Ann Inst Stat Math (2012) 64:751–764
DOI 10.1007/s10463-011-0332-y

The efficiency of the second-order nonlinear least squares estimator and its extension

Mijeong Kim · Yanyuan Ma

Received: 24 March 2010 / Revised: 21 September 2010 / Published online: 20 May 2011

Abstract We revisit the second-order nonlinear least square estimator proposed in Wang and Leblanc (Anne Inst Stat Math 60:883–900, 2008) and show that the estimator reaches the asymptotic optimality concerning the estimation variability. Using a fully semiparametric approach, we further modify and extend the method to the heteroscedastic error models and propose a semiparametric efficient estimator in this more general setting. Numerical results are provided to support the results and illustrate the finite sample performance of the proposed estimator.

Keywords Second-order least squares estimator · Heteroscedasticity · Moments · Semiparametric methods

1 Introduction

Wang and Leblanc (2008) considered a regression model with a parametric mean function and a constant variance:

$$Y = m(X; \beta) + \epsilon, \tag{1}$$

where they assumed that Y is a one-dimensional continuous response variable, and X is a covariate vector that can be continuous, discrete, or mixed. The mean function m is a known function up to the d-dimensional parameter β and the model error ϵ satisfies the usual mean zero assumption $E(\epsilon|X) = 0$. In addition, they also assumed

M. Kim (✉) · Y. Ma
Department of Statistics, Texas A&M University, College Station, TX 77843, USA
e-mail: mjkim@stat.tamu.edu

Y. Ma
e-mail: ma@stat.tamu.edu

that ϵ has a constant yet unknown variance σ^2, i.e. $E(\epsilon^2|X) = \sigma^2$. The observations are denoted $(X_1, Y_1), \ldots, (X_n, Y_n)$, each satisfies (1) and the n observations are independent of each other. Without the additional homoscedastic assumption, this is the usual semiparametric regression problem, or sometimes named the restricted moment model, and the consistent estimator family as well as the efficient estimator for β is known. See, for example, Tsiatis (2006, p. 53). With the additional assumption of homoscedasticity, Wang and Leblanc (2008) proposed a second-order least squares (SLS) type estimation procedure where they take advantage of the knowledge of the second moment of $Y - m(X; \beta)$. They showed that this SLS estimation of β indeed yields improvement over classical least squares estimation.

This naturally motivates us to ask: Can further improvement be obtained? In other words, we are curious to find out whether or not the Wang and Leblanc (2008) estimator reaches the optimal efficiency bound in the sense of Bickel et al. (1993). Studying the semiparametric efficiency bound is important in understanding a model. It provides an ultimate conclusion when searching for estimators or trying to improve existing procedures. Only when an efficient estimator is obtained, the procedure of estimation can be considered to have reached certain optimality. Researchers have been searching for optimal estimators in various problems, the most familiar example being the ordinary and weighted least square estimators in the regression setting. Efficiency issues are also considered in more complex problems such as the Cox model (Tsiatis 2006, Chapter 5.2, p 113), a class of general survival models (Zeng and Lin 2007), problems in case control designs (Rabinowitz 2000; Ma 2010) or involving auxiliary information (Chen et al. 2008), the partially linear models (Chamberlaine 1992; Ma et al. 2006), the latent variable models (Ma and Genton 2010), the functional estimation in semiparametric models (Maity et al. 2007; Müller 2009), the regression with missing covariates (Robins et al. 1994), the skewed distribution families (Ma et al. 2005; Ma and Hart 2007), the quantile regression models (Newey and Powell 1990) and the measurement error models (Tsiatis and Ma 2004; Ma and Carroll 2006). To answer the question of optimality in our problem, we view the model in (1) as a semiparametric problem and take a geometric approach. We construct the locally efficient semiparametric estimators, and proceed to identify the optimal semiparametric efficient (SE) estimator. The SE has the classical root-n convergence rate and is asymptotically normal. We further derive the estimation variance of the SE estimator, which reaches the semiparametric efficiency bound and compares with the result in Wang and Leblanc (2008). It demonstrates that the SLS estimator by Wang and Leblanc (2008) is semiparametrically efficient as well and is thus optimal asymptotically. The resulting estimator, which is asymptotically optimal, is new in literature.

In order to relax the homoscedasticity assumption on ϵ, we subsequently assume $E(\epsilon^2|X) = \sigma^2(X; \gamma)$, which is a function of X with unknown parameter γ. Note that here σ^2 has a known functional form. This model certainly includes the constant variance model as a special case. Although the model is more complex, we can easily adapt the analysis we performed and derive the optimal efficient estimator. The estimator and its asymptotic optimality is also new in literature.

The rest of the paper is organized as follows. We introduce the semiparametric method and show the efficiency of the SLS estimator in Sect. 2. In Sect. 3, we adapt our method to the heteroscedastic error models and propose the efficient estimators and the corresponding variance estimation. We also show the asymptotic optimality of the generalized estimator in this section. Numerical experiments are provided in Sect. 4, and we discuss some possible further extensions in Sect. 5. Technical details are provided in Appendix.

2 Efficiency results

For convenience, we denote the augmented parameter $\theta = (\beta^{\mathrm{T}}, \sigma^2)^{\mathrm{T}}$, and aim at finding the class of consistent semiparametric estimators for θ and identifying the most efficient one within this class. The probability density function (pdf) of a single observation (X, Y), ignoring the subscripts, can be written as

$$p_{X,Y}(x, y) = p_X(x) p_{\epsilon|X}\{y - m(x; \beta)|X = x\} = \eta_1(x)\eta_2\{y - m(x; \beta), x\}, \quad (2)$$

where $\eta_2(\cdot)$ satisfies $\int \epsilon\eta_2(\epsilon, x)\,\mathrm{d}\epsilon = 0$, $\int \epsilon^2\eta_2(\epsilon, x)\,\mathrm{d}\epsilon = \sigma^2$, and the third and fourth moments of ϵ conditional on X are constants. Here, we use $\eta_1(\cdot)$, $\eta_2(\cdot, \cdot)$ to denote the pdf of X and the conditional pdf of ϵ given X. This emphasizes that these pdfs are infinite-dimensional nuisance parameters. We sometimes write $p_{X,Y}(x, y)$ as $p_{X,Y}(x, y; \theta, \eta_1, \eta_2)$ to emphasize that the pdf contains a finite-dimensional parameter θ and infinite-dimensional parameters η_1, η_2. We also use ϵ and $Y - m(X; \beta)$ interchangeably, and use a subscript '0' to denote the true parameter value.

The geometric approach to semiparametric regression analysis consists of defining a Hilbert space $\mathcal{H}$ and finding two subspaces of $\mathcal{H}$, namely the nuisance tangent space Λ and its orthogonal complement $\Lambda^{\perp}$. Here, the Hilbert space is the space of all mean zero, length d, finite variance functions of (X, Y). Here and after, all the expectations are calculated under the true distribution. The subspace Λ is a space spanned by the nuisance score functions (the score function obtained through taking derivative of the logarithm of the pdf with respect to the nuisance parameter) of all the parametric submodels of (2) and their limiting points. The subspace $\Lambda^{\perp}$ is a space consisting of all the functions that are orthogonal to all the functions in Λ. See Tsiatis (2006, Chapter 4) for an elaborated explanation of these concepts. Once Λ and $\Lambda^{\perp}$ are obtained, we then project the score function $S_\theta = \partial \log p_{X,Y}/\partial\theta$ onto $\Lambda^{\perp}$ to obtain S_{eff}. The orthogonal projection of the score function onto $\Lambda^{\perp}$, i.e. S_{eff}, is usually referred to as the efficient score function. If it can be constructed, $\sum_{i=1}^{n} S_{\mathrm{eff}}(X_i, Y_i; \beta, \eta_{10}, \eta_{20}) = 0$ will then be the estimating equation that will yield the optimal estimator of θ.

For model (1), a careful analysis yields $\Lambda = \Lambda_{\eta_1} \oplus \Lambda_{\eta_2}$, where

$$\begin{aligned}
\Lambda_{\eta_1} &= \{f(X) : E(f) = 0\},\\
\Lambda_{\eta_2} &= \{g(\epsilon, X) : E(g|X) = 0,\ E(\epsilon g|X) = 0,\ E(\epsilon^2 g|X) = 0\}, \text{ and}\\
\Lambda^{\perp} &= \{h(\epsilon, X) : h = a(X)\epsilon + b(X)(\epsilon^2 - \sigma^2)\},
\end{aligned}$$

where a, b, f are all length d functions of X, and g, h are length d functions of ϵ, X. The derivation details are in Appendix A1. Taking derivative of the logarithm of the pdf with respect to θ gives us the score function

$$S_\theta(X,Y) = (S_\beta^{\mathrm{T}}, S_{\sigma^2})^{\mathrm{T}} = \left\{ -\frac{\partial \eta_2(\epsilon, X)}{\eta_2(\epsilon, X)\partial\epsilon} \frac{\partial m(X;\beta)}{\partial \beta^{\mathrm{T}}}, \frac{\partial \eta_2(\epsilon, X)}{\eta_2(\epsilon, X)\partial\sigma^2} \right\}^{\mathrm{T}}. \quad (3)$$

As pointed out in Appendix A2, the projection of an arbitrary function $h(X, Y) \in \mathcal{H}$ onto $\Lambda^\perp$ can be calculated as

$$\Pi(h|\Lambda^\perp) = \frac{E(\epsilon h|X)}{\sigma^2}\epsilon + \frac{E(Ch|X)}{E(C^2|X)}C, \quad (4)$$

where $C = \epsilon^2 - \sigma^2 - E(\epsilon^3 - \epsilon\sigma^2|X)\epsilon/\sigma^2$. Letting $h(X, Y) = S_\theta(X, Y)$, we can obtain $S_{\mathrm{eff}} = (S_{\beta,\mathrm{eff}}^{\mathrm{T}}, S_{\sigma^2,\mathrm{eff}})^{\mathrm{T}}$. Further solving $\sum_{i=1}^n S_{\mathrm{eff}}(X_i, Y_i) = 0$ gives the SE estimator, and the asymptotic covariance matrix of $\sqrt{n}\widehat{\theta}$ is

$$n\mathrm{cov}(\widehat{\theta}) = \{E(S_{\mathrm{eff}} S_{\mathrm{eff}}^{\mathrm{T}})\}^{-1}.$$

More specifically, we have

Theorem 1 *The efficient score functions for β and σ^2 have the form*

$$\begin{aligned} S_{\beta,\mathrm{eff}}(X,Y) &= \frac{\partial m(X;\beta)}{\partial\beta}\left\{\frac{\epsilon}{\sigma^2} - \frac{E(\epsilon^3|X)C}{\sigma^2 E(C^2|X)}\right\}, \\ S_{\sigma^2,\mathrm{eff}}(X,Y) &= \frac{C}{E(C^2|X)}. \end{aligned} \quad (5)$$

The estimation covariance matrix is

$$E(S_{\mathrm{eff}} S_{\mathrm{eff}}^{\mathrm{T}}|_{\theta=\theta_0})^{-1}$$
$$= \begin{bmatrix} \left(\sigma_0^2 - \frac{\mu_3^2}{\mu_4 - \sigma_0^4}\right)\left\{B - \frac{\mu_3^2}{\sigma_0^2(\mu_4-\sigma_0^4)}AA^{\mathrm{T}}\right\}^{-1} & \frac{\mu_3\{\sigma_0^2(\mu_4-\sigma_0^4)-\mu_3^2\}B^{-1}A}{\sigma_0^2(\mu_4-\sigma_0^4)-\mu_3^2 A^{\mathrm{T}}B^{-1}A} \\ \frac{\mu_3\{\sigma_0^2(\mu_4-\sigma_0^4)-\mu_3^2\}A^{\mathrm{T}}B^{-1}}{\sigma_0^2(\mu_4-\sigma_0^4)-\mu_3^2 A^{\mathrm{T}}B^{-1}A} & \frac{(\mu_4-\sigma_0^4)\{\sigma_0^2(\mu_4-\sigma_0^4)-\mu_3^2\}}{\sigma_0^2(\mu_4-\sigma_0^4)-\mu_3^2 A^{\mathrm{T}}B^{-1}A} \end{bmatrix}, \quad (6)$$

where $\mu_3 = E(\epsilon^3|X)$, $\mu_4 = E(\epsilon^4|X)$, and

$$A = E\left\{\frac{\partial m(X;\beta_0)}{\partial\beta}\right\}, \quad B = E\left\{\frac{\partial m(X;\beta_0)}{\partial\beta}\frac{\partial m(X;\beta_0)}{\partial\beta^T}\right\}.$$

The proof of Theorem 1 consists the derivation of the efficient score (5), given in Appendix A3, and the derivation of (6), given in Appendix A4. Comparing the variances in (6), and (7), (8) in Wang and Leblanc (2008), we obtain that their SLS estimator is indeed efficient.

We would like to point out that for convenience, we have assumed both μ_3 and μ_4 are constants, and we estimate them using the residuals from an initial OLS estimator. The same assumptions are made in Wang and Leblanc (2008). If, however, these assumptions are not valid, we still have $E(C|X) = E(\epsilon^2|X) - \sigma^2 - \mu_3^* E(\epsilon|X)/\sigma^2 = 0$, where we use μ_3^* to denote $E(\epsilon^3 - \epsilon\sigma^2|X)$ calculated under the wrong model. From the efficient score in (5), it is easily verified that we still have $E(S_{\rm eff}|X) = 0$, hence our estimator remains consistent. On the other hand, if these assumptions are indeed valid, then our procedure achieves the optimal efficiency.

3 Extension

In Sect. 2, we developed the efficient semiparametric estimator and its estimation variance in the model (1) with the homoscedasticity error. In this section, we extend the results to the heteroscedastic error case. We assume $E(\epsilon^2|X) = \sigma^2(X;\gamma)$, where the variance function σ^2 is known up to an unknown parameter γ. We denote the parameter $\theta = (\beta^{\rm T}, \gamma^{\rm T})^{\rm T}$. In this case, similar derivations show that the nuisance tangent space, now denoted Ω, can still be expressed as $\Omega = \Omega_{\eta_1} + \Omega_{\eta_2}$, where $\Omega_{\eta_1} = \Lambda_{\eta_1}$ is unchanged from the homoscedastic case, and

$$\Omega_{\eta_2} = \{g(\epsilon, X) : E(g|X) = 0, E(\epsilon g|X) = 0, E[\{\epsilon^2 - \sigma^2(X;\gamma)\}g|X] = 0\}.$$

Consequently,

$$\Omega^{\perp} = \{h(\epsilon, X) : h = a(X)\epsilon + b(X)\{\epsilon^2 - \sigma^2(X;\gamma)\}\}.$$

Here, a, b, g, h are all length d functions.

The score function

$$S_\theta(X, Y) = (S_\beta^{\rm T}, S_\gamma^{\rm T})^{\rm T} = \left\{-\frac{\partial\eta_2(\epsilon, X)}{\eta_2(\epsilon, X)\partial\epsilon}\frac{\partial m(X;\beta)}{\partial\beta^{\rm T}}, \frac{\partial\eta_2(\epsilon, X)}{\eta_2(\epsilon, X)\partial\sigma^2}\frac{\partial\sigma^2(X;\gamma)}{\partial\gamma^{\rm T}}\right\}^{\rm T}$$

can be similarly calculated by differentiating the logarithm of the pdf with respect to θ. The projection of an arbitrary function $h(X, Y) \in \mathcal{H}$ onto $\Omega^{\perp}$ also has a form similar to (4), i.e.

$$\Pi(h|\Omega^{\perp}) = \frac{E(\epsilon h|X)}{\sigma^2(X;\gamma)}\epsilon + \frac{E(Ch|X)}{E(C^2|X)}C,$$

where $C = \epsilon^2 - \sigma^2(X;\gamma) - E\{\epsilon^3 - \epsilon\sigma^2(X;\gamma)|X\}\epsilon/\sigma^2(X;\gamma)$. Projecting S_θ onto $\Omega^{\perp}$, we can obtain $S_{\rm eff} = (S_{\beta,{\rm eff}}^{\rm T}, S_{\gamma,{\rm eff}}^{\rm T})^{\rm T}$. The SE estimator can therefore be obtained through solving $\sum_{i=1}^n S_{\rm eff}(X_i, Y_i) = 0$, and the asymptotic covariance matrix of the resulting estimator satisfies $n\widehat{\rm cov(\theta)} = \{E(S_{\rm eff}S_{\rm eff}^{\rm T})\}^{-1}$ evaluated at $\theta = \theta_0$. We summarize the results parallel to Theorem 1 in Theorem 2, but omit the detailed proofs.

Theorem 2 *The efficient score functions for β and γ have the form*

$$S_{\beta,\text{eff}}(X,Y) = \frac{\partial m(X;\beta)}{\partial\beta}\left\{\frac{\epsilon}{\sigma^2(X;\gamma)} - \frac{E(\epsilon^3|X)C}{\sigma^2(X;\gamma)E(C^2|X)}\right\},$$
$$S_{\gamma,\text{eff}}(X,Y) = \frac{C}{E(C^2|X)}\frac{\partial\sigma^2(X;\gamma)}{\partial\gamma}.$$

The estimation covariance matrix is

$$E\left(S_{\text{eff}}S_{\text{eff}}^{\text{T}}\right) = E\begin{bmatrix} \frac{\partial m(X;\beta_0)}{\partial\beta}\frac{\partial m(X;\beta_0)}{\partial\beta^{\text{T}}}\frac{1}{\sigma^2(X;\gamma_0)}\left\{1+\frac{\mu_3^2}{\sigma^2(X;\gamma_0)E(C^2|X)}\right\} & -\frac{\mu_3}{\sigma^2(X;\gamma_0)E(C^2|X)}\frac{\partial m(X;\beta_0)}{\partial\beta}\frac{\partial\sigma^2(X;\gamma_0)}{\partial\gamma^{\text{T}}} \\ -\frac{\mu_3}{\sigma^2(X;\gamma_0)E(C^2|X)}\frac{\partial\sigma^2(X;\gamma_0)}{\partial\gamma}\frac{\partial m(X;\beta_0)}{\partial\beta^{\text{T}}} & \frac{1}{E(C^2|X)}\frac{\partial\sigma^2(X;\gamma_0)}{\partial\gamma}\frac{\partial\sigma^2(X;\gamma_0)}{\partial\gamma^{\text{T}}}\end{bmatrix},$$

where μ_3 is defined after (6).

The upper-left block of the inverse of $E(S_{\text{eff}}S_{\text{eff}}^{\text{T}})$ gives the covariance matrix of the efficient estimator $n\widehat{\beta}$. Specifically, denote

$$U_1 = m'_\beta(X;\beta_0)\mu_3\sigma^{-2}(X;\gamma_0)/\sqrt{E(C^2|X)},$$
$$U_2 = \frac{\partial\sigma^2(X;\gamma_0)}{\partial\gamma}/\sqrt{E(C^2|X)},$$

where $m'_\beta(X;\beta_0)$ denotes $\partial m(X;\beta)/\partial\beta^{\text{T}}$ evaluated at $\beta=\beta_0$, we have

$$n\text{cov}(\widehat{\beta}) = [E\{m'_\beta(X;\beta_0)m'_\beta(X;\beta_0)^{\text{T}}\sigma^{-2}(X;\gamma_0)\}$$
$$+E(U_1U_1^{\text{T}}) - E(U_1U_2^{\text{T}})\{E(U_2U_2^{\text{T}})\}^{-1}E(U_1U_2)^{\text{T}}]^{-1}.$$

In contrast to the efficient semiparametric estimator, we inspect the usual weighted least square (WLS) estimator where $\sum_{i=1}^n w_i\{Y_i - m(X_i;\beta)\}^2$ is minimized to obtain the WLS estimator $\widetilde{\beta}$. In this case, it is well known that the optimal weights are $w_i = 1/\sigma^2(X_i;\gamma_0)$, and the corresponding optimal estimator in the WLS family has asymptotic estimation covariance matrix

$$n\text{cov}(\widetilde{\beta}) = E\{\sigma^{-2}(X_i;\gamma_0)m'_\beta(X_i;\beta_0)m'_\beta(X_i;\beta_0)^{\text{T}}\}^{-1}.$$

The covariance matrices of the two estimators are obviously different. In fact, it is easy to verify that

$$\{n\text{cov}(\widehat{\beta})\}^{-1} - \{n\text{cov}(\widetilde{\beta})\}^{-1} = E(U_1U_1^{\text{T}}) - E(U_1U_2^{\text{T}})\{E(U_2U_2^{\text{T}})\}^{-1}E(U_1U_2)^{\text{T}}$$
$$= \text{cov}[U_1 - E(U_1U_2^{\text{T}})\{E(U_2U_2^{\text{T}})\}^{-1}U_2],$$

hence $n\text{cov}(\widetilde{\beta}) - n\text{cov}(\widehat{\beta})$ is nonnegative-definite. This shows that although the optimal WLS is the most efficient among the WLS estimator family, it is in general not as efficient as the SE estimator that we have derived. The practical improvement of the estimation variance will be demonstrated in the simulation studies in Sect. 4.

Unlike in the homoscedastic error case, it is no longer reasonable to assume μ_3 and μ_4 to be constants. Similar to σ^2, they are usually functions of the covariate X, say $\mu_3(X)$ and $\mu_4(X)$. Implementing our efficient estimator requires plugging in $\mu_3(X)$ and $\mu_4(X)$, which are generally unknown. In practice, we could first obtain the residual r_is of the model from an initial OLS estimator, then fit parametric or non-parametric models for (X_i, r_i^3) and (X_i, r_i^4) to obtain $\widehat{\mu}_3(X)$, $\widehat{\mu}_4(X)$. We then proceed with the estimation on β, γ. Similar to the homoscedastic case, even if the parametric models are misspecified, or the estimation of $\mu_3(X)$ and $\mu_4(X)$ is completely wrong, our estimator remains consistent. This is because $E(S_{\mathrm{eff}}|X) = 0$ is guaranteed by $E(\epsilon|X) = 0$ and $E(\epsilon^2|X) = \sigma^2(X;\gamma)$, it does not rely on the correctness of $\mu_3(X)$ and $\mu_4(X)$. However, when the model is correct, our estimator is efficient, even when $\mu_3(X)$, $\mu_4(X)$ are estimated rather crudely.

4 Numerical results

We carry out simulations to study the finite sample performance of the various estimators. The first two simulations focus on homoscedastic error models, as studied in Sect. 2, and the last two simulations on heteroscedastic error models as studied in Sect. 3.

The mean in simulation one has an exponential form, and the model is

$$Y = \beta_1 \exp(\beta_2 X) + \epsilon, \tag{7}$$

with the true parameter values $\beta_1 = 10$, $\beta_2 = -0.6$ and a constant error variance $\sigma^2 = 2$. In simulation two, we considered a growth model

$$Y = \frac{\beta_1}{1 + \exp(\beta_2 + \beta_3 X)} + \epsilon, \tag{8}$$

with the true parameter values $\beta_1 = 10$, $\beta_2 = 1.5$, $\beta_3 = -0.8$ and $\sigma^2 = 2$. Models (7) and (8) are identical to the simulation settings in Wang and Leblanc (2008). In both models, x_is are generated from a uniform distribution in (0, 20) and $\epsilon_i = (e_i - 3)/\sqrt{3}$, where e_is are generated from a $\chi^2(3)$ distribution. Thus, ϵ_is have mean zero and variance 2 but are asymmetrically distributed. The asymmetry insures that the third moment $E(\epsilon^3|X)$ does not vanish, hence the SE or SLS does not degenerate to the WLS estimator. We implemented the OLS, the SLS and the SE estimators, and report the sample mean and sample variance of these estimates. For the SE estimator, we also calculated the estimated variance.

We used a sample size $n = 200$, and generated 1,000 data sets. The simulation results are presented in Table 1. These results clearly indicate that all three estimators are consistent, while both the SLS and the SE estimators outperform the OLS in terms of estimation variance. The estimation variances of the SLS and the SE estimator are very close, which supports our claim that both are efficient. Finally, our variance

Table 1 Simulation results for exponential and growth mean model with homoscedastic error (results are based on a sample size of 200)

	OLS	VAR	SLS	VAR	SE	VAR	VAR1
Exponential model							
$\beta_1 = 10$	10.0768	0.5572	10.0938	0.3418	10.0889	0.3379	0.3721
$\beta_2 = -0.6$	−0.6068	0.0038	−0.6070	0.0024	−0.6068	0.0024	0.0024
$\sigma^2 = 2$	1.9597	0.1024	1.9418	0.0714	1.9319	0.0726	0.0641
Growth model							
$\beta_1 = 10$	9.9981	0.0146	9.9872	0.0125	9.9849	0.0107	0.0119
$\beta_2 = 1.5$	1.5236	0.0489	1.5173	0.0259	1.5166	0.0254	0.0259
$\beta_3 = -0.8$	−0.8104	0.0097	−0.8059	0.0047	−0.8060	0.0047	0.0045
$\sigma^2 = 2$	1.9469	0.1112	1.9179	0.1081	1.8920	0.0921	0.1092

The average and sample variance (VAR) of 1,000 OLS, SLS and SE estimators as well as the average of the 1,000 estimated variances (VAR1) for the SE estimator are presented

Table 2 Simulation results for exponential and growth mean model with homoscedastic error (results are based on a sample size of 500)

	OLS	VAR	SLS	VAR	SE	VAR	VAR1
Exponential model							
$\beta_1 = 10$	10.0185	0.2224	10.0221	0.1187	10.0228	0.1193	0.1193
$\beta_2 = -0.6$	−0.6025	0.0015	−0.6021	$8.73e^{-4}$	−0.6023	$8.76e^{-4}$	$8.74e^{-4}$
$\sigma^2 = 2$	1.9900	0.0483	1.9789	0.0303	1.9730	0.0303	0.0295
Growth model							
$\beta_1 = 10$	10.0005	0.0061	9.9933	0.0051	9.9953	0.0050	0.0051
$\beta_2 = 1.5$	1.5148	0.0181	1.5084	0.0107	1.5077	0.0105	0.0104
$\beta_3 = -0.8$	−0.8039	0.0032	−0.8029	0.0018	−0.8028	0.0018	0.0019
$\sigma^2 = 2$	1.9843	0.0466	1.9728	0.0463	1.9785	0.0468	0.0482

The average and sample variance (VAR) of 1,000 OLS, SLS and SE estimators, as well as the average of the 1,000 estimated variances (VAR1) for the SE estimator are presented

estimation is reasonably precise, in that the sample variance, calculated using the 1,000 estimates via standard sample variance calculation, and the estimated variance, calculated as the mean of the 1,000 estimated variances, are very close. To further demonstrate the impact of the sample size n, we increased n to 500. The numerical outcome in Table 2 further suggests the relevancy of our asymptotic results.

In Sect. 3, we have seen how the variance of the error can be allowed to depend on X. To experiment with the heteroscedastic error situation, we modified the error structure in the first two simulations to have a variance function $\sigma^2(X; \gamma) = \gamma_1 + \gamma_2 X^2$, while keeping the same mean functions and β values. For the exponential model (7), a true value $\gamma = (1, 0.1)$ was used and x_is are generated from a uniform distribution (0, 5). For the growth model (8), we use $\gamma = (2, 0.05)$ and generated x_is from a uniform distribution (0, 7). In both models, we set $\epsilon_i = e_i - k_i$, where $k_i = \sigma^2(x_i; \gamma)/2$ and

e_is are generated from a $\chi^2(k_i)$ distribution. Thus, the errors ϵ_is in both (7) and (8) have mean zero, variance $\sigma^2(X;\gamma)$ and have an asymmetric distribution.

In implementing the WLS estimators, we used the ideal weights $1/\sigma^2(X_i;\gamma_0)$, hence the WLS performance is optimal among all WLS estimators. Implementing the SE estimator requires plugging in the third and fourth conditional moment functions of the error. To test the optimality and robustness of our proposed estimator, we experimented with two different scenarios. In the first case, we calculated the true moment functions and plugged them into the SE estimator (SE1). In the second case, we adopted drastically different functions, and plugged them into the SE estimator as if they were the truth (SE2). To be specific, the true third and fourth conditional moment functions can be calculated to be $a_2X^2 + a_1$ and $a_5X^4 + a_4X^2 + a_3$, respectively, where $a_1 = 4\gamma_1, a_2 = 4\gamma_2, a_3 = 3\gamma_1^2 + 24\gamma_1, a_4 = 24\gamma_2 + 6\gamma_1\gamma_2$ and $a_5 = 3\gamma_2^2$. However, we used the wrong models $a_2X + a_1$ and $a_5X^2 + a_4X + a_3$ instead. Both results along with the optimal WLS results were reported in Table 3. These results are based on 1,000 simulations with a sample size $n = 400$. The results of Table 3 reflect the fact that all three estimators are consistent. Compared with the two SE estimators, the WLS estimator, although already optimal in its family, is much less efficient in that the sample variances in estimating βs are much larger than both SEs. We had expected to see SE1 to outperform SE2 substantially. However, to our surprise, the performance of two estimators is rather similar. This is a pleasant surprise, since modeling and estimating the third and fourth conditional moments usually need very large sample size and can be numerically unstable. Finally, the sample variance and estimated variance for both SEs match reasonably well, indicating the validity of our inference. We also increased the sample size to 500 and 1,000, and find the two get

Table 3 Simulation results for exponential and growth mean model with heteroscedastic error (results are based on a sample size of 400)

	WLS	VAR	SE1	VAR	VAR1	SE2	VAR	VAR1
Exponential model								
$\beta_1 = 10$	9.9931	0.0358	9.9980	0.0217	0.0255	10.0032	0.0229	0.0215
$\beta_2 = -0.6$	−0.5998	$3.51e^{-4}$	−0.5996	$2.27e^{-4}$	$2.32e^{-4}$	−0.5994	$2.98e^{-4}$	$2.14e^{-4}$
$\gamma_1 = 1$			1.0305	0.1328	0.1635	1.0232	0.1436	0.1474
$\gamma_2 = 0.1$			0.0967	0.0016	0.0019	0.0998	0.0019	0.0013
Growth model								
$\beta_1 = 10$	10.0086	0.0668	10.0018	0.0517	0.0497	9.9989	0.0512	0.0490
$\beta_2 = 1.5$	1.5081	0.0093	1.5023	0.0061	0.0061	1.5010	0.0068	0.0059
$\beta_3 = -0.8$	−0.8042	0.0039	−0.8020	0.0020	0.0018	−0.8028	0.0024	0.0018
$\gamma_1 = 2$			1.9768	0.3192	0.2523	1.9993	0.3386	0.2827
$\gamma_2 = 0.05$			0.0498	$9.43e^{-4}$	$7.87e^{-4}$	0.0498	0.0012	$6.68e^{-4}$

The average and sample variance (VAR) of 1,000 WLS, SE1 and SE2 estimators are presented. SE1 is the SE estimator with the true moment models, and SE2 the wrong moment models. Median of the 1,000 estimated variances (VAR1) for SE1 and SE2 are calculated

Table 4 Simulation results for exponential and growth mean model with heteroscedastic error (results are based on a sample size of 1,000)

	WLS	VAR	SE1	VAR	VAR1	SE2	VAR	VAR1
Exponential model								
$\beta_1 = 10$	10.0073	0.0151	10.0068	0.0076	0.0080	10.0107	0.0075	0.0076
$\beta_2 = -0.6$	−0.5999	1.42e^{-4}	−0.5999	7.96e^{-5}	8.09e^{-5}	−0.6001	8.73e^{-5}	7.85e^{-5}
$\gamma_1 = 1$			1.0379	0.0578	0.0555	1.0395	0.0632	0.0618
$\gamma_2 = 0.1$			0.0959	6.36e^{-4}	5.77e^{-4}	0.0963	7.39e^{-4}	5.15e^{-4}
Growth model								
$\beta_1 = 10$	10.0040	0.0258	10.0062	0.0192	0.0194	10.0057	0.0191	0.0192
$\beta_2 = 1.5$	1.5064	0.0034	1.5019	0.0023	0.0024	1.5006	0.0024	0.0023
$\beta_3 = -0.8$	−0.8036	0.0014	−0.8007	6.92e^{-4}	6.94e^{-4}	−0.8003	6.93e^{-4}	6.81e^{-4}
$\gamma_1 = 2$			1.9976	0.1207	0.1127	2.0088	0.1260	0.1242
$\gamma_2 = 0.05$			0.0496	3.77e^{-4}	3.51e^{-4}	0.0489	3.78e^{-4}	3.26e^{-4}

The average and sample variance (VAR) of 1,000 WLS, SE1 and SE2 estimators are presented. SE1 is the SE estimator with the true moment models, and SE2 the wrong moment models. Median of the 1,000 estimated variances (VAR1) for SE1 and SE2 are calculated

closer when the sample size increases, numerical results for $n = 500$ and $n = 1{,}000$ are given in Tables 2 and 4.

5 Discussion

We have derived a SE estimator in a regression model, where the regression error has conditional mean zero and conditional variance a constant. We have shown that this estimator achieves the optimal semiparametric efficiency bound and is equivalent to the SLS estimator proposed in Wang and Leblanc (2008), hence revealing an unknown optimality of their estimator. We further extended the model to the case where the second moment can be an arbitrary function of the covariates, and derived the SE estimator in this general case. The same kind of extension can also be made on the SLS estimator to handle heteroscedasticity. Simulation results demonstrated the significant improvement of the estimation variance in comparison to the classical WLS estimators and supported the inference procedure.

We have adopted fixed models for the third and fourth conditional moment functions of the error distribution, and demonstrated the consistency of the proposed estimator whether or not these higher moment models are misspecified. However, in reality, these moment functions need to be estimated. We caution that the estimation of the higher moments can be rather unstable, usually requiring a large sample size. Although the need to estimate higher order moments will not affect the estimation variance of the parameter of interest in the asymptotic sense, in finite samples, it is very likely to inflate the variance. Thus, we propose to adopt simple models for these higher moments. Finally, the same line of analysis can be extended to higher moments, although both the theoretical analysis and the implementation of the estimators will become increasingly complex.

6 Appendix

A1. Derivation of Λ and $\Lambda^{\perp}$

We consider Λ first. From Tsiatis (2006, Section 4.5), $\Lambda_{\eta_1} =$ (all length d mean zero functions of X).

We now derive Λ_{η_2}. As a model for $p_{\epsilon|X}(\epsilon|x)$, the pdf $\eta_2(\epsilon, x)$ satisfies the following conditions:

$$\int \eta_2(\epsilon, x)\,\mathrm{d}\epsilon = 1, \quad \int \epsilon\eta_2(\epsilon, x)\,\mathrm{d}\epsilon = 0, \quad \int \epsilon^2\eta_2(\epsilon, x)\,\mathrm{d}\epsilon = \sigma^2,$$

which can be equivalently written as

$$\int \eta_2(\epsilon, x)\,\mathrm{d}\epsilon = 1, \quad \int \epsilon\eta_2(\epsilon, x)\,\mathrm{d}\epsilon = 0, \quad \int (\epsilon^2 - \sigma^2)\eta_2(\epsilon, x)\,\mathrm{d}\epsilon = 0.$$

Following Tsiatis (2006, Section 4.5), the first constraint implies that any function $g(\epsilon, X)$ in Λ_{η_2} has to satisfy $E(g|X) = 0$, and the second constraint implies that g has to satisfy $E(\epsilon g|X) = 0$. Applying similar arguments to the third constraint, we can obtain that g has to also satisfy $E\{(\epsilon^2 - \sigma^2)g|X\} = 0$ and, consequently, $E(\epsilon^2 g|X) = 0$. These three requirements on g yield the desired form of the space Λ_{η_2}.

We point out that the space Λ_{η_1} is orthogonal to Λ_{η_2}, which justifies the notation $\Lambda = \Lambda_{\eta_1} \oplus \Lambda_{\eta_2}$. This is because for an arbitrary element $f_1(X) \in \Lambda_{\eta_1}$ and an arbitrary element $f_2(\epsilon, X) \in \Lambda_{\eta_2}$,

$$E\{f_1(X)f_2(\epsilon, X)\} = E[E\{f_1(X)f_2(\epsilon, X)|X\}] = E[f_1(X)E\{f_2(\epsilon, X)|X\}] = 0.$$

To show the form of $\Lambda^{\perp}$, we first define a space $K = \{a(X)\epsilon + b(X)(\epsilon^2 - \sigma^2)\}$, then show $K \subset \Lambda^{\perp}$ and $\Lambda^{\perp} \subset K$.

For any function $h(\epsilon, X) = a(X)\epsilon + b(X)(\epsilon^2 - \sigma^2) \in K$, we will show that $E(hf) = 0$ for all $f \in \Lambda_{\eta_1}$ and $E(hg) = 0$ for all $g \in \Lambda_{\eta_2}$. This would demonstrate that $h \in \Lambda^{\perp}$. We have

$$\begin{aligned}
E\{h(\epsilon, X)f^{\mathrm{T}}(X)\} &= E(E[\{a(X)\epsilon + b(X)(\epsilon^2 - \sigma^2)\}f^{\mathrm{T}}(X)|X]) \\
&= E\{a(X)f^{\mathrm{T}}(X)E(\epsilon|X)\} + E\{b(X)f^{\mathrm{T}}(X)E(\epsilon^2 - \sigma^2|X)\} \\
&= 0, \\
E\{h(\epsilon, X)g^{\mathrm{T}}(\epsilon, X)\} &= E(E[\{a(X)\epsilon + b(X)(\epsilon^2 - \sigma^2)\}g^{\mathrm{T}}(\epsilon, X)|X]) \\
&= E\{a(X)E(\epsilon g^{\mathrm{T}}|X)\} + E\{b(X)E(\epsilon^2 g^{\mathrm{T}}|X)\} \\
&\quad -\sigma^2 E\{b(X)E(g^{\mathrm{T}}|X)\} \\
&= 0.
\end{aligned}$$

Thus, $K \subset \Lambda^{\perp}$.

To show $\Lambda^{\perp} \subset K$, we consider an arbitrary $h \in \Lambda^{\perp}$. Let $\Lambda_{\eta 2} = \Lambda_a \cap \Lambda_b \cap \Lambda_c$, where

$$\Lambda_a = \{g : E(g|X) = 0\}, \quad \Lambda_b = \{g : E(\epsilon g|X) = 0\}, \quad \Lambda_c = \{g : E(\epsilon^2 g|X) = 0\}.$$

Lemma 4.3 of Tsiatis (2006) implies that $\Lambda_{\eta 1}^{\perp} = \Lambda_a$. It is then trivial to see that $h \in \Lambda^{\perp}$ implies $h \perp \Lambda_{\eta 1}$, which further implies $h \in \Lambda_a$. Thus $E(h|X) = 0$. Form $r(\epsilon, X) = E(\epsilon h|X)\epsilon/\sigma^2 + E(Ch|X)C/E(C^2|X)$, where C is defined after (4) and decompose h as

$$h = \{h - E(\epsilon h|X)\epsilon/\sigma^2 - E(Ch|X)C/E(C^2|X)\} + r.$$

Note $r \in K \subset \Lambda^{\perp}$, hence $h_1 = h - E(\epsilon h|X)\epsilon/\sigma^2 - E(Ch|X)C/E(C^2|X) = h - r \in \Lambda^{\perp}$ as well. However, we can easily verify that $h_1 \in \Lambda_{\eta 2}$ at the same time, by verifying that $E(h_1|X) = 0$, $E(\epsilon h_1|X) = 0$ and $E(\epsilon^2 h_1|X) = 0$. Hence, $h_1 = 0$. This indicates $h = r \in K$, thus $\Lambda^{\perp} \subset K$.

A2. Proof of (4)

Let $r(\epsilon, X) = \dfrac{E(\epsilon h|X)}{\sigma^2}\epsilon + \dfrac{E(Ch|X)}{E(C^2|X)}C$. Obviously $r(\epsilon, X) \in \Lambda^{\perp}$. Decompose $h - r$ as

$$h(\epsilon, X) - r(\epsilon, X) = E(h|X) + \{h(\epsilon, X) - r(\epsilon, X) - E(h|X)\}.$$

Note that $E(h|X) \in \Lambda_{\eta 1}$. We can also verify that $h(\epsilon, X) - r(\epsilon, X) - E(h|X) \in \Lambda_{\eta 2}$, by verifying that $E[\{h - r - E(h|X)\}|X] = 0$, $E[\epsilon\{h - r - E(h|X)\}|X] = 0$ and $E[\epsilon^2\{h - r - E(h|X)\}|X] = 0$.

Hence $h(\epsilon, X) - r(\epsilon, X) \in \Lambda$. Thus, we obtain that $\Pi(h|\Lambda) = h(\epsilon, X) - r(\epsilon, X)$ and $\Pi(h|\Lambda^{\perp}) = r(\epsilon, X)$.

A3. Calculation of S_{eff} given in (5)

$S_{\text{eff}}(X, Y)$ can be written as

$$S_{\text{eff}} = \Pi(S_\theta|\Lambda^{\perp}) = \frac{E(\epsilon S_\theta|X)}{\sigma^2}\epsilon + \frac{E(CS_\theta|X)}{E(C^2|X)}C.$$

Using the form of S_β and S_{σ^2} in (3), we can verify that $E(\epsilon S_\beta|X) = \dfrac{\partial m(X;\beta)}{\partial \beta}$ and $E(CS_\beta|X) = -\dfrac{\partial m(X;\beta)}{\partial \beta}\dfrac{\mu_3}{\sigma^2}$, thus

$$S_{\beta,\text{eff}} = \frac{\partial m(X;\beta)}{\partial \beta}\left\{\frac{\epsilon}{\sigma^2} - \frac{\mu_3}{\sigma^2 E(C^2|X)}C\right\}.$$

Similarly, we can verify that $E(\epsilon S_{\sigma^2}|X) = 0$ and $E(CS_{\sigma^2}|X) = 1$, hence $S_{\sigma^2,\mathrm{eff}} = C/E(C^2|X)$.

A4. Derivation of the variances in (6)

Using the explicit form of S_θ, we have

$$S_{\mathrm{eff}}S_{\mathrm{eff}}^{\mathrm{T}} = \begin{bmatrix} \frac{\partial m(X;\beta)}{\partial \beta}\frac{\partial m(X;\beta)}{\partial \beta^{\mathrm{T}}}\left\{\frac{\epsilon}{\sigma^2} - \frac{\mu_3 C}{\sigma^2 E(C^2|X)}\right\}^2 & \frac{\partial m(X;\beta)}{\partial \beta}\left\{\frac{\epsilon}{\sigma^2} - \frac{\mu_3 C}{\sigma^2 E(C^2|X)}\right\}\frac{C}{E(C^2|X)} \\ \frac{\partial m(X;\beta)}{\partial \beta^{\mathrm{T}}}\left\{\frac{\epsilon}{\sigma^2} - \frac{\mu_3 C}{\sigma^2 E(C^2|X)}\right\}\frac{C}{E(C^2|X)} & \frac{C^2}{E(C^2|X)^2} \end{bmatrix}.$$

Taking expectation of $S_{\mathrm{eff}}S_{\mathrm{eff}}^{\mathrm{T}}$ evaluated at the true parameter values, we have

$$\begin{aligned} &E(S_{\mathrm{eff}}S_{\mathrm{eff}}^{\mathrm{T}}|_{\theta=\theta_0}) \\ &= \begin{bmatrix} \frac{1}{\sigma_0^2}E\left\{\frac{\partial m(X;\beta_0)}{\partial \beta}\frac{\partial m(X;\beta_0)}{\partial \beta^{\mathrm{T}}}\right\} + \frac{1}{\sigma_0^4}E\left\{\frac{\partial m(X;\beta_0)}{\partial \beta}\frac{\partial m(X;\beta_0)}{\partial \beta^{\mathrm{T}}}\frac{\mu_3^2}{E(C^2|X)}\right\} & -\frac{1}{\sigma_0^2}E\left\{\frac{\partial m(X;\beta_0)}{\partial \beta}\frac{\mu_3}{E(C^2|X)}\right\} \\ -\frac{1}{\sigma_0^2}E\left\{\frac{\partial m(X;\beta_0)}{\partial \beta^{\mathrm{T}}}\frac{\mu_3}{E(C^2|X)}\right\} & E\left\{\frac{1}{E(C^2|X)}\right\} \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{\sigma_0^2}\left\{1 + \frac{\mu_3^2}{\sigma_0^2(\mu_4-\sigma_0^4)-\mu_3^2}\right\}B & -\frac{\mu_3}{\sigma_0^2(\mu_4-\sigma_0^4)-\mu_3^2}A \\ -\frac{\mu_3}{\sigma_0^2(\mu_4-\sigma_0^4)-\mu_3^2}A^{\mathrm{T}} & \frac{\sigma_0^2}{\sigma_0^2(\mu_4-\sigma_0^4)-\mu_3^2} \end{bmatrix}. \end{aligned}$$

Its inverse can then be calculated using the matrix inversion and is easy to verify to have the form in (6).

References

Bickel, P. J., Klaassen, C. A. J., Ritov, Y., Wellner, J. A. (1993). *Efficient and adaptive estimation for semiparametric models*. Baltimore: The Johns Hopkins University Press.

Chamberlaine, G. (1992). Efficiency bounds for semiparametric regression. *Econometrica, 60*, 567–596.

Chen, X. H., Hong, H., Tarozzi, A. (2008). Semiparametric efficiency in GMM models with auxiliary data. *Annals of Statistics, 36*, 808–843.

Ma, Y. (2010). A semiparametric efficient estimator in case–control studies. *Bernoulli, 16*, 585–603.

Ma, Y., Carroll, R. J. (2006). Locally efficient estimators for semiparametric models with measurement error. *Journal of the American Statistical Association, 101*, 1465–1474.

Ma, Y., Hart, J. (2007). Constrained local likelihood estimators for semiparametric skew-normal distributions. *Biometrika, 94*, 119–134.

Ma, Y., Genton, M. G., Tsiatis, A. A. (2005). Locally efficient semiparametric estimators for generalized skew-elliptical distributions. *Journal of the American Statistical Association, 100*, 980–989.

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

Ma, Y., Genton, M. G. (2010). Explicit semiparametric estimators for generalized linear latent variable models. *Journal of Royal Statistical Society Series B* (in press).

Maity, A., Ma, Y., Carroll, R. J. (2007). Efficient estimation of population-level summaries in general semiparametric regression models. *Journal of the American Statistical Association, 102*, 123–139.

Müller, U. U. (2009). Estimating linear functionals in nonlinear regression with responses missing at random. *Annals of Statistics, 37*, 2245–2277.

Newey, W., Powell, J. L. (1990). Efficient estimation of linear and type I censored regression models under conditional quantile restrictions. *Econometric Theory, 6*, 295–317.

Rabinowitz, D. (2000). Computing the efficient score in semi-parametric problems. *Statistica Sinica, 10*, 265–280.

Robins, J. M., Rotnitzky, A., Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association, 89*, 846–866.

Tsiatis, A. A. (2006). *Semiparametric theory and missing data*. New York: Springer.

Tsiatis, A. A., Ma, Y. (2004). Locally efficient semiparametric estimators for functional measurement error models. *Biometrika, 91*, 835–848.

Wang, L., Leblanc, A. (2008). Second-order nonlinear least squares estimation. *Annals of the Institute of Statistical Mathematics, 60*, 883–900.

Zeng, D., Lin, D. Y. (2007). Maximum likelihood estimation in semiparametric models with censored data (with discussion). *Journal of Royal Statistical Society Series B, 69*, 507–564.