
Article:
Bravo, Francesco (2019) Robust estimation and inference for general varying coefficients models with missing observations. TEST. pp. 1-23. ISSN: 1863-8260

https://doi.org/10.1007/s11749-019-00692-0

TEST
https://doi.org/10.1007/s11749-019-00692-0

ORIGINAL PAPER

Robust estimation and inference for general varying coefficient models with missing observations

Francesco Bravo[1]

Received: 31 August 2018 / Accepted: 22 November 2019

Abstract
This paper considers estimation and inference for a class of varying coefficient models in which some of the responses and some of the covariates are missing at random and outliers are present. The paper proposes two general estimators—and a computationally attractive and asymptotically equivalent one-step version of them—that combine inverse probability weighting and robust local linear estimation. The paper also considers inference for the unknown infinite-dimensional parameter and proposes two Wald statistics that are shown to have power under a sequence of local Pitman drifts and are consistent as the drifts diverge. The results of the paper are illustrated with three examples: robust local generalized estimating equations, robust local quasi-likelihood and robust local nonlinear least squares estimation. A simulation study shows that the proposed estimators and test statistics have competitive finite sample properties, whereas two empirical examples illustrate the applicability of the proposed estimation and testing methods.

Keywords Local linear estimation · MAR · M and Z estimators · Wald statistic

Mathematics Subject Classification 62E20 · 62G10

1 Introduction

This paper considers estimation and inference for a general class of varying coefficient models where some of the responses and possibly some of the covariates are not always observed and outliers can be present. In the absence of outliers and when all the variables are observable, the estimation of the unknown infinite-dimensional

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11749-019-00692-0) contains supplementary material, which is available to authorized users.

✉ Francesco Bravo
francesco.bravo@york.ac.uk; fb6@york.ac.uk

[1] Department of Economics, University of York, York, UK

Published online: 27 November 2019

parameter for specific examples of the model considered here can be carried out using a number of alternative methods, such as spline approximation [see, for example, Eubank et al. (2004) and Hastie and Tibshirani (1993) for varying coefficient models, and Verhasselt (2014) for generalized varying coefficient models], series estimation [see, for example, Huang et al. (2002)] or local smoothing [see, for example, Fan et al. (1998) for local maximum likelihood, Fan et al. (1995) for local quasi-likelihood and Ruppert and Wand (1994) for local least squares among many others]. However, the influence function of the resulting estimators is unbounded, and thus, outliers or large deviations from the response variable to its conditional mean can have potentially a very negative effect on them. Furthermore, ignoring the fact that some of the data is not always observable or simply excluding the missing observations, the so-called complete case analysis, may also negatively affect the estimators and cause a great loss of information. Clearly, these potential problems might also have negative effects on the quality of any inference about the unknown infinite-dimensional parameter, since these inferences are typically based on test statistics that rely on these estimators.

This paper uses local smoothing, which is flexible enough to accommodate the estimation of the unknown infinite-dimensional parameter of the general model considered here and it makes the computation of the asymptotic covariance matrices required for inference relatively easy—especially when some of the observations in the sample are missing and/or are characterized by the presence of outliers. The paper assumes that some of the responses and possibly some of the covariates are missing at random (MAR) and proposes estimators that combine the ideas of smoothing and robust estimation with the inverse probability weighting (IPW) method (Horvitz and Thompson 1952). Robust estimation in nonparametric and semiparametric models has been considered by Fan et al. (1994), Boente et al. (2006), Bianco et al. (2011) and Hu and Cui (2010) among many others. Nonparametric and semiparametric estimation with MAR observations has been considered by Cheng (1994), Liang et al. (2004), Chen et al. (2006), Liang (2008) and Bianco et al. (2019) among others. Robust nonparametric and semiparametric estimation with missing data has been considered recently by Boente et al. (2009) and Bravo (2015). The estimators of this paper use a real-valued function that downweights high leverage covariates and either a robustified loss function (M estimator) or a robustified set of estimating equations (Z estimators) that yield bounded influence functions. The unknown infinite-dimensional parameter is estimated using the local linear estimator (Fan and Gijbels 1996), whereas the probability of missing—the so-called selection probability—is estimated using either a robust parametric or a robust nonparametric estimator.

For inference, the paper focuses on a robust Wald statistic. A similar Wald statistic was used by Bianco and Spano (2019) in the context of parametric nonlinear regression models with MAR responses and by Bianco et al. (2006) for the finite-dimensional parameter in a partially linear model with all the variables observable. The Wald statistics considered here are different from those considered by Bianco and Spano (2019) and Bianco et al. (2006), because they use IPW-based robust local estimators. Furthermore, one of the proposed Wald statistics is characterized by a nonstandard asymptotic distribution. Alternatively, a robust "distance" type of statistic, which is in the same spirit of the robust deviance statistic proposed by Cantoni and Ronchetti (2001) in the context of parametric quasi-likelihood estimation, could be used for

inference. However, as opposed to the robust Wald statistics considered here, this statistic is not asymptotic distribution-free (that is it depends on nuisance parameters) under the null hypothesis [see Remark 4 in Sect. 3.2], which makes it less attractive for inferential purposes.

The new results of the paper are the following: first, it establishes the asymptotic normality of the proposed robust local M and Z estimators and it shows that the presence of MAR observations affects their asymptotic variance but not the asymptotic bias. This result is consistent with that obtained by Chen et al. (2006) and Bravo and Jacho-Chavez (2016) for semiparametric estimators in the presence of MAR responses and with some general results obtained by Robins and Rotnitzky (1995) (albeit for statistical models with finite-dimensional parameters). Second, it considers asymptotically equivalent one-step version of the proposed estimators that are computationally attractive and seem to perform well in the simulations. These results are rather general as can be applied to both single and multiple equation models (i.e., models for longitudinal or repeated outcomes data) and can be used to robustify a number of estimators including the local quasi-likelihood estimator for generalized linear models of Fan et al. (1995) and Chen et al. (2006), the local maximum likelihood estimator for varying coefficient models of Cai et al. (2000) and the local nonlinear least squares estimator for varying coefficient models of Kurum et al. (2013). They can also be applied to construct estimators for marginal parameters of the model, such as the marginal mean of the response, see Remark 3 in Sect. 3.1 for an example. Third, it considers two Wald statistics that can be used to test linear hypotheses about the infinite-dimensional parameter. Both statistics are shown to have power against a sequence of local alternatives and are consistent when the local alternatives diverge. The second Wald statistic has a nonstandard asymptotic distribution, but it is asymptotically distribution-free under the null hypothesis, which makes it easy to simulate and therefore appealing to be used in the applied research. Fourth, the paper considers three examples, that have been previously considered in the literature but not with outliers and MAR observations: estimation and inference for models defined by a quasi-likelihood function, for nonlinear regression models and for generalized estimating equation models. Finally, this paper presents Monte Carlo evidence about the finite sample performance for the proposed estimators and test statistics for the three examples and considers two real data applications that illustrate the applicability and usefulness of the proposed methods.

The rest of the paper is structured as follows: The next section introduces the statistical models and the estimators. Section 3 contains the main results of the paper; Sect. 4 presents one of the three examples (the robust local generalized estimation equations model) and reports the results of a simulation study. Section 5 contains one of the two real data applications, and Sect. 6 contains some concluding remarks. A supplemental appendix available online contains the other two examples (with related simulation studies), the other real data application and all the proofs of the results of the paper.

The following notation is used throughout the paper: "^{T}" and "$\otimes$" denote, respectively, transpose and the standard Kronecker product, $\|\cdot\|$ is the Euclidean norm and finally for any vector v $v^{\otimes 2} = vv^T$.

2 The statistical models and estimators

Let $\{Y_i, X_i, U_i\}_{i=1}^n$ denote a random sample from $[Y, X, U]$, where both Y and U are scalar random variables and $X = \left[X_1^T, X_2^T\right]^T$ is an $\mathbb{R}^k$ $(k = k_1 + k_2)$-valued random vector.[1] Assume that the response variable Y is related to the covariates X and U through the semiparametric specification $\eta(X, \alpha_0(U))$, where $\eta(\cdot) : \mathcal{X} \times \mathcal{U} \to \mathbb{R}$ is a known smooth function, that could represent, for example, a regression function or the link function in a generalized linear model, and $\alpha_0(\cdot)$ is an $\mathbb{R}^k$-valued unknown infinite-dimensional parameter assumed to be sufficiently smooth.

To introduce the M-type estimator for $\alpha_0(\cdot)$, let

$$\zeta(Y, \eta(X, \alpha(U)))\,\omega(X) \tag{1}$$

denote a loss function, where $\omega(\cdot)$ is a real-valued function that downweights high leverage covariates. Examples of $\zeta(\cdot)$ include Huber's ρ and Tukey's bisquare ρ functions and the loss functions used, for example, by Boente et al. (2006) and Bianco et al. (2011) to bound the deviances and/or the Pearson residuals. Let δ^Y and δ^{X_1} denote the binary indicators of missingness for Y and X_1, respectively; for $\delta = \delta^Y \delta^{X_1}$ let

$$\Pr(\delta = 1|Y, X, U) = \Pr(\delta = 1|X_2, U) := \pi(X_2, U) > 0 \quad a.s., \tag{2}$$

denote the selection probability, which allows for the observed responses Y_i and the covariates in the vector X_{1i} to be MAR for the same units i (i.e., $\delta_i^Y \delta_i^{X_1} = 0$) as well as for different units $i \neq j$ (i.e., $\delta_i^Y \delta_j^{X_1} = 0,$) in the sample $1 \leq i, j \leq n$.[2] Note that by the law of iterated expectations

$$\begin{aligned}
&E\left[\frac{\delta}{\pi(X_2, U)}\zeta(Y, \eta(X, \alpha_0(U)))\,\omega(X)\right] \\
&= E\left\{E\left[E\left(\frac{\delta}{\pi(X_2, U)}\zeta(Y, \eta(X, \alpha_0(U)))\,\omega(X)\,|Y, X, U\right)|U\right]\right\} \\
&= E\left\{E\left[\zeta(Y, \eta(X, \alpha_0(U)))\,\omega(X)\,|U\right]\right\},
\end{aligned}$$

which forms the basis for the estimators of this paper. Let

$$\alpha_0(U) = a_1 + a_2(U - u) := aW$$

denote the linear approximation of $\alpha_0(U)$ at the point u, where $a_1 = \alpha_0(u)$, $a_2 = h\partial\alpha_0(\cdot)/\partial u$, $W = [1, (U - u)/h]^T$, $h(n) := h$ is the bandwidth, and let $\widehat{\pi}(X_{2i}, U_i)$ denote an estimator for $\pi(X_{2i}, U_i)$, which can be either parametric or

[1] Note that the results of the paper are valid also for multivariate models, in which case Y and U are, say, $\mathbb{R}^m$-valued random vectors and X is an $\mathbb{R}^m \times \mathbb{R}^k$-valued random matrix; see Sect. 4 for an example.

[2] It is worth pointing out that the results of the paper could be easily modified to accommodate the cases where only the responses Y or only the covariates in X_1 are MAR, by changing the selection probability (2) and modifying the expressions appearing in the theorems of Sect. 3 accordingly.

nonparametric. For the former, let $\pi\left(X_{2i}, U_i, \gamma_0\right)$ denote a parametric specification for $\pi\left(X_{2i}, U_i\right)$ (for example, a logit or a probit model), where $\gamma_0 \in \Gamma$ is a vector of unknown finite-dimensional parameters, and let $\widehat{\gamma}$ denote a robust alternative to the maximum likelihood estimator such as the one suggested by Bianco and Yohai (1996) for the logistic regression; then, $\widehat{\pi}\left(X_{2i}, U_i\right) = \pi\left(X_{2i}, U_i, \widehat{\gamma}\right)$. For the latter

$$\widehat{\pi}\left(X_{2i}, U_i\right) = \frac{\sum_{j=1}^{n} \delta_j L_b\left(V_j - V_i\right) \omega\left(X_{2j}\right)}{\sum_{j=1}^{n} L_b\left(V_j - V_i\right) \omega\left(X_{2j}\right)},$$

where $\omega(\cdot)$ is real-valued function given in (1), $V_i = \left[X_{2i}^T, U_i\right]^T$ and $L_b(\cdot) = L(\cdot/b)/b^{k_2+1}$ is a product kernel function with bandwidth $b(n) := b$.

The IPW-based robust local M estimator for $\alpha_0(\cdot)$ evaluated at $U = u$ is $\widehat{a}_{\widehat{\pi}}^M = \left[\widehat{a}_{1\widehat{\pi}}^{M\,T}, \widehat{a}_{2\widehat{\pi}}^{M\,T}\right]^T$, where

$$\begin{aligned}\widehat{a}_{\widehat{\pi}}^M = \underset{a_1, a_2 \in \mathbb{R}^k}{\arg\min} \sum_{i=1}^{n} & \frac{\delta_i}{\widehat{\pi}\left(X_{2i}, U_i\right)} \zeta\left(Y_i, \eta\left(X_i, aW_i\right)\right) \\ & \omega\left(X_i\right) K_h\left(U_i - u\right), \end{aligned} \tag{3}$$

and $K_h(\cdot) = K(\cdot/h)/h$ is a kernel function with bandwidth $h(n) := h$. Alternatively, $\widehat{a}_{\widehat{\pi}}^M$ can be defined as the solution to the first-order conditions:

$$\sum_{i=1}^{n} \frac{\delta_i}{\widehat{\pi}\left(X_{2i}, U_i\right)} \frac{\partial \zeta\left(Y_i, \eta\left(X_i, \widehat{a}_{\widehat{\pi}}^M W_i\right)\right)}{\partial a} \omega\left(X_i\right) K_h\left(U_i - u\right) = 0.$$

The latter estimator suggests the second class of robust local estimators considered in this paper. Let

$$\mu\left(Y, \eta\left(X, \alpha(U)\right)\right) \omega(X) \tag{4}$$

denote an $\mathbb{R}^k$-valued vector of robust estimating equations. Estimating equations arise naturally in statistics, with the derivative of a quasi-likelihood (the quasi-score) (Wedderburn 1974) being a prominent example. Other important examples include generalized estimating equations of Liang and Zeger (1986), the variance function estimating equations of Carroll and Ruppert (1988) and the first-order conditions used in the Gauss–Newton method to solve nonlinear least squares problems. However, estimating equations are not robust to outlier and hence the use of their robust analog (4). The vector of robust estimating equations $\mu(\cdot)$ is such that $E\left[\mu_{G_\omega}\left(Y, \eta\left(X, \alpha(U)\right)\right) \omega(X)\right] = 0$ for a unique $\alpha(U) = \alpha_0(U)$, where $\mu_{G_\omega}(\cdot)$ is the centered robust estimating equations with the centering factor $G_\omega(\cdot)$ used to achieve Fisher consistency, see Sect. 4.1 for an example of $G_\omega(\cdot)$.

The IPW-based robust local Z estimator $\widehat{a}^Z$ for $\alpha_0(\cdot)$ evaluated at $U = u$ is defined as the solution to

$$\sum_{i=1}^{n} \frac{\delta_i}{\widehat{\pi}\left(X_{2i}, U_i\right)} \mu_{G_\omega}\left(Y_i, \eta\left(X_i, \widehat{a}^Z W_i\right)\right) \eta'\left(X_i, \widehat{a}^Z W_i\right) \omega\left(X_i\right) K_h\left(U_i - u\right) = 0, \tag{5}$$

where $\eta'(\cdot) = \partial \eta(\cdot) / \partial a$.

3 Asymptotic results

3.1 Estimation

Let $\mathcal{U}$ denote the support of U and, for simplicity of notation, let $\zeta(\cdot)\omega(\cdot) := \zeta_\omega(\cdot)$, $\mu_{G_\omega}(\cdot)\omega(\cdot) := \mu_\omega(\cdot)$, $H = diag[1, h] \otimes I_k$,

$$\begin{aligned}
&\frac{\partial \zeta_\omega(Y, \eta(X, aW))}{\partial a} = \frac{\partial \zeta_\omega(Y, \eta(X, aW))}{\partial \eta} \eta'(X, aW) := \zeta_{\omega 1}(Y, \eta(X, aW)), \\
&\frac{\partial^2 \zeta_\omega(Y, \eta(X, aW))}{(\partial a)^{\otimes 2}} = \frac{\partial^2 \zeta_\omega(Y, \eta(X, aW))}{\partial \eta^2} \eta'(X, aW)^{\otimes 2} \\
&\quad + \frac{\partial \zeta_\omega(Y, \eta(X, aW))}{\partial \eta} \eta''(X, aW) := \zeta_{\omega 2}(Y, \eta(X, aW)), \\
&\frac{\partial \left(\mu_\omega(Y, \eta(X, aW)) \eta'(X, aW)\right)}{\partial a} = \frac{\partial \mu_\omega(Y, \eta(X, aW))}{\partial \eta} \eta'(X, aW)^{\otimes 2} \\
&\quad + \mu_\omega(Y, \eta(X, aW)) \eta''(X, aW) := \mu_{\omega 1}(Y, \eta(X, aW)),
\end{aligned}$$

where $\eta''(\cdot) = \partial^2 \eta(\cdot) / (\partial a)^{\otimes 2}$.

Let $v_{1k} = \int u^k K(u)\, \mathrm{d}u$, $v_{2k} = \int u^k K^2(u)\, \mathrm{d}u$,

$$\begin{aligned}
\Gamma_{\zeta 0}(u) &= E\left[\zeta_{\omega 2}(Y, \eta(X, \alpha_0(U))) | U = u\right], \\
\Sigma_{\pi \zeta 0}(u) &= E\left[\frac{\zeta_{\omega 1}(Y, \eta(X, \alpha_0(U)))^{\otimes 2}}{\pi(X_2, U)} | U = u\right], \\
\Gamma_{\mu 0}(u) &= E\left[\mu_{\omega 1}(Y, \eta(X, \alpha_0(U))) | U = u\right], \\
\Sigma_{\pi \mu 0}(u) &= E\left[\frac{\left(\mu_\omega(Y, \eta(X, \alpha_0(U))) \eta'(X, \alpha_0(U))\right)^{\otimes 2}}{\pi(X_2, U)} | U = u\right],
\end{aligned}$$

$$\begin{aligned}
\Gamma_{\times 0}^{v_1}(u) &= \begin{bmatrix} 1 & v_{11} \\ v_{11} & v_{12} \end{bmatrix} \otimes \Gamma_{\times 0}(u) \text{ for } \times = \zeta \text{ or } \mu, \\
\Sigma_{\pi \times 0}^{v_2}(u) &= \begin{bmatrix} v_{20} & v_{21} \\ v_{21} & v_{22} \end{bmatrix} \otimes \Sigma_{\pi \times 0}(u) \text{ for} \times = \zeta \text{ or } \mu.
\end{aligned}$$

Theorem 1 *Under assumptions A1–A6 in the supplemental appendix*

$$(nh)^{1/2}\left(H\left(\widehat{a}_{\widehat{\pi}}^{M}-a_{0}\right)-\frac{h^{2}f\left(u\right)}{2\left(v_{12}-v_{11}^{2}\right)}\left[\begin{array}{c}\left(v_{12}^{2}-v_{11}v_{13}\right)\\ \left(v_{13}-v_{11}v_{12}\right)\end{array}\right]\otimes\frac{\partial^{2}\alpha_{0}\left(u\right)}{\partial u^{2}}\right)$$
$$\xrightarrow{d}N\left(0,\Gamma_{\zeta0}^{v_{1}}\left(u\right)^{-1}\frac{\Sigma_{\pi\zeta0}^{v_{2}}\left(u\right)}{f\left(u\right)}\Gamma_{\zeta0}^{v_{1}}\left(u\right)^{-1}\right).$$

Furthermore, if $K(\cdot)$ is symmetric

$$(nh)^{1/2}\left(\widehat{a}_{1\widehat{\pi}}^{M}-\alpha_{0}\left(u\right)-\frac{h^{2}f\left(u\right)v_{12}}{2}\frac{\partial^{2}\alpha_{0}\left(u\right)}{\partial u^{2}}\right)$$
$$\xrightarrow{d}N\left(0,\frac{\Gamma_{\zeta0}\left(u\right)^{-1}v_{20}\Sigma_{\pi\zeta0}\left(u\right)\Gamma_{\zeta0}\left(u\right)^{-1}}{f\left(u\right)}\right).$$

To reduce the computational cost of $\widehat{a}_{\widehat{\pi}}^{M}$, a one-step version of it is proposed. This procedure, which is effectively one iteration of the Raphson–Newton method, is appealing when the minimization of the loss function is difficult (or very time-consuming) to achieve to the desired degree of accuracy. If the initial estimator $\widehat{a}_{\widehat{\pi}}^{M}=\left[\widehat{a}_{1\widehat{\pi}}^{M\,T},\widehat{a}_{2\widehat{\pi}}^{M\,T}\right]^{T}$ is close enough to $\alpha_{0}\left(U\right)$ for $U\approx u$—see Assumption A7 in the supplemental appendix, then the estimator from applying one iteration will have the same asymptotic variance as that of the minimizer of the loss function. To be specific, the one-step IPW-based robust M local estimator has the form

$$\left[\begin{array}{c}\widetilde{a}_{1\widehat{\pi}}^{M}\\ \widetilde{a}_{2\widehat{\pi}}^{M}\end{array}\right]=\left[\begin{array}{c}\widehat{a}_{1\widehat{\pi}}^{M}\\ \widehat{a}_{2\widehat{\pi}}^{M}\end{array}\right]-\left[\sum_{i=1}^{n}\frac{\delta_{i}}{\widehat{\pi}_{i}}\zeta_{2\omega}\left(Y_{i},\eta\left(X_{i},\widehat{a}_{\widehat{\pi}}^{M}W_{i}\right)\right)K_{h}\left(U_{i}-u\right)\right]^{-1}$$
$$\times\sum_{i=1}^{n}\frac{\delta_{i}}{\widehat{\pi}_{i}}\zeta_{1\omega}\left(Y_{i},\eta\left(X_{i},\widehat{a}_{\widehat{\pi}}^{M}W_{i}\right)\right)\eta'\left(X_{i},\widehat{a}W_{i}\right)K_{h}\left(U_{i}-u\right). \quad (6)$$

Theorem 2 *Under the same assumptions of Theorem 1 and A7 in the supplemental appendix, the IPW-based one-step robust local M estimator given in (6) has the same asymptotic distribution as that given in Theorem 1. In particular, if $K(\cdot)$ is symmetric*

$$(nh)^{1/2}\left(\widetilde{a}_{1\widehat{\pi}}^{M}-\alpha_{0}\left(u\right)-\frac{h^{2}f\left(u\right)v_{12}}{2}\frac{\partial^{2}\alpha_{0}\left(u\right)}{\partial u^{2}}\right)$$
$$\xrightarrow{d}N\left(0,\frac{\Gamma_{\zeta0}\left(u\right)^{-1}v_{20}\Sigma_{\pi\zeta0}\left(u\right)\Gamma_{\zeta0}\left(u\right)^{-1}}{f\left(u\right)}\right).$$

The next theorem establishes the asymptotic normality of the IPW-based robust local Z estimator defined in (5).

Theorem 3 *Under assumptions A1–A3, A4–A6 in the supplemental appendix*

$$(nh)^{1/2}\left(H\left(\widehat{a}_{\widehat{\pi}}^{Z}-a_{0}\right)-\frac{h^{2}f\left(u\right)}{2\left(v_{12}-v_{11}^{2}\right)}\left[\begin{array}{c}\left(v_{12}^{2}-v_{11}v_{13}\right)\\ \left(v_{13}-v_{11}v_{12}\right)\end{array}\right]\otimes\frac{\partial^{2}\alpha_{0}\left(u\right)}{\partial u^{2}}\right)$$
$$\overset{d}{\rightarrow}N\left(0,\Gamma_{\mu0}^{v_{1}}\left(u\right)^{-1}\frac{\Sigma_{\pi\mu0}^{v_{2}}\left(u\right)}{f\left(u\right)}\Gamma_{\mu0}^{v_{1}}\left(u\right)^{-1}\right).$$

Furthermore, if $K(\cdot)$ is symmetric

$$(nh)^{1/2}\left(\widehat{a}_{1\widehat{\pi}}^{Z}-\alpha_{0}\left(u\right)-\frac{h^{2}f\left(u\right)v_{12}}{2}\frac{\partial^{2}\alpha_{0}\left(u\right)}{\partial u^{2}}\right)$$
$$\overset{d}{\rightarrow}N\left(0,\frac{\Gamma_{\mu0}\left(u\right)^{-1}v_{20}\Sigma_{\pi\mu0}\left(u\right)\Gamma_{\mu0}\left(u\right)^{-1}}{f\left(u\right)}\right).$$

As with the previous class of M estimators, it is possible to consider a one-step version of the IPW-based robust local Z estimator, which has the form

$$\left[\begin{array}{c}\widetilde{\alpha}_{1\widehat{\pi}}^{Z}\\ \widehat{a}_{2\widehat{\pi}}^{Z}\end{array}\right]=\left[\begin{array}{c}\widehat{a}_{1\widehat{\pi}}^{Z}\\ \widehat{a}_{2\widehat{\pi}}^{Z}\end{array}\right]-\left[\sum_{i=1}^{n}\frac{\delta_{i}}{\widehat{\pi}\left(X_{2i},U_{i}\right)}\mu_{1\omega}\left(Y_{i},\eta\left(X_{i},\widehat{a}^{Z}W_{i}\right)\right)K_{h}\left(U_{i}-u\right)\right]^{-1}$$
$$\times\sum_{i=1}^{n}\frac{\delta_{i}}{\widehat{\pi}\left(X_{2i},U_{i}\right)}\mu_{\omega}\left(Y_{i},\eta\left(X_{i},\widehat{a}^{Z}W_{i}\right)\right)K_{h}\left(U_{i}-u\right). \tag{7}$$

Theorem 4 *Under the same assumptions of Theorem 3 and A7 (with $\widehat{a}_{\widehat{\pi}}^{Z}$ replacing $\widehat{a}_{\widehat{\pi}}^{M}$), the IPW-based one-step version of the robust local Z estimator given in (7) has the same asymptotic distribution as that given in Theorem 3. In particular, if $K(\cdot)$ is symmetric*

$$(nh)^{1/2}\left(\widetilde{a}_{1\widehat{\pi}}^{Z}-\alpha_{0}\left(u\right)-\frac{h^{2}f\left(u\right)v_{12}}{2}\frac{\partial^{2}\alpha_{0}\left(u\right)}{\partial u^{2}}\right)$$
$$\overset{d}{\rightarrow}N\left(0,\frac{\Gamma_{\mu0}\left(u\right)^{-1}v_{20}\Sigma_{\pi\mu0}\left(u\right)\Gamma_{\mu0}\left(u\right)^{-1}}{f\left(u\right)}\right).$$

Remark 1 In the case where all the variables are observable, the resulting robust local M and Z estimators have the same asymptotic distributions as those given in Theorems 1–4 without the selection probability $\pi\left(X_{2},U\right)$.

Remark 2 In the case where all the X covariates are MAR, that is if the selection probability is $\Pr\left(\delta=1|Y,X,U\right)=\Pr\left(\delta=1|U\right):=\pi\left(U\right)$, the resulting IPW-based robust local M and Z estimators have the same asymptotic distributions as those given in Theorems 1–4 with $\Sigma_{\pi\times0}\left(u\right)=\Sigma_{\times0}\left(u\right)/\pi\left(u\right)$ and $\times=\zeta$ or μ.

Remark 3 The robust IPW estimation method of this paper can also be used to construct estimators for the unknown marginal mean μ_0 of the response Y. We first consider the case when only some of the observed responses Y_i are MAR, that is the selection probability is

$$\Pr\left(\delta^Y = 1|Y, X, U\right) = \Pr\left(\delta^Y = 1|X, U\right) := \pi\left(X, U\right) > 0 \quad a.s. \tag{8}$$

We consider two estimators: the first one is

$$\widehat{\mu}_{IPW} = \frac{1}{n}\sum_{i=1}^{n} \frac{\delta_i^Y Y_i}{\widehat{\pi}\left(X_i, U_i\right)},$$

whereas the second one is based on the assumption that $E\left(Y|X, U\right) = \eta\left(X, \alpha_0\left(U\right)\right)$ and is

$$\widehat{\mu}_{DR} = \frac{1}{n}\sum_{i=1}^{n} \frac{\delta_i^Y Y_i}{\widehat{\pi}\left(X_i, U_i\right)} + \frac{1}{n}\sum_{i=1}^{n}\left(1 - \frac{\delta_i^Y}{\widehat{\pi}\left(X_i, U_i\right)}\right)\eta\left(X_i, \widehat{\alpha}\left(U_i\right)\right),$$

where $\widehat{\pi}\left(X_i, U_i\right)$ is either the robust logit parametric or nonparametric estimate of the selection probability (8) and $\widehat{\alpha}\left(\cdot\right)$ can be either a robust M or Z estimator as discussed in Sect. 2. The first estimator is the standard IPW sample mean dating back to Horvitz and Thompson (1952); it is fully nonparametric in the sense that it does not include the additional information that the response Y is related to the covariates X and U through the function $\eta\left(\cdot\right)$ and is robust to the presence of outliers in the covariates X, but is not robust to a possibly misspecified parametric model $\pi\left(X_i, U_i, \gamma_0\right)$ for the selection probability (8). The second estimator is an imputation-type estimator, in the same spirit of the doubly robust estimators often used in the missing data literature, see, for example, Robins et al. (1994) and Scharfstein et al. (1999). By construction, it is robust to possible misspecification of the regression function $E\left(Y|X, U\right)$ or of the parametric model $\pi\left(X_i, U_i, \gamma_0\right)$ (i.e., it is doubly robust) but is sensitive to the presence of outliers in the covariates X, although this sensitivity is mitigated by the fact that $\widehat{\alpha}\left(\cdot\right)$ is a robust estimator. The following theorem establishes the asymptotic normality of $n^{1/2}\left(\widehat{\mu}_\bullet - \mu_0\right)$, where $\bullet$ is either IPW or DR.

Theorem 5 *Under Assumption A8*

$$\begin{aligned}
&n^{1/2}\left(\widehat{\mu}_\bullet^j - \mu_0\right) \xrightarrow{d} N\left(0, V_j\left(\mu_0\right)\right), \quad j = \text{“}NP\text{”}, \text{ or “}P_r\text{”}, \\
&V_{NP}\left(\mu_0\right) = E\left[\frac{\sigma^2\left(X, U\right)}{\pi\left(X, U\right)} + \left(E\left(Y|X, U\right) - \mu_0\right)^2\right], \\
&V_{P_r}\left(\mu_0\right) = V_{NP}\left(\mu_0\right) + V_{1P_r}\left(\gamma_0\right) - 2V_{2P_r}\left(\gamma_0\right),
\end{aligned}$$

where

$$
\begin{aligned}
V_{1P_r}(\gamma_0) &= E\left[\frac{E(Y|X,U)}{\pi(X,U,\gamma_0)}\frac{\partial\pi(X,U,\gamma_0)}{\partial\gamma^T}\right]\Omega(\gamma_0) \\
&\quad E\left[\frac{E(Y|X,U)}{\pi(X_2,U,\gamma_0)}\frac{\partial\pi(X,U,\gamma_0)}{\partial\gamma}\right], \\
V_{2P_r}(\gamma_0) &= E\left(E(Y|X,U)\,E\left(\frac{E(Y|X,U)}{\pi}\frac{\partial\pi}{\partial\gamma^T}\right)M^{-1}(\gamma_0)\,r_\omega(\gamma_0)\right), \\
\Omega(\gamma_0) &= M^{-1}(\gamma_0)\,Var(r_\omega(\gamma_0))\,M^{-1}(\gamma_0).
\end{aligned}
$$

Theorem 5 shows that the two proposed estimators are asymptotically equivalent. The asymptotic variance $V_{NP}(\mu_0)$ corresponds to the semiparametric efficiency bound of Hahn (1998); the asymptotic variance $V_{P_r}(\mu_0)$ will be smaller than $V_{NP}(\mu_0)$ if $V_{1P_r}(\gamma_0) \leq V_{2P_r}(\gamma_0)$. To this end, note that if the selection probability (8) was estimated by an ordinary parametric logit, the resulting asymptotic variance would be $V_P(\mu_0) = V_{NP}(\mu_0) - V_{1P}(\mu_0)$, where

$$
\begin{aligned}
V_{1P}(\mu_0) &= E\left[\frac{E(Y|X,U)}{\pi(X,U,\gamma_0)}\frac{\partial\pi(X,U,\gamma_0)}{\partial\gamma^T}\right]I(\gamma_0)^{-1} \\
&\quad E\left[\frac{E(Y|X,U)}{\pi(X_2,U,\gamma_0)}\frac{\partial\pi(X,U,\gamma_0)}{\partial\gamma}\right],
\end{aligned}
$$

and $I(\gamma_0)$ is the information matrix for a logit estimator, which implies that $\widehat{\mu}_\bullet^P$ would be more efficient than $\widehat{\mu}_\bullet^{NP}$. Thus, the closer (numerically) the influence function $M^{-1}(\gamma_0)\,r_\omega(\gamma_0)$ of the Bianco and Yohai (1996) estimator $\widehat{\gamma}$ is to that of the ordinary logit estimator, the more likely $\widehat{\mu}_\bullet^{P_r}$ will be more efficient than $\widehat{\mu}_\bullet^{NP}$.

We conclude this remark by briefly discussing the case where some of the observed covariates X_i, say X_{1i}, are also MAR. In this case, the IPW estimator can still be used, as it relies only on the observable covariates X_2 and U; on the other hand, the imputation estimator becomes unfeasible because of its dependence on the missing X_1. To obtain a feasible imputation estimator additional assumptions, such as specifying the joint distribution of X_1 and X_2, or the existence of an additional of set of covariates, say Z, that are related (parametrically or nonparametrically) to X_1, would be required.

3.2 Inference

The results of the previous section can be used to construct Wald statistics to test local statistical hypotheses about $\alpha(\cdot)$. To investigate the asymptotic properties of such statistics, we consider the following local hypothesis with a Pitman drift

$$
H_n : R\alpha(u) = r_0(u) + \gamma_n(u), \tag{9}
$$

where R is an $l \times k$ matrix of constants and $\gamma_n(\cdot)$ is a bounded continuous function that may depend on n. Let $S = [I_k, O_k]$ denote a selection matrix, where O_k is a $k \times k$ matrix of zeros, and let

$$W(u) = (nh)\left[\left(R\widetilde{\widehat{a}}^*_{1\widehat{\pi}} - r_0(u)\right)\right]^T \left(RS\widehat{\Gamma}^{v_1}_{\times}(u)^{-1}\widehat{\Sigma}^{v_2}_{\widehat{\pi}\times}(u)\widehat{\Gamma}^{v_1}_{\times}(u)^{-1}S^T R^T\right)^{-1} \times \left(R\widetilde{\widehat{a}}^*_{1\widehat{\pi}} - r_0(u)\right) \text{ for } \times = \zeta \text{ or } \mu,$$

denote the Wald statistic, where, for $* = M$ or Z, $\widetilde{\widehat{a}}^*_{1\widehat{\pi}}$ can be either the IPW-based robust local M or Z local estimator $\widehat{a}^*_{1\widehat{\pi}}$ or its one-step version $\widetilde{a}^*_{1\widehat{\pi}}$, $\widehat{\Gamma}^{v_1}_{\times}(\cdot)$ and $\widehat{\Sigma}^{v_2}_{\widehat{\pi}\times}(\cdot)$ are consistent estimators[3] of $\Gamma^{v_1}_{\times 0}(\cdot)$ and $\Sigma^{v_2}_{\pi\times 0}(\cdot)$, and $\widehat{\pi}(\cdot)$ is either the parametric or nonparametric estimator of $\pi(\cdot)$.

Proposition 1 *Under the assumptions of Theorems 1–4, if rank* $(R) = l$ $(l \leq k)$, *and* $nh^5 \to 0$, *then under* (9) (i) *for* $(nh)^{1/2}\gamma_n(u) \to \gamma(u) > 0$ *(for some* $\|\gamma(u)\| < \infty$*)*

$$W(u) \xrightarrow{d} \chi^2(\kappa, l),$$

where $\chi^2(\kappa, l)$ *is a noncentral Chi-squared distribution with l degrees of freedom and noncentrality parameter*

$$\kappa = f(u)\gamma(u)^T \left(RS\Gamma^{v_1}_{\times 0}(u)^{-1}\Sigma^{v_2}_{\pi\times 0}(u)\Gamma^{v_1}_{\times 0}(u)^{-1}S^T R^T\right)^{-1}\gamma(u) \quad (\times = \zeta \text{ or } \mu);$$

(ii) *for* $(nh)^{1/2}\gamma_n(u) \to \infty$,

$$W(u) \xrightarrow{p} \infty.$$

Proposition 1 shows that with undersmoothing the proposed test has power against local Pitman-type alternatives and is consistent against any fixed alternatives of the form $\gamma_n(\cdot) = \gamma(\cdot)$. Under the null hypothesis $H_0 : R\alpha(u) = r_0(u)$, the proposition can be used to construct confidence regions for $R\alpha(u)$ with nominal coverage $1-\alpha$, that is for $\Pr\left(\chi^2(l) \leq c_\alpha\right) = 1-\alpha$ and $C_\alpha(u) = \Pr\left(r_0(u) \,|\, W(u) \leq c_\alpha\right)$,

$$\Pr\left(r_0(u) \in C_\alpha(u)\right) = 1 - \alpha + o(1).$$

Note that in the case of $K(\cdot)$ being symmetric, the Wald statistic $W(u)$ simplifies to

$$W_s(u) = (nh)\left(R\widetilde{\widehat{a}}^*_{1\widehat{\pi}}(u) - r_0(u)\right)^T \left(R\widehat{\Gamma}_{\times}(u)^{-1}v_{20}\widehat{\Sigma}_{\widehat{\pi}\times}(u)\widehat{\Gamma}_{\times}(u)^{-1}R^T\right)^{-1} \times \left(R\widetilde{\widehat{a}}^*_{1\widehat{\pi}}(u) - r_0(u)\right). \tag{10}$$

[3] See the supplemental appendix for some examples of $\widehat{\Gamma}^{v_1}_{\times}(\cdot)$ and $\widehat{\Sigma}^{v_2}_{\widehat{\pi}\times}(\cdot)$.

Proposition 1 can also be used to test the important hypothesis of constancy of the varying coefficients $\alpha(\cdot)$, corresponding to

$$H_0 : \alpha_0(u) = \alpha_0. \tag{11}$$

The test can be implemented using the finite-dimensional analog of the IPW-based robust local M and Z estimators defined in (3) and (5), that is

$$\begin{aligned}
\widehat{\alpha}_{\widehat{\pi}}^{M} &= \arg\min_{\alpha\in A} \sum_{i=1}^{n} \frac{\delta_i}{\widehat{\pi}(X_{2i}, U_i)} \zeta(Y_i, \eta(X_i, \alpha))\, \omega(X_i), \\
\widehat{\alpha}_{\widehat{\pi}}^{Z} &= \sum_{i=1}^{n} \frac{\delta_i}{\widehat{\pi}(X_{2i}, U_i)} \mu\left(Y_i, \eta\left(X_i, \widehat{\alpha}^{Z}\right)\right) \eta'\left(X_i, \widehat{\alpha}^{Z}\right) \omega(X_i) = 0,
\end{aligned} \tag{12}$$

where A is a compact set and $\alpha_0 \in int(A)$. Let $\widehat{\alpha}_{\widehat{\pi}}^{*}$ $(* = M \text{ or } Z)$ denote either of the estimators defined in (12) and note that under the null hypothesis (11) and the assumption that $n^{1/2}\left(\widehat{\alpha}_{\widehat{\pi}}^{*} - \alpha_0\right) = O_p(1)$,

$$(nh)^{1/2}\left(\widetilde{a}_{1\widehat{\pi}}^{*} - \widehat{\alpha}_{\widehat{\pi}}^{*}\right) = (nh)^{1/2}\left(\widetilde{a}_{1\widehat{\pi}}^{*} - \alpha_0\right) + o_p(1),$$

hence by Proposition 1

$$\begin{aligned}
W_c(u) = (nh)\left(\widetilde{a}_{1\widehat{\pi}}^{*} - \alpha_0\right)^T &\left(S\widehat{\Gamma}_{\times}^{v_1}(u)^{-1}\, \widehat{\Sigma}_{\widehat{\pi}\times}^{v_2}(u)\, \widehat{\Gamma}_{\times}^{v_1}(u)^{-1} S^T\right)^{-1} \\
&\times \left(\widetilde{a}_{1\widehat{\pi}}^{*} - \alpha_0\right) \xrightarrow{d} \chi^2(p).
\end{aligned} \tag{13}$$

It is important to note that the test statistics $W(u)$, $W_s(u)$ and $W_c(u)$ are asymptotically valid at a single point u. To increase their power, one can consider them over a fixed range of values of u, say $\{u_j\}_{j=1}^{s}$, and use instead the test statistics $\max_j W(u_j)$ and $\max_j W_c(u_j)$ $(j = 1, \ldots, s)$, as the following proposition shows.

Proposition 2 *Under the assumptions of Proposition 1,* (i) *for* $(nh)^{1/2}\gamma_n(u_j) \to \gamma(u_j) > 0$ (*for some* $\|\gamma(u_j)\| < \infty$)

$$\begin{aligned}
\max_{1\le j\le s} W(u_j) &\xrightarrow{d} \max_j \chi_j^2(\kappa_j, l), \\
\max_{1\le j\le s} W_c(u_j) &\xrightarrow{d} \max_j \chi_j^2(\kappa_{jc}, l)
\end{aligned} \tag{14}$$

where

$$\begin{aligned}
\kappa_j &= f(u_j)\gamma(u_j)^T \left(RS\Gamma_{\times 0}^{v_1}(u_j)^{-1}\, \Sigma_{\pi\times}^{v_2}(u_j)\, \Gamma_{\times 0}^{v_1}(u)^{-1} S^T R^T\right)^{-1} \gamma(u_j), \\
\kappa_{jc} &= f(u_j)\gamma(u_j)^T \left(S\Gamma_{\times 0}^{v_1}(u_j)^{-1}\, \Sigma_{\pi\times 0}^{v_2}(u_j)\, \Gamma_{\times 0}^{v_1}(u)^{-1} S^T\right)^{-1} \gamma(u_j);
\end{aligned}$$

(ii) *for* $(nh)^{1/2}\gamma_n\left(u_j\right)\to\infty$

$$\max_{1\le j\le s} W\left(u_j\right),\quad \max_{1\le j\le s} W_c\left(u_j\right)\xrightarrow{p}\infty.$$

Note that the distribution of the test statistics in Proposition 2 is nonstandard, since it involves the maximum of s independent noncentral Chi-squared distributions. However, under the null hypothesis $R\alpha(u)=r_0(u)$, the test statistic is asymptotically distribution-free; hence, its distribution can be evaluated numerically or easily simulated.

Remark 4 As mentioned in Introduction, a robust distance statistic can also be used to test (9); however, the resulting statistic would not be asymptotically distribution-free as the following proposition shows for the simple null hypothesis $H_0:\alpha(u)=\alpha_0(u)$ and $K(\cdot)$ symmetric. Let

$$D=-2\sum_{i=1}^{n}\frac{\delta_i}{\widehat{\pi}\left(X_{2i},U_i\right)}\left(\zeta\left(Y_i,\eta\left(X_i,\widehat{a}^{M}_{1\widehat{\pi}}\right)\right)-\zeta\left(Y_i,\eta\left(X_i,\alpha_0(u)\right)\right)\right)$$
$$\omega\left(X_i\right)K_h\left(U_i-u\right).$$

Proposition 3 *Under the same assumptions of Proposition* 1

$$D\xrightarrow{d}\sum_{j=1}^{k}\lambda_j\chi^2_j(1),$$

where λ_j *are the eigenvalues of* $\Sigma_{\pi\zeta 0}(u)\,\Gamma_{\zeta 0}(u)^{-1}$.

4 Example and Monte Carlo

4.1 Robust generalized estimating equations (GEE) estimation

This section considers estimation of a varying coefficient GEE model for longitudinal data [see, for example, Liang and Zeger (1986) for GEE estimation with unknown finite-dimensional parameters]. Suppose that

$$E(Y|X)=\eta\left(X^T\alpha_0(U)\right),$$

where Y is an m-dimensional random vector (m is the number of subjects or clusters) and $\eta\left(X^T\alpha(U)\right)=\left[\eta\left(X_1^T\alpha(U_1)\right),\ldots,\eta\left(X_m^T\alpha(U_m)\right)\right]^T$. In this example, the robust estimating equations $\mu_\omega(\cdot)$ are

$$\begin{aligned}
&\frac{\delta}{\pi}\mu_{\omega}\left(Y,\eta\left(X,\alpha\left(U\right)\right)\right)\\
&\quad=\eta'\left(X^{T}\alpha\left(U\right)\right)^{T}V\left(\eta,\beta_{0}\right)^{-1}\Omega\Pi\left(\delta,X_{2},U\right)\\
&\qquad\times\left[\psi\left(A^{-1/2}\left(Y-\eta\left(X^{T}\alpha\left(U\right)\right)\right)\right)-G_{\omega}\left(Y,\eta\left(X^{T}\alpha\left(U\right)\right)\right)\right],
\end{aligned}$$

where $V\left(\eta,\beta_{0}\right)=R\left(\beta_{0}\right)A_{0}^{1/2}$, $R\left(\beta_{0}\right)$ is the working correlation matrix indexed by the q-dimensional unknown parameter β_{0}, $\psi\left(.\right)$ is a robust function such as the Huber function defined as

$$\psi_{c}\left(t\right)=\begin{cases}t & \text{if } |t|\leq c,\\ csign\left(t\right) & \text{if } |t|>c\end{cases}\tag{15}$$

with tuning constant c,

$$\begin{aligned}
A_{0}&=\phi_{0}\text{diag}\left(\text{Var}\left(\eta\left(X_{1}^{T}\alpha_{0}\left(U\right)\right)\right),\ldots,\text{Var}\left(\eta\left(X_{m}^{T}\alpha_{0}\left(U\right)\right)\right)\right),\\
\Omega&=\text{diag}\left(\omega\left(X_{1}\right),\ldots,\omega\left(X_{m}\right)\right),\\
\Pi\left(\delta,X_{2},U\right)&=\text{diag}\left(\delta_{1}/\pi\left(X_{12},U_{1}\right),\ldots,\delta_{m}/\pi\left(X_{m2},U_{m}\right)\right),
\end{aligned}$$

ϕ_{0} is the unknown dispersion parameter,

$$G_{\omega}\left(Y,\eta\left(X^{T}\alpha\left(u\right)\right)\right)=E\left[\psi\left(A^{-1/2}\left(Y-\eta\left(X^{T}\alpha\left(U\right)\right)\right)\right)\right]$$

is the correction factor used to achieve Fisher consistency and a monotone missing data pattern[4] is assumed, that is, for $\delta_{1}\geq\delta_{2}\geq\cdots\geq\delta_{m}$, $\delta_{1}=1$,

$$\pi\left(X_{k2},U_{k}\right)=\Pr\left(\delta_{k}=1|X_{k2},U_{k}\right)=\prod_{l=1}^{k}\Pr\left(\left(\delta_{l}=1|\delta_{l-1},X_{l2},U_{l}\right)\right).$$

Calculations show that

$$\begin{aligned}
&\mu_{\pi\omega1}\left(Y,\eta\left(X,aW\right)\right)\\
&\quad=\sum_{j=1}^{2k}\frac{\partial\eta'\left(Y,\eta\left(X^{T}aW\right)\right)}{\partial a_{j}}V\left(\eta,\beta\right)^{-1}\Omega\Pi\left(\delta,X_{2},U\right)\\
&\qquad\times\left[\psi\left(A^{-1/2}\left(Y-\eta\left(X^{T}aW\right)\right)\right)-G_{\omega}\left(Y,\eta\left(X^{T}aW\right)\right)\right]
\end{aligned}$$

[4] The monotone missing data pattern assumption is fairly common in missing data models for longitudinal studies, see, for example, Ibrahim and Molenberghs (2009) for a review. For nonmonotone missing data patters, the IPW estimation method of this paper is still valid; however, the estimation of the selection probabilities is substantially more challenging. One possible method is to use the randomized monotone missingness model proposed by Robins and Gill (1997), which is unfortunately quite complex to implement in practice and computationally intensive, see Lia et al. (2013) for further details.

$$
\begin{aligned}
&+\eta'\left(X^T aW\right)^T V(\eta,\beta)^{-1}\Omega\Pi(\delta,X_2,U)\\
&\times\left[\frac{\partial}{\partial a}\psi\left(A^{-1/2}\left(Y-\eta\left(X^T aW\right)\right)\right)-G'_\omega\left(Y,\eta\left(X^T aW\right)\right)\right],\\
&\Gamma_{\mu 0}(u)\\
&=E\left[\eta'\left(X^T\alpha_0(U)\right)^T V(\eta,\beta_0)^{-1}\Omega s_\alpha(X,U)\right.\\
&\left.\times\frac{\partial}{\partial\alpha}\psi\left(A^{-1/2}\left(Y-\eta\left(X^T\alpha_0(U)\right)\right)\right)|U=u\right]\\
&\Sigma_{\pi\mu 0}(u)\\
&=E\left\{\eta'\left(X^T\alpha_0(U)\right)^T V(\eta,\beta_0)^{-1}\Omega\Pi(\delta,X_2,U)^2\times\right.\\
&\left.-\left[\psi\left(A^{-1/2}\left(Y-\eta\left(X^T\alpha_0(U)\right)\right)\right)-G_\omega\left(Y,\eta\left(X^T\alpha_0(u)\right)\right)\right]|U=u\right\}^{\otimes 2},
\end{aligned}
$$

where $s_\alpha(X,U)=\partial\log f(Y|X,U)/\partial\alpha$ and $f(Y|X,U)$ is the joint conditional density of the response Y. Consistent estimators for $\Gamma_{\mu 0}(u)$ and $\Sigma_{\pi\mu 0}(u)$ can be found in the supplemental appendix.

4.2 Monte Carlo results

This section investigates the finite sample performance of the estimator and test statistic $\max_j W\left(u_j\right)$ given in Proposition 2 for the GEE model considered in the previous section using a varying coefficient logit regression

$$\operatorname{logit}(E(Y=1|X,U))=X_1\alpha_{10}(U)+X_2\alpha_{20}(U),$$

where Y is a three-dimensional binary response variable Y (i.e., $m=3$ is number of subjects or clusters), the covariates $X=[X_{k1},X_{k2}]^T$ $(k=1,2,3)$ are independently normally distributed with mean zero and unit variances, the three-dimensional covariate U is independent of X and uniformly distributed between 0 and 1 and $\alpha_0(U)=[\sin(\pi U/2),\cos(\pi U)]^T$. To generate the responses with an exchangeable covariance structure (with correlation coefficient set equal to 0.3), Parzen (2009) approach is used, in which a random effect is added to the marginal probability of success

$$m_{ki}=\frac{1}{1-\exp\left(X_{k1i}\alpha_{10}(U_{ki})+X_{k2i}\alpha_{20}(U_{ki})\right)}\quad(k=1,2,3).$$

The selection probabilities for subjects $k=2$ and $k=3$ are specified as

$$
\begin{aligned}
\Pr(\delta_2=1|X_{22},U_2,Y_1)&=\frac{\exp(\gamma_{10}+\gamma_{20}X_{22}+\gamma_{30}U_2+\gamma_{40}Y_1)}{1+\exp(\gamma_{10}+\gamma_{20}X_{22}+\gamma_{30}U_2+\gamma_{40}Y_1)},\\
\Pr(\delta_3=1|X_{32},U_3,Y_2,Y_1)&=\frac{\exp(\gamma_{10}+\gamma_{20}X_{32}+\gamma_{30}U_3+\gamma_{40}Y_1+\gamma_{50}Y_2)}{1+\exp(\gamma_{10}+\gamma_{20}X_{32}+\gamma_{30}U_3+\gamma_{40}Y_1+\gamma_{50}Y_2)}
\end{aligned}
\tag{16}
$$

with $\left[\gamma_{10}, \gamma_{20}, \gamma_{30}, \gamma_{40}, \gamma_{50}\right]^T = [1, 1.5, 0.3, 1, 0.7]^T$, which implies that the average percentages of missing are about 25% for $k = 2$ and 20% for $k = 3$. We consider three cases: the first one (case 0) has no missing values nor outliers; the second one (case C0) has missing values but no outliers; and finally the last case (case C3) has both missing values and outliers generated as

$$(C3)\ X_{k2j} \sim N(10, 1) \quad (k = 1, 2, 3; \quad j = 1, \ldots, 10),$$

that is in (C3) the first ten elements of the three covariates X_{k2} are outliers. The computation of $\widehat{a}_{\widehat{\pi}}^{Z}$ is carried out under the working independence assumption with $\phi_0 = 1$ using the Huber function (15) with $c = 1.2$ and the Newton–Raphson algorithm. In the simulations, the Epanechnikov kernel is used, whereas the weight function is

$$\omega(X) = \left(1 + \left\| S^{-1}(X - M) \right\|^2 / 2\right)^{-1/2}, \tag{17}$$

where M and S are robust location-scatter estimators such as the minimum covariance determinant estimator (MCD), see, for example, He et al. (2005). The MCD estimator is computed using the R routine `CovNAMcd`, which uses imputation to deal with missing observations. The bandwidth h is selected by a robust cross-validation procedure in which in the first step for a given h, $\widehat{t}_{-i} = \sum_{j \neq i}^{n} \Psi_{\widehat{\pi}c}\left(\widehat{t}_j, u\right)/n = 0$ and in the second step the robust bandwidth is chosen as $\widehat{h} = \arg\min_h \sum_{i=1}^{n} \delta_i \zeta_\omega\left(\widehat{t}_{-i}\right)/\widehat{\pi}(X_{2i}, U_i)$, that is $\widehat{h}$ is the minimizer of the robust quasi deviance. Note that the selection probabilities (16) are estimated only using the robust parametric logit estimator.

Table 1 reports the mean absolute bias (B)

$$B\left(\widehat{a}_{1kj}\right) = \frac{1}{n}\sum_{i=1}^{n} \left|\widehat{a}_{1kj}(U_i) - \alpha_{k0}(U_i)\right| \text{ for } k = 1, 2$$

and standard deviation (SD) of four different estimators of $\alpha_{k0}(\cdot)$: the robust local complete case estimator $\widehat{a}_{1kc}$—that is the estimator based on the sample where all the missing observations are dropped, the IPW local robust estimator $\widehat{a}_{1k\widehat{\pi}p}$ with robust logit estimation of $\pi(X_{2i}, U_i)$ and their nonrobust analogs computed with the Huber function (15) set to $c = \infty$, no correction term $G_\omega(\cdot)$ and an ordinary logit estimator for $\pi(X_{2i}, U_i)$.

Inference is based on the statistic $\max_{1 \leq j \leq s} W\left(u_j\right)$ (with $s = 5$) for the hypothesis

$$H_{\gamma k}: \begin{bmatrix} \alpha_1\left(u_{kj}\right) \\ \alpha_2\left(u_{kj}\right) \end{bmatrix} = \begin{bmatrix} \alpha_{10}\left(u_{kj}\right) \\ \alpha_{20}\left(u_{kj}\right) \end{bmatrix} + \gamma \begin{bmatrix} \widehat{\alpha}_{1\widehat{\pi}}^{Z} - \alpha_{10}\left(u_{kj}\right) \\ \widehat{\alpha}_{2\widehat{\pi}}^{Z} - \alpha_{20}\left(u_{kj}\right) \end{bmatrix} \quad (k = 1, 2, 3), \tag{18}$$

where $u_{kj} = [0.1, 0.3, 0.5, 0.7, 0.9]$, $\alpha_{10}\left(u_{kj}\right) = [0.156, 0.453, 0.707, 0.891, 0.987]$ and $\alpha_{20}\left(u_{kj}\right) = [0.951, 0.587, 0, 0.587, 0.951]$ $(k = 1, 2, 3)$ (that is, there are six parameters), $\widehat{\alpha}_{\widehat{\pi}}^{Z}$ is the parametric estimator defined in (12) and $\gamma \in \mathbb{R}$ index the departure from the null hypothesis (corresponding to $\gamma = 0$). The upper 10% and 5% critical values of the nonstandard distribution given in Proposition 2 are calculated

Table 1 Bias (B) and standard deviation (SD) for robust and nonrobust estimators in the GEE example

n	200		500		200		500	
	B	SD	B	SD	B	SD	B	SD
	$0^{(NR)}$				$0^{(R)}$			
$k=2$								
$\widehat{a}_{11c}$	.091	.327	.061	.163	.109	.357	.070	.175
$\widehat{a}_{11\widehat{\pi}p}$	–	–	–	–	–	–	–	–
$\widehat{a}_{12c}$	.103	.324	.069	.157	.109	.341	.074	.162
$\widehat{a}_{12\widehat{\pi}p}$	–	–	–	–	–	–	–	–
$k=3$								
$\widehat{a}_{11c}$	.109	.315	.077	.152	.119	.337	.082	.162
$\widehat{a}_{11\widehat{\pi}p}$	–	–	–	–	–	–	–	–
$\widehat{a}_{12c}$	.117	.310	.067	.145	.126	.326	.073	.156
$\widehat{a}_{12\widehat{\pi}p}$	–	–	–	–	–	–	–	–
	$\mathrm{C0}^{(NR)}$				$\mathrm{C0}^{(R)}$			
$k=2$								
$\widehat{a}_{11c}$	.139	.507	.087	.281	.145	.527	.092	.292
$\widehat{a}_{11\widehat{\pi}p}$	.110	.518	.072	.289	.120	.531	.074	.299
$\widehat{a}_{12c}$	.137	.499	.077	.256	.139	.521	.080	.273
$\widehat{a}_{12\widehat{\pi}p}$	.114	.510	.063	.265	.122	.527	.070	.281
$k=3$								
$\widehat{a}_{11c}$	.126	.495	.081	.269	.136	.502	.086	.279
$\widehat{a}_{11\widehat{\pi}p}$	.106	.510	.070	.277	.113	.524	.075	.288
$\widehat{a}_{12c}$	.130	.484	.080	.248	.132	.494	.081	.259
$\widehat{a}_{12\widehat{\pi}p}$	.110	.496	.066	.254	.118	.509	.069	.264
	$\mathrm{C3}^{(NR)}$				$\mathrm{C3}^{(R)}$			
$k=2$								
$\widehat{a}_{11c}$	.197	.628	.162	.456	.150	.538	.099	.301
$\widehat{a}_{11\widehat{\pi}p}$	.216	.657	.169	.470	.122	.543	.080	.310
$\widehat{a}_{12c}$	.193	.621	.162	.447	.143	.532	.084	.290
$\widehat{a}_{12\widehat{\pi}p}$	.204	.638	.159	.455	.124	.540	.074	.296
$k=3$								
$\widehat{a}_{11c}$	.199	.631	.145	.481	.146	.541	.090	.289
$\widehat{a}_{11\widehat{\pi}p}$	.214	.654	.153	.488	.128	.536	.080	.295
$\widehat{a}_{12c}$	.195	.625	.151	.470	.137	.533	.084	.273
$\widehat{a}_{12\widehat{\pi}p}$	.205	.640	.165	.483	.121	.531	.073	.278

NR nonrobust estimation, R robust estimation

Table 2 Finite sample size for robust and nonrobust tests in the GEE example

n	200		500		200		500	
	$0^{(NR)}$				$0^{(R)}$			
$\max_j W_{co}(u_j)$	.113	.061	.108	.057	.117	.065	.112	.062
	$C0^{(NR)}$				$C0^{(R)}$			
$\max_j W_{co}(u_j)$	.120	.065	.110	.060	.121	.068	.112	.061
$\max_j W_{\widehat{\pi}p}(u_j)$	.116	.062	.108	.058	.120	.065	.110	.059
	$C3^{(NR)}$				$C3^{(R)}$			
$\max_j W_{co}(u_j)$	.174	.125	.172	.120	.122	.068	.113	.062
$\max_j W_{\widehat{\pi}p}(u_j)$	.190	.138	.182	.121	.119	.063	.110	.059

NR nonrobust estimation, R robust estimation

using 10^5 simulations and are [11.307, 13.452] for $n = 200$ and [11.077, 13.168] for $n = 500$. Table 2 reports the finite sample size at the 0.10 and 0.05 nominal level of $\max_{1\leq j\leq 5} W(u_{kj})$ based on the four estimators and the three cases (0, C0 and C3) used in Table 1.

A full evaluation of the finite sample power of $\max_{1\leq j\leq 5} W(u_{kj})$ under (18) is not feasible as it would have to be calculated over the hypergrid $\gamma \times \cdots \times \gamma = [-1, -0.75, \ldots, 0, \ldots 0.75, 1]^4$ (6,561 evaluation points); therefore, Fig. 1 reports the contour plots at the level 0.45 of the size-adjusted finite sample power curves for the test of $H_{\gamma 2}$ (that is only for the second subject (or cluster)) over the grid $\gamma \times \gamma = [-1, -0.75, \ldots, 0, \ldots 0.75, 1] \times [-1, -0.75, \ldots, 0, \ldots 0.75, 1]$ for the C0 and C3 cases. Note that smaller contour plots indicate higher finite sample power.

The results of Tables 1 and 2 (together with those of Tables 3–6 in the supplemental appendix) can be summarized as follows: For estimation (Tables 1 and 3, 5 in the supplemental appendix), first, without outliers and missing observations (case 0) the nonrobust local estimators perform better than the robust ones both in terms of bias and standard deviation, with the bias and standard deviation up to, respectively, 16% and 8% smaller for the GEE example, 13% and 10% smaller for the Poisson regression example l and 7% and 11% smaller for the nonlinear regression example in the supplemental appendix. Note that for both estimators the bias and standard deviation decrease as the sample size increases, implying the validity of the asymptotic results of Sect. 3. Second, without outliers but with missing observations (case C0) the performance of the nonrobust and robust local estimators is fairly similar, with biases and standard deviations being, respectively, between 1% and 8% smaller and 2% and 6% (for the Poisson regression example—see Table 3 in the supplemental appendix, and the GEE example). Third, when outliers are present (case C3 and cases C1–C2 in the supplemental appendix) the robust estimators clearly outperform the nonrobust ones with biases being up to 60% smaller and standard deviations up to 50% smaller. Fourth, among the three robust local estimators, those based on the inverse probability weighting perform better in terms of bias (for the GEE example on average around

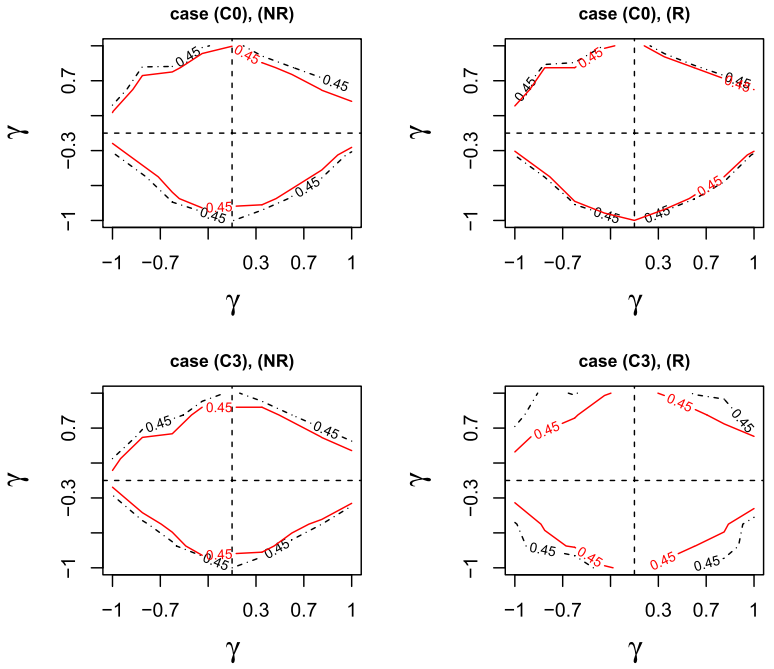

Fig. 1 Finite sample-adjusted power for the test statistic $\max_{1\leq j\leq 5} W\left(u_j\right)$; solid lines indicate $\max_{1\leq j\leq 5} W_{\widehat{\pi}p}\left(u_j\right)$ and dot dashed lines indicate $\max_{1\leq j\leq 5} W_{co}\left(u_j\right)$ for the GEE example

18% smaller, for the Poisson regression example on average around 9% smaller and for the nonlinear regression example on average around 10% smaller) but not in terms of standard deviation (for the GEE example on average around 2%, for the Poisson regression example on average around 1% and for the nonlinear regression example on average around 2%) than those based on the complete case analysis, which is not surprising because their asymptotic variance is affected by the estimated selection probabilities in the denominator. Note, however, that in terms of the mean-squared error (MSE) the local robust estimators have typically a smaller one than that of the nonrobust estimators, the exceptions being the GEE and the Poisson regression examples both with 200 observations, where the MSEs are, respectively, 0.272 and 0.279 and 0.312 and 0.315 (for the IPW estimator with nonparametric estimation of the selection probabilities). Finally, between the two inverse probability weighting estimators, those based on the parametric estimator of the selection probabilities seem to have an edge over those based on the nonparametric estimator in terms of both bias and standard deviation, which is again not surprising, given that the selection probabilities are estimated using the (robust) maximum likelihood estimator for a correctly specified parametric model. For inference (Tables 2 and 4 and 6 in the supplemental appendix), first, without outliers and missing observations (case 0) or without outliers but with missing observations (case C0), the tests based on the nonrobust local estimators are characterized by a slightly better (i.e., closer to the nominal level) size than that of the tests based on the robust local estimators (up to 3% smaller for the GEE-based

test with 200 observation). Second, when outliers are present (cases C1–C3) the size distortion of the tests based on the nonrobust local estimators worsen significantly (up to 30% larger size distortions in the GEE case with 200 observations), whereas that of the tests based on the robust local estimator remain similar to that of case C0. Third among the three test statistics, those based on the inverse probability weighting are more accurate (that is they have the smallest size distortion) with those based on the parametric estimator of the selection probabilities being the most accurate. Finally, Fig. 1 (combined with Figures 3 and 4 in the supplemental appendix) shows that in terms of finite sample power the tests based on the inverse probability weighting robust local estimators have typically larger power compared to those based on the complete case estimators both in the case of no outliers present (case C0) or with outliers (cases C2 and C3).

5 Empirical application

This section illustrates the applicability of the proposed estimation and testing methods by considering the New York air quality measurements data (from May to September 1973, available in the R package `datasets` which consists of 154 daily observations of mean ozone parts (per billion), solar radiations, wind speed (in mph) and temperature (in degrees F) and contains 37 missing ozone part observations and 7 missing solar radiation observations. The same data set was considered by Bianco and Spano (2019), who fitted an exponential growth regression model between the ozone parts and the temperature. Here, a linear varying coefficients specification is considered

$$Y = X^T \alpha_0 (U) + \varepsilon,$$

where Y represents the ozone parts, $X = [1, X_1, X_2]^T$ represent, respectively, the solar radiation and temperature, U is the wind speed and ε is a standard normal. The same Huber function with $c = 1.2$ and weight function $\omega(\cdot)$ as those given in (15) and (17) are used, with the computation of $\widehat{\alpha}_{\widehat{\pi}}(\cdot)$ based on the Newton–Raphson algorithm for the local estimating equations

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\delta_i}{\widehat{\pi}(X_{2i}, U_i)} \psi_c(\widehat{\varepsilon}_i) \begin{bmatrix} X_i \\ X_i (U_i - u)/h \end{bmatrix} \omega(X_i) K_h (U_i - u) = 0, \qquad (19)$$

where $\widehat{\varepsilon}_i = Y_i - X_i^T \widehat{a} W_i$.

Figure 2 shows the three varying coefficients $\widehat{\alpha}_j(u_i)$ estimated using (19) with the Bianco and Yohai (1996) robust logit estimator for the selection probabilities $\pi(X_{2i}, U_i)$ and their nonrobust analogs, together with their associated 95% confidence intervals.

The first estimated coefficient (intercept) represents the direct effect of the wind speed on the ozone parts and is a decreasing function of it. (The same decreasing relationship was found by Bianco and Spano (2019).) The other two estimated varying coefficients represent the combined effect on the ozone parts of the solar radiation and

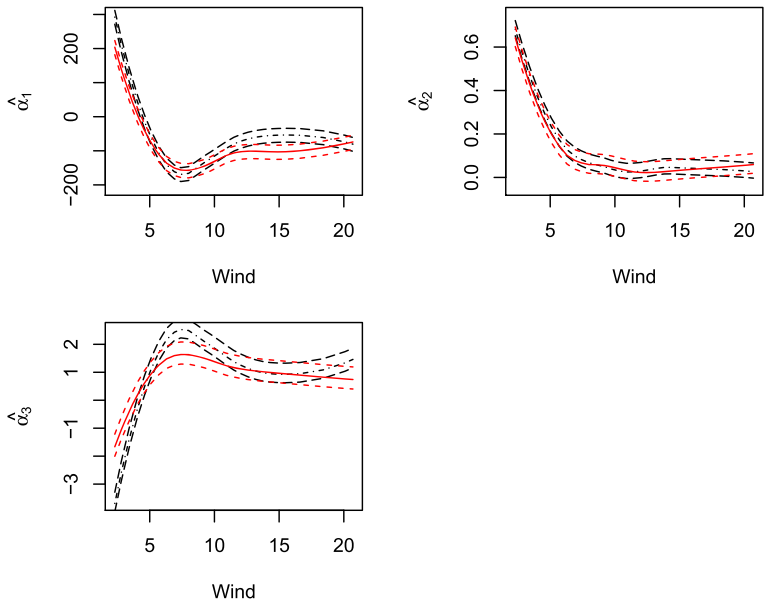

Fig. 2 Estimated varying coefficients for the New York air quality data. The solid line represents the robust local estimator with 95% confidence interval (dashed lines). The dot dashed line represents the nonrobust local estimator with 95% confidence interval (long dashed lines)

temperature with the wind speed. The former shows a clear decreasing relationship between the ozone parts and the wind speed combined with the solar radiations up to a speed of around 12 mph followed by a flatter relationship, whereas the latter shows an initial increasing relationship between the ozone parts and the wind speed and the temperature followed by a less clear relationship. Note that the robust local estimators are characterized by a more regular (and therefore easier to interpret) pattern. In terms of the mean effect (i.e., $\overline{\widehat{\alpha}_j} = \sum_{i=1}^{n} \widehat{\alpha}_j \left(U_i\right)/n$ for $j = 1, 2, 3$), the nonrobust estimation procedure has $\left[\overline{\widehat{\alpha}_1}, \overline{\widehat{\alpha}_2}, \overline{\widehat{\alpha}_3}\right]^T = [-0.94, 0.08, 1.56]^T$, whereas the robust one has $\left[\overline{\widehat{\alpha}_1}, \overline{\widehat{\alpha}_2}, \overline{\widehat{\alpha}_3}\right]^T = [-0.84, 0.07, 1.36]^T$. For inference, the null hypotheses of joint and individual constancy of the three varying coefficients are considered, that is

$$
\begin{aligned}
&H_0 : \left[\alpha_1\left(u_j\right) - \alpha_{10}, \alpha_2\left(u_j\right) - \alpha_{20}, \alpha_3\left(u_j\right) - \alpha_{30}\right]^T = 0^T \text{ and} \\
&H_0 : \left[\alpha_k\left(u_j\right) - \alpha_{k0}\right] = 0 \quad (k = 1, 2, 3)
\end{aligned}
$$

are tested using the $\max_{1 \le j \le 25} W_c\left(u_j\right)$ statistic (2) evaluated at 25 points u_j. The sample values of $\max_{1 \le j \le 25} W_c\left(u_j\right)$ for the null hypotheses of joint and individual constancy of the three varying coefficients are, respectively, 11.84, 7.84, 7.04 and 8.89 with corresponding p values of 0.028, 0.022, 0.028 and 0.016. Thus, the null hypotheses of joint and individual constancy of the varying coefficients are rejected at the 0.05 significance level.

6 Conclusions

This paper has considered robust local estimation and inference for a general class of varying coefficients models where some of the responses and covariates are missing at random and outliers might be present. The paper has proposed a general estimation method (and a computationally attractive one-step version of it) based on inverse probability of weighting of the selection probabilities that can be used to obtain both M and Z estimators and can accommodate longitudinal data. The paper has also proposed two Wald statistics that can be used to test hypotheses on the infinite-dimensional parameter, including that of constancy. A Monte Carlo study shows that the proposed estimators and Wald statistics perform well in finite samples, while two empirical applications illustrate their practical usefulness.

Acknowledgements I would like to thank an associate editor and a referee for very useful comments and suggestions.

References

Bianco A, Spano P (2019) Robust inference in nonlinear regression models. Test 28:369–398

Bianco A, Yohai V (1996) Robust estimation in the logistic regression model. In: Robust statistics, data analysis and computer intensive methods, Lecture Notes in Statistics 109, Springer, New York

Bianco A, Boente G, Martinez E (2006) Robust tests in semiparametric partly linear models. Scand J Stat 33:435–450

Bianco A, Boente G, Sombielle S (2011) Robust estimation for nonparametric generalized regression. Stat Probab Lett 81:1986–1994

Bianco A, Boente G, Gonzalez-Manteiga W, Perez A (2019) Plug-in marginal estimation under general regression model with missing responses and covariates. Test 28:106–146

Boente G, He X, Zhou J (2006) Robust estimates in generalized partially linear models. Ann Stat 34:2856–2878

Boente G, Gonzalez-Manteiga W, Perez-Gonzalez A (2009) Robust nonparametric estimation with missing data. J Stat Plan Inference 139:571–592

Bravo F (2015) Semiparametric estimation with missing covariates. J Multivar Anal 139:329–346

Bravo F, Jacho-Chavez D (2016) Semiparametric quasi-likelihood estimation with missing data. Commun Stat Theory Methods 46:1345–1369

Cai Z, Fan J, Li R (2000) Efficient estimation and inference for varying-coefficient models. J Am Stat Assoc 95:888–902

Cantoni E, Ronchetti E (2001) Robust inference for generalized linear models. J Am Stat Assoc 96:1022–1030

Carroll R, Ruppert D (1988) Transformation and weighting in regression. Chapman and Hall, London

Chen J, Fan J, Li K, Zhou H (2006) Local quasi-likelihood estimation with data missing at random. Statistica Sinica 16:1071–1100

Cheng P (1994) Nonparametric estimation of mean functionals with data missing at random. J Am Stat Assoc 89:81–87

Eubank R, Huang C, Munoz Maldonado Y, Wang N, Wang S, Buchanan R (2004) Smoothing spline estimation in varying-coefficient models. J R Stat Soc B 66:653–667

Fan J, Gijbels I (1996) Local polynomial modeling and its applications. Chapman and Hall, London

Fan J, Hu TC, Truong Y (1994) Robust non-parametric function estimation. Scand J Stat 21:433–446

Fan J, Heckman N, Wand M (1995) Local polynomial kernel regression for generalized linear models and quasilikelihood functions. J Am Stat Assoc 90:141–150
Fan J, Farmer M, Gijbels I (1998) Local maximum likelihood estimation and inference. J R Stat Soc B 60:591–608
Hahn J (1998) On the role of the propensity score in efficient semiparametric estimation of average treatment effects. Econometrica 66:315–331
Hastie T, Tibshirani R (1993) Varying-coefficient models (with discussion). J R Stat Soc 55:757–796
He X, Fung W, Zhu Z (2005) Robust estimation in generalized partial linear models for clustered data. J Am Stat Assoc 100:1176–1184
Horvitz D, Thompson D (1952) A generalization of sampling without replacement from a finite universe. J Am Stat Assoc 47:663–685
Hu T, Cui H (2010) Robust estimates in generalised varying coefficient partially linear models. J Nonparametric Stat 22:737–754
Huang J, Wu C, Zhou L (2002) Varying-coefficient models and basis function approximations for the analysis of repeated measurements. Biometrika 89:111–128
Ibrahim J, Molenberghs G (2009) Missing data methods in longitudinal studies: a review. Test 18:1–43
Kurum E, Li R, Senturk D, Wang Y (2013) Nonlinear varying coefficient models with application to a photosynthesis study. J Agric Biol Environ Stat 19:57–81
Lia L, Shen X, Li X, Robins J (2013) On weighting approaches for missing data. Stat Methods Med Res 22:14–30
Liang H (2008) Generalized partially linear models with missing covariates. J Multivar Anal 99:880–895
Liang K, Zeger S (1986) Longitudinal data analysis using generalised linear models. Biometrika 73:13–22
Liang H, Wang S, Robins J, Carroll R (2004) Estimation in partially linear models with missing covariates. J Am Stat Assoc 99:357–367
Parzen M (2009) A random effects model for simulating clustered binary data. Technical Report, Harvard University
Robins J, Gill R (1997) Non-response models for the analysis of non-monotone ignorable missing data. Stat Med 16:39–56
Robins J, Rotnitzky A (1995) Analysis of semiparametric models for repeated outcomes and missing data. J Am Stat Assoc 90:106–121
Robins J, Rotnitzky A, Zhao L (1994) Estimation of regression coefficients when some of the regressors are not always observed. J Am Stat Stat Assoc 89:846–866
Ruppert D, Wand P (1994) Multivariate locally weighted least squares regression. Ann Stat 22:1346–1370
Scharfstein D, Rotnitzky A, Robins J (1999) Adjusting for ignorable drop-out using semiparametric non-response models. J Am Stat Assoc 94:1096–1120
Verhasselt A (2014) Generalized varying coefficient models: a smooth variable selection technique. Statistica Sinica 24:147–171
Wedderburn R (1974) Quasi-likelihood functions, generalized linear models and the gauss-newton method. Biometrika 61:439–447