SHIFTING PARADIGMS: ON THE ROBUSTNESS OF ECONOMIC MODELS TO HEAVY-TAILEDNESS ASSUMPTIONS

Running title: Robustness of economic models

By Rustam Ibragimov[1] [2]

Yale University

Address for manuscript correspondence:

Rustam Ibragimov
Department of Economics
Yale University
28 Hillhouse Ave.
New Haven, CT 06511

Email address: rustam.ibragimov@yale.edu
Phone: +1-203-777-2750
Fax: +1-203-432-2128

ABSTRACT

The structure of many models in economics depends on majorization properties of convolutions of distributions. In this paper, we analyze robustness of these properties and the models based on them to heavy-tailedness assumptions. We show, in particular, that majorization properties of linear combinations of log-concavely distributed signals are reversed for very long-tailed distributions. As applications of the results, we study robustness of monotone consistency of the sample mean, value at risk analysis and the model of demand-driven innovation and spatial competition as well as that of optimal bundling strategies for a multiproduct monopolist in the case of an arbitrary degree of complementarity or substitutability among the goods. The implications of the models remain valid for not too heavy-tailed distributions. However, their main properties are reversed in the very thick-tailed setting.

KEYWORDS: Robustness, heavy-tailed distributions, innovation and spatial competition, firm growth, Gibrat's law, optimal bundling strategies, multiproduct monopolist, Vickrey auction, value at risk, coherent measures of risk, monotone consistency

[1]Please do not circulate without permission.

[2]The author is very grateful to Donald Andrews, Aydin Cecen, Peter Phillips, Herbert Scarf and the participants at the Prospectus Workshop at the Department of Economics, Yale University for many helpful comments and suggestions. The financial support from the Yale University Graduate Fellowship and the Cowles Foundation Prize is gratefully acknowledged.

0. INTRODUCTION AND DISCUSSION OF THE RESULTS

0.1. Robustness of models and tail probabilities of linear combinations. In recent years, many papers in economics and finance have focused on the analysis of the so-called "thick-tailed" paradigm. It was documented in numerous studies that the time series encountered in many fields in economics and finance appear to have heavy-tailed distributions for which the variance and higher moments fail to exist (see Axtell (2001), the discussions in Loretan and Phillips (1994), Duffie and Pan (1997) and Gabaix, Gopikrishnan, Plerou and Stanley (2003) and references therein). Motivated by these empirical findings, a number of studies in financial economics have focused on portfolio and value-at-risk modelling with heavy-tailed returns (see, e.g., the reviews in Duffie and Pan (1997), Uchaikin and Zolotarev (1999, Ch. 17) and Glasserman, Heidelberger and Shahabuddin (2002)). Several authors considered problems of statistical inference for data from thick-tailed populations (see Loretan and Phillips (1994), the papers in Adler, Feldman and Taqqu (1998) and references therein).

The results of many models in economics, finance, risk management and operations research, as well the analysis of their robustness to distributional assumptions, depend on majorization and dominance properties of tail probabilities and expectations of functions of random variables (r.v.'s) and their linear combinations. The concepts of stochastic dominance are of central importance in the portfolio choice problems and the analysis of risk measures (see Levy (1992) and Ruszczyński and Vanderbei (2003)). Furthermore, the study of robustness of option pricing formulae to assumptions on the distributions of the underlying assets' prices can be reduced to the analysis of extremal properties of linear functionals of probability measures with fixed moments and problems of deriving sharp semiparametric bounds on expectations of functions of random variables with prescribed distributional characteristics, such as moments or probability density norms (see the discussion in de la Peña, Ibragimov and Jordan (2003) and references therein). Moreover, the problems concerning the properties of cdf's, moments and tail probabilities of r.v.'s and their linear combinations also naturally appear in various microeconomic models. In monopoly theory, for instance, the tail probability of a sum of functions of consumers' valuations represents the probability that a monopolist will choose to produce and sell the good (see, e.g., Cornelli (1996)). The analysis of optimal bundling strategies for a multiproduct monopolist depends on stochastic comparisons between consumers' reservation prices (valuations) for goods provided by the seller and their bundles (see, among others, Adams and Yellen (1976), Palfrey (1983), McAfee, McMillan and Whinston (1989), Bakos and Brynjolfsson (1999), Fang and Norman (2003) and Venkatesh and Kamakura (2003)). According to Heckman and Honoré (1990), the majorization properties of (conditional) moments and tail probabilities of r.v.'s are central to the analysis of robustness of the Roy model of self-selection and earnings inequality to departures from the conditions of log-normality of skills. Similarly, a number of problems in theories of firm growth involve the analysis of stochastic comparisons for r.v.'s and their sums as well as the study of sharp bounds on their functionals (see Jovanovic and Rob (1987), Jovanovic and MacDonald (1994) and Sutton (1997, Section IV)).

0.2. Majorization properties of log-concavely distributed r.v.'s. Powerful tools for analyzing the ordering and extremal properties of expected values of functions of linear combinations of r.v.'s 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 more diverse than those of

b (see Marshall and Olkin (1979)). In this context, it is easy to see that the following relations hold:

$$(1/(n+1),...,1/(n+1),1/(n+1)) \prec (1/n,...,1/n,0), \quad n \geq 1. \tag{0.1}$$

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.

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 the latter 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 well-known result concerning majorization and peakedness properties linear combinations of log-concavely distributed r.v.'s:

Proposition 0.1 *(Proschan (1965)). 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 0.1 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 0.1 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.

0.3. Implications for monotone consistency of the sample mean and portfolio value at risk. Proposition 0.1 and its extensions have been applied to the analysis of a number of problems in economics, statistics and other fields. For instance, Eaton (1988) used generalizations of the results to obtain concentration inequalities for Gauss-Markov estimators. Several authors (see, e.g., Proschan (1965), Tong (1994) and Jensen (1997)) discussed implications of Proposition 0.1 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 (0.1), from Proposition 0.1 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.

Proposition 0.1 also has the following important interpretation in the framework of value-at-risk (VaR) analysis and portfolio choice theory. 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[3]. Let $Z_w = \sum_{i=1}^n w_i X_i$ be the return on a portfolio of risks $X_1, ..., X_n$ with weights $w = (w_1, ..., w_n) \in \mathbf{R}_+^n$, $\sum_{i=1}^n w_i = 1$. Further, let $\underline{w} = (1/n, 1/n, ..., 1/n)$ 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). According to Proposition 0.1, "diversification" of a portfolio of i.i.d. symmetric log-concave risks $X_1, ..., X_n$ (formalized 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. This implies that, for all $\alpha \in (0, 1/2)$, $VaR_\alpha(Z_v) \leq VaR_\alpha(Z_w)$ if $v \prec w$. Furthermore, by relations (0.1), the value at risk is minimal for the portfolio weights $\underline{w}$ and is maximal for the weights $\overline{w}$:

$$VaR_\alpha(Z_{\underline{w}}) \leq VaR_\alpha(Z_w) \leq VaR_\alpha(Z_{\overline{w}}). \tag{0.2}$$

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

$$VaR_\alpha(X_1 + X_2) \leq VaR_\alpha(X_1) + VaR_\alpha(X_2) \tag{0.3}$$

for all $\alpha \in (0, 1/2)$. Thus, the value at risk is a coherent risk measure in the sense of Artzner, Delbaen, Eber and Heath (1999) in the world of log-concave distributions (see also Embrechts, McNeil and Straumann (1999) and Section 3 in the present paper).

0.4. Implications for firm growth and Gibrat's law. 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 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. Using Proposition 0.1, Jovanovic and Rob (1987) showed that, in

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

the case of symmetric and log-concavely distributed signal shocks $\epsilon_1, \epsilon_2, ...$, 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.

0.5. Majorization comparisons and optimal bundling decisions for a multiproduct monopolist. Applying analytical and numerical techniques to derive stochastic comparisons related to those implied by Proposition 0.1 for prescribed distributions for reservation prices in the case of two products and their packages (such as bivariate uniform or Gaussian distributions), many of 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)). Palfrey (1983) obtained results that give conditions under which consumers prefer (*ex ante*) a single bundled Vickrey auction to separate provision of independently priced goods[4] under the MPSM property discussed in Subsection 0.3 (see Theorem 8 in Palfrey (1983)). Palfrey (1983) showed that, in the case of two bidders, 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. Palfrey's (1983) results also imply 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 (Theorems 5-7 in Palfrey (1983)). 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 (0.3). 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 MPSM property[5]. Among other results, Bakos and Brynjolfsson (1999) showed that, in the latter 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 as the number of goods in the bundle increases[6]. Recently,

[4]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 5 in the present paper).

[5]As discussed above, by Proposition 0.1, 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)).

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

applying Proposition 0.1, 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 of 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. 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)) is that, for light-tailed distributions, consumers' valuations per good for a bundle typically have a lower variance relative to the valuations for individual goods[7].

0.6. Extensions of Proposition 0.1 and its implications. A number of papers in probability and statistics have focused on extension of Proschan's results (see, e.g., Chan, Park and Proschan (1989), the review in Tong (1994), Jensen (1997) and Ma (1998)). These studies allow one to readily obtain extensions of the results discussed in Subsections 0.3-0.5 to more general classes of distributions[8]. One should emphasize, however, that in all the studies that dealt with generalizations of Proposition 0.1, 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[9].

One should also note here that departures from conditions of log-concavity of distributions are necessary in the study of robustness of models involving them to heavy-tailedness assumptions since all moments of a log-concave density are finite and thus any such density is very light-tailed (see An (1998)).

0.7. Main majorization results of the paper and their implications for monotone consistency and portfolio VaR. In this paper, we present results on robustness of majorization properties of tail probabilities of linear combinations of r.v.'s and models in economics and risk management based on them to thick-tailedness assumptions. In particular, we show that majorization properties of convex combinations of r.v.'s given by Proposition 0.1 continue to hold for not too heavy distributions, as modelled by convolutions of stable distributions with (different) characteristic exponents greater than the threshold value of one and log-concave distributions (Theorem 2.3). However, the properties are reversed for r.v.'s with very thick-tailed distributions, as modelled by convolutions of stable distributions with indices of stability less than one (Theorem 2.4). As discussed in the previous subsection, to our knowledge, the latter results are the first ones that show that the general majorization properties given by Proposition 0.1 are reversed for certain classes of distributions. Moreover, we obtain results that give analogues of Proposition 0.1 for heavy-tailed r.v.'s and majorization comparisons between powers of coefficients of their linear

[7] Further intuition behind the power of 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)).

[8] For instance, from Chan, Park and Proschan (1989) it follows that the results in Subsection 0.3-0.5 continue to hold for (dependent) r.v.'s (representing risks, consumers' signals or valuations) that have a sign-invariant and Schur-concave joint density. From Ma (1998) it follows that under certain additional assumptions, the results hold for non-identical log-concave distributions.

[9] One should note that the proof in Proschan (1965) can be reproduced word to word with respective changes of signs 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 0.1 for Schur-concavity of $\psi(a, x)$, $x > 0$, is hopeless.

combinations (Theorems 2.1 and 2.2). From our results it follows, in particular, that the implications of the proposition for monotone consistency of the sample mean and the portfolio value at risk discussed in Subsections 0.5 and 0.6 continue to hold for not too thick-tailed distributions (Corollaries 3.1 and 3.2). In addition, it is demonstrated that the VaR is a coherent measure of risk in the case of not very-heavy tailed returns (see the discussion in Section 3). The results on the portfolio VaR are reversed, however, for very long-tailed distributions of the risks (Corollary 3.3). From our results it follows that a diversification of a portfolio of very heavy tailed risks, as modelled by convolutions of stable distributions with characteristic exponents less than one, leads to an increase in the riskiness of the portfolio. More specifically, the signs of inequalities in (0.2) and (0.3) are reversed for very long-tailed risks. For instance, our results reveal that (see Corollary 3.3), in a world of very heavy-tailed risks, the value at risk always has a strict superadditivity property instead of subadditivity in (0.3) and thus is not a coherent risk measure in the sense of Artzner et. al. (1999). 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 (0.2) and their analogues in the very thick-tailed case (Corollaries 3.4 and 3.5).

0.8. Robustness of the model of demand-driven innovation and spatial competition to heavy-tailedness assumptions. In a similar context, using the general probability results obtained in 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 to the assumptions of heavy-tailedness of signals' distributions. In particular, we show that the properties of the model for log-concavely distributed signals (see Subsection 0.4) remain valid for the class of not too heavy-tailed distributions, as modelled by convolutions of stable distributions with the characteristic exponents in the interval $(1, 2)$ and log-concave distributions (Theorem 4.1). However, we prove that the above properties of the model of demand-driven innovation and spatial competition are reversed under the assumption that the distributions of the signals are very long-tailed (Theorem 4.2).

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 a 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.

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 4.3). The latter 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 very 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 with log-concavely distributed or not too heavy-tailed signals, in the model with arbitrary convex cost functions $C(y)$ and $K(z)$ and very 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 very heavy-tailed signals, there is negative autocorrelation in the size-difference. Essentially, in the case of very heavy-tailed signals, smaller firms, in fact, have an advantage over their larger counterparts. The underlying intuition is that in the presence of very 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 very heavy-tailed signals, the firm growth is likely to decrease with firm size. Similar to the discussion in Subsection 0.4, the above implies that Gibrat's law does not hold in the setup. Moreover, in the case of very-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.

0.9. Optimal bundling decisions for a multiproduct monopolist in the case of long-tailed reservation prices and interrelated goods. 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. Using the general majorization results obtained in the paper, 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. Among other results, we show that if the goods provided on a Vickrey auction are independently priced or are substitutes (or complements with not very high degree of complementarity) and bidders' tastes for the objects are not very heavy tailed, then the risk-neutral monopolist strictly prefers separate provision of the products to any other bundling decision (Theorem 5.1). The results are reversed, however, in the case of a risk-averse auctioneer providing independently priced goods or complements (or substitutes with not very high degree of substitutability) to consumers with very long-tailed tastes for the products (Theorem 5.2)[10]. According to our analysis, in the latter case, 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 (0.3), 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 very heavy-tailed, as modelled by positive stable distributions (see Section 1), then they unanimously prefer separate Vickrey auctions to any other bundling decision (Theorem 5.3). These results are at odds with a setting where valuations have a finite distributional support in which, according to Palfrey (1983), consumers never unanimously prefer separate provision of the products, as discussed in Subsection 0.5.

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

Using the main probabilistic results derived in this paper, 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 5.4-5.7). We show, in particular, that, for products with high marginal costs, the seller's optimal strategy is to provide complements with very 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 5.4 and 5.5). Contrary to the case where goods with very light-tailed valuations are considered, as in Bakos and Brynjolfsson (1999) and Fang and Norman (2003), if consumers' tastes for the products are very long-tailed, then the monopolist's optimal strategy is to provide independently priced goods or complements (or substitutes with not very high degree of substitutability) with relatively high marginal costs as a single bundle and those with sufficiently low marginal costs separately (Theorem 5.7). 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).

The underlying intuition that drives our results on bundling is closely related to that based on the behavior of variance in the world of light-tailed valuations (see Subsection 0.5). Namely, our majorization results imply, essentially, that, in the case of not very heavy-tailed reservation prices, the consumers' valuations for bundles of goods *always* have less spread relative to the valuations for component goods, as measured by peakedness. On the other hand, in the case of very heavy-tailed valuations, the spread of reservation prices for bundles is *always* greater than that of valuations for components (see Section 5 for more on the intuition).

0.10. Organization of the paper. The paper is organized as follows: Section 1 contains notations and definitions of classes of distributions used throughout the paper. In Section 2, we derive the main results of the paper on majorization properties of linear combinations of long-tailed r.v.'s. Section 3 presents implications of the majorization results in Section 2 in the study of monotone consistency of the sample mean and portfolio value at risk and in the analysis of coherency of the VaR. In Section 4, we obtain the applications of the general majorization results in analysis of the robustness of the model of demand-driven innovation and spatial competition described in Subsection 0.4 to the assumptions of heavy-tailedness of signals' distributions. In Section 5 the general majorization results are applied to the study of the optimal bundling strategies for a multiproduct monopolist in the case of heavy-tailed tastes for and an arbitrary degree of complementarity or substitutability among the goods produced. Finally, Section 6 contains the proofs of the results obtained in the paper.

1. 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 $S_\alpha(\sigma, \beta, \mu)$.

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)[11]; $\alpha = 1/2$ and $\beta \pm 1$ (Lévy distributions)[12]. Degenerate distributions correspond to the limiting case $\alpha = 0$.

The index of stability α characterizes the heaviness (the rate of decay) of the tails of stable distributions $S_\alpha(\sigma, \beta, \mu)$. In particular, the $p-$th absolute moments $E|X|^p$ of a r.v. $X \sim S_\alpha(\sigma, \beta, \mu)$, $\alpha \in (0, 2)$ are finite if $p < \alpha$ and are infinite otherwise. The symmetry index β characterizes the skewness of the distribution. The stable distributions with $\beta = 0$ are symmetric about the location parameter μ. The stable distributions with $\beta = \pm 1$ and $\alpha \in (0, 1)$ (and only they) are one-sided, the support of these distributions is the semi-axis $[\mu, \infty)$ for $\beta = 1$ and is $(-\infty, \mu]$ (in particular, the Lévy distribution with $\mu = 0$ is concentrated on the positive semi-axis for $\beta = 1$ and on the negative semi-axis for $\beta = -1$). In the case $\alpha > 1$ the location parameter μ is the mean of the distribution $S_\alpha(\sigma, \beta, \mu)$. The scale parameter σ is a generalization of the concept of standard deviation; it coincides with the latter in the special case of Gaussian distributions ($\alpha = 2$). 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[13], as defined in Subsection 0.2 in the introduction.

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[14] $\alpha \in (r, 2]$ and $\sigma > 0$. 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[15]. 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[16] $\alpha \in (0, r)$ and $\sigma > 0$. That is, $\underline{\mathcal{CS}}(r)$ consists of distributions of r.v.'s X such that, for some $k \geq 1$, $X = Y_1 + ... + Y_k$, where Y_i, $i = 1, ..., k$, are independent r.v.'s such that $Y_i \sim S_{\alpha_i}(\sigma_i, 0, 0)$, $\alpha_i \in (0, r)$, $\sigma_i > 0$, $i = 1, ..., k$.

A linear combination of independent stable r.v.'s with the *same* characteristic exponent α also has a stable distribution with the same α. However, in general, this does not hold true in the case of convolutions of stable distributions with *different* indices of stability. Therefore, the class $\overline{\mathcal{CS}}(r)$ of *convolutions* of symmetric stable

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

[12] 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.

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

[14] 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.

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

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

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 $0 < r_1 < r_2 \le 2$, 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 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[17] $\alpha < 2$.

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 $\underline{\mathcal{CS}}(r)$).

2. MAJORIZATION PROPERTIES OF HEAVY-TAILED DISTRIBUTIONS

Theorems 2.1-2.4 in this section give analogues of Proposition 0.1 in Subsection 0.2 in the introduction for heavy-tailed r.v.'s. In particular, according to the following Theorem 2.1, 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 0.1 with respect to the comparisons between the powers of the components of the vectors of weights of the combinations.

Theorem 2.1 *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 0.1 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 2.2 below, the majorization properties of the tail probabilities $\psi(a, x)$ in Theorem 2.1 are reversed in the case of r.v.'s from the classes $\underline{\mathcal{CS}}(r)$.

Theorem 2.2 *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 0.1 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 2.3 below, peakedness properties of linear combinations of r.v.'s with not too heavy-tailed distributions, as modelled, e.g., by convolutions of log-concave distributions and symmetric stable distributions with characteristic exponents greater than one, are the same as in the case of log-concave distributions in Proschan (1965).

[17] 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 2.3 *Proposition 0.1 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 2.4, peakedness properties given by Proposition 0.1 and Theorem 2.3 above are reversed in the case of r.v.'s with very heavy-tailed distributions, as modelled by convolutions of stable distributions with indices of stability less than one.

Theorem 2.4 *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 0.1 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 2.1. If r.v.'s $X_1, ..., X_n$ have a symmetric Cauchy distribution $S_1(\sigma, 0, 0)$ which is, as discussed in Section 1, exactly at the dividing boundary between the class $\underline{\mathcal{CS}}(1)$ in Theorem 2.4 and the class $\overline{\mathcal{CSLC}}$ in Theorem 2.3, 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 2.1 and 2.2 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 2.1-2.4, the above implies that Theorems 2.3 and 2.4 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 2.1 and 2.2 continue to hold for convolutions of distributions from the classes $\overline{\mathcal{CS}}(r)$ and $\underline{\mathcal{CS}}(r)$ with symmetric stable distributions $S_r(\sigma, 0, 0)$. The latter generalizations imply corresponding extensions in the applications of majorization properties of linear combinations of heavy-tailed r.v.'s throughout the rest of the paper.

3. MONOTONE CONSISTENCY, PORTFOLIO VALUE AT RISK AND COHERENCY OF THE VaR

Theorem 2.3 provides the following result concerning the monotone consistency properties of the sample mean for data from heavy-tailed population.

Corollary 3.1 *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 3.1, from the majorization results given by Theorem 2.3 it follows, similar to the case of log-concave distributions in Subsection 0.3 in the introduction, that diversification of a portfolio of not too 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 0.3 in the introduction, 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 3.2 *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 $\overline{w}$. For all $\alpha \in (0, 1/2)$,* one also has *$VaR_\alpha(X_1 + X_2) < VaR_\alpha(X_1) + VaR_\alpha(X_2)$.*

In contrast, the results in Theorem 2.4 imply that the results for the VaR of portfolios discussed in Subsection 0.3 in the introduction are reversed under the assumption that the distributions of the risks $X_1, ..., X_n$ are very long-tailed, as modelled by convolutions of stable distributions with indices of stability less than 1. In the latter setup, diversification of a portfolio of the risks increases riskiness of its return. We have the following

Corollary 3.3 *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 $\overline{w}$. For all $\alpha \in (0, 1/2)$,* one also has *$VaR_\alpha(X_1) + VaR_\alpha(X_2) < VaR_\alpha(X_1 + X_2)$.*

Theorems 2.1 and 2.2 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 3.2 and 3.3 in the worlds of not too heavy-tailed and very heavy-tailed risks.

Corollary 3.4 *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 3.5 *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 $\overline{w}$.

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 Fritelli 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., Fritelli 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[18] of a risk holder, in particular, from the regulatory point of view as well as from liquidity considerations (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 Fritelli 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 not very heavy-tailed i.i.d. r.v.'s $X_i \sim \overline{\mathcal{CSLC}}$ given by Corollary 3.2, imply that the value at risk is, in fact, a coherent measure of risk in the world of such risks.

Furthermore, from Corollary 3.3 it follows that axioms A2 and A5 are *always* violated for risks with very heavy-tailed distributions, (even) under their independence. Thus, the value at risk is not a coherent risk measure in the world of very long-tailed distributions.

Remark 3.1. It is well-known that if r.v.'s X and Y are such that $P(X > x) \leq P(Y > x)$ for all $x \in \mathbf{R}$, then $EU(X) \leq EU(Y)$ for all increasing functions $U : \mathbf{R} \to \mathbf{R}$ for which the expectations exist (see, e.g., Shaked and Shanthikumar (1994, pp. 3-4)). This fact and Theorems 2.1-2.4 imply corresponding results concerning majorization properties of expectations of (utility or payoff) functions of linear combinations of heavy-tailed r.v.'s. In particular, Theorems 2.1 and 2.2 give sharp bounds on the expected payoffs of contingent claims written on a portfolio of heavy-tailed risks similar to those in Corollaries 3.4 and 3.5. 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 2.1 and is Schur-concave in $(a_1^r, ..., a_n^r)$ under the assumptions of Theorem 2.2. 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 port-

[18]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.

folios of risks satisfying Theorem 2.1 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 2.2. 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 latter results complement those in Efron (1969) and Eaton (1970) (see also Marshall and Olkin (1979, pp. 361-365)) who studied classes of functions $U : \mathbf{R} \to \mathbf{R}$ and r.v.'s $X_1, ..., X_n$ for which Schur-concavity of $\varphi(a)$, $a \in \mathbf{R}_+^n$ in $(a_1^2, ..., a_n^2)$ holds. Further, we obtain that $\varphi(a)$ is Schur-convex in $a \in \mathbf{R}_+^n$ under the assumptions of Theorem 2.3 and is Schur-concave in $a \in \mathbf{R}_+^n$ under the assumptions of Theorem 2.4. 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 2.4, the expectations $EU(|\sum_{i=1}^n a_i X_i|)$ are infinite[19] for all $a \in \mathbf{R}_+^n$. 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.

4. DEMAND-DRIVEN INNOVATION AND SPATIAL COMPETITION

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 described in Subsection 0.6 in the introduction, 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

$$p_{\hat{\theta}} = p(\nu_{t-1}) - \rho(\hat{\theta}, \theta_t). \tag{4.1}$$

For $n > 0$, denote $\overline{s}_n = n^{-1} \sum_{i=1}^n s_i$ and $\overline{\epsilon}_n = n^{-1} \sum_{i=1}^n \epsilon_i$. Further, let $F(x; n) = P(|\overline{\epsilon}_n| \leq x)$, $x \geq 0$, $n = 1, 2, ...$, denote the cdf of $|\overline{\epsilon}_n|$, $n = 1, 2, ...$, 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)[20]) $\hat{\theta} = argmax_{\tilde{\theta}} N^{-1} \sum_{i=1}^N \rho(\tilde{\theta}, s_i) = argmax_{\tilde{\theta}} N^{-1} \sum_{i=1}^N (\tilde{\theta} - s_i)^2 = \overline{s}_N$. 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 $\rho(\hat{\theta}(\mathcal{S}), \theta) = \overline{\epsilon}_N^2$.

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

[20] 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.

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

$$\xi(x; y + z) = P(\rho(\hat{\theta}(\mathcal{S}), \theta) \leq x | y + z) = \sum_{n=0}^{\infty} F(\sqrt{x}; n)\pi(n; y + z), \tag{4.2}$$

$x \geq 0$ (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

$$p_{\hat{\theta}} - C'(y) + G'(y + z) = 0, \quad -K'(z) + G'(y + z) = 0. \tag{4.3}$$

The second-order conditions for a maximum are

$$G''(y + z) < C''(y), \quad C''(y)K''(z) > G''(y + z)(C''(y) + K''(z)) \tag{4.4}$$

(conditions (4.4) imply $G''(y + z) < K''(z)$). If

$$G'(y + z) \leq K'(z) \tag{4.5}$$

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

$$z = 0. \tag{4.6}$$

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

$$p_{\hat{\theta}} - C'(y) + G'(y) = 0, \tag{4.7}$$

$$G''(y) < C''(y). \tag{4.8}$$

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 the latter assumption, one can implicitly differentiate first-order conditions (4.3) and (4.7) (see Jovanovic and Rob (1987)).

Evidently, the condition $G'' < 0$ suffices for conditions (4.4) and (4.8) 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 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

$$\pi(n; y + z) = \pi_0(n; y + z) = \frac{[\mu(y + z)]^n}{n!} exp(-\mu(y + z)), \quad n = 1, 2, ... \tag{4.9}$$

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

$$\pi(n;y+z)=\pi_1(n;y+z)=\frac{[\mu(y+z)]^{n-1}}{(n-1)!}exp(-\mu(y+z)),\quad n=1,2,... \tag{4.10}$$

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

Lemma 4.1 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 4.1 *(Jovanovic and Rob (1987)). Suppose that, conditionally on $y+z$, N has a Poisson distribution $\pi_0(n;y+z)$ given by (4.9). 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 4.1 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 4.2. In the proposition and its analogues for heavy-tailed signals below (Theorems 4.1 and 4.2), $y_t^{(1)}$ 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 4.2 *(Jovanovic and Rob (1987)). Suppose that, conditionally on $y+z$, N has a Poisson distribution $\pi_0(n;y+z)$ in (4.9). 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 (4.3) and (4.4) or (4.6)-(4.8), 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 4.1 and Proposition 4.2 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 (4.10) as well as under the assumption that conditions (4.6)-(4.8) are satisfied.

The following theorem provides a generalization of Proposition 4.2 that shows that the results obtained by Jovanovic and Rob (1987) hold in the case of not very thick-tailed signals.

Theorem 4.1 *Suppose that, conditionally on $y+z$, N has a Poisson distribution $\pi_0(n;y+z)$ in (4.9) or a shifted Poisson distribution $\pi_1(n;y+z)$ in (4.10). 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 4.2 hold.*

Lemma 4.2 below shows that the sufficient conditions for an interior maximum in Lemma 4.1 which imply strict concavity of the function $G(y+z)$ are satisfied for shocks $\epsilon_1, \epsilon_2, ...$ with not very fat-tailed symmetric stable distributions.

Lemma 4.2 *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 4.2 and Theorem 4.1 are reversed in the case of shocks $\epsilon_1, \epsilon_2, ...$ with very fat tails.

Theorem 4.2 *Suppose that, conditionally on $y+z$, N has a shifted Poisson distribution $\pi_1(n; y+z)$ given by (4.10). 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 (4.3) and (4.4) or (4.6)-(4.8), 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 4.3 below, condition (4.5) and, thus, relation (4.6) is satisfied in the case of very 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 very heavy-tailed, then each firm chooses zero informational gathering effort: $z = 0$.

Theorem 4.3 *Under the assumptions of Theorem 4.2, $G'(y+z) \leq 0$. Therefore, if $K'(z) > 0$, then (4.5) is satisfied and the optimal choice of informational gathering effort is $z = 0$.*

Proposition 4.2 and Theorem 4.1 imply that, in the case of not very heavy-tailed signals, relatively large firms are likely to stay larger; in addition, the size-difference is positively autocorrelated. According to Theorem 4.2, these conclusions are reversed in a world of very 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 4.2 and Theorems 4.1-4.3 is that the larger is a firm's size, the greater is the amount of information the firm gets[21]. The samples of consumers' signals are informative about the ideal product θ if the signals' distributions are not very heavy tailed, as in Proposition 4.2 and Theorem 4.1. However, they are uninformative about θ in the case of very long-tailed distributions in Theorems 4.2 and 4.3. The larger firms that learn more are thus more likely to come up with a successful product if the signals are not very fat-tailed (see the discussion in Jovanovic and Rob (1987)). In

[21] One should also note that, under (4.3) and (4.4), 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)).

a world of very heavy-tailed signals, on the other hand, smaller firms that get less uninformative signals have an advantage over their larger counterparts (see Subsection 0.8 in the introduction). 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.

5. OPTIMAL BUNDLING DECISIONS FOR COMPLEMENTS AND SUBSTITUTES UNDER HEAVY-TAILEDNESS

Consider a setting with a single profit-maximizing risk-neutral[22] seller providing m goods to n consumers. 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))[23]. 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[24]. 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 disposal, the natural analogues of this property for interrelated goods are subadditivity $v(B) \leq \sum_{i \in B} v(\{i\})$ in the case of substitutes and superadditivity $\sum_{i \in B} v(\{i\}) \leq v(B)$ in the case of complements (see Dansby and Conrad (1984), Lewbel (1985) 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

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

[23] 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.

[24] 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)).

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 $r \in (0, 2]$, $g_r(x) = x^r I(x \geq 0)$, $h_r(x) = x|x|^{r-1}$, $x \in \mathbf{R}$, 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: $v(g_r, B) \geq 0$ for all $B \in 2^M$. It is easy to see that, for all $B \in 2^M$, $v(g_r, B) \leq \sum_{i\in B} v(g_r, \{i\})$, if $r \leq 1$, and $\sum_{i\in B} v(g_r, \{i\}) \leq v(g_r, B)$, if $r \geq 1$, and $X_i \geq 0$, $i \in B$. That is, consumers' reservation price $v(g_r, B)$ for a bundle is subadditive in those for the component products if $r \leq 1$, as it is typically required for substitutes, and is superadditive in the rectangle of non-negative tastes if $r \geq 1$, as it is usually assumed in the case of complements. Similarly, for $r \leq 1$, 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 $r \geq 1$, 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 $v(h_r, \{i\}) \geq 0$, $i \in B$, then $\sum_{i\in B} v(h_r, \{i\}) \leq v(h_r, B)$ for $r \geq 1$, and $v(h_r, B) \leq \sum_{i\in B} v(h_r, \{i\})$ for $r \leq 1$. If $v(h_r, \{i\}) \leq 0$, $i \in B$, then $v(h_r, B) \leq \sum_{i\in B} v(h_r, \{i\})$ for $r \geq 1$, and $\sum_{i\in B} v(h_r, \{i\}) \leq v(h_r, B)$ for $r \leq 1$. The above super- and subadditivity properties of $v(h_r, B)$ for $r \geq 1$ 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)).

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 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 5.1 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 valuations. According to the theorem, if consumers' tastes are not very heavy-tailed and the goods are independently priced or 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 5.1 *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 5.1. From the proof of Theorem 5.1 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. As in Remark 3.1, we get, therefore, by Shaked and Shanthikumar (1994, pp. 3-4), 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 5.2 below, this, in turn, implies that Theorem 5.1 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 5.1 is that, in the case of not very heavy-tailed tastes, similarly 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 very 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 5.1 for very 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 reversal of Theorem 5.1 holds.

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 5.2, in the latter case, the auctioneer strictly prefers providing all the items through one Vickrey auction to any other bundling decision, if consumers' tastes are very heavy-tailed and the goods are independently priced or are complements (or are substitutes with not very high degree of substitutability).

Theorem 5.2 *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.

Using the general majorization properties of long-tailed distributions presented in Section 2, one can also obtain the following Theorem 5.3 that characterizes buyers' preferences over the bundled auctions in the case of independently priced goods and very heavy-tailed 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 5.3, 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 very 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 5.3 does not contradict Theorem 6 in Palfrey (1983) since the support of heavy-tailed distributions in Theorem 5.3 is the infinite positive semi-axis $\mathbf{R}_+$).

Theorem 5.3 *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 optimal bundling decisions in Vickrey auctions in the case of very long-tailed tastes in Theorems 5.2 and 5.3 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, the latter implications hold even in the case of any consumer, whatever the values of her reservation prices are. Therefore, contrary to the case of very 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 5.2. 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, with two buyers, the bidders unanimously prefer separate provision of items *ex post* and thus *ex ante* and the seller, on the other hand, prefers a single auction. Since the above results are, essentially, deterministic, all they are robust with respect to risk attitudes of the seller and the buyer. However, 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. On the other hand, from Theorem 5.2 with $r = 1$ and Theorem 5.3 it follows that, in the case of an arbitrary number of buyers with (very heavy-tailed) positive stable reservation prices, 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.

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 5.4 and 5.5 characterize the optimal bundling strategies for a multiproduct monopolist in the latter 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 5.4 it follows that if consumers' tastes are not very heavy-tailed 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 0.5 in the introduction).

Theorem 5.4 *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 5.5, the patterns in the solutions to the seller's optimal bundling problem in Theorem 5.4 are reversed if consumers' tastes are very heavy-tailed and the goods are independently priced or are complements (or are substitutes with not very high degree of substitutability).

Theorem 5.5 *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 5.6 and 5.7 below give analogues of the results in Theorems 5.4 and 5.5 in the case of independently priced goods ($r = 1$).

Theorem 5.6 *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 5.4 holds.*

Theorem 5.7 *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 5.5 holds.*

Similar to the analysis in Bakos and Brynjolfsson (1999), the underlying intuition for Theorems 5.4 and 5.6 is that, for not very heavy-tailed distributions of 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 5.5 and 5.7 are driven by the fact that, in the case of very heavy-tailed reservation prices, peakedness of the valuations per good in bundles decreases with their size. Therefore, bundling of goods in the case of very 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 5.3. The assumptions of Theorem 5.5 with $r \geq 1$ (and those of Theorem 5.7) 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 5.4-5.7 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 5.5 and 5.7 are necessary since otherwise the monopolist would set an infinite price for each bundle of goods under very heavy-tailed distributions of consumers' tastes considered in the theorems.

Remark 5.4. It is important to note that Theorems 5.5 and 5.7 shed new light on marketing strategies involving 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 on 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. The latter 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 very 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 relative small fraction of consumers that have 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 under the free disposal assumption or symmetric long-tailed valuations 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 very extreme valuations.

6. PROOFS

Proof of Theorems 2.1 and 2.2. 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 $x \in \mathbf{R}$,

$$\psi(c, x) = P\Big(X_1 > x / \Big(\sum_{i=1}^n c_i^\alpha\Big)^{1/\alpha}\Big). \tag{6.1}$$

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

$\sum_{i=1}^n a_i^\alpha = \sum_{i=1}^n (a_i^r)^{\alpha/r} < \sum_{i=1}^n (b_i^r)^{\alpha/r} = \sum_{i=1}^n b_i^\alpha$, if $\alpha/r > 1$ and $\sum_{i=1}^n b_i^\alpha = \sum_{i=1}^n (b_i^r)^{\alpha/r} < \sum_{i=1}^n (a_i^r)^{\alpha/r} = \sum_{i=1}^n a_i^\alpha$, if $\alpha/r < 1$. This, together with (6.1), implies that

$$\psi(a, x) < \psi(b, x) \tag{6.2}$$

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

$$\psi(a, x) > \psi(b, x) \tag{6.3}$$

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 (6.2) and (6.3), for $j = 1, ..., k$, the r.v. $\sum_{i=1}^n b_i Y_{ij}$ is strictly more peaked than $\sum_{i=1}^n a_i Y_{ij}$, that is, for all $x > 0$ and all $j = 1, ..., k$,

$$P\Big(\Big|\sum_{i=1}^n a_i Y_{ij}\Big| > x\Big) > P\Big(\Big|\sum_{i=1}^n b_i Y_{ij}\Big| > x\Big). \tag{6.4}$$

The r.v.'s Y_{ij}, $i = 1, ..., n$, $j = 1, ..., k$, are symmetric and unimodal by Theorem 2.7.6 in Zolotarev (1986, p. 134). Therefore, from Theorem 1.6 in Dharmadhikari and Joag-Dev (1988, p. 13) it follows that the r.v.'s $\sum_{i=1}^n a_i Y_{ij}$, $j = 1, ..., k$, and $\sum_{i=1}^n b_i Y_{ij}$, $j = 1, ..., k$, are symmetric and unimodal as well. From Lemma in Birnbaum (1948) and its proof it follows that if X_1, X_2 and Y_1, Y_2 are independent absolutely continuous symmetric unimodal r.v.'s such that, for $i = 1, 2$, X_i is more peaked than Y_i, and one of the two peakedness comparisons is strict, then $X_1 + X_2$ is strictly more peaked than $Y_1 + Y_2$. This, together with (6.4) and symmetry and unimodality of $\sum_{i=1}^n a_i Y_{ij}$ and $\sum_{i=1}^n b_i Y_{ij}$, $j = 1, ..., k$, imply, by induction on k (see also Theorem 1 in Birnbaum (1948) and Theorem 2.C.3 in Dharmadhikari and Joag-Dev (1988)), that $\psi(a, x) = 1/2P\big(\big|\sum_{j=1}^k \sum_{i=1}^n a_i Y_{ij}\big| > x\big) > 1/2P\big(\big|\sum_{j=1}^k \sum_{i=1}^n b_i Y_{ij}\big| > x\big) = \psi(b, x)$ 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 2.4 for the class $\underline{\mathcal{CS}}(r)$ holds. The part of Theorem 2.1 for the class $\overline{\mathcal{CS}}(r)$ might be proven in a completely similar way. The proof is complete.

Proof of Theorems 2.3 and 2.4. Theorem 2.3 for the case of stable i.i.d. r.v.'s $X_i \sim S_\alpha(\sigma, \beta, 0)$, $i = 1, ..., n$, and Theorem 2.4 for both the cases of stable distributions $S_\alpha(\sigma, \beta, 0)$ and distributions from the class $\underline{CS}(1)$ are immediate consequences of Theorems 2.1 and 2.2 with $r = 1$. Let us prove Theorem 2.3 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 (6.2) and Proposition 0.1 in the introduction it follows that, for $j = 0, 1, ..., k$, the r.v. $\sum_{i=1}^n a_i Y_{ij}$ is strictly more peaked than $\sum_{i=1}^n b_i Y_{ij}$. Furthermore, from Theorem 2.7.6 in Zolotarev (1986, p. 134) and Theorems 1.6 and 1.10 in Dharmadhikari and Joag-Dev (1988, pp. 13 and 20) by induction it follows that the r.v.'s $\sum_{i=1}^n a_i Y_{ij}$ and $\sum_{i=1}^n b_i Y_{ij}$, $j = 0, 1, ..., k$, are symmetric and unimodal. Similar to the proof of Theorems 2.1 and 2.2, by Lemma in Birnbaum (1948) and its proof and induction, this implies that $\sum_{i=1}^n a_i X_i = \gamma \sum_{i=1}^n a_i Y_{i0} + \sum_{j=1}^k \sum_{i=1}^n a_i Y_{ij}$ is strictly more peaked than $\sum_{i=1}^n b_i X_i = \gamma \sum_{i=1}^n b_i Y_{i0} + \sum_{j=0}^k \sum_{i=1}^n b_i Y_{ij}$. This completes the proof of Theorem 2.3.

Proof of Corollary 3.1 . The corollary follows from Theorem 2.3 and relations (0.1) since $\overline{X}_n$ is weakly consistent for μ under its assumptions.

Proofs of Corollaries 3.2-3.5. It is easy to observe (see Marshall and Olkin (1979, p. 7)) that

$$\Big(\sum_{i=1}^n a_i/n, ..., \sum_{i=1}^n a_i/n\Big) \prec (a_1, ..., a_n) \prec \Big(\sum_{i=1}^n a_i, 0, ..., 0\Big), \tag{6.5}$$

for all $a \in \mathbf{R}_+^n$. From these relations it follows 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 latter comparisons and Theorem 2.2 it follows that, under the assumptions of Corollary 3.5, 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 3.5. Sharpness of the bounds in Corollaries 3.4 and 3.5 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 3.2-3.4 might be proven in a similar way, with the use of Theorems 2.1, 2.3 and 2.4 instead of Theorem 2.2 (the strict versions of inequalities (0.2) in Corollary 3.2 are consequences of bounds in Corollary 3.5 with $r = 1$).

Proof of Theorem 4.1. Let $j \in \{0, 1\}$, and let, conditionally on $y + z$, N have a distribution $\pi_j(n; y + z)$. Then from (4.2) it follows, similar to the proof of Lemma 2 in Jovanovic and Rob (1987), that, for $x \geq 0$,

$$\partial\xi(x; \lambda)/\partial\lambda = \mu \sum_{n=j}^{\infty} \pi_j(n; \lambda)(F(\sqrt{x}; n+1) - F(\sqrt{x}; n)) \tag{6.6}$$

(with $F(\sqrt{x}; 0) = 0$ if $j = 0$). Theorem 2.3 and relations (0.1) in the introduction imply that, under the assumptions of the theorem,

$$F(\sqrt{x}; n+1) = 1 - P(|\bar{\epsilon}_{n+1}| > \sqrt{x}) > 1 - P(|\bar{\epsilon}_n| > \sqrt{x}) = F(\sqrt{x}; n), \tag{6.7}$$

$x > 0$, $n = 1, 2, ...$ From (6.6) and (6.7) it follows that, under the assumptions of the theorem,

$$\partial\xi(x; \lambda)/\partial\lambda > 0 \tag{6.8}$$

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

$$\partial y/\partial p_{\hat{\theta}} = (1/C'')[1 + G''K''/(C''K'' - G''(C'' + K''))] > 0, \tag{6.9}$$

if (4.3) and (4.4) hold, and

$$\partial y/\partial p_{\hat{\theta}} = 1/(C'' - G'') > 0, \tag{6.10}$$

if (4.7) and (4.8) hold, that is, y is increasing in $p_{\hat{\theta}}$. Conclusion (c) of the theorem now follows from (6.8)-(6.10) and the property that, by (4.1), $p_{\hat{\theta}}$ is decreasing in $\rho(\hat{\theta}, \theta)$:

$$\partial p_{\hat{\theta}}/\partial \rho < 0. \tag{6.11}$$

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), the latter implies, by (6.9), that $p_{\hat{\theta}}^{(2)} > p_{\hat{\theta}}^{(1)}$ under (4.3) and (4.4). Since $y + z$ is increasing in $p_{\hat{\theta}}$ under (4.3) and (4.4) by Proposition 3 in Jovanovic and Rob (1987), we get, therefore, that, under the assumptions of the theorem, $\lambda^{(2)} > \lambda^{(1)}$ and thus

$$\xi^{(2)}(x) > \xi^{(1)}(x) \tag{6.12}$$

for all $x > 0$ by (6.8). As in the proof of Proposition 6 in Jovanovic and Rob (1987), we have

$$P(\rho^{(1)} > \rho^{(2)} | y^{(1)}, y^{(2)}) = \int \xi^{(2)}(x) d\xi^{(1)}(x) = \int \xi^{(1)}(x) d\xi^{(1)}(x) + \int (\xi^{(2)}(x) - \xi^{(1)}(x)) d\xi^{(1)}(x). \tag{6.13}$$

Since $\int \xi^{(1)}(x) d\xi^{(1)}(x) = 1/2$ using integration by parts, from (6.12) and (6.13) we get

$$P(\rho^{(1)} > \rho^{(2)} | y^{(1)}, y^{(2)}) > 1/2. \tag{6.14}$$

Relations (6.9), (6.11) and (6.14) 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 (6.9)-(6.11) and (6.13) since, by (6.8), holding $y^{(1)}$ constant and increasing $y^{(2)}$ or holding $y^{(2)}$ constant and decreasing $y^{(1)}$ increases $\xi^{(2)}(x) - \xi^{(1)}(x)$ for all $x > 0$. The proof is complete.

Proof of Lemma 4.2. We have that, under the assumptions of the lemma, $n^{-1/\alpha} \sum_{i=1}^{n} \epsilon_i \sim S_\alpha(\sigma, 0, 0)$. Furthermore, by Theorem 2.7.6 in Zolotarev (1986, p. 134), the distribution of the r.v.'s ϵ_i are unimodal. Therefore, the function $P(\epsilon_1 \leq x)$ is concave in $x > 0$. This, together with strict concavity of the function $x^{1-1/\alpha}$, $\alpha > 1$, in $x > 0$, implies that, 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 4.2 and 4.3. The proof is similar to the proof of Proposition 6 in Jovanovic and Rob (1987) and the proof of Theorem 4.1, with the use of Theorem 2.4 instead of Theorem 2.3 in this paper and Proposition

0.1 in Jovanovic and Rob (1987). Under the assumptions of Theorem 4.2, one has, by Theorem 2.4 and relations (0.1), that, similar to relation (6.7),

$$F(\sqrt{x}; n+1) = 1 - P(|\bar{\epsilon}_{n+1}| > \sqrt{x}) < 1 - P(|\bar{\epsilon}_n| > \sqrt{x}) = F(\sqrt{x}; n), \tag{6.15}$$

$x > 0$, $n = 1, 2, ...$ Relations (6.6) and (6.15) imply that, under the assumptions of Theorem 4.2,

$$\partial\xi(x; \lambda)/\partial\lambda < 0 \tag{6.16}$$

$x > 0$, 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 (6.16) it follows that $G'(\lambda) \leq 0$. This implies that conditions (4.5) 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 4.3 holds.

Relations (6.9)-(6.11) and (6.16) imply conclusion (c') of Theorem 4.2.

Let, as in the proof of Theorem 4.1, $\lambda^{(i)} = y^{(i)} + z^{(i)}$, $\rho^{(i)} = \rho(\hat{\theta}^{(i)}, \theta)$ and $\xi^{(i)}(x) = \xi(x; \lambda^{(i)})$, $i = 1, 2$, and let $y^{(2)} > y^{(1)}$. By (6.9) and Proposition 3 in Jovanovic and Rob (1987) we have $p_{\hat{\theta}}^{(2)} > p_{\hat{\theta}}^{(1)}$ under (4.3) and (4.4) and, therefore, $\lambda^{(2)} > \lambda^{(1)}$ under the assumptions of Theorem 4.2. This and (6.16) imply that

$$\xi^{(2)}(x) < \xi^{(1)}(x), \tag{6.17}$$

for all $x > 0$. From (6.13) and (6.17) it follows, similar to the proof Proposition 6 in Jovanovic and Rob (1987) and to the proof of Theorem 4.1 in the present paper, that

$$P(\rho^{(1)} > \rho^{(2)} | y^{(1)}, y^{(2)}) = 1/2 + \int (\xi^{(2)}(x) - \xi^{(1)}(x)) d\xi^{(1)}(x) < 1/2. \tag{6.18}$$

Relations (6.9)-(6.11) and (6.18) imply conclusion (a') of Theorem 4.2.

Conclusion (b') of Theorem 4.2 follows from (6.9)-(6.11) and (6.13) and the fact that, by (6.16), 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.

Proof of Theorem 5.1. 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

$$P(\pi_B \leq x) = P(v_{(n-1)}(g_r, B) \leq x) = n(H_k(x^{1/r}))^{n-1} - (n-1)(H_k(x^{1/r}))^n \tag{6.19}$$

(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

$$P(k\pi_1 \leq x) = P(v_{(n-1)}(g_r, \{1\}) \leq x/k) = n(H_1(x^{1/r}/k^{1/r}))^{n-1} - (n-1)(H_1(x^{1/r}/k^{1/r}))^n. \tag{6.20}$$

By Theorem 2.1 and comparisons (0.1), $H_k(xk^{1/r}) > H_1(x)$, $x > 0$, 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 (6.19) and (6.20) implies that $P(\pi_B \le x) > P(k\pi_1 \le x)$ for all $x > 0$, and, therefore (see Shaked and Shanthikumar (1994, pp. 3-4) and Remark 2), $E(\pi_B) < E(k\pi_1) = \sum_{i\in B} E(\pi_i)$. 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 $k_t \ge 2$ for at least one $t \in \{1, ..., l\}$,

$$E(\Pi_{\mathcal{B}}) = \sum_{s=1}^{l} E(\pi_{B_s}) < \sum_{s=1}^{l}\sum_{i\in B_s} E(\pi_i) = \sum_{i=1}^{m} E(\pi_i) = E(\Pi_{\underline{\mathcal{B}}}). \tag{6.21}$$

The proof is complete.

Proof of Theorem 5.2. 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 $card(B) = k \le m - 1$. With the same notations as in the proof of Theorem 5.1, comparisons (0.1) and Theorem 2.2 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 5.1, we get, therefore, that $P(\pi_B \le x) > P((k/m)\Pi_{\overline{\mathcal{B}}} \le x)$ 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 $EU(\pi_B) < EU((k/m)\Pi_{\overline{\mathcal{B}}}) \le (k/m)EU(\Pi_{\overline{\mathcal{B}}})$. Consequently, for any bundling decision $\mathcal{B} = \{B_1, ..., B_l\}$ such that $card(B_s) = k_s$, $s = 1, ..., l$, and $k_t \le m - 1$ for at least one $t \in \{1, ..., l\}$,

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

The proof is complete.

Proof of Theorem 5.3. Let $j \in J$. Let the vector $\tilde{X}^{(j)}$ of the jth buyer's reservation prices for goods in M take a value $\tilde{x}^{(j)} = (\tilde{x}_{1j}, ..., \tilde{x}_{nj}) \in \mathbf{R}^n_+$, $(\tilde{x}_{1j}, ..., \tilde{x}_{nj}) \ne (0, 0, ..., 0)$. Consider any bundle $B \in 2^M$ with $card(B) = k \ge 2$. 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 5.1, we get, similar to Palfrey (1983), that the expected surplus to the buyer when B is offered for sale is

$$ES_j(B, \tilde{x}^{(j)}) = \int_0^{v_j(B)} (H_k(x))^{n-1}dx = k\int_0^{v_j(B)/k} (H_k(kx))^{n-1}dx. \tag{6.22}$$

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 2.4 and (0.1), $H_k(kx) < H_1(x)$ for all $x > 0$. This, together with (6.22), implies

$$ES_j(B, \tilde{x}^{(j)}) < k\int_0^{v_j(B)/k} (H_1(x))^{n-1}dx \tag{6.23}$$

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 (6.5) it follows that $F(y_1, ..., y_k) \ge F(\sum_{i=1}^{k} y_i/k, ..., \sum_{i=1}^{k} y_i/k)$ for all $(y_1, ..., y_k) \in \mathbf{R}^k_+$ (see also the proof of Theorem 5 in Palfrey (1983)). In particular,

$$k\int_0^{v_j(B)/k} (H_1(x))^{n-1}dx \le \sum_{i\in B}\int_0^{\tilde{x}_{ij}} (H_1(x))^{n-1}dx = \sum_{i\in B} ES_j(\{i\}, \tilde{x}^{(j)}). \tag{6.24}$$

From (6.23) and (6.24) we get

$$ES_j(B, \tilde{x}^{(j)}) < \sum_{i\in B} ES_j(\{i\}, \tilde{x}^{(j)}) \tag{6.25}$$

if $v_j(B) > 0$ (clearly, (6.25) holds as equality if $v_j(B) = 0$). By (6.25), 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 5.4-5.7. 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 5.1, 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 5, 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 2.4 and relations (0.1) 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 (6.21) hold.

Suppose now that $p_B > \mu$ for all $B \in 2^M$. Then using again Theorem 2.4 and relations (0.1) 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 $t \in \{1, ..., l\}$,

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

$$J \sum_{s=1}^{l} k_s \big(p_{B_s} - (1/m) \sum_{i=1}^{m} c_i\big) P(\sum_{i=1}^{m} X_i \geq (mp_B)^{1/r}) \leq \sum_{s=1}^{l} (k_s/m) E(\Pi_{\overline{\mathcal{B}}}) = E(\Pi_{\overline{\mathcal{B}}}). \tag{6.26}$$

From (6.21) and (6.26) 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 5.5. Theorem 5.7 follows from Theorem 5.5 with $r = 1$. Theorems 5.4 and 5.6 could be proven in a similar way, with the use of Theorems 2.1 and 2.3 instead of Theorem 2.2. The proof is complete.

REFERENCES

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

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

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

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.

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.

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

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.

Cornelli, F. (1996): "Optimal selling procedures with fixed costs," *Journal of Economic Theory* **71**, 1-30.

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

de la Peña, V. H., Ibragimov, R. and Jordan, S. (2003): "Option bounds," To appear in *Journal of Applied Probability.*

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

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

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

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

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

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*, ed. M.A.H. Dempster, Cambridge University Press, Cambridge, 176-223.

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.

Heckman, J. J. and Honoré, B. E. (1990): "The empirical content of the Roy model," *Econometrica* **58**, 1121-1149.

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

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.

Jovanovic, B. and MacDonald, G. M. (1994): "Competitive diffusion," *Journal of Political Economy* **102**, 24-52.

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.

Levy, H. (1992): "Stochastic dominance and expected utility: survey and analysis," *Management Science* **38**, 555-593.

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.

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.

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.

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

Ruszczyński, A. and Vanderbey, R. J. (2003): "Frontiers of stochastically nondominated portfolios," *Econometrica* **71**, 1287-1297.

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.

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.

Department of Economics, Yale University, 28 Hillhouse Ave., New Haven, CT 06511, USA;
rustam.ibragimov@yale.edu