10-2015

Measure of location-based estimators in simple linear regression

Xijia LIU

Daniel P. A. PREVE

Singapore Management University, dpreve@smu.edu.sg

Citation

LIU, Xijia and PREVE, Daniel P. A.. Measure of location-based estimators in simple linear regression. (2015). *Journal of Statistical Computation and Simulation*. 86, (9), 1-14. Research Collection School Of Economics.

Available at: https://ink.library.smu.edu.sg/soe_research/2329

MEASURE OF LOCATION-BASED ESTIMATORS IN SIMPLE LINEAR REGRESSION

XIJIA LIU[†] AND DANIEL PREVE[‡]

Abstract. In this paper we consider certain measure of location-based estimators (MLBEs) for the slope parameter in a linear regression model with a single stochastic regressor. The median-unbiased MLBEs are interesting as they can be robust to heavy-tailed samples and, hence, preferable to the ordinary least squares estimator (LSE). Two different cases are considered as we investigate the statistical properties of the MLBEs. In the first case, the regressor and error are assumed to follow a symmetric stable distribution. In the second, other types of regressions, with potentially contaminated errors, are considered. For both cases the consistency and exact finite-sample distributions of the MLBEs are established. Some results for the corresponding limiting distributions are also provided. In addition, we illustrate how our results can be extended to include certain heteroscedastic regressions. Finite-sample properties of the MLBEs in comparison to the LSE are investigated in a simulation study.

1. Introduction

In regression analysis, an important question is how to obtain suitable estimators for the slope parameter β in the simple linear regression

$$y_i = \alpha + \beta x_i + u_i. \tag{1}$$

An example of such an estimator is the LSE for β, given by

$$\hat{\beta}_{LS} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \beta + \frac{\sum_{i=1}^n (x_i - \bar{x})(u_i - \bar{u})}{\sum_{i=1}^n (x_i - \bar{x})^2}, \tag{2}$$

which is consistent under quite general assumptions. A justification for the LSE is provided by the Gauss-Markow theorem which states that if the explanatory variable is non-stochastic and the regression errors are uncorrelated random variables with zero mean and common finite variance, then $\hat{\beta}_{LS}$ has the minimum variance of all linear unbiased estimators for β. However, the method of ordinary least squares is sensitive to large values of the error term. Estimators that are robust to heavy-tailed error distributions can be obtained using nonparametric (distribution free) techniques, an example being the Theil-Sen estimator (Sen, 1968b).

In this paper we consider robust MLBEs for the slope parameter in Equation (1) and investigate their finite-sample and asymptotic properties in a *parametric* setting. These estimators are based on measures of location, such as the sample median and trimmed mean. Although our results are more general, we focus on the case where the explanatory variable, which is assumed to be stochastic, follows a symmetric stable distribution and the error is either symmetric stable, with the same index of stability as the explanatory variable, or a normal mixture. We also consider a conditionally heteroscedastic specification. The MLBEs are similar to the estimator of

Key words and phrases. simple linear regression, robust estimators, measure of location, stable distribution, contaminated error, finite-sample, exact distribution.

[†] Uppsala University. [‡] City University of Hong Kong. Address correspondence to Xijia Liu, Department of Statistics, Uppsala University, Box 513, 751 20 Uppsala, Sweden; e-mail: xijia.liu@statistics.uu.se. We are grateful to seminar participants at Uppsala University for their comments and suggestions. All remaining errors are our own.

Preve and Medeiros (2011) in the sense that they are order statistics of successive ratios between the response and explanatory variable in a simple linear regression.

Regressions with symmetric stable errors have been considered by, for instance, Blattberg and Sargent (1971) and McCulloch (1998). A stable, potentially non-normal, error distribution can be motivated in a number of ways. For example, in economics, the error u_i may be thought of as the sum of a large number of independent, identically distributed (i.i.d.) stable random variables (say, decisions of investors). If the stability assumption is relaxed, in view of the generalized central limit theorem, then the distribution of u_i will be approximately stable. There is a large amount of evidence suggesting that many economic variables are best described by stable distributions with infinite variances. Classical examples include common stock price changes and changes in other speculative prices, cf. Mandelbrot (1963) and Fama (1965).[1]

For an example of an MLBE, consider the incomplete pairwise-slope estimator for β based on a sample of size n

$$\begin{aligned}\hat{\beta}_{PS} &= \text{med}\Big\{\frac{y_2-y_1}{x_2-x_1}, \frac{y_4-y_3}{x_4-x_3}, \dots, \frac{y_{2k}-y_{2k-1}}{x_{2k}-x_{2k-1}}\Big\} \\ &= \beta + \text{med}\{z_1, z_2, \dots, z_k\},\end{aligned} \tag{3}$$

where

$$z_i = \frac{u_{2i}-u_{2i-1}}{x_{2i}-x_{2i-1}},$$

and $\text{med}\{z_1, z_2, \dots, z_k\}$ is the sample median of $z_1, z_2, \dots, z_k$.[2] If the z_i are i.i.d. continuous random variables, standard results for order statistics show that the *exact* distribution of $\hat{\beta}_{PS}-\beta$ when k is odd can be expressed in terms of the incomplete beta function

$$\begin{aligned}G(z;k) &= \big[F_z(z)\big]^{r+1}\sum_{s=0}^{r}\binom{r+s}{r}\big[1-F_z(z)\big]^s \\ &= \frac{\Gamma(k+1)}{\Gamma(r+1)\Gamma(r+1)}\int_0^{F_z(z)} t^r(1-t)^r dt,\end{aligned} \tag{4}$$

where $\Gamma(\cdot)$ is the gamma function, $F_z(\cdot)$ is the cdf of the z_i and $k = 2r+1$.[3] See, for example, David and Nagaraja (2003, p. 10). The incomplete beta function has been tabled extensively and can easily be evaluated using standard mathematical software packages such as Mathematica and Matlab. Another example of an MLBE that we will consider is

$$\hat{\beta}_{UF} = \text{med}\Big\{\frac{y_1-\mu_y}{x_1-\mu_x}, \frac{y_2-\mu_y}{x_2-\mu_x}, \dots, \frac{y_n-\mu_y}{x_n-\mu_x}\Big\}, \tag{5}$$

where μ_y and μ_x are location parameters of the y_i and x_i, respectively. We will sometimes refer to this estimator as unfeasible as it requires both μ_y and μ_x to be known, which for most cases will not be realistic (cf. the $b(\alpha)$ estimators of Blattberg and Sargent, 1971).

Now consider *any* estimator $\hat{\beta}$ for β that can be decomposed into $\hat{\beta} = \beta + \text{med}\{z_1, z_2, \dots, z_k\}$, where the z_i are i.i.d. continuous random variables with zero median and k is odd. Then, in view of Equation (4), it is readily shown that the median of $\hat{\beta}-\beta$ is zero also. Hence, $\hat{\beta}$ is a median-unbiased estimator for β. If, in addition, the density of the z_i is symmetric about zero, then so is that of $\hat{\beta}-\beta$. This tells us that the distribution of $\hat{\beta}$ is centered about the

[1] See also the extensive bibliography on stable distributions compiled by J. P. Nolan, downloadable at `http://academic2.american.edu/~jpnolan`.

[2] The estimator $\hat{\beta}_{PS}$ is incomplete in the sense that it uses $k = \lfloor n/2 \rfloor$ differences, where $\lfloor n/2 \rfloor$ represents the integer part of $n/2$, instead of $n(n-1)/2$ (cf. Sen, 1968b).

[3] The corresponding expression when k is even is given in Desu and Rodine (1969).

unknown parameter β. Furthermore, if the median of z_i is unique (in general, the median may be an interval instead of a single number), then the sample median is a consistent estimator for the population median (e.g. Jiang, 2010, p. 5) and $\hat{\beta}$ converges in probability to β as k tends to infinity. Of course, as an alternative to the sample median, one could instead use a symmetrically trimmed mean in equations (3) and (5), cf. Section 5. Such an estimator could potentially have a higher asymptotic relative efficiency (ARE), see Oosterhoff (1994). The main focus of this note is to establish different conditions under which equations (3) and (5) are consistent, median-unbiased estimators with exact distributions that can be expressed in terms of Equation (4), and exact densities that are symmetric about β.

The remainder of this paper is organized as follows. In Section 2 we establish the consistency and exact finite-sample distribution of the MLBEs given by equations (3) and (5) in a symmetric stable regression. In doing so, we give conditions under which the median of the ratio of two independent symmetric stable random variables is unique and zero. In Section 3 we discuss how our results can be extended to include other types of regressions, with potentially contaminated errors. In Section 4 we illustrate how these results can be further extended to include certain types of conditionally heteroscedastic regressions. Section 5 reports the simulation results of a Monte Carlo study comparing the finite-sample performance of the MLBEs to each other, and to the LSE. In this study, we also consider feasible versions of (5). Section 6 concludes the paper. Mathematical proofs are collected in the Appendix. An extended Appendix available on request from the authors contains some results mentioned in the text but omitted from the note to save space.

2. A Symmetric Stable Regression

We will initially assume that both the explanatory variable and error in (1) are symmetric stable random variables. This also ensures that the response variable is symmetric stable. More specifically, for this specification both the unconditional and the (on x_i) conditional distribution of y_i follow a symmetric stable distribution, such as the normal or Cauchy distributions. As we shall see, although the conditional mean of y_i may not exist, the conditional median of y_i always exists for this specification.

The distribution of a stable random variable is described by four parameters, here denoted by a, b, c and d. The parameter a, the *index of stability*, is confined to the interval $(0, 2]$. The *skewness parameter* b is confined to $[-1, 1]$. The *scale parameter* $c > 0$, and the *location parameter* d can take on any real value. There exists a number of different parametrizations for symmetric stable distributions. Here we will use the $\mathcal{S}(a, b, c, d)$ parametrization in Definition 1.7 of Nolan (2012).

For the remainder of this section, we will focus our attention on the class of symmetric stable random variables. This class may be defined by the characteristic function,

$$\varphi(t) = E\left(e^{itv}\right) = e^{-c^a|t|^a + idt}, \tag{6}$$

where t is a real number. A random variable v is $\mathcal{S}(a, 0, c, d)$ distributed if its characteristic function is given by (6). While there is no general closed form expression for the density of a symmetric stable random variable, a great deal is known about their theoretical properties. Lemma 1 in the Appendix (given here without a proof) lists a selected few of these. The reader is referred to Nolan (2012), Nolan (2003) and Zolotarev (1986) for details.

There are only two known cases for which closed form expressions for the density of an $\mathcal{S}(a, 0, c, d)$ distributed random variable exists. These are the Gaussian ($a = 2$) and Cauchy

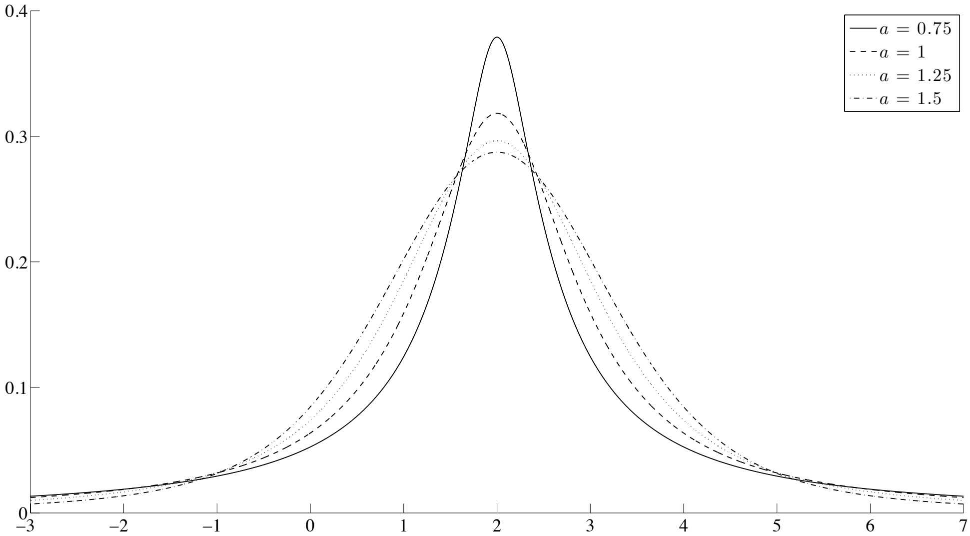

FIGURE 1. Densities of four $\mathcal{S}(a,0,1,d)$ distributed random variables, with common location parameter $d = 2$. Tails become progressively heavier as a decreases.

$(a = 1)$ densities, where the latter is given by

$$\frac{1}{\pi}\frac{c}{c^2+(v-d)^2}, \quad -\infty < v < \infty.$$

In general, all we have is integral representations of the density. Figure 1 shows the densities of four symmetric stable random variables with different indexes of stability.[4]

Assumption 1 holds throughout the section.

Assumption 1. *Let y_i $(i = 1, 2, \dots, n)$ be given by (1), with median μ_y. Suppose that*

- (i) *the x_i are independent $\mathcal{S}(a, 0, c_x, \mu_x)$,*
- (ii) *the u_i are independent $\mathcal{S}(a, 0, c_u, 0)$,*
- (iii) *for each i, x_i and u_i are mutually independent,*
- (iv) *the sample size is odd, $n = 2k + 1$.*

Here $a < 2$ is a typical setting in which the, on the explanatory variable, conditional mean and variance of the LSE for β may not exist. For example, if $a = 1$ the conditional distribution of (2) is *Cauchy.*

Proposition 1. *Let $G(\cdot)$ be given by (4). Under Assumption 1,*

- (i) *if μ_y and μ_x are known, $\hat{\beta}_{UF} \xrightarrow{p} \beta$ as $n \to \infty$ and the exact distribution of (5) is given by $P(\hat{\beta}_{UF} - \beta \le z) = G(z; n)$, with*

$$F_z(z) = \int_{-\infty}^{(c_x/c_u)z} \int_{-\infty}^{\infty} |t| f(st) f(t) dt ds,$$

where $f(\cdot)$ is the density of an $\mathcal{S}(a, 0, 1, 0)$ distributed random variable. For each k, the density of $\hat{\beta}_{UF} - \beta$ is symmetric about zero.

[4]Figure 1 was generated using the MATLAB function `stblpdf` of M. Veillette, downloadable at `http://math.bu.edu/people/mveillet/research.html`.

(ii) *if* $k = 2r + 1$ *is odd,* $\hat{\beta}_{PS} \xrightarrow{p} \beta$ *as* $n \to \infty$ *and the exact distribution of (3) is given by* $P(\hat{\beta}_{PS} - \beta \leq z) = G(z; k)$, *with* $F_z(z)$ *as in (i). For each* r, *the density of* $\hat{\beta}_{PS} - \beta$ *is symmetric about zero.*

The proof of the first part of Proposition 1 relies on Lemma 3 in the Appendix, which shows that the median of the ratio of two independent $\mathcal{S}(a, 0, 1, 0)$ distributed random variables is unique and zero. Although there is no closed form expression for $F_z(z)$ in Proposition 1 in general, just like the normal distribution, the cdf can be efficiently and accurately evaluated using numerical integration (Nolan, 1997). The values of a, c_x and c_u are not needed to estimate the slope parameter β, but would be to construct confidence intervals. In practice, these nuisance parameters can be estimated using the explanatory variable and the residuals $\hat{\varepsilon}_i = y_i - \hat{\beta}x_i$, where $\varepsilon_i = \alpha + u_i$, and consistent estimators for the index of stability and scale parameters of a stable distribution. See Fama and Roll (1971), McCulloch (1986) and more recently Garcia, Renault and Veredas (2011) for examples of consistent estimators for stable distributions.

We end this section with two examples of $F_z(z)$ in Proposition 1. Table 1 reports results for different symmetric stable ratio distributions. We consider Cauchy ($a = 1$) and Gaussian ($a = 2$) distributions for the error and explanatory variable. For the latter specification, $F_z(z)$ is the cdf of a Cauchy distribution. Here the limiting distribution of $\hat{\beta}_{UF}$ (and $\hat{\beta}_{PS}$) is normal.[5] The asymptotic variance of the ordinary least squares (and maximum likelihood) estimator for β is σ_u^2/σ_x^2 whereas that of $\hat{\beta}_{UF}$ is $(\pi/2)^2\sigma_u^2/\sigma_x^2$. Hence, the ARE of this MLBE with respect to the, asymptotically efficient, LSE is $(2/\pi)^2 \approx 0.405$ for the Gaussian specification. For the Cauchy specification, $F_z(z)$ is an integral which, for computational purposes, can be expressed in terms of the polylogarithm (dilogarithm) function.

TABLE 1. Symmetric stable ratio distributions ($a = 1, 2$).

Distribution u_i	Distribution x_i	Ratio Distribution $z_i = u_i/(x_i - \mu_x)$	Ratio Distribution $z_i = (u_{2i} - u_{2i-1})/(x_{2i} - x_{2i-1})$
$\mathcal{S}(1, 0, c_u, 0)$ Cauchy	$\mathcal{S}(1, 0, c_x, \mu_x)$ Cauchy	$F_z(z) = \frac{1}{\pi^2}\int_{-\infty}^{(c_x/c_u)z} \frac{\ln(t^2)}{t^2-1}dt$	$F_z(z) = \frac{1}{\pi^2}\int_{-\infty}^{(c_x/c_u)z} \frac{\ln(t^2)}{t^2-1}dt$
$\mathcal{S}(2, 0, \sigma_u/\sqrt{2}, 0)$ Gaussian	$\mathcal{S}(2, 0, \sigma_x/\sqrt{2}, \mu_x)$ Gaussian	$F_z(z) = \frac{1}{2} + \frac{1}{\pi}\arctan\left(\frac{\sigma_x}{\sigma_u}z\right)$	$F_z(z) = \frac{1}{2} + \frac{1}{\pi}\arctan\left(\frac{\sigma_x}{\sigma_u}z\right)$

3. A Contaminated Normal Regression

So far we have restricted our analysis to symmetric stable random variables. In this section we outline how our results can be extended to include other types of regressions, with potentially contaminated errors. As an important special case, we derive the exact distributions of (3) and (5) in a contaminated normal regression.

The regression we consider is

$$\begin{cases} y_i = \alpha + \beta x_i + u_i \\ u_i = (1 - b_i)v_i + b_i\sqrt{\gamma}v_i \end{cases} \tag{7}$$

where x_i is normally distributed, b_i is Bernoulli distributed with success parameter p, v_i is normally distributed with mean zero and variance σ_v^2, $\gamma > 1$, and x_i, b_i and v_i are mutually

[5]Cf. Proposition 2, with $p = 0$.

independent. For this specification,

$$E(u_i) = 0, \quad E(u_i^2) = [1 + (\gamma - 1)p]\sigma_v^2, \tag{8}$$

and the density of u_i is symmetric about zero. The contamination parameters p and γ are potentially unknown. Here a 'small' value of p and a 'large' value of γ is a typical setting in which the finite-sample performance of the LSE for β may be poor. For $p = 0$ there is no contamination and (7) is a special case of (1), with $x_i \sim \mathcal{S}(2, 0, \sigma_x/\sqrt{2}, \mu_x)$ and $u_i \sim \mathcal{S}(2, 0, \sigma_v/\sqrt{2}, 0)$.

Assumption 2 holds throughout the section.

Assumption 2. *Let y_i $(i = 1, 2, \ldots, n)$ be given by (7), with $\mu_y = E(y_i)$. Suppose that*

(i) the x_i are independent $\mathcal{N}(\mu_x, \sigma_x^2)$,
(ii) the b_i are independent Bernoulli distributed, with success parameter p,
(iii) the v_i are independent $\mathcal{N}(0, \sigma_v^2)$ and $\gamma > 1$,
(iv) for each i, x_i, b_i and v_i are mutually independent,
(v) the sample size is odd, $n = 2k + 1$.

Here both the finite-sample and limiting distributions of $\hat{\beta}_{UF}$ and $\hat{\beta}_{PS}$ can be obtained.

Proposition 2. *Let $G(\cdot)$ be given by (4). Under Assumption 2,*

(i) if μ_y and μ_x are known, $\hat{\beta}_{UF} \xrightarrow{p} \beta$ as $n \to \infty$ and the exact distribution of (5) is given by $P(\hat{\beta}_{UF} - \beta \le z) = G(z; n)$, with

$$F_z(z) = (1 - p)F_r(z) + pF_r(z/\sqrt{\gamma}), \quad F_r(z) = \frac{1}{2} + \frac{1}{\pi}\arctan\Big(\frac{\sigma_x}{\sigma_v} z\Big).$$

For each k, the density of $\hat{\beta}_{UF} - \beta$ is symmetric about zero. The limiting distribution of $\hat{\beta}_{UF}$ is normal,

$$\sqrt{n}(\hat{\beta}_{UF} - \beta) \xrightarrow{d} \mathcal{N}\big(0, [4f_z^2(0)]^{-1}\big), \quad f_z(0) = \left[1 + \left(\frac{1 - \sqrt{\gamma}}{\sqrt{\gamma}}\right) p\right] \frac{\sigma_x}{\pi\sigma_v}.$$

(ii) if $k = 2r + 1$ is odd, $\hat{\beta}_{PS} \xrightarrow{p} \beta$ as $n \to \infty$ and the exact distribution of (3) is given by $P(\hat{\beta}_{PS} - \beta \le z) = G(z; k)$, with

$$F_z(z) = (1 - p)^2 F_r(z) + 2p(1 - p)F_r(\sqrt{2/(\gamma + 1)}z) + p^2 F_r(z/\sqrt{\gamma}),$$

and $F_r(z)$ as in (i). For each r, the density of $\hat{\beta}_{PS} - \beta$ is symmetric about zero. The limiting distribution of $\hat{\beta}_{PS}$ is normal,

$$\sqrt{n}(\hat{\beta}_{PS} - \beta) \xrightarrow{d} \mathcal{N}(0, [4f_z^2(0)]^{-1}), \quad f_z(0) = \left[(1 - p)^2 + 2p(1 - p)\sqrt{\frac{2}{\gamma + 1}} + \frac{p^2}{\sqrt{\gamma}}\right] \frac{\sigma_x}{\pi\sigma_v}.$$

By the proof of Proposition 2, it is clear that similar results can be obtained for higher order mixtures and for a wide variety of cases where x_i and/or v_i are non-normally distributed, with finite first and second moments, and the density of v_i is symmetric about zero.

4. A Heteroscedastic Regression

In this section we illustrate how our results can be extended to include certain types of conditionally heteroscedastic regressions. The regression we consider is

$$\begin{cases} y_i = \alpha + \beta x_i + u_i \\ u_i = \lambda(x_i)v_i \end{cases} \tag{9}$$

with e.g. $\lambda(x) = (x - \mu_x)^2$, $\lambda(x) = |x - \mu_x|$ or $\lambda(x) = 1$. We will assume that the distribution of the i.i.d. v_i is symmetric about zero, which implies that the distribution of the u_i is also symmetric about zero.[6] For ease of exposition, we consider $\hat{\beta}_{UF}$ and note that similar results can be obtained for $\hat{\beta}_{PS}$. Of course, $\lambda(x) = x - \mu_x$ is trivial. Here $\mu_y = \alpha + \beta\mu_x$ and $\hat{\beta}_{UF} = \beta + \text{med}\{v_1, v_2, \ldots, v_n\}$, assuming all expectations exist. Thus, in this case the distribution of $\hat{\beta}_{UF} - \beta$ when n is odd is given by $G(z; n)$ with $F_z(z) = F_v(z)$, where $F_v(\cdot)$ is the cdf of the v_i. For another example, suppose $\lambda(x) = (x - \mu_x)^2$ and the x_i and v_i are independent $\mathcal{N}(\mu_x, \sigma_x^2)$ and $\mathcal{N}(0, \sigma_v^2)$, respectively. In this case, the distribution of $\hat{\beta}_{UF} - \beta = (x_i - \mu_x)v_i$ is a product normal distribution [7]. Then similar results can be derived, since $(x_i - \mu_x)v_i$ is symmetric about zero and has a unique median[8].

5. Simulation Study

In this section we report simulation results concerning the estimation of the slope parameter $\beta = 3$ in the regression $y_i = 7 + 3x_i + u_i$ $(i = 1, 2, \ldots, n)$. We consider sample sizes of $n = 27, 55, 111, 223$ and 447 to ensure that both n and $\lfloor n/2 \rfloor$ are odd numbers, cf. assumptions 1 and 2, where $\lfloor \cdot \rfloor$ is the *integer part* function. These sample sizes are used to illustrate the relation between k and r in propositions 1–2. We emphasize that the consistency of the estimators that we consider does not rely on the values of n or k, however, our exact distributional results in sections 2 through 4 do. Table 2 shows simulation results for various specifications of the explanatory variable and error. We report the empirical bias and mean squared error (MSE) of the estimators $\hat{\beta}_{UF}$, $\hat{\beta}_{FE}$, $\hat{\beta}_{PS}$ and $\hat{\beta}_{LS}$. The estimator $\hat{\beta}_{FE}$, described below, is a feasible version of $\hat{\beta}_{UF}$. Each table entry is based on 1 000 000 simulated samples. and rounded to three decimal places. Symmetric stable pseudo-random numbers were generated using Theorem 1.19 (a) in Nolan (2012). All of the reported experiments share a common initial state of the generator for pseudo-random number generation and were carried out using R.

Symmetric Stable Regression. Panels A-D of Table 2 report simulation results when the x_i and u_i are i.i.d. $\mathcal{S}(a, 0, 1, 1)$ and $\mathcal{S}(a, 0, 1, 0)$, respectively, for $a = 1, 1.25, 1.5$ and 1.75. To estimate the location parameters μ_y and μ_x when constructing a feasible version of $\hat{\beta}_{UF}$ for the symmetric stable regression in Section 2 we use the symmetrically trimmed mean

$$\hat{\mu}_x = \tfrac{1}{\lfloor np_2 \rfloor - \lfloor np_1 \rfloor} \sum_{i=\lfloor np_1 \rfloor + 1}^{\lfloor np_2 \rfloor} x_{(i)},$$

with $p_1 = 0.25$ and $p_2 = 1 - p_1$. Here $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$ is the ordered sample of size n. The proportions p_1 and $1 - p_2$ represent the proportion of the sample trimmed at either ends. According to Fama and Roll (1968), the symmetrically trimmed mean performs very well over the entire range $1 \leq a \leq 2$ for this choice of p_1 and p_2. In all four experiments, the bias and MSE of $\hat{\beta}_{UF}$ and $\hat{\beta}_{FE}$ is reasonable. $\hat{\beta}_{PS}$ also performs reasonably well, but has a much larger MSE. As expected, the performance of $\hat{\beta}_{LS}$ is unacceptable for values of a close to 1.

Contaminated Normal Regression. Panels E-F of Table 2 report simulation results when the errors $u_i = (1 - b_i)v_i + b_i\sqrt{\gamma}v_i$ are contaminated normal. In these two experiments the x_i and v_i are i.i.d. $\mathcal{N}(1, 1)$ and $\mathcal{N}(0, 1)$, respectively, and the b_i are i.i.d. Bernoulli with success parameter $p = 0.05$ and 0.1. The contamination parameter $\gamma = 36$, implying that the error variance, given by (8), is 2.75 for $p = 0.05$ and 4.5 for $p = 0.1$. To estimate μ_y and μ_x when constructing a feasible version of $\hat{\beta}_{UF}$ for the contaminated normal regression in Section 3 we

[6]See Lamma 2

[7]For this see Claim 5

[8]For this see Claim 5

TABLE 2. Each table entry, based on 1 000 000 simulated samples, reports the empirical bias/mean squared error of different estimators for the slope parameter $\beta = 3$ in the simple linear regression $y_i = 7 + 3x_i + u_i$. The following estimators are considered: The unfeasible (UF), feasible (FE), incomplete pairwise slope (PS) and ordinary least squares (LS) estimator. *Panels A-D:* $x_i \sim \mathcal{S}(a, 0, 1, 1)$ and $u_i \sim \mathcal{S}(a, 0, 1, 0)$ (Symmetric Stable Regression). *Panels E-F:* $x_i \sim \mathcal{N}(1, 1)$ and $u_i = (1 - b_i)v_i + b_i\sqrt{36}v_i$, where b_i is Bernoulli with success parameter p and $v_i \sim \mathcal{N}(0, 1)$ (Contaminated Normal Regression). *Panels G-H:* $x_i \sim \mathcal{N}(1, \sigma_x^2)$ and $u_i = (x_i - 1)^2 v_i$, where $v_i \sim \mathcal{N}(0, 1)$ (Heteroscedastic Normal Regression). Different sample sizes (n), indices of stability (a), success parameters (p) and regressor variances (σ_x^2) are considered.

	$\hat{\beta}_{UF}$		$\hat{\beta}_{FE}$		$\hat{\beta}_{PS}$		$\hat{\beta}_{LS}$		$\hat{\beta}_{UF}$		$\hat{\beta}_{FE}$		$\hat{\beta}_{PS}$		$\hat{\beta}_{LS}$	
n	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
							Panel A ($a = 1$)		*Panel E* ($p = 0.05$)							
27	0.000	0.055	-0.008	0.093	0.000	0.237	0.987	1.276×10^6	0.005	0.126	0.004	0.147	0.000	0.391	-0.001	0.193
55	0.000	0.016	-0.004	0.022	0.000	0.055	0.022	1.315×10^3	0.006	0.057	0.005	0.061	0.001	0.149	0.000	0.089
111	0.000	0.006	-0.002	0.007	0.000	0.016	-0.019	1.060×10^3	0.007	0.027	0.007	0.028	0.004	0.067	0.004	0.043
223	0.000	0.002	-0.001	0.002	0.000	0.006	1.077	1.374×10^6	0.003	0.013	0.003	0.014	0.001	0.032	0.002	0.021
447	0.000	0.001	-0.000	0.001	0.000	0.002	0.028	8.089×10^2	0.002	0.007	0.002	0.007	0.001	0.015	0.001	0.010
							Panel B ($a = 1.25$)		*Panel F* ($p = 0.1$)							
27	0.000	0.069	-0.012	0.089	0.000	0.230	0.023	8.684×10^2	0.006	0.157	0.005	0.213	0.001	0.635	-0.002	0.354
55	0.000	0.025	-0.005	0.029	0.000	0.069	0.002	9.824	0.007	0.070	0.006	0.079	0.002	0.222	0.001	0.163
111	0.000	0.010	-0.002	0.011	0.000	0.025	-0.002	1.503×10^1	0.008	0.033	0.007	0.035	0.004	0.095	0.005	0.079
223	0.000	0.004	-0.001	0.004	0.000	0.010	0.015	4.889×10^2	0.004	0.016	0.003	0.017	0.002	0.044	0.002	0.039
447	0.000	0.002	-0.001	0.002	0.000	0.004	0.002	6.390	0.002	0.008	0.002	0.008	0.001	0.021	0.001	0.019
							Panel C ($a = 1.5$)		*Panel G* ($\sigma_x = 1$)							
27	0.000	0.082	-0.011	0.095	0.000	0.233	0.001	4.986	0.003	0.010	0.000	0.144	0.001	0.127	-0.008	0.455
55	0.000	0.033	-0.005	0.036	0.000	0.081	0.001	0.540	0.002	0.004	0.002	0.056	0.003	0.050	-0.005	0.246
111	0.000	0.014	-0.002	0.015	0.000	0.033	-0.000	0.520	0.002	0.001	0.004	0.022	0.004	0.022	0.003	0.128
223	0.000	0.007	-0.001	0.007	0.000	0.014	0.001	2.006	0.001	0.001	0.002	0.009	0.002	0.010	0.001	0.066
447	0.000	0.003	-0.001	0.003	0.000	0.007	0.000	0.134	0.000	0.000	0.001	0.004	0.001	0.005	0.001	0.033
							Panel D ($a = 1.75$)		*Panel H* ($\sigma_x = \sqrt{2}$)							
27	0.000	0.093	-0.009	0.103	-0.000	0.239	0.000	0.219	0.004	0.020	0.001	0.289	0.002	0.254	-0.011	0.910
55	0.000	0.041	-0.004	0.043	-0.000	0.093	0.000	0.082	0.003	0.007	0.003	0.113	0.004	0.101	-0.007	0.492
111	0.000	0.019	-0.002	0.019	0.000	0.041	0.000	0.044	0.003	0.003	0.005	0.045	0.005	0.044	0.004	0.257
223	0.000	0.009	-0.001	0.009	0.000	0.019	0.000	0.040	0.001	0.001	0.002	0.018	0.002	0.020	0.002	0.131
447	0.000	0.004	-0.001	0.004	0.000	0.009	0.000	0.010	0.001	0.001	0.001	0.007	0.001	0.009	0.001	0.066

[a]

[a]We also did the simulation when $n = 30, 60, 120, 230$ and 450 and the convergence pattern of $\hat{\beta}_{UF}$, $\hat{\beta}_{FE}$ and $\hat{\beta}_{PS}$ is maintained.

use the sample mean, $\hat{\mu}_x = n^{-1}\sum_{i=1}^{n} x_i$. In both experiments, the results indicate that the MSE of $\hat{\beta}_{UF}$ and $\hat{\beta}_{FE}$ is considerably smaller than that of $\hat{\beta}_{LS}$. However, the MSE of $\hat{\beta}_{PS}$ is considerably higher than that of the LSE.

Heteroscedastic Regression. Finally, panels G-H of Table 2 report simulation results when the errors $u_i = \lambda(x_i)v_i$ are conditionally heteroscedastic. We consider the last example of Section 4, where $\lambda(x) = (x - \mu_x)^2$, when the x_i and v_i are i.i.d. $\mathcal{N}(1, \sigma_x^2)$ and $\mathcal{N}(0, 1)$, respectively, for $\sigma_x^2 = 1$ and 2. To estimate μ_y and μ_x when constructing a feasible version of $\hat{\beta}_{UF}$ we use the sample mean. In general, the bias of the MLBEs is similar in magnitude to that of the LSE, however, the MLBEs appear to have a much smaller MSE.

6. Conclusions and Extensions

In this note we have established the consistency and exact finite-sample distributions of two median-unbiased MLBEs for the slope parameter in a simple linear regression model when (1) the explanatory variable and error are symmetric stable random variables, and (2) the explanatory variable is normal and the error is contaminated normal. These exact distributions may be used for statistical inference. In addition, we have illustrated how our results can be extended to include certain heteroscedastic regressions. Our simulation results indicated that the MLBEs can have superior finite-sample properties compared to the LSE.

Because of their robustness and ease of computation, along the lines of Hinich and Talwar (1975), $\hat{\beta}_{PS}$ or $\hat{\beta}_{FE}$ can also be used as a starting point for a more sophisticated method. For example, in the context of numerical maximum likelihood estimation in a symmetric stable linear regression, to search for a global optimum, $\hat{\beta}_{PS}$ (or $\hat{\beta}_{FE}$) could be used as an easily computable starting point for the numerical method.[9] A well chosen starting point may lead to a drastic decrease in computational time.[10] Kadiyala and Murthy (1977), for example, use $\hat{\beta}_{LS}$ as a starting point. In light of our simulation results, this is a poor choice.

We have aimed for clarity at the expense of generality. For example, results analogous to those of Proposition 1 can be obtained for the slope parameters in a general, symmetric stable, linear regression with two or more statistically independent explanatory variables. For another example, it appears that our results can be extended to allow for serially correlated errors using existing results for m-dependent samples (e.g. Sen, 1968a).[11] Finally, results analogous to those of propositions 1 and 2 can be obtained for a simple unit root process, $y_t = y_{t-1} + u_t$, with symmetric stable or contaminated normal errors. This is work in progress.

APPENDIX

The following lemmas are applied in the proof of Propositions 1–2.

Lemma 1 (Properties of Symmetric Stable Variates). *If $v \sim \mathcal{S}(a, 0, c_v, d_v)$ and $w \sim \mathcal{S}(a, 0, c_w, d_w)$ are independent, then*

(i) v is absolutely continuous, with a continuous and unimodal density,

[9] In view of Lemma 1, a starting point for α (the regression intercept) given a sample of size $n = 2k + 1$ could then be $\hat{\alpha}_{PS} = y_{(k+1)} - \hat{\beta}_{PS}x_{(k+1)}$. More generally, α can be estimated by the sample median of the PS residuals $\hat{\varepsilon}_i = y_i - \hat{\beta}_{PS}x_i$ (cf. Hettmansperger, McKean and Sheather, 1997).

[10] Maximum likelihood estimation of the general linear regression model with symmetric stable errors has been considered by Kadiyala and Murthy (1977) and Barmi and Nelson (1997), among others. In most cases there is no closed form expression for the MLE (McCulloch, 1998). The maximization of the likelihood function then imposes a high computational burden even for small to moderate sample sizes.

[11] A sequence $u_1, u_2, \ldots$ of random variables is said to be m-dependent if and only if u_i and u_{i+k} are pairwise independent for all $k > m$. In the special case when $m = 0$, m-dependence reduces to independence.

(ii) *the density of v is symmetric about d_v, and the support of v is $(-\infty, \infty)$,*
(iii) *if $1 < a \leq 2$, the mean of v is finite and equal to d_v,*
(iv) *if $0 < a < 2$, the variance of v does not exist,*
(v) *for any $\alpha \neq 0$ and real β, $\alpha + \beta v \sim \mathcal{S}(a, 0, |\beta| c_v, \alpha + \beta d_v)$,*
(vi) *$v + w \sim \mathcal{S}(a, 0, c, d_v + d_w)$, where $c^a = c_v^a + c_w^a$.*

Lemma 2 (Symmetric Product and Ratio Distributions)**.** *Suppose v and w are two independent absolutely continuous random variables, and v is symmetrically distributed about zero. Then, the product $p = vw$ and ratio $r = v/w$ are absolutely continuous and symmetrically distributed about zero, with pdf's*

$$f_p(p) = \int_{-\infty}^{\infty} \frac{1}{|t|} f_v(p/t) f_w(t) dt \quad \text{and} \quad f_r(r) = \int_{-\infty}^{\infty} |t| f_v(rt) f_w(t) dt, \tag{10}$$

where $f_v(\cdot)$ and $f_w(\cdot)$ are the pdf's of v and w, respectively.

Proof. By theorems 3.1 and 7.1 in Curtiss (1941), the cdf's of p and r are absolutely continuous. Moreover, the pdf's of p and r exist almost everywhere and are given by (10). Since v is symmetrically distributed about zero, $f_v(s) = f_v(-s)$ for all real s. The result now follows by noting that $f_p(p) = f_p(-p)$ and $f_r(r) = f_r(-r)$ for all real p and r. □

Remark 1. *Lemma 2 holds for v symmetrically distributed about zero, but not for v symmetrically distributed about $\theta \neq 0$, $f_v(\theta + s) = f_v(\theta - s)$ for all real s.*

Lemma 3 (Uniqueness of the Median of a Ratio of Symmetric Stable Variates)**.** *Suppose that $v \sim \mathcal{S}(a, 0, 1, 0)$ and $w \sim \mathcal{S}(a, 0, 1, 0)$ are independent, then the median of $r = v/w$ is unique and zero.*

Proof. Let $\epsilon > 0$ be arbitrary. By Lemma 2, the density of r is symmetric about zero. Hence, to show that the median is not an interval, it is enough to show that

$$\int_0^{\epsilon} f_r(t) dt = F_r(\epsilon) - F_r(0) = F_r(\epsilon) - \tfrac{1}{2} > 0.$$

For $a = 2$ the ratio is standard Cauchy, hence, $F_r(\epsilon) - \frac{1}{2} = \frac{1}{\pi} \arctan(\epsilon) > 0$. For $0 < a < 2$ Theorem 1 in Shcolnick (1985) gives

$$r \overset{d}{=} xy, \quad y = \left[\frac{\sin\left(\frac{\pi a}{2} z\right)}{\sin\left(\frac{\pi a}{2}(1 - z)\right)} \right]^{\frac{1}{a}}$$

where $x \sim \mathcal{S}(1, 0, 1, 0)$ and y are independent, $\overset{d}{=}$ denotes equality in distribution, and z is uniformly distributed on $(0, 1)$. Hence,

$$P(0 < r < \epsilon) = P(0 < xy < \epsilon) \geq P(0 < x < \epsilon, 0 < y < 1) = P(0 < x < \epsilon) P(0 < y < 1).$$

As x is standard Cauchy, $P(0 < x < \epsilon) = \frac{1}{\pi} \arctan(\epsilon)$. Next we show that $P(0 < y < 1) = 1/2$. Since $P(0 < z < 1) = 1$, we only consider solutions $0 < z < 1$ to $0 < y(z) < 1$. For this subset, $0 < y < 1$ if and only if

$$\sin\left(\tfrac{\pi a}{2}(1 - z)\right) - \sin\left(\tfrac{\pi a}{2} z\right) = 2 \cos\left(\tfrac{\pi a}{4}\right) \sin\left(\tfrac{\pi a}{4} - \tfrac{\pi a}{2} z\right) > 0.$$

It follows that $P(0 < y < 1) = P(0 < z < 1/2) = 1/2$. Thus,

$$F_r(\epsilon) - \tfrac{1}{2} = P(0 < r < \epsilon) \geq \tfrac{1}{2\pi} \arctan(\epsilon) > 0.$$

□

Lemma 4 (Symmetric Ratio Mixture Distribution). *Suppose that*

$$z = (1-b)\frac{v}{w} + b\sqrt{\gamma}\frac{v}{w},$$

where b is Bernoulli distributed with success parameter p, v and w are independent absolutely continuous random variables, v is symmetrically distributed about zero and $\gamma > 0$. Then the ratio mixture z is absolutely continuous and symmetrically distributed about zero, with pdf

$$f_z(z) = (1-p)f_r(z) + p(1/\sqrt{\gamma})f_r(z/\sqrt{\gamma}),$$

where $f_r(\cdot)$ is the ratio density of Lemma 2.

Proof.

$$\begin{aligned} f_z(z) &= \sum_{k=0}^{1} f_{z,b}(z,k) = \sum_{k=0}^{1} (1-p)^{1-k} p^k f_{z|b=k}(z) \\ &= (1-p)f_r(z) + p(1/\sqrt{\gamma})f_r(z/\sqrt{\gamma}), \end{aligned}$$

where we have used Lemma 2 and that the pdf of $h = \sqrt{\gamma}v$ is $(1/\sqrt{\gamma})f_v(h/\sqrt{\gamma})$, which is symmetric about zero. Since $f_r(r) = f_r(-r)$ for all real r, the result now follows by noting that $f_z(z) = f_z(-z)$ for all real z. □

Proof of Proposition 1. First we will show that

$$z_i = \frac{y_{2i} - y_{2i-1}}{x_{2i} - x_{2i-1}} - \beta \stackrel{d}{=} \frac{y_i - \mu_y}{x_i - \mu_x} - \beta \stackrel{d}{=} \left(\tfrac{c_u}{c_x}\right) r_i, \tag{11}$$

where r_i is the ratio of two independent $\mathcal{S}(a,0,1,0)$ random variables. Since $\mu_y = \alpha + \beta\mu_x$, we have

$$\frac{y_i - \mu_y}{x_i - \mu_x} = \frac{\beta(x_i - \mu_x) + u_i}{x_i - \mu_x} = \beta + \frac{u_i}{x_i - \mu_x}.$$

In view of Lemma 1,

$$\frac{u_i}{x_i - \mu_x} \stackrel{d}{=} \frac{v_i}{w_i} \stackrel{d}{=} \left(\tfrac{c_u}{c_x}\right) r_i,$$

where v_i and w_i are independent $\mathcal{S}(a,0,c_u,0)$ and $\mathcal{S}(a,0,c_x,0)$ variates, respectively. Similarly,

$$\frac{y_{2i} - y_{2i-1}}{x_{2i} - x_{2i-1}} = \beta + \frac{u_{2i} - u_{2i-1}}{x_{2i} - x_{2i-1}},$$

where

$$\frac{u_{2i} - u_{2i-1}}{x_{2i} - x_{2i-1}} \stackrel{d}{=} \frac{v_i}{w_i} \stackrel{d}{=} \left(\tfrac{c_u}{c_x}\right) r_i,$$

and v_i and w_i here are independent $\mathcal{S}(a,0,2^{1/a}c_u,0)$ and $\mathcal{S}(a,0,2^{1/a}c_x,0)$ variates, respectively. This shows (11). By Lemma 2, the pdf of r_i is symmetric about zero and the cdf of r_i is given by

$$F_r(r) = \int_{-\infty}^{r} \int_{-\infty}^{\infty} |t| f(st) f(t)\,dt\,ds,$$

where $f(\cdot)$ is the pdf of a $\mathcal{S}(a,0,1,0)$ variate. Hence, the density of z_i is symmetric about zero and the distribution of z_i is given by

$$F_z(z) = F_r(c_x z/c_u) = \int_{-\infty}^{(c_x/c_u)z} \int_{-\infty}^{\infty} |t| f(st) f(t)\,dt\,ds.$$

It follows that $P(\hat{\beta}_{PS} - \beta \le z) = P(z_{(r+1)} \le z)$, where $k = 2r+1$ and $z_{(r+1)}$ is the sample median of the i.i.d. sequence $\{z_1, z_2, \ldots, z_k\}$. Standard results for order statistics gives us the exact distribution of $z_{(r+1)}$ in terms of $F_z(z)$. The consistency of $\hat{\beta}_{PS}$ follows from Lemma 3. This proves (ii). The proof of (i) is analogous. □

Proof of Proposition 2. Since $\mu_y = \alpha + \beta\mu_x$, we have

$$\frac{y_i - \mu_y}{x_i - \mu_x} = \frac{\beta(x_i - \mu_x) + u_i}{x_i - \mu_x} = \beta + z_i,$$

where

$$z_i = (1 - b_i)r_i + b_i\sqrt{\gamma}r_i,$$

and $r_i = v_i/(x_i - \mu_x)$. It follows that $P(\hat{\beta}_{UF} - \beta \leq z) = P(z_{(k+1)} \leq z)$, where $n = 2k + 1$ and $z_{(k+1)}$ is the sample median of the i.i.d. sequence $\{z_1, z_2, \ldots, z_n\}$. By Lemma 4, the pdf of z_i is symmetric about zero, and the cdf of z_i is given by

$$F_z(z) = \int_{-\infty}^{z} f_z(t)dt = (1 - p)F_r(z) + pF_r(z/\sqrt{\gamma}),$$

where $F_r(\cdot)$, the cdf of r_i, can be obtained using Lemma 2. For the particular case when both x_i and v_i are assumed to be normal, r_i is $\mathcal{S}(1, 0, \sigma_v/\sigma_x, 0)$ distributed, with

$$F_r(z) = \frac{1}{2} + \frac{1}{\pi}\arctan\left(\frac{\sigma_x}{\sigma_v}z\right).$$

Hence, since F_z is strictly increasing, the unique solution to $F_z(\xi) = 1/2$ is $\xi = 0$. Standard results for order statistics gives us the exact distribution of $z_{(k+1)}$ in terms of $F_z(z)$. This proves the first part of (i). For the second part, note that the continuous pdf $f_z(z)$ of z_i is given by

$$F_z'(z) = \frac{(1-p)}{\pi}\frac{\frac{\sigma_v}{\sigma_x}}{(\frac{\sigma_v}{\sigma_x})^2 + z^2} + \frac{p}{\pi}\frac{\frac{\sqrt{\gamma}\sigma_v}{\sigma_x}}{\left(\frac{\sqrt{\gamma}\sigma_v}{\sigma_x}\right)^2 + z^2},$$

with

$$f_z(0) = \left[1 + \left(\frac{1 - \sqrt{\gamma}}{\sqrt{\gamma}}\right)p\right]\frac{\sigma_x}{\pi\sigma_v}.$$

Since also the derivative of $f_z(z)$ is continuous, standard results (Cramér, 1946, p. 369) gives us the limiting distribution in terms of $f_z(0)$,

$$\sqrt{n}(\hat{\beta}_{UF} - \beta) = \sqrt{n}z_{(k+1)} \xrightarrow{d} \mathcal{N}(0, [4f_z^2(0)]^{-1}).$$

This proves the second part of (i). The proof of (ii) is analogous. □

References

Barmi, H. E. and Nelson, P. (1997). Inference from stable distributions, *Selected Proceedings of the Symposium on Estimating Functions. IMS Lecture Notes-Monograph Series* **32**: 439–456.

Bateman, H. and Erdelyi, A. (1953). Higher Transcendental Functions, Vol. I, New York: McGraw-Hill.

Blattberg, R. and Sargent, T. (1971). Regression with non-gaussian stable disturbances: some sampling results, *Econometrica* **39**(3): 501–510.

Cramér, H. (1946). *Mathematical Methods of Statistics*, Princeton University Press.

Curtiss, J. (1941). On the distribution of the quotient of two chance variables, *The Annals of Mathematical Statistics* **12**(4): 409–421.

David, H. and Nagaraja, H. (2003). *Order Statistics*, 3rd edn, John Wiley & Sons.

Desu, M. and Rodine, R. (1969). Estimation of the population median, *Skandinavisk Aktuarietidskrift* **28**: 67–70.

Fama, E. (1965). The behavior of stock-market prices, *Journal of Business* 34–105.

Fama, E. and Roll, R. (1968). Some properties of symmetric stable distributions, *Journal of the American Statistical Association* **63**(323): 817–836.

Fama, E. and Roll, R. (1971). Parameter estimates for symmetric stable distributions, *Journal of the American Statistical Association* **66**(331–338).

F.W.J.Olver, Daniel, W.L., R.F. Boisvert and C.W. Clark (2010). *NIST Handbook of Mathematical Functions*, Cambridge University Press.

Garcia, R., Renault, E. and Veredas, D. (2011). Estimation of stable distributions by indirect inference, *Journal of Econometrics* **161**: 325–337.

Hettmansperger, T., McKean, J. and Sheather, S. (1997). *Handbook of Statistics*, Vol. 15, Elsevier, chapter 7, Rank-based analyses of linear models, 145–173.

Hinich, M. and Talwar, P. (1975). A simple method for robust regression, *Journal of the American Statistical Association* **70**(349): 113–119.

Jiang, J. (2010). *Large Sample Techniques for Statistics*, Springer.

Kadiyala, K. and Murthy, K. (1977). Estimation of regression equation with cauchy disturbances, *The Canadian Journal of Statistics* **5**(1): 111–120.

Mandelbrot, B. (1963). The variation of certain speculative prices, *Journal of Business* pp. 394–419.

McCulloch, J. (1986). Simple consistent estimators of stable distribution parameters, *Communications in Statistics - Simulation and Computation* **15**(4): 1109–1136.

McCulloch, J. (1998). *A Practical Guide to Heavy Tails*, Birkhäuser, chapter IV, Linear regression with stable disturbances, pp. 359–376.

Nolan, J. (1997). Numerical calculation of stable densities and distribution functions, *Communications in Statistics - Stochastic Models* **13**(4): 759–774.

Nolan, J. (2003). *Handbook of Heavy Tailed Distributions in Finance*, Handbooks in Finance, Elsevier, chapter 3, pp. 106–129.

Nolan, J. (2012). *Stable Distributions - Models for Heavy Tailed Data*, Birkhauser, Boston. In progress, Chapter 1 online at academic2.american.edu/∼jpnolan.

Oosterhoff, J. (1994). Trimmed mean or sample median?, *Statistics & Probability Letters* **20**: 401–409.

Preve, D. and Medeiros, M. C. (2011). Linear programming-based estimators in simple linear regression, *Journal of Econometrics* **165**: 128–136.

Sen, P. (1968a). Asymptotic normality of sample quantiles for m-dependent processes, *The Annals of Mathematical Statistics* **39**(5): 1724–1730.

Sen, P. (1968b). Estimates of the regression coefficient based on kendall's tau, *Journal of the American Statistical Association* **63**(324): 1379–1389.

Shcolnick, S. (1985). On the ratio of independent stable random variables, *in* V. Kalashnikov and V. Zolotarev (eds), *Stability Problems for Stochastic Models*, Vol. 1155 of *Lecture Notes in Mathematics*, Springer, pp. 349–354.

Springer, M.D. and Thompson, W.E. (1970). The distribution of products of beta, gamma and Gaussian random variables, *Journal on Applied Mathematics* **18 (4)**: 721–737.

Zolotarev, V. (1986). *One-Dimensional Stable Distributions*, Vol. 65 of *Translations of Mathematical Monographs*, American Mathematical Society.

EXTENDED APPENDIX

Claim 1 (Section 1: Paragraph 5)**.** *Suppose that the estimator $\hat{\beta}$ for β can be decomposed into $\hat{\beta} = \beta + \text{med}\{z_1, z_2, \dots, z_k\}$, where the z_i are i.i.d. continuous random variables with zero median. Then,*

(*i*) *the median of $\hat{\beta} - \beta$ is zero (i.e. $\hat{\beta}$ is median-unbiased) and*
(*ii*) *if, in addition, the density of z_i is symmetric about zero, then so is the density of $\hat{\beta} - \beta$.*

Proof. Let $C = \Gamma(k+1)/[\Gamma(r+1)]^2$. Then, in view of (4), since $F_z(0) = 1/2$

$$G(0;k) = C\int_0^{1/2} t^r(1-t)^r dt \overset{s=1-t}{=} -C\int_1^{1/2}(1-s)^r s^r ds = C\int_{1/2}^{1} s^r(1-s)^r ds.$$

Seeing that the sum of the first and last integral is one, it follows that $G(0;k) = 1/2$. This proves (*i*). For the proof of (*ii*), note that $F_z(z) = 1 - F_z(-z)$ as the density of z_i is symmetric about zero. Hence,

$$\begin{aligned} G(z;k) &= C\int_0^{F_z(z)} t^r(1-t)^r dt = C\int_0^{1-F_z(-z)} t^r(1-t)^r dt \overset{s=1-t}{=} -C\int_1^{F_z(-z)}(1-s)^r s^r ds \\ &= C\int_{F_z(-z)}^{1} s^r(1-s)^r ds = 1 - C\int_0^{F_z(-z)} s^r(1-s)^r ds = 1 - G(-z;k). \end{aligned}$$

This proves (*ii*). □

Claim 2 (Section 2: Conditional Distribution of $\hat{\beta}_{LS}$)**.** *By (2), the LSE for β can be decomposed into*

$$\hat{\beta}_{LS} = \beta + \frac{\sum_{i=1}^n (x_i - \bar{x})(u_i - \bar{u})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \beta + (c_1 - \bar{c})u_1 + (c_2 - \bar{c})u_2 + \ldots + (c_n - \bar{c})u_n,$$

where

$$c_i = \frac{x_i - \bar{x}}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

Hence, conditional on the x_i, $\hat{\beta}_{LS} - \beta$ is a linear combination of independent, identically $\mathcal{S}(a, 0, c_u, 0)$ distributed random variables which, in view of Lemma 1, is symmetric stable with index of stability a. The claim now follows.

Claim 3 (Section 2: Results in Table 1)**.** *Since $\mathcal{S}(2, 0, \sigma/\sqrt{2}, \mu) = \mathcal{N}(\mu, \sigma^2)$, we have*

$$\frac{u_i}{x_i - \mu_x} \overset{d}{=} \Big(\frac{\sigma_u}{\sigma_x}\Big) r_i,$$

where r_i is the ratio of two independent $\mathcal{N}(0,1)$ random variables. Hence, r_i is $\mathcal{C}(0,1)$ where $\mathcal{C}(0,1)$ denotes the standard Cauchy distribution. Similarly,

$$\frac{u_{2i} - u_{2i-1}}{x_{2i} - x_{2i-1}} \overset{d}{=} \frac{v_i}{w_i} \overset{d}{=} \Big(\frac{\sigma_u}{\sigma_x}\Big) r_i,$$

where v_i and w_i are independent $\mathcal{N}(0, 2\sigma_u^2)$ and $\mathcal{N}(0, 2\sigma_x^2)$ variates, respectively. As the cdf of a standard Cauchy variate is $\frac{1}{2} + \frac{1}{\pi}\arctan(z)$, the results on the second row in Table 1 now follow. To show the results on the first row, note that $\mathcal{S}(1, 0, c, \mu) = \mathcal{C}(\mu, c)$. In view of Lemma 1,

$$\frac{u_i}{x_i - \mu_x} \overset{d}{=} \frac{v_i}{w_i} \overset{d}{=} \Big(\frac{c_u}{c_x}\Big) r_i,$$

where v_i and w_i are independent $\mathcal{C}(0, c_u)$ and $\mathcal{C}(0, c_x)$ variates, respectively, and r_i is the ratio of two independent $\mathcal{C}(0, 1)$ random variables. Similarly,

$$\frac{u_{2i} - u_{2i-1}}{x_{2i} - x_{2i-1}} \overset{d}{=} \frac{v_i}{w_i} \overset{d}{=} \Big(\frac{c_u}{c_x}\Big) r_i,$$

where v_i and w_i here are independent $\mathcal{C}(0, 2c_u)$ and $\mathcal{C}(0, 2c_x)$ variates, respectively. By Theorem 3.1 in Curtiss (1941), the density of the ratio of two independent standard Cauchy variates exists almost everywhere and is given by

$$\begin{aligned} f_r(r) &= \int_{-\infty}^{\infty} |t| \left(\frac{1}{\pi} \frac{1}{1 + (rt)^2}\right) \left(\frac{1}{\pi} \frac{1}{1 + t^2}\right) dt = \frac{1}{\pi^2} \int_0^{\infty} \frac{2t}{(1 + r^2t^2)(1 + t^2)} dt \\ &= \frac{1}{\pi^2} \int_0^{\infty} \frac{1}{(1 + r^2 s)(1 + s)} ds. \end{aligned} \tag{12}$$

From (12) it is readily seen that $f_r(r)$ is equal to $1/\pi^2$ for $r = \pm 1$, and is divergent for $r = 0$. For $r \neq \pm 1$ partial fraction decomposition yields,

$$\begin{aligned} f_r(r) &= \frac{1}{\pi^2} \int_0^{\infty} \frac{1}{(1 + r^2 s)(1 + s)} ds = \frac{1}{\pi^2} \int_0^{\infty} \frac{\left(\frac{r^2}{r^2 - 1}\right)}{1 + r^2 s} - \frac{\left(\frac{1}{r^2 - 1}\right)}{1 + s} ds \\ &= \lim_{t \to \infty} \frac{1}{\pi^2} \left(\frac{1}{r^2 - 1}\right) \ln\left(\frac{1 + r^2 t}{1 + t}\right) = \frac{1}{\pi^2} \frac{\ln(r^2)}{r^2 - 1}. \end{aligned}$$

A closer analysis shows that $f_r(r)$ is continuous at $r = \pm 1$. Hence, the ratio density f_r is continuous on $(-\infty, 0)$ and $(0, \infty)$, with a pole at zero. The results on the first row in Table 1 now follow. □

Claim 4 (Section 3: Part (*ii*) of Proposition 2)**.** *Since $\mu_y = \alpha + \beta\mu_x$, we have*

$$\frac{y_{2i} - y_{2i-1}}{x_{2i} - x_{2i-1}} = \frac{\beta(x_{2i} - x_{2i-1}) + u_{2i} - u_{2i-1}}{x_{2i} - x_{2i-1}} = \beta + z_i,$$

where

$$z_i = \frac{u_{2i} - u_{2i-1}}{x_{2i} - x_{2i-1}} = \frac{(1 - b_{2i})v_{2i} + b_{2i}\sqrt{\gamma}v_{2i} - (1 - b_{2i-1})v_{2i-1} - b_{2i-1}\sqrt{\gamma}v_{2i-1}}{x_{2i} - x_{2i-1}}.$$

It follows that $P(\hat{\beta}_{PS} - \beta \leq z) = P(z_{(r+1)} \leq z)$, where $k = 2r + 1$ and $z_{(r+1)}$ is the sample median of the i.i.d. sequence $\{z_1, z_2, \ldots, z_k\}$. The cdf of z_i is given by

$$F_z(z) = P(z_i \leq z) = \sum_{l,m} P(b_{2i} = l, b_{2i-1} = m) P(z_i \leq z | b_{2i} = l, b_{2i-1} = m),$$

where $l, m = 0, 1$ and

$$\begin{aligned} P(z_i \leq z | b_{2i} = 0, b_{2i-1} = 0) &= P\Big(\frac{v_{2i} - v_{2i-1}}{x_{2i} - x_{2i-1}} \leq z\Big) = \frac{1}{2} + \frac{1}{\pi} \arctan\Big(\frac{\sigma_x}{\sigma_v} z\Big), \\ P(z_i \leq z | b_{2i} = 0, b_{2i-1} = 1) &= P\Big(\frac{v_{2i} - \sqrt{\gamma}v_{2i-1}}{x_{2i} - x_{2i-1}} \leq z\Big) = \frac{1}{2} + \frac{1}{\pi} \arctan\Big(\frac{\sigma_x}{\sigma_v} \sqrt{\frac{2}{\gamma + 1}} z\Big), \\ P(z_i \leq z | b_{2i} = 1, b_{2i-1} = 0) &= P\Big(\frac{\sqrt{\gamma}v_{2i} - v_{2i-1}}{x_{2i} - x_{2i-1}} \leq z\Big) = \frac{1}{2} + \frac{1}{\pi} \arctan\Big(\frac{\sigma_x}{\sigma_v} \sqrt{\frac{2}{\gamma + 1}} z\Big), \\ P(z_i \leq z | b_{2i} = 1, b_{2i-1} = 1) &= P\Big(\frac{\sqrt{\gamma}v_{2i} - \sqrt{\gamma}v_{2i-1}}{x_{2i} - x_{2i-1}} \leq z\Big) = \frac{1}{2} + \frac{1}{\pi} \arctan\Big(\frac{\sigma_x}{\sigma_v} \frac{z}{\sqrt{\gamma}}\Big). \end{aligned}$$

Hence,

$$F_z(z) = (1 - p)^2 F_r(z) + 2p(1 - p) F_r(\sqrt{2/(\gamma + 1)} z) + p^2 F_r(z/\sqrt{\gamma}),$$

where

$$F_r(z) = \frac{1}{2} + \frac{1}{\pi}\arctan\Big(\frac{\sigma_x}{\sigma_v}z\Big).$$

Since F_z is strictly increasing, the unique solution to $F_z(\xi) = 1/2$ is $\xi = 0$. Standard results for order statistics gives us the exact distribution of $z_{(r+1)}$ in terms of $F_z(z)$. This proves the first part of (ii). For the second part, note that the continuous pdf $f_z(z)$ of z_i is given by

$$F_z'(z) = (1-p)^2 f_r(z) + 2p(1-p)\sqrt{2/(\gamma+1)}f_r(\sqrt{2/(\gamma+1)}z) + (p^2/\sqrt{\gamma})f_r(z/\sqrt{\gamma}),$$

where

$$f_r(z) = \frac{1}{\pi}\frac{\sigma_v/\sigma_x}{(\sigma_v/\sigma_x)^2 + z^2}.$$

It follows that $f_z(z) = f_z(-z)$ for all real z and, hence, that the density of z_i is symmetric about zero. Since also the derivative of $f_z(z)$ is continuous, standard results gives us the limiting distribution in terms of $f_z(0)$,

$$\sqrt{n}(\hat{\beta}_{PS} - \beta) = \sqrt{n}z_{(r+1)} \xrightarrow{d} \mathcal{N}(0, [4f_z^2(0)]^{-1}),$$

where

$$f_z(0) = \left[(1-p)^2 + 2p(1-p)\sqrt{\frac{2}{\gamma+1}} + \frac{p^2}{\sqrt{\gamma}}\right]\frac{\sigma_x}{\pi\sigma_v}.$$

This proves the second part of (ii).

Claim 5 (Section 4: Unique median of product normal distribution)**.** *Let X and Y are independent normal distribution with zero mean and variance σ_x^2 and σ_y^2, and $Z = XY$. Then the density of Z can be expressed by the modified Bessel function of the second kind. And the median of Z is unique.*

Proof. By Theorem 5 in Springer and Thompson (1970), the probability density function of Z can be expressed as a Meijer G-function multiplied by a normalizing constant H, i.e.

$$f_Z(z) = HG_{02}^{20}\left(\frac{z^2}{4\sigma_x^2\sigma_y^2}|0,0\right)$$

where $H = (2\pi\sigma_x\sigma_y)^{-1}$. By the results from Bateman and Erdelyi (1953), we know that

$$G_{02}^{20}\left(\frac{z^2}{4\sigma_x^2\sigma_y^2}|0,0\right) = 2K_0\left(\frac{|z|}{\sigma_x\sigma_y}\right)$$

where K_0 is the modified Bessel function of the second kind. Thus, the probability density function of Z can be expressed as $K_0\left(\frac{|z|}{\sigma_x\sigma_y}\right)/\pi\sigma_x\sigma_y$. By the results from Frank (2010), we know that the Bessel function of the second kind $K_0(z) > 0$ and is a monotone decreasing function on $(0,\infty)$, and $\lim_{z\to 0} K_0(z) = \infty$. Hence $\forall\epsilon > 0$, $\Pr(0 < Z < \epsilon) > 0$. By the same way in Lemma 3, this implies that the median of Z is 0 and unique since the distribution function of Z is symmetric about 0. □

Claim 6 (Section 6: Multiple Linear Regression Extension)**.** *Let $y_i = \alpha + \sum_{j=1}^{q}\beta_j x_{ji} + u_i$ $(i = 1, 2, \ldots, n)$, with median μ_y. Suppose that*

(i) *the x_{ji} are independent $\mathcal{S}(a, 0, c_{x_j}, \mu_{x_j})$,*
(ii) *the u_i are independent $\mathcal{S}(a, 0, c_u, 0)$,*
(iii) *for each i and j, x_{ji} and u_i are mutually independent,*
(iv) *the sample size is odd, $n = 2k + 1$.*

For ease of exposition, consider the extended unfeasible estimator and, for ease of notation, denote it by

$$\hat{\beta}_j = \operatorname{med}\Big\{\frac{y_1-\mu_y}{x_{j1}-\mu_{x_j}}, \frac{y_2-\mu_y}{x_{j2}-\mu_{x_j}}, \dots, \frac{y_n-\mu_y}{x_{jn}-\mu_{x_j}}\Big\}. \tag{13}$$

Then $\hat{\beta}_j \xrightarrow{p} \beta_j$ as $n \to \infty$ $(j = 1, \dots, q)$ and the exact distribution of (13) is given by

$$P(\hat{\beta}_j - \beta_j \le z) = \frac{\Gamma(n+1)}{\Gamma(k+1)\Gamma(k+1)} \int_0^{F_{z_j}(z)} t^k (1-t)^k dt,$$

with

$$F_{z_j}(z) = \int_{-\infty}^{(c_{x_j}/c_j)z} \int_{-\infty}^{\infty} |t| f(st) f(t) dt ds,$$

where $f(\cdot)$ is the density of a $\mathcal{S}(a,0,1,0)$ distributed random variable and $c_j^a = c_u^a + \sum_{m\ne j} |\beta_m|^a c_{x_m}^a$. For each k, the density of $\hat{\beta}_j - \beta_j$ is symmetric about zero. If $a = 2$ the limiting distribution of $\hat{\beta}_j$ is normal,

$$\sqrt{n}(\hat{\beta}_j - \beta_j) \xrightarrow{d} \mathcal{N}\big(0, [\pi(c_j/c_{x_j})/2]^2\big).$$

Proof. In view of Lemma 1, $\mu_y = \alpha + \sum_{j=1}^q \beta_j \mu_{x_j}$, hence,

$$\begin{aligned}\frac{y_i-\mu_y}{x_{ji}-\mu_{x_j}} &= \frac{\beta_1(x_{1i}-\mu_{x_1}) + \ldots + \beta_j(x_{ji}-\mu_{x_j}) + \ldots + \beta_q(x_{qi}-\mu_{x_q}) + u_i}{x_{ji}-\mu_{x_j}} \\ &= \beta_j + z_{ji}.\end{aligned}$$

Let r_{ji} denote the ratio of two independent $\mathcal{S}(a,0,1,0)$ random variables. By assumption,

$$z_{ji} = \frac{\sum_{m\ne j} \beta_m (x_{mi}-\mu_{x_m}) + u_i}{x_{ji}-\mu_{x_j}} \overset{d}{=} \frac{v_{ji}}{w_{ji}} \overset{d}{=} \Big(\frac{c_j}{c_{x_j}}\Big) r_{ji},$$

where v_{ji} and w_{ji} are independent $\mathcal{S}(a,0,c_j,0)$ and $\mathcal{S}(a,0,c_{x_j},0)$ variates, respectively, and $c_j^a = c_u^a + \sum_{m\ne j} |\beta_m|^a c_{x_m}^a$. By Lemma 2, the pdf of r_{ji} is symmetric about zero and the cdf of r_{ji} is given by

$$F_r(r) = \int_{-\infty}^{r} \int_{-\infty}^{\infty} |t| f(st) f(t) dt ds,$$

where $f(\cdot)$ is the pdf of a $\mathcal{S}(a,0,1,0)$ variate. Hence, the density of z_{ji} is symmetric about zero and the distribution of z_{ji} is given by

$$F_{z_j}(z) = F_r(c_{x_j} z/c_j) = \int_{-\infty}^{(c_{x_j}/c_j)z} \int_{-\infty}^{\infty} |t| f(st) f(t) dt ds.$$

It follows that $P(\hat{\beta}_j - \beta_j \le z) = P(z_{j(k+1)} \le z)$, where $z_{j(k+1)}$ is the sample median of the i.i.d. sequence $\{z_{j1}, z_{j2}, \dots, z_{jn}\}$. Standard results for order statistics gives us the exact distribution of $z_{j(k+1)}$ in terms of $F_{z_j}(z)$. The consistency of $\hat{\beta}_j$ follows from Lemma 3. This proves the first part of the claim. For the second part, note that if $a = 2$ the numerator and denominator of z_{ji} are independent Gaussian random variables. Hence, z_{ji} is Cauchy distributed with zero median, scale parameter c_j/c_{x_j} (cf. Nolan, 2012, p. 23), and cdf

$$F_{z_j}(z) = \frac{1}{2} + \frac{1}{\pi}\arctan\Big(\frac{c_{x_j}}{c_j} z\Big).$$

The continuous pdf $f_{z_j}(z)$ of z_{ji} is given by

$$F'_{z_j}(z) = \frac{(c_j/c_{x_j})}{\pi[(c_j/c_{x_j})^2 + z^2]},$$

with

$$f_{z_j}(0) = \frac{c_{x_j}}{\pi c_j}.$$

Since also the derivative of $f_{z_j}(z)$ is continuous, standard results (Cramér, 1946, p. 369) gives us the limiting distribution in terms of $f_{z_j}(0)$,

$$\sqrt{n}(\hat{\beta}_j - \beta_j) = \sqrt{n} z_{j(k+1)} \xrightarrow{d} \mathcal{N}(0, [4f_z^2(0)]^{-1}).$$

This proves the second part of the claim. $\square$