Biometrical Letters
Vol. 44 (2007), No. 2, 97-103

Non-observable regressors

João T. Mexia[1] and Manuela M. Oliveira[2]

[1] Department of Mathematics, Faculty of Science and Technology. New University of Lisbon. Monte da Caparica 2829-516 Caparica Portugal.
[2] Department of Mathematics. University of Évora. Colégio Luis António Verney, Rua Romão Ramalho 59, 7002. Évora. Portugal. Email: mmo@uevora.pt

SUMMARY

An example is used to show how, in clearly formulated problems, non-observable regressors may emerge. The model matrices and goal functions for the adjustment are discussed. The propose being the adjustment of not only the regression coefficients but also the non-observable regressors. Algorithms are presented to carry out the adjustments. It is discussed how the adjustment may be validated through analysis of the residues.

Key words: Non-observable regressor, regression coefficients, Algorithms.

1. Introduction

When a linear relation between a response variable and k regressors is assumed we are led to adjust a linear regression whenever we have corresponding values for all these variables. Let us now assume that one or more of the regressors are non-observable. We intend to show that if the problem under investigation has a sufficiently well-defined structure, non-observable regressors (NOR) may be estimated as well as the regression coefficients. We will start by discussing an example to see how non-observable regressors may emerge in clearly stated problems. Next we present some preliminary results that we will use. Then we consider the goal functions that are to be minimized. To carry out this minimization we present two algorithms which we discuss in a separate section. Lastly, we show how analysis of the residues may be used to validate the adjustments.

2. An example: Joint regression analysis

Joint regression analysis is a technique for the study of genotype × environment based on the use of regressions. In most applications the genotypes correspond to cultivars (e.g. Aastveit and Mejza, (1992)). Then, for each cultivar, a linear regression of yield on productivity is adjusted. This idea was easy to implement when one was analyzing a network of randomized block designs. To each block was, attributed, as a productivity level, it's environment index measured by it's mean yield (e.g. Gusmão (1985)). We now had only to adjust, for each cultivar, a linear regression of it's yields on the environment indexes. Nowadays randomized block designs have been largely replaced by α designs (e.g. John and Williams, (1995)). The blocks are now incomplete so we cannot measure the productivity of a block by it's average yield. To solve this problem, Mexia et al. (1999) introduced the $\pounds_2$ environment indexes. Let p_{ij} be 0 when the j-th cultivar is absent and 1 when the j-th cultivar is present, in the i-th block, with yield $y_{ij}, i=1,\ldots,b,\ j=1,\ldots,J$. Then, with $\boldsymbol{\theta}=(\alpha_1\beta_1,\ldots,\alpha_J\beta_J)$ and $\mathbf{x}=(x_1,\ldots,x_b)$ the vectors of regression coefficients and of environment indexes, we minimize

$$S(\boldsymbol{\theta},\mathbf{x})=\sum_{i=1}^{b}\sum_{j=1}^{J}p_{ij}(y_{ij}-\alpha_j-\beta_j x_i)^2.$$

The incidence matrix of the network gives us the weights $p_{ij}, i=1,\ldots,b,\ j=1,\ldots,J$, and thus enables joint estimation of regression coefficients and environmental indexes. This example shows how a change in the design led to emergence of a non-observable regressor and how the model structure enabled it to be estimated.

3. Preliminary results

If we have $\mathbf{a}=vec([\mathbf{a}_1,\ldots,\mathbf{a}_l])$ we will write $[\mathbf{a}_1,\ldots,\mathbf{a}_l]=vec^{-1}(\mathbf{a})$. Let us assume that we have h known regressors and l unobservable regressors. The model matrix would then be

$$X(\mathbf{u})=[X_0:X_1(\mathbf{u})],$$

where X_0 contains the values of the known regressors and

$$X_1(\mathbf{u}) = [\mathbf{u}_1, \ldots, \mathbf{u}_l] = vec_l^{-1}(\mathbf{u}),$$

the values of the non-observable regressors. Matrices $X(\mathbf{u})$, $\mathbf{X_0}$ and $X_1(\mathbf{u})$ will have n rows and $k = h+l$, h and l columns. With $\mathbf{A}^+$ the Moore-Penrose inverse of matrix $\mathbf{A}$, the orthogonal projection matrices on the range space $\Omega(\mathbf{u})$ of $X(\mathbf{u})$ and it's orthogonal complement $\Omega(\mathbf{u})^\perp$ will be $\varphi(\mathbf{u}) = X(\mathbf{u})X^+(\mathbf{u}) = X(\mathbf{u})(X^t(\mathbf{u})X(\mathbf{u}))^+ X^t(\mathbf{u})$ and $\varphi(\mathbf{u})^\perp = \mathbf{I}_n - \varphi((\mathbf{u})$, (e.g. Mexia, (1995)). Likewise the orthogonal projection matrices on the range space Ω_0 of $\mathbf{X_0}$ and it's orthogonal complement $\Omega_0^\perp$ will be $\varphi_0 = \mathbf{X}_0\mathbf{X}_0^t = \mathbf{X}_0(\mathbf{X}_0^t\mathbf{X}_0)^+\mathbf{X}_0^t$ and $\varphi_0^\perp = \mathbf{I}_n - \varphi_0$.

Moreover, we have the orthogonal direct sum $\Omega(\mathbf{u}) = \Omega_0 \cup w(\mathbf{u})$ of the range spaces Ω_0 and $w(\mathbf{u})$ of matrices $\mathbf{X_0}$ and $\varphi_0^\perp\mathbf{X}_1(\mathbf{u})$. Thus $dim(\Omega(\mathbf{u})) = dim(\Omega_0) + dim(w(\mathbf{u}))$. We have the upper bounds k, h and l for these discussions.

Let us assume that $dim(\Omega_0) = h$ and that $dim(\Omega) < k$, then the column vectors of $\mathbf{X_0}$ will be linearly independent and one, at least, of the column vectors of $X_1(\mathbf{u})$, say $\mathbf{u}_j$, will be a linear combination of the remaining column vectors of $X(\mathbf{u})$. There will also exist $\mathbf{z}\varepsilon\Omega(\mathbf{u})^\perp$, and if we add $\mathbf{z}$ to $\mathbf{u}_j$ we get a new model matrix $X_1(\mathbf{u}^\circ) = [u_1, \ldots, u_{j-1}, z + u_j, u_{j+1}, \ldots, u_l]$ and matrix $X_0(\mathbf{u}^\circ) = [\mathbf{X}_0 : X_1(\mathbf{u}^\circ)]$ will have the same number k=h+l of coluns as $X(\mathbf{u})$, moreover

$$dim(\Omega(\mathbf{u}^\circ)) = dim(\Omega(\mathbf{u})) + 1,$$

and

$$dim(w(\mathbf{u}^\circ)) = dim(w(\mathbf{u})) + 1.$$

The models we have considered have very interesting properties of scale invariance for the non-observable regressors. Thus, with $c \neq 0$, we have $\Omega(c\mathbf{u}) = \Omega(\mathbf{u})$ and $w(c\mathbf{u}) = w(\mathbf{u})$, thus, if $\mathbf{u}\varepsilon C$, $c\mathbf{u}\varepsilon C$.

4. Goal functions

In the absence of non-observable regressors we would have to minimize

$$S(\boldsymbol{\theta}) = \left\|\mathbf{Y} - \mathbf{X}\boldsymbol{\theta}\right\|^2,$$

while now we have to minimize

$$S(\boldsymbol{\theta}, \mathbf{u}) = \left\|\mathbf{Y} - X(\mathbf{u})\boldsymbol{\theta}\right\|^2.$$

Besides $S(\boldsymbol{\theta},\mathbf{u})$ we will also consider the partial goal function

$$S(\boldsymbol{\theta}\,|\,\mathbf{u}') = S(\boldsymbol{\theta},\mathbf{u}'),$$

and

$$S(\mathbf{u}\,|\,\boldsymbol{\theta}') = S(\boldsymbol{\theta}',\mathbf{u}),$$

obtained from $S(\boldsymbol{\theta},\mathbf{u})$ by fixing either $\mathbf{u}$ or $\boldsymbol{\theta}$. In the first case we take $\mathbf{u} = \mathbf{u}'$ and in the second $\boldsymbol{\theta} = \boldsymbol{\theta}'$.

Putting $W(\mathbf{u}) = (X^T(\mathbf{u})X(\mathbf{u})^+ X(\mathbf{u})$, we have $\varphi(\mathbf{u}) = X(\mathbf{u})W(\mathbf{u})$, thus with $\tilde{\boldsymbol{\theta}}(\mathbf{u}) = W(\mathbf{u})\mathbf{y}$ we will have $\varphi(\mathbf{u})\mathbf{y} = X(\mathbf{u})\tilde{\boldsymbol{\theta}}(\mathbf{u})$, as well as

$$\begin{aligned} S(\boldsymbol{\theta}\,|\,\mathbf{u}') &= \|\mathbf{y} - X(\mathbf{u}')\tilde{\boldsymbol{\theta}}(\mathbf{u}) + X(\mathbf{u}')(\tilde{\boldsymbol{\theta}}(\mathbf{u}) - \boldsymbol{\theta})\|^2 \\ &= \|\mathbf{y} - X(\mathbf{u}')\tilde{\boldsymbol{\theta}}(\mathbf{u})\|^2 + \|X(\mathbf{u}')(\tilde{\boldsymbol{\theta}}(\mathbf{u}) - \theta)\|^2 \geq \dot{S}(\mathbf{u}'), \end{aligned}$$

where

$$\begin{aligned} \dot{S}(\mathbf{u}') &= \|\mathbf{y} - X(\mathbf{u}')\tilde{\boldsymbol{\theta}}(\mathbf{u}')\|^2 \\ &= \|\varphi(\mathbf{u}')^{\perp}\mathbf{y}\|^2 = S(\tilde{\boldsymbol{\theta}}\mathbf{u}'\,|\,\mathbf{u}') \\ &= Min\{S(\boldsymbol{\theta}\,|\,\mathbf{u}')\}, \end{aligned}$$

since $\mathbf{y} - X(\mathbf{u}')\tilde{\boldsymbol{\theta}}(\mathbf{u}) \in \Omega(\mathbf{u}')$ and $X(\mathbf{u}')(\tilde{\boldsymbol{\theta}}(\mathbf{u}') - \theta) \in \Omega(\mathbf{u}')^{\perp}$ are mutually orthogonal.

Due to scale invariance, if $\mathbf{u} \neq \mathbf{0}$, $\dot{S}(\frac{1}{\|\mathbf{u}\|}\mathbf{u}) = \dot{S}(\mathbf{u})$, and, if $\Omega(\mathbf{u}) \subset \Omega(\dot{\mathbf{u}})$, $\dot{S}(\mathbf{u}) \geq \dot{S}(\dot{\mathbf{u}})$. Thus the minima of $\dot{S}(\mathbf{u})$ will be found in the family of vectors of $\mathscr{C}$ with norm 1.

5. Algorithms

We start with the $zig-zag$ iterative algorithm. To apply this algorithm it is necessary that whatever $\boldsymbol{\theta}'$ we may obtain an absolute minimum $\tilde{\mathbf{u}}(\boldsymbol{\theta}')$ for $S(\tilde{\mathbf{u}}(\boldsymbol{\theta}')\,|\,\boldsymbol{\theta}')$ thus we will have $S(\mathbf{u}\,|\,\boldsymbol{\theta}') \geq S(\tilde{\mathbf{u}}(\boldsymbol{\theta}')\,|\,\boldsymbol{\theta}')$ for all possible vectors $\mathbf{u}$.

For the first iteration we take the starting point $\mathbf{u} = \mathbf{u}_0$. For instance if we use α-designs the component of $\mathbf{u}_0$ for each block may be the average yield of the corresponding supra-block. Next we obtain the minimum $\tilde{\boldsymbol{\theta}}(\mathbf{u}_0) = w(\mathbf{u}_0)\mathbf{y}$ of $S(\boldsymbol{\theta}(\mathbf{u}_0)$ and the corresponding $\tilde{\mathbf{u}}(\tilde{\boldsymbol{\theta}}(\mathbf{u}_0))$. We then rescale $\tilde{\mathbf{u}}(\tilde{\boldsymbol{\theta}}(\mathbf{u}_0))$ to preserve the minimum and maximum of the environmental index.

Let u and $\bar{u}$ be the minimum and maximum initial values for the components

of $\mathbf{u}$. The kep these values invariant from iteration to iteration we take,

$$u_{1i} = \bar{u} + \frac{\bar{\bar{u}} - \bar{u}}{\bar{\bar{u}} - \bar{u}_1}(\tilde{u}_i - \bar{u}_1),\ i = 1,\ldots,n,$$

to obtain the component of vector $\mathbf{u}_1$. We now obtain $S_1 = \dot{S}(\mathbf{u}_1)$ and use $\mathbf{u}_1$ to start a second iteration. Rescaling, using u and u, will be carred out at the end of end iterations.

With $\mathbf{u}_i$ the vector obtained at the end of the i-th iteration and $S_i = \dot{S}(\mathbf{u}_i)$ we will have $S_0 > S_1 > S_2 > \ldots$. The iterations are stopped when $S_i - S_{i+1} < d$ with d a previously decided threshold.

With $\tilde{\mathbf{u}}$ the final vector adjusted we take $\tilde{\boldsymbol{\theta}} = \tilde{\boldsymbol{\theta}}(\tilde{\mathbf{u}})$ to complete the estimation. This algorithm performs well. Unfortunately convergence to the absolute minimum of the goal function has not been established.

Another algorithm is based on a double minimization. The first minimization replaces $S(\boldsymbol{\theta}, \mathbf{u})$ by $\dot{S}(\mathbf{u})$. Now, as we saw, we may assume that $\|\mathbf{u}\| = 1$. This point is important since the corresponding restriction $S^{\circ}{}_1(\mathbf{u})$ of $S^{\circ}(\mathbf{u})$ is defined in a compact set. We can now apply stochastic search algorithms, for instance see Appel *et a.* (2003) and Esquivel (2006), to carry out this sthochastic minimization. Since convergence of these algorithms is almost certain we can use the pair of estimators $(\hat{\mathbf{u}}, \hat{\boldsymbol{\theta}})$ with $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(\hat{\mathbf{u}})$ and $\hat{\mathbf{u}}$ the result of the second minimization.

Moreover $S(\boldsymbol{\theta}, \mathbf{u}) \geq \dot{S}(\mathbf{u}) \geq S(\hat{\mathbf{u}}) = S(\hat{\mathbf{u}}, \hat{\boldsymbol{\theta}}(\hat{\mathbf{u}})) = S(\hat{\mathbf{u}}, \hat{\boldsymbol{\theta}})$, so that $(\hat{\mathbf{u}}, \hat{\boldsymbol{\theta}})$ is an absolute minimum for $S(\boldsymbol{\theta}, \mathbf{u})$. In certain cases the second minimization can be carried out analytically (e.g. Sequeira, (2006) and Pereira, (2006)). The first of these cases was for logit models in which there were two additive factors. The second one was for Joint Regression Analysis and randomized block designs. These double minimization algorithms guarantee the absolute minimization of the goal function thus overcoming the theoretical problems the use of the $zig-zag$ algorithms.

6. Residues analysis

Sometimes it may be worthwhile to analyze the residues.

Let $(\tilde{\boldsymbol{\theta}}, \tilde{\mathbf{u}})$ be a pair of adjusted vectors the corresponding residues vector will be $\mathbf{R} = \mathbf{Y} - X(\tilde{\mathbf{u}})\tilde{\boldsymbol{\theta}}$. We now assume that $\mathbf{R}$ is normal with covariance matrix (approximately) equal to $\sigma^2\mathbf{W}$ with $\mathbf{W}$ a known positive semi-definite matrix. Matrix $\mathbf{W}$ will have eigenvalues $a_1 \geq, \ldots \geq a_r > 0, \ldots, 0$ associated with mutually

orthogonal eigenvectors $\boldsymbol{\alpha}_j, j = 1,\ldots,n$. Thus the matrix

$$\mathbf{G} = \mathbf{D}(a_1^{-1/2},\ldots,a_r^{-1/2})\begin{bmatrix}\boldsymbol{\alpha}_1^t\\ \vdots\\ \boldsymbol{\alpha}_r^t\end{bmatrix},$$

where $\mathbf{D}(a_1^{-1/2},\ldots,a_r^{-1/2})$ is the diagonal matrix with principal elements $a_1^{-1/2},\ldots,a_r^{-1/2}$, will be defined, and $\mathbf{GWG}^t = \mathbf{I}_r$.

If the model holds and the parameters were correctly estimated the mean vector of $\mathbf{R}$ will be null. Thus we may use the Shapiro-Wilk test to check if the hypothesis $H_0^\circ : \mathbf{GR} \sim N(\mathbf{0}, \sigma^2\mathbf{I}_r)$, holds. This procedure was used by Nunes (2006) in connection with Logit models. As an alternative approach we may adjust a secondary model to the residues to see if the initial model once adjusted, accounts for all relevant information. Absence of significant results for the residues model will validate the initial adjustment. A residues model was adjusted by Oliveira (2006) for Joint Regression Analysis.

Both these approaches may be used to validate adjustments obtained through $zig-zag$ algorithms.

Another method of validation for $zig-zag$ algorithms is to show that their adjustments agree with the double minimization results. This agreement was obtained by Sequeira (2006) for the already mentioned Logit model.

REFERENCES

Aastveit A. H., Mejza S. (1992). A selected bibliography on Statistical methods for the analysis of genotype × environment Interaction. Biuletyn Oceny Odmian, 24/25: 83-97.

Appel M. J., Labare R., Radulovic D. (2003). On accelerated random search - SIAM J. OPTIM- Vol. 14, 13 : 708 – 731.

Esquivel M. (2006). A conditional Gaussian martingale algorithm for global optimization. Lecture.

Gusmão L. (1985). An adequate design for regression analysis of yield trials. Theor. Appl. Genetics.p 71 : 314 – 319.

John J. A., Williams E. R. (1995). Cyclic and computer generated designs. Chaper and Hall.

Mexia, J. T. (1995). Introdução à Inferencia Estatística Linear. Edições lusofonas.

Mexia J. T., Pereira D., Baeta, J. (1999). L_2 environment indexes. Biometrical Letters, 38 : 33 – 40.

Oliveira A. (2006). Analysis of residues in Joint Regression Analysis. SCRA-FIM XIII-Thirteenth International Conference of the Forum for Interdisciplinary Mathematics on Interdisciplinary Mathematical and Statistical Techniques. September 1-4, New University of Lisbon- Tomar Polytechnic Institute Lisbon-Tomar, Portugal.

Pereira D. (2006). Double minimization for Joint Regression Analysis. SCRA-FIM XIII-Thirteenth

International Conference of the Forum for Interdisciplinary Mathematics on Interdisciplinary Mathematical and Statistical Techniques. September 1-4, New University of Lisbon- Tomar Polytechnic Institute Lisbon-Tomar, Portugal.

Schott J. R. (1997). Matrix analysis for Statistics. John Wiley & Sons.

Sequeira I. (2006). Double minimization for a logit model with two additive factors. SCRA-FIM XIII-Thirteenth International Conference of the Forum for Interdisciplinary Mathematics on Interdisciplinary Mathematical and Statistical Techniques. September 1-4, New University of Lisbon- Tomar Polytechnic Institute Lisbon-Tomar, Portugal.