Multivariate Nonlinear Least Squares: Direct and Beauchamp and Cornell Methodologies

Renato Guseo and Cinzia Mortarino

Abstract Simultaneous estimation in nonlinear multivariate regression contexts is a complex problem in inference. In this paper, we compare the generalized least squares approach, GLS, with the well-known methodology by Beauchamp and Cornell, B&C, and with the standard nonlinear least squares approach, NLS. In the first part of the paper, we contrast B&C versus standard NLS highlighting, from the theoretical point of view, how a model specification error could affect the estimation. A comprehensive simulation study is also performed in order to evaluate the effectiveness of B&C versus standard NLS both under correct or misspecified models.

Key words: nonlinear regression, Beauchamp and Cornell method, robustness

1 Introduction

Multiple linear regression is a central issue in statistical relationships modelling. Under weak conditions, ordinary least squares methodology (OLS) is usually applied. Due to the Gauss-Markov theorem, generalized least squares (GLS) give rise to the best linear unbiased estimator if observations are correlated with a known covariance matrix and the error term has a zero mean ([1]). Similar results are well-known for the multivariate case. Nevertheless, if the systematic relationships are nonlinear, previous optimality does not apply due to intrinsic curvatures of the *solution locus*. Following the suggestions of the GLS methodology, Beauchamp and Cornell [2] introduced an asymptotic approach (here denoted as B&C) for the multivariate nonlinear case when the corresponding covariance matrix is unknown.

Let us study a system with d nonlinear equations $y_{ij} = f_j(x_{ij};\vartheta) + \varepsilon_{ij},\ \ i = 1,2,\cdots,n,\ \ j = 1,2,\cdots,d$, where $\vartheta \in \mathbb{R}^p$ is a common vector of parameters with

Renato Guseo
Department of Statistical Sciences, via C. Battisti 241, Padua, Italy e-mail: renato.guseo@unipd.it

Cinzia Mortarino
Department of Statistical Sciences, via C. Battisti 241, Padua, Italy e-mail: mortarino@stat.unipd.it

$p << n$. In vector form we can write the model as follows,

$$y = f(\vartheta) + \varepsilon, \qquad y, f(\cdot), \varepsilon \in \mathbb{R}^{nd}, \tag{1}$$

where $y = (y_{11}, \ldots, y_{n1}, y_{12}, \ldots, y_{n2}, \ldots, y_{1d}, \ldots, y_{nd})'$ and the elements of $f(\vartheta)$ and ε are arranged consequently. We assume $\varepsilon \sim \mathcal{N}_{nd}(0, \Omega)$, where $\Omega = \Sigma \otimes I_n$, and $\otimes$ denotes the Kronecker product. In other words, for each component of the d-variate response, we consider zero mean, homoscedastic and uncorrelated errors. Both the requirement of normality and the Kronecker structure imposed on Ω are necessary to the B&C approach, while of course they are not required for the NLS method. In the rest of the paper, we will compare the two methods both in the case of a correct (1) and wrong (2) model specification,

$$y = g(\vartheta) + u, \tag{2}$$

where $g(\vartheta) = f(\vartheta) - \xi(\vartheta)$, $u = \xi(\vartheta) + \varepsilon$ and, in general, $\xi(\vartheta) \neq 0$, a.e..

2 Estimation methods: GLS, B&C, NLS

If the matrix Σ is known, we can apply the standard GLS approach, that minimizes, with respect to ϑ the Minkowski metric, ${}_f\delta^2_{GLS}(\vartheta) = [y - f(\vartheta)]' (\Sigma^{-1} \otimes I_n) [y - f(\vartheta)]$. Whenever Σ is unknown, [2] suggest to substitute Σ with a consistent estimate obtained from the residuals of marginal models estimated with direct NLS. In the rest of the paper, we will denote by B&C this two-stage procedure (for details about Σ estimation see also [3]). The ϑ estimate minimizing the GLS is asymptotically the same optimizing

$${}_f\delta^2_{BC}(\vartheta) = [y - f(\vartheta)]'(\hat{\Sigma}^{-1} \otimes I_n)[y - f(\vartheta)]. \tag{3}$$

Our aim is to compare the B&C approach with NLS method that ignores the covariance structure minimizing the Euclidean metric ${}_f\delta^2_S(\vartheta) = [y - f(\vartheta)]'[y - f(\vartheta)]$. Let us denote ${}_f\hat{\vartheta}_S$ the direct NLS estimate, and ${}_f\hat{\vartheta}_{BC}$ the B&C estimate.

3 Estimation methods in case of uncorrect model specification

The B&C approach starting from model (1) leads to a consistent estimate of $\Sigma \otimes I_n$. Conversely, if the wrong model (2) is specified, through the first step residuals, we obtain a consistent estimate of $E(uu') = \Sigma \otimes I_n + \xi(\vartheta)\xi(\vartheta)' = \Phi$. If we denote by $\widehat{\Phi}$ the estimate of the covariance structure obtained with model (2), the B&C approach leads to the minimization, with respect to ϑ, of the function

$${}_g\delta^2_{BC}(\vartheta) = [y - g(\vartheta)]'\widehat{\Phi}^{-1}[y - g(\vartheta)].$$

The optimal value, ${}_g\hat{\vartheta}_{BC}$, is asymptotically the same that minimizes

$$
\begin{aligned}
{}_g\delta^2_{GLS}(\vartheta) &= [y-g(\vartheta)]'\Phi^{-1}[y-g(\vartheta)] = [y-f(\vartheta)]'(\Sigma^{-1}\otimes I_n)[y-f(\vartheta)]+\\
&+ 2\xi(\vartheta)'(\Sigma^{-1}\otimes I_n)[y-f(\vartheta)] + +\xi(\vartheta)'(\Sigma^{-1}\otimes I_n)\xi(\theta)+\\
&- \frac{1}{1+\xi(\theta)'(\Sigma^{-1}\otimes I_n)\xi(\theta)}\Big\{\big(\xi(\theta)'(\Sigma^{-1}\otimes I_n)[y-f(\vartheta)]\big)^2+ \qquad (4)\\
&+ 2\xi(\vartheta)'(\Sigma^{-1}\otimes I_n)\xi(\theta)\xi(\theta)'(\Sigma^{-1}\otimes I_n)[y-f(\vartheta)] + \big(\xi(\theta)'(\Sigma^{-1}\otimes I_n)\xi(\theta)\big)^2\Big\}.
\end{aligned}
$$

From Eq. (4), we see that the function ${}_g\delta^2_{GLS}(\vartheta)$ differs from ${}_f\delta^2_{GLS}(\vartheta)$ and, in particular, there are some "interaction" terms between $\xi(\vartheta)$, which is the specification error, and the actual covariance structure Σ. At the first step in the application of the B&C procedure, the estimate $\hat{\Phi}$ misunderstands the variability of the stochastic component, Σ, for the specification error, $\xi(\vartheta)$, modifying the objective function.

Conversely, if we ignore the covariance structure, we minimize

$$
{}_g\delta^2_{S}(\vartheta) = [y-f(\vartheta)]'[y-f(\vartheta)] + 2\xi(\vartheta)'[y-f(\vartheta)] + \xi(\vartheta)'\xi(\vartheta). \qquad (5)
$$

In other words, we aim at evaluating to what extent the B&C approach really gives us an advantage with respct to the simpler NLS method.

4 A simulation study

Starting from the example originally proposed by [2] (and later by [3]), we consider as system f (true model) the following compartmental model with two response components, d=2, and three free parameters, p=3,

$$
\begin{aligned}
f_1(\vartheta) &= \vartheta_1 e^{-\vartheta_2 x} + (1-\vartheta_1)e^{-\vartheta_3 x}\\
f_2(\vartheta) &= 1-(\vartheta_1+\vartheta_4)e^{-\vartheta_2 x} + (\vartheta_1+\vartheta_4-1)e^{-\vartheta_3 x}
\end{aligned}
$$

where $\vartheta_4 = \frac{(\vartheta_3-\vartheta_2)\vartheta_1(1-\vartheta_1)}{(\vartheta_3-\vartheta_2)\vartheta_1+\vartheta_2}$. The true ϑ value, was selected as ϑ_o =(0.047, 0.002, 0.066). After choosing different specifications for Σ matrix, for each of them 100 values for the vector ε were generated, leading to corresponding response vectors, y. Each of them was used to fit the regression model f. In order to assess the effect of a wrong specification, 5 different g models were also fitted. In detail:

$$
\begin{aligned}
&g_1 = \begin{cases} \alpha x^2+\beta x+\gamma & d{=}1\\ -\alpha x^2-\beta x+\zeta & d{=}2 \end{cases}, \quad g_2(\vartheta) = f(\vartheta)|_{\vartheta_1=\vartheta_3}, \quad g_3(\vartheta) = f(\vartheta)|_{\vartheta_1=\vartheta_2},\\
&g_4(\vartheta) = f(\vartheta_o)+F(\vartheta_o)(\vartheta-\vartheta_o),\; g_5(\vartheta) = f(\vartheta_o^*)+F(\vartheta_o^*)(\vartheta-\vartheta_o^*)
\end{aligned}
$$

where ϑ_o^* =(0.0423, 0.0022, 0.0726). In other words, g_1 represents an interpolating polynomial, g_2 and g_3 are restricted versions of the correct f function (g_2 imposes a plausible link, while g_3 sets a strong constraint). Finally, g_4 and g_5 are first order approximations of f evaluated at the true ϑ_o point and at a close ϑ_o^* point. In order to compare the two procedures we had to choose a sensible criterion. Since the different g models did not depend upon the same parameter set, we focused on a measure pertaining to a common element, i.e., the predicted response, $\hat{y}$. Due to Eq. (4) we feared that ${}_g\hat{\vartheta}_{BC}$ could be heavily affected by the confounding between Σ and $\xi(\vartheta)$, moving ${}_g\hat{\vartheta}_{BC}$ far from the optimal value (and thus moving ${}_g\hat{y}_{BC} = g({}_g\hat{\vartheta}_{BC})$ far from the observed value y). For this reason we evaluated $\rho^2_{y,\hat{y}}$, the squared Pearson correlation coefficient between vector y (observed values) and vector $\hat{y}$ (fitted values with both procedures). In Figure 1 we compare for each y vector, the ρ^2 coefficients for B&C and NLS fitted values. We observe that when the correct model is specified (f), or when the specification error is narrow $(g_4,\ g_5)$, the two estimation methods provide very similar values. Conversely, alternative misspecification errors give rise to mild (g_2) or very strong $(g_1,\ g_3)$ preference towards the NLS procedure. Analogous patterns can be observed for different Σ configurations. These features are even more evident when the two response components have a different number of observations. As a concluding remark, and within the limitations of the present simulation study, our opinion is that whenever the specification of the model used cannot be fully trusted, the B&C method might be very misleading and due to its robustness the NLS approach should be preferred.

References

1. Aitken, A.: On least squares and linear combinations of observations. Proceedings of the Royal Society of Edinburgh **A, 55**, 42–48 (1935)
2. Beauchamp, J., Cornell, R.: Simultaneous nonlinear estimation. Technometrics **8**(2), 319–326 (1966)
3. Seber, G., Wild, C.: Nonlinear Regression. Wiley, New York (2003)

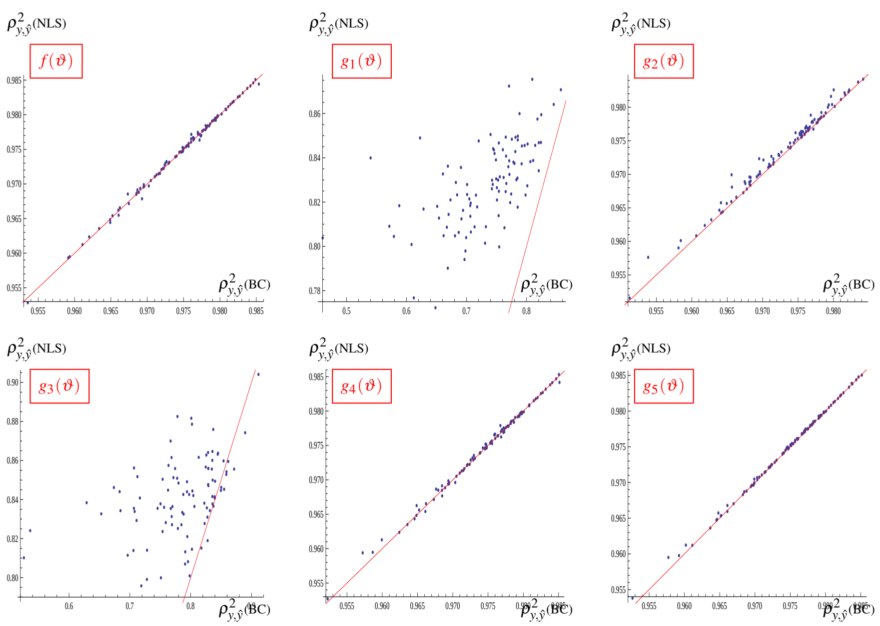

Fig. 1 Comparison between ρ^2 coefficients for B&C and NLS fitted values for alternative model specifications ($\sigma_{11} = 0.05,\ \sigma_{22} = 0.03, \sigma_{12} = -0.6\sigma_{11}\sigma_{22}$). The red line of each sub-plot is the bisector of the first quadrant.