

ON THE ROBUSTNESS OF LOCATION ESTIMATORS IN MODELS OF FIRM GROWTH UNDER HEAVY-TAILEDNESS ¹

Running title: Robustness of location estimators and heavy-tailedness

By Rustam Ibragimov²

Department of Economics, Harvard University

October 2006

Address for manuscript correspondence:

Rustam Ibragimov
Department of Economics
Harvard University
Littauer Center
1805 Cambridge St.
Cambridge, MA 02138

Email: ribragim@fas.harvard.edu

Phone: +1-617-496-4795

Fax: +1-617-495-7730

¹The results in this paper constitute a part of the author's dissertation "New majorization theory in economics and martingale convergence results in econometrics" presented to the faculty of the Graduate School of Yale University in candidacy for the degree of Doctor of Philosophy in Economics in March, 2005. Some of the results were originally contained in the work circulated in 2003-2006 under the titles "On the robustness of economic models to heavy-tailedness assumptions" and "Demand-driven innovation and spatial competition over time under heavy-tailed signals".

²I am indebted to my advisors, Donald Andrews, Peter Phillips and Herbert Scarf, for all their support and guidance in all stages of the current project. I also thank Donald Brown, Aydin Cecen, Gary Chamberlain, Paul Embrechts, Frank Fabozzi, Xavier Gabaix, Wolfgang Härdle, Samuel Karlin, Benoît Mandelbrot, Ingram Olkin, Gustavo Soares, Kevin Song, Johan Walden and the participants at seminars at the Departments of Economics at Yale University, University of British Columbia, the University of California at San Diego, Harvard University, the London School of Economics and Political Science, Massachusetts Institute of Technology, the Université de Montréal, McGill University and New York University, the Division of the Humanities and Social Sciences at California Institute of Technology, Nuffield College, University of Oxford, and the Department of Statistics at Columbia University as well as the participants at the 18th New England Statistics Symposium at Harvard University, April 2004, the International Conference on Stochastic Finance, Lisbon, Portugal, September 2004, and Deutsche Bundesbank Conference on Heavy Tails and Stable Paretian Distributions in Finance and Macroeconomics, Eltville, Germany, November 2005, for helpful comments and discussions.

ABSTRACT

Focusing on the model of demand-driven innovation and spatial competition over time, we study the effects of the robustness of estimators employed by firms to make inferences about their markets on the firms' growth patterns. We show that if consumers' signals in the model are not extremely thick-tailed and the firms use the sample mean of the signals to estimate the ideal product, then the firms' output levels exhibit positive persistence. In such a setting, large firms have an advantage over their smaller counterparts. These properties are reversed for signals with extremely heavy-tailed distributions. In such a case, the model implies anti-persistence in output levels, together with a surprising pattern of oscillations in firm sizes, with smaller firms being likely to become larger ones next period, and vice versa. We further show that the implications of the model under not extreme heavy-tailedness continue to hold under the only assumption of symmetry of consumers' signals if the firms use a more robust estimator of the ideal product, the sample median.

KEYWORDS: robustness, location estimators, heavy-tailed distributions, demand-driven innovation, spatial competition, firm growth, signals, investment, information, sample mean, sample median, majorization

JEL Classification: C13, C22, D83, D92

1 Introduction and discussion of the results

1.1 Objectives

The goal of this paper is to demonstrate that robustness of statistical procedures to heavy-tailedness can have important effects on the properties of firm growth models. Focusing on the model of demand-driven innovation and spatial competition over time, we show that the firms' growth patterns depend crucially on the degree of heavy-tailedness of consumers' signals and on the choice of estimators employed by the firms to make inferences about their markets. If consumers' signals in the model are extremely heavy-tailed and the firms use the sample means of the signals as product designs, then the firms' output levels exhibit anti-persistence and smaller firms have an advantage over their larger counterparts. These properties are reversals of those that hold under not extremely thick-tailed signals or in the case when the firms switch to more robust estimators of the ideal product, such as sample medians, in the presence of extreme heavy-tailedness.

1.2 Output persistence and demand-driven innovation and spatial competition over time

Since the seminal work of Nelson and Plosser (1982), many studies in economics have focused on models that could account for positive persistence in levels of output, among other “stylized facts” on output dynamics. Most of the models proposed in this stream of literature focus primarily on technology shocks as the driving force of economic fluctuations and usually rely on capital accumulation, intertemporal substitution, capital irreversibility or different types of capital adjustments costs or lags as sources of shock propagation to generate persistence.

Jovanovic and Rob (1987), hereafter JR, develop a model of demand-driven innovation and spatial competition over time in which the source of output persistence is, in contrast, private information alone. The model is based on the idea that larger firms get better information about their markets. The firms choose their products and then make output decisions based on how successful their product design is (in terms of the closeness to the ideal product). In the model, output decision has two effects. One is to maximize contemporaneous profits. The other is that output generates signals and thus information about the next period's ideal product. The greater is the output the more signals are likely to be received regarding the next period's ideal product and more information about the firm's market is likely to be collected.

JR further focus on the analysis of the properties of the model in the case where the distribution of consumers' signals is log-concave which implies that the tails of signals' distributions decline at least exponentially fast and, thus, the distributions are extremely light-tailed, see An

(1998) and Section 2 in this paper. From the results in JR it follows that if the signals are log-concavely distributed and the firms use the sample mean of consumers' signals to estimate the ideal product (the center of the signals' distribution) and choose it as the product design, the model implies positive persistence in output levels. Furthermore, in such a setting, large firms always have an advantage over their smaller counterparts. More precisely, according to JR, under the above assumptions, the model has the following properties: the probability of rank reversals in adjacent periods (that is, the probability of the smaller of two firms becoming the larger one next period) is always less than one half; this probability diminishes as the current size-difference increases; and the distribution of future size is stochastically increasing as a function of current size. The intuition for the results is that the larger is a firm's size, the greater is the amount of information the firm gets. The larger firms that learn more are thus more likely to come up with a successful product.

One should note here that the analysis of (arbitrary) log-concavely distributed signals in JR implicitly makes the assumption that the firms choose the sample mean of consumers' signals as the product design. This is because the optimal product design in the setting employed in JR is the posterior median of the ideal product θ given a sample of signals \mathcal{S} (see equation (6) and the discussion at the beginning of Section 5 in JR). Although the posterior median coincides with the sample mean in the case of normal signals and diffuse priors, it is not the case in general, see, e.g., Subsections 4.2.1, 4.2.3 and 4.3.1 in Berger (1985), Subsections 3.1 and 3.2 in Box and Tiao (1973) and Chapter 2 and Section 4.2 in Carlin and Louis (2000) (under the normality assumption for the sample of observed signals \mathcal{S} and a diffuse prior for θ , the sample mean of signals in \mathcal{S} is the posterior median, mode and mean of θ).

1.3 Heavy-tailedness

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

$$P(|X| > x) \asymp x^{-\alpha} \tag{1}$$

(here, $g(x) \asymp h(x)$ denotes that there are constants, c and C such that $0 < c \leq g(x)/h(x) \leq C < \infty$ for large $x > 0$). 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 Gabaix, Gopikrishnan, Plerou and Stanley, 2003; Loretan and Phillips, 1994; Ibragimov, 2005; Rachev, Menn and Fabozzi, 2005,

and references therein).³ 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); $\alpha \approx 3$ (Gabaix et al., 2003). Power laws (1) with $\alpha \approx 1$ (Zipf laws) have been found to hold for sizes of firms and largest mutual funds (see Axtell, 2001; Gabaix et al., 2003) and city sizes (see Gabaix, 1999*a,b*, for the discussion and explanations of the Zipf law for cities).

De Vany and Walls (2004) show that stable distributions with tail indices $1 < \alpha < 2$ provide a good model for distributions of profits in motion pictures. Chapter 11 in Rachev et al. (2005) discusses and reviews the vast literature that supports heavy-tailedness and the stable Paretian hypothesis (with $1 < \alpha < 2$) for equity and bond return distributions. Some studies indicate that the tail exponent is close to one or 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). Scherer, Harhoff and Kukies (2000) and Silverberg and Verspagen (2004) report the tail indices α to be considerably less than one for financial returns from technological innovations. As discussed by Nešlehova, Embrechts and Chavez-Demoulin (2006), tail indices less than one are observed for loss distributions of a number of operational risks.

Many works in econometrics, statistics and finance have focused on the analysis of robust estimation procedures and financial models, including those that perform well in the presence of heavy-tailedness and extreme outliers (see, among others, the reviews in Huber, 1981, Rousseeuw and Leroy, 1987, Huber, 1996, and Jurečková and Sen, 1996; the papers in Maddala and Rao, 1997; Part II in Fabozzi, Focardi and Kolm, 2006, and references therein).

1.4 Main results of the paper on the robustness to heavy-tailedness

The fact that a number of economic and financial time series have the tail exponents of approximately equal to or (slightly or even substantially) less than one discussed in the previous subsection is important in the context of the results in this paper. As we demonstrate, the value of the tail index $\alpha = 1$ (that is, non-existence of the first moment) is exactly the critical boundary between robustness of implications of the model of demand-driven innovation and spatial competition over time to heavy-tailedness assumptions and their reversals. According to the results

³Several 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 modeling based on certain above distributions is still under way in empirical literature.

obtained in this paper, the implications of the model are robust to thick-tailedness assumptions with tail indices $\alpha > 1$ (Theorem 1). But its conclusions are reversed for extremely heavy-tailed distributions with $\alpha < 1$ and infinite first moments (Theorem 2).

We prove *inter alia* that if consumers' signals are independent and extremely thick-tailed and the firms choose the sample mean of the signals as the product design then relatively large firms are not likely to stay larger and the model thus implies anti-persistence in output levels. In this case, a surprising pattern of oscillations in firm sizes emerges, with smaller firms being likely to become larger ones next period, and vice versa. Moreover, it is likely that very small firms will become very large next period, and the size of very large firms will shrink to very small.

More precisely, under the above assumptions, the probability of rank reversals in adjacent periods (that is, the probability of the smaller of the two firms becoming the larger one next period) is always greater than one half; this probability increases as the current size-difference increases; and the distribution of future size is stochastically decreasing as a function of current size.

Essentially, in the case of extremely heavy-tailed signals, smaller firms, in fact, have an advantage over their larger counterparts if the sample mean is employed as the product design. The driving force for this conclusion is that in the presence of extremely heavy-tailed shocks, the sample mean of signals is not informative about the ideal product (population center) θ since the sample of signals is very likely to contain extreme outliers. Sensitivity of the sample mean to the presence of extreme outliers also implies, according to our results, that if consumers' signals are extremely long-tailed, then it is optimal for the firms to switch to employing more robust estimators of the next period's product such as the sample median.

The assumption that the sample mean of signals is employed to approximate the ideal product (estimate the population center) and is chosen as the product design in the case of extremely thick-tailed signals is appropriate in the setting where the firms do not realize that they are in the presence of extreme heavy-tailedness and utilize the same inference methods as in the case of distributions with light tails. The firms might not be able to make inferences about thick-tailedness of consumers' signals on their markets because of time or data availability constraints. The presence of heavy-tailedness and extreme signals, together with constraints on making inferences about it, is likely to be the case for industries with very uncertain consumer perception of new products or constantly changing environments and new industries in which business decisions on the basis of former experience are impossible and the risk facing the firms is higher than in other sectors. Many high-tech industries, together with the Net economy, exhibit the above patterns. The results in this paper provide new insights concerning firm size and growth patterns in such settings. In particular, the rapid rise of Internet businesses during the late 1990's and their sudden fall following an extreme event, the fall of NASDAQ by 10% in April, 2000, might illustrate the oscillation patterns in the firm sizes predicted by the results for growth models with extremely heavy-tailed signals obtained in this paper.

As follows from the results in Theorem 3 in the paper, if the firms know that they are in the presence of extreme heavy-tailedness and employ robust inference methods, namely, use the sample median instead of the sample mean as the product design, then the counterintuitive conclusions discussed above disappear. According to Theorem 3, if the sample median is employed as the product design, then larger firms have an advantage over their smaller counterparts in the case of *arbitrary symmetric* consumers' signals. That is, in any such setting, the implications of the model of demand-driven innovation and spatial competition over time for the sample mean and log-concavely distributed signals in JR continue to hold.

The results obtained in this paper highlight, therefore, the necessity of making inferences about the presence or absence of heavy-tailedness and extreme outliers before making business decisions, if possible, and of employing robust estimation methods, such as the use of the sample median instead of the sample mean in the presence of thick-tailed signals. According to the results, having more information is always advantageous if robust inference methods are employed; this advantage, however, can be completely lost and even become a disadvantage if the decisions are made using non-robust estimators in the presence of extreme heavy-tailedness.

1.5 Extensions

The proof of the results in JR discussed in Subsection 1.1 is based on majorization properties of log-concave distributions obtained in the seminal work by Proschan (1965). The proof of the main results in this paper uses the analogues of the majorization results in Proschan (1965) in the case of heavy-tailedness obtained recently in Ibragimov (2005, 2007) (see Appendix A1).⁴ The proof of the results in the paper on firm growth with sample medians employed as product designs is based on peakedness properties of these estimators presented in Karlin (1992).

As follows from the extensions of the majorization results in Appendix A1 to the dependent case obtained in Ibragimov (2005, 2007), the main results of this paper also hold for convolutions of consumers' signals with joint α -symmetric distributions that exhibit both heavy-tailedness and dependence.⁵ Convolutions of α -symmetric distributions contain, as subclasses, convolutions of certain models with common shocks affecting all heavy-tailed signals (such as macroe-

⁴Besides the analysis of firm growth models for firms that invest into information about their markets considered in this paper and the study of efficiency for linear estimators under heavy-tailedness in Ibragimov (2007), the majorization results obtained in Ibragimov (2005, 2007) have several other applications, including the study of value at risk models for portfolios of thick-tailed risks, optimal bundling strategies for a multiproduct monopolist and the analysis of inheritance models (see Ibragimov, 2005).

⁵An n -dimensional distribution is called α -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 $\alpha > 0$ (see Fang, Kotz and Ng, 1990, Ch. 7). Such distributions should not be confused with multivariate spherically symmetric stable distributions, which have characteristic functions $\exp[-\lambda(\sum_{i=1}^n t_i^2)^{\beta/2}]$, $0 < \beta \leq 2$. Obviously, spherically symmetric stable distributions are particular examples of α -symmetric distributions with $\alpha = 2$ (that is, of spherical distributions) and $\phi(x) = \exp(-x^\beta)$.

conomic or political ones, see the discussion in Andrews, 2005) as well as spherical distributions which are α -symmetric with $\alpha = 2$. Spherical distributions, in turn, include such examples as Kotz type, multinormal, logistic and multivariate α -stable distributions. In addition, they include a subclass of mixtures of normal distributions as well as multivariate t -distributions that were used in the literature to model heavy-tailedness phenomena with dependence and finite moments up to a certain order.

The results for independent signals with tail indices $\alpha < 1$ discussed in Subsection 1.4 continue to hold for dependent signals that have α -symmetric distributions with $\alpha < 1$. Similarly, the results in the paper for independent signals with tail exponents $\alpha > 1$ also hold for α -symmetric distributions with indices α in the same range. As discussed in Remark 2, using the majorization results established in Ibragimov and Walden (2006*a,b*), the above conclusions can be further extended to the case of bounded signals with distributions generated by truncated α -symmetric densities as well as to the case of signals with distributions in the domain of attraction of a stable law.

1.6 Organization of the paper

The paper is organized as follows: Section 2 introduces notation and definitions of classes of distributions used throughout the paper and reviews their basic properties. To make the paper self-contained, in Sections 3 and 4 we review the setup of the model of demand-driven innovation and spatial competition over time and its properties under log-concavity of signals' distributions derived in JR. Section 5 presents the main results of the paper on the robustness of growth theory for firms investing into information about their markets to heavy-tailedness of consumers' signals. In Section 6, we make some concluding remarks. Appendix A1 discusses majorization properties of linear combinations of log-concavely distributed r.v.'s used in JR to establish the results discussed in Subsection 1.4. The appendix also presents the analogs of the majorization phenomena in the case of heavy-tailed distributions obtained in Ibragimov (2005, 2007) and the results on majorization properties of sample medians in Karlin (1992) that are needed for the proof of the main results in Section 5. Appendix A2 contains proofs of the results obtained in the paper.

2 Notation and definitions

We say that a r.v. X (a distribution) 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 (resp., log-concave) if $\log f(x)$ is concave in $x \in \Omega$, that is, if for all $x_1, x_2 \in \Omega$, and any $\lambda \in [0, 1]$, $f(\lambda x_1 + (1 - \lambda)x_2) \geq (f(x_1))^\lambda (f(x_2))^{1-\lambda}$ (see An, 1998). Examples of log-concave distributions include (see, for instance, Marshall and Olkin, 1979, p. 493) the normal distribution, the uniform density, the exponential density, the

Gamma distribution $\Gamma(\alpha, \beta)$ with the shape parameter $\alpha \geq 1$, the Beta distribution $\mathcal{B}(a, b)$ with $a \geq 1$ and $b \geq 1$; the Weibull distribution $\mathcal{W}(\gamma, \alpha)$ with the shape parameter $\alpha \geq 1$.

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

Throughout the paper, \mathcal{LC} denotes the class of symmetric log-concave distributions (\mathcal{LC} stands for “log-concave”).

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

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

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

As is well-known, a closed form expression for the density $f(x)$ of the distribution $S_\alpha(\sigma, \beta, \mu)$ is available in the following cases (and only in those cases): $\alpha = 2$ (Gaussian distributions); $\alpha = 1$ and $\beta = 0$ (Cauchy distributions with densities $f(x) = \sigma/(\pi(\sigma^2 + (x - \mu)^2))$, $\sigma > 0$, $\mu \in \mathbf{R}$); $\alpha = 1/2$ and $\beta = \pm 1$ (Lévy distributions with densities $f(x) = (\sigma/(2\pi))^{1/2} \exp(-\sigma/(2x)) x^{-3/2}$, $x \geq 0$; $f(x) = 0$, $x < 0$, where $\sigma > 0$, and their shifted versions). Degenerate distributions correspond to the limiting case $\alpha = 0$.

The index of stability α characterizes the heaviness (the rate of decay) of the tails of stable distributions $S_\alpha(\sigma, \beta, \mu)$. In particular, if $X \sim S_\alpha(\sigma, \beta, \mu)$, then its distribution satisfies power law (1). 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 . 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 $\beta = 0$ 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 $a_i \geq 0$, $i = 1, \dots, n$,

$\sum_{i=1}^n a_i X_i / \left(\sum_{i=1}^n a_i^\alpha \right)^{1/\alpha} \sim S_\alpha(\sigma, \beta, \mu)$. The monographs by Zolotarev (1986), Embrechts, Klupperberg and Mikosch (1997), Uchaikin and Zolotarev (1999), Rachev and Mittnik (2000) and Rachev et al. (2005) contain detailed reviews of properties of stable distributions.

Denote by $\overline{\mathcal{CS}}$ the class of distributions which are k -fold convolutions of symmetric stable distributions $S_\alpha(\sigma, 0, 0)$ with characteristic exponents $\alpha \in (1, 2]$ and $\sigma > 0$ for some $k \geq 1$ (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 of one are taken). 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 $Y_i \sim S_{\alpha_i}(\sigma_i, 0, 0)$, $\alpha_i \in (1, 2]$, $\sigma_i > 0$, $i = 1, \dots, k$.

Further, $\overline{\mathcal{CSLC}}$ stands for the class of two-fold convolutions of distributions from the classes \mathcal{LC} and $\overline{\mathcal{CS}}$. That is, $\overline{\mathcal{CSLC}}$ is the class of two-fold convolutions of symmetric distributions which are either log-concave or stable with characteristic exponents greater than one (\mathcal{CSLC} is the abbreviation of “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 $\overline{\mathcal{CS}}$.

Finally, we denote by $\underline{\mathcal{CS}}$ the class of distributions which are k -fold convolutions of symmetric stable distributions $S_\alpha(\sigma, 0, 0)$ with indices of stability $\alpha \in (0, 1)$ and $\sigma > 0$ for some $k \geq 1$ (the underline indicates considering 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 $Y_i \sim S_{\alpha_i}(\sigma_i, 0, 0)$, $\alpha_i \in (0, 1)$, $\sigma_i > 0$, $i = 1, \dots, k$.

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 in the case of convolutions of stable distributions with *different* indices of stability. Therefore, 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 $\sigma > 0$.

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

Clearly, $\overline{\mathcal{CS}} \subset \overline{\mathcal{CSLC}}$ 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 $\sigma > 0$. The properties of stable distributions discussed above imply that the r.v.’s $X \sim \overline{\mathcal{CS}}$ have finite first moments: $E|X| < \infty$. However, the means are infinite for the r.v.’s $X \sim \overline{\mathcal{CS}}$: $E|X| = \infty$.

3 Demand-driven innovation and spatial competition over time

In this and in the next section, we review the setup of the model of demand-driven innovation and spatial competition over time developed by JR and its properties under log-concavity of signals' distributions. Throughout the paper, \mathbf{R}_+ stands for $\mathbf{R}_+ = [0, \infty)$.

Consider a market for a differentiated commodity. Let $\hat{\theta} \in \mathbf{R}_+$ be a location variable which differentiates the firm's product, let $\theta \in \mathbf{R}_+$ be an "ideal" product, and let $\rho(x, \theta) = |x - \theta|$, $x \in \mathbf{R}_+$, denote the absolute loss function.⁶ A consumer of type $u \in \mathbf{R}_+$ has the utility function $u - \rho(\hat{\theta}, \theta) - p_{\hat{\theta}}$, if she purchases one unit of good produced by the firm, and 0, if not, where $p_{\hat{\theta}}$ is the price the consumer pays for the good. Consumers are assumed to be perfectly informed about all price-quality combinations offered by various sellers and the firm is assumed to be a price taker. In what follows, we suppose that the price p of the ideal product θ is unity in terms of some "outside good": $p = 1$.

Under the above assumptions, a necessary condition for an equilibrium is that $\rho(\hat{\theta}, \theta) + p_{\hat{\theta}} = 1$ for all $\hat{\theta} \in \mathbf{R}$.

Simplifying the setting of the model considered in JR, we suppose that each period the firm makes two decisions. First, it chooses the product design $\hat{\theta}$, and does so before knowing what θ prevails for that period. The commitment to a particular $\hat{\theta}$ is costless but irreversible until next period. Having committed to $\hat{\theta}$, the firm then learns θ . Being of measure zero, the firm will be a price taker and its price is

$$p_{\hat{\theta}} = 1 - \rho(\hat{\theta}, \theta). \quad (1)$$

The firm then chooses the level of output y , with $C(y)$ denoting the corresponding convex and twice differentiable cost function.

Each period, the firm observes a sample \mathcal{S} of signals $s_i = \theta + \epsilon_i$, $i = 1, \dots, N$, about the next period's ideal product $\theta \in \mathbf{R}$, where ϵ_i , $i = 1, \dots, N$, are i.i.d. unimodal shocks with mode 0 and N is a (random) sample size. The size N of the sample \mathcal{S} of signals about the next period's ideal product observed by the firm follows a distribution $\pi(n; y)$ conditionally on y : $\pi(n; y) = P(N = n|y)$, $n = 0, 1, 2, \dots$. The function $\pi(n; y)$ is assumed to be increasing in y for all n , so that N is stochastically increasing in y and larger firms are likely to get more signals each period and to learn more about the next period's ideal product.

Below, we denote by \mathcal{S}_t , $\hat{\theta}_t$, θ_t and y_t the values of the variables in period t . In the model, the sequence of events is as follows: in period t , first \mathcal{S}_t is observed, next $\hat{\theta}_t$ is chosen; then θ_t is observed and y_t is chosen; the period then ends.

⁶From the proof of the results in this paper it follows that they continue to hold in the case of arbitrary loss functions $\rho(x, y) = \psi(|x - y|)$, where ψ is nonnegative and increasing on \mathbf{R}_+ .

Throughout the paper, for $v = (v_1, v_2, \dots, v_n) \in \mathbf{R}^n$, we denote by $\bar{v}_n = g_1(v_1, \dots, v_n) = n^{-1} \sum_{i=1}^n v_i$ the sample mean of v_i 's. In the case when n is odd, $n = 2k - 1$, we further denote by \tilde{v}_n the sample median (that is, the k th order statistic) of v_1, \dots, v_n : $\tilde{v}_n = g_2(v_1, \dots, v_n) = v_{(k)} = \text{median}(v_1, v_2, \dots, v_n)$ (here and in what follows, $v_{(1)} \leq v_{(2)} \leq \dots \leq v_{(n)}$ stand for components of x in nondecreasing order).

Let $g(v) = g(v_1, v_2, \dots, v_n)$ be an estimator based on a sample of observations $v = (v_1, v_2, \dots, v_n) \in \mathbf{R}^n$ that satisfies the translation equivariance condition: $g(v_1 + a, v_2 + a, \dots, v_n + a) = g(v_1, v_2, \dots, v_n) + a$ for all $a \in \mathbf{R}$ (see Bickel and Lehmann, 1975*a,b*, Ch. 4 in Rousseeuw and Leroy, 1987, and Sections 2.3 and 2.4 in Jurečková and Sen, 1996). Evidently, this condition holds for the sample mean $g_1(v_1, \dots, v_n) = \bar{v}_n$, $n \geq 1$, and the sample median $g_2(v_1, \dots, v_n) = \tilde{v}_n$, $n = 2k - 1$, $k = 1, 2, \dots$

Let $F(x; n) = P(|g(\epsilon_1, \epsilon_2, \dots, \epsilon_n)| \leq x)$, $x \geq 0$, $n = 1, 2, \dots$, denote the cdf of $|g(\epsilon_1, \epsilon_2, \dots, \epsilon_n)|$, $n = 1, 2, \dots$, on \mathbf{R}_+ , so that $F(x; n) = P(|\bar{\epsilon}_n| \leq x)$, $n = 1, 2, \dots$, for the sample mean $g_1(\epsilon_1, \dots, \epsilon_n) = \bar{\epsilon}_n$, and $F(x; n) = P(|\tilde{\epsilon}_n| \leq x)$, $n = 2k - 1$, $k = 1, 2, \dots$, for the sample median $g_2(\epsilon_1, \dots, \epsilon_n) = \tilde{\epsilon}_n$.

Suppose that, for $N > 0$, the firm chooses the estimator $\hat{\theta} = \hat{\theta}(S) = g(s_1, \dots, s_N)$ of θ as the product design. The loss associated with this choice of $\hat{\theta}$ for $N > 0$ is $\rho(\hat{\theta}(S), \theta) = |g(\epsilon_1, \epsilon_2, \dots, \epsilon_N)|$. In the case when $N = 0$ belongs to the support of N , so that $\pi(0; y) \neq 0$, it is usually assumed that $\rho(\hat{\theta}(S), \theta) = \infty$ for $N = 0$. The cdf of $\rho(\hat{\theta}(S), \theta)$ (on \mathbf{R}_+) conditional on y is

$$\xi(x; y) = P(\rho(\hat{\theta}(S), \theta) \leq x | y) = \sum_{n=0}^{\infty} F(x; n) \pi(n; y), \quad (2)$$

$x \geq 0$ (with $F(x; 0) = 0$ if $N = 0$ belongs to the support of N under the above convention).

The dynamic programming formulation of the firm's problem of choosing y following a realization $\rho(\hat{\theta}, \theta) = x$, is $V(x) = \max_y \left\{ y(1 - x) - C(y) + \beta \int V(\tilde{x}) d\xi(\tilde{x}; y) \right\}$.

Let $G(y) = \beta \int V(\tilde{x}) d\xi(\tilde{x}; y)$. The first-order and second-order conditions for an interior maximum in y are

$$p_{\hat{\theta}} - C'(y) + G'(y) = 0, \quad (3)$$

$$G''(y) < C''(y). \quad (4)$$

We assume that, for any continuous $f : \mathbf{R} \rightarrow \mathbf{R}$, the expression $\int f(\tilde{x}) d\xi(\tilde{x}; \lambda)$ is differentiable in λ . Under this assumption, one can implicitly differentiate first-order condition (3) (see JR).

Evidently, condition (4) holds if the function $G(y)$ is strictly concave: $G'' < 0$. However, $G'' > 0$ is also consistent with maxima being interior.⁷

⁷By Proposition 4 in JR, in the model of demand-driven innovation and spatial competition over time involving

4 Log-concave signals and demand-driven innovation and spatial competition over time

Throughout the paper, the distribution $\pi(n; y) = P(N = n|y)$ of N conditional on y will be assumed to be one of the following: a Poisson distribution with the mean μy : $\pi_0(n; y) = \frac{(\mu y)^n}{n!} \exp(-\mu y)$, $n = 0, 1, \dots$ (with the convention that $\rho(\hat{\theta}, \theta) = \infty$ for $N = 0$); a shifted Poisson distribution $\pi_1(n; y) = \frac{(\mu y)^{n-1}}{(n-1)!} \exp(-\mu y)$, $n = 1, 2, \dots$; or a Poisson-type distribution concentrated on odd numbers $\pi_2(n; y) = \frac{(\mu y)^{k-1}}{(k-1)!} \exp(-\mu y)$ for $n = 2k - 1$, $k = 1, 2, \dots$, $\pi_2(n; y) = 0$ for $n = 2k$, $k = 0, 1, 2, \dots$ (note that there is no ambiguity concerning the value of $\rho(\hat{\theta}, \theta)$ in the case $N = 0$ for distributions π_1 and π_2). The supports of the distributions π_j , $j = 0, 1, 2$, are, respectively, $M_0 = \{1, 2, 3, \dots\}$, $M_1 = \{0, 1, 2, \dots\}$ and $M_2 = \{1, 3, 5, \dots\}$.

In this paper, we consider the conclusions of the model of demand-driven innovation and spatial competition over time in the case where the firm employs the sample means \bar{s}_N as in JR (see the discussion in Subsection 1.2) or sample medians \tilde{s}_N of consumers' signals s_1, \dots, s_N as product designs: $\hat{\theta} = \hat{\theta}(\mathcal{S}) = g_1(s_1, \dots, s_N) = \bar{s}_N$ or $\hat{\theta} = \hat{\theta}(\mathcal{S}) = g_2(s_1, \dots, s_N) = \tilde{s}_N$ for $N = 2k - 1$, $k = 1, 2, \dots$

JR obtained the following Proposition 1.⁸ In the proposition and its analogues for heavy-tailed signals obtained below (Theorems 1 and 2), $y_t^{(1)}$ and $y_t^{(2)}$ are sizes of two firms at period t ; $y_{t+1}^{(1)}$ and $y_{t+1}^{(2)}$ stand for their sizes next period.

Proposition 1 (JR). *Suppose that, conditionally on y , N has one of the distributions $\pi_i(n; y)$, $i = 0, 1, 2$. Let the shocks $\epsilon_1, \epsilon_2, \dots$ be i.i.d. r.v.'s such that $\epsilon_i \sim \mathcal{LC}$, $i = 1, 2, \dots$. If the optimal level y_t of output satisfies the first- and second-order conditions for an interior maximum (3) and (4) and the firm chooses the sample mean $\hat{\theta} = g_1(s_1, \dots, s_N) = \bar{s}_N$ as the product design for $N > 0$, then the following conclusions (a)-(c) hold.*

(a) *The probability of rank reversals in adjacent periods $P(y_{t+1}^{(1)} > y_{t+1}^{(2)} | y_t^{(2)} > y_t^{(1)})$ is always less than 1/2.*

(b) *This probability diminishes as the current size-difference $y_t^{(2)} - y_t^{(1)}$ increases (holding constant the size of one of the firms).*

(c) *The distribution of future size is stochastically increasing as a function of current size y_t , that is, $P(y_{t+1} > y | y_t)$ is increasing in y_t for all $y \geq 0$.*

the choice of informational gathering effort z in addition to the choice of output y , larger firms always invest more in information if the function G is convex ($G'' > 0$). Thus, under this condition, investment z into gathering information in JR is secondary with respect to persistence results comparing to y . One should note that, according to empirical studies, there is a positive relationship between R&D expenditures and firm size, that suggests that G is indeed convex (see Kamien and Schwartz, 1982, and the discussion following Proposition 4 in JR)

⁸In JR, the proposition is formulated for the Poisson distribution π_0 . The argument for the distributions π_j , $j = 1, 2$, is completely similar to that case.

Lemma 1 in JR and its proof imply the following *sufficient* conditions for concavity of the function $G(y)$; under the assumptions of the lemma, therefore, the second-order condition (4) for an interior maximum with respect to y is satisfied.

Lemma 1 (JR). *Suppose that, conditionally on y , N has one of the distributions $\pi_i(n; y)$, $i = 0, 1, 2$. The function $G(y)$ is strictly concave in y if the sequence $\{F(x; n+1) - F(x; n)\}_{n=0}^{\infty}$ is strictly decreasing in n for all $x > 0$.*

As noted in JR, the conditions of Lemma 1 are satisfied for normal r.v.'s $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$, $i = 1, 2, \dots$ and the sample means \bar{s}_N employed as product designs.

5 Main results: robustness to heavy-tailedness assumptions

In this section, we present the main results of the paper on the robustness of the model of demand-driven innovation and spatial competition over time to heavy-tailedness assumptions.

The following theorem provides a generalization of Proposition 1 that shows that the results obtained by JR continue to hold in the case of not extremely thick-tailed signals.

Theorem 1 *Suppose that, conditionally on y , N has one of the distributions $\pi_j(n; y)$, $j = 0, 1, 2$. Let the shocks $\epsilon_1, \epsilon_2, \dots$ be i.i.d. r.v.'s such that $\epsilon_i \sim S_\alpha(\sigma, \beta, 0)$, $i = 1, 2, \dots$, for some $\sigma > 0$, $\beta \in [-1, 1]$ and $\alpha \in (1, 2]$, or $\epsilon_i \sim \overline{\text{CSLC}}$, $i = 1, 2, \dots$. Then conclusions (a), (b) and (c) in Proposition 1 hold.*

Lemma 2 shows that strict concavity of the function $G(y)$ in Lemma 1 and, consequently, the second-order condition (4) are satisfied for shocks $\epsilon_1, \epsilon_2, \dots$ with not extremely fat-tailed symmetric stable distributions and the sample mean of signals employed as the product design.

Lemma 2 *Suppose that the firm chooses the sample mean $\hat{\theta} = g_1(s_1, \dots, s_N) = \bar{s}_N$ as the product design for $N > 0$ and the shocks $\epsilon_1, \epsilon_2, \dots$ are i.i.d. r.v.'s such that $\epsilon_i \sim S_\alpha(\sigma, 0, 0)$, $i = 1, 2, \dots$, for some $\sigma > 0$, and $\alpha \in (1, 2]$. Then the sequence $\{F(x; n+1) - F(x; n)\}_{n=0}^{\infty}$ in Lemma 1 is strictly decreasing in n for all $x > 0$. Thus, the function $G(y)$ is strictly concave in y if, conditionally on y , N has one of the distributions $\pi_j(n; y)$, $j = 0, 1, 2$.*

As the following theorem shows, the conclusions of Proposition 1 and Theorem 1 are reversed in the case of shocks $\epsilon_1, \epsilon_2, \dots$ with extremely fat tails.

Theorem 2 *Suppose that, conditionally on y , N has one of the distributions $\pi_i(n; y)$, $i = 1, 2$. Let the shocks $\epsilon_1, \epsilon_2, \dots$ be i.i.d. r.v.'s such that $\epsilon_i \sim S_\alpha(\sigma, \beta, 0)$, $i = 1, 2, \dots$, for some $\sigma > 0$, $\beta \in [-1, 1]$ and $\alpha \in (0, 1)$, or $\epsilon_i \sim \underline{\mathcal{CS}}$, $i = 1, 2, \dots$. If the optimal level y_t of output satisfies the first- and second-order conditions for an interior maximum (3) and (4) and the firm chooses the sample mean $\hat{\theta} = g_1(s_1, \dots, s_N) = \bar{s}_N$ as the product design for $N > 0$, then the following conclusions (a')-(c') hold.*

(a') *The probability of rank reversals in adjacent periods $P(y_{t+1}^{(1)} > y_{t+1}^{(2)} | y_t^{(2)} > y_t^{(1)})$ is always greater than $1/2$.*

(b') *This probability increases as the current size-difference $y_t^{(2)} - y_t^{(1)}$ increases (holding constant the size of one of the firms).*

(c') *The distribution of future size is stochastically decreasing as a function of current size y_t , that is, $P(y_{t+1} > y | y_t)$ is decreasing in y_t for all $y \geq 0$.*

Remark 1 *From the proof of Theorem 2 it follows that, under its assumptions, $G' \leq 0$. It is not difficult to see that this implies that, in the setting of JR' model with the choice of investment z into information gathering in addition to the choice of quantity y , the optimal level of z is zero if the investment cost $K(z)$ is increasing and the first- and second-order conditions for an interior maximum are satisfied.*

According to our results, there is no informational advantage in the presence of extremely heavy-tailed signals if the sample mean is used as the product design. As the following theorem shows, having more signals is, however, always advantageous if a more robust estimator of θ , namely, the sample median, is used as the product design instead of the sample mean.

Theorem 3 *Suppose that, conditionally on y , N has the Poisson-type distribution $\pi_2(n; y)$ and $\epsilon_1, \epsilon_2, \dots$ are i.i.d. r.v.'s with a symmetric density $f(x)$. If the optimal level y_t of output satisfies (3) and (4) and the firm chooses the sample median $\hat{\theta} = g_2(s_1, \dots, s_N) = \tilde{s}_N$ as the product design for $N = 2k - 1$, $k = 1, 2, \dots$, then conclusions (a)-(c) of Proposition 1 hold.*

6 Concluding remarks

The results obtained in this paper illustrate that robustness of statistical techniques to heavy-tailedness can have important effects on the properties of economic models. In particular, the crossover through the value of the tail index $\alpha = 1$ in consumers' signals is crucial for the conclusions of the model of demand-driven innovation and spatial competition over time in which the firms use sample means of the signals as product designs. The conclusions of the model in such a setting are exactly the opposites of one another in the worlds of signals with tail indices

$\alpha < 1$ and $\alpha > 1$. However, the model's implications hold regardless of the degree of heavy-tailedness if the firms switch to the use of sample medians as estimators of the ideal product. These conclusions emphasize the dangers in the use of non-robust estimators in the presence of heavy-tailedness or statistical techniques that tend to overestimate the tail index. This may lead to making inferences and decisions on the base of models' predictions that are, in fact, wrong and, even more, the correct conclusions must be precisely the opposite ones.

As discussed in the introduction, the results obtained in the paper can be extended to more general settings, including the case of dependence among consumers' signals as well as signals with distributions in the domain of attraction of a stable law. Other extensions seem also be of interest, including analogues of the results for sample medians in Theorem 3 for distributions π_0 and π_1 that we conjecture to hold. In addition, one can use the efficiency comparisons between sample means and sample medians, together with inequalities between their tail probabilities for certain classes of distributions in Karlin (1992), to further analyze the differences in growth patterns for different types of firms.

Thick-tailed distributions appear naturally in economic models with herding behavior (see Cont and Bouchaud, 2000). These results may be helpful in the analysis of the important problem of how heavy-tailedness of consumers' signals can arise endogenously in firm growth models.

7 Appendix A1. Majorization properties of heavy-tailed distributions

As before, for a vector $v \in \mathbf{R}^n$, we denote by $v_{(1)} \geq \dots \geq v_{(n)}$ its components in nondecreasing order.

Definition 1 (Marshall and Olkin, 1979). *Let $v, w \in \mathbf{R}^n$. The vector v is said to be majorized by the vector w , written $v \prec w$, if $\sum_{i=1}^k v_{(i)} \geq \sum_{i=1}^k w_{(i)}$, $k = 1, \dots, n-1$, and $\sum_{i=1}^n v_{(i)} = \sum_{i=1}^n w_{(i)}$.*

The relation $v \prec w$ implies that the components of the vector v are less diverse than those of w (see Marshall and Olkin, 1979). In this context, it is easy to see that the following relations hold:

$$(1/(n+1), \dots, 1/(n+1), 1/(n+1)) \prec (1/n, \dots, 1/n, 0), \quad n \geq 1. \quad (5)$$

Definition 2 (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 $(v \prec w) \implies (\phi(v) \leq \phi(w))$ (resp. $(v \prec w) \implies (\phi(v) \geq \phi(w))$) for all $v, w \in A$. If, in addition, $\phi(v) < \phi(w)$ (resp., $\phi(v) > \phi(w)$) whenever $v \prec w$ and v is not a permutation of w , then ϕ is said to be strictly Schur-convex (resp., strictly Schur-concave) on A .*

The following Theorems 4 and 5 concerning general majorization properties of arbitrary convex combinations of heavy-tailed r.v.'s follow from the results obtained in Ibragimov (2005, 2007) (see Theorems 3.1 and 3.2 in Ibragimov, 2007). These theorems provide extensions of majorization results for light-tailed log-concavely distributed r.v.'s in Proschan (1965).

Theorem 4 (Ibragimov, 2005, 2007). *If X_1, \dots, X_n are i.i.d. r.v.'s such that $X_1 \sim S_\alpha(\sigma, \beta, 0)$ for some $\sigma > 0$, $\beta \in [-1, 1]$ and $\alpha \in (1, 2]$, or $X_1 \sim \overline{\mathcal{CSLC}}$, then the function $\psi(w, x) = P(|\sum_{i=1}^n w_i X_i| > x)$ is strictly Schur-convex in $w = (w_1, \dots, w_n) \in \mathbf{R}_+^n$ for all $x > 0$.*

According to Theorem 5, the majorization properties given by Theorem 4 above are reversed in the case of extremely thick-tailed distributions with tail indices less than one.

Theorem 5 (Ibragimov, 2005, 2007). *If X_1, \dots, X_n are i.i.d. r.v.'s such that $X_1 \sim S_\alpha(\sigma, \beta, 0)$ for some $\sigma > 0$, $\beta \in [-1, 1]$ and $\alpha \in (0, 1)$, or $X_1 \sim \underline{\mathcal{CS}}$, then the function $\psi(w, x)$ in Theorem 4 is strictly Schur-concave in $(w_1, \dots, w_n) \in \mathbf{R}_+^n$ for all $x > 0$.*

The following theorem obtained by Karlin (1992) gives the results on peakedness properties of sample medians $\tilde{X}_n = g_2(X_1, X_2, \dots, X_n) = \text{median}(X_1, X_2, \dots, X_n) = X_{(k)}$ of $n = 2k - 1$ symmetric r.v.'s X_i .

Theorem 6 Karlin (1992)⁹ *If X_i , $i = 1, 2, \dots$, are i.i.d. r.v.'s with a symmetric density f , then $P(|g_2(X_1, X_2, \dots, X_{2k+1})| > x) < P(|g_2(X_1, X_2, \dots, X_{2k-1})| > x)$ for all $x > 0$ and all $k = 1, 2, \dots$*

Remark 2 *In the arguments for the main results in the paper it is important that the comparisons for the tail probabilities of estimators implied by Theorems 4-6 (such as relations (7) and (14) in Appendix A2 below) hold for all values $x > 0$. The property that such comparisons hold for all $x > 0$ is a qualitative difference of the distributions in the classes $\overline{\mathcal{CSLC}}$ and $\underline{\mathcal{CS}}$ from distributions in the domain of attraction of stable laws. For instance, let $n \geq 1$ and let X_1, X_2, \dots, X_n be i.i.d. r.v.'s with regularly varying heavy tails: $P(X_i > x) = L(x)/x^\alpha$, where $\alpha > 0$ and $L(x)$ is a slowly varying at infinity function (so that the distribution of X_i 's is in the domain of attraction of a stable distribution with the characteristic exponent α for $0 < \alpha < 2$ and in the domain of attraction of a normal distribution for $\alpha > 2$). Then there exists a sufficiently large $x_0 = x_0(n)$ such that, for all $x > x_0$, $P(|\bar{X}_n| > x) < P(|X_1| > x)$ if $\alpha > 1$ and $P(|\bar{X}_n| > x) > P(|X_1| > x)$ if $\alpha < 1$. However, as follows from the discussion in Ibragimov and Walden (2006a,b), the latter comparisons for X_i 's, in general, do not hold for all $x > 0$; moreover, in general, they do not hold*

⁹The argument for the strict inequalities is as follows (see Karlin (1992)). Let $F(x) = \int_{-\infty}^x f(v)dv$ denote the cdf of X_i . The densities of the r.v.'s $\epsilon^{(k+1)}$ and $\epsilon^{(k)}$ are then $f^{(k+1)}(v) = [(2k+1)!/(k!)^2]f(v)(F(v))^k(1-F(v))^k$ and $f^{(k)}(v) = [(2k-1)!/((k-1)!)^2]f(v)(F(v))^{k-1}(1-F(v))^{k-1}$. It is easy to see that the difference $f^{(k+1)}(v) - f^{(k)}(v)$ has one sign change from + to - over the positive semi-axis. Since $\int_0^\infty (f^{(k+1)}(v) - f^{(k)}(v))dv = 0$, this implies that $\int_{-\infty}^x (f^{(k+1)}(v) - f^{(k)}(v))dv > 0$ for all $x > 0$ and, therefore, Theorem 6 indeed holds.

for all $x > 0$ even in the case of distributions with bounded support. The results in Ibragimov and Walden (2006b) provide analogues of Theorems 4 and 5 with comparisons, for all $x > 0$, between the tail probabilities of sample means for certain classes of r.v.'s in the domain of attraction of a stable distribution, including the averages of nonlinear transformations of stable random variates. Using these comparisons similar to the proof of the main results in this paper, one can obtain their extensions to the case of such classes of consumers' signals. Similarly, extensions of the majorization comparisons to distributions with a bounded support established in Ibragimov and Walden (2006a) allow one to obtain analogues of some of the results in this paper in the case of bounded signals.

8 Appendix A2. Proofs

Let $j \in \{0, 1, 2\}$, and let, conditionally on y , N have a distribution $\pi_j(n; y)$. Then from (2) it follows, similar to the proof of Lemma 2 in JR, that if the firm chooses $\hat{\theta} = \hat{\theta}(S) = g(s_1, \dots, s_N)$ as the product design, then, for all $x \geq 0$,

$$\partial \xi(x; y) / \partial y = \mu \sum_{n \in M_j} \pi_j(n; y) (F(x; n+1) - F(x; n)) \quad (6)$$

(with $F(x; 0) = 0$ if $j = 0$), where, as in Section 4, $M_0 = \{1, 2, 3, \dots\}$, $M_1 = \{0, 1, 2, \dots\}$ and $M_2 = \{1, 3, 5, \dots\}$ denote the supports of the distributions π_j , $j = 0, 1, 2$.

Proof of Theorem 1. Theorem 4 and relations (5) in Appendix A1 imply that, under the assumptions of the theorem,

$$F(x; n+1) = 1 - P(|\bar{\epsilon}_{n+1}| > x) > 1 - P(|\bar{\epsilon}_n| > x) = F(x; n), \quad (7)$$

$x > 0$, $n = 1, 2, \dots$ From (6) and (7) it, therefore, follows that

$$\partial \xi(x; y) / \partial y > 0 \quad (8)$$

for all $x > 0$, that is, $\xi(x, y)$ is increasing in y for all $x > 0$. As in JR, implicitly differentiating first-order condition (3), we have

$$\partial y / \partial p_{\hat{\theta}} = 1 / (C'' - G'') > 0, \quad (9)$$

if (3) and (4) hold, that is, y is increasing in $p_{\hat{\theta}}$. Conclusion (c) of the theorem now follows from (8) and (9) and the property that, by (1), $p_{\hat{\theta}}$ is decreasing in $\rho(\hat{\theta}, \theta)$:

$$\partial p_{\hat{\theta}} / \partial \rho < 0. \quad (10)$$

Let $\xi^{(i)}(x) = \xi(x; y^{(i)})$, $i = 1, 2$, and let $y^{(2)} > y^{(1)}$. By (8), we get that

$$\xi^{(2)}(x) > \xi^{(1)}(x) \quad (11)$$

for all $x > 0$. As in the proof of Proposition 6 in JR, we have

$$P(\rho^{(1)} > \rho^{(2)} | y^{(1)}, y^{(2)}) = \int \xi^{(2)}(x) d\xi^{(1)}(x) = \int \xi^{(1)}(x) d\xi^{(1)}(x) + \int (\xi^{(2)}(x) - \xi^{(1)}(x)) d\xi^{(1)}(x). \quad (12)$$

Since $\int \xi^{(1)}(x) d\xi^{(1)}(x) = 1/2$ using integration by parts, from (11) and (12) we get

$$P(\rho^{(1)} > \rho^{(2)} | y^{(1)}, y^{(2)}) > 1/2. \quad (13)$$

Relations (9), (10) and (13) imply conclusion (a) of the theorem.

As in the proof of Proposition 6 in JR, conclusion (b) of the theorem follows from (9), (10) and (12) since, by (8), holding $y^{(1)}$ constant and increasing $y^{(2)}$ or holding $y^{(2)}$ constant and decreasing $y^{(1)}$ increases $\xi^{(2)}(x) - \xi^{(1)}(x)$ for all $x > 0$. ■

Proof of Lemma 2. We have that, under the assumptions of the lemma, $n^{-1/\alpha} \sum_{i=1}^n \epsilon_i \sim S_\alpha(\sigma, 0, 0)$. Furthermore, by Theorem 2.7.6 in (Zolotarev, 1986, p. 134), the distribution of the r.v.'s ϵ_i are unimodal. Therefore, the function $P(\epsilon_1 \leq x)$ is concave in $x > 0$. This, together with strict concavity of the function $x^{1-1/\alpha}$, $\alpha > 1$, in $x > 0$, implies that, under the assumptions of the lemma, for all $n \geq 2$ and $x > 0$, $F(x, n) = P(|\bar{\epsilon}_n| \leq x)$ satisfies

$$F(x; n) = 2P(\epsilon_1 \leq xn^{1-1/\alpha}) - 1 > 2P(\epsilon_1 \leq x/2((n+1)^{1-1/\alpha} + (n-1)^{1-1/\alpha})) - 1 \geq P(\epsilon_1 \leq x(n+1)^{1-1/\alpha}) + P(\epsilon_1 \leq x(n-1)^{1-1/\alpha}) - 1 = 1/2(F(x; n+1) + F(x; n-1)).$$

For $n = 1$, using again unimodality of ϵ_1 and ϵ_2 , we get that, for all $x > 0$,

$$F(x; 1) = 2P(\epsilon_1 \leq x) - 1 \geq 2[2^{-(1-1/\alpha)}P(\epsilon_1 \leq 2^{1-1/\alpha}x) + (1 - 2^{-(1-1/\alpha)})1/2] - 1 > P(\epsilon_1 \leq 2^{1-1/\alpha}x) - 1/2 = 1/2F(x; 2).$$

■

Proof of Theorem 2. The proof is similar to the proof of Proposition 6 in JR and the proof of Theorem 1, with the use of Theorem 5 instead of Theorem 4 in this paper and the majorization results for log-concave distributions due to Proschan (1965) in JR. Under the assumptions of Theorem 2, one has, by Theorem 5 and relations (5), that, similar to relation (7),

$$F(x; n+1) = 1 - P(|\bar{\epsilon}_{n+1}| > x) < 1 - P(|\bar{\epsilon}_n| > x) = F(x; n), \quad (14)$$

$x > 0$, $n = 1, 2, \dots$ Relations (6) and (14) imply that, under the assumptions of Theorem 2,

$$\partial \xi(x; y) / \partial y < 0 \quad (15)$$

$x > 0$, that is, $\xi(x, y)$ is decreasing in y for all $x > 0$. Relations (9), (10) and (15) imply conclusion (c') of Theorem 2.

Let, as in the proof of Theorem 1, $\xi^{(i)}(x) = \xi(x; y^{(i)})$, $i = 1, 2$, and let $y^{(2)} > y^{(1)}$. Property (15) implies that

$$\xi^{(2)}(x) < \xi^{(1)}(x), \quad (16)$$

for all $x > 0$. From (12) and (16) it follows, similar to the proof of Proposition 6 in JR and to the proof of Theorem 1 in the present paper, that

$$P(\rho^{(1)} > \rho^{(2)} | y^{(1)}, y^{(2)}) = 1/2 + \int (\xi^{(2)}(x) - \xi^{(1)}(x)) d\xi^{(1)}(x) < 1/2. \quad (17)$$

Relations (9), (10) and (17) imply conclusion (a') of Theorem 2.

Conclusion (b') of Theorem 2 follows from (9), (10) and (12) and the fact that, by (15), increase in the current size-difference $y^{(2)} - y^{(1)}$ (holding constant $y^{(1)}$ or $y^{(2)}$) decreases $\xi^{(2)}(x) - \xi^{(1)}(x)$ for all $x > 0$ under the assumptions of the theorem. ■

Proof of Theorem 3. From Theorem 6 it follows that, under the assumptions of the theorem,

$$F(x; 2k + 1) = P(|\tilde{\epsilon}_{2k+1}| \leq x) > P(|\tilde{\epsilon}_{2k-1}| \leq x) = F(x; 2k - 1) \quad (18)$$

for all $k \geq 1$ and all $x > 0$. From (6) and (18) it follows that (8) is true for all $x > 0$. As in the proof of Proposition 1 in JR and Theorem 1 in the present section, this implies that conclusions (a)-(c) of Proposition 1 hold. ■

References

- An, M. Y. (1998), 'Logconcavity versus logconvexity: a complete characterization', *Journal of Economic Theory* **80**, 350–369.
- Andrews, D. W. K. (2005), 'Cross-section regression with common shocks', *Econometrica* pp. 1551–1585.
- Axtell, R. L. (2001), 'Zipf distribution of u.s. firm sizes', *Science* **293**, 1818–1820.
- Berger, J. O. (1985), *Statistical decision theory and Bayesian analysis*, Springer-Verlag, New York.
- Bickel, P. J. and Lehmann, E. L. (1975*a*), 'Descriptive statistics for nonparametric models. I. Introduction', *Annals of Statistics* **3**, 1038–1044.
- Bickel, P. J. and Lehmann, E. L. (1975*b*), 'Descriptive statistics for nonparametric models. II. Location', *Annals of Statistics* **3**, 1045–1069.
- Box, G. E. P. and Tiao, G. C. (1973), *Bayesian inference in statistical analysis*, Addison-Wesley Publishing Co., Reading, Massachusetts-London-Don Mills, Ontario.

- Carlin, B. P. and Louis, T. A. (2000), *Bayes and empirical Bayes methods for data analysis*, Chapman & Hall, London.
- Cont, R. and Bouchaud, J. P. (2000), ‘Herd behavior and aggregate fluctuations in financial markets’, *Macroeconomic Dynamics* **4**, 170–196.
- De Vany, A. S. and Walls, W. D. (2004), ‘Motion picture profit, the stable paretian hypothesis and the curse of the superstar’, *Journal of Economic Dynamics and Control* **28**, 1035–1057.
- Embrechts, P., Klupperberg, C. and Mikosch, T. (1997), *Modelling extremal events for insurance and finance*, Springer, New York.
- Fabozzi, F. J., Focardi, S. M. and Kolm, P. N. (2006), *Financial Modeling of the Equity Market: CAPM to Cointegration*, Wiley, Hoboken, NJ.
- Fama, E. F. (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 & Hall, New York.
- 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.
- Huber, P. J. (1981), *Robust statistics*, Wiley, New York.
- Huber, P. J. (1996), *Robust statistical procedures*, second edn, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.
- Ibragimov, R. (2005), *New majorization theory in economics and martingale convergence results in econometrics*. Ph.D. dissertation, Yale University.
- Ibragimov, R. (2007), ‘Efficiency of linear estimators under heavy-tailedness: Convolutions of α -symmetric distributions’, *Econometric Theory* **23**(3).
- Ibragimov, R. and Walden, J. (2006a), ‘The limits of diversification when losses may be large’, *Harvard Institute of Economic Research Discussion Paper No. 2104*. Forthcoming in *Journal of Banking and Finance*. Available at <http://www.economics.harvard.edu/hier/2006papers/HIER2104.pdf>.

- Ibragimov, R. and Walden, J. (2006*b*), ‘Portfolio diversification under local, moderate and global deviations from power laws’, *Working paper, Harvard University and University of California at Berkeley*.
- 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.
- Jovanovic, B. and Rob, R. R. (1987), ‘Demand-driven innovation and spatial competition over time’, *Review of Economic Studies* **54**, 63–72.
- Jurečková, J. and Sen, P. K. (1996), *Robust statistical procedures: Asymptotics and interrelations*, Wiley, New York.
- Kamien, M. and Schwartz, N. (1982), *Market structure and innovation*, Cambridge University Press, New York.
- Karlin, S. (1968), *Total positivity*, Stanford University Press, Stanford.
- Karlin, S. (1992), Stochastic comparisons between means and medians for i.i.d. random variables, in e. K. V. Mardia, ed., ‘The Art of Statistical Science. A Tribute to G. S. Watson’, Wiley, Chichester, pp. 261–274.
- Loretan, M. and Phillips, P. C. B. (1994), ‘Testing the covariance stationarity of heavy-tailed time series’, *Journal of Empirical Finance* **1**, 211–248.
- Maddala, G. S. and Rao, C. R., eds (1997), *Robust inference*, Handbook of statistics **15**, North Holland, Amsterdam.
- 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*, Springer-Verlag, New York.
- Marshall, A. W. and Olkin, I. (1979), *Inequalities: theory of majorization and its applications*, Academic Press, New York.
- McCulloch, J. H. (1996), Financial applications of stable distributions, in G. S. Maddala and C. R. Rao, eds, ‘Handbook of statistics, Vol. 14’, Elsevier, Amsterdam, pp. 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.
- Nelson, C. R. and Plosser, C. I. (1982), ‘Trends and random walks in macroeconomic time series’, *Journal of Monetary Economics* **10**, 139–162.

- Nešlehova, J., Embrechts, P. and Chavez-Demoulin, V. (2006), ‘Infinite mean models and the l₁ norm for operational risk’, *Journal of Operational Risk* **1**, 3–25.
- Proschan, F. (1965), ‘Peakedness of distributions of convex combinations’, *Annals of Mathematical Statistics* **36**, 1703–1706.
- Rachev, S. T., Menn, C. and Fabozzi, F. J. (2005), *Fat-tailed and skewed asset return distributions: implications for risk management, portfolio selection, and option pricing*, Wiley, Hoboken, NJ.
- Rachev, S. T. and Mittnik, S. (2000), *Stable Paretian Models in Finance*, Wiley, New York.
- Rousseeuw, P. J. and Leroy, A. M. (1987), *Robust regression and outlier detection*, Wiley, New York.
- Scherer, F. M., Harhoff, D. and Kukies, J. (2000), ‘Uncertainty and the size distribution of rewards from innovation’, *Journal of Evolutionary Economics* **10**, 175–200.
- Silverberg, G. and Verspagen, B. (2004), ‘The size distribution of innovations revisited: an application of extreme value statistics to citation and value measures of patent significance’, *Working paper, Maastricht University*. Forthcoming in *Journal of Econometrics*.
- Uchaikin, V. V. and Zolotarev, V. M. (1999), *Chance and stability. Stable distributions and their applications*, VSP, Utrecht.
- Zolotarev, V. M. (1986), *One-dimensional stable distributions*, Providence: American Mathematical Society.