DOI: 10.5604/01.3001.0014.0825

PRZEGLĄD STATYSTYCZNY
R. LXIV – ZESZYT 3 – 2017

ALICJA OLEJNIK[1], JAKUB OLEJNIK[2]

AN ALTERNATIVE TO PARTIAL REGRESSION IN MAXIMUM LIKELIHOOD ESTIMATION OF SPATIAL AUTOREGRESSIVE PANEL DATA MODEL[3] [4]

1. INTRODUCTION

Partial regression, developed by Frisch, Waugh (1933), is a popular method of elimination of nuisance slope parameters. It is widely used in *inter alia* panel data analysis. One of its special cases is commonly used to estimate regression parameters in, so called, fixed effects models. Partial regression allows one to find slope parameters without the need of estimating actual levels of fixed effects. In this form it is referred to as the *demeaning* procedure (cf. Baltagi, 2005).

As the Maximum Likelihood (ML) estimation procedure is one of the most popular estimation methods for Spatial Autoregressive Model (SAR) many researchers have also used the technique of demeaning in ML estimation of the SAR model. However, validity of this approach has been occasionally subjected to doubt (e.g. Anselin et al., 2006) on the grounds that the demeaning procedure yields singular variance of the error term. As Pace (2014) rightly points out, maximising demeaned likelihood can still produce consistent estimates[5] of regression parameters, as the demeaned likelihood can be interpreted as a concentrated likelihood. However, estimates of their variances are likely to be invalid.

A procedure to overcome this problem was first proposed by Lee, Yu (2010) for a reasonably general spatial fixed time/individual fixed effect model. They noticed that applying certain transformation of data, prior to conducting ML procedure, can effectively eliminate fixed effects and, at the same time, properly account for the singularity. In this paper we generalise this approach and show that, contrary to a statement included in Lee, Yu (2010), ML estimation with demeaning of SAR model is feasible in a larger class of settings than originally described.

---

[1] University of Lodz, Faculty of Economics and Sociology, Department of Spatial Econometrics, 37 Rewolucji 1905 r. St., 90-214 Lodz, Poland, corresponding author – e-mail: olejnika@uni.lodz.pl.

[2] University of Lodz, Faculty of Mathematics and Computer Science, Department of Applied Computer Science, 22 Banacha St., 90-238 Lodz, Poland.

[3] This research has been supported by the NCN grant no. 2011/03/D/HS4/04305.

[4] We would like to thank the anonymous referees for their constructive comments, which helped improve this paper and suggested interesting topics for authors' further research.

[5] This is, naturally, to say that the ML estimator is consistent.

Our invariant subspace framework allows, under some assumptions, to effectively deal with large class of fixed effects designs in panel and non-panel data models. Designs handled by the framework range from group-specific fixed effects with non-uniform cardinality to multiple levels of group-specific effects with possibly overlapping groups and non-constant (yet known) effect sizes within those groups. This can be done under the assumption that the Krylov subspace[6] for spatially lagged fixed effects is of incomplete dimension. The crucial requirement expresses certain degree of compatibility of the fixed effects design with assumed spatial weight matrix. In the original paper of Lee, Yu (2010) the considered model specification also includes spatially correlated error term, however in our paper, for simplicity of presentation, we employ only autoregressive scheme. Therefore, the aim of our paper is to develop extension of the fixed effect eliminating transformation of Lee, Yu (2010), so that effectively a larger class of fixed effect designs can be handled.

Unless specified differently, throughout the paper we use the short term SAR model to actually describe the panel-data SAR model. All statements applicable to non-panel data SAR model are also valid in the panel case. Whenever n is used to denote sample size, it can be read $n = NT$, moreover $\mathbb{R}^n = \mathbb{R}^N \otimes \mathbb{R}^T$, $\mathrm{I}_n = \mathrm{I}_N \otimes \mathrm{I}_T$ etc. This notation can also cover the case of either spatial unit unbalanced or time unbalanced panel data set, that is if $N = n_1 + \cdots + n_T$ and $\mathbb{R}^n = \mathbb{R}^{n_1} \oplus \ldots \oplus \mathbb{R}^{n_T}$, or if $T = t_1 + \cdots + t_N$, etc., respectively.

The rest of this paper is structured as follows. Section 2 introduces the concept of partial regression by Frisch, Waugh (1933). Section 3 introduces Spatial Autoregressive Model specification and describes the well-known naive approach of demeaning in ML estimation. Section 4 presents our original approach. Section 5 formulates statements on asymptotic behaviour of our estimator. Finally, section 6 presents a summary and conclusions.

2. PARTIAL REGRESSION

Although for our purposes it is enough to consider the basic form of the Frisch-Waugh (F-W) theorem, it is worthwhile to mention some of its interesting extensions. In particular, Fiebig, Bartels (1996) develop an extension of the F-W procedure that is able to handle model specifications with non-spherical disturbances, that is where the variance covariance matrix of the error term is not proportional to identity matrix.

Another interesting extension to partial regression has been recently developed in Yamada (2016). It has been shown that the F-W theorem is invariant under certain modification of the least squares objective function. Namely, if instead of the usual least squares optimization problem

$$\hat{\beta} = \arg\min_{\beta \in \mathbb{R}^k} \|Y - X\beta\|^2$$

[6] To be defined in section 4, can be found also in e.g. Liesen, Strakoš (2013).

we consider the LASSO (least absolute shrinkage and selection operator) regression. This can be described as solution to the modified problem $\hat{\beta} = \arg\min_{\beta\in\mathbb{R}^k}\|Y - X\beta\|^2 + + \lambda\|\beta\|_1$, where λ is a tuning parameter and $\|\beta\|_1 = \sum_{i<k}|\beta_i|$. Similarly, the F-W theorem still holds if the usual least squares is replaced with ridge regression i.e. $\hat{\beta} = \arg\min_{\beta\in\mathbb{R}^k}\|Y - X\beta\|^2 + \lambda\|\beta\|^2$.

Those results suggest that partial regression might be a technique applicable in a variety of estimation schemes. The maximum likelihood estimation procedure is one of them. This is implied by fact of equivalence of the estimates form OLS and ML approaches under normality of error term. However, the question of applicability of partial regression becomes far more difficult if one considers the spatially autoregressive term in model specification. In our paper we show that, under some assumptions, the Maximum Likelihood estimation procedure in case of a spatially autoregressive DGP can also benefit from virtues of F-W theorem.

In the reminder of this section we present the concept of partial regression developed by Frisch, Waugh (1933). Let us consider a standard linear model

$$Y = X_1\,\beta_1 + X_2\,\beta_2 + \varepsilon, \quad \varepsilon \sim \mathrm{N}(0, \sigma^2\mathrm{I}_n),$$

where Y is a $n \times 1$ vector of observations, X_1 and X_2 are respectively $k_1 \times n$ and $k_2 \times n$ design matrices, $\theta = (\beta_1, \sigma^2)$ is the unknown parameter of interest and $\tau = \beta_2$ is a nuisance parameter. The partial regression technique allows us to find θ without actually estimating τ (c.f. Greene, 2008), Section 3.3). Let us denote $M_{X_2} = \mathrm{I}_n - X_2(X_2^{\mathrm{T}}X_2)^{-1}X_2^{\mathrm{T}}$. The partial regression estimator $\hat{\theta} = \left(\hat{\beta}, \widehat{\sigma^2}\right)$ is given by

$$\hat{\beta} = \left(X_1^{\mathrm{T}}M_{X_2}X_1\right)^{-1}M_{X_2}X_1^{\mathrm{T}}Y \tag{1}$$

and asymptotically unbiased[7]

$$\widehat{\sigma^2} = n^{-1}\left(M_{X_2}Y - M_{X_2}X_1\hat{\beta}\right)^{\mathrm{T}}\left(M_{X_2}Y - M_{X_2}X_1\hat{\beta}\right). \tag{2}$$

It turns out that $\hat{\beta}$ coincides with the corresponding element of slope estimator in the full Ordinary Least Squares scheme, i.e. $\hat{\beta} = \widehat{\beta_1}^{\mathrm{OLS}}$ with

$$\left[\widehat{\beta_1}^{\mathrm{OLS}}, \widehat{\beta_2}^{\mathrm{OLS}}\right]^{\mathrm{T}} = ([X_1\ X_2]^{\mathrm{T}}[X_1\ X_2])^{-1}[X_1\ X_2]^{\mathrm{T}}Y.$$

Moreover, the Frisch-Waugh theorem also states that

$$\widehat{\beta_2} = (X_2^{\mathrm{T}}X_2)^{-1}X_2^{\mathrm{T}}\left(Y - X_1\hat{\beta}\right) = \widehat{\beta_2}^{\mathrm{OLS}} \tag{3}$$

[7] Provided that $k_1 + k_2 = O(1)$.

and the variance of $\hat{\beta}$ can be obtained through the partitioned inverse[8] of the design moment matrix $[X_1\ X_2]^{\mathrm{T}}[X_1\ X_2]$, which is $\left(X_1^{\mathrm{T}} M_{X_2} X_1\right)^{-1}$. In the context of panel data model, by substituting a time or individual effect dummy variable for X_2 we obtain the well-known demeaning procedure.

3. DEMEANING IN ML ESTIMATION OF SPATIAL AUTOREGRESSIVE MODEL

In this section we introduce the Spatial Autoregressive Model specification and describe the well-known naive approach of demeaning in ML estimation, as used in e.g. Elhorst and Fréret (2008). Let us consider a standard spatial autoregressive linear model

$$Y = \rho \mathbf{W}\, Y + X_1\, \beta_1 + X_2\, \beta_2 + \varepsilon,\ \varepsilon \sim \mathrm{N}(0, \sigma^2 \mathrm{I}_n), \tag{4}$$

where $\mathbf{W}$ is an arbitrary spatial weight $n \times n$ matrix (with zero diagonal) and ρ is the scalar autoregressive parameter. Moreover, as previously, Y is a $n \times 1$ vector of observations, X_1 and X_2 are $k_1 \times n$, $k_2 \times n$ respectively design matrices, $\theta = (\rho, \beta_1, \sigma^2)$ is the unknown parameter of interest and $\tau = \beta_2$ is the nuisance parameter. The $\rho \mathbf{W} Y$ term is referred to as the spatial autoregressive term. The elements $\left(w_{ij}\right)_{ij \leq n}$ of $\mathbf{W}$ have the common interpretation of spatial weights, i.e. a measure of influence of j-th unit on unit i. Since $\rho \mathbf{W} Y = \rho\left(\sum_{j=1}^{n} w_{ij} Y_j\right)_{i \leq n}$, the spatial autoregressive term conveys information on weighted averages of influences from other spatial[9] units ($w_{ii} = 0$, $i \leq n$) on a given unit.

It is a well-known fact that the specification (4) cannot be estimated with the use of classical Ordinary Least Squares (see Anselin, 1988). Instead, the ML estimation procedure is a commonly suggested feasible alternative. To implement the maximum likelihood estimation procedure for the SAR specification it is enough to notice that, using the form of Gaussian density of ε, we can obtain the following formula for log likelihood function

$$\begin{aligned} &\log L(Y, \theta, \tau) = \\ &= -\frac{n}{2}\log(2\pi\sigma^2) + \log\det(\mathrm{I}_n - \rho\mathbf{W}) - \frac{1}{2\sigma^2}\|Y - \rho\mathbf{W}Y - X_1\beta_1 - X_2\beta_2\|^2, \end{aligned} \tag{5}$$

with the assumption that $\det(\mathrm{I}_n - \rho\mathbf{W})$ is positive for all ρ in its parameter space[10]. A straightforward implementation of the idea of partial regression consists in applying

[8] I.e. the relevant part of the inverse.

[9] Or spatio-temporal in dynamic panel case.

[10] It is a common practice to assume that the parameter space for spatial autoregressive parameter ρ is an interval $\mathcal{I} \subset \mathbb{R}$ such that $0 \in \mathcal{I}$. The endpoints of $\mathcal{I}$ are established from a condition ensuring invertibility of the spatial lag $\mathrm{I}_n - \rho\mathbf{W}$, for example $\|\mathbf{W}\| < 1$, for a matrix (submultiplicative) norm or

the demeaning operator M_{X_2} to the formula under the norm in (5). This, widely used, approach is supported by the fact that first order differential optimality condition on (5) is consistent with (3), i.e.

$$\tau_{\max} = \tau_{\max}(\rho, \beta, \sigma^2) = (X_2^{\mathrm{T}} X_2)^{-1} X_2^{\mathrm{T}} (Y - \rho \mathbf{W} Y - X_1 \beta)$$

and, as a result of simple algebra, we get the concentrated log likelihood

$$\begin{aligned} &\log L(Y, \theta, \tau_{\max}) = \\ &= -\frac{n}{2}\log(2\pi\sigma^2) + \log\det(\mathrm{I_n} - \rho\mathbf{W}) - \frac{1}{2\sigma^2}\left\|M_{X_2} Y - \rho M_{X_2}\mathbf{W}Y - M_{X_2}X_1\beta_1\right\|^2. \end{aligned} \tag{6}$$

Unfortunately, the operator M_{X_2} is not unitary, let alone invertible thus the formula above cannot be interpreted as a regular likelihood function. Nonetheless, the estimation approach of maximising (6) with respect to θ yields reasonably good estimates, provided that $k_2 = O(1)$, c.f. Elhorst (2009). However, the estimate of the asymptotic variance of the resulting ML estimator may be invalid. Moreover, the estimates of θ, when $\mathrm{limsup}_{\mathrm{n}\to\infty}\, k_1/n > 0$, are not consistent.

4. PARTIAL REGRESSION IN ML ESTIMATION OF SAR MODEL

In this section we derive our original approach to the problem of eliminating fixed effects in case of ML estimation of the SAR model by employing an alternative to the idea of partial regression. We will consider two cases. In Case I we assume that the spatial weight matrix is in some sense consistent with the nuisance slope parameter design X_2. In Case II an assumption about dimension of certain invariant subspace is made instead.

One approach to the issue of singularity of the M_{X_2} operator can be to pre-multiply ε by an orthogonal (i.e. transformation of coordinates) matrix $\boldsymbol{E}$ which maps the range of M_{X_2} onto $\mathbb{R}^{n-k_2}$ interpreted as a natural subset of $\mathbb{R}^n = \boldsymbol{E}(\mathbb{R}^n) = \mathbb{R}^{n-k_2} \oplus \mathbb{R}^{k_2}$, where $\oplus$ is the coordinate-wise direct sum of linear spaces. Then, we could integrate over unnecessary degrees of freedom, at the same time eliminating τ from the likelihood function (5). Effectively, this is the same as using the transformation $\pi\boldsymbol{E}M_{X_2}$, with $\pi = \pi_{n-k_2}$ being the natural projection $\mathbb{R}^n \xrightarrow{\pi} \mathbb{R}^{n-k_2}$ preserving $n - k_2$ first coordinates. Indeed, it can be observed that $(\pi\boldsymbol{E})^{\mathrm{T}}\pi\boldsymbol{E} = M_{X_2}$, thus $\pi\boldsymbol{E}M_{X_2} = \pi\boldsymbol{E}$.

Obviously, the transformations $\boldsymbol{E}$ and $\pi\boldsymbol{E}$ are not uniquely defined. In fact any such transformation $\boldsymbol{E}$, as described in previous paragraph can be equally useful. Here, we propose a method of construction of a possible candidate. Let $c_1, \ldots, c_n$ be rows of the matrix M_{X_2}. Let $i_1 = 1$. For $j = 2, \ldots, n - k_2$, once i_k, for $k < j$, have been defined,

---

$\max(\mathcal{J} \cup -\mathcal{J}) < 1/\lambda_{\mathfrak{R}}$, where $\lambda_{\mathfrak{R}}$ is a modulus-maximal real eigenvalue of $\mathbf{W}$ (c.f. Appendix in Olejnik, Özyurt, 2016).

we can set $i_j = \min\left\{k > i_{j-1} : M_{[c_1 \ldots c_{j-1}]} c_k \neq 0\right\}$. Having chosen vectors $c_{i_1}, \ldots, c_{i_{n-k_2}}$ as a basis of the range of M_{X_2}, we can proceed with Gram-Schmidt ortho-normalization process and thus obtain an orthonormal system of vectors $\tilde{c}_i$, for $1 \leq i \leq n - k_2$. We apply the same procedure for matrix $\mathrm{I} - M_{X_2}$ and obtain orthonormal system $\tilde{d}_i$, for $1 \leq i \leq k_2$. Finally, it is enough to set $\boldsymbol{E} = \left[\tilde{c}_1 \ldots \tilde{c}_{n-k_2} \tilde{d}_1 \ldots \tilde{d}_{k_2}\right]^{\mathrm{T}}$.

Case I

Let us assume that $\pi\boldsymbol{E}\mathbf{W} = \pi\boldsymbol{E}\mathbf{W}M_{X_2}$. Since $\varepsilon \sim \mathrm{N}(0, \sigma^2\mathrm{I}_n)$, we can immediately conclude that $\boldsymbol{E}\varepsilon \sim \mathrm{N}(0, \sigma^2\mathrm{I}_n)$, so that $\pi\boldsymbol{E}\varepsilon \sim \mathrm{N}\left(0, \sigma^2\mathrm{I}_{n-k_2}\right)$. Notice that, by denoting $\mathbf{W}_{\pi\boldsymbol{E}} = \pi\boldsymbol{E}\mathbf{W}\boldsymbol{E}^{\mathrm{T}}\pi^{\mathrm{T}}$ and observing that $\pi\pi^{\mathrm{T}} = \mathrm{I}_{n-k_2}$, we have

$$\pi\boldsymbol{E}\varepsilon = \pi\boldsymbol{E}Y - \rho\pi\boldsymbol{E}\mathbf{W}Y - \pi\boldsymbol{E}\boldsymbol{X}_1\beta_1 - \pi\boldsymbol{E}X_2\beta_2 =$$

$$= \pi\boldsymbol{E}Y - \rho\pi\boldsymbol{E}\mathbf{W}M_{X_2}Y - \boldsymbol{\pi E X}_1\beta_1 = \left(\mathrm{I}_{n-k_2} - \rho\,\mathbf{W}_{\pi\boldsymbol{E}}\right)\pi\boldsymbol{E}Y + \boldsymbol{\pi E X}_1\beta_1.$$

It can be noticed that $\left(\mathrm{I}_{n-k_2} - \rho\mathbf{W}_{\pi\boldsymbol{E}}\right) = \pi\boldsymbol{E}(\mathrm{I}_n - \rho\mathbf{W})\boldsymbol{E}^{\mathrm{T}}\pi^{\mathrm{T}}$ is invertible if $(\mathrm{I}_n - \rho\mathbf{W})$ is invertible. Indeed, it is enough either to observe that, by a simple algebra, we have $\left(\mathrm{I}_{n-k_2} - \rho \cdot \mathbf{W}_{\pi\boldsymbol{E}}\right)\pi\boldsymbol{E}(\mathrm{I}_n - \rho \cdot \mathbf{W})^{-1}\boldsymbol{E}^{\mathrm{T}}\pi^{\mathrm{T}} = \mathrm{I}_{n-k_2}$, since $\boldsymbol{E}\mathbf{W}(\mathrm{I}_n - (\pi\boldsymbol{E})^{\mathrm{T}}\pi\boldsymbol{E}) = \mathbf{0}$. Thus, for each value of ρ in its parameter space we can properly define the transformation

$$T(\epsilon) = \left(\mathrm{I}_{n-k_2} - \rho\mathbf{W}_{\pi\boldsymbol{E}}\right)^{-1}\epsilon - \pi\boldsymbol{E}X_1\beta_1, \text{ for } \epsilon \in \mathbb{R}^{n-k_2}.$$

Since $\pi\boldsymbol{E}Y = T(\varepsilon)$ and $\frac{\partial}{\partial\epsilon}T^{-1} = \mathrm{I}_{n-k_2} - \rho\mathbf{W}_{\pi\boldsymbol{E}}$ we obtain the following form of logarithm of likelihood function for θ, based on observable values of $\pi\boldsymbol{E}Y$

$$\log L(\pi\boldsymbol{E}Y, \theta) = -\frac{n-k_2}{2}\log(2\pi\sigma^2) + \\ + \log\det\left(\mathrm{I}_{n-k_2} - \rho\mathbf{W}_{\pi\boldsymbol{E}}\right) - \frac{1}{2\sigma^2}\|\pi\boldsymbol{E}Y - \rho\mathbf{W}_{\pi\boldsymbol{E}}\pi\boldsymbol{E}Y - \pi\boldsymbol{E}X_1\beta_1\|^2. \tag{7}$$

Now, we can differentiate $\log L(\pi\boldsymbol{E}Y, \theta)$ with respect to (β, σ^2) and equate the result to zero, thus get the optimal relations between β, σ^2 and ρ

$$\beta_1 = (X_1E^T\pi^T\pi EX_1)^{-1}X_1E^T\pi^T(\pi\boldsymbol{E}Y - \rho\mathbf{W}_{\pi\boldsymbol{E}}\pi\boldsymbol{E}Y), \tag{8}$$

$$\sigma^2 = \frac{1}{n-k_2}\|\pi\boldsymbol{E}Y - \rho\mathbf{W}_{\pi\boldsymbol{E}}\pi\boldsymbol{E}Y - \pi\boldsymbol{E}X_1\beta_1\|^2. \tag{9}$$

Relations (8) and (9) are SAR counterparts of partial regression estimators (1), (2). In order to obtain the ML estimates β_1 and σ^2 we need to evaluate the above formulas a the maximum likelihood estimate of ρ.

Substituting the above equations, (8) and (9), into (7) we get the concentrated log likelihood function $\log L_{\mathrm{Conc}}$

$$\log L_{\mathrm{Conc}}(\pi \boldsymbol{E}Y, \rho) = -\tfrac{n-k_2}{2}\log(\boldsymbol{a}\rho^2 + \boldsymbol{b}\rho + \boldsymbol{c}) + \log\det\left(\mathrm{I}_{n-k_2} - \rho \mathbf{W}_{\pi \boldsymbol{E}}\right) + \mathrm{Const},$$

where the coefficients of the quadratic polynomial in ρ are given by

$$\boldsymbol{a} = \|\mathbb{e}(\pi \boldsymbol{E}Y, \pi \boldsymbol{E}X_1)\|^2,$$

$$\boldsymbol{b} = 2\mathbb{e}(\pi \boldsymbol{E}Y, \pi \boldsymbol{E}X_1)^{\mathrm{T}}\mathbb{e}(\mathbf{W}_{\pi \boldsymbol{E}}\pi \boldsymbol{E}Y, \pi \boldsymbol{E}X_1),$$

$$\boldsymbol{c} = \|\mathbb{e}(\mathbf{W}_{\pi \boldsymbol{E}}\pi \boldsymbol{E}Y, \pi \boldsymbol{E}X_1)\|^2,$$

and $\mathbb{e}(V_1, V_2)$ is the column vector of OLS residuals obtained by regressing V_1 on V_2. Lastly, it is clear that simply maximising (presumably numerically) $\log L_{\mathrm{Conc}}(\pi \boldsymbol{E}Y, \rho)$ with respect to its single parameter ρ gives the desired value of $\rho_{\max}$, which can be further substituted into (8) and (9).

Case II

Now, let us assume that $\pi \boldsymbol{E}\mathbf{W} - \pi \boldsymbol{E}\mathbf{W}M_{X_2} \neq 0$. The approach we present further is based on the concept of Krylov subspace. The Krylov subspace for X_2 with respect to $\mathbf{W}$ is the minimal $\mathbf{W}$-invariant subspace containing X_2. We will denote it by H, i.e.

$$H = \mathrm{span}(\mathbf{W}^k X_2 \beta : k = 0,1, \dots, n \wedge \beta \in \mathbb{R}^{k_2}).$$

We will assume that H is a proper subspace of $\mathbb{R}^n$, so that $n_* = n - k_* > 0$, with $k_* = \dim H$.

With the notation of M_H being orthogonal projection on orthogonal complement of H, as previously, we define linear isometry $\boldsymbol{F}$ to be an operator that takes range of M_H onto $\mathbb{R}^{n_*} \backsimeq \boldsymbol{F}(H^{\perp})$ with coordinate-wise $\mathbb{R}^n = \boldsymbol{F}(H^{\perp}) \oplus \mathbb{R}^{k_*}$. Furthermore, let $\pi_* : \mathbb{R}^n \to \mathbb{R}^{n_*}$ be orthogonal projection preserving first n_* coordinates. We have $\pi_* \boldsymbol{F} = \pi_* \boldsymbol{F} M_H$ and $\pi_* \boldsymbol{F}\mathbf{W}(\mathrm{I} - M_H) = \pi_* \boldsymbol{F}(\mathrm{I} - M_H) = 0$.

Again, considering the fact that $\varepsilon \sim \mathrm{N}(0, \sigma^2 \mathrm{I}_n)$, we can immediately conclude that $\boldsymbol{F}\varepsilon \sim \mathrm{N}(0, \sigma^2 \mathrm{I}_n)$, and further that $\pi_* \boldsymbol{F}\varepsilon \sim \mathrm{N}\left(0, \sigma^2 \mathrm{I}_{n_*}\right)$. Notice that, denoting $\mathbf{W}_{\pi_* \boldsymbol{F}} = \pi_* \boldsymbol{F}\mathbf{W}\boldsymbol{F}^{\mathrm{T}}\pi_*^{\mathrm{T}}$ and observing that $\pi_* \pi_*^{\mathrm{T}} = \mathrm{I}_{n_*}$ we have

$$\pi_*\boldsymbol{F}\varepsilon = \pi_*\boldsymbol{F}Y - \rho\pi_*\boldsymbol{F}\mathbf{W}Y - \pi_*\boldsymbol{F}X_1\beta_1 - \pi_*\boldsymbol{F}X_2\beta_2 =$$

$$= \pi_*\boldsymbol{F}Y - \rho\pi_*\boldsymbol{F}\mathbf{W}(M_H Y + (\mathrm{I} - M_H)Y) - \pi_*\boldsymbol{F}X_1\beta_1 =$$

$$= \left(\mathrm{I}_{n_*} - \rho\mathbf{W}_{\pi_*\boldsymbol{F}}\right)\pi_*\boldsymbol{E}Y - \pi_*\boldsymbol{F}\mathbf{W}(\mathrm{I} - M_H)Y + \pi_*\boldsymbol{F}X_1\beta_1 =$$

$$= \left(\mathrm{I}_{n_*} - \rho\mathbf{W}_{\pi_*\boldsymbol{F}}\right)\pi_*\boldsymbol{E}Y + \pi_*\boldsymbol{F}X_1\beta_1.$$

It can be also noticed that $\left(\mathrm{I}_{n_*} - \rho\mathbf{W}_{\pi_*\boldsymbol{F}}\right) = \pi_*\boldsymbol{F}(\mathrm{I}_n - \rho\mathbf{W})\boldsymbol{F}^{\mathrm{T}}\pi_*^{\mathrm{T}}$ is invertible if $(\mathrm{I}_n - \rho\mathbf{W})$ is invertible, since $\left(\mathrm{I}_{n_*} - \rho\mathbf{W}_{\pi_*\boldsymbol{F}}\right)\pi_*\boldsymbol{F}(\mathrm{I}_n - \rho\mathbf{W})^{-1}\boldsymbol{F}^{\mathrm{T}}\pi_*^{\mathrm{T}} = \mathrm{I}_{n_*}$, since $\boldsymbol{F}\mathbf{W}(\mathrm{I}_n - M_H) = \mathbf{0}$.

As a result, for each value of ρ in its parameter space we can properly define the transformation

$$T(\epsilon) = \left(\mathrm{I}_{n_*} - \rho\mathbf{W}_{\pi_*\boldsymbol{F}}\right)^{-1}\epsilon - \pi_*\boldsymbol{F}X_1\beta_1, \text{ for } \epsilon \in \mathbb{R}^{n_*}.$$

Since $\pi_*\boldsymbol{F}Y = T(\varepsilon)$ and $\frac{\partial}{\partial\epsilon}T^{-1} = \mathrm{I}_{n_*} - \rho\mathbf{W}_{\pi_*\boldsymbol{F}}$ we obtain the following form of logarithm of likelihood function for θ based on observable values of $\pi_*\boldsymbol{F}Y$

$$\log L(\pi\boldsymbol{E}Y, \theta) =$$

$$= -\frac{n_*}{2}\log(2\pi\sigma^2) + \log\det\left(\mathrm{I}_{n_*} - \rho\mathbf{W}_{\pi_*\boldsymbol{F}}\right) - \frac{1}{2\sigma^2}\left\|\pi_*\boldsymbol{F}Y - \rho\mathbf{W}_{\pi_*\boldsymbol{F}}\pi_*\boldsymbol{F}Y - \pi_*\boldsymbol{F}X_1\beta_1\right\|^2. \tag{10}$$

Now, we can differentiate with respect to (β, σ^2) and equate to zero to get optimal relations between β, σ^2 and ρ

$$\beta_1 = (X_1\boldsymbol{F}^T\pi_*^T\pi_*\boldsymbol{F}X_1)^{-1}X_1\boldsymbol{F}^T\pi_*^T\left(\pi_*\boldsymbol{F}Y - \rho\mathbf{W}_{\pi_*\boldsymbol{F}}\pi_*\boldsymbol{F}Y\right), \tag{11}$$

$$\sigma^2 = \frac{1}{n_*}\left\|\pi_*\boldsymbol{F}Y - \rho\mathbf{W}_{\pi_*\boldsymbol{F}}\pi_*\boldsymbol{F}Y - \pi_*\boldsymbol{F}X_1\beta_1\right\|^2, \tag{12}$$

which are SAR counterparts of partial regression estimators (1), (2). In order to be able to evaluate the above formulas we need to obtain maximum likelihood estimate of ρ.

Substituting the above equations, (11) and (12), to (10) we get the concentrated likelihood function L_{Conc}

$$\log L_{\mathrm{Conc}}(\pi\boldsymbol{E}Y, \rho) = -\frac{n_*}{2}\log(\boldsymbol{a}\rho^2 + \boldsymbol{b}\rho + \boldsymbol{c}) + \log\det\left(\mathrm{I}_{n_*} - \rho\mathbf{W}_{\pi_*\boldsymbol{F}}\right) + \mathrm{Const},$$

where the coefficients of quadratic polynomial in ρ are given by

$$\boldsymbol{a} = \|\mathbb{e}(\pi_* \boldsymbol{F} Y, \pi_* \boldsymbol{F} X_1)\|^2,$$

$$\boldsymbol{b} = 2\mathbb{e}(\pi_* \boldsymbol{F} Y, \pi_* \boldsymbol{F} X_1)^{\mathrm{T}} \mathbb{e}\left(\mathbf{W}_{\pi_* \boldsymbol{F}} \pi_* \boldsymbol{F} Y, \pi_* \boldsymbol{F} X_1\right),$$

$$\boldsymbol{c} = \left\|\mathbb{e}\left(\mathbf{W}_{\pi_* \boldsymbol{F}} \pi_* \boldsymbol{F} Y, \pi_* \boldsymbol{F} X_1\right)\right\|^2$$

and $\mathbb{e}(V_1, V_2)$ is the column vector of OLS residuals obtained by regressing V_1 on V_2. By maximising $\log L_{\mathrm{Conc}}(\pi_* \boldsymbol{F} Y, \rho)$ with respect to its single parameter ρ we obtain the desired value of $\rho_{\max}$, which can be further substituted into (11) and (12).

Finally, let us notice that Case II simplifies to Case I if $M_H = M_{X_2}$ thus the remaining part of the paper considers Case II only. Still, in this section, we have retained the presentation of both cases separately since Case I is considerably simpler in application and should be used instead of Case II whenever possible.

5. ASYMPTOTICS OF THE PARAMETER ESTIMATES

In this section we formulate two statements on asymptotic behaviour of the ML estimator presented in section 3. First of those statements concerns consistency, second concerns limiting variance of the estimates from our ML estimator.

Large sample theory for maximum likelihood estimation establishes, under some assumptions, two important facts about the ML estimator $\hat{\theta} = \hat{\theta}_n$. Firstly, it is the consistency of ML estimates and secondly the limiting distribution for the quantity $\sqrt{n}\left(\hat{\theta} - \theta_0\right)$, where θ_0 is the true parameter value. Clearly, in the case of the SAR model, given by equation (4), the observed sample $Y = (Y_1, \ldots, Y_n)^{\mathrm{T}}$ is not independent thus the classical textbook results are not applicable. Nonetheless, it has been a commonplace since the early days of applied spatial econometrics (c.f. Anselin, 1988) to assume that $\hat{\theta}$ is consistent and the its deviation $\sqrt{n}\left(\hat{\theta} - \theta_0\right)$ is asymptotically normal with zero mean and variance $\left[-\frac{1}{n}\mathbb{E}\frac{\partial^2}{\partial \theta^2}\log L(Y, \theta)\right]^{-1}$. This popular belief was supported by the fact that any sensible asymptotic theory (covering at least the increasing domain scheme) would definitely have to give asymptotics of the form mentioned. This is because, such a theory would have to include the simple asymptotic setting, in which there exists a sequence of parallel spatial domains, independent and unrelated to one another, being included in the sample as n increases. Notice that this simple setting is subject to vector-valued independent sample ML asymptotics theorem. If one further assures identifiability and uniqueness of the maximiser, the above-mentioned asymptotics follow.

With the papers of Kelejian and Prucha (2001) as well as Lee (2004) it became apparent that an asymptotic theory covering more sophisticated asymptotic settings is possible. Using general tools for consistency and asymptotic normality proofs (described in e.g. Pötscher, Prucha, 1997) one can construct asymptotic theory for ML estimates covering both infill and increasing domain schemes. The crucial assumptions

that have to be made concern the spatial weight $\mathbf{W}$ and the design matrix X. From those assumptions identifiable uniqueness of parameters and a certain uniform law of large numbers for the log likelihood function can be deduced. Those two elements allow one to utilize the theory of general M-estimators from Pötscher, Prucha (1997) to obtain the desired results.

Below we describe (after Lee, Yu, 2010) a set of possible assumptions which assure fairly general statement about asymptotics of ML estimates. Apart from the natural postulates of the zero diagonal of spatial weight matrix $\mathbf{W} = \mathbf{W}(n)$ we mention the following.

Assumption 0. The error term ε in (4) follows multivariate normal distribution with uncorrelated, homoscedastic components. In particular, this implies that all moments of ε are finite.

Assumption 1[11]. For elements ρ of its parameter space[12] $\mathcal{R}$ the spatial lag operator $\mathrm{I}_n - \rho\mathbf{W}$ is invertible and the true value of ρ_0 is an interior point of $\mathcal{R}$.

Assumption 2[13]. There exists a constant C such that for any rows or columns, say v, of any of the matrices $\mathbf{W}$, $\mathbf{W}_{\pi_* F}$ and $(\mathrm{I}_n - \rho\mathbf{W})^{-1}$, $\left(\mathrm{I}_{n_*} - \rho\mathbf{W}_{\pi_* F}\right)^{-1}$, $\rho \in \mathcal{R}$, $n \in \mathbb{N}$, its ℓ_1-norm[14] $\|v\|_1$ does not exceeds C.

Assumption 3[15]. The elements of non-stochastic design matrix $X = X(n)$ are bounded and the sequence $\frac{1}{n}X^T M_H X$ converges to a non-singular limit.

Assumption 4. The ratio n_*/n converges as $n \to \infty$ and $\alpha_* = \liminf_{n\to\infty} \frac{n_*}{n} > 0$.

Assumption 5[16]. Estimated parameters are uniquely identified[17].

Let us note that the natural Assumption 1 is crucial not only for identifiability of the parameter ρ but it is also necessary for our ability to present a closed form of Y from (4), thus for effective interpretation of the model. Assumption 2 limits spatial dependence to 'manageable degree'. This means, in particular, that the amount of information obtained from a larger sample is sufficient to decrease variance of estimates. Assumption 3 assures that the design matrix is well-behaved and in particular through non-singularity of the corresponding limit conveys sufficient information on the slope parameters of interest.

Assumption 4 guarantees that the dimension of appropriate Krylov space does not reduce the number of available degrees of freedom excessively. Assumption 5 assures that the hypothetical probability distributions for different parameter values remain clearly distinguishable by ML estimation procedure as sample increases. This assumption is typically expanded into a highly technical statement involving terms from log likelihood function, so that it implies unique identification. In our paper, to

[11] C.f. Lee, Yu, page 167, Assumption 3 therein.

[12] See footnote 10.

[13] C.f. Lee, Yu, page 167, Assumption 5 therein.

[14] For either a column or row vector $v = (v_1, \ldots, v_n)$ its ℓ_1-norm is $\|v\|_1 = \sum_{i=1}^n |v_i|$.

[15] C.f. Lee, Yu, page 167, Assumption 4 therein.

[16] C.f. Lee, Yu, page 168, Assumption 7 therein.

[17] In the sense of Definition 3.1 in Pötscher, Prucha (1997).

avoid unnecessary complexity we decide to readably assume the unique identification of parameters.

Lastly, for completeness of presentation, we conclude with the asymptotic distribution of $\hat{\theta}$. Namely, under the assumptions 1–5 we can state that $\sqrt{n_*}\left(\hat{\theta} - \theta_0\right)$ converges in distribution to $\mathrm{N}(0, n_* I(\theta_0)^{-1})$. More precisely, conducting differentiation and applying expectation in the score matrix we obtain a formula for the Fisher information $I(\theta, n)$ for parameter $\theta = (\beta, \rho, \sigma^2)$. Namely, denoting $\bar{\mathbf{W}} = \mathbf{W}(\mathrm{I}_n - \rho\mathbf{W})^{-1}$ and $\bar{\mathbf{W}}_* = \mathbf{W}_{\pi_* \boldsymbol{F}}\left(\mathrm{I}_{n_*} - \rho\mathbf{W}_{\pi_* \boldsymbol{F}}\right)^{-1}$ we obtain[18]

$$I(\theta, n) = \begin{bmatrix} \sigma^{-2} X_1^T M_H X_1 & I_{\beta,\rho}(\theta) & 0 \\ I_{\beta,\rho}(\theta)^{\mathrm{T}} & I_{\rho,\rho}(\theta) & \sigma^{-2}\mathrm{trace}(\bar{\mathbf{W}}_*) \\ 0 & \sigma^{-2}\mathrm{trace}(\bar{\mathbf{W}}_*) & \dfrac{n_*}{2}\sigma^{-4} \end{bmatrix},$$

$$I_{\beta,\rho}(\theta) = \sigma^{-2} X_1^{\mathrm{T}} M_H \bar{\mathbf{W}} M_H X_1 \beta,$$

$$I_{\rho,\rho}(\theta) = \mathrm{trace}\left(\bar{\mathbf{W}}_* \bar{\mathbf{W}}_* + \bar{\mathbf{W}}_*{}^{\mathrm{T}} \bar{\mathbf{W}}_*\right) + \beta^{\mathrm{T}} X_1^{\mathrm{T}} M_H \bar{\mathbf{W}}^{\mathrm{T}} M_H \bar{\mathbf{W}} M_H X_1 \beta.$$

Setting $\Sigma = \lim_{n\to\infty} n_* I(\theta_0, n)^{-1}$, we have convergence in distribution of $\sqrt{n_*}\left(\hat{\theta} - \theta_0\right)$ to $\mathrm{N}(0, \Sigma)$ provided that the limit Σ exists.

6. DISCUSSION OF THE ADOPTED ASSUMPTION

The Assumptions 1–3 and 5 are well known in spatial econometric literature. Extensive discussion on the topic has been given in numerous papers e.g. Kelejian, Prucha (2001), Lee (2004), Lee, Yu (2010). The new assumption introduced in this paper is the Assumption 4, which connects n – the increasing sample size, with the amount of degrees of freedom lost due to the use of the generalized demeaning procedure. In terms of a standard fixed effects setting, where $n = N \cdot T$, Assumption 4 is equivalent to requiring that $T \to \infty$, when N spatial fixed effects are present, and requiring that $N \to \infty$, whenever T temporal fixed effects are included in the model.

Obviously, in a fully general case it cannot be guaranteed that the requirement in Assumption 4 is satisfied. Then, a natural question arises: is Assumption 4 often met in practice? It turns out that some "rules of thumb" can be formulated which imply affirmative answer in many practical settings. We will present them in the following examples as well as in an empirical illustration described in next section. For simplicity, we consider the case of $n = N \cdot T$, that is a standard balanced panel

[18] C.f. Elhorst (2014), page 46.

data set. Moreover, the spatial $n \times n$ weight matrix is purely spatial, i.e. it does not contain any dynamic references.

For arbitrary $m \in \mathbb{N}$ let us denote $\mathbf{1}_m = (1, \ldots, 1)^{\mathrm{T}} \in \mathbb{R}^m$. If the spatial weight matrix $\mathbf{W}$ is constant in time, i.e. $\mathbf{W} = \overline{\mathbf{W}} \otimes \mathrm{I}_N$, and the fixed effect design X_2 is constant in time (i.e. each column of X_2 is of the form $v \otimes \mathbf{1}_T$, for some $v \in \mathbb{R}^N$) then $\mathbf{W}X_2$ is also time-constant. By induction, we infer that the Krylov space H for X_2 contains only time-constant vectors. Finally, $\dim H \leq N$, thus $\frac{n_*}{n} = 1 - \frac{\dim H}{NT}$, which converges to 1 as $T \to \infty$.

Another example is when, as it is very often found in practice, $\mathbf{W}$ is row-standardized, and when the matrix X_2 contains purely temporal effect of arbitrary shape. That is, each column of X_2 is of the form $\mathbf{1}_N \otimes v$, for some $v \in \mathbb{R}^T$, then $\mathbf{W}(\mathbf{1}_N \otimes v) = \mathbf{1}_N \otimes v$. This implies that $\dim H = k_2$, thus $\alpha_* = 1$. Even if k_2 is not bounded, Assumption 4 is satisfied if $N \to \infty$, since obviously $k_2 \leq T$.

7. AN EMPIRICAL ILLUSTRATION

The background for this empirical illustration is a theoretical model developed by Fingleton (2001, 2004), which is based on the NEG theory and Verdoorn's law (c.f. Verdoorn, 1949). This law links the increase in labour productivity with an increase in production. More precisely, the Verdoorn's law states that in a long run productivity grows proportionally to the square root of output. According to e.g. Fingleton (2001), the exponential growth rate of productivity can be modelled by the use of the following specification

$$p = \alpha_0 + \rho \mathbf{W} p + \alpha_1 H + \alpha_2 G_0 + \alpha_3 q + \varepsilon, \; \varepsilon \sim N(0, \sigma^2)$$

where: p represents the exponential growth rate of productivity, $\mathbf{W}$ is a spatial weight matrix, H refers to human capital, G_0 is the initial level of technology, and q is the exponential growth rate. As described in Olejnik and Olejnik (2017) the specification can be further transformed into the following Spatial Panel Durbin Model

$$p = \rho \mathbf{W} p + \pi_1 q + \pi_2 \mathbf{W} q + \eta_1 H + \mathbf{FE} + \varepsilon, \tag{13}$$

with α_0, ρ, π_1, π_2, η_1 being model parameters, $\mathbf{W}$ is spatial weight matrix. The term G_0 does not appear in (13) as they have been incorporated into fixed effects $\mathbf{FE}$. In our example the fixed effects are $2N$ dummy variables of the form $e_i \otimes v_{2004}$ and $e_i \otimes v_{2008}$, $i = 1, \ldots, N$, where v_{2004} and v_{2008} are fixed effects distinguishing periods after EU enlargement and global financial crises in 2008, respectively. Notice that the groups of observations distinguished by this fixed effect design are overlapping, thus the standard demeaning procedure cannot be used. Moreover, the additional term of spatially lagged exogenous variable q has been introduced into (13) to account for additional externalities.

The data for the example covers 261 regions of EU for the years 2000–2013. The productivity growth p for the years 2001–2013 is approximated by the exponential rate of change of regional productivity (quotient of regional production over the number of economically active population) related to regional productivity in the initial year 2000. Similarly, the exponential growth rate is approximated by logarithm of the ratio of regional production in years 2001–2013 to the base year 2000. The matrix $\mathbf{W}$ is a row-standardised spatial weight matrix of three nearest neighbours (c.f. Anselin, 1988). The human capital H is approximated by employment in technology and knowledge-intensive sectors expressed as a percentage of economically active population, expressed in logarithms.

For the purpose of empirical comparison we apply both standard Maximum Likelihood estimation procedure using dummy variables and our modified approach. Results are presented in table 1.

Table 1.

Comparison of standard ML and the augmented ML approach

Parameter	Corresponding variable	Standard ML		New ML	
		Coeff.	t-stat	Coeff.	t-stat
ρ	$\mathbf{W}p$	0.64	49.44	0.64	45.48
π_1	q	0.74	56.50	0.74	51.97
π_2	$\mathbf{W}q$	−0.45	25.66	−0.45	23.61
η_1	H	0.09	9.69	0.09	8.92
σ^2	Error variance	1.0542		1.2459	
$\overline{R^2}$	Goodness of fit	0.9517		0.9429	

Source: own calculation.

Table 1 shows that both estimation procedures yield virtually the same values for both autoregressive and regressive parameters (ρ and β_1 respectively in notation from previous sections). However, there is a considerable difference in estimates of the σ^2 parameter. As expected, our procedure yields a consistent estimates of the error variance, which turns out to be rather cautious. This is because the standard ML estimate does not properly reflect the loss of degrees of freedom related to the use of general fixed effect dummies. In contrast our estimation scheme, through consideration of the dimension of Krylov space for fixed effects design matrix, allows one to estimate σ^2 and also goodness-of-fit measures more reliably. Moreover, if the size of fixed effect design grows with sample size (e.g. in our example $2N$ might grow with n), then the standard ML estimate of σ^2 might even turn out to be inconsistent.

8. SUMMARY

Since the early days of spatial econometrics it has been known that ordinary least squares procedure for estimating model parameters in the case of spatial autoregressive specification leads to inconsistent estimates. This is because, the specification incorporates the lagged dependant variable term as one of the regressors. Maximal likelihood procedure has been long considered a remedy for this endogeneity problem. Although, new alternatives to ML have been found (e.g. generalized method of moments) the original procedure of maximal likelihood remains widely used by practitioners.

In this paper we have proposed an alternative to partial regression in a spatial autoregressive econometric model when the maximum likelihood procedure is used. Under certain assumptions on the dimension of some invariant space associated with spatial weight matrix we have managed to formulate a feasible procedure, which can be used to handle large class of fixed effect designs. This can be done at the expense of possibly decreased number of degrees of freedom in Gaussian log likelihood function.

Our result contradicts the conjecture, expressed in a celebrated paper by Lee, Yu (2010) on bias correction in the case of incidental parameter problem, that such scheme would not be possible except cases of individual fixed effects and time fixed effects with row standardized spatial weight matrix. As in our reasoning we carefully manage the degrees of freedom at the step of sample transformation (demeaning), estimator bias in the case of incidental parameter problem does not occur in our setting, cf. Elhorst (2014).

REFERENCES

Anselin L., (1988), *Spatial Econometrics: Methods and Models*, Kluwer Academic Publishers, Dordrecht, The Netherlands.

Anselin L., Le Gallo J., Jayet H., (2006), *Spatial Panel Econometrics*, in: Matyas L., Sevestre P., (eds.), *The Econometrics of Panel Data, Fundamentals and Recent Developments in Theory and Practice*, 3rd edition, Kluwer, Dordrecht.

Baltagi B. H., (2005), *Econometric Analysis of Panel Data*, John Wiley & Sons, Chichester West Sussex, England.

Elhorst J. P., (2014), *Spatial Econometrics: From Cross-Sectional Data to Spatial Panels*, Springer, Heidelberg.

Elhorst J. P., Fréret S., (2009), Evidence of Political Yardstick Competition in France Using a Two-Regime Spatial Durbin Model with Fixed Effects, *Journal of Regional Science*, 49 (5), 931–951.

Fiebig D. G., Bartels R., (1996), The Frisch-Waugh Theorem and Generalized Least Squares, *Econometric Reviews*, 15 (4), 431–443.

Frisch R., Waugh F. V., (1933), Partial Time Regressions as Compared with Individual Trends, *Econometrica*, 1 (4), 387–401.

Greene W. H., (2008), *Econometric Analysis*, 6th edition, Prentice Hall, Upper Saddle River, N. J.

Kelejian H. H., Prucha I. R., (2001), On the Asymptotic Distribution of the Moran I Test Statistic with Applications, *Journal of Econometrics*, 104, 219–257.

Lee L.-F., (2004), Asymptotic Distributions of Quasi-Maximum Likelihood Estimators for Spatial Autoregressive Models, *Econometrica*, 72 (6), 1899–1925.

Lee L.-F., Yu J., (2010), Estimation of Spatial Autoregressive Panel Data Models with Fixed Effects, *Journal of Econometrics*, 154 (2), 165–85.

Liesen J., Strakoš Z., (2013), *Krylov Subspace Methods: Principles and Analysis*, Oxford University Press, Oxford UK.

Olejnik A., Olejnik J., (2017), Increasing Returns to Scale, Productivity and Economic Growth – A Spatial Analysis of the Contemporary EU Economy, submitted to *Argumenta Oeconomica*.

Olejnik A., Özyurt S., (2016), Multi-Dimensional Spatial Auto-regressive Models: How Do They Perform in an Economic Growth Framework?, working paper.

Pace R. K., (2014), *Maximum Likelihood Estimation*, in: Fisher M. M., Nijkamp P., (eds.), *Handbook of Regional Science*, Springer-Verlag, Berlin Heidelberg.

Pötscher B. M., Prucha I., (1997), *Dynamic Nonlinear Econometric Models*, Springer-Verlag, Berlin Heidelberg.

Yamada H., (2016), The Frisch–Waugh–Lovell Theorem for the Lasso and the Ridge Regression, *Communications in Statistics – Theory and Methods*, Available online at: http://dx.doi.org/10.1080/03610926.2016.1252403.

PROCEDURA ALTERNATYWNA DO REGRESJI CZĘŚCIOWEJ W ESTYMACJI MODELU PRZESTRZENNEGO AUTOREGRESYJNEGO METODĄ NAJWIĘKSZEJ WIAROGODNOŚCI

Streszczenie

W niniejszej pracy wprowadzono procedurę alternatywną do procedury regresji częściowej. Opisane postępowanie może być zastosowane w przypadku estymowania parametrów modelu przestrzennego autoregresyjnego metodą największej wiarogodności. Przy pewnych założeniach dotyczących wymiaru pewnej przestrzeni niezmienniczej związanej z macierzą wag przestrzennych sformułowany jest schemat postępowania obejmujący szeroką klasę macierzy efektów stałych. W pewnych przypadkach opisana procedura może eliminować efekty stałe kosztem obniżonej liczby stopni swobody. Dodatkowo, przedstawiono własności asymptotyczne zaprezentowanego estymatora.

Słowa kluczowe: regresja częściowa, metoda największej wiarogodności, model przestrzenny autoregresyjny, model z efektami stałymi

AN ALTERNATIVE TO PARTIAL REGRESSION IN MAXIMUM LIKELIHOOD ESTIMATION OF SPATIAL AUTOREGRESSIVE MODEL

Abstract

In this paper an alternative procedure to partial regression is introduced. The presented procedure can be used in maximum likelihood estimation of spatial autoregressive model. Under certain assumptions on the dimension of certain invariant space associated with spatial weight matrix a feasible procedure is formulated, which can be used to handle large class of fixed effect designs. This is done at the expense of possibly decreased number of degrees of freedom in the Gaussian log likelihood function. Additionally, a statement on asymptotic behaviour of presented estimator is given.

Keywords: partial regression, maximum likelihood estimation, spatial autoregressive model, fixed effects model