spatial context, is scarce. This scenario is highly common when using observational data.

The development of spatial statistics has given rise to several forecasting techniques that can be useful in handling missing data. Cressie (1993) describes several forecasting tools in a spatial context, based on Kriging's methods. Kelejian and Prucha (2007) analyze the relative efficiency of five different predictors for a SARAR model based on different information sets. Although these works are related with missing data problems, as they provide imputation options which could be useful for handling this problem, they make no reference to estimating parameters under missing data.

This work presents three alternative estimators for the Spatial Lag Model based on the Two-Stage Least Squares estimation method. Estimators are efficient within their type and consistent under random missing data in the dependent variable. Kelejian and Prucha (2010) present three Two-Stage estimators considering different missing data configurations; and show that, in some cases, missing data effects can be asymptotically neglected. LeSage and Pace (2004) analyze housing prices using the EM algorithm for predicting missing  $y$ 's in an SLM with correlated errors. Wang and Lee (2013) propose three different estimators for an SLM under missing data in the dependent variable, based on Non-Linear Least Squares, the General Method of Moments, and Two-Stage Least Squares, respectively.

Our proposals are mainly related with the I2SLS estimator presented in Wang and Lee (2013); but the key difference is the number of imputations. In SLM, the dependent (or explained) variable also acts as explanatory variable through the spatial lag. In this work, in order to impute as few observations as possible, we propose to impute only the spatial lag,  $Wy$ . We consider only those equations for which we observe the dependent variable; and only impute the

unobserved  $y$ 's in their spatial lag. Partial imputation (in contrast with total imputation) gives rise to the possibility of working only with complete data.

Following Kelejian et al. (2004), we derive an optimal instrument matrix to minimize the estimators' variances.

In the model with missing data described in the following section, we make imputation explicit and show its impact on the error term. Section 3 presents the proposed estimators, as well as their asymptotic distributions. Section 4 presents the Monte Carlo results, where we assess the estimators' performance and compare their behaviour with the BI2SLS estimator proposed in Wang and Lee (2013). Finally, section 5 provides the main conclusions. All proofs are provided in the appendices.

## 2. Missing Data Model

### 2.1 Model

Consider the following SLM,

$$y_n = \lambda_0 W_n y_n + X_n \beta_0 + e_n, \quad (1)$$

where  $y_n$  is an  $(n \times 1)$  vector of dependent variables,  $X_n$  is an  $(n \times k)$  matrix of exogenous variables,  $\beta_0$  is a  $(k \times 1)$  vector of parameters,  $e_n$  is an  $(n \times 1)$  vector of disturbances such that  $e_n \sim (0, \sigma_e^2 I_n)$ ,  $W_n$  is an  $(n \times n)$  contiguity matrix, and  $\lambda_0$  is the spatial autoregressive parameter.

Model (1) can be expressed in terms of observed and unobserved variables as follows:

$$\begin{pmatrix} y_n^o \\ y_n^u \end{pmatrix} = \lambda_0 W_n \begin{pmatrix} y_n^o \\ y_n^u \end{pmatrix} + X_n \beta_0 + e_n. \quad (2)$$

In this model,  $y_n^o$  is an  $(n^o \times 1)$  sub-vector of observed outcomes from  $y_n$ , where  $n^o$  is the total number of observed  $y$ 's. In the same line,  $y_n^u$  is an  $(n^u \times 1)$  sub-vector of unobserved outcomes from  $y_n$ , where  $n^u$  is the total number of unobserved  $y$ 's, with  $n = n^o + n^u$ .

Let  $J_n^u = [0_{n^u \times n^o}, I_{n^u \times n^u}]$  be an  $(n^u \times n)$  selection matrix that chooses the unobserved elements from  $y_n$ . In the same line, let  $J_n^o = [I_{n^o \times n^o}, 0_{n^o \times n^u}]$  be an  $(n^o \times n)$  selection matrix that chooses the observed elements from  $y_n$ . Thus,  $y_n^o = J_n^o y_n$  and  $y_n^u = J_n^u y_n$ .

As mentioned above, we work only with those equations for which the dependent variable is observed. We derive such equations by pre-multiplying (2) by  $J_n^o$ :

$$y_n^o = \lambda_0 J_n^o W_n \begin{pmatrix} y_n^o \\ y_n^u \end{pmatrix} + J_n^o X_n \beta_0 + J_n^o e_n. \quad (3)$$

We can think of model (1) as a structural model; and we estimate the parameters using (3).

Little (1992) states that, as data set  $X$  is complete, and  $y$  are missing at random, imputing unobserved  $y$ 's does not add information into the regression of  $y$  on  $X$ . Based on this statement, and trying to impute as few observations as possible, we only impute missing data in  $J_n^o W_n \begin{pmatrix} y_n^o \\ y_n^u \end{pmatrix}$ , which is the spatial lag of observed dependent variables. Imputing the spatial

lag not necessary implies imputing all unobserved data. The level of imputation is the main difference with the estimators proposed in Wang and Lee (2013).

From equation (1),  $y_n = (I_n - \lambda_0 W_n)^{-1} X_n \beta_0 + (I_n - \lambda_0 W_n)^{-1} e_n$ . Applying expectations, we obtain  $E(y_n^u) = J_n^u E(y_n) = J_n^u (I_n - \lambda_0 W_n)^{-1} X_n \beta_0 = F_n(\theta_0)$ , with  $\theta_0 = (\lambda_0, \beta_0')$ .

Therefore, we can express equation (3) as:

$$y_n^o = J_n^o \lambda_0 W_n \begin{pmatrix} y_n^o \\ F_n(\theta_0) \end{pmatrix} + J_n^o X_n \beta_0 + J_n^o u_n, \quad (4)$$

where  $u_n = [\lambda_0 W_n J_n^{u'} J_n^u S_n^{-1}(\lambda_0) + I_n] e_n$ , with  $S_n(\lambda_0) = (I_n - \lambda_0 W_n)$ .<sup>1</sup>

The expectation of  $y_n$  on the right hand side of equation (4) depends on unknown parameters, which must be estimated. Assuming that  $\hat{\theta}$  is a consistent estimator and replacing in (4), we obtain:

$$y_n^o = J_n^o \lambda_0 W_n \begin{pmatrix} y_n^o \\ F_n(\hat{\theta}) \end{pmatrix} + J_n^o X_n \beta_0 + J_n^o \tilde{u}_n, \quad (5)$$

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

With some algebra (see appendix A) we arrive at

$$\tilde{u}_n = u_n - \lambda_0 W_n J_n^{u'} J_n^u S_n^{-1}(\lambda_0) C_n (\hat{\theta} - \theta_0) + \lambda_0 W_n J_n^{u'} J_n^u R_n (\hat{\lambda} - \lambda_0), \quad (6)$$

with  $R_n = [S_n^{-1}(\hat{\lambda}) G_n X_n \hat{\beta} - S_n^{-1}(\lambda_0) G_n X_n \beta_0]$  and  $C_n = (G_n X_n \beta_0 X_n)$ .

Following Wang and Lee (2013), we replace  $\hat{\theta}$  for a Non-Linear Least Squares estimator, which is consistent with an asymptotic distribution given by:

$$\sqrt{n}(\hat{\theta}_{nls,n} - \theta) = \left[ \frac{1}{n} C_n' B_n' B_n C_n \right]^{-1} \frac{1}{\sqrt{n}} C_n' B_n' B_n e_n + o_p(1) \xrightarrow{d} N(0, \Sigma_{nls}). \quad (7)$$

Replacing (7) into (6) and using some algebra (see appendix A), we obtain:

$$y_n^o = J_n^o \tilde{Z}_n \theta_0 + J_n^o (H_n(\lambda_0) e_n + R^*), \quad (8)$$

where:  $\tilde{Z}_n = (W_n \tilde{y} X_n)$ , with  $\tilde{y} = \begin{pmatrix} y_n^o \\ F_n(\hat{\theta}_{nls}) \end{pmatrix}$ ;  $\theta_0 = (\lambda_0 \beta_0)'$  and

$$H_n(\lambda_0) = \left\{ \lambda_0 W_n J_n^{u'} J_n^u S_n^{-1}(\lambda_0) + I_n - \lambda_0 W_n J_n^{u'} J_n^u S_n^{-1}(\lambda_0) C_n [C_n' B_n' B_n C_n]^{-1} C_n' B_n' B_n \right\}.$$

### 3. Estimators

This section presents three alternative estimators for model (8). The first one is based on the IBG2SLS estimator proposed in Wang and Lee (2013). The second one is similar, except for the use of an approximation to the optimal instrument matrix, as presented in Kelejian et al. (2004). The last one is  $\sqrt{n}$  equivalent to the initial one.

<sup>1</sup>See appendix A for more details.

### 3.1 Assumptions

**Assumption 1:**  $e_i \sim i.i.d. (0, \sigma_e^2)$ ,  $\forall i = 1, \dots, n$ , with  $0 < \sigma_e^2 < \infty$ .

**Assumption 2:** The elements of  $X_n$  are uniformly bounded constants;  $X_n$  has full rank  $k$ ; and  $\lim_{n \rightarrow \infty} \frac{1}{n} X_n' X_n$  exists and is non-singular.

**Assumption 3:** Spatial weight matrices  $W_n$  and  $S_n^{-1}(\lambda)$  are uniformly bounded in both row and column sums in absolute value, for  $|\lambda| < 1$ .

This assumption ensures that  $S_n(\lambda)$  is not singular. Furthermore, this allows to express  $S_n^{-1}(\lambda) = \sum_{k=0}^{\infty} \lambda^k W_n^k$ .

**Assumption 4:** Proportion  $\frac{n^0}{n}$  of observations tends to  $c$ , where  $c$  is a finite positive constant, as  $n$  approaches infinity.

This assumption indicates that the number of non-missing observations should not be too small relative to  $n$ . Since  $n^0$  remains a share of  $n$  (for a large value of  $n$ ), the rate of convergence of an estimator can be expressed as a function of  $n$ .

**Assumption 5:** The instrument matrix  $Q_n^o$  is a function of  $J_n^o X_n$  and  $J_n^o W_n$ . The elements of  $Q_n^o$  should be uniformly bounded. Furthermore,  $\lim_{n \rightarrow \infty} \frac{1}{n} Q_n^{o'} Q_n^o$  is non-singular and  $\lim_{n \rightarrow \infty} \frac{1}{n} Q_n^{o'} [J_n^o G_n X_n \beta_0 J_n^o X_n]$  has full column rank.

### Two-Stage Least Square Estimator With Imputation (I2SLS)

As mentioned above, unlike the estimator proposed in Wang and Lee (2013), in the I2SLS estimator we only impute  $J_n^o W_n y_n$ , which does not necessarily require to impute all  $y_n$ , but only those unobserved  $y$ 's belonging to the spatial lag of  $y^o$ . This partial imputation gives rise to the possibility of working only with complete data; i.e., if we observe  $J_n^o W_n y_n$ , we do not need to perform any imputations.

The I2SLS estimator is given by:

$$\hat{\theta}_{i2sls,n} = \left[ \tilde{Z}_n^{o'} Q_n^o (Q_n^{o'} Q_n^o)^{-1} Q_n^{o'} \tilde{Z}_n^o \right]^{-1} \tilde{Z}_n^{o'} Q_n^o (Q_n^{o'} Q_n^o)^{-1} Q_n^{o'} y_n^o, \quad (9)$$

where  $Q_n^o$  is an instrument matrix with rank  $\geq k+1$ , which is a function of  $J_n^o W_n$  and  $J_n^o X_n$ ; and  $\tilde{Z}_n^o = J_n^o \tilde{Z}_n$ .

**Proposition 1:** Under assumptions 1-5,  $\hat{\theta}_{i2sls,n}$  is a consistent estimator of  $\theta_0$  and  $\sqrt{n}(\hat{\theta}_{i2sls,n} - \theta_0) \xrightarrow{d} N(0, \Sigma_{i2sls})$ , where

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

## Generalized Two-Stage Least Squares Estimator with Imputation (IG2SLS)

Here we propose a generalized version of the previous estimator using the error term structure for weighting observations.

The asymptotic variance of model (8) is:

$$Avar(\tilde{u}_n^o) = Avar(H_n^o(\lambda_0)e_n + J_n^o R^*) = \sigma_e^2 \left( H_n^o(\lambda_0) H_n^{o'}(\lambda_0) \right),$$

with  $H_n^o(\lambda_0) = J_n^o H_n(\lambda_0)$  and  $J_n^o R^* \xrightarrow{p} 0$ . Given that  $\Omega(\lambda_0) = H_n^o(\lambda_0) H_n^{o'}(\lambda_0)$ , then  $Avar(\tilde{u}_n^o) = \sigma_e^2 \Omega(\lambda_0)$ .

**Assumption 6:** Let  $\Omega_n(\lambda_0)$  be an  $(n^o \times n^o)$  uniformly positive-definite and symmetric matrix, with  $\lim_{n \rightarrow \infty} \frac{1}{n} \Omega_n(\lambda_0)$  positive-definite and symmetric. Furthermore,  $\lim_{n \rightarrow \infty} \frac{1}{n} Q_n^{o'} \Omega_n^{-1}(\lambda_0) Q_n^o$  is finite and non-singular; and  $\lim_{n \rightarrow \infty} \frac{1}{n} Q_n^{o'} \Omega_n^{-1}(\lambda_0) [J_n^o G_n X_n \beta_0 J_n^o X_n]$  has full column rank.

The IG2SLS estimator is given by:

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

**Proposition 2:** Under assumptions 1-4 and 6,  $\hat{\theta}_{ig2sls,n}$  is a consistent estimator of  $\theta_0$  and  $\sqrt{n}(\hat{\theta}_{ig2sls,n} - \theta_0) \xrightarrow{d} N(0, \Sigma_{ig2sls})$ , where

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

## Best Generalized Two-Stage Least Squares Estimator with Imputation (IBG2SLS)

Here we derive an optimal instrument matrix that can be used to obtain the best estimator, where “best” is defined as the most efficient estimator of its kind.<sup>2</sup>

For the variance of IG2SLS to be minimal, the term in brackets in  $\Sigma_{ig2sls}$  should be maximal.

Replacing  $\Omega_n^{-1}(\lambda_0)$  by  $K_n' K_n$ , we obtain  $C_n^{o'} K_n' K_n Q_n^o \left( Q_n^{o'} K_n' K_n Q_n^o \right)^{-1} Q_n^{o'} K_n' K_n C_n^o$ . Defining  $A_n = (K_n C_n^o)'$ ,  $X_n = K_n Q_n^o$  and  $V_n = I_n$ , and applying the Generalized Cauchy-Schwarz Inequality,<sup>3</sup> we obtain:

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

where  $(K_n C_n^o)' (K_n C_n^o)$  is the upper bound of the term we want to maximize. Such bound is reached when  $Q_n^o = C_n^o$ , which implies that the instrument matrix that minimizes the variance of IG2SLS is  $Q_n^{o*} = C_n^o$ .

<sup>2</sup>See appendix C for details.

<sup>3</sup>The Generalized Cauchy-Schwarz Inequality states that if  $V_n$  is an  $(n \times n)$  positive-definite and symmetric matrix,  $A_n$  and  $X_n'$  are  $(n \times k)$  matrices, and  $X_n$  is of rank  $k$ , then  $A_n V_n A_n' \geq A_n X_n \left( X_n' V_n^{-1} X_n \right)^{-1} X_n' A_n$ .

Considering the above, the IBG2SLS estimator is given by:

$$\begin{aligned} \hat{\theta}_{ibg2sls,n} &= \left[ \tilde{Z}_n^{o'} \Omega_n^{-1}(\lambda_0) Q_n^{o*} \left( Q_n^{o*'} \Omega_n^{-1}(\lambda_0) Q_n^{o*} \right)^{-1} Q_n^{o*'} \Omega_n^{-1}(\lambda_0) \tilde{Z}_n^o \right]^{-1} \\ &\quad \tilde{Z}_n^{o'} \Omega_n^{-1}(\lambda_0) Q_n^{o*} \left( Q_n^{o*'} \Omega_n^{-1}(\lambda_0) Q_n^{o*} \right)^{-1} Q_n^{o*'} \Omega_n^{-1}(\lambda_0) y_n^o. \end{aligned} \quad (10)$$

The Asymptotic distribution of IBG2SLS is the same as that of IG2SLS, the only difference being that we define  $Q_n^{o*} = C_n^o$  as instrument matrix, which causes some terms to vanish asymptotically.

**Proposition 3:** Under assumptions 1-4 and 6,  $\hat{\theta}_{ibg2sls,n}$  is a consistent estimator of  $\theta_0$ , and  $\sqrt{n}(\hat{\theta}_{ibg2sls,n} - \theta_0) \xrightarrow{d} N(0, \Sigma_{ibg2sls})$ , where

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

### Series-Type Efficient Two-Stage Estimator with Imputation (IST2SLS)

The optimal instrument matrix derived above is the same optimal instrument matrix proposed in Kelejian and Prucha (1998). In a later work (Kelejian et al., 2004), the authors developed an estimator based on an approximation of such optimal matrix, named ‘‘Series-Type Efficient IV estimator’’. We propose to use that approximation to obtain an alternative estimator named ‘‘Series-Type Efficient Two- Stage Estimator with Imputation’’, IST2SLS.

**Approximation to the optimal instrument matrix:** Given assumption 3, we can express  $E(y_n) = \sum_{k=0}^{\infty} \lambda_0^k W_n^k X_n \beta_0$ . Replacing the latter into the optimal instrument matrix, we obtain  $E(Z_n) = [W_n E(y_n) X_n] = [\sum_{k=0}^{\infty} \lambda_0^k W_n^{k+1} X_n \beta_0 X_n]$ . Kelejian et al. (2004) state that, if  $r_n$  is a sequence of natural numbers, such that  $r_n \uparrow \infty$ , the optimal instrument matrix can be approximated by  $\hat{Q}_n^{okp} = [\sum_{k=0}^{r_n} \lambda_0^k W_n^{k+1} X_n \hat{\beta}_0 X_n]$ , where  $\hat{\theta} = (\hat{\lambda}_0 \hat{\beta}_0)'$  should be a  $\sqrt{n}$ -consistent estimator of  $\theta$ .

Following the above, the IST2SLS estimator is given by:

$$\begin{aligned} \hat{\theta}_{ist2sls,n} &= \left[ \tilde{Z}_n^{o'} \Omega_n^{-1}(\lambda_0) \hat{Q}_n^{okp} \left( \hat{Q}_n^{okp'} \Omega_n^{-1}(\lambda_0) \hat{Q}_n^{okp} \right)^{-1} \hat{Q}_n^{okp'} \Omega_n^{-1}(\lambda_0) \tilde{Z}_n^o \right]^{-1} \\ &\quad \tilde{Z}_n^{o'} \Omega_n^{-1}(\lambda_0) \hat{Q}_n^{okp} \left( \hat{Q}_n^{okp'} \Omega_n^{-1}(\lambda_0) \hat{Q}_n^{okp} \right)^{-1} \hat{Q}_n^{okp'} \Omega_n^{-1}(\lambda_0) y_n^o, \end{aligned}$$

where  $\hat{Q}_n^{okp} = J_n^o \hat{Q}_n^{kp} = [J_n^o \sum_{k=0}^{r_n} \lambda_0^k W_n^{k+1} X_n \hat{\beta}_{nls} J_n^o X_n]$ .

The asymptotic distribution of IST2SLS is the same as that of IG2SLS, but considering that we use  $Q_n^o = \hat{Q}_n^{okp}$  as instrument matrix, and given that  $\frac{1}{n} \hat{Q}_n^{okp} \xrightarrow{p} \frac{1}{n} C_n^o$ , we arrive at the following proposition:

**Proposition 4:** Under assumptions 1-4 and 6, and considering that  $\sqrt{n}(\hat{\lambda}_{nls} - \lambda_0) = O_p(1)$  with  $|\lambda_0| < 1$ ; that  $(\hat{\beta}_{nls} - \beta_0) = o_p(1)$ ; and that  $r_n$  is a sequence of natural numbers with  $0 \leq r_n \leq n$ ,  $r_n \uparrow \infty$  and  $r_n = o(\sqrt{n})$ , then  $\hat{\theta}_{ist2sls,n}$  is a consistent estimator of  $\theta_0$  and  $\sqrt{n}(\hat{\theta}_{ist2sls,n} - \theta_0) \xrightarrow{d} N(0, \Sigma_{ist2sls})$ , where

$$\Sigma_{ist2sls} = \sigma_e^2 \lim_{n \rightarrow \infty} \left[ \frac{1}{n} C_n^{o'} \Omega_n^{-1}(\lambda_0) C_n^o \right]^{-1}.$$

## Asymptotic Best Generalized Two-Stage Least Squares Estimator with Imputation (AIBG2SLS)

The following estimator results from an asymptotic property known as *asymptotic equivalence*. The AIBG2SLS estimator is given by:

$$\hat{\theta}_{aibg2sls,n} = \left( C_n^{o'} \Omega_n^{-1}(\lambda_0) C_n^o \right)^{-1} C_n^{o'} \Omega_n^{-1}(\lambda_0) y_n^o.$$

Given that<sup>4</sup>  $\sqrt{n}(\hat{\theta}_{ibg2sls,n} - \hat{\theta}_{aibg2sls,n}) = o_p(1)$ , the estimators IBG2SLS and AIBG2SLS are both  $\sqrt{n}$ -equivalent, and therefore have the same asymptotic distribution.

**Proposition 5:** *Under assumptions 1-4 and 6,  $\hat{\theta}_{aibg2sls,n}$  is a consistent estimator of  $\theta_0$ , and  $\sqrt{n}(\hat{\theta}_{aibg2sls,n} - \theta_0) \xrightarrow{d} N(0, \Sigma_{aibg2sls})$ , where*

$$\Sigma_{aibg2sls} = \sigma_e^2 \lim_{n \rightarrow \infty} \left[ \frac{1}{n} C_n^{o'} \Omega_n^{-1}(\lambda_0) C_n^o \right]^{-1}.$$

### 3.2 Error Variance Estimator

In this section we present an estimator of the error variance,  $\sigma_e^2$ , which is based on a non-linear least squares estimation.

The model  $J_n^o y_n = J_n^o S_n^{-1}(\lambda_0) X_n \beta_0 + J_n^o S_n^{-1}(\lambda_0) e_n$  can be estimated using NLS. Its variance is given by  $\sigma_e^2 \Sigma_n(\lambda_0)$ , where  $\Sigma_n(\lambda_0) = [J_n^o S_n^{-1}(\lambda_0) S_n^{-1'}(\lambda_0) J_n^{o'}]$ .

Using the Cholesky decomposition, we obtain  $\Sigma_n^{-1}(\lambda_0) = T_n'(\lambda_0) T_n(\lambda_0)$ , where  $T_n(\lambda_0)$  is an  $(n^o \times n^o)$  lower triangular matrix. Multiplying the error term by  $T_n(\lambda_0)$ , we obtain:

$$E[T_n(\lambda_0) J_n^o S_n^{-1}(\lambda_0) e_n] = 0, \text{Var}[T_n(\lambda_0) J_n^o S_n^{-1}(\lambda_0) e_n] = \sigma_e^2 I_n, \text{ and} \\ \text{cov}[T_n(\lambda_0) J_n^o S_n^{-1}(\lambda_0) e_n] = 0.$$

Given that  $v_i$  is the  $i$ -th element of  $v_n = T_n(\lambda_0) J_n^o S_n^{-1}(\lambda_0) e_n$ , by the weak law of large numbers we obtain  $\frac{1}{n} \sum_{i=1}^n v_i^2 \xrightarrow{p} \frac{1}{n} \sum_{i=1}^n E(v_i^2) = \sigma_e^2$ . Therefore, the proposed estimator for the error term variance is:

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

which is a consistent estimator of  $\sigma_e^2$ . See appendix F.

## 4. Finite Sample Properties

To investigate the performance of the proposed estimators with finite samples, we conduct a Monte Carlo experiment<sup>5</sup> designed as follows. The model is  $y = \lambda W y + \beta_1 + \beta_2 X_2 + e$ , with  $\beta_1 = \beta_2 = 1$ ,  $\lambda = 0.4$ , and both  $X_2$  and  $e \sim \mathcal{N}(0, 1I_n)$ . We consider three sample sizes: small ( $n = 67$ ), medium ( $n = 217$ ), and large ( $n = 417$ ). For all sample sizes we fixed three missing rates, namely, 10%, 25%, and 50%. The number of repetitions is 500.

<sup>4</sup>See appendix E for details.

<sup>5</sup>We use the R software, which can be freely downloaded from <http://www.r-project.org/>.



The contiguity matrices  $W$  were constructed as follows. For each repetition we simulate pairs of latitude and longitude points ( $n$  points) using two uniform distributions  $U(0, 1)$ . Once we have the  $n$  coordinates (spatial locations), we create  $W$  following the  $k$  nearest neighborhood criteria, with  $k = \{4, 8\}$ . We set two different values for  $k$  to assess the density impact.<sup>6</sup>

We compare the estimation results for IBG2SLS, IST2SLS, and IBG2SLSA. For comparison, we also report the results of one of the estimators proposed in Wang and Lee (2013), the IBG2SLS(W-L). As pointed out previously, the only difference between our proposed estimators and IBG2SLS(W-L) is that we impute the unobserved  $y$ 's in the spatial lag,  $J_n^o W_n y_n$ , while IBG2SLS(W-L) imputes all unobserved  $y$ 's; i.e., all  $y_n^u$ .

Regarding the IST2SLS estimator, we need to set a value for  $r_n$ . In their simulations, Kelejian et al. (2004) try different values for  $r_n$  and show that setting  $r_n = n^\alpha$ , with  $\alpha = 0.25$ , yields good enough results, so we follow this criterion.

## 4.1 Results

For the IBG2SLS, IST2SLS, and IBG2SLSA estimations we report bias, root mean square error (RMSE), and true and estimated standard deviation (except for  $\sigma_e$ ). For comparison we also report bias and RMSE for IBG2SLS(W-L).<sup>7</sup>

Table 1 presents the results for  $k = 4$ . Regarding samples sizes, the results are as expected; i.e., the larger the sample size, the smaller the bias and variability. One thing to note is that a missing rate appears not to be as important as the total number of observed data; e.g., if we test for bias and RSME for  $n = 417$ , we see that the larger missing rate has better results than any missing rate for  $n = 217$ .

For the smaller sample size, the AIBG2SLS shows slightly better results in terms of variability, although for the rest of samples sizes and missing rates, the results for all estimations are quite similar.

For medium and large sample sizes, the estimated standard deviation and the true one are mostly equal; and as the observed data grows, they mostly match with RSME.

Table 2 reports results for  $k = 8$ . First, as the literature suggests, density is positively and highly related with bias and variability; and all results are greater than for  $k = 4$ . Despite this fact, the results are quite similar to the previous ones; i.e., as sample sizes grow, bias and variability decrease, and the total number of observed data becomes more relevant than the missing rate. For the smaller sample size, AIBG2SLS shows better results in terms of both bias and variability.

Summarizing, the only significant difference between estimators is that AIBG2SLS performs better for smaller samples, while for larger samples the performance of all four estimators is practically the same.

<sup>6</sup>Some works show that density negatively impacts on bias and inference. See Smith (2009); Farber et al. (2010), among others.

<sup>7</sup>We do not report the standard deviation for IBG2SLS(W-L), as we only seek to compare its empirical variability.

**Table 1**  
Simulation for  $k = 4$ .

$n$	$n^o$		IBG2SLS(W-L)		IBG2SLS			IST2SLS			AIBG2SLS		
			Bias	RMSE	Bias	RMSE	SD	Bias	RMSE	SD	Bias	RMSE	SD
67	33	$\lambda_0$	0.073	0.357	0.072	0.363	0.276 (0.263)	0.072	0.338	0.276 (0.263)	0.069	0.352	0.277 (0.263)
		$\beta_1$	-0.126	0.614	-0.124	0.626	0.483 (0.461)	-0.124	0.582	0.483 (0.461)	-0.119	0.607	0.484 (0.461)
		$\beta_2$	0.012	0.205	0.013	0.206	0.185 (0.190)	0.008	0.201	0.185 (0.190)	0.012	0.204	0.185 (0.190)
		$\sigma_e$	-	-	0.055	0.267	-	0.055	0.267	-	0.055	0.267	-
	51	$\lambda_0$	0.048	0.295	0.047	0.299	0.237 (0.228)	0.047	0.247	0.237 (0.228)	0.048	0.301	0.238 (0.228)
		$\beta_1$	-0.090	0.534	-0.088	0.542	0.424 (0.404)	-0.088	0.499	0.424 (0.404)	0.090	0.546	0.425 (0.404)
		$\beta_2$	0.017	0.149	0.017	0.150	0.146 (0.148)	0.015	0.148	0.146 (0.148)	0.017	0.150	0.146 (0.148)
		$\sigma_e$	-	-	0.020	0.217	-	0.020	0.217	-	0.020	0.217	-
	61	$\lambda_0$	0.024	0.260	0.023	0.263	0.215 (0.212)	0.023	0.241	0.215 (0.212)	0.025	0.264	0.215 (0.212)
		$\beta_1$	-0.029	0.457	-0.028	0.462	0.381 (0.377)	-0.031	0.425	0.381 (0.377)	-0.031	0.464	0.382 (0.377)
		$\beta_2$	0.018	0.137	0.019	0.137	0.133 (0.133)	0.018	0.134	0.133 (0.133)	0.018	0.137	0.133 (0.133)
		$\sigma_e$	-	-	0.004	0.198	-	0.004	0.198	-	0.004	0.198	-
109	$\lambda_0$	0.009	0.141	0.008	0.141	0.137 (0.136)	0.008	0.140	0.137 (0.136)	0.009	0.141	0.137 (0.136)	
	$\beta_1$	-0.014	0.248	-0.012	0.248	0.242 (0.242)	-0.012	0.246	0.242 (0.242)	-0.013	0.248	0.243 (0.242)	
	$\beta_2$	0.004	0.100	0.005	0.100	0.100 (0.101)	0.005	0.101	0.100 (0.101)	0.004	0.100	0.100 (0.101)	
	$\sigma_e$	-	-	0.015	0.146	-	0.015	0.146	-	0.015	0.146	-	
217	$\lambda_0$	0.005	0.120	0.004	0.120	0.118 (0.117)	0.004	0.122	0.118 (0.117)	0.004	0.120	0.118 (0.117)	
	$\beta_1$	-0.009	0.214	-0.008	0.214	0.210 (0.208)	-0.008	0.217	0.210 (0.208)	-0.008	0.214	0.210 (0.208)	
	$\beta_2$	0.007	0.083	0.007	0.083	0.080 (0.080)	0.008	0.085	0.080 (0.080)	0.007	0.083	0.080 (0.080)	
	$\sigma_e$	-	-	0.001	0.121	-	0.001	0.121	-	0.001	0.121	-	
195	$\lambda_0$	-0.003	0.109	-0.003	0.109	0.111 (0.110)	-0.001	0.110	0.111 (0.110)	-0.003	0.109	0.111 (0.110)	
	$\beta_1$	0.007	0.195	0.007	0.195	0.198 (0.196)	0.005	0.199	0.198 (0.196)	0.007	0.195	0.198 (0.196)	
	$\beta_2$	0.010	0.075	0.010	0.075	0.073 (0.073)	0.010	0.075	0.073 (0.073)	0.010	0.075	0.073 (0.073)	
	$\sigma_e$	-	-	-0.019	0.108	-	-0.019	0.108	-	-0.019	0.108	-	
209	$\lambda_0$	0.004	0.098	0.003	0.098	0.097 (0.097)	0.005	0.098	0.097 (0.097)	0.004	0.098	0.097 (0.097)	
	$\beta_1$	-0.008	0.165	-0.007	0.165	0.171 (0.170)	-0.009	0.167	0.171 (0.170)	-0.007	0.165	0.171 (0.170)	
	$\beta_2$	0.000	0.075	0.001	0.075	0.072 (0.073)	0.001	0.076	0.072 (0.073)	0.001	0.075	0.072 (0.073)	
	$\sigma_e$	-	-	0.007	0.103	-	0.007	0.103	-	0.007	0.103	-	
417	$\lambda_0$	0.005	0.089	0.005	0.089	0.083 (0.083)	0.005	0.090	0.083 (0.083)	0.005	0.089	0.083 (0.083)	
	$\beta_1$	-0.007	0.159	-0.006	0.159	0.149 (0.148)	-0.008	0.161	0.149 (0.148)	-0.006	0.159	0.149 (0.148)	
	$\beta_2$	0.000	0.059	0.000	0.059	0.058 (0.058)	0.000	0.060	0.058 (0.058)	0.000	0.059	0.058 (0.058)	
	$\sigma_e$	-	-	-0.009	0.087	-	-0.009	0.087	-	-0.009	0.087	-	
376	$\lambda_0$	0.002	0.078	0.002	0.079	0.078 (0.078)	0.003	0.079	0.078 (0.078)	0.002	0.078	0.078 (0.078)	
	$\beta_1$	-0.006	0.139	-0.005	0.139	0.140 (0.139)	-0.006	0.140	0.140 (0.139)	-0.005	0.139	0.140 (0.139)	
	$\beta_2$	0.000	0.056	0.000	0.056	0.053 (0.052)	0.000	0.056	0.053 (0.052)	0.000	0.056	0.053 (0.052)	
	$\sigma_e$	-	-	-0.017	0.078	-	-0.017	0.078	-	-0.017	0.078	-	

Notes: True standard errors in parenthesis.

**Table 2**  
Simulation for  $k = 8$ .

$n$	$n^o$		IBG2SLS(W-L)		IBG2SLS			IST2SLS			AIBG2SLS		
			Bias	RMSE	Bias	RMSE	SD	Bias	RMSE	SD	Bias	RMSE	SD
67	33	$\lambda_0$	0.162	0.627	0.177	0.722	0.427 (0.402)	0.143	0.549	0.427 (0.402)	0.176	0.966	0.438 (0.402)
		$\beta_1$	-0.285	1.084	-0.304	1.170	0.738 (0.691)	-0.255	0.973	0.738 (0.691)	-0.298	1.468	0.756 (0.691)
		$\beta_2$	0.026	0.204	0.028	0.210	0.178 (0.185)	0.026	0.199	0.178 (0.185)	0.032	0.231	0.179 (0.185)
		$\sigma_e$	-	-	0.090	0.259	-	0.090	0.259	-	0.090	0.259	-
	51	$\lambda_0$	0.114	0.491	0.109	0.476	0.359 (0.339)	0.100	0.417	0.359 (0.339)	0.129	0.704	0.360 (0.339)
		$\beta_1$	-0.203	0.843	-0.197	0.829	0.623 (0.584)	-0.185	0.755	0.623 (0.584)	-0.223	1.060	0.626 (0.584)
		$\beta_2$	0.022	0.151	0.022	0.150	0.143 (0.146)	0.020	0.147	0.143 (0.146)	0.024	0.165	0.143 (0.146)
		$\sigma_e$	-	-	0.049	0.206	-	0.049	0.206	-	0.049	0.206	-
	61	$\lambda_0$	0.067	0.417	0.067	0.420	0.320 (0.314)	0.056	0.355	0.320 (0.314)	0.070	0.436	0.321 (0.315)
		$\beta_1$	-0.102	0.685	-0.102	0.687	0.548 (0.540)	-0.090	0.590	0.548 (0.540)	-0.108	0.716	0.550 (0.540)
		$\beta_2$	0.003	0.144	0.003	0.144	0.131 (0.133)	0.002	0.142	0.131 (0.133)	0.003	0.144	0.131 (0.133)
		$\sigma_e$	-	-	0.025	0.194	-	0.025	0.194	-	0.025	0.194	-
109	$\lambda_0$	0.029	0.220	0.028	0.221	0.195 (0.190)	0.030	0.218	0.195 (0.190)	0.029	0.220	0.195 (0.190)	
	$\beta_1$	-0.047	0.380	-0.048	0.381	0.338 (0.328)	-0.051	0.377	0.338 (0.328)	-0.048	0.380	0.338 (0.328)	
	$\beta_2$	0.003	0.097	0.004	0.097	0.098 (0.099)	0.003	0.097	0.098 (0.099)	0.003	0.097	0.098 (0.099)	
	$\sigma_e$	-	-	0.018	0.145	-	0.018	0.145	-	0.018	0.145	-	
217	$\lambda_0$	0.036	0.189	0.035	0.189	0.168 (0.163)	0.035	0.188	0.168 (0.163)	0.035	0.189	0.168 (0.163)	
	$\beta_1$	-0.066	0.331	-0.065	0.332	0.291 (0.282)	-0.064	0.330	0.291 (0.282)	-0.065	0.332	0.291 (0.282)	
	$\beta_2$	0.000	0.078	0.000	0.078	0.078 (0.078)	-0.001	0.078	0.078 (0.078)	0.000	0.078	0.078 (0.078)	
	$\sigma_e$	-	-	-0.002	0.116	-	-0.002	0.116	-	-0.002	0.116	-	
195	$\lambda_0$	0.014	0.168	0.014	0.168	0.157 (0.154)	0.016	0.168	0.157 (0.154)	0.014	0.168	0.157 (0.154)	
	$\beta_1$	-0.017	0.281	-0.016	0.281	0.270 (0.267)	-0.021	0.282	0.270 (0.267)	-0.017	0.281	0.270 (0.267)	
	$\beta_2$	0.005	0.073	0.005	0.073	0.072 (0.072)	0.005	0.074	0.072 (0.072)	0.005	0.073	0.072 (0.072)	
	$\sigma_e$	-	-	-0.004	0.104	-	-0.004	0.104	-	-0.004	0.104	-	
209	$\lambda_0$	0.014	0.143	0.013	0.143	0.135 (0.133)	0.014	0.142	0.135 (0.133)	0.013	0.143	0.135 (0.133)	
	$\beta_1$	-0.024	0.245	-0.023	0.245	0.232 (0.230)	-0.024	0.245	0.232 (0.230)	-0.023	0.245	0.232 (0.230)	
	$\beta_2$	-0.001	0.073	-0.002	0.073	0.070 (0.071)	-0.001	0.074	0.070 (0.071)	-0.002	0.073	0.070 (0.071)	
	$\sigma_e$	-	-	0.014	0.100	-	0.014	0.100	-	0.014	0.100	-	
417	$\lambda_0$	0.003	0.118	0.002	0.118	0.115 (0.115)	0.004	0.118	0.115 (0.115)	0.003	0.118	0.115 (0.115)	
	$\beta_1$	-0.007	0.207	-0.007	0.207	0.200 (0.200)	-0.009	0.208	0.200 (0.200)	-0.007	0.207	0.200 (0.200)	
	$\beta_2$	0.001	0.060	0.001	0.060	0.057 (0.057)	0.000	0.060	0.057 (0.057)	0.001	0.060	0.057 (0.057)	
	$\sigma_e$	-	-	0.006	0.082	-	0.006	0.082	-	0.006	0.082	-	
376	$\lambda_0$	0.006	0.121	0.006	0.121	0.109 (0.109)	0.009	0.122	0.109 (0.109)	0.006	0.121	0.109 (0.109)	
	$\beta_1$	-0.009	0.209	-0.008	0.209	0.189 (0.188)	-0.013	0.210	0.189 (0.188)	-0.009	0.209	0.189 (0.188)	
	$\beta_2$	0.003	0.052	0.003	0.052	0.052 (0.052)	0.003	0.052	0.052 (0.052)	0.003	0.052	0.052 (0.052)	
	$\sigma_e$	-	-	0.000	0.071	-	0.000	0.071	-	0.000	0.071	-	

Notes: True standard errors in parenthesis.

At the same time, although we do not present results for different values of  $\sigma_e$ , we can expect that, *given the model*, the smaller the  $\sigma_e$ , the better the estimations; and, given that the goodness of fit is negatively related with  $\sigma_e$ , the estimation performance improves with a better model fit.<sup>8</sup>

We do not find any regularities between the value of the spatial correlation coefficient and the estimators' performance.

## 5. Conclusions

The aim of this work is to present a series of estimators for dealing with random missing data in the dependent variable within a Spatial Lag Model. We propose three alternatives based on a Two-Stage Least Squares estimation methodology; derive the optimal instrument matrix to maximize efficiency; and present an approximation for it.

All estimators require to impute *some* missing data and this *partial* imputation (in contrast with total imputation) is the key difference relative to the estimators proposed in Wang and Lee (2013). Partial imputation gives rise to the possibility of working only with complete data, since our proposed estimators only require knowing the spatial lag for the observed dependent variable. If we observe  $J_n^o W_n y_n$ , we do not need to perform any imputations.

Given that the analytical results are for large samples (asymptotic results), we carry out a Monte Carlo simulation to assess the estimators' performance for finite samples sizes. The results show that there is no gain in imputing all unobserved data. Although in small samples our estimator AIBG2SLS shows better results in terms of bias and variability, for medium and large samples all estimators (including the one with total imputation) show a similar behaviour.

Imputation requires a previous estimation of the parameters. In this case we use Non-Linear Least Squares to obtain an initial estimation. This is not a constraint: many results are maintained if we use another initial estimation; the only requirement is that the estimator should be consistent.

A key assumption is that missing data are random. In some empirical applications, missing data are spatially grouped (e.g., missing data for a complete region or state). If we can assume that the missing data process is random (i.e., not related to the problem under analysis), the proposed estimators can be applied.

At the same time, as stated in Mur et al. (2008), spatial heterogeneity has not received much attention in literature, even considering that there is wide evidence that spatial behavioral relations tend to be unstable. Further research should seek to develop estimators under missing data conditions considering spatial heterogeneity.

---

<sup>8</sup>Thinking in the goodness of fit as the correlation between  $y$  and  $\hat{y}$ .

## Appendix A - Model

**Equation (4):** The vector of observed and unobserved variables can be expressed as:

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

Replacing this equality into (3) and distributing terms, we obtain:

$$y_n^o = J_n^o \lambda_0 W_n \begin{pmatrix} y_n^o \\ F_n(\theta_0) \end{pmatrix} + J_n^o X_n \beta_0 + J_n^o \lambda_0 W_n \begin{pmatrix} 0 \\ y_n^u - F_n(\theta_0) \end{pmatrix} + J_n^o e_n.$$

Given that  $(y_n^u - F_n(\theta_0)) = J_n^u S_n^{-1}(\lambda_0) e_n$ , with  $S_n(\lambda_0) = (I_n - \lambda_0 W_n)$ , and

$$\begin{pmatrix} 0 \\ y_n^u - F_n(\theta_0) \end{pmatrix} = J_n^{u'} J_n^u S_n^{-1}(\lambda_0) e_n,$$

then:

$$J_n^o \lambda_0 W_n \begin{pmatrix} 0 \\ y_n^u - F_n(\theta_0) \end{pmatrix} + J_n^o e_n = [\lambda_0 W_n J_n^{u'} J_n^u S_n^{-1}(\lambda_0) + I_n] e_n = u_n,$$

and so we obtain equation (4).

**Equation (5):** Given that  $\begin{pmatrix} y_n^o \\ F_n(\theta_0) \end{pmatrix} = \left[ \begin{pmatrix} y_n^o \\ F_n(\hat{\theta}) \end{pmatrix} - \begin{pmatrix} 0 \\ F_n(\hat{\theta}) - F_n(\theta_0) \end{pmatrix} \right]$ ,

and replacing on the right hand side of equation (4), we obtain:

$$y_n^o = J_n^o \lambda_0 W_n \begin{pmatrix} y_n^o \\ F_n(\hat{\theta}) \end{pmatrix} + J_n^o X_n \beta_0 - J_n^o \lambda_0 W_n \begin{pmatrix} 0 \\ F_n(\hat{\theta}) - F_n(\theta_0) \end{pmatrix} + J_n^o u_n.$$

Given that  $\begin{pmatrix} 0 \\ F_n(\hat{\theta}) - F_n(\theta_0) \end{pmatrix} = J_n^{u'} J_n^u [(S_n^{-1}(\hat{\lambda}) X_n \hat{\beta}) - (S_n^{-1}(\lambda_0) X_n \beta_0)]$ , we arrive at (5).

**Equation (6):** Consider  $S_n(\lambda_0) - S_n(\hat{\lambda}) = (I_n - \lambda_0 W_n) - (I_n - \hat{\lambda} W_n) = (\hat{\lambda} - \lambda_0) W_n$ .

Based on this equality, we obtain:

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

Given that  $W_n S_n^{-1}(\lambda_0) = G_n$ , and considering that  $(\hat{\lambda} - \lambda_0)$  is a scalar, we obtain:

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

Given the above, we arrive at:

$$\begin{aligned} [S_n^{-1}(\hat{\lambda}) X_n \hat{\beta} - S_n^{-1}(\lambda_0) X_n \beta_0] &= [(\hat{\lambda} - \lambda_0) S_n^{-1}(\hat{\lambda}) G_n] X_n \hat{\beta} + S_n^{-1}(\lambda_0) X_n (\hat{\beta} - \beta_0) \\ &= [S_n^{-1}(\hat{\lambda}) G_n X_n \hat{\beta} S_n^{-1}(\lambda_0) X_n] \begin{bmatrix} (\hat{\lambda} - \lambda_0) \\ (\hat{\beta} - \beta_0) \end{bmatrix} \\ &= S_n^{-1}(\lambda_0) C_n (\hat{\theta} - \theta_0) + R (\hat{\lambda} - \lambda_0), \end{aligned} \quad (11)$$

with  $(\hat{\theta} - \theta_0) = \begin{bmatrix} (\hat{\lambda} - \lambda_0) \\ (\hat{\beta} - \beta_0) \end{bmatrix}$ ,  $C_n = (G_n X_n \beta_0 X_n)$  and  $R = [S_n^{-1}(\hat{\lambda}) G_n X_n \hat{\beta} - S_n^{-1}(\lambda_0) G_n X_n \beta_0]$ .

Replacing the term in brackets for  $\tilde{u}_n$  in (5) by its equivalent (11), distributing, and ordering terms, we obtain (6).

**Equation (8):** Replacing (7) into (6), we arrive at:

$$\tilde{u}_n = u_n - \lambda_0 W_n J_n^{u'} J_n^u S_n^{-1}(\lambda_0) C_n \left[ [C_n' B_n' B_n C_n]^{-1} C_n' B_n' B_n e_n + o_p\left(\frac{1}{\sqrt{n}}\right) \right] + \lambda_0 W_n J_n^{u'} J_n^u R_n (\hat{\lambda} - \lambda_0).$$

Replacing  $u_n$  by its equivalent, reordering, and distributing terms, we obtain:

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

with  $R^* = \lambda_0 W_n J_n^{u'} J_n^u \left[ R_n (\hat{\lambda} - \lambda_0) - S_n^{-1}(\lambda_0) C_n l_k o_p\left(\frac{1}{\sqrt{n}}\right) \right]$ , where  $l_k$  is a  $k+1$  vector of unit values.

Given the above, by imputing  $\theta$  with  $\hat{\theta}_{nls}$ , model (5) can be expressed as (8).

## Appendix B - I2SLS Estimator

### Proof of proposition 1: Consistency

Considering that:

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

and  $\tilde{u}_n^o = J_n^o \tilde{u}_n$ , we need to prove that  $\left[ \frac{1}{n} \tilde{Z}_n^{o'} Q_n^o \left( \frac{1}{n} Q_n^{o'} Q_n^o \right)^{-1} \frac{1}{n} Q_n^{o'} \tilde{Z}_n^o \right]^{-1}$  converges to a well-defined limit with full rank.

Given assumption 5,  $\lim_{n \rightarrow \infty} \frac{1}{n} Q_n^{o'} Q_n^o$  is well-defined, so we only need to prove that  $\frac{1}{n} Q_n^{o'} \tilde{Z}_n^o = \left[ \frac{1}{n} Q_n^{o'} J_n^o W_n \tilde{y}_n \frac{1}{n} Q_n^{o'} J_n^o X_n \right]$  converges to a well-defined limit with full rank.

Given that  $\tilde{y}_n = J_n^{o'} J_n^o S_n^{-1}(\lambda_0) (X_n \beta_0 + e_n) + J_n^{u'} J_n^u S_n^{-1}(\hat{\lambda}) X_n \hat{\beta}$  and  $J_n^{o'} J_n^o = I_n - J_n^{u'} J_n^u$ , we obtain:

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

with  $A_n = [S_n^{-1}(\hat{\lambda}) X_n \hat{\beta} - S_n^{-1}(\lambda_0) X_n \beta_0]$ . Given the consistency of  $(\hat{\lambda}, \hat{\beta})$ ,  $A_n = o_p(1)$ . Furthermore,  $\frac{1}{n} Q_n^{o'} J_n^o W_n J_n^{o'} J_n^o S_n^{-1}(\lambda_0) e_n = o_p(1)$ .

Given the above,

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} Q_n^{o'} \tilde{Z}_n^o &= \lim_{n \rightarrow \infty} \frac{1}{n} Q_n^{o'} [J_n^o G_n X_n \beta_0 + o_p(1) J_n^o X_n] \\ &\xrightarrow{p} \lim_{n \rightarrow \infty} \frac{1}{n} Q_n^{o'} [J_n^o G_n X_n \beta_0 J_n^o X_n] \end{aligned} \quad (12)$$

Under assumption 5 this limit is well-defined.

With the previous results,  $\left[\frac{1}{n}\tilde{Z}_n'Q_n^o\left(\frac{1}{n}Q_n^{o'}Q_n^o\right)^{-1}\frac{1}{n}Q_n^{o'}\tilde{Z}_n^o\right]^{-1}$  converges in probability to a non-singular matrix.

At the same time:

$$\begin{aligned}\frac{1}{n}Q_n^{o'}\tilde{u}_n^o &= \frac{1}{n}Q_n^{o'}\left\{\left[\lambda_0W_nJ_n^{u'}J_n^uS_n^{-1}+I_n\right]e_n-\lambda_0W_nJ_n^{u'}J_n^u\left[\left(S_n^{-1}(\hat{\lambda})X_n\hat{\beta}\right)-\left(S_n^{-1}(\lambda_0)X_n\beta_0\right)\right]\right\} \\ &= o_p(1).\end{aligned}\quad (13)$$

Given (12) and (13), then  $(\hat{\theta}_{2sls,n}-\theta_0)\xrightarrow{p}0$ .

## Proof of proposition 1: Asymptotic Distribution

Given that:

$$\sqrt{n}(\hat{\theta}_{2sls,n}-\theta_0)=\left[\frac{1}{n}\tilde{Z}_n'Q_n^o\left(Q_n^{o'}Q_n^o\right)^{-1}Q_n^{o'}\tilde{Z}_n^o\right]^{-1}\tilde{Z}_n'Q_n^o\left(Q_n^{o'}Q_n^o\right)^{-1}\frac{1}{\sqrt{n}}Q_n^{o'}\tilde{u}_n^o,\quad (14)$$

in order to derive the asymptotic distribution of  $\sqrt{n}(\hat{\theta}_{2sls,n}-\theta_0)$  we need to obtain first the asymptotic distribution of  $\frac{1}{\sqrt{n}}Q_n^{o'}\tilde{u}_n^o$ .

$$\frac{1}{\sqrt{n}}Q_n^{o'}\tilde{u}_n^o=\frac{1}{\sqrt{n}}Q_n^{o'}J_n^o\left(H_n(\lambda_0)e_n+R^*\right)=\frac{1}{\sqrt{n}}Q_n^{o'}J_n^oH_n(\lambda_0)e_n+\frac{1}{\sqrt{n}}Q_n^{o'}J_n^oR^*.\quad (15)$$

Working on the second term in the last equality, we obtain:

$$\begin{aligned}\frac{1}{\sqrt{n}}Q_n^{o'}J_n^oR^* &= \frac{1}{\sqrt{n}}Q_n^{o'}J_n^o\left\{\lambda_0W_nJ_n^{u'}J_n^u\left[R_n(\hat{\lambda}-\lambda_0)-S_n^{-1}(\lambda_0)C_nl_ko_p\left(\frac{1}{\sqrt{n}}\right)\right]\right\} \\ &= \frac{1}{n}Q_n^{o'}J_n^o\lambda_0W_nJ_n^{u'}J_n^uR_n\sqrt{n}(\hat{\lambda}-\lambda_0)-\frac{1}{n}Q_n^{o'}J_n^oS_n^{-1}(\lambda_0)C_nl_ko_p(1) \\ &= o_p(1).\end{aligned}$$

Given  $R_n$ , the consistency of  $\hat{\theta}$ , and considering that the terms  $\frac{1}{n}Q_n^{o'}J_n^o\lambda_0W_nJ_n^{u'}J_n^u$  and  $\sqrt{n}(\hat{\lambda}-\lambda_0)$  are  $O_p(1)$ , then  $\frac{1}{n}Q_n^{o'}J_n^o\lambda_0W_nJ_n^{u'}J_n^uR_n=O_p\left(\frac{1}{n}\right)$ .

Additionally,  $\frac{1}{n}Q_n^{o'}J_n^oS_n^{-1}(\lambda_0)C_nl_k=O_p(1)$ ; thus,  $\frac{1}{n}Q_n^{o'}J_n^oS_n^{-1}(\lambda_0)C_nl_ko_p(1)=o_p(1)$ .

Given the above, we obtain:

$$\frac{1}{\sqrt{n}}Q_n^{o'}\tilde{u}_n^o=\frac{1}{\sqrt{n}}Q_n^{o'}H_n^o(\lambda_0)e_n+o_p(1),$$

with  $H_n^o(\lambda_0)=J_n^oH_n(\lambda_0)$ . Therefore,  $\frac{1}{\sqrt{n}}Q_n^{o'}\tilde{u}_n^o-\frac{1}{\sqrt{n}}Q_n^{o'}H_n^o(\lambda_0)e_n\xrightarrow{p}0$ .

Given this, based on the asymptotic equivalence lemma, the asymptotic distribution of  $\frac{1}{\sqrt{n}}Q_n^{o'}H_n^o(\lambda_0)e_n$  is the same as that of  $\frac{1}{\sqrt{n}}Q_n^{o'}\tilde{u}_n^o$ .

Given that  $H_n(\lambda_0)$  is absolutely summable, then  $Q_n^{o'}H_n^o(\lambda_0)=D_n^o$  is bounded. Assuming that  $\lim_{n\rightarrow\infty}\frac{1}{n}D_n^oD_n^{o'}$  is finite and positive definite,<sup>9</sup> given assumption 1, and using the Central Limit Theorem for triangular arrays presented in Kelejian and Prucha (1999) or Lee (2003), we obtain:

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

<sup>9</sup>Kelejian and Prucha (1999) prove that, given that  $A$  and  $B$  are two absolutely summable matrices, so is  $AB$ . Furthermore, given that  $Z$  is a bounded matrix, then  $Z'AZ=O(n)$ . Therefore,  $D_n^oD_n^{o'}=O(n)$ .

Given the above, and considering (14), then:

$$\sqrt{n}(\hat{\theta}_{i2sls,n} - \theta_0) \xrightarrow{d} N(0, \Sigma_{i2sls}),$$

with

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

In the variance equation we consider that  $\lim_{n \rightarrow \infty} \frac{1}{n} Q_n^{o'} \tilde{Z}_n^o \xrightarrow{p} \lim_{n \rightarrow \infty} \frac{1}{n} Q_n^{o'} C_n^o$ .

## Appendix C - IG2SLS Estimator

Under assumption 6,  $\Omega_n(\lambda_0) = H_n^o(\lambda_0) H_n^{o'}(\lambda_0)$  is an  $(n^o \times n^o)$  definite positive and invertible matrix, so we can apply the Cholesky decomposition and express it as  $\Omega_n^{-1}(\lambda_0) = K_n' K_n$ , where  $K_n$  is an  $(n^o \times n^o)$  invertible lower triangular matrix. Multiplying both sides of model (8) by  $K_n$ , we obtain:

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

Since our transformed model is a function of  $y_n^{o*} = K_n y_n^o$ ,  $\tilde{Z}_n^{o*} = K_n \tilde{Z}_n^o$ , and  $\tilde{u}_n^{o*} = K_n \tilde{u}_n^o$ , the instrument matrix will not be the same as before.

Given that:

$$K_n y_n^o = K_n (J_n^o X_n \beta_0 + \lambda_0 J_n^o W_n X_n \beta_o + \lambda_0^2 J_n^o W_n^2 X_n \beta_o + \dots) + K_n J_n^o (I_n - \lambda_0 W_n)^{-1} e_n,$$

the instrument matrix is now given by  $Q_n^{o*} = K_n Q_n^o$ . The IG2SLS estimator is:

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

## Proof of proposition 2: Consistency and Asymptotic Distribution

We provide the complete proof in appendix B. The difference between IG2SLS and I2SLS is that the former considers a weights matrix,  $\Omega_n^{-1}(\lambda_0)$ . As before, we need to prove that  $\frac{1}{n} Q_n^{o'} \Omega_n^{-1}(\lambda_0) \tilde{Z}_n^o$  converges to a well-defined limit with full rank. Introducing  $\Omega_n^{-1}(\lambda_0)$  into the derivation of (12), we arrive at  $\frac{1}{n} Q_n^{o'} \Omega_n^{-1}(\lambda_0) \tilde{Z}_n^o = \frac{1}{n} Q_n^{o'} \Omega_n^{-1}(\lambda_0) [J_n^o G_n X_n \beta_0 + o_p(1) J_n^o X_n]$ ; and from this equality we obtain  $\frac{1}{n} Q_n^{o'} \Omega_n^{-1}(\lambda_0) \tilde{Z}_n^o \xrightarrow{p} \lim_{n \rightarrow \infty} \frac{1}{n} Q_n^{o'} \Omega_n^{-1}(\lambda_0) [J_n^o G_n X_n \beta_0 J_n^o X_n]$ . Under assumption 6, this limit is well-defined. Following the derivations in appendix B, but considering  $\Omega_n^{-1}(\lambda_0)$  and assuming it is an absolutely summable matrix, we can derive that



$\frac{1}{\sqrt{n}} Q_n^{o'} \Omega_n^{-1}(\lambda_0) \tilde{u}_n^o \xrightarrow{d} N(0, \sigma_e^2 \lim_{n \rightarrow \infty} \frac{1}{n} Q_n^{o'} \Omega_n^{-1}(\lambda_0) Q_n^o)$ . This implies that  $\sqrt{n}(\hat{\theta}_{ig2sls,n} - \theta_0) \xrightarrow{d} N(0, \Sigma_{ig2sls})$ , where

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

## Appendix D - IST2SLS Estimator

### Proof of proposition 4:

Given assumptions 1-4 and 6 and the states in proposition 4, Kelejian et al. (2004) prove that  $n^{-1} a_n' (\tilde{y}_n - E\tilde{y}_n) = o_p(1)$ , where  $E\tilde{y}_n = W_n (I_n - \lambda_0 W_n)^{-1} X_n \beta$ , and  $\tilde{y}_n = \sum_{k=0}^{r_n} \hat{\lambda}_0^k W_n^{k+1} X_n \hat{\beta}_0$ .

Given that  $a_n = (a_{1,n}, \dots, a_{n,n})'$  is a sequence of  $(n \times 1)$  constant vectors whose elements are uniformly bounded in absolute value, then, given that  $a_n = (a_{i,n}, \dots, a_{n,n})'$ , where  $a_{i,n}$  is a vector whose  $i$ -th element is one and zero otherwise,

$$n^{-1} (\tilde{y}_n - E\tilde{y}_n) = (o_p(1) \dots o_p(1))'. \quad (16)$$

Given that  $\hat{Q}_n^{kp} = (\tilde{y}_n' X_n)$ , and  $C_n = (E\tilde{y}_n' X_n)$ , then  $\frac{1}{n} \hat{Q}_n^{kp} \xrightarrow{p} \frac{1}{n} C_n$ . Therefore, it is also true that  $\frac{1}{n} J_n^o \hat{Q}_n^{kp} \xrightarrow{p} \frac{1}{n} J_n^o C_n$ , since  $J_n^o$  is a non-stochastic selection matrix.

The asymptotic distribution variance of IG2SLS, using the instrument matrix  $\hat{Q}_n^{kp}$ , is equivalent to the following expression:

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

## Appendix E - Asymptotic Equivalence of $\hat{\theta}_{ibg2sls,n}$ and $\hat{\theta}_{aibg2sls,n}$

In order to prove that  $\hat{\theta}_{ibg2sls,n}$  and  $\hat{\theta}_{aibg2sls,n}$  are  $\sqrt{n}$ -equivalent, we need to prove that  $\sqrt{n}(\hat{\theta}_{ibg2sls,n} - \hat{\theta}_{aibg2sls,n}) = o_p(1)$ . Since IBG2SLS uses the optimal instrument matrix  $Q_n^{o*} = C_n^o$ , we obtain:

$$\begin{aligned} \sqrt{n}(\hat{\theta}_{ibg2sls,n} - \hat{\theta}_{aibg2sls,n}) &= \left[ \frac{1}{n} \tilde{Z}_n^{o'} \Omega_n^{-1}(\lambda_0) C_n^o \left( \frac{1}{n} C_n^{o'} \Omega_n^{-1}(\lambda_0) C_n^o \right)^{-1} \frac{1}{n} C_n^{o'} \Omega_n^{-1}(\lambda_0) \tilde{Z}_n^o \right]^{-1} \\ &\quad \frac{1}{n} \tilde{Z}_n^{o'} \Omega_n^{-1}(\lambda_0) C_n^o \left( \frac{1}{n} C_n^{o'} \Omega_n^{-1}(\lambda_0) C_n^o \right)^{-1} \frac{1}{\sqrt{n}} C_n^{o'} \Omega_n^{-1}(\lambda_0) y_n^o \\ &\quad - \left( \frac{1}{n} C_n^{o'} \Omega_n^{-1}(\lambda_0) C_n^o \right)^{-1} \frac{1}{\sqrt{n}} C_n^{o'} \Omega_n^{-1}(\lambda_0) y_n^o. \end{aligned}$$

From proposition 2 we know that  $\frac{1}{n} Q_n^{o'} \Omega_n^{-1}(\lambda_0) \tilde{Z}_n^o \xrightarrow{p} \frac{1}{n} Q_n^{o'} \Omega_n^{-1}(\lambda_0) C_n^o$ . Replacing  $Q_n^o$  by  $Q_n^{o*} = C_n^o$ , we obtain  $\frac{1}{n} C_n^{o'} \Omega_n^{-1}(\lambda_0) \tilde{Z}_n^o \xrightarrow{p} \frac{1}{n} C_n^{o'} \Omega_n^{-1}(\lambda_0) C_n^o$ . Replacing the previous equalities and cancelling terms, we obtain:  $\sqrt{n}(\hat{\theta}_{ibg2sls,n} - \hat{\theta}_{aibg2sls,n}) \xrightarrow{p} 0$ .

## Appendix F - Variance Error Estimator

We need to prove that  $J_n^o S_n^{-1}(\hat{\lambda}) \hat{e}_n = J_n^o y_n - J_n^o S_n^{-1}(\hat{\lambda}) X_n \hat{\beta}$  is a consistent estimator of  $J_n^o S_n^{-1}(\lambda_0) e_n$ .

Consider that:

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

These equivalences are based on the consistence of  $\hat{\theta} = (\hat{\lambda}, \hat{\beta}')$ , which makes the term in brackets be  $o_p(1)$ . Furthermore,  $\Sigma_n^{-1}(\hat{\lambda}) \xrightarrow{p} \Sigma_n^{-1}(\lambda_0)$ .

Given the above derivations, we obtain that

$$\hat{v}_n = T_n(\hat{\lambda}) J_n^o S_n^{-1}(\hat{\lambda}) \hat{e}_n \xrightarrow{p} v_n = T_n(\lambda_0) J_n^o S_n^{-1}(\lambda_0) e_n,$$

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

## References

- Anselin, L. (1988). *Spatial Econometrics: Methods and Models*, vol. 4. Boston: Kluwer.
- Cressie, N. (1993). *Statistics for Spatial Data*. New York: John Wiley & Sons.
- Farber, S., Páez, A., and Volz, E. (2010). Topology, Dependency Tests and Estimation Bias in Network Autoregressive Models. In A. Páez, J. Gallo, R. N. Buliung, and S. Dall'erba (Eds.), *Progress in Spatial Analysis - Methods and Applications* (pp. 29–57). Berlin, Heidelberg: Springer.
- Florax, R., and 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., and Venables, A. (1999). *The Spatial Economics: Cities, Regions and International Trade*. Cambridge, MA: MIT Press.
- Kelejian, H., and 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., and Prucha, I. (1999). A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model. *International Economic Review* 40(2), 509–533.
- Kelejian, H., and 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., and 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., and Yuzefovich, Y. (2004). Instrumental Variable Estimation of a Spatial Autoregressive Model with Autoregressive Disturbances: Large and Small Sample Results. In J. LeSage and R. Pace (Eds.), *Advances in Econometrics: Spatial and Spatiotemporal Econometrics*, vol. 18 (pp. 163–198). Bingley: Emerald Group Publishing Limited.
- Krugman, P. (1991). *Geography and trade*. Cambridge, MA: 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.
- LeSage, J., and 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.
- Mur, J., López, F., and Angulo, A. (2008). Symptoms of Instability in Models of Spatial Dependence. *Geographical Analysis* 40(2), 189–211.
- Smith, T. (2009). Estimation Bias in Spatial Models with Strongly Connected Weight Matrices. *Geographical Analysis* 41(3), 307–332.
- Wang, W., and Lee, L. F. (2013). Estimation of spatial autoregressive models with randomly missing data in the dependent variable. *The Econometrics Journal* 16(1), 73–102.