Comment on "Wang et al. (2005), Robust Estimating Functions and Bias Correction for Longitudinal Data Analysis"

Nicola Lunardon*   Giovanna Menardi†

Abstract

This note provides a discussion on the manuscript by Wang et al. (2005), who aim to robustify inference for longitudinal data analysis by replacing the ordinary generalized estimating function with an influence-bounded, possibly biased, version. To adjust for the bias of the ensuing robust estimator, the authors provide its analytic approximation by means of asymptotic expansions, and estimate it by pluging-in a non-robust estimate of the parameter of interest. In this letter, we argue that the proposed bias-corrected estimator is, in fact, non-robust.

Keywords: Bias; Estimating function; M-estimation; Robustness.

1   Framework and notation

Consider a response variable $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{in_i})$ and a set of predictors $(\mathbf{x}_{i1}, \ldots, \mathbf{x}_{in_i})^T$ observed over time on a subject i, i.e. $\mathbf{x}_{it}$ is a p-dimensional vector ($t = 1, \ldots, n_i$, $i = 1, \ldots, K$). Assume that $g(\boldsymbol{\mu}_i) = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{in_i})^T \boldsymbol{\beta}$, for some link function $g(\cdot)$, $\boldsymbol{\mu}_i = \mu_i(\boldsymbol{\beta}) = \mathbb{E}(\mathbf{Y}_i)$, and $\boldsymbol{\beta}$ the p-dimensional parameter of interest, with true value $\boldsymbol{\beta}_0$. Let $F(\mathbf{y}; \boldsymbol{\beta}, \phi, \boldsymbol{\alpha}) = \prod_{i=1}^{K} F_i(\mathbf{y}_i; \boldsymbol{\beta}, \phi, \boldsymbol{\alpha})$ be the distribution function of the central model, assumed to underlie $\mathbf{Y} = (Y_1, \ldots, Y_K)$, $F_{it}(y_{it}; \boldsymbol{\beta}, \phi, \boldsymbol{\alpha})$ the marginal distributions of $\mathbf{Y}_i$, and ϕ, $\boldsymbol{\alpha}$ governing the dispersion and the working intra-subject correlation matrix, respectively. To ease notation, hereafter we remove dependence on ϕ and

---
*Department of Economics, Quantitative Methods and Business Strategy, University of Milano-Bicocca, Italy, email: nicola.lunardon@unimib.it

†Department of Statistical Sciences, University of Padova, Italy, email: menardi@stat.unipd.it

$\boldsymbol{\alpha}$, which are supposed to be known. To account for outliers, the true distribution function is assumed to lie in a λ-neighborhood of F, i.e. $H(\mathbf{y}; \boldsymbol{\beta}) = (1 - \lambda)F(\mathbf{y}; \boldsymbol{\beta}) + \lambda\delta(\mathbf{y})$, with $\delta(\cdot)$ an arbitrary distribution function and $\lambda \in [0, 1]$.

Within this framework, Wang et al. (2005) aim to robustify the inference based on the unbounded generalized estimating function $U_G(\boldsymbol{\beta})$ via an influence-bounded, possibly biased, version $U_R(\boldsymbol{\beta})$. The associated M-estimators, roots of the estimating equations $U_l(\boldsymbol{\beta}) = 0$, $l \in \{G, R\}$, are $\widehat{\boldsymbol{\beta}}_G$ and $\widehat{\boldsymbol{\beta}}_R$. Since the bias of an estimating function reflects on the bias of the associated estimator, which, in turn, may not even vanish as $K \to \infty$, a long-standing practice is to correct for the bias of the estimating function. Wang et al. (2005) follow a different route and correct directly for the bias of $\widehat{\boldsymbol{\beta}}_R$ by deriving its analytic approximation

$$\mathbb{E}(\widehat{\boldsymbol{\beta}}_R) - \boldsymbol{\beta}_0 \simeq \Delta(\boldsymbol{\beta}_0) = \Delta[\mathbb{E}\{U_R(\boldsymbol{\beta}_0)\}] = O(1), \tag{1}$$

where $\Delta(\cdot)$ takes over the dependence of the bias of $\widehat{\boldsymbol{\beta}}_R$ on the bias of $U_R(\cdot)$. The authors define the robust bias-corrected estimator

$$\widehat{\boldsymbol{\beta}}_N = \widehat{\boldsymbol{\beta}}_R - \Delta(\widehat{\boldsymbol{\beta}}_G), \tag{2}$$

where $\Delta(\widehat{\boldsymbol{\beta}}_G) = \Delta\{U_R(\widehat{\boldsymbol{\beta}}_G)\}$ is meant to estimate $\Delta(\boldsymbol{\beta}_0)$. The estimator $\widehat{\boldsymbol{\beta}}_G$ is assumed asymptotically unbiased for $\boldsymbol{\beta}_0$, i.e. $\mathbb{E}(\mathbf{Y}_i) = \mu_i(\boldsymbol{\beta}_0)$, otherwise its estimation would introduce a bias of the same order of the estimand $\Delta(\boldsymbol{\beta}_0)$.

We argue that the aim of robustifying generalized estimating functions and simultaneously limiting the bias of the robust estimator risks to be vanished when the bias is estimated by means of a non-robust estimator. Intuitively, since $U_R(\boldsymbol{\beta})$ is biased, $\widehat{\boldsymbol{\beta}}_R$ is asymptotically unbiased for $\boldsymbol{\beta}_0^R \neq \boldsymbol{\beta}_0$ so that the leftmost approximate equivalence in (1) results in

$$\boldsymbol{\beta}_0^R - \boldsymbol{\beta}_0 \simeq \Delta(\boldsymbol{\beta}_0).$$

With the same argument of Wang et al., once estimands are replaced by their asymptotically unbiased estimators, we get

$$\widehat{\boldsymbol{\beta}}_R - \widehat{\boldsymbol{\beta}}_G \simeq \Delta(\widehat{\boldsymbol{\beta}}_G).$$

Hence, from (2), $\widehat{\boldsymbol{\beta}}_N$ is roughly equivalent to $\widehat{\boldsymbol{\beta}}_G$, thus there is no guarantee about its robustness. A more detailed proof is available upon request.

2 Empirical study: epileptic seizure count data

In the empirical work which follows we provide some evidence of our claim. We first show via simulations a gross equivalence of $\widehat{\boldsymbol{\beta}}_N$ and $\widehat{\boldsymbol{\beta}}_G$. Then, we illustrate, through a sensitivity analysis, that $\widehat{\boldsymbol{\beta}}_N$ is as exposed to contaminations as $\widehat{\boldsymbol{\beta}}_G$.

Data are simulated by mimicking the features of the epileptic seizure count data analyzed, among others, by Diggle et al. (1994, Ex. 1.6), and Wang et al. (2005). The study was meant to assess the effectiveness of a progabide drug, based on data collected over 4 periods from a placebo and a treatment group. Data exhibited extra Poisson variation and the presence of at least one outlier within the treatment group (patient 207). Hereafter, we consider the setting in Diggle et al. (1994) who carried out the analysis by assuming an exchangeable correlation structure and by specifying $\log(\mu_{it}) = o_{it} + \beta_1 + \beta_2 x_{it}^{(1)} + \beta_3 x_{it}^{(2)} + \beta_4 x_{it}^{(1)} x_{it}^{(2)}, t = 1, \ldots, 5, i = 1, \ldots, K$, with $x_{it}^{(1)} = 1$ if $t = 1$ and 0 otherwise, $x_{it}^{(2)} = 1$ for treatment and 0 for placebo, $o_{it} = \log 8$ if $t = 1$ and $\log 2$ otherwise. It follows that the treatment effect is gauged by β_4.

Data have been simulated according to a joint distribution H_i with marginal distributions H_{it}; details are provided in Table 1, along with the complete simulation setting. Coherently with Wang et al. (2005), the contamination is set to guarantee that $\mathbb{E}(\mathbf{Y}_i) = \mu_i(\boldsymbol{\beta}_0)$, so that $\widehat{\boldsymbol{\beta}}_G$ is asymptotically unbiased. To mimic the mechanism giving rise to outlier 207 in the real data, contamination occurs within the treatment group and after the baseline period; hence, $\widehat{\beta}_{G4}$, i.e. the β_4-component of $\widehat{\boldsymbol{\beta}}_G$, is the only affected by contamination. The estimators $\widehat{\boldsymbol{\beta}}_G$ and $\widehat{\boldsymbol{\beta}}_R$ are computed from the associated estimating functions with correct specification of the mean, correlation structure, and $\mathrm{Var}(Y_{it}) = \phi \mu_{it}$. See details in the online Supporting Information.

Table 2 displays the Monte Carlo correlations between pairs of associated components of $\widehat{\boldsymbol{\beta}}_G$, $\widehat{\boldsymbol{\beta}}_R$, and $\widehat{\boldsymbol{\beta}}_N$ and shows that all the considered components are almost perfectly correlated. While the correlation between $(\widehat{\boldsymbol{\beta}}_R, \widehat{\boldsymbol{\beta}}_N)$ would be expected to be higher than the one between $(\widehat{\boldsymbol{\beta}}_G, \widehat{\boldsymbol{\beta}}_N)$, the inspection of the table reveals that the opposite is in fact true.

As a further step of the numerical work, we have performed a sensitivity analysis to illustrate the exposure of the bias-corrected estimator $\widehat{\boldsymbol{\beta}}_N$ to data contamination, compared to $\widehat{\boldsymbol{\beta}}_G$ and $\widehat{\boldsymbol{\beta}}_R$. For a specific random sample generated from F, one observation has been replaced with a point mass contamination having increasing value. Figure 1 displays the percentage

Table 1: Simulation settings. The true value of $\boldsymbol{\beta}, \boldsymbol{\alpha}$, and ϕ have been set equal to the corresponding estimates of Diggle et al. (1994, Example 1.6) on the real seizure count data when patient 207 is removed from the sample.

$$H_{it} = (1 - \lambda)F_{it} + (\lambda/2)\delta_{it}^+ + (\lambda/2)\delta_{it}^-$$
$$F_{it} = \text{NegBin}\{\omega_{it}, \omega_{it}/(\omega_{it} + \mu_{it})\}$$
$$\delta_{it}^+ = \text{NegBin}\{\omega_{it}^+, \omega_{it}^+/(\omega_{it}^+ + \mu_{it}^+)\}$$
$$\delta_{it}^- = \text{NegBin}\{\omega_{it}^-, \omega_{it}^-/(\omega_{it}^- + \mu_{it}^-)\}$$
$$\log(\mu_{it}) = o_{it} + \beta_1 + \beta_2 x_{it}^{(1)} + \beta_3 x_{it}^{(2)} + \beta_4 x_{it}^{(1)} x_{it}^{(2)}$$
$$\log(\mu_{it}^+) = o_{it} + \beta_1 + \beta_2 x_{it}^{(1)} + \beta_3 x_{it}^{(2)} + \beta_4^+ x_{it}^{(1)} x_{it}^{(2)}$$
$$\log(\mu_{it}^-) = o_{it} + \beta_1 + \beta_2 x_{it}^{(1)} + \beta_3 x_{it}^{(2)} + \beta_4^- x_{it}^{(1)} x_{it}^{(2)}$$
$$\lambda = 0.03, \ \alpha = 0.598, \ \phi = 10.385$$
$$\beta_1 = 1.348, \ \beta_2 = 0.112, \ \beta_3 = -0.107$$
$$\beta_4 = -0.302, \ \beta_4^+ = 0.389, \ \beta_4^- = -5.958$$
$$\omega_{it} = \mu_{it}/(\phi - 1) \ \Rightarrow \ \text{Var}_{F_{it}}(Y_{it}) = \phi \mu_{it}$$
$$\omega_{it}^+ = \mu_{it}^+/(\phi - 1) \ \Rightarrow \ \text{Var}_{\delta_{it}^+}(Y_{it}) = \phi \mu_{it}^+$$
$$\omega_{it}^- = \mu_{it}^-/(\phi - 1) \ \Rightarrow \ \text{Var}_{\delta_{it}^-}(Y_{it}) = \phi \mu_{it}^-$$
$$K = 60, \ n_0 = n_1 = K/2$$
$$\# \text{ Monte Carlo replications} = 20000$$

variation of the estimates of β_4 based on contaminated data with respect to the associated estimates based on original data versus the value of the point mass contamination. It is immediate to note how, while $\widehat{\beta}_{R4}$ remains broadly constant for increasing contamination, the bias-corrected estimate $\widehat{\beta}_{N4}$ is sensitive to the outlier, sharing a behavior similar to $\widehat{\beta}_{G4}$.

References

Diggle, P., Liang, K., and Zeger, S. L. (1994). *Longitudinal Data Analysis.* New York: Oxford University Press.

Wang, Y., Lin, X., and Zhu, M. (2005). Robust estimating functions and bias correction for longitudinal data analysis. *Biometrics* **61,** 684–691.

Supporting Information

Additional information to reproduce the analyses, along with further simulations, may be found online in the Supporting Information Section at the

Table 2: Monte Carlo estimates of correlation for pairs of components of estimators referring to the same component of $\boldsymbol{\beta}$. For greater clarity, the remaining correlations are not reported.

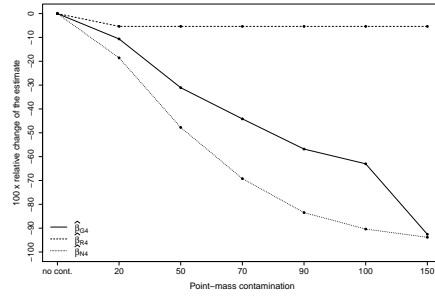|	$\widehat{\beta}_{G1}$	$\widehat{\beta}_{G2}$	$\widehat{\beta}_{G3}$	$\widehat{\beta}_{G4}$	$\widehat{\beta}_{N1}$	$\widehat{\beta}_{N2}$	$\widehat{\beta}_{N3}$	$\widehat{\beta}_{N4}$	$\widehat{\beta}_{R1}$	$\widehat{\beta}_{R2}$	$\widehat{\beta}_{R3}$	$\widehat{\beta}_{R4}$
$\widehat{\beta}_{G1}$	1.00	·	·	·	1.00	·	·	·	0.96	·	·	·
$\widehat{\beta}_{G2}$	·	1.00	·	·	·	1.00	·	·	·	0.90	·	·
$\widehat{\beta}_{G3}$	·	·	1.00	·	·	·	1.00	·	·	·	0.96	·
$\widehat{\beta}_{G4}$	·	·	·	1.00	·	·	·	1.00	·	·	·	0.91
$\widehat{\beta}_{N1}$	·	·	·	·	1.00	·	·	·	0.95	·	·	·
$\widehat{\beta}_{N2}$	·	·	·	·	·	1.00	·	·	·	0.90	·	·
$\widehat{\beta}_{N3}$	·	·	·	·	·	·	1.00	·	·	·	0.95	·
$\widehat{\beta}_{N4}$	·	·	·	·	·	·	·	1.00	·	·	·	0.91
$\widehat{\beta}_{R1}$	·	·	·	·	·	·	·	·	1.00	·	·	·
$\widehat{\beta}_{R2}$	·	·	·	·	·	·	·	·	·	1.00	·	·
$\widehat{\beta}_{R3}$	·	·	·	·	·	·	·	·	·	·	1.00	·
$\widehat{\beta}_{R4}$	·	·	·	·	·	·	·	·	·	·	·	1.00

Figure 1: Sensitivity analysis. Plot of $100\{\widehat{\beta}_{l4}(\delta_y)/\widehat{\beta}_{l4} - 1\}$ versus different point mass contaminations δ_y, with $\widehat{\beta}_{l4}$ the estimate of β_{l4} for uncontaminated data, and $\widehat{\beta}_{l4}(\delta_y)$ the estimate of β_{l4} when data are contaminated, $l \in \{G, N, R\}$.

end of the letter.