Efficiency for heteroscedastic regression with responses missing at random

Ursula U. Müller[1]

Department of Statistics
Texas A&M University
College Station, TX 77843-3143
USA

Anton Schick

Department of Mathematical Sciences
Binghamton University
Binghamton, NY 13902-6000
USA

Abstract

Heteroscedastic regression models are commonly used when the error variance differs across observations, i.e. when the error distribution depends on covariate values. We consider such models with responses possibly missing at random and show that functionals of the conditional distribution of the response given the covariates can be estimated efficiently using complete case analysis. We provide a formula for the efficient influence function in the general semiparametric heteroscedastic regression model and discuss special cases and examples.

Keywords: Efficiency transfer, partially linear, random coefficient, semiparametric regression, single index, transfer principle.
2010 MSC: 62G05, 62G08, 62G20, 62J02

1. Introduction

Missing data are common in applied research studies, and numerous methods are used to deal with this challenge, e.g. maximum likelihood estimation, single value imputation and multiple imputation. The simplest approach is probably *complete case analysis*, also known as *listwise deletion*, i.e. a statistical analysis that uses only cases that are completely observed, and ignores those with one or more missing entries.

Email addresses: uschi@stat.tamu.edu (Ursula U. Müller), anton@math.binghamton.edufn2 (Anton Schick)
URL: http://www.stat.tamu.edu/∼uschi/ (Ursula U. Müller), http://www.math.binghamton.edu/anton/ (Anton Schick)
[1]Corresponding author

Here we consider the situation when certain variables Y are possibly missing, but other variables X are always observed, i.e. our observations are independent copies $(X_1, \delta_1 Y_1, \delta_1), \dots, (X_n, \delta_n Y_n, \delta_n)$ of a base observation $(X, \delta Y, \delta)$, where the indicator variable δ equals 1 if Y is observed, and 0 otherwise. We assume that the variables Y are *missing at random* (MAR), i.e. the probability that Y is observed depends only on data X that are always available,

$$P(\delta = 1|X, Y) = P(\delta = 1|X) = \pi(X).$$

For further reading on missing data models and on methods for handling those data we refer to the books by Little and Rubin (2002), Tsiatis (2006), and Kim and Shao (2013). The MAR assumption makes it possible to draw valid statistical conclusions from the data without using auxiliary information from other sources. The (unknown) function $\pi(\cdot)$ is sometimes called the "propensity score". In order to exclude the case that no Y is observed we assume

$$E[\delta] = E[\pi(X)] > 0.$$

We will use a result by Müller and Schick (2016) (MS for short), the *efficiency transfer for regression models* with responses missing at random, and specialize it to semiparametric heteroscedastic regression. In MS the authors showed that under the above MAR assumption (the "MAR model") general functionals $\tau(Q)$ of the *conditional* distribution $Q \in \mathscr{Q}$ of Y given X can be estimated efficiently by an estimator that uses only the completely observed cases. More precisely, complete case versions of least dispersed regular estimators of such functionals in the full model remain least dispersed regular in the corresponding MAR model. Although the complete case analysis is known to suffer from bias problems, this is not the case when one is interested in functionals of this type.

To explain the results in MS more rigorously, let

$$T_n = t_n(X_1, Y_1, \dots, X_n, Y_n)$$

denote an efficient estimator in the "full model" with no missing data. The complete case version of T_n is

$$T_{n,c} = t_N(X_{i_1}, Y_{i_1}, \dots, X_{i_N}, Y_{i_N}),$$

with $N = \sum_{j=1}^n \delta_j$ denoting the number of complete cases and $\{i_1, \dots, i_N\} \subset \{1, \dots, n\}$ the indices of the complete cases. Write $G \in \mathscr{G}$ for the distribution of the covariate X and G_1 for the conditional distribution of X given $\delta = 1$. Then $G \otimes Q$ is the joint distribution of the pair (X, Y), while $G_1 \otimes Q$ is the joint distribution of (X, Y) given $\delta = 1$ under the MAR assumption. Suppose T_n is semiparametrically efficient (as laid out in Bickel et al., 1998) for estimating the functional $\tau(Q)$ in the full model, i.e. it is asymptotically linear with efficient influence function $\gamma_G(X, Y)$,

$$T_n = \tau(Q) + \frac{1}{n} \sum_{j=1}^n \gamma_G(X_j, Y_j) + o_P(n^{-1/2}). \tag{1.1}$$

Then its complete case version $T_{n,c}$ is efficient in the MAR model, with influence function $\delta\gamma_{G_1}(X,Y)/E[\delta]$,

$$T_{n,c} = \tau(Q) + \frac{1}{n}\sum_{j=1}^{n} \frac{\delta_j}{E[\delta]}\gamma_{G_1}(X_j, Y_j) + o_P(n^{-1/2}), \tag{1.2}$$

provided G_1 belongs to the model $\mathscr{G}$. Of course, the efficient influence function γ_G depends also on the transition distribution Q, and a more precise notation would be $\gamma_{G\otimes Q}$. As we only encounter changes in the marginal distribution, from G to G_1, we have chosen to suppress the dependence on Q for the sake of simplicity in notation, here and in the following. For example, we shall write E_G instead of $E_{G\otimes Q}$ to denote the expectation associated with the distribution $G \otimes Q$ and E_{G_1} for the expectation associated with $G_1 \otimes Q$.

The *efficiency transfer* was derived in two steps. The first step uses the *transfer principle* for complete case statistics by Koul, Müller and Schick (2012), which states in particular that the complete case version of an asymptotically linear estimator with (not necessarily efficient) influence function γ_G has expansion (1.2) in the MAR model if G_1 belongs to the model $\mathscr{G}$. The second step establishes that if γ_G is the efficient influence function of the original estimator then $\delta\gamma_{G_1}(X,Y)/E[\delta]$ is the efficient influence function of the complete case estimator in the MAR model. Together these two steps show that efficiency is preserved by complete case analysis. This finding simplifies the derivation of efficient estimators considerably: it suffices to find an efficient estimator in the full model and to verify that the assumptions for that estimator are also satisfied with G replaced by G_1.

Sections 2 and 3 of MS prove the efficiency transfer outlined above for general functionals of the conditional distribution and general random vectors (X,Y) that satisfy the MAR assumption. Section 4 of MS provides the specific details for the important case of a *homoscedastic* semiparametric regression model, i.e. when Y is a response variable that is linked to a covariate vector X via

$$Y = r(X, \vartheta, \varrho) + \varepsilon,$$

where the errors ε and X are independent. This covers parametric and nonparametric regression as important examples, namely when the regression function r is specified by a finite-dimensional parameter ϑ or by an infinite-dimensional parameter ϱ. The efficient influence for homoscedastic semiparametric regression models is provided in Theorem 4.1 of MS.

In this paper we work out the specific details of the efficiency transfer for the *heteroscedastic* semiparametric model, i.e. for the above model without the assumption that the error ε and the covariate X are independent. This situation is quite common in applications. We consider functionals of the regression parameters ϑ and ϱ only. This differs from the homoscedastic case, where we studied functionals of these regression parameters and the error density. The analogue would be to consider functionals that also include the conditional error density given the covariate, but functionals of this quantity are typically not estimable at the root-n rate.

This article is organized as follows. We present our main results, the efficient influence functions for general functionals of the conditional distribution Q in the full heteroscedastic regression model and under the MAR assumption in Section 2. In Example 1 and Example 2 we specialize the influence function for two functionals of interest: the finite-dimensional parameter and linear functionals of the regression function. In Section 3 we study parametric regression and nonparametric regression as special cases. Section 4 discusses four examples of semiparametric regression functions: partially linear regression, single index models, partially linear random coefficient models and partially linear single index models. In this section we only consider estimation of the finite parameter ϑ, which is of particular interest for applications. The proofs are in Section 5.

2. Main results

We consider the semiparametric heteroscedastic regression model

$$Y = r(X, \vartheta, \varrho) + \varepsilon, \quad E[\varepsilon|X] = 0,$$

where r is a known function of the covariate vector X, ϑ is a p-dimensional parameter, and ϱ is an infinite-dimensional parameter. In addition to our assumption that the errors are conditionally centered, we assume that ε given $X = x$ has a conditional density $y \mapsto f(y|x)$, which has a finite variance $\sigma^2(x)$ and finite Fisher information $J(x)$ for location. We also require that $\sigma^2(x)$ and $J(x)$ are bounded and bounded away from zero.

We write $\ell_f(\cdot|x) = -f'(\cdot|x)/f(\cdot|x)$ for the score function of the density $f(\cdot|x)$ and denote the distribution of X by G. Let F denote the transition distribution defined by $F(x, dy) = f(y|x)\, dy$. Note that $F(X, \cdot)$ is the conditional distribution of ε given X.

In this model the conditional distribution of Y given $X = x$ is

$$Q(x, dy) = Q_{\vartheta,\varrho,f}(x, dy) = f(y - r(x, \vartheta, \varrho)|x)\, dy.$$

For the efficiency considerations including notation we refer to MS Section 2. In MS the efficient influence function for a general functional of the conditional distribution Q of Y given X in the MAR model (without assuming a regression structure) is derived as a projection of any gradient of the functional onto the tangent space. MS follows the approach described in Chapter 3 of Bickel et al. (1998), which is suitable for estimating such general functionals; see also Tsiatis (2006) on estimating the finite-dimensional parameter. In this special case the more familiar method that projects the score function of the parameter onto the nuisance tangent space can be used to obtain the efficient score function, which is also Tsiatis' approach.

The MAR model contains the full model as a special case, with $\pi(\cdot) = 1$ and G in place of G_1. We therefore only need to determine the efficient influence function for the heteroscedastic regression model with fully observed data. This suffices thanks to the efficiency transfer, as explained in the introduction.

The tangent space in MS is written as an orthogonal sum of subspaces. Since we are interested in functionals of the conditional distribution Q, we limit our attention to the subspace that is relevant for estimating such functionals, namely the tangent space $\mathscr{V}(G)$ of the model $\mathscr{M}(G) = \{G \otimes R : R \in \mathscr{Q}\}$ at $G \otimes Q$. It consists of functions v that satisfy $\int v(x,y)Q(x,dy) = 0$ and $\int v^2(x,y)G(dx)Q(x,dy) < \infty$, and for which there is a sequence Q_{nv} in $\mathscr{Q}$ such that

$$\iint \Big(n^{1/2}(dQ_{nv}^{1/2}(x,\cdot) - dQ^{1/2}(x,\cdot)) - \frac{1}{2}v(x,\cdot)dQ^{1/2}(x,\cdot)\Big)^2 G(dx) \to 0. \quad (2.1)$$

As in Schick (1993), the set $\mathscr{V}(G)$ consists of the functions

$$v(X,Y) = [a^\top h(X) + b(X)]\ell_f(\varepsilon|X) + c(X,\varepsilon) \quad (2.2)$$

where a belongs to $\mathbb{R}^p$ and comes in because we do not know ϑ, h is the $L_2(G)$ derivative of $t \mapsto r(\cdot,t,\varrho)$ at ϑ, b belongs to some closed linear subspace $\mathscr{B}(G)$ of $L_2(G)$ and comes in because r involves ϱ, and c is a member of $\mathscr{C}(G)$, where

$$\mathscr{C}(G) = \Big\{c \in L_2(G\otimes F) : \int c(x,y)f(y|x)\,dy = \int yc(x,y)f(y|x)\,dy = 0\Big\}.$$

From now on we abbreviate $\vartheta + n^{-1/2}a$ by ϑ_{na} for $a \in \mathbb{R}^p$. To guarantee that (2.1) is met, we assume that for each b in $\mathscr{B}(G)$ there is a sequence ϱ_{nb} such that

$$\int \Big(n^{1/2}(r(x,\vartheta_{na},\varrho_{nb}) - r(x,\vartheta,\varrho)) - a^\top h(x) - b(x)\Big)^2 dG(x) = o(1) \quad (2.3)$$

holds for all $a \in \mathbb{R}^p$. Note that $\mathscr{C}(G)$ is the tangent space for the conditional error densities. Indeed, for each $c \in \mathscr{C}(G)$, there is a sequence f_{nc} of such densities such that

$$\iint \Big(n^{1/2}(f_{nc}^{1/2}(y|x) - f^{1/2}(y|x)) - (1/2)c(x,y)f^{1/2}(y|x)\Big)^2 dy\,dG(x) = o(1). \quad (2.4)$$

We then have (2.1) if we take $Q_{nv} = Q_{\vartheta_{na},\varrho_{nb},f_{nc}}$ with $v(X,Y)$ as in (2.2).

We are interested in estimating a functional

$$\tau(Q_{\vartheta,\varrho,f}) = \tau_0(\vartheta,\varrho)$$

of the regression parameters ϑ and ϱ. We assume that there is a sequence ϱ_{nb} such that (2.3) is satisfied and, additionally, that the functional τ_0 is differentiable, i.e.

$$n^{1/2}(\tau_0(\vartheta_{na},\varrho_{nb}) - \tau_0(\vartheta,\varrho)) \to a_*^\top a + \int b_G b\,dG \quad (2.5)$$

holds for all $a \in \mathbb{R}^p$, $b \in \mathscr{B}(G)$ and for some $a_* \in \mathbb{R}^p$ and $b_G \in \mathscr{B}(G)$.

We will provide the efficient influence function in Theorem 1 below. To describe it we need to introduce some additional notation. We let M denote the measure with density $1/\sigma^2$ with respect to G and let $\bar{\Pi}_G$ denote the projection operator onto $\mathscr{B}(G)$ in $L_2(M)$. Now introduce

$$h_G = (h_1 - \bar{\Pi}_G h_1, \dots, h_p - \bar{\Pi}_G h_p)^\top,$$

where $h_1, \dots, h_p$ are the components of the derivative h from above, and the matrix

$$H_G = \int h_G h_G^\top \, dM.$$

Theorem 1. *Suppose conditions (2.3)–(2.5) are satisfied and the matrix H_G is positive definite. Then the efficient influence function for estimating $\tau_0(\vartheta, \varrho)$ in model $\mathscr{M}(G)$ is*

$$\gamma_G(X, Y) = k_G(X) \frac{\varepsilon}{\sigma^2(X)}$$

with

$$k_G(X) = (a_* - \alpha_G)^\top H_G^{-1} h_G(X) + \bar{\Pi}_G(\sigma^2 b_G)(X)$$

and

$$\alpha_G = \int \bar{\Pi}_G(\sigma^2 b_G) h \, dM.$$

The proof of Theorem 1 is in Section 5. Note that γ_G is simpler than the corresponding influence function in Theorem 4.1 of MS for homoscedastic regression. Its structure is similar to that in the homoscedastic case when the error distribution happens to be normal. The structure of γ_G is explained by the fact that the tangent space $\mathscr{C}(G)$ for the conditional error density is much larger than the tangent space for the error density in the homoscedastic case. The latter consists of functions of the error variable only, while the former consists of all functions of ε and X that are orthogonal to $\{a(X) + b(X)\varepsilon : a, b \in L_2(G)\}$. Thus the efficient influence function needs to be of the form $b(X)\varepsilon$ for some $b \in L_2(G)$ which turns out to be k_G/σ^2.

In order to formulate the corresponding result for the MAR model, we need versions of (2.3)–(2.5) with G_1 in place of G, i.e. we need to replace $\mathscr{B}(G)$ by $\mathscr{B}(G_1)$, $\mathscr{C}(G)$ by $\mathscr{C}(G_1)$, $dG(x)$ by $dG_1(x)$ and b_G by b_{G_1}, and set

$$H_{G_1} = \int h_{G_1} h_{G_1}^\top \, dM_1 \quad \text{and} \quad \alpha_{G_1} = \int \bar{\Pi}_{G_1}(\sigma^2 b_{G_1}) h \, dM_1,$$

where M_1 is now the measure with density $1/\sigma^2$ with respect to G_1. Note that G_1 has density $\pi/E[\delta]$ with respect to G. If π is bounded away from zero then $L_2(G_1)$ equals $L_2(G)$ and (2.3)–(2.5) hold with G_1 replacing G, $\mathscr{B}(G_1) = \mathscr{B}(G)$, $\mathscr{C}(G_1) = \mathscr{C}(G)$, and b_{G_1} the projection of $E[\delta] b_G/\pi$ onto $\mathscr{B}(G_1)$ in $L_2(G_1)$. The latter follows from the fact that $\int b_G b \, dG = \int (E[\delta] b_G/\pi) b \, dG_1$ for b in $\mathscr{B}(G)$. If π is not bounded away from zero then $L_2(G)$ is a subset of $L_2(G_1)$, $\mathscr{B}(G_1)$ contains $\mathscr{B}(G)$, $\mathscr{C}(G_1)$ contains $\mathscr{C}(G)$, and (2.3) and (2.4) hold with $dG(x)$ replaced by $dG_1(x)$. Moreover, (2.5) is no longer guaranteed. Given Theorem 1 and the efficiency transfer explained in the introduction, we derive the following corollary for the MAR model.

Corollary 1. *Suppose that assumptions (2.3)–(2.5) are satisfied with G_1 replacing G and that the matrix H_{G_1} is positive definite. Then the influence function for the MAR model is*

$$\frac{\delta}{E[\delta]}\gamma_{G_1}(X,Y) = \frac{\delta}{E[\delta]}k_{G_1}(X)\frac{\varepsilon}{\sigma^2(X)}$$

with

$$k_{G_1}(X) = (a_* - \alpha_{G_1})^\top H_{G_1}^{-1} h_{G_1}(X) + \bar{\Pi}_{G_1}(\sigma^2 b_{G_1})(X),$$

for some $a_* \in \mathbb{R}^p$ and $b_{G_1} \in \mathscr{B}(G_1)$.

Remark 1. Constructing estimators with the above influence function typically requires an estimator of $\sigma^2(\cdot)$. Estimators of the variance function are unreliable if the dimension of the covariate vector is moderate to large. A common approach to overcome this difficulty is to model the variance function as a function of a low-dimensional transformation of X,

$$\sigma^2(X) = \tau(\xi(X)).$$

Here τ is some unknown function and ξ some known transformation. Examples are $\xi(X) = \|X\|$ and $\xi(X) = a^\top X$ for some known vector a; the latter covers the case when $\xi(X)$ denotes a fixed component of X. Working with dimension reducing transformations is useful for practical applications since they bypass the curse of dimensionality. However, under such a structural assumption the above influence function is typically no longer the efficient one. We will demonstrate this in Example 3 in Section 4.1.

The following two examples provide the efficient influence functions for two special cases of functionals $\tau(Q_{\vartheta,\varrho,f}) = \tau_0(\vartheta,\varrho)$, namely of the finite dimensional parameter ϑ and of integrated regression functions.

Example 1. *Estimating the finite-dimensional parameter.* The functional

$$\tau(Q_{\vartheta,\varrho,f}) = a_0^\top \vartheta,$$

for some $a_0 \in \mathbb{R}^p$, satisfies (2.5) with $a_* = a_0$ and $b_G = 0$. Hence the efficient influence function in model $\mathscr{M}(G)$ reduces to $a_0^\top H_G^{-1} h_G(X)\varepsilon/\sigma^2(X)$. Thus the influence function for estimating ϑ in this model is

$$H_G^{-1} h_G(X)\frac{\varepsilon}{\sigma^2(X)}.$$

The efficient influence function for estimating ϑ in the MAR model is therefore

$$\frac{\delta}{E[\delta]} H_{G_1}^{-1} h_{G_1}(X)\frac{\varepsilon}{\sigma^2(X)} \tag{2.6}$$

provided the matrix

$$H_{G_1} = E_{G_1}[h_{G_1}(X) h_{G_1}(X)^\top \sigma^{-2}(X)] = \int h_{G_1} h_{G_1}^\top \, dG_1$$

is positive definite.

Example 2. *Estimating a linear functional of the regression function.* We want to estimate

$$\tau(Q_{\vartheta,\varrho,f}) = \tau_0(\vartheta,\varrho) = \int w(x) r(x,\vartheta,\varrho)\,dx,$$

for example a certain area under the regression curve if w is an indicator. For this we assume that G has a density g and that w/g belongs to $L_2(G)$. With the aid of (2.3) we obtain that

$$n^{1/2}(\tau_0(\vartheta_{na},\varrho_{nb}) - \tau_0(\vartheta,\varrho)) = \int \frac{w(x)}{g(x)} n^{1/2}(r(x,\vartheta_{na},\varrho_{nb}) - r(x,\vartheta,\varrho))\,dG(x)$$

converges to

$$\int \frac{w}{g}(h^\top a + b)\,dG = \Big(\int w(x)h(x)\,dx\Big)^\top a + \int \frac{w}{g} b\,dG, \quad a \in \mathbb{R}^p, b \in \mathscr{B}(G).$$

From this we see that (2.5) is satisfied with $a_* = \int w(x)h(x)\,dx$ and b_G the projection of w/g onto $\mathscr{B}(G)$ in $L_2(G)$. In Section 3 we provide explicit formulas for the parametric and nonparametric heteroscedastic regression model, with and without missing responses.

3. Special cases: parametric and nonparametric regression

In this section we treat two important special cases where either the infinite-dimensional parameter ϱ or the finite-dimensional parameter ϑ is absent.

3.1. Parametric regression function

Consider the parametric regression model $Y = r_\vartheta(X) + \varepsilon$, where ε and X are as before and $r_\vartheta(X)$ is a regression function that is known except for an unknown parameter vector ϑ. One typically assumes that $r_t(x)$ is differentiable in t with gradient $\dot r_t(x)$. We will further require that

$$\int (r_{\vartheta+a} - r_\vartheta - a^\top \dot r_\vartheta)^2\,dG = o(|a|^2)$$

and that the matrix

$$R_G = \int \dot r_\vartheta \dot r_\vartheta^\top\,dG$$

is positive definite. This model does not involve ϱ, so we write $r_\vartheta(X)$ instead of $r(X,\vartheta,\varrho)$. Hence we have

$$Q(x,dy) = Q_{\vartheta,f}(x,dy) = f(y - r_\vartheta(x))\,dy,$$

and the functional of interest is

$$\tau(Q_\vartheta) = \tau_0(\vartheta).$$

For the efficiency considerations we need to assume τ_0 is differentiable in the sense that

$$n^{1/2}(\tau_0(\vartheta_{na}) - \tau_0(\vartheta)) \to a_*^\top a$$

for all $a \in \mathbb{R}^p$ and some $a_* \in \mathbb{R}^p$. This is (2.5) for our special case without the nonparametric part ϱ. The tangent associated with the perturbed version $Q_{\vartheta_{na}, f_{nc}}$ of $Q_{\vartheta,f}$ is

$$a^\top \dot{r}_\vartheta(X)\ell_f(\varepsilon|X) + c(X,\varepsilon);$$

see (2.1) ff. Here a belongs to $\mathbb{R}^p$ and c to $\mathscr{C}(G)$. So we have $h = \dot{r}_\vartheta$, $\mathscr{B}(G) = \{0\}$ and $b_G = 0$. This implies that α_G must be zero and that $h_G = h$ which is the gradient $\dot{r}_\vartheta$ in this model, i.e. $h(X) = \dot{r}_\vartheta(X)$. This complies with the linear regression model with regression function $r_\vartheta(X) = \vartheta^\top h(X)$.

The efficient influence function for model $\mathscr{M}(G)$ from Theorem 1 here simplifies to

$$a_*^\top \Big[\int \dot{r}_\vartheta \dot{r}_\vartheta^\top \frac{1}{\sigma^2}\, dG \Big]^{-1} \dot{r}_\vartheta(X) \frac{\varepsilon}{\sigma^2(X)}.$$

From this we conclude, as in Example 1, that the efficient influence function for estimating ϑ in the MAR model is

$$\frac{\delta}{E[\delta]} \Big[\int \dot{r}_\vartheta \dot{r}_\vartheta^\top \frac{1}{\sigma^2}\, dG_1 \Big]^{-1} \dot{r}_\vartheta(X) \frac{\varepsilon}{\sigma^2(X)}.$$

Combining Example 2 and the above yields the efficient influence function for $\int w(x) r_\vartheta(x)\, dx$ in the MAR model:

$$\frac{\delta}{E[\delta]} \int w(x) \dot{r}_\vartheta^\top(x)\, dx \Big[\int \dot{r}_\vartheta \dot{r}_\vartheta^\top \frac{1}{\sigma^2}\, dG_1 \Big]^{-1} \dot{r}_\vartheta(X) \frac{\varepsilon}{\sigma^2(X)}.$$

The efficient influence function for estimating ϑ in the MAR model was already derived by Müller and Van Keilegom (2012), who discuss a more general class of models defined by conditional constraints, with the parametric regression model as a special case. We refer to that paper for the construction of optimally weighted least squares estimators. Their construction involves estimators of the variance function σ^2. A related conditionally constrained model was studied by Robins and Rotnitzky (1995), who propose an inverse probability weighted estimating equation. These authors work with a parametric model for $\pi(X)$, so their statistical model is structurally different from our model since we do not assume a specific structure for $\pi(X)$. Robins, Rotnitzky and Zhao (1994) propose efficient estimators for ϑ in the parametric regression model, but also work with a parametric model for $\pi(X)$. Tsiatis (2006) considers parametric regression as well. He proposes a weighted estimating equation for ϑ, which is essentially the same approach as in Müller and Van Keilegom (2012), but he considers only the full model. Thanks to the efficiency transfer it is now clear that a complete case version of his estimator can be used in the MAR model to efficiently estimate ϑ (which is the approach of Müller and Van Keilegom).

In linear regression, i.e. $r_\vartheta(X) = \vartheta^\top h(X)$, the efficient influence function in the MAR model simplifies to

$$\frac{\delta}{E[\delta]}\Big[\int h\, h^\top \frac{1}{\sigma^2}\, dG_1\Big]^{-1} h(X)\frac{\varepsilon}{\sigma^2(X)}.$$

In the full model this is the influence function of the weighted least squares estimator

$$\hat\vartheta_{WLSE} = \Big(\frac{1}{n}\sum_{j=1}^n h(X_j)h(X_j)^\top \frac{1}{\sigma^2(X_j)}\Big)^{-1}\frac{1}{n}\sum_{j=1}^n h(X_j)Y_j\frac{1}{\sigma^2(X_j)}$$

with known variance function. Carroll (1982) has shown that a plug-in estimator, where σ^2 is replaced by a kernel estimator, is asymptotically equivalent to $\hat\vartheta_{WLSE}$, i.e. it has influence function

$$\Big[\int h\, h^\top \frac{1}{\sigma^2}\, dG\Big]^{-1} h(X)\frac{\varepsilon}{\sigma^2(X)}$$

and is therefore efficient. Similar results with different estimators of the variance functions were obtained by Schick (1987), Robinson (1987), Müller and Stadtmüller (1987). Complete case versions of these estimators are therefore efficient, as long as the matrix $\int h\, h^\top \sigma^{-2} dG_1$ is positive definite. Schick (2013) uses the empirical likelihood approach to obtain an efficient estimator in the MAR model that does not require estimating σ^2.

3.2. Nonparametric regression function

In this model we have no finite-dimensional parameter, i.e. the regression function is $r(X,\varrho) = \varrho(X)$ with ϱ smooth, and the functional of interest is

$$\tau(Q_{\varrho,f}) = \tau_0(\varrho).$$

In this model we have $a_* = 0$, $h = 0$, $\alpha_G = 0$ and $\mathscr{B}(G) = L_2(G)$. Thus the projection $\bar\Pi_G(\sigma^2 b_G)$ equals $\sigma^2 b_G$ and the influence function from Theorem 1 for model $\mathscr{M}(G)$ reduces to

$$b_G(X)\varepsilon.$$

Therefore, by the efficiency transfer, the efficient influence function in the MAR model is

$$\frac{\delta}{E[\delta]} b_{G_1}(X)\varepsilon.$$

The functional $\int w(x)\varrho(x)\, dx$ can be treated as a special case with $b_{G_1} = w/g_1$. As in Example 2 we again assume that w/g_1 belongs to $L_2(G_1)$.

A candidate for an efficient estimator in the full model is the plug-in estimator $\int w(x)\hat\varrho(x)\, dx$, where $\hat\varrho$ is a nonparametric estimator of the regression function ϱ, such as a kernel estimator or a locally linear smoother. The proof for a locally linear smoother can be carried out along the lines of the proof of Lemma 3.5 in Müller, Schick and Wefelmeyer (2007).

4. Examples of semiparametric regression functions

In this section we discuss some specific semiparametric regression functions and describe the efficient influence function for estimating the finite-dimensional parameter. In the examples below, the set $\mathscr{B}(G)$ will consist of all functions $b(X)$ of the form

$$b(X) = k(W)^\top Z$$

where $Z = t(X)$ and $W = s(X)$ are fixed measurable functions of X into $\mathbb{R}^m$ and k is a measurable function into $\mathbb{R}^m$ that varies subject to b belonging to $L_2(G)$. In this case, $h_G(X) = h(X) - \bar{\Pi}_G[h(X)]$ is of the form

$$h_G(X) = h(X) - E_G[\sigma^{-2}(X)h(X)Z^\top|W](E_G[\sigma^{-2}(X)ZZ^\top|W])^{-1}Z \qquad (4.1)$$

provided the random matrix

$$E_G[\sigma^{-2}(X)ZZ^\top|W]$$

is well defined and almost surely invertible. The matrix H_G can only be singular if there is a unit vector u and a function k such that $P(u^\top h(X) = k(W)^\top Z) = 1$.

4.1. Partially linear regression

In the partially linear regression model the covariate vector X is of the form $(U^\top, V^\top)^\top$ and the regression function is

$$r(X, \vartheta, \varrho) = \vartheta^\top U + \varrho(V),$$

with ϑ a vector of the same dimension as U and ϱ a smooth function. We assume that $E_G[\|U\|^2]$ is finite. To identify ϑ, we require, as in the homoscedastic case, that

$$A(G) = E_G[(U - E_G[U|V])(U - E_G[U|V])^\top]$$

is positive definite.

Here the differentiability assumption (2.3) on the regression function holds with $h(X) = U$ and $b(X)$ of the form $k(V)$ with k in $L_2(\varGamma)$ and $\varGamma$ the distribution of V. Thus (4.1), applied with $W = V$ and $Z = 1$, implies $h_G(X) = U - \nu_G(V)$ with

$$\nu_G(V) = \frac{1}{E_G[\sigma^{-2}(X)|V]} E_G[\sigma^{-2}(X)U|V].$$

Since $A(G)$ is positive definite, so is

$$H_G = E_G\Big[(U - \nu_G(V))(U - \nu_G(V))^\top \frac{1}{\sigma^2(X)}\Big].$$

Then the efficient influence function for estimating ϑ is

$$\gamma_G(X, Y) = H_G^{-1}(U - \nu_G(V)) \frac{\varepsilon}{\sigma^2(X)}.$$

This result was obtained in Ma, Chiou and Wang (2006).

We only required that $A(G)$ is positive definite. Thus, if $A(G_1)$ is positive definite, then the efficient influence function for the MAR model is

$$\frac{\delta}{E[\delta]} H_{G_1}^{-1}(U - \nu_{G_1}(V)) \frac{\varepsilon}{\sigma^2(X)},$$

with

$$\nu_{G_1}(V) = \frac{1}{E_{G_1}[\sigma^{-2}(X)|V]} E_{G_1}[\sigma^{-2}(X)U|V]$$

and

$$H_{G_1} = E_{G_1}\Big[(U - \nu_{G_1}(V))(U - \nu_{G_1}(V))^\top \frac{1}{\sigma^2(X)}\Big].$$

Estimators with influence function $\gamma_G(X, Y)$ have been constructed by Schick (1996) and Ma et al. (2006). The latter authors do so under the assumption $\sigma^2(X) = \tau(\xi(X))$ for some known transformation ξ. Since this is an additional structural assumption, the estimator is typically no longer efficient. This can be seen by means of the following example.

Example 3. *Dimension reducing transformations and efficiency.* Consider the common situation where the variance depends on just one covariate, say on V, i.e. the transformation $\xi(X)$ equals V. For simplicity we assume that U and V have dimension one. Then the variance is $\sigma^2(X) = \tau(V)$ and $h_G(X)$ simplifies to $h_G(X) = U - m(V)$ with $m(V)$ short for $m_G(V) = E_G(U|V)$. Under this structural assumption on the variance function, we need to modify $\mathscr{C}(G)$ to incorporate this information. In view of the identity $E[\varepsilon^2|X] = \tau(V)$, we have

$$E_G[\varepsilon^2(z(X) - E_G[z(X)|V])] = 0, \quad z \in L_2(G),$$

and thus need to impose the constraints

$$E_G[c(X,\varepsilon)\varepsilon^2(z(X) - E_G[z(X)|V])] = 0, \quad z \in L_2(G),$$

on members c of $\mathscr{C}(G)$. The tangents are now of the form

$$[aU + b(V)]\ell_f(\varepsilon|X) + c(X,\varepsilon),$$

where $a \in \mathbb{R}$, $b \in L_2(\Gamma)$ and $c \in \bar{\mathscr{C}}(G)$, which are the members of $\mathscr{C}(G)$ that also satisfy the additional constraints. For the case where the third and fourth conditional moments of ε given X are also functions of V only, i.e. $E_G[\varepsilon^k|X] = \mu_k(V)$, $k = 3, 4$, we were able to get an explicit form for the influence function for estimating ϑ,

$$\bar{\gamma}_G(X,Y) = \frac{U - m(V)}{E_G[(U - m(V))^2 w(V)]}\Big[\frac{\varepsilon}{\tau(V)} - \frac{\mu_3(V)}{\tau(V)\Delta(V)}\Big(\varepsilon^2 - \tau(V) - \frac{\mu_3(V)}{\tau(V)}\varepsilon\Big)\Big],$$

where

$$\Delta(V) = \mu_4(V) - \tau^2(V) - \frac{\mu_3^2(V)}{\tau(V)} = E_G\Big[\Big(\varepsilon^2 - \tau(V) - \frac{\mu_3(V)}{\tau(V)}\varepsilon\Big)^2\Big|X\Big]$$

and

$$w(V) = \frac{1}{\tau(V)} + \frac{\mu_3^2(V)}{\tau^2(V)\Delta(V)};$$

see Section 5 for the derivation. This influence function equals our influence function

$$\gamma_G(X,Y) = \frac{U - m(V)}{E_G[(U - m(V))^2/\tau(V)]} \frac{\varepsilon}{\tau(V)}$$

for the original model only if $\mu_3(V)$ is zero. The asymptotic variance of an estimator with influence function γ_G is

$$\frac{1}{E_G[(U - m(V))^2/\tau(V)]},$$

and the asymptotic variance of an estimator with influence function $\bar{\gamma}_G$ is

$$\frac{1}{E_G[(U - m(V))^2 w(V)]},$$

which is smaller than the previous variance, unless $\mu_3(V) = 0$.

4.2. Single index model

We now consider a single index model with

$$r(X, \vartheta, \varrho) = \varrho(U + \vartheta^\top V)$$

and $X = (U, V^\top)^\top$, where U is one-dimensional, V is p-dimensional, and ϱ is a smooth function that is not constant. To identify the p-dimensional parameter ϑ we require that the matrix $E_G[XX^\top]$ is positive definite. Here one verifies that (2.3) holds with $h(X)$ equal to $\varrho'(U+\vartheta^\top V)V$ and $b(X)$ of the form $k(U+\vartheta^\top V)$ for some k in $L_2(\Gamma)$, where Γ is the distribution of the index $U + \vartheta^\top V$. It follows from (4.1) applied with $Z = 1$ and $W = U + \vartheta^\top V$, that $h_G(X)$ is given by

$$h_G(X) = \varrho'(W)(V - \mu_G(W))$$

with

$$\mu_G(W) = \frac{1}{E_G[\sigma^{-2}(X)|W]} E_G[\sigma^{-2}(X)V|W]\Big).$$

For the construction of estimators in single index models, we refer to Cui, Härdle and Zhu (2011), who consider an extended class of single index models and provide many references. Their model class covers the homoscedastic single index, i.e. the model specified above but with constant variance, which is typically considered in the literature; see, for example, Powell, Stock and Stoker (1989), and Härdle, Hall and Ichimura (1993). Cui et al. model heteroscedasticity by assuming that the conditional variance is a known function of the mean function. Hence their model is different from that above, since it has further structural assumptions. For the construction of efficient estimators it will certainly make sense to incorporate some dimension reducing transformation to estimate the variance function. However, this will have an impact on the form of the influence function, which will be more complicated than the one considered here. Using (2.6), the efficient influence function for estimating ϑ in the MAR model is

$$\frac{\delta}{E[\delta]} H_{G_1}^{-1} h_{G_1}(X) \frac{\varepsilon}{\sigma^2(X)}$$

with

$$h_{G_1}(X) = \varrho'(W)\Big(V - \tau(G_1) E_{G_1}[\sigma^{-2}(X) V | W]\Big)$$

and

$$H_{G_1} = E_{G_1}[h_{G_1}(X) h_{G_1}(X)^\top \sigma^{-2}(X)].$$

This requires that the matrices H_{G_1} and $E_{G_1}[XX^\top]$ are positive definite.

4.3. Partially linear random coefficient model

Another useful flexible semiparametric model is the partially linear random coefficient model. Here we consider the version with regression function

$$r(X, \vartheta, \varrho) = \vartheta^\top U + \varrho(T)^\top V$$

and covariate vector $X = (T, U^\top, V^\top)^\top$. We assume that U and ϑ and also V and $\varrho(T)$ have matching dimensions, and that ϱ is a smooth function. To identify ϑ we have to rule out that for any unit vector v, $v^\top U$ equals almost surely $k(T)^\top V$ for some function k. To identify ϱ we impose that the eigenvalues of the random matrix $D_G(T) = E_G[VV^\top | T]$ fall into a compact subset of $(0, \infty)$. We also require that $\|U\|$ has a finite second moment. Then (2.3) holds true with $h(X) = U$ and $b(X)$ of the form $k(T)^\top V$ with $E_G[\|k(T)\|^2]$ finite.

From (4.1) applied with $W = T$ and $Z = V$ we obtain

$$h_G(X) = U - E_G[\sigma^{-2}(X) U V^\top | T](E_G[\sigma^{-2}(X) V V^\top | T)^{-1} V.$$

As before, we can insert this into formula (2.6) to obtain the efficient influence function for estimating ϑ in the MAR model.

There is not much literature on efficient estimation in the partially linear random coefficient model. The most pertinent reference seems to be Long, Ouyang and Shang (2013), in which an efficient sequential estimation method involving kernel estimators for the variance function is suggested. An early paper treating this model is Ahmad, Leelahanon and Li (2005), who appear to be the first to propose efficient (series) estimators. However, these authors demonstrate efficiency for the special case when $\sigma^2(\cdot)$ is constant, the "conditionally homoscedastic" model, which is structurally different from our model.

4.4. Partially linear single index model

In this model the regression function is of the form

$$r(X, \vartheta, \varrho) = \vartheta_1^\top U + \varrho(T + \vartheta_2^\top V),$$

with covariate vector $X = (T, U^\top, V^\top)^\top$, where U and V have the same dimensions as the components ϑ_1 and ϑ_2 of the partitioned parameter vector $\vartheta = (\vartheta_1^\top, \vartheta_2^\top)^\top$, respectively. Here we verify (2.3) with

$$h(X) = \begin{pmatrix} U \\ \varrho'(T + \vartheta_2^\top V)V \end{pmatrix}$$

and with $b(X)$ of the form $k(T + \vartheta_2^\top V)$. Thus, (4.1) applied with $W = T + \vartheta_2^\top V$ and $Z = 1$ yields

$$\begin{aligned} h_G(X) &= h(X) - \tau_G(W) E_G[\sigma^{-2}(X)h(X)|W] \\ &= \begin{pmatrix} U - \tau_G(W)E_G[\sigma^{-2}(X)U|W] \\ \varrho'(W)\big(V - \tau_G(W)E_G[\sigma^{-2}(X)V|W]\big) \end{pmatrix} \end{aligned}$$

with $\tau_G(W) = 1/E_G[\sigma^{-2}(X)|W]$. Again we can use (2.6) to obtain the efficient influence function for estimating ϑ in the MAR model. There are not many articles that consider the partially linear single index model with heteroscedasticity to which we can refer here. Ma and Zhu (2013) discuss efficiency and provide the same efficient influence function we do. However, as in Ma et al. (2006), they use a dimension reducing transformation to make the estimation of the variance function feasible. This is a structural constraint that typically changes the efficient influence function; see Example 3 in Section 4.1. Lai, Wang and Zhou (2014) also refer to Ma et al. (2006) and propose an estimation approach that is supposed to be efficient, but also involves a dimension reducing transformation as in Ma et al. (2006).

5. Proofs

Here we give the proof of Theorem 1 and provide further details for Example 3 in Section 4.1. For the proof of Theorem 1 we will repeatedly use the fact that

$$\int \ell_f(y|x) f(y|x)\, dy = 0, \quad \int y \ell_f(y|x) f(y|x)\, dy = 1, \tag{5.1}$$

and

$$\int y^2 \ell_f(y|x) f(y|x)\, dy = 0 \tag{5.2}$$

hold for all x.

Proof of Theorem 1. We need to show that $\gamma_G(X, Y)$ belongs to the tangent space $\mathscr{V}(G)$ and is a gradient. The latter means that

$$E_G[\gamma_G(X,Y)\{(a^\top h + b(X))\ell_f(\varepsilon|X) + c(X,\varepsilon)\}] = a_*^\top a + \int b_G b\, dG \tag{5.3}$$

must hold for all $a \in \mathbb{R}^p$, $b \in \mathscr{B}(G)$, and $c \in \mathscr{C}(G)$.

Let $K = \{a^\top h + b : a \in \mathbb{R}^p, b \in \mathscr{B}(G)\}$. Then the tangent space is

$$\mathscr{V}(G) = \{k(X)\ell_f(\varepsilon|X) + c(X,\varepsilon) : k \in K, c \in \mathscr{C}(G)\}.$$

Using (5.1) we see that the map $(x,y) \mapsto u(x)\ell_f(y|x) - y/\sigma^2(x)$ belongs to $\mathscr{C}(G)$ for every u in $L_2(G)$. This and the fact that k_G belongs to K show that

$$\gamma_G(X,Y) = k_G(X)\frac{\varepsilon}{\sigma^2(X)} = k_G(X)\ell_f(\varepsilon|X) - k_G(X)\Big(\ell_f(\varepsilon|X) - \frac{\varepsilon}{\sigma^2(X)}\Big)$$

is a tangent. It is easy to see that $\gamma_G(X,Y)$ is orthogonal to $c(X,\varepsilon)$ for every c in $\mathscr{C}(G)$, i.e.

$$E_G[\gamma_G(X,Y)c(X,\varepsilon)] = 0.$$

Using (5.1), we obtain

$$E_G[\gamma_G(X,Y)k(X)\ell_f(\varepsilon|X)] = \int k_G k\, dM, \quad k \in K.$$

By the definition of h_G we have $\int h_G h^\top\, dM = H_G$ and $\int h_G b\, dM = 0$ for all b in $\mathscr{B}(G)$. Using these properties of h_G we obtain

$$\begin{aligned}\int k_G k\, dM &= (a_* - \alpha_G)^\top a + \int \bar{\Pi}_G(\sigma^2 b_G)h^\top a\, dM + \int \sigma^2 b_G b\, dM \\ &= a_*^\top a + \int b_G b\, dG, \quad k = a^\top h + b \in K.\end{aligned}$$

This shows that (5.3) holds. Consequently, $\gamma_G(X,Y)$ is the efficient influence function. □

Technical details for Example 3, Section 4.1. Here we show that the function $\bar{\gamma}_G(X,Y)$ is the efficient influence function for estimating ϑ when U and V are one-dimensional, $\sigma^2(X) = \tau(V)$, and $E_G[\varepsilon^3|X]$ and $E_G[\varepsilon^4|X]$ happen to be functions of V only. The tangent space for this model is

$$\bar{\mathscr{V}}(G) = \{(aU + b(V))\ell_f(\varepsilon|X) + c(X,\varepsilon) : a \in \mathbb{R}, b \in L_2(\Gamma), c \in \bar{\mathscr{C}}(G)\}.$$

Thus it suffices to show that $\bar{\gamma}_G$ belongs to $\bar{\mathscr{V}}(G)$ and is a gradient,

$$E_G[\gamma_G(X,Y)\{(aU + b(V))\ell_f(\varepsilon|X) + c(X,\varepsilon)\}] = a, \tag{5.4}$$

for all $a \in \mathbb{R}$, $b \in L_2(\Gamma)$ and $c \in \bar{\mathscr{C}}(G)$.

To simplify notation, we abbreviate $E_G[(U - m(V))^2 w(V)]$ by $\bar{H}$. Then we can write

$$\bar{\gamma}_G(X,Y) = \frac{U - m(V)}{\bar{H}}\Big[\frac{\varepsilon}{\tau(V)} - \frac{\mu_3(V)}{\tau(V)\Delta(V)}\Big(\varepsilon^2 - \tau(V) - \frac{\mu_3(V)}{\tau(V)}\varepsilon\Big)\Big].$$

To see that $\bar{\gamma}_G(X,Y)$ is a tangent, we express it as

$$\bar{\gamma}_G(X,Y) = \frac{U - m(V)}{\bar{H}}\ell_f(\varepsilon|X) - c_0(X,\varepsilon)$$

with

$$c_0(X,\varepsilon) = \frac{U - m(V)}{\bar{H}}\Big[\ell_f(\varepsilon|X) - \frac{\varepsilon}{\tau(V)} + \frac{\mu_3(V)}{\tau(V)\Delta(V)}\Big(\varepsilon^2 - \tau(V) - \frac{\mu_3(V)}{\tau(V)}\varepsilon\Big)\Big],$$

and verify that c_0 belongs to $\bar{\mathscr{C}}(G)$. Using (5.1) and (5.2) we obtain the identities

$$\int c_0(x,y)f(y|x)\,dy = 0, \quad \int c_0(x,y)yf(y|x)\,dy = 0, \quad \int c_0(x,y)y^2 f(y,x)\,dy = 0,$$

from which we can conclude that c_0 belongs to $\bar{\mathscr{C}}(G)$. These identities also yield

$$E_G[\bar\gamma_G(X,Y)c(X,\varepsilon)] = 0, \quad c \in \bar{\mathscr{C}}(G).$$

Using (5.1) and (5.2) we obtain

$$E_G[\bar\gamma_G(X,Y)b(V)\ell_f(\varepsilon|X)] = E_G\Big[\frac{U-m(V)}{\bar H}b(V)w(V)\Big] = 0, \quad b \in L_2(\Gamma).$$

Finally, we have

$$\begin{aligned} E_G[\bar\gamma_G(X,Y)aU\ell_f(\varepsilon|X)] &= \frac{a}{\bar H}E_G[(U-m(V)Uw(V)] \\ &= \frac{a}{\bar H}E_G[(U-m(V))^2 w(V)] = a, \quad a \in \mathbb{R}. \end{aligned}$$

The last three identities show that (5.4) holds. This completes the proof. □

Acknowledgements

The authors thank the referees for helpful suggestions and comments that improved the manuscript.

References

[1] Ahmad, I., Leelahanon, S. and Li, Q. (2005). Efficient estimation of a semiparametric partially linear varying coefficient model. *Ann. Statist.*, **33**, 258-283.

[2] Bickel, P.J., Klaassen, C.A.J., Ritov, Y. and Wellner, J.A. (1998). *Efficient and Adaptive Estimation for Semiparametric Models.* Springer.

[3] Carroll, R.J. (1982). Adapting for Heteroscedasticity in Linear Models. *Ann. Statist.*, **10**, 1224-1233.

[4] Cui, X., Härdle, W. and Zhu, L. (2011). The EFM approach for single-index models. *Ann. Statist.*, **39**, 1658-1688.

[5] Härdle, W., Hall P. and Ichimura H. (1993). Optimal smoothing in single-index models. *Ann. Statist.*, **21**, 157-178.

[6] Kim, K.K. and Shao, J. (2013). *Statistical Methods for Handling Incomplete Data.* Chapman and Hall/CRC.

[7] Koul, H.L., Müller, U.U. and Schick, A. (2012). The transfer principle: a tool for complete case analysis. *Ann. Statist.*, **40**, 3031-3049.

[8] Lai P., Wang Q. and Zhou X.-H. (2014). Variable selection and semiparametric efficient estimation for the heteroscedastic partially linear single-index model. *Comput. Statist. Data Anal.*, **70**, 241-256.

[9] Little, R.J.A. and Rubin, D.B. (2002). *Statistical analysis with missing data.* Second edition. Wiley-Interscience.

[10] Long, W., Ouyang M. and Shang Y. (2013). Efficient estimation of partially linear varying coefficient models. *Economics Letters*, **121**, 79-81.

[11] Ma, Y., Chiou, J.M. and Wang, N. (2006). Efficient semiparametric estimator for heteroscedastic partially linear models. *Biometrika*, **93**, 75-84.

[12] Ma, Y. and Zhu, L. (2013). Doubly robust and efficient estimators for heteroscedastic partially linear single-index models allowing high dimensional covariates. *J. R. Stat. Soc. Series B Stat. Methodol.*, **75**, 305-322

[13] Müller, H.-G. and Stadtmüller, U. (1987). Estimation of heteroscedasticity in regression analysis. *Ann. Statist.*, **15**, 610-625.

[14] Müller, U.U. and Schick, A. (2016). Efficiency transfer for regression models with responses missing at random. To appear in: *Bernoulli.*

[15] Müller, U.U., Schick, A. and Wefelmeyer, W. (2007). Estimating the error distribution function in semiparametric regression. *Statist. Decisions*, **25**, 1-18.

[16] Müller, U.U. and Van Keilegom, I. (2012). Efficient parameter estimation in regression with missing responses. *Electron. J. Stat.*, **6**, 1200-1219.

[17] Powell, J.L., Stock, J.H. and Stoker T.M. (1989). Semiparametric estimation of index coefficients. *Econometrica*, **56**, 1403-1430.

[18] Robins, J.M. and Rotnitzky, A. (1995). Semiparametric Efficiency in Multivariate Regression Models with Missing Data, *J. Am. Stat. Assoc.*, **90**, 122-129.

[19] Robins, J.M., Rotnitzky, A. and Zhao L.P. (1994). Estimation of Regression Coefficients When Some Regressors Are Not Always Observed, *J. Am. Stat. Assoc.*, **89**, 846-866.

[20] Robinson, P. (1987). Asymptotically efficient estimation in the presence of heteroskedasticity of unknown form. *Econometrica*, **55**, 875-891.

[21] Schick, A. (1987). A note on the construction of asymptotically linear estimators. *J. Statist. Plann. Inference*, **16**, 89-105. Correction: **22** (1989), 269-270.

[22] Schick, A. (1993). On efficient estimation in regression models. *Ann. Statist.*, **21**, 1486-1521. Correction and addendum: **23** (1995) 1862-1863.

[23] Schick, A. (1996). Weighted least squares estimates in partly linear regression models. *Statist. Probab. Lett.*, **27**, 281-287.

[24] Schick, A. (2013). Weighted least squares estimation with missing responses: an empirical likelihood approach. *Electron. J. Stat.*, **7**, 932-945.

[25] Tsiatis, A. (2006). *Semiparametric Theory and Missing Data*, Springer, New York.