

**EFFICIENCY OF LINEAR ESTIMATORS UNDER HEAVY-TAILEDNESS:
CONVOLUTIONS OF α -SYMMETRIC DISTRIBUTIONS ¹**

Running title: EFFICIENCY OF LINEAR ESTIMATORS

By Rustam Ibragimov²

Department of Economics, Yale University

Address for manuscript correspondence:

Rustam Ibragimov
Department of Economics
Harvard University
Littauer Center
1875 Cambridge St.
Cambridge, MA 02138

Email address: ribragim@fas.harvard.edu

Phone: +1-617-496-4795

Fax: +1-617-495-7730

ABSTRACT

The present paper focuses on the analysis of efficiency, peakedness and majorization properties of linear estimators under heavy-tailedness assumptions. The main results show that peakedness and majorization properties of log-concavely distributed random samples established by Proschan (1965) continue to hold for convolutions of

¹The results in this paper constitute a part of the author's dissertation "New majorization theory in economics and martingale convergence results in econometrics" presented to the faculty of the Graduate School of Yale University in candidacy for the degree of Doctor of Philosophy in Economics in March, 2005. Some of the results were originally contained in the work circulated in 2003-2005 under the titles "Shifting paradigms: On the robustness of economic models to heavy-tailedness assumptions" and "On the robustness of economic models to heavy-tailedness assumptions"

²I am indebted to my advisors, Donald Andrews, Peter Phillips and Herbert Scarf, for all their support and guidance in all stages of the current project. I also thank the Associate Editor, two anonymous referees, Donald Brown, Aydin Cecen, Gary Chamberlain, Brian Dineen, Darrell Duffie, Xavier Gabaix, Tilmann Gneiting, Philip Haile, Wolfgang Härdle, Boyan Jovanovic, Samuel Karlin, Benoît Mandelbrot, Alex Maynard, Marcelo Morreira, Ingram Olkin, Ben Polak, Gustavo Soares, Kevin Song and the participants at seminars at the Departments of Economics at Yale University, University of British Columbia, the University of California at San Diego, Harvard University, the London School of Economics and Political Science, Massachusetts Institute of Technology, the Université de Montréal, McGill University and New York University, the Division of the Humanities and Social Sciences at California Institute of Technology, Nuffield College, University of Oxford, and the Department of Statistics at Columbia University as well as the participants at the 18th New England Statistics Symposium at Harvard University, April 2004, the International Conference on Stochastic Finance, Lisbon, Portugal, September 2004, and the Conference on "Heavy tails and stable Paretian distributions in finance and macroeconomics" in celebration of the 80th birthday of Benoît Mandelbrot, Deutsche Bundesbank, Eltville, Germany, November 2005, for many helpful comments and discussions.

α -symmetric distributions with $\alpha > 1$. However, these properties are reversed in the case of convolutions of α -symmetric distributions with $\alpha < 1$.

Among other results, the paper shows that the sample mean is the best linear unbiased estimator of the population mean for not extremely heavy-tailed populations in the sense of its peakedness properties. In addition, in such a case, the sample mean exhibits the property of monotone consistency and, thus, an increase in the sample size always improves its performance. However, efficiency of the sample mean in the sense of its peakedness decreases with the sample size if the sample mean is used to estimate the population center under extreme heavy-tailedness. The paper also provides applications of the main efficiency and majorization comparison results in the study of concentration inequalities for linear estimators.

KEYWORDS: Linear estimators, efficiency, peakedness, majorization, robustness, heavy-tailed distributions, dependence, α -symmetric distributions, sample mean, monotone consistency

JEL Classification: C12, C13, C16

1 Introduction and discussion of the results

1.1 Efficiency and peakedness of estimators

A number of problems in econometrics and statistics involve comparisons of estimators' performance. The present paper focuses on comparisons of linear estimators under heavy-tailedness and obtains characterizations of optimal linear estimators for heavy-tailed data.

Let $\hat{\theta}^{(1)}$ and $\hat{\theta}^{(2)}$ be two estimators of a population parameter $\theta \in \mathbf{R}$. In the case when $\hat{\theta}^{(i)}$, $i = 1, 2$, are unbiased for θ and have finite second moments, their comparisons are traditionally based on quadratic loss functions leading to comparisons of the variances $Var(\hat{\theta}^{(i)})$, $i = 1, 2$: $\hat{\theta}^{(1)}$ is preferred to $\hat{\theta}^{(2)}$ if $Var(\hat{\theta}^{(1)}) < Var(\hat{\theta}^{(2)})$ (in other words, if $\hat{\theta}^{(1)}$ is more efficient than $\hat{\theta}^{(2)}$).

This approach breaks down, however, in the case of heavy-tailed estimators $\hat{\theta}^{(i)}$ for which variances do not exist and one has to rely on loss functions more general than quadratic ones. In the case of an increasing loss function $U : \mathbf{R}_+ = [0, \infty) \rightarrow \mathbf{R}$, $\hat{\theta}^{(1)}$ is preferred to $\hat{\theta}^{(2)}$ if (provided that the expectations exist)

$$EU(|\hat{\theta}^{(1)} - \theta|) < EU(|\hat{\theta}^{(2)} - \theta|) \quad (1.1)$$

(in the efficiency literature, it is common to consider loss functions that satisfy additional assumptions of boundedness). Orderings of estimators based on comparisons (1.1) are, of course, dependent on the choice of the loss functions U .

A natural approach to comparison of performance of estimators is to order them by the likelihood of observing their large deviations from the true parameter. This approach corresponds to the choice of indicator functions $U_\epsilon(x) = I(x > \epsilon)$, $\epsilon > 0$, in (1.1) and relies on the concept of peakedness of random variables (r.v.'s) introduced by Birnbaum (1948).

Definition 1.1 (Birnbaum, 1948). *A r.v. X is more peaked about $\theta \in \mathbf{R}$ than is Y if $P(|X - \theta| > \epsilon) \leq P(|Y - \theta| > \epsilon)$ for all $\epsilon \geq 0$. If this inequality is strict whenever the two probabilities are not both zero or both one, then the r.v. X is said to be strictly more peaked about θ than is Y . In case $\theta = 0$, X is simply said to be (strictly) more peaked than Y .*

The following definition introduces a peakedness-based analogue of the concept of efficiency for estimators that will be explored throughout the paper.

Definition 1.2 *The estimator $\hat{\theta}^{(1)}$ is said to be more efficient than $\hat{\theta}^{(2)}$ in the sense of peakedness (P -more efficient than $\hat{\theta}^{(2)}$ for short) if $\hat{\theta}^{(1)}$ is strictly more peaked about θ than is $\hat{\theta}^{(2)}$.*

The property of being P -less efficient is defined in a similar way. Roughly speaking, $\hat{\theta}^{(1)}$ is P -more efficient than $\hat{\theta}^{(2)}$ if the distribution of $\hat{\theta}^{(1)}$ is more concentrated about the true parameter θ than is that of $\hat{\theta}^{(2)}$. As follows from well-known properties of first-order stochastic dominance (see Shaked and Shanthikumar, 1994, pp. 3-4, and Remark 3.2 in this paper), if $\hat{\theta}^{(1)}$ is P -more efficient than $\hat{\theta}^{(2)}$, then comparisons (1.1) are independent of the choice of U and hold for any increasing loss function.

Comparisons of estimators are closely related to the analysis of the problem of whether having more data improves performance of an estimator of a population parameter. Indeed, obviously, an increase in the sample size always

improves performance of the estimator $\hat{\theta}_n$ of a population parameter θ if $\hat{\theta}_{n+1}$ is P-more efficient than $\hat{\theta}_n$ for all $n \geq 1$. In contrast, having larger samples is disadvantageous for performance of the estimator if P-efficiency of $\hat{\theta}_n$ decreases with n . Increasing P-efficiency is the basis for the following definition of monotone consistency, the concept studied by, e.g., Proschan (1965), Tong (1994) and Jensen (1997).

Definition 1.3 *A weakly consistent estimator $\hat{\theta}_n$ of a population parameter θ is said to exhibit monotone consistency for θ if $\hat{\theta}_{n+1}$ is P-more efficient than $\hat{\theta}_n$ for all $n \geq 1$ and, thus, $P(|\hat{\theta}_n - \theta| > \epsilon)$ converges to zero monotonically in n for all $\epsilon > 0$.*

1.2 Objectives and key results

The present paper focuses on the analysis of efficiency and peakedness properties of linear estimators under heavy-tailedness assumptions and on the study of efficiency and monotone consistency properties of the sample mean for heavy-tailed data. The main results show that peakedness and majorization properties of log-concavely distributed random variables established by Proschan (1965) continue to hold for convolutions of α -symmetric distributions with $\alpha > 1$ (Theorem 3.1 and 3.3). However, these properties are reversed in the case of convolutions of α -symmetric distributions with $\alpha < 1$ (Theorem 3.2 and 3.4).

Among other results, we show that the sample mean $\bar{X}_n = (1/n) \sum_{i=1}^n X_i$ is the best linear unbiased estimator of the population mean in the sense of P-efficiency for not extremely heavy-tailed populations (see Theorem 3.1). However, according to our results, P-efficiency of the sample mean is smallest among all linear estimators $\hat{\theta}(a) = \sum_{i=1}^n a_i X_i$ with weights $a_i \geq 0$, $i = 1, \dots, n$, $\sum_{i=1}^n a_i = 1$, of the population center in the case of extremely heavy-tailed data (Theorem 3.2). The above results imply that P-efficiency of \bar{X}_n is increasing in n for not extremely heavy-tailed populations and, thus, an increase in the sample size always improves performance of the sample mean in such a case. In the case of data from extremely heavy-tailed populations, P-efficiency of the sample mean is, however, decreasing in the sample size n .

Convolutions of α -symmetric distributions considered in the paper exhibit both heavy-tailedness and dependence (see Section 2). In particular, they contain, as subclasses, convolutions of certain models with common shocks affecting all heavy-tailed risks as well as spherical distributions which are α -symmetric with $\alpha = 2$. Spherical distributions, in turn, include such examples as Kotz type, multinormal, logistic and multivariate α -stable distributions. In addition, they include a subclass of mixtures of normal distributions as well as multivariate t -distributions that were used in the literature to model heavy-tailedness phenomena with dependence and finite moments up to a certain order. The results in the paper are also obtained for skewed stable distributions (such as, for instance, extremely heavy-tailed Lévy distributions with $\alpha = 1/2$ concentrated on the positive semi-axis) as well as for r.v.'s with non-identical one-dimensional distributions.

The law of large numbers (LLN) provides conditions (such as existence of first moments) under which the sample mean \bar{X}_n converges in probability to the population mean μ : $P(|\bar{X}_n - \mu| > \epsilon) \rightarrow 0$ for all $\epsilon > 0$. However, as discussed in, e.g., Proschan (1965), nothing is said by the LLN about the probability $P(|\bar{X}_n - \mu| > \epsilon)$ of a given size deviation of \bar{X}_n from μ decreasing monotonically as the sample size n increases. From the results in Proschan (1965) (see Remark 3.1) it follows that $P(|\bar{X}_n - \mu| > \epsilon)$ converges to zero monotonically in the case of i.i.d. symmetric r.v.'s X_i with log-concave distributions which are, as discussed in the next section, extremely light-tailed. The results in the present paper imply that monotone decrease of $P(|\bar{X}_n - \mu| > \epsilon)$, where μ is the population center, continues

to hold in the case of convolutions of α -symmetric distributions with $\alpha > 1$. On the other hand, according to the results in this paper, the tail probabilities $P(|\bar{X}_n - \mu| > \epsilon)$, where μ is the population center, monotonically diverge from zero in the case of convolutions of α -symmetric distributions with $\alpha < 1$ that have infinite first moments. According to these results and their more general analogues established in the paper, the sample mean and other linear estimators perform poorly in inference about the population center under extreme heavy-tailedness. Therefore, more robust statistical procedures, such as those based on sample medians, must be employed in such a setting.

The results obtained in the paper have applications in the study of robustness of model of firm growth theory for firms that can invest into information about their markets, value at risk analysis, optimal strategies for a multiproduct monopolist as well that of inheritance models in mathematical evolutionary theory (see Ibragimov, 2004a, b, c, d, 2005).

1.3 Organization of the paper

The paper is organized as follows. Section 2 introduces the classes of distributions considered throughout the paper and discusses their structure and the main properties. Section 3 presents the main results of the paper on efficiency properties of linear estimators for convolutions of α -symmetric distributions. Section 4 contains the proofs of the results obtained.

2 Notations and distributional assumptions

A r.v. X with density $f : \mathbf{R} \rightarrow \mathbf{R}$ and the convex distribution support $\Omega = \{x \in \mathbf{R} : f(x) > 0\}$ is said to be log-concavely distributed if $\log f(x)$ is concave in $x \in \Omega$, that is, if for all $x_1, x_2 \in \Omega$, and any $\lambda \in [0, 1]$, $f(\lambda x_1 + (1 - \lambda)x_2) \geq (f(x_1))^\lambda (f(x_2))^{1-\lambda}$ (see An, 1998). A distribution is called log-concave if its density f satisfies the above inequalities. Examples of log-concave distributions include the normal distribution, the uniform density, the exponential density, the Gamma distribution $\Gamma(\alpha, \beta)$ with the shape parameter $\alpha \geq 1$, the Beta distribution $\mathcal{B}(a, b)$ with $a \geq 1$ and $b \geq 1$; the Weibull distribution $\mathcal{W}(\gamma, \alpha)$ with the shape parameter $\alpha \geq 1$.

If a r.v. X is log-concavely distributed, then its density has at most an exponential tail, that is, $f(x) = O(\exp(-\lambda x))$ for some $\lambda > 0$, as $x \rightarrow \infty$ and all the power moments $E|X|^\gamma$, $\gamma > 0$, of the r.v. exist (see Corollary 1 in An, 1998). The reader is referred to Karlin (1968), Marshall and Olkin (1979) and An (1998) for a survey of many other properties of log-concave distributions.

For $0 < \alpha \leq 2$, $\sigma > 0$, $\beta \in [-1, 1]$ and $\mu \in \mathbf{R}$, we denote by $S_\alpha(\sigma, \beta, \mu)$ the stable distribution with the characteristic exponent (index of stability) α , the scale parameter σ , the symmetry index (skewness parameter) β and the location parameter μ . That is, $S_\alpha(\sigma, \beta, \mu)$ is the distribution of a r.v. X with the characteristic function

$$E(e^{ixX}) = \begin{cases} \exp\{i\mu x - \sigma^\alpha |x|^\alpha (1 - i\beta \operatorname{sign}(x) \tan(\pi\alpha/2))\}, & \alpha \neq 1, \\ \exp\{i\mu x - \sigma|x|(1 + (2/\pi)i\beta \operatorname{sign}(x) \ln|x|)\}, & \alpha = 1, \end{cases}$$

$x \in \mathbf{R}$, where $i^2 = -1$ and $\operatorname{sign}(x)$ is the sign of x defined by $\operatorname{sign}(x) = 1$ if $x > 0$, $\operatorname{sign}(0) = 0$ and $\operatorname{sign}(x) = -1$ otherwise. In what follows, we write $X \sim S_\alpha(\sigma, \beta, \mu)$, if the r.v. X has the stable distribution $S_\alpha(\sigma, \beta, \mu)$ and write $X \sim \mathcal{LC}$ if the distribution of X is symmetric and log-concave (\mathcal{LC} stands for ‘‘log-concave’’).

A closed form expression for the density $f(x)$ of the distribution $S_\alpha(\sigma, \beta, \mu)$ is available in the following cases (and

only in those cases): $\alpha = 2$ (Gaussian distributions); $\alpha = 1$ and $\beta = 0$ (Cauchy distributions with densities $f(x) = \sigma/(\pi(\sigma^2+(x-\mu)^2))$); $\alpha = 1/2$ and $\beta \pm 1$ (Lévy distributions that have densities $f(x) = (\sigma/(2\pi))^{1/2} \exp(-\sigma/(2x))x^{-3/2}$, $x \geq 0$; $f(x) = 0$, $x < 0$, where $\sigma > 0$, and their shifted versions). Degenerate distributions correspond to the limiting case $\alpha = 0$.

The index of stability α characterizes the heaviness (the rate of decay) of the tails of stable distributions $S_\alpha(\sigma, \beta, \mu)$. The distribution of a stable r.v. $X \sim S_\alpha(\sigma, \beta, \mu)$ with $\alpha \in (0, 2)$ obeys power law $P(|X| > x) \sim x^{-\alpha}$ and thus the p -th absolute moments $E|X|^p$ of X are finite if $p < \alpha$ and are infinite otherwise. The symmetry index β characterizes the skewness of the distribution. The stable distributions with $\beta = 0$ are symmetric about the location parameter μ . The stable distributions with $\beta = \pm 1$ and $\alpha \in (0, 1)$ (and only they) are one-sided, the support of these distributions is the semi-axis $[\mu, \infty)$ for $\beta = 1$ and is $(-\infty, \mu]$ (in particular, the Lévy distribution with $\mu = 0$ is concentrated on the positive semi-axis for $\beta = 1$ and on the negative semi-axis for $\beta = -1$). In the case $\alpha > 1$ the location parameter μ is the mean of the distribution $S_\alpha(\sigma, \beta, \mu)$. The scale parameter σ is a generalization of the concept of standard deviation; it coincides with the latter in the special case of Gaussian distributions ($\alpha = 2$).

Distributions $S_\alpha(\sigma, \beta, \mu)$ with $\mu = 0$ for $\alpha \neq 1$ and $\beta \neq 0$ for $\alpha = 1$ are called strictly stable. If $X_i \sim S_\alpha(\sigma, \beta, \mu)$, $\alpha \in (0, 2]$, are i.i.d. strictly stable r.v.'s, then, for all $a_i \geq 0$, $i = 1, \dots, n$, such that $\sum_{i=1}^n a_i \neq 0$, one has

$$\sum_{i=1}^n a_i X_i / \left(\sum_{i=1}^n a_i^\alpha \right)^{1/\alpha} \sim X_1. \quad (2.1)$$

For a detailed review of properties of stable distributions the reader is referred to, e.g., the monographs by Zolotarev (1986) and Uchaikin and Zolotarev (1999).

According to the definition introduced by Cambanis, Keener and Simons (1983), an n -dimensional distribution is called α -symmetric if its characteristic function (c.f.) can be written as $\phi((\sum_{i=1}^n |t_i|^\alpha)^{1/\alpha})$, where $\phi: \mathbf{R}_+ \rightarrow \mathbf{R}$ is a continuous function (with $\phi(0) = 1$) and $\alpha > 0$. An important property of α -symmetric distributions is that, similar to strictly stable laws, they satisfy property (2.1). The number α is called the index and the function ϕ is called the c.f. generator of the α -symmetric distribution. The class of α -symmetric distributions contains, as a subclass, spherical distributions corresponding to the case $\alpha = 2$ (see Fang, Kotz and Ng, 1990, p. 184). Spherical distributions, in turn, include such examples as Kotz type, multinormal, multivariate t and multivariate spherically symmetric α -stable distributions (Fang et. al., 1990, Ch. 3). Spherically symmetric stable distributions have characteristic functions $\exp[-\lambda(\sum_{i=1}^n t_i^2)^{\gamma/2}]$, $0 < \gamma \leq 2$, and are, thus, examples of α -symmetric distributions with $\alpha = 2$ and the c.f. generator $\phi(x) = \exp(-x^\gamma)$.

For any $0 < \alpha \leq 2$, the class of α -symmetric distributions includes distributions of risks X_1, \dots, X_n that have the common factor representation

$$(X_1, \dots, X_n) = (ZY_1, \dots, ZY_n), \quad (2.2)$$

where $Y_i \sim S_\alpha(\sigma, 0, 0)$ are i.i.d. symmetric stable r.v.'s with $\sigma > 0$ and the index of stability α and $Z \geq 0$ is a nonnegative r.v. independent of Y_i 's (see Bretagnolle, Dacuhna-Castelle and Krivine, 1966, and Fang et. al., 1990, p. 197). Although the dependence structure in model (2.2) alone is restrictive, convolutions of such vectors provide a natural framework for modeling of random environments with different common shocks Z , such as macroeconomic or political ones, that affect all risks X_i (see Andrews, 2003). In the case $Z = 1$ (a.s.), model (2.2) represents vectors with i.i.d. symmetric stable components that have c.f.'s $\exp[-\lambda \sum_{i=1}^n |t_i|^\alpha]$ which are particular cases of c.f.'s of α -symmetric distributions with the generator $\phi(x) = \exp(-\lambda x^\alpha)$.

According to the results in Bretagnolle et. al. (1966) and Kuritsyn and Shestakov (1984), the function $\exp(-(|t_1|^\alpha + |t_2|^\alpha)^{1/\alpha})$ is a c.f. of two α -symmetric r.v.'s for all $\alpha \geq 1$ (the generator of the function is $\phi(u) = \exp(-u)$). Zastavnyi (1993) demonstrates that the class of more than two α -symmetric r.v.'s with $\alpha > 2$ consists of degenerate variables (so that their c.f. generator $\phi(u) = 1$). For further review of properties and examples of α -symmetric distributions the reader is referred to Fang et. al. (1990, Ch. 7) and Gneiting (1998).

Convolutions of α -symmetric distributions are symmetric and unimodal. These convolutions also exhibit both heavy-tailedness in marginals and dependence among them. Both the classes of convolutions of α -symmetric distributions with $\alpha < 1$ and those with $\alpha > 1$ can be used to model heavy-tailedness of an arbitrary order in marginals. For instance, the class of convolutions of models (2.2) with $\alpha < 1$ has extremely heavy-tailed marginal distributions with infinite means. On the other hand, convolutions of such models with $1 < \alpha \leq 2$ can have marginals with power moments finite up to a certain positive order (or finite exponential moments) depending on the choice of the r.v.'s Z . For instance, convolutions of models (2.2) with $1 < \alpha < 2$ and $E|Z| < \infty$ have finite means but infinite variances, however, marginals of such convolutions have infinite means if the r.v.'s Z satisfy $E|Z| = \infty$. Moments $E|ZY_i|^p$, $p > 0$, of marginals in models (2.2) with $\alpha = 2$ (that correspond to Gaussian r.v.'s Y_i) are finite if and only if $E|Z|^p < \infty$. In particular, all marginal power moments in models (2.2) with $\alpha = 2$ are finite if $E|Z|^p < \infty$ for all $p > 0$. Similarly, marginals of spherically symmetric (that is, 2-symmetric) distributions range from extremely heavy-tailed to extreme lighted-tailed ones. For example, marginal moments of spherically symmetric α -stable distributions with c.f.'s $\exp[-\lambda(\sum_{i=1}^n t_i^2)^{\gamma/2}]$, $0 < \gamma < 2$, are finite if and only if their order is less than γ . Marginal moments of a multivariate t -distribution with k degrees of freedom which is an example of a spherical distribution are finite if and only if the order of the moments is less than k .

Let Φ stand for the class of c.f. generators ϕ such that $\phi(0) = 1$, $\lim_{t \rightarrow \infty} \phi(t) = 0$, and the function $\phi'(t)$ is concave. In what follows, we consider the following distributional assumptions (A1)-(A4) and (B1)-(B4).

Let $r \in (0, 2)$.

(A1) The random vector $(X_1 - \mu, \dots, X_n - \mu)$ is a sum of i.i.d. random vectors (Y_{1j}, \dots, Y_{nj}) , $j = 1, \dots, k$, where (Y_{1j}, \dots, Y_{nj}) has an absolutely continuous α -symmetric distribution with the c.f. generator $\phi_j \in \Phi$ and the index $\alpha_j \in (r, 2]$;

(A2) The random vector $(X_1 - \mu, \dots, X_n - \mu)$ is a sum of i.i.d. random vectors $(Y_{1j}, \dots, Y_{nj}) = (Z_j V_{1j}, \dots, Z_j V_{nj})$, $j = 1, \dots, k$, where $V_{ij} \sim S_{\alpha_j}(\sigma_j, 0, 0)$, $i = 1, \dots, n$, $j = 1, \dots, k$, with $\sigma_j > 0$ and $\alpha_j \in (r, 2]$ and Z_j are positive absolutely continuous r.v.'s independent of V_{ij} .

(A3) The random vector $(X_1 - \mu, \dots, X_n - \mu)$ has an α -symmetric distribution with a continuous c.f. generator $\phi: \mathbf{R}_+ \rightarrow \mathbf{R}$ and the index $\alpha \in (r, 2]$.

(A4) $(X_1, \dots, X_n) = (ZV_1, \dots, ZV_n)$, where V_i , $i = 1, \dots, n$, are i.i.d. r.v.'s such that $V_1 \sim S_\alpha(\sigma, \beta, \mu)$ for some $\sigma > 0$, $\beta \in [-1, 1]$, and $\alpha \in (r, 2]$, with $\beta = 0$ for $\alpha = 1$, and Z is a positive r.v. independent of V_i 's.

We will also need the following assumption (A2') which is more general than assumption (A2) with $r = 1$.

(A2') The random vector $(X_1 - \mu, \dots, X_n - \mu)$ is a sum of i.i.d. random vectors $(Y_{1j}, \dots, Y_{nj}) = (Z_j V_{1j}, \dots, Z_j V_{nj})$, $j = 1, \dots, k$, where $V_{ij} \sim \mathcal{LC}$ or $V_{ij} \sim S_{\alpha_j}(\sigma_j, 0, 0)$, $i = 1, \dots, n$, $j = 1, \dots, k$, with $\sigma_j > 0$ and $\alpha_j \in (1, 2]$ and Z_j are positive absolutely continuous r.v.'s independent of V_{ij} .

The following distributional assumptions (B1)-(B4) involve conditions which are the opposite of those in (A1)-

(A4). Let $r \in (0, 2]$.

(B1) The random vector $(X_1 - \mu, \dots, X_n - \mu)$ is a sum of i.i.d. random vectors (Y_{1j}, \dots, Y_{nj}) , $j = 1, \dots, k$, where (Y_{1j}, \dots, Y_{nj}) has an absolutely continuous α -symmetric distribution with the c.f. generator $\phi_j \in \Phi$ and the index $\alpha_j \in (0, r)$;

(B2) The random vector $(X_1 - \mu, \dots, X_n - \mu)$ is a sum of i.i.d. random vectors $(Y_{1j}, \dots, Y_{nj}) = (Z_j V_{1j}, \dots, Z_j V_{nj})$, $j = 1, \dots, k$, where $V_{ij} \sim S_{\alpha_j}(\sigma_j, 0, 0)$, $i = 1, \dots, n$, $j = 1, \dots, k$, with $\sigma_j > 0$ and $\alpha_j \in (0, r)$ and Z_j are positive absolutely continuous r.v.'s independent of V_{ij} .

(B3) The random vector $(X_1 - \mu, \dots, X_n - \mu)$ has an α -symmetric distribution with a continuous c.f. generator $\phi : \mathbf{R}_+ \rightarrow \mathbf{R}$ and the index $\alpha \in (0, r)$.

(B4) $(X_1, \dots, X_n) = (ZV_1, \dots, ZV_n)$, where V_i , $i = 1, \dots, n$, are i.i.d. r.v.'s such that $V_1 \sim S_\alpha(\sigma, \beta, \mu)$ for some $\sigma > 0$, $\beta \in [-1, 1]$ and $\alpha \in (0, r)$, with $\beta = 0$ for $\alpha = 1$, and Z is a positive r.v. independent of V_i 's.

It is easy to see that if X_1, \dots, X_n satisfy (2.1) with $\alpha < 1$, then $E|\bar{X}_n| > E|X_1|$ that, evidently, cannot hold in the case $E|X_1| < \infty$. Consequently, first moments of such r.v.'s are infinite. It is not difficult to see that this implies that marginal first moments of r.v.'s X_1, \dots, X_n satisfying one of assumptions (B1)-(B4) with $r \leq 1$ are infinite.

The indices of stability α_j and the scale parameters σ_j in assumptions (A2) and (B2) are different among the vectors (Y_{1j}, \dots, Y_{nj}) . A linear combination of independent stable r.v.'s with the *same* characteristic exponent α also has a stable distribution with the same α . However, in general, this does not hold true in the case of convolutions of stable distributions with *different* indices of stability. Therefore, the class of random vectors (X_1, \dots, X_n) satisfying assumption (A2) (resp., assumption (B2)) with $Z_j = Z$, where Z is a positive absolutely continuous r.v. independent of symmetric stable r.v.'s V_{ij} , is wider than the class of random vectors (X_1, \dots, X_n) satisfying assumption (A4) (resp., assumption (B4)) with $\beta = 0$.

3 Main results: efficiency properties of linear estimators under heavy-tailedness and dependence

In what follows, for a vector $c \in \mathbf{R}^n$, we denote by $c_{[1]} \geq \dots \geq c_{[n]}$ its components in decreasing order. A vector $a \in \mathbf{R}^n$ is said to be majorized by a vector $b \in \mathbf{R}^n$, written $a \prec b$, if $\sum_{i=1}^k a_{[i]} \leq \sum_{i=1}^k b_{[i]}$, $k = 1, \dots, n-1$, and $\sum_{i=1}^n a_{[i]} = \sum_{i=1}^n b_{[i]}$. The relation $a \prec b$ implies that the components of the vector a are more diverse than those of b (see Marshall and Olkin, 1979). In this context, it is easy to see that the following relations hold:

$$\left(\sum_{i=1}^n a_i/n, \dots, \sum_{i=1}^n a_i/n \right) \prec (a_1, \dots, a_n) \prec \left(\sum_{i=1}^n a_i, 0, \dots, 0 \right), \quad a \in \mathbf{R}_+^n, \quad (3.1)$$

for all $a \in \mathbf{R}_+^n$. In particular,

$$(1/(n+1), \dots, 1/(n+1), 1/(n+1)) \prec (1/n, \dots, 1/n, 0), \quad n \geq 1. \quad (3.2)$$

A function $\phi : A \rightarrow \mathbf{R}$ defined on $A \subseteq \mathbf{R}^n$ is called *Schur-convex* (resp., *Schur-concave*) on A if $(a \prec b) \implies (\phi(a) \leq \phi(b))$ (resp. $(a \prec b) \implies (\phi(a) \geq \phi(b))$) for all $a, b \in A$. If, in addition, $\phi(a) < \phi(b)$ (resp., $\phi(a) > \phi(b)$) whenever $a \prec b$ and a is not a permutation of b , then ϕ is said to be *strictly Schur-convex* (resp., *strictly Schur-concave*) on A .

In what follows, given a random sample X_1, \dots, X_n from a population with center μ , and weights $a = (a_1, \dots, a_n) \in \mathbf{R}_+^n$, we denote by $\hat{\theta}_n(a)$ the linear estimator $\hat{\theta}_n(a) = \sum_{i=1}^n a_i X_i$ and by $\psi(a, \epsilon)$, $\epsilon > 0$, its tail probability $\psi(a, \epsilon) = P(|\hat{\theta}_n(a) - \mu| > \epsilon)$. We also denote by \mathcal{I}_n the simplex $\mathcal{I}_n = \{a = (a_1, \dots, a_n) \in \mathbf{R}_+^n : \sum_{i=1}^n a_i = 1\}$.

Theorem 3.1 concerns efficiency comparisons for linear estimators in the case of convolutions of α -symmetric distributions with $\alpha > 1$. It shows that, for convolutions of α -symmetric distributions with $\alpha > 1$, the sample mean is the best linear unbiased estimator of the population mean in the sense of P-efficiency. In addition, according to the theorem, the sample mean exhibits monotone consistency under such distributional assumptions.

Theorem 3.1 *Let $\mu \in \mathbf{R}$. Suppose that, for $n \geq 1$, the random samples X_1, \dots, X_n satisfy assumption (A2') or one of assumptions (A1), (A3) or (A4) with $r = 1$. Then the following conclusions hold.*

(i) *Let $a, b \in \mathcal{I}_n$. The linear estimator $\hat{\theta}_n(a)$ is P-more efficient than $\hat{\theta}_n(b)$ if $a \prec b$ and a is not a permutation of b (equivalently, $\psi(a, \epsilon)$ is strictly Schur-convex in $a = (a_1, \dots, a_n) \in \mathbf{R}_+^n$ for all $\epsilon > 0$).*

(ii) *The sample mean $\bar{X}_n = (1/n) \sum_{i=1}^n X_i$ is P-more efficient than any other linear unbiased estimator $\hat{\theta}_n(a) = \sum_{i=1}^n a_i X_i$, $a \in \mathcal{I}_n$. In particular, \bar{X}_n exhibits monotone consistency for μ and $P(|\bar{X}_n - \mu| > \epsilon)$ converges to zero strictly monotonically in n for all $\epsilon > 0$.*

According to the following theorem, the conclusions of Theorem 3.1 are reversed for convolutions of α -symmetric distributions with $\alpha < 1$. In this case, peakedness of the sample mean about the population center decreases with the sample size. In addition, under the above distributional assumptions, P-efficiency of the sample mean is smallest among all linear estimators $\hat{\theta}_n(a)$ with $a \in \mathcal{I}_n$.

Theorem 3.2 *Let $\mu \in \mathbf{R}$. Suppose that, for $n \geq 1$, the random samples X_1, \dots, X_n satisfy one of assumptions (B1)-(B4) with $r = 1$. Then the following conclusions hold.*

(i) *Let $a, b \in \mathcal{I}_n$. The linear estimator $\hat{\theta}_n(a)$ is P-less efficient than $\hat{\theta}_n(b)$ if $a \prec b$ and a is not a permutation of b (equivalently, $\psi(a, \epsilon)$ is strictly Schur-concave in $a = (a_1, \dots, a_n) \in \mathbf{R}_+^n$ for all $\epsilon > 0$).*

(ii) *The sample mean $\bar{X}_n = (1/n) \sum_{i=1}^n X_i$ is P-less efficient than any other linear estimator $\hat{\theta}_n(a) = \sum_{i=1}^n a_i X_i$ with $a \in \mathcal{I}_n$. In particular, P-efficiency of \bar{X}_n decreases with n , that is, $P(|\bar{X}_{n+1} - \mu| > \epsilon) > P(|\bar{X}_n - \mu| > \epsilon) > P(|X_1 - \mu| > \epsilon)$ for all $n \geq 1$ and all $\epsilon > 0$.*

The following Theorem 3.3 shows that efficiency comparisons for linear estimators for population with distributions satisfying one of assumptions (A1)-(A4) are of the same type as in Theorem 3.1 with respect to the comparisons between the powers of the components of the vectors of weights of the combinations. Theorem 3.3 also provides concentration inequalities for linear estimators in the case of such distributions that refine and complement the efficiency and peakedness comparisons implied by Theorem 3.1.

Theorem 3.3 *Suppose that, for $n \geq 1$, the random samples X_1, \dots, X_n satisfy one of assumptions (A1)-(A4) with $r \in (0, 2)$. Then the following conclusions hold.*

(i) *Let $\mu = 0$ and $a, b \in \mathbf{R}_+^n$. Then $\hat{\theta}_n(a)$ is strictly more peaked than $\hat{\theta}_n(b)$ if $(a_1^r, \dots, a_n^r) \prec (b_1^r, \dots, b_n^r)$ and (a_1^r, \dots, a_n^r) is not a permutation of (b_1^r, \dots, b_n^r) (equivalently, $\psi(a, \epsilon)$ is strictly Schur-convex in $(a_1^r, \dots, a_n^r) \in \mathbf{R}_+^n$ for all $\epsilon > 0$).*

(ii) Let $\mu \in \mathbf{R}$. The linear estimators $\hat{\theta}_n(a) = \sum_{i=1}^n a_i X_i$, $a \in \mathcal{I}_n$, satisfy the following concentration inequalities for all $\epsilon > 0$: $P\left(|\bar{X}_n - \mu| > n^{1/r-1}\epsilon / \left(\sum_{i=1}^n a_i^r\right)^{1/r}\right) \leq P(|\hat{\theta}_n(a) - \mu| > \epsilon) \leq P\left(|X_1 - \mu| > \epsilon / \left(\sum_{i=1}^n a_i^r\right)^{1/r}\right)$, with strict right-hand side inequality if $a = (a_1, a_2, \dots, a_n)$ is not a permutation of $(1, 0, \dots, 0)$ and strict left-hand side inequality if $a \neq (1/n, 1/n, \dots, 1/n)$.

As follows from Theorem 3.4 below, the efficiency properties of linear estimators in Theorem 3.3 are reversed in the case of populations with distributions satisfying one of assumptions (B1)-(B4), in particular, for convolutions of α -symmetric distributions with $\alpha < r$. The concentration inequalities in Theorem 3.4 refine and complement the efficiency orderings for linear estimators given by Theorem 3.2.

Theorem 3.4 Suppose that, for $n \geq 1$, the random samples X_1, \dots, X_n satisfy one of assumptions (B1)-(B4) with $r \in (0, 2]$. Then the following conclusions hold.

(i) Let $\mu = 0$ and $a, b \in \mathbf{R}_+^n$. Then $\hat{\theta}_n(a)$ is strictly more peaked than $\hat{\theta}_n(b)$ if $(a_1^r, \dots, a_n^r) \prec (b_1^r, \dots, b_n^r)$ and (a_1^r, \dots, a_n^r) is not a permutation of (b_1^r, \dots, b_n^r) (equivalently, $\psi(a, \epsilon)$ is strictly Schur-concave in $(a_1^r, \dots, a_n^r) \in \mathbf{R}_+^n$ for all $\epsilon > 0$).

(ii) Let $\mu \in \mathbf{R}$. The linear estimators $\hat{\theta}_n(a) = \sum_{i=1}^n a_i X_i$, $a \in \mathcal{I}_n$, satisfy the following concentration inequalities for all $\epsilon > 0$: $P\left(|X_1 - \mu| > \epsilon / \left(\sum_{i=1}^n a_i^r\right)^{1/r}\right) \leq P(|\hat{\theta}_n(a) - \mu| > \epsilon) \leq P\left(|\bar{X}_n - \mu| > n^{1/r-1}\epsilon / \left(\sum_{i=1}^n a_i^r\right)^{1/r}\right)$, with strict left-hand side inequality if $a = (a_1, \dots, a_n)$ is not a permutation of $(1, 0, \dots, 0)$ and strict right-hand side inequality if $a \neq (1/n, 1/n, \dots, 1/n)$.

The following Proposition 3.1 provides analogues of the results in this section for linear estimators $\hat{\tau}_n(a) = \sum_{i=1}^n a_{[i]} W_i$ for not necessarily identically distributed r.v.'s W_i (a certain ordering in the components of the vector a is necessary for the extensions of the majorization results in this section to the case of non-identically distributed r.v.'s W_i since Schur-convexity and Schur-concavity of a function $f(a)$ in a imply its symmetry in the components of a). Let $\sigma_1, \dots, \sigma_n > 0$ and $\mu_1, \dots, \mu_n < 0$.

Proposition 3.1 Let $W_i \sim S_\alpha(\sigma_i, \beta, \mu_i)$, $\alpha \in (0, 2]$, $\beta \in [-1, 1]$, $\beta = 0$ for $\alpha = 1$, be independent non-identically distributed stable r.v.'s. Then the following conclusions hold.

(i) The function $\vartheta(a, \epsilon) = P(\hat{\tau}_n(a) > \epsilon)$ is strictly Schur-concave in $a = (a_1, \dots, a_n) \in \mathbf{R}^n$ if $\alpha > 1$, $\sigma_1 \geq \dots \geq \sigma_n > 0$ and $\mu_1 \leq \dots \leq \mu_n \leq 0$ and is strictly Schur-convex in $a = (a_1, \dots, a_n) \in \mathbf{R}^n$ if $\alpha < 1$, $\sigma_n \geq \dots \geq \sigma_1 > 0$ and $0 \geq \mu_1 \geq \dots \geq \mu_n$.

(ii) Let $\mu_i = \mu$, $i = 1, \dots, n$. Theorems 3.1 and 3.3 hold (in the same range of parameters r and α) for $\hat{\tau}_n(a)$ if $\sigma_1 \geq \dots \geq \sigma_n > 0$. Theorems 3.2 and 3.4 hold for $\hat{\tau}_n(a)$ if $\sigma_n \geq \dots \geq \sigma_1 > 0$.

The following corollary provides analogues of the results in the paper for convolutions of stable distributions with different location (and scale) parameters.

Corollary 3.1 Let X_1, \dots, X_n be i.i.d. r.v.'s with a common distribution which is a convolution of stable distributions $S_{\alpha_j}(\sigma_j, 0, \mu_j)$, $j = 1, \dots, k$, with different scale parameters $\sigma_j > 0$ and different location parameters $\mu_j \in \mathbf{R}$ such that $\sum_{j=1}^k \mu_j < 0$: $X_i = \sum_{j=1}^k Y_{ij}$, where $Y_{ij} \sim S_{\alpha_j}(\sigma_j, 0, \mu_j)$, are independent stable r.v.'s. Then the function $\tilde{\psi}(a, \epsilon) = P\left(\sum_{i=1}^n a_i X_i > \epsilon\right)$ is strictly Schur-convex in $a = (a_1, \dots, a_n) \in \mathbf{R}_+^n$ for all $\epsilon > 0$.

Remark 3.1 *Theorem 3.1 provides generalizations of the results in Proschan (1965) who showed that the tail probabilities $\psi(a, \epsilon) = P(|\sum_{i=1}^n a_i X_i - \mu| > \epsilon)$ are Schur-convex in $a = (a_1, \dots, a_n) \in \mathbf{R}_+^n$ for all $\epsilon > 0$ for random samples X_1, \dots, X_n from symmetric log-concavely distributed populations ($X_1 - \mu \sim \mathcal{LC}$).³ Proschan's (1965) results and their extensions have been applied to the analysis of many problems in statistics, econometrics, economic theory, mathematical evolutionary theory and other fields (see the review in Ibragimov, 2004a, b, c, d, 2005, and references therein). A number of papers in probability and statistics have focused on extension of Proschan's results (see, among others, the review in Tong, 1994, Jensen, 1997, and Ma, 1998). However, that in all the studies that dealt with generalizations of the results, the majorization properties of the tail probabilities were of the same type as in Proschan (1965). Namely, the results gave extensions of Proschan's results concerning Schur-convexity of the tail probabilities $\psi(a, \epsilon)$, $\epsilon > 0$, to classes of r.v.'s more general than those considered in Proschan (1965). We are not aware of any general results concerning Schur-concavity of the tail probabilities $\psi(a, \epsilon)$, $\epsilon > 0$. Such general results are provided by Theorems 3.2 and 3.4.*

Remark 3.2 *It is well-known that if r.v.'s X and Y are such that $P(X > x) \leq P(Y > x)$ for all $x \in \mathbf{R}$, then $EU(X) \leq EU(Y)$ for all increasing functions $U : \mathbf{R} \rightarrow \mathbf{R}$ for which the expectations exist (see, e.g., Shaked and Shanthikumar, 1994, pp. 3-4). This fact and Theorems 3.1-3.4 imply corresponding results concerning majorization properties of expectations of loss functions of linear estimators under heavy-tailedness. For instance, we get that if $U : \mathbf{R}_+ \rightarrow \mathbf{R}$ is an increasing function, then, assuming existence of the expectations, the function $\varphi(a) = EU(|\hat{\theta}_n(a) - \mu|)$, $a \in \mathbf{R}_+^n$ is Schur-convex in (a_1^r, \dots, a_n^r) under the assumptions of Theorem 3.3 and is Schur-concave in (a_1^r, \dots, a_n^r) under the assumptions of Theorem 3.4. We also get that the function $\varphi(a)$, $a \in \mathbf{R}_+^n$ is Schur-concave in (a_1^2, \dots, a_n^2) under assumptions (B1)-(B4) with $r = 2$. These results complement those in Efron (1969) and Eaton (1970) (see also Marshall and Olkin, 1979, pp. 361-365) who studied classes of functions $U : \mathbf{R} \rightarrow \mathbf{R}$ and r.v.'s X_1, \dots, X_n for which Schur-concavity of $\varphi(a)$, $a \in \mathbf{R}_+^n$ in (a_1^2, \dots, a_n^2) holds. Further, we obtain that $\varphi(a)$ is Schur-convex in $a \in \mathbf{R}_+^n$ under the assumptions of Theorem 3.1 and is Schur-concave in $a \in \mathbf{R}_+^n$ under the assumptions of Theorem 3.2. Since $E|X_i| = \infty$ for r.v.'s X_1, \dots, X_n satisfying the assumptions of Theorem 3.2 (see Section 2), we get that, in the case of such r.v.'s and increasing convex loss functions $U : \mathbf{R}_+ \rightarrow \mathbf{R}$, the expectations $EU(|\hat{\theta}_n(a) - \mu|)$ are infinite for all $a \in \mathbf{R}_+^n$, $\sum_{i=1}^n a_i \neq 0$. Therefore, the last result does not contradict the well-known fact that (see Marshall and Olkin, 1979, p. 361) the function $Ef(\sum_{i=1}^n a_i Y_i)$ is Schur-convex in $(a_1, \dots, a_n) \in \mathbf{R}$ for all i.i.d. r.v.'s Y_1, \dots, Y_n and convex functions $f : \mathbf{R} \rightarrow \mathbf{R}$ as it might seem on the first sight.*

Remark 3.3 *Similar to the proof of Proposition 3.1, one can also obtain analogues of the results in the present section in the case of distributions with dependent and not necessarily identically distributed marginals, including convolutions of shifted and scaled α -symmetric distributions.*

4 Proofs

In the proofs below, we provide the complete argument for the main majorizations results that provide a reversal of those available in the literature, namely for Theorems 3.2 and 3.4. The proof of Theorem 3.3 that gives the results on Schur-convexity of the tail probabilities of linear combinations of r.v.'s follows the same lines as that of Theorem

³Proschan (1965) notes that similar majorization orderings also hold for (two-fold) convolutions of log-concave distributions with symmetric Cauchy distributions and shows that peakedness comparisons implied by them are reversed for $n = 2^k$, vectors $a = (1/n, 1/n, \dots, 1/n) \in \mathbf{R}^n$ with identical components and certain transforms of symmetric Cauchy r.v.'s.

3.4, with respective changes in the signs of inequalities. We also provide the complete proof of Theorem 3.2 in the case of assumption (A2') since, in this case, it is not implied by Theorem 3.3 alone, but needs to combine the results in that theorem with those for log-concave distributions in Proschan (1965).

Proof of Theorems 3.3 and 3.4. Let $r, \alpha \in (0, 2], \sigma > 0, \beta \in [-1, 1], \beta = 0$ for $\alpha = 1$, and let $a = (a_1, \dots, a_n) \in \mathbf{R}_+^n$ and $b = (b_1, \dots, b_n) \in \mathbf{R}_+^n$ be such that $(a_1^r, \dots, a_n^r) \prec (b_1^r, \dots, b_n^r)$ and (a_1^r, \dots, a_n^r) is not a permutation of (b_1^r, \dots, b_n^r) (clearly, $\sum_{i=1}^n a_i \neq 0$ and $\sum_{i=1}^n b_i \neq 0$). Let (X_1, \dots, X_n) be a random vector satisfying one of the assumptions (A3), (A4), (B3) or (B4) with $\mu = 0$. As follows from the discussion in Section 2, property (2.1) holds for $X_i, i = 1, \dots, n$. Consequently, if $c = (c_1, \dots, c_n) \in \mathbf{R}_+^n, \sum_{i=1}^n c_i \neq 0$, and $\epsilon > 0$, then

$$\psi(c, \epsilon) = P\left(|X_1| > \epsilon / \left(\sum_{i=1}^n c_i^\alpha\right)^{1/\alpha}\right). \quad (4.3)$$

According to Proposition 3.C.1.a in Marshall and Olkin (1979), the function $\phi(c_1, \dots, c_n) = \sum_{i=1}^n c_i^\alpha$ is strictly Schur-convex in $(c_1, \dots, c_n) \in \mathbf{R}_+^n$ if $\alpha > 1$ and is strictly Schur-concave in $(c_1, \dots, c_n) \in \mathbf{R}_+^n$ if $\alpha < 1$. Therefore, we have $\sum_{i=1}^n a_i^\alpha = \sum_{i=1}^n (a_i^r)^{\alpha/r} < \sum_{i=1}^n (b_i^r)^{\alpha/r} = \sum_{i=1}^n b_i^\alpha$, if $\alpha/r > 1$ and $\sum_{i=1}^n b_i^\alpha = \sum_{i=1}^n (b_i^r)^{\alpha/r} < \sum_{i=1}^n (a_i^r)^{\alpha/r} = \sum_{i=1}^n a_i^\alpha$, if $\alpha/r < 1$. This, together with (4.3), implies that $\psi(a, \epsilon) < \psi(b, \epsilon)$ if either (A3) or (A4) is satisfied, and $\psi(a, \epsilon) > \psi(b, \epsilon)$ if either (B3) or (B4) is satisfied. Consequently, part (i) of Theorem 3.3 holds if (X_1, \dots, X_n) satisfies (A3) or (A4) and part (i) of Theorem 3.4 holds for (X_1, \dots, X_n) satisfying (B3) or (B4).

Let now $\mu \in \mathbf{R}$ and suppose that the random vector $(X_1 - \mu, \dots, X_n - \mu)$ is a sum of i.i.d. random vectors $(Y_{1j}, \dots, Y_{nj}), j = 1, \dots, k$, satisfying the assumptions in (B1) or (B2). By Theorem 3.4 for distributions satisfying (B3) or (B4), for $j = 1, \dots, k$, the r.v. $\sum_{i=1}^n b_i Y_{ij}$ is strictly more peaked than $\sum_{i=1}^n a_i Y_{ij}$, that is, for all $\epsilon > 0$ and all $j = 1, \dots, k$,

$$P\left(\left|\sum_{i=1}^n a_i Y_{ij}\right| > \epsilon\right) > P\left(\left|\sum_{i=1}^n b_i Y_{ij}\right| > \epsilon\right). \quad (4.4)$$

The r.v.'s $\sum_{i=1}^n a_i Y_{ij}, j = 1, \dots, k$, and $\sum_{i=1}^n b_i Y_{ij}, j = 1, \dots, k$, are symmetric and unimodal if one of the conditions (A1), (A2), (B1) or (B2) is satisfied. In the case of (A1) and (B1) this easily follows from a result due to R. Askey, see Theorem 4.1 in Gneiting, 1998. In the case of assumptions (A2) and (B2), symmetry and unimodality of $\sum_{i=1}^n a_i Y_{ij}, j = 1, \dots, k$, and $\sum_{i=1}^n b_i Y_{ij}$ follows from symmetry and unimodality of $\sum_{i=1}^n a_i V_{ij}$ and $\sum_{i=1}^n b_i V_{ij}$ implied by Theorem 2.7.6 in Zolotarev, 1986, p. 134, and Theorem 1.6 in Dharmadhikari and Joag-Dev, 1988, p. 13, the definition of unimodality and conditioning arguments.

From Lemma in Birnbaum (1948) and its proof it follows that if X_1, X_2 and Y_1, Y_2 are independent absolutely continuous symmetric unimodal r.v.'s such that, for $i = 1, 2, X_i$ is more peaked than Y_i , and one of the two peakedness comparisons is strict, then $X_1 + X_2$ is strictly more peaked than $Y_1 + Y_2$. This, together with (4.4) and symmetry and unimodality of $\sum_{i=1}^n a_i Y_{ij}$ and $\sum_{i=1}^n b_i Y_{ij}, j = 1, \dots, k$, imply, by induction on k (see also Theorem 1 in Birnbaum, 1948, and Theorem 2.C.3 in Dharmadhikari and Joag-Dev, 1988), that $\psi(a, \epsilon) = P\left(\left|\sum_{j=1}^k \sum_{i=1}^n a_i Y_{ij}\right| > \epsilon\right) > P\left(\left|\sum_{j=1}^k \sum_{i=1}^n b_i Y_{ij}\right| > \epsilon\right) = \psi(b, \epsilon)$ for $\epsilon > 0$. Therefore, part (i) of Theorem 3.4 holds if either (B1) or (B2) is satisfied. Part (i) of Theorem 3.3 for random vectors (X_1, \dots, X_n) satisfying one of conditions (A1) or (A2) with $\mu \in \mathbf{R}$ might be proven in a completely similar way, with the reversal of inequality signs in (4.4). Parts (ii) of Theorems 3.3 and 3.4 follow from their parts (i) and majorization comparisons (3.1). The proof is complete.

Proof of Theorems 3.1 and 3.2. Theorems 3.3 and 3.4 imply that part (i) of Theorem 3.2 holds if one of assumptions (B1)-(B4) with $r = 1$ is satisfied and part (i) of Theorem 3.1 holds if one of assumptions (A1)-(A4) with $r = 1$ is

satisfied. Let us prove that part (i) of Theorem 3.1 holds under assumption (A2'). Let vectors $a = (a_1, \dots, a_n) \in \mathcal{I}_n$ and $b = (b_1, \dots, b_n) \in \mathcal{I}_n$ be such that $a \prec b$ and a is not a permutation of b . Suppose that the vector of r.v.'s $(X_1 - \mu, \dots, X_n - \mu)$ is a sum of i.i.d. random vectors $(Y_{1j}, \dots, Y_{nj}) = (Z_j V_{1j}, \dots, Z_j V_{nj})$, $j = 1, \dots, k$, such that, for $i = 1, \dots, n$ and $j = 1, \dots, k$, $V_{ij} \sim \mathcal{LC}$ or $V_{ij} \sim S_{\alpha_j}(\sigma_j, 0, 0)$, where $\sigma_j > 0$ and $\alpha_j \in (1, 2]$, and Z_j are absolutely continuous positive r.v.'s independent of V_{ij} . From part (i) of Theorem 3.3 and the results in Proschan (1965) it follows that, for $j = 1, \dots, k$, the r.v. $\sum_{i=1}^n a_i Y_{ij}$ is strictly more peaked than $\sum_{i=1}^n b_i Y_{ij}$. Furthermore, similar to the proof of Theorems 3.3 and 3.4, from Theorem 2.7.6 in Zolotarev (1986, p. 134) and Theorems 1.6 and 1.10 in Dharmadhikari and Joag-Dev (1988, pp. 13 and 20), together with the definition of unimodality and conditioning arguments, it follows that the r.v.'s $\sum_{i=1}^n a_i V_{ij}$ and $\sum_{i=1}^n b_i V_{ij}$, $j = 0, 1, \dots, k$, are symmetric and unimodal. As in the proof of Theorems 3.3 and 3.4, by Lemma in Birnbaum (1948) and its proof and induction, this implies that $\sum_{i=1}^n a_i X_i = \sum_{j=1}^k \sum_{i=1}^n a_i V_{ij}$ is strictly more peaked than $\sum_{i=1}^n b_i X_i = \sum_{j=1}^k \sum_{i=1}^n b_i V_{ij}$. This completes the proof of part (i) of Theorem 3.1.

As easy to see, under assumptions of Theorem 3.1, the characteristic function $E \exp(it\bar{X}_n)$ of \bar{X}_n converges to $E \exp(it\mu)$ as $n \rightarrow \infty$ for all $t \in \mathbf{R}$, that is, \bar{X}_n is weakly consistent for μ . This, together with parts (i) of Theorems 3.1 and 3.2 and majorization comparisons (3.1) and (3.2) imply parts (ii) of the theorems.

Proof of Proposition 3.1 and Corollary 3.1. Under the assumptions of the Proposition 3.1, according to (2.1), $\sum_{i=1}^n a_{[i]} W_i \sim \sum_{i=1}^n a_{[i]} \mu_i + \left(\sum_{i=1}^n \sigma_i^\alpha a_{[i]}^\alpha \right)^{1/\alpha} Q_1$, where $Q_1 \sim S_\alpha(1, \beta, 0)$. Consequently, the function $\vartheta(a, \epsilon)$ in part (i) of the proposition satisfies $\vartheta(a, \epsilon) = P\left(Q_1 > \left(\epsilon - \sum_{i=1}^n a_{[i]} \mu_i\right) / \left(\sum_{i=1}^n \sigma_i^\alpha a_{[i]}^\alpha\right)^{1/\alpha}\right)$ and, if $\mu_i = 0$, then $P(|\hat{\tau}_n(a)| > \epsilon) = P\left(|Q_1| > \epsilon / \left(\sum_{i=1}^n \sigma_i^\alpha a_{[i]}^\alpha\right)^{1/\alpha}\right)$. By Theorem 3.A.4 in Marshall and Olkin (1979), the function $\chi_1(c_1, \dots, c_n) = \sum_{i=1}^n \sigma_i^\alpha c_{[i]}^\alpha$ is strictly Schur-convex in $(c_1, \dots, c_n) \in \mathbf{R}_+^n$ if $\alpha > 1$ and $\sigma_1 \geq \dots \geq \sigma_n > 0$ and is strictly Schur-concave in $(c_1, \dots, c_n) \in \mathbf{R}_+^n$ if $\alpha < 1$ and $\sigma_n \geq \dots \geq \sigma_1 \geq 0$. In addition, by the same theorem, $\chi_2(c_1, \dots, c_n) = \sum_{i=1}^n \mu_i c_{[i]}$ is Schur-convex in $(c_1, \dots, c_n) \in \mathbf{R}_+^n$ if $0 \geq \mu_1 \geq \dots \geq \mu_n$ and is Schur-concave in $(c_1, \dots, c_n) \in \mathbf{R}_+^n$ if $0 \geq \mu_n \geq \dots \geq \mu_1$. Similar to the proof of Theorems 3.1 and 3.2, the above implies that Proposition 3.1 holds. Since the sum $\sum_{i=1}^n a_i$ is fixed under majorization comparisons, the results in Theorems 3.1 and 3.2 imply that, under the assumptions of Proposition 3.1, the function $\tilde{\psi}(a, \epsilon) = P\left[\sum_{i=1}^n a_i \left(\sum_{j=1}^k (Y_{ij} - \mu_j)\right) > \epsilon - \left(\sum_{i=1}^n a_i\right) \left(\sum_{j=1}^k \mu_j\right)\right]$ is strictly Schur-convex in $a = (a_1, \dots, a_n) \in \mathbf{R}_+^n$ if $\alpha_j < 1$, $j = 1, \dots, k$, and is strictly Schur-concave in $a = (a_1, \dots, a_n) \in \mathbf{R}_+^n$ if $\alpha_j > 1$, $j = 1, \dots, k$. Consequently, Corollary 3.1 holds.

REFERENCES

- An, M. Y. (1998). Logconcavity versus logconvexity: a complete characterization. *Journal of Economic Theory* **80**, 350-369.
- Andrews, D. W. K. (2003). Cross-section regression with common shocks. *Cowles Foundation Discussion Paper No. 1428*, Yale University.
- Birnbaum, Z. W. (1948). On random variables with comparable peakedness. *Annals of Mathematical Statistics* **19**, 76-81.
- Bretagnolle, J., Dacunha-Castelle, D. and Krivine, J. L. (1966). Lois stables et espaces L^p . *Annales de l'Institut H. Poincaré. Section B. Calcul des Probabilités et Statistique*. **2**, 231-259.
- Cambanis, S., Keener, R. and Simons, G. (1983). On α -symmetric distributions. *Journal of Multivariate Analysis* **13**, 213-233.

- Dharmadhikari, S. W. and Joag-Dev, K. (1988). *Unimodality, convexity and applications*. Boston: Academic Press.
- Eaton, M. L. (1970). A note on symmetric Bernoulli random variables. *Annals of Mathematical Statistics* **41**, 1223-1226.
- Efron, B. (1969). Student's t -test under symmetry conditions. *Journal of the American Statistical Association* **64**, 1278-1302.
- Fang, K.-T., Kotz, S. and Ng, K. W. (1990). *Symmetric multivariate and related distributions*. Chapman and Hall, New York.
- Gneiting, T. (1998). On α -symmetric multivariate characteristic functions. *Journal of Multivariate Analysis* **64**, 131-147.
- Ibragimov, R. (2004a). A tale of two tails: peakedness properties in inheritance models of evolutionary theory. *Working paper*, Yale University.
- Ibragimov, R. (2004b). Demand-driven innovation and spatial competition over time under heavy-tailed signals. Under revision for the *Journal of Economic Theory*.
- Ibragimov, R. (2004c). Optimal bundling strategies for complements and substitutes with heavy-tailed valuations. *Working paper*, Yale University.
- Ibragimov, R. (2004d). Portfolio diversification and value at risk under thick-tailedness. *Working paper*, Yale University.
- Ibragimov, R. (2005). *New majorization theory in economics and martingale convergence results in econometrics*. Ph.D. dissertation, Yale University.
- Jensen, D. R. (1997). Peakedness of linear forms in ensembles and mixtures. *Statistics and Probability Letters* **35**, 277-282.
- Karlin, S. (1968). *Total positivity*. Stanford University Press.
- Kuritsyn, Yu. G. and Shestakov, A. V. (1984). On α -symmetric distributions. *Theory of Probability and Its Applications* **29**, 804-806.
- Ma, C. (1998). On peakedness of distributions of convex combinations. *Journal of Statistical Planning and Inference* **70**, 51-56.
- Marshall, A. W. and Olkin, I. (1979). *Inequalities: theory of majorization and its applications*. New York: Academic Press.
- Proschan, F. (1965). Peakedness of distributions of convex combinations. *Annals of Mathematical Statistics* **36**, 1703-1706.
- Shaked, M. and Shanthikumar, J. G. (1994). *Stochastic orders and their applications*. Boston: Academic Press.
- Tong, Y. L. (1994). Some recent developments on majorization inequalities in probability and statistics. *Linear Algebra and its Applications* **199**, 69-90.
- Uchaikin, V. V. and Zolotarev, V. M. (1999). *Chance and stability. Stable distributions and their applications*. Utrecht: VSP.
- Zastavnyi, V. P. (1993). Positive definite functions depending on the norm. *Russian Journal of Mathematical Physics* **1**, 511-522.
- Zolotarev, V. M. (1986). *One-dimensional stable distributions*. Providence: American Mathematical Society.