387(12): 2889-2908, 2008" which should be cited to refer to this work.

Market model with heterogeneous buyers

Matúš Medo*, Yi-Cheng Zhang

Physics Department, University of Fribourg, 1700 Fribourg, Switzerland

Abstract

In market modeling, one often treats buyers as a homogeneous group. In this paper we consider buyers with heterogeneous preferences and products available in many variants. Such a framework allows us to successfully model various market phenomena. In particular, we investigate how is the vendor's behavior influenced by the amount of available information and by the presence of correlations in the system.

Keywords: Market model; Supply–demand law; Correlations; Matching problem

1. Introduction

The standard economics textbooks make the supply–demand law as one of the pillars of the modern economic theory. However, many people, especially economists (see for example Ref. [1]), gradually realize that the most important factor is missing in the traditional supply–demand law. The study of complex systems [2–4] has already led to novel approaches to market phenomena. In a previous work [5], one of us introduced a simple framework to treat both quality and information capability, yielding a generalized supply–demand law. However, in the previous paper, a product is simply characterized by a single scalar variable: quality. In the modern economy we face a much more complex world, where the products have many attributes and consumers have heterogeneous tastes [6]. These preferences cannot be simply represented as price and quality alone. We therefore generalize the previous work to allow multiple variants of each product as well as many different tastes among consumers.

Thus the producers face a dilemma: whether to target the average taste by producing a single or a few variants to leverage the economy of scale, or to match precisely each consumer's taste [7]. We shall see that the answer depends on the information level that the producers may access: whether they know, and how well they know the consumers' preferences. In addition, producers face also the nonlinear production costs. All the factors have to compromise to yield a combined result that gives various degrees of product diversity. With our approach, the supply–demand problem of producers with the capability of producing variations and consumers' diverse tastes becomes a matching problem [8,9], where many mathematical and statistical mechanic tools are available to handle the complexity of the combinatorial problem.

* Corresponding author. Tel.: +41 26 300 9139; fax: +41 26 300 9758.
E-mail address: matus.medo@unifr.ch (M. Medo).

In this paper we build a market model and investigate its behavior under various circumstances. In the first part of the paper we do not consider correlations between preferences of the parties included in the system. While unrealistic, this assumption allows us to discover the basic properties of the model and outline the way of reasoning which can be used also in later, more realistic considerations. In the second part of the paper we discuss the correlations and the ways how they can be introduced to the system. The last part of the paper deals with the consequences of the correlations for the model.

2. General framework: One vendor with many buyers

Let us start with a market where only one vendor and M buyers are present. The vendor can produce N different variants of a product (e.g. many different shoes). With regard to the market, he has to decide which variants it is optimal to produce. We assume that all buyers satisfied with the offer by one item, others stay out of the trading. Buyers in the market we label with lowercase Latin letters ($i = 1, \dots, M$). The different variants the vendor can produce we label with Greek letters ($\alpha = 1, \dots, N$). The price of variant α we label as P_α. We assume that every variant can be produced in as many pieces as it is needed and as fast as it is needed.

The simple structure sketched above offers us enough space to model the basic features of real markets. To establish a mathematical model for the market we have to introduce some assumptions about participants' preferences and their consequences on the trading process. To keep complexity of the model at a minimum we assume that the buyer's opinion about a variant can be represented by one scalar quantity, which we call *cost* and label it with x; we assume that $x \in [0; 1]$. The smaller is the cost $x_{i,\alpha}$, the bigger is the probability that the buyer i is satisfied with variant α when asked. Preferences of the vendor are easier to introduce; they are represented by costs which he suffers during production and sale of a particular variant. The cost for variant α we label y_α and after a proper rescaling of monetary units $y_\alpha \in [0; 1]$. To simplify our considerations, we arrange the variants in the order of cost: $y_1 < y_2 < \cdots < y_N$. We stress a conceptual difference between vendor's and buyer's costs. The seller's cost y_α is strictly monetary—it represents a real amount of money (although in arbitrary units). In contrast, the buyer's cost $x_{i,\alpha}$ has no tangible interpretation, it simply represents something as airy as happiness with the given variant.

The vendor is able to produce N different variants. However, when he is producing more variants, his expenses grow due to the need of an additional investment. The vendor's tendency to produce only few different variants can be modeled e.g. by a nonlinear of expenses (doubling production of one single variant does not require double expenses). We adopt another approach; we assume that to initiate the production of a variant, the vendor has to pay an additional charge $Z > 0$ which is refered to as initial cost.

Now let us assume that the vendor offered k most favorable variants (thus $\alpha = 1, \dots, k$, $k \le N$) to customers and the number of units sold of variant α is n_α. The total vendor's profit is

$$X(k, \{n_\alpha\}) = \sum_{\alpha=1}^{k} n_\alpha (P_\alpha - y_\alpha) - kZ. \tag{1}$$

Here the last term kZ comes for the initial costs of k produced variants, $P_\alpha - y_\alpha$ is the profit for one sold unit of variant α. Due to the monetary rescaling used to confine y_α to the range [0; 1], units for profit, initial costs and prices are arbitrary.

It is natural to assume that when buyer i is asked about interest to buy variant α, the decision is based on the cost $x_{i,\alpha}$. We formalize this by the assumption that the probability of acceptance is a function of the variant cost; we call this function acceptance function. Obviously, $f(x)$ is a decreasing function of the cost x. Moreover, we assume that $f(0) = 1$. This means that if a buyer considers a variant to be the perfect one, she surely buys it.

When we offer a random variant to one buyer, the acceptance probability is

$$\int_0^1 \pi(x) f(x) \mathrm{d}x \equiv p. \tag{2}$$

Here $\pi(x)$ is the probability distribution of cost x (i.e. it defines what "to offer a random variant" really means). The probability p of accepting a random proposal is an important parameter of the model. From our everyday life we know that largely we do not agree to such an offer. For this reason we assume that $p \ll 1$ in our calculations.

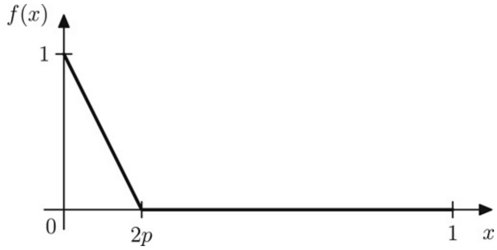

Fig. 1. One particularly simple choice for the buyers' acceptance function $f(x)$.

One example of a reasonable choice for the acceptance function is (see Fig. 1)

$$f(x;\,p) = \begin{cases} 1 - x/2p & (0 \le x \le 2p), \\ 0 & (2p < x) \end{cases} \tag{3}$$

with $p < 0.5$. This choice is especially convenient due to its simplicity. If we now assume that uniform distribution of the buyer's costs, $\pi(x) = 1$ for $0 \le x \le 1$, parameter p of the acceptance function (3) is just the probability p of accepting a random offer introduced in the previous paragraph.

In the rest of this paper we assume that the prices of all variants are equal to 1, $P_\alpha = 1$. This relieves us from many technicalities, and helps us to highlight the important features of the model. Nevertheless, generalization to various prices is straightforward.

3. No correlations in costs

We begin our investigation with the simplest case of the presented model—the market without correlations, where all costs y_α and $x_{i,\alpha}$ are mutually independent. We model this by costs uniformly distributed in the range [0; 1]. To keep the variants ordered, we first draw their costs and then we renumber all variants achieve $y_1 < y_2 < \cdots < y_N$. It follows that after averaging over realizations, the formula $\langle y_\alpha \rangle = \alpha/(N+1)$ holds.

3.1. A vendor without knowledge of buyers' preferences

If a vendor wants to discover which variants are most acceptable for buyers, in a market without correlations each buyer has to be asked for preferences. This cannot be done in big markets, thus it is natural to investigate the case with no information about buyers' preferences on the vendor's side. In Section 3.4 we show that without correlations even an expensive global opinion survey brings only a negligible contribution to the vendor's income.

Without any information about preferences, the vendor is not able to discover which variants are most favored by buyers. Therefore the best strategy is to offer the variants that are most favorable from his point of view. Let us label the number of variants the vendor is willing to offer as k. We assume that all these variants are available to buyers simultaneously, similarly to different types of shoes available in a shoe shop. Every buyer goes through the offered variants and decides whether some of them are suitable or not.

From the buyer's point of view, the vendor makes random proposals; the probability of accepting one particular offer is thus by definition equal to p. The probability P_A that one particular buyer accepts one of the k proposed variants is complementary to the probability $(1-p)^k$ of denying all offered variants. Thus we have

$$P_A = 1 - (1-p)^k \approx 1 - \mathrm{e}^{-pk}, \tag{4}$$

where the approximation used is valid for $pk \ll 1$, i.e. for very choosy consumers (then p is a small quantity) and a small number of offered variants. Now the average number of items sold by the vendor to all M buyers is MP_A. Since no correlations are present, the average number of items sold of variant α is $\langle n_\alpha \rangle = MP_A/k$, it is a decreasing function of k.

The quantity of vendor's interest is the total profit X introduced in (1). Its expected value can be found using $\langle n_\alpha \rangle$, $\langle y_\alpha \rangle$, and P_A. We obtain

$$X_U(k) = M\left(1 - \mathrm{e}^{-pk}\right)\left(1 - \frac{1+k}{2(N+1)}\right) - kZ. \tag{5}$$

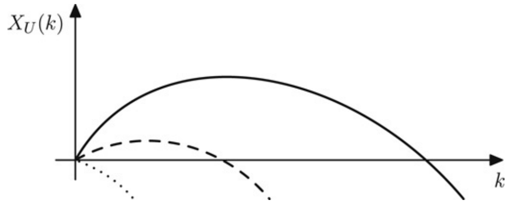

Fig. 2. Expected profit of the uninformed vendor, $X_U(k)$, drawn against k for small initial costs (solid line), medium initial costs (dashed line) and high initial costs (dotted line). In the last case, the condition $Z > Mp$ is fulfilled and the optimal vendor's strategy is to stop the production.

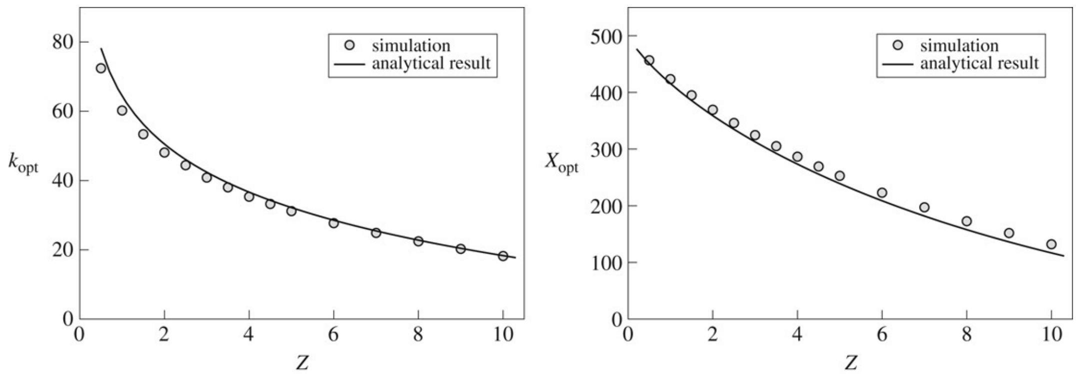

Fig. 3. The optimal number of offered variants and the optimal profit as functions of the initial cost Z for $M = 500$, $N = 2\,000$, and $p = 0.05$. Numerical results (empty circles) are averages over 1 000 realizations, analytical results (solid lines) come from Eq. (6). For the optimal profit arbitrary units are used.

Here the subscript U reminds that we are dealing with an "uninformed" vendor. This function is sketched in Fig. 2 for three different choices of the initial cost Z. The optimal number of variants the vendor should offer maximizes his profit. One can easily show that when $X_U'(0) < 0$, $X_U(k) < 0$ for all $k > 0$. Thus the condition $X_U'(0) < 0$, which can be rewritten as $Z > Mp$, characterizes a market where the optimal vendor's strategy is to stop the production and stay idle.

Since for every product numerous variations can be made, the total number of variants the vendor can offer, N, is large. Thus we are allowed to assume that the optimal number of offered variants satisfies the condition $k_{\rm opt} \ll N$ and solve the maximization condition $X_U'(k) = 0$ approximately. We obtain

$$k_{\rm opt} = \frac{1}{p} \ln \frac{Mp}{Z}, \qquad X_{\rm opt} = M - \frac{Z}{p}\left(1 + \ln \frac{Mp}{Z}\right), \tag{6}$$

where $X_{\rm opt}$ is the optimal expected profit, $X_{\rm opt} = X_U(k_{\rm opt})$. The used approximations are valid when $Z \gg M/N$ and $p \ll 1$. In Fig. 3, these results are shown to match a numerical treatment of the problem. In the figure we see how the initial cost Z influences diversity of the vendor's production: decreasing Z increases differentiation of the vendor's supply in full agreement with expectations. We can examine this feature in detail if we plot the optimal number of offered variants against Z for one particular realization of the model as it is shown in Fig. 4 (thickness of the lines is proportional to the number of buyers of a variant).

3.2. Improvement of the vendor's profit by a sequential offering of variants

So far we dealt with a very passive approach of the vendor. While offering k variants to the market, he had no influence on the sale. In consequence, due to the absence of correlations in the system, every offered variant had the same average number of items sold. In a big market this is a natural approach. (While the use of advertising can promote some variants, its treatment exceeds our scope.)

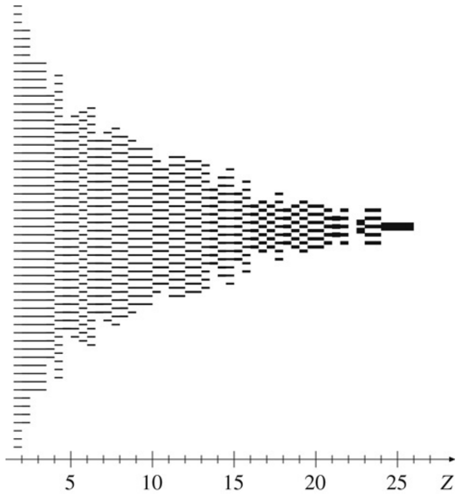

Fig. 4. Differentiation of the vendor's production for various initial costs Z: single realization of the model (no averaging present) with $M = 500$, $N = 2000$, $p = 0.05$. Vertical axis has no significant meaning, it serves purely to distinguish different variants.

In a small market a personal offering is possible. The vendor can promote favorable variants to increase the profit simply by offering the most favorable variant first. If a buyer is not interested, the second most favorable variant follows, etc. The average sale of the first variant is $\langle n'_1 \rangle = Mp$, for the second variant it is $\langle n'_2 \rangle = M(1-p)p$ and in general we have $\langle n'_\alpha \rangle = Mp(1-p)^{\alpha-1}$. Hence the expected total sale is

$$\sum_{i=1}^{k} Mp(1-p)^{i-1} = M\left[1-(1-p)^k\right].$$

This is equal to the expected total sale MP_A of the uninformed vendor in the previous section. We can conclude that the vendor's profit improvement (if any) does not come from an increased total sale but rather from an increased sale of the variants that are more profitable for the vendor.

Now we investigate the optimal number of variants to offer in this case, k'_{opt}. Since $\langle n'_\alpha \rangle$ decreases with α, at some moment it is not profitable to offer one more variant and the vendor's profit is maximized. The corresponding equation $\langle n'_\alpha \rangle = \langle n'_\alpha \rangle \langle y_\alpha \rangle + Z$ can be solved with respect to α, leading to k'_{opt}. When the total number of possible variants N is big, $\langle y_\alpha \rangle \ll 1$ and the term $\langle n'_\alpha \rangle \langle y_\alpha \rangle$ can be neglected. The approximate solution is then

$$k'_{\text{opt}} \approx \frac{\ln(Z/Mp)}{\ln(1-p)}. \tag{7}$$

This optimal number of variants to offer is smaller than k_{opt} given by Eq. (6). We can also notice that when p is small, using the approximation $\ln(1-p) \approx -p$ we are left with $k'_{\text{opt}} \approx k_{\text{opt}}$. This is an intriguing property—by the two different approaches we obtained the same result. To compare k'_{opt} in markets with different sizes, we plot it as a function of Z/M in Fig. 5. As can be seen, in a big market ($M \gtrsim 100\,000$) Eq. (7) fits well a numerical simulation of the system.

One can examine also the increase of the vendor's profit caused by the change of the sale method. Using previous results, the approximate formula $\Delta X_{\text{opt}} \approx Z[1 + \ln(Mp/Z)]/(Np^2)$ can be obtained. We see that when the total number of variants N is big, sequential offering results in a small growth of the vendor's profit. Nevertheless, in a system with a limited offer (small N) or with very choosy buyers (very small p), the improvement can be substantial.

We should notice that the stopping condition "income greater than expenses" introduced above can be hard to use in practice. It is because n'_α is a random quantity and can drop to the disadvantageous region $n'_\alpha < Z + n'_\alpha y_\alpha$ even when $\langle n'_\alpha \rangle$ is big enough to cover the expenses. Thus for the vendor it is not enough to simply check profitability of the sale of one particular variant n'_α. Rather he has to take into account sales of all previously offered variants. This is especially important in systems with a small number of buyers M where relative fluctuations are bigger. This effect

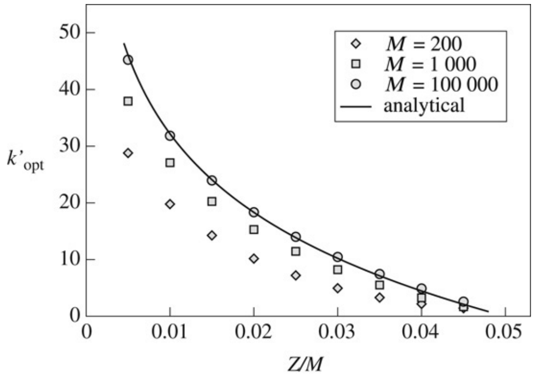

Fig. 5. Successive offering: numerical and analytical results for the vendor using stopping condition described in the text in the markets with various sizes (on horizontal axis we have $q \equiv Z/M$). All numerical results are obtained as average of 10 000 realizations with $p = 0.05$, $N = 2000$; solid line represents Eq. (7).

is shown in Fig. 5 where numerical results for the vendor blindly using the stopping condition are shown for various market sizes. Clearly as M increases, numerical results approach the analytical result (7).

3.3. Competition of two vendors

In real markets we seldom find a monopolist vendor; competition and partition of the market is a natural phenomenon. To investigate the model behavior in such a case we introduce the second vendor to the market. We assume that the vendors differ by initial costs which are Z_1 and Z_2 respectively. Again we do not consider the influence of advertisements and reputation, albeit they are vital in a market competition.

The course of the solution is similar to the one leading to Eq. (5). We label the number of variants offered by vendor 1 as k_1, the number of variants offered by vendor 2 as k_2, and we assume that there is no overlap between offered variants. The aggregate sale of two buyers is MP'_A where

$$P'_A = 1 - (1-p)^{k_1+k_2} \approx 1 - \exp[-p(k_1+k_2)].$$

With our assumptions about the equal status of the vendors, every offered variant has the same average sale. Therefore both vendors gain the share proportional to the number of variants they offer. Thus vendor 1 takes $k_1/(k_1+k_2)$ of the total sale and vice versa. When $k_1/N, k_2/N \ll 1$, we can simplify the expected profits to the form

$$X_1(k_1,k_2) = M\left(1 - \mathrm{e}^{-p(k_1+k_2)}\right)\frac{k_1}{k_1+k_2} - k_1 Z_1, \tag{8a}$$

$$X_2(k_1,k_2) = M\left(1 - \mathrm{e}^{-p(k_1+k_2)}\right)\frac{k_2}{k_1+k_2} - k_2 Z_2. \tag{8b}$$

Both parties maximize their profits by adjusting k_1 and k_2. The corresponding system $\partial_{k_1} X_1(k_1,k_2) = 0$, $\partial_{k_2} X_2(k_1,k_2) = 0$ cannot be solved analytically but its numerical treatment is straightforward. The result is shown in Fig. 6 where we have fixed the initial cost Z_2 to investigate how k_{opt} and X_{opt} for both vendors vary with Z_1.

We see that at $Z_1 \approx 12$ vendor 1 stops the production for he cannot stand the competition of vendor 2. By putting $k_2 = 0$ in the equations $\partial_{k_1} X_1(k_1,k_2) = 0$ and $\partial_{k_2} X_2(k_1,k_2) = 0$ we obtain the expression for the value Z_1^* when this price-out occurs

$$Z_1^* = \frac{Mp}{\ln(Mp/Z_2)}\left(1 - \frac{Z_2}{Mp}\right). \tag{9}$$

It is in a good agreement with the values found by a numerical simulation of the model. Important feature of this result is that it depends on the initial price Z_2 of the competitive vendor—decreasing the production costs can expel others from the market.

One can notice that when vendor 1 tries to increase the profit by deliberately increasing k_1 (with the intention to increase the sale), the term $-k_1 Z_1$ prevents the success of this strategy. As a result, the vendors have to adapt to each

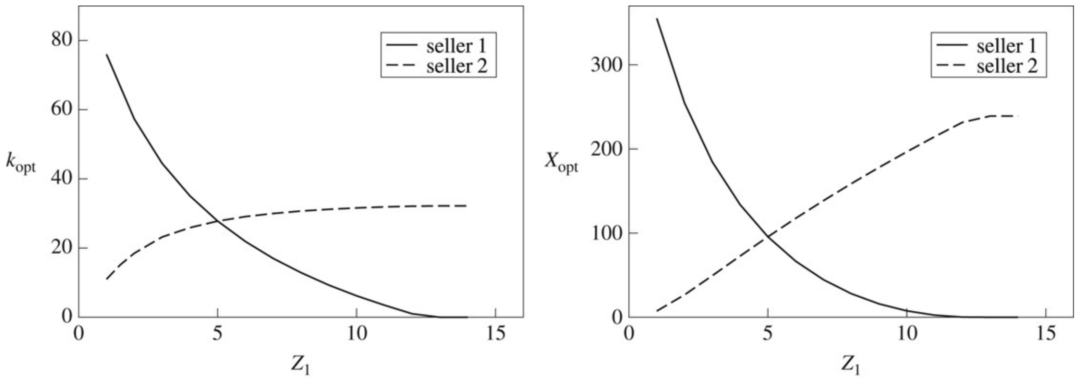

Fig. 6. The optimal number of variants to offer (left) and the optimal profit (right) for vendor 1 (solid line) and for vendor 2 (dashed line) against Z_1. The initial cost of the second vendor is $Z_2 = 5.0$, $M = 500$, and $p = 0.05$.

other. In mathematical terms, $X_1(k_{1\text{opt}}, k_{2\text{opt}}) \geq X_1(k_1, k_{2\text{opt}})$. At the same time, the sum of profits is not maximized at $k_{1\text{opt}}$ and $k_{2\text{opt}}$. It is more profitable to remove the less efficient producer (the one with the higher value of initial costs). This is an analogy of a real market where ruining (or taking over) of a competitor can improve the company profit.

3.4. *An informed vendor*

Now we would like to investigate the artificial case of the market where the vendor knows costs $x_{i,\alpha}$ of all buyers. This knowledge can be used to increase the optimal profit. We start with a simpler question: if the vendor offers only one variant, how much the sale can be increased by a good choice of the variant? The probability that buyer i is agreeable to buy variant α is $f(x_{i,\alpha})$. Since costs $x_{i,\alpha}$ are random and independent, only the average acceptance probability p plays a role and the number of users willing to buy this variant, n_α, is thus binomially distributed with the mean $\langle n_\alpha \rangle = Mp$ and the variance $\sigma^2 = Mp(1-p)$. When the number of buyers M is big, we can pass to a continuous approximation and assume the normal distribution of n_α

$$f(n_\alpha) \approx \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(n_\alpha - Mp)^2}{2\sigma^2}\right]. \tag{10}$$

The biggest value from the set $\{n_\alpha\}$ ($\alpha = 1, \dots, N$) we label as m. This is the number of potential buyers for the most accepted variant and the vendor does the best when by offering this variant. The probability density $f_N(m)$ (often called *extremal distribution*) is

$$f_N(m) = \frac{N}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(m - Mp)^2}{2\sigma^2}\right] \left(\frac{1}{2} + \frac{1}{2}\,\mathrm{Erf}\left[\frac{m - Mp}{\sigma\sqrt{2}}\right]\right)^{N-1}. \tag{11}$$

The multiplication by N appears because we do not care which one of all N variants is "the most accepted" one and the error function term represents the probability that the remaining $N-1$ variants are less accepted.

Since we are interested in big values of N, we expect that the difference $\langle m \rangle - M$ is big in comparison with σ. Therefore we use the approximation $\mathrm{Erf}(x) \approx 1 - \exp[-x^2]/\sqrt{\pi x^2}$, which is valid for $x \gg 1$. When the error function value is close to one, we can also use the approximation $(1-x)^N \approx \exp[-Nx]$ ($x \ll 1$) to obtain

$$f_N(m) \approx \frac{N}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(m - Mp)^2}{2\sigma^2} - \frac{\sigma N}{\sqrt{2\pi}} \frac{\exp\left[-(m - Mp)^2/2\sigma^2\right]}{m - Mp}\right].$$

This form is too complicated to obtain an analytical result for $\langle m \rangle$. Instead we compute the most probable value $\tilde{m}$

$$\tilde{m} \approx Mp + \sigma\sqrt{2\ln\frac{\sigma^3 N}{\sqrt{2\pi}}}.$$

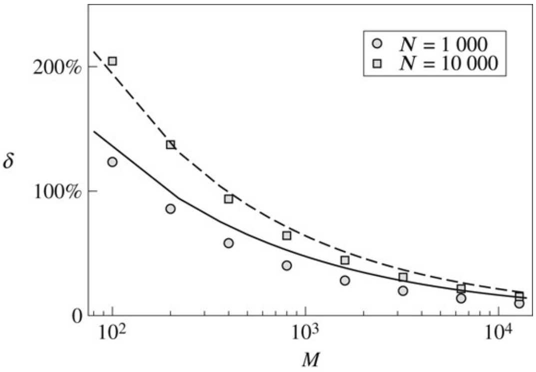

Fig. 7. The relative growth of the vendor's sale δ is drawn against the total number of buyers M. Solid lines represent the analytical result, outcomes from numerical simulations are shown as symbols.

Here the first term Mp represents the average value of the sale and the additional term represents the gain arising from the additional vendor's knowledge. To get a better notion about the sale growth we use the relative sale growth

$$\delta \equiv \frac{\tilde{m} - Mp}{Mp} \approx \sqrt{\frac{1}{Mp} \ln \frac{pMN^2}{2\pi}}. \tag{12}$$

To simplify the formula, the assumption $p \ll 1$ has been used. A comparison of this result with a numerical simulation of the model is shown in Fig. 7. As can be seen, a good agreement is obtained.

When $\delta \ll 1$, all vendor's information is indeed useless and the average sale improvement is negligible. The inequality $\delta \ll 1$ leads to the condition

$$N^2 \ll \frac{2\pi}{Mp} \mathrm{e}^{Mp}. \tag{13}$$

Thus when the number of variants is not large enough, buyers' opinions in the uncorrelated market cannot be used to increase the vendor's sale and profit.

From the previous results we can draw useful implications about the vendor with perfect information, offering more than only one variant. When the total number of variants N is big, the number of variants offered by the vendor is small in comparison with N. Therefore the average sale of all offered variants is increased at most by δ given by Eq. (12) and the same applies to the total sale. However, the vendor is interested mainly in his profit. When we take into account different costs y_α of variants, the resulting growth of the income due to the informations is even smaller than δ because the variant with the highest sale can have a high cost for the vendor. Thus condition (13) has more general consequences. It specifies the circumstances when even the perfect information about buyers' preferences do not help the vendor to achieve a significant improvement of his profit.

4. Correlations in the system

Now we would like to add one important flavor to the model—correlations. They arise from conformity of people's tastes (buyer–buyer correlations) and from the fact that high quality preferred by buyers results in high costs on the vendor's side (buyer–vendor anticorrelations). To approach the behavior of a real market, we investigate how these correlations influence our results obtained so far. Before doing so, we briefly discuss correlations from a general point of view.

4.1. Measures of correlations

Correlation is the degree to which two or more quantities are associated. We shall discuss different ways how to measure the correlations and how to introduce them to the system. In particular, we would like to measure the correlation between two lists (vectors) of costs: $\vec{x}_i$ and $\vec{x}_j$ (two buyers) or $\vec{x}_i$ and $\vec{y}$ (a buyer and the vendor). All lists of our interest have length N and contain real numbers between 0 and 1. A common choice for the correlation measure is Pearson's correlation coefficient r. For lists $\vec{x}$ and $\vec{y}$ it is defined as

$$r^2 = \frac{\left[\sum_{\alpha=1}^{N}(x_\alpha - \overline{x})(y_\alpha - \overline{y})\right]^2}{\sum_{\alpha=1}^{N}(x_\alpha - \overline{x})^2 \sum_{\alpha=1}^{N}(y_\alpha - \overline{y})^2}. \tag{14}$$

This measure is sensitive to nonlinear transformations of values in lists $\vec{x}$ and $\vec{y}$. In addition, since it originates in the least-square fitting of the data by a straight line, it measures only a linear correlation. For this reasons, in this work we use another correlation measure, Kendall's tau. For lists $\vec{x}$ and $\vec{y}$ it is given by the formula

$$\tau = \frac{2}{N(N-1)} \sum_{\alpha<\beta} \sigma_{\alpha\beta}, \quad \sigma_{\alpha\beta} = \operatorname{sgn}[(x_\alpha - x_\beta)(y_\alpha - y_\beta)] \tag{15}$$

and it ranges from $+1$ (exactly the same ordering of lists $\vec{x}$ and $\vec{y}$) to -1 (reverse ordering of lists); uncorrelated lists have $\tau = 0$. Notably, Kendall's tau is insensitive to all monotonic mappings of the data. This is the strongest property we can expect from a correlation measure—more general transformations, nonmonotonic mappings, can sweep out any structure present in the data.

4.2. *Lists with a given correlation degree*

Now we would like to construct a set of lists that have mutual values of Kendall's tau equal to τ_0. Such a set would represent lists of buyers' preferences in an equally dispersed society. Since the buyers' tastes are to a certain extent similar, we expect positive correlations with $\tau_0 > 0$. Nevertheless, in the following discussion we do not confine ourself to this region.

First we address a different question. Let us assume that between lists 1 and 2 there is τ_{12}, between lists 1 and 3 there is τ_{13}. Does it imply any constraints on τ_{23}? The answer is yes. It can be shown (see Appendix B) that τ_{23} fulfills the inequality

$$|\tau_{12} + \tau_{13}| - 1 \le \tau_{23} \le 1 - |\tau_{12} - \tau_{13}|, \tag{16}$$

which is an analogy of the triangular inequality for side lengths of a triangle. From (16) we can draw various simple conclusions. First, if we want to construct three lists which have pairwisely τ_0, it is possible only for $-1/3 \le \tau_0 \le 1$.[1] Thus it is impossible to have more than two lists which are perfectly anticorrelated. Another simple result is that when $\tau_{12} = -1$, inevitably $\tau_{23} = -\tau_{13}$.

Now the question is whether we are able to create the whole system of M lists which all have pairwisely Kendall's tau equal to τ_0. The answer depends on the magnitude of M. It can be shown (see Appendix B) that the upper bound for M is

$$M_m = \begin{cases} 2 + \log_2 \dfrac{(1-\tau_0)(N-1)N}{4} & (\tau_0 \ge 0), \\ \min\left[2 + \log_2 \dfrac{(1-\tau_0)(N-1)N}{4},\ 2\log_2 \dfrac{1-\tau_0}{-\tau_0}\right] & (\tau_0 < 0). \end{cases} \tag{17}$$

As can be seen in Fig. 8, this quantity grows slowly with the list length N. Therefore to model a market with a large number of equally correlated buyers we would need an enormous number of possible variants.

4.3. *Generation of correlated lists*

In the previous paragraphs we found that the society with a fixed mutual correlation degree of buyers is limited in its size. Therefore to introduce correlations to the presented market model we need a different approach. While *copulas* represent a general tool (see e.g. [10,11]), they are useful mainly for numerical simulations and offer only small possibilities for analytical results. Here we adopt a simpler way to generate correlated lists.

[1] An example for lists with the pairwise value $\tau_0 = -1/3$: $x_1 = \{3, 2, 1\}$, $x_2 = \{2, 1, 3\}$ and $x_3 = \{1, 3, 2\}$.

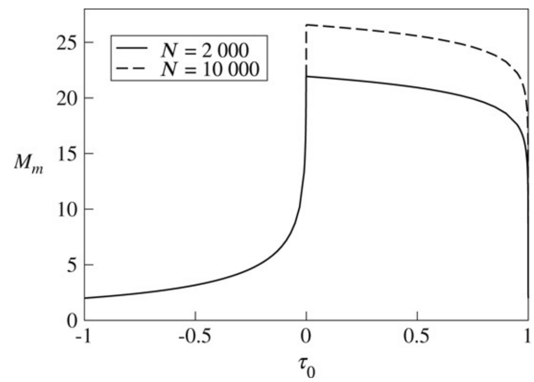

Fig. 8. The upper bound M_m as a function of τ_0 for two different lengths of lists N. In both cases the upper bound M_m is the same over a large part of the region $\tau_0 < 0$ and drops to 1 when $\tau_0 \to 1$.

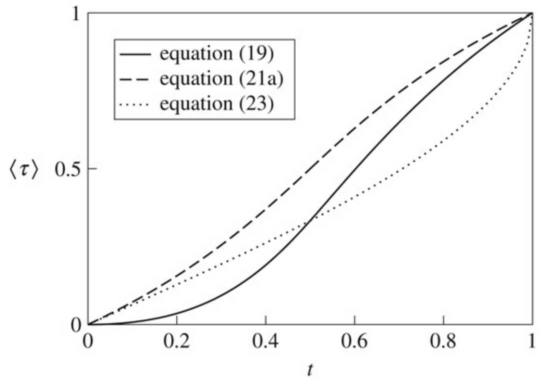

Fig. 9. The dependence of $\langle\tau\rangle$ on t for the proposed constructions of correlated lists.

Let us consider the lists of variables

$$x_{i,\alpha} = (1-t)\,a_{i,\alpha} + t\,c_\alpha, \tag{18a}$$

$$y_\alpha = (1-t)\,b_\alpha + st\,c_\alpha + \frac{t}{2}(1-s), \tag{18b}$$

where $a_{i,\alpha}$, b_α and c_α are independent random variables uniformly distributed in the range [0; 1]. Here $s = \pm 1$ indicates correlation/anticorrelation between $\vec{x}_i$ and $\vec{y}$ and $t \in [0; 1]$ is the binding parameter controlling strength of the correlation: $t = 0$ leads to uncorrelated lists, $t = 1$ to perfectly correlated ($s = 1$) or anticorrelated ($s = -1$) lists. In all cases, values $x_{i,\alpha}$, y_α lie in the range [0; 1].

For the lists defined above, it can be shown that (see Appendix C)

$$\langle\tau_{xy}\rangle = s\,\langle\tau_{xx}\rangle, \quad \langle\tau_{xx}\rangle = \begin{cases} \dfrac{u^2}{15}\left(10 - 6u + u^2\right) & (u \le 1), \\ \dfrac{1}{15}\left(15 - \dfrac{14}{u} + \dfrac{4}{u^2}\right) & (u > 1), \end{cases} \tag{19}$$

where $t/(1-t) \equiv u$. Plots of $\langle\tau_{xx}\rangle$ and $\langle\tau_{xy}\rangle$ are shown in Fig. 9. Since buyers' lists are prepared using the same formula, the average value of their correlation is nonnegative. Notably, for any value of τ we can find suitable s and t that produce lists with the expected correlation equal to τ.

Lists created using Eq. (18) do not have fixed mutual correlation, its actual value fluctuates around the mean value given by (19). According to the law of large numbers, $f(\tau)$ is normally distributed. In Appendix B it is shown that the variance of τ is proportional to $1/N$. Such fluctuations are negligible for long lists. We can conclude that Eq. (18) present a way to create a system with the desired amount of correlation τ for any τ, M and $N \gg 1$. Yet there is a hitch in the proposed construction of correlated lists. The parameter t influences the distribution of costs: for $t = 0$ or $t = 1$ they are distributed uniformly, for $t = 0.5$ the distribution $f(x)$ has a tent shape. This is an implausible property: the changes of the cost distributions can drive or distract the phenomena we are interested in. To fix this problem we propose the following two solutions.

First, to obtain correlated lists we can use the formulae

$$x_{i,\alpha} = \frac{1}{2} + st\left(\frac{\alpha-1}{N-1} - \frac{1}{2}\right) + (1-t)\left(s_{i,\alpha} - \frac{1}{2}\right), \quad y_\alpha = \frac{\alpha-1}{N-1}, \tag{20}$$

where $s_{\alpha,j}$ is a random quantity distributed uniformly in the range [0; 1]. The complicated form of $x_{i,\alpha}$ has a simple meaning. The vendor's costs grow uniformly with α and buyers' costs are connected to the vendor's by the parameter $t \in [0; 1]$. The term proportional to $1-t$ introduces a noise to the system, resulting in differences between buyers' and vendor's lists. Finally, the term $1/2$ represents the average value of buyers' costs. It is easy to check that $x_{i,\alpha}$ given by (20) is confined to the range [0; 1] for every $t \in [0; 1]$ and $s = \pm 1$. The overall distribution of costs is uniform in the range [0; 1] and thus we avoid the problems of Eq. (18). Moreover, this construction is simple enough to tract the proposed model analytically.

Using the techniques shown in Appendix B we can find Kendall's tau in this case. In the limit $N \to \infty$ one obtains

$$\langle \tau_{xy} \rangle = \begin{cases} \frac{s}{6}\left(4u - u^2\right) & (u \le 1), \\ \frac{s}{6}\left(6 - \frac{4}{u} + \frac{1}{u^2}\right) & (u > 1), \end{cases} \tag{21a}$$

$$\langle \tau_{xx} \rangle = \begin{cases} \frac{u^2}{15}\left(10 - 6u + u^2\right) & (u \le 1), \\ \frac{1}{15}\left(15 - \frac{14}{u} + \frac{4}{u^2}\right) & (u > 1), \end{cases} \tag{21b}$$

where again $u \equiv t/(1-t)$. The form of $\langle \tau_{xx} \rangle$ is identical with (19) found before for a different construction of correlated lists.

As we will see later, Eq. (20) is not appropriate to produce anticorrelated lists. Hence we present one more approach here—less accessible to analytical computation but more robust. The normal distribution is stable with respect to addition of random variables and this motivates us to make the following choice

$$x_{i,\alpha} = \sqrt{1-t}\, a_{i,\alpha} + \sqrt{t}\, c_\alpha, \tag{22a}$$

$$y_\alpha = \sqrt{1-t}\, b_\alpha + s\sqrt{t}\, c_\alpha \tag{22b}$$

where $a_{i,\alpha}, b_\alpha, c_\alpha$ are drawn from the standard normal distribution. It can be shown[2] that in this case

$$\langle \tau_{xx} \rangle = \frac{2}{\pi}\arcsin t, \qquad \langle \tau_{xy} \rangle = s\,\langle \tau_{xx} \rangle. \tag{23}$$

The course of $\langle \tau_{xx} \rangle$ is shown in Fig. 9. As our market model assumes costs confined to the range [0; 1], the costs given by (22) have to be transformed using the cumulative distribution function of the standard normal distribution $\Phi(x)$. In this way we obtain

$$\hat{x}_{i,\alpha} = \Phi^{-1}(x_{i,\alpha}), \qquad \hat{y}_\alpha = \Phi^{-1}(y_\alpha). \tag{24}$$

Since this transformation is monotonic, it does not affect the value of $\langle \tau \rangle$ and we can use Eq. (23) for transformed lists of costs.

[2] In the derivation, the following formula is useful

$$\int_0^\infty e^{-px^2}\,\mathrm{Erf}(ax)\,\mathrm{Erf}(bx)\,\mathrm{d}x = \frac{1}{\sqrt{\pi p}}\arctan\left[ab/\sqrt{p(a^2+b^2+p)}\right].$$

5. A market with correlations

When we discussed the market without correlations, the probability distribution of the variant cost $\pi_\alpha(x_{i,\alpha})$ was independent of α. Consequently, the probability of accepting variant α

$$P_A(\alpha) = \int_D \pi_\alpha(x_{i,\alpha}) f(x_{i,\alpha}) \mathrm{d}x_{i,\alpha} \tag{25}$$

was also independent of α (we labeled $P_A \equiv p$). As a result, when we change the acceptance function $f(x)$ while preserving the quantity $\int_0^1 \pi(x) f(x)\mathrm{d}x$, the derived results remain unchanged. In the presence of correlations we witness a very different picture: the detailed shape of the acceptance function $f(x)$ is important. To keep the algebra as simple as possible, from now on we adopt the simplest choice for $f(x)$: the step function $f(x) = 1 - \Theta(x - p)$. This means that a buyer accepts a proposed variant only when its cost is smaller than p.

In the following we first deal with the market where costs are given by Eq. (20) for it is more accessible to analytical treatment. Then we shortly present analytical results for the market where to introduce correlations, Eq. (22) is used.

5.1. An uninformed vendor in a market with correlations

Here we assume cost correlations created using Eq. (20). When the vendor has no information about the preferences of buyers, similarly to Section 3.1 the best strategy is to produce vendor's most favorable variants. First we focus on the case of positive correlations; in (20) we set $s = 1$ and $0 \le t \le 1$. Using (25) and the chosen step acceptance function $f(x)$, the probability that one buyer accepts variant α is

$$P_A(\alpha) = \begin{cases} 0 & 1 + (N-1)\,\dfrac{p}{t} < \alpha, \\ \dfrac{1}{1-t}\left[p - t\dfrac{\alpha - 1}{N-1}\right] & 1 + (N-1)\dfrac{p+t-1}{t} < \alpha < 1 + (N-1)\,\dfrac{p}{t}, \\ 1 & \alpha < 1 + (N-1)\dfrac{p+t-1}{t}. \end{cases} \tag{26}$$

Since we expect the total number of variants N to be very large and p rather small, the second region makes the major contribution and thus we simplify Eq. (26) to $P_A(\alpha) \approx (p - t\alpha/n)/(1-t)$.

We assume that the vendor is simultaneously offering his k most favorable variants. The probability that one buyer denies all offered variants is

$$P_D(k) = \prod_{\alpha=1}^{k} [1 - P_A(\alpha)] = \prod_{\alpha=1}^{k} \left(1 - \frac{p}{1-t}\right)\left(1 + \frac{t\alpha}{N(1-t+p)}\right). \tag{27}$$

Since N is big, we use the approximation $1 - x \approx \exp[-x]$ to evaluate this expression analytically, leading to

$$P_D(k) \approx \left(1 - \frac{p}{1-t}\right)^k \prod_{\alpha=1}^{k} \exp\left[-\frac{t\alpha}{N(1-t+p)}\right] \approx \exp\left[-\frac{pk}{1-t} + \frac{tk^2}{2N(1-t+p)}\right]. \tag{28}$$

Here we used also $1 - p/(1-t) \approx \exp[-p/(1-t)]$ which is valid when $p/(1-t)$ is small. When this is not the case, the denying probability $P_D(k)$ approaches zero and thus the accepting probability is virtually one regardless to the approximation used.

With respect to (20), the sum of the vendor's expected costs can be written as

$$kZ + \sum_{\alpha=1}^{k} M P_S(\alpha) \frac{\alpha - 1}{N-1} \approx kZ + \sum_{\alpha=1}^{k} M P_S(\alpha) \frac{\alpha}{N}. \tag{29}$$

Here the first term represents fixed costs for producing k different variants, $P_S(\alpha)$ is the probability that to one buyer variant α is sold. Since the probability that of the successful trade is $1 - P_D(k)$, from the condition $\sum_{\alpha=1}^{k} P_S(\alpha) =$

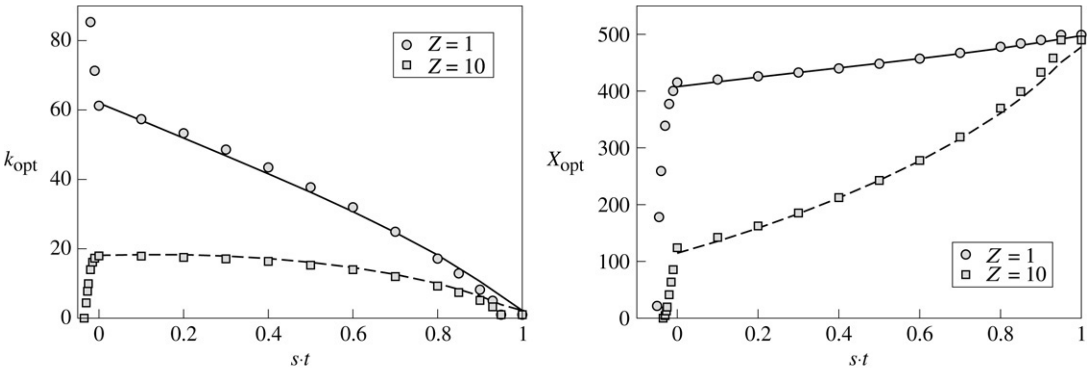

Fig. 10. The optimal number of variants to produce (left) and the optimal profit drawn against st for two different values of the initial cost Z. Lines show the analytical results derived above, symbols represent numerical simulations (averages of 1000 realizations), model parameters are set to $N = 2000$, $M = 500$, $p = 0.05$. The decay of both quantities for $st < 0$ is in agreement with Eq. (31).

$1 - P_D(k)$ we can deduce

$$P_S(\alpha) = \frac{P_A(\alpha)}{\sum_{\alpha=1}^{k} P_A(\alpha)} \left[1 - P_D(k)\right]. \tag{30}$$

This corresponds to the portioning of the probability $1 - P_D(k)$ among k variants according to their probability of acceptance.

Now we can use (28)–(30) to write down the expected profit of the vendor offering his k topmost variants $\overline{X}(k)$. It is not possible to carry out the maximization of this expression analytically—numerical techniques have to be used to find k_{opt} and X_{opt}. Results are shown in Fig. 10 as lines together with outcomes from a numerical simulation of the model; a good agreement is found for $st > 0$. Results confirm that positive correlations between buyers and the vendor increase the vendor's profit. This pattern is most obvious in the case $t = 1$ when the vendor can offer only the most favorable variant and still every buyer buys it.

In the numerical results shown in Fig. 10 we can notice one striking feature. When $st < 0$, k_{opt} changes rapidly and X_{opt} falls to zero quickly. Such a behavior is rather surprising for one does not expect abrupt changes in the region $st < 0$ when there were none in the opposite region $st > 0$. The reason for this behavior is simple—when $s = -1$, vendor's most preferred variants have costs too high to be accepted by buyers. This effect can be quantified. When buyers' costs are generated by (20), the inequality $x_{i,\alpha} \geq t(N-\alpha)/(N-1)$ holds. Due to the acceptance function only the variants with cost smaller than p are accepted. Therefore only variants with $\alpha \geq \alpha_{\min}$ can be possibly accepted, where

$$\alpha_{\min} = 1 + (N-1)\,\frac{t-p}{t} \approx N(1 - p/t). \tag{31}$$

Thus with negative correlations in the market, the vendor is able to sell the most favorable variant (the one with $\alpha = 1$) only if $p \geq t$. When $p < t$, the vendor sells no variants $\alpha = 1, \dots, \alpha_{\min} - 1$. Since $\alpha_{\min}$ grows steeply with t (already with $t = 2p$ one obtain $\alpha_{\min} = N/2$), the vendor offering his top k variants has to offer too many of them and he suffers both big initial costs and big costs y_α. As a result, the vendor is pushed out of the market.

Without detailed investigation we can infer the system behavior when the step acceptance function is replaced by a different choice. In the limit of careless buyers with $f(x) = C$, the influence of correlations vanishes and both k_{opt} and X_{opt} do not depend on st and the model simplifies to the case investigated in Section 3.1. Thus as $f(x)$ gradually changes from the step function to a constant function, the dependence on st gets weaker. In particular, if the largest cost x for which $f(x) > 0$ is x_0 (for the step function $x_0 = p$), in Eq. (31) p is replaced by x_0. As a consequence, $\alpha_{\min} =$ decreases and the steep decline of X_{opt} in Fig. 10 shifts to a lower value of st.

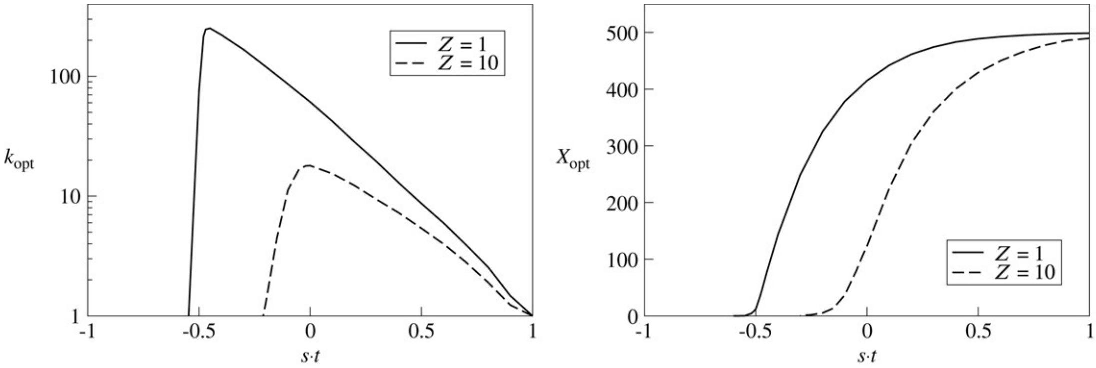

Fig. 11. The optimal number of variants and the optimal profit of the uninformed vendor in the market with correlations given by (22) and (24). Numerical results are averages of 1000 repetitions, $p = 0.05$, $N = 2000$, $M = 500$.

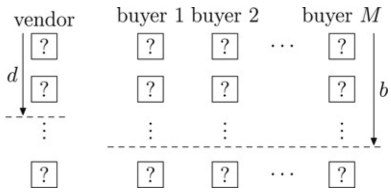

Fig. 12. The layout of the introduced trading model. Each column represents a sorted list of variants' costs (the most preferred at top). The vendor is willing to go down his list by d, in consequence buyers are forced to go down by some value b. Question marks signalize that after sorting of all lists we do not know standing of variants in the lists.

5.2. An uninformed vendor in a different market with correlations

Now we switch to the market costs drawn using Eq. (22) and transformed to the range [0; 1] by Eq. (24). As we have already mentioned, this case is not allowable for an analytical treatment—hence we present only numerical results in Fig. 11. They agree with our expectations: when $st = 1$, for the vendor it is sufficient to produce only one variant. As the positive correlations diminish, the optimal number of offered variants grows and the profit shrinks. A closer investigation of the vendor's behavior in this case exceeds the scope of this paper and remains as a future challenge.

6. Another trading model

In previous sections we presented a way how to deal with the trading process. Here we shortly present a different model which arises from the same playground as our previous reasonings but highlight slightly different aspects of the market phenomenon.

Let us have a market with M buyers and N different variants that the vendor can produce. Preferences of the interested parties are again represented by the scalar costs $x_{i,\alpha}$, y_α ($i = 1, \ldots, M$, $\alpha = 1, \ldots, N$) uniformly distributed in the range [0; 1] (thus again we have no correlations in the system).

In a market, a vendor is aware that when some buyer is not satisfied with the offer, she can choose a different vendor. Therefore every vendor tries to induce as small cost as possible to the customers. This can be done by offering of the variants highly preferred by many buyers. We can visualize the process by sorting preference lists of all interested parties. Now when some variant is near the top of a buyer's list, its cost is small and it is favorable for this buyer.

We have to specify the criterion for the "variant preferred by many buyers". First, it can be the variant that is not too deep in nobody's list. Thus, if we label the position of variant α in the list of buyer i as $k_{i,\alpha}$ and $\max_i k_{i,\alpha}$ as b_α, the vendor chooses the variant α that has the smallest b_α. The selection process is visualized in Fig. 12. Now the question is: how far buyers have to go down their lists? In other words: if we label $b \equiv \min_\alpha b_\alpha$, what is $\langle b \rangle$?

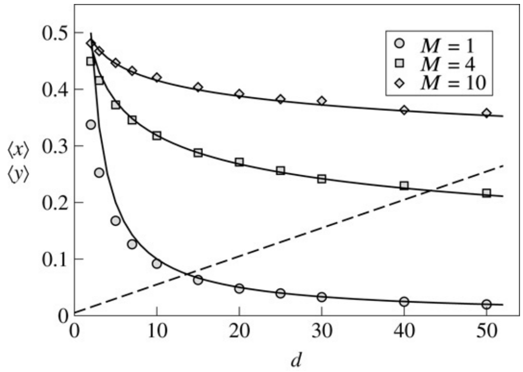

Fig. 13. Numerical and analytical results for $\langle x \rangle$ plotted against d for various values of M ($N = 1000$, numerical results are averages of 1000 repetitions). The analytical result for $\langle y \rangle$ is shown by the broken line.

Since we have M buyers in the market, the probability of a particular value b is approximately given by the formula

$$P(b) \approx \left[1 - \left(\frac{b-1}{N}\right)^M\right]^d \times \left[M \frac{d}{N} \left(\frac{b}{N}\right)^{M-1}\right]. \tag{32}$$

Here the first term denotes the probability that there is no such a variant which is among topmost $b-1$ for every buyer and among topmost d for the vendor.[3] The second term responds to the fact that there is some variant which is exactly on bth place in the list of a buyer (we do not care who it is, thus the multiplication by M appears), among b topmost variants in lists of other buyers and among d topmost variants in the vendor's list.

In Eq. (32) we can use the approximation $(1-x)^s \approx \exp[-xd]$ which is valid when $x \ll 1$. To calculate $\langle b \rangle = \sum_{b=1}^{N} bP(b)$ we replace the summation by the integration in the range $[0; \infty]$ which yields

$$\langle b \rangle \approx \frac{N\Gamma(1/M)}{M} d^{-1/M}. \tag{33}$$

Here we dropped terms vanishing in the limit $N \to \infty$ (for there is a big number of variants that the vendor can produce).

Since there are no correlations in the system, when the vendor offers his d topmost variants, every variant has the same probability to be chosen by a buyer. Thus the vendor has to go down his list on average by $(1+d)/2$. On average this corresponds to the cost

$$\langle y \rangle = \frac{1+d}{2N}. \tag{34}$$

With the probability $1/M$, a particular buyer has the sold variant on the bth place of his list. With the complementary probability $1 - 1/M$ he has this variant somewhere between the 1st and bth place. Thus we have

$$\langle x \rangle = \frac{1}{N}\left(\frac{1}{M}\langle b \rangle + \frac{M-1}{M}\frac{1+\langle b \rangle}{2}\right) \approx \frac{(M+1)\Gamma(1/M)}{2M^2} d^{-1/M}. \tag{35}$$

When the number of buyers M is large, this approaches $\frac{1}{2} d^{-1/M}$. A comparison of results (34) and (35) with numerical simulations is shown in Fig. 13; a good agreement is found. A small discrepancy for $M = 1$ can be corrected using the result $\langle y \rangle \approx 1/(d+1)$ which we develop in the following section.

To discover the scaling behavior of $\langle b \rangle$, one can follow a shorter path. The probability that one particular offered variant is among topmost b in one buyer's list is b/N. For all buyers simultaneously the probability is $(b/N)^M$. Since the vendor offers d items, the probability that at least one of them is above the line is approximately $d(b/N)^M$. If this

[3] Since in one list each variant appears only once, this is only an approximate form of the probability.

equal to $O(1)$, the trading is successful. Thus we obtain $\langle b\rangle = N\,O(1)\,d^{-1/M}$. This result scales with M and d in the same way as the previous outcome of the detailed derivation.

At the end we have to mention that this model of transactions between the vendor and buyers is not relevant for a high number of buyers because $\langle x\rangle$ decreases very slowly in this case. In other words: when M is high, the probability that there is a variant which is not worst in any buyer list approaches zero.

6.1. A vendor producing more than one variant

From the previous discussion we know that in a big market vendor cannot insist on selling only one variant. Therefore we would like to investigate a more relaxed case where the vendor offers simultaneously d different variants. Then every buyer can choose the most suitable one for him or her. We would like to investigate, how much the buyers suffer in this case. To do so, we label the most favorable variant from the vendor's offer for one particular buyer b and compute the average value of this quantity.

The probability that one particular value b occurs is

$$P(b) = \frac{d}{N-b+1}\prod_{j=1}^{b-1}\left(1-\frac{d}{N-j+1}\right). \tag{36}$$

In this formula the product represents the probability that all d variants offered by the vendor are in the buyer's list lower than $b-1$, the first term represents the probability that one of the offered variants is on bth place in the buyer's list. Now we can derive $\langle b\rangle$

$$\langle b\rangle = \sum_{b=0}^{N} b\,P(b) = \frac{N+1}{d+1} - d\binom{N}{d}^{-1} \approx \frac{N+1}{d+1}. \tag{37}$$

Consequently, the average cost suffered by a buyer is given by $\langle x\rangle = \langle b\rangle/N \approx 1/(d+1)$. We see that M does not appear in $\langle x\rangle$. This means that the problem with the improper behavior of the model in big markets do not appear in this variation. Since in the calculation we did not make any approximations, no numerical simulation is needed to check the result.

7. Conclusion

The aim of this paper is to explore the modeling of a market with heterogeneous buyers and a vendor producing multiple variants. We see that the outcomes depend on whether one or both sides have adequate information about the other side or not. In standard microeconomics, Pigou [12] has introduced the concept of price or demand elasticity. Vendors, knowing the buyers' reserve prices to pay and thus pricing individually, can reap significant profit—this is usually called the first-degree price differentiation. Our analysis can be considered as a generalization in this direction. We show that if individual tastes are taken into account, there is much complexity in the system; treating individual tastes with a large number of buyers presents a considerable mathematical challenge. Our models point out a convenient way to tackle this type of problems and we expect that many real economy-motivated problems can be analyzed in a similar way. Vendors and buyers have many ways to improve their welfare.

In this study we have proposed two simple market models. While accessible to analytical solutions, they exhibit many features of real markets. In particular, diversification of the vendor's production and market competition are used as examples. The diversification is presented as an interplay between the vendor's pursuit to follow the buyers' tastes and the costs growing with the number of produced variants. We also show that in a market with many buyers without preferences correlations, the knowledge of these preferences does not increase the vendor's profit. When the correlations are introduced to the system, many technical complications arise. Nevertheless, the results are consistent with the expectations: a positive correlation between the buyers' and vendor's costs improves the vendor's profit. Also, when the interests of the two parties diverge (the correlation are negative), the vendor is able to make only a small or even no profit. In addition, in Section 6 a similarly aimed model based on the well-known matching problem is investigated.

As many other directions can be explored further, we do not consider this topic exhausted. First of all, in a correlated market the vendor strategies and the influence of information deserve attention. Furthermore, while in the present work we investigated the influence of tastes on the market, the product quality and price were excluded from the analysis.

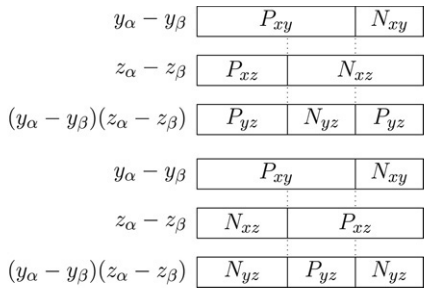

Fig. A.1. An illustration of the proof. The first case (first three lines) has the biggest possible value of P_{yz}, the second case has the smallest possible value of P_{yz}.

Eventually, the framework established herein can be used to raise the law of supply and demand from a microscopical point of view.

Acknowledgments

We acknowledge the partial support from the Swiss National Science Foundation (project 205120-113842) as well as STIPCO (European exchange program).

Appendix A. Proof of τ-inequality

Let us have three lists $\vec{x}, \vec{y}, \vec{z}$ consisting of N mutually different real numbers. Kendall's τ for lists $\vec{x}$ and $\vec{y}$ can be written as $\tau_{xy} = (P_{xy} - N_{xy})/T$ where P_{xy} is the number of pairs $\alpha < \beta$ that satisfy $(x_\alpha - x_\beta)(y_\alpha - y_\beta) > 0$, N_{xy} is the same with a negative result of the product, and $T = N(N-1)/2$ is the total number of different pairs α, β. For the given values τ_{xy} and N it follows that

$$P_{xy} = T(1 + \tau_{xy})/2, \qquad N_{xy} = T(1 - \tau_{xy})/2. \tag{A.1}$$

We would like to find bounds for τ_{yz} when τ_{xy} and τ_{xz} are given. First we reorder lists $\vec{x}, \vec{y}, \vec{z}$ so that lists $\vec{x}$ is sorted in the descending order and for $\alpha < \beta$ it is $x_\alpha - x_\beta > 0$. Such a rearrangement does not affect the values of $P_{xy}, P_{xz}, P_{yz}, N_{xy}, N_{xz}, N_{yz}$ and thus the values of Kendall's tau between lists remain also unchanged.

Since now all differences $x_\alpha - x_\beta$ are positive, from τ_{xy} we can deduce that there are P_{xy} positive differences $y_\alpha - y_\beta$ and N_{xy} negative differences. Similarly, P_{xz} differences $z_\alpha - z_\beta$ are positive and N_{xz} are negative. The values of P_{yz} and N_{yz} depend on the relative ordering of lists $\vec{y}$ and $\vec{z}$. The biggest possible value of P_{yz} occurs when positive differences $y_\alpha - y_\beta$ are aligned with positive differences $z_\alpha - z_\beta$ (see Fig. A.1). By contrast, the smallest value of P_{yz} (and thus the smallest value of τ_{yz}) occurs when positive differences $y_\alpha - y_\beta$ are aligned with negative differences $z_\alpha - z_\beta$.

From Fig. A.1 we see that P_{yz} and N_{yz} fulfill the inequalities

$$N_{yz} \geq |P_{xy} - N_{xz}|, \qquad P_{yz} \geq |P_{xy} - P_{xz}|.$$

Using $P_{ab} + N_{ab} = T$ and (A.1) we obtain

$$\frac{T}{2}|\tau_{xy} - \tau_{yz}| \leq P_{yz} \leq T - \frac{T}{2}|\tau_{xy} + \tau_{yz}|,$$
$$-T + \frac{T}{2}|\tau_{xy} - \tau_{yz}| \leq -M_{yz} \leq -\frac{T}{2}|\tau_{xy} + \tau_{yz}|.$$

These two inequalities summed together and divided by T yield the desired inequality (16).

Appendix B. System's upper bound with a given τ_0

To obtain the upper limit M_m for the number of lists that have pairwise Kendall's tau equal to τ_0 we use a constructive way of reasoning. Without loss of generality we assume that in the first list all $N(N-1)/2 \equiv T$

1	+ + + + + + + + + + + +
2	+ + + + + + + + − − − −
3	+ + + + + + − − + + − −
4	+ + + + + + − − − − + +
5	+ + + + + − + − + − − +
6	+ + + + + − + − − + + −

Fig. B.1. Construction of the set of lists with the given value of Kendall's tau for $T = 10$ (thus $N = 5$) and $W = 4$ (thus $\tau_0 = 0.2$).

differences $x_\alpha - x_\beta$ ($\alpha < \beta$) are positive. From Eq. (A.1) it follows that to maintain $\tau_{12} = \tau_0$, in the second list exactly $N_2 = (1-\tau_0)T/2$ pairs have to be negative. The same holds for all other lists which we would like to add to the system, $N_i = (1-\tau_0)T/2 \equiv W$ is required for $i \geq 2$ ($N_1 = 0$ is fixed by the chosen ordering of the first list).

To represent the way of construction we use Fig. B.1. The first list is represented by T positive pairs there, second one by $T - N_2$ positive pairs and by N_2 negative pairs. We would like to find a third list satisfying $\tau_{13} = \tau_{23} = \tau_0$. We already know that $N_3 = W$ negative pairs have to be accommodated there. To achieve $\tau_{23} = \tau_0$ it is necessary that exactly half of the negative pairs in the second list meets with negative pairs in the third list. This can be fulfilled in two ways (lines 3 and 4 in Fig. B.1) which also have mutually $\tau_{34} = \tau_0$.

Now there are two important points to notice. First, the construction of the third and the fourth list is impossible when N_3 is an odd number. Thus for corresponding values of τ_0 there are not more than two lists that have mutually Kendall's tau equal to this τ_0. Second, we have used purely combinatorial arguments here without taking care whether described set of lists (e.g. lists 1, 2, 3 and 4 from Fig. B.1) do exist. Thus we estimate an upper bound which cannot be exceeded but which can dwarf the real maximum by far.

Now we can continue with the fifth list where again $N_5 = W$ negative pairs are present. Among them exactly $N_5/2$ have to meet with negative pairs in the second list and also half of them have to meet with negative pairs in the third and fourth list. This can be achieved by lists 5 and 6 in Fig. B.1, their relative Kendall's tau is also τ_0.

During the construction process we divide W negative pairs present in the second lists into two groups with size $W/2$ (lists 3 and 4), then we divide again and obtain lists 5 and 6 with groups of negative pairs with size $W/4$. Clearly, this sequence ends when we divide the original number of negative pairs W so many times that we arrive at 1 which cannot be divided further. Consequently, the upper bound we are looking for is the following

$$M_m \approx 2 + 2\log_2 w = 2 + 2\log_2 \frac{(1-\tau_0)(N-1)N}{4}. \tag{B.1}$$

This formula holds when $\tau_0 < 1$ (for $\tau_0 = 1$, clearly $M_m = 1$).

In the previous construction, there is one hidden flaw. Since only $W/2$ negative pairs in the third list meet negative pairs in second list, to prepare lists 1, 2 and 3 we use altogether $W + W/2$ pairs. If this number is greater than T, the third list cannot be constructed in the described way and the limiting M_m differs from the previously found result. To construct lists 5 and 6 we need together $W + W/2 + W/4$ pairs which has to be smaller than T. By generalizing previous argument we can write down the inequality for the maximum number of lists M_m

$$W + W/2 + \cdots + W/2^{M_m/2-1} \leq T.$$

For M_m itself it follows that

$$M_m \leq 2\log_2 \frac{1-\tau_0}{-\tau_0}. \tag{B.2}$$

This limit is relevant only for $\tau_0 < 0$ (with $\tau_0 > 0$ we have $W < T/2$ and $W + W/2 + \cdots$ is less than T). As an actual upper bound, the smaller value from (B.1) and (B.2) applies.

Appendix C. Expected values of $\langle\tau\rangle$

For lists created using (18) we can rearrange (15) as follows

$$\langle\tau_{xx}\rangle = \frac{2}{N(N-1)}\sum_{\alpha<\beta}\langle\sigma_{\alpha\beta}\rangle = \langle\sigma_{\alpha\beta}\rangle.$$

Moreover, $\sigma_{\alpha\beta}$ can be rewritten as

$$\langle\sigma_{\alpha\beta}\rangle = P_{++} + P_{--} - P_{-+} - P_{+-} = 1 - 2P_{-+} - 2P_{+-} = 1 - 4P_{+-}.$$

Here P_{++} is the probability that both $x_\alpha - x_\beta$ and $x'_\alpha - x'_\beta$ are positive and so forth, the formulae $P_{+-} = P_{-+}$, $P_{++} + P_{--} + P_{-+} + P_{+-} = 1$ are used. According to (18) we write

$$\begin{aligned} x_\alpha - x_\beta &= (1-t)(a_\alpha - a_\beta) + t(c_\alpha - c_\beta) \equiv (1-t)A + tC, \\ x'_\alpha - x'_\beta &= (1-t)(b_\alpha - b_\beta) + t(c_\alpha - c_\beta) \equiv (1-t)B + tC, \end{aligned}$$

where A, B, C lie in the range $[-1; 1]$ and are equally distributed with the density $\varrho(A) = 1 - |A|$. Now we have $(t/(1-t) \equiv u)$

$$P(+-|C) = \begin{cases} \frac{1}{2}(uC)^2\left[1 - \frac{1}{2}(uC)^2\right] & (C \le 1/u), \\ 0 & (C > 1/u). \end{cases}$$

If $u \le 1$, the first case applies to all possible values of C, $P(+-|C) = 0$ is possible only if $u > 1$. Finally, using

$$P(+-) = \int_{-1}^{1} P(+-|C)\varrho(C)\,\mathrm{d}C$$

with $\varrho(C) = 1 - |C|$ it follows that

$$\langle\tau_{xx}\rangle = \begin{cases} \frac{u^2}{15}\left(10 - 6u + u^2\right) & (u \le 1), \\ \frac{1}{15}\left(15 - \frac{14}{u} + \frac{4}{u^2}\right) & (u > 1). \end{cases}$$

The quantity $\langle\tau_{xy}\rangle$ can be derived in the same way.

The variance of τ_{xx} can be found by a direct computation of $\langle\tau_{xx}^2\rangle$. We have

$$\tau_{xx}^2 = \frac{1}{N^2(N-1)^2}\left(\sum_{\alpha\neq\beta}\sigma_{\alpha\beta}^2 + \sum_{\alpha\neq\beta}\sum_{\gamma\neq\delta}\sigma_{\alpha\beta}\sigma_{\gamma\delta}\right).$$

The averaging procedure is straightforward. At the end we obtain

$$\sigma_\tau^2 = \langle\tau_{xx}^2\rangle - \langle\tau_{xx}\rangle^2 \approx \frac{4}{N}\left(\langle\sigma_{\alpha\gamma}\sigma_{\gamma\beta}\rangle - \langle\sigma_{\alpha\beta}\rangle^2\right),$$

where the terms proportional to higher powers of $1/N$ were neglected. The variance is largest when $t = 0$, for $t = \pm 1$ obviously $\sigma_\tau = 0$.

References

[1] J. Stiglitz, Information and the change in the paradigm in economics, in: T. Frangsmyr Les (Ed.), Prix Nobel: The Nobel Prizes 2001, The Nobel Foundation, 2002, pp. 472–540.

[2] W.B. Arthur, S.N. Durlauf, D.A. Lane (Eds.), The Economy as an Evolving Complex System II, in: Proceedings Vol. XXVII, Santa Fe Institute Studies in the Sciences of Complexity, Addison-Wesley, Reading, MA, 1997.

[3] D. Challet, M. Marsili, Y.-C. Zhang, Minority Games: Interacting Agents in Financial Markets, Oxford University Press, UK, 2004.

[4] A. De Martino, M. Marsili, Statistical mechanics of socio-economic systems with heterogeneous agents, Journal of Physics A 39 (2006) 465–540.

[5] Y.-C. Zhang, Supply and demand law under limited information, Physica A 350 (2005) 500–532.
[6] Y.-C. Zhang, Happier world with more information, Physica A 299 (2001) 104–120.
[7] J.Y. Bakos, Reducing buyer search costs: Implications for electronic marketplaces, Management Science 43 (1997) 1676–1692.
[8] M.-J. Oméro, M. Dzierzawa, M. Marsili, Y.-C. Zhang, Scaling behavior in the stable marriage problem, J. Phys. I France 7 (1997) 1723–1732.
[9] A. Lage-Castellanos, R. Mulet, The marriage problem: From the bar of appointments to the agency, Physica A 364 (2006) 389–402.
[10] F. Lindskog, Modelling dependence with copulas and applications to risk management, Master Thesis, ETH Zurich, available online from: http://www.risklab.ch/Papers.html, 2000.
[11] R.B. Nelsen, An Introduction to Copulas, Springer-Verlag, New York, 1999.
[12] A.C. Pigou, The Economics of Welfare, 4th ed., Transaction Publishers, London, UK, 2002.