Statistica Sinica **19** (2009), 1013-1036

t-TYPE ESTIMATORS FOR A CLASS OF LINEAR ERRORS-IN-VARIABLES MODELS

Tao Hu and Hengjian Cui

Beijing Normal University

Abstract: This paper discusses t-type regression estimators of parameters for linear errors-in-variables model, and the EM algorithm for t-type estimators in linear errors-in-variables model is given. When the error variables for both the response and the manifest variables have a joint distribution that is spherically symmetric, but otherwise is unknown, the influence functions of the t-type regression estimates based on orthogonal residuals are calculated and the proposed estimators are shown to be consistent and asymptotically normal under some mild conditions. Simulation studies are conducted to examine the small-sample properties of the proposed estimates and a dataset is used to illustrate our approach.

Key words and phrases: Asymptotics, errors-in-variables, influence function, M-estimate, t-type regression estimator.

1. Introduction

The ordinary linear model $Y = X^{\tau}\beta_0 + \epsilon$ is one of the most mature and widely applied ways to explain how a dependent variable Y relates to independent variables X, where $\beta_0 \in R^p$ is a p-dimensional parameter, the errors ϵ are assumed to be independent and identically distributed (i.i.d.) with $E\epsilon = 0$. The popular choice is least square estimators (LSE) which corresponds to the maximum likelihood estimator (MLE) when the error ϵ is $N(0, \sigma^2)$. On the other hand, a robust estimator of β (for example, M-estimator) is more interesting in applications. In many situations, however, there exist covariate measurement errors. For example, it has been well documented that covariates such as blood pressure, urinary sodium chloride level, and exposure to pollutants are subject to measurement errors, and these cause difficulties in conducting a statistical analysis that involves them. A detailed study of linear models with measurement errors is in Fuller (1987). Carroll, Ruppert and Stefanski (1995) summarized much of the recent work for non-linear regression models with measurement errors. In the present paper we consider t-type regression estimators(He, Simpson and Wang (2000)) based on orthogonal residuals for the linear errors-in-variables model under spherically symmetry.

Consider the linear errors-in-variables (EV) model

$$\begin{cases} Y = x^\tau \beta_0 + \epsilon, \\ X = x + u, \end{cases} \tag{1.1}$$

where X and x are observable and unobservable random vectors on R^p, respectively. In this paper, $(\epsilon, u^\tau)^\tau$ is assumed to be $(p+1)$-dimensional spherically symmetric (this means that $(\epsilon, u^\tau)^\tau \stackrel{d}{=} RU_{p+1}$ where R is a nonnegative random variable, U_{p+1} is a uniform random vector on $\Omega_p = \{a : a \in R^{p+1}, \|a\| = 1\}$, where R and U_{p+1} are independent), and we suppose that $\sigma^2 = ER^2/(p+1) > 0$ is unknown, and that $(\epsilon, u^\tau)^\tau$ and x are independent. Note that $\stackrel{d}{=}$ means equal in distribution. For simplicity, we assume the intercept is zero. Spherical symmetry implies that ϵ and each component of u have the same distribution; this ensures model identifiability. A special case of such EV models with Gaussian errors and known variance ratio is frequently considered in the literature. Multivariate t-distributions are additional examples for this error structure (see, e.g., Cui (1997), and He and Liang (2000)).

We restrict ourselves to structural models in which x has independent and identically distributed coordinates. If x_i stems from non-stochastic designs, the model is said to have a functional relationship; see Fuller (1987) for details.

It is well known that in the EV model, ordinary least-squares (OLS) estimators are biased and inconsistent, and that orthogonal regression is better in that case; see Fuller (1987). However, both methods are very sensitive to outliers in the data and some robust alternatives have been proposed. Brown (1982) and Ketellapper and Ronner (1984) applied robust ordinary regression techniques in the EV model; Zamar (1989) proposed robust orthogonal regression M-estimators and showed that it outperforms the robust ordinary regression; Cheng and Van Ness (1992) generalize the proposal of Zamar by defining robust orthogonal Generalized M-estimators that have a bounded influence function in the simple case. In a more general context, He and Liang (2000) proposed a regression quantile approach in the EV model to allow for heavier-tailed errors distribution than Gaussian. Fekria and Ruiz-Gazen (2004) considered a class of weighted orthogonal regression estimators derived from robust estimators of multivariate location and scatter in the structural errors-in-variables (EV) model.

More recently, He, Simpson and Wang (2000) and He, Cui and Simpson (2004) proposed the weighted t-type regression estimator for linear model; that is, they viewed the components of the error vector ϵ_i as independent and identically distributed with a common t-distribution whose scale parameter and degrees of freedom are σ and ν, respectively. Then the t-type regression estimator $\hat{\beta}$ of β_0 is obtained by maximizing marginal likelihood of such a scaled t-type error distribution.

For the specific model (1.1), we consider t-type regression estimation for the linear EV model by applying the t-type likelihood to orthogonal residuals. In this situation, the t-type estimator of (β_0, σ_0) is

$$(\hat{\beta}, \hat{\sigma}) = \arg \min_{\beta \in R^p, \sigma > 0} \sum_{i=1}^{n} \Big[\rho\Big(\frac{Y_i - X_i^\tau \beta}{\sigma \sqrt{1 + \|\beta\|^2}} \Big) + \log(\sigma) \Big], \tag{1.2}$$

where $\rho(\cdot)$ could be $((\nu+1)/2)\log(1 + x^2/\nu)$, for examples, or another function that satisfies the conditions A1$-$A3 in Section 2.2. This weighted t-type estimation method is based on orthogonal residuals rather than on vertical distances in regression space. For the usefulness of orthogonal residual, see the examples and discussion found in Cheng and Van Ness (1992), Cui (1997), and He and Liang (2000), and the references therein.

If $\psi(\cdot) = \rho^{(1)}(\cdot)$ is the derivative of ρ and $\chi(\cdot) = x\psi(\cdot) - 1$, then

$$\begin{cases} \displaystyle\sum_{i=1}^{n} \psi\Big(\frac{Y_i - X_i^\tau \beta}{\sigma\sqrt{1 + \|\beta\|^2}} \Big) \Big(X_i + \frac{Y_i - X_i^\tau \beta}{1 + \|\beta\|^2} \Big) = 0, \\ \displaystyle\sum_{i=1}^{n} \chi\Big(\frac{Y_i - X_i^\tau \beta}{\sigma\sqrt{1 + \|\beta\|^2}} \Big) = 0. \end{cases} \tag{1.3}$$

If $(\epsilon_i, u_i^\tau)^\tau$ has t-distribution with degrees of freedom ν , the t-type regression estimator is just the maximum likelihood estimator (MLE) of β, σ. When $(\epsilon_i, u_i^\tau)^\tau$ is not a t, the t-type regression estimator corresponds to ordinary least square estimator. With a suitable $\rho(\cdot)$, the t-type regression estimator corresponds to the ordinary M-estimator.

What are the advantages of the t-type regression estimator over the ordinary M-estimator?

(i) Under the normal model, the t-type regression estimator is still quite efficient, and is quite robust even if the errors are far from normal (note that the degrees of freedom $\nu = 1 \sim 5$ may be used; see He, Simpson and Wang (2000), and He, Cui and Simpson (2004)).

(ii) With M-estimation equations, one can usually overcome the computing complexity of minimizing the objective function, but there may be misleading or non-unique solutions. The t-type estimator, however, can be obtained rapidly and stably by using EM algorithms directly to optimize the objective function. Such a solution is even unique for (1.3) (see Dempster, Laird and Rubin (1977), Little (1988), Lange, Little and Taylor (1989), Wu (1983), and He, Cui and Simpson (2004)).

Among developments in t-type regression estimator theory, He, Simpson and Wang (2000) proved that $\hat{\beta}$ is asymptotically normal under the assumption that

$\hat{\beta}$ is consistent; He, Cui and Simpson (2004) used the t-type regression estimator after de-correlating data in order to enhance the efficiency of the estimator, and Cui (2004) showed that a t-type estimator is consistent, which implies its normality. In this paper, we consider a robust estimator (t-type regression estimator) for linear errors-in-variables (EV) models. The influence functions of the proposed estimators are calculated. We also establish their consistency and asymptotic normality by applying modern empirical process theory (van der Vaart and Wellner (1996)).

The t-type regression estimation problem we consider differs from the robust M-estimation literature in several ways. First, one of our objectives is to estimate the regression parameters β and the scale parameter σ simultaneously in linear errors-in-variables (EV) models in order to keep a balance between robustness and efficiency. M-estimators with more general loss functions, such as those considered in Cheng and Van Ness (1992) are not scale equivariant unless a preliminary scale estimate is available. Second, Cheng and Van Ness (1992) derived bounded influence robust estimates for parameters in the univariate, Gaussian, structural EV model with the ratio of the error variances known. In this paper, $(\epsilon, u^{\tau})^{\tau}$ is assumed to be $(p+1)$-dimensional spherically symmetric. Simultaneous estimation of regression and scale with spherically symmetric errors brings out some technical complications in proving asymptotic results. Finally, the EM algorithm for t-type estimators in linear EV model is applied. Simulations show that performance of the t-type estimators proposed by the EM algorithm is quite good. Fekria and Ruiz-Gazen (2004) considered robust estimation in the structural errors-in-variables (EV) model from another angle and derived weighted orthogonal regression estimators from robust estimators of multivariate location and scatter such as M-estimators, S estimators, and the MCD estimator that may involve more complex computation than ours.

The rest of the paper is organized as follows. In Section 2 the influence functions of the proposed estimators are calculated. We also give some conditions under which the proposed method lead to strongly consistent estimators, and we derive the asymptotic distribution of those estimators. In Section 3, we derive the EM algorithm of the t-type estimator in a linear EV model. In Section 4, simulation studies and a data set are utilized to assess the robustness and efficiency of the proposals. All the proofs are postponed to the Appendix.

2. Influence Function and Asymptotic Properties

2.1. Influence function

The influence function measures the sensitivity of the functional to small amounts of contamination in the distribution. By definition, the influence function of the functional version of an estimator T at F is

$$IF(x,y;T,F)=\lim_{\delta\downarrow 0}\frac{T((1-\delta)F+\delta\Delta_{(x,y)})-T(F)}{\delta},$$

where $\Delta_{(x,y)}$ is a Dirac measure putting all its mass on (x,y). In other words, the influence function describes the effect of an infinitesimal contamination at the point (x,y) on the estimator, standardized by the mass of the contamination. If its influence function is bounded, the estimator is said to be robust. For more details and interpretation of the influence function, see Hampel, Ronchetti, Rousseeuw and Stahel (1986).

Suppose there are n independent and identically distributed observations $(X_1,Y_1),\ldots,(X_n,Y_n)$ from the model (1.1) with distribution F. It is convenient to write the estimators $\hat\beta$ $\hat\sigma$ in functional form as $\hat\beta=T(F_n)$ and $\hat\sigma=S(F_n)$, where F_n is the empirical distribution of $(X_1,Y_1),\ldots,(X_n,Y_n)$ and $T(F)$, $S(F)$ achieve

$$\min_{S,T}\int\Big[\rho\Big(\frac{y-x^\tau T(F)}{S(F)\sqrt{1+\|T(F)\|^2}}\Big)+\log(S(F))\Big]dF(x,y), \tag{2.1}$$

with F the joint distribution of (X,Y). Hereafter T and S are shorthand for $T(F)$ and $S(F)$. A defining relationship for β and σ can be obtained from (2.1) by differentiation:

$$\begin{cases}\int\psi\Big(\frac{v_1(x,y,T)}{S}\Big)v_2(x,y,T)dF(x,y)=0,\\ \int\chi\Big(\frac{v_1(x,y,T)}{S}\Big)dF(x,y)=0.\end{cases} \tag{2.2}$$

where

$$v_1(x,y,T)=\frac{y-x^\tau T}{\sqrt{1+\|T\|^2}} \tag{2.3}$$

$$v_2(x,y,T)=\Big(I-\frac{TT^\tau}{1+\|T\|^2}\Big)x+\frac{T}{1+\|T\|^2}y. \tag{2.4}$$

Since all the estimators discussed in this paper are Fisher consistent(see Lemma 5 in Appendix C.4), the notations T, S and β, σ will be used interchangeably as is common practice in the literature. Let $F_{(x,y),\epsilon}=(1-\epsilon)F+\epsilon\Delta_{(x,y)}$, where $\Delta_{x,y}$ is a Dirac measure putting all its mass on (x,y). The influence functions can be found straightforwardly by inserting $F_{(x,y),\epsilon}$ for F in (2.2), and then taking the derivative with respect to ϵ at $\epsilon=0$. We obtain that the two influence curves $IC(x,y;F,T)$ and $IC(x,y;F,S)$ satisfy the system of equations

$$\begin{pmatrix}D_{11} & D_{12}\\ D_{21} & D_{22}\end{pmatrix}\begin{pmatrix}IC(x,y;F,T)\\ IC(x,y;F,S)\end{pmatrix}=-\begin{pmatrix}\psi\Big(\frac{v_1(x,y,T)}{S}\Big)v_2(x,y,T)\\ \chi\Big(\frac{v_1(x,y,T)}{S}\Big)\end{pmatrix}S,$$

where

$$D_{11} = \int \psi'\Big(\frac{v_1(x,y,T)}{S}\Big) v_2(x,y,T)\frac{\partial v_1(x,y,T)}{\partial T^\tau} dF$$
$$+S\int \psi\Big(\frac{v_1(x,y,T)}{S}\Big)\frac{\partial v_2(x,y,T)}{\partial T^\tau} dF,$$
$$D_{12} = -\int \psi'\Big(\frac{v_1(x,y,T)}{S}\Big)\frac{v_1(x,y,T)}{S} v_2(x,y,T)^\tau dF,$$
$$D_{21} = \int \chi'\Big(\frac{v_1(x,y,T)}{S}\Big)\frac{\partial v_1(x,y,T)}{\partial T} dF,$$
$$D_{22} = -\int \chi'\Big(\frac{v_1(x,y,T))}{S}\Big)\frac{v_1(x,y,T)}{S} dF.$$

From (A.3.1) in the Appendix C.2, and Lemma 7 in the Appendix C.4, we can establish that $D_{11} = -E[\psi'(v_1(x,y,T)/S)]\Sigma_x/\sqrt{1+\|T\|^2}$. The spherically symmetric error distribution assumption implies $D_{12}=0$ and $D_{21}=0$ so some integrals vanish for reasons of symmetry and there are considerable simplifications:

$$IC(x,y;F,T) = -S\psi\Big(\frac{v_1(x,y,T)}{S}\Big) D_{11}^{-1} v_2(x,y,T),$$
$$IC(x,y;F,S) = -\frac{S}{D_{22}}\chi\Big(\frac{v_1(x,y,T)}{S}\Big).$$

Under regularity conditions described in Section 2.2 below, the estimators T and S defined in (1.2) are strongly consistent and asymptotically normal with asymptotic variance $\int IC(x,y;F,T)^2 dF(x,y)$ and $\int IC(x,y;F,S)^2 dF(x,y)$, respectively. Note that the influence function for scale is bounded and the influence function for regression is only bounded for the response variable. Therefore, we could consider the generalized M-estimator for EV models that was studied by Cheng and Van Ness (1992) in the univariate Gaussian structural EV model, but leave this for further study.

2.2. Asymptotic properties

First, some conditions are listed.

A1. $\rho(\cdot)$ is symmetric about 0 and continuously increasing from 0 to infinity on $[0,\infty)$.

A2. ψ has a bounded and continuous second-order derivative $\psi^{(2)}$, with $\sup_{t\in R}(1+|t|^i)\|\psi^{(i-1)}\| < \infty$ for $i=1,2$.

A3. $\chi(\cdot)$ is an increasing function of $|\cdot|$ with $-1=\chi(0) < \lim_{|t|\to\infty}\chi(t) = \kappa > 0$, and there exists a $\delta>0$ such that $|\chi'(y)|>0$ for all $y\in(-\delta,\delta)\backslash 0$.

B1. x is non-degenerate, and $E\rho[\alpha(R+\|x\|)] < +\infty$, $\forall|\alpha| > 1$.

B2. $P\{R=0\}=0$, $\Sigma_x = E(xx^\tau) > 0$ and $E(\|x\|+R^2)[1+\sup_{|t|\le R+|x|}\psi^2(\epsilon/\sigma+t)] < +\infty$.

C1. $f_0(\cdot)$ is a non-increasing function, where $f_0(\cdot)$ is the true probability density function.

C2. The discontinuous points of $\psi(\cdot)$ are at most countable, $E\psi(\epsilon/\sigma)=0$, and $E\psi^2(\epsilon/\sigma) < \infty$ for any $\sigma > 0$.

Conditions A1$-$A3 concern ρ, conditions B1$-$B2 concern x; and C1$-$C2 bring in the error distribution. It is easy to verify that $\rho(x) = ((\nu+1)/2)\log(1+x^2/\nu)$ satisfies A1$-$A3.

Theorem 1. *Suppose* A1$-$A3, B1, *and* C1$-$C2 *are satisfied. Then* $\hat\beta \xrightarrow{a.s.} \beta_0$, *and* $\hat\sigma \xrightarrow{a.s.} \sigma_0$ *as* $n\to\infty$, *where* σ_0 *satisfies* $E[\chi(\epsilon/\sigma_0)]=0$.

Theorem 2. *Suppose* A1$-$A3, B2, *and* C1$-$C2 *hold. Then, as* $n\to\infty$,

$$\frac{E\psi'(\epsilon/\sigma_0)}{\sigma_0\sqrt{1+\|\beta_0\|^2}}\sqrt{n}\Sigma^{-1/2}\Sigma_x(\hat\beta-\beta_0) \xrightarrow{d} N(0,I_p),$$

$$\sigma_0^{-2}A^{-1/2}E(e(\beta_0)\chi'(e(\beta_0)/\sigma_0))n^{-1/2}(\hat\sigma-\sigma_0) \xrightarrow{d} N(0,1),$$

where $\xrightarrow{d}$ *denotes convergence in distribution,* $\Sigma = E\psi^2(\epsilon/\sigma_0)\Sigma_x + E[\psi^2(\epsilon/\sigma_0)u_{11}^2]$ $(I_p-\beta_0\beta_0^\tau/1+\|\beta_0\|^2)$, $e_i(\beta_0) = (\epsilon_i - u_i^\tau\beta_0)/\sqrt{1+\|\beta_0\|^2}$, $A = E\chi^2(e(\beta_0)/\sigma_0)$, u_{11}^2 *is the first component of* u, *and* σ_0 *satisfies* $E[\chi(\epsilon/\sigma_0)]=0$.

The proofs of Theorems 1 and 2 are given in the Appendix. One may prove Theorem 2 based on the general result of He and Shao (1996); however, for the specific $\rho(x)$, one may use the direct proof as given in the Appendix.

3. EM Algorithm for t-type Estimators in the Linear EV Model

In this section, we use an EM algorithm to compute the t-type estimators in the linear EV model defined at (1.2). Consider the linear EV model

$$Y_i = x_i^\tau\beta + \epsilon_i, \quad X_i = x_i + u_i, \tag{3.1}$$

where the (X_i, Y_i), $i=1,\ldots,n$ are observed. Assume

$$e_i = (\epsilon_i, u_i^\tau)^\tau = (Y_i - x_i^\tau\beta, (X_i-x_i)^\tau)^\tau \overset{i.i.d.}{\sim} t_{p+1}(0,\sigma^2 I_{p+1},\nu),\ i=1,\ldots,n,$$

with density denoted by
$f_{p+1}(y|0,\sigma^2 I_{p+1},\nu) = C(v)(\sigma^2)^{-(p+1)/2}(1+\|y\|^2/(\nu\sigma^2))^{-(\nu+p+1)/2}$, and $C(\nu) =$

$\Gamma((v+p+1)/2)/[\Gamma(1/2)^{p+1}\Gamma(\nu/2)\nu^{(p+1)/2}]$. Then

$$\begin{aligned}\log(f_{p+1}(y|0,\sigma^2 I_{p+1},\nu)) &\propto -(p+1)\log(\sigma)-\frac{\nu+p+1}{2}\log\Big(1+\frac{\|y\|^2}{\nu\sigma^2}\Big)\\ &= -(p+1)\Big[\rho(\|y\|/\sigma)+\log(\sigma)\Big].\end{aligned}$$

The MLE of β and the scale parameter σ is

$$\begin{aligned}(\hat\beta,\hat\sigma)_{MLE} &= \arg\max_{\beta,x_i\in R^p,\sigma>0}\Pi_{i=1}^n f_{p+1}(Y_i-x_i^\tau\beta, X_i-x_i|0,\sigma^2,\nu)\\ &= \arg\max_{\beta,x_i\in R^p,\sigma>0}\Big[\sum_{i=1}^n \log f_{p+1}(Y_i-x_i^\tau\beta, X_i-x_i|0,\sigma^2,\nu)\Big]\\ &= \arg\min_{\beta,x_i\in R^p,\sigma>0}\Big[\rho\Big(\sqrt{\frac{(Y_i-x_i^\tau\beta)^2+\|X_i-x_i\|^2}{\sigma}}\Big)+\log(\sigma)\Big]\\ &= \arg\min_{\beta\in R^p,\sigma>0}\sum_{i=1}^n\Big[\rho\Big(\frac{Y_i-X_i^\tau\beta}{\sigma\sqrt{1+\|\beta\|^2}}\Big)+\log(\sigma)\Big]. \qquad (3.2)\end{aligned}$$

Write $e_i \stackrel{d}{=} \sigma z_i\sqrt{q_i}$, where $z_i \sim N_{p+1}(0,I_{p+1})$, $q_i\sim\chi^2_\nu/\nu$, z_i, and the q_i are independent. Then

$$e_i|q_i \sim N_{p+1}\Big(0,\frac{\sigma^2 I_{p+1}}{q_i}\Big)\propto \frac{q_i^{(p+1)/2}}{\sigma^{p+1}}\exp\Big\{-\frac{q_i[(Y_i-x_i^\tau\beta)^2+\|X_i-x_i\|^2]}{2\sigma^2}\Big\}=Q_i,$$

$$L_i=\log Q_i=\frac{p+1}{2}\log(q_i)-\frac{p+1}{2}\log(\sigma^2)-\frac{q_i[(Y_i-x_i^\tau\beta)^2+\|X_i-x_i\|^2]}{2\sigma^2}.$$

Because $q_i|e_i\sim\chi^2_{\nu+p+1}/(\nu+\delta_i^2)$, where $\delta_i^2=[(Y_i-x_i^\tau\beta)^2+\|X_i-x_i\|^2]/\sigma^2$, we get

$$E(q_i)=\frac{\nu+p+1}{\nu+\delta_i^2}=\frac{\nu+p+1}{\nu+[(Y_i-x_i^\tau\beta)^2+\|X_i-x_i\|^2]/\sigma^2}.$$

We use an EM algorithm to get the MLE of (β,σ) in model (3.1). For the E-step,

$$\begin{aligned}F(\beta,\sigma^2,x_1,\cdots,x_n) &= \sum_{i=1}^n E[L_i|\beta^{(t)},\sigma^{(t)},x_i^{(t)}]\\ &= \sum_{i=1}^n\Big\{\frac{p+1}{2}E[\log(q_i)|\beta^{(t)},\sigma^{(t)},x_i^{(t)}]\\ &\quad -\frac{p+1}{2}E[\log(\sigma^2)|\beta^{(t)},\sigma^{(t)},x_i^{(t)}]\\ &\quad -E\Big[\frac{q_i[(Y_i-x_i^\tau\beta)^2+\|X_i-x_i\|^2]}{2\sigma^2}\Big|\beta^{(t)},\sigma^{(t)},x_i^{(t)}\Big]\Big\}\end{aligned}$$

$$= \sum_{i=1}^{n} \left\{ \frac{p+1}{2} E\Big[\log(q_i) | \beta^{(t)}, \sigma^{(t)}, x_i^{(t)} \Big] - \frac{p+1}{2} \log(\sigma^2) - \frac{h_i^{t+1}[(Y_i - x_i^\tau \beta)^2 + \|X_i - x_i\|^2]}{2\sigma^2} \right\},$$

where

$$h_i^{(t+1)} = E[q_i | \beta^{(t)}, \sigma^{(t)}, x_i^{(t)}] = \frac{\nu + p + 1}{\nu + [(Y_i - x_i^{(t)\tau} \beta^{(t)})^2 + \|X_i - x_i^{(t)}\|^2]/(\sigma^2)^{(t)}}.$$

The M-step, differentiating $F(\beta, \sigma^2, x_1, \ldots, x_n)$ with respect to β, σ^2, and x_i, respectively, and using the formulas $\partial(AX)/\partial X = A^\tau$ and $\partial(A^\tau X A)/\partial X = (A + A^\tau)X$, we can derive the EM equations

$$\begin{cases} \beta^{(t+1)} = \left[\sum_{i=1}^{n} h_i^{(t+1)} x_i^{(t)} x_i^{(t)\tau} \right]^{-1} \sum_{i=1}^{n} h_i^{(t+1)} Y_i x_i^{(t)}, \\ x_i^{(t+1)} = \left(I_p - \dfrac{\beta^{(t+1)} \beta^{(t+1)\tau}}{1 + \|\beta^{(t+1)}\|^2} \right) [Y_i \beta^{(t+1)} + X_i], \\ (\sigma)^{(t+1)} = \left\{ \sum_{i=1}^{n} \dfrac{h_i^{(t+1)}[(Y_i - x_i^{(t+1)\tau} \beta^{(t+1)})^2 + \|X_i - x_i^{(t+1)}\|^2]}{(p+1)n} \right\}^{1/2}, \end{cases} \tag{3.3}$$

for $i = 1, \ldots, n$, where the expression of $x_{(t+1)}$ is from $\partial F/\partial x_i = h_i^{(t+1)} \partial[(Y_i - x_i^\tau \beta)^2 + \|X_i - x_i\|^2] \partial x_i = 0$, i.e., $(Y_i - x_i^\tau \beta)\beta + (X_i - x_i) = 0$ and

$$x_i^{(t+1)} = [\beta^{(t+1)} \beta^{(t+1)\tau} + I_p]^{-1} (Y_i \beta^{(t+1)} + X_i) = \Big(I_p - \frac{\beta^{(t+1)} \beta^{(t+1)\tau}}{1 + \|\beta^{(t+1)}\|^2} \Big) [Y_i \beta^{(t+1)} + X_i].$$

As long as the initial values are given, we can use (3.3) for iteration and the MLE of (β, σ) is figured out approximately. If the distribution $(\epsilon_i, u_i^\tau)^\tau$ is not a t-distribution with ν degrees of freedom at (1.1), we still use (3.2) to obtain the estimator of β, σ as the t-type estimator defined in (1.2). The computations of such t-type estimator can use (3.3). We suggest the L_1 estimator, $\arg\min_{\beta \in R^p} \sum_{i=1}^{n} |Y_i - X_i^\tau \beta|$, as the initial value $\beta^{(0)}$ and $MAD\{Y_i - X_i^\tau \beta^{(0)}\}/0.6745$ as the initial value $\sigma^{(0)}$ of σ.

4. Simulations, Comparisons and Examples

4.1. Simulation 1

As an example, we took the linear EV model (1.1) to be $Y_i = x_i^\tau \beta + \epsilon_i$, $X_i = x_i + u_i$, with $p = 2$, $n = 100$, $\nu = 5$, $\beta = (1, 2)^\tau$ and $x = (i, i^2)^\tau/(3n)$, where $(\epsilon_i, u_i^\tau)^\tau$ were i.i.d. generated from $t_3(0, \sigma^2 I_3, 5)$ with $\sigma = 0.2$. We used the EM algorithm (the standard program can be packaged by Splus and R) described

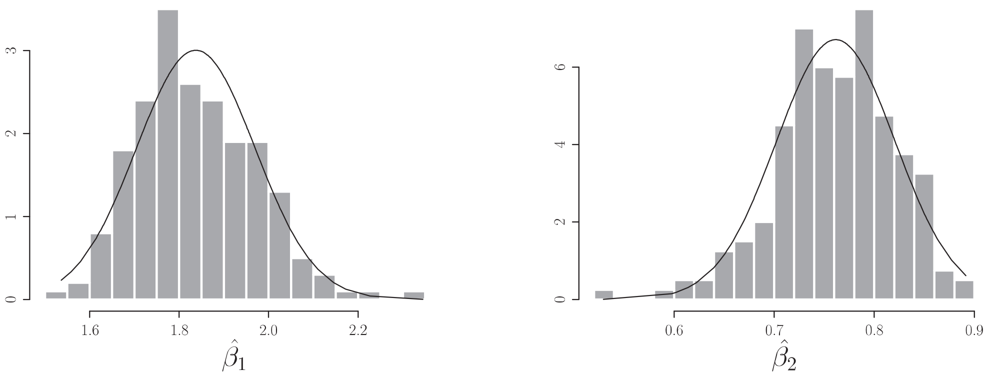

Figure 1. The histograms of $\hat{\beta}_1$ and $\hat{\beta}_2$.

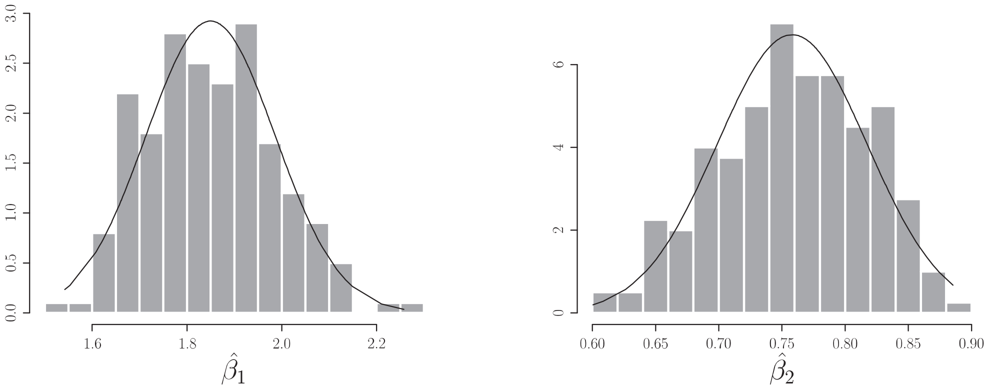

Figure 2. The histograms of $\hat{\beta}_1$ and $\hat{\beta}_2$.

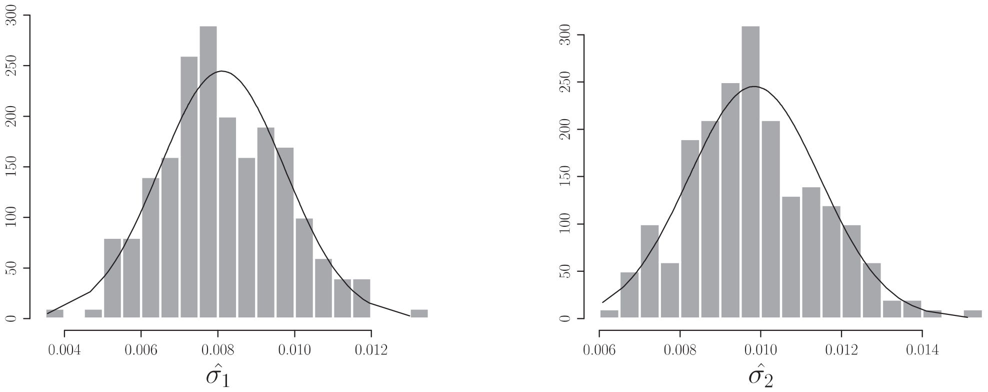

Figure 3. The histograms of $\hat{\sigma_1}$ and $\hat{\sigma_2}$.

Table 1. Comparison of MSE, bias, std with LSE for Simulation 1.

Error model of ϵ	Estimator	MSE($\hat{\beta}$)	bias($\hat{\beta}_1$)	std($\hat{\beta}_1$)	bias($\hat{\beta}_2$)	std($\hat{\beta}_2$)
$N(0, 0.2/3)$	T-type	0.0301	-0.0227	0.1580	-0.0016	0.0696
	(LSE)	(0.0243)	(-0.0419)	(0.1383)	(0.0109)	(0.0593)
T_0	T-type	0.0251	-0.0304	0.1423	0.0003	0.0649
	(LSE)	(0.0272)	(-0.0485)	(0.1462)	(0.0120)	(0.0603)
T_1	T-type	0.0328	0.0140	0.1633	-0.0068	0.0766
	(LSE)	(0.1190)	(0.2315)	(0.2203)	(-0.0937)	(0.0902)
T_2	T-type	0.0788	0.0760	0.2242	-0.0368	0.1147
	(LSE)	(0.2989)	(0.3626)	(0.2714)	(-0.1503)	(0.1149)
T_3	T-type	0.4128	0.1846	0.5439	-0.0899	0.2736
	(LSE)	(2.2302)	(1.1691)	(0.7269)	(-0.4881)	(0.3111)

in (3.3) to get the t-type estimators $\hat{\beta}$ and $\hat{\sigma}$. We replicated 200 times to get the histograms for $\hat{\beta}$, see Figure 1, and to find the mean and std of β to be $(0.996, 2.002)^\tau$ and (0.155, 0.065), respectively. With $(\epsilon_i, u_i^\tau)^\tau$ i.i.d. generated from $t_3(0, \sigma^2 I_3, 10)$, the histograms for $\hat{\beta}$ are in Figure 2, while the mean and std of $\hat{\beta}$ were (1.011, 1.994) and (0.150, 0.065), respectively. The histograms for $\hat{\sigma}$ are in Figure 3. From Figures 1−3, we see that the t-type estimators of $\hat{\beta}$ and $\hat{\sigma}$ by EM algorithm were close to the true values and the distributions were approximately normal, thus performance looks quite good.

To see how the robust estimators protect us from gross errors in the data, we considered the model $Y_i = x_i^\tau \beta + \epsilon_i$, $X_i = x_i + u_i$ with $p = 2$, $n = 100$, $\beta = (1, 2)^\tau$ and $x = (10i, 30i^2)^\tau/(3n)$, where $(\epsilon_i, u_i^\tau)^\tau$ are i.i.d. and u_i generated from $t_2(0, \sigma^2 I_2)$ with $\sigma = 0.2$, where the error variable ϵ takes one of the following forms:

$$N\Big(0, \frac{0.2}{3}\Big), \quad T_0 = t(0, 0.2^2, 5), \quad T_1 = 0.90t(0, 0.2^2, 5) + 0.10t(0, 8^2, 5),$$

$$T_2 = 0.80t(0, 0.2^2, 5) + 0.20t(0, 12^2, 5), \quad T_3 = 0.70t(0, 0.2^2, 5) + 0.30t(0, 16^2, 5).$$

Note that $N(0, 0.2/3)$ has the same variance as T_0. We used the ρ function from the t with $\nu = 5$ degrees of freedom. So when ϵ is from T_0, the t-type estimates are just the MLE. To measure performance, we used the bias of $\hat{\beta}$ and standard deviation for the $\hat{\beta}$ as well as the mean square error of the $\hat{\beta}$, MSE($\hat{\beta}$). The bias, standard errors (std), and MSE of these estimators was computed from 100 Monte Carlo samples and compared with those of the ordinary least squares estimator (LSE). The results of the simulation study are given in Table 1. Our studies showed good performance of t-type estimators in the presence of outliers.

4.2. Simulation 2

The data were generated from the model $Y_i = x_i^\tau \beta + \epsilon_i$, $X_i = x_i + u_i$, with the covariate vector $x = (x_1, x_2)^\tau$ simulated from a normal distribution with mean

Table 2. Comparison of MSE with LAD-OR for Simulation 2.

Error model of ϵ	Estimator	MSE($\hat{\beta}$)
$t(0, 0.2^2, 5)$	T-type	0.0081
	(LAD-OR)	(0.0104)
$N(0, 0.2/3)$	T-type	0.0115
	(LAD-OR)	(0.0144)

Table 3. Regression estimation for Body height and Arm length data.

	LSE-OR	LSE-OR without 18, 24	LAD-OR	T-type
$\hat{\beta}_0$	0.9594	1.005	1.002	1.003

zero and $\text{cov}(x_i, x_j) = 0.5^{|i-j|}$, where ϵ_i were i.i.d. generated from $t(0, \sigma^2, 5)$ with $\sigma = 0.2$ and $N(0, 0.2/3)$. In this section, we also used the ρ function from a t with $\nu = 5$ degrees of freedom.

For Simulation 2, we give a comparison with the estimators proposed by He and Liang (2000) that minimize $n^{-1}\sum_i \rho_\tau((Y_i - X_i\beta)/(\sqrt{1 + \|\beta\|^2})$, where $\rho_\tau(r) = \tau \max\{r, 0\} + (1 - \tau)\max\{-r, 0\}$. In this simulation, we took $\tau = 0.5$. The sample size n was set to 100. The number of simulated realizations was 200. To measure performance, we use MSE($\hat{\beta}$), see Table 2. The estimates are labeled LAD-OR and T-type. It can be seen that the T-type are comparable with LAD-OR. We also notice the quantile estimators based on orthogonal residuals involve complicated functional optimization because the criterion function is nondifferentiable, while the t-type estimators are faster by virtue of the EM-type algorithm.

4.3. Example

We consider the relationship between body height and arm span using a sample of $n = 39$ observations. This data set comes from the junior class of statistics majors in Beijing Normal University, China. Students measured their own body height and arm span. We take Y_i as the height, and X_i as the arm length of the ith person. Here, we observe Y and X as data with measurement errors. Figure 4 gives the scatter-plot of Y and X and it shows an approximate linear relationship between Y and X, where X is the observed variable of x with error u, whose variance is assumed to be approximately equal to the variance of error due to the same measurement tool being used. In Figure 4, note the outlying points #18 and #24. We model the relationship between Y and X as $Y = x\beta_0 + \epsilon$, $X = x + u$. Table 3 gives the estimated parameters for LSE based on orthogonal residuals with complete data(labeled as LSE-OR in Table 3) and after removing the two points #18 and #24 (labeled as LSE-OR without #18,

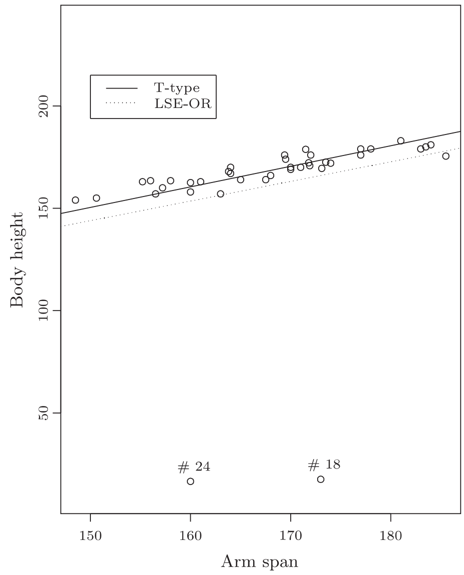

Figure 4. Comparison of the fitted straight line using LSE-OR and T-type.

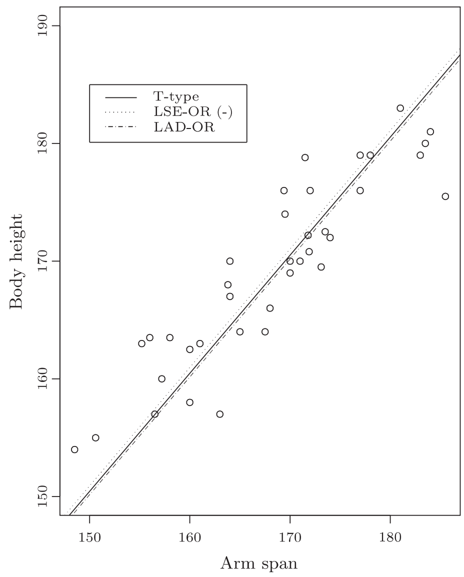

Figure 5. Comparison of the fitted straight line using LSE-OR without outliers (LSE-OR(-)), T-type and LAD-OR.

#24 in Table 3). We compared quantile estimates based on orthogonal residuals with $\tau = 0.5$ as proposed by He and Liang (2000) (labeled as LAD-OR in Table 3 and Figures 4, 5) and the t-type estimates studied here (labeled as T-type in

Table 3 and Figures 4, 5). From Table 3, we can see that the t-type estimates are comparable to LAD-OR; LAD-OR, and t-type the are very similar to those of the LSE-OR without #18, #24, and they give a good fit to the data without being perturbed by a small proportion of outliers. Figure 4 also gives the comparison of the fitted straight line using LSE-OR and t-type estimates and shows t-type estimates may have better performance than LSE-OR in the presence of outliers. Furthermore, Figure 5 gives the comparison of the fitted straight line using LSE without outliers (#18, #24), t-type, and LSE-OR. From Figure 5, we can see that the t-type estimates are basically consistent with LSE based on orthogonal residuals without outliers.

Appendix. Proofs

In this section, we give the proofs of Theorems 1 and 2.

A.1. Outline of Proofs

The proof of consistency can be established by verifying steps (i), (ii). (i) We use the Maximal Inequality of Pollard (1990) and the concept of polynomial class of set (VC subgraph class), see Pollard (1984), and van der Vaart and Wellner (1996), to show

$$\sup_{\substack{\beta\in R^p\\ \sigma\in[\sigma_1,\sigma_2]}}\left|\frac{1}{n}\sum_{i=1}^{n}\left[\rho\Big(\frac{Y_i-X_i^{\tau}\beta}{\sigma\sqrt{1+\|\beta\|^2}}\Big)+\log(\sigma)\right]-E\left[\rho\Big(\frac{Y_i-X_i^{\tau}\beta}{\sigma\sqrt{1+\|\beta\|^2}}\Big)+\log(\sigma)\right]\right|\xrightarrow{\text{a.s.}}0,$$

where σ_1 and σ_2 to be specified in Lemma 1. (ii) We show that (β_0,σ_0) is a unique minimum of $E[\rho((Y_i-X_i^{\tau}\beta)/(\sigma\sqrt{1+\|\beta\|^2})+\log(\sigma)]$. The proof of the consistency in this paper is similar to the proof of the consistency of the M-estimator in van der Vaart and Wellner (1996). The proof of asymptotic normality is based on Huber's Z-theorem (see Pollard (1985), and van der Vaart and Wellner (1996)).

A.2. Proof of Theorem 1

Let $K_0=\{(a,b,c):\|a\|\le 1/\sigma_1,\ b\in R^p,\ \|b\|\le(1+\|\beta_0\|)/\sigma_1,|c|\le\log(\sigma_2/\sigma_1)\}$,

$$\mathcal{F}=\Big\{\rho[(\epsilon,u^{\tau})^{\tau}a+x^{\tau}b]+c:(a,b,c)\in K_0\Big\},$$

where σ_1 and σ_2 to be specified in Lemma 1. Since $\rho(t)$ is a non-increasing function when $t<0$, and is a non-decreasing function when $t\ge 0$, the graphs of functions in $\mathcal{F}$ form a polynomial class of set (VC subgraph class), see Pollard (1984), and van der Vaart and Wellner (1996), with envelope $F=\rho[(1+\|\beta_0\|)(R+$

$\|x\|)/\sigma_1] + \rho[-(1+\|\beta_0\|)(R+\|x\|)/\sigma_1] + \log(\sigma_2/\sigma_1)$, and $E(F) < \infty$. Therefore, by Theorem 24 of Chapter II.5 in Pollard (1984), we can get that

$$\sup_{(a,b,c)\in K_0}\Bigg|n^{-1}\sum_{i=1}^{n}\Big[\rho\Big(\frac{(\epsilon_i, u_i^\tau)a + x_i^\tau b}{\sigma}\Big) + \log(\sigma)\Big] - E\Big[\rho\Big(\frac{(\epsilon_i, u_i^\tau)a + x_i^\tau b}{\sigma}\Big) + \log(\sigma)\Big]\Bigg| \to 0 \text{ a.s.}$$

as $n \to \infty$.

It follows that

$$\begin{aligned}
&\sup_{\substack{\beta\in R^p\\ \sigma\in[\sigma_1,\sigma_2]}}\Bigg|\frac{1}{n}\sum_{i=1}^{n}\Big[\rho\Big(\frac{Y_i - X_i^\tau\beta}{\sigma\sqrt{1+\|\beta\|^2}}\Big) + \log(\sigma)\Big] - E\Big[\rho\Big(\frac{Y_i - X_i^\tau\beta}{\sigma\sqrt{1+\|\beta\|^2}}\Big) + \log(\sigma)\Big]\Bigg| \\
&= \sup_{\substack{\beta\in R^p\\ \sigma\in[\sigma_1,\sigma_2]}}\Bigg|\frac{1}{n}\sum_{i=1}^{n}\Big[\rho\Big(\frac{Y_i - X_i^\tau\beta}{\sigma\sqrt{1+\|\beta\|^2}}\Big) + \log(\sigma)\Big] \\
&\quad - E\rho\Big[\frac{1}{\sigma}\Big(\epsilon - \frac{x^\tau(\beta-\beta_0)}{\sqrt{1+\|\beta\|^2}}\Big) + \log(\sigma)\Big]\Bigg| \\
&= \sup_{\substack{\beta\in R^p\\ \sigma\in[\sigma_1,\sigma_2]}}\Bigg|\frac{1}{n}\sum_{i=1}^{n}\Big[\rho\Big(\frac{\epsilon_i - u_i^\tau\beta}{\sigma\sqrt{1+\|\beta\|^2}} - \frac{x_i^\tau(\beta-\beta_0)}{\sigma\sqrt{1+\|\beta\|^2}}\Big) + \log(\sigma)\Big] \\
&\quad - E\Big[\rho\Big(\frac{\epsilon - u^\tau\beta}{\sigma\sqrt{1+\|\beta\|^2}} - \frac{x^\tau(\beta-\beta_0)}{\sigma\sqrt{1+\|\beta\|^2}}\Big) + \log(\sigma)\Big]\Bigg| \\
&\le \sup_{K_0}\Bigg|n^{-1}\sum_{i=1}^{n}\Big[\rho\Big(\frac{(\epsilon_i, u_i^\tau)a + x_i^\tau b}{\sigma}\Big) + c\Big] \\
&\quad - E\Big[\rho\Big(\frac{(\epsilon_i, u_i^\tau)a + x_i^\tau b}{\sigma}\Big) + c\Big]\Bigg| \xrightarrow{\text{a.s.}} 0. \qquad \text{(A.2.1)}
\end{aligned}$$

With $K_\epsilon = (\{\beta : \|\beta - \beta_0\| \ge \epsilon\} \otimes \{\sigma : |\sigma - \sigma_0| \ge \epsilon\}) \cap (R^p \otimes [\sigma_1, \sigma_2])$,

$$\zeta_{1n} = \sup_{\substack{\beta\in R^p\\ \sigma\in[\sigma_1,\sigma_2]}}\Bigg|\frac{1}{n}\sum_{i=1}^{n}\Big[\rho\Big(\frac{Y_i - X_i^\tau\beta}{\sigma\sqrt{1+\|\beta\|^2}}\Big) + \log(\sigma)\Big] - E\Big[\rho\Big(\frac{Y_i - X_i^\tau\beta}{\sigma\sqrt{1+\|\beta\|^2}}\Big) + \log(\sigma)\Big]\Bigg|,$$

$$\zeta_{2n} = \sup\Bigg|\frac{1}{n}\sum_{i=1}^{n}\Big[\rho\Big(\frac{Y_i - X_i^\tau\beta_0}{\sigma_0\sqrt{1+\|\beta_0\|^2}}\Big) + \log(\sigma_0)\Big]$$

$$-E\Big[\rho\Big(\frac{Y_i - X_i^\tau\beta_0}{\sigma_0\sqrt{1+\|\beta\|^2}}\Big) + \log(\sigma_0)\Big]\Bigg|.$$

By (A.2.1), we have $\zeta_{1n} = o(1)$ a.s. and $\zeta_{2n} = o(1)$ a.s. But

$$\begin{aligned}
&\inf_{K_\epsilon} E\Big[\rho\Big(\frac{Y_i - X_i^\tau\beta}{\sigma\sqrt{1+\|\beta\|^2}}\Big) + \log(\sigma)\Big]\\
&\quad= \inf_{K_\epsilon} E\Big[\rho\Big(\frac{Y_i - X_i^\tau\beta}{\sigma\sqrt{1+\|\beta\|^2}}\Big) + \log(\sigma)\Big] - \inf_{K_\epsilon}\frac{1}{n}\sum_{i=1}^n\Big[\rho\Big(\frac{Y_i - X_i^\tau\beta}{\sigma\sqrt{1+\|\beta\|^2}}\Big) + \log(\sigma)\Big]\\
&\qquad+ \inf_{K_\epsilon}\frac{1}{n}\sum_{i=1}^n\Big[\rho\Big(\frac{Y_i - X_i^\tau\beta}{\sigma\sqrt{1+\|\beta\|^2}}\Big) + \log(\sigma)\Big]\\
&\quad\le \zeta_{1n} + \inf_{K_\epsilon}\frac{1}{n}\sum_{i=1}^n\Big[\rho\Big(\frac{Y_i - X_i^\tau\beta}{\sigma\sqrt{1+\|\beta\|^2}}\Big) + \log(\sigma)\Big].
\end{aligned}$$

If $(\hat\beta,\hat\sigma)\in K_\epsilon$, we have

$$\inf_{K_\epsilon}\frac{1}{n}\sum_{i=1}^n\Big[\rho\Big(\frac{Y_i - X_i^\tau\beta}{\sigma\sqrt{1+\|\beta\|^2}}\Big) + \log(\sigma)\Big] = \frac{1}{n}\sum_{i=1}^n\Big[\rho\Big(\frac{Y_i - X_i^\tau\hat\beta}{\hat\sigma\sqrt{1+\|\hat\beta\|^2}}\Big) + \log(\hat\sigma)\Big].$$

(1.2) yields

$$\inf_{K}\frac{1}{n}\sum_{i=1}^n\Big[\rho\Big(\frac{Y_i - X_i^\tau\hat\beta}{\hat\sigma\sqrt{1+\|\hat\beta\|^2}}\Big) + \log(\hat\sigma)\Big] \le \frac{1}{n}\sum_{i=1}^n\Big[\rho\Big(\frac{Y_i - X_i^\tau\beta_0}{\sigma_0\sqrt{1+\|\beta_0\|^2}}\Big) + \log(\sigma_0)\Big].$$

By Lemma 5, we can get that

$$\inf_{K_\epsilon} E\Big[\rho\Big(\frac{Y_i - X_i^\tau\beta}{\sigma\sqrt{1+\|\beta\|^2}}\Big) + \log(\sigma)\Big] - E\Big[\rho\Big(\frac{Y_i - X_i^\tau\beta_0}{\sigma_0\sqrt{1+\|\beta_0\|^2}}\Big) + \log(\sigma_0)\Big] := \delta_\epsilon > 0.$$

Hence

$$\begin{aligned}
\inf_{K_\epsilon}\frac{1}{n}\sum_{i=1}^n\Big[\rho\Big(\frac{Y_i - X_i^\tau\hat\beta}{\hat\sigma\sqrt{1+\|\hat\beta\|^2}}\Big) + \log(\hat\sigma)\Big] &\le \frac{1}{n}\sum_{i=1}^n\Big[\rho\Big(\frac{Y_i - X_i^\tau\beta_0}{\sigma_0\sqrt{1+\|\beta_0\|^2}}\Big) + \log(\sigma_0)\Big]\\
&= \zeta_{2n} + E\Big[\rho\Big(\frac{Y_i - X_i^\tau\beta_0}{\sigma_0\sqrt{1+\|\beta_0\|^2}}\Big) + \log(\sigma_0)\Big],
\end{aligned}$$

which leads to

$$\inf_{K_\epsilon} E\Big[\rho\Big(\frac{Y_i - X_i^\tau\beta}{\sigma\sqrt{1+\|\beta\|^2}}\Big) + \log\sigma\Big] \le \zeta_{1n} + \zeta_{2n} + E\Big[\rho\Big(\frac{Y_i - X_i^\tau\beta_0}{\sigma_0\sqrt{1+\|\beta_0\|^2}}\Big) + \log(\sigma_0)\Big]$$

$$:= \zeta_n + E\Big[\rho\Big(\frac{Y_i - X_i^\tau \beta_0}{\sigma_0\sqrt{1+\|\beta_0\|^2}}\Big) + \log(\sigma_0)\Big].$$

Hence, we can get that $\zeta_n \geq \delta_\epsilon$. Furthermore, we have $\{(\hat\beta,\hat\sigma)\in K_\epsilon\} \subseteq \{\zeta_n \geq \delta_\epsilon\}$. Then, $\cup_{k=1}^\infty \cap_{n=k}^\infty \{(\hat\beta,\hat\sigma)\in K_\epsilon\} \subseteq \cup_{k=1}^\infty \cap_{n=k}^\infty \{\zeta_n \geq \delta_\epsilon\}$. This completes the proof.

A.3. Proof of Theorem 2

Let

$$e_i(\beta) = \frac{\epsilon_i - u_i^\tau\beta - x_i^\tau(\beta-\beta_0)}{\sqrt{1+\|\beta\|^2}}, \quad e(\beta) = \frac{\epsilon - u^\tau\beta - x^\tau(\beta-\beta_0)}{\sqrt{1+\|\beta\|^2}},$$

$$\widetilde{x}_i(\beta) = x_i + u_i + \frac{\epsilon_i - u_i^\tau\beta - x_i^\tau(\beta-\beta_0)}{1+\|\beta\|^2}\beta.$$

Then we can write $g_n(\beta,\sigma) = n^{-1}\widetilde{x}_i(\beta)\psi(e_i(\beta)/\sigma) = P_n\widetilde{x}(\beta)\psi(e_1(\beta)/\sigma) := P_n g(\beta,\sigma)$ and $f_n(\beta,\sigma) = n^{-1}\sum_{i=1}^n \chi(e_i(\beta)/\sigma) = P_n\chi(e_1(\beta)/\sigma) := P_n f(\beta,\sigma)$.

Let $\Psi_n(\beta,\sigma) = (g_n(\beta,\sigma), f_n(\beta,\sigma))^\tau$. Then by (1.3) we get that $\Psi_n(\hat\beta,\hat\sigma) = 0$. Let $\Psi(\beta,\sigma) = E\Psi_n(\beta,\sigma)$. It easy to see that $\Psi(\beta_0,\sigma_0) = 0$.

By Theorem 1, we conclude that $\hat\beta \to \beta_0$, $\hat\sigma \to \sigma_0$ a.s.. Finally, our goal is to use Huber's Z-theorem, see Pollard (1985), and van der Vaart and Wellner (1996), to establish asymptotic normality. It suffices for this to verify the conditions A.1−A.4 of Huber's Z-theorem in van der Vaart and Wellner (1996).

For A.1, note that $\sqrt{n}(\Psi_n(\beta,\sigma) - \Psi(\beta,\sigma)) = \sqrt{n}(P_n - P)(g(\beta,\sigma), f(\beta,\sigma))^\tau$. It follows from Lemma 6 and Lemma 7 that $\sqrt{n}(P_n - P)g(\beta_0,\sigma_0) \xrightarrow{d} N(0,\Sigma)$ as $n\to\infty$, where Σ is defined in (A.4.2) below.

By the Central Limit Theorem, we have $\sqrt{n}(P_n - P)f(\beta_0,\sigma_0) \xrightarrow{d} N(0,A)$, where $A = E\chi^2(e(\beta_0)/\sigma_0)$, and $\sqrt{n}(\Psi_n(\beta_0,\sigma_0) - \Psi(\beta_0,\sigma_0)) \xrightarrow{d} (N(0,\Sigma), N(0, A))^\tau$.

It follows from **A**1 and **A**2 in Section 2 and the expression for $\Psi(\beta,\sigma)$, that A.1 in van der Vaart and Wellner (1996) holds.

For A.4, notice that

$$\frac{\partial g_n(\beta_0,\sigma_0)}{\partial\beta}$$
$$= \frac{1}{n}\sum_{i=1}^n \Bigg\{\Big(x_i + u_i + \frac{\epsilon_i - u_i^\tau\beta_0 - x_i^\tau(\beta_0-\beta_0)}{1+\|\beta\|^2}\beta_0\Big)\psi'\Big(\frac{\epsilon_i - u_i^\tau\beta_0 - x_i^\tau(\beta_0-\beta_0)}{\sigma_0\sqrt{1+\|\beta_0\|^2}}\Big)$$
$$\cdot\Big[\frac{x_i+u_i}{\sqrt{1+\|\beta_0\|^2}} + \frac{\epsilon_i - u_i^\tau\beta_0 - x_i^\tau(\beta_0-\beta_0)}{1+\|\beta_0\|^{3/2}}\beta_0\Big]^\tau + \psi\Big(\frac{\epsilon_i - u_i^\tau\beta_0 - x_i^\tau(\beta_0-\beta_0)}{\sigma_0\sqrt{1+\|\beta_0\|^2}}\Big)$$

$$\times\Big[\frac{-(x_i+u_i)\beta_0^\tau+(\epsilon_i-u_i^\tau\beta_0-x_i^\tau(\beta_0-\beta_0))I_p}{\sqrt{1+\|\beta_0\|^2}}-\frac{2\epsilon_i-u_i^\tau\beta_0-x_i^\tau(\beta_0-\beta_0)}{(1+\|\beta_0\|^2)^2}\beta_0\beta_0^\tau\Big]\Big\}.$$

It then follows from that

$$\begin{aligned}
E\Big(\frac{\partial g_n(\beta_0,\sigma_0)}{\partial\beta_0}\Big) &= E\Big\{\psi'\Big(\frac{\epsilon-u^\tau\beta_0-x^\tau(\beta_0-\beta_0)}{\sigma_0\sqrt{1+\|\beta_0\|^2}}\Big)\Big\} \\
&\quad\times\Big(x+u+\frac{\epsilon-u^\tau\beta_0-x^\tau(\beta_0-\beta_0)}{1+\|\beta_0\|^2}\beta_0\Big)\Big(\frac{\epsilon-u^\tau\beta_0-x^\tau(\beta_0-\beta_0)}{\sigma_0(1+\|\beta_0\|^{3/2})}\Big)^\tau \\
&\quad+\psi\Big(\frac{\epsilon-u^\tau\beta_0-x^\tau(\beta_0-\beta_0)}{\sigma_0\sqrt{1+\|\beta_0\|^2}}\Big)\Big[\frac{\epsilon-u^\tau\beta_0-x^\tau(\beta_0-\beta_0)}{\sqrt{1+\|\beta_0\|^2}}I_p \\
&\quad-\frac{(x+u)\beta_0^\tau}{\sqrt{1+\|\beta\|^2}}-\frac{2\epsilon-u^\tau\beta_0-x^\tau(\beta_0-\beta_0)}{(1+\|\beta_0\|^2)^2}\beta_0\beta_0^\tau\Big] \\
&= E\Big\{-\frac{1}{\sigma_0\sqrt{1+\|\beta\|^2}}\psi'\Big(\frac{\xi_1}{\sigma_0}\Big)\Big[x+\Gamma^*\xi+\frac{\xi_1}{\sqrt{1+\|\beta_0\|^2}}\beta_0\Big] \\
&\quad\times\Big[x+\Gamma^*\xi+\frac{\xi_1}{\sqrt{1+\|\beta_0\|^2}}\beta_0\Big]^\tau \\
&\quad+\frac{1}{\sigma_0\sqrt{1+\|\beta_0\|^2}}\psi'\Big(\frac{\xi_1}{\sigma_0}\Big)\xi_1 I_p-\frac{(x+\Gamma^*\xi)\beta_0^\tau}{\sqrt{1+\|\beta_0\|^2}}-\frac{2\xi_1\beta_0\beta_0^{\tau\ 2}}{1+\|\beta_0\|^2}\Big]\Big\} \\
&= -\frac{E[\psi'(\xi_1/\sigma_0)]\Sigma_x}{\sigma_0\sqrt{1+\|\beta_0\|^2}}-\frac{1}{\sigma_0\sqrt{1+\|\beta_0\|^2}}\ E\Big[\psi'\Big(\frac{\xi_1}{\sigma_0}\Big)\Big(\Gamma^*\xi\xi^\tau\Gamma^{*\tau} \\
&\quad+\Gamma^*\xi\frac{\xi_1\beta_0^\tau}{\sqrt{1+\|\beta_0\|^2}}+\frac{\beta_0\xi_1}{\sqrt{1+\|\beta_0\|^2}}\xi^\tau\Gamma^*+\frac{\xi_1^2\beta_0\beta_0^\tau}{1+\|\beta_0\|^2}\Big)\Big] \\
&\quad+E\frac{[\psi'(\xi_1/\sigma_0)\xi_1]}{\sqrt{1+\|\beta_0\|^2}}I_p-\frac{1}{\sqrt{1+\|\beta_0\|^2}}\Big(\Gamma^*E\Big[\psi\Big(\frac{\xi_1}{\sigma_0}\Big)\Big]\beta_0^\tau \\
&\quad+\frac{E[\psi(\xi_1/\sigma_0)\xi_1]\beta_0\beta_0^\tau}{\sqrt{1+\|\beta_0\|^2}}\Big) \\
&= -\frac{E[\psi'(\xi_1/\sigma_0)]\Sigma_x}{\sigma_0\sqrt{1+\|\beta_0\|^2}}-\frac{1}{\sigma_0\sqrt{1+\|\beta_0\|^2}}\Big(E\Big[\psi'\Big(\frac{\xi_1}{\sigma_0}\Big)\xi_2^2 \\
&\quad-\sigma_0E[\psi(\xi_1)\xi_1]\Big]\Big)\Big(I_p-\frac{\beta_0\beta_0^\tau}{1+\|\beta_0\|^2}\Big) \\
&= -\frac{E[\psi'(\xi_1/\sigma_0)]\Sigma_x}{\sigma_0\sqrt{1+\|\beta_0\|^2}}. \qquad (A.3.1)
\end{aligned}$$

Moreover, we have $\partial f(\beta,\sigma)/\partial\sigma=-e_1(\beta_0)\chi'(e_1(\beta_0)/\sigma_0)/\sigma_0^2$. Now Ψ is differentiable at β_0, σ_0 with derivative (matrix)

$$\dot{\Psi}(\beta_0,\sigma_0) = \begin{pmatrix} -\dfrac{E[\psi'(\xi_1/\sigma_0)]\Sigma_x}{\sigma_0\sqrt{1+\|\beta_0\|^2}} & 0 \\ 0 & -E\Big[\dfrac{e_1(\beta_0)}{\sigma_0^2}\chi'\Big(\dfrac{e_1(\beta_0)}{\sigma_0}\Big)\Big] \end{pmatrix}.$$

So $\dot{\Psi}(\beta_0,\sigma_0)$ is negative definite by the conditions B1$-$B2 in Section 2. Thus it remains only to verify the asymptotic equicontinuity condition of Huber's Z-theorem. For this A.2, note that

$$\begin{aligned}&\sqrt{n}(\Psi_n(\beta,\sigma)-\Psi(\beta,\sigma))-\sqrt{n}(\Psi_n(\beta_0,\sigma_0)-\Psi(\beta_0,\sigma_0))\\ &\quad=\sqrt{n}(P_n-P)(g(\beta,\sigma)-g(\beta_0,\sigma_0),f(\beta,\sigma)-f(\beta_0,\sigma_0))^\tau,\end{aligned}$$

so, according to Theorem 2.1 in van de Geer (2000), we need only show that the two function classes $\mathcal{F}_j(\beta,\sigma)=\{f_j(\beta,\sigma):\|\beta-\beta_0\|\le\delta,\|\sigma-\sigma_0\|\le\delta\}$, $j=1,2$, where $f_1(\beta,\sigma)=g(\beta,\sigma)=\widetilde{x}(\beta)\psi(e_1(\beta)/\sigma)$, $f_2(\beta,\sigma)=f(\beta,\sigma)=-\chi(e_1(\beta)/\sigma)/\sigma^2$ are both VC subgraph classes.

With A3 in Section 2, using the same argument in Lemma 9, we can get that the class $\mathcal{F}_2(\beta,\sigma)$ is VC subgraph class. So $\mathcal{F}_1(\beta,\sigma)$ is also a VC subgraph classes. Then the asymptotic equicontinuity condition of Huber's Z theorem holds.

From the fact that $-E[\psi'(\xi_1/\sigma_0)]\Sigma_x\sqrt{n}(\hat{\beta}-\beta_0)/(\sigma_0\sqrt{1+\|\beta_0\|^2})$ and $\sqrt{n}g_n(\beta_0,\sigma_0)=n^{-1/2}\sum_{i=1}^n\widetilde{x}_i(\beta_0)\psi(e_i(\beta_0)/\sigma_0)$ have the same asymptotic distribution, by Lemma 6 and Lemma 7 we can conclude from Huber's Z-theorem that

$$\frac{E\psi'(\epsilon/\sigma_0)}{\sigma_0\sqrt{1+\|\beta_0\|^2}}\sqrt{n}\Sigma^{-1/2}\Sigma_x(\hat{\beta}-\beta_0)\stackrel{d}{\longrightarrow}N(0,I_p),$$

$$\sigma_0^{-2}A^{-1/2}E(e(\beta_0)\chi'\Big(\frac{e(\beta_0)}{\sigma_0}\Big)n^{-1/2}(\hat{\sigma}-\sigma_0)\stackrel{d}{\longrightarrow}N(0,1).$$

This completes the proof.

A.4. Technical Lemmas

Lemma 1. *Under* A1$-$A3, *there exists constants* $\sigma_1,\sigma_2>0$ *such that* $\sigma_1\le\sigma_0\le\sigma_2$ *for sufficient large* n.

Proof. If there exists a subsequence $0<\sigma_{0n'}\to 0$ as $n'\to\infty$, then by the continuity and boundedness of $\chi(\cdot)$ we have $\lim_{n'\to\infty}E[\chi\{\epsilon-u^\tau\beta_0/\sigma_{0n'}(\sqrt{1+\|\beta_0\|^2})\}]=\kappa$ uniformly for $1\le i\le n$. Therefore, $\lim_{\sigma_{0n'}\to 0}E\{\chi[(\epsilon_i-u^\tau\beta_0)/(\sigma_{0n'}\sqrt{1+\|\beta_0\|^2})]\}=\kappa>0$, which contracts the definition of σ_0. If for some subsequence $\sigma_{0n''}\to\infty$ as $n''\to\infty$, then $\lim_{\sigma_{0n''}\to\infty}E\{\chi[(\epsilon-u^\tau\beta_0)/(\sigma_{0n''}\sqrt{1+\|\beta_0\|^2})]\}=-1$, which contracts the definition of σ_0. The proof is complete.

Lemma 2. *Under* A1−A3 *and* B1, *we have* $\underline{\lim}_{\sigma\to 0}\inf_{s\geq 0} E\chi[(\epsilon_i+s)/\sigma] > 0$.

Proof. For all $\delta_1 > 0$,

$$\begin{aligned} E\chi\Big[\frac{(\epsilon_i+s)}{\sigma}\Big] &= E\chi\Big[\frac{(\epsilon_i+s)}{\sigma}\Big] I(|(\epsilon_i+s)|\geq\delta_1) + E\chi\Big[\frac{(\epsilon_i+s)}{\sigma}\Big] I(|(\epsilon_i+s)|<\delta_1) \\ &\geq E\bigg[\chi\Big[\frac{\delta_1}{\sigma}\Big] P\{(|(\epsilon_i+s)|\geq\delta_1)\}\bigg] - P\{(|(\epsilon_i+s)|<\delta_1)\} \\ &\geq \chi\Big[\frac{\delta_1}{\sigma}\Big] - (1+\kappa)P\{(|(\epsilon_i+s)|<\delta_1)\} \geq \chi\Big[\frac{\delta_1}{\sigma}\Big] - C_1\delta_1, \end{aligned}$$

where $C_1 > 0$ is a constant depending on f_0 only. Thus $\underline{\lim}_{\sigma\to 0}\inf_{s\geq 0} E\chi[((\epsilon_i + s)/\sigma] > [\kappa - C_1\delta_1]$. The proof of Lemma 2 is completed by taking $\delta_1 = \kappa/2C_1$.

Lemma 3(Maximal inequality). (Pollard (1990)). *If independent random processes* $\{f_i(\omega,t)\}_{i=1}^{\infty}$ *are manageable with respect to the envelopes* $\{F_i(\omega)\}_{i=1}^{\infty}$, *then there exists a constant* $K_q \geq 0$ *such that* $E\sup_t|\sum_{i=1}^n(f_i(\omega,t)-Ef_i(\omega,t))|^q \leq K_q\|F_n(\omega)\|^q$, *for* $q>1$, *where* $F_n = (F_1,\ldots,F_n)^t$.

Lemma 4. *Under* A1−A3 *and* B1, *there exist two constants* $\sigma_1 > 1$, *and* $0 < \sigma_2 < 1$ *such that* $\overline{\lim}_{n\to\infty}P\{\hat\sigma \leq \sigma_2, \hat\sigma \geq \sigma_1\} = 0$.

Proof. Let $\xi_n(\beta,\sigma) = n^{-1}\sum_{i=1}^n \chi\{[\epsilon_i - x_i^\tau(\beta-\beta_0)]/(\sigma\sqrt{1+\|\beta_0\|^2}))\}$. Since $\chi(\cdot)$ is continuous and bounded, it follows from Lemma 3 that $Z_{1n} =: \sup_{\beta>0,\sigma>0} |\xi_n(\beta,\sigma) - E(\xi_n(\beta,\sigma))| = O_p(1/\sqrt{n})$. By Lemma 2, there exists a constant $0 < \sigma_1 < 1$, such that $\inf_{s\geq 0} E\chi[(\epsilon_i+s)/\sigma] \geq \delta_2/4$, as n is large sufficiently. Hence, on one hand, if $\hat\sigma \leq \sigma_2$ then $0 = \xi_n(\hat\beta,\hat\sigma)) \geq \delta_2/4 - Z_{1n}$. This implies that $P\{Z_{1n} \geq \delta_2/4\} \to 0$ as $n\to\infty$.

On the other hand, with $Z_{2n} =: |n^{-1}\sum_{i=1}^n[\rho(\epsilon_i) - E\rho(\epsilon_i)]| = O_p(1/\sqrt{n})$, if $\hat\sigma \geq \sigma_2$,

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^n[\rho(\epsilon_i) - E\rho(\epsilon_i)] &\geq \frac{1}{n}\sum_{i=1}^n\left\{\bigg[\rho\bigg[\frac{1}{\hat\sigma}\Big(\epsilon_i - \frac{x_i(\hat\beta-\beta_0)}{\sqrt{1+\|\hat\beta\|^2}}\Big)\bigg] + \log(\hat\sigma)\bigg] - E\rho(\epsilon_i)\right\} \\ &\geq \log(\sigma_2) - E\rho(\epsilon_i) \geq \log(\sigma_2) - C_3. \end{aligned}$$

Therefore, $Z_{2n} \geq \log(\sigma_1) - C_3$, where $C_3 > 0$ is a constant, and $P\{\hat\sigma \geq \sigma_2\} \leq P\{Z_{2n} \geq \log(\sigma_1) - C_3\} \to 0$ as $n\to\infty$ by taking σ_2 large enough. This competes the proof.

Lemma 5. *Under the conditions of Theorem* 1,

$$(\beta_0,\sigma_0) = \arg\min_{\beta\in R^p,\sigma>0} E\rho\bigg[\frac{1}{\sigma}\Big(\epsilon - \frac{x^\tau(\beta-\beta_0)}{\sqrt{1+\|\beta\|^2}}\Big) + \log(\sigma)\bigg]$$

and (β_0,σ_0) *is unique.*

Proof. Let

$$\begin{aligned} h(\beta,\sigma) \,=\, & E\Big\{\rho\Big[\frac{1}{\sigma}\Big(\epsilon-\frac{x^\tau(\beta-\beta_0)}{\sqrt{1+\|\beta\|^2}}\Big)\Big]-\rho\Big(\frac{\epsilon}{\sigma}\Big)\Big\}+\Big[E\rho\Big(\frac{\epsilon}{\sigma}\Big)+\log(\sigma)\Big] \\ :=\,& h_1(\beta,\sigma)+h_2(\beta,\sigma). \end{aligned}$$

First, we have

$$\begin{aligned} h_1(\beta,\sigma) &= E\int_{\frac{\epsilon}{\sigma}}^{\frac{\epsilon}{\sigma}-\frac{x^\tau(\beta-\beta_0)}{\sigma\sqrt{1+\|\beta\|^2}}}\psi(t)dt = E\int_0^{-\frac{x^\tau(\beta-\beta_0)}{\sigma\sqrt{1+\|\beta\|^2}}}\psi\Big(\frac{\epsilon}{\sigma}+t\Big)dt \\ &= E_x\int_0^{-\frac{x^\tau(\beta-\beta_0)}{\sigma\sqrt{1+\|\beta\|^2}}}E\psi\Big(\frac{\epsilon}{\sigma}+t\Big)dt \ge 0 \end{aligned}$$

and $h_1(\beta_0,\sigma)=0$. Next, if there exists $\widetilde{\beta}$ such that $h_1(\widetilde{\beta},\sigma)=0$, then $\widetilde{\beta}=\beta_0$. In fact, if $h_1(\widetilde{\beta},\sigma)=0$ and $|E\psi(\epsilon/\sigma+t)dt|>0$, then

$$E_x\int_0^{-\frac{x^\tau(\widetilde{\beta}-\beta_0)}{\sigma\sqrt{1+\|\beta\|^2}}}\min\Big\{E\psi(\frac{\epsilon}{\sigma}+t),E\psi(\frac{\epsilon}{\sigma}-t)\Big\}dt=0.$$

Thus, $x^\tau(\widetilde{\beta}-\beta_0)/\sigma\sqrt{1+\|\widetilde{\beta}\|^2}=0$, a.s. P_x. It follows from x being nondegenerate that $\widetilde{\beta}=\beta_0$.

We need only prove σ_0 is the unique maximum of $h_2(\beta,\sigma)$ and since σ_0 satisfies $E[\chi\{c\epsilon/\sigma_0\}]=0$, we need only prove σ_0 is unique solution of $E[\chi(\epsilon/\sigma_0)]=0$.

Assume σ_1 is another solution, and without loss of generality that $\sigma_0<\sigma_1$, hence $E[\chi(\epsilon/\sigma_0)-\chi(\epsilon/\sigma_1)]=0$. By an elementary calculation, we have

$$\int_0^\infty\Big[\chi\Big(\frac{y}{\sigma_0}\Big)-\chi\Big(\frac{y}{\sigma_1}\Big)\Big]f(y)dy+\int_0^\infty\Big[\chi\Big(\frac{-y}{\sigma_0}\Big)-\chi\Big(\frac{-y}{\sigma_1}\Big)\Big]f(y)dy=J_1+J_2=0.$$

Notice that both J_1 and J_2 are nonnegative, so we get that $J_1=0$ and $J_2=0$.

From $J_1=0$ by A3 and C1, we have $\sigma_0=\sigma_1$. From $J_2=0$ by A3 and C1, we have $\sigma_0=\sigma_1$.

For the t distribution, without loss of generality, let $\nu=1$, $\rho(x)=2^{-1}\log(1+x^2)$. Then $\psi(x)=x/(1+x^2)$, $\chi(x)=x\psi(x)-1=-1/(1+x^2)$, $\chi'(x)=2x/(1+x^2)^2$. Then, we have $E\chi'(\epsilon/\sigma)\epsilon/\sigma=\int 2y^2f(y)/(1+y^2)^2dy>0$. And the solution is unique.

Lemma 6. *If*

$$g(\beta,\sigma)=E\Big\{\Big[\Big(x+u+\frac{\epsilon-u^\tau\beta-x^\tau(\beta-\beta_0)}{1+\|\beta\|^2}\beta\Big)\Big]\psi\Big(\frac{\epsilon-u^\tau\beta-x^\tau(\beta-\beta_0)}{\sigma\sqrt{1+\|\beta\|^2}}\Big)\Big\},$$

then $g(\beta_0,\sigma_0)=0$.

Proof. From the expression for $g(\beta,\sigma)$ we have

$$g(\beta_0,\sigma_0)=E\Big\{\Big(x+u+\frac{\epsilon-u^\tau\beta_0}{1+\|\beta_0\|^2}\beta_0\Big)\psi\Big(\frac{\epsilon-u^\tau\beta_0}{\sigma_0\sqrt{1+\|\beta_0\|^2}}\Big)\Big\}.$$

Consider the $(p+1)\times(p+1)$ orthogonal matrix

$$\Gamma=\begin{pmatrix}\frac{1}{\sqrt{1+\|\beta_0\|^2}} & \frac{-\beta_0}{\sqrt{1+\|\beta_0\|^2}}\\ \star & \Gamma_1\end{pmatrix},$$

where $\star$ denotes a submatrix that not need be specified. Let

$$\xi=\Gamma(\epsilon,u^\tau)^\tau\sim EC_{p+1}(0,I_{p+1}), \tag{A.4.1}$$

where $EC_{p+1}(0,I_{p+1})$ represents the spherical distribution.

Then we have $u=-(\beta_0/\sqrt{1+\|\beta_0\|^2},\Gamma_1^\tau)\xi=\Gamma^*\xi$. From the above facts, we have

$$\begin{aligned}g(\beta_0,\sigma_0)&=E\Big\{\Big(x+u+\frac{\epsilon-u^\tau\beta_0}{1+\|\beta_0\|^2}\beta_0\Big)\psi\Big(\frac{\epsilon-u^\tau\beta_0}{\sigma_0\sqrt{1+\|\beta_0\|^2}}\Big)\Big\}+E\Big[x\psi\Big(\frac{\epsilon}{\sigma_0}\Big)\Big]\\&=E\Big\{\Big[\Big(\frac{-\beta_0}{\sqrt{1+\|\beta_0\|^2}},\Gamma_1^\tau\Big)\xi+\frac{\beta_0\xi_1}{\sqrt{1+\|\beta_0\|^2}}\Big]\psi\Big(\frac{\xi_1}{\sigma_0}\Big)\Big\}\\&=\frac{-\beta_0}{\sqrt{1+\|\beta_0\|^2}}E\Big(\xi_1\psi\Big(\frac{\xi_1}{\sigma_0}\Big)\Big)+\frac{E(\xi_1\psi(\xi_1/\sigma_0))}{\sqrt{1+\|\beta_0\|^2}}\cdot\beta_0=0.\end{aligned}$$

Lemma 7. *Under the conditions of Theorem* 1*, we have* $E[\psi'(\xi_1/\sigma)\xi_2^2]=\sigma E[\psi(\xi_1/\sigma)\xi_1]$*, and*

$$\begin{aligned}\Sigma&=\mathrm{Cov}\Big[\Big(x+u+\frac{\epsilon-u^\tau\beta_0}{1+\|\beta_0\|^2}\cdot\beta_0\Big)\psi\Big(\frac{\epsilon-u^\tau\beta_0}{\sigma_0\sqrt{1+\|\beta_0\|^2}}\Big)\Big]\\&=E\psi^2\Big(\frac{\epsilon}{\sigma_0}\Big)\Sigma_x+E\Big[\psi^2\Big(\frac{\epsilon}{\sigma_0}\Big)u_{11}^2\Big]\Big(I_p-\frac{\beta_0\beta_0^\tau}{1+\|\beta_0\|^2}\Big),\end{aligned}\tag{A.4.2}$$

where $\xi=(\xi_1,\ldots,\xi_{p+1})^\tau$ *is the same as in* (A.4.1).

Proof. Since $\xi\overset{d}{\sim}EC_{p+1}(0,I_{p+1})$, $\xi\overset{d}{=}R\cdot Z/\|Z\|$, where $Z=Z_1,\cdots,Z_{p+1}^\tau\overset{d}{\sim}N(0,I_{p+1})$, $R\ge 0$ and Z are independent. Denote by $G(r)$ the distribution function of random variable R and, invoking the independence of $\|Z\|$ and $Z/\|Z\|$, we have

$$\begin{aligned}&E[\psi'(\frac{\xi_1}{\sigma})\xi_2^2]\\&=(2\pi)^{-(p+1)/2}\int_0^\infty\int_{R^{p+1}}\psi'\Big(\frac{rz_1}{\sigma\|z\|}\Big)\frac{r^2z_2^2}{\|z\|^2}\exp\Big\{-\frac{\|z\|^2}{2}\Big\}dz_1\cdots dz_{p+1}dG(r)\end{aligned}$$

$$= \frac{(2\pi)^{-(p+1)/2}}{p}\int_0^\infty\int_{R^{p+1}}\psi'\Big(\frac{rz_1}{\sigma\|z\|}\Big)r^2\|z\|^{-2}\sum_{i=2}^{p+1}z_i^2\exp\Big\{-\frac{\|z\|^2}{2}\Big\}dz_1\cdots dz_{p+1}dG(r)$$

$$= \sigma\frac{(2\pi)^{-(p+1)/2}}{p}\int_0^\infty\int_{R^p}\exp\Big\{-\frac{1}{2}\sum_{i=2}^{p+1}z_i^2\Big\}\Big(\int_{R^1}r\|z\|\exp\Big\{-\frac{1}{2}z_1^2\Big\}d\psi\Big(\frac{rz_1}{\sigma\|z\|}\Big)\Big)$$

$$\cdot dz_2\cdots dz_{p+1}dG(r)$$

$$= -\sigma\frac{(2\pi)^{-(p+1)/2}}{p}\int_0^\infty\int_{R^p}\exp\Big\{-\frac{1}{2}\sum_{i=2}^{p+1}z_i^2\Big\}\Big(\int_{R^1}r\psi\Big(\frac{rz_1}{\sigma\|z\|}\Big)\Big[\frac{z_1}{\|z\|}-z_1\|z\|\Big]$$

$$\cdot\exp\Big\{-\frac{1}{2}z_1^2\Big\}dz_1\Big)dz_2\cdots dz_{p+1}dG(r)$$

$$= \sigma\frac{(2\pi)^{-(p+1)/2}}{p}\int_0^\infty\int_{R^{p+1}}(\|z\|^2-1)\frac{rz_1}{\sigma\|z\|}\psi\Big(\frac{rz_1}{\sigma\|z\|}\Big)\exp\Big\{-\frac{\|z\|^2}{2}\Big\}dz_1\cdots dz_{p+1}dG(r)$$

$$= \sigma\frac{1}{p}E\Big\{[E(\chi_p^2-1)]R\psi\Big(R\frac{Z_1}{\sigma\|Z_1\|}\Big)\Big\} = \sigma E\Big[\psi\Big(\frac{\xi_1}{\sigma}\Big)\xi_1\Big].$$

Acknowledgement

The authors are very grateful to the editor, an associate editor, an consulting editor and one anonymous referee for helpful comments and constructive advice, which lead to a substantially improved manuscript. The authors are grateful to Professor Xuming He for very valuable comments and suggestions. The project was supported by the Natural Science Foundation of China(No: 10771017,10231030) and Key Project of Ministry of Education of the People's Republic of China(No: 309007).

References

Brown, M. (1982). Robust line estimation with errors in both variables, *J. Amer. Statist. Assoc.* **77**, 71- 79.

Cheng, C. L. and Van Ness, J. W. (1992). Generalized M-estimators for errors-in-variables regression. *Ann. Statist.* **20**, 385-397.

Carroll, R. J., Ruppert, D. and Stefanski, L. A. (1995). *Measurement Error in Nonlinear Models.* Chapman and Hall, New York.

Cui, H. J. (1997). Asymptotic normality of M-estimates in the EV model. *J. Sys. Sci. and Math. Scis.* **10**, 225-236.

Cui, H. J. (2004). On asymptotic of t-type regression estimation in multiple linear model. *Sci. China Ser. A* **47**, 628-639.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with Discussion). *J. Roy. Statist. Soc. Ser. B* **39**, 1-38.

Fekria, M. and Ruiz-Gazen, A. (2004). Robust weighted orthogonal regression in the errors-in-variables model. *J. Multivariate Anal.* **88**, 89-108.

Fuller, W. A. (1987). *Measurement Error Model*. John Wiley, New York.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. and Stahel, W. A. (1986). *Robust Statistics*. Wiley, New York.

He, X. M., Simpson, D. and Wang, G. Y. (2000). Breakdown points of t-type regression estimators. *Biometrica* **87**, 675-687.

He, X. M., Cui, H. J. and Simpson, D. (2004). Longitudinal data analysis using t-type repression. *J. Statist. Plann. Inference* **122**, 253-269.

He, X. and Liang, H. (2000). Quantile regression estimates for a class of linear and partially linear errors-in-variables models. *Statist. Sinica* **1**, 129-140.

He, X. and Shao, Q. M. (1996). A general Bahadur representation of M-estimators and its application to linear regression with nonstochatic designs. *Ann. Statist.* **24**, 2608-2630.

Ketellapper, R. H. and Ronner, A. E. (1984). Are robust estimation methods useful in the structural errors-in variables model. *Metrika* **31**, 33-41.

Lange, K. L., Little, R, J. A. and Taylor, J. M. G. (1989). Robust statistical modeling using the t distribution. *J. Amer. Statist. Assoc.* **84**, 881-896.

Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *J. Amer. Statist. Assoc.* **83**, 1198-1202.

Pollard, D. (1984). *Convergence of Stochastic Processes.* Springer, New York.

Pollard, D. (1985). New ways to prove central limit theorems. *Econometric Theory* **1**, 295-314.

Pollard, D. (1990). *Empirical Processes: Theory and Applications.* NSF-CBMS Regional Conference Series in Probability and Statistics. Vol 2. Institute of Mathematical Statistics,Hayward, California.

van de Geer, S. A. (2000). *Empirical Processes in M-Estimation*. Cambridge University Press, Cambridge.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes, With Applications to Statistics*. Springer-Verlag, New York.

Wu, C. F. J. (1983). On the convergence properties of the EM algorithm. *Ann. Statist.* **7**, 95-103.

Zamar, R. H. (1989). Robust estimation in the errors-in-variables model. *Biometrika* **76**, 149-160.

School of Mathematical Sciences, Beijing Normal University, Beijing, P. R. China, 100875.

E-mail: hutao@mail.bnu.edu.cn

School of Mathematical Sciences, Beijing Normal University, Beijing, P. R. China, 100875.

E-mail: hjcui@bnu.edu.cn

(Received January 2008; accepted September 2008)