Available online at www.sciencedirect.com
SCIENCE DIRECT®
Economics Letters 82 (2004) 1–7
economics letters
ELSEVIER
www.elsevier.com/locate/econbase

First-order serial correlation in seemingly unrelated regressions

Bertrand M. Koebel*

Faculty of Economics and Management, Otto-von-Guericke University, P.O. Box 4120, D-39016 Magdeburg, Germany

Abstract

This note shows that the positive definiteness of the variance matrix of seemingly unrelated residuals with first-order serial correlation implies that the roots of the serial correlation matrix have a modulus smaller than 1.

Keywords: System of regressions; Panel data

JEL classification: C30

1. Motivation

This note considers the estimation of seemingly unrelated regressions (SUR) with first-order serial correlation in the residuals. The main objective is to show that the positive definiteness of the residual covariance matrix implies that the roots of the serial correlation matrix have a modulus smaller than 1, which in turns implies that the error process is stationary. A related result was obtained by Beach and MacKinnon (1979), in the context of maximum likelihood estimation, for serial correlation matrices with real and nonrepeated characteristic roots. This note generalizes Beach and MacKinnon's finding to arbitrary correlation matrices. The occurrence of complex characteristic roots is quite frequent for nonsymmetric matrices and is also found in empirical investigations (see Koebel, 2003).

There already exist procedures to deal with first-order serial correlation in SUR (see Greene, 2003, Chapter 14 for an overview), but most of them lead to a loss of the first period observation(s). Guilkey and Schmidt (1973) and Judge et al. (1985, Chapter 12.3) develop a method for avoiding this loss. A byproduct of this note is to propose a simpler alternative. This method is interesting for time series and panel data with a short-time dimension in which case loosing observations is costly.

* Tel.: +49-391-671-8431; fax: +49-391-671-1218.
E-mail address: koebel@ww.uni-magdeburg.de (B.M. Koebel).

0165-1765/$ - see front matter
doi:10.1016/j.econlet.2003.10.001

2. Notations and result

Consider the system of J regressions:

$$\boldsymbol{y}_{nt} = \boldsymbol{f}(\boldsymbol{x}_{nt}, \boldsymbol{\alpha}_n) + \boldsymbol{\nu}_{nt}, \tag{1}$$

where $\boldsymbol{y}_{nt}$, $\boldsymbol{f}$ and $\boldsymbol{\nu}_{nt}$ are $(J \times 1)$ vectors, $\boldsymbol{x}_{nt}$ comprises the explanatory variables of the system and $\boldsymbol{\alpha}_n$ is an unknown parameter vector. The subscript $n = 1,\ldots, N$ denotes individuals and $t = 1, 2,\ldots, T$ denotes time. The residual vector $\boldsymbol{\nu}_{nt}$ satisfies $\mathrm{E}[\boldsymbol{\nu}_{nt}] = 0$, $\mathrm{E}[\boldsymbol{\nu}_{nt}\boldsymbol{\nu}'_{nt}] = \boldsymbol{\Psi}$ and may be subject to first-order serial correlation:

$$\boldsymbol{\nu}_{nt} = \mathbf{R}\boldsymbol{\nu}_{n,t-1} + \boldsymbol{\zeta}_{nt}, \qquad \lim_{j \to +\infty} \mathbf{R}^j = 0 \tag{2}$$

with $\mathrm{E}[\boldsymbol{\zeta}_{nt}] = 0$, $\mathrm{E}[\boldsymbol{\nu}_{n,t-1}\boldsymbol{\zeta}'_{nt}] = 0$, $\mathrm{E}[\boldsymbol{\zeta}_{nt}\boldsymbol{\zeta}'_{nt}] = \boldsymbol{\Psi} - \mathbf{R}\boldsymbol{\Psi}\mathbf{R}'$ and $\mathrm{E}[\boldsymbol{\zeta}_{nt}\boldsymbol{\zeta}'_{is}] = 0$ for all $n \neq i$ or $t \neq s$. The matrix $\mathbf{R}$ $(J \times J)$ comprises unknown parameters. A first question is:

(1) Given that $\boldsymbol{\Psi}$ is positive definite, which conditions on the matrix $\mathbf{R}$ ensure that variance matrix $\boldsymbol{\Psi} - \mathbf{R}\boldsymbol{\Psi}\mathbf{R}'$ be also positive definite?

Notice that the equation system is not singular here as was the case in Berndt and Savin (1975). Hence, in contrast to their situation, we have no further restrictions on $\mathbf{R}$. In the special case where $J = 1$, the condition $\mathbf{R} < 1$ is necessary and sufficient for $(1 - \mathbf{R}^2)\boldsymbol{\Psi}$ to be positive. What about the cases where $J > 1$?

For simplifying the presentation, it is helpful to stack up Eq. (1) over the J equations first, and then over time periods and individuals:

$$\mathbf{y} = \mathbf{f}(\mathbf{x}, \boldsymbol{\alpha}) + \boldsymbol{\nu}. \tag{3}$$

Applying the ordinary nonlinear SUR estimator to Eq. (3) takes into account contemporary but not serial correlation of the residuals. Thus, this procedure yields consistent but inefficient estimates. Indeed, the variance matrix of $\boldsymbol{\Psi}$ $(JTN \times 1)$ in the regression is not $\mathbf{I}_{NT} \otimes \boldsymbol{\Phi}$ but $\mathrm{E}[\boldsymbol{\nu}\boldsymbol{\nu}'] = \mathbf{I}_N \otimes \boldsymbol{\Phi}$ where the $\boldsymbol{\nu}$ $(JT \times JT)$ is given by:[1]

$$\Phi = \begin{bmatrix} \boldsymbol{\Psi} & \boldsymbol{\Psi}\mathbf{R}' & \cdots & \boldsymbol{\Psi}\mathbf{R}'^{T-1} \\ \mathbf{R}\boldsymbol{\Psi} & \boldsymbol{\Psi} & \cdots & \boldsymbol{\Psi}\mathbf{R}'^{T-2} \\ \vdots & \vdots & & \vdots \\ \mathbf{R}^{T-1}\boldsymbol{\Psi} & \mathbf{R}^{T-2}\boldsymbol{\Psi} & \cdots & \boldsymbol{\Psi} \end{bmatrix}. \tag{4}$$

Now, a second question arises:

(2) Given that $\boldsymbol{\Psi}$ is positive definite, when is the variance matrix $\boldsymbol{\Phi}$ positive definite?

[1] The notation $\mathbf{I}_N$ is used to represent the $(N \times N)$ identity matrix.

Below, it is first shown that both above questions are related in the sense that $\mathbf{\Phi}$ is positive definite if and only if $\mathbf{\Psi}$ and $\mathbf{\Psi} - \mathbf{R\Psi R}'$ are positive definite, and that when $\mathbf{\Phi}$ is positive definite, the eigenvalues of $\mathbf{R}$ automatically have eigenvalues with modulus inferior to 1. It is well known that this, in turn, ensures that matrix $\mathbf{R}$ converges: $\lim_{j \to +\infty} \mathbf{R}^j = 0$. The converse of this result, however, is not true.[2]

3. Proof of the claims

•It is shown that the matrix $\mathbf{\Phi}$ defined in Eq. (4) is positive definite iff $\mathbf{\Psi}$ and $\mathbf{\Psi} - \mathbf{R\Psi R}'$ is positive definite. For

$$\mathbf{S} \equiv \begin{bmatrix} \mathbf{I}_J & 0 & 0 & \cdots & 0 \\ -\mathbf{R} & \mathbf{I}_J & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{I}_j & 0 \\ 0 & \cdots & 0 & -\mathbf{R} & \mathbf{I}_j \end{bmatrix},$$

it follows that

$$\mathbf{S\Phi S}' = \begin{bmatrix} \mathbf{\Psi} & 0 & \cdots & 0 \\ 0 & \mathbf{\Psi} - \mathbf{R\Psi R}' & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{\Psi} - \mathbf{R\Psi R}' \end{bmatrix},$$

implying that $\mathbf{\Phi}$ is positive definite iff $\mathbf{\Psi}$ and $\mathbf{\Psi} - \mathbf{R\Psi R}'$ are positive definite.

•When $\mathbf{\Psi}$ and $\mathbf{\Psi} - \mathbf{R\Psi R}'$ are positive definite, the eigenvalues of $\mathbf{R}$ have modulus less than 1.

In order to prove this assertion, the real Jordan canonical form of R as given by Horn and Johnson (1991) in Theorem 3.4.5, is useful.

[2] For the following choice of $\mathbf{\Psi}$ (positive definite) and $\mathbf{R}$ (with eigenvalues smaller than 1), the matrix $\mathbf{\Psi} - \mathbf{R\Psi R}'$ is not positive definite.

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} 1/3 & -1 \\ 0 & 1/2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1/3 & 0 \\ -1 & 1/2 \end{pmatrix} = \begin{pmatrix} -1/9 & 1/2 \\ 1/2 & 3/4 \end{pmatrix}.$$

Theorem. *To each real matrix* $\boldsymbol{R}$ $(J \times J)$, *there exists a real nonsingular matrix* $\boldsymbol{Q}$ *such that* $\boldsymbol{R} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^{-1}$, *where* $\boldsymbol{\Lambda}$ $(J \times J)$ *is a block diagonal real matrix of the form:*

$$\boldsymbol{\Lambda} = \begin{bmatrix} \mathbf{C}_1 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & & & \vdots \\ \vdots & \ddots & \mathbf{C}_G & \ddots & & \vdots \\ \vdots & & \ddots & \mathbf{D}_{G+1} & \ddots & \vdots \\ \vdots & & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & \cdots & 0 & \mathbf{D}_H \end{bmatrix}.$$

The block $\mathbf{C}_g$ is associated to the complex eigenvalues $\lambda_g = a_g + b_g i$ and $\bar{\lambda}_g = a_g - b_g i$ of matrix $\mathbf{R}$ and has the form

$$\mathbf{C}_g = \begin{bmatrix} \mathbf{B}_g & I_2 & 0 & \cdots & 0 \\ 0 & \mathbf{B}_g & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & \ddots & I_2 \\ 0 & \cdots & \cdots & 0 & \mathbf{B}_g \end{bmatrix}, \text{ where } \mathbf{B}_g = \begin{bmatrix} a_g & b_g \\ -b_g & a_g \end{bmatrix}.$$

The block $\mathbf{D}_g$ is associated to the real eigenvalue λ_g of matrix $\mathbf{R}$ and has the form

$$\mathbf{D}_g = \begin{bmatrix} \lambda_g & 1 & 0 & \cdots & 0 \\ 0 & \lambda_g & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & \ddots & 1 \\ 0 & \cdots & \cdots & 0 & \lambda_g \end{bmatrix}.$$

Using the real Jordan canonical form of $\mathbf{R}$,

$$\mathbf{\Psi} - \mathbf{R}\mathbf{\Psi}\mathbf{R}' = \mathbf{\Psi} - \mathbf{Q}\Lambda\mathbf{Q}^{-1}\mathbf{\Psi}\mathbf{Q}'^{-1}\Lambda'\mathbf{Q}'.$$

The positive definiteness of $\mathbf{\Psi}$ and $\mathbf{\Psi} - \mathbf{R}\mathbf{\Psi}\mathbf{R}'$ implies that

$$\boldsymbol{e}_k'\mathbf{Q}^{-1}\mathbf{\Psi}\mathbf{Q}'^{-1}\boldsymbol{e}_k > \boldsymbol{e}_k'\Lambda\mathbf{Q}^{-1}\mathbf{\Psi}\mathbf{Q}'^{-1}\Lambda'\boldsymbol{e}_k > 0 \tag{5}$$

where $\boldsymbol{e}_k$ is the k^{th} column of the $(J \times J)$ identity matrix. Notice that $\boldsymbol{e}_k'\mathbf{\Lambda}$ gives the k^{th} line of $\mathbf{\Lambda}$. Thus, the vector $\boldsymbol{e}_k$ can be chosen to pick up the last lines of $\mathbf{C}_g$ and $\mathbf{D}_g$ (given by $(0,\ldots,0,\ a_g,\ b_g)$, $(0,\ldots,0,\ -b_g,\ a_g)$ and $(0,\ldots,\ 0,\ \lambda_g)$). *For $\boldsymbol{e}_k'\mathbf{\Lambda}=(0,\ldots,0,\ \lambda_g,\ 0,\ldots,\ 0)$, Eq. (5) becomes

$$\boldsymbol{e}_k'Q^{-1}\mathbf{\Psi}\mathbf{Q}'^{-1}\boldsymbol{e}_k > \lambda_g^2\boldsymbol{e}'_k\mathbf{Q}^{-1}\mathbf{\Psi}\mathbf{Q}'^{-1}\boldsymbol{e}_k,$$

from which $|\lambda_g| < 1$ directly follows.. *In the case where $\boldsymbol{e}_k'\mathbf{\Lambda}=(0,\ldots,\ 0,\ a_g,\ b_g,\ 0,\ldots,\ 0)$ and $\boldsymbol{e}_{k+1}'\mathbf{\Lambda}=(0,\ldots,\ 0,\ -b_g,\ a_g,\ 0,\ldots,\ 0)$, Eq. (5) yields

$$\boldsymbol{e}_k'\mathbf{Q}^{-1}\mathbf{\Psi}\mathbf{Q}'^{-1}\boldsymbol{e}_k > (a_g\boldsymbol{e}_k + b_g\boldsymbol{e}_{k+1})'\mathbf{Q}^{-1}\mathbf{\Psi}\mathbf{Q}'^{-1}(a_g\boldsymbol{e}_k + b_g\boldsymbol{e}_{k+1}) > 0$$

and

$$\boldsymbol{e}_{k+1}'\mathbf{Q}^{-1}\mathbf{\Psi}\mathbf{Q}'^{-1}\boldsymbol{e}_{k+1} > (a_g\boldsymbol{e}_{k+1} - b_g\boldsymbol{e}_k)'\mathbf{Q}^{-1}\mathbf{\Psi}\mathbf{Q}'^{-1}(a_g\boldsymbol{e}k + 1 - b_g\boldsymbol{e}_k) > 0.$$

Adding up these inequalities leads to $a_g^2 + b_g^2 < 1$ or equivalently $|\lambda_g| < 1$.

4. Consequence for estimation

A useful method that can be used to obtain a consistent estimate of $\mathrm{E}[\boldsymbol{\nu}\boldsymbol{\nu}']$ consists in replacing the matrices $\mathbf{\Psi}$ and $\mathbf{R}\mathbf{\Psi}$ by their sample analogue

$$\hat{\mathbf{\Psi}} = \frac{1}{N(T-1)}\sum_{n=1}^{N}\sum_{t=2}^{T}\hat{\boldsymbol{\nu}}_{n,t-1}\hat{\boldsymbol{\nu}}'_{n,t-1}$$

and

$$\widehat{\mathbf{R}\mathbf{\Psi}} = \frac{1}{N(T-1)}\sum_{n=1}^{N}\sum_{t=2}^{T}\hat{\boldsymbol{\nu}}_{nt}\hat{\boldsymbol{\nu}}'_{n,t-1}$$

The vector $\hat{\boldsymbol{\nu}}_{n\,t}$ $(J \times 1)$ denotes the estimated least squares residuals for observation (n,t) and is obtained by solving

$$\min_{\alpha} v(\alpha)'\ v(\alpha),$$

with $\boldsymbol{\nu}(\alpha)$ defined as in Eq. (3).

This allows to identify $\hat{\mathbf{R}} = \widehat{\mathbf{R}\Psi}\hat{\Psi}^{-1}$ or equivalently $(\hat{\mathbf{R}}')_j = \hat{\Psi}^{-1}\widehat{\Psi\mathbf{R}}'_j$ which is the least square estimator of the j^{th} line of $\mathbf{R}$ in the regression

$$\hat{v}_{j,n,t} = \hat{v}'_{n,t-1}\mathbf{R}_j + \zeta_{j,n,t}, \qquad n = 1,\ldots,N,\ t = 2,\ldots,T.$$

Using $\hat{\mathbf{R}}$ and $\hat{\Psi}$, it becomes possible to build up $\hat{\Phi}$ according to Eq. (4). It is also possible to iterate upon α and $\hat{\Phi}_g$ by solving at g^{th} iteration

$$\min_{\alpha} v(\alpha)' \left(\mathbf{I}_N \otimes \hat{\Phi}_g\right)^{-1} v(\alpha), \tag{6}$$

and then replacing $\hat{\Phi}_g$ by $\hat{\Phi}_{g+1}$ in Eq. (6), where $\hat{\Phi}_{g+1}$ is obtained from the estimates of the residual vector $\hat{\boldsymbol{v}}_{g+1}$ solving Eq. (6), and so on until convergence is achieved.

This approach has two main drawbacks: it does not ensure that the estimated matrix $\hat{\mathbf{R}}$ has eigenvalues with modulus inferior to 1, and there is no guarantee for $\hat{\Phi}$ to be positive definite (even though $\hat{\mathbf{R}}$ had eigenvalues with modulus inferior to 1). However, the above results indicate that imposing positive (semi-) definiteness on $\hat{\Phi}$ by methods outlined by Koebel et al. (2003) for example, automatically ensures that $\hat{\mathbf{R}}$ converges.

The advantage of this method is that the parameter vector α is estimated using the full set of observations, whereas the regression

$$(-\hat{\mathbf{R}}\ \ \mathbf{I}_J)\begin{pmatrix} \boldsymbol{y}_{n,t-1} \\ \\ \boldsymbol{y}_{nt} \end{pmatrix} = (-\hat{\mathbf{R}}\ \ \mathbf{I}_J)\begin{pmatrix} \boldsymbol{f}(x_{n,t-1},\alpha_n) \\ \\ \boldsymbol{f}(x_{nt},\alpha_n) \end{pmatrix} + (-\hat{\mathbf{R}}\ \mathbf{I}_J)\begin{pmatrix} \boldsymbol{\nu}_{n,t-1} \\ \\ \boldsymbol{\nu}_{nt} \end{pmatrix}$$

analogous to the one considered by Guilkey and Schmidt (1973), leads to loosing the JN observations for which $t = 1$. Thus, this method should be most useful in cases where JN is big in comparison to T, which is the case in most empirical contributions. See Koebel (2003) for an empirical implementation. In contrast, when both J and T are small, considering N regressions as a system of JT seemingly unrelated regressions, in the spirit of MaCurdy (1981), has the advantage of allowing arbitrary covariance structures across time and equations.

Acknowledgements

I am indebted to François Laisney, Alban Thomas and to an anonymous referee for helpful comments.

References

Beach, C.M., MacKinnon, J.G., 1979. Maximum likelihood estimates of singular equation systems with autoregressive disturbances. International Economic Review 20, 459–464.

Berndt, E.R., Savin, N.E., 1975. Estimation and hypothesis testing in singular equation systems with autoregressive disturbances. Econometrica 43, 937–957.

Guilkey, D.K., Schmidt, P., 1973. Estimation of seemingly unrelated regressions with vector autoregressive errors. Journal of the American Statistical Association 68, 642–647.

Horn, R.A., Johnson, C.R., 1991. Matrix Analysis. Cambridge Univ. Press, Cambridge.
Judge, G.G., Griths, W.E., Hill, R.C., Lütkepohl, H., Lee, T.-C., 1985. The Theory and Practice of Econometrics, 2nd ed. John Wiley & Sons, New York.
Koebel, B.M., 2003. Searching for Functional Structure in Multi-Output Multi-Skill Technologies, Faculty of Economics and Management Discussion Paper No. 2. Otto-von-Guericke University, Magdeburg.
Koebel, B.M., Falk, M., Laisney, F., 2003. Imposing and testing curvature conditions on a Box–Cox Cost Function. Journal of Business and Economic Statistics 21, 319–335.
MaCurdy, T.E., 1981. An empirical model of labor supply in a life-cycle setting. Journal of Political Economy 89, 1059–1085.