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 $y_{ij} = f_j(x_{ij}; \theta) + \epsilon_{ij}, \quad i = 1, 2, \ldots, n, \quad j = 1, 2, \ldots, d$, where $\theta \in \mathbb{R}^p$ 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
\[ y = f(\theta) + \epsilon, \quad y, f(\cdot), \epsilon \in \mathbb{R}^{nd}, \] (1)

where \( y = (y_{11}, \ldots, y_{n1}, y_{12}, \ldots, y_{n2}, \ldots, y_{id}, \ldots, y_{nd})^\prime \) and the elements of \( f(\theta) \) and \( \epsilon \) are arranged consequently. We assume \( \epsilon \sim \mathcal{N}_{nd}(0, \Omega) \), 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 \( \Omega \) 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,

\[ y = g(\theta) + u, \] (2)

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

2 Estimation methods: GLS, B&C, NLS

If the matrix \( \Sigma \) is known, we can apply the standard GLS approach, that minimizes, with respect to \( \theta \) the Minkowski metric, \( f^2_{\text{GLS}}(\theta) = [y - f(\theta)]'(\Sigma^{-1} \otimes I_n)[y - f(\theta)]. \) Whenever \( \Sigma \) is unknown, [2] suggest to substitute \( \Sigma \) 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 \( \Sigma \) estimation see also [3]). The \( \hat{\theta} \) estimate minimizing the GLS is asymptotically the same optimizing

\[ f^2_{\text{BC}}(\theta) = [y - g(\theta)]'\hat{\Sigma}^{-1}(\hat{\Sigma}^{-1} \otimes I_n)[y - g(\theta)]. \] (3)

Our aim is to compare the B&C approach with NLS method that ignores the covariance structure minimizing the Euclidean metric \( f^2_\delta(\theta) = [y - f(\theta)]'[y - f(\theta)]. \) Let us denote \( f^2_\delta \) the direct NLS estimate, and \( f^2_{\text{BC}} \) 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 \( E(uu') = \Sigma \otimes I_n + \xi(\theta)\xi(\theta)' = \Phi \). If we denote by \( \hat{\Phi} \) the estimate of the covariance structure obtained with model (2), the B&C approach leads to the minimization, with respect to \( \theta \), of the function

\[ g^2_{\text{BC}}(\theta) = [y - g(\theta)]'\hat{\Phi}^{-1}[y - g(\theta)]. \]

The optimal value, \( g^2_{\text{BC}} \), is asymptotically the same that minimizes
\[ s^2_{\text{GLS}}(\theta) = [y - g(\theta)]' \Phi^{-1} [y - g(\theta)] = [y - f(\theta)]' (\Sigma^{-1} \otimes I_n) [y - f(\theta)] + \\
+ 2 \xi(\theta)' (\Sigma^{-1} \otimes I_n) [y - f(\theta)] + \xi(\theta)' (\Sigma^{-1} \otimes I_n) \xi(\theta) + \\
- \frac{1}{1 + \xi(\theta)' (\Sigma^{-1} \otimes I_n) \xi(\theta)} \left\{ (\xi(\theta)' (\Sigma^{-1} \otimes I_n) [y - f(\theta)])^2 + \\
+ 2 \xi(\theta)' (\Sigma^{-1} \otimes I_n) \xi(\theta) \xi(\theta)' (\Sigma^{-1} \otimes I_n) \xi(\theta) \right\}. \tag{4} \]

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

Conversely, if we ignore the covariance structure, we minimize

\[ s^2_{\xi}(\theta) = [y - f(\theta)]' [y - f(\theta)] + 2 \xi(\theta)' (y - f(\theta)) + \xi(\theta)' \xi(\theta). \tag{5} \]

In other words, we aim at evaluating to what extent the B&C approach really gives us an advantage with respect 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, \( p=3 \),

\[
\begin{align*}
f_1(\theta) &= \theta_1 e^{-\theta_2 x} + (1 - \theta_1) e^{-\theta_3 x} \\
f_2(\theta) &= 1 - (\theta_1 + \theta_2) e^{-\theta_4 x} + (\theta_1 + \theta_2 - 1) e^{-\theta_5 x},
\end{align*}
\]

where \( \theta_4 = \frac{(\theta_1 - \theta_2) \theta_1 (1 - \theta_1)}{(\theta_1 - \theta_2) \theta_1 + \theta_2} \). The true \( \theta \) value, was selected as \( \theta_0 = (0.047, 0.002, 0.066) \). After choosing different specifications for \( \Sigma \) matrix, for each of them 100 values for the vector \( e \) 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:

\[
\begin{align*}
g_1 &= \begin{cases} \alpha x^2 + \beta x + \gamma & d=1, \\
- \alpha x^2 - \beta x + \xi & d=2, 
\end{cases} \\
g_2(\theta) &= f(\theta)|_{\theta_1=\theta_3}, \quad g_3(\theta) = f(\theta)|_{\theta_1=\theta_2}, \\
g_4(\theta) &= f(\theta) + F(\theta_0) (\theta - \theta_0), \quad g_5(\theta) = f(\theta^*_0) + F(\theta^*_0) (\theta - \theta^*_0).
\end{align*}
\]

where \( \theta^*_0 = (0.0423, 0.0022, 0.0726) \). 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 \( \theta_0 \) point and at a close \( \theta^*_0 \) 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 $g \hat{\theta}_{BC}$ could be heavily affected by the confounding between $\Sigma$ and $\xi(\theta)$, moving $g \hat{\theta}_{BC}$ far from the optimal value (and thus moving $g \hat{\theta}_{BC} = g(g \hat{\theta}_{BC})$ 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 $\rho^2$ 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 $\Sigma$ 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


Fig. 1 Comparison between $\rho^2$ 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.