Journal of Statistical Software

January 2015, Volume 63, Issue 18. *http://www.jstatsoft.org/*

Comparing Implementations of Estimation Methods for Spatial Econometrics

Roger Bivand
Norwegian School of Economics

Gianfranco Piras
Regional Research Institute
West Virginia University

Abstract

Recent advances in the implementation of spatial econometrics model estimation techniques have made it desirable to compare results, which should correspond between implementations across software applications for the same data. These model estimation techniques are associated with methods for estimating impacts (emanating effects), which are also presented and compared. This review constitutes an up-to-date comparison of generalized method of moments and maximum likelihood implementations now available. The comparison uses the cross-sectional US county data set provided by Drukker, Prucha, and Raciborski (2013d). The comparisons will be cast in the context of alternatives using the MATLAB Spatial Econometrics toolbox, Stata's user-written **sppack** commands, Python with **PySAL** and R packages including **spdep**, **sphet** and **McSpatial**.

Keywords: spatial econometrics, maximum likelihood, generalized method of moments, estimation, R, Stata, Python, MATLAB.

1. Introduction

Researchers applying spatial econometric methods to empirical economic questions now have a wide range of tools, and a growing literature supporting these tools. In the 1970s and 1980s, it was typical for researchers to use tools coded in Fortran or other general programming languages, or to seek to integrate functions into existing statistical and/or matrix language environments (Anselin 2010). The use of spatial econometrics tools was widened by the ease with which methods and examples presented in Anselin (1988b) could be reproduced using **SpaceStat** (Anselin 1992), written in GAUSS (Aptech Systems, Inc. 2007), and shipped as a built runtime module. It was rapidly complemented by the Spatial Econometrics toolbox for MATLAB (The MathWorks, Inc. 2011), provided as source code together with extensive

documentation (see also LeSage and Pace 2009, http://www.spatial-econometrics.com/). This toolbox is under active development, and accepts contributed functions, thus broadening its appeal. In addition, Griffith and Layne (1999) gave code listings for model fitting techniques using SAS (SAS Institute Inc. 2008) and SPSS (IBM Corporation 2010). A suite of commands for spatial data analysis for use with Stata (StataCorp. 2011) was provided by Maurizio Pisati, and distributed using the standard contributed command system (Pisati 2001).[1]

The thrust of **SpaceStat** was largely been taken over by **GeoDa** (Anselin, Syabri, and Kho 2006).[2] The Python (van Rossum 1995) spatial analysis library (**PySAL**; Rey and Anselin 2010, http://pysal.org/) has also been launched. Since the R language and environment (R Core Team 2014) became available in the 1990s, collaborative code development has proceeded with varying speed. Initial attempts to implement spatial econometrics techniques in R in the **spdep** (Bivand 2013) package were checked against **SpaceStat**, and subsequently against Maurizio Pisati's Stata code and **GeoDa** by comparing results for the same input data and spatial weights (Bivand and Gebhardt 2000; Bivand 2002). A broad survey of the analysis of spatial data in the R environment is given by Bivand (2006) and Bivand, Pebesma, and Gómez-Rubio (2013b).

In the spirit of Rey (2009), this comparison will attempt to examine some features of the implementation of user-written **sppack** commands for fitting spatial econometrics models in Stata, with those in the Spatial Econometrics toolbox for MATLAB, in R and in Python. We have chosen only to compare implementations for which the estimations can be scripted, and from which the output can be transferred back to R in binary form. A consequence of this restriction is that we have not included **GeoDa**. Because Millo and Piras (2012) provide recent comparative results for implementations of spatial panel models in R and MATLAB, we restrict our consideration to cross-sectional models, and within that to cross-sectional models implemented in at least two application languages. This results in our putting Bayesian methods for spatial econometric models aside until they become available in other languages than MATLAB. Finally, we have chosen not to consider tests for residual autocorrelation or for model specification,[3] or other diagnostic or exploratory techniques or measures, feeling that model estimation is of more immediate importance.

Initially, we describe the framework used for our comparative study, and the data set chosen for use. Next we define the models to be compared, and then move to comparisons for generalized method of moments (GMM) estimators. The GMM presentation is a substantial extension of Piras (2010), as many theoretical results have been published since then, and have been incorporated into the **sphet** package, as well as made available in Stata and **PySAL**. Following the comparison of GMM implementations, we examine implementations of maximum likelihood estimators, focussing on the consequences of details in the choices of numerical methods across the alternatives. Before concluding, we compare the provision of functions for calculating emanating effects (Kelejian, Tavlas, and Hondroyiannis 2006), also

[1] Most of the comparison in this paper will be made against the user-written **sppack** commands Drukker, Prucha, and Raciborski (2013c). For the sake of simplicity, we sometimes will refer in the main text simply to Stata. The reader should keep in mind that this is a shortcut for user-written **sppack** commands in Stata.

[2] https://geodacenter.asu.edu/projects/opengeoda, source code: http://code.google.com/p/opengeoda/.

[3] See e.g., Anselin, Bera, Florax, and Yoon (1996); Anselin (1988a); Anselin and Bera (1998); Florax and Folmer (1992); Cliff and Ord (1972); Florax, Folmer, and Rey (2003); Kelejian and Piras (2011); Burridge (1980).

known as impacts (LeSage and Fischer 2008; LeSage and Pace 2009), simultaneous spatial reaction function/reduced form (Anselin and Lozano-Gracia 2008), equilibrium effects (Ward and Gleditsch 2008), and unwittingly touched upon by Bivand (2002) in trying to create a prediction method for spatial econometrics models.

1.1. Comparative study

The comparative study was constructed around unified R scripts. The first script prepared the data from the input data set for export to MATLAB in a text file, to Stata as a `dta` file and to Python as a `dbf` file written as part of an ESRI Shapefile. The data to be exported were run through `model.frame` first to generate the intercept where needed and any dummy variables (none were needed with the data set used here). Next, the first script read a GAL-format file of county neighbors from which to form spatial weights; a row-standardized weights object was then formed for export and use in R. Weights were exported to MATLAB in a three-column sparse matrix text file, to Stata in GWT-format and to Python in GAL-format.

This R script was then used to run R code to estimate chosen spatial econometrics models, and to write scripts for MATLAB, Stata and Python. Keeping all of the code in a single R script was intended to ensure that attention was focussed on estimating the same models across the different implementations. The scripts were run chunk-by-chunk by hand or by highlighting code for execution in the script editors of the different applications, but finally for MATLAB and Stata they were run from start to termination continuously. The scripts output is a binary objects containing the estimated model results; in the R case, `save` was used for the objects from a given class of models. In MATLAB, use was made of the analogous `save` function; in Stata the `file` command with `write binary` options was used; in Python `save` imported from **NumPy** (Oliphant 2006). With these mechanisms, and after careful investigation of the output estimation objects, we were able to ensure that we were comparing estimates of the same model components.

A second unified script was used to coordinate and document the collation of results from the four applications into tabular form for this article, typically by setting up the results from one application in a column vector, using `cbind` to combine these columns. The binary output from R was read using `load`; from Stata using the R function `readBin`; from MATLAB using `readMat` in the **R.Matlab** package (Bengtsson 2005); and from Python using the `npyLoad` function from the **RcppCNPy** package. The tables for presentation below were then formatted using the same rounding arguments either for the whole table or row-wise, so that no differences could be introduced by rounding results from implementations in different ways. The remaining differences, if any, come from differences in the implementations, and it is these we intend to account for as far as possible.

The analysis has been carried out on an Intel Core i7 64-bit system with 8GB RAM under Windows 7 Enterprise SP1. The software used was Stata 12.1 with the **sppack** version "st0292", MATLAB R2011b with the March 2010 version of the Spatial Econometrics toolbox,[4] R 2.15.2 (R Core Team 2014) with packages **spdep** 0.5-56, **sphet** 1.5, and **McSpatial** 1.1.1 (McMillen 2012) and their contemporary dependencies, and Python 2.7 (32-bit) with **PySAL** 1.4. Some of the **PySAL** model estimation functions may also be accessed from **GeoDaSpace**, but we

[4]Local modifications were made in a copy kept by agreement with its authors as a subdirectory on `http://R-Forge.R-project.org/projects/spdep2/`; these changes will be mentioned in the comparison of maximum likelihood methods, as they permitted additional options and returned values to be used.

have used **PySAL** directly here.

We can see from the comparison of OLS results for the selected data set shown in Table 2 that the linear algebra output of the applications used is identical, and we can assume that any differences found in comparing spatial econometrics techniques will then not be caused by divergences in linear algebra implementations. There are, however, other underlying technologies that may differ, in particular numerical optimization. From examining source code where available, it appears that the GM methods in **PySAL** use the **SciPy** (Jones, Oliphant, Peterson, and others 2001) `fmin_l_bfgs_b` function in the optimize module, based on L_BFGS_B version 2.1 from 1997,[5] a quasi-Newton function for bound-constrained optimization. Nash and Varadhan (2011) provide a helpful guide to available numerical optimization technologies, and to those available in R. In **sphet**, use is made of the `nlminb` function, which is a reverse-communication trust-region quasi-Newton method from the **Port** library (Gay 1990),[6] with enhanced scaling features. The same function is used by default for fitting in **spdep** when more than one parameter is to be optimized; optionally, other optimizers, including the R implementation of L_BFGS_B version 2.1. For bounded line search in **spdep**, use is made of the `optimize` function, based on Brent (1973). In the R **McSpatial** package (McMillen 2012), `optimize` is used first to conduct a line search, and then `nlm`, a Newton-type algorithm (Schnabel, Koontz, and Weiss 1985) is used to optimize all the parameters in a maximum likelihood model, possibly modifying the outcome of the line search. The GM functions in the Spatial Econometrics toolbox use an included function `minz` contributed by Michael Cliff; for bounded line search, the MATLAB `fminbnd` function also based on Brent (1973) is used, while when more than one parameter is to be optimized, the MATLAB `fminsearch` function is used – it is an implementation of the Nelder-Mead simplex algorithm. Finally, the Stata implementations of spatial econometric models use `optimize` mechanisms described in the help page for `maximize`, referred to in Drukker *et al.* (2013c,d). The default numerical optimizer is `"nr"`, a Stata-modified Newton-Raphson algorithm, but other algorithms may be chosen (Gould, Pitblado, and Poi 2010).

1.2. Data set

We were fortunate to be able to use the simulated US Driving Under the Influence (DUI) county data set used in Drukker *et al.* (2013b,c,d) for our comparison. The data used is simulated for 3109 counties (omitting Alaska, Hawaii, and US territories), and uses simulations from variables used by Powers and Wilson (2004). The counties are taken from an ESRI Shapefile downloaded from the US Census.[7] The dependent variable `dui` is defined as the alcohol-related arrest rate per 100,000 daily vehicle miles traveled (DVMT). The explanatory variables include `police` (number of sworn officers per 100,000 DVMT); `nondui` (non-alcohol-related arrests per 100,000 DVMT); `vehicles` (number of registered vehicles per 1,000 residents), and `dry` (a dummy for counties that prohibit alcohol sale within their borders, about 10% of counties). A further dummy variable `elect` takes values of 1 if a county government faces an election, 0 otherwise, and has 295 non-zero entries. Descriptive statistics for the simulated DUI data set are shown in Table 1.

Table 2 shows the stylized results from estimating the relationship between the simulated

[5] `http://www.scipy.org/doc/api_docs/SciPy.optimize.lbfgsb.html`, `http://users.eecs.northwestern.edu/~nocedal/lbfgsb.html`.

[6] `http://netlib.sandia.gov/port/`.

[7] `ftp://ftp2.census.gov/geo/tiger/TIGER2008/tl_2008_us_county00.zip`.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
dui	15.01	19.88	20.83	20.84	21.82	26.62
police	25.28	29.73	30.72	30.70	31.67	36.78
nondui	18.01	34.41	40.19	40.98	46.74	76.50
vehicles	390.40	479.90	501.30	501.80	523.60	625.90

Table 1: Descriptive statistics for the dependent variable and the non-dummy explanatory variables, simulated DUI data set.

	R `lm`	Stata `reg`	MATLAB SE `ols`	Python **PySAL** `OLS`
Intercept	−5.4428237	−5.4428237	−5.4428237	−5.4428237
	(0.229431)	(0.229431)	(0.229431)	(0.229431)
police	0.5990957	0.5990957	0.5990957	0.5990957
	(0.014935)	(0.014935)	(0.014935)	(0.014935)
nondui	0.0002746	0.0002746	0.0002746	0.0002746
	(0.001088)	(0.001088)	(0.001088)	(0.001088)
vehicles	0.0156842	0.0156842	0.0156842	0.0156842
	(0.000670)	(0.000670)	(0.000670)	(0.000670)
dry	0.1060904	0.1060904	0.1060904	0.1060904
	(0.035011)	(0.035011)	(0.035011)	(0.035011)

Table 2: OLS estimation results for four implementations, simulated DUI data set (standard errors in parentheses).

dependent variable and four explanatory variables using least squares. Powers and Wilson (2004, p. 331) found that "there is no significant relationship between prohibition status and the DUI arrest rate when controlling for the proportionate number of sworn officers and the non-DUI arrest rate per officer" when examining data for 75 counties in Arkansas. Their best model has an adjusted R^2 of 0.397, while the simulated data has a higher value of 0.850, and the explanatory variables with the exception of `nondui` are all significant. As can be seen easily, all four implementations, R `lm`, Stata `reg`, MATLAB Spatial Econometrics (SE) toolbox `ols`, and Python **PySAL** `OLS` give identical results, so confirming that all four applications are handling the same data.

Drukker *et al.* (2013d) do not specify how spatial dependence was introduced into the dependent variable and/or residuals. Using the 3109 county `SpatialPolygons` read from the ESRI Shapefile mentioned above, we recreated the Queen contiguity list of neighbors with `poly2nb` in **spdep**. The descriptive statistics for the neighbor object shown by Drukker *et al.* (2013b) match ours exactly:[8]

```
R> library("rgdal")
R> county <- readOGR("tl_2008_us_county00", "tl_2008_us_county00")
```

[8]We were perplexed to find that the results from fitting the ML SARAR model given by Drukker *et al.* (2013d), did not match those we obtained in R or Stata. We initially adopted row-standardized spatial weights, $\mathbf{W}$, where the county contiguities c_{ij}, taking values of 1 if contiguous, sharing at least one boundary point, and 0 otherwise, are row-standardized by dividing by row sums. It turned out (personal communication, Rafal Raciborski) that the spatial weights used in estimation in Drukker *et al.* (2013d) were in fact minmax-normalized, where the required operation is:

$$w_{ij} = \frac{c_{ij}}{m}. \tag{1}$$

```
R> drop <- county$STATEFP00 == "02" | county$STATEFP00 == "15" |
+    as.character(county$STATEFP00) > "56"
R> ccounty <- county[!drop, ]
R> library("rgeos")
R> strt <- gUnarySTRtreeQuery(ccounty)
R> library("spdep")
R> nblist <- poly2nb(ccounty, foundInBox = strt)
R> nblist

Neighbour list object:
Number of regions: 3109
Number of nonzero links: 18474
Percentage nonzero weights: 0.1911259
Average number of links: 5.942104
```

2. A general spatial model

The present discussion is almost entirely based on Kelejian and Prucha (2010), Drukker, Egger, and Prucha (2013a), Arraiz, Drukker, Kelejian, and Prucha (2010) and Drukker *et al.* (2013c) that provide some important extensions to Kelejian and Prucha (1998, 1999). The model presented is quite general and allows for some of the right hand side variables to be endogenous. Specifically, the point of departure will be the following Cliff-Ord spatial model:

$$\mathbf{y} = \mathbf{Y}\pi + \mathbf{X}\beta + \rho_{\text{Lag}}\mathbf{W}\mathbf{y} + \mathbf{u} \tag{3}$$

where $\mathbf{y}$ is an $n \times 1$ vector of observations on the dependent variable, $\mathbf{Y}$ is an $n \times p$ matrix of observations on p endogenous variables, $\mathbf{X}$ is a $n \times k$ matrix of observations on k exogenous variables, $\mathbf{W}$ is an $n \times n$ observed and non-stochastic spatial weighting matrix and, consequently, $\mathbf{W}\mathbf{y}$ is an $n \times 1$ variable that is generally referred to as the spatially lagged dependent variable; π and β are corresponding parameters; and ρ_{Lag} is the spatial autoregressive coefficient. Given the presence of $\mathbf{Y}$, the model can be viewed as a representation of a single equation of a system of equations.[9]

The error vector $\mathbf{u}$ follows a spatial autoregressive process of the form:

$$\mathbf{u} = \rho_{\text{Err}}\mathbf{M}\mathbf{u} + \varepsilon \tag{4}$$

where m is the minmax value, the minimum of the pair of maxima of row and column sums:

$$m = \min(\max \sum_{i=1}^{n} c_{ij}, \max \sum_{j=1}^{n} c_{ij}). \tag{2}$$

Using these weights, we could reproduce the ML SARAR estimation results given by Drukker *et al.* (2013d), and could confirm that the underlying contiguities were the same as those used in the Stata documentation.

[9] A spatial lag of the matrix of observations on the exogenous variables $\mathbf{WX}$ may be added to the model, however Kelejian and Prucha (1997) warn against this. To avoid complications, we will not consider it in the present section. To avoid complications, we will not consider it in the present section. However, in the maximum likelihood implementation we will discuss the estimation of a model that includes $\mathbf{WX}$ among the set of regressors.

where ρ_{Err} is a scalar parameter generally referred to as the spatial autoregressive parameter, $\mathbf{M}$ is an $n \times n$ spatial weighting matrix that may or may not be the same as $\mathbf{W}$,[10] $\mathbf{Mu}$ is an $n \times 1$ vector of observation on the spatially lagged vector of residuals.

An alternative, more compact way to express the same model is:

$$\mathbf{y} = \mathbf{Z}\delta + \mathbf{u} \tag{5}$$

where $\mathbf{Z} = [\mathbf{Y}, \mathbf{X}, \mathbf{Wy}]$ is the set of all (endogenous and exogenous) explanatory variables, and $\delta = [\pi^\top, \beta^\top, \rho_{\text{Lag}}]^\top$ is the corresponding vector of parameters. Finally, the assumption on which the maximum likelihood relies is that $\varepsilon \sim N(0, \sigma^2)$. In the GMM approach, the estimation theory is developed both under the assumptions that the innovations ε are homoskedastic, and heteroskedastic of unknown form.

2.1. Notation

Here we adopt the notation ρ_{Lag} for the spatial autoregressive parameter on the spatially lagged dependent variable $\mathbf{y}$, and ρ_{Err} for the spatial autoregressive parameter on the spatially lagged residuals. In Ord (1975), ρ is used for both parameters, but subsequently two schools have developed, with Anselin (1988b) and LeSage and Pace (2009) (and many others) using ρ for the spatial autoregressive parameter on the spatially lagged dependent variable $\mathbf{y}$, and λ for the spatial autoregressive parameter on the spatially lagged residuals. Kelejian and Prucha (1998, 1999) (and many others particularly in the econometrics literature) adopt the opposite notation, using λ for the spatial autoregressive parameter on the spatially lagged dependent variable $\mathbf{y}$, and ρ for the spatial autoregressive parameter on the spatially lagged residuals. Our choice is an attempt to disambiguate in this comparison, not least because the different software implementations use one or other scheme, so inviting confusion.

The names used for models also vary between software implementations, so that the general model given in Equation 3 is also known as a SARAR model, with first order autoregressive processes in the dependent variable and the residuals. The same model is termed a SAC model in LeSage and Pace (2009) and the MATLAB Spatial Econometrics toolbox. This model and all its derivatives are simultaneous autoregressive (SAR) models in the sense used in spatial statistics (see also Ripley 1981, p. 88), in contrast to conditional autoregressive (CAR) models often used in disease mapping and other application areas. Other issues in the naming of models will be mentioned in the next section.

2.2. Restrictions on the general model

The general model (Equation 3) may be restricted by setting $\pi = 0$ to remove the endogenous variables. All of the models considered when comparing maximum likelihood implementations, and many GMM implementations, impose this restriction. The spatial lag model is formed as a special case with $\rho_{\text{Err}} = 0$, and the spatial error model with $\rho_{\text{Lag}} = 0$. The spatial error model with no endogenous variables is:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{u}, \mathbf{u} = \rho_{\text{Err}}\mathbf{Mu} + \varepsilon \tag{6}$$

[10] In many empirical applications $\mathbf{W}$ and $\mathbf{M}$ are assumed to be identical. However, both R and Stata allow them to differ. From an implementation perspective, the main complication is in the definition of the instrument matrix. We will discuss this issue in the next section.

This model is known as SEM (spatial error model) in LeSage and Pace (2009), and SARE (spatial autoregressive error model) in Drukker *et al.* (2013d).

The spatial lag model with no endogenous variables is:

$$\mathbf{y} = \mathbf{X}\beta + \rho_{\text{Lag}}\mathbf{W}\mathbf{y} + \varepsilon \tag{7}$$

This model is known as SAR (spatial autoregressive model) in both LeSage and Pace (2009) and Drukker *et al.* (2013d).

In fitting spatial lag models and by extension any model including the spatially lagged dependent variable, it has emerged over time that, unlike the spatial error model, the spatial dependence in the parameter ρ_{Lag} feeds back, obliging analysts to base interpretation not on the fitted parameters β, but rather on correctly formulated impact measures, as discussed in references given on page 3.

This feedback comes from the data generation process of the spatial lag model (and by extension in the general model). Rewriting:

$$\begin{aligned} \mathbf{y} - \rho_{\text{Lag}}\mathbf{W}\mathbf{y} &= \mathbf{X}\beta + \varepsilon \\ (\mathbf{I} - \rho_{\text{Lag}}\mathbf{W})\mathbf{y} &= \mathbf{X}\beta + \varepsilon \\ \mathbf{y} &= (\mathbf{I} - \rho_{\text{Lag}}\mathbf{W})^{-1}\mathbf{X}\beta + (\mathbf{I} - \rho_{\text{Lag}}\mathbf{W})^{-1}\varepsilon \end{aligned}$$

where $\mathbf{I}$ is the $n \times n$ identity matrix. This means that the expected impact of a unit change in an exogenous variable r for a single observation i on the dependent variable y_i is no longer equal to β_r, unless $\rho_{\text{Lag}} = 0$. The awkward $n \times n$ $S_r(\mathbf{W}) = ((\mathbf{I} - \rho_{\text{Lag}}\mathbf{W})^{-1}\mathbf{I}\beta_r)$ matrix term is needed to calculate impact measures for the spatial lag model, and similar forms for other models including the general model, when $\rho_{\text{Lag}} \neq 0$.[11]

3. Comparing GMM implementations

The estimation of the general SARAR(1,1)[12] model consists of various steps. Given the simultaneous presence of the endogenous variables on the right hand side of Equation 3 and the spatially autocorrelated residuals, these steps will alternate instrumental variables (IV) and GMM estimators. These estimators will be based on a set of linear and quadratic moment conditions of the form:

$$E\mathbf{H}^{\top}\varepsilon = 0 \tag{8}$$

$$E\varepsilon^{\top}\mathbf{A}\varepsilon = 0 \tag{9}$$

where $\mathbf{H}$ is an $n \times p$ non-stochastic matrix of instruments, and $\mathbf{A}$ is an $n \times n$ weighting matrix whose specification will be introduced later. At this point, it is useful to introduce the spatial Cochrane-Orcutt transformation of the model, that is:

$$\mathbf{y}^{\star} = \mathbf{Z}^{\star}\delta + \varepsilon \tag{10}$$

[11] When an additional endogenous variable is included, one should make additional (and generally restrictive) assumptions to be able to calculate this impact measures. Since, to the best of our knowledge, no theory is available for this case, we will exclude it from our discussion.

[12] This notation is simply to emphasize that the model has only one spatial lag for the dependent variable, and one lag of the error term. Of course, a researcher can include as many lags as he wishes.

where $\mathbf{y}^\star = \mathbf{y} - \rho_{\text{Err}}\mathbf{M}\mathbf{y}$ and $\mathbf{Z}^\star = \mathbf{Z} - \rho_{\text{Err}}\mathbf{M}\mathbf{Z}$.

As a preview of the estimation steps, an initial IV estimator of δ leads to a set of consistent residuals. This vector of residuals constitutes the basis for the derivation of the quadratic moment conditions that provide a first consistent estimate for the autoregressive parameter ρ_{Err}.[13] An estimate of δ is obtained from the transformed model after replacing the true value of ρ_{Err} with its consistent estimate obtained in the previous step. Finally, in a new GM iteration, it is possible to obtain a consistent and efficient estimate of ρ_{Err} based on generalized spatial two stage least square residuals. The asymptotic variance-covariance matrix for the coefficients is then calculated using the estimate of δ, the residuals, and the estimate of the spatial coefficient ρ_{Err}.

3.1. SARAR model

For the case of no additional endogenous variables other than the spatial lag, the "ideal" instruments should be expressed in terms of $E(\mathbf{W}\mathbf{y})$. This is simply because the best instruments for the right hand side variables are the conditional means and, since $\mathbf{X}$ and $\mathbf{M}\mathbf{X}$ are non-stochastic, we can simply focus on the spatial lags $\mathbf{W}\mathbf{y}$ and $\mathbf{M}\mathbf{W}\mathbf{y}$ (see Equation 10). Given the reduced form of the model

$$\mathbf{y} = (\mathbf{I} - \rho_{\text{Lag}}\mathbf{W})^{-1}(\mathbf{X}\beta + \mathbf{u}) \tag{11}$$

it follows that the best instruments can be expressed in terms of the $E(\mathbf{W}\mathbf{y}) = \mathbf{W}(\mathbf{I} - \rho_{\text{Lag}}\mathbf{W})^{-1}\mathbf{X}\beta$ (Lee 2003, 2007; Kelejian, Prucha, and Yuzefovich 2004; Das, Kelejian, and Prucha 2003). Given that the roots of $\rho_{\text{Lag}}\mathbf{W}$ are less than one in absolute value, Kelejian and Prucha (1999) suggested to generate an approximation to the best instruments (say $\mathbf{H}$) as the subset of the linearly independent columns of

$$\mathbf{H} = (\mathbf{X}, \mathbf{W}\mathbf{X}, \mathbf{W}^{\mathbf{2}}\mathbf{X}, \ldots, \mathbf{W}^{\mathbf{q}}\mathbf{X}, \mathbf{M}\mathbf{X}, \mathbf{M}\mathbf{W}\mathbf{X}, \ldots, \mathbf{M}\mathbf{W}^{\mathbf{q}}\mathbf{X}) \tag{12}$$

where q is a pre-selected finite constant and is generally set to 2 in applied studies.[14] The inclusion of instruments involving $\mathbf{M}$ in the instrument matrix $\mathbf{H}$ is only needed for the formulation of instrumental variable estimators applied to the spatially Cochrane-Orcutt transformed model. At a minimum the instruments should include the linearly independent columns of $\mathbf{X}$ and $\mathbf{M}\mathbf{X}$. In a more general setting where additional endogenous variables are present, since the system determining $\mathbf{y}$ and $\mathbf{Y}$ is not completely specified, the optimal instruments are not known. In this case it may be reasonable to use a set of instruments as above where $\mathbf{X}$ is augmented by other exogenous variables that are expected to determine $\mathbf{Y}$. Unless additional information is available, it is not recommended to take the spatial lag of these additional exogenous variables.

The starting point for the estimation of ρ_{Err} are the two following quadratic moment conditions expressed as functions of the innovation ε

$$E[\varepsilon^\top \mathbf{A}_s \varepsilon] = 0 \tag{13}$$

[13] The estimate of ρ_{Err} could be, in turn, used to define a weighting matrix for the moment conditions in order to obtain a consistent and efficient estimator. While possible, this would imply some computational complications. However, the ρ_{Err} is already consistent and, therefore, can be used to define the transformed model.

[14] While R and Stata set $q = 2$ by default, **PySAL** leaves the choice open to users.

where the matrices $\mathbf{A}_s$ are such that $\text{tr}(\mathbf{A}_s) = 0$. Furthermore, under heteroskedasticity it is also assumed that the diagonal elements of the matrices $\mathbf{A}_s$ are zero. The reason for this is that simplifies the formulae for the variance-covariance matrix (i.e., only depends on second moments of the innovations and not on third and fourth moments). Specific suggestions for $\mathbf{A}_s$ are given below. In general, such choices will depend on whether or not the model assumes heteroskedasticity.

As we already mentioned, the actual estimation procedure consists of several steps, which we will review next. Our review of the estimation procedure will be intentionally brief, as we refer the interested reader to more specific literature (Kelejian and Prucha 2010; Drukker *et al.* 2013a; Arraiz *et al.* 2010; Drukker *et al.* 2013c).

Step 1.1 – 2SLS estimator

The initial set of estimates are obtained from the untransformed model using the matrix of instruments H, yielding to:

$$\tilde{\delta} = (\tilde{\mathbf{Z}}^\top \tilde{\mathbf{Z}})^{-1} \tilde{\mathbf{Z}}^\top \mathbf{y} \tag{14}$$

where $\tilde{\mathbf{Z}} = \mathbf{P}\mathbf{Z} = [\mathbf{X}, \widehat{\mathbf{W}\mathbf{y}}]$, $\widehat{\mathbf{W}\mathbf{y}} = \mathbf{P}\mathbf{W}\mathbf{y}$, $\mathbf{P} = \mathbf{H}(\mathbf{H}^\top\mathbf{H})^{-1}\mathbf{H}^\top$ and $\mathbf{H}$ was specified above. The estimate $\tilde{\delta}$ can then be used to obtain a initial estimate of the regression residuals, say $\tilde{\mathbf{u}}$. The GMM estimator of the next step is based on the assumption that these regression residuals are consistent.

Step 1.2 – Initial GMM estimator

Let $\tilde{\delta}$ be the initial estimator of δ obtained from Step 1.1, let $\tilde{\mathbf{u}} = \mathbf{y} - \mathbf{Z}\tilde{\delta}$, and $\bar{\tilde{\mathbf{u}}} = M\tilde{\mathbf{u}}$. Then, we can operationalize the sample moments corresponding to Equation 13, that is:

$$m(\tilde{\delta}, \rho_{\text{Err}}) = n^{-1} \begin{bmatrix} (\tilde{\mathbf{u}} - \rho_{\text{Err}}\bar{\tilde{\mathbf{u}}})^\top \mathbf{A}_1 (\tilde{\mathbf{u}} - \rho_{\text{Err}}\bar{\tilde{\mathbf{u}}}) \\ \vdots \\ (\tilde{\mathbf{u}} - \rho_{\text{Err}}\bar{\tilde{\mathbf{u}}})^\top \mathbf{A}_s (\tilde{\mathbf{u}} - \rho_{\text{Err}}\bar{\tilde{\mathbf{u}}}) \end{bmatrix}$$

It is convenient to rewrite $m(\tilde{\delta}, \rho_{\text{Err}})$ as an explicit function of the parameters ρ_{Err} and ρ^2_{Err} as in the following expression

$$m(\tilde{\delta}, \rho_{\text{Err}}) = \tilde{\mathbf{\Gamma}} \begin{bmatrix} \rho_{\text{Err}} \\ \rho^2_{\text{Err}} \end{bmatrix} - \tilde{\gamma}$$

where

$$\tilde{\mathbf{\Gamma}} = n^{-1} \begin{bmatrix} \tilde{\mathbf{u}}^\top(\mathbf{A}_1 + \mathbf{A}_1^\top)\bar{\tilde{\mathbf{u}}} & -\bar{\tilde{\mathbf{u}}}\mathbf{A}_1\bar{\tilde{\mathbf{u}}} \\ \vdots & \vdots \\ \tilde{\mathbf{u}}^\top(\mathbf{A}_s + \mathbf{A}_s^\top)\bar{\tilde{\mathbf{u}}} & -\bar{\tilde{\mathbf{u}}}\mathbf{A}_s\bar{\tilde{\mathbf{u}}} \end{bmatrix} \quad \text{and} \quad \tilde{\gamma} = n^{-1} \begin{bmatrix} \tilde{\mathbf{u}}^\top\mathbf{A}_1\tilde{\mathbf{u}} \\ \vdots \\ \tilde{\mathbf{u}}^\top\mathbf{A}_s\tilde{\mathbf{u}} \end{bmatrix}$$

Therefore, the initial GMM estimator for ρ_{Err} is obtained by simply minimizing the following relation:

$$\tilde{\rho}_{\text{Err}} = \text{argmin} \left\{ \left[\tilde{\mathbf{\Gamma}} \begin{pmatrix} \rho_{\text{Err}} \\ \rho^2_{\text{Err}} \end{pmatrix} - \tilde{\gamma} \right]^\top \left[\tilde{\mathbf{\Gamma}} \begin{pmatrix} \rho_{\text{Err}} \\ \rho^2_{\text{Err}} \end{pmatrix} - \tilde{\gamma} \right] \right\}.$$

The expression above can be interpreted as a nonlinear least squares system of equations. The initial estimate is thus obtained as a solution of the above system.

We finally need an expression for the matrices $\mathbf{A}_s$. Drukker *et al.* (2013a) suggest, for the homoskedastic case, the following expressions:[15]

$$\mathbf{A}_1 = \left\{1 + [n^{-1}\text{tr}(\mathbf{M}^\top\mathbf{M})]^2\right\}^{-1} [\mathbf{M}^\top\mathbf{M} - n^{-1}\text{tr}(\mathbf{M}^\top\mathbf{M})\mathbf{I}] \tag{15}$$

and

$$\mathbf{A}_2 = \mathbf{M} \tag{16}$$

On the other hand, when heteroskedasticity is assumed, Kelejian and Prucha (2010) recommend the following expressions for $\mathbf{A}_1$ and $\mathbf{A}_2$:[16]

$$\mathbf{A}_1 = \mathbf{M}^\top\mathbf{M} - diag(\mathbf{M}^\top\mathbf{M}) \tag{17}$$

and

$$\mathbf{A}_2 = \mathbf{M} \tag{18}$$

Step 2.1 – GS2SLS estimator

Using the estimate of ρ_{Err} obtained in Step 1.2, one can take a Cochrane-Orcutt transformation of the model as in Equation 10 and estimate it again by 2SLS. The entire procedure is known in the literature as generalized spatial two stage least square (GS2SLS):

$$\hat{\delta} = [\widehat{\mathbf{Z}}^{\star\top}\mathbf{Z}]^{-1}\widehat{\mathbf{Z}}^{\star\top}\mathbf{y} \tag{19}$$

where $\widehat{\mathbf{Z}}^\star = \mathbf{P}\mathbf{Z}^\star$, $\mathbf{Z}^\star = \mathbf{Z} - \tilde{\rho}_{\text{Err}}\mathbf{W}\mathbf{Z}$, $\mathbf{y}^\star = \mathbf{y} - \tilde{\rho}_{\text{Err}}\mathbf{W}\mathbf{y}$ and $\mathbf{P} = \mathbf{H}(\mathbf{H}^\top\mathbf{H})^{-1}\mathbf{H}^\top$.

Step 2.2 – Consistent and efficient GMM estimator

In the second sub-step of the second step, an efficient GMM estimator of ρ_{Err} based on the GS2SLS residuals is obtained by minimizing the following expression:

$$\hat{\rho}_{\text{Err}} = \text{argmin}\left\{\left[\hat{\mathbf{\Gamma}}\begin{pmatrix}\rho_{\text{Err}}\\ \rho^2_{\text{Err}}\end{pmatrix} - \hat{\gamma}\right]^\top (\hat{\mathbf{\Psi}}^{\rho_{\text{Err}}\rho_{\text{Err}}})^{-1}\left[\hat{\mathbf{\Gamma}}\begin{pmatrix}\rho_{\text{Err}}\\ \rho^2_{\text{Err}}\end{pmatrix} - \hat{\gamma}\right]\right\}$$

where $\hat{\mathbf{\Psi}}^{\rho_{\text{Err}}\rho_{\text{Err}}}$ is an estimator for the variance-covariance matrix of the (normalized) sample moment vector based on GS2SLS residuals. This estimator differs for the cases of homoskedastic and heteroskedastic errors.

For the homoskedastic case, the r, s (with r, s=1,2) element of $\hat{\mathbf{\Psi}}^{\rho_{\text{Err}}\rho_{\text{Err}}}$ is given by:

$$\begin{aligned}\hat{\mathbf{\Psi}}^{\rho_{\text{Err}}\rho_{\text{Err}}}_{rs} &= [\tilde{\sigma}^2]^2(2n)^{-1}\text{tr}[(\mathbf{A}_r + \mathbf{A}_\mathbf{r}^\top)(\mathbf{A}_s + \mathbf{A}_\mathbf{s}^\top)]\\ &+ \tilde{\sigma}^2 n^{-1}\tilde{\mathbf{a}}_r^\top\tilde{\mathbf{a}}_s^\top\\ &+ n^{-1}(\tilde{\mu}^{(4)} - 3[\tilde{\sigma}^2]^2)\text{vec}_D(\mathbf{A}_r)^\top\text{vec}_D(\mathbf{A}_s)\\ &+ n^{-1}\tilde{\mu}^{(3)}[\tilde{\mathbf{a}}_r^\top\text{vec}_D(\mathbf{A}_s) + \tilde{\mathbf{a}}_s^\top\text{vec}_D(\mathbf{A}_r)]\end{aligned} \tag{20}$$

[15]The scaling factor $v = \left\{1 + [n^{-1}\text{tr}(\mathbf{M}^\top\mathbf{M})]^2\right\}^{-1}$ is needed to obtain the same estimator of Kelejian and Prucha (1998, 1999)

[16]For the heteroskedastic case, Arraiz *et al.* (2010) derive a consistent and efficient GMM estimator based on 2SLS residuals. While we will not review this estimator in the present section, the R and **PySAL** implementations allow for this additional step.

where $\tilde{\mathbf{a}}_r = \hat{\mathbf{T}}\tilde{\alpha}_r$, $\hat{\mathbf{T}} = \mathbf{H}\hat{\mathbf{P}}$, $\hat{\mathbf{P}} = \hat{\mathbf{Q}}_{\mathbf{HH}}^{-1}\hat{\mathbf{Q}}_{\mathbf{HZ}}[\hat{\mathbf{Q}}_{\mathbf{HZ}}^{\top}\hat{\mathbf{Q}}_{\mathbf{HH}}^{-1}\hat{\mathbf{Q}}_{\mathbf{HZ}}^{\top}]^{-1}$, $\hat{\mathbf{Q}}_{\mathbf{HH}}^{-1} = (n^{-1}\mathbf{H}^{\top}\mathbf{H})$, $\hat{\mathbf{Q}}_{\mathbf{HZ}} = [n^{-1}\mathbf{H}^{\top}\mathbf{Z}]$, $\mathbf{Z} = (\mathbf{I} - \tilde{\rho}_{\text{Err}}\mathbf{M})\mathbf{Z}$, $\tilde{\alpha}_r = -n^{-1}[\mathbf{Z}^{\top}(\mathbf{A}_r + \mathbf{A}_r^{\top})\hat{\varepsilon}]$, $\hat{\sigma}^2 = n^{-1}\hat{\varepsilon}^{\top}\hat{\varepsilon}$, $\hat{\mu}^{(3)} = n^{-1}\sum_{i=1}^{n}\hat{\varepsilon}^3$, and $\hat{\mu}^{(3)} = n^{-1}\sum_{i=1}^{n}\hat{\varepsilon}^4$.

For the heteroskedastic case, the r, s (with r, s=1,2) element of $\hat{\mathbf{\Psi}}^{\rho_{\text{Err}}\rho_{\text{Err}}}$ is given by:

$$\hat{\mathbf{\Psi}}_{rs}^{\rho_{\text{Err}}\rho_{\text{Err}}} = (2n)^{-1}\text{tr}[(\mathbf{A}_r + \mathbf{A}_{\mathbf{r}}^{\top})\hat{\mathbf{\Sigma}}(\mathbf{A}_s + \mathbf{A}_{\mathbf{s}}^{\top})\hat{\mathbf{\Sigma}}] + n^{-1}\tilde{\mathbf{a}}_r^{\top}\hat{\mathbf{\Sigma}}\tilde{\mathbf{a}}_s^{\top} \tag{21}$$

where, $\hat{\mathbf{\Sigma}}$ is a diagonal matrix whose ith diagonal element is $\hat{\varepsilon}^2$, and $\hat{\varepsilon}$, and $\tilde{\mathbf{a}}_r$ are defined as above.[17]

Having completed the previous steps, a consistent estimator for the asymptotic VC matrix $\mathbf{\Omega}$ can be derived. The estimator is given by $n\widehat{\mathbf{\Omega}}$ where:

$$\widehat{\mathbf{\Omega}} = \begin{pmatrix} \widehat{\mathbf{\Omega}}_{\delta\delta} & \widehat{\mathbf{\Omega}}_{\delta\rho_{\text{Err}}} \\ \widehat{\mathbf{\Omega}}_{\delta\rho_{\text{Err}}^{\top}} & \widehat{\mathbf{\Omega}}_{\rho_{\text{Err}}\rho_{\text{Err}}} \end{pmatrix}$$

where $\widehat{\mathbf{\Omega}}_{\delta\delta} = \hat{\mathbf{P}}^{\top}\widehat{\mathbf{\Psi}}_{\delta\delta}\hat{\mathbf{P}}$, $\widehat{\mathbf{\Omega}}_{\delta\rho_{\text{Err}}} = \hat{\mathbf{P}}^{\top}\widehat{\mathbf{\Psi}}_{\delta\rho_{\text{Err}}}[\widehat{\mathbf{\Psi}}_{\rho_{\text{Err}}\rho_{\text{Err}}}]^{-1}\widehat{\mathbf{J}}[\widehat{\mathbf{J}}^{\top}[\widehat{\mathbf{\Psi}}_{\rho_{\text{Err}}\rho_{\text{Err}}}]^{-1}\widehat{\mathbf{J}}]^{-1}$, $\widehat{\mathbf{\Omega}}_{\rho_{\text{Err}}\rho_{\text{Err}}} = [\widehat{\mathbf{J}}^{\top}[\widehat{\mathbf{\Psi}}_{\rho_{\text{Err}}\rho_{\text{Err}}}]^{-1}\widehat{\mathbf{J}}]^{-1}$, $\widehat{\mathbf{J}} = \widehat{\mathbf{\Gamma}}[1, 2\hat{\rho}_{\text{Err}}]^{\top}$.

Some of the element of the VC matrix were defined before. Once more, the estimators for $\widehat{\mathbf{\Psi}}_{\delta\delta}(\hat{\rho}_{\text{Err}})$, and $\widehat{\mathbf{\Psi}}_{\delta\rho_{\text{Err}}}(\hat{\rho}_{\text{Err}})$ are different for the homoskedastic and the heteroskedastic cases.

In the homoskedastic case: $\widehat{\mathbf{\Psi}}_{\delta\delta} = \hat{\sigma}^2\widehat{\mathbf{Q}}_{\mathbf{HH}}$, $\widehat{\mathbf{\Psi}}_{\delta\rho_{\text{Err}}} = \hat{\sigma}^2 n^{-1}\mathbf{H}^{\top}[\mathbf{a}_1, \mathbf{a}_2] + \mu^{(3)}n^{-1}\mathbf{H}^{\top}[\text{vec}_D(\mathbf{A}_1), \text{vec}_D(\mathbf{A}_2)]$.

In the heteroskedastic case: $\widehat{\mathbf{\Psi}}_{\delta\delta} = n^{-1}\mathbf{H}^{\top}\hat{\mathbf{\Sigma}}\mathbf{H}$, $\widehat{\mathbf{\Psi}}_{\delta\rho_{\text{Err}}} = n^{-1}\mathbf{H}^{\top}\hat{\mathbf{\Sigma}}[\mathbf{a}_1\mathbf{a}_2]$.

Homoskedasticity with and without additional endogenous variables

There are various implementations of the GMM general model (under homoskedasticity) that are presented in Table 3.[18] Some of them are based on the Kelejian and Prucha (1999) moment conditions (i.e., the R function `gstsls` available from **spdep**, the Spatial Econometrics toolbox function `sac_gmm`, and **PySAL** `GM_Combo`). The remaining implementations are based on the Drukker *et al.* (2013a) moment conditions (i.e., the function `spreg` available from **sphet**, the Stata command `spreg gs2sls`, and **PySAL** `GM_Combo_Hom`).

The results for all of them are reported in Table 3. Specifically, the results based on the Kelejian and Prucha (1999) moment conditions correspond to columns one, four, and six. A glance at this columns reveals that, while the estimated coefficients obtained from the function `gstsls` in R and **PySAL** `GM_Combo` are identical (up to the sixth digit), those obtained from the Spatial Econometrics toolbox function `sac_gmm` are slightly different. The discrepancies with `sac_gmm` in MATLAB are related to a different sets of instruments. The SE toolbox uses two different sets of instruments: one for estimating the "original" model, and one for estimating the model after the Cochrane-Orcutt transformation. In the first step of the procedure, the matrix of instruments contains all of the linearly independent columns of $\mathbf{X}$, $\mathbf{WX}$, and $\mathbf{W}^2\mathbf{X}$ and it is the same for all three implementations. However, to estimate the transformed model, the matrix of instruments employed by MATLAB includes, along with

[17] It is worth noticing here that the second term in Equation 21 limits to zero when there are only exogenous variables in the model. As we will demonstrate later, while the implementations in R and Stata produce an estimate of this term, **PySAL** simply assumes it to be zero.

[18] For simplicity, in all models it is assumed that $\mathbf{W}$ and $\mathbf{M}$ are the same. As mentioned, this assumption will be relaxed later in the paper.

	R `gstsls`	R `spreg`	Stata `spreg`	**PySAL** `GM_Combo`	**PySAL** `GM_Combo_Hom`	SE `sac_gmm`
Intercept	−6.409919	−6.409969	−6.409968	−6.409919	−6.409969	−6.403747
	(0.418363)	(0.416359)	(0.416359)	(0.417959)	(0.416359)	(0.417963)
`police`	0.598107	0.598102	0.598102	0.598107	0.598102	0.598107
	(0.014918)	(0.014907)	(0.014907)	(0.014903)	(0.014907)	(0.014918)
`nondui`	0.000247	0.000247	0.000247	0.000247	0.000247	0.000247
	(0.001087)	(0.001086)	(0.001086)	(0.001086)	(0.001086)	(0.001087)
`vehicles`	0.015712	0.015713	0.015713	0.015712	0.015713	0.015712
	(0.000669)	(0.000668)	(0.000668)	(0.000668)	(0.000668)	(0.000669)
`dry`	0.106088	0.106098	0.106098	0.106088	0.106098	0.106088
	(0.034962)	(0.034927)	(0.034927)	(0.034929)	(0.034927)	(0.034962)
ρ_{Lag}	0.046928	0.046932	0.046932	0.046928	0.046932	0.046926
	(0.016982)	(0.016927)	(0.016927)	(0.016966)	(0.016927)	(0.016982)
ρ_{Err}	0.000957	−0.005621	−0.005621	0.000957	−0.005621	0.000957
	(–)	(0.034984)	(0.034984)	(–)	(0.034984)	(0.009316)

Table 3: SARAR estimation results for six implementations, DUI data set (standard errors in parentheses).

	R `spreg`	Stata `spivreg`	**PySAL** `GM_Endog_Combo_Hom`
Intercept	11.605968	11.605968	11.605968
	(1.666744)	(1.666744)	(1.666744)
`nondui`	−0.000196	−0.000196	−0.000196
	(0.002759)	(0.002759)	(0.002759)
`vehicles`	0.092996	0.092996	0.092996
	(0.005649)	(0.005649)	(0.005649)
`dry`	0.398260	0.398260	0.398260
	(0.090902)	(0.090902)	(0.090902)
`police`	−1.351308	−1.351308	−1.351308
	(0.141018)	(0.141018)	(0.141018)
ρ_{Lag}	0.193190	0.193190	0.193190
	(0.044310)	(0.044310)	(0.044310)
ρ_{Err}	−0.085975	−0.085975	−0.085975
	(0.030183)	(0.030183)	(0.030183)

Table 4: SARAR estimation when `police` is treated as endogenous. Results for three implementations, DUI data set (standard errors in parentheses).

the intercept (untransformed), the transformed exogenous variables (say $\mathbf{X}^\star$), and the spatial lags of them (say, $\mathbf{WX}^\star$ and $\mathbf{W}^2\mathbf{X}^\star$).[19] The R and **PySAL** implementations use $\mathbf{X}^\star$ (which may or may not include an intercept), and then spatial lags of $\mathbf{X}$. Finally, there are also differences in reported standard errors between the three implementations. These differences pertain to a degrees of freedom correction in the variance-covariance matrix operated in R and MATLAB. However, the same correction is available as an option from **PySAL**.[20]

It should also be noted that of the three available implementations, the SE toolbox is the only one that produces inference for ρ_{Err}.[21]

Let us turn now to the results shown in columns two, three, and five. From Table 3 it can be observed that, apart from a different decimal for the intercept calculated in Stata, all implementations otherwise match exactly. The only major differentiation among the three implementations is the possibility of setting a different matrix $\mathbf{A}_1$ in **PySAL**. As noted in Anselin (2013), there may be a problem with one of the sub-matrix (i.e., $\mathbf{\Omega}_{\delta\rho_{\text{Err}}}$) of the variance-covariance matrix of the estimated coefficients. The standard result that the variance-covariance matrix must be block-diagonal between the model coefficients and the error parameter may be invalidated by certain choices of $\mathbf{A}_1$ (e.g., the one used by Drukker *et al.* 2013a). The options of choosing alternative $\mathbf{A}_1$ in **PySAL** are meant to avoid this problem.

As in Drukker *et al.* (2013c), the set of explanatory variables includes `police`. Undoubtably, the size of the `police` force may be related with the arrest rates (`dui`). As a consequence, `police` is treated as an endogenous variable. Drukker *et al.* (2013c) also assume that the

[19] The present discussion is assuming that $\mathbf{W}$ and $\mathbf{M}$ are the same, otherwise $\mathbf{M}$ should be considered instead of $\mathbf{W}$.

[20] Unless otherwise stated, we tried to use the default values of the various implementations.

[21] Kelejian and Prucha (1999) do not derive the joint asymptotic distribution of the parameters of the model but consider ρ_{Err} as a nuisance parameter. Instructions on how the inference on the spatial parameter is calculated in the SE toolbox can be found on Ingmar Prucha's web page at `http://econweb.umd.edu/~prucha/STATPROG/OLS/desols.pdf`

dummy variable `elect` (where elect is 1 if a county government faces an election, 0 otherwise) constitutes a valid instrument for `police`.

Table 4 presents the results when this additional endogeneity is taken into account. Results from three different implementations are presented, namely the function `spreg` available from **sphet** under R, the command `spreg` available from Stata (setting the option to `het`), and the function `GM_Combo_Het` available from **PySAL**. A glance at Table 4 shows that all implementations give the same results. As pointed out in LeSage and Pace (2009), the estimated vector of parameters in a spatial model does not have the same interpretation as in a simple linear regression model. The estimate of ρ_{Lag} is positive and significant thus indicating spatial autocorrelation in the dependent variable. Drukker *et al.* (2013c) try to find some possible explanation for this. On the one hand, the positive coefficient may be explained in terms of coordination effort among police departments in different counties. On the other hand, it might well be that an enforcement effort in one of the counties leads people living close to the border to drink in neighboring counties. The estimate of ρ_{Err} is also significant but negative. As we will see later, those results are also in line with the evidence from the maximum likelihood estimation of the model.

One thing should be noted here. When we discussed the instrument matrix, we anticipated that when additional endogenous variables were present (as it is in this case), the optimal instruments are unknown. We also anticipated that, unless additional information was available, we did not recommended the inclusion of the spatial lag of these additional exogenous variables in the matrix of instruments. However, results reported in Table 4 do consider the spatial lags of `elect`. This is because, unlike R and **PySAL**, there is no option in Stata not to include them. For comparison purposes then, the default values of the corresponding arguments in R and **PySAL** were set to include those spatial lags.

Heteroskedasticity with and without additional endogenous variables

When the errors are assumed to be heteroskedastic of unknown form, the results are those presented in Table 5 (when no additional endogeneity is assumed) and Table 6 (when `police` is treated as endogenous). From Table 5 it can be noticed that the implementations in R and **PySAL** are identical (up to the sixth decimal), and that Stata only presents very minor differences. Those differences relate to the estimated value of the ρ_{Err} coefficient (obtained through the non-linear least square algorithm), and to the standard error of the intercept.

When the endogeneity of the `police` variable is taken into account, the implementations in R and **PySAL** are again identical, and Stata presents the same minor differences (see Table 6).

W *and* **M** *are different*

As we anticipated, **W** and **M** do not need to be the same in all applications. Since **PySAL** does not include this feature, results are limited to R and Stata. From an implementation perspective, the main complication in this case is the definition of the instrument matrix. While if **W** and **M** are identical the instruments need to be specified only in terms of **W**, when the two spatial weighting matrix are different the instruments should combine them as in Equation 12.

Table 7 summarizes the estimation results when `police` is endogenous and **W** and **M** are

	R `spreg`	Stata `spreg`	**PySAL** `GM_Combo_Het`
Intercept	−6.410088	−6.410088	−6.410088
	(0.445961)	(0.445958)	(0.445961)
`police`	0.598088	0.598088	0.598088
	(0.018154)	(0.018154)	(0.018154)
`nondui`	0.000247	0.000247	0.000247
	(0.001097)	(0.001097)	(0.001097)
`vehicles`	0.015713	0.015713	0.015713
	(0.000784)	(0.000784)	(0.000784)
`dry`	0.106121	0.106121	0.106121
	(0.033807)	(0.033807)	(0.033807)
ρ_{Lag}	0.046944	0.046944	0.046944
	(0.017928)	(0.017928)	(0.017928)
ρ_{Err}	−0.000366	−0.000378	−0.000366
	(0.036803)	(0.036803)	(0.036803)

Table 5: SARAR estimation with heteroskedasticity. Results for three implementations, DUI data set (standard errors in parentheses).

	R `spreg`	Stata `spivreg`	**PySAL** `GM_Combo_Het`
Intercept	11.649298	11.649298	11.649298
	(1.873178)	(1.873179)	(1.873178)
`nondui`	−0.000155	−0.000155	−0.000155
	(0.002843)	(0.002843)	(0.002843)
`vehicles`	0.093058	0.093058	0.093058
	(0.005967)	(0.005967)	(0.005967)
`dry`	0.398707	0.398707	0.398707
	(0.094791)	(0.094791)	(0.094791)
`police`	−1.352871	−1.352871	−1.352871
	(0.149223)	(0.149223)	(0.149223)
ρ_{Lag}	0.192149	0.192149	0.192149
	(0.051833)	(0.051833)	(0.051833)
ρ_{Err}	−0.050266	−0.050263	−0.050266
	(0.039931)	(0.039931)	(0.039931)

Table 6: SARAR estimation with heteroskedasticity when `police` is treated as endogenous. Results for three implementations, DUI data set (standard errors in parentheses).

different.[22] Since the endogeneity of the `police` variable is accounted for, the default value to compute the lagged "additional" instruments (i.e., `lag.instr`) was changed in R. The results from the two implementations are almost identical except for a difference in the last decimal digit for the estimate of ρ_{Err}. This is due to small differences in the optimization routines adopted in the two environments. When the residuals are assumed to be heteroskedastic, a similar evidence is observed.[23]

[22] **M** is defined as a row-standardized six nearest neighbors matrix, treating the county centroid coordinates as projected, not geographical.

[23] Results are not reported but are available from the authors upon request.

	R spreg	Stata spivreg
Intercept	9.210831	9.210831
	(1.454592)	(1.454592)
nondui	−0.000238	−0.000238
	(0.002480)	(0.002480)
vehicles	0.083249	0.083249
	(0.004799)	(0.004799)
dry	0.361584	0.361584
	(0.081570)	(0.081570)
police	−1.105622	−1.105622
	(0.119709)	(0.119709)
ρ_{Lag}	0.180395	0.180395
	(0.040716)	(0.040716)
ρ_{Err}	−0.011908	−0.011906
	(0.033255)	(0.033255)

Table 7: SARAR estimation when police is treated as endogenous and **W** and **M** are different. Results for two implementations, DUI data set (standard errors in parentheses).

3.2. Spatial lag model

The estimation of the spatial lag model in Equation 7 can be easily undertaken using two stage least squares. If the only restriction is $\rho_{\text{Err}} = 0$ and there are exogenous variables other than the spatial lag in the model (i.e., if $\pi \neq 0$), additional instruments are then needed in order to estimate the model.

When homoskedasticity is assumed, the variance covariance matrix can be estimated consistently by

$$\tilde{\sigma}^2(\tilde{\mathbf{Z}}^\top\tilde{\mathbf{Z}})^{-1} \tag{22}$$

where $\tilde{\sigma}^2 = n^{-1}\sum_{i=1}^{n}\tilde{\mathbf{u}}_i^2$ and $\tilde{\mathbf{u}} = \mathbf{y} - \mathbf{Z}\tilde{\delta}$. On the other hand, when heteroskedasticity is assumed the estimation of the coefficients variance covariance matrix takes the following sandwich form:

$$(\tilde{\mathbf{Z}}^\top\tilde{\mathbf{Z}})^{-1}(\tilde{\mathbf{Z}}^\top\tilde{\Sigma}\tilde{\mathbf{Z}})(\tilde{\mathbf{Z}}^\top\tilde{\mathbf{Z}})^{-1} \tag{23}$$

where $\hat{\Sigma}$ is a diagonal matrix whose elements are the squared residuals.

Homoskedasticity with and without additional endogenous variables

From Table 8 we can see that five implementations are available of the spatial lag model when considering homoskedastic innovations and no additional endogenous variables. The first two columns of the table report results obtained from R. There are multiple functions that allow the estimation of the spatial lag model available from R under the **spdep** and **sphet** packages. stsls was the first function made available as part of the package **spdep**.[24] With the development of **sphet**, the wrapper function spreg also allows the estimation of the spatial lag model. The other three columns of Table 8 report the results obtained from Stata, **PySAL**, and the SE toolbox, respectively.

[24] stsls was originally contributed by Luc Anselin.

	R `stsls`	R `spreg`	Stata `spreg gs2sls`	**PySAL** `GM_Lag`	SE `sar_gmm`
Intercept	−6.410152	−6.410152	−6.410152	−6.410152	−6.410152
	(0.418129)	(0.418129)	(0.417725)	(0.417725)	(0.418129)
`police`	0.598081	0.598081	0.598081	0.598081	0.598081
	(0.014918)	(0.014918)	(0.014904)	(0.014904)	(0.014918)
`nondui`	0.000247	0.000247	0.000247	0.000247	0.000247
	(0.001087)	(0.001087)	(0.001086)	(0.001086)	(0.001087)
`vehicles`	0.015714	0.015714	0.015714	0.015714	0.015714
	(0.000669)	(0.000669)	(0.000668)	(0.000668)	(0.000669)
`dry`	0.106134	0.106134	0.106134	0.106134	0.106134
	(0.034962)	(0.034962)	(0.034928)	(0.034928)	(0.034962)
ρ_{Lag}	0.046950	0.046950	0.046950	0.046950	0.046950
	(0.016977)	(0.016977)	(0.016960)	(0.016960)	(0.016977)

Table 8: Lag model estimation results for five implementations. DUI data set (standard errors in parentheses).

	R `spreg`	Stata `spivreg`	**PySAL** `GM_Lag`
Intercept	11.507606	11.507606	11.507606
	(1.686222)	(1.684594)	(1.686222)
`nondui`	−0.000293	−0.000293	−0.000293
	(0.002771)	(0.002768)	(0.002771)
`vehicles`	0.092866	0.092866	0.092866
	(0.005663)	(0.005657)	(0.005663)
`dry`	0.397357	0.397357	0.397357
	(0.091419)	(0.091331)	(0.091419)
`police`	−1.348024	−1.348024	−1.348024
	(0.141410)	(0.141273)	(0.141410)
ρ_{Lag}	0.195595	0.195595	0.195595
	(0.045906)	(0.045862)	(0.045906)

Table 9: Lag model estimation when `police` is treated as endogenous. Results for three implementations. DUI data set (standard errors in parentheses).

Given that we are considering the same matrix of instruments, the coefficient values of all implementations agree exactly. This must be expected given that the OLS results also matched. There are differences, though, in the reported standard errors. In the two (R and SE toolbox) functions, the error variance is calculated with a degrees of freedom correction (i.e., dividing by $n - k$), while in the other two implementations is simply divided by n.

Table 9 shows the results for the lag model when `police` is treated as endogenous. While all the coefficients match, differences in the standard error are produced by the same degrees of freedom correction.

Heteroskedasticity with and without additional endogenous variables

Apart from MATLAB SE toolbox, all other implementations (including the two R functions `stsls` and `spreg`) allow the estimation of the lag model under heteroskedastic innovations. Of

course, the estimated coefficients are not different from the homoskedastic case, and the only variation is in the standard errors. However, the standard errors under heteroskedasticity are equal across the four models, and, therefore, we are not reporting them in the paper.[25]

HAC estimation

A different specification of the model can be based on the assumptions that the error term follows

$$u = \mathbf{R}\varepsilon \tag{24}$$

where $\mathbf{R}$ is an $n \times n$ unknown non-stochastic matrix, and ε is a vector of innovations. The asymptotic distribution of the corresponding IV estimators involves the VC matrix:

$$\mathbf{\Psi} = n^{-1}\mathbf{H}^\top \mathbf{\Omega} \mathbf{H} \tag{25}$$

where $\mathbf{\Omega} = \mathbf{R}\mathbf{R}^\top$ denotes the variance covariance matrix of ε. Kelejian and Prucha (2007) suggest estimating the individual r, s elements of $\mathbf{\Psi}$ as

$$\tilde{\psi}_{rs} = n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} h_{ir} h_{js} \hat{\varepsilon}_i \hat{\varepsilon}_j K(d^*_{ij}/d) \tag{26}$$

where the subscripts refer to the elements of the matrix of instruments $\mathbf{H}$ and the vector of estimated residuals $\hat{\varepsilon}$. The Kernel function $K()$ is defined in terms of the distance measure d^*_{ij}, the distance between observations i and j. The bandwidth d is such that if $d^*_{ij} \geq d$, the associated Kernel is set to zero ($K(d^*_{ij}/d) = 0$). In other words, the bandwidth together with the Kernel function limits the number of sample covariances.

Based on Equation 26, the asymptotic variance covariance matrix ($\tilde{\mathbf{\Phi}}$) of the S2SLS estimator of the parameters vector is given by:

$$\tilde{\mathbf{\Phi}} = n^2 (\tilde{\mathbf{Z}}^\top \tilde{\mathbf{Z}})^{-1} \mathbf{Z}^\top \mathbf{H} (\mathbf{H}^\top \mathbf{H})^{-1} \tilde{\mathbf{\Psi}} (\mathbf{H}^\top \mathbf{H})^{-1} \mathbf{H}^\top \mathbf{Z} (\tilde{\mathbf{Z}}^\top \tilde{\mathbf{Z}})^{-1} \tag{27}$$

Piras (2010) mentioned that the implementation was compared with code used by Anselin and Lozano-Gracia (2008) and that the two implementations gave the same results. However, at the time Piras (2010) was published, no other public implementation was available. Here we compare standard error estimates using a Triangular kernel with a variable bandwidth of the six nearest neighbors.[26]

The example in Table 10 refers to the case in which `police` is treated as endogenous. Of course, the coefficients are the same and are comparable to those in Table 9. It is interesting to note that also the standard errors are the same between the two implementations. Also, as one would expect, they are a little bit higher than those in Table 9. When there are no additional endogenous variables other than the spatial lag, the results from the two implementation match well and are not reported here. Of course, the model can also be estimated by OLS when no form of endogeneity is present. Results are in accordance also in this case.

[25] These and other results are available from the authors upon request.

[26] After reading the coordinates for each location, we used the function `knearneigh` to generate the distances and, after some transformation between objects, the final kernel using the R function `read.gwt2dist`. In **PySAL**, the function `pysal.kernelW_from_shapefile` was used to directly read the shape file in and generate the kernel. In both cases, the centroids of the counties were used as point locations, but the distances for the nearest neighbor procedure were calculated as though the coordinates were projected, while they are in fact geographical. There are many available options for the kernel both in R and **PySAL**.

	R `spreg`	**PySAL** `GM_Lag`
Intercept	11.507606	11.507606
	(1.842620)	(1.842620)
`nondui`	−0.000293	−0.000293
	(0.002805)	(0.002805)
`vehicles`	0.092866	0.092866
	(0.005980)	(0.005980)
`dry`	0.397357	0.397357
	(0.095334)	(0.095334)
`police`	−1.348024	−1.348024
	(0.149460)	(0.149460)
ρ_{Lag}	0.195595	0.195595
	(0.054681)	(0.054681)

Table 10: HAC estimation when `police` is treated as endogenous. Results for two implementations. DUI data set (standard errors in parentheses).

3.3. Spatial error model

In the spatial error model (i.e., $\rho_{\text{Lag}} = 0$), we can still use most of the formulae above. The first step of the estimation procedure is either OLS (when $\pi = 0$), or IV, when $\pi \neq 0$ and there are endogenous variable (other than the spatial lag) in the model. After estimating ρ_{Err} in the GMM step, we can then take the spatial Cochrane-Orcutt transformation. The resulting model can be then estimated by two stage least squares using the matrix of instruments $\mathbf{H}$, where $\mathbf{H}$ is made up of, at least, the linearly independent columns of $\mathbf{X}$, and $\mathbf{MX}$.

Homoskedasticity with and without additional endogenous variables

Table 11 shows the spatial error model estimation results for six different implementations. These implementations include two R functions (`GMerrorsar` available from **spdep**,[27] and `spreg` available from **sphet**), one Stata command (`spreg gs2sls`), two different implementation based on **PySAL** (`GM_Error` and `GM_Error_Hom`), and, finally, the SE toolbox for MATLAB (`sem_gmm`).

Three of them are based on the Kelejian and Prucha (1999) moment conditions (columns one, four, and six), the others are based on Drukker *et al.* (2013a) moment conditions (columns two, three, and five). Let us focus first on the results obtained using Kelejian and Prucha (1999) moment conditions. All three implementations produce the same identical estimated coefficients. There are differences in terms of the standard errors. We refer to the fact that, while `GMerrorsar` and `sem_gmm` produce an estimate for the standard error of the spatial coefficient, the `GM_Error` function in **PySAL** does not. The reason is that in their original contribution Kelejian and Prucha (1999) do not derive the joint asymptotic distribution of the parameters of the model and, therefore, `GM_Error` does not produce any inference on ρ_{Err}.[28]

Moving now to the three implementations that are based on Drukker *et al.* (2013a) moment conditions, we observe that, while Stata and `spreg` (available from R) present exactly the

[27] `GMerrorsar` was originally contributed by Luc Anselin.

[28] Instructions on how to produce inference are again available from Ingmar Prucha's website at `http://econweb.umd.edu/~prucha/STATPROG/OLS/desols.pdf`.

	R `GMerrorsar`	R `spreg`	Stata `spreg gs2sls`	**PySAL** `GM_Error`	**PySAL** `GM_Error_Hom`	SE `sem_gmm`
Intercept	−5.431921	−5.431959	−5.431959	−5.431921	−5.431960	−5.431921
	(0.229056)	(0.229067)	(0.229067)	(0.229052)	(0.229050)	(0.229056)
`police`	0.599854	0.599851	0.599851	0.599854	0.599851	0.599854
	(0.014888)	(0.014890)	(0.014890)	(0.014888)	(0.014887)	(0.014888)
`nondui`	0.000257	0.000257	0.000257	0.000257	0.000257	0.000257
	(0.001086)	(0.001086)	(0.001086)	(0.001086)	(0.001086)	(0.001086)
`vehicles`	0.015612	0.015612	0.015612	0.015612	0.015612	0.015612
	(0.000667)	(0.000667)	(0.000667)	(0.000667)	(0.000667)	(0.000667)
`dry`	0.103654	0.103663	0.103663	0.103654	0.103663	0.103654
	(0.034966)	(0.034967)	(0.034967)	(0.034966)	(0.034965)	(0.034966)
ρ_{Err}	0.050883	0.047050	0.047050	0.050883	0.051491	0.050883
	(0.080487)	(0.029543)	(0.029543)	(–)	(0.028809)	(0.080487)

Table 11: Error model estimation results for six implementations. DUI data set (standard errors in parentheses).

	R spreg	Stata spivreg	**PySAL** GM_Endog_Error_Hom
Intercept	15.484115	15.484115	17.326391
	(1.578958)	(1.578959)	(1.748870)
nondui	−0.000208	−0.000208	−0.000226
	(0.002755)	(0.002755)	(0.002962)
vehicles	0.092430	0.092430	0.099270
	(0.005655)	(0.005655)	(0.006272)
dry	0.395797	0.395797	0.421893
	(0.090901)	(0.090901)	(0.097922)
police	−1.337080	−1.337080	−1.508904
	(0.141153)	(0.141153)	(0.156655)
ρ_{Err}	−0.004487	−0.004483	−0.005472
	(0.025467)	(0.025467)	(0.025496)

Table 12: Error model estimation when police is treated as endogenous. Results for three implementations. DUI data set (standard errors in parentheses).

same results, some distinctions are observed in **PySAL**.[29] The reason for these differences was anticipated before. In a spatial error model, the second term in Equation 21 limits to zero (when there are only exogenous variables in the model). The implementations in R and Stata produce an estimate of this term, while **PySAL** does not. In fact, we modified the R code for the error model setting the second term of Equation 21 to zero. As a consequence, we obtain the same results of **PySAL**.[30] However, as it can be appreciated from Table 11, the differences (with this specific dataset) are very minor.

Table 12 displays the results produced by three implementations when police is instrumented. The only three implementations in which this feature is available are R, Stata, and **PySAL**. A glance at the table reveals that the results across implementations are very different. The differences between R and Stata are very minor and they can be attributable to differences in optimization routines. The differences with **PySAL** relates to a different specification of the instrument matrix. The instrument matrix in **PySAL** is specified in terms of the exogenous variables (i.e., $\mathbf{X}$), which are augmented with the only additional instrument (elect). However, one could argue that while the initial step only involves the $\mathbf{X}$ matrix, the third step (after the GMM estimator is obtained) also involve the $\mathbf{MX}$ (through $\mathbf{X}^{\star}$).

Heteroskedasticity with and without additional endogenous variables

The same evidence is confirmed also when heteroskedasticity is assumed and, therefore, results are not reported in the paper but are available from the authors. We will make one final remark. In our theoretical review of the general spatial model we anticipated the possibility of obtaining a consistent and efficient estimator of ρ_{Err} based on 2SLS rather than GS2SLS.

Table 13 reports the results when considering this additional step (i.e., step1.c) while taking all the variables as exogenous. Stata does not present this feature. Results from R and **PySAL**

[29] Anselin, Amaral, and Arribas-Bel (2012) report different results between the implementations of Stata and **PySAL**. We tried to replicate the results in Table 2 of Anselin *et al.* (2012) and noticed that the latest release of Stata gives results that are different from those reported in the table. However, the new Stata results match exactly those given in the third column of Table 2 (i.e., sphet2).

[30] Results are available from the authors upon request.

	R spreg	PySAL GM_Endog_Error_Het
Intercept	5.432185	5.431451
	(0.255239)	(0.255217)
police	0.599834	0.599890
	(0.018086)	(0.018084)
nondui	0.000257	0.000256
	(0.001096)	(0.001096)
vehicles	0.015614	0.015609
	(0.000781)	(0.000780)
dry	0.103715	0.103545
	(0.033712)	(0.033710)
Err	0.051595	0.056201
	(0.030924)	(0.029984)

Table 13: Error estimation results. Results for two implementations (step 1.c). DUI data set (standard errors in parentheses).

are very similar and again di erences are due to the estimation of ( ).[31]

4. Comparing maximum likelihood estimation

Having compared implementations of GMM estimators, we move to cross-section maximum likelihood (ML) estimation, recalling that ML estimation for spatial panel models was compared for MATLAB and R implementations in Millo and Piras (2012). Since Python PySAL has no ML implementations, it will not be considered. None of the ML implementations make provision for instrumenting endogenous right hand side variables, nor for accommodating heteroskedasticity { this is provided for in Bayesian implementations in the Spatial Econometrics toolbox.

In Section 1.1 (page 4), we described the numerical optimizers used in the various applications. In many cases, the numerical optimization functions can return numerical Hessians for use as estimators of the covariance matrix of model coe cients, which may be used instead of analytical, asymptotic covariance matrices. In other cases, the numerical Hessian may be found by examining the form of the function being optimized around the optimum, for example using nite-di erence Hessian algorithms. In implementations in the MATLAB Spatial Econometrics toolbox, use is made ofdhess taken from the CompEcon toolbox by Paul Fackler, but with the relative step size hard-coded to 10 10^{-8} in sar , sdmand sac, but to 1:0 10^{-5} in semas in the original CompEcon function. In R, use is made offdHess from the nlme (Pinheiro, Bates, DebRoy, Sarkar, and R Core Team 2013) package with a default relative step size of 6055 10^{-6} used without modi cation. In these comparisons in R, we will usually use analytical, asymptotic covariance matrices, but numerical Hessians are used sometimes for comparison.

An extensive treatment of maximum likelihood estimation for spatial models is given by LeSage and Pace (2009, pp. 45{75). Further discussion of the general model, and issues arising in tting two parameters when the surface of the function shows little structure may

[31] Evidence for this can be obtained from the authors upon request.

Anselin L (2010). "Thirty Years of Spatial Econometrics." *Papers in Regional Science*, **89**(1), 3–25.

Anselin L (2013). "GMM Estimation of Spatial Error Autocorrelation with and without Heteroskedasticity." Manuscript.

Anselin L, Amaral PV, Arribas-Bel D (2012). "Technical Aspects of Implementing GMM Estimation of the Spatial Error Model in **PySAL** and **GeoDaSpace**." Technical report, URL https://geodacenter.asu.edu/drupal_files/sperrorgmm_wp2.pdf.

Anselin L, Bera AK (1998). "Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics." In A Ullah, DE Giles (eds.), *Handbook of Applied Economic Statistics*, pp. 237–289. Marcel Dekker, New York.

Anselin L, Bera AK, Florax RJGM, Yoon MJ (1996). "Simple Diagnostic Tests for Spatial Dependence." *Regional Science and Urban Economics*, **26**, 77–104.

Anselin L, Lozano-Gracia N (2008). "Errors in Variables and Spatial Effects in Hedonic House Price Models of Ambient Air Quality." *Empirical Economics*, **34**(1), 5–34.

Anselin L, Syabri I, Kho Y (2006). "**GeoDa**: An Introduction to Spatial Data Analysis." *Geographical Analysis*, **38**(1), 5–22.

Aptech Systems, Inc (2007). *GAUSS User Guide*. Maple Valley. URL http://www.aptech.com/wp-content/uploads/2012/08.

Arraiz I, Drukker DM, Kelejian HH, Prucha IR (2010). "A Spatial Cliff-Ord-Type Model with Heteroskedastic Innovations: Small and Large Sample Results." *Journal of Regional Science*, **50**(2), 592–614.

Bengtsson H (2005). "**R.matlab**: Local and Remote Matlab Connectivity in R." Mathematical Statistics, Centre for Mathematical Sciences, Lund University, Sweden. (manuscript in progress).

Bivand RS (2002). "Spatial Econometrics Functions in R: Classes and Methods." *Journal of Geographical Systems*, **4**, 405–421.

Bivand RS (2006). "Implementing Spatial Data Analysis Software Tools in R." *Geographical Analysis*, **38**(1), 23–40. ISSN 0016-7363.

Bivand RS (2012). "After 'Raising the Bar': Applied Maximum Likelihood Estimation of Families of Models in Spatial Econometrics." *Estadística Española*, **54**, 71–88.

Bivand RS (2013). ***spdep**: Spatial Dependence: Weighting Schemes, Statistics and Models*. R package version 0.5-56, URL http://CRAN.R-project.org/package=spdep.

Bivand RS, Gebhardt A (2000). "Implementing Functions for Spatial Statistical Analysis Using the R Language." *Journal of Geographical Systems*, **2**, 307–317.

Bivand RS, Hauke J, Kossowski T (2013a). "Computing the Jacobian in Gaussian Spatial Autoregressive Models: An Illustrated Comparison of Available Methods." *Geographical Analysis*, **45**(2), 150–179.

Bivand RS, Pebesma E, Gómez-Rubio V (2013b). *Applied Spatial Data Analysis with R*. 2nd edition. Springer-Verlag, New York.

Brent R (1973). *Algorithms for Minimization without Derivatives*. Prentice-Hall, Englewood Cliffs.

Burridge P (1980). "On the Cliff-Ord Test for Spatial Correlation." *Journal of the Royal Statistical Society B*, **42**, 107–108.

Cliff A, Ord JK (1972). "Testing for Spatial Autocorrelation among Regression Residuals." *Geographical Analysis*, **4**, 267–284.

Das D, Kelejian HH, Prucha IR (2003). "Finite Sample Properties of Estimators of Spatial Autoregressive Models with Autoregressive Disturbances." *Papers in Regional Science*, **82**, 1–27.

Drukker DM, Egger P, Prucha IR (2013a). "On Two-Step Estimation of a Spatial Autoregressive Model with Autoregressive Disturbances and Endogenous Regressors." *Econometric Reviews*, **32**(5-6), 686–733.

Drukker DM, Peng IRPH, Raciborski R (2013b). "Creating and Managing Spatial-Weighting Matrices with the `spmat` Command." *Stata Journal*, **13**(2), 242–286.

Drukker DM, Prucha IR, Raciborski R (2013c). "A Command for Estimating Spatial-Autoregressive Models with Spatial-Autoregressive Disturbances and Additional Endogenous Variables." *Stata Journal*, **13**(2), 287–301.

Drukker DM, Prucha IR, Raciborski R (2013d). "Maximum Likelihood and Generalized Spatial Two-Stage Least-Squares Estimators for a Spatial-Autoregressive Model with Spatial-Autoregressive Disturbances." *Stata Journal*, **13**(2), 221–241.

Florax RJGM, Folmer H (1992). "Specification and Estimation of Spatial Linear Regression Models: Monte Carlo Evaluation of Pre-Test Estimators." *Regional Science and Urban Economics*, **22**, 405–432.

Florax RJGM, Folmer H, Rey SJ (2003). "Specification Searches in Spatial Econometrics: The Relevance of Hendry's Methodology." *Regional Science and Urban Economics*, **33**, 557–579.

Gay DM (1990). "Usage Summary for Selected Optimization Routines." Computing Science Technical Report No. 153.

Gould W, Pitblado JS, Poi B (2010). *Maximum Likelihood Estimation with Stata*. StataCorp LP, College Station.

Griffith DA, Layne LJ (1999). *A Casebook for Spatial Statistical Data Analysis*. Oxford University Press, New York.

IBM Corporation (2010). *IBM SPSS Statistics 19*. Armonk. IBM Corporation, URL http://www-01.ibm.com/software/analytics/spss/.

Jones E, Oliphant T, Peterson P, others (2001). "**SciPy**: Open Source Scientific Tools for Python." URL http://www.scipy.org/.

Kelejian HH, Piras G (2011). "An Extension of Kelejian's J-Test for Non-Nested Spatial Models." *Regional Science and Urban Economics*, **41**, 281–292.

Kelejian HH, Prucha IR (1997). "Estimation of Spatial Regression Models with Autoregressive Errors by Two-Stage Least Squares Procedures: A Serious Problem." *International Regional Science Review*, **20**(1&2), 103–111.

Kelejian HH, Prucha IR (1998). "Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances." *Journal of Real Estate Finance and Economics*, **17**(1), 99–121.

Kelejian HH, Prucha IR (1999). "A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model." *International Economic Review*, **40**, 509–533.

Kelejian HH, Prucha IR (2007). "HAC Estimation in a Spatial Framework." *Journal of Econometrics*, **140**(1), 131–154.

Kelejian HH, Prucha IR (2010). "Specification and Estimation of Spatial Autoregressive Models with Autoregressive and Heteroskedastic Disturbances." *Journal of Econometrics*, **157**(1), 53–67.

Kelejian HH, Prucha IR, Yuzefovich Y (2004). "Instrumental Variable Estimation of a Spatial Autoregressive Model with Autoregressive Disturbances: Large and Small Sample Results." In JP LeSage, KR Pace (eds.), *Advances in Econometrics: Spatial and Spatiotemporal Econometrics*, pp. 163–198. Elsevier, Oxford.

Kelejian HH, Tavlas GS, Hondroyiannis G (2006). "A Spatial Modelling Approach to Contagion among Emerging Economies." *Open Economies Review*, **17**(4-5), 423–441.

Lee LF (2003). "Best Spatial Two-Stage Least Squares Estimators for a Spatial Autoregressive Model with Autoregressive Disturbances." *Econometric Reviews*, **22**, 307–335.

Lee LF (2007). "GMM and 2SLS Estimation of Mixed Regressive, Spatial Autoregressive Models." *Journal of Econometrics*, **137**, 489–514.

LeSage JP, Fischer MM (2008). "Spatial Growth Regression: Model Specification, Estimation and Interpretation." *Spatial Economic Analysis*, **3**, 275–304.

LeSage JP, Pace KR (2009). *Introduction to Spatial Econometrics*. CRC Press, Boca Raton, FL.

McMillen D (2012). ***McSpatial**: Nonparametric Spatial Data Analysis*. R package version 1.1.1, URL http://CRAN.R-project.org/src/contrib/Archive/McSpatial/McSpatial_1.1.1.tar.gz.

Millo G, Piras G (2012). "**splm**: Spatial Panel Data Models in R." *Journal of Statistical Software*, **47**(1), 1–38. URL http://www.jstatsoft.org/v47/i01/.

Nash JC, Varadhan R (2011). "Unifying Optimization Algorithms to Aid Software System Users: **optimx** for R." *Journal of Statistical Software*, **43**(9), 1–14. URL http://www.jstatsoft.org/v43/i09/.

Oliphant TE (2006). "Guide to **NumPy**." Trelgol Publishing, URL http://numpy.scipy.org/.

Ord JK (1975). "Estimation Methods for Models of Spatial Interaction." *Journal of the American Statistical Association*, **70**(349), 120–126.

Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team (2013). ***nlme**: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-109, URL http://CRAN.R-project.org/package=nlme.

Piras G (2010). "**sphet**: Spatial Models with Heteroskedastic Innovations in R." *Journal of Statistical Software*, **35**(1), 1–21. URL http://www.jstatsoft.org/v35/i01/.

Pisati M (2001). "sg162. Tools for Spatial Data Analysis." *Stata Technical Bulletin*, **10**(60), 21–37.

Powers EL, Wilson JK (2004). "Access Denied: The Relationship between Alcohol Prohibition and Driving under the Influence." *Sociological Inquiry*, **74**(3), 318–337.

R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Rey SJ (2009). "Show Me the Code: Spatial Analysis and Open Source." *Journal of Geographical Systems*, **11**, 191–207.

Rey SJ, Anselin L (2010). "**PySAL**: A Python Library of Spatial Analytical Methods." In MM Fischer, A Getis (eds.), *Handbook of Applied Spatial Analysis*, pp. 175–193. Springer-Verlag, Berlin.

Ripley BD (1981). *Spatial Statistics*. John Wiley & Sons, New York.

SAS Institute Inc (2008). *SAS/IML 9.2 User's Guide*. SAS Institute Inc., Cary. URL http://www.sas.com/.

Schnabel RB, Koontz JE, Weiss BE (1985). "A Modular System of Algorithms for Unconstrained Minimization." *ACM Transactions on Mathematical Software*, **11**(4), 419–440.

StataCorp (2011). *Stata Data Analysis Statistical Software: Release 12*. StataCorp LP, College Station, TX. URL http://www.stata.com/.

The MathWorks, Inc (2011). *MATLAB – The Language of Technical Computing, Version 7.13 (R2011b)*. The MathWorks, Inc., Natick, Massachusetts. URL http://www.mathworks.com/products/matlab/.

van Rossum G (1995). *Python Reference Manual*. CWI Report, URL http://ftp.cwi.nl/CWIreports/AA/CS-R9526.ps.Z.

Ward MD, Gleditsch KS (2008). *Spatial Regression Models*. Sage, Thousand Oaks.

Affiliation:

Roger Bivand
Department of Economics
Norwegian School of Economics
Helleveien 30
5045 Bergen, Norway
E-mail: Roger.Bivand@nhh.no

Gianfranco Piras
Regional Research Institute
West Virginia University
886 Chestnut Ridge Road
P.O. Box 6825
Morgantown, WV 26506-6825, United States of America
E-mail: Gianfranco.Piras@mail.wvu.edu

Journal of Statistical Software http://www.jstatsoft.org/
published by the American Statistical Association http://www.amstat.org/
Volume 63, Issue 18 *Submitted:* 2013-05-23
January 2015 *Accepted:* 2014-10-13