ON THE ROBUSTNESS OF ECONOMIC MODELS TO HEAVY-TAILEDNESS ASSUMPTIONS

Rustam Ibragimov[1]

Department of Economics, Yale University

ABSTRACT

Many economic models have a structure that depends on majorization phenomena for convolutions of distributions. The present paper develops a unified approach to the analysis of majorization properties of linear combinations of random variables. It further studies the robustness of these properties and the implications of a number of important economic models to heavy-tailedness assumptions. The paper shows that majorizations for log-concave distributed random variables established in the seminal work by Proschan (1965) continue to hold for not extremely thick-tailed distributions. However, the majorization properties are reversed in the case of distributions with extremely heavy-tailed densities. This is the first probabilistic result that shows that majorization properties of log-concave densities are reversed for a wide class of distributions and is the key to reversals of properties of many economic models built upon the popular log-concavity assumption.

In a series of applications of the main probabilistic results, we study robustness of monotone consistency of the sample mean, value at risk (VaR) analysis for financial portfolios, growth theory for firms that can invest in information about their market, as well as that of bundling theory for sellers of baskets of goods with an arbitrary degree of complementarity or substitutability. The main results show that many economic models are robust to heavy-tailedness assumptions as long as the distributions entering these assumptions are not extremely thick-tailed. But the implications of these models are reversed for distributions with sufficiently long-tailed densities. The following list summarizes some of the main results.

i) Using the general majorization results obtained, the paper shows, 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. Namely, in the case of risks with extremely thick-tailed distributions, diversification of a portfolio always leads to an increase in riskiness of the portfolio's return. The paper further demonstrates that the stylized facts on diversification are robust to thick-tailedness of risks or returns as long as their distributions are not extremely long-tailed. Moreover, we show that, in the world of not extremely heavy-tailed risks, the value at risk satisfies the important condition of coherency. However, coherency of the value at risk is always violated if distributions of risks are extremely thick-tailed. In addition, the paper shows that the sample mean exhibits monotone consistency in the case of data from not extremely thick-tailed populations. Thus, an increase in the sample size always improves performance of the sample mean.

[1]Job market paper. I am indebted to Donald Andrews, Peter Phillips and Herbert Scarf for all their support and guidance in all stages of the current project. I also thank Aydin Cecen, Brian Dineen, Darrell Duffie, Xavier Gabaix, Philip Haile, Samuel Karlin, Alex Maynard, Ingram Olkin, Ben Polak, Gustavo Soares, Kevin Song and the participants at the International Conference on Stochastic Finance, 2004 and the 18th New England Statistics Symposium at Harvard University for helpful comments and discussions. The financial support from the Yale University Graduate Fellowship and the Cowles Foundation Prize is gratefully acknowledged.

(ii) The paper develops a framework that allows one to model the optimal bundling problem of a multiproduct monopolist providing interrelated goods with an arbitrary degree of complementarity or substitutability. Characterizations of optimal bundling strategies are derived for the seller in the case of long-tailed valuations and tastes for the products. We show, in particular, that if the 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. In addition, we obtain characterizations of optimal bundling strategies for a monopolist who provides complements or substitutes for profit-maximizing prices to buyers with long-tailed tastes.

(iii) Another application that is explored in depth concerns the analysis of growth of firms that invest into learning about the next period's optimal product. The paper presents 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 extremely thick-tailed signals.

(iv) Several extensions of the above results to the case of dependence are obtained, including convolutions of $\alpha-$symmetric and spherical distributions and models with common shocks which are of great importance in economics and finance.

Keywords and phrases: Robustness, heavy-tailed distributions, majorization, monotone consistency, value at risk, coherent measures of risk, portfolio diversification, optimal bundling, multiproduct monopolist, Vickrey auction, ascending auction, descending auction, substitutes, complements, interrelated goods, independently priced goods, firm growth, innovation and spatial competition, Gibrat's law, $\alpha-$symmetric distributions, spherical distributions, models with common shocks

CONTENTS

1 Introduction

1.1 Objectives and key results

The present paper develops a unified approach to the analysis of a number of models in economics that depend on majorization and stochastic dominance properties of convolutions of distributions. The main results show that many economic models are robust to heavy-tailedness assumptions as long as the distributions entering these assumptions are not extremely thick-tailed.[2] But the implications of these models are reversed for distributions with extremely long-tailed densities.

Many models in economics, finance and risk management have a structure that depends on majorization phenomena for linear combinations of random variables (r.v.'s). The majorization relation is a formalization of the concept of diversity in the components of vectors. Over the past decades, majorization theory, which focuses on the study of this relation and functions that preserve it, has found applications in disciplines ranging from statistics, probability theory and economics to mathematical genetics, linear algebra and geometry. This paper presents a unified framework for the analysis of the majorization properties of linear combinations of random variables. It further studies the robustness of these majorization properties and the implications of a number of important economic models to thick-tailedness assumptions. The paper shows, in particular, that majorizations for log-concavely distributed signals established in the seminal work by Proschan (1965) continue to hold for r.v.'s with not extremely heavy-tailed densities. More precisely, the tails of distributions of linear combinations of such long-tailed r.v.'s continue to exhibit the property of Schur-convexity, as in the case of log-concave distributions. However, the majorization properties are reversed for extremely thick-tailed distributions, in which case Schur-convexity of the tails is replaced by their Schur-concavity. This is the first probabilistic result that shows that majorization properties of log-concave densities are reversed for a wide class of distributions and is the key to reversals of properties of many economic models built upon the popular log-concavity assumption. One should emphasize here that, although log-concave distributions have many appealing properties that have been utilized in a number of works in economics, they cannot be used in the study of thick-tailedness phenomena since any log-concave density is extremely light-tailed: in particular, its tails decline at least exponentially fast and all its moments exist.

In a series of applications of the main probabilistic results, we study robustness of monotone consistency of the sample mean, value at risk (VaR) analysis for financial portfolios, growth theory for firms that can invest in information about their market, as well as that of bundling theory for sellers of baskets of goods with an arbitrary degree of complementarity or substitutability. The following list summarizes some of the main results.

(i) The paper provides a precise formalization of the concept of portfolio diversification on the basis of majorization ordering. Using the general majorization results obtained in the paper, it further shows, 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 obtained in the paper is that the stylized facts on diversifi-

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

cation 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.

In addition, the paper shows that the sample mean exhibits monotone consistency in the case of data from not extremely thick-tailed populations. Thus, an increase in the sample size always improves performance of the sample mean.

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

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

(iv) Several extensions of the above results to the case of dependence are obtained, including convolutions of $\alpha-$symmetric and spherical distributions and models with common shocks which are of great importance in economics and finance.

1.2 Heavy-tailedness in economic and financial data and its modelling

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 the papers in Mandelbrot, 1997) and Fama (1965) who pioneered the study of heavy-tailed distributions with tails declining as $x^{-\alpha}$, $\alpha > 0$, in these fields.[3] It was documented in numerous studies that the time series encountered in many fields in economics and finance are heavy-tailed (see the discussion in Loretan and Phillips, 1994, Meerschaert and Scheffler, 2000, Gabaix, Gopikrishnan, Plerou and Stanley, 2003, and references therein). 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).

One should note that a few studies report the index of stability to be close to one or even slightly less than one for certain financial time series (e.g., Bulgarian lev/US dollar exchange spot rates and increments of the market time process for Deutsche Bank price record in Rachev and Mittnik, 2000).

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

1.3 Implications for economic models: robustness and reversals

As demonstrated in the present paper, the value of the tail index $\alpha = 1$ (that is, the existence of the first moment) is exactly the critical boundary between robustness of implications of many economic models to heavy-tailedness assumptions and their reversals. More precisely, according to the results, the implications of most of the models are robust to thick-tailedness assumptions with tail indices $\alpha > 1$. But the conclusions of the models are reversed for extremely heavy-tailed distributions with $\alpha < 1$.

The above results have several implications concerning the use of the models considered in this paper in the real-world settings that are likely to be heavy-tailed. The crossover through the value $\alpha = 1$ is crucial for the models' conclusions, since those are exactly the opposites of one another in the worlds with $\alpha > 1$ and $\alpha < 1$. Therefore, obviously, it is necessary to have accurate and robust estimates of the tail parameters for the data observed in the heavy-tailed settings to which models are applied. The results, in particular, emphasize the danger in the use of statistical techniques that tend to overestimate the tail index. This may lead to making inferences and decisions on

[3] If a model involves a r.v. X with such thick-tailed distribution, then $P(|X| > x) = x^{-\alpha}$. The r.v. for which this is the case has finite moments $E|X|^p$ of order $p < \alpha$. However, the moments are infinite for $p > \alpha$.

the base of models' predictions that are, in fact, wrong and, even more, the correct conclusions must be precisely the opposite ones.

One should also emphasize here that several results in this paper do not require the distributions entering their assumptions to be extremely heavy-tailed with $\alpha < 1$ in order to exhibit reversals. This is the case, for instance, for the results on optimal bundling of interrelated goods. As demonstrated in Part II of the paper, optimal bundling strategies of a multiproduct monopolist depend crucially on both thick-tailedness of the tastes (characterized by the tail parameters α) as well as on the degree r of complementarity or substitutability among the products provided, with $r > 1$ corresponding to the case of complements and $r < 1$ modelling the case of substitutes. For instance, the characterizations of the optimal bundling strategies for the seller of baskets of complements or substitutes derived in Part II depend on comparisons between α and r. Therefore, the strategies exhibit reversal patterns even in the case when consumers have tastes with $\alpha > 1$. Namely, the optimal bundling strategies of a multiproduct monopolist providing complements with relatively high degree of complementarity $r > \alpha$ on auctions or for profit-maximizing prices are reversals of her optimal strategies in the case of substitutes (for which $r < 1$) and of those in the case of complements with relatively low degree of complementarity $1 < r < \alpha$.

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

To illustrate the main ideas of the proof and in order to simplify the presentation of the main results, we first model heavy-tailedness using the framework of independent stable distributions and their convolutions. More precisely, the class of not extremely thick-tailed distributions is first modelled using convolutions of stable distributions with (different) indices of stability greater than one. Similarly, the results of the paper for extremely heavy-tailed case are first presented and proven using the framework of convolutions of stable distributions with characteristic exponents less than one. The former class has tail exponents $\alpha > 1$ (and thus, as discussed in the previous subsection, the implications of economic models continue to hold) and for the latter class one has $\alpha < 1$ (so that the models exhibit reversals). In some places throughout the paper, we will omit the words "not extremely" in the discussion of the results for distributions with $\alpha > 1$ (or with relatively large α) and refer to such distributions as just "heavy-tailed" or "thick-tailed", if this does not lead to a confusion. The class of "not extremely heavy-tailed" or, with this convention on the terminology, of "heavy-tailed" distributions is thus opposed to the class of "extremely thick-tailed" distributions with $\alpha < 1$ (or with relatively small α).

At the end of the paper we show, however, that all the results obtained of the paper continue to hold for a wide class of multivariate distributions for which marginals are dependent and can be non-identical and, in addition to that, can have finite variances, unlike stable distributions and their convolutions.

More precisely, we show all the results continue to hold for convolutions of dependent r.v.'s with joint $\alpha-$symmetric distributions and their analogues with non-identical marginals.[4] The class of $\alpha-$symmetric distributions is very wide and includes, in particular, spherical distributions corresponding to $\alpha = 2$. Important examples of

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

spherical distributions, in turn, are given by Kotz type, multinormal and logistic distributions and multivariate stable laws. In addition, they include a subclass of mixtures of normal distributions as well as multivariate $t-$distributions that were used in a number of papers to model heavy-tailedness phenomena with dependence and finite moments up to a certain order (see, among others, Praetz, 1972, Blattberg and Gonedes, 1974, and Glasserman, Heidelberger and Shahabuddin, 2002). Moreover, the class of $\alpha-$symmetric distributions includes a wide class of convolutions of models with common shocks affecting all risks (such as macroeconomic or political ones, see Andrews, 2003) which are of great importance in economics and finance.

Similar to the framework based on stable distributions, the implications of the economic models considered in the paper continue to hold for convolutions of $\alpha-$symmetric distributions with $\alpha > 1$. The implications of the models are reversed in the case of convolutions of $\alpha-$symmetric distributions with $\alpha < 1$.

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

1.5 Optimistic implications

The main message of the results in this paper is that the presence of heavy-tailedness can either reinforce or reverse the implications of economic models, depending on the degree of thick-tailedness. This message is optimistic since, as discussed before, many economic models are robust to heavy-tailedness (and dependence) as long as the tail indices $\alpha > 1$ and empirical studies observe such values for α in most of economic and financial time series.

However, as the main results demonstrate, the reversals of the models are possible for a wide class of extremely thick-tailed distributions. Therefore, the models should be applied with care in presence of very heavy-tailed signals, especially in the case of the tail indices close to the critical boundary $\alpha = 1$.

Moreover, the dual patterns in the models' implications for thick-tailed distributions and for extremely heavy-tailed distributions illustrate and shed a new light on several phenomena observed in the real world, such as in the case of dual patterns in bundling strategies in real-world markets (see Part II). The results in this paper also provide new insights concerning firm growth and size patterns in high-tech sectors, in particular, in the Internet economy (see the discussion in Part III).

The applications considered in the paper require not only inferences about the asymptotic behavior of tails of heavy-tailed r.v.'s but also tail comparisons for their linear combinations with different weights. As demonstrated in this paper, majorization theory provides a natural and powerful tool for this.

The paper demonstrates that majorization results for linear combinations of r.v.'s provide a natural unifying framework for the analysis of a great number of models in a wide range of areas in economics, finance, risk management, econometrics and statistics. Applications of the main majorization results obtained in the paper cover all the areas in economics and econometrics where Proschan's (1965) seminal results on peakedness properties of log-concave densities have been used in the literature and provide precise conditions for robustness and reversals of models in those areas. Furthermore, the results provide a unifying approach to the analysis of important models in several other fields where applicability of majorization theory has not yet been recognized so far, even in the case of log-concave distributions. These fields include, for instance, value at risk analysis for financial portfolios and bundling theory for an auctioneer of interrelated goods. Interestingly, the main majorization results are also related to the study of properties of inheritance models that have been a subject of an increasing interest in economics in recent years. Applications of the main majorization results of this paper in that area, together with several their extensions to the case of transforms of heavy-tailed r.v.'s, are presented in Ibragimov (2004).

1.6 Organization of the paper

The main body of the paper consists of three parts, with each part exploring in depth each of the areas of applications of the main majorization results obtained. Part I presents the analysis of majorization properties of heavy-tailed distributions and applies it to the study of monotone consistency of the sample mean and VaR for financial portfolios under heavy-tailedness. Section 3 in Part I contains notations and definitions of classes of distributions used throughout the paper. Part II develops a framework for the analysis of the optimal bundling problem with complements and substitutes under heavy-tailedness of consumers' valuations and tastes. The general majorizations results obtained in Part I are applied in Part II to derive complete characterizations of the optimal bundling strategies in such a setting. In Part III, we obtain applications of the majorization properties derived in Part I in the analysis of the robustness of the model of demand-driven innovation and spatial competition over time to heavy-tailedness of signals' distributions. Part IV discusses extensions of the results in the paper to the case of dependence, including convolutions of α-symmetric and spherical distributions and models with common shocks, and makes some concluding remarks.

Part I

MAJORIZATION THEORY UNDER THICK-TAILEDNESS AND APPLICATIONS TO MONOTONE CONSISTENCY AND VALUE AT RISK

2 Discussion of the results in Part I

2.1 Majorization

The present paper demonstrates that powerful tools for the study of robustness of many important economic models to heavy-tailedness assumptions are given by majorization theory. A vector $a \in \mathbf{R}^n$ is said to be majorized by a vector $b \in \mathbf{R}^n$, written $a \prec b$, if $\sum_{i=1}^{k} a_{[i]} \leq \sum_{i=1}^{k} b_{[i]}$, $k = 1, ..., n-1$, and $\sum_{i=1}^{n} a_{[i]} = \sum_{i=1}^{n} b_{[i]}$, where $a_{[1]} \geq \ldots \geq a_{[n]}$ and $b_{[1]} \geq \ldots \geq b_{[n]}$ denote components of a and b in decreasing order. The relation $a \prec b$ implies that the components of the vector a are less diverse than those of b (see Marshall and Olkin, 1979). In this context, it is easy to see that the following relations hold:

$$(1/(n+1), ..., 1/(n+1), 1/(n+1)) \prec (1/n, ..., 1/n, 0), \quad n \geq 1. \tag{2.1}$$

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

2.2 Log-concave distributions and their majorization properties

A r.v. X with density $f : \mathbf{R} \to \mathbf{R}$ and the convex distribution support $\Omega = \{x \in \mathbf{R} : f(x) > 0\}$ is said to be log-concavely distributed 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).

Following Birnbaum (1948), we say that a r.v. X is more peaked about $\mu \in \mathbf{R}$ than is Y if $P(|X - \mu| > x) \leq P(|Y - \mu| > x)$ for all $x \geq 0$. If this inequality is strict whenever the two probabilities are not both zero, the r.v. X is said to be *strictly* more peaked about μ than is Y. In case $\mu = 0$, X is simply said to be (strictly) more peaked than Y. 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.

Throughout the paper, $\mathbf{R}_+$ stands for $\mathbf{R}_+ = [0, \infty)$. Proschan (1965) obtains the following seminal result concerning majorization and peakedness properties of linear combinations of log-concavely distributed r.v.'s:

Proposition 2.1 *(Proschan, 1965).*[5] *If $X_1, ..., X_n$ are i.i.d. symmetric log-concavely distributed r.v.'s, then the function $\psi(a, x) = P\big(\sum_{i=1}^n a_i X_i > x\big)$ is strictly Schur-convex in $a = (a_1, ..., a_n) \in \mathbf{R}_+^n$ for $x > 0$ and is strictly Schur-concave in $a = (a_1, ..., a_n) \in \mathbf{R}_+^n$ for $x < 0$.*

Clearly, from Proposition 2.1 it follows that $\sum_{i=1}^n a_i X_i$ is strictly more peaked than $\sum_{i=1}^n b_i X_i$ if $a \prec b$ and a is not a permutation of b.

Proschan (1965) notes that Proposition 2.1 also holds for (two-fold) convolutions of log-concave distributions with symmetric Cauchy distributions and shows that comparisons implied by the proposition are reversed for $n = 2^k$, vectors $a = (1/n, 1/n, ..., 1/n) \in \mathbf{R}^n$ with identical components and certain transforms of symmetric Cauchy r.v.'s.

2.3 Applications of Proschan's (1965) results for log-concave densities and monotone consistency

Proposition 2.1 obtained by Proschan (1965) and its extensions have been applied to the analysis of many problems in statistics, econometrics, economic theory and other fields. For instance, Eaton (1988) used generalizations of the results to obtain concentration inequalities for Gauss-Markov estimators. Karlin (1984, 1992) applied Proposition 2.1 in the study of environmental sex determination models. Jovanovic and Rob (1987) used the result in the analysis of the model of demand-driven innovation and spatial competition over time. Fang and Norman (2003) applied Proposition 2.1 in the study of optimal bundling strategies for a multiproduct monopolist.

Several authors (see, e.g., Proschan, 1965, Tong, 1994, and Jensen, 1997) discussed implications of Proposition 2.1 and its extensions in the study of monotone consistency of estimators in econometrics and statistics. A weakly consistent estimator $\hat{\theta}_n$ of a population parameter θ is said to exhibit *monotone consistency* for θ if $\hat{\theta}_n$ becomes successively more peaked about θ as n increases, that is, if $P(|\hat{\theta}_{n+1} - \theta| > x) \leq P(|\hat{\theta}_n - \theta| > x)$ for all $x \geq 0$. By majorization comparisons (2.1), from Proposition 2.1 it follows that samples $X_1, ..., X_n$, $n \geq 1$, from a log-concavely distributed population symmetric about $\mu \in \mathbf{R}$, have the *monotone peakedness of the sample mean (MPSM)* property, that is, the sample mean $\overline{X}_n = (1/n) \sum_{i=1}^n X_i$ becomes increasingly more peaked about μ as n gets larger. Thus, $\overline{X}_n$ exhibits monotone consistency for μ. This implies that an increase in the sample size always improves performance of the sample mean.

2.4 Implications for value at risk analysis

Proposition 2.1 also has the following important interpretation in the framework of value-at-risk analysis and portfolio choice theory that, to our knowledge, has not yet been recognized in the literature. In what follows, given a loss probability $\alpha \in (0, 1/2)$ and a r.v. (risk) Z, we denote by $VaR_\alpha(Z)$ the value at risk (VaR) of Z at level α, that is, its $(1-\alpha)-$quantile.[6]

[5]This proposition is formulated as Theorem 12.J.1 in Marshall and Olkin (1979) and is the main result in Section 12.J in that book. Proschan's (1979) work is also presented, in a rearranged form, in Section 11 of Chapter 7 in Karlin (1968), where peakedness results are formulated for "PF2 densities," which is the same as "log-concave densities."

[6]That is, in the case of an absolutely continuous risk Z, $P(Z > VaR_\alpha(Z)) = \alpha$.

For $w = (w_1, ..., w_n) \in \mathbf{R}_+^n$, $\sum_{i=1}^n w_i = 1$, denote by Z_w the return on the portfolio of risks $X_1, ..., X_n$ with weights w. Further, denote $\underline{w} = (1/n, 1/n, ..., 1/n)$ and $\overline{w} = (1, 0, ..., 0)$. The expressions $VaR_\alpha(Z_{\underline{w}})$ and $VaR_\alpha(Z_{\overline{w}})$ are, thus, the values at risk of the portfolio with equal weights and of the portfolio consisting of only one return (risk).

Suppose that $v = (v_1, ..., v_n) \in \mathbf{R}_+^n$ and $w = (w_1, ..., w_n) \in \mathbf{R}_+^n$, $\sum_{i=1}^n v_i = \sum_{i=1}^n w_i = 1$, are the weights of two portfolios of risks (or assets' returns). If $v \prec w$, it is natural to think about the portfolio with weights v as being more diversified than that with weights w so that, for example, the portfolio with equal weights $\underline{w}$ is more diversified than the portfolio with weights $\overline{w}$ consisting of one risk (in this regard, the notion of one portfolio being more or less diversified than another one is, in some sense, the opposite of that for vectors of weights for the portfolio).

According to Proposition 2.1, diversification of a portfolio of extremely light-tailed i.i.d. symmetric log-concave risks $X_1, ..., X_n$ (formalized, as above, by majorization properties of the vector of portfolio weights w) decreases riskiness of its return Z_w in the sense of (first-order) stochastic dominance. More precisely, for all $\alpha \in (0, 1/2)$, $VaR_\alpha(Z_v) \le VaR_\alpha(Z_w)$ if $v \prec w$. Furthermore, by relations (2.1), the value at risk is minimal for the portfolio weights $\underline{w}$ and is maximal for the weights $\overline{w}$:

$$VaR_\alpha(Z_{\underline{w}}) \le VaR_\alpha(Z_w) \le VaR_\alpha(Z_{\overline{w}}). \tag{2.2}$$

These comparisons further imply that, in the case of i.i.d. log-concavely distributed risks X_1 and X_2, the VaR has the following subadditivity property:

$$VaR_\alpha(X_1 + X_2) \le VaR_\alpha(X_1) + VaR_\alpha(X_2) \tag{2.3}$$

for all $\alpha \in (0, 1/2)$. Thus, in the world of extremely light-tailed distributions, the value at risk exhibits the important coherency property in the sense of Artzner, Delbaen, Eber and Heath (1999).[7] One should note here that subadditivity is one of the natural requirements to be imposed on a risk measure, together with other axioms of the coherency concept. In particular, as discussed in Artzner et. al. (1999), if this condition does not hold for a risk measure used, then this may have severe consequences for regulators, exchange and society.

2.5 Extensions of Proschan's (1965) results

A number of papers in probability and statistics have focused on extension of Proschan's results (see, among others, Chan, Park and Proschan, 1989, the review in Tong, 1994, Jensen, 1997, and Ma, 1998). One should emphasize, however, that in all the studies that dealt with generalizations of Proposition 2.1, the majorization properties of the tail probabilities were of the same type as in Proschan (1965). Namely, the results gave extensions of Proschan's results concerning *Schur−convexity* of the tail probabilities $\psi(a, x)$, $x > 0$, to classes of r.v.'s more general than those considered in Proschan (1965). We are not aware of any general results concerning *Schur−concavity* of the tail probabilities $\psi(a, x)$, $x > 0$, for certain classes of r.v.'s.[8]

[7]see also Embrechts, McNeil and Straumann, 1999, and Section 6 in this part of the paper for more on the concept of coherency.

[8]One should note that the proof in Proschan (1965) can be reproduced word to word with respective changes of signs of inequalities under the "assumptions" that $X_1, ..., X_n$ are i.i.d. symmetric log-convexily distributed r.v.'s. However, as it is easy to see, the later objects do not exist, namely, there does not exist a symmetric r.v.'s with a log-convex density (see also An, 1998). Therefore, this approach to obtaining counterparts of Proposition 2.1 for Schur-concavity of $\psi(a, x)$, $x > 0$, is hopeless.

We emphasize here again that departures from conditions of log-concavity of distributions are necessary in the study of robustness of models involving them to heavy-tailedness assumptions because any log-concave distribution is extremely light-tailed. In particular, the tails of log-concave distributions decline at least exponentially fast and all their moments are finite (see An, 1998).[9]

2.6 Main majorization results of the paper

In this part, we present the main probabilistic results of the paper on majorization properties of linear combinations of thick-tailed r.v.'s. In particular, we show that the majorization properties of convex combinations of r.v.'s given by Proposition 2.1 continue to hold for heavy-tailed distributions (Theorem 4.3). However, these properties are reversed for distributions with extremely thick tails (Theorem 4.4). As discussed before, to our knowledge, this is the first result in the probabilistic and statistics literature that shows that majorization properties of log-concave densities are reversed for a wide class of distributions. Moreover, as we demonstrate throughout the paper, this result provides the key to reversals of properties of many economic models built upon the popular log-concavity assumption.

In addition, we also obtain results that give analogues of Proposition 2.1 for heavy-tailed r.v.'s and majorization comparisons between powers of coefficients of their linear combinations (Theorems 4.1 and 4.2). These results are central in the development of the framework for the analysis of optimal bundling strategies with interrelated goods in Part II of the paper.

2.7 Monotone consistency and portfolio VaR under heavy-tailedness

From the main majorization results obtained in this part of the paper it follows that the implications of Proposition 2.1 for monotone consistency of the sample mean and the portfolio value at risk discussed in Subsections 2.3 and 2.4 continue to hold for thick-tailed distributions (Corollaries 5.1 and 5.2). In particular, diversification of a portfolio of heavy-tailed risks (formalized by majorization properties of the vector of portfolio weights) decreases riskiness of its return in the sense of (first-order) stochastic dominance. In addition, it is demonstrated that the VaR is a coherent measure of risk in the case of heavy-tailed distributions (Corollary 6.1).

The majorization results obtained further imply that the results on the VaR analysis are reversed for extremely long-tailed risks. For instance, this part of the paper shows that the stylized fact that portfolio diversification is always preferable is reversed for a wide class of distributions of risks (Corollary 5.3). The class of distributions for which this is the case is the class of extremely heavy-tailed distributions: a diversification of a portfolio of extremely

[9] The reader is referred to Karlin (1968), Marshall and Olkin (1979) and An (1998) for a survey of examples many other properties of log-concave distributions. Some of these properties are the following:

Any log-concave density is unimodal. Moreover, it has the property of strong unimodality, that is, its convolution with any other unimodal density is again unimodal;

The survivor and distribution functions of log-concave densities are both log-concave and, thus, a log-concavely distributed r.v. has the new-better-than-used property;

A log-concave density is of Pólya frequency of order 2 (PF-2);

The hazard function of a log-concave density is monotonically increasing.

Examples of log-concave distributions include the normal distribution, the uniform density, the exponential density, the Gamma distribution $\Gamma(\alpha, \beta)$ with the shape parameter $\alpha \geq 1$, the Beta distribution $\mathcal{B}(a, b)$ with $a \geq 1$ and $b \geq 1$; the Weibull distribution $\mathcal{W}(\gamma, \alpha)$ with the shape parameter $\alpha \geq 1$.

thick-tailed risks always leads to an increase in the riskiness of their portfolio. In particular, the signs of inequalities in (2.2) and (2.3) are reversed for extremely long-tailed risks. These reversals also reveal that (see Corollary 6.2), in a world of extremely heavy-tailed risks, the value at risk always has a strict superadditivity property instead of subadditivity in (2.3) and thus is not a coherent risk measure in such a setting. One should indicate here that, so far, only a few particular counterexamples that show that VaR is not, in general, a coherent measure of risk were available in the literature (see Artzner et. al., 1999, and Embrechts et. al., 1999). Our results demonstrate, on the other hand, that it is *always* violated for a wide class of risks with extremely heavy-tailed distributions.

Using the general results on majorization properties of the tail probabilities of linear combinations of r.v.'s derived in the paper, we also obtain sharp bounds on the VaR of portfolios of heavy-tailed risks that give refinements of estimates (2.2) and their analogues in the extremely thick-tailed case (Corollaries 5.4 and 5.5).

2.8 Organization of Part I

This part of the paper is organized as follows: Section 3 contains notations and definitions of classes of distributions used throughout the paper. In Section 4, we derive the main results of the paper on majorization properties of linear combinations of long-tailed r.v.'s. Section 5 presents implications of the majorization results in Section 4 for the study of monotone consistency of the sample mean and portfolio value at risk. Section 6 discusses the implications of the results for coherency of the VaR under thick-tailedness. Finally, Section 7 contains the proofs of the results obtained in this part of the paper.

3 Notations

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

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

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

As it 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); $\alpha = 1/2$ and $\beta \pm 1$ (Lévy distributions).[10] 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 $S_\alpha(\sigma, \beta, \mu)$. For a stable r.v. $X \sim S_\alpha(\sigma, \beta, \mu)$ with $\alpha \in (0, 2)$ one has $P(|X| > x) = x^{-\alpha}$ and thus the $p-$th absolute moments $E|X|^p$ of X are finite if $p < \alpha$ and are infinite otherwise.

[10]The densities of Cauchy distributions are $f(x) = \sigma/(\pi(\sigma^2 + (x - \mu)^2))$. Lévy distributions have densities $f(x) = (\sigma/(2\pi))^{1/2}exp(-\sigma/(2x))x^{-3/2}$, $x \geq 0$; $f(x) = 0$, $x < 0$, where $\sigma > 0$, and their shifted versions.

The 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 $(-\infty, \mu]$ for $\beta = -1$ (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 standard deviation in the special case of Gaussian distributions ($\alpha = 2$). For a detailed review of properties of stable distributions the reader is referred to, e.g., the monographs by Zolotarev (1986) and Uchaikin and Zolotarev (1999).

Throughout the paper, $\mathcal{LC}$ denotes the class of symmetric log-concave distributions, as defined in Subsection 2.2.[11]

For $0 < r < 2$, we denote by $\overline{\mathcal{CS}}(r)$ the class of distributions which are convolutions of symmetric stable distributions $S_\alpha(\sigma, 0, 0)$ with characteristic exponents $\alpha \in (r, 2]$ and $\sigma > 0$.[12] That is, $\overline{\mathcal{CS}}(r)$ consists of distributions of r.v.'s X such that, for some $k \geq 1$, $X = Y_1 + ... + Y_k$, where Y_i, $i = 1, ..., k$, are independent r.v.'s such that $Y_i \sim S_{\alpha_i}(\sigma_i, 0, 0)$, $\alpha_i \in (r, 2]$, $\sigma_i > 0$, $i = 1, ..., k$.

Further, $\overline{\mathcal{CSLC}}$ stands for the class of convolutions of distributions from the classes $\mathcal{LC}$ and $\overline{\mathcal{CS}}(1)$. That is, $\overline{\mathcal{CSLC}}$ is the class of convolutions of symmetric distributions which are either log-concave or stable with characteristic exponents greater than one.[13] In other words, $\overline{\mathcal{CSLC}}$ consists of distributions of r.v.'s X such that $X = Y_1 + Y_2$, where Y_1 and Y_2 are independent r.v.'s with distributions belonging to $\mathcal{LC}$ or $\overline{\mathcal{CS}}(1)$.

Finally, for $0 < r \leq 2$, we denote by $\underline{\mathcal{CS}}(r)$ the class of distributions which are convolutions of symmetric stable distributions $S_\alpha(\sigma, 0, 0)$ with indices of stability $\alpha \in (0, r)$ and $\sigma > 0$.[14] That is, $\underline{\mathcal{CS}}(r)$ consists of distributions of r.v.'s X such that, for some $k \geq 1$, $X = Y_1 + ... + Y_k$, where Y_i, $i = 1, ..., k$, are independent r.v.'s such that $Y_i \sim S_{\alpha_i}(\sigma_i, 0, 0)$, $\alpha_i \in (0, r)$, $\sigma_i > 0$, $i = 1, ..., k$.

A linear combination of independent stable r.v.'s with the *same* characteristic exponent α also has a stable distribution with the same α. However, in general, this does not hold in the case of convolutions of stable distributions with *different* indices of stability. Therefore, the class $\overline{\mathcal{CS}}(r)$ of *convolutions* of symmetric stable distributions with *different* indices of stability $\alpha \in (r, 2]$ is wider than the class of *all* symmetric stable distributions $S_\alpha(\sigma, 0, 0)$ with $\alpha \in (r, 2]$ and $\sigma > 0$. Similarly, the class $\underline{\mathcal{CS}}(r)$ is wider than the class of *all* symmetric stable distributions $S_\alpha(\sigma, 0, 0)$ with $\alpha \in (0, r)$ and $\sigma > 0$.

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

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

In some sense, symmetric (about $\mu = 0$) Cauchy distributions $S_1(\sigma, 0, 0)$ are at the dividing boundary between

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

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

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

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

the classes $\underline{\mathcal{CS}}(1)$ and $\overline{\mathcal{CS}}(1)$ (and between the classes $\underline{\mathcal{CS}}(1)$ and $\overline{\mathcal{CSLC}}$). Similarly, for $r \in (0,2)$, symmetric stable distributions $S_r(\sigma, 0, 0)$ with the characteristic exponent $\alpha = r$ are at the dividing boundary between the classes $\underline{\mathcal{CS}}(r)$ and $\overline{\mathcal{CS}}(r)$. Further, symmetric normal distributions $S_2(\sigma, 0, 0)$ are at the dividing boundary between the class $\mathcal{LC}$ of log-concave distributions and the class $\underline{\mathcal{CS}}(2)$ of convolutions of symmetric stable distributions with indices of stability $\alpha < 2$.[15]

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

4 Main results on majorization properties of heavy-tailed distributions

Theorems 4.1-4.4 in this section give analogues of Proposition 2.1 in Subsection 2.2 for heavy-tailed r.v.'s. In particular, according to the following Theorem 4.1, the majorization properties of convex combinations of r.v.'s in the classes $\overline{\mathcal{CS}}(r)$ are of the same type as in Proposition 2.1 with respect to the comparisons between the powers of the components of the vectors of weights of the combinations.

Theorem 4.1 *Let $r \in (0,2)$. If $X_1, ..., X_n$ are i.i.d. r.v.'s such that $X_i \sim S_\alpha(\sigma, \beta, 0)$, $i = 1, ..., n$, for some $\sigma > 0$, $\beta \in [-1,1]$ and $\alpha \in (r, 2]$, or $X_i \sim \overline{\mathcal{CS}}(r)$, $i = 1, ..., n$, then the function $\psi(a, x)$, $a \in \mathbf{R}^n_+$ in Proposition 2.1 is strictly Schur-convex in $(a_1^r, ..., a_n^r)$ for $x > 0$ and is strictly Schur-concave in $(a_1^r, ..., a_n^r)$ for $x < 0$.*

As follows from Theorem 4.2 below, the majorization properties of the tail probabilities $\psi(a, x)$ in Theorem 4.1 are reversed in the case of r.v.'s from the classes $\underline{\mathcal{CS}}(r)$.

Theorem 4.2 *Let $r \in (0,2]$. If $X_1, ..., X_n$ are i.i.d. r.v.'s such that $X_i \sim S_\alpha(\sigma, \beta, 0)$, $i = 1, ..., n$, for some $\sigma > 0$, $\beta \in [-1,1]$ and $\alpha \in (0, r)$, or $X_i \sim \underline{\mathcal{CS}}(r)$, $i = 1, ..., n$, then the function $\psi(a, x)$, $a \in \mathbf{R}^n_+$ in Proposition 2.1 is strictly Schur-concave in $(a_1^r, ..., a_n^r)$ for $x > 0$ and is strictly Schur-convex in $(a_1^r, ..., a_n^r)$ for $x < 0$.*

According to Theorem 4.3 below, peakedness properties of linear combinations of r.v.'s with heavy-tailed distributions are the same as in the case of log-concave distributions in Proschan (1965).

Theorem 4.3 *Proposition 2.1 holds if $X_1, ..., X_n$ are i.i.d. r.v.'s such that $X_i \sim S_\alpha(\sigma, \beta, 0)$, $i = 1, ..., n$, for some $\sigma > 0$, $\beta \in [-1,1]$ and $\alpha \in (1, 2]$, or $X_i \sim \overline{\mathcal{CSLC}}$, $i = 1, ..., n$.*

As follows from Theorem 4.4, peakedness properties given by Proposition 2.1 and Theorem 4.3 above are reversed in the case of r.v.'s with extremely heavy-tailed distributions.

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

Theorem 4.4 *If $X_1, ..., X_n$ are i.i.d. r.v.'s such that $X_i \sim S_\alpha(\sigma, \beta, 0)$, $i = 1, ..., n$, for some $\sigma > 0$, $\beta \in [-1, 1]$ and $\alpha \in (0, 1)$, or $X_i \sim \underline{\mathcal{CS}}(1)$, $i = 1, ..., n$, then the function $\psi(a, x)$ in Proposition 2.1 is strictly Schur-concave in $(a_1, ..., a_n) \in \mathbf{R}_+^n$ for $x > 0$ and is strictly Schur-convex in $(a_1, ..., a_n) \in \mathbf{R}_+^n$ for $x < 0$.*

Remark 4.1. If r.v.'s $X_1, ..., X_n$ have a symmetric Cauchy distribution $S_1(\sigma, 0, 0)$ which is, as discussed in Section 3, exactly at the dividing boundary between the class $\underline{\mathcal{CS}}(1)$ in Theorem 4.4 and the class $\overline{\mathcal{CSLC}}$ in Theorem 4.3, 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). Similarly, the function $\psi(a, x)$, $a \in \mathbf{R}_+^n$, in Theorems 4.1 and 4.2 depends only on $\sum_{i=1}^n a_i^r$ and x and so is *both* Schur-concave and Schur-convex in $(a_1^r, ..., a_n^r)$ for all $x \in \mathbf{R}$ if the r.v.'s $X_1, ..., X_n$ in the theorems have a symmetric stable distribution $S_r(\sigma, 0, 0)$ with the index of stability $\alpha = r$ which is at the dividing boundary between the classes $\overline{\mathcal{CS}}(r)$ and $\underline{\mathcal{CS}}(r)$. As follows from the proof of Theorems 4.1-4.4, the above implies that Theorems 4.3 and 4.4 continue to hold for convolutions of distributions from the classes $\overline{\mathcal{CSLC}}$ and $\underline{\mathcal{CS}}(1)$ with symmetric Cauchy distributions $S_1(\sigma, 0, 0)$. Similarly, Theorem 4.1 and 4.2 continue to hold for convolutions of distributions from the classes $\overline{\mathcal{CS}}(r)$ and $\underline{\mathcal{CS}}(r)$ with symmetric stable distributions $S_r(\sigma, 0, 0)$. These generalizations imply corresponding extensions in all the applications of majorization properties of linear combinations of heavy-tailed r.v.'s throughout the rest of the paper.

5 Monotone consistency and portfolio value at risk under heavy-tailedness

Theorem 4.3 provides the following result concerning the monotone consistency properties of the sample mean for data from heavy-tailed population.

Corollary 5.1 *Let $\mu \in \mathbf{R}$. If $X_1, ..., X_n$, $n \geq 1$, are i.i.d. r.v.'s such that $X_i \sim S_\alpha(\sigma, \beta, \mu)$, $i = 1, ..., n$, for some $\sigma > 0$, $\beta \in [-1, 1]$ and $\alpha \in (1, 2]$, or $X_i - \mu \sim \overline{\mathcal{CSLC}}$, then the sample mean $\overline{X}_n = (1/n) \sum_{i=1}^n X_i$ exhibits monotone consistency for μ, that is, $P(|\overline{X}_n - \mu| > x)$ converges to zero monotonically in n for all $x \geq 0$.*

In addition to Corollary 5.1, from the majorization results given by Theorem 4.3 it follows, similar to the case of extremely light-tailed log-concave distributions in Subsection 2.4, that diversification of a portfolio of thick-tailed risks $X_i \sim \overline{\mathcal{CSLC}}$, $i = 1, ..., n$, with weights $w = (w_1, ..., w_n) \in \mathbf{R}_+^n$, $\sum_{i=1}^n w_i = 1$, leads to a decrease in the riskiness of its return $Z_w = \sum_{i=1}^n w_i X_i$ in the sense of (first-order) stochastic dominance. Let, as in Subsection 2.4, for $\alpha \in (0, 1/2)$, $VaR_\alpha(Z_w)$ be the value at risk of Z_w associated with the loss probability α. We obtain the following result.

Corollary 5.2 *Let X_i, $i = 1, ..., n$, be i.i.d. r.v.'s such that $X_i \sim \overline{\mathcal{CSLC}}$, $i = 1, ..., n$. Then $VaR_\alpha(Z_v) < VaR_\alpha(Z_w)$ if $v \prec w$ and v is not a permutation of w. In particular, $VaR_\alpha(Z_{\underline{w}}) < VaR_\alpha(Z_w) < VaR_\alpha(Z_{\overline{w}})$ for all $\alpha \in (0, 1/2)$ and all weights w such that $w \neq \underline{w}$ and w is not a permutation of $\overline{w}$.*

In contrast, the results in Theorem 4.4 imply that the results for portfolio VaR discussed in Subsection 2.4 in the introduction are reversed under the assumption that the distributions of the risks $X_1, ..., X_n$ are extremely long-tailed. In such a setting, diversification of a portfolio of the risks increases riskiness of its return. We have the following

Corollary 5.3 *Let X_i, $i = 1, ..., n$, be i.i.d. r.v.'s such that $X_i \sim \underline{\mathcal{CS}}(1)$, $i = 1, ..., n$. Then $VaR_\alpha(Z_v) > VaR_\alpha(Z_w)$ if $v \prec w$ and v is not a permutation of w. In particular, $VaR_\alpha(Z_{\overline{w}}) < VaR_\alpha(Z_w) < VaR_\alpha(Z_{\underline{w}})$ for all $\alpha \in (0, 1/2)$ and all weights w such that $w \neq \underline{w}$ and w is not a permutation of $\overline{w}$.*

Theorems 4.1 and 4.2 imply the following results that give sharp bounds on the value at risk of portfolios of heavy-tailed returns (risks). These bounds refine and complement the estimates given by Corollaries 5.2 and 5.3 in the world of heavy-tailed risks.

Corollary 5.4 *Let $r \in (0, 2)$ and let $X_1, ..., X_n$ be i.i.d. risks such that $X_i \sim \overline{\mathcal{CS}}(r)$, $i = 1, ..., n$. Then the following sharp bounds hold:*

$$n^{1-1/r}(\sum_{i=1}^n w_i^r)^{1/r} VaR_\alpha(Z_{\underline{w}}) < VaR_\alpha(Z_w) < (\sum_{i=1}^n w_i^r)^{1/r} VaR_\alpha(Z_{\overline{w}})$$

for all $\alpha \in (0, 1/2)$ and all weights w such that $w \neq \underline{w}$ and w is not a permutation of $\overline{w}$.

Corollary 5.5 *Let $r \in (0, 2]$, $\alpha \in (0, 1/2)$ and let $X_1, ..., X_n$ be i.i.d. risks such that $X_i \sim \underline{\mathcal{CS}}(r)$, $i = 1, ..., n$. Then the following sharp bounds hold :*

$$(\sum_{i=1}^n w_i^r)^{1/r} VaR_\alpha(Z_{\overline{w}}) < VaR_\alpha(Z_w) < n^{1-1/r}(\sum_{i=1}^n w_i^r)^{1/r} VaR_\alpha(Z_{\underline{w}})$$

for all $\alpha \in (0, 1/2)$ and all weights w such that $w \neq \underline{w}$ and w is not a permutation of $\overline{w}$.

6 (Non-)Coherency of the value at risk under thick-tailedness

Let $\mathcal{X}$ be a certain linear space of r.v.'s X defined on a probability space $(\Omega, \Im, P)$. We assume that $\mathcal{X}$ contains all degenerate r.v.'s $X \equiv a \in \mathbf{R}$. According to the definition in Artzner et. al. (1999) (see also Embrechts et. al., 1999, and Frittelli and Gianin, 2002), a functional $\mathcal{R} : \mathcal{X} \to R$ is said to be a *coherent* measure of risk if it satisfies the following axioms:

A1. (Monotonicity) $\mathcal{R}(X) \geq \mathcal{R}(Y)$ for all $X, Y \in \mathcal{X}$ such that $Y \leq X$ (a.s.), that is, $P(X \leq Y) = 1$.

A2. (Translation invariance) $\mathcal{R}(X + a) = \mathcal{R}(X) + a$ for all $X \in \mathcal{X}$ and any $a \in \mathbf{R}$.

A3. (Positive homogeneity) $\mathcal{R}(\lambda X) = \lambda \mathcal{R}(X)$ for all $X \in \mathcal{X}$ and any $\lambda \geq 0$.

A4. (Subadditivity) $\mathcal{R}(X + Y) \leq \mathcal{R}(X) + \mathcal{R}(Y)$ for all $X, Y \in \mathcal{X}$.

In some papers (see, e.g., Frittelli and Gianin, 2002, and Fölmer and Schied, 2002), the axioms A3 and A4 were replaced by the following weaker axiom of convexity:

A5. (Convexity) $\mathcal{R}(\lambda X + (1 - \lambda)Y) \leq \lambda \mathcal{R}(X) + (1 - \lambda)\mathcal{R}(Y)$ for all $X, Y \in \mathcal{X}$ and any $\lambda \in [0, 1]$

(clearly, A5 follows from A3 and A4). The above axioms are natural conditions to be imposed on measures of risk in the setting where positive values of r.v.'s $X \in \mathcal{X}$ represent losses of a risk holder.[16] For instance, subadditivity property is important, among others, from the regulatory point of view because if a firm were forced to meet the requirement of extra capital which is not subadditive, it might be motivated to break up into several separately incorporated affiliates (see the discussion in Artzner et. al., 1999). In addition to that, the properties A1-A5 are important because, as follows from Huber (1981, Ch. 10) (see also Artzner et. al., 1999), in the case of a finite Ω, a risk measure $\mathcal{R}$ is coherent (that is, it satisfies A1-A4) if and only if it is representable as $\mathcal{R}(X) = \sup_{Q \in \mathcal{P}} E_Q(X)$, where $\mathcal{P}$ is some set of probability measures on Ω and, for $Q \in \mathcal{P}$, E_Q denotes the expectation with respect to Q. In other words, the risk measure $\mathcal{R}$ is the worst result of computing the expected loss $E_Q(X)$ over a set $\mathcal{P}$ of "generalized scenarios" (probability measures) Q. A similar representation holds as well in the case of an arbitrary Ω and the space $\mathcal{X} = L^\infty(\Omega, \Im, P)$ of bounded r.v.'s (see Fölmer and Schied, 2002); moreover, as discussed in Frittelli and Gianin (2002), by duality theory, the convexity axiom A5 alone implies analogues of such characterizations for an arbitrary Ω and the space $\mathcal{X} = L_p(\Omega, \Im, P)$, $p \geq 1$, of r.v.'s X with a finite $p-$th moment $E|X|^p < \infty$.

It is easy to verify that the value at risk $VaR_\alpha(X)$ satisfies the axioms of monotonicity, positive homogeneity and translation invariance A1, A3 and A4. However, as follows from the counterexamples constructed by Artzner et. al. (1999) and Embrechts et. al. (1999), in general, it fails to satisfy the subadditivity and convexity properties A2 and A5, in particular, for certain Pareto distributions (Examples 6 and 7 in Embrechts et. al., 1999).

On the other hand, our comparisons for i.i.d. r.v.'s $X_i \sim \overline{\mathcal{CSLC}}$ given by Corollary 5.2, imply that the value at risk satisfies subadditivity and convexity axioms A4 and A5 and is, thus, a coherent measure of risk in the world of long-tailed risks from the class $\overline{\mathcal{CSLC}}$:

Corollary 6.1 *Let X_1 and X_2 be i.i.d. risks such that $X_i \sim \overline{\mathcal{CSLC}}$, $i = 1, 2$. For all $\alpha \in (0, 1/2)$ and any $\lambda \in (0, 1)$, one has $VaR_\alpha(X_1 + X_2) < VaR_\alpha(X_1) + VaR_\alpha(X_2)$ and $VaR_\alpha(\lambda X_1 + (1-\lambda)X_2) < \lambda VaR_\alpha(X_1) + (1-\lambda)VaR_\alpha(X_2)$. That is, subadditivity and convexity axioms A4 and A5 are satisfied for VaR and it is a coherent measure of risk for the class $\overline{\mathcal{CSLC}}$.*

Furthermore, from Corollary 5.3 it follows that axioms A2 and A5 are *always* violated for risks with extremely heavy-tailed distributions. Thus, the value at risk is not a coherent risk measure in the world of extremely long-tailed distributions:

Corollary 6.2 *Let X_1 and X_2 be i.i.d. risks such that $X_i \sim \underline{\mathcal{CS}}(1)$, $i = 1, 2$. For all $\alpha \in (0, 1/2)$ and any $\lambda \in (0, 1)$, one has $VaR_\alpha(X_1) + VaR_\alpha(X_2) < VaR_\alpha(X_1 + X_2)$ and $\lambda VaR_\alpha(X_1) + (1-\lambda)VaR_\alpha(X_2) < VaR_\alpha(\lambda X_1 + (1-\lambda)X_2)$. That is, subadditivity and convexity axioms A4 and A5 are violated for VaR and it is not a coherent measure of risk for the class $\underline{\mathcal{CS}}(1)$.*

<u>Remark 6.1</u>. It is well-known that if r.v.'s X and Y are such that $P(X > x) \leq P(Y > x)$ for all $x \in \mathbf{R}$, then $EU(X) \leq EU(Y)$ for all increasing functions $U : \mathbf{R} \to \mathbf{R}$ for which the expectations exist (see Shaked and Shanthikumar, 1994, pp. 3-4). This fact and Theorems 4.1-4.4 imply corresponding results concerning majorization properties

[16]This interpretation of losses follows that in Embrechts et. al. (1999) and is in contrast to Artzner et. al. (1999) who interpret *negative* values of risks in $\mathcal{X}$ as losses.

of expectations of (utility or payoff) functions of linear combinations of heavy-tailed r.v.'s. In particular, Theorems 4.1 and 4.2 give sharp bounds on the expected payoffs of contingent claims written on a portfolio of heavy-tailed risks similar to those in Corollaries 5.4 and 5.5. For instance, we get that if $U : \mathbf{R}_+ \to \mathbf{R}$ is an increasing function, then, assuming existence of the expectations, the function $\varphi(a) = EU(|\sum_{i=1}^n a_i X_i|)$, $a \in \mathbf{R}_+^n$, is Schur-convex in $(a_1^r, ..., a_n^r)$ under the assumptions of Theorem 4.1 and is Schur-concave in $(a_1^r, ..., a_n^r)$ under the assumptions of Theorem 4.2. In particular, $EU\big(\big|n^{1-1/r}\big(\sum_{i=1}^n w_i^r\big)^{1/r} Z_{\underline{w}}\big|\big) \leq EU(|Z_w|) \leq EU\big(\big|\big(\sum_{i=1}^n w_i^r\big)^{1/r} Z_{\overline{w}}\big|\big)$ for all portfolios of risks satisfying Theorem 4.1 and $EU\big(\big|\big(\sum_{i=1}^n w_i^r\big)^{1/r} Z_{\overline{w}}\big|\big) \leq EU(|Z_w|) \leq EU\big(\big|n^{1-1/r}\big(\sum_{i=1}^n w_i^r\big)^{1/r} Z_{\underline{w}}\big|\big)$ for all portfolios of risks satisfying Theorem 4.2. We also get that the function $\varphi(a)$, $a \in \mathbf{R}_+^n$ is Schur-concave in $(a_1^2, ..., a_n^2)$ if $X_i \sim S_\alpha(\sigma, \beta, 0)$, $i = 1, ..., n$, for some $\sigma > 0$, $\beta \in [-1, 1]$ and $\alpha \in (0, 2)$, or $X_i \sim \underline{\mathcal{CS}}(2)$. The above results extend and complement those in Efron (1969) and Eaton (1970) (see also Marshall and Olkin, 1979, pp. 361-365) who studied classes of functions $U : \mathbf{R} \to \mathbf{R}$ and r.v.'s $X_1, ..., X_n$ for which Schur-concavity of $\varphi(a)$, $a \in \mathbf{R}_+^n$, in $(a_1^2, ..., a_n^2)$ holds. Further, we obtain that $\varphi(a)$ is Schur-convex in $a \in \mathbf{R}_+^n$ under the assumptions of Theorem 4.3 and is Schur-concave in $a \in \mathbf{R}_+^n$ under the assumptions of Theorem 4.4. It is important to note here that in the case of increasing *convex* functions $U : \mathbf{R}_+ \to \mathbf{R}$ and r.v.'s $X_1, ..., X_n$ satisfying the assumptions of Theorem 4.4, the expectations $EU(|\sum_{i=1}^n a_i X_i|)$ are infinite for all $a \in \mathbf{R}_+^n$.[17] Therefore, the last result does not contradict the well-known fact that (see Marshall and Olkin, 1979, p. 361) the function $Ef(\sum_{i=1}^n a_i X_i)$ is Schur-convex in $(a_1, ..., a_n) \in \mathbf{R}$ for all i.i.d. r.v.'s $X_1, ..., X_n$ and convex functions $f : \mathbf{R} \to \mathbf{R}$ as it might seem on the first sight.

7 Proofs of the results in Part I

In the proofs below, we provide the complete argument for the main majorizations results that provide a reversal of those available in the literature, namely for Theorem 4.2 and Theorem 4.4. The proof of Theorem 4.1 that gives the results on Schur-convexity of the tail probabilities of linear combinations of r.v.'s follows the same lines as that of Theorem 4.2, with respective changes in the signs of inequalities. We also provide the complete proof of Theorem 4.4 since it is not implied by Theorem 4.1 alone, but needs to combine the results in that theorem with those given by Proposition 2.1.

Proof of Theorems 4.1 and 4.2. Let $r, \alpha \in (0, 2]$, $\sigma > 0$, $\beta \in [-1, 1]$, and let $a = (a_1, ..., a_n) \in \mathbf{R}_+^n$ and $b = (b_1, ..., b_n) \in \mathbf{R}_+^n$ be such that $(a_1^r, ..., a_n^r) \prec (b_1^r, ..., b_n^r)$ and $(a_1^r, ..., a_n^r)$ is not a permutation of $(b_1^r, ..., b_n^r)$ (clearly, $\sum_{i=1}^n a_i \neq 0$ and $\sum_{i=1}^n b_i \neq 0$). Let $X_1, ..., X_n$ be independent r.v.'s such that $X_i \sim S_\alpha(\sigma, \beta, 0)$, $i = 1, ..., n$. It is not difficult to see that if $c = (c_1, ..., c_n) \in \mathbf{R}_+^n$, $\sum_{i=1}^n c_i \neq 0$, then $\sum_{i=1}^n c_i X_i / \left(\sum_{i=1}^n c_i^\alpha\right)^{1/\alpha} \sim S_\alpha(\sigma, \beta, 0)$. Consequently, for $x \in \mathbf{R}$,

$$\psi(c, x) = P\Big(X_1 > x / \Big(\sum_{i=1}^n c_i^\alpha\Big)^{1/\alpha}\Big). \tag{7.1}$$

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

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

$\sum_{i=1}^n a_i^\alpha$, if $\alpha/r < 1$. This, together with (7.1), implies that

$$\psi(a,x) < \psi(b,x) \tag{7.2}$$

if $x > 0$, $\alpha > r$ or $x < 0$, $\alpha < r$, and

$$\psi(a,x) > \psi(b,x) \tag{7.3}$$

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

Suppose now that $X_1,...,X_n$ are i.i.d. r.v.'s such that $X_i \sim \underline{\mathcal{CS}}(r)$, $i = 1,...,n$. By definition of the class $\underline{\mathcal{CS}}(r)$, there exist independent r.v.'s Y_{ij}, $i = 1,...,n$, $j = 1,...,k$, such that $Y_{ij} \sim S_{\alpha_i}(\sigma_i,0,0)$, $\alpha_i \in (0,r)$, $\sigma_i > 0$, $i = 1,...,n$, $j = 1,...,k$, and $X_i = \sum_{j=1}^k Y_{ij}$, $i = 1,...,n$. By (7.2) and (7.3), for $j = 1,...,k$, the r.v. $\sum_{i=1}^n b_iY_{ij}$ is strictly more peaked than $\sum_{i=1}^n a_iY_{ij}$, that is, for all $x > 0$ and all $j = 1,...,k$,

$$P\Big(\Big|\sum_{i=1}^n a_iY_{ij}\Big| > x\Big) > P\Big(\Big|\sum_{i=1}^n b_iY_{ij}\Big| > x\Big). \tag{7.4}$$

The r.v.'s Y_{ij}, $i = 1,...,n$, $j = 1,...,k$, are symmetric and unimodal by Theorem 2.7.6 in Zolotarev (1986, p. 134). Therefore, from Theorem 1.6 in Dharmadhikari and Joag-Dev (1988, p. 13) it follows that the r.v.'s $\sum_{i=1}^n a_iY_{ij}$, $j = 1,...,k$, and $\sum_{i=1}^n b_iY_{ij}$, $j = 1,...,k$, are symmetric and unimodal as well. From Lemma in Birnbaum (1948) and its proof it follows that if X_1, X_2 and Y_1, Y_2 are independent absolutely continuous symmetric unimodal r.v.'s such that, for $i = 1,2$, X_i is more peaked than Y_i, and one of the two peakedness comparisons is strict, then $X_1 + X_2$ is strictly more peaked than $Y_1 + Y_2$. This, together with (7.4) and symmetry and unimodality of $\sum_{i=1}^n a_iY_{ij}$ and $\sum_{i=1}^n b_iY_{ij}$, $j = 1,...,k$, imply, by induction on k (see also Theorem 1 in Birnbaum, 1948, and Theorem 2.C.3 in Dharmadhikari and Joag-Dev, 1988), that $\psi(a,x) = 1/2P\big(\big|\sum_{j=1}^k\sum_{i=1}^n a_iY_{ij}\big| > x\big) > 1/2P\big(\big|\sum_{j=1}^k\sum_{i=1}^n b_iY_{ij}\big| > x\big) = \psi(b,x)$ for $x > 0$ and $\psi(a,x) = 1 - \psi(a,-x) < 1 - \psi(b,-x) = \psi(b,x)$ for $x < 0$. Therefore, the conclusion of Theorem 4.4 for the class $\underline{\mathcal{CS}}(r)$ holds. The part of Theorem 4.1 for the class $\overline{\mathcal{CS}}(r)$ might be proven in a completely similar way. The proof is complete.

Proof of Theorems 4.3 and 4.4. Theorem 4.3 for the case of stable i.i.d. r.v.'s $X_i \sim S_\alpha(\sigma,\beta,0)$, $i = 1,...,n$, and Theorem 4.4 for both the cases of stable distributions $S_\alpha(\sigma,\beta,0)$ and distributions from the class $\underline{\mathcal{CS}}(1)$ are immediate consequences of Theorems 4.1 and 4.2 with $r = 1$. Let us prove Theorem 4.3 for the case of the class $\overline{\mathcal{CSLC}}$. Let vectors $a = (a_1,...,a_n) \in \mathbf{R}_+^n$ and $b = (b_1,...,b_n) \in \mathbf{R}_+^n$ be such that $a \prec b$ and a is not a permutation of b. Let $X_1,...,X_n$ be i.i.d. r.v.'s such that $X_i \sim \overline{\mathcal{CSLC}}$, $i = 1,...,n$. By definition, $X_i = \gamma Y_{i0} + \sum_{j=1}^k Y_{ij}$, $i = 1,...,n$, where $\gamma \in \{0,1\}$, $k \geq 0$ and $(Y_{1j},...,Y_{nj})$, $j = 0,1,...,k$, are independent vectors with i.i.d. components such that $Y_{i0} \sim \mathcal{LC}$, $i = 1,...,n$, and $Y_{ij} \sim S_{\alpha_i}(\sigma_i,0,0)$, $\alpha_i \in (1,2]$, $\sigma_i > 0$, $i = 1,...,n$, $j = 1,...,k$. From (7.2) and Proposition 2.1 in the introduction it follows that, for $j = 0,1,...,k$, the r.v. $\sum_{i=1}^n a_iY_{ij}$ is strictly more peaked than $\sum_{i=1}^n b_iY_{ij}$. Furthermore, from Theorem 2.7.6 in Zolotarev (1986, p. 134) and Theorems 1.6 and 1.10 in Dharmadhikari and Joag-Dev (1988, pp. 13 and 20) by induction on i it follows that the r.v.'s $\sum_{i=1}^n a_iY_{ij}$ and $\sum_{i=1}^n b_iY_{ij}$, $j = 0,1,...,k$, are symmetric and unimodal. Similar to the proof of Theorems 4.1 and 4.2, by Lemma in Birnbaum (1948) and its proof and induction, this implies that $\sum_{i=1}^n a_iX_i = \gamma\sum_{i=1}^n a_iY_{i0} + \sum_{j=1}^k\sum_{i=1}^n a_iY_{ij}$ is strictly more peaked than $\sum_{i=1}^n b_iX_i = \gamma\sum_{i=1}^n b_iY_{i0} + \sum_{j=0}^k\sum_{i=1}^n b_iY_{ij}$. This completes the proof of Theorem 4.3.

Proof of Corollary 5.1 . The corollary follows from Theorem 4.3 and relations (2.1) since $\overline{X}_n$ is weakly consistent for μ under its assumptions.

Proofs of Corollaries 5.2-5.5. It is easy to observe (see Marshall and Olkin, 1979, p. 7) that

$$\Big(\sum_{i=1}^n a_i/n, ..., \sum_{i=1}^n a_i/n\Big) \prec (a_1, ..., a_n) \prec \Big(\sum_{i=1}^n a_i, 0, ..., 0\Big), \tag{7.5}$$

for all $a \in \mathbf{R}^n_+$. These relations imply that

$$\big(\sum_{i=1}^n w_i^r/n, ..., \sum_{i=1}^n w_i^r/n\big) \prec (w_1^r, ..., w_n^r) \prec \big(\sum_{i=1}^n w_i^r, 0, ..., 0\big)$$

for all portfolio weights w and all $r \in (0, 2]$. From the above majorization comparisons and Theorem 4.2 it follows that, under the assumptions of Corollary 5.5, for all $\alpha \in (0, 1/2)$ and all w such that $w \neq \underline{w}$ and w is not a permutation of $\overline{w}$,

$$P(Z_{\underline{w}} > VaR_\alpha(Z_{\underline{w}})) = \alpha = P(Z_w > VaR_\alpha(Z_w)) < P\big(Z_{\underline{w}} > n^{1/r-1} VaR_\alpha(Z_w)/\big(\sum_{i=1}^n w_i^r\big)^{1/r}\big),$$

$$P(Z_{\overline{w}} > VaR_\alpha(Z_{\overline{w}})) = \alpha = P(Z_w > VaR_\alpha(Z_w)) > P\big(Z_{\overline{w}} > VaR_\alpha(Z_w)/\big(\sum_{i=1}^n w_i^r\big)^{1/r}\big).$$

This implies the bounds in Corollary 5.5. Sharpness of the bounds in the corollary follows from the fact that, as it is not difficult to see, the bounds become equalities in the limit as $\alpha \to r$ for symmetric stable r.v.'s $X_i \sim S_\alpha(\sigma, 0, 0)$, $i = 1, ..., n$. Corollaries 5.2-5.4 might be proven in a similar way, with the use of Theorems 4.1, 4.3 and 4.4 instead of Theorem 4.2 (the strict versions of inequalities (2.2) in Corollary 5.2 are consequences of bounds in Corollary 5.5 with $r = 1$). Corollaries 6.1 and 6.2 immediately follow from Corollaries 5.2 and 5.3.

Part II

OPTIMAL BUNDLING STRATEGIES FOR COMPLEMENTS AND SUBSTITUTES WITH HEAVY-TAILED VALUATIONS

8 Discussion of the results in Part II

8.1 Optimal bundling decisions for a multiproduct monopolist

Recent years have witnessed a surge in the interest in the analysis of optimal bundling strategies for a multiproduct monopolist. Many studies in the bundling literature emphasized that a monopoly's bundling decisions depend on correlations between consumers' valuations for the products (see Adams and Yellen, 1976, McAfee et. al., 1989, Schmalensee, 1984 and Salinger, 1995), the degrees of complementarity and substitutability between the goods (e.g., Lewbel, 1985, and Venkatesh and Kamakura, 2003) and the marginal costs for the products (see, among others, Salinger, 1995, and Venkatesh and Kamakura, 2003). Most of the papers in this stream of research have focused on prescribed distributions for reservation prices in the case of two products and their packages (such as bivariate uniform or Gaussian distributions) and only a few general results are available for larger bundles. Palfrey (1983) obtained results that give conditions under which the market participants prefer (*ex ante*) a single bundled Vickrey auction to separate provision of an arbitrary number of independently priced goods.[18] Palfrey (1983) also showed that, in the case of two bidders with additive valuations for bundles, the seller maximizes her profit by selling the goods in a single bundle; the two buyers, however, unanimously prefer separate provision of objects to any other bundling decision. From the results in Palfrey (1983) it further follows that, if stand-alone valuations are concentrated on a finite interval, then consumers never unanimously prefer separate provision of items to a single Vickrey auction, *ex ante*, if there are more than two buyers. In a related paper, Chakraborty (1999) obtained characterizations of optimal bundling strategies for a monopolist providing two goods on Vickrey auctions under a regularity condition on quantiles of reservation prices which is implied, in the case of symmetry, by subadditivity property (2.3). Bakos and Brynjolfsson (1999) investigated the optimal bundling decisions for a multiproduct monopolist providing bundles of independently priced goods with zero marginal costs (information goods) for profit-maximizing prices to consumers with valuations that have the monotone peakedness of the sample mean (MPSM) property discussed in Subsection 2.3.[19] Among other results, Bakos and Brynjolfsson (1999) showed that, in such a setting, if the seller prefers bundling a certain number of goods to selling them separately and if the optimal price per good for the bundle is less than the mean valuation, then bundling any greater number of goods will further increase the seller's profits, compared to the case when the additional goods are sold separately. According to the result, if consumers' valuations have the MPSM property, then a form of superadditivity for bundling decisions holds: the benefits to the seller grow

[18] The goods provided by the monopolist are said to be *independently priced* if consumers' valuations for their bundles are additive in those for the component goods, as opposed to the case of *interrelated goods*, e.g., substitutes or complements (see Dansby and Conrad (1984), Lewbel (1985), Venkatesh and Kamakura (2003) and Section 9 in the present paper).

[19] As discussed above, by Proposition 2.1, this condition is satisfied for log-concavely distributed valuations symmetric about the mean reservation price. In particular, the condition holds for valuations with a finite support $[\underline{v}, \overline{v}]$ distributed as the truncation $XI(|X-\mu| < h)$ of an arbitrary r.v. X with a log-concave density symmetric about $(\underline{v} + \overline{v})/2$, where $I(\cdot)$ is the indicator function (see also Remark 2 in An (1998)).

as the number of goods in the bundle increases.[20] Recently, applying Proposition 2.1, Fang and Norman (2003) showed that a multiproduct monopolist providing bundles of independently priced goods to consumers with log-concavely distributed valuations prefers selling them separately to any other bundling decision if the marginal costs of all the products are greater than the mean valuation; under some additional distributional assumptions, the seller prefers providing the goods as a single bundle to any other bundling decision if the marginal costs of the goods are identical and are less than the mean reservation price. To our knowledge, modelling of the optimal bundling problem in the case of an arbitrary number of goods in the literature was based on the assumptions that the products are independently priced and consumers' valuations for them have extremely light-tailed distributions (e.g., distributions with log-concave densities). The main intuition behind the analysis of optimal bundling decisions (see the discussion in Palfrey, 1983, Schmalensee, 1984, Salinger, 1995, Bakos and Brynjolfsson, 1999, and Fang and Norman, 2003) in such a setting is that, for light-tailed distributions such as those with log-concave densities, consumers' valuations per good for a bundle typically have a lower variance relative to the valuations for individual goods.[21]

8.2 The main results of Part II: optimal bundling decisions for a multiproduct monopolist in the case of long-tailed reservation prices and interrelated goods

This part of the paper contributes to the existing literature on bundling and thick-tailedness in economics and finance in a number of ways. First, we develop a framework that allows one to model the optimal bundling problem of a multiproduct monopolist providing large bundles of interrelated goods with an arbitrary degree of complementarity or substitutability. Second, we derive characterizations of optimal bundling strategies for the seller in this setup in the case of long-tailed valuations and tastes for the products, where no results in the literature are available even in the case of two goods, to our knowledge. Third, our analysis provides a unified approach to the study of optimal bundling problems in the case where the goods are provided on an auction as well as in the case where the prices for the products are set by the monopolist (see Sections 10 and 11 that present the main results of this part of the paper along the above three lines). In particular, the approach reveals that the analysis of optimal bundling strategies in both of the above cases is based on the same probabilistic concepts and results discussed in Part I of this paper.

Moreover, our study shows that patterns in the optimal bundling strategies are the opposites of one another, depending on the degrees of thick-tailedness of consumers' valuations and the degrees of complementarity and substitutability among the goods provided (e.g., Theorems 10.1 and 10.2 and Theorems 11.1 and 11.2). In that, for instance, the solutions to the optimal bundling problem with not extremely long-tailed valuations are the opposites of those in the case of extremely light-tailed reservation prices, even in the case of independently priced goods (Theorems 11.3 and 11.4). In particular, our study shows that a number of results available in the literature for independently priced goods with extremely light-tailed valuations (such as those with log-concave densities) are reversed in the case of valuations with very thick-tailed distributions. However, the results for extremely light-tailed reservation prices continue to hold under the assumption that distributions of the reservation prices are not too thick-tailed (see Theorems 11.1-11.4 and the discussion below). In other words, the optimal bundling strategies analyzed in the

[20]This property is similar to the case of Vickrey auctions with two buyers (see Remark 2 in Palfrey (1983)).

[21]Further explanation for bundling is that, for light-tailed distributions, it reduces uncertainty about consumers' valuations and leads to a decrease in extreme values of the distribution of valuations per good, thereby reducing buyer diversity and increasing the predictive power of the selling strategy (see Schmalensee, 1984, and Bakos and Brynjolfsson, 1999). Recently, Nalebuff (2004) showed that, in addition to being an effective tool of price discrimination, bundling is also a credible tool to protect a monopolist against entry.

literature in the case of light-tailed valuations are robust to thick-tailedness assumptions on consumers' valuations as long as the distributions entering the assumptions are not extremely heavy-tailed. However, they are reversed to assumptions that involve extremely long-tailed distributions.

In addition, the approach developed in this part of the paper allows one to study, in a unified way, both the seller's and consumers' preferences over bundling decisions. According to our analysis, the seller's and the buyers' *ex-ante* preferences over bundles continue to be the opposites of one another in the case of thick-tailed valuations and an arbitrary number of bidders, similar to the *ex-post* results for two buyers in Palfrey, 1983 (see Corollary 10.2 and Theorem 10.3). However, in the framework with extremely long-tailed valuations, consumers' surplus is maximized under separate provision of independently priced goods, regardless of the number of buyers. This conclusion established in the present paper is in contrast with the results available in the literature for extremely light-tailed case (see Remark 10.3).

We obtain complete characterizations of solutions to the seller's optimal bundling problem in the case when complements or substitutes are provided through a Vickrey (second price sealed bid) or any other revenue equivalent auction[22] and consumers' tastes for the goods are heavy-tailed. Among other results, we show that if the goods provided on a Vickrey auction are substitutes (or complements with not very high degree of complementarity) and bidders' tastes for the objects are not extremely heavy-tailed, then the risk-neutral monopolist strictly prefers separate provision of the products to any other bundling decision (Theorem 10.1). The results are reversed, however, in the case of a risk-averse auctioneer providing complements (or substitutes with not very high degree of substitutability) to consumers with extremely long-tailed tastes for the products (Theorem 10.2).[23] According to our analysis, in such a setting, regardless of the number of consumers, the seller always strictly prefers providing the goods on a single Vickrey auction to any other bundling decision, as in the setting with two buyers in Palfrey (1983). This conclusion provides, in particular, a reversal of the results in Chakraborty (1999) from which it follows that, in the case of symmetric valuations satisfying comparisons that hold for distributions with log-concave densities, provision of independently priced goods through separate Vickrey auctions generates larger expected profits to the seller than any other bundling decision if the number of buyers is sufficiently large. We also obtain a characterization of consumers' preferences over the monopolist's bundling decision in a Vickrey auction in the case of heavy-tailed valuations for the products. We show, for instance, that if bidders' reservation prices for independently priced goods are extremely heavy-tailed, as modelled by positive stable distributions (see Section 3), then they unanimously prefer separate Vickrey auctions to any other bundling decision (Theorem 10.3). These results are at odds with a setting where valuations have a finite distributional support in which case, according to Palfrey (1983), consumers never unanimously prefer separate auctions if there are more than two buyers, as discussed in Subsection 8.1. By the revenue equivalence principle, a number of results on the seller's optimal bundling strategies continue to hold for first price sealed bid auction as well as for ascending and descending auctions. The characterizations of consumers' preferences over Vickrey auctions obtained in this part of the paper continue to hold in the case of ascending auctions with private values, due to their weak equivalence.

[22]More precisely, through any other auction over bundles that generates the same expected revenues in the case of a risk-neutral seller and the same utilities of wealth in the case when the seller is risk-averse or risk-neutral.

[23]The assumption of seller's risk aversion is necessary in the case of extremely heavy-tailed tastes and valuations since otherwise the monopolist's expected profit is infinite for any bundling decision.

We also obtain characterizations of optimal bundling strategies for a monopolist who provides goods with an arbitrary degree of complementarity or substitutability to consumers with heavy-tailed tastes for profit-maximizing prices (Theorems 11.1-11.4). We show, in particular, that, for products with high marginal costs, the seller's optimal strategy is to provide complements with extremely heavy-tailed consumers' tastes for them separately and those with sufficiently light-tailed valuations as a single bundle. For relatively low marginal costs, these conclusions are reversed (Theorems 11.1 and 11.2). In particular, contrary to the case where goods with extremely light-tailed valuations are considered, as in Bakos and Brynjolfsson (1999) and Fang and Norman (2003), if consumers' tastes for the products are extremely long-tailed, then the monopolist's optimal strategy is to provide independently priced goods with relatively high marginal costs as a single bundle and those with sufficiently low marginal costs separately (Theorem 11.4). Our results imply, for instance, that for positive stable distributions of tastes, irrespective of the marginal costs of producing the goods in question, the optimal strategy is to provide the goods as a single bundle if the goods are independently priced or are complements (or if the goods are substitutes with not very high degree of substitutability).

It is important to note that the results in this part of the paper shed new light on bundling strategies observed in real-world markets, in particular, those that involve exclusion of goods for which observations of extreme (both positive and negative) valuations are more likely from the bundle and selling them separately. Such strategies are often observed in the market, in particular, in the bundling decisions of cable and direct satellite broadcast television firms that have marginal costs of reproduction close to zero. These firms typically offer a "basic" bundle and use such strategies as pay-per-view approach for unusual special events such as boxing matches (see Bakos and Brynjolfsson, 1999). The high valuations for the special events are concentrated among a small fraction of consumers and thus are likely to be extremely heavy-tailed. Therefore, the optimal bundling strategies for the special events are likely to be the opposites of those for light-tailed distributions of valuations and thus, in contrast to the basic bundles, the events are likely to be provided on pay-per-view basis. Season tickets for entertainment performances offered by sporting and cultural organizations that have sufficiently high marginal costs of production might illustrate the dual pattern in bundling. It seems plausible that most of the demand for season tickets is concentrated around a relatively small fraction of consumers that have extremely high valuations for performances offered by the entertainment organization. The optimal strategy is to offer tickets to such consumers as a bundle, as predicted by our results for heavy-tailed tastes in the case of sufficiently large marginal costs. This strategy is the opposite of separate provision of the most of tickets to performances to consumers who are likely not to have extreme valuations.

The underlying intuition that drives our results on bundling is closely related to that based on the behavior of variance in the world of extremely light-tailed valuations (see Subsection 8.1). Namely, our majorization results established in Part I imply, essentially, that, in the case of reservation prices with not extremely heavy tails, the consumers' valuations per good for bundles of products *always* have less spread relative to the valuations for component goods, as measured by their peakedness. On the other hand, in the case of extremely heavy-tailed valuations, the spread of reservation prices per good for bundles is *always* greater than that of valuations for components (the reader is referred to Sections 10 and 11 for more on the intuition).

8.3 Organization of Part II

This part of the paper is organized as follows: Section 9 develops a framework for the analysis of optimal bundling with interrelated goods. Section 10 presents the main results of Part II on the optimal bundling strategies for a multiproduct auctioneer in the case of heavy-tailed tastes for and an arbitrary degree of complementarity or substitutability among the goods provided. Section 11 contains the main results of this part of the paper on characterizations of optimal bundling strategies for a seller providing interrelated goods to consumer with thick-tailed valuation for profit-maximizing prices. Section 12 contains proofs of the results obtained in this part of the paper.

9 A model for optimal bundling of interrelated goods

Consider a setting with a single profit-maximizing risk-neutral seller providing m goods to n consumers.[24] Let $M = \{1, 2, ..., m\}$ be the set of goods sold on the market and let $J = \{1, 2, ..., n\}$ denote the set of buyers. Let 2^M be the set of all subsets of M. As in Palfrey (1983), the seller's bundling decisions $\mathcal{B}$ are defined as partitions of the set of items M into a set of subsets, $\{B_1, ..., B_l\} = \mathcal{B}$, where l is the cardinality of $\mathcal{B}$; the subsets $B_s \in 2^M$, $s = 1, ..., l$, are referred to as bundles. That is, $B_s \neq \emptyset$ for $s = 1, ..., l$; $B_s \cap B_t =$ for $s \neq t$, $s, t = 1, ..., l$; and $\cup_{s=1}^{l} B_s = M$ (see Palfrey, 1983, and Fang and Norman, 2003). It is assumed that the seller can offer one (and only one) partition $\mathcal{B}$ for sale on the market (this referred to as pure bundling, see Adams and Yellen, 1976).[25] We denote by $\underline{\mathcal{B}} = \{\{1\}, \{2\}, ..., \{n\}\}$ and $\overline{\mathcal{B}} = \{1, 2, ..., n\}$ the bundling decisions corresponding, respectively, to the cases when the goods are sold separately (that is, on separate auctions or using unbundled sales) and as a single bundle M. For a bundle $B \in 2^M$, we write $card(B)$ for a number of elements in B and denote by π_B the seller's profit resulting from selling the bundle. For a bundling decision $\mathcal{B} = \{B_1, ..., B_l\}$, we write $\Pi_{\mathcal{B}}$ for the seller's total profit resulting from following $\mathcal{B}$, that is, $\Pi_{\mathcal{B}} = \sum_{s=1}^{l} \pi_{B_s}$. The risk-neutral seller (strictly) prefers a bundling decision $\mathcal{B}_1$ to a bundling decision $\mathcal{B}_2$ *ex ante* if $E\Pi_{\mathcal{B}_1} \geq E\Pi_{\mathcal{B}_2}$ (resp., if $E\Pi_{\mathcal{B}_1} > E\Pi_{\mathcal{B}_2}$). The seller prefers a bundling decision $\mathcal{B}_1$ to a bundling decision $\mathcal{B}_2$ *ex post* if $\Pi_{\mathcal{B}_1} \geq \Pi_{\mathcal{B}_2}$ (a.s.), that is, if $P(\Pi_{\mathcal{B}_1} \geq \Pi_{\mathcal{B}_2}) = 1$.

A representative consumer's preferences over the bundles $B \in 2^M$, on the other hand, are determined by her reservation prices (valuations) $v(B)$ for the bundles and, in particular, by their valuations $v(\{i\})$ for goods $i \in M$ (when the goods are sold separately) which are referred to as stand-alone reservation prices. In the case when the reservation prices for bundles are nonnegative: $v(B) \geq 0$, $B \in 2^M$, it is said that the goods in M and their bundles satisfy the *free disposal* condition.[26] The free disposal assumption is particularly important in the case of information goods and in the economics of the Internet (see Bakes and Brynjolfsson, 1999, 2000). If consumers' valuations for a bundle of goods is additive in those of component goods: $v(B) = \sum_{i \in B} v(\{i\})$, then the products provided by the monopolist are said to be *independently priced* (see Venkatesh and Kamakura, 2003). Under free

[24] So that the seller's utility of wealth function is linear.

[25] The analysis of *mixed* bundling, in which consumers can choose among *all* bundling decisions available (see Adams and Yellen, 1976, and McAfee et. al., 1989) is beyond the scope of this paper.

[26] The case when the support of the valuations $v(B)$ intersects with $(-\infty, 0)$ corresponds to the situation when the goods have negative value to some consumers (e.g., articles exposing certain political views, advertisements or pornography in the case of information goods, see Bakos and Brynjolfsson, 1999).

disposal, the natural analogues of this property for interrelated goods are subadditivity $v(B) \le \sum_{i\in B} v(\{i\})$ in the case of substitutes and superadditivity $\sum_{i\in B} v(\{i\}) \le v(B)$ in the case of complements (see Dansby and Conrad, 1984, Lewbel, 1985, Section 16.3 in Krishna, 2002, and Venkatesh and Kamakura, 2003).

Throughout this section, X_i, $i \in M$, denote i.i.d. r.v.'s representing the distribution of consumers' tastes for goods $i \in M$ that determine their reservation prices for the goods and their bundles. We suppose that a representative consumer's reservation price $v(B)$ for a bundle B of goods produced by the monopolist is a function of her tastes for the component goods in the bundle. More precisely, we model the setting with interrelated goods by assuming that a representative consumer's valuations for bundles $B \in 2^M$ are given by $v(g_r, B) = g_r\big(\sum_{i\in B} X_i\big)$ or $v(h_r, B) = h_r\big(\sum_{i\in B} X_i\big)$ where, for $r \in (0, 2]$, $g_r(x) = x^r I(x \ge 0)$, $h_r(x) = x|x|^{r-1}$, $x \in \mathbf{R}$, and $I(\cdot)$ denotes the indicator function. The valuations for goods $i \in M$ in the case when they are sold separately are thus $v(g_r, \{i\}) = g_r(X_i)$ or $v(h_r, \{i\}) = h_r(X_i)$, $i \in M$. Clearly, in the case $r = 1$, one has $v(h_1, \{i\}) = h_1(X_i) = X_i$, $i \in M$. Also, the reservation prices $v(g_r, B)$ satisfy the free-disposal condition: $v(g_r, B) \ge 0$ for all $B \in 2^M$. It is easy to see that, for all $B \in 2^M$, $v(g_r, B) \le \sum_{i\in B} v(g_r, \{i\})$, if $r \le 1$, and $\sum_{i\in B} v(g_r, \{i\}) \le v(g_r, B)$, if $r \ge 1$, and $X_i \ge 0$, $i \in B$. That is, consumers' reservation price $v(g_r, B)$ for a bundle is subadditive in those for the component products if $r \le 1$, as it is typically required for substitutes, and is superadditive in the rectangle of non-negative tastes if $r \ge 1$, as it is usually assumed in the case of complements. Similarly, for $r \le 1$, the reservation prices $v(h_r, B)$ are subadditive in those for component products in the rectangle of non-negative stand-alone valuation $v(h_r, \{i\})$ $i \in M$, and are superadditive in the components' valuations in the case when all the stand-alone valuations are non-positive. For $r \ge 1$, the valuations for bundles $v(h_r, B)$ are superadditive in those for the components if all the stand-alone reservation prices are non-negative and are subadditive if the valuations for all component products are non-positive. More precisely, if $v(h_r, \{i\}) \ge 0$, $i \in B$, then $\sum_{i\in B} v(h_r, \{i\}) \le v(h_r, B)$ for $r \ge 1$, and $v(h_r, B) \le \sum_{i\in B} v(h_r, \{i\})$ for $r \le 1$. If $v(h_r, \{i\}) \le 0$, $i \in B$, then $v(h_r, B) \le \sum_{i\in B} v(h_r, \{i\})$ for $r \ge 1$, and $\sum_{i\in B} v(h_r, \{i\}) \le v(h_r, B)$ for $r \le 1$. The above super- and subadditivity properties of $v(h_r, B)$ for $r \ge 1$ are consistent with the assumption typically imposed on the value function of (complementary) gains and losses in mental accounting and prospect theory (see, e.g., Kahneman and Tversky, 1979, and Thaler, 1985). The case $r = 1$ with reservation prices for bundles $v(h_1, B) = \sum_{i\in B} X_i$ models the case of independently priced goods.

For $j \in J$, the jth consumer's tastes for goods in M are assumed to be $\tilde{X}_{ij}$, $i \in M$, where $\tilde{X}^{(j)} = (\tilde{X}_{1j}, ..., \tilde{X}_{nj})$, $j \in M$, are independent copies of the vector $(X_1, ..., X_n)$, and her reservation prices $v_j(B)$ for bundles $B \in 2^M$ of goods in M are given by $v_j(g_r, B) = g_r\big(\sum_{i\in B} \tilde{X}_{ij}\big)$ or $v_j(h_r, B) = h_r\big(\sum_{i\in B} \tilde{X}_{ij}\big)$. The seller is assumed to know only the distribution of consumers' reservation prices for goods in M and their bundles. The valuations $v_j(g_r, B)$ ($v_j(h_r, B)$) for bundles $B \in 2^M$, are known to buyer j, however, the buyer has only the same incomplete information about the other consumers' reservation prices as does the seller (see Palfrey, 1983).

10 Optimal bundled auctions for complements and substitutes with thick-tailed tastes

Let us consider first the case in which the goods in M and their bundles are provided by a risk-neutral seller through Vickrey auctions (see Palfrey, 1983). In this setting, the buyers submit simultaneous sealed bids for bundles of goods sold by the seller. The bidder with the highest bid wins the auction and pays the seller the second highest bid. It is well-known that, in such a setup, a dominant strategy for each bidder is to bid her true reservation prices. In accordance with the assumption of nonnegativity of bids and valuations usually imposed in the auction theory, we suppose first that, for $j \in J$, the jth consumer's reservation price for a bundle $B \in 2^M$ of goods sold is given by $v_j(g_r, B) = g_r(\sum_{i \in B} \tilde{X}_{ij}) \geq 0$. The seller's profit from following a bundling decision $\mathcal{B} = \{B_1, ..., B_l\}$ is, evidently, $\sum_{s=1}^{l} v_{(n-1)}(g_r, B_s)$, where, for $s = 1, ..., l$, $v_{(n-1)}(g_r, B_s)$ denotes the second highest of consumers' reservation prices for the bundle B_s (that is, the second highest order statistic of the reservation prices for the bundle). The following Theorem 10.1 extends the results in Palfrey (1983) and Chakraborty (1999) to the case of interrelated goods (with an arbitrary degree of complementarity or substitutability) and consumers with long-tailed tastes. According to the theorem, if consumers' tastes are not extremely heavy-tailed and the goods are substitutes (or are complements with not very high degree of complementarity) then the auctioneer strictly prefers separate provision of goods to any other bundling decision.

Theorem 10.1 *Let $r \in (0, 2)$, and let the reservation prices for bundles $B \in 2^M$ of goods from M be given by $v(g_r, B)$. Suppose that the tastes X_i, $i \in M$, are i.i.d. r.v.'s such that $X_i \sim S_\alpha(\sigma, \beta, 0)$, $i \in M$, for some $\sigma > 0$, $\beta \in [-1, 1]$ and $\alpha \in (r, 2]$, or $X_i \sim \overline{\mathcal{CS}}(r)$, $i \in M$. Then, for all $n \geq 2$, the seller strictly prefers (ex ante) $\underline{\mathcal{B}}$ (that is, n separate Vickrey auctions) to any other bundling decision.*

Remark 10.1. From the proof of Theorem 10.1 it follows that, under its assumptions, for any bundle $B \in 2^M$ with the number of elements $card(B) = k \geq 2$, the seller's profit π_B from selling B on a Vickrey auction is strictly (first-order) stochastically dominated by the profit from selling one of goods in B, say good $i \in B$, separately k times, that is, by the r.v. $k\pi_i$, where $\pi_i = \pi_{B_i}$ with $B_i = \{i\}$. Namely, for all $x > 0$, one has $P(\pi_B > x) < P(k\pi_i > x)$ that means that selling one of goods in B k times separately is always likely to generate higher profits to the seller than selling the bundle B. By Shaked and Shanthikumar (1994, pp. 3-4), we get, therefore, similar to Remark 6.1, that $EU(\pi_B) \leq EU(k\pi_i)$ for all increasing functions $U : \mathbf{R}_+ \to \mathbf{R}$ for which the expectations exist. Similar to the proof of Theorem 10.2 below, this, in turn, implies that Theorem 10.1 holds as well in the case of a *risk-loving* seller with any increasing *convex* utility of wealth function U such that $U(0) = 0$.

The intuition behind the results given by Theorem 10.1 is that, in the case of tastes with not extremely heavy-tailed tastes, similar to the case of log-concave distributions, the valuations per good become increasingly more concentrated about the mean valuations with the size of bundles. In particular, in the case of not extremely long-tailed reservation prices, buyers with high valuations for the bundle are more likely to win the bundled auction and the next highest bidder is likely to have relatively lower valuations than in the case of separate auctions. Since it is increasingly likely that at least one of the buyers will have valuations on the upper tail of the distribution as the number of bidders gets larger, it becomes more likely that the winner of the auction prefers bundled auctions (see Palfrey, 1983).

There are no counterparts of Theorem 10.1 for extremely heavy-tailed distributions of consumers' valuations (such as $\underline{\mathcal{CS}}(r)$) if the seller's utility of wealth is linear since, as it is not difficult to see, in this case, the seller's expected profits from following any bundling decision are infinite. However, in the case of a risk-averse seller with a concave utility of wealth function, the following Theorems 10.2 and 10.2 that give reversals of the results in Theorem

10.1 hold.

Suppose that the seller (strictly) prefers a bundling decision $\mathcal{B}_1$ to a bundling decision $\mathcal{B}_2$ if $EU(\Pi_{\mathcal{B}_1}) \geq EU(\Pi_{\mathcal{B}_2})$ (resp., if $EU(\Pi_{\mathcal{B}_1}) > EU(\Pi_{\mathcal{B}_2})$), where $U : \mathbf{R}_+ \to \mathbf{R}$ is an increasing concave function with $U(0) = 0$ (that represents the seller's utility of wealth satisfying the property of diminishing returns). According to the following Theorem 10.2, in this case, the auctioneer strictly prefers providing all the items through one Vickrey auction to any other bundling decision, if consumers' tastes are extremely heavy-tailed and the goods are complements (or are substitutes with not very high degree of substitutability).

Theorem 10.2 *Let $r \in (0, 2]$, and let the reservation prices for bundles $B \in 2^M$ of goods from M be given by $v(g_r, B)$. Suppose that the tastes X_i, $i \in M$, are i.i.d. r.v.'s such that $X_i \sim S_\alpha(\sigma, \beta, 0)$, $i \in M$, for some $\sigma > 0$, $\beta \in [-1, 1]$ and $\alpha \in (0, r)$, or $X_i \sim \underline{\mathcal{CS}}(r)$, $i \in M$. If the seller's utility of wealth is concave, then, for all $n \geq 2$, the seller strictly prefers (ex ante) $\overline{\mathcal{B}}$ (that is, a single Vickrey auction) to any other bundling decision.*

Theorems 10.1 and 10.2 demonstrate that the seller's optimal bundling decisions depend crucially on both tail indices α of the tastes as well as on the degree r of complementarity or substitutability among the products provided. In addition, it is important to note that the classes of extremely thick-tailed distributions covered by the theorems contain many examples with non-negative tastes for individual goods. Namely, as discussed in Section 3, any extremely heavy-tailed distribution with $\alpha < 1$ and $\beta = 1$ is concentrated on the positive semi-axis. In particular, such extremely thick-tailed distributions include the Lévy distribution with the density $f(x) = (\sigma/(2\pi))^{1/2} exp(-\sigma/(2x))x^{-3/2}$, $x \geq 0$, for which $\alpha = 1/2$ and $\beta = 1$. Evidently, in the case of nonnegative tastes and $r = 1$, one has $v(g_1, B) = \sum_{i \in B} v(\{i\})$, so that this corresponds to the case of independently priced goods. Due to importance of the case of nonnegative stand-alone tastes and the case of additive valuations, we formulate the following immediate corollaries of Theorems 10.1 and 10.2 in such settings.

The conclusions of Corollaries 10.1 and 10.2, are of course, similar to those of Theorems 10.1 and 10.2: if consumers' tastes are nonnegative and extremely thick-tailed, then the seller prefers providing independently priced goods and complements (and substitutes with relatively low degree of substitutability) on a single auction. However, she prefers providing substitutes separately if the degree of substitutability among the goods is relatively high.

Corollary 10.1 *Let the tastes X_i, $i \in M$, for goods in M be nonnegative extremely heavy-tailed i.i.d. r.v.'s such that $X_i \sim S_\alpha(\sigma, 1, 0)$ for some $\sigma > 0$ and $\alpha \in (0, 1)$. Further, let the reservation prices for bundles $B \in 2^M$ of goods from M be given by $v(g_r, B)$, $r \in (0, 2]$. If $r > \alpha$, then, for all $n \geq 2$, the risk-neutral seller strictly prefers (ex ante) $\underline{\mathcal{B}}$ (that is, n separate Vickrey auctions) to any other bundling decision. If $r < \alpha$ and the seller's utility of wealth is concave, then, for all $n \geq 2$, the seller strictly prefers (ex ante) $\overline{\mathcal{B}}$ (that is, a single Vickrey auction) to any other bundling decision.*

Corollary 10.2 *Let the reservation prices for bundles $B \in 2^M$ be given by $v(B) = \sum_{i \in B} X_i$. Suppose that the stand-alone reservation prices X_i, $i \in M$, for goods in M are nonnegative extremely heavy-tailed i.i.d. r.v.'s such that $X_i \sim S_\alpha(\sigma, 1, 0)$ for some $\sigma > 0$ and $\alpha \in (0, 1)$. If the seller's utility of wealth is concave, then, for all $n \geq 2$, the seller strictly prefers (ex ante) $\overline{\mathcal{B}}$ (that is, a single Vickrey auction) to any other bundling decision.*

Using the general majorization properties of long-tailed distributions presented in Part I of the paper, one can also obtain the following Theorem 10.3 that characterizes buyers' preferences over the bundled auctions in the case of independently priced goods and extremely heavy-tailed nonnegative stand-alone reservation prices.

Let $j \in J$ and let $\tilde{x}^{(j)} = (\tilde{x}_{1j}, ..., \tilde{x}_{nj}) \in \mathbf{R}^n_+$. If a bundle B consisting of independently priced goods is offered for sale on a Vickrey auction then the expectation of the surplus $S_j(B, \tilde{x}^{(j)})$ to consumer j with the values of stand-alone reservation prices $\tilde{X}^{(j)} = \tilde{x}^{(j)}$ and induced valuations for bundles $v_j(B) = \sum_{i \in B} \tilde{x}_{ij}$, $B \in 2^M$, is (see Palfrey, 1983)

$$ES_j(B, \tilde{x}^{(j)}) = P\Big(\max_{s \in J, s \neq j} v_s(B) < v_j(B)\Big)\Big(v_j(B) - E\big(\max_{s \in J, s \neq j} v_s(B) | \max_{s \in J, s \neq j} v_s(B) < v_j(B)\big)\Big),$$

where $v_t(B) = \sum_{i \in B} \tilde{X}_{it}$, $B \in 2^M$, $t \in J$, $t \neq j$. If the seller follows a bundling decision $\mathcal{B} = \{B_1, ..., B_l\}$, then the expectation of the surplus $S_j(\mathcal{B}, \tilde{x}^{(j)})$ to the jth buyer with the vector of stand-alone valuations $\tilde{X}^{(j)} = \tilde{x}^{(j)}$ is $ES_j(\mathcal{B}, \tilde{x}^{(j)}) = \sum_{s=1}^{l} ES_j(B_s, \tilde{x}^{(j)})$. The jth buyer with $\tilde{X}^{(j)} = \tilde{x}^{(j)}$ is said to (strictly) prefer a bundling decision $\mathcal{B}_1$ to a bundling decision $\mathcal{B}_2$, *ex ante*, if $ES_j(\mathcal{B}_1, \tilde{x}^{(j)}) \geq ES_j(\mathcal{B}_2, \tilde{x}^{(j)})$ (resp., if $ES_j(\mathcal{B}_1, \tilde{x}^{(j)}) > ES_j(\mathcal{B}_2, \tilde{x}^{(j)})$). If all buyers $j \in J$ (strictly) prefer a bundling decision $\mathcal{B}_1$ to a bundling decision $\mathcal{B}_2$ *ex ante* for *almost all* realizations of their reservation prices $\tilde{X}^{(j)}$, it is said that buyers *unanimously* (strictly) prefer $\mathcal{B}_1$ to $\mathcal{B}_2$ *ex ante*. More precisely, buyers unanimously (strictly) prefer a partition $\mathcal{B}_1$ to a partition $\mathcal{B}_2$ if, for all $j \in J$, $P[E(S_j(\mathcal{B}_1, \tilde{X}^{(j)})|\tilde{X}^{(j)}) \geq E(S_j(\mathcal{B}_2, \tilde{X}^{(j)})|\tilde{X}^{(j)})] = 1$ (resp., $P[E(S_j(\mathcal{B}_1, \tilde{X}^{(j)})|\tilde{X}^{(j)}) > ES_j(\mathcal{B}_2, \tilde{X}^{(j)})|\tilde{X}^{(j)})] = 1$), where, as usual, $E(\cdot|\tilde{X}^{(j)})$ stands for the expectation conditional on $\tilde{X}^{(j)}$. Clearly, in the case of absolutely continuous reservation prices X_i, $i \in M$, consumers unanimously prefer $\mathcal{B}_1$ to $\mathcal{B}_2$ *ex ante* if each of them prefers $\mathcal{B}_1$ to $\mathcal{B}_2$ for all but a finite number of realizations of her stand-alone valuations.

According to Theorem 10.3, consumers unanimously prefer (*ex ante*) separate provision of goods in Vickrey auctions to any other bundling decision in the case of an arbitrary number of buyers, if their valuations are extremely long-tailed, as modelled by positive stable distributions. These results are reversals of those given by Theorem 6 in Palfrey (1983) from which it follows that if consumers' valuations are concentrated on a finite interval, then buyers never unanimously prefer separate provision auctions if there are more than two buyers on the market (Theorem 10.3 does not contradict Theorem 6 in Palfrey (1983) since the support of heavy-tailed distributions in Theorem 10.3 is the infinite positive semi-axis $\mathbf{R}_+$).

Theorem 10.3 *Let the reservation prices for bundles $B \in 2^M$ be given by $v(B) = \sum_{i \in B} X_i$. Suppose that the stand-alone reservation prices X_i, $i \in M$, for goods in M are i.i.d. r.v.'s such that $X_i \sim S_\alpha(\sigma, 1, 0)$ for some $\sigma > 0$ and $\alpha \in (0, 1)$. Then buyers unanimously strictly prefer (ex ante) $\underline{\mathcal{B}}$ (that is, n separate auctions) to any other bundling decision.*

The underlying intuition behind the reversals of the results on preferences over bundling decisions in Vickrey auctions in the case of extremely long-tailed tastes in Theorems 10.2 and 10.3 and Corollary 10.2 is that the distributions of the valuations for individual goods are more peaked than those of the valuations per good in bundles. This implies, in particular, that buyers who are on the upper tail of the distributions for the goods are more likely to win separate auctions and the next highest bidder is likely to have relatively lower valuations than in the case of a bundled auction. In the case of positive stable valuations, these implications hold even for any consumer, whatever the values of her reservation prices are. Therefore, contrary to the case of extremely light-tailed valuations (see the discussion preceding Theorem 5 in Palfrey, 1983), as the number of buyers gets larger, the winner of the auction is likely to prefer separate provision of the products.

Remark 10.2. In this paper's setting with private values and the assumptions that tastes for each good as well as each agent are independently distributed[27], first price sealed bid auction, Dutch and English auctions over bundles are revenue equivalent to the Vickrey auction for a risk-neutral seller (see Krishna, 2002). Therefore, Theorem 10.1 and the results in Corollaries 10.1 and 10.2 for a risk-neutral monopolist continue to hold for the above auctions as well. In addition to that, under the above assumptions, English ascending auctions over bundles are weakly equivalent to Vickrey auctions. Therefore, the results given by Theorem 10.3 continue to hold for English auctions as well.

Remark 10.3. As shown by Palfrey (1983), in Vickrey auctions with independently priced goods and an arbitrary number of bidders, the total surplus (that is, the sum of the seller's profit and buyers' surplus) is always maximized in the case when the goods are provided on separate auctions. Palfrey (1983) also proves that, under nonnegative valuations for individual goods and additive valuations for bundles, the seller prefers a single bundled Vickrey auction to any other bundling decision, *ex post* and thus *ex ante*, if there are two buyers. The two buyers, on the other hand, unanimously prefer separate provision of items *ex post* and thus *ex ante*. Since these results are, essentially, deterministic, all they are robust with respect to risk attitudes of the seller and the buyer.

The results in the present section for additive valuations and nonnegative extremely heavy-tailed stand-alone reservation prices are in accordance with the above findings for the two-buyer setting. Indeed, from Corollary 10.2 and Theorem 10.3 it follows that, in the case of an arbitrary number of buyers (in particular, in the case with two bidders) with extremely heavy-tailed nonnegative valuations for independently priced goods, the market participants' *ex ante* preferences over the bundling decisions are the same as in the case of the *ex post* analysis for two-buyer setting in Palfrey (1983). Namely, the seller's expected utility of wealth is maximized in the case of a single auction and the buyers unanimously prefer separate provision of goods to any other bundling decision. Thus, the effects of bundling on the seller's *expected* utility of wealth and the buyers' *expected* surplus continue to be the opposites of one another, although (by Palfrey, 1983) the expected total surplus is still maximized under the separate provision.

One should also note that, as discussed in Palfrey (1983), the *ex post* results on the seller's and the buyers' preferences available in the two-buyer setup cannot be extended in any way to the case when there are more than two buyers. Subsequently, Chakraborty (1999) obtained results on optimal bundling under conditions on quantiles similar to the VaR subadditivity (2.3). Chakraborty showed that, under these conditions, provision of independently priced goods through separate Vickrey auctions generates larger expected profits to the seller than any other bundling decision if the number of buyers is sufficiently large. As follows from the results in Part II of the paper, the assumptions on quantiles are satisfied for log-concave distributions and for not extremely heavy-tailed distributions.

On the other hand, Corollary 10.2 and Theorem 10.3 show that the *ex ante* results on optimal bundling with nonnegative stand-alone reservation prices and additive valuations for bundles hold *regardless* of the number of buyers, if their valuations are *extremely* thick-tailed. These conclusions do not contradict the above results in

[27] So that, interdependent values and affiliated signals (see Milgrom and Weber, 1982, and Krishna, 2002) are ruled out.

Palfrey (1983) and Chakraborty (1999) because they concern *ex ante* analysis and hold for extremely thick-tailed distributions, for which the conditions in Chakraborty (1999) (including the VaR subadditivity, see Part II) are not satisfied.[28]

11 Optimal bundling of interrelated goods under tastes' heavy-tailedness: the case of profit-maximizing prices

Let us now turn to the case in which the prices for goods on the market and their bundles are set by the monopolist. Let c_i, $i \in M$, be the marginal costs of goods in M. Suppose that the seller can provide bundles B of goods in M for prices per good $p \in [0, p_{max}]$, where p_{max} is the (regulatory) maximum price, with the convention that p_{max} can be infinite. For a bundle of goods $B \in 2^M$, denote by p_B the profit-maximizing price per good for the bundle, so that the seller's expected profit from selling B (at the price p_B) is $\pi_B = J(kp_B - \sum_{i \in B} c_i)P(v(B) \geq kp_B)$, where $k = card(B)$. Clearly, in the case when the marginal costs are identical for goods produced by the seller, that is, $c_i = c$ for all $i \in M$, the values of p_B are the same for all bundles B that consist of the same number $card(B)$ of goods: $p_B = p_{B'}$, if $card(B) = card(B')$. With identical marginal costs, we denote by $\overline{p}$ the profit maximizing price per good in the case when all the goods in M are sold as a single bundle and by $\underline{p}$ the profit maximizing price of each good $i \in M$ under unbundled sales. That is, in the case when $c_i = c$ for all $i \in M$, $\overline{p} = p_B$ with $B = M$, and $\underline{p} = p_B$ with $B = \{i\}$, $i \in M$.

The following Theorems 11.1 and 11.2 characterize the optimal bundling strategies for a multiproduct monopolist in the above setting with an arbitrary degree of complementarity or substitutability for goods in M (the cases of valuations $v(g_r, B)$ and $v(h_r, B)$ with an arbitrary $r \in (0, 2]$). From Theorem 11.1 it follows that if the tails of consumers' tastes are extremely heavy and the goods are independently priced or are substitutes (or are complements with not very high degree of complementarity), then the patterns in seller's optimal bundling strategies are the same as in the case of independently priced goods with log-concavely distributed valuations (see Bakos and Brynjolfsson (1999) and Fang and Norman (2003) and the discussion in Subsection 8.1).

Theorem 11.1 *Let $\mu \in \mathbf{R}$, $r \in (0, 2)$, and let the reservation prices for bundles $B \in 2^M$ of goods from M be given by $v(g_r, B)$ or by $v(h_r, B)$. Suppose that the tastes X_i, $i \in M$, are i.i.d. r.v.'s such that $X_i \sim S_\alpha(\sigma, \beta, \mu)$, $i \in M$, for some $\sigma > 0$, $\beta \in [-1, 1]$ and $\alpha \in (r, 2]$, or $X_i - \mu \sim \overline{\mathcal{CS}}(r)$, $i \in M$. The seller strictly prefers $\overline{\mathcal{B}}$ to any other bundling decision (that is, the goods are sold as a single bundle), if $c_i = c$, $i \in M$, and $\underline{p} < \mu$. The seller strictly prefers $\underline{\mathcal{B}}$ to any other bundling decision (that is, the goods are sold separately), if $c_i \geq \mu$, $i \in M$, or if $c_i = c$, $i \in M$, and $\overline{p} > \mu$.*

According to Theorem 11.2, the patterns in the solutions to the seller's optimal bundling problem in Theorem

[28] The results in this section that show that, in certain cases, separate auctions are optimal for the seller irrespective of the number of buyers do not contradict the *ex-post* analysis with two buyers in Palfrey (1983) either. This is because the assumptions under which these results hold involve valuation functions $v(B)$ for bundles B which are not additive over some subset of the support of distribution of the tastes X_i. Note also that payment rules for bundles in the auction setting considered in this paper are different from those under the VCG mechanism in Krishna (2002) who showed that, in that case, it is always better to sell the objects as a single bundle. In the VCG mechanism, this holds regardless of whether the objects are substitutes or complements. In addition to that, as Krishna notes, bundling may not be advantageous *ex post* even when the goods are complements. On the other hand, according to the results given by Corollary 10.1, for the second-price sealed bid auctions, bundling of complements is always disadvantageous for the seller, *ex ante*, regardless of the number of buyers, if the consumers' tastes are extremely thick-tailed. However, in this case, the seller always prefers, *ex ante*, a single auction for substitutes with sufficiently high degree of substitutability.

11.1 are reversed if consumers' tastes are extremely heavy-tailed and the goods are independently priced or are complements (or are substitutes with not very high degree of substitutability).

Theorem 11.2 *Let $\mu \in \mathbf{R}$, $r \in (0, 2]$, $p_{max} < \infty$, and let the reservation prices for bundles $B \in 2^M$ of goods from M be given by $v(g_r, B)$ or by $v(h_r, B)$. Suppose that the tastes X_i, $i \in M$, are i.i.d. r.v.'s such that $X_i \sim S_\alpha(\sigma, \beta, \mu)$, $i \in M$, for some $\sigma > 0$, $\beta \in [-1, 1]$ and $\alpha \in (0, r)$, or $X_i - \mu \sim \underline{\mathcal{CS}}(r)$, $i \in M$. The seller strictly prefers $\underline{\mathcal{B}}$ to any other bundling decision (that is, the goods are sold separately), if $c_i = c$, $i \in M$, and $\overline{p} < \mu$. The seller strictly prefers $\overline{\mathcal{B}}$ to any other bundling decision (that is, the goods are sold as a single bundle), if $c_i \geq \mu$, $i \in M$, or if $c_i = c$, $i \in M$, and $\underline{p} > \mu$.*

Theorem 11.3 and 11.4 below give analogues of the results in Theorems 11.1 and 11.2 in the case of independently priced goods ($r = 1$).

Theorem 11.3 *Let $\mu \in \mathbf{R}$, and let the reservation prices for bundles $B \in 2^M$ be given by $v(h_1, B) = \sum_{i \in B} X_i$. Suppose that the stand-alone reservation prices $v(h_1, \{i\}) = X_i$, $i \in M$, for goods in M are i.i.d. r.v.'s such that $X_i \sim S_\alpha(\sigma, \beta, \mu)$, $i \in M$, for some $\sigma > 0$, $\beta \in [-1, 1]$ and $\alpha \in (1, 2]$, or $v_i - \mu \sim \overline{\mathcal{CSLC}}$, $i \in M$. Then the conclusion of Theorem 11.1 holds.*

Theorem 11.4 *Let $\mu \in \mathbf{R}$, $p_{max} < \infty$, and let the reservation prices for bundles $B \in 2^M$ be given by $v(h_1, B) = \sum_{i \in B} X_i$. Suppose that the stand-alone reservation prices $v(h_1, \{i\}) = X_i$, $i \in M$, for goods in M are i.i.d. r.v.'s such that $X_i \sim S_\alpha(\sigma, \beta, \mu)$, $i \in M$, for some $\sigma > 0$, $\beta \in [-1, 1]$ and $\alpha \in (0, 1)$, or $v_i - \mu \sim \underline{\mathcal{CS}}(1)$, $i \in M$. Then the conclusion of Theorem 11.2 holds.*

Similar to the analysis in Bakos and Brynjolfsson (1999), the underlying intuition for Theorems 11.1 and 11.3 is that, for not extremely heavy-tailed reservation prices and the marginal costs of goods on the right of the mean valuation, bundling decreases profits since it reduces peakedness of the valuation per good and thereby decreases the fraction of buyers with valuations for bundles greater than their total marginal costs. For the identical marginal costs of goods less than the mean valuation, bundling is likely to have the opposite effect on the profit.

On the other hand, the results in Theorems 11.2 and 11.4 are driven by the fact that, in the case of extremely thick-tailed reservation prices, peakedness of the valuations per good in bundles decreases with their size. Therefore, bundling of goods in the case of extremely long-tailed valuations and marginal costs of goods higher than the mean reservation price increases the fraction of buyers with reservation prices for bundles greater than their total marginal costs and thereby leads to an increase in the monopolist's profit. This effect is reversed in the case of the identical marginal costs on the left of the mean valuation.

Remark 11.1. The assumptions of Theorem 11.2 with $r \geq 1$ (and those of Theorem 11.4) are satisfied, in particular, for positive stable tastes (stand-alone reservation prices) $X_i \sim S_\alpha(\sigma, 1, \mu)$, $i \in M$, where $\sigma > 0$ and $\alpha \in (0, 1)$, for which thus the free disposal condition holds, including the Lévy distributions $S_{1/2}(\sigma, 1, \mu)$. Furthermore, from the proof of Theorems 11.1-11.4 it follows that the first parts (second parts) of conclusions in the theorems hold as well in the case of arbitrary marginal costs c_i if the price per good p_B in *each* bundle $B \in 2^M$ is less than (greater than) μ. One should also note here that the conditions $p_{max} < \infty$ in Theorems 11.2 and 11.4 are necessary since otherwise the monopolist would set an infinite price for each bundle of goods under extremely heavy-tailed distributions of consumers' tastes considered in the theorems.

12 Proofs of the results in Part II

Proof of Theorem 10.1. Let $r \in (0, 2)$ and let X_i, $i \in M$, be i.i.d. r.v.'s such that $X_i \sim S_\alpha(\sigma, \beta, 0)$, $i \in M$, for some $\sigma > 0$, $\beta \in [-1, 1]$ and $\alpha \in (r, 2]$, or $X_i \sim \overline{\mathcal{CS}}(r)$, $i \in M$. Consider any bundle $B \in 2^M$ with $card(B) = k \geq 2$. Denote $H_k(x) = P(\sum_{i=1}^k X_i \leq x)$, $x \in \mathbf{R}$. Clearly, the cdf of the r.v. $v(g_r, B) = g_r(\sum_{i\in B} X_i)$ is $P(v(g_r, B) \leq x) = H_k(x^{1/r})$ for $x \geq 0$, $P(v(g_r, B) \leq x) = 0$ otherwise. Therefore, we have that, for all $x > 0$, the cdf of the seller's profit π_B resulting from selling B is

$$P(\pi_B \leq x) = P(v_{(n-1)}(g_r, B) \leq x) = n(H_k(x^{1/r}))^{n-1} - (n-1)(H_k(x^{1/r}))^n \tag{12.1}$$

(this cdf is zero for $x < 0$). For $i \in M$, let π_i be the seller's profit resulting from selling good i separately, that is, $\pi_i = \pi_{B_i}$ with $B_i = \{i\}$. For $x > 0$, the cdf of the r.v. $k\pi_1$ (that represents the seller's profit resulting from selling good 1 k times) is

$$P(k\pi_1 \leq x) = P(v_{(n-1)}(g_r, \{1\}) \leq x/k) = n(H_1(x^{1/r}/k^{1/r}))^{n-1} - (n-1)(H_1(x^{1/r}/k^{1/r}))^n. \tag{12.2}$$

By Theorem 4.1 and comparisons (2.1), $H_k(xk^{1/r}) > H_1(x)$, $x > 0$, and, therefore, $H_k(x^{1/r}) > H_1(x^{1/r}/k^{1/r})$, $x > 0$. Since the function $ny^{n-1} - (n-1)y^n$ is increasing in $y \in (0, 1)$, this, together with (12.1) and (12.2) implies that $P(\pi_B \leq x) > P(k\pi_1 \leq x)$ for all $x > 0$, and, therefore (see Shaked and Shanthikumar (1994, pp. 3-4) and Remark 10.1), $E(\pi_B) < E(k\pi_1) = \sum_{i\in B} E(\pi_i)$. Consequently, we get that for any bundling decision $\mathcal{B} = \{B_1, ..., B_l\}$ such that $card(B_s) = k_s$, $s = 1, ..., l$, and $k_t \geq 2$ for at least one $t \in \{1, ..., l\}$,

$$E(\Pi_{\mathcal{B}}) = \sum_{s=1}^{l} E(\pi_{B_s}) < \sum_{s=1}^{l} \sum_{i\in B_s} E(\pi_i) = \sum_{i=1}^{m} E(\pi_i) = E(\Pi_{\underline{\mathcal{B}}}). \tag{12.3}$$

The proof is complete.

Proof of Theorem 10.2. Let $r \in (0, 2)$ and let X_i, $i \in M$, be i.i.d. r.v.'s such that $X_i \sim S_\alpha(\sigma, \beta, 0)$, $i \in M$, for some $\sigma > 0$, $\beta \in [-1, 1]$ and $\alpha \in (0, r)$, or $X_i \sim \underline{\mathcal{CS}}(r)$, $i \in M$. Consider any bundle $B \in 2^M$ with $card(B) = k \leq m - 1$. With the same notations as in the proof of Theorem 10.1, comparisons (2.1) and Theorem 4.2 imply that $H_k(xk^{1/r}) > H_m(xm^{1/r})$, $x > 0$, and, therefore, $H_k(x^{1/r}) > H_m(x^{1/r}m^{1/r}/k^{1/r})$, $x > 0$. Similar to the proof of Theorem 10.1, we get, therefore, that $P(\pi_B \leq x) > P((k/m)\Pi_{\overline{\mathcal{B}}} \leq x)$ for all $x > 0$. By Shaked and Shanthikumar (1994, pp. 3-4) and the property that U is an increasing concave function with $U(0) = 0$, we get, therefore, that $EU(\pi_B) < EU((k/m)\Pi_{\overline{\mathcal{B}}}) \leq (k/m)EU(\Pi_{\overline{\mathcal{B}}})$. Consequently, for any bundling decision $\mathcal{B} = \{B_1, ..., B_l\}$ such that $card(B_s) = k_s$, $s = 1, ..., l$, and $k_t \leq m - 1$ for at least one $t \in \{1, ..., l\}$,

$$EU(\Pi_{\mathcal{B}}) = EU(\sum_{s=1}^{l} \pi_{B_s}) \leq \sum_{s=1}^{l} EU(\pi_{B_s}) < \sum_{s=1}^{l} EU((k_s/m)\Pi_{\overline{\mathcal{B}}}) \leq \sum_{s=1}^{l} (k_s/m)EU(\Pi_{\overline{\mathcal{B}}}) = EU(\Pi_{\overline{\mathcal{B}}}).$$

The proof is complete.

Proof of Corollaries 10.1 and 10.2. The corollaries are immediate consequences of Theorems 10.1 and 10.2 in the case of positive stable distributions.

Proof of Theorem 10.3. Let $j \in J$. Let the vector $\tilde{X}^{(j)}$ of the jth buyer's reservation prices for goods in M take a value $\tilde{x}^{(j)} = (\tilde{x}_{1j}, ..., \tilde{x}_{nj}) \in \mathbf{R}_+^n$, $(\tilde{x}_{1j}, ..., \tilde{x}_{nj}) \neq (0, 0, ..., 0)$. Consider any bundle $B \in 2^M$ with $card(B) = k \geq 2$. The $j-$th buyer's reservation price for the bundle is $v_j(B) = \sum_{i \in B} \tilde{x}_{ij}$. Using the same notations as in the proof of Theorem 10.1, we get, similar to Palfrey (1983), that the expected surplus to the buyer when B is offered for sale is

$$ES_j(B, \tilde{x}^{(j)}) = \int_0^{v_j(B)} (H_k(x))^{n-1} dx = k \int_0^{v_j(B)/k} (H_k(kx))^{n-1} dx. \tag{12.4}$$

On the other hand, the expected surplus to consumer j when good $i \in B$ is offered for sale separately is $ES_j(\{i\}, \tilde{x}^{(j)})$ $= \int_0^{\tilde{x}_{ij}} (H_1(x))^{n-1} dx$. By Theorem 4.4 and (2.1), $H_k(kx) < H_1(x)$ for all $x > 0$. This, together with (12.4), implies

$$ES_j(B, \tilde{x}^{(j)}) < k \int_0^{v_j(B)/k} (H_1(x))^{n-1} dx \tag{12.5}$$

if $v_j(B) > 0$. Since the function $(H_1(y))^{n-1}$ is increasing in $y \in \mathbf{R}_+$, from Theorem 3.C.1 in Marshall and Olkin (1979) we get that the function $F(y_1, ..., y_k) = \sum_{i=1}^k \int_0^{y_i} (H_1(x))^{n-1} dx$ is Schur-convex in $(y_1, ..., y_k) \in \mathbf{R}_+^k$. Therefore, from majorization comparisons (7.5) it follows that $F(y_1, ..., y_k) \geq F(\sum_{i=1}^k y_i/k, ..., \sum_{i=1}^k y_i/k)$ for all $(y_1, ..., y_k) \in \mathbf{R}_+^k$ (see also the proof of Theorem 5 in Palfrey, 1983). In particular,

$$k \int_0^{v_j(B)/k} (H_1(x))^{n-1} dx \leq \sum_{i \in B} \int_0^{\tilde{x}_{ij}} (H_1(x))^{n-1} dx = \sum_{i \in B} ES_j(\{i\}, \tilde{x}^{(j)}). \tag{12.6}$$

From (12.5) and (12.6) we get

$$ES_j(B, \tilde{x}^{(j)}) < \sum_{i \in B} ES_j(\{i\}, \tilde{x}^{(j)}) \tag{12.7}$$

if $v_j(B) > 0$ (clearly, (12.7) holds as equality if $v_j(B) = 0$). By (12.7), we have that if the seller follows a bundling decision $\mathcal{B} = \{B_1, ..., B_l\}$ such that $card(B_s) = k_s$, $s = 1, ..., l$, and $k_t \geq 2$ for at least one $t \in \{1, ..., l\}$, then the expected surplus $ES_j(\mathcal{B}, \tilde{x}^{(j)})$ to buyer j satisfies $ES_j(\mathcal{B}, \tilde{x}^{(j)}) = \sum_{s=1}^l ES_j(B_s, \tilde{x}^{(j)}) < \sum_{i=1}^n ES_j(\{i\}, \tilde{x}^{(j)}) =$ $ES_j(\underline{\mathcal{B}}, \tilde{x}^{(j)})$. The proof is complete.

Proofs of Theorems 11.1-11.4. Let $r \in (0, 2]$ and let c_i, $i \in M$, be arbitrary marginals costs of goods in M. Let the reservation prices $v(B)$ for bundles $B \in 2^M$ be given by $v(B) = v(g_r; B) = g_r\big(\sum_{i \in B} X_i\big)$ or by $v(B) = v(h_r; B) =$ $h_r\big(\sum_{i \in B} X_i\big)$. Further, let $\mu \in \mathbf{R}$ and $p_{max} < \infty$. Suppose that the tastes X_i, $i \in M$, are i.i.d. r.v.'s such that $X_i \sim S_\alpha(\sigma, \beta, \mu)$, $i \in M$, for some $\sigma > 0$, $\beta \in [-1, 1]$ and $\alpha \in (0, r)$, or $X_i - \mu \sim \underline{\mathcal{CS}}(r)$, $i \in M$. We will show that the seller's profit maximizing bundling decision is $\underline{\mathcal{B}}$ if the prices per good $p_B < \mu$ for all bundles $B \in 2^M$, and is $\overline{\mathcal{B}}$ if $p_B > \mu$ for all $B \in 2^M$. For a bundle $B \in 2^M$, the profit maximizing price per good in the bundle is $p_B = \arg\max_{p \in [0, p_{max}]} \big(p - (1/k) \sum_{i \in B} c_i\big) P(v(B) \geq kp)$ and the seller's profit per good resulting from selling the bundle B (at the price per good p_B) is $E(\pi_B) = Jk\big(p_B - \sum_{i \in B} c_i\big) P(v(B) \geq kp_B)$, where $k = card(B)$ is the number of goods in B. For $i \in M$, let p_i be the price of good i in the case when the goods are sold separately (that is, in the case of the bundling decision $\underline{\mathcal{B}}$) and let, as in the proof of Theorem 10.1, π_i be the monopolist's profit from selling the good, namely, $p_i = p_{B_i}$ and $\pi_i = \pi_{B_i}$ with $B_i = \{i\}$. As in the setup of the optimal bundling problem in Section 11, in the case when $c_i = c$ for all $i \in M$, we write $\overline{p} = p_M$ for the price per good in the case when all the n goods are sold as a single bundle $B = M$ (that is, in the case of the bundling decision $\overline{\mathcal{B}}$) and $\underline{p}$ for the price of each good under unbundled sales (that is, $\underline{p} = p_B$ with $B = \{i\}$, $i \in M$).

Suppose that $p_B < \mu$ for all $B \in 2^M$. Then from Theorem 4.4 and relations (2.1) it follows that, for any bundle $B \in 2^M$ with the number of goods $card(B) = k \geq 2$, $E(\pi_B) = J\big(kp_B - \sum_{i\in B} c_i\big)P(v(B) \geq kp_B) = J\big(kp_B - \sum_{i\in B} c_i\big)P(\sum_{i\in B} X_i \geq (kp_B)^{1/r}) < J\sum_{i\in B}\big(p_B - c_i\big)P\big(X_i \geq (p_B)^{1/r}\big) \leq \sum_{i\in B} E(\pi_i)$. This implies that for any bundling decision $\mathcal{B} = \{B_1, ..., B_l\}$ such that $card(B_s) = k_s$, $s = 1, ..., l$, and $k_t \geq 2$ for at least one $t \in \{1, ..., l\}$, comparisons (12.3) hold.

Suppose now that $p_B > \mu$ for all $B \in 2^M$. Then using again Theorem 4.4 and relations (2.1) we get that, for any bundle $B \in 2^M$ with $card(B) = k \leq m - 1$, $E(\pi_B) = J\big(kp_B - \sum_{i\in B} c_i\big)P(\sum_{i\in B} X_i \geq (kp_B)^{1/r}) < J\big(kp_B - \sum_{i\in B} c_i\big)P(\sum_{i=1}^m X_i \geq (mp_B)^{1/r})$. Therefore, for any bundling decision $\mathcal{B} = \{B_1, ..., B_l\}$ such that $card(B_s) = k_s$, $s = 1, ..., l$, and $k_t \leq m - 1$ for at least one $t \in \{1, ..., l\}$,

$$E(\Pi_{\mathcal{B}}) = \sum_{s=1}^{l} E(\pi_{B_s}) < J\sum_{s=1}^{l}\big(k_s p_{B_s} - \sum_{i\in B_s} c_i\big)P(\sum_{i=1}^{m} X_i \geq (mp_B)^{1/r}) =$$

$$J\sum_{s=1}^{l} k_s\big(p_{B_s} - (1/m)\sum_{i=1}^{m} c_i\big)P(\sum_{i=1}^{m} X_i \geq (mp_B)^{1/r}) \leq \sum_{s=1}^{l}(k_s/m)E(\Pi_{\overline{\mathcal{B}}}) = E(\Pi_{\overline{\mathcal{B}}}). \tag{12.8}$$

From (12.3) and (12.8) we get that the profit maximizing bundling decision is $\underline{\mathcal{B}}$ if $p_B > \mu$ for all $B \in 2^M$ and is $\overline{\mathcal{B}}$ if $p_B < \mu$ for all $B \in 2^M$.

Clearly, the condition that $p_B > \mu$ for all $B \in 2^M$ is satisfied if $c_i \geq \mu$ for all $i \in M$. Furthermore, in the case of identical marginal costs $c_i = c$, $i \in M$, the condition that $p_B > \mu$ for all $B \in 2^M$ holds if $\underline{p} > \mu$. Indeed, suppose this not the case and that there exists a bundle $B \in 2^M$ with $card(B) = k > 1$ and $p_B \leq \mu$. Then, as above, we get $kE(\pi_1) = Jk(\underline{p} - c)P(X_1 \geq (\underline{p})^{1/r}) < Jk(\underline{p} - c)P(\sum_{i=1}^k X_i \geq (k\underline{p})^{1/r}) \leq E(\pi_B)$. On the other hand, $E(\pi_B) = Jk(p_B - c)P(\sum_{i=1}^k X_i \geq (k\underline{p})^{1/r}) < Jk(p_B - c)P(X_1 \geq (p_B)^{1/r}) \leq kE(\pi_1)$, which is a contradiction. Similarly, we get that if $c_i = c$, $i \in M$, then $\overline{p} < \mu$ implies that $p_B < \mu$ for all $B \in 2^M$. This completes the proof of Theorem 11.2. Theorem 11.4 follows from Theorem 11.2 with $r = 1$. Theorems 11.1 and 11.3 could be proven in a similar way, with the use of Theorems 4.1 and 4.3 instead of Theorem 4.2. The proof is complete.

Part III

DEMAND-DRIVEN INNOVATION AND SPATIAL COMPETITION OVER TIME UNDER HEAVY-TAILED SIGNALS

13 Discussion of the results in Part III

13.1 Gibrat's law and demand-driven innovation and spatial competition over time

A voluminous empirical literature on firm growth has focused on testing the validity of Gibrat's law according to which firm growth rates are independent of their sizes and are non-autocorrelated over time. Many papers in the field have observed deviations from Gibrat's law in data, including the patterns of positive or negative dependence between firm growth and size and autocorrelation in firm growth rates (see, e.g., the reviews in McCloughan, 1995, and Sutton, 1997). Motivated, in part, by these empirical findings, several studies have proposed models that could account for such phenomena. E.g., Jovanovic (1982) developed a Bayesian learning model of firm growth in which firms uncover their relative efficiency with time. The general learning model predicts negative dependence between age and firm growth and suggests, therefore, that a similar pattern in correlation between the growth rates and firm size holds. Jovanovic and Rob (1987) proposed a model of demand-driven innovation and spatial competition over time based on the idea that larger firms get better information about the design of future products. The model implies departures from Gibrat's law in heterogenous markets, with firms' size being autocorrelated over time. Jovanovic and Rob' model assumes that each period, the firm observes a sample $\mathcal{S}$ of signals $s_i = \theta + \epsilon_i$, $i = 1, ..., N$, about the next period's ideal product $\theta \in \mathbf{R}$, where ϵ_i, $i = 1, ..., N$, are i.i.d. unimodal shocks with mode 0 and N is a (random) sample size. The firm then chooses a product design $\hat{\theta} \in \mathbf{R}$, a level of output y and an amount of investment in information $z \geq 0$, with $C(y)$ and $K(z)$ denoting the corresponding convex and twice differentiable cost functions. Applying Proschan (1965)'s result given by Proposition 2.1 in Part I of the paper, Jovanovic and Rob (1987) showed that, in the case of signal shocks $\epsilon_1, \epsilon_2, ...$, with a symmetric log-concave density, the model has the following properties: If the optimal levels (y, z) of the firm's output and informational gathering effort satisfy the first- and second-order conditions for a maximum, then

• 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;

• The distribution of future size is stochastically increasing as a function of current size.

The above properties imply that, under the assumption that the density of signals is log-concave (which implies, as discussed before, that the tails of signals' distributions decline at least exponentially fast and, thus, the distributions are extremely light-tailed), relatively large firms are likely to stay larger; in addition, the size-difference is positively autocorrelated. 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 (see the discussion in Jovanovic and Rob, 1987).

13.2 Main results of Part III on the robustness of firm growth theory to heavy-tailedness

In this part of the paper, we focus on the analysis of robustness of the properties of Jovanovic and Rob's (1987) model of demand-driven innovation and spatial competition over time to the assumptions of heavy-tailedness of signals' distributions. Using the main majorization results established in Part I of the paper, we show that the properties of the model for signals with log-concave densities remain valid for heavy-tailed distributions (Theorem 15.1). However, we prove that the above properties of the model are reversed under the assumption that the distributions of the signals are extremely long-tailed (Theorem 15.2).

We prove *inter alia* that the following results hold: Suppose that in Jovanovic and Rob (1987), the signal shocks $\epsilon_1, \epsilon_2, ...$ are i.i.d. r.v.'s with an extremely heavy-tailed distribution which is a convolution of symmetric stable distributions with indices of stability in the interval $(0, 1)$. If the optimal levels (y, z) of the firm's output and informational gathering effort satisfy the first- and second-order conditions for a maximum, then

- 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;

- The distribution of future size is stochastically decreasing as a function of current size.

In other words, properties of the model of demand-driven innovation and spatial competition over time are robust to thick-tailedness assumptions as long as the distributions entering the assumptions are not extremely heavy-tailed. However, they are reversed if the assumptions involve extremely thick-tailed distributions.

According to the results, in the case of extremely long-tailed signals, relatively large firms are not likely to stay larger. 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, surprisingly, 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.[29]

Furthermore, we show that if the cost $K(z)$ of engaging in the informational gathering effort is increasing in $z \geq 0$, and the shocks $\epsilon_1, \epsilon_2, ...$ are i.i.d. r.v.'s with a distribution which is a convolution of symmetric stable distributions with characteristic exponents in the interval $(0, 1)$, then the optimal choice of investment z is zero: $z = 0$ (Theorem 15.3). This result is quite intuitive and implies that, if the cost to the firms of gathering information is increasing in z and the sample of signals consists of extremely long-tailed r.v.'s and is, therefore, uninformative about the next period's ideal product θ, then all firms choose not to invest in information gathering. Furthermore, in contrast to the model of demand-driven innovation and spatial competition over time with signals with a log-concave density or with not extremely heavy-tailed signals, in the model with arbitrary convex cost functions $C(y)$ and $K(z)$ and

[29] Interestingly, similar oscillation patterns are observed in the propagation of the sex ratio in several inheritance models under heavy-tailedness, see Ibragimov (2004).

extremely fat-tailed shocks $\epsilon_1, \epsilon_2, ...$, it turns out that large firms are not likely to stay larger. In addition to that, under the assumptions of extremely heavy-tailed signals, there is negative autocorrelation in the size-difference.

Essentially, in the case of extremely heavy-tailed signals, smaller firms, in fact, have an advantage over their larger counterparts. The underlying intuition is that in the presence of extremely heavy-tailed shocks, the sample of signals is not informative about the ideal product since it is likely to contain extreme outliers. Hence, it is sheer luck in choosing the product design $\hat{\theta}$ close to θ, and not the informational advantage that matters. Smaller firms which get less useless information and spend less in its gathering and processing are more likely to be more successful. It is important to note that from these results it follows that, in the model with extremely heavy-tailed signals, the firm growth is likely to decrease with firm size. The above implies that Gibrat's law does not hold in the setup. Moreover, in the case of extremely heavy-tailed signals, both the implications of Gibrat's law fail. First, firm growth and size appear to be dependent; second, the implication of Gibrat's law that firm growth rates are non-autocorrelated over time does not hold either.

Firm growth models with heavy-tailedness seem to provide an approximation for industries with very uncertain consumer perception of new products or constantly changing environments and new industries in which business decisions on the base of former experience are impossible and the risk facing the firms is higher than in other sectors. Many high-tech industries, especially those in the Net economy, exhibit the above patterns. The results in this part of the paper provide new insights concerning firm size and growth patterns in such settings. For instance, Agarwal and Gort (1994) have observed that entrants in markets for high-technology products tend to have higher survival rates than incumbents. These results for "technical" products are consistent with findings in Audretsch (1991) who showed that new firm survival rates tend to be higher in sectors with high innovative activity by small firms, which are more likely to be recent entrants. On the other hand, according to the results in Agarwal and Gort (1994), survival rates for incumbents are higher than for entrants in markets for "nontechnical" products, where advantages of experience and learning by doing are greater. Agarwal and Gort (1994) note that their results are consistent with entry in high technology industries being accompanied by breakthroughs in knowledge or innovations by inventors and firms initially outside the market. Such breakthroughs in knowledge are extreme events that yield superior knowledge to entrants and give them an advantage over incumbents, similar to an advantage of small firms over their large counterparts in the model of demand-driven model of innovation and spatial competition over time with extremely heavy-tailed signals.

The launch, rapid rise and sudden fall of Internet businesses during the late 1990's and the first half of 2000 (see, among others, Thornton and Marche, 2003, for the account of the development and failure of the e-commerce industry and the analysis of factors that have lead to it) might illustrate the oscillation patterns in the firm sizes predicted by the results for growth models with extremely heavy-tailed signals obtained in this part of the paper. The Net economy grew 174.5% from 1995 to 1998 and 68% from 1998 to 1999, and, similarly, many dotcom businesses experienced extremely fast growth after their opening. For instance, as discussed in Thornton and Marche (2003), boo.com grew from 5 to 420 employees in just a year; eToys.com grew from 13 to 235 full time employees in its first year, then in the 7 months after its IPO, its staff increased to 940 employees; the number of employees at garden.com increased from 149 when it went public in May, 1999 to 267 by the end of June, 2000; staff of streamline.com grew from 189 to 350 in just a year after its IPO in March, 1999. The profits and stock prices of many Internet companies skyrocketed as well, with millions of dollars in revenues of Amazon.com and eToys.com during the holiday season of 1998. The high growth of e-tail businesses and the Internet economy continued until an extreme event, the fall of NASDAQ by 10% in April, 2000, "has sent prices in the high-tech sectors tumbling, and prompted investors to re-evaluate and demand profitability" (Thornton and Marche, 2003). This extreme signal has lead to crashes of hundreds of Internet businesses in 2000 and the following year, with 234 "dot com deaths" by the end of 2000 and 330 in the first half of 2001 and businesses losing hundreds of millions of dollars in just a few years (see Thornton and Marche, 2003).

13.3 Organization of Part III

This part of the paper is organized as follows: Section 14 reviews the properties of the model of demand-driven innovation and spatial competition over time with log-concavely distributed signals derived by Jovanovic and Rob (1987). Section 15 presents the main results of the paper on robustness of the model of demand-driven innovation and spatial competition over time to assumptions of heavy-tailedness of consumers' signals. Section 16 contains proofs of the results obtained in this part.

14 Demand-driven innovation and spatial competition over time with log-concavely distributed signals

Let $\rho(x, y) = (x - y)^2$, $x, y \in \mathbf{R}$, denote the quadratic loss function. In the setting of Jovanovic and Rob's (1987) model of demand-driven innovation and spatial competition over time described in Subsection 13.1, let a consumer of type $u \in \mathbf{R}$ have 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. Under the former assumption, a necessary condition for an equilibrium is that $\rho(\hat{\theta}, \theta) + p_{\hat{\theta}} = p$ for all $\hat{\theta} \in \mathbf{R}$, where p is the price of the ideal product θ. The size N of the sample $\mathcal{S}$ of signals about the next period's ideal product observed by the firm follows a distribution $\pi(n; y + z)$ conditionally on $y + z$: $\pi(n; y + z) = P(N = n|y + z)$, $n = 0, 1, 2, ...$ Below, we denote by $\mathcal{S}_t$, $\hat{\theta}_t$, θ_t, y_t and z_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 and z_t are chosen; the period then ends.

Let L be the set of measures on the set $\mathbf{R}^2$ of pairs of decisions (y, z) among firms; we consider Markovian equilibria with the aggregate state being the distribution of decisions $\nu_t \in L$ such that $\nu_t = \alpha(\theta_t, \nu_{t-1})$ (see Brock and Mirman, 1972, Jovanovic and Rob, 1987, and Stokey and Lucas, 1989). In such an equilibrium, the price p of the ideal product at t can be expressed as a function of ν_{t-1} (see Jovanovic and Rob, 1987); this equilibrium relationship will be denoted $p(\nu_{t-1})$. The price of the product for a firm that locates as $\hat{\theta}$ at t is

$$p_{\hat{\theta}} = p(\nu_{t-1}) - \rho(\hat{\theta}, \theta_t). \tag{14.9}$$

For $n > 0$, denote $\bar{s}_n = n^{-1} \sum_{i=1}^{n} s_i$ and $\bar{\epsilon}_n = n^{-1} \sum_{i=1}^{n} \epsilon_i$. Further, let $F(x; n) = P(|\bar{\epsilon}_n| \leq x)$, $x \geq 0$, $n = 1, 2, ...$, denote the cdf of $|\bar{\epsilon}_n|$, $n = 1, 2, ...$, on $\mathbf{R}_+$. Assuming a diffuse prior for $\theta \in \mathbf{R}$, the optimal choice of $\hat{\theta} = \hat{\theta}(\mathcal{S})$ in the case $N > 0$ is (see Jovanovic and Rob, 1987) $\hat{\theta} = argmax_{\tilde{\theta}}\, N^{-1}\sum_{i=1}^N \rho(\tilde{\theta}, s_i) = argmax_{\tilde{\theta}}\, N^{-1}\sum_{i=1}^N (\tilde{\theta} - s_i)^2 = \overline{s}_N$.[30] It is not difficult to see that the loss associated with the choice of the product design $\hat{\theta}(\mathcal{S})$ for $N > 0$ is $\rho(\hat{\theta}(\mathcal{S}), \theta) = \overline{\epsilon}_N^{\,2}$. In the case when $N = 0$ belongs to the support of N, so that $\pi(0; y+z) \neq 0$, it is usually assumed that $\rho(\hat{\theta}(\mathcal{S}), \theta) = \infty$ for $N = 0$. The cdf of $\rho(\hat{\theta}(\mathcal{S}), \theta)$ (on $\mathbf{R}_+$) conditional on $y + z$ is

$$\xi(x; y + z) = P(\rho(\hat{\theta}(\mathcal{S}), \theta) \leq x | y + z) = \sum_{n=0}^{\infty} F(\sqrt{x}; n)\pi(n; y + z), \tag{14.10}$$

$x \geq 0$ (with $F(\sqrt{x}; 0) = 0$ 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 and z, following a realization $\rho(\hat{\theta}, \theta) = x$, is $V(x, \nu_{-1}) = \max_{y,z}\Big\{y(p(\nu_{-1}) - x) - C(y) - K(z) + \beta \int V(\tilde{x}, \alpha(\nu_{-1}))d\xi(\tilde{x}; y + z)\Big\}$ (see Jovanovic and Rob, 1987).

Let $G(y + z) = \beta \int V(\tilde{x}, \alpha(\nu_{-1}))d\xi(\tilde{x}; y + z)$. The first-order necessary conditions for an interior maximum (y, z) are

$$p_{\hat{\theta}} - C'(y) + G'(y + z) = 0, \quad -K'(z) + G'(y + z) = 0. \tag{14.11}$$

The second-order conditions for a maximum are

$$G''(y + z) < C''(y), \quad C''(y)K''(z) > G''(y + z)(C''(y) + K''(z)) \tag{14.12}$$

(conditions (14.12) imply $G''(y + z) < K''(z)$). If

$$G'(y + z) \leq K'(z) \tag{14.13}$$

for all (y, z), then the optimal level of informational gathering effort is zero:

$$z = 0. \tag{14.14}$$

In that case, the first- and second-order conditions for a point $(y, 0)$ in the interior of $\{(y, 0)\}$ to be optimal are

$$p_{\hat{\theta}} - C'(y) + G'(y) = 0, \tag{14.15}$$

$$G''(y) < C''(y). \tag{14.16}$$

We assume that, for any continuous $f : \mathbf{R} \to \mathbf{R}$, the expression $\int f(\tilde{x})d\xi(\tilde{x}; \lambda)$ is differentiable in λ. Under this assumption, one can implicitly differentiate first-order conditions (14.11) and (14.15) (see Jovanovic and Rob, 1987).

Evidently, the condition $G'' < 0$ suffices for conditions (14.12) and (14.16) to hold. However, $G'' > 0$ is also consistent with maxima being interior. By Proposition 4 in Jovanovic and Rob (1987), if the function G is convex ($G'' > 0$), then larger firms invest more in information. One should note that, according to empirical studies, there

[30] In the setting of Jovanovic and Rob (1987), the absolute deviation $\rho(\theta, \hat{\theta}) = |\theta - \hat{\theta}|$ needs to be replaced by the quadratic loss $\rho(\hat{\theta}, \theta) = (\hat{\theta} - \theta)^2$, as in the present section, for the product design to be given by the sample mean of signals: $\hat{\theta} = \overline{s}_N$. No conclusion derived in Jovanovic and Rob (1987) is affected by this small modification.

is a positive relationship between R&D expenditures and firm size, that suggests that $G(y+z)$ is indeed convex (see Kamien and Schwartz, 1982, and the discussion following Proposition 4 in Jovanovic and Rob, 1987).

Suppose that, conditionally on $y+z$, N has a Poisson distribution with

$$\pi(n; y+z) = \pi_0(n; y+z) = \frac{[\mu(y+z)]^n}{n!} exp(-\mu(y+z)), \quad n = 1, 2, ... \tag{14.17}$$

(with the convention that $\rho(\hat{\theta}, \theta) = \infty$ for $N = 0$) or a shifted Poisson distribution

$$\pi(n; y+z) = \pi_1(n; y+z) = \frac{[\mu(y+z)]^{n-1}}{(n-1)!} exp(-\mu(y+z)), \quad n = 1, 2, ... \tag{14.18}$$

(this distribution allows one to avoid the ambiguity concerning the value of $\rho(\hat{\theta}, \theta)$ in the case $N = 0$).

Lemma 14.1 obtained by Jovanovic and Rob (1987) gives *sufficient* conditions for concavity of the function $G(y+z)$; under the assumptions of the lemma, therefore, the second-order conditions for an interior maximum with respect to y and z are satisfied.

Lemma 14.1 *(Jovanovic and Rob, 1987). Suppose that, conditionally on $y+z$, N has a Poisson distribution $\pi_0(n; y+z)$ given by (14.17). The function $G(y+z)$ is strictly concave in $y+z$ 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 Jovanovic and Rob (1987), the conditions of Lemma 14.1 are satisfied for normal r.v.'s $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$, $i = 1, 2...$

Jovanovic and Rob (1987) obtained the following Proposition 14.1. In the proposition and its analogues for heavy-tailed signals below (Theorems 15.1 and 15.2), $y_t^{(1)}$ 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 14.1 *(Jovanovic and Rob, 1987). Suppose that, conditionally on $y+z$, N has a Poisson distribution $\pi_0(n; y+z)$ in (14.17). Let the shocks $\epsilon_1, \epsilon_2, ...$ be i.i.d. r.v.'s such that $\epsilon_i \sim \mathcal{LC}$, $i = 1, 2, ...$ If the optimal levels (y_t, z_t) of output and informational gathering effort satisfy (14.11) and (14.12) or (14.14)-(14.16), then*

(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 $y \geq 0$.

Note that, using the arguments completely similar to the proof of above Lemma 14.1 and Proposition 14.1 in Jovanovic and Rob (1987), one has that the lemma and the proposition also hold under the assumption that N has a shifted Poisson distribution $\pi_1(n; y+z)$ given by (14.18) as well as under the assumption that conditions (14.14)-(14.16) are satisfied.

15 Main results of Part III on the robustness of firm growth to heavy-tailedness

The following theorem provides a generalization of Proposition 14.1 that shows that the results obtained by Jovanovic and Rob (1987) hold in the case of thick-tailed signals.

Theorem 15.1 *Suppose that, conditionally on $y+z$, N has a Poisson distribution $\pi_0(n; y+z)$ in (14.17) or a shifted Poisson distribution $\pi_1(n; y+z)$ in (14.18). Let the shocks $\epsilon_1, \epsilon_2, ...$ be i.i.d. r.v.'s such that $\epsilon_i \sim S_\alpha(\sigma, \beta, 0)$, $i = 1, 2, ...$, for some $\sigma > 0$, $\beta \in [-1, 1]$ and $\alpha \in (1, 2]$, or $\epsilon_i \sim \overline{\mathcal{CSLC}}$, $i = 1, 2, ...$ Then conclusions (a), (b) and (c) in Proposition 14.1 hold.*

Lemma 15.1 below shows that the sufficient conditions for an interior maximum in Lemma 14.1 which imply strict concavity of the function $G(y+z)$ are satisfied for shocks $\epsilon_1, \epsilon_2, ...$ with not extremely fat-tailed symmetric stable distributions.

Lemma 15.1 *If the shocks $\epsilon_1, \epsilon_2, ...$ are i.i.d. r.v.'s such that $\epsilon_i \sim S_\alpha(\sigma, 0, 0)$, $i = 1, 2, ...$, for some $\sigma > 0$, and $\alpha \in (1, 2]$, then the sequence $\{F(x; n+1) - F(x; n)\}_{n=0}^{\infty}$ is decreasing in n for all $x > 0$.*

As the following theorem shows, the conclusions of Proposition 14.1 and Theorem 15.1 are reversed in the case of shocks $\epsilon_1, \epsilon_2, ...$ with extremely fat tails.

Theorem 15.2 *Suppose that, conditionally on $y+z$, N has a shifted Poisson distribution $\pi_1(n; y+z)$ given by (14.18). Let the shocks $\epsilon_1, \epsilon_2, ...$ be i.i.d. r.v.'s such that $\epsilon_i \sim S_\alpha(\sigma, \beta, 0)$, $i = 1, 2, ...$, for some $\sigma > 0$, $\beta \in [-1, 1]$ and $\alpha \in (0, 1)$, or $\epsilon_i \sim \underline{\mathcal{CS}}(1)$, $i = 1, 2, ...$ If the optimal levels (y_t, z_t) of output and informational gathering effort satisfy (14.11) and (14.12) or (14.14)-(14.16), then*

(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 $y \geq 0$.

According to Theorem 15.3 below, condition (14.13) and, thus, relation (14.14) is satisfied in the case of extremely fat-tailed shocks $\epsilon_1, \epsilon_2, ...$ and the increasing costs $K(z)$ of engaging in the informational gathering effort. That is, if the function $K(z)$ is increasing in $z \geq 0$ and the distribution of the signal shocks is extremely heavy-tailed, then each firm chooses zero informational gathering effort: $z = 0$.

Theorem 15.3 *Under the assumptions of Theorem 15.2, $G'(y+z) \leq 0$. Therefore, if $K'(z) > 0$, then (14.13) is satisfied and the optimal choice of informational gathering effort is $z = 0$.*

Proposition 14.1 and Theorem 15.1 imply that, in the case of not extremely heavy-tailed signals, relatively large firms are likely to stay larger; in addition, the size-difference is positively autocorrelated. According to Theorem 15.2, these conclusions are reversed in a world of extremely fat-tailed signals: relatively large firms are not likely to stay larger and the size-difference exhibits negative autocorrelation. The intuition for the results given by Proposition 14.1 and Theorems 15.1-15.3 is that the larger is a firm's size, the greater is the amount of information the firm gets.[31] The samples of consumers' signals are informative about the ideal product θ if the signals' distributions are not extremely heavy-tailed, as in Proposition 14.1 and Theorem 15.1. However, they are uninformative about θ in the case of extremely long-tailed distributions in Theorems 15.2 and 15.3. The larger firms that learn more are thus more likely to come up with a successful product if the signals are not extremely fat-tailed (see the discussion in Jovanovic and Rob, 1987). In a world of extremely heavy-tailed signals, on the other hand, smaller firms that get less uninformative signals have an advantage over their larger counterparts (see Subsection 13.2). The fact that heavy-tailed samples are uninformative about the next period's ideal product also drives the conclusion that it is optimal not to invest into the informational gathering if the cost $K(z)$ of the investment is increasing in z.

16 Proofs of the results in Part III

Proof of Theorem 15.1. Let $j \in \{0, 1\}$, and let, conditionally on $y+z$, N have a distribution $\pi_j(n; y+z)$. Then from (14.10) it follows, similar to the proof of Lemma 2 in Jovanovic and Rob (1987), that, for $x \geq 0$,

$$\partial\xi(x;\lambda)/\partial\lambda = \mu \sum_{n=j}^{\infty} \pi_j(n;\lambda)(F(\sqrt{x}; n+1) - F(\sqrt{x}; n)) \quad (16.1)$$

(with $F(\sqrt{x}; 0) = 0$ if $j = 0$). Theorem 4.3 and relations (2.1) in Part I imply that, under the assumptions of the theorem,

$$F(\sqrt{x}; n+1) = 1 - P(|\bar{\epsilon}_{n+1}| > \sqrt{x}) > 1 - P(|\bar{\epsilon}_n| > \sqrt{x}) = F(\sqrt{x}; n), \quad (16.2)$$

$x > 0$, $n = 1, 2, ...$ From (16.1) and (16.2) it follows that, under the assumptions of the theorem,

$$\partial\xi(x;\lambda)/\partial\lambda > 0 \quad (16.3)$$

for all $x > 0$, that is, $\xi(x, \lambda)$ is increasing in λ for all $x > 0$. As in Jovanovic and Rob (1987) we have

$$\partial y/\partial p_{\hat{\theta}} = (1/C'')[1 + G''K''/(C''K'' - G''(C'' + K''))] > 0, \quad (16.4)$$

if (14.11) and (14.12) hold, and

$$\partial y/\partial p_{\hat{\theta}} = 1/(C'' - G'') > 0, \quad (16.5)$$

if (14.15) and (14.16) hold, that is, y is increasing in $p_{\hat{\theta}}$. Conclusion (c) of the theorem now follows from (16.3)-(16.5) and the property that, by (14.9), $p_{\hat{\theta}}$ is decreasing in $\rho(\hat{\theta}, \theta)$:

$$\partial p_{\hat{\theta}}/\partial\rho < 0. \quad (16.6)$$

[31] One should also note that, under (14.11) and (14.12), large firms will not reduce their investment in information to the point where their informational advantage disappears (Proposition 3 in Jovanovic and Rob, 1987) and, with an additional assumption of convexity of G, they always invest more according to Proposition 4 in Jovanovic and Rob (1987) (see also Nelson and Winter, 1978).

Let $\lambda^{(i)} = y^{(i)} + z^{(i)}$, $\rho^{(i)} = \rho(\hat{\theta}^{(i)}, \theta)$ and $\xi^{(i)}(x) = \xi(x; \lambda^{(i)})$, $i = 1, 2$, and let $y^{(2)} > y^{(1)}$. As in the proof of Proposition 6 in Jovanovic and Rob (1987), this implies, by (16.4), that $p_{\hat{\theta}}^{(2)} > p_{\hat{\theta}}^{(1)}$ under (14.11) and (14.12). Since $y + z$ is increasing in $p_{\hat{\theta}}$ under (14.11) and (14.12) by Proposition 3 in Jovanovic and Rob (1987), we get, therefore, that, under the assumptions of the theorem, $\lambda^{(2)} > \lambda^{(1)}$ and thus

$$\xi^{(2)}(x) > \xi^{(1)}(x) \tag{16.7}$$

for all $x > 0$ by (16.3). As in the proof of Proposition 6 in Jovanovic and Rob (1987), we have

$$P(\rho^{(1)} > \rho^{(2)} | y^{(1)}, y^{(2)}) = \int \xi^{(2)}(x) d\xi^{(1)}(x) = \int \xi^{(1)}(x) d\xi^{(1)}(x) + \int (\xi^{(2)}(x) - \xi^{(1)}(x)) d\xi^{(1)}(x). \tag{16.8}$$

Since $\int \xi^{(1)}(x) d\xi^{(1)}(x) = 1/2$ using integration by parts, from (16.7) and (16.8) we get

$$P(\rho^{(1)} > \rho^{(2)} | y^{(1)}, y^{(2)}) > 1/2. \tag{16.9}$$

Relations (16.4), (16.6) and (16.9) imply conclusion (a) of the theorem.

As in the proof of Proposition 6 in Jovanovic and Rob (1987), conclusion (b) of the theorem follows from (16.4)-(16.6) and (16.8) since, by (16.3), 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 $x > 0$. The proof is complete.

Proof of Lemma 15.1. 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, for $n \geq 2$ and $x > 0$,

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

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

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

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

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

The proof is complete.

Proof of Theorems 15.2 and 15.3. The proof is similar to the proof of Proposition 6 in Jovanovic and Rob (1987) and the proof of Theorem 15.1, with the use of Theorem 4.4 instead of Theorem 4.3 in this paper and Proposition 2.1 in Jovanovic and Rob (1987). Under the assumptions of Theorem 15.2, one has, by Theorem 4.4 and relations (2.1), that, similar to relation (16.2),

$$F(\sqrt{x}; n+1) = 1 - P(|\overline{\epsilon}_{n+1}| > \sqrt{x}) < 1 - P(|\overline{\epsilon}_n| > \sqrt{x}) = F(\sqrt{x}; n), \tag{16.10}$$

$x > 0$, $n = 1, 2, ...$ Relations (16.1) and (16.10) imply that, under the assumptions of Theorem 15.2,

$$\partial \xi(x; \lambda) / \partial \lambda < 0 \tag{16.11}$$

$x > 0$, that is, $\xi(x, \lambda)$ is decreasing in λ for all $x > 0$. Similar to the proof of Lemma 2 in Jovanovic and Rob (1987), from (16.11) it follows that $G'(\lambda) \leq 0$. This implies that conditions (14.13) is satisfied and the optimal choice of informational gathering effort is $z = 0$ if the cost function $K(z)$ is increasing: $K'(z) > 0$. Thus, Theorem 15.3 holds.

Relations (16.4)-(16.6) and (16.11) imply conclusion (c') of Theorem 15.2.

Let, as in the proof of Theorem 15.1, $\lambda^{(i)} = y^{(i)} + z^{(i)}$, $\rho^{(i)} = \rho(\hat{\theta}^{(i)}, \theta)$ and $\xi^{(i)}(x) = \xi(x; \lambda^{(i)})$, $i = 1, 2$, and let $y^{(2)} > y^{(1)}$. By (16.4) and Proposition 3 in Jovanovic and Rob (1987) we have $p_{\hat{\theta}}^{(2)} > p_{\hat{\theta}}^{(1)}$ under (14.11) and (14.12) and, therefore, $\lambda^{(2)} > \lambda^{(1)}$ under the assumptions of Theorem 15.2. This and (16.11) imply that

$$\xi^{(2)}(x) < \xi^{(1)}(x), \tag{16.12}$$

for all $x > 0$. From (16.8) and (16.12) it follows, similar to the proof Proposition 6 in Jovanovic and Rob (1987) and to the proof of Theorem 15.1 in the present paper, that

$$P(\rho^{(1)} > \rho^{(2)} | y^{(1)}, y^{(2)}) = 1/2 + \int (\xi^{(2)}(x) - \xi^{(1)}(x)) d\xi^{(1)}(x) < 1/2. \tag{16.13}$$

Relations (16.4)-(16.6) and (16.13) imply conclusion (a') of Theorem 15.2.

Conclusion (b') of Theorem 15.2 follows from (16.4)-(16.6) and (16.8) and the fact that, by (16.11), 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. The proof is complete.

Part IV

EXTENSIONS AND CONCLUSION

17 Generalizations to dependence and non-identical distributions

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

According to the definition introduced by Cambanis, Keener and Simons (1983), an $n-$dimensional distribution is called $\alpha-$symmetric if its characteristic function (c.f.) can be written as $\phi((\sum_{i=1}^{n}|t_i|^{\alpha})^{1/\alpha})$, where $\phi : \mathbf{R}_+ \rightarrow \mathbf{R}$ is a continuous function and $\alpha > 0$. The number α is called the index and the function ϕ is called the c.f. generator of the $\alpha-$symmetric distribution. The class of $\alpha-$symmetric distributions is very broad and contains, in particular, spherical distributions corresponding to the case $\alpha = 2$ (see Fang, Kotz and Ng, 1990, p. 184). Spherical distributions, in turn, include such important examples as Kotz type, multinormal, multivariate t and multivariate stable laws (Fang et. al., 1990, Ch. 3). Furthermore, for any $0 < \alpha \leq 2$, the class of $\alpha-$symmetric distributions includes distributions of risks $X_1, ..., X_n$ that have the representation

$$(X_1, ..., X_n) = (ZY_1, ..., ZY_n) \tag{17.14}$$

where $Y_i \sim S_\alpha(\sigma, 0, 0)$ are i.i.d. symmetric stable r.v.'s with $\sigma > 0$ and the index of stability α and $Z \geq 0$ is a nonnegative r.v. independent of $Y_i's$ (see Fang et. al., 1990, p. 197). Models (17.14) and their convolutions belong to the class of models with common shocks Z, such as macroeconomic or political ones, that affect all risks Y_i.

It is important to emphasize here that the necessity in the study of effects of common shocks arises in many areas of economics and finance (see Andrews, 2003). The extensions of the results in the paper to such models provides, in fact, a new approach to the analysis of robustness of many economic models to both heavy-tailedness and to common shocks.

In addition, one should indicate here that the extensions of the results to $\alpha-$symmetric and, in particular, spherical distributions cover many thick-tailed models with finite variances and finite higher moments. For instance, multivariate $t-$distributions that belong to the class of spherical distributions, provide one of now well-established approaches to modelling heavy-tailedness phenomena with moments up to some order (see Praetz, 1972, Blattberg and Gonedes, 1974, and Glasserman, 2002).

The following theorems provide precise formulations of the extensions of the results obtained in the paper to the dependent case. According to the theorems, all the results established in the paper for convolutions of i.i.d. stable distributions with indices of stability α belonging to a certain range (and convolutions of those with log-concave distributions in the case of the class $\overline{\mathcal{CSLC}}$) continue to hold for convolutions of $\alpha-$symmetric distributions and models with common shocks (17.14) with parameters α in the same range.

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

Theorem 17.1 *Theorem 4.1, Corollary 5.4 and Theorems 10.1 and 11.1 continue to hold under any of the following two assumptions:*

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

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

Theorem 17.2 *Theorem 4.2, Corollary 5.5 and Theorems 10.2 and 11.2 continue to hold if any of the following assumptions is satisfied:*

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

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

Theorem 17.3 *Proposition 2.1, Theorem 4.3, Corollaries 5.1 and 5.2, Theorem 11.3, Proposition 14.1 and 15.1 continue to hold if any of the following is satisfied:*

the random vector $(X_1, ..., X_n)$ *or* $(\epsilon_1, ..., \epsilon_n)$ *entering their assumptions is a sum of independent random vectors* $(Y_{1j}, ..., Y_{nj})$, $j = 1, ..., k$, *where* $(Y_{1j}, ..., Y_{nj})$ *has an absolutely continuous* $\alpha-$*symmetric distribution with the c.f. generator* $\phi_j \in \Phi$ *and the index* $\alpha_j \in (1, 2]$. *In particular, the results hold when the vector of r.v.'s entering their assumptions is a sum of independent random vectors* $(Y_{1j}, ..., Y_{nj})$, $j = 1, ..., k$, *that have absolutely continuous spherical distributions with c.f. generators* $\phi_j \in \Phi$ *(the case* $\alpha_j = 2$ *for all* j*).*

the vector of r.v.'s $(X_1, ..., X_n)$ *or* $(\epsilon_1, ..., \epsilon_n)$ *entering the assumptions of the results is a sum of random vectors* $(Z_jY_{1j}, ..., Z_jY_{nj})$, $j = 1, ..., k$, *where* Y_{ij}, $i = 1, ..., n$, $j = 1, ..., k$, *are independent r.v.'s such that* $Y_{ij} \sim S_{\alpha_j}(\sigma_j, 0, 0)$, $\sigma_j > 0$, $\alpha_j \in (1, 2]$, *and* Z_j *are independent positive absolutely continuous r.v.'s independent of* Y_{ij}.

Theorem 17.4 *Theorem 4.4, Corollary 5.3 and Theorems 11.4, 15.2 and 15.3 continue to hold if any of the following is satisfied:*

the vector of r.v.'s $(X_1, ..., X_n)$ *or* $(\epsilon_1, ..., \epsilon_n)$ *entering their assumptions is a sum of independent random vectors* $(Y_{1j}, ..., Y_{nj})$, $j = 1, ..., k$, *where* $(Y_{1j}, ..., Y_{nj})$ *has an absolutely continuous* $\alpha-$*symmetric distribution with the c.f. generator* $\phi_j \in \Phi$ *and the index* $\alpha_j \in (0, 1)$;

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

As for generalizations of the main majorization results in Part I to the case of non-identical distributions, the following their analogues hold.

Let $\sigma_1,...,\sigma_n \geq 0$ be some scale parameters and let $X_i \sim S_\alpha(\sigma_i, \beta, 0)$, $\alpha \in (0,2]$, $\beta > 0$, be independent non-identically distributed stable r.v.'s. Further, let $\varphi(a,x)$ denote the tail probability $\varphi(a,x) = P\big(\sum_{i=1}^n a_{[i]}X_i > x\big)$, where, as in Subsection 2.1, $a_{[1]} \geq ... \geq a_{[n]}$ denote the components of the vector $a = (a_1,...,a_n) \in \mathbf{R}^n$ in decreasing order.[32] Similar to the proof of Theorems 4.1 and 4.3, one can show that the theorems continue to hold (in the same range of parameters r and α) for the function $\varphi(a,x)$, $x > 0$, if $\sigma_1 \geq ... \geq \sigma_n \geq 0$. Similarly, Theorems 4.2 and 4.4 continue to hold for the tail probabilities $\varphi(a,x)$, if $\sigma_n \geq ... \geq \sigma_1 \geq 0$.[33] Using conditioning arguments, one gets that these extensions also hold in the case of random scale parameters σ_i.

The proof of the results in Parts I and III on monotone consistency and on the robustness of the model of demand-driven innovation and spatial competition over time reveals that the results depend only on comparisons between $\varphi(a,x) = P(\overline{X}_n > x) = P\big(\sum_{i=1}^{n+1} a_{[i]}X_i > x\big)$ and $\varphi(b,x) = P(\overline{X}_n > x) = P\big(\sum_{i=1}^{n+1} b_{[i]}X_i > x\big)$, where $a = (1/n,...,1/n,0)$ and $b = (1/(n+1),...,1/(n+1),1/(n+1))$. From (2.1) and the above discussion it thus follows that analogues of the results of the paper on monotone consistency and growth theory also hold in the case of dependent non-identically distributed data and signals. The generalizations of the main majorization results also imply analogues of the results of the paper on the value at risk for financial portfolios and on optimal bundling with substitutes and complements for the case of dependent risks, valuations and tastes with non-identical distributions.

<u>Proof of Theorems 17.1- 17.4.</u> The proof of the extensions of Theorems 4.1-4.4 (and that of Proposition 2.1) to the dependent case follows the same lines as the proof of the above theorems since the following properties hold:

$\sum_{i=1}^n c_iX_i/\big(\sum_{i=1}^n c_i^\alpha\big)^{1/\alpha}$ has the same distribution as that of X_1 if $(X_1,...,X_n)$ has an $\alpha-$symmetric distribution (see, e.g., Fang, Kotz and Ng, 1990, Ch. 7);

The r.v.'s $\sum_{i=1}^n a_iY_{ij}$, $j = 1,...,k$, and $\sum_{i=1}^n b_iY_{ij}$, $j = 1,...,k$, are symmetric and unimodal if $(Y_{1j},...,Y_{nj})$ has an $\alpha-$symmetric distribution with the c.f. generator $\phi_j \in \Phi$ (this easily follows from a result due to R. Askey, see Theorem 4.1 in Gneiting, 1998);

The r.v.'s $Z_j\sum_{i=1}^n a_iY_{ij}$, $j = 1,...,k$, and $Z_j\sum_{i=1}^n b_iY_{ij}$, $j = 1,...,k$, are symmetric and unimodal if $Y_{ij} \sim S_{\alpha_j}(\sigma_j,0,0)$, $i = 1,...,n$, $j = 1,...,k$, and Z_j are positive r.v.'s independent of Y_{ij} (this follows by symmetry and unimodality of $\sum_{i=1}^n a_iY_{ij}$ and $\sum_{i=1}^n b_iY_{ij}$ implied by Theorem 2.7.6 in Zolotarev, 1986, p. 134, and Theorem 1.6 in

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

[33] These results for $\varphi(a,x)$ can be established in the same way as Theorems 4.1-4.4 using the fact that, by Theorem 3.A.4 in Marshall and Olkin (1979), the function $\chi_1(c_1,...,c_n) = \sum_{i=1}^n \sigma_i c_{[i]}^\alpha$ is strictly Schur-convex in $(c_1,...,c_n) \in \mathbf{R}_+^n$ if $\alpha > 1$ and $\sigma_1 \geq ... \geq \sigma_n \geq 0$ and the function $\chi_2(c_1,...,c_n) = \sum_{i=1}^n \sigma_i c_{[i]}^\alpha$ is Strictly Schur-concave in $(c_1,...,c_n) \in \mathbf{R}_+^n$ if $\alpha < 1$ and $\sigma_n \geq ... \geq \sigma_1 \geq 0$.

Dharmadhikari and Joag-Dev, 1988, p. 13, the definition of unimodality and conditioning arguments).

The proof of the rest of the extensions given by Theorems 17.1-17.4 is completely similar to that of the corresponding results in the paper, with the use of the analogues of Theorems 4.1-4.4 for convolutions of $\alpha-$symmetric distributions and models with common shocks (17.14).

18 Concluding remarks

The paper developed a unified approach to the analysis of robustness of a number of economic models using new majorization theory for heavy-tailed distributions. According to the main results obtained, many economic models are robust to heavy-tailedness assumptions as long as the distributions entering these assumptions are not extremely thick-tailed. But the conclusions of the models are reversed for sufficiently heavy-tailed distributions.

Furthermore, as follows from the extensions obtained in this part of the paper, in addition to heavy-tailedness, the paper accomplishes the unification of the analysis of robustness of economic models to such important distributional phenomena as dependence, skewness and the case of non-identical marginals.

As demonstrated in the paper, the main majorization results obtained in Part I provide a unification of the study of many problems in a number of areas in econometrics, statistics, finance, risk management and economic theory. Furthermore, the approach developed in this paper and the main majorization results obtained are applicable in all fields where the necessity in the analysis of concentration for linear combinations of r.v.'s arises. New applications of the approach and the majorization results in several such areas, including reliability theory, optimal voting, and bundling of financial securities, are currently under way by the author.

REFERENCES

Adams, W. J. and Yellen, J. L. (1976). Commodity bundling and the burden of monopoly. *Quarterly Journal of Economics* **90**, 475-498.

Agarwal, R. and Gort, M. (1996). The evolution of markets and entry, exit and survival of firms. *Review of Economics and Statistics* **78**, 489-498.

An, M. Y. (1998). Logconcavity versus logconvexity: a complete characterization. *Journal of Economic Theory* **80**, 350-369.

Andrews, D. W. K. (2003). Cross-section regression with common shocks. *Cowles Foundation Discussion Paper No. 1428*, Yale University.

Artzner, P., Delbaen, F., Eber, J.-M. and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance* **9**, 203-228.

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

Audretsch, D. B. (1991). New-firm survival and the technological regime. *Review of Economics and Statistics* **73**, 441-450.

Bakos, Y. and Brynjolfsson, E. (1999). Bundling information goods: pricing, profits and efficiency. *Management Science* **45**, 1613-1630.

Bakos, Y. and Brynjolfsson, E. (2000). Bundling and competition on the internet. *Marketing Science* **19**, 63-82.

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

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

Brock, W. A. and Mirman, L. J. (1972). Optimal economic growth and uncertainty: the discounted case. *Journal of Economic Theory* **4**, 479-513.

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

Chakraborty, I. (1999). Bundling decisions for selling multiple objects. *Economic Theory* **13**, 723-733.

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

Dansby, R. E. and Conrad, C. (1984). Commodity bundling. *American Economic Review* **74**, 377-381.

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

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

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

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

Embrechts, P., McNeil, A., and Straumann, D. (2002). Correlation and dependence in risk management: properties and pitfalls. In: *Risk Management: Value at Risk and Beyond* (M. A. H. Dempster, ed.), Cambridge University

Press, Cambridge, 176-223.

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

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

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

Fölmer, H. and Schied, A. (2002). Convex measures of risk and trading constraints. *Finance and Stochastics* **6**, 429-447.

Frittelli, M. and Gianin, E. R. (2002). Putting order in risk measures. *Journal of Banking & Finance* **26**, 1473-1486.

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

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

Gneiting, T. (1998). On $\alpha-$symmetric multivariate characteristic functions. *Journal of Multivariate Analysis* **64**, 131-147.

Huber, P. (1981). *Robust statistics.* New York: Wiley.

Ibragimov, R. (2004). A tale of two tails: peakedness properties in inheritance models of evolutionary theory. Mimeo, Yale University. Available at http://pantheon.yale.edu/˜ri23/Papers/TaleOfTwoTales.pdf

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

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

Jovanovic, B. (1982). Selection and the evolution of industry. *Econometrica* **50**, 649-670.

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

Kahneman, D. and Tversky, A. (1979). Prospect theory: an analysis of decisions under risk. *Econometrica* **47**, 263-291.

Kamien, M. and Schwartz, N. (1982). *Market structure and innovation.* New York: Cambridge University Press.

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

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

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

Krishna, V. (2002). *Auction theory.* San Diego: Academic Press.

Lewbel, A. (1985). Bundling of substitutes or complements. *International Journal of Industrial Organization* **3**, 101-107.

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

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

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

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

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

McAfee, R., McMillan, J. and Whinston, M. (1989). Multiproduct monopoly, commodity bundling and correlation of values. *Quarterly Journal of Economics* **114**, 371-384.

McCloughan, P. (1995). Simulation of concentration development from modified Gibrat growth-entry-exit processes. *Journal of Industrial Economics* **43**, 405-433.

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

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

Milgrom, P. and Weber, R. (1982). A theory of auctions and competitive bidding. *Econometrica* **50**, 1089-1122.

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

Nalebuff, B. (2004). Bundling as an entry barrier. *The Quarterly Journal of Economics* **119**, 159-187.

Nelson, R. R. and Winter, S. G. (1978). Forces generating and limiting concentration under Schumpeterian competition. *The Bell Journal of Economics* **9**, 524-548.

Palfrey, T. R. (1983). Bundling decisions by a multiproduct monopolist with incomplete information. *Econometrica* **51**, 463-483.

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

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

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

Salinger, M. A. (1995). A graphical analysis of bundling. *Journal of Business* **68**, 85-98.

Schmalensee, R. L. (1984). Gaussian demand and commodity bundling. *Journal of Business* **57**, S211-S230.

Shaked, M. and Shanthikumar, J. G. (1994). *Stochastic orders and their applications.* Boston: Academic Press.

Stockey, N. L. and Lucas, R. E. (1989). *Recursive methods in economic dynamics.* Cambridge: Harvard University Press.

Sutton, J. (1997). Gibrat's legacy. *Journal of Economic Literature* **35**, 40-59.

Thaler, R. (1985). Mental accounting and consumer choice. *Marketing Science* **3**, 199-214.

Thornton, J. and Marche, S. (2003). Sorting through the dot bomb rubble: how did the high-profile e-tailers fail? *International Journal of Information Management* **23**, 121-138.

Tong, Y. L. (1994). Some recent developments on majorization inequalities in probability and statistics. *Linear Algebra and its Applications* **199**, 69-90.

Uchaikin, V. V. and Zolotarev, V. M. (1999). *Chance and stability. Stable distributions and their applications.* Utrecht: VSP.

Venkatesh, R. and Kamakura, W. (2003). Optimal bundling and pricing under a monopoly: contrasting complements and substitutes from independently valued products. *Journal of Business* **76**, 211-231.

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