

A tail of two tails: peakedness properties in inheritance models of evolutionary theory

Rustam Ibragimov

Department of Economics, Harvard University, Littauer Center,
1805 Cambridge St., Cambridge, MA 02138 (ribragim@fas.harvard.edu)

Abstract. In this paper, 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.

The proof of the results in the paper is based on the general results on majorization properties of heavy-tailed distributions obtained recently in Ibragimov (2004) and also presented in the author's Ph.D. dissertation Ibragimov (2005) and several their extensions derived in this work.

Keywords: Multifactorial inheritance models – Phenotypic traits – Heritability – Human capital.

JEL Classification: C10, C32, I10, Q50, Q57

1 Introduction and discussion of the results

1.1 Motivation and review of the literature

In recent years, there has been a growing interest in inheritance models and related problems in economics. This strand of research is motivated, in part, by significant influence of an individual's genetic endowment on achievable economic outcomes and strong dependence of the distribution of human capital on the distribution of abilities and rates of illness in population and on the transmission of genes through generations (see, among others, Becker, 1993, Ch. 4, 5, Currie, 2000, Frank and McGuire, 2000, Haveman and Wolfe, 2000, Zak, 2002, and references therein). It was demonstrated in a number of studies that heritability is significant and often exceeds environmental effects for many human psychological, psychiatric and neurological phenotypes affecting human capital in an economy (e.g., Ehrman and Parsons, 1977, the contributions in Fuller and Simmel, 1983, Plomin, deFries and McClearn, 1990, Plomin, Owen and McGuffin, 1994, Rowe, 1994, and Gilger, 2000). Significant heritability has been found for such traits as intelligence (IQ), scholastic achievement, risk-taking behavior, learning and learning disabilities, socioeconomic status, memory, nonverbal thinking skills, aggressive behavior, delinquent or criminal behavior, for many behavioral and mental disorders including, e.g., autism, schizophrenia, depression, Alzheimer's disease and reading disability as well as for smoking behavior and drug and alcohol abuse and dependence. Additionally, significant genetic influence was found for a number of medical disorders, for example, for hypertension, ischemic heart disease, tuberculosis, arterial hypertension, bronchial asthma, rheumatoid arthritis, peptic ulcer and epilepsy and predisposition to at least several types of cancer (see Ehrman and Parsons, 1977, Plomin et. al., 1994, Lichtenstein, Holm, Verkasalo, et. al., 2000, and Risch, 2001). Motivated by the analysis of economic effects of behavioral traits and their evolution and intergenerational transmission, many authors have focused on the study of models in economics with altruism, inequity aversion and standards of fairness as well as of evolution of risk attitudes underlying strategic behavior (see Becker, 1974, 1976, Kahneman, Knetsch and Thaler, 1986, Bernheim and Ray, 1987, Simon, 1990, 1993, Samuelson, 1993, Bergstrom, 1995, 2002, Robson, 1995, 2002, and references therein). Moreover, a series of studies in economics analyzed closely related models of intergenerational mobility in various contexts (see, among others, Goldberger, 1989, Peters, 1992, Zimmerman, 1992, Becker, 1993, Ch. 10, and Mulligan, 1999).

Formally, Galtonian-type multifactorial (polygenic) inheritance models (e.g., Karlin, 1984, 1992, and Karlin and Lessard, 1986) with sex- and time-dependent heritability and purely

parental transmission have the form

$$X_{t+1}(\lambda) = \lambda_t X_t^p + (1 - \lambda_t) X_t^m, t = 0, 1, \dots, \quad (1)$$

where X_{t+1} is the trait value of the offspring, X_t^p and X_t^m , $t = 0, 1, 2, \dots$, are, respectively, paternal and maternal contributions and $\lambda = \{\lambda_t\}_{t=0}^\infty$ is a sequence of numbers such that $0 \leq \lambda_t \leq 1$, $t = 0, 1, \dots$ (several patterns of the models considered in this paper are more general than those in Karlin, 1984, 1992, and Karlin and Lessard, 1986, and our notations differ from those in the above works). The values λ_t and $1 - \lambda_t$ are, respectively, paternal and maternal heritability coefficients; we assume that heritability can change with time t .¹

Throughout the paper, we focus on the standard case where X_t^p and X_t^m do not depend on the future values of λ_s , $s = t + 1, t + 2, \dots$. In such a setting, the trait values X_{t+1} depend only on the vectors $\lambda^{(t)} = (\lambda_0, \lambda_1, \dots, \lambda_t)$ of the first t elements in the infinite sequences λ (“histories” of the heritability coefficients λ_s up to time t). To simplify notation, we will continue to use λ to denote the arguments at X_t , as above.

Process (1) with $\lambda_t = 1/2$, $t = 0, 1, 2, \dots$, corresponds to the Galtonian blending model

$$X_{t+1} = (X_t^p + X_t^m)/2, \quad (2)$$

while the case $\lambda_t \neq 1/2$ represents asymmetric transmission from parents.

Let, for $t = 0, 1, 2, \dots$, X_t' and X_t'' denote independent realizations (copies) of the random variable (r.v.) $X_t = X_t(\lambda)$ and let the trait X_0 have a sex-independent distribution in the population at time $t = 0$. In the case where

$$(X_t^p, X_t^m) = (X_t', X_t'') \quad a.s., \quad (3)$$

$t = 0, 1, 2, \dots$, time series (1) model transmission through generations of the trait X_0 with time-dependent and asymmetric heritability. Time series (2), (3) and, more generally, (1), (3) are the main models for transmission of phenotypic traits in humans and other mammals: in the models, the time- $(t + 1)$ offspring receives trait contributions from both parents; the distribution of the trait is assumed to be the same among males and females, as it is the case for most of human phenotypes.² In period $t + 1$, the phenotypic contributions from males and females combine again to propagate into the time- $(t + 2)$ offspring’s trait distribution. The intertemporal propagation of the phenotype X and the main distributional properties of the time series $\{X_t\}$ are determined by the sequence of heritability coefficients $\{\lambda_t\}$. The density f_t of the r.v.’s X_t that follow model (1) and (3) evolves over time according to the following equation: $f_{t+1}(z) = \int_{-\infty}^{\infty} f_t\left(\frac{z-y}{\lambda_t}\right) \left(\frac{1}{\lambda_t}\right) f_t\left(\frac{y}{1-\lambda_t}\right) \left(\frac{1}{1-\lambda_t}\right) dy$.

A problem of interest in inheritance models (1) and, in particular, in models (2) is how the distributional characteristics of the trait X transmit through generations. In particular, the

question as to whether the trait X becomes increasingly more peaked (concentrated) about some value $\mu \in \mathbf{R}$ over time is important - for instance, whether there appears to be a risk group for a trait representing a behavioral or medical disorder or whether genetic diversity or inequality in the distribution of a phenotype affecting human capital in an economy increases with time.

In recent years, a number of studies in human genetics and psychology found departures from normality assumptions in many phenotypic data, including (moderate) thick-tailedness of distribution of many human traits, in particular, of different achievement and psychometric measures (see Micceri, 1989, and the discussion in Allison, Neale, Zannolli, et. al., 1999, and Allison, Fernandez, Heo, et. al., 2000) as well as sex differences in the distribution of extreme outliers for several traits related to, e.g., intellectual abilities (see Hedges and Nowell, 1995). These findings prompted many authors to focus on developing statistical procedures for biometric data robust to non-Gaussianity and heavy-tailedness assumptions, including robust techniques for detection of genes influencing complex quantitative traits (see Allison et. al., 1999, and Allison et. al., 2000, and references therein).

1.2 Discussion of the results

In this paper, we study transmission of the distributional properties of traits through generations in polygenic inheritance models. Motivated by the above-mentioned recent findings of departures from Gaussianity for many phenotypes' distributions, we further focus on the analysis of implications of these models under heavy-tailedness of traits. We obtain results concerning the transmission of peakedness (concentration) properties of fat-tailed traits in general inheritance model (1) with sex- and time-dependent heritability. For instance, from our results it follows that the following conclusions hold (see Theorem 1 and Remark 1 following Theorem 2).

In what follows, we denote by $\bar{\lambda}$ the sequence $\{\lambda_t\}_{t=0}^{\infty}$ with $\lambda_t = 1/2$ for all $t \geq 0$: $\bar{\lambda} = \{1/2, 1/2, \dots\}$. Consider model (1) with the parental contributions given by (3). Let $X_0 - \mu$ have a not extremely heavy-tailed distribution with a finite first moment. More precisely, let the distribution of $X_0 - \mu$ be a convolution of symmetric log-concave distributions and symmetric stable distributions with characteristic exponents in the interval $[1, 2)$ (see Section 2 for definitions of log-concave and stable distributions and a review of their basic properties). Then for all $t = 0, 1, 2, \dots$, the r.v. $X_{t+1}(\lambda)$ is more peaked about μ than is $X_t(\lambda)$, but is less peaked than is $X_{t+1}(\bar{\lambda})$. That is, $P(|X_{t+1}(\bar{\lambda}) - \mu| > x) \leq P(|X_{t+1}(\lambda) - \mu| > x) \leq P(|X_t(\lambda) - \mu| > x)$ for all $x \geq 0$ and all $t = 0, 1, 2, \dots$. Suppose now that the distribution of $X_0 - \mu$ is extremely thick-tailed and has an infinite mean. More precisely, suppose that the distribution of $X_0 - \mu$

is a convolution of symmetric stable distributions with characteristic exponents in the interval $(0, 1]$. Then for all $t = 0, 1, 2, \dots$, the r.v. $X_{t+1}(\lambda)$ is less peaked about μ than is $X_t(\lambda)$, but is more peaked than is $X_{t+1}(\bar{\lambda})$. That is, $P(|X_t(\lambda) - \mu| > x) \leq P(|X_{t+1}(\lambda) - \mu| > x) \leq P(|X_{t+1}(\bar{\lambda}) - \mu| > x)$ for all $x \geq 0$ and all $t = 0, 1, 2, \dots$

According to the above results, if the initial distribution of the trait X (say, a behavioral or medical disorder or an ability for which heritability is significant) in the population is not extremely heavy-tailed and has a finite mean, then the trait distribution becomes increasingly more peaked over time. Roughly speaking, concentration of the distribution of the disorder about some risk group in the population and inequality in the distribution of the ability becomes increasingly more pronounced. Furthermore, at any given time, peakedness of the trait is maximal (the spread of the trait in the population is minimal) in the case of symmetric heritability. In the case of a trait with an extremely heavy-tailed initial distribution with an infinite first moment (say, a medical or behavioral disorder for which there is no strongly expressed risk group or a relatively equally distributed ability with significant genetic influence), the situation is reversed: the trait distribution becomes less peaked with time and increasingly more spread in the population. Moreover, peakedness of the trait is minimal (the spread of the trait in the population is maximal) in the case of symmetric heritability.

Similarly, we obtain analogues of our results on multifactorial inheritance models in a more general setting with traits' distributions given by convolutions of a wide class of transforms of stable r.v.'s.

The proof of the results in this paper is based on general results on peakedness properties of convolutions of distributions and majorization phenomena for tail probabilities of linear combinations of r.v.'s presented in Appendix A1. These properties and phenomena were first analyzed, under the assumptions of log-concavity of distributions, in the seminal paper by Proschan (1965) that found applications in the study of many problems in statistics, econometrics, economic theory, mathematical biology and other fields (see the discussion in Ibragimov, 2004, 2005). The proof of the main results in this paper is based on analogues of the results in Proschan (1965) in the case of heavy-tailed distributions recently obtained by Ibragimov (2004) and also presented in Ibragimov (2005). To our knowledge, the results in Ibragimov (2004, 2005) are the first ones in the literature that give extensions of those in Proschan (1965) for the paradigm of thick-tailedness and also show that general majorization properties of convex combinations of symmetric log-concavely distributed r.v.'s derived by Proschan (1965) are reversed for certain wide classes of distributions (see the discussion in Ibragimov, 2004, 2005). These results provide the key to the analysis of inheritance models under traits' heavy-tailedness and to obtaining contrasting results for the classes of not extremely thick-tailed and extremely long-tailed phenotypes, similar to the results on robustness vs. reversals of properties of many

of economic models in Ibragimov (2004) and Chapter 1 in Ibragimov (2005).

Besides the analysis of multifactorial inheritance models considered in this paper, the majorization results obtained in Ibragimov (2004, 2005) have many other applications. These applications include the analysis of models of environmental sex determination, the study of efficiency of linear estimators and the robustness of the model of demand-driven innovation and spatial competition over time, portfolio value at risk analysis as well as the study of optimal strategies for a multiproduct monopolist providing interrelated goods.³

The paper is organized as follows: Section 2 contains notations and definitions of classes of distributions used throughout the paper and reviews their basic properties. In Section 3, we present the main results on the properties of polygenic inheritance models under heavy-tailedness of traits' distributions. Appendix A1 reviews peakedness properties of log-concavely distributed r.v.'s derived by Proschan (1965) and their analogues for thick-tailed distributions obtained in Ibragimov (2004). Finally, Appendix A2 contains proofs of the main results obtained in the paper.

2 Notations and classes of distributions

In this section, we introduce certain classes of distributions we will be dealing with throughout the paper. The notations for these classes are similar to those in Ibragimov (2004).

We say that 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 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 $\lambda \in [0, 1]$,

$$f(\lambda x_1 + (1 - \lambda)x_2) \geq f^\lambda(x_1)f^{1-\lambda}(x_2). \quad (4)$$

(see An, 1998). A distribution is said to be log-concave if its density f satisfies (4).

Examples of log-concave distributions include (see, for instance, Marshall and Olkin, 1979, p. 493) the normal distribution $\mathcal{N}(\mu, \sigma^2)$, the uniform density $\mathcal{U}(\theta_1, \theta_2)$, the exponential density, the logistic distribution, 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). This implies, in particular, that distributions with log-concave densities *cannot* be used to model heavy-tailed phenomena.

As in Ibragimov (2004), we denote by \mathcal{LC} the class of symmetric log-concave distributions (\mathcal{LC} stands for “log-concave”).

In the studies based on models incorporating fat-tailed r.v.’s, it is usually assumed that the distributions of the r.v.’s belong to the class of stable laws. Although there are several alternatives to the stable modeling of heavy-tailed data, focusing on stable distribution models is justified in many cases and has a number of advantages, as discussed in, e.g., Adler, Feldman and Gallagher, 1998. In particular, the statistical methods for stable laws work as well for the data in the domain of attraction of stable distributions. Furthermore, stable laws and the long-tailed distributions in the domain of their attraction behave similarly at the tails of the distributions which is usually the region of interest for heavy-tailed techniques. Finally, there are few reliable approaches available in the case of heavy-tailed r.v.’s not in a stable domain of attraction (Adler, Feldman and Gallagher, 1998).

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

$$E(e^{ixX}) = \begin{cases} \exp\{i\mu x - \sigma^\alpha |x|^\alpha (1 - i\beta \operatorname{sign}(x) \tan(\pi\alpha/2))\}, & \alpha \neq 1, \\ \exp\{i\mu x - \sigma |x| (1 + (2/\pi)i\beta \operatorname{sign}(x) \ln|x|)\}, & \alpha = 1, \end{cases} \quad (5)$$

$x \in \mathbf{R}$, where $\operatorname{sign}(x)$ is the sign of x defined by $\operatorname{sign}(x) = 1$ if $x > 0$, $\operatorname{sign}(0) = 0$ and $\operatorname{sign}(x) = -1$ otherwise. For a detailed review of properties of stable distributions the reader is referred to, e.g., the monograph by Zolotarev (1986).

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). 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. In particular, if X has the stable distribution $S_\alpha(\sigma, \beta, \mu) : X \sim S_\alpha(\sigma, \beta, \mu)$, then there exists a constant $C > 0$ such that

$$\lim_{x \rightarrow +\infty} x^\alpha P(|X| > x) = C. \quad (6)$$

This implies that the p -th absolute moments $E|X|^p$ of a r.v. $X \sim S_\alpha(\sigma, \beta, \mu)$, $\alpha \in (0, 2)$ 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 . 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 standard deviation 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 $a_i \geq 0$, $i = 1, \dots, n$,

$$\sum_{i=1}^n a_i X_i / \left(\sum_{i=1}^n a_i^\alpha \right)^{1/\alpha} \sim S_\alpha(\sigma, \beta, \mu). \quad (7)$$

Further, we consider the class $\overline{\mathcal{CS}}$ of distributions which are convolutions of symmetric stable distributions $S_\alpha(\sigma, 0, 0)$ with characteristic exponents $\alpha \in [1, 2]$ and $\sigma > 0$ (here and below, \mathcal{CS} stands for ‘‘convolutions of stable’’; the overline indicates relation to stable distributions with indices of stability *greater* than the threshold value 1). That is, $\overline{\mathcal{CS}}$ consists of distributions of r.v.'s X such that, for some $k \geq 1$, $X = Y_1 + \dots + Y_k$, where Y_i , $i = 1, \dots, k$, are independent r.v.'s such that $Y_i \sim S_{\alpha_i}(\sigma_i, 0, 0)$, $\alpha_i \in (1, 2]$, $\sigma_i > 0$, $i = 1, \dots, k$.

By $\overline{\mathcal{CSLC}}$, we denote the class of convolutions of distributions from the classes \mathcal{LC} and $\overline{\mathcal{CS}}$. 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 (\mathcal{CSLC} stands for ‘‘convolutions of stable and log-concave’’). 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 $\overline{\mathcal{CS}}$.

$\underline{\mathcal{CS}}$ stands for the class of distributions which are convolutions of symmetric stable distributions $S_\alpha(\sigma, 0, 0)$ with indices of stability $\alpha \in (0, 1)$ and $\sigma > 0$ (the underline indicates relation to stable distributions with indices of stability *less* than the threshold value 1). That is, $\underline{\mathcal{CS}}$ consists of distributions of r.v.'s X such that, for some $k \geq 1$, $X = Y_1 + \dots + Y_k$, where Y_i , $i = 1, \dots, k$, are independent r.v.'s such that $Y_i \sim S_{\alpha_i}(\sigma_i, 0, 0)$, $\alpha_i \in (0, 1)$, $\sigma_i > 0$, $i = 1, \dots, k$.

Let $\mathbf{R}_+ = [0, \infty)$. Throughout the paper, \overline{M} denotes the class of differentiable odd functions $f : \mathbf{R} \rightarrow \mathbf{R}$ such that f is concave and increasing on \mathbf{R}_+ and \underline{M} denotes the class of odd functions $f : \mathbf{R} \rightarrow \mathbf{R}$ such that f is convex and increasing on \mathbf{R}_+ .

By $\overline{\mathcal{CTSLC}}$, we denote the class of convolutions of log-concave distributions and distributions of transforms $f(Y)$, $f \in \overline{M}$, of symmetric stable r.v.'s $Y \sim S_\alpha(\sigma, 0, 0)$ with characteristic exponents $\alpha \in (1, 2]$ and $\sigma > 0$ (\mathcal{CTSLC} stands for ‘‘convolutions of transforms of stable and log-concave’’). That is, $\overline{\mathcal{CTSLC}}$ consists of distributions of r.v.'s X such that, for some $k \geq 1$,

$$X = \gamma Y_0 + f_1(Y_1) + \dots + f_k(Y_k), \quad (8)$$

where $\gamma \in \{0, 1\}$, $f_i \in \overline{M}$, $i = 1, \dots, k$, and Y_i , $i = 0, 1, \dots, k$, are independent r.v.'s such that $Y_0 \sim \mathcal{LC}$ and $Y_i \sim S_{\alpha_i}(\sigma_i, 0, 0)$, $\alpha_i \in (1, 2]$, $\sigma_i > 0$, $i = 1, \dots, k$.

We note that (see Ibragimov, 2004) the class $\overline{\mathcal{CS}}$ of *convolutions* of symmetric stable distributions with *different* indices of stability $\alpha \in (1, 2]$ is wider than the class of *all* symmetric stable distributions $S_\alpha(\sigma, 0, 0)$ with $\alpha \in (1, 2]$ and $\sigma > 0$. Similarly, the class $\underline{\mathcal{CS}}$ is wider than the class of *all* symmetric stable distributions $S_\alpha(\sigma, 0, 0)$ with $\alpha \in (0, 1)$ and $\sigma > 0$.

Clearly, one has $\mathcal{LC} \subset \overline{\mathcal{CSLC}}$, $\overline{\mathcal{CS}} \subset \overline{\mathcal{CSLC}}$ and $\overline{\mathcal{CSLC}} \subset \overline{\mathcal{CTSLC}}$. Note also 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 $\sigma > 0$.

In some sense, symmetric (about 0) Cauchy distributions $S_1(\sigma, 0, 0)$ are at the dividing boundary between the classes $\underline{\mathcal{CS}}$ and $\overline{\mathcal{CSLC}}$.

In what follows, we write $X \sim \mathcal{LC}$ (resp., $X \sim \overline{\mathcal{CSLC}}$, $X \sim \underline{\mathcal{CS}}$ or $X \sim \overline{\mathcal{CTSLC}}$) if the distribution of the r.v. X belongs to the class \mathcal{LC} (resp., $\overline{\mathcal{CSLC}}$, $\underline{\mathcal{CS}}$ or $\overline{\mathcal{CTSLC}}$). In addition to that, the notation $X =^d Y$ for two r.v.'s X and Y will mean that their distributions are the same.

3 Main results

The following concept of peakedness of r.v.'s was introduced by Birnbaum (1948).

Definition 1 (Birnbaum, 1948, see also Proschan, 1965, and Marshall and Olkin, 1979, p. 372). *A r.v. X is more peaked about $\mu \in \mathbf{R}$ than is Y , written $Y \leq_\mu^p X$, if $P(|X - \mu| > x) \leq P(|Y - \mu| > x)$ for all $x \geq 0$. If these inequalities are strict whenever the two probabilities are not both 0 or both 1, then the r.v. X is strictly more peaked about μ than is Y , written $Y <_\mu^p X$.*

In the case $\mu = 0$, we simply say that the r.v. X is more peaked (strictly more peaked) than Y and write $Y \leq^p X$ ($Y <^p X$).

Roughly speaking, a r.v. X is more peaked about $\mu \in \mathbf{R}$ than is Y , if the distribution of X is more concentrated about μ than is that of Y .

Theorem 1 below provides results on the peakedness properties of the distribution of the trait $\{X_t\}$ in general model (1) with the parental contributions determined by (3) and sex- and time-dependent heritability. Let, as in the introduction, for $t = 0, 1, 2, \dots$, $\lambda^{(t)} = (\lambda_0, \lambda_1, \dots, \lambda_t)$ be the vectors of ‘‘histories’’ of the coefficients in model (1) up to time t . Further, for $t = 0, 1, 2, \dots$, denote $\bar{\lambda}^{(t)} = (1/2, 1/2, \dots, 1/2) \in \mathbf{R}^{t+1}$.

Theorem 1 *Consider model (1) with the parental contributions determined by (3). Let $\mu \in \mathbf{R}$, $t \in \{0, 1, 2, \dots\}$ and let $\lambda_t \notin \{0, 1\}$ and $\lambda^{(t)} \neq \bar{\lambda}^{(t)}$. If $X_0 \sim S_\alpha(\sigma, \beta, \mu)$ for some $\sigma > 0$, $\beta \in [-1, 1]$*

and $\alpha \in (1, 2]$, or $X_0 = \mu + Y$, where $Y \sim \overline{\mathcal{CSLC}}$, then $X_t(\lambda) \prec_{\mu}^p X_{t+1}(\lambda) \prec_{\mu}^p X_{t+1}(\bar{\lambda})$. That is,

$$P(|X_{t+1}(\bar{\lambda}) - \mu| > x) < P(|X_{t+1}(\lambda) - \mu| > x) < P(|X_t(\lambda) - \mu| > x) \quad (9)$$

for all $x > 0$. If $X_0 \sim S_{\alpha}(\sigma, \beta, \mu)$ for some $\sigma > 0$, $\beta \in [-1, 1]$ and $\alpha \in (0, 1)$, or $X_0 = \mu + Y$, where $Y \sim \underline{\mathcal{CS}}$, then $X_{t+1}(\bar{\lambda}) \prec_{\mu}^p X_{t+1}(\lambda) \prec_{\mu}^p X_t(\lambda)$. That is,

$$P(|X_t(\lambda) - \mu| > x) < P(|X_{t+1}(\lambda) - \mu| > x) < P(|X_{t+1}(\bar{\lambda}) - \mu| > x) \quad (10)$$

for all $x > 0$.

According to the following theorem, in the case of Galtonian blending model (2) with symmetric heritability, peakedness comparisons in (9) continue to hold in the case of the more general class $\overline{\mathcal{CTSLLC}}$ of thick-tailed initial distributions of the phenotype X than the class $\overline{\mathcal{CSLC}}$ in Theorem 1.

Theorem 2 Consider model (2) with the parental contributions determined by (3). Let $\mu \in \mathbf{R}$ and $t \in \{0, 1, 2, \dots\}$. If $X_0 = \mu + Y$, where $Y \sim \overline{\mathcal{CTSLLC}}$, then $X_t(\bar{\lambda}) \prec_{\mu}^p X_{t+1}(\bar{\lambda})$. That is,

$$P(|X_{t+1}(\bar{\lambda}) - \mu| > x) < P(|X_t(\bar{\lambda}) - \mu| > x) \quad (11)$$

for all $x > 0$.

Let us emphasize again the meaning of the results given by Theorems 1 and 2. According to comparisons (9) and (11), if the initial distribution of the trait X in the population is less heavy-tailed than Cauchy distribution (and thus, in particular, the first moment of X_0 is finite), then the trait distribution becomes increasingly more peaked and concentrated about μ over time. Moreover, at any given time t , asymmetry in heritability decreases peakedness and concentration of the trait. According to inequalities (10) the situation is reversed in the case of traits with initial distributions that have tails thicker than Cauchy and thus have infinite means. In the setting with such an extremely heavy-tailed initial density, the trait distribution becomes less peaked and concentrated with time and increasingly more spread in the population. Moreover, peakedness and concentration of the trait increases with asymmetry in heritability.

Remark 1. From Remark 2 in Appendix A1 and the proof of the theorems in this section it follows that Theorem 1 continues to hold for convolutions of the distributions in the classes $\underline{\mathcal{CS}}$ and $\overline{\mathcal{CSLC}}$ with symmetric Cauchy distributions $S_1(\sigma, 0, 0)$; Theorem 2 continues to hold for (two-fold) convolutions of distributions in the class $\overline{\mathcal{CTSLLC}}$ with the distributions of transforms $X_0 \sim f(Y_0)$ of symmetric Cauchy r.v.'s $Y_0 \sim S_1(\sigma, 0, 0)$, where $f \in \overline{M}$ is strictly concave on \mathbf{R}_+ .

Appendix A1: Majorization properties of log-concave and heavy-tailed distributions

For a vector $a \in \mathbf{R}^n$, denote by $a_{[1]} \geq \dots \geq a_{[n]}$ its components in decreasing order.

Definition 2 (Marshall and Olkin, 1979). Let $a, b \in \mathbf{R}^n$. The vector a is said to be majorized by the vector b , 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 . In this context, it is easy to see that, for all $a \in \mathbf{R}_+^n$, the following relations hold:

$$\left(\sum_{i=1}^n a_i/n, \dots, \sum_{i=1}^n a_i/n \right) \prec (a_1, \dots, a_n) \prec \left(\sum_{i=1}^n a_i, 0, \dots, 0 \right). \quad (12)$$

Definition 3 (Marshall and Olkin, 1979). 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 .

Proschan (1965) obtains the following seminal result concerning majorization properties of tail probabilities of linear combinations of log-concavely distributed r.v.'s:

Proposition 1 (Proschan, 1965). If X_1, \dots, X_n are i.i.d. symmetric log-concavely distributed r.v.'s, then the function $\psi(a, x) = P(\sum_{i=1}^n a_i X_i > x)$ is strictly Schur-convex in $a = (a_1, \dots, a_n) \in \mathbf{R}_+^n$ for $x > 0$ and is strictly Schur-concave in $a = (a_1, \dots, a_n) \in \mathbf{R}_+^n$ for $x < 0$.

Clearly, from Proposition 1 it follows that $\sum_{i=1}^n b_i X_i <^p \sum_{i=1}^n a_i X_i$ if $a \prec b$ and a is not a permutation of b .

Proschan (1965) notes that Proposition 1 also holds for (two-fold) convolutions of log-concave distributions with symmetric Cauchy distributions and obtained results on peakedness properties of averages $(f(Y_1) + f(Y_2))/2$ of transforms of symmetric Cauchy r.v.'s Y_1 and Y_2 for $f \in \underline{M}$ and $f \in \overline{M}$ (see Lemmas 2.7 and 2.8 in Proschan, 1965).

The following Lemmas 1 and 2 concerning general majorization properties of arbitrary convex combinations of heavy-tailed r.v.'s were obtained in Ibragimov (2004) (see Theorems 4.3 and 4.4 and Remark 4.1 in that paper) and also presented in the author's Ph.D. dissertation

Ibragimov (2005). According to Lemma 1, peakedness properties of linear combinations of r.v.'s with not extremely heavy-tailed distributions are the same as in the case of log-concave distributions in Proschan (1965).

Lemma 1 (Ibragimov, 2004). *Proposition 1 holds if X_1, \dots, X_n are i.i.d r.v.'s such that $X_i \sim S_\alpha(\sigma, \beta, 0)$, $1 \leq i \leq n$, for some $\sigma > 0$, $\beta \in [-1, 1]$ and $\alpha \in (1, 2]$, or $X_i \sim \overline{\mathcal{CSLC}}$, $1 \leq i \leq n$.*

According to Lemma 2, the peakedness properties given by Proposition 1 and Theorem 1 above are reversed in the case of r.v.'s with very heavy-tailed distributions, as modeled by convolutions of stable distributions with indices of stability not greater than one.

Lemma 2 (Ibragimov, 2004). *If X_1, \dots, X_n are i.i.d. r.v.'s such that $X_i \sim S_\alpha(\sigma, \beta, 0)$, $1 \leq i \leq n$, for some $\sigma > 0$, $\beta \in [-1, 1]$ and $\alpha \in (0, 1)$, or $X_i \sim \underline{\mathcal{CS}}$, $1 \leq i \leq n$, then the function $\psi(a, x)$ in Proposition 1 is strictly Schur-concave in $(a_1, \dots, a_n) \in \mathbf{R}_+^n$ for $x > 0$ and is strictly Schur-convex in $(a_1, \dots, a_n) \in \mathbf{R}_+^n$ for $x < 0$.*

The following lemmas generalize Lemmas 2.7 and 2.8 in Proschan (1965) and provide new results on peakedness properties of averages of transforms of arbitrary stable r.v.'s and their convolutions. For r.v.'s X_1, X_2, \dots , and $n \geq 1$, we denote by \overline{X}_n the sample mean $\overline{X}_n = (1/n) \sum_{i=1}^n X_i$ (in particular, \overline{X}_2 denotes $\overline{X}_2 = (X_1 + X_2)/2$).

Lemma 3 *If X_1 and X_2 are i.i.d r.v.'s such that $X_i = f(Y_i)$, $i = 1, 2$, where $f \in \overline{M}$ and $Y_i \sim S_\alpha(\sigma, \beta, 0)$, $\alpha \in (1, 2]$, $\sigma > 0$, $\beta \in [-1, 1]$, then $P(\overline{X}_2 > x) < P(X_1 > x)$ for $x > 0$ and $P(\overline{X}_2 > x) > P(X_1 > x)$ for $x < 0$. If $n = 2^k$, $k \geq 1$, and X_1, \dots, X_n are i.i.d r.v.'s such that $X_i \sim \overline{\mathcal{CTSLC}}$, $1 \leq i \leq n$, then $\overline{X}_{n/2} <^p \overline{X}_n$, that is, $P(|\overline{X}_n| > x) < P(|\overline{X}_{n/2}| > x)$ for all $x > 0$.*

Lemma 4 *If X_1 and X_2 are i.i.d r.v.'s such that $X_i = f(Y_i)$, $i = 1, 2$, where $f \in \underline{M}$ and $Y_i \sim S_\alpha(\sigma, \beta, 0)$, $\alpha \in (0, 1)$, $\sigma > 0$, $\beta \in [-1, 1]$, then $P(X_1 > x) < P(\overline{X}_2 > x)$ for $x > 0$ and $P(X_1 > x) > P(\overline{X}_2 > x)$ for $x < 0$.*

Remark 2. If r.v.'s X_1, \dots, X_n have a symmetric Cauchy distribution $S_1(\sigma, 0, 0)$ (with $\alpha = 1$) which is, as discussed in Section 2, exactly at the dividing boundary between the class $\overline{\mathcal{CSLC}}$ in Theorem 1 and the class $\underline{\mathcal{CS}}$ in Theorem 2, then the function $\psi(a, x)$ in the theorems depends only on $\sum_{i=1}^n a_i$ and x and so is *both* Schur-concave and Schur-convex in $a \in \mathbf{R}_+^n$ for all $x \in \mathbf{R}$ (see Proschan, 1965, and Remark 4.1 in Ibragimov, 2004). As noted in Ibragimov (2004), this implies that Theorems 1 and 2 continue to hold for convolutions of distributions from the classes

$\overline{\mathcal{CSLC}}$ and $\underline{\mathcal{CS}}$ with symmetric Cauchy distributions. As follows from Proschan (1965), Lemma 3 holds for i.i.d. r.v.'s X_1, X_2, \dots such that $X_i = f(Y_i)$, $i \geq 1$, where Y_i have the Cauchy ($\alpha = 1$) distribution $Y_i \sim S_1(\sigma, 0, 0)$ and $f \in \overline{M}$ is strictly concave on \mathbf{R}_+ , and Lemma 4 holds for i.i.d. r.v.'s X_1, X_2 such that $X_i = f(Y_i)$, $i = 1, 2$, where $Y_i \sim S_1(\sigma, 0, 0)$ and $f \in \underline{M}$ is strictly convex on \mathbf{R}_+ . As in Proschan (1965), this implies that Lemma 3 continues to hold for convolutions of distributions from the class $\overline{\mathcal{C}\mathcal{T}\mathcal{S}\mathcal{L}\mathcal{C}}$ with the distributions of transforms $f(Y_i)$, $f \in \underline{M}$, of symmetric Cauchy r.v.'s $Y_i \sim S_1(\sigma, 0, 0)$, where f is strictly concave on \mathbf{R}_+ .

4 Conclusion

As demonstrated in the paper, the evolution of peakedness and concentration properties of traits in multifactorial inheritance models depends crucially on heavy-tailedness of the traits' initial distributions. In this work, we focused on the analysis of multifactorial inheritance models under the assumption of purely parental transmission of phenotypes through generations. However, as indicated before the approach developed in the paper is also directly applicable, in particular, in the study of inheritance models with threshold-type propagation of traits, e.g., in the analysis of polygenic or temperature-dependent sex determination with heavy-tailed sex-determining traits; the results in this direction are presented in Chapter 2 of the author's dissertation Ibragimov (2005). In addition, the methods developed in this paper can also be applied in the study of multifactorial inheritance models that involve both parental and environmental contributions as well as of models of intergenerational mobility. Furthermore, from the extensions of the main majorization results in the case of dependence in Ibragimov (2004, 2005) it follows that the analogues of the results in this paper hold as well for inheritance models with wide classes of dependence structures for paternal contributions X_t^p and X_t^m and environmental shocks. Generalizations of the results in this work to the above settings are left for further research.

Appendix A2: Proofs

Proof of Theorems 1 and 2. Let $X_0 \sim S_\alpha(\beta, \sigma, \mu)$ for some $\sigma > 0$, $\beta \in [-1, 1]$ and $\alpha \in (0, 1)$ or $X_0 = \mu + Y$, where $Y \sim \underline{\mathcal{CS}}$. For $t = 0, 1, \dots$, denote $N_t = 2^t$, $0^{(t)} = (0, 0, \dots, 0) \in \mathbf{R}^{N_t}$ and $\bar{a}^{(t)} = (1/N_t, 1/N_t, \dots, 1/N_t) \in \mathbf{R}^{N_t}$. Let us define recursively the following vectors. Set $a^{(0)} = (1) \in \mathbf{R}^1$. For $t = 1, 2, \dots$, let $b^{(t)} = (a^{(t-1)}, 0^{(t-1)}) \in \mathbf{R}^{N_t}$ and $c^{(t)} = (0^{(t-1)}, a^{(t-1)}) \in \mathbf{R}^{N_t}$ be the vectors with the components $b_i^{(t)} = a_i^{(t-1)}$, $c_i^{(t)} = 0$, $i = 1, 2, \dots, N_{t-1}$; $b_i^{(t)} = 0$, $c_i^{(t)} = a_{i-N_{t-1}}^{(t-1)}$, $i = N_{t-1} + 1, \dots, N_t$. Moreover, let $a^{(t)} = \lambda_t b^{(t)} + (1 - \lambda_t) c^{(t)}$ be the vector with the

components $a_i^{(t)} = \lambda_t b_i^{(t)} = \lambda_t a_i^{(t-1)}$, $i = 1, 2, \dots, N_{t-1}$, $a_i^{(t)} = (1 - \lambda_t) c_i^{(t)} = (1 - \lambda_t) a_{i-N_{t-1}}^{(t-1)}$, $i = N_{t-1} + 1, \dots, N_t$. Let Y_1, Y_2, \dots, Y_{N_t} be independent copies of the r.v. X_0 . Denote $Y^{(t)} = (Y_1, Y_2, \dots, Y_{N_t})$. It is easy to see that if, for $t \geq 1$, $Y_1^{(t-1)}$ and $Y_2^{(t-1)}$ are independent copies of $Y^{(t-1)}$, then $\lambda_t a^{(t-1)}(Y_1^{(t-1)})' + (1 - \lambda_t) a^{(t-1)}(Y_2^{(t-1)})'$ has the same distribution as $a^{(t)}(Y^{(t)})'$. By induction, this implies that, for $t = 0, 1, 2, \dots$, $X_t(\bar{\lambda})$ is distributed as $\bar{a}^{(t)}(Y^{(t)})'$ and, similarly, the distributions of $X_t(\lambda)$ and $a^{(t)}(Y^{(t)})'$ are the same:

$$X_t(\bar{\lambda}) =^d \bar{a}^{(t)}(Y^{(t)})', \quad (13)$$

$$X_t(\lambda) =^d a^{(t)}(Y^{(t)})'. \quad (14)$$

Since for $t = 1, 2, \dots$, $\sum_{i=1}^{N_t} a_i^{(t)} = \lambda_t \sum_{i=1}^{N_{t-1}} a_i^{(t-1)} + (1 - \lambda_t) \sum_{i=1}^{N_{t-1}} a_i^{(t-1)} = \sum_{i=1}^{N_{t-1}} a_i^{(t-1)} = \dots = a_1^{(0)} = 1$ ($a_1^{(0)} = 1$ is the only component of the vector $a^{(0)}$), from relations (12) in Appendix A1 it follows that

$$\bar{a}^{(t)} \prec a^{(t)}. \quad (15)$$

Since the components of the vector $c^{(t)}$ are permutations of those of $b^{(t)}$, one has $c^{(t)} \prec b^{(t)}$. Further, evidently, $b^{(t)} \prec b^{(t)}$. 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), from the above majorization comparisons we get

$$a^{(t)} = \lambda_t b^{(t)} + (1 - \lambda_t) c^{(t)} \prec b^{(t)}. \quad (16)$$

Lemma 2 in Appendix A1 and relations (13) and (15) imply that, for all $t = 0, 1, 2, \dots$, and all $x > 0$,

$$\begin{aligned} P(|X_t(\lambda) - \mu| > x) &= P(|a^{(t)}(Y^{(t)})' - \mu| > x) < \\ P(|\bar{a}^{(t)}(Y^{(t)})' - \mu| > x) &= P(|X_t(\bar{\lambda}) - \mu| > x), \end{aligned} \quad (17)$$

if $\lambda^{(t)} \neq \bar{\lambda}^{(t)}$. Similarly, from Lemma 2 and relations (14) and (16) it follows that, for all $t = 0, 1, 2, \dots$, and $x > 0$,

$$\begin{aligned} P(|X_{t+1}(\lambda) - \mu| > x) &= P(|a^{(t+1)}(Y^{(t+1)})' - \mu| > x) > P(|b^{(t+1)}(Y^{(t+1)})' - \mu| > x) = \\ P(|(a^{(t)}, 0^{(t)})(Y^{(t+1)})' - \mu| > x) &= P(|a^{(t)}(Y^{(t)})' - \mu| > x) = P(|X_t(\lambda) - \mu| > x), \end{aligned} \quad (18)$$

$\lambda_t \notin \{0, 1\}$. Relations (17) and (18) thus imply that inequalities (10) hold. Inequalities (9) might be proven in a similar way, with the use of Lemma 1 instead of Lemma 2. Thus, Theorem 1 holds. Using Lemma 3 instead of Lemma 2, we obtain Theorem 2. The proof is complete.

Proof of Lemmas 3 and 4. Let $\alpha_1 \in (1, 2]$, $\alpha_2 \in (0, 1)$, and let $f_1 \in \overline{M}$, $f_2 \in \underline{M}$. For $j = 1, 2$, let $Y_1^{(j)}$ and $Y_2^{(j)}$ be i.i.d. r.v.'s such that $Y_i^{(j)} \sim S_{\alpha_j}(\sigma, \beta, 0)$, $\sigma > 0$, $\beta \in [-1, 1]$, $i = 1, 2$, and let $X_i^{(j)} = f_j(Y_i^{(j)})$, $i = 1, 2$, $j = 1, 2$. As in the proof of Lemmas 2.7 and 2.8 in Proschan (1965), we have $|f_1((y_1 + y_2)/2)| \geq |(f_1(y_1) + f_1(y_2))/2|$, $|f_2((y_1 + y_2)/2)| \leq |(f_2(y_1) + f_2(y_2))/2|$ for all $y_1, y_2 \in \mathbf{R}$. Since the functions $|f_j(x)|$ are increasing in $|x|$, we get that $|f_1((y_1 + y_2)/2^{1/\alpha_1})| \geq |(f_1(y_1) + f_1(y_2))/2|$, $|f_2((y_1 + y_2)/2^{1/\alpha_2})| \leq |(f_2(y_1) + f_2(y_2))/2|$, with strict inequalities for $y_1 + y_2 \neq 0$. Since, by (7), $2^{-1/\alpha_1}(Y_1^{(1)} + Y_2^{(1)}) \sim S_{\alpha_1}(\sigma, \beta, 0)$ and the function f_1 is odd, this implies that, for all $x > 0$, $P((X_1^{(1)} + X_2^{(1)})/2 > x) = P((f_1(Y_1^{(1)}) + f_1(Y_2^{(1)}))/2 > x) < P(f_1((Y_1^{(1)} + Y_2^{(1)})/2^{1/\alpha_1}) > x) = P(f_1(Y_1^{(1)}) > x) = P(X_1^{(1)} > x)$ and, for all $x < 0$, $P((X_1^{(1)} + X_2^{(1)})/2 > x) = 1 - P((f_1(Y_1^{(1)}) + f_1(Y_2^{(1)}))/2 < x) > 1 - P(f_1((Y_1^{(1)} + Y_2^{(1)})/2^{1/\alpha_1}) < x) = 1 - P(f_1(Y_1^{(1)}) < x) = P(X_1^{(1)} > x)$. Similarly, we get that $P((X_1^{(2)} + X_2^{(2)})/2 > x) < P(X_1^{(2)} > x)$ for all $x > 0$ and $P((X_1^{(2)} + X_2^{(2)})/2 > x) > P(X_1^{(2)} > x)$ for all $x < 0$. This proves Lemma 4 and the first part of Lemma 3.

Let now $n = 2^k$, $k \geq 2$, and let X_1, \dots, X_n be i.i.d. r.v.'s such that $X_i \sim \overline{\mathcal{CTSLC}}$, $1 \leq i \leq n$. By definition of the class $\overline{\mathcal{CTSLC}}$, there exist i.i.d. r.v.'s Y_{ij} , $j = 0, 1, \dots, k$, $i = 1, \dots, n$, and functions $f_j \in \overline{M}$, $j = 1, \dots, k$, such that $Y_{i0} \sim \mathcal{LC}$ and $Y_{ij} \sim S_{\alpha_j}(\sigma_j, 0, 0)$, $\alpha_j \in (1, 2]$, $\sigma_j > 0$, $j = 1, \dots, k$, and $X_i = \gamma Y_{i0} + f_1(Y_{i1}) + \dots + f_k(Y_{ik})$, $\gamma \in \{0, 1\}$, $i = 1, \dots, n$. From the above we have that for all $i = 1, \dots, n/2$ and $j = 1, \dots, k$, $f_j(Y_{ij}) <^p (f_j(Y_{ij}) + f_j(Y_{n/2+i,j}))/2$. In addition, by Proposition 1, $Y_{i0} <^p (Y_{i0} + Y_{n/2+i,0})/2$.

According to Theorem 2.7.6 in Zolotarev (1986, p. 134) and Theorem 1.10 in Dharmadhikari and Joag-Dev (1988, p. 20), the densities of the r.v.'s Y_{ij} , $j = 0, 1, \dots, k$, $i = 1, \dots, n$, are symmetric and unimodal. This implies, as it is not difficult to see, symmetry and unimodality of the densities of the r.v.'s $f_j(Y_{ij})$, $f_j \in \overline{M}$, $j = 1, \dots, k$, $i = 1, \dots, n$. By Theorem 1.6 in Dharmadhikari and Joag-Dev (1988, p. 13), we get, in turn, that the densities of the r.v.'s $(Y_{i0} + Y_{n/2+i,0})/2$ and $(f_j(Y_{ij}) + f_j(Y_{n/2+i,j}))/2$, $j = 1, \dots, k$, $i = 1, \dots, n$, are symmetric and unimodal.

From Lemma in Birnbaum (1948) and its proof it follows that if V_1, V_2 and W_1, W_2 are independent absolutely continuous symmetric unimodal r.v.'s such that $W_i \leq^p V_i$, $i = 1, 2$, then $W_1 + W_2 \leq^p V_1 + V_2$; furthermore, this peakedness comparison is strict if $W_1 <^p V_1$ or $W_2 <^p V_2$. This implies by induction (see also Theorem 1 in Birnbaum, 1948, and Theorem 2.C.3 in Shaked and Shanthikumar, 1994) that

$$\overline{X}_{n/2} = (2/n) \sum_{i=1}^{n/2} [\gamma Y_{i0} + f_1(Y_{i1}) + \dots + f_k(Y_{ik})] <^p$$

$$(1/n) \sum_{i=1}^{n/2} [\gamma (Y_{i0} + Y_{n/2+i,0}) + (f_1(Y_{i1}) + f_1(Y_{n/2+i,1})) \dots + (f_k(Y_{ik}) + f_k(Y_{n/2+i,k}))] = \overline{X}_n.$$

This completes the proof of Lemma 3.

5 Acknowledgements

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. 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 thank two anonymous referees, Donald Brown, Aydin Cecen, Gary Chamberlain, Brian Dineen, Darrell Duffie, Xavier Gabaix, Philip Haile, Samuel Karlin, Alex Maynard, Ingram Olkin, Ben Polak, Gustavo Soares and Kevin Song for helpful comments. I am also grateful to the participants at seminars at Yale University, University of British Columbia, California Institute of Technology, the University of California at San Diego, Columbia University, Harvard University, the London School of Economics and Political Science, Massachusetts Institute of Technology, the Université de Montréal, McGill University, New York University and Nuffield College, University of Oxford, 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 useful discussions. The financial support from the Yale University Graduate Fellowship and the Cowles Foundation Prize is gratefully acknowledged.

Notes

¹More general analogues of models (1) include, in addition to parental contributions X_t^p and X_t^m , an independent residual (environmental) contribution ϵ_t : $X_{t+1}(\lambda, \kappa) = \lambda_t X_t^p + \kappa_t X_t^m + (1 - \lambda_t - \kappa_t)\epsilon_t$, where $\lambda_t, \kappa_t \geq 0$ and $\lambda_t + \kappa_t \leq 1$. In the case of symmetric and time-independent heritability, $\lambda_t = \kappa_t = h^2/2$, $t = 0, 1, 2, \dots$, where h^2 is the heritability coefficient of Galton (1886) given by the coefficient at the regression of an offspring on the midparent value at an equilibrium (see Roughgarden, 1979, Ch. 9, Bulmer, 1980, Ch. 6, and Becker, 1993, Ch. 10). The above extensions of (1) with $\lambda_t = 0$, $t = 0, 1, 2, \dots$ (or with $X_t^p = X_t^m = X_t$ a.s., $t = 0, 1, 2, \dots$) have exactly the same form as the models of intergenerational mobility $X_{t+1} = h^2 X_t + \epsilon_t$, $t = 0, 1, 2, \dots$, where X_t is the measure of economic status such as earnings or income and ϵ_t is an independent error term; in mathematical evolutionary theory, such settings represent intergenerational transmission of a phenotype maternally affecting itself, see Roff (1997, pp. 250-254)

²One should note that, in contrast to humans and other mammals, the assumption that the paternal and maternal phenotypic contributions X_t^p and X_t^m have the same distributions is not appropriate for a number of other species. For example, in several reptile species sex determination mechanism is temperature dependent: the sex of an embryo is determined by incubation temperature (see Bull, 1981, Cherfas and Gribbin, 1985, Ch. 5, Bull and Charnov, 1989, and Janzen and Paukstis, 1991). In many turtles embryos hatch as males in cool and as females in warm conditions, with a sharp transition from all-male to all-female broods. Alligators, crocodiles and some lizards exhibit the opposite pattern in sex determination: males develop at warm and females at cool

temperatures. The inheritance mechanisms where an offspring sex is determined by environmental conditions after conception are referred to as environmental mechanisms of sex determination (e.g., Bulmer and Bull, 1982, Karlin, 1984, Karlin and Lessard, 1986, and Janzen and Paukstis, 1991).

It is interesting that some theories have suggested that environmental sex-determination could have been the cause of dinosaur extinction. If sex determination mechanism in dinosaurs was temperature-dependent, like in modern reptiles, then they might have gone extinct because one sex was no longer produced due a major temperature change on Earth (Cherfas and Gribbin, 1985, Ch. 5). One should note here that, since the length of the temperature interval at which both sexes are produced might be as small as 8 degrees C, as in the case of turtles, a relatively small change in environmental conditions might be sufficient for extinction of some species. This is particularly important for conservation of threatened species living today (see Cherfas and Gribbin, 1985).

Traditionally (see Bulmer and Bull, 1982, Karlin, 1984, and Karlin and Lessard, 1986), environmental sex determination is modeled by time series (2) with the parental contributions X_t^p and X_t^m given by independent r.v.'s with the non-identical cdf's $P(X_t^p \leq x) = P(X_t \leq x | X_t > K)$, $P(X_t^m \leq x) = P(X_t \leq x | X_t \leq K)$, $K \in \mathbf{R}$, $t = 0, 1, \dots$. The settings considered by Bulmer and Bull (1982), Karlin (1984) and Karlin and Lessard (1986) thus model a situation where a sex response trait is determined by a continuous phenotype or environmental variable X (such as size, fitness, exposure to sunlight, food resources, temperature, humidity, etc.). An individual with $X = \tilde{x}$ becomes a male if the value of \tilde{x} is greater than the threshold level K , and a female otherwise.

The above models of threshold sex-determination can also be used as first approximations in the analysis of the part of the variation of sex ratio in humans controlled by parental hormonal levels as well as in the study of the properties of extreme cases of dependence of sex determination on such traits as, e.g., socioeconomic status or parental income, with complete segregation in the choice of sex of the offspring (in these cases, the threshold values K represent cut-off points of the parental hormonal levels, wealth or income). It is important to emphasize here that the dependence of the sex ratio of the offspring on the hormone levels of parents, socioeconomic status and related traits was found in many papers in the literature (see the discussion in Ibragimov, 2005, for details).

³The following list summarizes some of other applications of the main majorization results in Ibragimov (2004) presented in the author's Ph.D. dissertation Ibragimov (2005).

(i) From the majorization results obtained in Ibragimov (2004, 2005) it follows 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, 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. The main majorization results in Ibragimov (2004, 2005) also provide sharp concentration inequalities for linear estimators as well as their extensions to the case of wide classes of dependent data.

(ii) Using the general majorization results, 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.

(iii) 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. Char-

acterizations 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.

(iv) Another application of the main majorization results explored in depth in Ibragimov (2004, 2005) 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.

(v) We also provide the study of the intergenerational transmission of the sex ratio in models of threshold (e.g., polygenic or temperature-dependent) sex determination with long-tailed sex-determining traits. Among other results, we show that if the distribution of the sex determining trait is not very thick-tailed, then several properties of these models are the same as in the case of log-concave densities analyzed by Karlin (1984, 1992). In particular, the excess of males (females) among parents leads to the same pattern for the population of the offspring. Thus, the excess of one sex over the other one accumulates with time and the sex ratio in the total alive population cannot stabilize at the balanced sex ratio value of $1/2$. We further show that the above properties are reversed for sufficiently heavy-tailed distributions of sex determining traits. In such settings, the sex ratio of the offspring oscillates around the balanced sex ratio value and an excess of males (females) in the initial period leads to an excess of females (males) offspring next period. Therefore, the sex ratio in the total living population can, in fact, stabilize at $1/2$. Interestingly, these results are related, in particular, to the analysis of correlation between human sex ratios and socioeconomic status of parents as well as to the study of the variation of the sex ratio due to parental hormonal levels.

REFERENCES

- Adler, R. J., Feldman, R. E. and Gallagher, C. (1998) Analysing stable time series. In *A practical guide to heavy tails. Statistical techniques and applications* (eds R. J. Adler, R. E. Feldman and M. S. Taqqu), pp. 133-158. Boston: Birkhauser.
- Allison, D. B., Fernandez J. R., Heo M. and Beasley T. M. (2000) Testing the robustness of the new Haseman-Elston quantitative-trait loci-mapping procedure. *American Journal of Human Genetics* **67**, 249-252.
- Allison, D. B., Neale, M. C., Zannolli, R., Schork, N. J., Amos, C. I. and Blangero, J. (1999) Testing the robustness of the likelihood-ratio test in a variance-component quantitative-trait loci-mapping procedure. *American Journal of Human Genetics* **65**, 531-544.
- An, M. Y. (1998) Logconcavity versus logconvexity: a complete characterization. *Journal of Economic Theory* **80**, 350-369.
- Becker, G. S. (1974) A theory of social interactions. *Journal of Political Economy* **82**, 1063-1091.
- Becker, G. S. (1976) Altruism, egoism and genetic fitness: economics and sociobiology. *Journal of Economic Literature* **14**, 817-826.

- Becker, G. S. (1993) *Human capital: a theoretical and empirical analysis, with special reference to education*. Chicago: Chicago University Press.
- Bergstrom, T. C. (1995) On the evolution of altruistic ethical rules for siblings. *American Economic Review* **85**, 58-81.
- Bergstrom, T. C. (2002) Evolution of social behavior: Individual and group selection. *Journal of Economic Perspectives* **16**, 67-88.
- Bernheim, B. D. and Ray, D. (1987) Economic growth with intergenerational altruism. *Review of Economic Studies* **54**, 227-241.
- Birnbaum, Z. W. (1948) On random variables with comparable peakedness. *Annals of Mathematical Statistics* **19**, 76-81.
- Bull, J. J. (1981) Evolution of environmental sex determination from genotypic sex determination. *Heredity* **47**, 173-184.
- Bull, J. J. and Charnov, E. L. (1989) Enigmatic reptilian sex ratios. *Evolution* **43**, 1561-1566.
- Bulmer, M. G. (1980) *The mathematical theory of quantitative genetics*. New York: Oxford University Press.
- Bulmer, M. G. and Bull, J. J. (1982) Models in polygenic sex determination and sex ratio evolution. *Evolution* **36**, 13-26.
- Cherfas, J. and Gribbin, J. (1985) *The redundant male*. New York: Pantheon Books.
- Currie, J. (2000) Child health in developed countries. In *Handbook of health economics* (eds A. J. Culyer and J. P. Newhouse), vol. 1B, pp. 1053-1139. New York: Elsevier.
- Dharmadhikari, S. W. and Joag-Dev, K. (1988): *Unimodality, convexity and applications*. Boston: Academic Press.
- Ehrman, L. and Parsons, P. A. (1977) *The genetics of behavior*. Sunderland: Sinauer Associates.
- Frank, R. G. and McGuire, T. G. (2000) Economics and mental health. In *Handbook of health economics* (eds A. J. Culyer and J. P. Newhouse), vol. 1B, pp. 892-954. New York: Elsevier.
- Fuller, J. L. and Simmel, E. C., eds. (1983) *Behavior genetics: principles and applications*. Hillsdale: Lawrence Erlbaum Associates.
- Galton, F. (1886) Regression towards mediocrity in hereditary stature. *Journal of the Anthropological Institute of Great Britain and Ireland* **15**, 246-263.
- Gilger, J. W. (2000) Contributions and promise of human behavioral genetics. *Human Biology* **72**, 229-255.
- Goldberger, A. S. (1989) Economic and mechanical models of intergenerational transmission. *American Economic Review* **79**, 504-513.
- Haveman, R. and Wolfe, B. (2000) The economics of disability and disability policy. In *Handbook of health economics* (eds A. J. Culyer and J. P. Newhouse), vol. 1B, pp. 995-1051. New York: Elsevier.

- Hedges, L. V. and Nowell, A. (1995) Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science* **269**, 41-45.
- Ibragimov, R. (2004) On the robustness of economic models to heavy-tailedness assumptions. Mimeo, Yale University. Available at <http://post.economics.harvard.edu/faculty/ibragimov/Papers/HeavyTails.pdf>
- Ibragimov, R. (2005). *New majorization theory in economics and martingale convergence results in econometrics*. Ph.D. dissertation, Yale University.
- Janzen, F. J. and Paukstis, G. L. (1991) Environmental sex determination in reptiles: ecology, evolution, and experimental design. *The Quarterly Review of Biology* **66**, 149-179.
- Jovanovic, B. and Rob, R. R. (1987). Demand-driven innovation and spatial competition over time. *Review of Economic Studies* **54**, 63-72.
- Kahneman, D., Knetsch, J. L. and Thaler, R. (1986). Fairness as a constraint on profit seeking: entitlements in the market. *American Economic Review* **76**, 728-741.
- 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.
- Karlin, S. and Lessard, S. (1986) *Theoretical studies on sex ratio evolution*. Princeton: Princeton University Press.
- Lichtenstein, P., Holm, N. V., Verkasalo, P. K., Iliadou, A., Kaprio, J., Koskenvuo, M., Pukkala, E., Skytthe, A. and Hemminki, K. (2000) Environmental and heritable factors in the causation of cancer analyses of cohorts of twins from Sweden, Denmark, and Finland. *The New England Journal of Medicine* **343**, 78-85.
- Marshall, A. W. and Olkin, I. (1979) *Inequalities: theory of majorization and its applications*. New York: Academic Press.
- Micceri, T. (1989) The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin* **105**, 156-166.
- Mulligan, C. B. (1999) Galton versus the human capital approach to inheritance. *Journal of Political Economy* **107**, S184-S224.
- Peters, H. E. (1992) Patterns of intergenerational mobility in income and earnings. *Review of Economics and Statistics* **74**, 456-466.
- Plomin, R., deFries, J. C. and McClearn, G. E. (1990) *Behavioral genetics: a primer*. New York: W. H. Freeman.
- Plomin, R., Owen, M. J. and McGuffin, P. (1994) The genetic basis of complex behaviors. *Science* **264**, 1733-1739.
- Proschan, F. (1965) Peakedness of distributions of convex combinations. *Annals of Mathematical Statistics* **36**, 1703-1706.

- Risch, N. (2001) The genetic epidemiology of cancer: Interpreting family and twin studies and their implications for molecular genetic approaches. *Cancer epidemiology, biomarkers & prevention* **10**, 733-741.
- Robson, A. J. (1995) The evolution of strategic behavior. *Canadian Journal of Economics - Revue Canadienne D'Economique* **28**, 17-41.
- Robson, A. J. (2002) Evolution and human nature. *Journal of Economic Perspectives* **16**, 89-106.
- Roff, D. A. (1997) *Evolutionary quantitative genetics*. New York: Chapman & Hall.
- Roughgarden, J. (1979) *Theory of population genetics and evolutionary ecology. An introduction*. New York: Macmillan.
- Rowe, D. C. (1994) *The limits of family influence: genes, experience and behavior*. New York: Guilford Press.
- Samuelson, P. A. (1993) Altruism as a problem involving group versus individual selection in economics and biology. *American Economic Review* **83**, 143-148.
- Shaked, M. and Shanthikumar, J. G. (1994). *Stochastic orders and their applications*. Probability and Mathematical Statistics. Academic Press, Boston, MA.
- Simon, H. A. (1990) Mechanism for social selection and successful altruism. *Science*, **250**, 1665-1668.
- Simon, H. A. (1993) Altruism and economics. *American Economic Review* **83**, 156-161.
- Zak, P. J. (2002) Genetics, family structure, and economic growth. *Journal of Evolutionary Economics* **12**, 343-365.
- Zimmerman, D. J. (1992) Regression toward mediocrity in economic stature. *American Economic Review* **82**, 409-429.
- Zolotarev, V. M. (1986) *One-dimensional stable distributions*. Providence: American Mathematical Society.