Journal of Statistical Planning and Inference 139 (2009) 125 – 135

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Second-order least squares estimation of censored regression models

Taraneh Abarin, Liqun Wang*

Department of Statistics, University of Manitoba, Winnipeg, Manitoba, Canada R3T 2N2

ARTICLE INFO

Article history:
Received 31 January 2007
Received in revised form
9 April 2008
Accepted 11 April 2008
Available online 13 May 2008

Keywords:
Censored regression model
Tobit model
Asymmetric errors
M-estimator
Consistency
Asymptotic normality
Weighted (nonlinear) least squares

ABSTRACT

This paper proposes the second-order least squares estimation, which is an extension of the ordinary least squares method, for censored regression models where the error term has a general parametric distribution (not necessarily normal). The strong consistency and asymptotic normality of the estimator are derived under fairly general regularity conditions. We also propose a computationally simpler estimator which is consistent and asymptotically normal under the same regularity conditions. Finite sample behavior of the proposed estimators under both correctly and misspecified models are investigated through Monte Carlo simulations. The simulation results show that the proposed estimator using optimal weighting matrix performs very similar to the maximum likelihood estimator, and the estimator with the identity weight is more robust against the misspecification.

1. Introduction

Regression models with censored response variables arise frequently in econometrics, biostatistics, and many other areas. In economics, such a model was first used by Tobin (1958) to analyze household expenditure on durable goods. Other econometric applications of Tobit model include Heckman and MaCurdy (1986) and Killingsworth and Heckman (1986).

In the last decades, various methods have been introduced to estimate the parameters for various censored regression models. In particular, Amemiya (1973) and Wang (1998) investigated maximum likelihood and method of moment estimators when the regression error distribution is normal, while Powell (1984, 1986) proposed the semiparametric estimators when the error distribution is symmetric or satisfies certain quantile restrictions.

In practice, many variables of interests have asymmetric distributions, e.g., income and expenditures, insurance claim and premiums, survival or failure times, etc. Regression with asymmetric error distributions has been considered by many authors, e.g., Williams (1997), Austin et al. (2003), Marazzi and Yohai (2004), and Bianco et al. (2005). Most estimators considered belong to a large family of the so-called M-estimators which maximize or minimize certain criteria.

Recently, Wang and Leblanc (2007) studied a second-order least squares (SLS) estimator for general nonlinear regression models. They have shown that the SLS estimator is asymptotically more efficient than the ordinary least squares (OLS) estimator when the error distribution has a nonzero third moment. However, the framework used in Wang and Leblanc (2007) does not cover the censored regression model considered in the current paper. Hence our goal here is to generalize the SLS approach to this important model.

In particular, we introduce the censored model and the SLS estimator in Section 2. There we also show that the SLS estimators are strongly consistent and asymptotically normally distributed under some general regularity conditions. The SLS estimator is based on the first two conditional moments of the response variable given the predictor variables. In practice, it is not always

* Corresponding author. Tel.: +1 204 474 6270; fax: +1 204 474 7621.
E-mail address: liqun_wang@umanitoba.ca (L. Wang).

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2008.04.016

straightforward to calculate the closed forms of these moments. In this case, the objective function will involve multiple integrals which is difficult to minimize. To overcome this numerical difficulty, in Section 3 we propose a simulation-based SLS estimator and give its asymptotic properties. Further, in Section 4 we carry out substantial Monte Carlo simulations to study finite sample behavior of the SLS estimators and compare them with the maximum likelihood estimator (MLE). Finally, Section 5 contains conclusions, and theoretical assumptions and proofs are given in the Appendix.

2. SLS estimator

Consider the censored regression model

$$Y^* = X'\beta + U, \quad Y = \max(Y^*, 0), \tag{1}$$

where $Y^*, Y \in \mathbb{R}$ are latent and observed response variables, $X \in \mathbb{R}^p$ is a vector of observed predictors, and U is the random error with density $f(u;\theta)$ and satisfies $E(U|X)=0$ and $E(U^2|X)<\infty$. The unknown parameters are $\beta \in \Omega \subset \mathbb{R}^p$ and $\theta \in \Theta \subset \mathbb{R}^q$. Although we consider zero as the left censoring value, it can be easily replaced by any other constant.

Following Wang and Leblanc (2007), we construct the SLS estimator based on the first two conditional moments of Y given X. Under model (1), these two conditional moments are, respectively, given by

$$E(Y|X) = \int I(X'\beta_0 + u)(X'\beta_0 + u) f(u;\theta_0)\,\mathrm{d}u, \tag{2}$$

$$E(Y^2|X) = \int I(X'\beta_0 + u)(X'\beta_0 + u)^2 f(u;\theta_0)\,\mathrm{d}u, \tag{3}$$

where $I(u)=1$ for $u>0$ and $I(u)=0$ for $u \leqslant 0$. Let $\gamma = (\beta', \theta')'$ denote the vector of model parameters and $\Gamma = \Omega \times \Theta \subset \mathbb{R}^{p+q}$ the parameter space. For every $x \in \mathbb{R}^p$ and $\gamma \in \Gamma$, define

$$m_1(x;\gamma) = \int I(x'\beta + u)(x'\beta + u) f(u;\theta)\,\mathrm{d}u\,, \tag{4}$$

$$m_2(x;\gamma) = \int I(x'\beta + u)(x'\beta + u)^2 f(u;\theta)\,\mathrm{d}u. \tag{5}$$

Now suppose $(Y_i, X_i')',\ i = 1, 2, \ldots, n$ is an *i.i.d.* random sample and let

$$\rho_i(\gamma) = (Y_i - m_1(X_i;\gamma), Y_i^2 - m_2(X_i;\gamma))'.$$

Then the SLS estimator for γ is defined as

$$\hat{\gamma}_n = \operatorname*{argmin}_{\gamma \in \Gamma} Q_n(\gamma),$$

where

$$Q_n(\gamma) = \sum_{i=1}^{n} \rho_i'(\gamma) W_i \rho_i(\gamma) \tag{6}$$

and $W_i = W(X_i)$ is a 2×2 nonnegative definite matrix which may depend on X_i. The asymptotic properties of the SLS estimator is given below where the assumptions and proofs are in the Appendix.

Theorem 1. *As* $n \to \infty$,

1. *under assumptions* A1–A5, *the SLS* $\hat{\gamma}_n \xrightarrow{\text{a.s.}} \gamma_0$;
2. *under assumptions* A1–A7, $\sqrt{n}(\hat{\gamma}_n - \gamma_0) \xrightarrow{\text{L}} \mathrm{N}(0, B^{-1}AB^{-1})$, *where*

$$A = E\left[\frac{\partial \rho_1'(\gamma_0)}{\partial \gamma} W_1 \rho_1(\gamma_0) \rho_1'(\gamma_0) W_1 \frac{\partial \rho_1(\gamma_0)}{\partial \gamma'}\right] \tag{7}$$

and

$$B = E\left[\frac{\partial \rho_1'(\gamma_0)}{\partial \gamma} W_1 \frac{\partial \rho_1(\gamma_0)}{\partial \gamma'}\right]. \tag{8}$$

The SLS estimator $\hat{\gamma}_n$ and its asymptotic covariance matrix depend on the weighting matrix $W(X)$. As is shown in Wang and Leblanc (2007) or Abarin and Wang (2006),

$$B^{-1}AB^{-1} \geqslant E\left[\frac{\partial \rho_1'(\gamma_0)}{\partial \gamma} V^{-1} \frac{\partial \rho_1(\gamma_0)}{\partial \gamma'}\right]^{-1} \tag{9}$$

and the lower bound is attained for $W = V^{-1}$ in A and B, where $V = E[\rho_1(\gamma_0)\rho_1'(\gamma_0)|X]$.

In practice V depends on unknown parameter γ_0, and it must be estimated before the SLS $\hat{\gamma}_n$ using $W = V^{-1}$ is computed. This can be done using the following two-stage procedure. First, minimize $Q_n(\gamma)$ using the identity matrix $W(X) = I_2$ to obtain the first-stage estimator $\hat{\gamma}_n$. Secondly, estimate V using $\hat{\gamma}_n$ and then minimize $Q_n(\gamma)$ again with $W = \hat{V}^{-1}$ to obtain the two-stage estimator $\hat{\hat{\gamma}}_n$. Since $\hat{V}$ is consistent for V, the asymptotic covariance matrix of $\hat{\hat{\gamma}}_n$ is given by the lower bound in (9).

In the rest of this section we consider some examples, where explicit forms of the first two moments (2) and (3) can be obtained. The third and fourth conditional moments for these examples can be calculated similarly but are not presented here because they are quite tedious.

Example 1. First consider the model (1) where U has a normal distribution $N(0, \sigma_u^2)$. This is a standard Tobit model. For this model $\gamma = (\beta', \sigma_u^2)'$, and the first two conditional moments are given by

$$E(Y|X) = X'\beta\Phi\left(\frac{X'\beta}{\sigma_u}\right) + \sigma_u\phi\left(\frac{X'\beta}{\sigma_u}\right)$$

and

$$E(Y^2|X) = [(X'\beta)^2 + \sigma_u^2]\Phi\left(\frac{X'\beta}{\sigma_u}\right) + \sigma_u(X'\beta)\phi\left(\frac{X'\beta}{\sigma_u}\right),$$

where Φ and ϕ are the standard normal distribution and density function, respectively.

Example 2. Now consider the model (1) where $(U/\sigma_u)\sqrt{k/(k-2)}$ has a t distribution $t(k)$ with $k > 2$. Then we have

$$E(Y|X) = (X'\beta)F_k\left(\frac{X'\beta}{\sigma_u}\right) + \frac{\sigma_u\sqrt{k}\Gamma\left(\frac{k-1}{2}\right)}{2\sqrt{\pi}\Gamma\left(\frac{k}{2}\right)}\left(1 + \frac{(X'\beta)^2}{k\sigma_u^2}\right)^{-(k-1)/2}$$

and

$$E(Y^2|X) = [(X'\beta)^2 + k\sigma_u^2]F_k\left(\frac{X'\beta}{\sigma_u}\right) + \frac{\sigma_u(X'\beta)\sqrt{k}\Gamma\left(\frac{k-1}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{k}{2}\right)}\left(1 + \frac{(X'\beta)^2}{k\sigma_u^2}\right)^{-(k-1)/2} + \frac{k(k-1)\sigma_u^2}{k-2}F_{k-2}\left(\sqrt{\frac{k-2}{k}}\frac{X'\beta}{\sigma_u}\right),$$

where F_k is the distribution function of $t(k)$.

Example 3. Now consider the model (1) where $U\sqrt{2k}/\sigma_u + k$ has a chi-square distribution $\chi^2(k)$, $k > 1$. Then we have

$$E(Y|X) = (X'\beta + k\sigma_u) - (1 - I(X'\beta))\left[(X'\beta)F_k\left(\frac{-X'\beta}{\sigma_u}\right) + k\sigma_u F_{k+2}\left(\frac{-X'\beta}{\sigma_u}\right)\right]$$

and

$$E(Y^2|X) = (X'\beta + k\sigma_u)^2 + 2k\sigma_u^2 - (1 - I(X'\beta))\left[(X'\beta)^2F_k\left(\frac{-X'\beta}{\sigma_u}\right) + 2k\sigma_u(X'\beta)F_{k+2}\left(\frac{-X'\beta}{\sigma}\right) + \sigma_u^2k(k+2)F_{k+4}\left(\frac{-X'\beta}{\sigma_u}\right)\right],$$

where F_k is the distribution function of $\chi^2(k)$.

In Section 4, the above three models and the corresponding moments will be used in our Monte Carlo simulation studies.

3. Simulation-based estimator

The numerical computation of the SLS estimator of the last section can be done using standard numerical procedures when closed forms of the first two conditional moments are available. However, sometimes explicit forms of the integrals in (4) and (5) may be difficult or impossible to derive. In this case numerical minimization of $Q_n(\gamma)$ will be troublesome, especially when the dimension of parameters $p+q$ is greater than two or three. To overcome this computational difficulty, in this section we consider a simulation-based approach in which the integrals are simulated by Monte Carlo methods. First note that by a change of variables the integrals in (4) and (5) can be written as

$$m_1(x;\gamma)=\int I(u)uf(u-x'\beta;\theta)\,\mathrm{d}u, \tag{10}$$

$$m_2(x;\gamma)=\int I(u)u^2f(u-x'\beta;\theta)\,\mathrm{d}u. \tag{11}$$

The simulation-based estimator (SBE) can be constructed in the following way. First, choose a known density $g(t)$ with support in $[0,+\infty)$, and generate an *i.i.d.* random sample $\{t_{ij}, j=1,2,\ldots,2S, i=1,2,\ldots,n\}$ from $g(t)$. Then approximate $m_1(x;\gamma)$ and $m_2(x;\gamma)$ by the Monte Carlo simulators

$$m_{1,1}(x_i;\gamma)=\frac{1}{S}\sum_{j=1}^{S}\frac{t_{ij}f(t_{ij}-x_i'\beta;\theta)}{g(t_{ij})},\quad m_{1,2}(x_i;\gamma)=\frac{1}{S}\sum_{j=S+1}^{2S}\frac{t_{ij}f(t_{ij}-x_i'\beta;\theta)}{g(t_{ij})},$$

$$m_{2,1}(x_i;\gamma)=\frac{1}{S}\sum_{j=1}^{S}\frac{t_{ij}^2f(t_{ij}-x_i'\beta;\theta)}{g(t_{ij})},\quad m_{2,2}(x_i;\gamma)=\frac{1}{S}\sum_{j=S+1}^{2S}\frac{t_{ij}^2f(t_{ij}-x_i'\beta;\theta)}{g(t_{ij})}.$$

Hence, a simulated version of the objective function $Q_n(\gamma)$ can be defined as

$$Q_{n,S}(\gamma)=\sum_{i=1}^{n}\rho_{i,1}'(\gamma)W_i\rho_{i,2}(\gamma), \tag{12}$$

where

$$\rho_{i,1}(\gamma)=(Y_i-m_{1,1}(X_i;\gamma), Y_i^2-m_{2,1}(X_i;\gamma))',$$

$$\rho_{i,2}(\gamma)=(Y_i-m_{1,2}(X_i;\gamma).Y_i^2-m_{2,2}(X_i;\gamma))'.$$

Since, $\rho_{i,1}(\gamma)$ and $\rho_{i,2}(\gamma)$ are conditionally independent given Y_i, X_i, $Q_{n,S}(\gamma)$ is an unbiased simulator for $Q_n(\gamma)$. Finally, the SBE for γ can be defined by

$$\hat{\gamma}_{n,S}=\operatorname*{argmin}_{\gamma\in\Gamma} Q_{n,S}(\gamma).$$

For the SBE, we have the following results.

Theorem 2. *Suppose that* $\mathrm{Supp}[g(t)]\supseteq[0,+\infty)\cap\mathrm{Supp}[f(u-x'\beta;\theta)]$ *for all* $\gamma\in\Gamma$ *and* $x\in\mathbb{R}^p$. *Then as* $n\to\infty$,

1. *under assumptions* A1–A5, $\hat{\gamma}_{n,S}\xrightarrow{\text{a.s.}}\gamma_0$;
2. *under assumptions* A1–A5 *and* A7–A8, $\sqrt{n}(\hat{\gamma}_{n,S}-\gamma_0)\xrightarrow{L}\mathrm{N}(0,B^{-1}A_SB^{-1})$, *where*

$$\begin{aligned}2A_S=&\,E\left[\frac{\partial\rho_{1,1}'(\gamma_0)}{\partial\gamma}W_1\rho_{1,2}(\gamma_0)\rho_{1,2}'(\gamma_0)W_1\frac{\partial\rho_{1,1}(\gamma_0)}{\partial\gamma'}\right]\\&+E\left[\frac{\partial\rho_{1,1}'(\gamma_0)}{\partial\gamma}W_1\rho_{1,2}(\gamma_0)\rho_{1,1}'(\gamma_0)W_1\frac{\partial\rho_{1,2}(\gamma_0)}{\partial\gamma'}\right].\end{aligned} \tag{13}$$

The proof for Theorem 2 is analogous to Theorem 3 in Wang (2004) and is therefore omitted. In general, the SBE $\hat{\gamma}_{n,S}$ is less efficient than the SLS $\hat{\gamma}_n$, due to the simulation approximation of $\rho_i(\gamma)$ by $\rho_{i,1}(\gamma)$ and $\rho_{i,2}(\gamma)$. Wang (2004) showed that the efficiency loss caused by simulation has a magnitude of $\mathrm{O}(1/S)$. Therefore, the larger the simulation size S, the smaller the efficiency loss. Note also that the above asymptotic results do not require the simulation size S tends to infinity.

4. Simulation studies

In this section we study finite sample behavior of the SLS estimator with both identity and optimal weighting matrices, and compare them with the MLE. We conducted substantial simulation studies using a variety of configurations with respect to sample size, degree of censoring, and error distribution. However, for the sake of saving space, we present here a subset of representative results.

In particular, we simulate the three models in Examples 1–3 with $\beta=(\beta_0,\beta_1)'$ and $\theta=\sigma_u^2$. For each model, we consider the amount of censoring of 31% and 60%, respectively. The covariate X has a normal distribution. For each model, 1000 Monte Carlo repetitions are carried out for each of the sample sizes $n=50,70,100,200,300,500$. We computed the Monte Carlo means of the SLS estimator using the identity weight (SLSIDEN) and the optimal weight (SLSOPT), together with their root mean squared errors (RMSE).

The simulation results for all sample sizes are summarized in Figs. 1 and 2. In particular, Fig. 1 contains the estimates of the model with normal errors, 60% censoring, and the true parameter values $\beta_0=-6$, $\beta_1=1.5$ and $\sigma_u^2=16$. Fig. 2 compares the RMSE of the estimators of σ_u^2 for the three models with 60% censoring. Both graphs show that the SLSOPT and MLE perform very similarly. Further, Fig. 1 shows that all three estimators achieve their large-sample properties with moderate sample sizes, even for a relatively high amount of censoring. Fig. 2 shows that the SLSIDEN clearly has smaller RMSE than the other two estimators in all models.

Next we present more detailed numerical results for the sample size n=200. In particular, Tables 1 and 2 present the simulation results for the models with normal, t, and χ^2 error distributions and 31% and 60% censoring, respectively. Moreover, we report 95% confidence intervals for the parameters as well.

Again, these results show that in general the SLS with optimal weight (SLSOPT) performs very closely to the MLE. This pattern holds for both amount of censoring. Generally, the bias of the estimators increases as the error distribution changes from symmetric to asymmetric. In the case of χ^2 error distribution, MLE shows more bias in both β_0 and β_1. Comparing Tables 1 and 2 reveals that as the proportion of censored observation declines, the RMSE of the estimators decreases. This is because of decrease in the variance of the estimators. It is also apparent that the 95% confidence intervals for β_1 are generally shorter than the confidence intervals for β_0 and σ_u^2. As usual, estimating σ_u^2 with a similar accuracy or precision as regression coefficients would need more Monte Carlo iterations.

We also examine the behavior of the estimators under misspecified distributions. In particular, we calculate the estimators assuming the normal random errors, while the data are generated using t or χ^2 distributions. In each case, the error term U is normalized to have zero mean and variance σ_u^2. The simulated mean estimates, the RMSE and the biases for the sample size $n=200$ and 31% censoring are presented in Table 3. As we can see from Table 3 that, under misspecification, both the MLE and

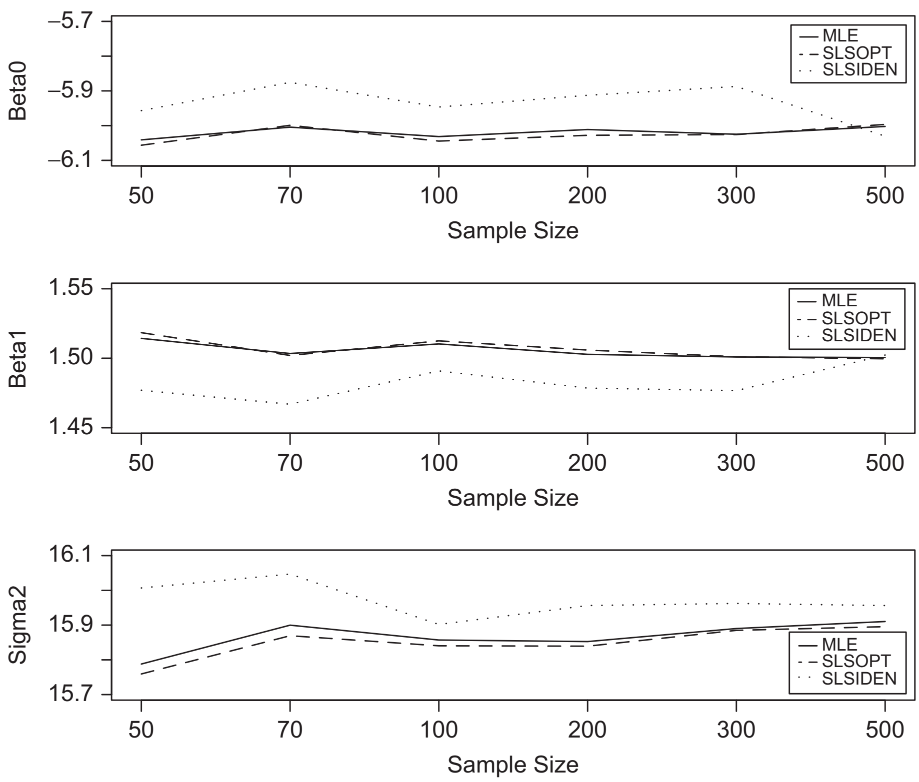

Fig. 1. Monte Carlo estimates of the normal model with 60% censoring and various sample sizes. The true parameter values are $\beta_0=-6$, $\beta_1=1.5$ and $\sigma_u^2=16$.

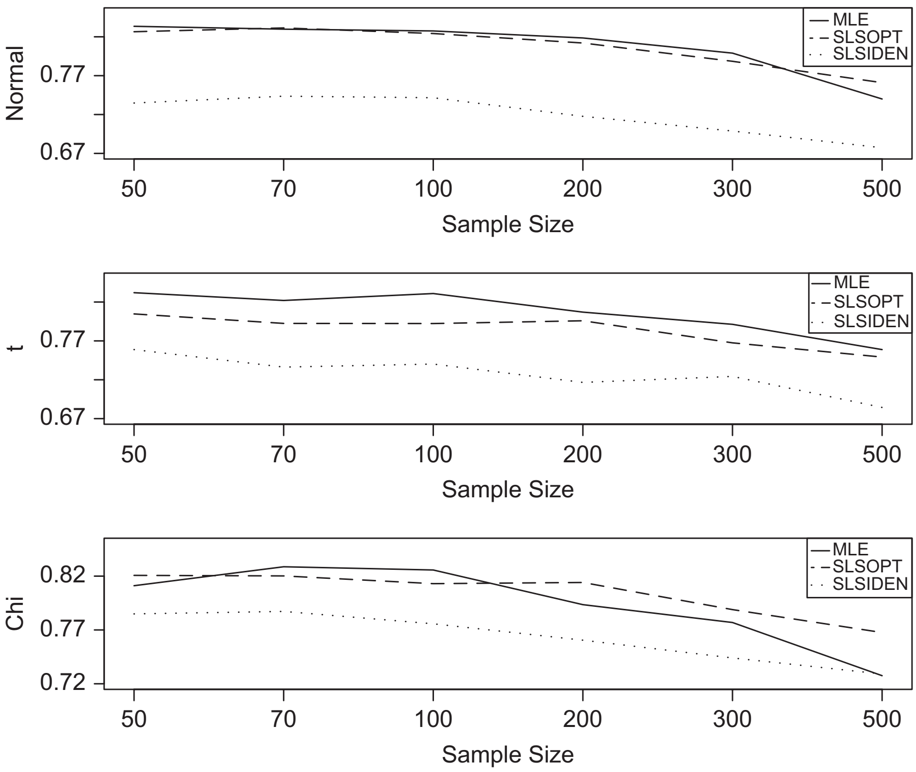

Fig. 2. RMSE of the estimators of σ_u^2 in three models with 60% censoring and various sample sizes.

Table 1
Simulation results of three models with sample size $n = 200$ and 31% censoring

Error	Normal	$t(5)$	$\chi^2(4)$
$\beta_0 = -1.5$			
SLSIDEN	−1.4860	−1.4483	−1.4806
RMSE	0.7428	0.7754	0.8069
95% C.I.	(−1.532, −1.440)	(−1.496, −1.400)	(−1.531, −1.431)
MLE	−1.5097	−1.5188	−1.5695
RMSE	0.4614	0.4298	0.4814
95% C.I.	(−1.538, −1.481)	(−1.545, −1.492)	(−1.599, −1.540)
SLSOPT	−1.5016	−1.5155	−1.5472
RMSE	0.4749	0.4959	0.5770
95% C.I.	(−1.531, −1.472)	(−1.546, −1.485)	(−1.583, −1.512)
$\beta_1 = 1.5$			
SLSIDEN	1.4924	1.4869	1.4947
RMSE	0.1632	0.1538	0.1829
95% C.I.	(1.482, 1.503)	(1.477, 1.496)	(1.483, 1.506)
MLE	1.4999	1.5026	1.5219
RMSE	0.1091	0.0948	0.0923
95% C.I.	(1.493,1.507)	(1.497,1.508)	(1.516,1.527)
SLSOPT	1.4988	1.5021	1.5092
RMSE	0.1106	0.1103	0.1161
95% C.I.	(1.492, 1.506)	(1.495, 1.509)	(1.502, 1.516)
$\sigma_u^2 = 16$			
SLSIDEN	15.9419	15.9363	15.9210
RMSE	0.6625	0.6878	0.7301
95% C.I.	(15.901,15.983)	(15.894,15.979)	(15.876,15.966)
MLE	15.9249	15.9388	15.9287
RMSE	0.7835	0.7879	0.7549
95% C.I.	(15.877, 15.973)	(15.890, 15.987)	(15.882, 15.975)
SLSOPT	15.9075	15.8561	15.9408
RMSE	0.7830	0.7968	0.7995
95% C.I.	(15.859, 15.956)	(15.807, 15.905)	(15.891, 15.990)

Table 2
Simulation results of three models with sample size $n = 200$ and 60% censoring

Error	Normal	$t(5)$	$\chi^2(4)$
$\beta_0 = -6$			
SLSIDEN	−5.9127	−5.9714	−5.9146
RMSE	0.8471	0.8548	0.8592
95% C.I.	(−5.965, −5.860)	(−6.024, −5.918)	(−5.968, −5.862)
MLE	−6.0113	−6.0388	−6.1176
RMSE	0.5910	0.6093	0.6660
95% C.I.	(−6.048, −5.975)	(−6.077, −6.001)	(−6.158, −6.077)
SLSOPT	−6.0279	−6.0418	−6.0361
RMSE	0.6102	0.6775	0.6992
95% C.I.	(−6.066, −5.990)	(−6.084, −6.000)	(−6.079, −5.993)
$\beta_1 = 1.5$			
SLSIDEN	1.4785	1.4957	1.4752
RMSE	0.2024	0.1796	0.2136
95% C.I.	(1.466, 1.491)	(1.485, 1.507)	(1.462, 1.488)
MLE	1.5028	1.5051	1.5287
RMSE	0.1325	0.1180	0.1086
95% C.I.	(1.495, 1.511)	(1.498, 1.512)	(1.522, 1.535)
SLSOPT	1.5059	1.5064	1.5073
RMSE	0.1346	0.1288	0.1271
95% C.I.	(1.498, 1.514)	(1.498, 1.514)	(1.499, 1.515)
$\sigma_u^2 = 16$			
SLSIDEN	15.9563	15.9154	16.0025
RMSE	0.7177	0.7166	0.7605
95% C.I.	(15.912, 16.001)	(15.871, 15.960)	(15.955, 16.050)
MLE	15.8526	15.8927	15.9100
RMSE	0.8185	0.8070	0.7935
95% C.I.	(15.803, 15.903)	(15.843, 15.942)	(15.861, 15.959)
SLSOPT	15.8393	15.8426	15.9053
RMSE	0.8120	0.7959	0.8141
95% C.I.	(15.790, 15.889)	(15.794, 15.891)	(15.855, 15.955)

Table 3
Simulation results for misspecified models with sample size $n = 200$ and 31% censoring

Error	$t(5)$			$\chi^2(4)$		
	Mean	RMSE	Bias	Mean	RMSE	Bias
$\beta_0 = -1.5$						
SLSIDEN	−1.7644	0.7664	0.2644	−1.3148	0.8223	0.1852
MLE	−1.7792	0.5409	0.2792	−1.8527	0.6024	0.3527
SLSOPT	−1.9290	0.6285	0.4290	−2.0917	0.7539	0.5917
$\beta_1 = 1.5$						
SLSIDEN	1.5153	0.1583	0.0153	1.4671	0.2041	0.0329
MLE	1.5560	0.1204	0.0560	1.5430	0.1309	0.0430
SLSOPT	1.5748	0.1447	0.0748	1.5582	0.1377	0.0582
$\sigma_u^2 = 16$						
SLSIDEN	15.7071	0.6819	0.2929	16.1360	0.6900	0.1360
MLE	15.2305	0.9006	0.7695	16.6462	0.8713	0.6462
SLSOPT	15.3186	0.8878	0.6814	16.7860	0.8889	0.7860

the SLSOPT have relatively high biases. In contrast to previous simulation results, in this case the SLSIDEN performs dramatically better in terms of bias for all the parameters and RMSE for σ_u^2. This is due to the fact that in SLSIDEN, the weighting matrix does not depend on the parameters which are poorly estimated because of misspecification. Although SLSIDEN is not as efficient as MLE and SLSOPT in the correct specification cases, it shows more robustness in misspecified cases.

5. Conclusions

This paper proposes the second-order least squares estimators for the censored regression models where the error term has a general parametric distribution. The asymptotic properties of the proposed estimators are derived. The finite sample properties of the estimators with two different weighting matrices are studied and compared with the maximum likelihood estimator over a range of error distributions and censoring degrees. We also compared the efficiency of the estimators with the MLE, in both correctly specified and misspecified error distributions. The results show that the proposed estimator with optimal weighting matrix performs very similar to the MLE, and the estimator with the identity weight is more robust against the misspecification. From a practical point of view, numerical computation of the proposed estimators is always feasible.

Acknowledgments

The authors are grateful to the editor and a referee for their comments and suggestions. This research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada (NSERC).

Appendix A. Assumptions

In order to prove the consistency and the asymptotic normality of the SLS $\hat{\gamma}_n$, we assume the following regularity conditions, where μ denotes the Lebesgue measure and $\|\cdot\|$ denotes the Euclidean norm.

A1. $f(u;\theta)$ is continuous in $\theta \in \Theta$ for μ-almost all u.

A2. The parameter space $\Gamma \subset \mathbb{R}^{p+q}$ is compact.

A3. The weight $W(X)$ is nonnegative definite with probability one and satisfies $E\|W(X)\|(Y^4 + \|X\|^4 + 1) < \infty$.

A4. $E\|W(X)\| \int (u^4 + 1)\sup_{\Theta} f(u,\theta)\,\mathrm{d}u < \infty$.

A5. $E[\rho(\gamma) - \rho(\gamma_0)]'W(X)[\rho(\gamma) - \rho(\gamma_0)] = 0$ if and only if $\gamma = \gamma_0$.

A6. There exists an open subset $\theta_0 \in \Theta_0 \subset \Theta$ in which $f(u;\theta)$ is twice continuously differentiable with respect to θ, for μ-almost all u. Furthermore, there exists positive function $\xi(u)$, such that the first two partial derivatives of $f(u;\theta)$ with respect to θ are bounded by $\xi(u)$, and $E\|W(X)\| \int (u^4 + 1)\xi^2(u)\,\mathrm{d}u < \infty$.

A7. The matrix

$$B = E\left[\frac{\partial \rho'(\gamma_0)}{\partial \gamma} W(X) \frac{\partial \rho(\gamma_0)}{\partial \gamma'}\right]$$

is nonsingular, where

$$\frac{\partial \rho'(\gamma_0)}{\partial \gamma} = -\left(\frac{\partial m_1(x;\gamma_0)}{\partial \gamma}, \frac{\partial m_2(x;\gamma_0)}{\partial \gamma}\right).$$

In the above, assumptions A1–A4 ensure that the objective function $Q_n(\gamma)$ is continuous and uniformly converges in γ. Additionally, assumption A5 means that the objective function $Q_n(\gamma)$ for large n attains a unique minimum at the true parameter value γ_0. Assumption A7 is necessary for the existence of the variance of the SLS $\hat{\gamma}_n$. Finally, assumption A6 guarantees uniform convergence of the second derivative of $Q_n(\gamma)$. This assumption and the Dominated Convergence Theorem together imply that the first derivatives $\partial m_1(X;\gamma)/\partial\gamma$ and $\partial m_2(X;\gamma)/\partial\gamma$ exist and their elements are, respectively, given by

$$\frac{\partial m_1(x;\gamma)}{\partial \theta} = \int_{-x'\beta}^{\infty} (x'\beta + u)\frac{\partial f(u;\theta)}{\partial \theta}\,\mathrm{d}u,$$

$$\frac{\partial m_1(x;\gamma)}{\partial \beta} = x\int_{-x'\beta}^{\infty} f(u;\theta)\,\mathrm{d}u$$

and

$$\frac{\partial m_2(x;\gamma)}{\partial \theta} = \int_{-x'\beta}^{\infty} (x'\beta + u)^2\frac{\partial f(u;\theta)}{\partial \theta}\,\mathrm{d}u,$$

$$\frac{\partial m_2(x;\gamma)}{\partial \beta} = 2x\int_{-x'\beta}^{\infty} (x'\beta + u) f(u;\theta)\,\mathrm{d}u.$$

With the modified expression (10) and (11), assumption A6 can be substituted by the following assumption.

A8. There exist an open subset $\gamma_0 \in \Gamma_0 \subset \Gamma$, in which $f(u - x'\beta;\theta)$ is twice continuously differentiable with respect to γ and u. Furthermore, there exists positive function $\xi(u,x)$, such that the first two partial derivatives of $f(u - x'\beta;\theta)$ with respect to θ and β are bounded by $\xi(u,x)$, and $E\|W(X)\| \int (u^4+1)\xi^2(u,X)\,du < \infty$.

Note that $Q_{n,S}(\gamma)$ is continuous in, and differentiable with respect to, γ, as long as function $f(u - x'\beta;\theta)$ has these properties. In particular, the first derivative of $\rho_{i,1}(\gamma)$ becomes

$$\frac{\partial \rho'_{i,1}(\gamma)}{\partial\gamma} = -\left(\frac{\partial m_{1,1}(x_i;\gamma)}{\partial\gamma}, \frac{\partial m_{2,1}(x_i;\gamma)}{\partial\gamma}\right),$$

where $\partial m_{1,1}(x_i;\gamma)/\partial\gamma$ is the column vector with elements

$$\frac{\partial m_{1,1}(x_i;\gamma)}{\partial\theta} = \frac{1}{S}\sum_{j=1}^{S}\frac{t_{ij}}{g(t_{ij})}\frac{\partial f(t_{ij}-x_i'\beta;\theta)}{\partial\theta},$$

$$\frac{\partial m_{1,1}(x_i;\gamma)}{\partial\beta} = \frac{1}{S}\sum_{j=1}^{S}\frac{t_{ij}}{g(t_{ij})}\frac{\partial f(t_{ij}-x_i'\beta;\theta)}{\partial\beta},$$

and $\partial m_{2,1}(x_i;\gamma)/\partial\gamma$ is the column vector with elements

$$\frac{\partial m_{2,1}(x_i;\gamma)}{\partial\theta} = \frac{1}{S}\sum_{j=1}^{S}\frac{t_{ij}^2}{g(t_{ij})}\frac{\partial f(t_{ij}-x_i'\beta;\theta)}{\partial\theta},$$

$$\frac{\partial m_{2,1}(x_i;\gamma)}{\partial\beta} = \frac{1}{S}\sum_{j=1}^{S}\frac{t_{ij}^2}{g(t_{ij})}\frac{\partial f(t_{ij}-x_i'\beta;\theta)}{\partial\beta}.$$

The derivatives $\partial m_{1,2}(x_i;\gamma)/\partial\gamma$ and $\partial m_{2,2}(x_i;\gamma)/\partial\gamma$ can be given similarly.

Appendix B. Proof of Theorem 1.1

First, assumption A1 and the Dominated Convergence Theorem imply that $m_1(X;\gamma)$, $m_2(X;\gamma)$ and therefore $Q_n(\gamma)$ are continuous in $\gamma \in \Gamma$. Let $Q(\gamma) = E\rho_1'(\gamma)W_1\rho_1(\gamma)$. Since by Hölder's inequality and assumptions A2–A4

$$\begin{aligned} E\|W_1\|\sup_\Gamma[Y_1 - m_1(X_1;\gamma)]^2 &\leqslant 2E\|W_1\|Y_1^2 + 2E\|W_1\|\sup_\Gamma m_1^2(X_1;\gamma) \\ &\leqslant 2E\|W_1\|Y_1^2 + 2E\|W_1\|\int \sup_{\Omega\times\Theta}(X_1'\beta+u)^2 f(u;\theta)\,du \\ &< \infty \end{aligned}$$

and

$$\begin{aligned} E\|W_1\|\sup_\Gamma[Y_1^2 - m_2(X_1;\gamma)]^2 &\leqslant 2E\|W_1\|Y_1^4 + 2E\|W_1\|\sup_\Gamma m_2^2(X_1;\gamma) \\ &\leqslant 2E\|W_1\|Y_1^4 + 2E\|W_1\|\int \sup_{\Omega\times\Theta}(X_1'\beta+u)^4 f(u;\theta)\,du \\ &< \infty, \end{aligned}$$

we have

$$\begin{aligned} E\sup_\Gamma\|\rho_1'(\gamma)W_1\rho_1(\gamma)\| &\leqslant E\|W_1\|\sup_\Gamma\|\rho_1(\gamma)\|^2 \\ &\leqslant E\|W_1\|\sup_\Gamma[Y_1 - m_1(X_1;\gamma)]^2 + E\|W_1\|\sup\nolimits_\Gamma[Y_1^2 - m_2(X_1;\gamma)]^2 \\ &< \infty. \end{aligned} \tag{14}$$

It follows from Jennrich (1969, Theorem 2) that $(1/n)Q_n(\gamma)$ converges almost surely to $Q(\gamma)$ uniformly in $\gamma \in \Gamma$. Further, since

$$E[\rho_1'(\gamma_0)W_1(\rho_1(\gamma) - \rho_1(\gamma_0))] = E[E(\rho_1'(\gamma_0)|X_1)W_1(\rho_1(\gamma) - \rho_1(\gamma_0))] = 0,$$

we have

$$Q(\gamma) = Q(\gamma_0) + E[(\rho_1(\gamma) - \rho_1(\gamma_0))' W_1 (\rho_1(\gamma) - \rho_1(\gamma_0))].$$

It follows that $Q(\gamma) \geqslant Q(\gamma_0)$ and, by assumption A5, equality holds if and only if $\gamma = \gamma_0$. Thus all conditions of Amemiya (1973, Lemma 3) hold and, therefore, $\hat{\gamma}_n \xrightarrow{\text{a.s.}} \gamma_0$ follows.

Appendix C. Proof of Theorem 1.2

By assumption A6 the first derivative $\partial Q_n(\gamma)/\partial\gamma$ exists and has a first-order Taylor expansion in a neighborhood $\Gamma_0 \subset \Gamma$ of γ_0. Since $\partial Q_n(\hat{\gamma}_n)/\partial\gamma = 0$ and $\hat{\gamma}_n \xrightarrow{\text{a.s.}} \gamma_0$, for sufficiently large n we have

$$0 = \frac{\partial Q_n(\gamma_0)}{\partial\gamma} + \frac{\partial^2 Q_n(\tilde{\gamma}_n)}{\partial\gamma\partial\gamma'}(\hat{\gamma}_n - \gamma_0), \tag{15}$$

where $\|\tilde{\gamma}_n - \gamma_0\| \leqslant \|\hat{\gamma}_n - \gamma_0\|$. The first derivative of $Q_n(\gamma)$ in (15) is given by

$$\frac{\partial Q_n(\gamma)}{\partial\gamma} = 2\sum_{i=1}^{n} \frac{\partial\rho_i'(\gamma)}{\partial\gamma} W_i \rho_i(\gamma),$$

where

$$\frac{\partial\rho_i'(\gamma)}{\partial\gamma} = -\left(\frac{\partial m_1(X_i;\gamma)}{\partial\gamma}, \frac{\partial m_2(X_i;\gamma)}{\partial\gamma}\right)$$

and the first derivatives of $m_1(X_i;\gamma)$ and $m_2(X_i;\gamma)$ with respect to γ are given after assumption A7. Therefore, by the Central Limit Theorem we have

$$\frac{1}{\sqrt{n}}\frac{\partial Q_n(\gamma_0)}{\partial\gamma} \xrightarrow{L} \mathrm{N}(0, 4A), \tag{16}$$

where

$$A = E\left[\frac{\partial\rho_1'(\gamma_0)}{\partial\gamma} W_1 \rho_1(\gamma_0)\rho_1'(\gamma_0) W_1 \frac{\partial\rho_1(\gamma_0)}{\partial\gamma'}\right].$$

The second derivative of $Q_n(\gamma)$ in (15) is given by

$$\frac{\partial^2 Q_n(\gamma)}{\partial\gamma\,\partial\gamma'} = 2\sum_{i=1}^{n}\left[\frac{\partial\rho_i'(\gamma)}{\partial\gamma} W_i \frac{\partial\rho_i(\gamma)}{\partial\gamma'} + (\rho_i'(\gamma)W_i \otimes I_{p+q})\frac{\partial \mathrm{vec}(\partial\rho_i'(\gamma)/\partial\gamma)}{\partial\gamma'}\right],$$

where

$$\frac{\partial\,\mathrm{vec}(\partial\rho_i'(\gamma)/\partial\gamma)}{\partial\gamma'} = -\left(\frac{\partial^2 m_1(X;\gamma)}{\partial\gamma\,\partial\gamma'}, \frac{\partial^2 m_2(X;\gamma)}{\partial\gamma\,\partial\gamma'}\right)'.$$

The elements in $\partial^2 m_1(x;\gamma)/\partial\gamma\,\partial\gamma'$ are

$$\frac{\partial^2 m_1(x;\gamma)}{\partial\theta\,\partial\theta'} = \int_{-x'\beta}^{\infty}(x'\beta + u)\frac{\partial^2 f(u;\theta)}{\partial\theta\,\partial\theta'}\,\mathrm{d}u,$$

$$\frac{\partial^2 m_1(x;\gamma)}{\partial\beta\,\partial\theta'} = x\int_{-x'\beta}^{\infty}\frac{\partial f(u;\theta)}{\partial\theta'}\,\mathrm{d}u,$$

$$\frac{\partial^2 m_1(x;\gamma)}{\partial\beta\,\partial\beta'} = xx' f(-x'\beta;\theta),$$

and the elements in $\partial^2 m_2(x;\gamma)/\partial\gamma\,\partial\gamma'$ are

$$\frac{\partial^2 m_2(x;\gamma)}{\partial\theta\,\partial\theta'} = \int_{-x'\beta}^{\infty}(x'\beta + u)^2\frac{\partial^2 f(u;\theta)}{\partial\theta\,\partial\theta'}\,\mathrm{d}u,$$

$$\frac{\partial^2 m_2(x;\gamma)}{\partial\beta\,\partial\theta'} = 2x\int_{-x'\beta}^{\infty}(x'\beta+u)\frac{\partial f(u;\theta)}{\partial\theta'}\,\mathrm{d}u,$$

$$\frac{\partial^2 m_2(x;\gamma)}{\partial\beta\,\partial\beta'} = 2xx'\int_{-x'\beta}^{\infty} f(u;\theta)\,\mathrm{d}u.$$

Analogous to the proof of Theorem 1.1, in the following we verify by assumption A6 and Jennrich (1969, Theorem 2) that $(1/n)\partial^2 Q_n(\gamma)/\partial\gamma\,\partial\gamma'$ converges almost surely to $\partial^2 Q(\gamma)/\partial\gamma\,\partial\gamma'$ uniformly in $\gamma\in\Gamma_0$. First, assumption A6 implies

$$E\sup_{\Gamma}\left\|\frac{\partial\rho_1'(\gamma)}{\partial\gamma}W_1\frac{\partial\rho_1(\gamma)}{\partial\gamma'}\right\|<\infty$$

and

$$\begin{aligned}
&E\sup_{\Gamma}\left\|(\rho_1'(\gamma)W_1\otimes I_{p+q})\frac{\partial\,\mathrm{vec}(\partial\rho_1'(\gamma)/\partial\gamma)}{\partial\gamma'}\right\|\\
&\leqslant\sqrt{2(p+q)}E\|W_1\|\sup_{\Gamma}\|\rho_1(\gamma)\|\left\|\frac{\partial\,\mathrm{vec}(\partial\rho_1'(\gamma)/\partial\gamma)}{\partial\gamma'}\right\|\\
&\leqslant\sqrt{2(p+q)}\left(E\|W_1\|\sup_{\Gamma}\|\rho_1(\gamma)\|^2 E\|W_1\|\sup_{\Gamma}\left\|\frac{\partial\,\mathrm{vec}(\partial\rho_1'(\gamma)/\partial\gamma)}{\partial\gamma'}\right\|^2\right)^{1/2}<\infty,
\end{aligned}$$

where the last inequality holds, because

$$E\left(\|W_1\|\sup_{\Gamma}\left\|\frac{\partial\,\mathrm{vec}(\partial\rho_1'(\gamma)/\partial\gamma)}{\partial\gamma'}\right\|\right)^2<\infty$$

and, by (14), $E\|W_1\|\sup_{\Gamma}\|\rho_1(\gamma)\|^2<\infty$. Therefore by Amemiya (1973, Lemma 4) we have

$$\begin{aligned}
\frac{1}{n}\frac{\partial^2 Q_n(\tilde\gamma_n)}{\partial\gamma\partial\gamma'} &\xrightarrow{\text{a.s.}} 2E\left[\frac{\partial\rho_1'(\gamma_0)}{\partial\gamma}W_1\frac{\partial\rho_1(\gamma_0)}{\partial\gamma'}+(\rho_1'(\gamma_0)W_1\otimes I_{p+q})\frac{\partial\,\mathrm{vec}(\partial\rho_1'(\gamma_0)/\partial\gamma)}{\partial\gamma'}\right]\\
&=2B,
\end{aligned}\tag{17}$$

where the second equality holds, because

$$\begin{aligned}
E\left[(\rho_1'(\gamma_0)W_1\otimes I_{p+q})\frac{\partial\,\mathrm{vec}(\partial\rho_1'(\gamma_0)/\partial\gamma)}{\partial\gamma'}\right] &= E\left[(E(\rho_1'(\gamma_0)|X_1)W_1\otimes I_{p+q})\frac{\partial\,\mathrm{vec}(\partial\rho_1'(\gamma_0)/\partial\gamma)}{\partial\gamma'}\right]\\
&=0.
\end{aligned}$$

It follows then from (15)–(17), assumption A7 and the Slutsky Theorem, that $\sqrt{n}(\hat\gamma_n-\gamma_0)\xrightarrow{L}\mathrm{N}(0,B^{-1}AB^{-1})$.

References

Abarin, T., Wang, L., 2006. Comparison of GMM with second-order least squares estimation in nonlinear models. Far East J. Theoret. Statist. 20 (2), 179–196.
Amemiya, T., 1973. Regression analysis when the dependent variable is truncated normal. Econometrica 41, 997–1016.
Austin, P.C., Ghali, W.A., Tu, J.V., 2003. A comparison of several regression models for analyzing cost of CABG surgery. Statist. Medicine 22, 2799–2816.
Bianco, A.M., Garcia Ben, M., Johai, V.J., 2005. Robust estimation for linear regression with asymmetric errors. Canad. J. Statist. 33, 511–528.
Heckman, J.J., MaCurdy, T.E., 1986. Labor econometrics. In: Griliches, Z., Intriligator, M.D. (Eds.), Handbook of Econometrics III. North-Holland, Amsterdam.
Jennrich, R., 1969. Asymptotic properties of non-linear least squares estimators. Ann. Math. Statist. 40, 633–643.
Killingsworth, M.R., Heckman, J.J., 1986. Female labor supply: a survey. In: Ashenfelter, O., Layard, R. (Eds.), Handbook of Labor Economics I. North-Holland, Amsterdam.
Marazzi, A., Yohai, V.J., 2004. Adaptive truncated maximum likelihood regression with asymmetric errors. J. Statist. Plann. Inference 122, 271–291.
Powell, J., 1984. Least absolute deviations estimation for the censored regression model. J. Econometrics 25, 303–325.
Powell, J., 1986. Symmetrically trimmed least squares estimation for Tobit models. Econometrica 54, 1435–1460.
Tobin, J., 1958. Estimation of relationships for limited dependent variables. Econometrica 26, 24–36.
Wang, L., 1998. Estimation of censored linear errors-in-variables models. J. Econometrics 84, 383–400.
Wang, L., 2004. Estimation of nonlinear models with Berkson measurement errors. Ann. Statist. 32, 2559–2579.
Wang, L., Leblanc, A., 2007. Second-order nonlinear least squares estimation. Ann. Inst. Statist. Math., to appear, doi:10.1007/s10463-007-0139-z.
Williams, M., 1997. A regression technique accounting for heteroscedastic and asymmetric errors. J. Agricultural Biol. Environ. Statist. 2, 108–129.