HIER

Harvard Institute of Economic Research

Discussion Paper Number 2085

On Efficiency of Linear Estimators
Under Heavy-Tailedness

by

Rustam Ibragimov

August 2005

HARVARD UNIVERSITY
Cambridge, Massachusetts

ON EFFICIENCY OF LINEAR ESTIMATORS UNDER HEAVY-TAILEDNESS[1]

Running title: EFFICIENCY OF LINEAR ESTIMATORS

By Rustam Ibragimov[2]

Department of Economics, Yale University

First draft: March 2003

This version: April 2005

Address for manuscript correspondence:

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

Email: ribragim@fas.harvard.edu
Phone: +1-203-887-1631
Fax: +1-617-495-7730 .

ABSTRACT

The present paper develops a new unified approach to the analysis of efficiency, peakedness and majorization properties of linear estimators. It further studies the robustness of these properties to heavy-tailedness assumptions. The main results show that peakedness and majorization phenomena for random samples from log-concavely

[1]The 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"

[2]I 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 Donald Brown, Aydin Cecen, Gary Chamberlain, Brian Dineen, Darrell Duffie, Xavier Gabaix, Philip Haile, Samuel Karlin, Alex Maynard, 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, and the International Conference on Stochastic Finance, Lisbon, Portugal, September 2004, for many helpful comments and discussions.

distributed populations established in the seminal work by Proschan (1965) continue to hold for not extremely thick-tailed distributions. However, these phenomena are reversed in the case of populations with extremely heavy-tailed densities.

Among other results, we show 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. Moreover, in such a case, the sample mean exhibits the important property of monotone consistency and, thus, an increase in the sample size always improves its performance. However, as we demonstrate, 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 thick-tailedness. We also provide applications of the main efficiency and majorization comparison results in the study of concentration inequalities for linear estimators as well as their extensions to the case of wide classes of dependent data.

The main results obtained in the paper provide the basis for the analysis of many problems in a number of other areas, in addition to econometrics and statistics, and, in particular, have applications in the study of robustness of model of firm growth 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.

KEYWORDS: Linear estimators, efficiency, peakedness, majorization, robustness, heavy-tailed distributions, sample mean, monotone consistency, linear combinations of random variables, tail probabilities

JEL Classification: C12, C13, C16

1 Introduction and discussion of the results

1.1 Efficiency and peakedness of estimators

A problem of great importance in econometrics and statistics is that of comparison of estimators' performance. The present paper develops a new unified approach to comparisons of linear estimators under heavy-tailedness and obtains complete characterizations of the optimal linear estimators for thick-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 $U : \mathbf{R}_+ = [0, \infty) \to \mathbf{R}$, $\hat{\theta}^{(1)}$ is preferred to $\hat{\theta}^{(2)}$ if (provided that the expectations exist)

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

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 $P(|X-\theta| > \epsilon) \le P(|Y-\theta| > \epsilon)$ for all $\epsilon \ge 0$. 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 with finite second moments 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, e.g., Shaked and Shanthikumar, 1994, pp. 3-4, and Remark 3.4 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 fundamental problem in statistics and econometrics of whether having more data is always better for inference. 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 always 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.

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 strictly monotonically in n for all $\epsilon \geq 0$.*

1.2 Objectives and key results

The questions of key interest in statistical inference and in the study of the law of large numbers are those of efficiency comparisons for linear estimators and efficiency and monotone consistency properties of the sample mean. Besides econometrics, statistics and probability theory, the necessity in the analysis of efficiency properties of linear estimators and closely related problems in the study of tail probabilities of linear combinations of r.v.'s naturally arises in a number of other fields, including economics, finance, risk management and mathematical biology. In particular, this analysis turns out to be crucial for the study of models of firm growth for firms that can invest into information about their markets, optimal bundling strategies for a multiproduct monopolist, value at risk models for financial portfolios as well as for the analysis of multifactorial inheritance models of mathematical evolutionary theory (see the discussion in the author's Ph.D. dissertation Ibragimov, 2005, as well as in Ibragimov, 2004a, b, c, d, and in Subsection 1.5 below).

Efficiency comparisons for linear estimators are examples of many problems in economics and statistics that depend on majorization phenomena for linear combinations of r.v.'s. The majorization relation is a formalization of the concept of diversity in the components of vectors. Over the past decades, majorization theory, which focuses on the study of this relation and functions that preserve it, has found applications in disciplines ranging from probability theory, statistics and economic theory to mathematical genetics, linear algebra and geometry. This paper presents a unified framework for the analysis of the majorization properties of linear combinations of random variables. It further studies the robustness of these majorization properties and their implications for efficiency comparisons of linear estimators to thick-tailedness assumptions. The main results show that the properties and their implications are robust to heavy-tailedness assumptions as long as the distributions entering these assumptions are not extremely thick-tailed (Theorem 3.1 and 3.3).[3] But the properties and their conclusions are reversed for distributions with extremely long-tailed densities (Theorem 3.2 and 3.4).

Among other results, we show that the sample mean $\overline{X}_n = (1/n)\sum_{i=1}^n X_i$ 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 $a_i \geq 0$, $i = 1, ..., n$, $\sum_{i=1}^n a_i = 1$, of the population center in the case of extremely heavy-tailed data (Theorem 3.2). The above results imply that P-efficiency of $\overline{X}_n$ 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 n. Therefore, having more data is always disadvantageous for inference if the sample mean is used to estimate the population center under extreme thick-tailedness.

Furthermore, we obtain extensions of the above results for a wide class of dependent data (see Section 4). Namely, we show all the results in the paper continue to hold for convolutions of dependent r.v.'s with joint $\alpha-$symmetric distributions and their analogues with non-identical marginals.[4] The class of $\alpha-$symmetric distributions is very wide and includes, in particular, spherical distributions corresponding to $\alpha = 2$. Important examples of spherical distributions, in turn, are given by Kotz type, multinormal and logistic distributions and multivariate stable laws. In addition, they include a subclass of mixtures of normal distributions as well as multivariate $t-$distributions that were used in a number of papers to model heavy-tailedness phenomena with dependence and finite moments up to a certain order (see, among others, Praetz, 1972, Blattberg and Gonedes, 1974, and Glasserman et. al., 2002). Moreover, the class of $\alpha-$symmetric distributions includes a wide class of convolutions of models with common shocks affecting all risks (such as macroeconomic or political ones, see Andrews, 2003) which are of great importance in economics and finance.

The results on efficiency and peakedness comparisons of linear estimators obtained in this paper results show that majorizations for log-concavely distributed signals established in the seminal work by Proschan (1965) continue to hold for r.v.'s with not extremely heavy-tailed densities. More precisely, the tails of distributions of linear combinations of such long-tailed r.v.'s continue to exhibit the property of Schur-convexity, as in the case of log-concave distributions (Theorem 3.1). However, as we demonstrate, the majorization properties are reversed for extremely thick-tailed distributions, in which case Schur-convexity of the tails is replaced by their Schur-concavity (Theorem 3.2). This is the first probabilistic result that shows that majorization properties of log-concave densities are reversed for a wide class of distributions and is the key to the analysis of efficiency of linear estimators under extreme heavy-tailedness and, in particular, to the reversals of the stylized fact that having more data is always advantageous for inference obtained in Theorem 3.2. One should emphasize here that, although log-concave distributions have many appealing properties that have been utilized in a number of works in economics, statistics and probability theory, they cannot be used in the study of thick-tailedness phenomena since any log-concave density is extremely

[3]According to well-established parlance in the many scientific literatures, robustness is understood to mean insensitivity to deviations from distributional assumptions. In this paper, the use of the term "robustness" accords with this tradition.

[4]An $n-$dimensional distribution is called $\alpha-$symmetric if its characteristic function can be written as $\phi((\sum_{i=1}^n |t_i|^\alpha)^{1/\alpha})$, where ϕ is a continuous function and $\alpha > 0$. Such distributions should not be confused with multivariate spherically symmetric stable distributions, which have characteristic functions $\exp\big[-\lambda\big(\sum_{i=1}^n t_i^2\big)^{\beta/2}\big]$, $0 < \beta \leq 2$. Obviously, spherically symmetric stable distributions are particular examples of $\alpha-$symmetric distributions with $\alpha = 2$ (that is, of spherical distributions) and $\phi(x) = \exp(-x^\beta)$.

light-tailed: in particular, its tails decline at least exponentially fast and all its moments exist (see An, 1998, and Section 2 in the present paper).

1.3 Heavy-tailedness paradigm

This paper belongs to a large stream of literature in economics and finance that has focused on the analysis of thick-tailed phenomena. This stream of literature goes back to Mandelbrot (1963) (see also the papers in Mandelbrot, 1997, and Fama, 1965), who pioneered the study of heavy-tailed distributions with tails declining as $x^{-\alpha}$, $\alpha > 0$, in these fields. If a model involves a r.v. X with such thick-tailed distribution, then

$$P(|X| > x) \sim x^{-\alpha}. \tag{1.2}$$

The r.v. X for which this is the case has finite moments $E|X|^p$ of order $p < \alpha$. However, the moments are infinite for $p \geq \alpha$.

It was documented in numerous studies that the time series encountered in many fields in economics and finance are heavy-tailed (see the discussion in Loretan and Phillips, 1994, Meerschaert and Scheffler, 2000, Gabaix, Gopikrishnan, Plerou and Stanley, 2003, and references therein). Motivated by these empirical findings, a number of studies in financial economics have focused on portfolio and value-at-risk modelling with heavy-tailed returns (see, e.g., the reviews in Duffie and Pan, 1997, Uchaikin and Zolotarev, 1999, Ch. 17, and Glasserman, Heidelberger and Shahabuddin, 2002). Several authors considered problems of statistical inference for data from thick-tailed populations (see Loretan and Phillips, 1994, the papers in Adler, Feldman and Taqqu, 1998, and references therein).

Mandelbrot (1963) presented evidence that historical daily changes of cotton prices have the tail index $\alpha \approx 1.7$, and thus have infinite variances. Using different models and statistical techniques, subsequent research reported the following estimates of the tail parameters α for returns on various stocks and stock indices: $3 < \alpha < 5$ (Jansen and de Vries, 1991); $2 < \alpha < 4$ (Loretan and Phillips, 1994); $1.5 < \alpha < 2$ (McCulloch, 1996, 1997); $0.9 < \alpha < 2$ (Rachev and Mittnik, 2000). Recent studies (see Gabaix et. al., 2003, and references therein) have found that the returns on many stocks and stock indices have the tail exponent $\alpha \approx 3$, while the distributions of trading volume and the number of trades on financial markets obey power laws (1.2) with $\alpha \approx 1.5$ and $\alpha \approx 3.4$, respectively. As discussed in Gabaix et. al. (2003), these estimates of the tail indices α are robust to different types and sizes of financial markets, market trends and are similar for different countries. Motivated by these empirical findings, Gabaix et. al. (2003) proposed a model that demonstrates that the above power laws for stock returns, trading volume and the number of trades are explained by trading of large market participants, namely, the largest mutual funds whose sizes have the tail exponent $\alpha \approx 1$. Power laws (1.2) with $\alpha \approx 1$ (Zipf laws) have also been found to hold for firm sizes (see Axtell, 2001) and city sizes (see Gabaix, 1999a, b for discussion and explanations of the Zipf law for cities). One should also note that some studies also report the tail exponent to be close to one or even slightly less than one for such financial time series as Bulgarian lev/US dollar exchange spot rates and increments of the market time process for Deutsche Bank price record (see Rachev and Mittnik, 2000).

The fact that a number of economic and financial time series have the tail exponents of approximately one is very important in the context of the results in this paper: as we demonstrate, majorization phenomena for linear combinations of r.v.'s and their applications in the case of distributions with the tail exponents $\alpha < 1$ and infinite means are the opposites of those for distributions with $\alpha > 1$ for which the first moment is finite.

Several frameworks have been proposed to model heavy-tailedness phenomena, including stable distributions, Pareto distributions, multivariate $t-$distributions, mixtures of normals, power exponential distributions, ARCH processes, mixed diffusion jump processes, variance gamma and normal inverse Gamma distributions. However, the debate concerning the values of the tail indices for different heavy-tailed financial data and on appropriateness of their modelling based on certain above distributions is still under way in empirical literature. In particular, as indicated before, a number of studies continue to find tail parameters less than two in different financial data sets and also argue that stable distributions are appropriate for their modelling.

1.4 Thick tails and extremely thick tails and extensions to the case of dependence

To illustrate the main ideas of the proof and in order to simplify the presentation of the main results in this paper, we first model heavy-tailedness using the framework of independent stable distributions and their convolutions. More precisely, the class of not extremely thick-tailed distributions is first modelled using convolutions of stable distributions with (different) indices of stability greater than one. Similarly, the results of the paper for extremely heavy-tailed case are first presented and proven using the framework of convolutions of stable distributions with characteristic exponents less than one. The former class has tail exponents $\alpha > 1$ and for the latter class one has $\alpha < 1$.

In Section 4 we show, however, that all the results obtained in the paper continue to hold for a wide class of multivariate distributions for which marginals are dependent and can be non-identical and, in addition to that, can have finite variances, unlike stable distributions and their convolutions. As indicated before, according to these extensions, all the results in the paper continue to hold for convolutions of $\alpha-$symmetric distributions and their analogues with non-identical one-dimensional marginals (see Subsection 1.2). Similar to the framework based on stable distributions, the majorization properties of log-concave densities and their implications for monotone consistency of the sample mean and for efficiency comparisons of linear estimators continue to hold for convolutions of $\alpha-$symmetric distributions with $\alpha > 1$. These properties and their implications are reversed in the case of convolutions of $\alpha-$symmetric distributions with $\alpha < 1$.

One should also note here that all the results in the paper are available for the case of skewed distributions, including skewed stable distributions (such as, for instance, extremely heavy-tailed Lévy distributions with $\alpha = 1/2$ concentrated on the positive semi-axis) and, according to the extensions discussed above, $\alpha-$symmetric distributions with skewed marginals. Therefore, this paper, in fact, succeeds in the unification of the robustness of majorization properties of convolutions of distributions and their implications for efficiency comparisons of linear estimators to all the main distributional properties: heavy-tailedness, dependence, skewness and the case of non-identical one-dimensional distributions.

1.5 Optimistic implications

The main majorization results obtained in this paper provide the basis for the analysis of many problems in a number of other areas, in addition to econometrics and statistics, and have many applications besides the study of efficiency properties of linear estimators dealt with in this paper. These applications, presented in the author's Ph.D. dissertation Ibragimov (2005) and, in the form focusing on each particular area and the audience of readers, in Ibragimov (2004a, b, c, d), include the study of robustness of the model of demand-driven innovation and spatial competition over time, value at risk analysis, optimal strategies for a multiproduct monopolist as well that of inheritance models in mathematical evolutionary theory.[5]

The main message of the results in this paper and of their applications is that the presence of heavy-tailedness can either reinforce or reverse the implications of models in econometrics, statistics, probability theory, economics, finance and risk management, depending on the degree of thick-tailedness. Similar to the properties of P-efficiency of the sample mean in this paper, the standard implications of models in the above fields continue to hold for not extremely heavy-tailed distributions. However, these properties are reversed under the assumptions of extreme thick-tailedness.

This message is optimistic since, according to the results in this paper and those in Ibragimov (2005), the models in the above fields are robust to heavy-tailedness (and dependence) as long as the tail indices $\alpha > 1$ and empirical studies observe such values for α in most of economic and financial time series. However, the reversals of the models are possible for a wide class of extremely thick-tailed distributions. Therefore, the models should be applied with care in presence of very heavy-tailed signals, especially in the case of the tail indices close to the critical boundary

[5] The following list summarizes some of the applications of the main majorization results in this paper presented in Ibragimov (2005) (see also Ibragimov, 2004a, b, c, d).

(i) We show, for the first time in the literature, that the stylized fact that portfolio diversification is always preferable is reversed for a wide class of distributions of risks. The class of distributions for which this is the case is the class of extremely heavy-tailed distributions. The encouraging message of the results is that the stylized facts on diversification are nevertheless robust to thick-tailedness of risks or returns as long as their distributions are not extremely long-tailed.

Moreover, we demonstrate that, in the world of not extremely heavy-tailed risks, VaR satisfies the important condition of coherency, which is a natural requirement to be imposed on a measure of risk from the points of view of exchange, regulators and society. However, coherency of the value at risk is always violated if distributions of risks are extremely thick-tailed. We also obtain sharp bounds on the VaR of the returns on portfolios of risks with long-tailed returns.

(ii) Using the general majorization results obtained in this paper, we develop a framework that allows one to model the optimal bundling problem of a multiproduct monopolist providing interrelated goods with an arbitrary degree of complementarity or substitutability. Characterizations of optimal bundling strategies are derived for the seller in the case of long-tailed valuations and tastes for the products. We show, in particular, that if goods provided in a Vickrey auction or any other revenue equivalent auction are substitutes and bidders' tastes for the objects are not extremely heavy-tailed, then the monopolist prefers separate provision of the products. However, if the goods are complements and consumers' tastes are extremely thick-tailed, then the seller prefers providing the products on a single auction. We also present results on consumers' preferences over bundled auctions in the case when their valuations exhibit heavy-tailedness. In addition, we obtain characterizations of optimal bundling strategies for a monopolist who provides complements or substitutes for profit-maximizing prices to buyers with long-tailed tastes.

(iii) Another application of the main majorization results in this paper concerns the analysis of growth of firms that invest into learning about the next period's optimal product. We present a study of robustness of the model of demand-driven innovation and spatial competition over time with log-concavely distributed signals developed by Jovanovic and Rob (1987) to heavy-tailedness assumptions. The implications of the model remain valid for not extremely long-tailed distributions of consumers' signals. However, again these properties are reversed for signals with extremely thick-tailed densities.

(iv) We study transmission of traits through generations in multifactorial inheritance models with sex- and time-dependent heritability. We further analyze the implications of these models under heavy-tailedness of traits' distributions. Among other results, we show that in the case of a trait (for instance, a medical or behavioral disorder or a phenotype with significant heritability affecting human capital in an economy) with not very thick-tailed initial density, the trait distribution becomes increasingly more peaked, that is, increasingly more concentrated and unequally spread, with time. But these patterns are reversed for traits with sufficiently heavy-tailed initial distributions (e.g., a medical or behavioral disorder for which there is no strongly expressed risk group or a relatively equally distributed ability with significant genetic influence). Such traits' distributions become less peaked over time and increasingly more spread in the population.

$\alpha = 1$.

2 Notations

In this section, we introduce classes of distributions we will be dealing with throughout the paper.

We say that a r.v. X with density $f : \mathbf{R} \to \mathbf{R}$ and the convex distribution support $\Omega = \{x \in \mathbf{R} : f(x) > 0\}$ is 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 $\lambda \in [0,1]$,

$$f(\lambda x_1 + (1-\lambda)x_2) \geq (f(x_1))^{\lambda}(f(x_2))^{1-\lambda}. \tag{2.1}$$

(see An, 1998). A distribution is said to be log-concave if its density f satisfies (2.1).

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 $x \to \infty$ 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.[6]

Throughout the paper, $\mathcal{LC}$ denotes the class of symmetric log-concave distributions.[7]

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

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

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

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); $\alpha = 1/2$ and $\beta \pm 1$ (Lévy distributions).[8] 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 (1.2) and thus

[6] Some of these properties are the following:
Any log-concave density is unimodal. Moreover, it has the property of strong unimodality, that is, its convolution with any other unimodal density is again unimodal;
The survivor and distribution functions of log-concave densities are both log-concave and, thus, a log-concavely distributed r.v. has the new-better-than-used property;
A log-concave density is of Pólya frequency of order 2 (PF-2);
The hazard function of a log-concave density is monotonically increasing.
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 $\alpha \geq 1$.

[7] $\mathcal{LC}$ stands for "log-concave".

[8] The densities of Cauchy distributions are $f(x) = \sigma/(\pi(\sigma^2 + (x-\mu)^2))$. Lévy distributions 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.

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 $a_i \geq 0$, $i = 1, ..., n$, $\sum_{i=1}^n a_i X_i / \Big(\sum_{i=1}^n a_i^\alpha \Big)^{1/\alpha} \sim S_\alpha(\sigma, \beta, \mu)$.

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).

For $0 < r < 2$, we denote by $\overline{\mathcal{CS}}(r)$ the class of distributions which are convolutions of symmetric stable distributions $S_\alpha(\sigma, 0, 0)$ with characteristic exponents $\alpha \in (r, 2]$ and $\sigma > 0$.[9] That is, $\overline{\mathcal{CS}}(r)$ consists of distributions of r.v.'s X such that, for some $k \geq 1$, $X = Y_1 + ... + Y_k$, where Y_i, $i = 1, ..., k$, are independent r.v.'s such that $Y_i \sim S_{\alpha_i}(\sigma_i, 0, 0)$, $\alpha_i \in (r, 2]$, $\sigma_i > 0$, $i = 1, ..., k$.

Further, $\overline{\mathcal{CSLC}}$ stands for the class of convolutions of distributions from the classes $\mathcal{LC}$ and $\overline{\mathcal{CS}}(1)$. That is, $\overline{\mathcal{CSLC}}$ is the class of convolutions of symmetric distributions which are either log-concave or stable with characteristic exponents greater than one.[10] In other words, $\overline{\mathcal{CSLC}}$ consists of distributions of r.v.'s X such that $X = Y_1 + Y_2$, where Y_1 and Y_2 are independent r.v.'s with distributions belonging to $\mathcal{LC}$ or $\overline{\mathcal{CS}}(1)$.

Finally, for $0 < r \leq 2$, we denote by $\underline{\mathcal{CS}}(r)$ the class of distributions which are convolutions of symmetric stable distributions $S_\alpha(\sigma, 0, 0)$ with indices of stability $\alpha \in (0, r)$ and $\sigma > 0$.[11] That is, $\underline{\mathcal{CS}}(r)$ consists of distributions of r.v.'s X such that, for some $k \geq 1$, $X = Y_1 + ... + Y_k$, where Y_i, $i = 1, ..., k$, are independent r.v.'s such that $Y_i \sim S_{\alpha_i}(\sigma_i, 0, 0)$, $\alpha_i \in (0, r)$, $\sigma_i > 0$, $i = 1, ..., k$.

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 $\overline{\mathcal{CS}}(r)$ of *convolutions* of symmetric stable distributions with *different* indices of stability $\alpha \in (r, 2]$ is wider than the class of *all* symmetric stable distributions $S_\alpha(\sigma, 0, 0)$ with $\alpha \in (r, 2]$ and $\sigma > 0$. Similarly, the class $\underline{\mathcal{CS}}(r)$ is wider than the class of *all* symmetric stable distributions $S_\alpha(\sigma, 0, 0)$ with $\alpha \in (0, r)$ and $\sigma > 0$.

Clearly, $\overline{\mathcal{CS}}(1) \subset \overline{\mathcal{CSLC}}$ and $\mathcal{LC} \subset \overline{\mathcal{CSLC}}$. It should also be noted that the class $\overline{\mathcal{CSLC}}$ is wider than the class of (two-fold) convolutions of log-concave distributions with stable distributions $S_\alpha(\sigma, 0, 0)$ with $\alpha \in (1, 2]$ and $\sigma > 0$.

By definition, for $0 < r_1 < r_2 \leq 2$, the following inclusions hold: $\overline{\mathcal{CS}}(r_2) \subset \overline{\mathcal{CS}}(r_1)$ and $\underline{\mathcal{CS}}(r_1) \subset \underline{\mathcal{CS}}(r_2)$.

[9]Here and below, $\mathcal{CS}$ stands for "convolutions of stable"; the overline indicates that convolutions of stable distributions with indices of stability *greater* than the threshold value r are taken.

[10]$\mathcal{CSLC}$ is the abbreviation of "convolutions of stable and log-concave".

[11]The underline indicates considering stable distributions with indices of stability *less* than the threshold value r.

In some sense, symmetric (about $\mu = 0$) Cauchy distributions $S_1(\sigma, 0, 0)$ are at the dividing boundary between the classes $\underline{\mathcal{CS}}(1)$ and $\overline{\mathcal{CS}}(1)$ (and between the classes $\underline{\mathcal{CS}}(1)$ and $\overline{\mathcal{CSLC}}$). Similarly, for $r \in (0, 2)$, symmetric stable distributions $S_r(\sigma, 0, 0)$ with the characteristic exponent $\alpha = r$ are at the dividing boundary between the classes $\underline{\mathcal{CS}}(r)$ and $\overline{\mathcal{CS}}(r)$. Further, symmetric normal distributions $S_2(\sigma, 0, 0)$ are at the dividing boundary between the class $\mathcal{LC}$ of log-concave distributions and the class $\underline{\mathcal{CS}}(2)$ of convolutions of symmetric stable distributions with indices of stability $\alpha < 2$.[12]

In what follows, we write $X \sim \mathcal{LC}$ (resp., $X \sim \overline{\mathcal{CSLC}}$, $X \sim \overline{\mathcal{CS}}(r)$ or $X \sim \underline{\mathcal{CS}}(r)$) if the distribution of the r.v. X belongs to the class $\mathcal{LC}$ (resp., $\overline{\mathcal{CSLC}}$, $\overline{\mathcal{CS}}(r)$ or $\underline{\mathcal{CS}}(r)$).

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

The present paper demonstrates that powerful tools for the analysis of efficiency of linear estimators are given by majorization theory. 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 $\sum_{i=1}^k a_{[i]} \le \sum_{i=1}^k b_{[i]}$, $k = 1, ..., n-1$, and $\sum_{i=1}^n a_{[i]} = \sum_{i=1}^n b_{[i]}$, where $a_{[1]} \ge \ldots \ge a_{[n]}$ and $b_{[1]} \ge \ldots \ge b_{[n]}$ denote components of a and b in decreasing order. 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:

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

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

$$(1/(n+1), ..., 1/(n+1), 1/(n+1)) \prec (1/n, ..., 1/n, 0), \quad n \ge 1. \tag{3.2}$$

A function $\phi : A \to \mathbf{R}$ defined on $A \subseteq \mathbf{R}^n$ is called *Schur-convex* (resp., *Schur-concave*) on A if $(a \prec b) \Longrightarrow (\phi(a) \le \phi(b))$ (resp. $(a \prec b) \Longrightarrow (\phi(a) \ge \phi(b))$ 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.

The following theorems present our main results on efficiency comparisons of linear estimators.

In what follows, given a random sample $X_1, ..., X_n$ from a population with center μ, and weights $a = (a_1, ..., a_n) \in \mathbf{R}_+^n$, 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 $\mathcal{I}_n = \{a = (a_1, ..., a_n) \in \mathbf{R}_+^n : \sum_{i=1}^n a_i = 1\}$.

Theorem 3.1 concerns efficiency comparisons for linear estimators in the case of not extremely heavy-tailed data. It provides majorization comparisons that generalize the results in the seminal work of Proschan (1965) for linear combinations of log-concavely distributed r.v.'s (see Subsection 1.2 in the introduction and Remark 3.1 in the present

[12]More precisely, the symmetric Cauchy distributions are the only ones that belong to all the classes $\underline{\mathcal{CS}}(r)$ with $r > 1$ and all the classes $\overline{\mathcal{CS}}(r)$ with $r < 1$. Symmetric stable distributions $S_r(\sigma, 0, 0)$ are the only ones that belong to all the classes $\underline{\mathcal{CS}}(r')$ with $r' > r$ and all the classes $\overline{\mathcal{CS}}(r')$ with $r' < r$. Symmetric normal distributions are the only distributions belonging to the class $\mathcal{LC}$ and all the classes $\overline{\mathcal{CS}}(r)$ with $r \in (0, 2)$.

section) to a wide class of long-tailed distributions. Theorem 3.1 further shows that the sample mean is the best linear unbiased estimator of the population mean in the sense of P-efficiency and, moreover, an increase in the sample size always improves performance of this estimator.

Theorem 3.1 *Let $\mu \in \mathbf{R}$. Suppose that $X_1, ..., X_n$, $n \geq 1$, are random samples such that $X_1 \sim S_\alpha(\sigma, \beta, \mu)$ for some $\sigma > 0$, $\beta \in [-1, 1]$ and $\alpha \in (1, 2]$, or $X_1 - \mu \sim \overline{\mathcal{CSLC}}$. Then the following conclusions hold.*

(i) Let $a, b \in \mathbf{R}_+^n$. 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 $a = (a_1, ..., a_n) \in \mathbf{R}_+^n$ for all $\epsilon > 0$).

(ii) The sample mean $\overline{X}_n = (1/n)\sum_{i=1}^n X_i$ 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, $\overline{X}_n$ exhibits monotone consistency for μ and $P(|\overline{X}_n - \mu| > \epsilon)$ 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 extremely heavy-tailed populations. In this case, having larger samples is, in fact, always disadvantageous if the sample mean is used to estimate the population center. Theorem 3.2 provides the first general results in the literature that show that the results of Proschan (1965) on majorization properties of linear combinations of log-concavely distributed r.v.'s are reversed for a wide class of distributions. The class of distributions for which this is the case is precisely the class of distributions with extremely thick tails.

Theorem 3.2 *Let $\mu \in \mathbf{R}$. Suppose that $X_1, ..., X_n$, $n \geq 1$, are random samples such that $X_1 \sim S_\alpha(\sigma, \beta, \mu)$ for some $\sigma > 0$, $\beta \in [-1, 1]$ and $\alpha \in (0, 1)$, or $X_1 - \mu \sim \underline{\mathcal{CS}}(1)$. Then the following conclusions hold.*

(i) Let $a, b \in \mathbf{R}_+^n$. 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 $a = (a_1, ..., a_n) \in \mathbf{R}_+^n$ for all $\epsilon > 0$).

(ii) The sample mean $\overline{X}_n = (1/n)\sum_{i=1}^n X_i$ 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 $\overline{X}_n$ decreases with n, that is, $P(|\overline{X}_{n+1} - \mu| > \epsilon) > P(|\overline{X}_n - \mu| > \epsilon) > P(|X_1 - \mu| > \epsilon)$ 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 in the classes $\overline{\mathcal{CS}}(r)$ 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 3.1 further provides concentration inequalities for linear estimators in the case of classes $\overline{\mathcal{CS}}(r)$ that refine and complement the efficiency and peakedness comparisons implied by Theorem 3.1.

Theorem 3.3 *Let $r \in (0, 2)$. Suppose that $X_1, ..., X_n$ are random samples such that $X_1 \sim S_\alpha(\sigma, \beta, \mu)$ for some $\sigma > 0$, $\beta \in [-1, 1]$, and $\alpha \in (r, 2]$, where $\beta = 0$ for $\alpha = 1$, or $X_1 \sim \overline{\mathcal{CS}}(r)$. Then the following conclusions hold.*

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

(ii) $\hat{\theta}_n(a) = \sum_{i=1}^n a_i X_i,\ a \in \mathcal{I}_n$, *satisfies the following concentration inequalities for all* $x > 0$:

$$P\Big(|\overline{X}_n - \mu| > n^{1/r-1}x/\big(\sum_{i=1}^n a_i^r\big)^{1/r}\Big) \leq P(|\hat{\theta}_n(a) - \mu| > x) \leq P\Big(|X_1 - \mu| > x/\big(\sum_{i=1}^n a_i^r\big)^{1/r}\Big),$$

with strict right-hand side inequality if $a = (a_1, a_2, ..., a_n)$ *is not a permutation of* $(1, 0, ..., 0)$ *and strict left-hand side inequality if* $a \neq (1/n, 1/n, ..., 1/n)$.

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 from the classes $\underline{\mathcal{CS}}(r)$. The concentration inequalities in Theorem 3.4 refine and complement the efficiency orderings for linear estimators given by Theorem 3.2.

Theorem 3.4 *Let* $r \in (0, 2]$. *Suppose that* $X_1, ..., X_n$ *are random samples such that* $X_1 \sim S_\alpha(\sigma, \beta, \mu)$ *for some* $\sigma > 0$, $\beta \in [-1, 1]$ *and* $\alpha \in (0, r)$, *where* $\beta = 0$ *for* $\alpha = 1$, *or* $X_1 \sim \underline{\mathcal{CS}}(r)$. *Then the following conclusions hold.*

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

(ii) $\hat{\theta}_n(a) = \sum_{i=1}^n a_i X_i,\ a \in \mathcal{I}_n$, *satisfies the following concentration inequalities for all* $x > 0$:

$$P\Big(|X_1 - \mu| > x/\big(\sum_{i=1}^n a_i^r\big)^{1/r}\Big) \leq P(|\hat{\theta}_n(a) - \mu| > x) \leq P\Big(|\overline{X}_n - \mu| > n^{1/r-1}x/\big(\sum_{i=1}^n a_i^r\big)^{1/r}\Big),$$

with strict left-hand side inequality if $a = (a_1, ..., a_n)$ *is not a permutation of* $(1, 0, ..., 0)$ *and strict right-hand side inequality if* $a \neq (1/n, 1/n, ..., 1/n)$.

Remark 3.1 *Theorem 3.1 provides a substantial generalization of the results in the seminal work of 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* $a = (a_1, ..., a_n) \in \mathbf{R}_+^n$ *for all* $\epsilon > 0$ *for random samples* $X_1, ..., X_n$ *from symmetric log-concavely distributed populations* $(X_1 - \mu \sim \mathcal{LC})$.[13] [14]*). 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. For instance, Eaton (1988) used generalizations of the results to obtain concentration inequalities for Gauss-Markov estimators. Karlin (1984, 1992) applied them in the study of environmental sex determination models. Jovanovic and Rob (1987) used majorization properties of log-concavely distributed r.v.'s derived by Proschan (1965) in the analysis of the model of demand-driven innovation and spatial competition over time. Fang and Norman (2003) applied them in the study of optimal bundling strategies for a multiproduct monopolist. Several authors (see, e.g., Proschan, 1965, Tong, 1994, and Jensen, 1997) discussed implications of the majorization results for log-concave distributions and their extensions in the study of monotone consistency of estimators. A number of papers in probability and statistics have focused on extension of*

[13] Proschan (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 $a = (1/n, 1/n, ..., 1/n) \in \mathbf{R}^n$ with identical components and certain transforms of symmetric Cauchy r.v.'s.

[14] The main results in Proschan (1965) are reviewed in Section 12.J in Marshall and Olkin (1979). Proschan's (1979) work is also presented, in a rearranged form, in Section 11 of Chapter 7 in Karlin (1968). Peakedness results in Proschan (1965) and Karlin (1968) are formulated for "PF2 densities," which is the same as "log-concave densities."

Proschan's results (see, among others, Chan, Park and Proschan, 1989, the review in Tong, 1994, Jensen, 1997, and Ma, 1998). One should emphasize, 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 $\psi(a, \epsilon)$, $\epsilon > 0$, for certain wide classes of r.v.'s.[15] *As indicated before, such general results are provided, to our knowledge, for the first time in the literature, by Theorems 3.2 and 3.4.*

Remark 3.2 *Using the approach presented in the paper, one can easily obtain, in complete similarity with the proof of Theorems 3.1-3.4, their analogues for i.i.d. r.v.'s with skewed Cauchy-type distributions $X_i \sim S_1(\sigma, \beta, \mu)$ with $\alpha = 1$ and $\beta \neq 0$. For instance, similar to the proof of the theorems, it is not difficult to show that if X_i, $i = 1, 2, ..., n$, are i.i.d. r.v.'s such that $X_1 \sim S_1(\sigma, \beta, \mu)$, then, for all $x \in \mathbf{R}$, the tail probability $f(a, x) = P\Big(\sum_{i=1}^n a_i X_i > x\Big)$ is strictly Schur-concave in $a = (a_1, ..., a_n) \in \mathbf{R}_+^n$ if $\beta > 0$ and is strictly Schur-convex in $a = (a_1, ..., a_n) \in \mathbf{R}_+^n$ if $\beta < 0$.*[16]

Remark 3.3 *If the population has a symmetric Cauchy distribution $S_1(\sigma, 0, 0)$ which is, as discussed in Section 1, exactly at the dividing boundary between the class $\underline{\mathcal{CS}}(1)$ in Theorem 3.2 and the class $\overline{\mathcal{CSLC}}$ in Theorem 3.1, then the tail probability $\psi(a, \epsilon)$ in the theorems depends only on $\sum_{i=1}^n a_i$ and ϵ and so is both Schur-concave and Schur-convex in $a \in \mathbf{R}_+^n$ for all $\epsilon > 0$ (see Proschan, 1965). Similarly, the function $\psi(a, \epsilon)$ in Theorems 3.3 and 3.4 depends only on $\sum_{i=1}^n a_i^r$ and ϵ and so is both Schur-concave and Schur-convex in $(a_1^r, ..., a_n^r)$ for all $\epsilon > 0$ if the population in the theorems has a symmetric stable distribution $S_r(\sigma, 0, 0)$ with the index of stability $\alpha = r$ which is at the dividing boundary between the classes $\overline{\mathcal{CS}}(r)$ and $\underline{\mathcal{CS}}(r)$. As follows from the proof of Theorems 3.1-3.4, the above implies that Theorems 3.1 and 3.2 continue to hold for convolutions of distributions from the classes $\overline{\mathcal{CSLC}}$ and $\underline{\mathcal{CS}}(1)$ with symmetric Cauchy distributions $S_1(\sigma, 0, 0)$. Similarly, Theorem 3.3 and 3.4 continue to hold for convolutions of distributions from the classes $\overline{\mathcal{CS}}(r)$ and $\underline{\mathcal{CS}}(r)$ with symmetric stable distributions $S_r(\sigma, 0, 0)$.*

Remark 3.4 *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 $U : \mathbf{R} \to \mathbf{R}$ for which the expectations exist (see, e.g., Shaked and Shanthikumar, 1994, pp. 3-4). This fact and Theorems 3.3-3.2 imply corresponding results concerning majorization properties of expectations of loss functions of linear estimators under heavy-tailedness. For instance, we get that if $U : \mathbf{R}_+ \to \mathbf{R}$ 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 $(a_1^r, ..., a_n^r)$ under the assumptions of Theorem 3.3 and is Schur-concave in $(a_1^r, ..., a_n^r)$ under the assumptions of Theorem 3.4. We also get that the function $\varphi(a)$, $a \in \mathbf{R}_+^n$ is Schur-concave in $(a_1^2, ..., a_n^2)$ if $X_i \sim S_\alpha(\sigma, \beta, \mu)$, $i = 1, ..., n$, for some $\sigma > 0$, $\beta \in [-1, 1]$ and $\alpha \in (0, 2)$, where $\beta = 0$ for $\alpha = 1$, or $X_i \sim \underline{\mathcal{CS}}(2)$.*

[15] One should note that the proof in Proschan (1965) can be reproduced word to word with respective changes of signs of inequalities under the "assumptions" that $X_1, ..., X_n$ are i.i.d. symmetric log-convexily distributed r.v.'s. However, as it is easy to see, the later objects do not exist, namely, there does not exist a symmetric r.v.'s with a log-convex density (see also An, 1998). Therefore, this approach to obtaining counterparts of the results in Proschan (1965) for Schur-concavity of $\psi(a, \epsilon)$, $\epsilon > 0$, is hopeless.

[16] Since, as it is easy to see, for all $c_i \geq 0$, $i = 1, 2, ..., n$, $\sum_{i=1}^n c_i X_i$ has the same distribution as $\Big(\sum_{i=1}^n c_i\Big) X_1 - \frac{2\sigma\beta}{\pi}\Big(\sum_{i=1}^n c_i \ ln \ c_i - (\sum_{i=1}^n c_i) \ ln \ (\sum_{i=1}^n c_i)\Big)$ and, by Proposition 3.C.1.a in Marshall and Olkin (1979), the function $\sum_{i=1}^n c_i \ ln \ c_i$ is strictly Schur-convex in $c = (c_1, ..., c_n) \in \mathbf{R}_+^n$.

These results complement those in Efron (1969) and Eaton (1970) (see also Marshall and Olkin, 1979, pp. 361-365) who studied classes of functions $U : \mathbf{R} \to \mathbf{R}$ *and r.v.'s* $X_1, ..., X_n$ *for which Schur-concavity of* $\varphi(a)$, $a \in \mathbf{R}_+^n$ *in* $(a_1^2, ..., a_n^2)$ *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. It is important to note here that in the case of increasing convex loss functions* $U : \mathbf{R}_+ \to \mathbf{R}$ *and r.v.'s* $X_1, ..., X_n$ *satisfying the assumptions of Theorem 3.2, the expectations* $EU(|\hat{\theta}_n(a) - \mu|)$ *are infinite for all* $a \in \mathbf{R}_+^n$.[17] *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* $(a_1, ..., a_n) \in \mathbf{R}$ *for all i.i.d. r.v.'s* $Y_1, ..., Y_n$ *and convex functions* $f : \mathbf{R} \to \mathbf{R}$ *as it might seem on the first sight.*

Remark 3.5 *According to Lemma in Birnbaum (1948) and its proof, P-efficiency is preserved under convolutions of symmetric unimodal absolutely continuous estimators. In particular, the following result holds. Suppose that* $\hat{\theta}^{(1)}$ *and* $\hat{\theta}^{(2)}$ *are independent and are each P-more efficient than an estimator* $\hat{\theta}^{(3)}$ *of the same population parameter* θ*. Then the convex combinations* $\alpha\hat{\theta}^{(1)} + (1-\alpha)\hat{\theta}^{(2)}$, $\alpha \in [0,1]$, *of* $\hat{\theta}^{(1)}$ *and* $\hat{\theta}^{(2)}$ *are also P-more efficient than* $\hat{\theta}^{(3)}$ *if* $\hat{\theta}_i$, $i = 1, 2, 3$, *are absolutely continuous, unimodal and symmetric about* θ*. Since for any* $b \in \mathbf{R}^n$, *the set* $\{\tilde{b} \in \mathbf{R}^n : \tilde{b} \prec b\}$ *is convex (see, e.g., Proposition 4.C.1 in Marshall and Olkin, 1979), part (i) of Theorem 3.1 implies the following result that shows that, in the case of not extremely heavy-tailed populations, P-efficiency comparisons of linear estimators continue to hold for their convex combinations even if the estimators are dependent. Suppose that, under the assumptions of Theorem 3.1,* $a \prec b$ *and* $c \prec b$, *and* a *and* c *are not permutations of* b *and, thus, according to part (i) of the theorem, the estimators* $\hat{\theta}^{(1)} = \hat{\theta}_n(a)$ *and* $\hat{\theta}^{(2)} = \hat{\theta}_n(c)$ *are both P-more efficient than* $\hat{\theta}^{(3)} = \hat{\theta}_n(b)$*. Then the convex combinations* $\alpha\hat{\theta}^{(1)} + (1-\alpha)\hat{\theta}^{(2)} = \alpha\hat{\theta}(a) + (1-\alpha)\hat{\theta}(c) = \hat{\theta}(\alpha a + (1-\alpha)c)$, $\alpha \in [0,1]$, *of* $\hat{\theta}^{(1)}$ *and* $\hat{\theta}^{(2)}$ *are also P-more efficient than* $\hat{\theta}^{(3)} = \hat{\theta}_n(b)$*. In contrast, under the assumptions of Theorem 3.2, the estimators* $\hat{\theta}^{(1)} = \hat{\theta}_n(a)$ *and* $\hat{\theta}^{(2)} = \hat{\theta}_n(c)$ *and their convex combinations* $\alpha\hat{\theta}^{(1)} + (1-\alpha)\hat{\theta}^{(2)}$, $\alpha \in [0,1]$, *are all P-less efficient than* $\hat{\theta}^{(3)} = \hat{\theta}_n(b)$.

4 Extensions to the dependent case

As indicated in Subsection 1.4 in the introduction, the results obtained in this paper continue to hold for wide classes of dependent and non-identically distributed r.v.'s. More precisely, the results continue to hold for convolutions of r.v.'s with joint $\alpha-$symmetric and spherical distributions and their non-identically distributed versions as well as for a wide class of models with common shocks.

According to the definition introduced by Cambanis, Keener and Simons (1983), an $n-$dimensional distribution is called $\alpha-$symmetric if its characteristic function (c.f.) can be written as $\phi((\sum_{i=1}^n |t_i|^\alpha)^{1/\alpha})$, where $\phi : \mathbf{R}_+ \to \mathbf{R}$ is a continuous function and $\alpha > 0$. The number α is called the index and the function ϕ is called the c.f. generator of the $\alpha-$symmetric distribution. The class of $\alpha-$symmetric distributions is very broad and contains, in particular, spherical distributions corresponding to the case $\alpha = 2$ (see Fang, Kotz and Ng, 1990, p. 184). Spherical distributions, in turn, include such important examples as Kotz type, multinormal, multivariate t and multivariate stable laws (Fang et. al., 1990, Ch. 3). Furthermore, for any $0 < \alpha \leq 2$, the class of $\alpha-$symmetric distributions

[17] Since the function $(f(x) - f(0))/x$ is increasing in $x > 0$ by, e.g., Marshall and Olkin (1979), p. 453.

includes distributions of risks $X_1, ..., X_n$ that have the representation

$$(X_1, ..., X_n) = (ZY_1, ..., ZY_n) \tag{4.1}$$

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 Fang et. al., 1990, p. 197). Models (4.1) and their convolutions belong to the class of models with common shocks Z, such as macroeconomic or political ones, that affect all risks Y_i.

It is important to emphasize here that the necessity in the study of effects of common shocks arises in many areas of economics and finance (see Andrews, 2003). The extensions of the results in this paper to such models provide a new approach to the analysis of robustness of many models in those fields as well as statistics, econometrics and risk management to both heavy-tailedness and to common shocks.

In addition, one should indicate here that the extensions of the results to $\alpha-$symmetric and, in particular, spherical distributions cover many thick-tailed models with finite variances and finite higher moments. For instance, multivariate $t-$distributions that belong to the class of spherical distributions, provide one of now well-established approaches to modelling heavy-tailedness phenomena with moments up to some order (see Praetz, 1972, Blattberg and Gonedes, 1974, and Glasserman, 2002). The following theorems provide precise formulations of the extensions of the results in Section 3 to the dependent case. According to the theorems, all the results presented in that section for convolutions of i.i.d. stable distributions with indices of stability α belonging to a certain range (and convolutions of those with log-concave distributions in the case of the class $\overline{\mathcal{CSLC}}$) continue to hold for convolutions of $\alpha-$symmetric distributions and models with common shocks (4.1) with parameters α in the same range.

Let Φ denote the class of c.f. generators ϕ such that $\phi(0) = 1$, $\lim_{t\to\infty} \phi(t) = 0$, and the function $\phi'(t)$ is concave.

Theorem 4.1 *Theorem 3.1 continues to hold if any of the following is satisfied:*

the random vector $(X_1, ..., X_n)$ is a sum of i.i.d. random vectors $(Y_{1j}, ..., Y_{nj})$, $j = 1, ..., k$, where $(Y_{1j}, ..., Y_{nj})$ has an absolutely continuous $\alpha-$symmetric distribution with the c.f. generator $\phi_j \in \Phi$ and the index $\alpha_j \in (1, 2]$. In particular, the theorem holds when the vector of r.v.'s $(X_1, ..., X_n)$ is a sum of i.i.d. random vectors $(Y_{1j}, ..., Y_{nj})$, $j = 1, ..., k$, that have absolutely continuous spherical distributions with c.f. generators $\phi_j \in \Phi$ (the case $\alpha_j = 2$ for all j).

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

Theorem 4.2 *Theorem 3.2 continue to hold if any of the following is satisfied:*

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

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

Theorem 4.3 *Theorem 3.3 continues to hold under any of the following two assumptions:*

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

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

Theorem 4.4 *Theorem 3.4 continues to hold if any of the following assumptions is satisfied:*

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

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

As for generalizations of the main majorization results of the paper and their applications to the case of non-identical distributions, the following conclusions hold.

Let $\sigma_1, ..., \sigma_n \geq 0$ be some scale parameters and let $X_i \sim S_\alpha(\sigma_i, \beta, 0)$, $\alpha \in (0, 2]$, $\beta \in [-1, 1]$, $\beta = 0$ for $\alpha = 1$, be independent non-identically distributed stable r.v.'s. Further, let $\varphi(a, \epsilon)$ denote the tail probability $\varphi(a, \epsilon) = P\big(\sum_{i=1}^n a_{[i]}X_i > \epsilon\big)$, where, as in Section 3, $a_{[1]} \geq ... \geq a_{[n]}$ denote the components of the vector $a = (a_1, ..., a_n) \in \mathbf{R}^n$ in decreasing order.[18] Similar to the proof of Theorems 3.1 and 3.3, one can show that parts (i) of the theorems continue to hold (in the same range of parameters r and α) for the function $\varphi(a, \epsilon)$, $\epsilon > 0$, if $\sigma_1 \geq ... \geq \sigma_n \geq 0$. Similarly, parts (i) of Theorems 3.2 and 3.4 continue to hold for the tail probabilities $\varphi(a, \epsilon)$, if $\sigma_n \geq ... \geq \sigma_1 \geq 0$.[19] Using conditioning arguments, one gets that these extensions also hold in the case of random scale parameters σ_i. The generalizations of the main majorization results also imply analogues of the results of the paper on efficiency comparisons of linear estimators and monotone consistency of the sample mean for dependent non-identically distributed data.

[18] A certain ordering in the components of the vector a is necessary for the extensions of the majorization results in this paper to the case of non-identically distributed r.v.'s X_i since Schur-convexity and Schur-concavity of a function $f(a)$ in a imply its symmetry in the components of a.

[19] These results for $\varphi(a, x)$ can be established in the same way as Theorems 3.3-3.2 using the fact that, by Theorem 3.A.4 in Marshall and Olkin (1979), the function $\chi(c_1, ..., c_n) = \sum_{i=1}^n \sigma_i^\alpha c_{[i]}^\alpha$ is strictly Schur-convex in $(c_1, ..., c_n) \in \mathbf{R}_+^n$ if $\alpha > 1$ and $\sigma_1 \geq ... \geq \sigma_n \geq 0$ and is strictly Schur-concave in $(c_1, ..., c_n) \in \mathbf{R}_+^n$ if $\alpha < 1$ and $\sigma_n \geq ... \geq \sigma_1 \geq 0$.

5 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 3.4, with respective changes in the signs of inequalities. We also provide the complete proof of Theorem 3.2 since 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) (see Remark 3.1).

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 $a = (a_1, ..., a_n) \in \mathbf{R}_+^n$ and $b = (b_1, ..., b_n) \in \mathbf{R}_+^n$ be such that $(a_1^r, ..., a_n^r) \prec (b_1^r, ..., b_n^r)$ and $(a_1^r, ..., a_n^r)$ is not a permutation of $(b_1^r, ..., b_n^r)$ (clearly, $\sum_{i=1}^n a_i \neq 0$ and $\sum_{i=1}^n b_i \neq 0$). Let $X_1, ..., X_n$ be independent r.v.'s such that $X_i \sim S_\alpha(\sigma, \beta, 0)$, $i = 1, ..., n$. It is not difficult to see that if $c = (c_1, ..., c_n) \in \mathbf{R}_+^n$, $\sum_{i=1}^n c_i \neq 0$, then $\sum_{i=1}^n c_i X_i / \left(\sum_{i=1}^n c_i^\alpha\right)^{1/\alpha} \sim S_\alpha(\sigma, \beta, 0)$. Consequently, for $\epsilon > 0$,

$$\psi(c, \epsilon) = P\Big(|X_1| > \epsilon / \Big(\sum_{i=1}^n c_i^\alpha\Big)^{1/\alpha}\Big). \tag{5.1}$$

According to Proposition 3.C.1.a in Marshall and Olkin (1979), the function $\phi(c_1, ..., c_n) = \sum_{i=1}^n c_i^\alpha$ is strictly Schur-convex in $(c_1, ..., c_n) \in \mathbf{R}_+^n$ if $\alpha > 1$ and is strictly Schur-concave in $(c_1, ..., c_n) \in \mathbf{R}_+^n$ 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 (5.1), implies that

$$\psi(a, \epsilon) < \psi(b, \epsilon) \tag{5.2}$$

if $\alpha > r$, and

$$\psi(a, \epsilon) > \psi(b, \epsilon) \tag{5.3}$$

if $\alpha < r$. This completes the proof of parts (i) of the theorems in the case of stable distributions $S_\alpha(\sigma, \beta, 0)$.

Suppose now that $X_1, ..., X_n$ are i.i.d. r.v.'s such that $X_i \sim \underline{\mathcal{CS}}(r)$, $i = 1, ..., n$. By definition of the class $\underline{\mathcal{CS}}(r)$, there exist independent r.v.'s Y_{ij}, $i = 1, ..., n$, $j = 1, ..., k$, such that $Y_{ij} \sim S_{\alpha_j}(\sigma_j, 0, 0)$, $\alpha_j \in (0, r)$, $\sigma_j > 0$, $i = 1, ..., n$, $j = 1, ..., k$, and $X_i = \sum_{j=1}^k Y_{ij}$, $i = 1, ..., n$. By (5.2) and (5.3), for $j = 1, ..., k$, 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 $j = 1, ..., k$,

$$P\Big(\Big|\sum_{i=1}^n a_i Y_{ij}\Big| > \epsilon\Big) > P\Big(\Big|\sum_{i=1}^n b_i Y_{ij}\Big| > \epsilon\Big). \tag{5.4}$$

The r.v.'s Y_{ij}, $i = 1, ..., n$, $j = 1, ..., k$, are symmetric and unimodal by Theorem 2.7.6 in Zolotarev (1986, p. 134). Therefore, from Theorem 1.6 in Dharmadhikari and Joag-Dev (1988, p. 13) it follows that the r.v.'s $\sum_{i=1}^n a_i Y_{ij}$, $j = 1, ..., k$, and $\sum_{i=1}^n b_i Y_{ij}$, $j = 1, ..., k$, are symmetric and unimodal as well. 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 (5.4) and symmetry and unimodality of $\sum_{i=1}^n a_i Y_{ij}$ and $\sum_{i=1}^n b_i Y_{ij}$,

$j = 1, ..., k$, imply, by induction on k (see also Theorem 1 in Birnbaum, 1948, and Theorem 2.C.3 in Dharmadhikari and Joag-Dev, 1988), that $\psi(a, \epsilon) = P\big(\big|\sum_{j=1}^k \sum_{i=1}^n a_i Y_{ij}\big| > \epsilon\big) > P\big(\big|\sum_{j=1}^k \sum_{i=1}^n b_i Y_{ij}\big| > \epsilon\big) = \psi(b, \epsilon)$ for $\epsilon > 0$. Therefore, part (i) of Theorem 3.4 for the class $\underline{\mathcal{CS}}(r)$ holds. Part (i) of Theorem 3.3 for the class $\overline{\mathcal{CS}}(r)$ might be proven in a completely similar way. 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. Parts (i) of Theorem 3.1 for the case of stable i.i.d. r.v.'s $X_i \sim S_\alpha(\sigma, \beta, 0)$, $i = 1, ..., n$, and Theorem 3.2 for both the cases of stable distributions $S_\alpha(\sigma, \beta, 0)$ and distributions from the class $\underline{\mathcal{CS}}(1)$ are immediate consequences of parts (i) of Theorems 3.3 and 3.4 with $r = 1$. Let us prove part (i) of Theorem 3.1 for the case of the class $\overline{\mathcal{CSLC}}$. Let vectors $a = (a_1, ..., a_n) \in \mathbf{R}_+^n$ and $b = (b_1, ..., b_n) \in \mathbf{R}_+^n$ be such that $a \prec b$ and a is not a permutation of b. Let $X_1, ..., X_n$ be i.i.d. r.v.'s such that $X_i \sim \overline{\mathcal{CSLC}}$, $i = 1, ..., n$. By definition, $X_i = \gamma Y_{i0} + \sum_{j=1}^k Y_{ij}$, $i = 1, ..., n$, where $\gamma \in \{0, 1\}$, $k \geq 0$ and $(Y_{1j}, ..., Y_{nj})$, $j = 0, 1, ..., k$, are independent vectors with i.i.d. components such that $Y_{i0} \sim \mathcal{LC}$, $i = 1, ..., n$, and $Y_{ij} \sim S_{\alpha_j}(\sigma_j, 0, 0)$, $\alpha_j \in (1, 2]$, $\sigma_j > 0$, $i = 1, ..., n$, $j = 1, ..., k$. From (5.2) and the results in Proschan (1965) (see Remark 3.1) it follows that, for $j = 0, 1, ..., k$, 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, 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) by induction it follows that the r.v.'s $\sum_{i=1}^n a_i Y_{ij}$ and $\sum_{i=1}^n b_i Y_{ij}$, $j = 0, 1, ..., k$, are symmetric and unimodal. Similar to the proof of Theorems 3.3 and 3.4, by Lemma in Birnbaum (1948) and its proof and induction, this implies that $\sum_{i=1}^n a_i X_i = \gamma \sum_{i=1}^n a_i Y_{i0} + \sum_{j=1}^k \sum_{i=1}^n a_i Y_{ij}$ is strictly more peaked than $\sum_{i=1}^n b_i X_i = \gamma \sum_{i=1}^n b_i Y_{i0} + \sum_{j=0}^k \sum_{i=1}^n b_i Y_{ij}$. This completes the proof of part (i) of Theorem 3.1.

Parts (ii) of Theorems 3.1 and 3.2 follow from their parts (i), majorization comparisons (3.1) and (3.2) and the fact that, under assumptions of Theorem 3.1, $\overline{X}_n$ is weakly consistent for μ.

Proof of Theorem 4.1-4.4. The proof of the theorems follows exactly the same lines as that of Theorems 3.1-3.4 in the paper, using the fact that the property that $\sum_{i=1}^n c_i X_i / \left(\sum_{i=1}^n c_i^\alpha\right)^{1/\alpha}$ has the same distribution as that of X_1 continues to hold if $(X_1, ..., X_n)$ has an $\alpha-$symmetric distribution (see, e.g., Fang, Kotz and Ng, 1990, Ch. 7).

REFERENCES

Adler, R. J., Feldman, R. E. and Taqqu, M. S., eds. (1998). *A practical guide to heavy tails. Statistical techniques and applications.* Boston: Birkhauser.

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.

Axtell, R. L. (2001). Zipf distribution of U.S. firm sizes. *Science* **293**, 1818-1820.

Birnbaum, Z. W. (1948). On random variables with comparable peakedness. *Annals of Mathematical Statistics* **19**, 76-81.

Blattberg, R. C. and Gonedes, R. C. (1974). Comparison of stable and student distributions as statistical models for stock prices. *Journal of Business* **47**, 244-280.

Cambanis, S., Keener, R. and Simons, G. (1983). On $\alpha-$symmetric distributions. *Journal of Multivariate Analysis* **13**, 213-233.

Chan, W., Park, D. H. and Proschan, F. (1989). Peakedness of weighted averages of jointly distributed random variables. In: *Contributions to Probability and Statistics* (Gleser, L. J., Perlman, M. D., Press, S. J. and Sampson, A. R., eds.), 58-62. New York: Springer.

Dharmadhikari, S. W. and Joag-Dev, K. (1988). *Unimodality, convexity and applications.* Boston: Academic Press.

Duffie, D. and Pan, J. (1997). An overview of value at risk. *Journal of Derivatives* **4**, 7-49.

Eaton, M. L. (1970). A note on symmetric Bernoulli random variables. *Annals of Mathematical Statistics* **41**, 1223-1226.

Eaton, M. L. (1988). Concentration inequalities for Gauss-Markov estimators. *Journal of Multivariate Analysis* **25**, 119-138.

Efron, B. (1969). Student's t-test under symmetry conditions. *Journal of the American Statistical Association* **64**, 1278-1302.

Fama, E. (1965). The behavior of stock market prices. *Journal of Business* **38**, 34-105.

Fang, K.-T., Kotz, S. and Ng, K. W. (1990). *Symmetric multivariate and related distributions.* Chapman and Hall, New York.

Fang, H. and Norman, P. (2003). To bundle or not to bundle. *Cowles Foundation Discussion Paper No. 1440*, Yale University, and *SSRI Working Paper No. 2003-18*, Department of Economics, University of Wisconsin-Madison.

Gabaix, X. (1999a). Zipf's law and the growth of cities. *American Economic Review* **89**, 129-132.

Gabaix X. (1999b). Zipf's law for cities: An explanation. *Quarterly Journal of Economics* **114**, 739-767.

Gabaix, X., Gopikrishnan, P., Plerou, V. and Stanley, H. E. (2003). A theory of power-law distributions in financial market fluctuations. *Nature* **423**, 267-270.

Glasserman, P., Heidelberger, P., Shahabuddin, P. (2002). Portfolio value-at-risk with heavy-tailed risk factors. *Mathematical Finance* **12**, 239–269.

Gneiting, T. (1998). On $\alpha-$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.

Jansen, D. W. and De Vries, C. G. (1991). On the frequency of large stock returns: Putting booms and busts into perspective. *Review of Economics and Statistics* **73**,18-32.

Jensen, D. R. (1997). Peakedness of linear forms in ensembles and mixtures. *Statistics and Probability Letters* **35**, 277-282.

Jovanovic, B. and Rob, R. R. (1987). Demand-driven innovation and spatial competition over time. *Review of Economic Studies* **54**, 63-72.

Karlin, S. (1968). *Total positivity.* Stanford University Press.

Karlin, S. (1984). Mathematical models, problems, and controversies of evolutionary theory. *American Mathematical Society. Bulletin. New Series* **10**, 221-274.

Karlin, S. (1992). Stochastic comparisons between means and medians for i.i.d. random variables. In *The art of statistical science. A tribute to G. S. Watson. (Ed. K. V. Mardia).*, Wiley, Chichester.

Loretan, M. and Phillips, P. C. B. (1994). Testing the covariance stationarity of heavy-tailed time series. *Journal of Empirical Finance* **1**, 211-248.

Ma, C. (1998). On peakedness of distributions of convex combinations. *Journal of Statistical Planning and Inference* **70**, 51-56.

Mandelbrot, B. (1963). The variation of certain speculative prices. *Journal of Business* **36**, 394-419.

Mandelbrot, B. (1997). *Fractals and scaling in finance. Discontinuity, concentration, risk.* New York: Springer-Verlag.

Marshall, A. W. and Olkin, I. (1979). *Inequalities: theory of majorization and its applications.* New York: Academic Press.

McCulloch, J. H. (1996). Financial applications of stable distributions. In: *Handbook of statistics* (G. S. Maddala and C. R. Rao, eds.), Vol. 14 (Amsterdam: Elsevier), 393-425.

McCulloch, J. H. (1997). Measuring tail thickness to estimate the stable index alpha: A critique. *Journal of Business and Economic Statistics* **15**, 74-81.

Meerschaert, M. M. and Scheffler, H.-P. (2000). Portfolio modelling with heavy tailed random vectors. In: *Handbook of Heavy-Tailed Distributions in Finance* (S. T. Rachev, ed.), Elsevier North-Holland, New York, 595-640.

Praetz, P. (1972). The distribution of share price changes. *Journal of Business* **45**, 49-55.

Proschan, F. (1965). Peakedness of distributions of convex combinations. *Annals of Mathematical Statistics* **36**, 1703-1706.

Rachev, S. T. and Mittnik, S. (2000). *Stable Paretian Models in Finance.* New York: Wiley.

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.

Zolotarev, V. M. (1986). *One-dimensional stable distributions.* Providence: American Mathematical Society.