Emergent Behavior and Criticality in Online Auctions

Inaugural-Dissertation
zur
Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Universität zu Köln

vorgelegt von

Alireza Namazi

aus Teheran

Köln 2005

Berichterstatter	Priv.-Doz. Dr. A. Schadschneider Prof. Dr. R. Schrader
Vorsitzender der Prüfungskomission	Prof. Dr. M. Abd-Elmeguid
Tag der mündlichen Prüfung	11.07.2005

Contents

1. Introduction

During the last century it has been realized that power-law distributions are surprisingly ubiquitous in natural and social sciences. Examples are the power-law distribution of individual wealth or people's annual incomes [1] and the frequency of use of words in different languages [2] known as Pareto and Zipf's laws, respectively. Similar power-law distributions were observed in physics when studying the behavior of systems at their critical points, related to second order phase transitions, in which fluctuations on all length scales are important. Those power laws are linked to the fractal structure and scale-free property of a system at its critical point. To understand the physics underlying the properties of such systems one has to take into account effects of all length scales down to the microscopic level [3].

Since the works of Wilson and introduction of the renormalization group theory [4–6] physicists have tools for such studies. A great amount of work has been done within the last three decades and a nice classification known as universality classes has been found, which gives physicists the ability of understanding a variety of different systems by identifying the universality class they belong to.

This paradigm of statistical physics has always been tempting for physicists encouraging them towards solving problems of nonphysical systems having a large number of interacting agents known as many body systems. Interactions in nonphysical systems are typically unknown or can not be precisely defined, making the theoretical investigations more difficult. Even by approximations and assumptions the nonlinearity of interactions presents serious problems. These systems have typically open boundary conditions, i.e. energy and information can constantly be imported and exported across the system boundaries driving them into a state far from equilibrium. The theory of nonequilibrium systems is up to now not as well developed as the theory of equilibrium systems, for which Gibbs ensembles and the Hamiltonian of the system can be used for the calculation of macroscopic averages.

Recent numerical techniques like Monte Carlo simulations, related to the achievements in computer and electronic sciences, have opened the possibility of working on analytically unsolvable problems. Examples are the Ising model in an external magnetic field in two dimensions and the Ising model in three dimensions.

Numerical methods are also applicable by studying problems arising in systems far from equilibrium and those belonging originally to other scientific disciplines like sociology, economics, traffic engineering, biology, etc. For instance the methods used to understand the physics of nuclei and elementary particles are applicable for analyzing stock market data. The new born interdisciplinary science is called econophysics [7, 8].

Efforts towards understanding systems showing this kind of properties lead to the development of methods, concepts and classifications now known as the science of complex systems. Complex systems are typically many body interacting systems with long-range

correlations between their agents.
The theory of complex systems was first made popular by the Santa Fe Institute[1]. Sherman and Shultz [9] give the following popular definition:

> Complexity refers to the condition of the universe which is integrated and yet too rich and varied for us to understand in simple common mechanistic or linear ways. We can understand many parts of the universe in these ways, but the larger and more intricately related phenomena can only be understood by principles and patterns not in detail. Complexity deals with the nature of emergence, innovation, learning and adaptation.

The self-similar and scale-invariant (both in space and time) behavior of complex systems compared with findings of statistical physics suggests that they should be at a critical point. But since no external tuning ever occurs one should think of a self-organization mechanism, i.e. those systems organize themselves towards their critical points[2].
Many physicists believe that this kind of process (named "self organized criticality" after P. Bak [10]) is responsible for many of the empirically found power-law distributions and 1/f noises.
Statistical analysis of data related to a complex system gives hints towards understanding interactions between agents and the structure of agents. The common way of understanding complex systems with a nonlinear interaction is to define models and make use of simulations. The predictions of the simulations and the comparison with the real system provide the possibility of 1) a better understanding of the system through its global rules, number of degrees of freedom, stochastic elements etc. and 2) tuning and calibrating the model for practical purposes.
The common approach of a physicist is to find the simplest model, i.e. a model with the minimum number of microscopic rules. The aim is to find a generic model with the ability of reproducing as much as possible results known from the real system.
This method was used by physicists to introduce a variety of different models describing problems like traffic flow [11], pedestrian dynamics, economics [8], stock markets [7], biological systems [12, 13], social sciences [14], etc. [15].
One question arising is how many rules and stochastic parameters a model needs to represent important features of the real system. For example, consider the Nagel-Schreckenberg (NaSch) model as one of the simplest microscopic traffic flow models [16, 17]. This cellular automaton model has two parameters (maximum velocity and the probability of braking) and four update rules (acceleration, safety distance control, randomization braking and driving).
The NaSch model reproduces some, but not all of the empirical observations. To make the model able to reproduce these other properties of the traffic flow one needs extra rules [18]. Similar models are now used for forecasting traffic on highways in North Rhine-Westphalia (NRW) in Germany [19]. The NaSch model does not use any assumptions and constraints

[1] http://www.santafe.edu/
[2] This is debated. Labor experiments require additional conditions, which can be considered as tuning.

for the higher order fluctuations, like fluctuations in the velocity or time headway distributions. This demonstrates the ability of minimal simple models to produce realistic results for systems having few number of decision freedom (where the separation between different rational strategies is not necessary).
One of the difficulties one encounters by studying more complicated systems, especially systems with humans as agents, is the lack of empirical data. This data are necessary for comparing the predictions of a model both qualitatively and quantitatively with the real system.
However a rather philosophical question is whether human based systems[3] could be treated as systems of unintelligent particles.
Approaches of game theory to understand such systems based on the assumption of "rational strategies" agents choose to optimize their payoff. The main method of the game theory is to search for the Nash equilibrium of the system, in which no agent could improve his payoff better by changing his strategy [20]. However there are many phenomena observed, where rational strategies do not provide a correct explanation [21]. The choice of strategy seems to be a combination of random and rational selections.
Online auction houses like eBay[4] can be considered as complex systems with a large number of agents interacting through auctions. The data related to auctions is huge and well formatted, making it suitable for statistical analysis. The agents participating in online auctions are humans. These two aspects of online auctions are the motivation for this study.
The auction house eBay has similarities with other social structures like cities. The network economies provided by eBay has caused an ever growing number of agents. The large number of bidders encourage more sellers. This will encourage more bidders, which in turn encourage more sellers, etc. making the whole system grow, which give the eBay company possibility of providing better services, more advertising, etc. This encourages again both bidders and sellers, making eBay a good candidate for studying dynamical many body interactive systems.
The aim of this study is not to measure the properties of the system quantitatively in order to find and calibrate a model capable to reproduce these properties. We are rather interested in the important microscopic rules, which can describe the empirical observations. There is a hope to find a quantitative description of roles of rationality of the agents in the system. It is interesting to examine whether the empirical findings could be separated in those, requiring rational agents and those, having their roots in totally stochastic fluctuations.

1.1. How this study is organized

Chapter 2 gives an overview of complex systems and complexity and discusses power-law generating processes and ideas related to complex systems. Since statistical relations and

[3]Like pedestrian dynamics, traffic flows, economy and social systems.
[4]eBay is a trademark of eBay Inc [22].

distributions use the language of probability theory and stochastic processes, important terms and methods are given also in this chapter.
In chapter 3, common terms and their definition used in the context of auctions and online auctions is given, different type of auctions are briefly listed and the mechanism of eBay auctions is discussed.
Game-theoretical approaches to the problem of auctions and online auctions, definition of strategies and an overview of recent researches and opinions are given in chapter 4.
The main part of this work studies the empirical findings and their interpretation. The method of data collection and empirical results are presented in chapter 5. Our interpretations are discussed in chapter 6. It turned out that many of the observed behavior can be explained without considering rational agents. In this chapter an application of our findings in form of a method of detecting shill bidding is given and a minimal model for simulating auctions is discussed.
The work concludes with a summary.

2. Complex Systems

As we will show eBay can be considered as a complex system, where large numbers of agents interact with each other during auctions.
As mentioned, the network economies provided by eBay cause an ever growing number of agents, where the large number of bidders encourage more sellers, which in turn encourage more bidders, etc. An analogy is known from cities, where the number of population cause better economic conditions, which make the city more attractive for those changing their resident city. These kind of growth processes manifest themselves often in power-law distributions known as Zipf's law [2], which is a log-log plot of the rank of a variable versus its value (size) and show a line of the form $y = a - bx$, where b is close to 1. Zipf's law is similar to the cumulative frequency distribution of a variable.
Although there is at present no accepted precise definition of complexity, complex systems share a number of properties and behaviors. These systems are not only complicated but show also long-range correlations, power laws, 1/f noise, scale invariance, self-similarity, self-organization, criticality, etc. Some of these properties, e.g. power laws and 1/f noise are believed to be fingerprints of complex systems.
Properties of complex systems are not properties of any single agent in the system. They usually can not be predicted or deduced from the behavior of the lower-level agents and demonstrate a new stage in the evolution of the system.
A number of these properties are understood but many of them are subjects of new studies. Studying complex systems is rather new, yet there exists a huge literature. In this section we give a brief discussion of the most important properties of complex systems and other related topics named above. Wherever such properties are observed in our study we refer to descriptions in this section.

2.1. Emergence

Perhaps the most important property of complex systems is the so-called emergence. In such systems some behaviors emerge as a result of relationships between the elements of the system. An emergent behavior or emergent property is present when a number of simple agents in a system form more complex behaviors as a whole, i.e. the system made of several agents shows properties which the agents themselves do not have. For example, consider traffic flow as a system of moving cars and the occurrence of jams, which are not a property of cars themselves, but exist in the relation between cars. Emergent properties can also arise between other emergent properties.
The conditions for which a system shows emergent behavior are still subject of modern studies but typically emergent properties arise when the system reaches a combination of

interactivity, organization and diversity [23]. As mentioned, in a complex system it would be impossible for an element or part of the system to control the whole system. If it would be possible, all the complexity would have to be present in that element. For example, consider the stock market. Stock market regulates the relative prices of companies around the world but there exists no leader or one investor which controls the entire market and the complexity of the system emerges through interactions of individual investors [23].

In physics, emergence is used to describe phenomena which occur at macroscopic scale but not at microscopic levels. Examples are color and temperature. Even some basic structures like mass, space, and time are believed as emergent phenomena in some theories (emerging from more fundamental concepts like the Higgs boson).

Another property seen often in complex systems is that the relationships between elements (at microscopic level) are short-range and nonlinear. It means the interaction occurs just between near neighbors and may cause both large and small results.

Nonlinear relationships and interactions are also seen in many chaotic systems in which observed behaviors are random-like. To distinguish between chaotic and complex systems one should measure the order of complexity of the system. This could be done by using algorithmic complexity, which is related to the length of the shortest computer program, which can reproduce the measurements of the system. There exists other methods of measuring the complexity of a system, examples are Gell-Mann's effective complexity [24] and Bennett's logical depth [25].

Typically complex systems can not be considered as Markovian stochastic processes. They have a large history and are very sensible against even small changes in the initial configuration. Any small change can lead to large deviations in the future. This property is also similar to properties of chaotic systems (butterfly effect).

Another important typical property of complex systems is the open boundary condition. It means the energy and information are imported and exported across the system boundaries. Because of this, complex systems are usually far from equilibrium. But it is possible that the system shows stability, where expectation values of observables appear to be time-independent. This state of the system is called stationary state or stationary equilibrium.

2.2. Self-organization

In some systems, normally open ones, under especial conditions it is observed that the internal order (organization) of the system increases automatically and without any external tuning or management. In this case we talk about a self-organizing system and the process is named self-organization. Self-organizing systems show typically emergent properties, however, the relation between emergence and self-organization is still an open question[1]. This idea seems to challenge the second law of thermodynamics, which suggests that the entropy of a system should always increase (or the order should always decrease). However, as far as considering open systems, these two ideas are not in contradiction. There exists the possibility of reducing the entropy by transferring it to the environment, because of the

[1] For further discussion please see [26, 27].

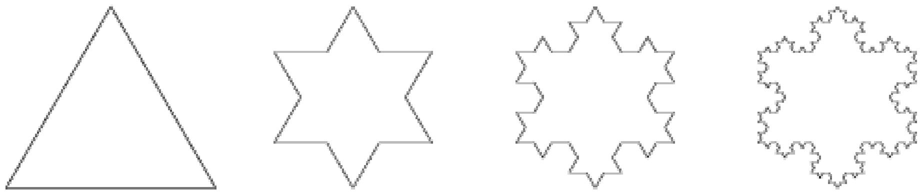

Figure 2.1.: This graph shows an example of Koch's snowflake. From [29].

flow of matter and energy through the system boundaries. It appears, as Ilya Prigogine [28] believed, that because isolated (closed) systems can not transfer energy and matter they cannot decrease their entropy and only open systems far away from equilibrium can exhibit self-organization. However, it is possible for an isolated system to increase macroscopic order by increasing its overall (macroscopic and microscopic) entropy. Which means that some macroscopic degrees of freedom can become more ordered at the expense of system's microscopic disorder.
There exist several classes of physical processes that can be viewed as self-organization. An overview is given below (see [30] for a detailed one):

i. Systems in thermodynamic equilibrium:

   a) structural phase transitions, and spontaneous symmetry breaking:
      i. crystallization and spontaneous magnetization.
      ii. superconductivity, Bose-Einstein condensation and laser.

   b) second-order phase transitions related to critical points and associated with scale-invariant and fractal structures:
      i. percolation in random media.
      ii. critical opalescence of fluids at the critical point.

ii. Thermodynamic systems away from equilibrium (the theory of dissipative structures was developed to unify the understanding of these phenomena), examples include:

   a) structure formation in astrophysics and cosmology including star and galaxy formation.
   b) self-similar expansion.
   c) turbulence and convection in fluid dynamics.
   d) percolation.

iii. Self-organizing dynamical systems:

   a) self-organized criticality, which claims that many systems exhibit critical and scale-invariant behavior similar to the one observed by equilibrium systems undergoing a second-order phase transition at their critical points. Examples are avalanches, earthquakes and forest fires. This theory has been successfully applied to a wide range of different systems. Other possible examples are traffic jams, size of cities and size of companies.

2.3. Self-similarity, Fractals, Recursion

A self-similar object is similar to a part of itself (exactly, approximately or statistically, which means parts of them show the same statistical properties at many scales). One of the simplest examples is Koch's snowflake. An example for statistical self-similarity are coastlines. Self-similarity is a property of fractals which can typically be defined by recursive procedures. Fig. 2.1 shows an example of Koch's snowflake taken from [29].

2.4. Correlation, Power spectrum and 1/f noise

As mentioned, complex systems have typically a long-range memory. This property could be measured in terms of the temporal autocorrelation function of the stochastic process $X(t)$ (which could be any time signal) and is defined as:

$$R(\tau) = \langle X(t_1)X(t_2)\rangle - \langle X(t_1)\rangle \langle X(t_2)\rangle \equiv \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} X_1 X_2 P(X_1, X_2; t_1, t_2) dX_1 dX_2, \quad (2.1)$$

where $\tau \equiv t_2 - t_1$ and $P(X_1, X_2; t_1, t_2)$ is the joint probability density that X_1 is observed at time t_1 and X_2 is observed at time t_2. If there is no correlation between the signal at t_1 and at τ time units later, we have $R(\tau) = 0$. The duration of memory can be measured using the speed with which $R(\tau)$ decreases. Short-range correlated random processes are characterized by a typical time memory τ_c called the correlation time of the process. An example is an exponential decaying autocorrelation function:

$$R(\tau) \propto \exp(-\tau/\tau_c). \quad (2.2)$$

Long-range correlated random processes are in contrast characterized by the lack of a typical temporal scale. This is the case if the autocorrelation function follows a power law (see sec. 2.5.1).
Short-range, long-range or white noise processes can be distinguished by using the so-called power spectrum of the process. For a given signal, the power spectrum gives a plot of the portion of a signal's power (energy per unit time) falling within given frequency bins and can be expressed in terms of the Fourier transform of the autocorrelation function:

$$S(f) = \int_{-\infty}^{\infty} R(\tau) \exp(-2\pi i f \tau) d\tau, \quad (2.3)$$

Consider a stochastic process with a power spectrum of the form:

$$S(f) \sim |f|^{-\beta}, \quad (2.4)$$

quantity x	minimum $x_{\min}$	exponent α
frequency of use of words	1	2.20(1)
number of citations to papers	100	3.04(2)
number of hits on web sites	1	2.40(1)
copies of books sold in the US	2 000 000	3.51(16)
telephone calls received	10	2.22(1)
magnitude of earthquakes	3.8	3.04(4)
diameter of moon craters	0.01	3.14(5)
intensity of solar flares	200	1.83(2)
intensity of wars	3	1.80(9)
net worth of Americans	$600m	2.09(4)
frequency of family names	10 000	1.94(1)
population of US cities	40 000	2.30(5)

Table 2.1.: Parameters for some empirically found power-law distributions with the form $p(x) = Cx^{-\alpha}$. Numbers in parentheses give the standard errors. (From [35])

with $0 < \beta < 2$.
When $\beta = 0$, the power spectrum is frequency-independent and the stochastic process is approximately white noise. White noise $\eta(\tau)$ is the formal derivative of a Wiener process (formal derivative because Wiener process is not differentiable) and has the following properties:

$$\begin{aligned} \langle \eta(\tau) \rangle &= 0 \\ \langle \eta(\tau)\eta(\tau') \rangle &= \delta(\tau - \tau'). \end{aligned} \tag{2.5}$$

So the integral of white noise is Wiener process and is characterized by $\beta = 2$. Short-range processes share the property of having $\beta = 2$. When $\beta \approx 1$ the autocorrelation function lacks a typical time scale and the stochastic process is long-range correlated. In this case the stochastic process is called $1/f$ noise. $1/f$ noise is observed in many different phenomena [31–34] . $1/f$ noise showing self-similar temporal structures are examples of temporal fractals. In contrast to spatial fractals, temporal fractals cannot be observed directly. They are also called "flicker noise"or "pink noise".

2.5. Power laws, Stable laws, Scaling

Power-law distributions of the form $p(x) = Cx^{-\alpha}$ are ubiquitous and appear in many different scientific disciplines like physics, biology, social sciences, economics, computer science, seismology, etc. In Table 2.1 we have listed some examples.
As mentioned in the previous section, power laws could be considered as fingerprint of complex systems. Here we give an overview of the properties of power laws. In section 2.6

simple processes which produce power laws, and proposed theories explaining conditions for which power laws occur are discussed.
In physics, power-law distributions are linked to continuous phase transitions and critical points. They are known as one of the critical phenomena. At a critical point the length scale of a system diverges and the system shows no length scale at all. Such a system is scale-free. For example consider a magnet with a correlation length, a parameter which measures the typical size of magnetic domains. Under certain conditions the correlation length diverges (this is indeed seen experimentally, predicted by simulations and solved exactly theoretically). The diverging correlation length means that all system elements should be in connection to each other and the size of the magnetic domain (cluster) will be the size of the system. In this case the system percolates and we name this one cluster, the spanning cluster. In sec. 2.5.1 we will show that at a critical point (due to the scale-invariance property), the observable quantities in the system should follow a power-law distribution. Usually (and also for the example explained above) the circumstances under which the divergence of the correlation length happens are very specific ones and the parameters of the system have to be tuned precisely to produce the power-law behavior. This makes, however, the divergence of length-scales an unlikely explanation for generic power-law distributions. This is a motivation for the theory of self organized criticality (SOC), which suggests that a large class of different systems move themselves towards their critical points without any external tuning.

2.5.1. Scale-free distributions

Power-law distributions are the only distributions showing scale-free behavior (see Appendix A.1). The form of a power-law distribution is the same at any scale one looks at it. That is the reason why these distributions are also called scale-free. This can be shown by means of an example. Consider the population of cities is observed to follow a power-law distribution and consider we find that cities with a population of 100.000 are 5 times as common as cities with a population of 200.000. Switching to measuring population in millions we will find again that cities with a population of 1 Million are 5 times as common as cities with a population of 2 Millions. The scale-free property of a distribution $p(x)$ can be captured mathematically in the following condition:

$$p(ax) = g(a)p(x), \tag{2.6}$$

where a is an arbitrary constant. This equation means if one increases the scale by which one measures x by a factor a, the form of the distribution $p(x)$ remains unchanged, except for an overall multiplicative constant.
It is easy to show (see Appendix A.1) that a power-law $p(x)$ is the only distribution satisfying the condition of Eq. (2.6).

2.5.2. Stable laws, Invariance properties

Consider $S_n = x_1 + x_2 + ... + x_n$ as the sum of n independent identically distributed (i.i.d.) random variables x_i with probability density function (pdf) $P_1(x)$. In general S_n will have a different pdf $P_n(x)$ given by n convolutions of $P_1(x)$. If the functional form of $P_n(x)$ is the same as the functional form of $P_1(x)$, this pdf and the corresponding stochastic process are said to be stable. Within the formalism of the renormalization group, a stable law is connected to a fixed point of the renormalization group (RG) transformation. Attractive fixed points of RG describe the macroscopic behavior in the $n \to \infty$ limit. Phase transitions as a global change of regime at the macroscopic level under tuning of a control parameter to change the strength of the correlations, are linked to repulsive fixed points [36].
One of the properties of stable distributions is self-similarity. We will show this in section 2.5.3.
The central limit theorem (CLT) states that in the $n \to \infty$ limit, and if x_i has a zero mean and a finite variance σ^2, the normalized sum $\frac{S_n}{\sqrt{n}}$ will be a random variable with a pdf converging to the Gaussian distribution with variance σ^2. This means Gaussian distributions are stable. The stable laws have been studied and classified by Paul Lévy (see [37]). He discovered that in addition to Gaussian distributions, there exists a large number of other pdfs sharing the stability condition:

$$P_n(y)dy = P_1(x)dx, \tag{2.7}$$

with $y = a_n x + b_n$. All these pdfs share the property of having an asymptotic power-law behavior. The asymptotic approximation of non-Gaussian stable distributions valid for large values of $|x|$ can be written as:

$$P(x) \sim |x|^{-(1+\mu)}, \tag{2.8}$$

This power-law behavior has deep consequences for the moments of the distribution. For instance, all Lévy stable processes with $\mu < 2$ have infinite variance and with $\mu < 1$, have not only infinite variance, but also infinite mean.
Power-law distributions are not only stable under addition (aggregation), but also share other invariance properties. As Mandelbrot shows in [38] (see [39] for further discussion), power-law distributions are the only distributions that are invariant under mixture and maximization transformations. These properties can be seen as a reason of ubiquity of power laws. We refer to Mandelbrot [38], who suggests power-law distributions are the norm for high variability data (i.e. data with infinite variance) and need no extra special explanation (just like the Gaussian distribution as a norm for low variability data, i.e. data with finite variance). Invariance properties under observation transformations, such as addition or mixture of several data sets or using a subset of maximums/minimums of the data, are quite important for practical purpose. For example in economics, aggregate incomes are easier to collect than data including each type of income separately; file sizes in the Internet are a mixture of different distributions of the file sizes existing on the various Web servers; historical recorded data like earthquakes, wars, etc. are typically recorded as exceptional events (e.g. largest).

2.5.3. Scaling

As mentioned, Gaussian and Lévy distributions are stable. They also present similar scaling properties. As we have shown in sec. 2.5.2 the pdf of the sum of N Gaussian i.i.d. random variables with mean $\langle x \rangle$ and variance σ^2 is also a Gaussian with mean $N\langle x \rangle$ and variance $N\sigma^2$. It means the rescaled variable $\frac{S-N\langle x \rangle}{\sigma\sqrt{N}}$ has the same pdf as the initial variables and is independent of N. This is also the case if one starts with power-law distributed variables of the form Eq. (2.8) with $1 < \mu < 2$. For the case $0 < \mu \leq 1$ the rescaled variable is $\frac{S}{\sigma\sqrt{N}}$ (in this case $\langle x \rangle$ is not defined). This property provide a testing strategy for the existence of Gaussian and power laws. One can compare different data sets (e.g. obtained from different time windows or system sizes) and try to collapse the different pdfs onto one universal curve by varying the exponent μ. When such a reascaling exists, we say that the pdf exhibits scaling properties. Scaling is observed in a wide variety of different systems. Important examples are equilibrium systems at their critical points and self-organizing non-equilibrium systems.

2.5.4. Log-normal versus power law

If the logarithm of a variable x is distributed according to a Gaussian pdf, x is said to be distributed according to a log-normal pdf:

$$P(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\frac{1}{x}\ [\exp{(-\ln x - \mu)^2/2\sigma^2}], \tag{2.9}$$

Log-normal distributions are also very often observed in natural and social sciences. They can be mistaken locally for a power law because they can look very similar over a relatively large interval [40].
It is not easy to distinguish between power law and log-normal if the pdf curve is not extended over several orders of magnitude. Log-normal distributions are not stable under addition, but stable under multiplication, i.e. the product of log-normal distributions is again log-normal.

2.6. Mechanisms for producing power laws

In this section we give an overview and describe a number of important physical mechanisms that have been proposed to explain the occurrence of power laws.

2.6.1. The Yule process

The Yule growth process [41] is one of the most important general mechanisms that can produce power-law distributions with a wide range of exponents to match the observations by a suitable tuning of its parameters. For several of the distributions listed in Table 2.1, especially city populations, citations and personal income, it is now the most widely accepted theory.

Consider a system as a collection of objects, such as papers (other examples include cities, web pages, etc.). New objects appear every once in a while as people publish new papers. Each paper has some property z associated with it, such as citations to the paper (or number of people in a city etc.), that is observed to be distributed according to a power law. We wish to explain this measured power-law distribution. Newly appearing objects have some initial value z_0 of z, for the case of papers and citations $z_0 = 0$. This could be different if we consider other objects. Suppose in the time span between the appearance of one paper and the next, l new citations (people) are added to the entire system. So some papers (cities) will get new citations (people), but not necessarily all. Suppose the simplest case that these new citations are added to papers in proportion to the number of citations that the paper already has (i.e. we assume, that a paper that already has many citations is more likely to be discovered during a literature search and hence more likely to be cited again). This type of "rich-get-richer" process seems to be plausible for a wide range of systems. Simon [42] named this process the "Gibrat principle", but it is used under names of the "Matthew effect" [43], "cumulative advantage" [44] and "preferential attachment" [45].

There is, however, a problem when $z_0 = 0$. For example, new papers appear with zero citations and will never get any citations. To overcome this problem one assigns new citations in proportion to $z + c$, where c is a constant. Thus there are three parameters z_0, c and l that control the behavior of the model.

Let us measure the time in terms of steps, by each one new paper appears. so the total number of citations is $n(l+z_0+c)$. Let us denote by $p_{z,n}$ that fraction of papers that have z citations when the total number of papers is n. The probability that paper i receives a new citation in the interval between nth and $n+1$th step is given by $l(z_i+c)/(n(l+z_0+c))$ and the total expected number of papers with z citations that receive a new citation will be:

$$\frac{l(z+c)}{n(l+z_0+c)} \times n p_{z,n} = \frac{l(z+c)}{l+z_0+c} p_{z,n}. \tag{2.10}$$

The number of papers with z citation will decrease on each time step by this number. At the same time this number increases because of those papers that previously had $z-1$ citations and now have z. Thus one can solve the master equation for the new number $(n+1)p_{z,n+1}$. In Appendix A.2 we show that the Yule process generates a power-law distribution $p_z \sim z^{-\alpha}$ ($p_z = \lim_{n\to\infty} p_{n,z}$) with an exponent related to the three parameters of the process according to

$$\alpha = 2 + \frac{z_0 + c}{l}. \tag{2.11}$$

For citations of papers or links to world wide web pages we have $z_0 = 0$ and we must have $c > 0$ (to get any citations or links at all). So $\alpha = 2 + c/l$. Price [44] assumed that $c = 1$ so that paper citations have the same exponent $\alpha = 2 + 1/l$ as the standard Yule process. The most widely studied model of links on the web (Barabási and Albert [45]) assumes $c = l$ so that $\alpha = 3$. This assumption can, however, not really be justified. The actual exponent for numbers of links to web sites is measured to be $\alpha = 2.2$. So if the Yule process is to predict the measurements we should put $c \simeq 0.2l$.

2.6.2. Multiplicative processes

Levy and Solomon [46] have found that random multiplicative processes $x_t = \lambda_1 \lambda_2 ... \lambda_t$ (with $\lambda_j > 0$) lead, in the presence of a boundary constraint, to a distribution $P(x_t)$ in the form of a power law $x_t^{-(1+\mu)}$. This process is discussed in detail by Sornette and Cont [47], who found that this result applies to the asymptotic distribution of x_t if the necessary conditions 1) $\langle \log \lambda_j \rangle < 0$ (corresponding to a drift $x_t \to 0$) and 2) x_t not be allowed to become too small, are satisfied.
They show that a class of convergent multiplicative processes with repulsion from the origin of the form:

$$x(t+1) = e^{F(x(t),\{b(t),f(t),...\})} \; b(t) \; x(t) \;, \tag{2.12}$$

share the same power-law pdf

$$P(x) = Cx^{-1-\mu} \tag{2.13}$$

for large x with μ solution of

$$\langle b(t)^{\mu} \rangle = 1 \;. \tag{2.14}$$

with $F \to 0$ for large $x(t)$, leading to a pure multiplicative process for large $x(t)$ and $F \to \infty$ for $x(t) \to 0$ (repulsion from the origin). F must obey some additional constraints such as monotonicity which ensures that no measure is concentrated over a finite interval.
The fundamental reason for the existence of the power-law pdf (2.13) is that $\ln x(t)$ undergoes a random walk with drift to the left and which is repelled from $-\infty$. A simple Boltzmann argument [47] shows that the stationary concentration profile is exponential, leading to the power-law pdf in the $x(t)$ variable.
These results were proved for the process $x(t+1) = b(t)x(t) + f(t)$ by Kesten [48] using renewal theory and was then revisited by several authors in the differing contexts of ARCH processes in econometry [49] and 1D random-field Ising models [50] using Mellin transforms, extremal properties of the $G-$*harmonic* functions on non-compact groups [46] and the Wiener-Hopf technique [47].

2.6.3. Random walk

The widely used term "random walk" was first used by the biologist Karl Pearson in 1905 [51], but the history of random walks goes back to two earlier observations. One of these observations was the irregular movement of small pollen grains in a liquid observed by Brown in 1828 and known in physics as Brownian motion. The other one was the observation of irregular series produced in gambling (coin tossing). This observation rose the interest of Pascal, Bernoulli and Fermat in mid 16th century.
In physics, random walks are used as simplified models of Brownian motion and also as models for understanding some power-law distributions observed in the nature. A number of properties of random walks are distributed according to power laws. For instance the distribution of return times t (the walker returns to position 0 for the first time at time t) follows a power law [35]:

$$f_t \sim t^{-3/2}. \tag{2.15}$$

Random walks are also used for studying the avalanche dynamics typically occurring in the theory of self-organized criticality. By using an unbiased first-return random walk process and defining the number of steps the walker takes to return to position 0 as the lifetime of the avalanche T, and the number of different sites the walker has visited as spatial size of the avalanche S, Yang *et al.* [52] show that T and S both follow power-law distributions. An early analytical work of Montroll *et al.* [53] shows that for $T \to \infty$ the relation between $\langle S \rangle$ and T follows a power law:

$$\langle S \rangle \sim (16T/\pi)^{0.5}. \tag{2.16}$$

2.6.4. Combination of exponentials

Consider a process in which items grow exponentially in time. An example is the population of organisms reproducing without any resource constraint. The size of the population is then given by $x \sim e^{\alpha t}$ with $\alpha > 0$. Suppose the items have a fixed probability of dying per unit time, so that the times t at which they die are exponentially distributed $p(t) \sim e^{\beta t}$ with $\beta < 0$. This process has been discussed by Reed and Hughes [54], who found that the distribution of the sizes x of the items at the time they die follows a power law. This is actually an example of the general mechanism of combination of two exponentials, which produces power-law distributions. Exponential distributions are very common and arise in many different phenomena. In physics, examples of exponential distributions include survival times of β decay and the Boltzmann distribution of energies.
We describe this mechanism mathematically as follows. Suppose some quantity y has an exponential distribution:

$$p(y) \sim e^{gy}, \tag{2.17}$$

with g as a constant (for $g > 0$ there must exist a cutoff on the distribution). Suppose, we are interested in some other quantity x which is exponentially related to y:

$$x \sim e^{hy}, \tag{2.18}$$

(h is another constant).
The probability density function of x can be written as:

$$p(x) = p(y)\frac{dy}{dx} \sim \frac{e^{gy}}{he^{hy}} = \frac{x^{-1+g/h}}{h}. \tag{2.19}$$

And $p(x)$ will be a power-law distribution with exponent $\alpha = 1 - g/h$.
The power-law distribution of the frequencies of words was explained by using this mechanism [55].

2.6.5. Self-organized criticality

At their critical points systems develop power-law distributions because of the divergence of some characteristic scale. It was first proposed by Bak *et al.* [56], that many dynamical systems arrange themselves so that they always sit at their critical point. One says that such systems display self-organized criticality (SOC). The original purpose of their work was to explain why spatial and temporal fractals (1/f noise. See sec. 2.4) are found so frequently in the nature. Despite initial enthusiasm and a large number of investigations related to this problem, up to know there exist no accepted precise definition of SOC and necessary conditions under which SOC behavior arises. Although self-organized critical models have been used to understand many phenomena like forest fires [57], evolution [58], avalanches [56], earthquakes [10, 59] and solar flares [60], SOC suffers from the lack of a detailed formalism and often does not provide a detailed explanation for the origin of power laws in these systems.

This theory explains, however, how a number of interactive systems (observed in nature) generate power-law relationships from simple interaction rules. As mentioned, the theory of SOC claims that such systems self-organize themselves into a critical state. In this state small perturbations might start chain-reactions named avalanches affecting a great number of agents of the system.

An example of a naturally occurring SOC phenomena is the well known Gutenberg-Richter-law, which gives the magnitude-frequency relation of earthquakes. However the standard example of the theory is established to be the so called sandpile model. We will explain this model briefly here.

Imagine we have a quadratic lattice sandbox and add one sand grain at each time unit to the box. First grains form a sandpile with a stable slope but by increasing number of grains the slope increases until it reaches locally a critical value such that the addition of an extra grain results in an avalanche. These avalanches fill the empty space of the lattice little by little.

In both cases, whether the slope is larger or smaller than the critical slope, the processes bring it back to its critical value through avalanches or addition of grains respectively leaving the system in a critical state. In this state, the distribution of many variables is found to follow power laws. Examples are the lifetime, size and linear extent of the avalanches. The relations between these variables are also found to be described by scale-independent relationships [61].

The most important elements of the theory are listed below:

- Avalanches can have any size[2]. Yet, the distribution of the size of avalanches follows a power law.
- Avalanches happen in a stochastic manner and there exists no periodicity.
- The surface of the sandpile shows a fractal structure.
- Global physical laws determine and dictate the interactions between the grains.

It is important to mention that sandpile experiments require extra rules, e.g. the addition-rate should be such that the system has enough time to relax. Those rules could be considered as tuning, which challenges the concept of self-organization.

[2]Ranging from one grain to the total number of grains building the sandpile.

2.7. Mechanisms for producing log-normals

2.7.1. Multiplicative processes

We describe a multiplicative process which can produce log-normal distribution [62]. This process was first introduced 1930 by the economist Gibrat [63] and is also used in biology to describe the growth of organisms.
Suppose we start with an organism of size X_0. At each step t, the organism may grow or shrink, according to a random variable R_t, so that $X_t = R_t X_{t-1}$. If the R_k (with $1 < k \leq t$), have all log-normal distributions, then the X_t will also have log-normal distribution for all t due to the property of stability under production of log-normal distributions (sec 2.5.2). More generally, X_t will have a log-normal distribution even if the R_t are not themselves log-normal distributed. Consider $\ln X_t = \ln X_0 + \sum_{k=1}^{t} \ln R_k$. If the $\ln R_t$ are independent and identically distributed variables with finite mean and variance, the Central Limit Theorem says that $\sum_{k=1}^{t} \ln R_k$ converges to a Gaussian (normal) distribution (sec 2.5.2), and X_t is well approximated by a log-normal distribution.

3. Auctions, Online auctions, eBay

3.1. Auctions

An auction is a process of buying and selling things. In an auction an item is offered up for sale by a seller. After taking bids, the item will be sold to the highest bidder. There exists a special terminology used in the context of auctions (and online auctions). The most important terms and their definitions are listed below. Definitions used by eBay is given if a term is used only in online auctions (like "proxy bidding").
Different type of auctions are briefly listed and the mechanism of eBay auctions is discussed in this chapter.

- **Ask Price**: The minimum bid accepted for the item. In online auctions, ask price is the current or listed price plus the bid increment.
- **Auctioneer**: An agent (person or a company) who conducts an auction.
- **Bid**: An offered price from a bidder.
- **Bidder**: Buyer. An agent who puts offers (bids) on the auctioned item. In the context of online auctions the term "agent"is often used for "bidder".
- **Bid Increment**: The minimum amount by which a bid must exceed the current price to be accepted by the auctioneer. This amount is usually determined by the current price, i.e. it might not be constant during an auction.
- **Bid Retraction**: When a bidder cancels his own bid.
- **Bid Shilling (also Shill Bidding)**: Fraudulent bidding by the seller by using an alternate account or an associate of the seller in order to inflate the price or increase the attractivity of an item.
- **Common Value Auctions**: In a common value auction, the actual value of the object being auctioned is the same to all bidders (but the actual value is perhaps not known to anyone). Bidders place their bids on the basis of their own estimate of the actual value of the item, but these values are strongly positively correlated (in contrast to private value auctions).
- **Current Price (also Listed Price or Current Bid)**: The last listed price of the item. In second-price auctions current price is the second highest bid to date plus the bid increment (if this amount is smaller than the maximum bid, unless the current price would be the maximum bid).

- **Duration**: Online auctions run on eBay have fixed time lengths of 1,3,5,7 or 10 days.
- **Feedback**: Publicly-available ratings and comments that buyers and sellers leave on each others profile. eBay only allow a user to provide feedback on another user with whom he has completed a transaction.
- **Final Sale Price (or Final Value)**: Final listed price.
- **Final Value Fee**: The fee a seller has to pay the auctioneer, depending on the final sale price.
- **Hard Close**: The auction has a fixed duration and ending time known to everyone.
- **High Bidder**: The buyer who placed the bid which would win the auction if it were to close immediately.
- **Insertion Fee (or Listing Fee)**: The fee a seller has to pay the auctioneer for listing his item.
- **Listing**: The act of inserting/submitting an item for sale.
- **Maximum Bid (or Threshold)**: The maximum amount the buyer is willing to pay for an item. This amount is not revealed to other users and is used in proxy bidding.
- **Minimum Bid**: The minimum bid at any time is bid increment over the current ask price of the auction. (If a bid is the first bid, the minimum bid is the starting price.)
- **Private Value Auctions**: In private value auctions, in contrast to common value auctions, the actual value of the object being auctioned is not the same to all bidders. Every bidder has his own estimation of the actual value. Examples are rare works of art, antiques and stamp collections.
- **Proxy Bidding (or Automatic Bidding)**: In online auctions: An option whereby the auction site (eBay) automatically increases the buyer's current bid (called also proxy bid) for the item over any subsequent bids placed by other buyers by the lowest possible amount necessary (determined by price-dependent bid increments) to maintain the buyer as the high bidder until the buyer's maximum bid is reached.
- **Proxy Bid**: The bid placed by the proxy system as bidder's.
- **Reserve Price**: The minimum price the seller will accept for the item to be sold. If the final price is lower than the reserve price, no sale will take place. Reserve price is not revealed to other users.
- **Soft Close**: The auction does not have a fixed ending time.
- **Sniping**: Placing a winning bid just before an auction closes. This is a strategy used to prevent other bidders from outbidding the sniper or driving the price higher.

- **Starting Bid (also Starting Price or Opening Bid)**: The smallest amount that can be entered as a bid for an auction. This is set by the seller. (Note that, in addition to starting price, there may also be a reserve price on the item.)

- **User**: In online auctions, a user is a registered member, who can participate in auctions as buyer or list items for sale as seller.

- **Winner**: Highest bidder at the end of the auction, who pays the final listed price for the item.

- **Winning Bid**: A bid with which an auction ends.

3.2. Different types of auctions

There exists different types of auctions. The features and mechanisms of most common auction types are briefly listed below:

- English auction: Buyers bid against one another openly, by bidding higher than the previous bid. The auction ends when there exists no bidder, who is willing to bid further, or when a pre-determined price is reached. The highest bidder pays the price. The seller may set a reserve price, so if the auctioneer fails to raise a bid higher than this reserve the sale will not go ahead.

- Dutch auction (traditional): In the traditional Dutch auction the auctioneer begins with a high asking price and lower the price until a bidder is willing to accept the auctioneer's price, or a pre-determined minimum price is reached. In this kind of auctions a sale requires only one bid.

- Dutch auction (by online auctions): Online auctions, in which more than one identical good is sold simultaneously to a number of high bidders.

- Sealed first-price auction: All bidders simultaneously submit bids. No bidder knows the bid of any other participant. The highest bidder pays the price he submitted.

- Sealed second-price auction (Vickrey auction): This auction type is identical to the sealed first-price auction, except the winning bidder pays the amount of the second highest bid.

- Silent auction: Participants submit bids normally on paper with or without the knowledge how many other people are bidding or what their bids are. The highest bidder pays the price he has submitted.

3.3. Properties of online auctions in comparison to traditional auctions

In contrast to traditional auctions, in online auctions bids can be placed at any time. Items are listed for a number of days, so potential buyers have enough time to search, decide, and bid. There are also no geographical constraints, sellers and bidders with internet access can participate from anywhere in the world. This makes them more accessible and reduces at the same time the cost of attending the auction. So the number of listed items and the number of bids for each item may increase[1]. The items do not need to be shipped to a central location, reducing costs, and reducing the seller's minimum acceptable price. Due to the broad scope of products available, potential for a relatively low price, reduced selling costs and the ease of access, there exist a large number of bidders and sellers building a network.
A very recent work [64] studies the network of economic interactions that forms on eBay.

3.4. eBay

3.4.1. Some facts about eBay

With 105 million members worldwide (42 million in the US) and nearly $24 billion (worth of goods exchanged in year 2004) eBay is the leading global e-commerce website. Each day, there are millions of items sold on eBay. People come to eBay to buy all kinds of practical, unique, and normal items, such as computers, automobiles, musical instruments, real estate, jewellery, cameras, furniture, boats, scientific equipments, sporting goods, etc. eBay reports that every three hours one Corvette is sold on its site , a diamond ring is sold every six minutes, a digital camera every 90 seconds, and an article of clothing every three seconds.
Fig. 3.1 shows one of the most recent advertisements of eBay just before the election (May 2005) in North Rhine-Westphalia (NRW) in Germany.

3.4.2. eBay auctions

The following description concerns rules used for items listed on eBay Germany (www.eby.de) in the time between May 2003 and May 2004. Some of these rules and specially the detail amount of fees could be different if one uses other eBay sites (like eBay.com or eBay.co.uk) for listing. The rules may change with the time so the information provided in the present work are not necessarily valid in the future.

Listing items

To list an item for sale, the seller enters an auction category, a title, a description, shipping conditions, payment methods, a starting price, auction's starting date, duration of the auc-

[1] In comparison to traditional auctions

Figure 3.1.: One of the most recent advertisements of eBay just before the election (May 2005) in North Rhine-Westphalia (NRW) in Germany.

tion and some listing options which have influences on layout or position of the item when it appears on the internet pages of eBay. The seller pays eBay an "Insertion Fee" (minimum 0.25 Euro for an item with starting price below 5 Euro and without any extra listing option), which depends on the listing options, the starting price and the insertion category of the item. The seller is also charged a "Final Value Fee" which is based on the final sale price of the item (between 2.5 and 5 percent of the final sale price).

Publicly available information

Following data are publicly available as an auction proceeds:

- current bid
- total number of bids to date
- identity of all bidders and the time of their bid[2] (but not the amount of their bid as long as the auction has not ended).
- remaining time of the auction

[2]If the auction is set to be "private" by the seller, the identity of the bidders are nor revealed.

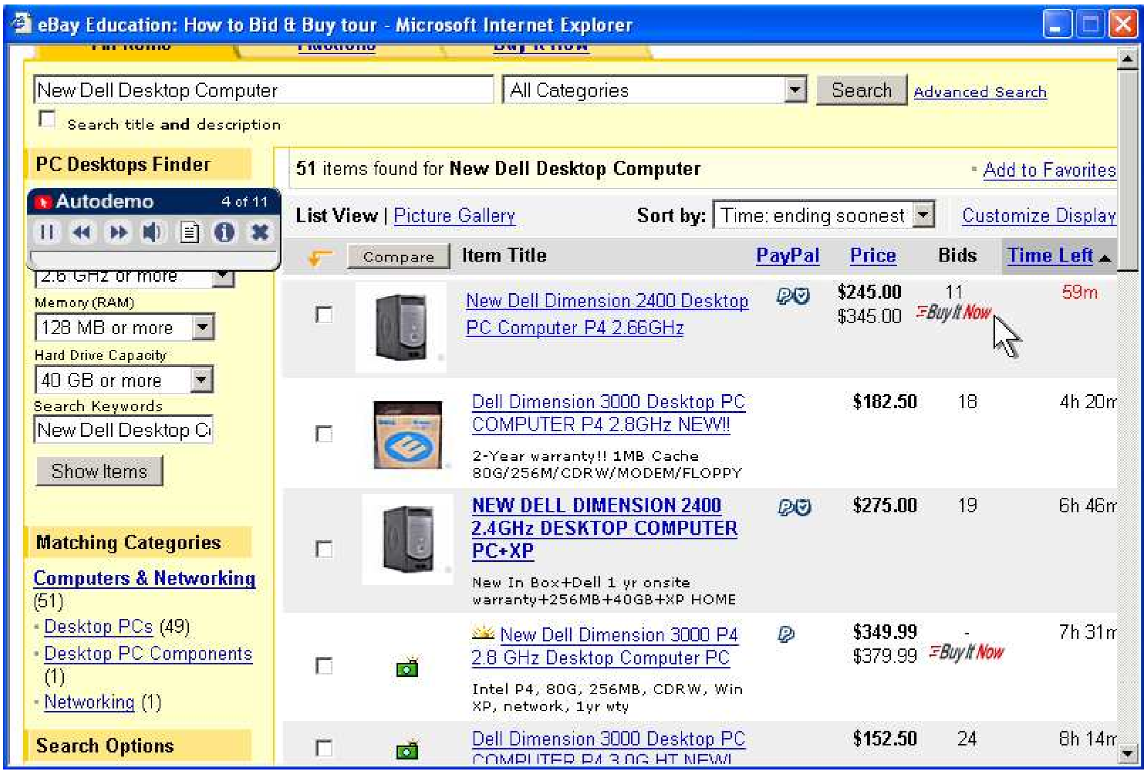

Figure 3.2.: Screenshot of an eBay webpage. This Screenshot is taken from eBay's course center [65].

- feedback information of seller and bidders
- static information, like starting date and time of the auction, starting price, insertion category, etc.

Fig. 3.2 and Fig.3.3 show screenshots of the items-listing and the auction pages. These screenshots are taken from eBay's course center [65].

Bidding

To participate in an auction run on eBay, one has to be a registered user. It is free of costs to register on eBay. Any registered user can bid in an auction of his choice. Generally, the user searches for the item of his interest and checks if the item matches. Before bidding he might consider the seller's reputation[3]. Placed bids are considered binding but bid retraction is possible in certain circumstances.

[3] Available in the form of feedback comments and rating.

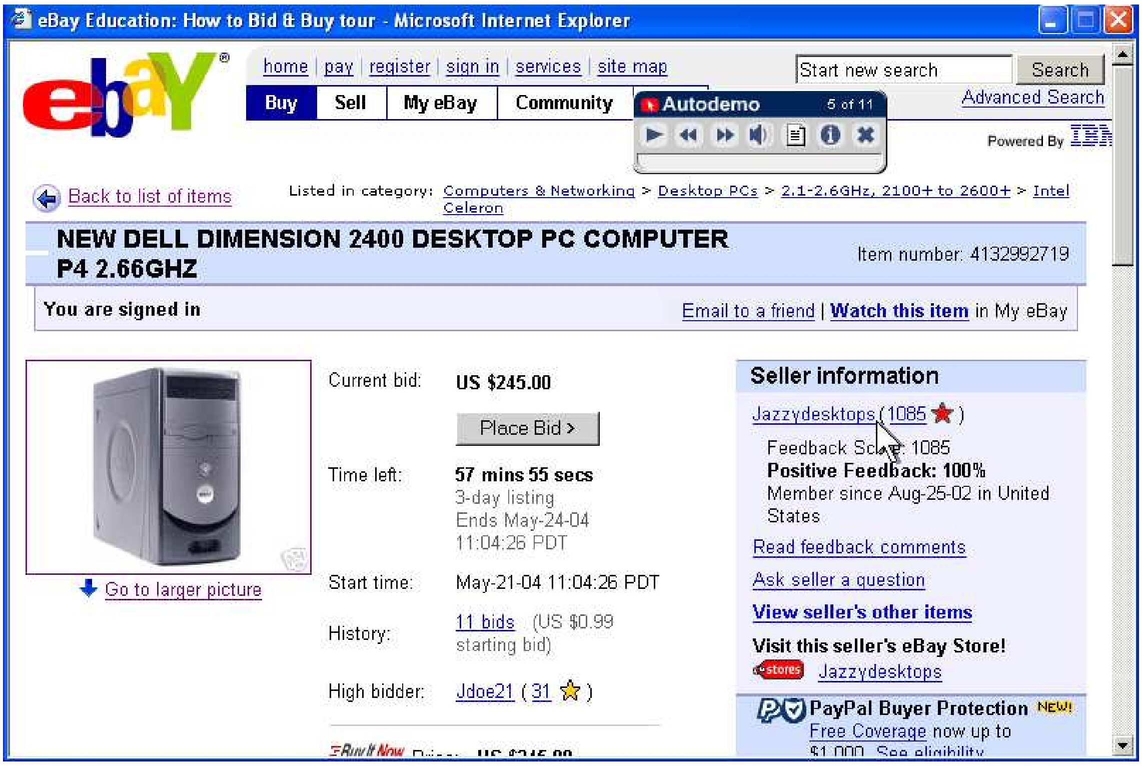

Figure 3.3.: Screenshot of an eBay webpage showing the publicly available data for an auction. This Screenshot is taken from eBay's course center [65].

Mechanism of auctions

All eBay auctions are similar to second-price auctions (see sec. 4.1). They use an ascending-bid format with the important difference that there is a fixed ending time set by the seller. There are a variety of different auction types run on eBay. Although we only study the standard single item auction type (mostly with starting price of 1 Euro), here you find a brief description of different types of listing:

- Single Item: This is the most common type of listing. Here a single item is being offered. The highest bidder pays the amount of the second highest bid plus a price-dependent increment (see table 3.1).
- Dutch Auction: The seller offers more than one of the exact same item. The bidders enter the desired quantity of the items with the price they want to pay per item. All winners pay the lowest winning bid price.
- Buy It Now: A bidder can immediately win the item by choosing the "Buy It Now option". This option can be selected by the seller during the listing process. The "Buy It Now" price is also set by the seller during the listing process and is known to bidders. This option is available to bidders until the first bid.

Current Bid	Bid Increment
EUR 1,00 - EUR 49,99	EUR 0,50
EUR 50,00 - EUR 499,99	EUR 1,00
EUR 500,00 - EUR 999,99	EUR 5,00
EUR 1000,00 - EUR 4.999,99	EUR 10,00
EUR 5000,00 and up	EUR 50,00

Table 3.1.: Bid increments used by eBay (From [66]).

Bid increments

Bid increment increases as the current bid increases. Table 3.1 shows how eBay determines the bid increments.

Feedbacks

Every eBay user has a feedback profile, which includes a feedback score and comments left by their trading partners from previous transactions. This comments are classified as positive (+1), negative (-1) or neutral (0). The feedback score is the sum of these scores, in which only comments from distinct users are used.

Proxy bidding

eBay uses a proxy mechanism for all submitted bids. Each bid on eBay is interpreted as the bidder's maximum bid. An automated proxy system would bid for the bidder as the auction proceeds, bidding only enough to outbid other bidders. If someone outbids the bid, the system automatically increases his bid. This continues until someone exceeds this maximum bid, or the auction ends and this bidder wins the auction. By using the proxy system the user doesn't have to keep coming back to re-bid every time another bid is placed. The description below is a brief overview how bidding on eBay works (from [67]):

i. When a user places a bid, he enters the maximum amount he would be willing to pay for the item. This maximum amount is not revealed to other bidders and the seller and is kept confidential.

ii. The eBay system compares this bid with bids of the other bidders and places bids on behalf of the user, using only as much of his bid as is necessary to maintain his high bid position. The system will bid up until the user's maximum amount is reached.

iii. If another bidder has a higher maximum, the user will be outbid. But, if no other bidder has a higher maximum, the user wins the item and could pay less than his maximum price.

eBay's proxy bidding system plays a central rule for the auction process. The algorithm is used in our study for reverse engineering and finding out about the current price at any given time. To make the mechanism of eBay's proxy bidding system better understood, we give an example. Consider there is a "Stamp Collection"with the starting bid of 1.00 Euro offered. You see it but you are not sure that you really want it, so you bid 2.00 Euro. You will be the high bidder with 1.00 Euro (since you have not outbid any other bidder, the current amount is the same as the starting bid). The total number of bids will be 1. Later another interested bidder sees the item. He enters a bid of 20.00 Euro. In this moment, because he is the new high bidder and because he took the lead from someone else, he needs to beat your max bid. He takes the lead for 2.50 Euro (that is 2.00 Euro to match you plus 0.50 Euro for the increment to pass you). The total number of bids will be 2. It is not possible for you to know about his true maximum bid in this moment. Later, he decides that 20.00 Euro may not be enough to hold the item, so he bids again for 30.00 Euro. He is bidding basically against himself but eBay recognize this, so he still remains the high bidder and the current bid is still 2.50 Euro, but the total number of bids will be 3. You see this and decide that you really do want the item, and you are willing to bid more. You bid 5.00 Euro. That is not enough to overtake him, so he remains the high bidder. eBay moves his proxy bid up to 5.50 Euro. At this moment the total number of bids is 4. Now you bid 25.00 Euro. He remains the leader at 26.50 Euro. Still later you bid 50.00 Euro. You now take the lead at 30.50 Euro. Later, a stamp collector comes along and is interested in this item. She bids 400.00 Euro for it. She takes the lead at 51.00 Euro, since the increment is now 1.00 Euro instead of the former 0.50 Euro. If there are no other bids, the collector wins for 51.00 Euro. Her 349.00 safety margin would never be seen by anybody. The total number of bids will be 7 and the distinct number of bidders will be 3 at the end.

4. Game-theoretical Approaches

Game theorists study the behavior of individuals and optimal strategies in games. Although the first formalization of the game theory by John von Neumann and Oskar Morgenstern [68] in 1944 had the aim of explaining economic behaviors, it found also applications in other fields like biology [69], political science [70], psychology [71], social science [69], warfare and operations research.

By using the methods of game theory one seeks to find rational strategies, in which situations are taken into consideration where the outcome depends on the strategies chosen by all players (not only on one's own strategy). These strategies might have different or overlapping goals.

The analysis and design of auctions are successful applications of the game theory. Auction theory was pioneered by W. Vickrey in 1961 [72].

In this chapter we give an overview of results and raising questions when using game theoretical approaches to understand second-price auctions.

4.1. Second-price auctions

The most familiar traditional type of auction is the open ascending-bid auction. This type of auction is also called English auction (see sec. 3.1). A complete analysis of the English auction as a game is rather complicated.

Another auction type similar to English auction but simpler for analysis is the second-price (Vickrey) auction. Here the winner in the English auction pays the amount of the second-highest bid. In a second-price auction, each potential buyer submits his bid to the auctioneer, at the end of the auction, the auctioneer awards the auctioned object to the bidder with the highest bid and charges him the amount of the second-highest bid.

One of the problems appearing when dealing with English auctions (in private value auctions and in common value auctions with incomplete information) is the so-called winner's curse. In common value auctions with incomplete information, the object being sold have a similar value for all bidders, but the bidders are uncertain about this value at the time they submit their bids. Each bidder has his private estimation of the value of the auctioned object. It is plausible to assume that the average estimation of the bidders is close to the actual value of the object. It means, the person with the highest bid has almost certainly overestimated the value of the object. This person is the winner of the English auction. Thus, the winner, who wins after bidding his true valuation has almost certainly overpaid.

This result can be obtained formally by using the conditional probability. If we calculate the bidder's expected payoff from the auction conditioned on the assumption that he won the auction, it turns out that for bidders bidding their true estimate, the expected profit is

negative. It means that on average the winning bidder is overpaying.
Second-price auction was originally created by W. Vickrey [72] to reduce this problem (winner's curse) and encourage people to bid what they believe the object is truly worth. It turns out that this type of auction can reduce, but not eliminate, the winner's curse.
Let us answer the question of how one should bid in a second-price auction by using the language of the game theory through an example. Consider a private value auction. It means each bidder's estimation of the value is based on his personal opinions and tests for the object (see sec. 3.1).
First consider the situation that a bidder bids less than the object was worth to him. In this case, if he wins the auction, he pays the second-highest bid, so bidding less than his true valuation does not change the result. But he risks that the object is being sold to someone else at a lower price than what the object is worth to him, which makes him worse off. Now assume he bids more than his true value, the only case where this can make a difference is when there exists, below this new bid, another bid exceeding his own true value. In this case he, if he wins the object, must pay that price, which he prefers less than loosing the auction. In other cases the outcome will be the same.
So in a second-price auction the optimal bidding strategy is to bid one's true value for the object. Such strategy is called weakly dominant strategy in game theory. That is the strategy with the best outcome, irrespective of what the other bidders are doing.
Second-price auctions provide insight into a Nash equilibrium (where no bidder can make his payoff better by changing his strategy [20]) of the English auction. There exists a strategy in the English auction which is equivalent to the weakly dominant strategy in the second-price auction. In this strategy, a bidder remains active until the price exceeds his true value, and then drops out. If all bidders use this strategy, no bidder can make his payoff better by switching to a different one.
eBay online auctions have features of both English and second-price auctions, in which, the current price is observable to everybody. However, a bidder, instead of frequently checking for the current price, can use the proxy bidding system provided by eBay, to stay in until the price exceeds a given amount. If the high bidder is another one and his bid is below that amount, then the system only increases the price enough so that it has the new high bid. Operationally this gives bidders the possibility to act as if they are participating in a second-price auction and bid their true value just once.

4.2. Game theoretical approaches by eBay auctions

Conventional game theory assumes that each player acts fully rational. However this assumption does not hold in many cases, whether the players are humans or animals. This counts as one of the major problems of applying game theory in the real world.
As we have shown in the previous section, given a second-price auction and fully rational agents, the timing of bids plays no role and there is no inducement to bid less than one's own value. However, empirical observations provided by Roth and Ockenfels [73] show that internet auctions with a hard close (like eBay auctions) do not show the properties of the perfectly functioning second-price auctions.

In hard close auctions, the majority of bids is found to be placed just at the end of an auction. These strategies are called late bidding and sniping. One of the arising questions is the explanation of the difference between predictions of game theory and empirical findings. Several groups have tried to understand this puzzle by defining and explaining new rational strategies [73–76]. In this section we give a list of used strategies and an overview of existing answers. Possible different strategies are:

- Sniping: Bidding in the last few seconds.
- Late Bidding: Bidding in the last few minutes.
- Evaluation: Placing one's true value. In most cases the bids are placed early and are significantly greater than the minimum required value.
- Incremental Bidding: Placing minimum acceptable bids. Usually more than one bid is placed close to each other.
- Seller's Shill Bidding: See sec. 3.1
- Squeezing: By squeezing, a seller uses any second eBay account to bid in his favour, in order to uncover the sealed bid of potential buyers. By learning the threshold of the highest bidder, the seller either retracts or cancels a shill bid and then he will submit another bid, this time matching the threshold (he learned from the highest bidder before). The unsuspecting bidder already has placed his bid, if he is not out-bidden by a higher price he will pay his maximum price, and gain no profit from the auction. The potential payoff has been squeezed from him by the seller, so that the buyer makes zero profits, instead of gaining the difference between the second-highest bid and his threshold (from [76]).
- Unmasking: Bidding as long as someone else is the high bidder.
- Bid shielding: When a bidder puts a false high bid (which is then retracted) to reveal the current highest value or to deter other bidders from competing. Then the high bid can be retracted allowing the bidder to win the auction at a low price.

Some papers show that sniping is rational in common value auctions. Wilcox [74] argues that potential buyers try to get additional information from other bidders, especially from experts, those who frequently place bids on similar items. The potential buyers observe the bids of experts as an indication for the market price of the auctioned goods. In response to this behavior experts place their bids late.
Roth and Ockenfels [73] have observed similar behavior in auctions of antiques. But this behavior has not been observed with other auctioned goods. It seems that the expert bidders will place their bids late on goods where the market value is hard to determine.
It is plausible to assume that the bidders try to protect themselves against the incremental bidding strategies of others. This seems to be a reason for sniping and late bidding. Ariely *et al.* [77] found evidence for incremental bidding behavior in second-price internet auctions.

A theoretical explanation for incremental bidding is provided by Roth and Ockenfels [75]. They argue that by late bidding one may face the risk of not being able to bid at all, due to technological problems e.g. internet connection and network delay uncertainties. They also show that in case of only one bidder, the expected payoff of an early bid exceeds that of a late bid. They conclude that there exists no dominant strategy (incremental bidding is not dominated strategically), however, late bidding can be the best response to incremental bidding strategy.
Another group of papers consider sniping as the best response to shill bidding, which has solely the purpose of inflating the final price.
Wang *et al.* [78] discuss under which circumstances shilling would be a utility-maximizing strategy. Chakraborty and Kosmopoulou [79] show that in a common value auction it is only the auctioneer who could gain from shilling activities. They show that shill bidding reduces the surplus of the bidders and conclude that the auctioneer has an incentive to encourage shilling.
Barbaro and Bracht [76] demonstrate that sniping is a dominant strategy if one takes into account additional rules of online auctions. They considered the possibilities of retracting a potential buyer's bid or canceling a bid by the seller themselves. They called this strategy squeezing. They suggested that the best response to the squeezing strategy is the late bidding, because squeezing involves time. They show that sniping is part of a Bayesian-Nash equilibrium.

4.3. Revenue equivalence theorem

There are many interesting results found by theoretical analysis of auctions. One important result is given here. It is important that one takes both the auction format (English, second-price, Dutch, etc.) and the valuation environment (common values, private values) into consideration. For example it is found that under independent private values, English and Vickrey auctions are theoretically equivalent.
The seller's revenue from English, Vickery, first-price and Dutch auctions with private values[1] is found to be the same. This leads to the "Revenue equivalence theorem", which states that all auction formats that award the item to the highest bidder and lead to the same bidder participation will generate on average the same revenue for the seller [80].

[1]There exist some additional assumptions.

5. Data Collection and Empirical Results

In this chapter we explain the method used to collect the data from eBay.de. Furthermore we give results of the statistical analysis of this data such as important relations between different variables and their distributions. For the statistical analysis we use some standard software like SPSS and GNU Plot, but to have more detailed and effective results for our specific problems and questions, we had to develop our own individual tools. These tools are developed in C++ by using several compilers such as GNU g++, GNU gcc and MS-Visual C++ under both SUN and MS-Windows platforms.
As mentioned in chaper 2, the study of complex systems has to deal with high variability data related to phenomena where large events and small events both occur. Normally large events occur rarely. In contrast small events occur frequently and are common. This property is captured in ubiquitous power-law distributions.
Statistical analysis of such high variability data is, however, tricky. Low frequency of large events causes a poor statistics in the tail of histograms, making them noisy and difficult to work with. For example it is very difficult to fit a curve to such noisy histograms in order to find the probability density function (pdf). A common way of solving this problem is to use logarithmic binning, which has the disadvantage of loosing information. The better way is to plot the data by calculating the cumulative density functions (cdf) instead of plotting the relative frequencies (and trying to find the probability density functions directly). In our work we use mainly cumulative frequency histograms and fit curves to find cdf of distributions. Knowing the cdf it is possible to find the pdf. It is simply the derivative of the cdf. There exist also simple relations between exponents of both functions. We will discuss this in sec. 5.1.
In this work we use the terms heavy-tailed, scaling and power-law distributions interchangeable.

5.1. Density functions

5.1.1. Probability density function

Having a set of random variables x, e.g. time series or any statistical study like exam scores, one of the most common statistical analysis is to find the probability distribution of x. Probability distributions are typically defined in terms of the probability density function (pdf) or the cumulative distribution function (cdf).
The pdf $p(x)$, is defined such that the probability of finding x in an interval Δx around x is given by $p(x)\Delta x$. $p(x)$ is by definition non-negative and normalized

$$\int_{x_{\min}}^{x_{\max}} p(x)\,\mathrm{d}x = 1, \tag{5.1}$$

where $x_{\min}$ is the smallest and $x_{\max}$ the largest possible value for x.
The probability density function is also called probability distribution.

5.1.2. Cumulative density function

The cumulative density function (cdf), $P(x)$, describes the probability that x takes on a value greater than or equal to x:

$$P(x) = \int_{x}^{\infty} p(x')\,\mathrm{d}x'. \tag{5.2}$$

The cdf of a distribution with a power-law pdf, $p(x) = Cx^{-\mu}$, also follows a power law, with an exponent $\beta = \mu - 1$:

$$P(x) = C\int_{x}^{\infty} x'^{-\mu}\,\mathrm{d}x' = \frac{C}{\mu-1}x^{-(\mu-1)} = C'x^{-\beta}. \tag{5.3}$$

The cdf of a distribution with an exponential pdf, $p(x) = Ce^{-\alpha x}$, is also an exponential function with the same parameter α:

$$P(x) = C\int_{x}^{\infty} e^{-\alpha x'}\,\mathrm{d}x' = \frac{C}{\alpha}e^{-\alpha x} = C'e^{-\alpha x}. \tag{5.4}$$

5.2. Data collection

To collect the data we used different self developed software and scripts, which automate monitoring eBay auctions through internet interfaces (HTTP protocol). The entire information about eBay transactions are well protocolled and saved in eBay databases. This information is however not publicly available. The method we used is to our knowledge, despite of its weakness, the only one, one can use to collect this kind of data. All results achieved by analysing/studying the data gathered in this fashion could be more precise if one has the opportunity of using eBay databases directly.
The response of eBay to any HTTP request is in the form of HTML files. To mine out particular patterns of content and extracte formatted data from unformatted HTML files we use the screen scraping[1] technique and advantages of regular expressions provided by many programming languages.
The screen scraping is an inelegant method of consuming data from a web page, depending on a consistent format of the web page. Nevertheless the huge data mining possibility this technique provides has turn it into a science and many leading companies like Microsoft have built it in their web services products.

[1]Screen scraping is the act of parsing the HTML in generated web pages by using programs designed to mine out particular patterns of content.

The programming language we use is PERL due to high flexibility of regular expressions and comfortable HTML parsing techniques.
The extracted formatted data was stored in a MS-Access database into several tables.
Because many of the information we are interested in, are available just after an auction is ended, the auction web-pages should be visited both before and after ending of the auction (note that an ended auction is not listed or linked on eBay web-pages directly). There are some serious problems linked to the use of the screen-scraping method, for example by calculating the time intervals between events, one should take the summer-winter clock change into consideration. This is not done automatically by standard packages known to us (the format of time used by eBay.de is the common "dd.mm.yy hh:mm:ss"format). To gain dynamic information, like current price of an item at a specified time or the feedback score of a seller at the time a given item was running by him, from the raw data, one should find reverse engineering methods by using the same algorithms used by eBay. There exist in most cases no clear explanation of those algorithms and one has to test several algorithms to find the right one.
To minimize errors in the process of data extraction and filter out wrong data from the database, one needs access to the information originally published on the web-pages, to control suspicious items, e.g. those items showing extreme behavior.
So we needed to capture the whole published information of a number of items. This information includes plain text, graphics, java-scripts and all other data formats, which are integrable in HTML files. To gain this data we used the Microsoft "MHT"web-archive single file format by using MS-ADO and Visual C++.
eBay's answer to each automated HTTP-request takes about 2 seconds. With about 1 million auctions ending every day, it is impossible to collect all the transaction information by using the method explained above. So one has the problem of collecting a good subset of data, which captures many different features and includes as many as possible cases happening in eBay. To solve this problem we decided to work with two different sets of data. One of them should be gathered from as many different categories as possible, with large variety in final prices (see Fig. 5.1). For this purpose we decided to gather all auctioned items listed by eBay as the result to a widely used search-query. The other data set should include all transactions in one or more categories over a large period of time. We decided to monitor categories "Web Projects"and "websites & domains"because of the variety of offers and the possibility of studying both common value and private value strategies (see sec. 3.1). In these categories one finds both standard offers, like web spaces, with a common value and also rare valuable offers (projects or domain names) without any standard or common value.
In Fig. 5.1 we show the cumulative probability distribution of final prices of all auctions. The final price varies between 1 Euro and 10000 Euro, but the number of auctions ending with a price higher than 1000 Euro is not big enough to offer a good statistical analysis, i.e. the results of our study are applicable to auctions ending with a final price below 1000 Euro.
Many of our results are found to be valid for both different data sets. These data sets have the following properties:

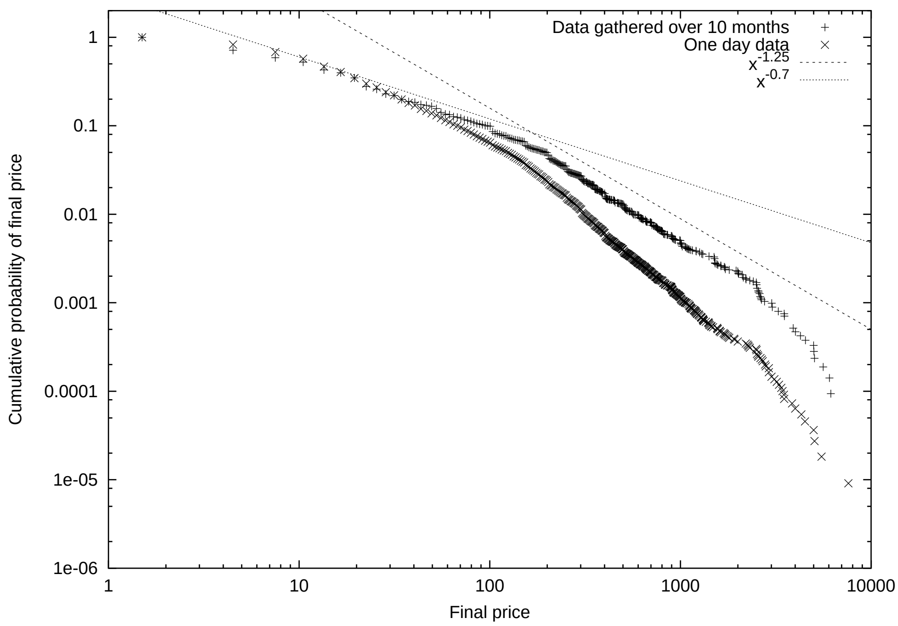

Figure 5.1.: Cumulative probability distribution of final price for both data sets. Power-law lines are guides for the eye and have slopes of -1.25 and -0.7.

i. **DB-1:** a data set collected from auctions existing on March 22, 2004 on eBay.de. We focus on auctions with the label "OVP"[2] in the title, indicating a new product, and find 173,315 auctioned items, grouped in 9904 subcategories by eBay. 262,508 distinct agents bidding on items and 43,500 sellers offering auctioned items are identified.

ii. **DB-2:** all auctions in subcategories "web projects"[3] and "websites & domains"[4] are collected over 10 months, involving 11,145 agents that bid on 52,373 items (some of these items are listed also in a second category).

5.3. Empirical results

As mentioned, distributions of many variables and also functionalities of expectation values of some variables on other variables are measured to follow simple and common functions like exponentials and power laws. In this section we give some of our most important empirical findings. In the next chapter we will give our interpretations and explanations of some of these relations.

[2] Originalverpackung
[3] Computer—Domainnamen— Web-Projekte
[4] Business & Industrie—Geschäfts- & Firmenverkäufe—Websites & Domains

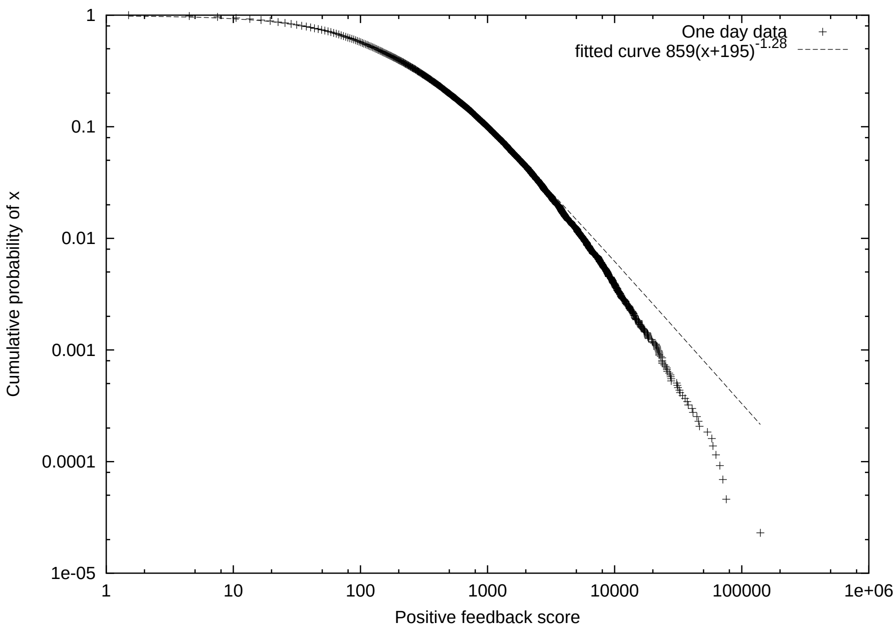

Figure 5.2.: Distribution of the positive feedback scores of sellers. Cumulative probability distribution of the positive feedback scores of each seller is plotted. The data set DB-1 is used. The solid line corresponds to a fit function of the form $f(x) \propto (x+c)^{-\mu}$ with $\mu = -1.28$ and $c = 195$.

5.3.1. Distribution of positive feedback scores of sellers

The distribution of the positive feedback scores has a power-law tail. We have studied this distribution for both data sets and found similar results. This kind of distribution is known from scale-free networks, for instance similar distributions are found when studying the total number of internet web-pages linked to a specified web-page (see Table 2.1 and [45]).
We use a modified scaling law to describe these distributions. A fit function of the form $f(x) \propto (x+c)^{-\mu}$ is used. For the exponent μ we find values $\mu = -1.28$ and $\mu = -1.42$ for the one day gathered data set, DB-1 (Fig. 5.2) and the data set gathered over 10 months, DB-2 (Fig. 5.3) respectively. This indicates that the exponents of pdfs of the positive feedback scores are -2.28 and -2.42 for DB-1 and DB-2 respectively.

5.3.2. Distribution of positive feedback scores of bidders

The distribution of positive feedback scores of bidders has also a power-law tail. This distribution is studied by using a subset of 126152 randomly chosen agents of the data set DB-1. Fig. 5.4 shows the cumulative distribution of positive feedback scores. A power law $f(x) \propto (x)^{-2.2}$ is plotted as a guide for eyes.

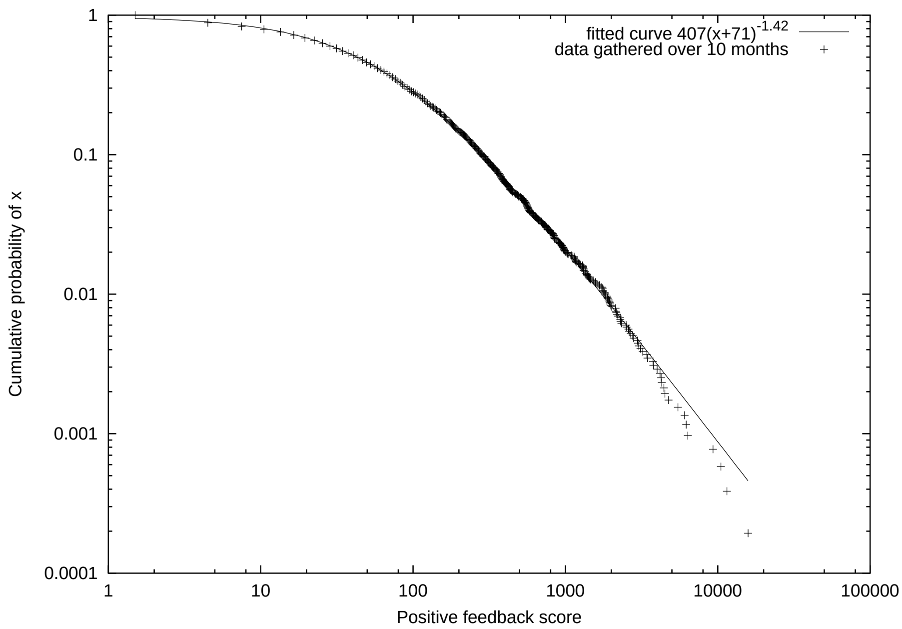

Figure 5.3.: Distribution of positive feedback scores of sellers. Cumulative probability distribution of positive feedback scores of each seller is plotted. The data set DB-2 is used. The solid line corresponds to a fit function of the form $f(x) \propto (x+c)^{-\mu}$ with $\mu = -1.42$ and $c = 71$.

5.3.3. Distribution of agents bidding on a certain item

The probability distribution of distinct number of agents n_{agent} simultaneously bidding on a certain item is exponential and is given by:

$$P(n) \propto \exp(-n/n_0), \tag{5.5}$$

where $n_0 = 2.9$.
This is in agreement with the behavior found in a previous study [81] for eBay.com and the Korean eBay, where the authors found $n_0 = 2.5$ for eBay.com and $n_0 = 7.4$ for Korean eBay.
In Fig.5.5 the histogram of number of agents n_{agent} simultaneously bidding on a certain item is plotted.

5.3.4. Distribution of total number of bids placed on one item

The probability distribution of the total number of bids n_{bids} received for an item is exponential and is given by Eq. (5.5) where $n_0 = 6.5$.
This is also in agreement with the behavior found on eBay.com and Korean eBay [81], where the authors found $n_0 = 5.6$ for eBay.com and $n_0 = 10.8$ for Korean eBay.

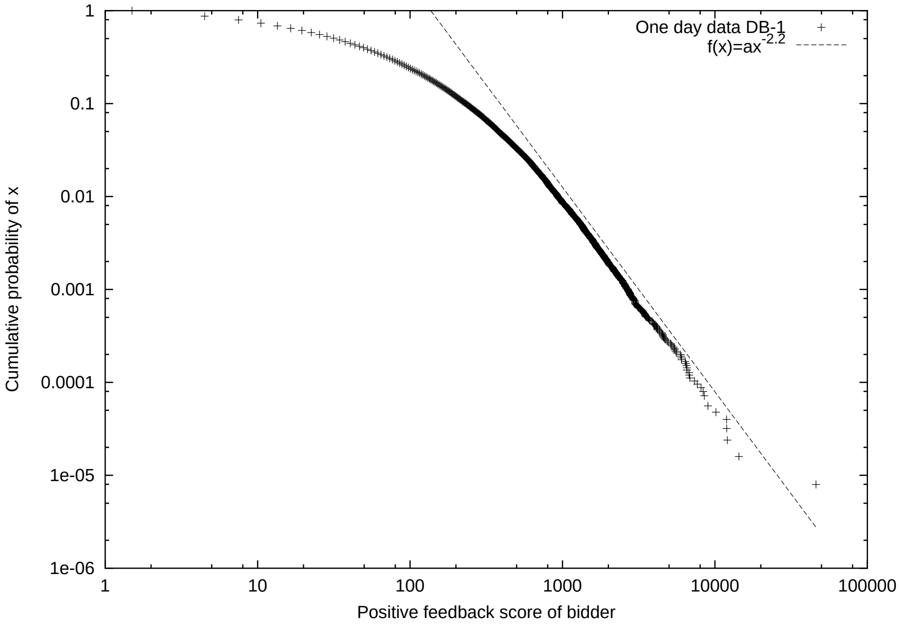

Figure 5.4.: Distribution of positive feedback scores of bidders. Cumulative probability distribution of positive feedback scores of each bidder is plotted. A subset of 126152 randomly chosen agents of the data set DB-1 is used. The solid line corresponds to the power law $f(x) \propto (x)^{-2.2}$.

In Fig. 5.6 the cumulative probability distribution $P(n_{\rm bids})$ of bids $n_{\rm bids}$ received for an item is plotted.

5.3.5. Total activity of individual agents as bidder or seller

We found that activities of individual agents as bidder or seller follow power-law distributions. The total number of bids placed and the total number of auctions offered by agents are studied by using the data set gathered over 10 months (DB-2). One can find that the probability distribution of the total number of bids placed by the same agent, $n_{\rm bids}$, follows a power law:

$$P(n_{\rm bids}) \propto n_{\rm bids}^{-\gamma}, \tag{5.6}$$

where $\gamma = 1.9$ (Fig. 5.7).
In Fig. 5.7 both cumulative probability and probability distributions of $n_{\rm bids}$ are shown.
The probability distribution of total number of auctions offered by the same agent, denoted by $n_{\rm auct}$, is also characterized by a similar power law:

$$P(n_{\rm auct}) \propto n_{\rm auct}^{-\beta}, \tag{5.7}$$

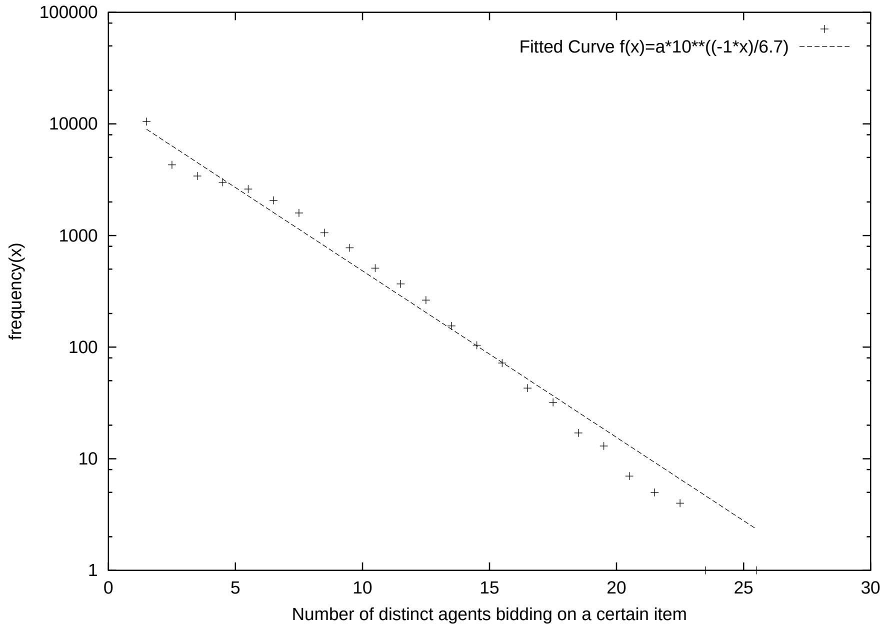

Figure 5.5.: Histogram of agents. Histogram of number of agents n_{agent} simultaneously bidding on an item is plotted.

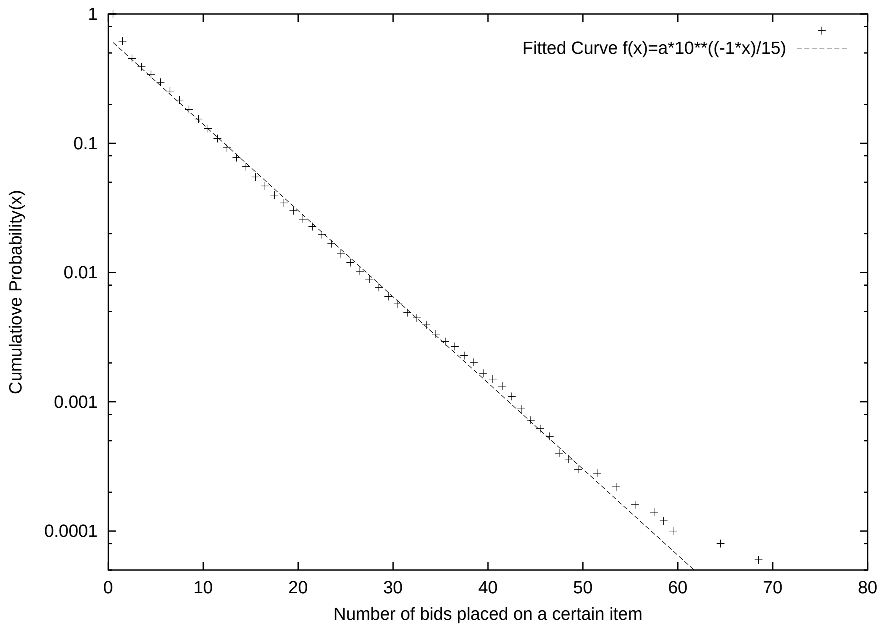

Figure 5.6.: Distribution of bids. Cumulative probability distribution of number of bids n_{bids} received for an item is plotted.

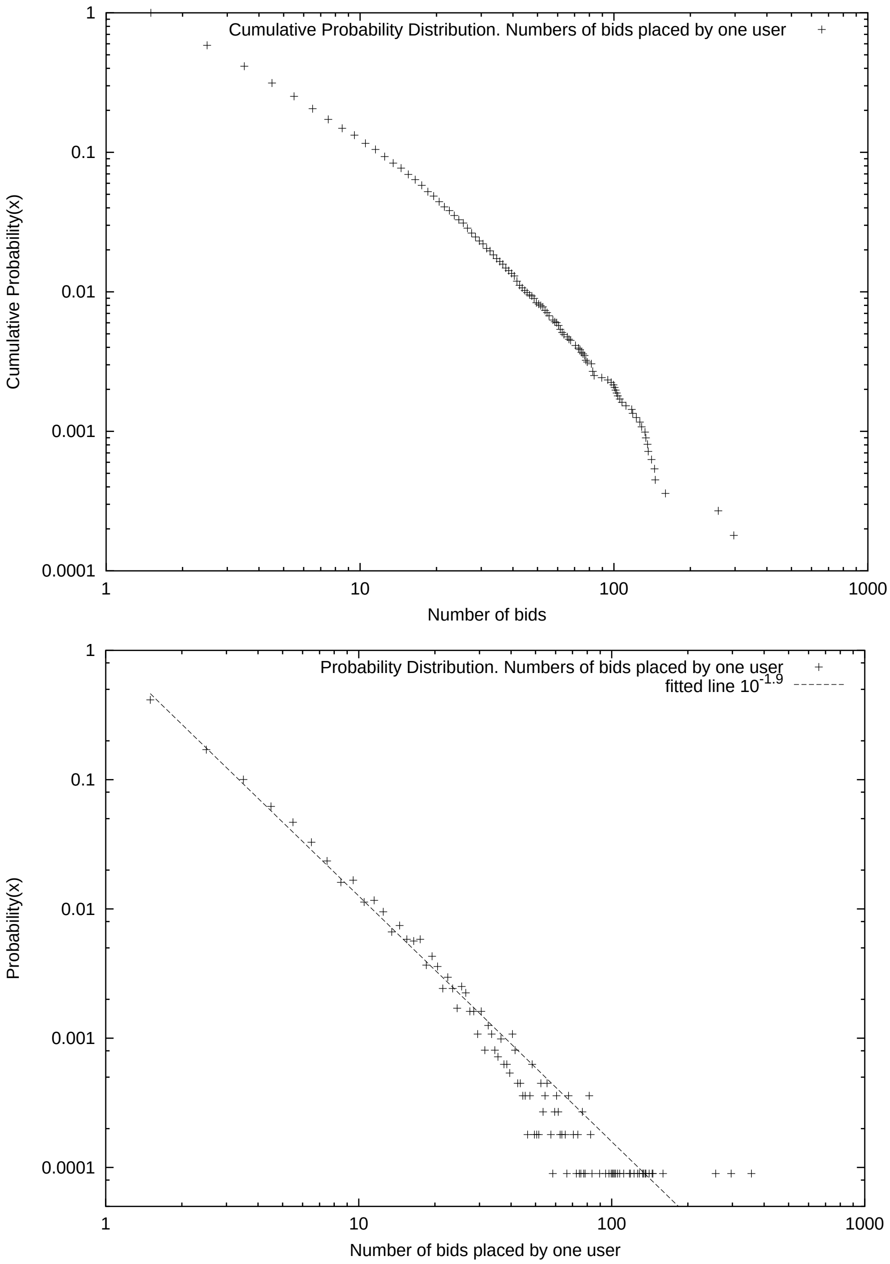

Figure 5.7.: Distribution of the total number of bids placed by the same agent follows a power law. Cumulative probability and probability distributions of the total number of bids n_{bid} placed by the same agent are plotted.

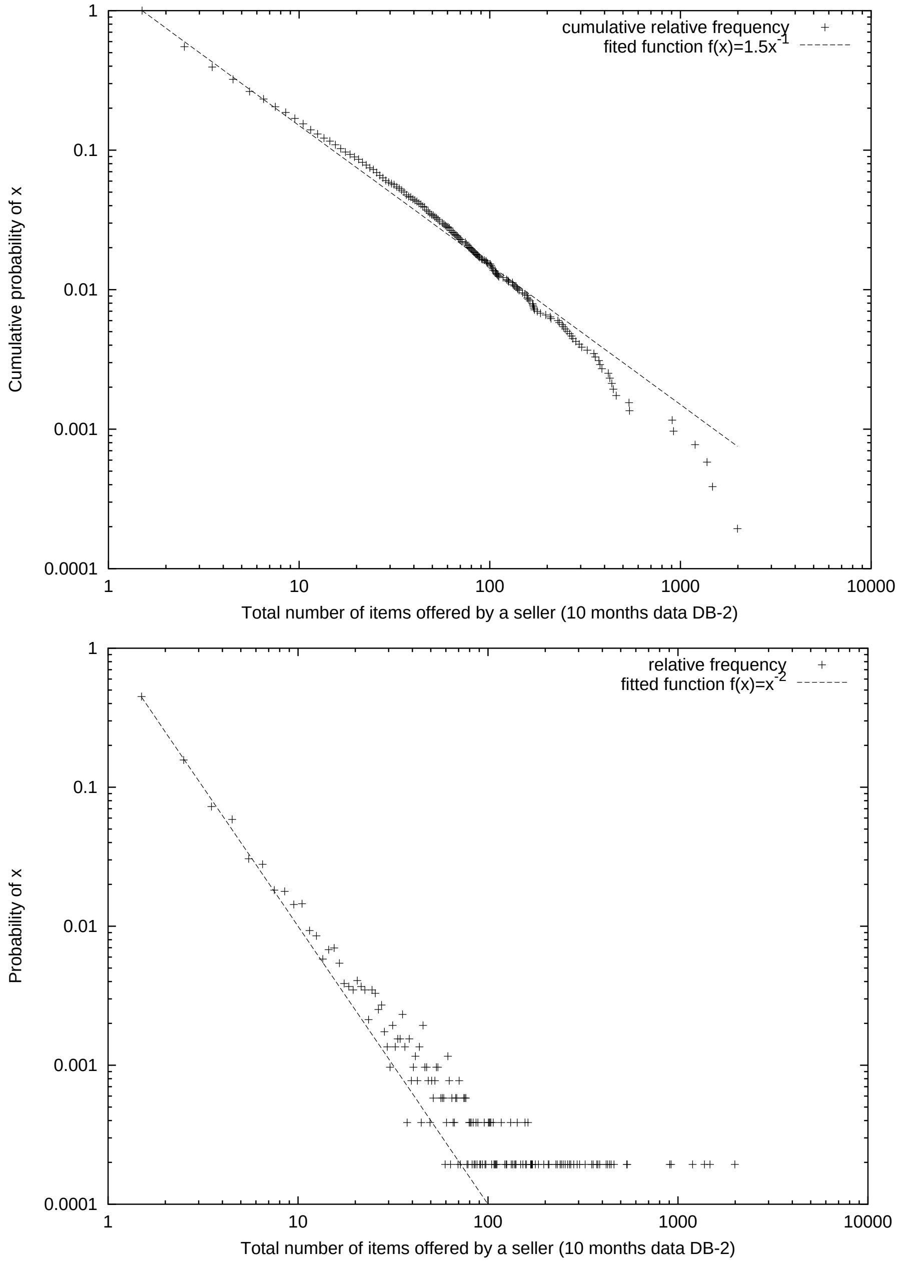

Figure 5.8.: Distribution of the total number of different auctions offered by the same agent follows a power law. Cumulative probability and probability distributions of the total number of different auctions, n_{auct}, offered by the same agent are plotted.

where $\beta = 2$ (Fig. 5.8).
In Fig. 5.8 both cumulative probability and probability distributions of n_{auct} are shown. This is an important result and the value of $\beta = 2$ is understandable. We will refer to and discuss this result in the sec. 6.1.
To find the pdf of this distribution a fit function of the form $f(x) = \alpha x^{\beta}$ is used. The parameters are found to have values $\alpha = 1.00487 \pm 0.0062$ and $\beta = -1.99832 \pm 0.01056$.

5.3.6. Distribution of returns

We define the dimensionless variable return, ϱ, as the relative increase of the submitted bid, b:

$$\varrho = (b - p_{\mathrm{auct}})/p_{\mathrm{auct}}, \tag{5.8}$$

where p_{auct} is the current or listed price just before the bid is placed.
There are several reasons why the definition of returns does make sense. It is common, that the human perception or response to a physical stimulus increases with the relative changes of the stimulus. This functionality is known as the Weber-Fechner law [82]. The Weber-Fechner law states that in order that the intensity of a sensation may achieve an arithmetic progression, the stimulus itself must achieve a geometric progression. Examples are brightness, sound intensity and pitch.
The distribution of ϱ is found to follow a power law for almost 3 orders of magnitude with exponent -2.44 (Fig. 5.9). Similar surprising distributions are also found by statistical analyzing of music pieces, focusing on the distribution of pitch appearances [83].

5.3.7. Correlation of returns and bids

Although the bids are correlated, the returns show a very short range correlation. By computing the correlation function $c_{\mathrm{ij}} = \langle \varrho_{\mathrm{i}} \varrho_{\mathrm{j}} \rangle - \langle \varrho_{\mathrm{i}} \rangle \langle \varrho_{\mathrm{j}} \rangle$ of the returns (indices denote the chronological order of arriving bids; averaging is done over all auctions) one finds that c_{ij} has non-vanishing values just for $i = j$ (Fig. 5.10). For comparison we have computed the correlation function of bids $k_{\mathrm{ij}} = \langle b_{\mathrm{i}} b_{\mathrm{j}} \rangle - \langle b_{\mathrm{i}} \rangle \langle b_{\mathrm{j}} \rangle$ too. The non-vanishing correlation is expected because of the drift in the listed price and the minimum accepted bid. We found an interesting behavior, k_{ij} is not monotonically decreasing. This behavior could not be understood and needs more studies (Fig. 5.11).

5.3.8. Distribution of bid submission times

By analyzing the collected data we found that bidders prefer to bid close to auction ending times. Fig. 5.12 shows the cumulative probability distribution of bid submission times as a function of the time remaining until the end of the auction. Two regimes with exponential behavior can be observed related to the most common auction lengths of 7 and 10 days. Both parts are well described by $P(\Delta t) \propto \exp(-\Delta t/T_{\mathrm{o}})$ with $T_{\mathrm{o}} = 68.94$ (Fig. 5.12).
Close to the end of the auctions sniping (see sec. 4.2) leads to a power-law distribution $P(\Delta t) \propto (\Delta t)^{-\gamma}$ with $\gamma = 1.1$ (Fig. 5.13).

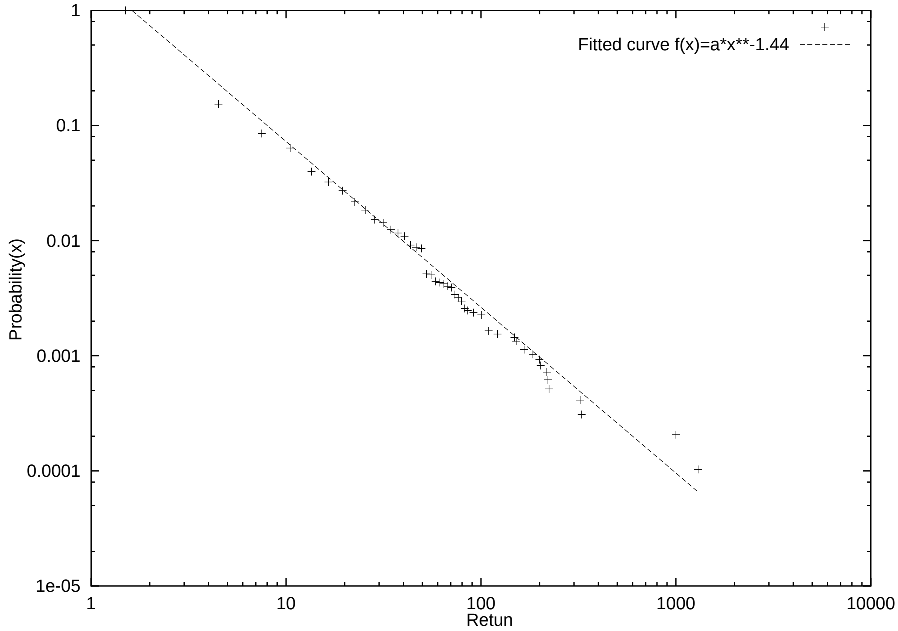

Figure 5.9.: Distribution of returns defined as $\varrho = (b - p_{\mathrm{auct}})/p_{\mathrm{auct}}$ follows a power law. The cumulative probability distribution of ϱ is plotted.

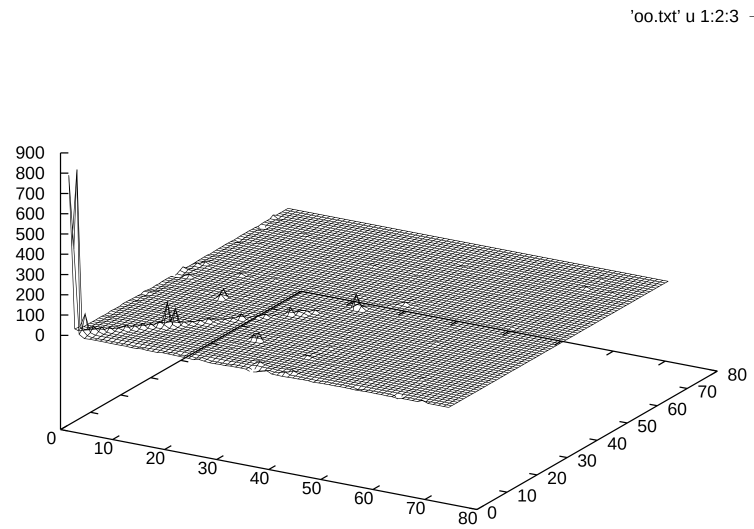

Figure 5.10.: Correlation function of returns $c_{\mathrm{ij}} = \left\langle \varrho_{\mathrm{i}}\varrho_{\mathrm{j}} \right\rangle - \left\langle \varrho_{\mathrm{i}} \right\rangle \left\langle \varrho_{\mathrm{j}} \right\rangle$. The flat landscape is expected and indicates a short-range correlation.

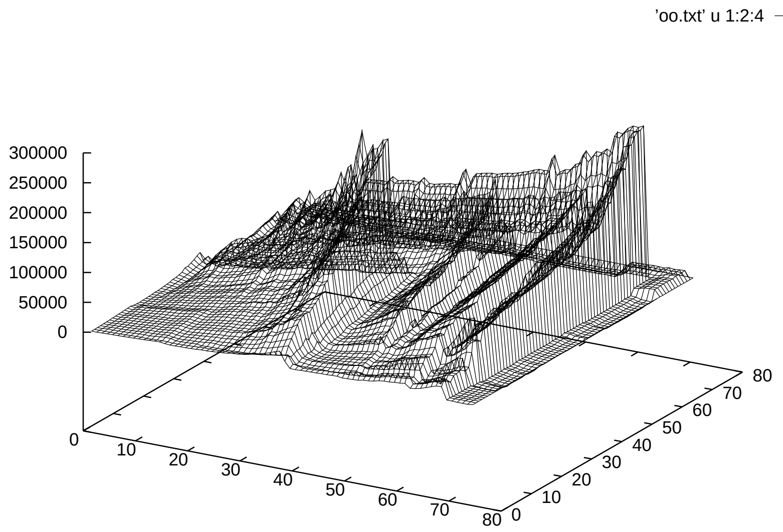

Figure 5.11.: Correlation function of bids, $k_{\rm ij} = \left\langle b_{\rm i} b_{\rm j} \right\rangle - \left\langle b_{\rm i} \right\rangle \left\langle b_{\rm j} \right\rangle$. The non-vanishing correlation is expected because of the drift in the listed price and the minimum accepted bid. The non-monotonic decreasing behavior of $k_{\rm ij}$ could not be understood and needs more studies.

5.3.9. Relation between price and number of bids

Each bid increases the listed price of an auctioned item (for an exception of the case, if a high bidder places a new bid. See sec. 3.4.2). So an interesting questions is: How does the price increases in average when a new bid arrives?

There are several studies describing the influence of static parameters like the ending time (which day of the week, on which daytime), start price, etc. on the final price, less is known about relations between dynamic parameters.

We looked for the functionality of the price on other dynamic parameters (such as number of bids) and studied the relation between price $p_{\rm auct}$ and number of bids $n_{\rm bid}$ for each closed auction. i.e. we look for the relation between the final price and the total number of bids placed of an item. Fig. 5.14 shows the result for the data set DB-2.

The mean achieved price $\langle p_{\rm auct} \rangle$ for auctioned items with the same number of bids placed on them ($n_{\rm bid}$) is found to increase in a power law with the number of bids.

This kind of functionality for conditional expectation values is rather common in systems with very nonlinear dynamics or with stochastic multiplicative amplification effects, which are discussed in sec. 2.6.

This result seems to be plausible. Properties like $p_{\rm auct}$, $n_{\rm bid}$ and bidding time $t_{\rm bid}$ are connected and to find the expectation values, one needs a large number of conditional

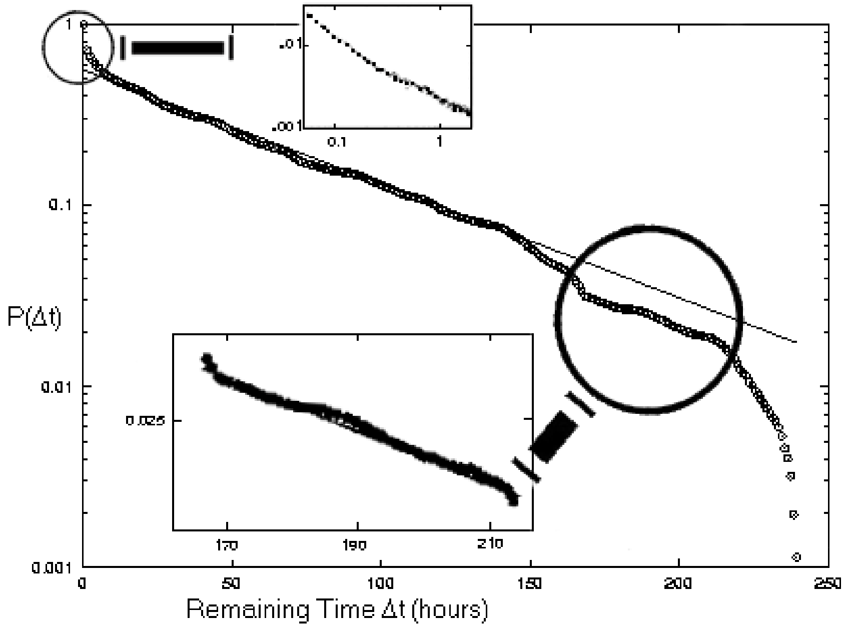

Figure 5.12.: Distribution of bid submission times. The graph shows the cumulative probability distribution of bid submission times as a function of the time remaining until the end of the auction. The Insets show the distribution of very small remaining times and for remaining time more than 7 days.

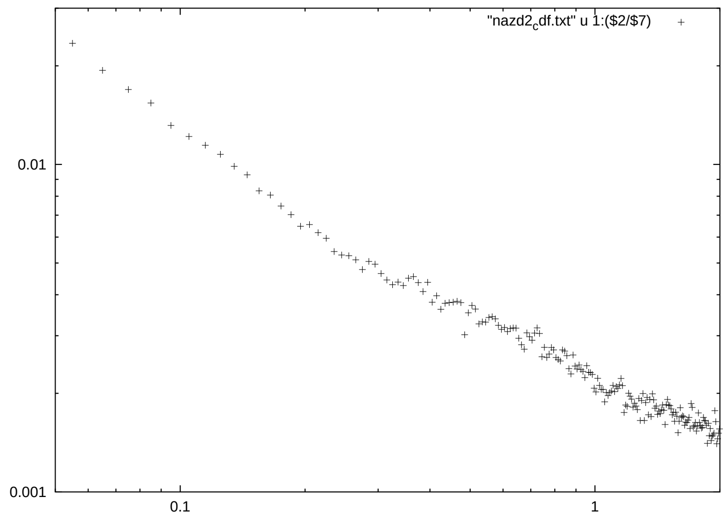

Figure 5.13.: Cumulative probability distribution of bid submission times close to the end of the auctions, where sniping leads to a power-law distribution $P(\Delta t) \propto (\Delta t)^{-\gamma}$ with $\gamma = 1.1$.

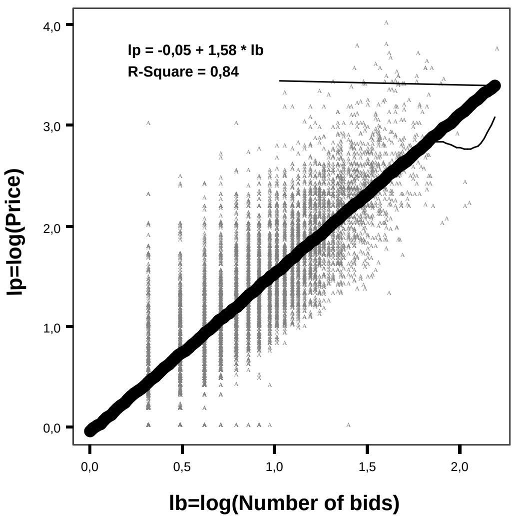

Figure 5.14.: Relation between final price and total number of bids n_{bid}. Scatter plot and linear regression (see Appendix A.3) of $\log(p_{\text{auct}})$ versus $\log(n_{\text{bid}})$ is shown. The data set DB-2 is used.

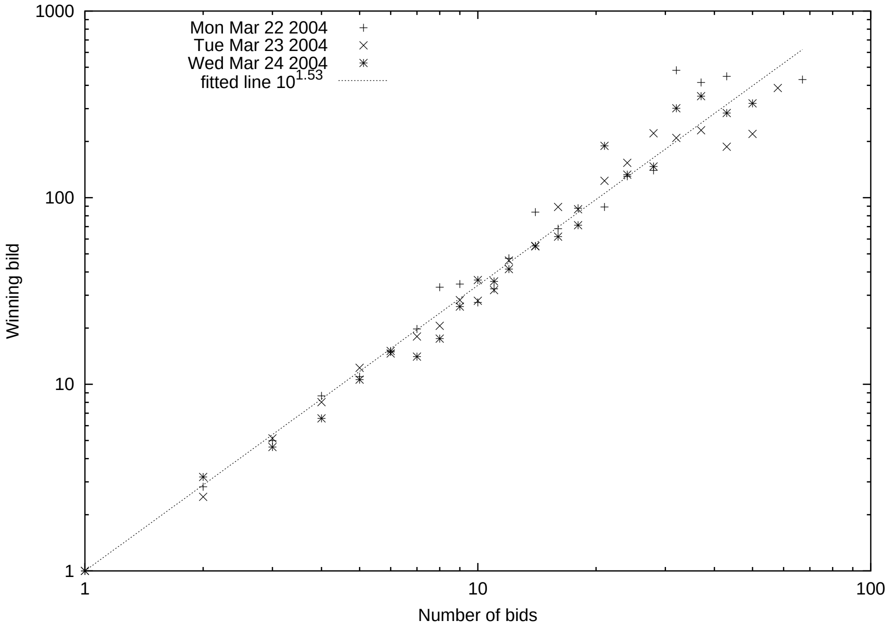

Figure 5.15.: Relation between final price $\langle p_{\text{auct}} \rangle$ and number of bids n_{bid}. Subsets of the data set DB-1 is used. One-day-collected data, collected on different days of week are used.

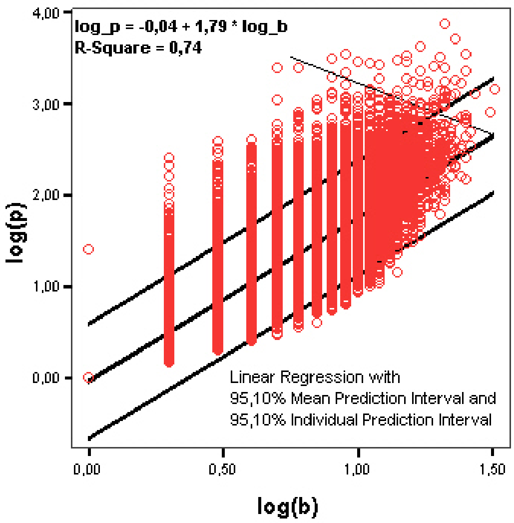

Figure 5.16.: Relation between final price and distinct number of bidders. Scatter plot and linear regression of $\log(p_{\text{auct}})$ versus $\log(n_{\text{dist-bidder}})$ is shown.

probability functions, but if no scale characterize the system then we expect that for example the relation between $\langle p_{\text{auct}} \rangle$ and n_{bid} to be described by a scale-free relationship, that is a power law as shown in sec. 2.5.1.
The relation between $\langle p_{\text{auct}} \rangle$ and n_{bid} is shown in Fig. 5.14 and Fig. 5.15. A power-law relation can be seen

$$\langle p_{\text{auct}} \rangle \propto n_{\text{bid}}^{\alpha}, \tag{5.9}$$

where $\alpha = 1.58$ for the data set DB-2 (Fig. 5.14) and $\alpha = 1.53$ for the data set DB-1 (Fig. 5.15).
This functionality seems to be universal and not depending on subcategories or time intervals one uses to collect the data. As shown, this functionality is examined both with data collected in large time intervals with focusing on selected subcategories (DB-2 Fig. 5.14) and with just one-day-collected data on a very large domain of subcategories, collected on different days of week (DB-1).
The minimum starting price on ebay.de is 1 Euro. An interesting open question is if this functionality is different on ebay.com or ebay.uk, where the minimum starting prices are lower than 1 (independent of currency).
Another interesting question is the relation between price and the number of distinct bidders. To find this relation we have analyzed all auctions starting with 1 Euro and ending after at least one bid. The data set DB-1 is used. We found 520449 bids placed on 59305 auctions. The mean and standard deviation of the number of distinct bidders per auction

are found to be 5.141 and 3.22 respectively. The relation between $\langle p_{\mathrm{auct}} \rangle$ and $n_{\mathrm{dist-bidder}}$ is again a power law:

$$\langle p_{\mathrm{auct}} \rangle \propto n^{\alpha}_{\mathrm{dist-bidder}}, \tag{5.10}$$

where $\alpha = 1.79$ for the data set DB-1.
The relation between p_{auct} and $n_{\mathrm{dist-bidder}}$ is shown in Fig. 5.16.

6. Interpretation of Empirical Results

In this chapter we give our interpretation and explanations of the behavior found in online auctions. Some of the empirical findings provided in chapter 5 could be understood or even predicted quantitatively by means of existing theories/mechanisms and stochastic processes discussed in sec. 2.6. Other interesting results could be understood only qualitatively so far.

We will present our interpretations and compare our results, where possible, with other similar systems and processes. In some cases it is important to have a better knowledge of the network architecture of eBay as a community, where several types of interactions can take place. We will discuss these features and our ideas whenever they are necessary for a better understanding. Some terms and definitions given in chapters 3, 4 are used and referred in this chapter.

6.1. Number of items offered by sellers

As presented in sec. 5.3.5, the total number of items offered by the same agent follows a power-law distribution:

$$P(n_{\mathrm{auct}}) \propto n_{\mathrm{auct}}^{-\beta}, \tag{6.1}$$

where $\beta = 2$ (Fig. 5.8).

This empirically found result could be explained by using the Yule process (discussed in sec. 2.6.1).

Consider at time t there exists $n_{\mathrm{x,t}}$ agents who have offered exactly x items up for sale until this time. We are interested in the distribution of x, i.e. we would like to know the probability that an agent has offered exactly x items in a given time span. Let us measure the time such that at each time step one agent becomes a seller by offering up a single item for sale. This is a natural assumption because it is common that agents who start to act as a seller in eBay should first learn about offering items in practice.

Suppose in addition to the item offered by the new seller, j new items are offered at each time step by people who have previously offered other items. So the total number of offered items up to time t is given by

$$n_{\mathrm{t}} = tj + t. \tag{6.2}$$

It is plausible to assume that agents who offered more items in the past are more likely to offer new items. So we assume that these j new items are offered by sellers in proportion to the number of items they have already offered, so $P_{\mathrm{i}} \propto x_{\mathrm{i}}$, where P_{i} denotes the probability that the seller i with total number of offered items x_{i} offers a new item up for sale.

Let $P_{\mathrm{x,t}}$ denote the probability of finding a seller with exactly x offered items up to the time t. So we will have $n_{\mathrm{x,t}} = tP_{\mathrm{x,t}}$ and the total number of sellers with x offers which offer a new item in time t will be:

$$j\frac{x}{n_{\mathrm{t}}}n_{\mathrm{x,t}} = \frac{jx}{j+1}P_{\mathrm{x,t}} \tag{6.3}$$

Note that the total number of sellers at time t, denoted by n, is selected to be t, i.e. $n = t$. By using the Eq. (6.3) we can write the master equation as following:

$$(t+1)P_{\mathrm{x,t+1}} = tP_{\mathrm{x,t}} + \frac{j}{j+1}((x-1)P_{\mathrm{x-1,t}} - xP_{\mathrm{x,t}}), \qquad \text{for } x > 1. \tag{6.4}$$

and

$$(t+1)P_{\mathrm{1,t+1}} = tP_{\mathrm{1,t}} + 1 - \frac{j}{j+1}P_{\mathrm{1,t}}, \qquad \text{for } x = 1. \tag{6.5}$$

We are interested in the stationary state, where $P_{\mathrm{x,t}}$ does not depend on t anymore. So we solve these equations in the limit $t \to \infty$, and use $P_{\mathrm{x}} = \lim_{t\to\infty} P_{\mathrm{x,t}}$.
Solving Eq. (6.5) one gains

$$P_1 = \frac{j+1}{2j+1}, \tag{6.6}$$

and Eq. (6.4) becomes

$$P_x = \frac{j}{j+1}\big[(x-1)P_{x-1} - xP_x\big], \tag{6.7}$$

and we will have

$$P_x = \frac{x-1}{x+1+1/j}\,P_{x-1}. \tag{6.8}$$

This equation can be iterated to get

$$\begin{aligned} P_x &= \frac{(x-1)(x-2)\dots 1}{(x+1+1/j)(x+1/j)\dots(3+1/j)}\,P_1 \\ &= (1+1/j)\frac{(x-1)\dots 1}{(x+1+1/j)\dots(2+1/j)}. \end{aligned} \tag{6.9}$$

One can simplify this equation by making use of the property $\Gamma(a) = (a-1)\Gamma(a-1)$ of the Γ-function:

$$P_x = (1+1/j)\frac{\Gamma(x)\Gamma(2+1/j)}{\Gamma(x+2+1/j)}, \tag{6.10}$$

and by using the notation of Legendre beta-function $\beta(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ we obtain

$$P_{\mathrm{x}} = (1+\frac{1}{j})\mathrm{B}(x, 2+1/j). \tag{6.11}$$

Since the beta-function has a power-law tail $\mathrm{B}(a,b) \propto a^{-b}$, P_{x} can be written as

$$P_{\mathrm{x}} \propto x^{-\alpha}, \tag{6.12}$$

where the exponent is given by $\alpha = 2 + 1/j$.
As mentioned we use the data set which includes all auctioned items in two different subcategories over 10 months. The other data set is not suitable for this study because it is just a snapshot of an extreme dynamical system and can provide no information about activity of agents over a large time span directly. The only parameter which contains information about the history of agents activities and could be used for similar studies is the total number of feedback scores. But this variable (gathered with the method we have used) can not be divided into scores gained from sell and scores gained from buy activities. We measured the total number of offered items to be $n_t = 52373$ and the total number of distinct sellers to be $n = 5165$ (when using DB-2). By using Eq. (6.2) we find

$$j = \frac{n_t}{n} - 1 = 9.14. \tag{6.13}$$

The exponent of the pdf (Eq. (6.12)) is predicted to be $\alpha = 2.1094$ and the P_1 is predicted to be $\frac{j+1}{2j+1} = 0.526$.
The actual exponent for the distribution and P_1 are empirically found to be 2 and 0.48 respectively (see Fig. 5.8) which are both in good agreement with the theoretical predictions. For the other data set (DB-1) we find $n_t = 173315$ and $n = 43501$. So we can calculate j to be 2.98 and $\alpha = 2.33$. These results are not compared with the data we have collected because the number of distinct sellers is huge and the database SQL-queries are very time-consuming.
The difference between the two predicted exponents, related to different data sets ($\alpha_{\mathrm{DB-1}} = 2.1094$ and $\alpha_{\mathrm{DB-2}} = 2.33$), can be understood when one takes into consideration, that in subcategories like "web-projects" there exists a clear difference between seller and buyer communities, i.e. there exists a number of sellers, which use eBay as a market place. They offer their products and in many cases the same or similar product for several times. On the other hand there exist a community of buyers, who are interested in products offered by the sellers in these subcategories, but never offer an item up for sale in the same subcategory. This argumentation is not valid if one take a large variety of subcategories into consideration because an agent, which has act as a buyer once, could act as a seller in some quite different class of categories. For instance an agent, who buys a computer for his personal use (and not in order to trade with it) can offer an old book or his car up for sale in eBay. We believe this is the reason why the relative distinct number of sellers are larger for the data set including many different subcategories (DB-1 includes 9904 different subcategories), and the reason why the predicted exponents are different for these two data sets.
The observed power-law and explanation given above suggest that the offered items in selected subcategories are dominated by a few number of power-sellers, who offer new items proportional to their previous offered items.

6.2. Number of positive feedback scores

After the end of an auction the participating buyer and the seller (transaction partners) are encouraged to leave feedback comments on their profiles by eBay. This policy provides a nice information for other users and can be used by future activities of the seller or the buyer. The method used in the previous section predicts the distribution of the total number of items offered by a seller in a large period of time. If after any transaction all sellers would receive a positive feedback, the distribution of the positive feedback scores should also be the same as the distribution of the total number of offers. Our statistical analysis shows a different result. Both these distributions follow power laws but the exponents are different. For the data set DB-2 these distributions are given by

$$P(n_{\mathrm{auct}}) \propto n_{\mathrm{auct}}^{-\beta}, \tag{6.14}$$

where $\beta = 2$ (Fig. 5.8).
for the total number of offered items and

$$P(n_{\mathrm{auct}}) \propto n_{\mathrm{auct}}^{-\gamma}, \tag{6.15}$$

where $\gamma = 2.42$ (Fig. 5.3).
for the positive feedback scores.
This can be understood when one considers that the probability of receiving a positive feedback is smaller than one and decreases with increasing number of items a seller offers. This seems to be plausible for great values of n_{auct}, where the seller should use automated methods of working out emails (as the standard method, eBay uses for sending information to sellers and buyers) and sending his sold items, which would increase the probability that buyers are unsatisfied. In contrast, it is also plausible to assume that the power-sellers are more familiar with the problems a seller may have and should be able to have satisfied customers (in average).
To verify these assumptions and find the reason why these two distributions have different exponents one needs to know more about the relation between number of offers and number of negative and neutral feedbacks. This is not done yet.

6.3. Bid submission times

As we have shown in sec. 5.3.8 the distribution of bid submission times as a function of the remaining time until the end of the auction is exponential except for the final seconds and is well described by $P(\Delta t) \propto \exp(-\Delta t/T_0)$ with $T_0 = 68.94$ (Fig. 5.12). This is subject of discussion of many studies based on game theoretical approaches [73–76]. These studies, in contrast to our quantitative description, include more or less just qualitative discussions, where people search for rational strategies leading to a nonuniform distribution for bid submission times. As mentioned and discussed in sec 4.1 in a Vickrey auction[1] (note that eBay auctions can be viewed as Vickrey auctions) there is no reason for bidders to submit

[1] Second-price auction.

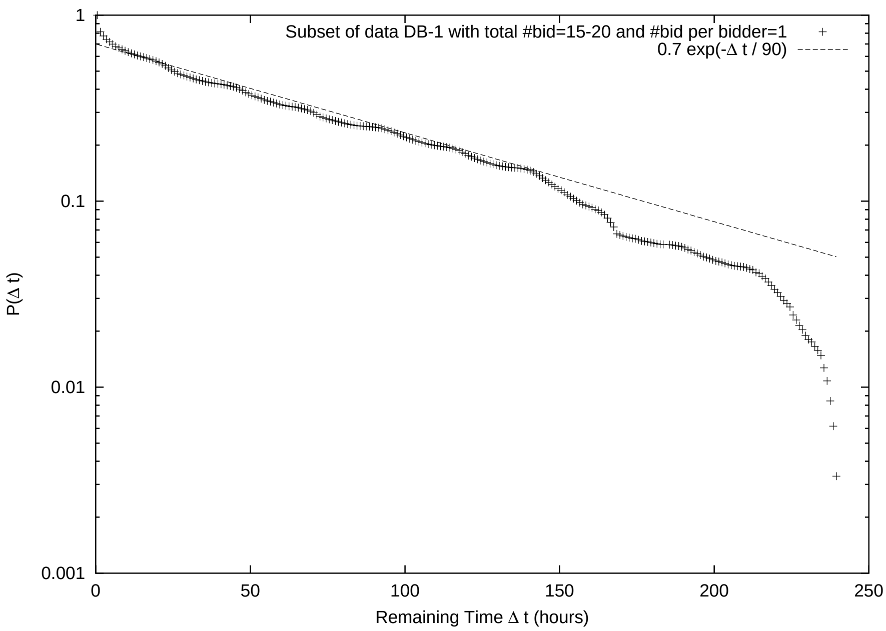

Figure 6.1.: Distribution of bid submission times. The graph shows the cumulative probability distribution of bid submission times for a subset of the data set DB-1 as a function of the time remaining until the end of the auction. The distribution hast the form $P(\Delta t) \propto \exp(-\Delta t/T_0)$ with $T_0 = 90$. Comparison with Fig. 5.12 shows the same functional form for both distributions.

bids less than their true values (values they are willing to pay) and also there exists no reason to wait and submit bids later. But recent studies [76], [75] show that there are reasons why a bidder should bid later, for example to avoid bidding-wars, being outbidden, being victim of shill bidding, being squeezed, etc. The idea is that there exist other rational strategies which lead to gain better payoffs. These strategies suggest a preferred time-span for bid submissions. Descriptions of several ideas are given in an overview in sec. 4.2.
The study of strategies and their impact on other observable parameters is somehow complicated. There exists no exact definition and conditions under which one distinguishes between two or more strategies, i.e. having the bid history, the exact identification of strategies bidders have used is not easy and not accurate and rather impossible by means of automatic methods.
Consider the following three strategies: sniping, late bidding and evaluation (see sec. 4.2 for the definition). They are all characterized by the number of bids the bidder places, which is exactly one. However having the information, how many bids a bidder places on an auction, is not enough for the identification of the strategy. Sniping strategy needs, for example, submission of exactly one bid in the last 3 or 5 seconds ($\Delta t < 3$ or $\Delta t < 5$). This one bid should be the winning bid, i.e. the return, as defined in Eq. (5.8) and explained in sec. 5.3.6 should be small (is equal to the increment divided by the price).

N = Number of cases	Δt	N /(Total number of events)
182	≤ 5 (sec)	0.019
425	≤ 10 (sec)	0.025
1387	≤ 1 (min)	0.082
2185	≤ 10 (min)	0.130
3140	≤ 1 (hour)	0.186
5379	$\frac{\Delta t}{T} \geq 0.5$	0.319

Table 6.1.: Number of cases depending on the remaining time for the subset of data DB-1. All auctions starting with 1 Euro, ending with the total number of bids between 15-20 are studied. Bid-events of bidders who have bidden just once on an item are studied. T is the duration of the auction.

We are interested to know which strategies are used and how frequently they are used (how are strategies weighted). To find out the weight of different strategies and to examine if the distributions of bid submission times is related to the choice of strategy, we analyzed a subset of the data set DB-1 with the following properties:

- To avoid the unknown effect of parameters on the choice of strategy, we decided to work with a subset of data including items with similar parameters like the total number of bids placed on them and starting price. We select all auctions with starting price of 1 Euro, which have ended with a total number of bids between 15 and 20. We find the total number of distinct bidders to be 45575 and the total number of bids to be 90923. The reason for the choice of total number of bids between 15 and 20 is that if the total number of bids is too small some strategies might not have been used. On the other hand, if we select items with a large total number of bids we will not have enough data for a good statistical analysis.

- We are first interested in the simplest strategies. They all share the same property of having exactly one bid. These strategies are as named above sniping, late bidding and evaluation. To identify these strategies we focus on bid-events of bidders who have bidden just once on an item. We find 16855 such events. The number of cases depending on Δt is given it Table 6.1.

To identify a strategy we need the bid submission time. The eBay auctions have different durations of 1,3,5,7 and 10 days and it is better to work with $\tau = \frac{\Delta t}{T}$ instead of ΔT, where T denotes the duration of the auction.
Using this subset of data, one should expect a different distribution for bid submission times (different from the distribution found for the whole data set), because strategies like squeezing, shill bidding, unmasking and warrior strategies are filtered out. This is however not the case. It is a surprising result that the distribution of bid submission times for this subset of data is the same (approximately) as the one measured by using the whole database. In Fig. 6.1 the distribution of bid submission times is shown. This distribution has also the form $P(\Delta t) \propto \exp(-\Delta t/T_0)$ with $T_0 = 90$.

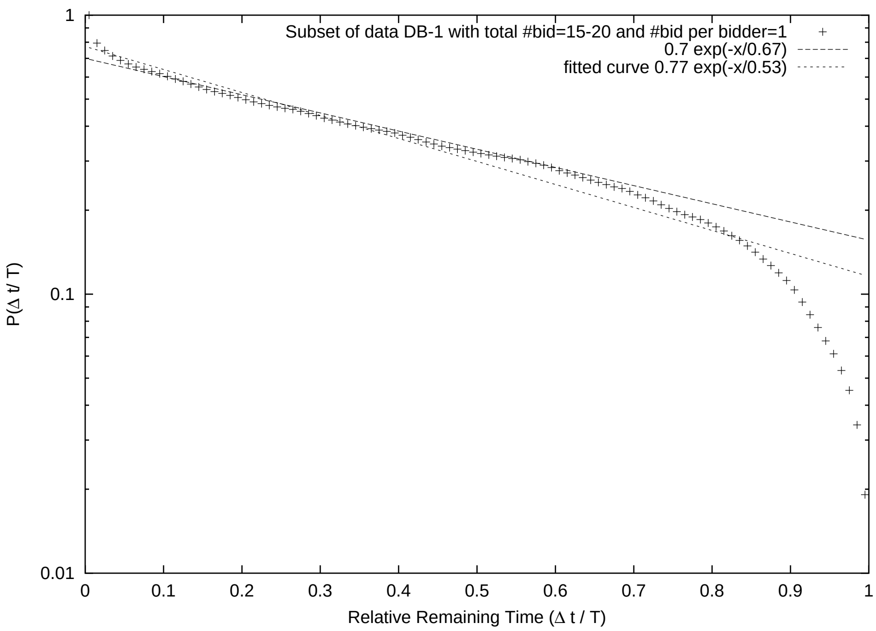

Figure 6.2.: Distribution of relative bid submission times. The graph shows the cumulative probability distribution of relative submission times for a subset of the data set DB-1 as a function of the relative time remaining until the end of the auction $\tau = \frac{\Delta t}{T}$. The distribution has the form $P(\tau) \propto \exp(-\tau/\tau_0)$ with $\tau_0 = 0.53$.

In Fig. 6.2 we show the distribution of $\frac{\Delta t}{T}$ for this subset of data. The distribution has the form $P(\tau) \propto \exp(-\tau/\tau_0)$ with $\tau_0 = 0.53$. The constants T_0 and τ_0 are related. The mean value of auction duration of this subset of data is 6.59 days. So $\frac{T_0}{6.59 \times 24} = 0.56$, which is very close to the value of τ_0.

These results suggests that the exponential nature of the distribution of the bid submission times is not to be understood and explained only by means of the chosen strategy of bidders and has it roots in other properties of eBay. One of these properties is the listing method eBay uses. The items are listed by using the so called "ending soonest" criterion by default (if a user has not changed the standard sort criterion). This means, the auctions closer to their end are listed first. Normally there are 50 items listed on each page by eBay (to see the remaining items one has to click the "Next" navigator-link in the bottom of the page). So the auctions not close to their end are found only after several clicks and can not be seen directly. This means, the probability of being seen for an item, increases with the decreasing remaining auction time. We do not have any suggestion which functional form this probability should have, but if the pdf of this probability has an exponential form, the distribution of bid submission times could be explained without considering alternative rational strategies.

Other sort possibilities are as called by eBay the following, "newly listed", "lowest price first", "highest price first", "closest distance first" and criteria depending on payment pos-

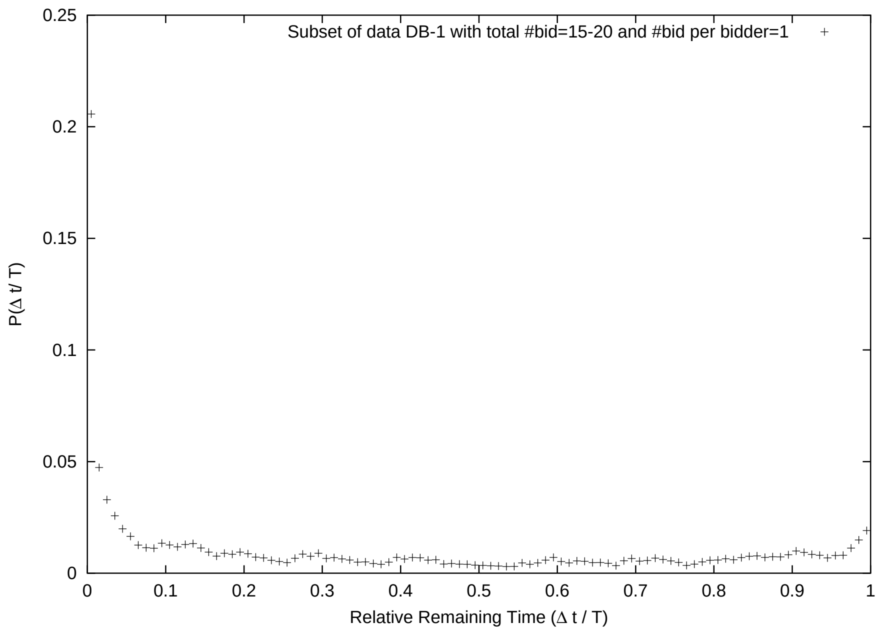

Figure 6.3.: Distribution of bid submission times. The graph shows the probability distribution of relative submission times for a subset of the data set DB-1 as a function of the relative time remaining until the end of the auction $\frac{\Delta t}{T}$. An uniform binning method with size of bins equal to 0.01 is used. The approximately symmetric form is an evidence for the role of sorting criteria provided by eBay and suggests that the exponential behavior of bid submission times is caused partly by the sorting criterion eBay uses for the listing of items.

sibilities.

We can imagine that the "newly listed" criterion is the most used one after the standard "ending soonest" criterion. This assumption can not be verified here because the corresponding data are not available. We have found however evidence which shows the probability of bidding increases not only at the end of the auction, but also at the beginning. The distribution seems to show a very nice symmetric behavior around $\frac{\Delta t}{T} = 0.5$ with the exception of the regime where $\frac{\Delta t}{T} < 0.1$. To show this we have measured the probability $P(\frac{\Delta t}{T})$ by using an uniform binning method with the size of bins equal to 0.01 (and using the subset of data explained above). The result is shown in Fig. 6.3. To observe the mentioned symmetric behavior better, we show the frequency of relative submission times of the same data in Fig. 6.5, in which we use bins of 0.1. The frequency of bids with a relative submission times $\frac{\Delta t}{T} \in [0, 0.1)$ has the value 6682 and is not shown in the graph (see also Fig. 6.4).

If we assume that the choice of the strategy does not depend on the time the item is observed by the bidder, the probability that a bid is placed on it at time t (remaining time Δt) will be the product of two probabilities: the probability that the item is first observed

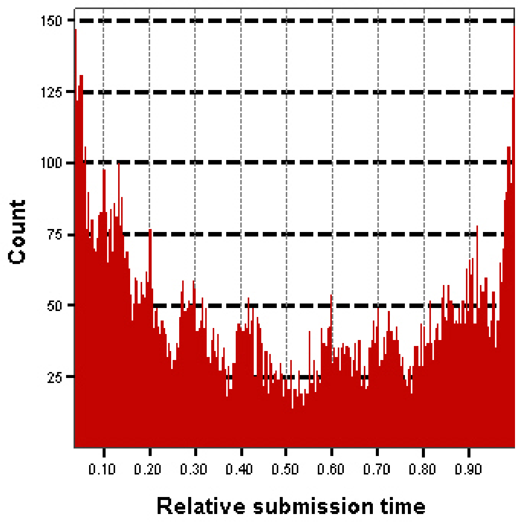

Figure 6.4.: Frequency of bid submission times. The graph shows the frequency of relative submission times for a subset of the data set DB-1 as a function of the relative time remaining until the end of the auction $\frac{\Delta t}{T}$. An uniform binning method with size of bins equal to 0.004 is used. The frequency of bids with a relative submission times $\frac{\Delta t}{T} \in [0, 0.04)$ is not shown in the graph.

by a bidder at this time and the probability that the bidder places a bid on that item after he has seen it:

$$P(\Delta t) = P(S)P(B). \tag{6.16}$$

$P(S)$ denotes the probability of being seen and is a function of Δt. $P(B)$ denotes the probability of placing a bid and depends on the chosen strategy of the bidder.
To reveal the influence of the sort criterion explained above, and focus on the effect of strategies bidders use for bidding, an interesting way is to look at the items listed by using the eBay's "Featured Plus!" option. This option can be selected by the seller and costs 12,95 Euro. The items including this option appear at the top of the listing page and can be seen as long as the auction runs independent of the remaining time of the auction.
eBay describe this option as follows: *"How Featured Plus! works in Search: Search results can be organized by the listing end date, price, or other options that the viewer can choose. There are 50 results per page. Your Featured Plus! item will appear at the top of the page it naturally falls on in the search results list."* [22].
Another reason why the distribution of bid submission times is exponential could be related to the "Watch this item" option provided by eBay. This option gives an interested buyer the possibility of monitoring an auctioned item in his "My eBay" environment. So the bidder can take his time for making his decision without loosing the information about the interested

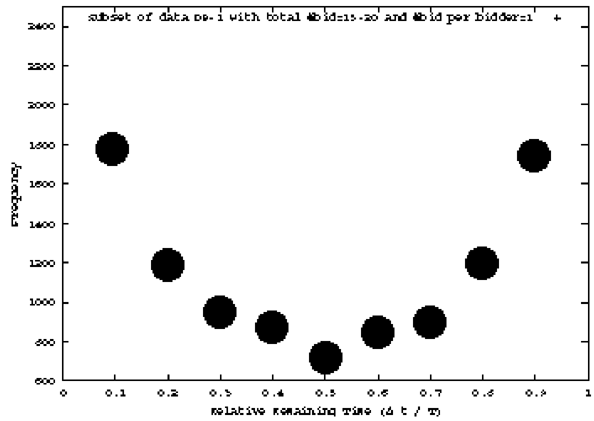

Figure 6.5.: Distribution of bid submission times. The graph shows the frequency of relative submission times for a subset of the data set DB-1 as a function of the relative time remaining until the end of the auction $\frac{\Delta t}{T}$. An uniform binning method with size of bins equal to 0.1 is used. The frequency of bids with a relative submission times $\frac{\Delta t}{T} \in [0, 0.1)$ has the value 6682 and is not shown in the graph. The approximately symmetric form is an evidence for the role of sorting criteria provided by eBay. And suggest that the exponential behavior of bid submission times is caused partly by the sorting criterion eBay uses for the listing of items.

item (like web-page of the item, etc.). Unfortunately, we do not know how many bidders use this option.

It would be interesting to know how many times a bidder visits the item's web-page in average before he places his bid. Unfortunately this kind of information is not to be gained from the listing data.

So far we have shown that the choice of strategy does not explain the exponential behavior of the bid submission times. But this does not mean that the bid submission time is independent of the choice of strategy. To examine this, one should measure the order of relation between "strategies"and bid submission times. This is, as explained above, not easy due to unspecific definition and conditions of strategies. Instead of measuring the relation between strategy and bid submission time one can measure the relation/correlation between return and bid submission time. One expects that towards the end of the auction, sniping and late bidding should cause small values for the returns because the actual bid of the winning bid is not being listed and the winning bid will be the second highest bid plus a small increment ($\varrho \approx \frac{\text{Increment}}{p_{\text{auct}}}$). In contrast early bids, mostly coming from evaluators, should be independent of the actual listed price and should cause greater values for returns (this should be the case at least for auctions beginning with 1 Euro).

We studied this and tried to find the relation between return ϱ and relative bid submission time τ. A scatter plot of the logarithm of returns versus the logarithm of relative submission times τ is shown in Fig. 6.6. The regression line shows a clear dependence

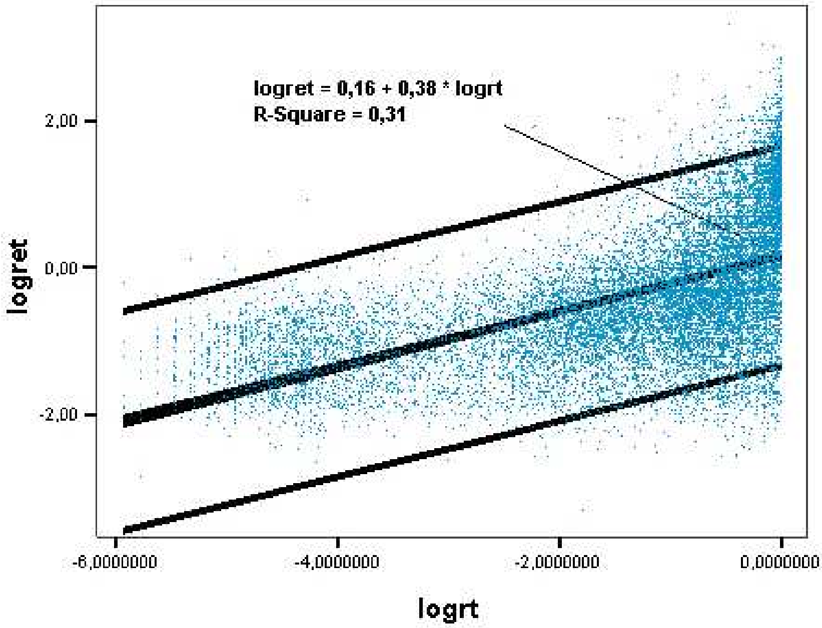

Figure 6.6.: Scatter plot of $\log(\varrho)$ versus $\log(\tau)$ with $\tau = \frac{\Delta t}{T}$. The graph shows the relation between logarithm of returns and logarithm of relative submission times for a subset of the data set DB-1. The regression line shows a clear dependence of both parameters. The form of relation is $\langle \varrho \rangle \propto \tau^{0.38}$. A confidence interval of 95% is used.

of both parameters. The bold lines show a confidence interval of 95%, which is used for determination of the regression. The form of this relation is $\langle \varrho \rangle \propto \tau^{0.38}$. It follows again a power law, similar to the relation between $\langle p_{\mathrm{auct}} \rangle$ and n_{bid} (Fig. 5.14 and Fig. 5.15).

6.4. Achieved price and its relation to the number of bids

One of our most interesting findings is the relation between the final price p_{auct} and number of bids placed n_{bid}. As presented in sec. 5.3.9, this relation between $\langle p_{\mathrm{auct}} \rangle$ and n_{bid} follows a power law:

$$\langle p_{\mathrm{auct}} \rangle \propto n_{\mathrm{bid}}^{\alpha}, \tag{6.17}$$

where $\alpha = 1.58$ for the data set DB-2 (Fig. 5.14) and $\alpha = 1.53$ for DB-1 and its subsets (Fig. 5.15).

Suppose the starting bid is $P_0 = 1$ and suppose we forget about the second maximum rule for the determination of the price. We assume that after each bid the price increases and reaches the amount of the new placed bid. Then $P_1 = (\varrho_1 + 1)P_0$ and

$$P_n = (\varrho_n + 1) \times \ldots \times (\varrho_1 + 1)P_0, \tag{6.18}$$

Number of bids	μ	σ	Number of samples
10	1.5938	0.39061	849
20	2.0931	0.37168	204
30	2.3555	0.39811	64
40	2.4231	0.48846	31
> 30	2.5831	0.46509	683
> 40	2.7566	0.47132	282

Table 6.2.: Statistics of frequency of logarithm of end price for fixed total number of bids for all auctions starting with 1 Euro. The data set DB-2 is used.

and we will have

$$\ln(P_n) = \ln(\varrho_n + 1) + \ldots + \ln(\varrho_1 + 1). \tag{6.19}$$

If the $\ln(\varrho_n+1)$ (for $n = 1, 2, \ldots$) are independent and identically distributed (i.i.d.) random variables with finite mean and variance, the Central Limit Theorem says that $\sum_{k=1}^{n} \ln \varrho_k$ converges to a Gaussian (normal) distribution (sec 2.5.2), and P_n is well approximated by a log-normal distribution.
Although our assumption is not what really happens in eBay, the prediction of Eq. (6.19) is observed (approximately) empirically. In Fig. 6.7 we show the frequencies of logarithm of end prices for fixed total number of bids by using the data set DB-2 and auctions starting with 1 Euro, for the total number of bids equal to 10, 20, 30 and 40. For better statistics we have studied this distribution also for a given interval of the total number of bids. In Fig. 6.8 we show frequencies of the logarithm of end prices for the total number of bids greater than 30 and total number of bids greater than 40. Means and standard deviations are given in Table 6.2.
The distribution functions of the logarithm of prices seem to be similar for both data sets. We have studied this distribution also for the other data set (DB-1). The larger number of data in this database makes it possible to have better fitted curves. The model discussed above makes use of the central limit theorem, which is only valid in the limit where $n \to \infty$. Since we use it for $10 < n < 70$, deviations from the predictions of the central limit theorem can be expected. However, since we watch many realizations of the same stochastic process, these deviations will be reduced because the averaging over many realizations is the same as averaging one realization over a long time.
The Gaussian distribution of the logarithm of end prices seems to be valid also for the other data set (DB-1). In Fig. 6.9 we show frequencies of logarithm of end prices for fixed total number of bids by using the data set DB-1 and auctions starting with 1 Euro, for the total number of bids equal to 10 and 30. The mean and standard deviations found by using this data set are very close to those found by using the data base DB-2.
As mentioned the stochastic process described above is not what happens in reality in eBay. As a matter of fact one should expect different distributions for ϱ_n. The ϱ_n depends on the actual price and its minimum value is determined by the minimum increment, which depends itself on the actual price:

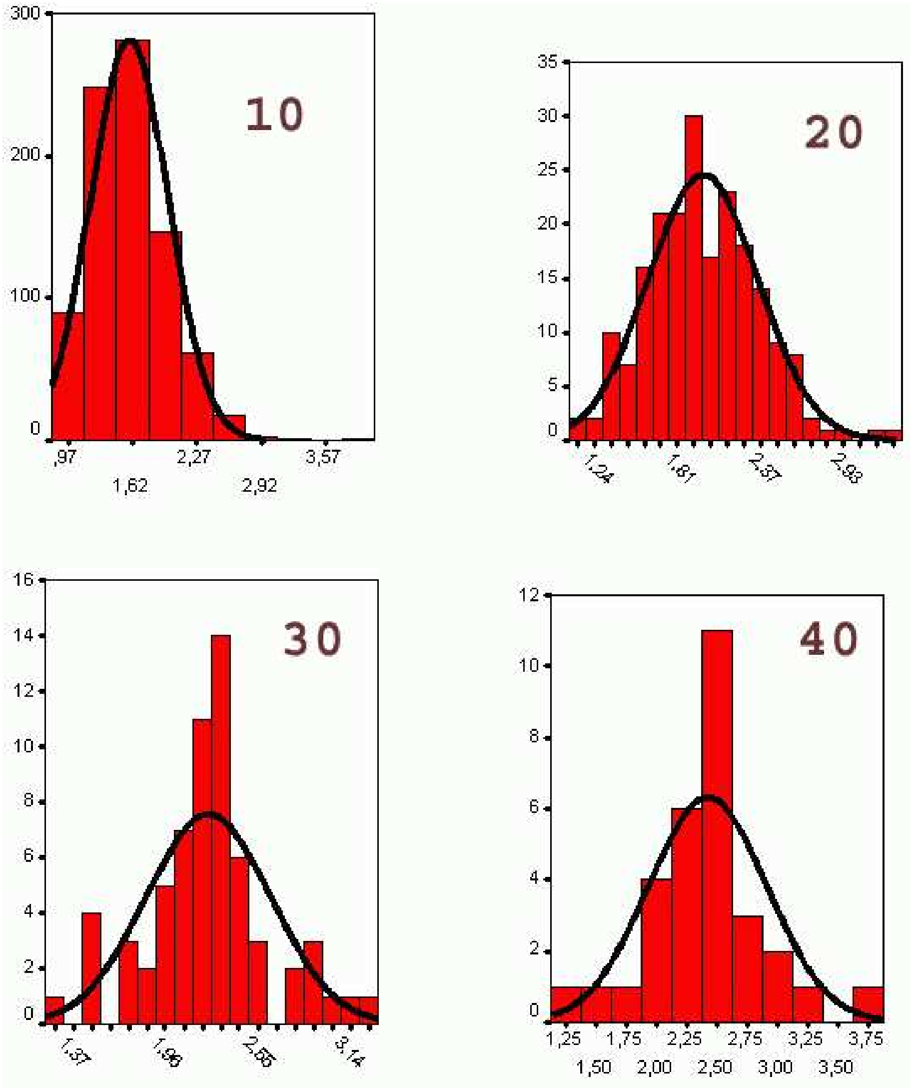

Figure 6.7.: Distribution of the end price for fixed total number of bids by using the data set DB-2. Distribution of the logarithm of end price for all auctions starting with 1 Euro and ending after exactly 10, 20, 30 and 40 bids. Means and standard deviations are given in Table 6.2.

$$\varrho_n^{min} = \frac{F(P_{n-1})}{P_{n-1}}, \tag{6.20}$$

where F is the price dependent bid-increment (given in Table 3.1).
So the expected value for the price by using the stochastic process of Eq. (6.18) will be

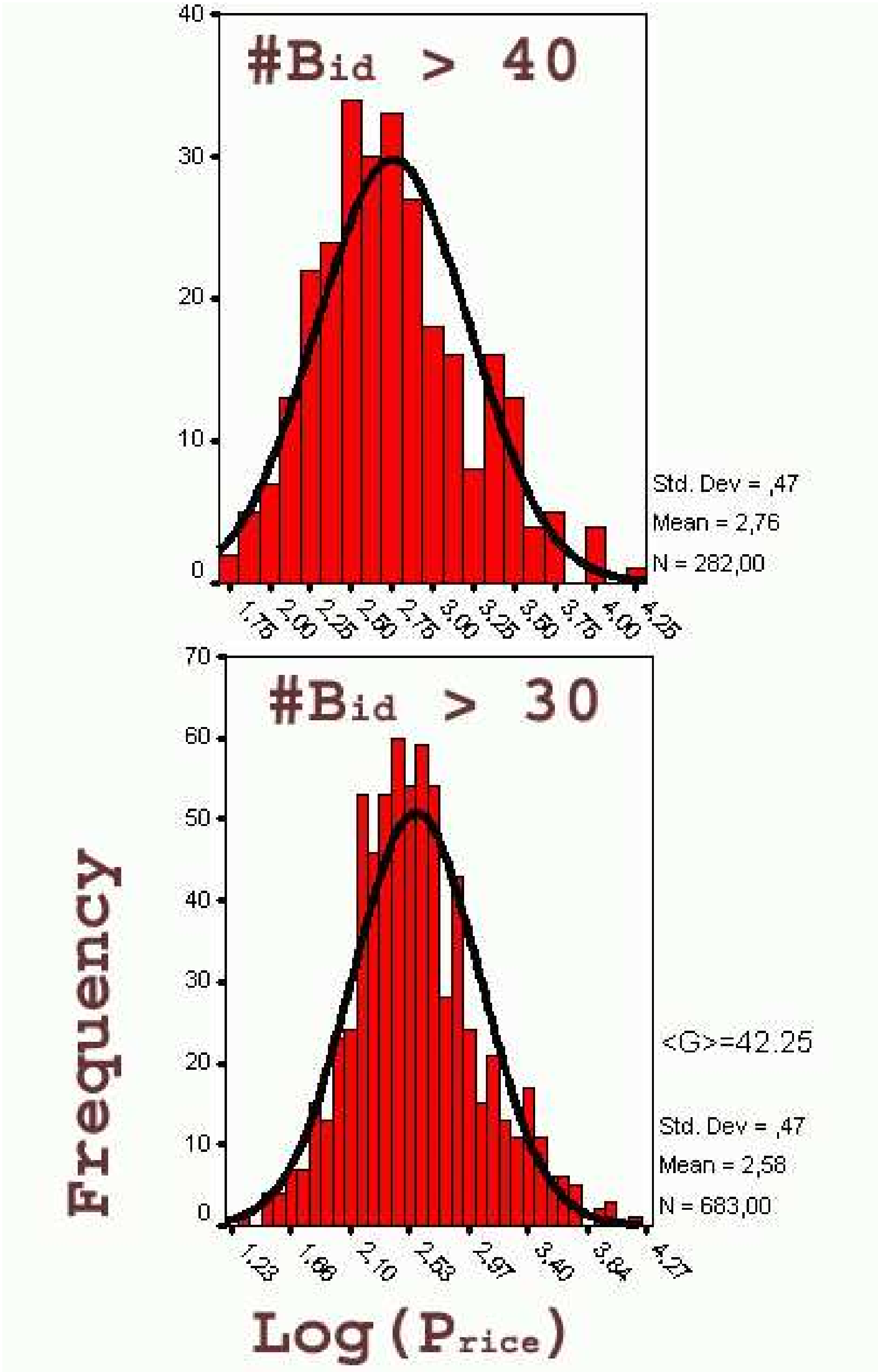

Figure 6.8.: Distribution of the end price for a fixed total number of bids by using the data set DB-2. Distribution of the logarithm of end price for all auctions starting with 1 Euro and ending after minimum 30 bids or minimum 40 bids are plotted. The fitted curves are Gaussian (normal) distributions with μ values of 2.58 and 2.76 respectively and $\sigma = 0.47$.

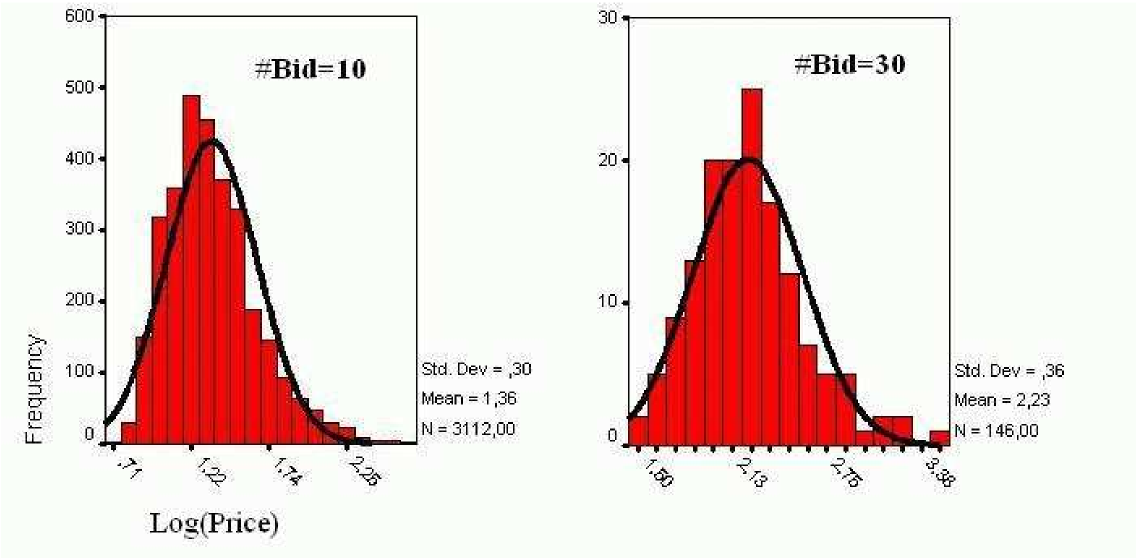

Figure 6.9.: Distribution of the end price for a fixed total number of bids by using the data set DB-1. Distribution of the logarithm of end price for all auctions starting with 1 Euro and ending after exactly 10 and 30 bids. Means and standard deviations are close to those found by using the data base DB-2 (given in Table 6.2).

$$\langle P_n \rangle = \int_{\varrho_1^{min}}^{\varrho_1^{max}} \ldots \int_{\varrho_n^{min}}^{\varrho_n^{max}} P(\varrho_1)(1+\varrho_1)\ldots P(\varrho_n)(1+\varrho_n) \times P_0 \, \mathrm{d}\varrho_1 \ldots \mathrm{d}\varrho_n, \tag{6.21}$$

which is not solvable without assumptions, making it tractable.
The log-normal distribution, as discussed in sec. 2.5.4, has similarities with power laws. Locally these distributions can be mistaken for each other. To find the actual distribution of the end price for items ending with a given number of bids, we started a systematic study. We used another subset of the data set DB-1 including all auctions starting with 1 Euro and ending after n_{bid} bids with $n_{bid} \in (10, 15]$, $n_{bid} \in (20, 25]$, $n_{bid} \in (30, 35]$ and $n_{bid} \in (40, 45]$.
We show the results in Fig. 6.10. All these distributions can be well fitted with a curve of the form $f(x) \propto (x + c)^b$, which has a power-law asymptotic. Values of c and b are given in the graphs.
The cumulative distributions and the corresponding pdfs seem to follow power laws. This is, as mentioned in sec. 2.5.3, to be verified by searching for an universal scaling law. For this purpose we measure the scaled distribution $P(\frac{p_{\mathrm{auct}}}{n_{\mathrm{bid}}^{\gamma}})$ for several values of γ and try to find a value for γ for which all distributions of $\frac{p_{\mathrm{auct}}}{n_{\mathrm{bid}}^{\gamma}}$ are similar. These distributions for γ equal to 0.8, 1.1, 1.5, 2 , 3 and 4 are shown in Fig. 6.11. One finds that for $1.5 \geq \gamma \leq 2$ all distributions are similar, where $\gamma < 1.5$ and $\gamma > 2$ are bad candidates.
So far we have studied the relation between the number of bids and the end price of

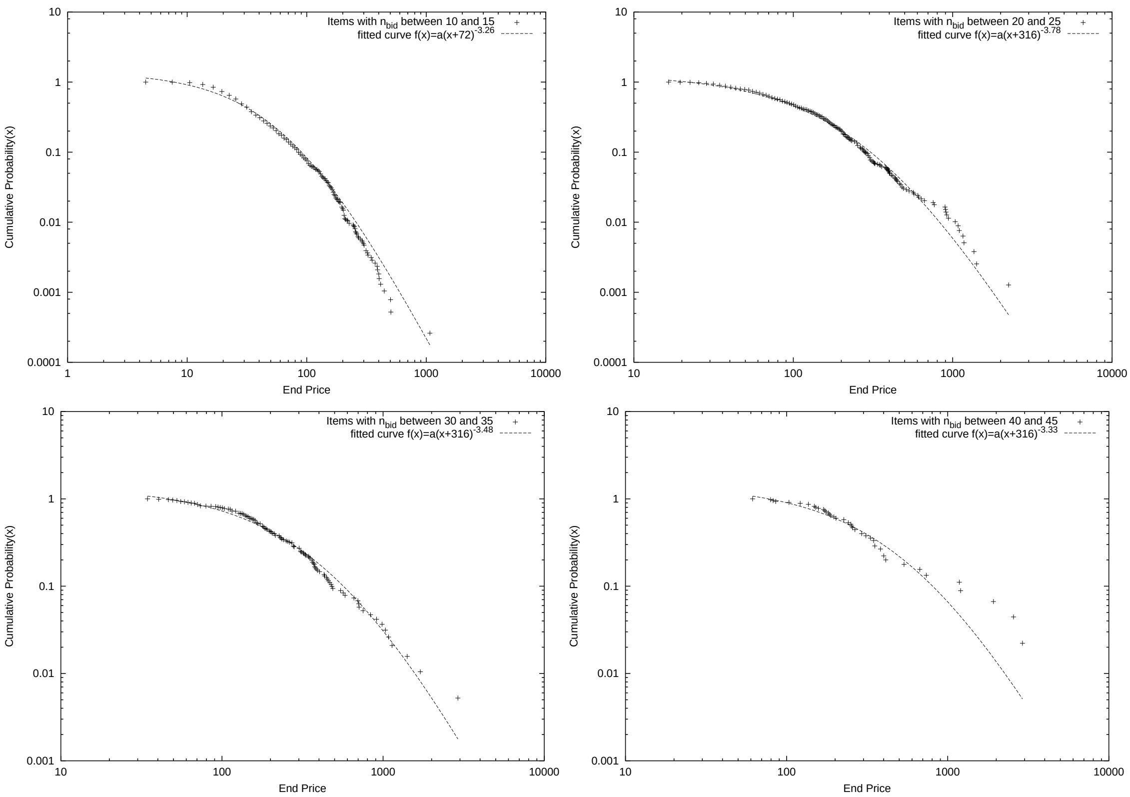

Figure 6.10.: Cumulative distribution of the end price for fixed total number of bids by using a subset of the data set DB-1 including all auctions starting with 1 Euro and ending after n_{bid} bids with $n_{bid} \in (10, 15]$, $n_{bid} \in (20, 25]$, $n_{bid} \in (30, 35]$ and $n_{bid} \in (40, 45]$. Fitted curves have the form $f(x) \propto (x + c)^b$.

auctions starting with 1 Euro. We showed that a simple stochastic process predicts a log-normal distribution for prices reached after n_{bids} bids. This approximation is close to the observed actual distribution but is not exactly the right one. The actual distribution has the important scale-free property related to the critical state of eBay as a complex system.

6.5. Auctions with other starting prices

It would be interesting to know the relation between the end price and the number of bids for auctions with starting prices other than 1 Euro. The data we have gathered -and are working with- is however not suitable for this kind of studies. The problem we encounter when running such studies is the poor statistics we have for auctions starting with higher prices.
In Fig. 6.12 we show the histogram of the number of auctions starting with different prices (in Euro). The data we use for this histogram is a subset of 10% randomly chosen auctions

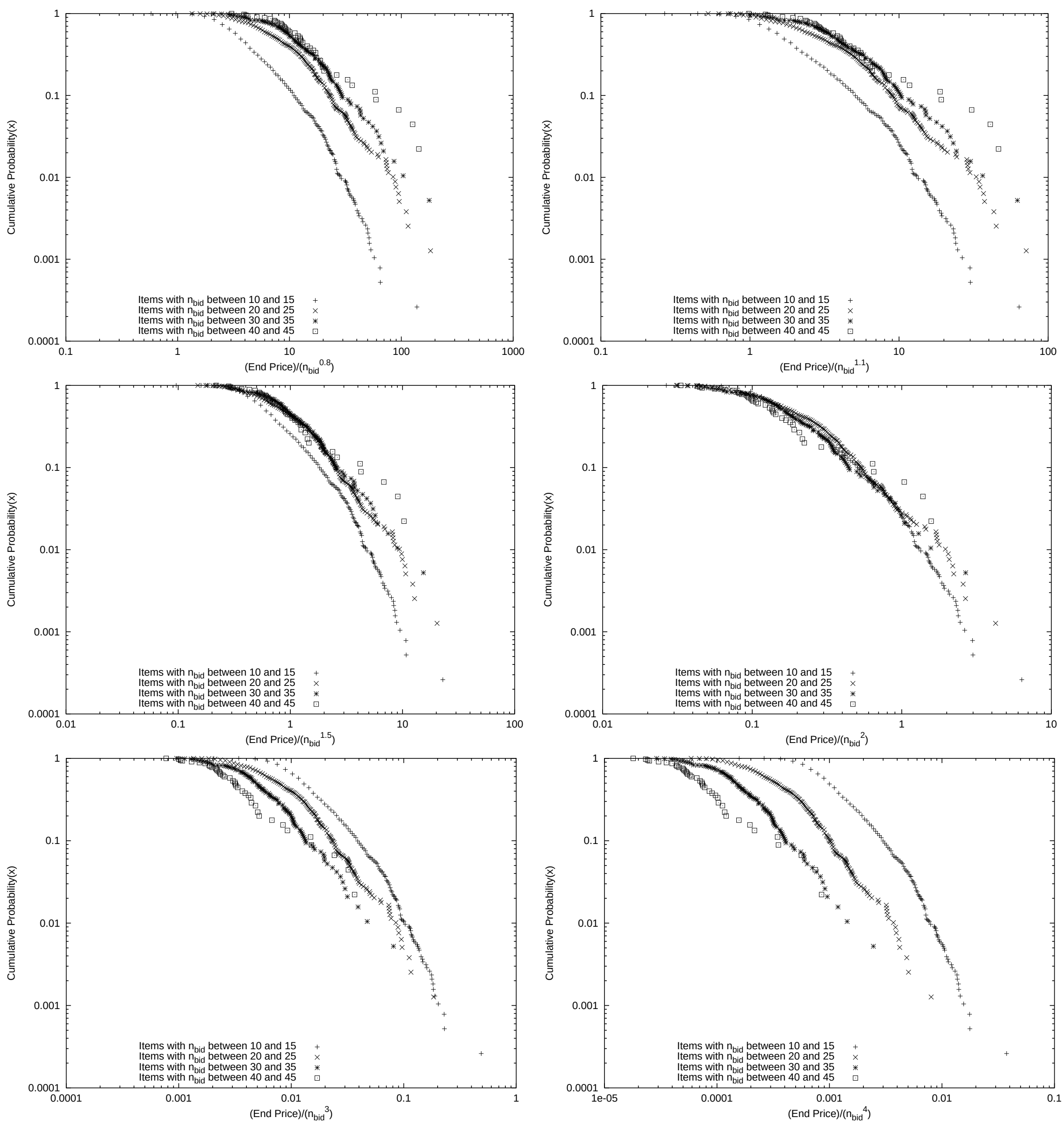

Figure 6.11.: Universal scaling law. Distributions of $\frac{p_{\text{auct}}}{n_{\text{bid}}^{\gamma}}$ for several values of $\gamma = 0.8, 1.1, 1.5, 2, 3$ and 4. For $1.5 \geq \gamma \leq 2$ all distributions are similar. The data set DB-1 is used for this study.

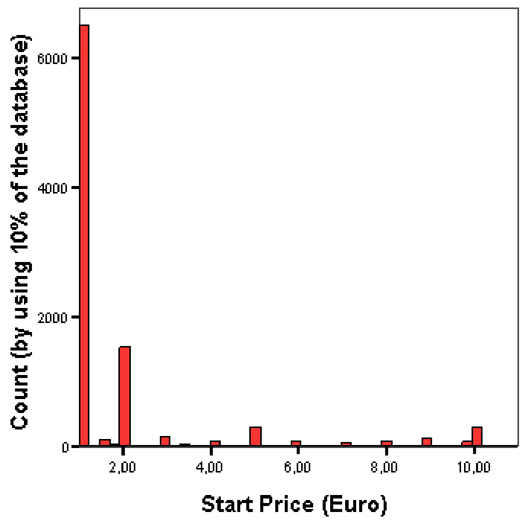

Figure 6.12.: Histogram of the number of auctions starting with different prices (in Euro). The data we use for this histogram is a subset of 10% randomly chosen auctions of the data set DB-1. This diagram shows that the number of auctions decreases strongly with the increasing starting price.

of the data set DB-1. Fig. 6.12 shows that the number of auctions decreases strongly with the increasing starting price. We limit this study to two different intervals of starting prices and study all auctions ending with at least one bid, with starting prices $p_{\mathrm{start}} \in (1,3)$ including 18041 auctions and $p_{\mathrm{start}} \in (9,11)$ including 3952 auctions.
There exists different possibilities for the functionality of $\langle p_{\mathrm{auct}} \rangle$ on n_{bid} (for the case the starting price is higher than 1 Euro), even if we assume that the relation found empirically (Eq. (6.17)) should hold. Depending on the assumption whether the simple stochastic process introduced in Eq. (6.18) is applicable or not one yields different results.
Let us first start with Eq. (6.18), where $P_0 > 1$. In this case the equation will be

$$\ln(P_n) = \ln(\varrho_n + 1) + \ldots + \ln(\varrho_1 + 1) + \ln(P_0), \tag{6.22}$$

so

$$\ln(\frac{P_n}{P_0}) = \ln(\varrho_n + 1) + \ldots + \ln(\varrho_1 + 1). \tag{6.23}$$

By using the assumption that ϱ_n are i.i.d. random variables, the right side of this equation will represent the price achieved after n bids when starting from 1 Euro. By averaging both sides of the Eq. (6.23), and by setting $\langle \ln(P_n) \rangle \approx \ln(\langle P_n \rangle)$ one obtains

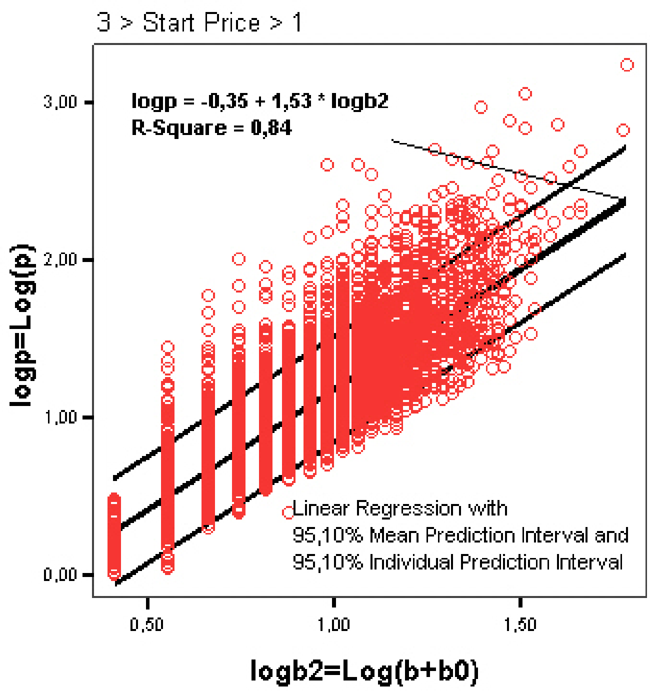

Figure 6.13.: Scatter plot and linear regression (curve fitting) of $\log(p_{\text{auct}})$ as a function of $\log(n_{\text{bid}} - n_{\text{start}})$ for data with $p_{\text{start}} \in (1, 3)$. The exponent α is found to be 1.53 equal to the one found by analyzing auctions starting with 1 Euro.

$$\left\langle \ln(\frac{P_n}{P_0}) \right\rangle = \langle \ln(P_n) \rangle \,|_{p_{\text{start}}=1}, \tag{6.24}$$

and by using Eq. (6.17) we will have

$$\left\langle \ln(\frac{P_n}{P_0}) \right\rangle \propto \alpha n_{\text{bid}}. \tag{6.25}$$

Using the assumption $\left\langle \ln(\frac{P_n}{P_0}) \right\rangle \approx \ln \left\langle \frac{P_n}{P_0} \right\rangle$ again, one should expect

$$\left\langle \frac{P_n}{P_0} \right\rangle \propto n_{\text{bid}}^{\alpha}. \tag{6.26}$$

On the other hand one can imagine that the starting price does not have an important influence on the evolution of the price, so that the transformed value $p_{\text{auct}} - p_{\text{start}}$ should follow the empirically found relation (Eq. (6.17)).
Another possibility is that the starting price imagined as the price achieved after n' bids, if the auction would start from 1 Euro and the price would follow Eq. (6.17).

These possibilities could be verified/falsified by means of statistical analysis of the data. We consider three possibilities for the functionality of $\langle p_{\mathrm{auct}} \rangle$ on n_{bid} as follows:

i. $\left\langle \frac{p_{\mathrm{auct}}}{p_{\mathrm{start}}} \right\rangle \propto n_{\mathrm{bid}}^{\alpha}$,

ii. $\langle p_{\mathrm{auct}} - p_{\mathrm{start}} \rangle \propto n_{\mathrm{bid}}^{\alpha}$,

iii. $\langle p_{\mathrm{auct}} \rangle \propto (n_{\mathrm{bid}} + n_{\mathrm{start}})^{\alpha}$,

where α can be different from the one known for auctions starting with 1 Euro. In the third possibility we make use of n_{start}, which is defined as the number of bids an auction should receive if starting with 1 Euro to reach its starting price p_{start}, by assuming that the relation $\langle p_{\mathrm{auct}} \rangle \propto n_{\mathrm{bid}}^{\alpha}$ is also valid for auctions starting with a given price higher than 1 Euro. We calculate n_{start} by using the Eq. (6.17) as

$$n_{\mathrm{start}} \propto e^{(\ln(p_{\mathrm{start}})/\alpha)}. \tag{6.27}$$

We use the empirical results gained by analyzing auctions starting with 1 Euro (presented in sec. 5.3.9) and set $\alpha = 1.53$. For the auctions starting with $p_{\mathrm{start}} \in (1, 3)$ we calculate a mean value for n_{start} to be 1.57 and for the auctions starting with $p_{\mathrm{start}} \in (9, 11)$ a mean value of n_{start} to be 4.5.
Fig. 6.13 shows the linear regression (curve fitting) of $\log(p_{\mathrm{auct}})$ as a function of $\log(n_{\mathrm{bid}} + n_{\mathrm{start}})$ for data with $p_{\mathrm{start}} \in (1, 3)$ and Fig. 6.14 shows the same graph for data with $p_{\mathrm{start}} \in (9, 11)$.
Fig. 6.15 shows the scatter plot and the linear regression of $\log(p_{\mathrm{auct}} - 9)$ as function of $\log(n_{\mathrm{bid}})$ for data with $p_{\mathrm{start}} \in (9, 11)$. The exponent α is measured to be 1.31. This seems to be in better agreement with results known from analyzing auctions starting with 1 Euro compared with the result shown in Fig. 6.14, where $\alpha = 1.20$. To verify this result and make a better comparison we analyze all auctions starting with $p_{\mathrm{start}} > 1$ and ending after at least one bid. The result of this study is presented in Fig. 6.16 which shows the scatter plot and the linear regression of $\log(p_{\mathrm{auct}} - p_{\mathrm{start}} + 1)$ as a function of $\log(n_{\mathrm{bid}})$. The exponent α is measured to be 1.41, which is close to 1.53.
This study shows that the relation between the expectation value of the rescaled/transformed price, achieved after n bids and number of bids n_{bid} ($n = n_{\mathrm{bid}}$) is scale-free. The exponents turn out to be similar and approximately independent of the starting price if one considers the transformed achieved price equal to $p_{\mathrm{auct}} - p_{\mathrm{start}} + 1$.

6.6. Fraudulent bidding

The interesting observation explained in the previous section combined with the scale-free distribution of the end prices for a fixed number of bids, which can be fitted by using a log-normal distribution, can be used for detecting auctions showing an abnormal behavior. Independent of the strategy chosen by a bidder, it is plausible to assume that he would like to win the interested auctioned item with a price below a maximum budget he has and with

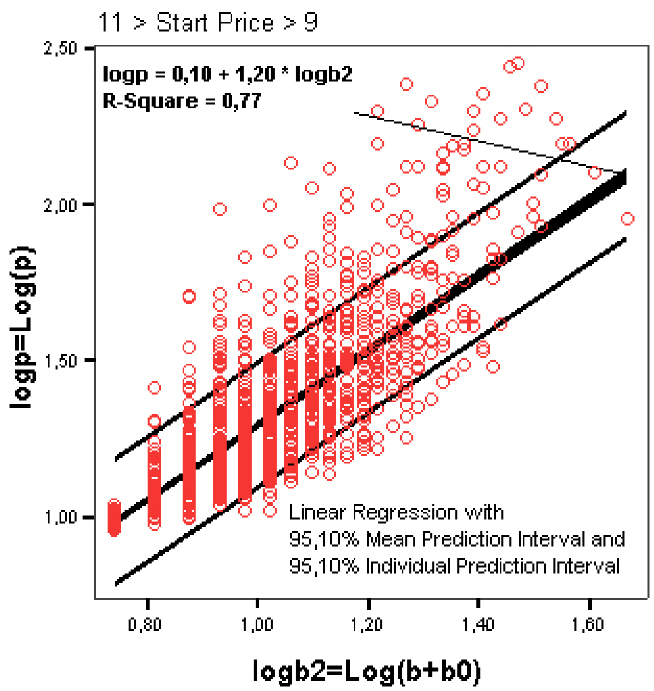

Figure 6.14.: Scatter plot and linear regression of $\log(p_{\text{auct}})$ as a function of $\log(n_{\text{bid}} + n_{\text{start}})$ for data with $p_{\text{start}} \in (9, 11)$. The exponent α is found to be 1.2 which is not in good agreement with the result found by analyzing the auctions starting with 1 Euro.

a winning bid as low as possible. There exists, as mentioned in sec. 4.2, strategies leading to a higher end price, which is just in the interest of the seller. Shill bidding as a fraudulent bidding by a seller using an alternate account or a friend in order to inflate the price is one of these strategies. Shill bidding is strictly forbidden by eBay [84], but nevertheless happens quite frequently. The other strategy is the so called squeezing, in which a seller uses a second account in order to uncover the bid of a potential buyer. Then he retracts his bid and uses shill bidding to push the price as high as possible without winning.
Both these strategies cause a statistically measurable deviation from the average behavior. The simple relation of Eq. (6.17) provides a tool for searching for such strategies. This will be discussed in the following section.

6.6.1. Shill bidding

It is not uncommon that certain agents try to manipulate the sell price of the items they offer. Therefore we distinguish two different types of bidding behavior:

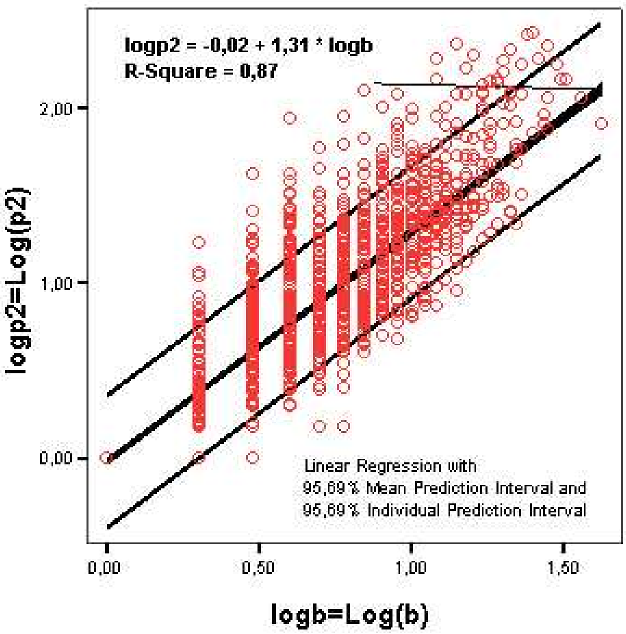

Figure 6.15.: Scatter plot and linear regression of $\log(p_{\mathrm{auct}} - 9)$ as a function of $\log(n_{\mathrm{bid}})$ for data with $p_{\mathrm{start}} \in (9, 11)$.

i. All bidders try to keep the price as low as possible,

ii. At least one bidder tries to push the price higher.

The second type of behavior normally corresponds to shill bidding, which happens typically in private value auctions. Usually this sort of manipulation can be identified only after the auction has ended because the whole purpose of shill bidding is increasing the price without winning in the end.
Online auction fraud is according to the National Consumer League [85] and Federal Trade Commission [86] the number one Internet fraud and is increasing rapidly [87]. Because of the huge number of online auctions and users, and the ease of creating new accounts, it seems to be very difficult to detect shill bidding without any computer aided method.
Previous trials have shown that shill bidding is difficult to recognize in an automated way. We have tried to identify this sort of manipulation through the discussed statistical properties. Indeed, successful shill bidding leads to auctions which show clear deviations from the observed simple statistical laws.
By using the relation between $\langle p_{\mathrm{auct}} \rangle$ and n_{bid} (Eq. (6.17)) and the fact that the variances of log-normal fits of the distributions of the prices depend very weakly on n_{bid}, ($\sigma^2 \approx 0.15$)

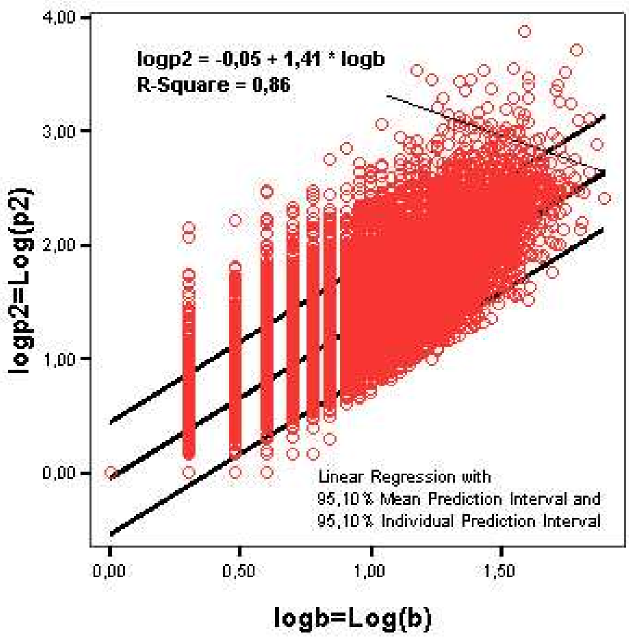

Figure 6.16.: Scatter plot and linear regression of $\log(p_{\text{auct}} - p_{\text{start}} + 1)$ as function of $\log(n_{\text{bid}})$ for all auctions with $p_{\text{start}} > 1$ and ending after at least one bid.

we define a method of finding indications of shill bidding automatically.
Assume that the distribution of logarithm of prices with a fixed number of bids is a Gaussian distribution with mean μ and variance σ^2. The probability to find a price $(\log(\text{price}))$ that falls within one standard deviation σ of the average price is theoretically 68.3%. This theoretical confidence interval is indicated in Fig. 6.17 by the two straight lines with slopes 1.32 and 1.92, respectively. Therefore one expects that the probability P_σ that the actual price p_{auct} is larger than $\langle p_{\text{auct}} \rangle + \sigma$ is about 15.9%. However, for our data we find $P_\sigma = 7\%$ which indicates deviations from the log-normal distribution (see inset of Fig. 6.17).
We assume that auctions using shill bidding or other similar strategies leading to a higher price should fall out of the confidence interval defined above. To verify this assumption we have performed two tests by:

i. investigating the statistical properties of auctions identified as shill bidding,

ii. checking whether randomly chosen auctions outside the confidence interval of the price-bid relationship discussed above show indications of shill bidding.

Both tests require an extensive amount of work, e.g. the investigation of the trading history of the seller over a long time or monitoring eBay discussion forums [88, 89].

Hints and criteria we use for finding indications of shill bidding are as follows:

- A member bids several times under the highest bidder towards the end of an auction, incrementing the current price by small amounts.
- A member bids several times towards the end of an auction and retracts if he bids more than the high bidder.
- A member bidding several times even when there are no other bidders.
- A bidder bids exclusively or nearly exclusively on one or two seller's items, yet rarely wins.
- Similarity of seller and bidder IDs.
- Seller relist quickly the item. When items are accidentally won by a shill account, they are often relisted soon after auction closes.
- Prompt feedback after an item is accidentally won by a shill account.

Fig. 6.17 shows a comparison between these shill-auctions and average behavior of all auctions, clearly indicating the deviations from the average behavior.
For test (i) we have chosen 9 auctions that clearly have been identified as manipulated by shill bidding, e.g. through information from discussion forums. Only after that we have investigated the bidding history of these auctions in more detail. Fig. 6.17 shows that all of those, except for one, are clearly outside the confidence interval. For test (ii) 10 auctions outside the confidence interval have been chosen randomly. These have been checked thoroughly for indications of shill bidding. This also required investigating other auctions by the same seller etc. In this way we have found a clear indication for shill bidding in 7 of the 10 auctions.

6.7. Returns

As shown in sec. 6.4, the distribution of returns ϱ defined as the relative increase of the submitted bid b, $\varrho = (b - p_{\mathrm{auct}})/p_{\mathrm{auct}}$ plays a central role for the evolution of the price.
As presented in sec. 5.3.6 and in Fig. 5.9 the distribution of ϱ is found to follow a power law for almost 3 orders of magnitude with exponent -2.44.
It seems that the distribution of returns does not depend on variables like the number of bids or actual price. The returns are found to be very short-ranged correlated and approximately independent (see sec. 5.3.7).
Although there exist evidences that the human perception to a physical stimulus increases with the relative changes of the stimulus (Weber-Fechner law [82]) and for a wide range of sensations the generalized Stevens' power law [96, 97], (which states that the intensity of a stimulus relates to its perceived strength according to the relation $S = cI^{\alpha}$, where S is the amount of sensation, C a constant, I the stimulus intensity, and α the exponent) is found to hold, we do not have an explanation of the power-law distribution of returns up

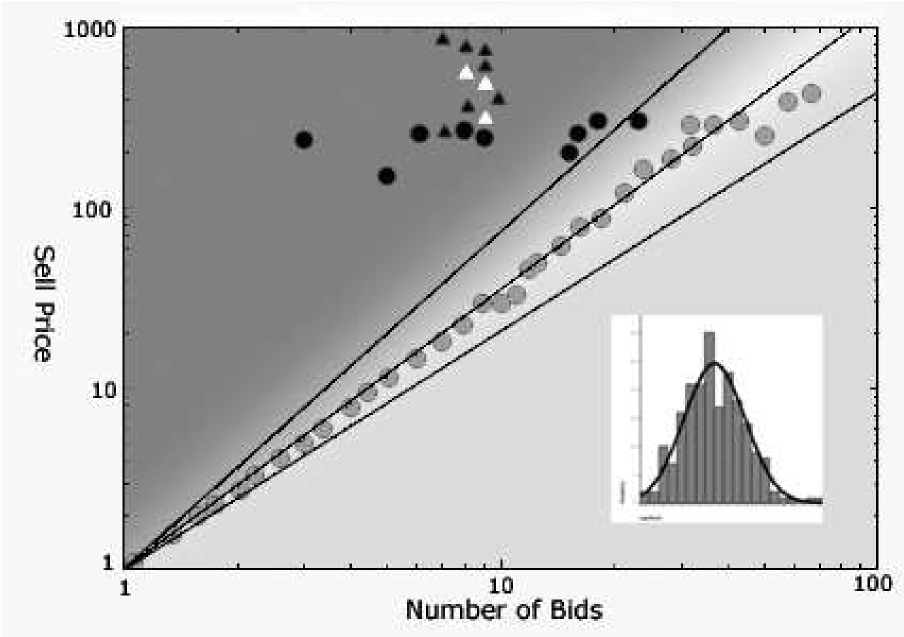

Figure 6.17.: Shill Bidding. Distribution of sell prices as function of the number of bids placed. Only auctions with a starting price of 1 Euro are considered. We use both data sets DB-1 and DB-2 for our -in the text discussed- test. Gray dots correspond to the average sell price for fixed number of bids. The broken lines specify the confidence interval (one standard deviation) determined from a log-normal fit. This is shown for $n_{\rm bid} = 20$ in the inset. For other values of $n_{\rm bid}$ very similar results are obtained. Black dots indicate auctions identified as shill bidding using the criteria of [90–95]. Triangles denote auctions that have been tested for possible shill bidding. For black triangles strong indications for shill bidding have been found. Light-gray and dark-gray colors denote regimes with high or low probability of shill bidding respectively.

to now. We believe that this power-law distribution can be viewed as an emergent behavior of the eBay as a complex system, giving indications for the criticality of the system.
The only parameter with influence on returns seems to be the bid submission time. As shown in sec. 6.3, return and relative bid submission time are correlated. The relation is described by a linear regression and using the common form of least squares fitting and is given by $\langle \varrho \rangle \propto \tau^{0.38}$ (Fig. 6.6).
We found that the power-law form of the distribution does not change when considering different time spans for the bid submissions. We analyzed the data set DB-1 and selected all bids related to auctions with a duration of 7 days. We divided the bid-event data in several subsets and looked at bids placed in time spans of last 6 minutes, last 30 minutes, last one hour, last 3 hours, last 9 hours, first 5 hours, first 10 hours, first 15 hours and first 50 hours and measured the cumulative frequency of returns. The result is shown in

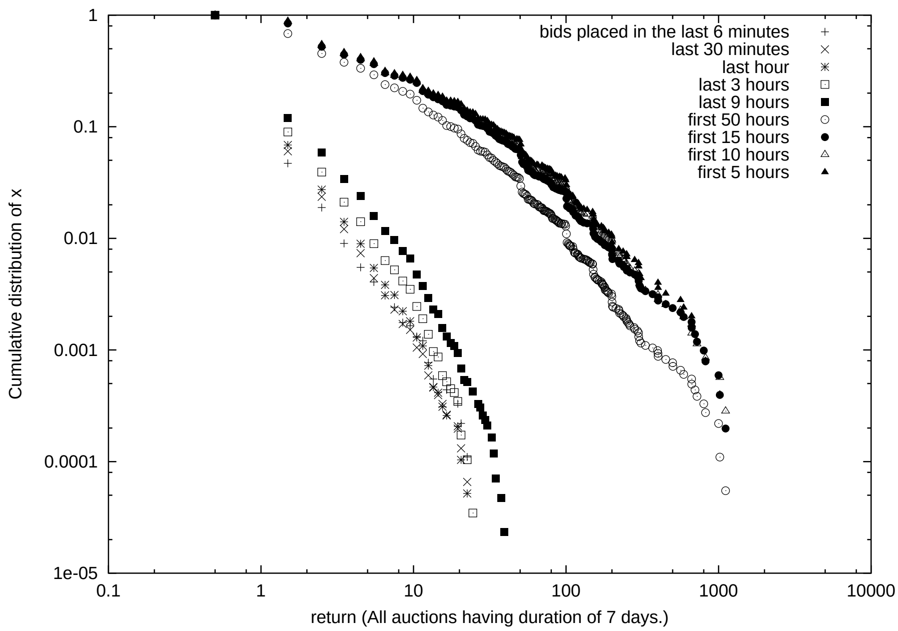

Figure 6.18.: Cumulative distribution of returns for several time intervals of the arriving bid. The power-law form of the distribution seems to be stable.

Fig. 6.18. The power-law form of the distribution seems to be stable.
If we assume that the exponents of the power-law distributions are close to γ independent of the strategy a bidder may choose, the cumulative probability density function of returns related to the bids with the same strategy j (denoted by $P_j(\varrho >)$) will be a power law with exponent $\gamma + 1$

$$P_j(\varrho >) = C_j \varrho^{\gamma+1}. \tag{6.28}$$

Let a bidder choose a strategy j among n possible strategies with a given probability p_j. The summation over all p_j should be 1:

$$\sum_{j=1}^{n} p_j = 1. \tag{6.29}$$

Now imagine we have a sample data of bids without any knowledge about the strategy the bidders have chosen by placing their bid and measure the cumulative frequency of returns of these bids. The cumulative probability density function should follow a power law due to the stability under mixture discussed in sec. 2.5.2. This is easy to show:

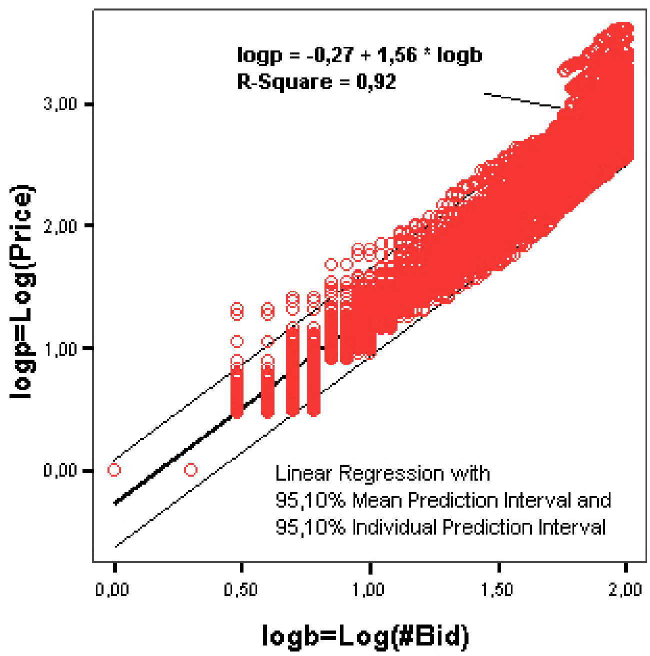

Figure 6.19.: Relation between $\langle p_{\rm auct}\rangle$ and $n_{\rm bid}$ by using the described model. The model is able to reproduce the empirically found relation.

$$
\begin{aligned}
P(\varrho >) &= p_1 P_1(\varrho >) + \cdots + p_n P_n(\varrho >) = \sum_{i=1}^{n} p_j P_j(\varrho >) \\
&= \sum_{i=1}^{n} p_j C_j \varrho^{\gamma+1} = (\sum_{j=1}^{n} p_j C_j)\varrho^{\gamma+1} = C\varrho^{\gamma+1} \qquad (6.30)
\end{aligned}
$$

where $C = \sum_{j=1}^{n} p_j C_j$.

By using the assumption that the distribution of returns does not depend strongly on the choice of the strategy we defined a model and used Monte Carlo simulations to test the behavior like the relation between $\langle p_{\rm auct}\rangle$ and $n_{\rm bid}$.

A simple model with following rules provides results similar to our empirical observations.

For $(t+1)$th iteration:

i. pick a random number $\varrho(t+1)$ with a power-law distribution in interval $[a, b]$ and exponent $\alpha = -2.5$, where $a = 10 \times ({\rm increment}/p_{\rm auct}(t))$ and $b = 1000$,

ii. calculate bid $b(t+1) = (\varrho(t+1)+1)p_{\rm auct}(t)$,

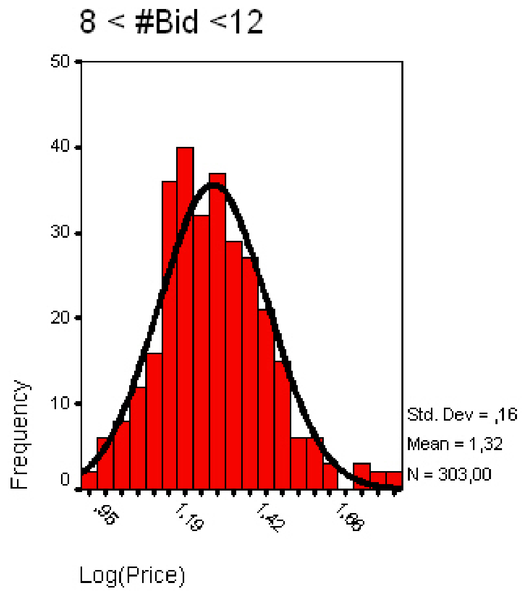

Figure 6.20.: Simulation results. Distribution of the end price for a fixed number of bids by using the described model. Distribution of the logarithm of end price for auctions starting with 1 Euro and ending after minimum 9 and maximum 11 bids are plotted. The fitted curve is a Gaussian (normal) distribution with $\mu = 1.32$ and $\sigma = 0.16$.

iii. $p_{\text{auct}}(t+1)$ is selected as the second maximum of the set $\{b(1), \ldots, b(t+1)\}$

In Figs. 6.19, 6.20 and 6.21 we show the results of the simulations. We have simulate 10.000 auctions starting with 1 Euro.
In Fig. 6.19 the relation between $\log(p_{\text{auct}})$ and $\log(n_{\text{bid}})$ is plotted. The model seems to generate results in a very good agreement with the empirically found relation.
Figs. 6.20 and 6.21 show frequencies of logarithm of the end prices for fixed number of bids. The means are close to the empirical values. This is not the case for standard deviations.

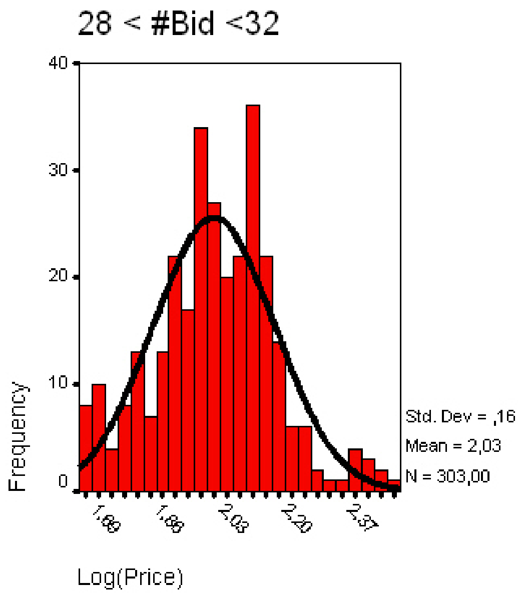

Figure 6.21.: Simulation results. Distribution of the end price for a fixed total number of bids by using the described model. Distribution of the logarithm of end price for auctions starting with 1 Euro and ending after minimum 29 and maximum 31 bids are plotted. The fitted curve is a Gaussian (normal) distribution with $\mu = 2.03$ and $\sigma = 0.16$.

7. Discussion

7.1. Summary

The present work studies eBay online auctions as a complex system, where agents in form of sellers and bidders interact in a large number. The first chapter gives reasons why studying such a system is interesting also for physicists. The main argument is the existence of different rational strategies an agent can choose. The question whether human based systems could be analyzed by means of methods developed for unintelligent agents is still open. We discuss this in section 6.3 and find that the assumption of rational agents can be neglected at least for a large number of behaviors and properties.
The main part of the present work studies the empirical findings and their interpretation. These findings are the result of statistical analysis of two major sets of data gathered by using the internet interfaces. The screen-scraping and HTTP parsing methods used provide well formatted data over 200,000 auctions. The auctions are chosen such that the data includes a large variety of final prices and other dynamical variables. Some of our studies need a long time monitoring of the activity of agents. That is the reason why we gathered two different sets of data. One of them represents a randomly chosen part of all auctions of eBay for a fixed date. The other one monitors some selected categories over a large time span.
Results of statistical analysis of the data can be classified in two main groups. One includes the probability distributions of selected variables. The other one includes relations between different variables which allow to identify correlations. These results are briefly listed below[1]:

- Probability distribution of variables
  - Final price (for the complete data set and for a variety of fixed number of bids).
  - Starting price.
  - Duration of auctions.
  - Number of positive feedbacks received (both for sellers and bidders).
  - Total number of bids placed on an item.
  - Total number of bids placed by a bidder.
  - Total number of items offered by a seller.
  - Bid submission time (both with respect to the real remaining time to the end of an auction and relative remaining time to the end of an auction).

[1] Note that some of these results are not explicitly included in the present work.

  - Return, defined as the relative difference between the placed bid and the actual price.

- Correlated variables
  - Relation between final price and number of bids placed on an item. Also as a function of the starting price.
  - Relation between final price and distinct number of bidders participating in an auction.
  - Relation between return and relative bid submission time.

Statistical analysis of the eBay data shows that the probability density functions for a wide range of quantities follow rather simple functionalities like exponential and power laws.
Some of these power-law distributions are to be explained by means of simple growth processes. In section 6.1 we present an explanation of the distribution of the total number of items offered by the same seller by using the Yule process (discussed in sec. 2.6.1). The result predicted by the theory is very close to the measured one.
It is found that the role of rational agents is not crucial and some observed power laws are to be understood as results of some simple microscopic rules, providing a self-similar structure.
Similar arguments can be used for explaining the distribution of the number of positive feedbacks (sec. 5.3.3).
One of the most interesting findings of this study is the scale-invariant relation between the average price and the number of bids placed on the item. This nontrivial relation was studied thoroughly by means of simple stochastic processes and by searching for scaling conditions in sec. 6.4. Similar relations are found between other quantities. For instance, the relation between return and relative bid submission time[2] is a similar scale-invariant relationship. It is also found that the relation between the average price and the distinct number of bidders obeys the same scale-independent functionality[3]. These relations are evidences, which indicate that the system should be in the critical state. Scale-invariance is connected to the fractal structure of the system making the system show self-similar properties on any scale, which is observed by many human based systems.
Another interesting result of this study is the finding that the bidders prefer to bid close to the auction ending times. The probability of bid submission as a function of the time remaining until the end of the auction is found to decrease exponentially with increasing remaining time. Close to the end of the auctions the distribution of bid submission times is found to follow a power law.
These findings challenge the prediction of the game theory, which suggests that in second-price auctions[4] with fully rational agents the timing of bids should not play any role and there should exist no incentive to bid less than one's own private value. Several studies and efforts towards understanding and explaining the difference between theoretical prediction

[2]The remaining time to the end of auction divided by its duration.
[3]With other exponent.
[4]eBay auctions can be considered as second-price auctions due to the proxy-bidding mechanism.

of equally distributed bid submission times and actual observed exponential and power-law distributions are done by using methods of the game theory. A brief presentation of results and explanations of these approaches is given in chapter 4.
Our own study indicates that the choice of the bidding strategy is not the only parameter responsible for the exponential behavior of the distribution of bid submission times. Instead also other mechanisms like the sorting criteria offered by eBay are relevant.
The items are listed on eBay pages by using the so called "ending soonest" criterion by default. This means, auctions closer to their end are listed first. So the probability of being seen for an item, increases with the decreasing remaining auction time due the limited number of items listed on each page. The analysis of the data shows that the probability of bidding increases not only at the end of the auction, but also at the beginning. The distribution shows a symmetric behavior around the mid-duration of the auction (with the exception of the regime very close to the end). This finding suggests that the other sort criterion called "newly listed" should have been used commonly by potential buyers and that the probability of placing a bid should be directly connected to the probability of finding of the auction on eBay pages.
The empirically found power-law distribution of bid submission times close to the end of auctions (Fig. 5.13) can be understood by considering an avalanche like dynamics known from the theory of self-organized criticality. Close to the end, placing a new bid results in reaction of other potential buyers causing them placing more new bids, this activity stays on until the system has reached a relaxed state. Self-organized criticality is one of the mechanisms of producing power-law distributions, which are observed almost everywhere in natural and human sciences and are related to large interactive systems known as complex systems. Chapter 2 gives an overview of complex systems and power-law generating mechanisms.
Simple and robust relations, like that between the average price and the number of bids, can provide the possibility to identify auctions showing a statistically abnormal dynamical progress. For example the average price versus number of bids relation can be successfully used for the detection of the forbidden shill bidding, which happens frequently. Shill bidding is the fraudulent bidding by a seller on his own item by using an alternate account or a friend in order to inflate the price. Shill bidding is difficult to recognize in an automated way. We found that shill auctions show statistically measurable deviations from the average behavior and used this finding for construction of a method, which identifies suspicious auctions automatically. This method is introduced in sec. 6.6.1.
We have tested the method by studying the history of selected suspicious auctions in more detail. Indeed we found that a large fraction of them shows strong evidence for manipulations by shill bidding.
The criticality of eBay, and perhaps one of the main findings of the present study, can be summarized in the power-law distribution of returns[5]. It is found that a variety of properties of the system can be captured and explained through this distribution showing the possibility of finding a variable, which mirrors a great part of the complexity of the system.

[5] Definition of returns and explanations are given in sections 5.3.6 and 6.7.

7.2. Open problems

It is shown that the choice of strategy does not explain the exponential behavior of the bid submission times. But this does not mean that the bid submission time is independent of the choice of strategy. Indeed there are evidences that both distribution and average value of returns depend on the bid submission time (return can be used for a automated separation between some different strategies, for example consider sniping and evaluation strategies, which result in average relative small and large values of return, respectively). How the choice of strategy quantitatively determines the bid submission time is still an open question.

To answer this question, however, one needs other data than used in the present work. One possibility is to analyze the auctions listed by using the so-called "Featured Plus!" option. These items appear at the top of the listing page and can be seen as long as the auction runs independent of the remaining time of the auction. By using this kind of data, the influence of sorting criteria would be unimportant and the chosen strategy could be identified more precise and simpler.

Another open question concerns the power law distribution of the returns. Although there exist evidences that the human perception to a physical stimulus increases according to exponential and power-law relations with the stimulus[6], we do not have an explanation why the distribution of stimulus (in this case, return) should follow a power law.

Some interesting relations are not studied in the present work due to technical restrictions. We believe, answering some arising questions, especially when searching for the influence of rationality, are only possible by using detailed information about events like number of visits of a potential buyer before bidding, etc.

Although lots of relations and distributions could be understood qualitatively and quantitatively, there exists, up to now, no unifying model, which describes all of the observations. It is debatable if such a unifying model could exist. It seems to be possible to model many of features discussed in this work because of the stochastic nature of them. But many other features are results of rational adaptive decisions of agents. These features are not easily understandable, nor the conditions on which an agent would select a specified decision are clearly available.

eBay regarded as a complex system seems to have features of totally stochastic processes, irrational agents and fully rational agents altogether.

[6] Weber-Fechner law and Stevens' power law [96, 97].

A. Appendix

A.1. Power-law distributions and scale-free condition

Here we show[1] that a power-law $p(x)$ is the only distribution satisfying the scale free condition of Eq. (2.6).
Starting from Eq. (2.6), we first set $x = 1$ to gain $p(a) = g(a)p(1)$ and $g(a) = p(a)/p(1)$. The Eq. 2.6 can then be written as

$$p(ax) = \frac{p(a)p(x)}{p(1)}. \tag{A.1}$$

Now we differentiate both sides with respect to a

$$xp'(ax) = \frac{p'(a)p(x)}{p(1)}, \tag{A.2}$$

where p' is the derivative of p with respect to its argument. Let us set $a = 1$ now to get

$$x\frac{\mathrm{d}p}{\mathrm{d}x} = \frac{p'(1)}{p(1)}p(x). \tag{A.3}$$

This first-order differential equation has the following solution

$$\ln p(x) = \frac{p(1)}{p'(1)}\ln x + \text{constant}. \tag{A.4}$$

By setting $x = 1$ we find that the constant is $\ln p(1)$. And we will have

$$p(x) = p(1)\, x^{-\alpha}, \tag{A.5}$$

where $\alpha = -p(1)/p'(1)$.

[1] Further discussion in [35].

A.2. The Yule process

Here we show[2] how the Yule process generates power-law distributions. Using the process definition in sec. 2.6.1, one can derive the master equation as follows

$$(n+1)p_{z,n+1} = np_{z,n} + l\frac{z-1+c}{z_0+c+l}p_{z-1,n} - l\frac{z+c}{z_0+c+l}p_{z,n}, \qquad \text{for } z > z_0, \tag{A.6}$$

and

$$(n+1)p_{z_0,n+1} = np_{z_0,n} + 1 - l\frac{z_0+c}{z_0+c+l}p_{z_0,n}, \qquad \text{for } z = z_0. \tag{A.7}$$

(Note that z can not be less than z_0.)
By searching for stationary solutions of these equations and using $p_z = \lim_{n\to\infty} p_{n,z}$ one finds that:

$$p_{z_0} = \frac{z_0+c+l}{(l+1)(z_0+c)+l}, \tag{A.8}$$

and

$$\begin{aligned} p_z &= \frac{(z-1+c)(z-2+c)\dots(z_0+c)}{(z-1+c+\alpha)(z-2+c+\alpha)\dots(z_0+c+\alpha)}\, p_{z_0} \\ &= \frac{\Gamma(z+c)\Gamma(z_0+c+\alpha)}{\Gamma(z_0+c)\Gamma(z+c+\alpha)}\, p_{z_0}, \end{aligned} \tag{A.9}$$

where the Γ-function notation and defined $\alpha = 2 + (z_0 + c)/l$ are used. By using the beta-function one can simplify this expression:

$$p_z = \frac{\mathrm{B}(z+c,\alpha)}{\mathrm{B}(z_0+c,\alpha)}\, p_{z_0}. \tag{A.10}$$

The beta-function has a power law in its tail[3], causing the Yule process generating a power-law distribution of the form $p_z \sim z^{-\alpha}$ with the exponent:

$$\alpha = 2 + \frac{z_0+c}{l}. \tag{A.11}$$

[2] Further discussion in [35].
[3] $\mathrm{B}(a,b) \sim a^{-b}$.

A.3. Linear regression

Here we present[4] the linear least squares fitting technique as a mathematical procedure for finding the best-fitting curve to a given set of points by minimizing the sum of the squares of the offsets (also called residuals) of the points from the curve. The least squares fitting is the most commonly applied form of linear regression. In this fitting technique the sum of the squares of the offsets is used instead of the offset absolute values because this allows the residuals to be treated as a continuous differentiable quantity. However, because squares of the offsets are used, outlying points can have a disproportionate effect on the fit.
Assume a linear fit of a set of n data points

$$y = a + bx. \tag{A.12}$$

The sum of the squares of the offsets will be

$$L^2(a,b) = \sum_{i=1}^{n} [y_i - (a + bx_i)]^2. \tag{A.13}$$

The condition for L^2 to be a minimum is that

$$\frac{\partial(L^2)}{\partial a} = -2\sum_{i=1}^{n} [y_i - (a + bx_i)] = 0, \tag{A.14}$$

and

$$\frac{\partial(L^2)}{\partial b} = -2\sum_{i=1}^{n} [y_i - (a + bx_i)]x_i = 0. \tag{A.15}$$

These lead to the equations

$$na + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i, \tag{A.16}$$

and

$$a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} {x_i}^2 = \sum_{i=1}^{n} x_i y_i. \tag{A.17}$$

And we will have

$$a = \frac{\sum_{i=1}^{n} y_i \sum_{i=1}^{n} {x_i}^2 - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i y_i}{n\sum_{i=1}^{n} {x_i}^2 - \left(\sum_{i=1}^{n} x_i\right)^2}, \tag{A.18}$$

and

[4] See [98] for further discussion.

$$b = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} {x_i}^2 - \left(\sum_{i=1}^{n} x_i\right)^2}. \tag{A.19}$$

This can be simplified by means of the definition of variances ${\sigma_x}^2$ and ${\sigma_y}^2$ and covariance $\mathrm{Co}(x,y) = \frac{\sum_{i=1}^{n}(x_i - \langle x \rangle)(y_i - \langle y \rangle)}{n}$ as

$$b = \frac{\mathrm{Co}(x,y)}{{\sigma_x}^2}. \tag{A.20}$$

A very useful quantity is the so-called correlation coefficient R, which is defined by

$$R = \sqrt{bb'}, \tag{A.21}$$

where b' is the coefficient in

$$x = a' + b'y, \tag{A.22}$$

and is given by

$$b = \frac{\mathrm{Co}(x,y)}{{\sigma_y}^2}. \tag{A.23}$$

So R^2 will be

$$R^2 = \frac{\mathrm{Co}^2(x,y)}{{\sigma_x}^2{\sigma_y}^2}. \tag{A.24}$$

This correlation coefficient will be 1 if there is complete correlation (then the lines coincide since all data points lie on them), otherwise it will be smaller than 1 and decreases toward 0 for linear fits to increasingly noisy data.
The correlation coefficient is independent of origin and scale.

Bibliography

[1] V. Pareto. Cours d'economie politique. *Droz, Geneva*, 1896.

[2] G. K. Zipf. *Human behaviour and the principle of least effort.* Addison-Wesley, Reading, MA, 1949.

[3] J. M. Yeomans. *Statistical mechanics of phase transitions*. Oxford: Clarendon Press, 1992.

[4] K. G. Wilson. The renormalization group and critical phenomena. *Rev. Mod. Phys.*, 55:583, 1983.

[5] K. G. Wilson. The renormalization group: Critical phenomena and the kondo problem. *Rev. Mod. Phys.*, 47:773, 1975.

[6] K. G. Wilson and J. Kogut. The renormalization group and the epsilon-expansion. *Physics Reports*, 12:75–199, August 1974.

[7] Johannes Voit. *The Statistical Mechanics of Financial Markets.* Springer-Verlag, May 2001.

[8] R.N. Mantegna and H.E. Stanley. *An Introduction to econophysics.* Cambridge University Press, November 1999.

[9] H. Sherman and R. Shultz. *Open Boundaries: Creating Business Innovation through Complexity.* Perseus Books Group, October 1999.

[10] P. Bak and C. Tang. Earthquakes as a self-organized critical phenomenon. *Journal of Geophysical Research*, 94:15635–15637, 1989.

[11] D. Chowdhury, L. Santen, and A. Schadschneider. Statistical physics of vehicular traffic and some related systems. *Phys. Repts.*, 329, No. 4-6:199–329, 2000.

[12] H. Frauenfelder, P. G. Wolynes, and R. H. Austin. Biological physics. *Rev. Mod. Phys*, 71:419, 1999.

[13] M. Lässig and A. Valleriani. *Biological Evolution and Statistical Physics (Lecture Notes in Physics).* Springer, 2002.

[14] W. Weidlich. *Sociodynamics; A Systematic Approach to Mathematical Modelling in the Social Sciences.* CRC Press, 2000.

[15] D. Stauffer. Introduction to statistical physics outside physics. *Cond-mat/0310037*, 2003.

[16] D. Chowdhury, L. Santen, and A. Schadschneider. Statistical physics of vehicular traffic and some related systems. *Physics Reports*, 199:329, 2000.

[17] K. Nagel and M. Schreckenberg. A cellular automaton model for freeway traffic. *Physics Reports*, 2:2221, 1992.

[18] W. Knospe, L. Santen, A. Schadschneider, and M. Schreckenberg. Empirical test for cellular automaton models of traffic flow. *Phys. Rev. E.*, 70:016115, 2004.

[19] http://www.autobahn-nrw.de. 2005.

[20] J. F. Nash. Non-cooperative games. *J. Physique I*, 54:286–295, 1951.

[21] A. Ockenfels. Private communications.

[22] http://www.ebay.com. 2005.

[23] http://en.wikipedia.org/wiki/emergence. 2005.

[24] Murray Gell-Mann. *The Quark and the Jaguar : Adventures in the Simple and Complex.* W. H. Freeman; Reprint edition, September 1995.

[25] CH. Bennett. *How to Define Complexity in Physics, and Why. In: Complexity, Entropy and the Physics of Information.* Addison Wesley, July 1990.

[26] Hermann Haken. *Information and Self-Organization: A Macroscopic Approach to Complex Systems.* Springer; 2nd edition, 2000.

[27] B. Eckhardt. BOOK REVIEW: Synergetics. *Journal of Physics A Mathematical General*, 38:773–774, January 2005.

[28] Wil. Lepkowski. The social thermodynamics of ilya prigogine. *Chemical and Engineering News*, 57:30, 1979.

[29] Eric W. Weisstein. Koch snowflake. *From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/KochSnowflake.html.*, 2005.

[30] http://en.wikipedia.org/wiki/self-organization. 2005.

[31] W. H. Press. Flicker noise in astronomy and elsewhere. *Comments Astrophys.*, 7:103–119, 1978.

[32] M. S. Keshner. 1/f noise. *Proc. IEEE*, 70:212–218, 1982.

[33] P. Dutta and P. Horn. Low-frequency fluctuations in solids: 1/f noise. *Rev. Mod. Phys.*, 53:497–516, 1981.

[34] B. B. Mandelbrot. *The Fractal Geometry of Nature.* W. H. Freeman, San Francisco, 1982.

[35] M. E. J. Newman. Power laws, pareto distributions and zipf's law. *cond-mat/0412004*, 2004.

[36] D. Sornette. *Critical phenomena in Natural Sciences.* Springer.

[37] B. B. Mandelbrot. *The Paul Lévy I knew. Lévy flights and related topics in physics, Lecture Notes in Phys. 450.* Berlin, 1995.

[38] B. B. Mandelbrot. *Fractals and Scaling in Finance.* Springer, 1997.

[39] W. Willinger, D. Alderson, J. Doyle, and L. Li. More normal than normal: Scaling distributions and complex systems. *Proc. 2004 Winter Sim. Conf.*, (To appear).

[40] E. W. Montroll and M. F. Shlesinger. On 1/f noise and other other distributions with long tails. *Proc. Natl. Acad. Sci. USA*, 78:3380–3383, 1982.

[41] G. U. Yule. A mathematical theory of evolution based on the conclusions of Dr. J. C. Willis. *Philos. Trans. R. Soc. London B*, 213:21–87, 1925.

[42] H. A. Simon. On a class of skew distribution functions. *Biometrika*, 42:425–440, 1955.

[43] R. K. Merton. The matthew effect in science. *Science*, 159:56–63, 1968.

[44] D. J. de S. Price. A general theory of bibliometric and other cumulative advantage processes. *J. Amer. Soc. Inform. Sci.*, 27:292–306, 1976.

[45] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 1999.

[46] M. Levy and S. Solomon. Power laws are logarithmic boltzmann laws. *Int. J. Mod. Phys.*, C7:65:596–601, 1996.

[47] D. Sornette and R. Cont. Convergent multiplicative processes repelled from zero:power laws and truncated power laws. *J. Phys. I France*, 7:431, 1997.

[48] H. Kesten. Random difference equations and renewal theory for products of random matrices. *Acta Math.*, 131:207–248, 1973.

[49] L. de Haan, S.I. Resnick, H. Rootzén, and C.G. de Vries. Extremal behaviour of solutions to a stochastic difference equation, with applications to arch-processes. *Stochastic Processes and Applics.*, 32:213–224, 1989.

[50] C. de Calan, J.-M. Luck, T.M. Nieuwenhuizen, and D. Petritis. On the distribution of a random variable occurring in 1d disordered systems. *J. Phys. A*, 18:501, 1985.

[51] K. Pearson. The problem of the random walk. *Nature*, 72:294, 1905.

[52] C. B. Yang and R. C. Hwa. The origin of power-law distributions in self-organized criticality. *J. Phys. A: Math. Gen.*, 37 No 42:L523–L529, 2004.

[53] E. W. Montroll and G. H. Weiss. *J. Math. Phys.*, 6:167, 1965.

[54] W. J. Reed and B. D. Hughes. From gene families and genera to incomes and internet file sizes: Why power laws are so common in nature. *Phys. Rev. E*, 66:067103, 2002.

[55] G. A. Miller. Some effects of intermittent silence. *American Journal of Psychology*, 70:311–314, 1957.

[56] P. Bak, C. Tang, and K. Wiesenfeld. Self-organized criticality: An explanation of the 1/f noise. *Phys. Rev. Lett.*, 59:381–384, 1987.

[57] B. Drossel and F. Schwabl. Self-organized critical forest-fire model. *Phys. Rev. Lett.*, 69:1629–1632, 1992.

[58] P. Bak and K. Sneppen. Punctuated equilibrium and criticality in a simple model of evolution. *Phys. Rev. Lett.*, 74:4083–4086, 1993.

[59] Z. Olami, H. J. S. Feder, and K. Christensen. Self-organized criticality in a continuous, nonconservative cellular automaton modeling earthquakes. *Phys. Rev. Lett.*, 68:1244–1247, 1992.

[60] E. T. Lu and R. J. Hamilton. Avalanches of the distribution of solar flares. *Astrophysical Journal*, 380:89–92, 1991.

[61] H.J. Jensen. *Self-Organized Criticality.* Cambridge University Press, New York, 1998.

[62] M. Mitzenmacher. A brief history of generative models for power law and lognormal distributions. *Internet Mathematics*, 1:2:226–251, 2004.

[63] R. Gibrat. Une loi des réparations économiques: léffet proportionnel. *Bull. Statist. gén Fr.*, 19:469, 1930.

[64] Joerg Reichardt and Stefan Bornholdt. Economic networks and social communities in online-auction sites. *physics/0503138*, 2005.

[65] http://pages.ebay.com/education/buying.html. 2004.

[66] http://pages.ebay.de/help/buy/bid-increments.html. 2004.

[67] http://pages.ebay.com/help/buy/proxy-bidding.html. 2004.

[68] Oskar Morgenstern and John Von Neumann. *Theory of Games and Economic Behavior.* Princeton University Press, 1980.

[69] Andrew M. Colman. *Game Theory and its Applications in the Social and Biological Sciences.* Routledge 2nd Rev edition, 1995.

[70] James D. Morrow. *Game Theory for Political Scientists.* Princeton University Press, 1994.

[71] David P. Barash. *The Survival Game:How Game Theory Explains the Biology of Cooperation and Competition.* Owl Books, 2004.

[72] W. Vickrey. Counterspeculation and competitive sealed tenders. *Journal of Finance*, 16:8–37, 1961.

[73] A. E. Roth and A. Ockenfels. Last-minute bidding and the rules for ending second-price auctions: Evidence from ebay and amazon auctions on the internet. *American Economic Review*, 92:1093–1103, 2002.

[74] R. T. Wilcox. Experts and amateurs: The role of experience in internet auctions. *Marketing Letters*, 11:363–374, 2000.

[75] A. E. Roth and A. Ockenfels. Late bidding in second price internet auctions: Theory and evidence concerning different rules for ending an auction. *Games and Economic Behavior*, 92:1093–1103, 2005 (forthcoming).

[76] S. Barbaro and B. Bracht. Shilling, squeezing, sniping: Explaining late bidding in online second-price auctions. *http://www.staff.uni-mainz.de/barbaro/node5.html*, 2004.

[77] D. Ariely, A. Ockenfels, and A. E. Roth. An experimental analysis of ending rules in internet auctions. *RAND Journal of Economics*, (forthcoming).

[78] W. Wang, Z. Hidv'egi, and A. B. Whinston. Shill bidding in multi-round online auction. 2001.

[79] I. Chakraborty and G. Kosmopoulou. Auctions with shill bidding. *Economic Theory*, 24:271–287, 2004.

[80] Paul Klemperer. *Auctions: Theory and Practice.* Princeton University Press, 2004.

[81] I. Yang, H. Jeong, B. Kahng, and A.-L. Barabási. Emerging behavior in electronic bidding. *Phys. Rev. E*, 68(1):016102, July 2003.

[82] M. Copelli, A. C. Roque, R. F. Oliveira, and O. Kinouchi. Physics of psychophysics: Stevens and weber-fechner laws are transfer functions of excitable media. *Phys. Rev. E*, 65(6):060901, June 2002.

[83] B. Manaris, T. Purewal, and C. McCormick. Progress towards recognizing and classifying beautiful music with computers-midi-encoded music and the zipf-mandelbrot law. *Proceedings of the IEEE SoutheastCon 2002 Columbia, SC*, pages 52–57, 2002.

[84] http://pages.ebay.com/help/community/shillbidding.html. 2004.

[85] http://www.natlconsumersleague.org/internetscamfactsheet.html. 2005.

[86] http://www.ftc.gov/. 2005.

[87] Harshit S. Shah, Neeraj R. Joshi, Ashish Sureka, and Peter R. Wurman. Mining for bidding strategies on ebay. *Lecture Notes on Artificial Intelligence. http://www.csc.ncsu.edu/faculty/wurman/publications.html*, 2003.

[88] http://answercenter.ebay.de/ search.jsp?search=true&q=shill. 2003.

[89] http://answercenter.ebay.de/ search.jsp?search=true&q=pusher&date=any. 2003.

[90] A. Gronen. *eBay Dirty Tricks.* Data Becker, 2003.

[91] http://www.darlingtontown.co.uk/hintsntips/ shill_bidding_on_ebay.shtml. 2004.

[92] http://www.ukauctionhelp.co.uk/shill.php. 2004.

[93] http://www.basestealer.com/shilling.html. 2004.

[94] D. A. Karp and K. Lichtenberg. *eBay Hacks.* O'Reilly, 2004.

[95] http://www.fraudbureau.com/articles/online/article6.html. 2004.

[96] S. S. Stevens. *Psychophysics: Introduction to its perceptual, neural and social prospects.* Wiley, New York, 1975.

[97] S. S. Stevens. A scale for the measurement of a psychological magnitude: Loudness. *Psychological Review*, 43:405–416, 1936.

[98] Eric W. Weisstein. Least squares fitting. *From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/LeastSquaresFitting.html*, 2005.

Deutsche Zusammenfassung

Die vorliegende Arbeit beschäftigt sich mit eBay-Auktionen als einem komplexen System. Das Onlineauktionshaus eBay bildet ein System aus vielen Agenten, die als Käufer und Verkäufer durch Auktionen in Wechselwirkung zu einander treten.
Solche menschbasierten Systeme unterscheiden sich von anderen in der Physik bekannten Systemen durch das rationale Handeln ihrer Agenten. Von daher ist die Möglichkeit, diese Systeme mittels der Methoden der Statistischen Physik zu untersuchen umstritten. Anders als in Systemen unintelligenter (unbelebter) Agenten können Menschen in ihrem sozialen System zwischen mehreren existierenden rationalen Strategien wählen. Diese Konzepte werden in Abschnitt 6.3 diskutiert.
In vorliegender Arbeit wird jedoch evident, dass die Rationalität der Agenten bei einem großen Teil der Eigenschaften des Systems keine Rolle spielt und vernachlässigt werden kann.
Der Hauptteil dieser Arbeit befasst sich mit empirischen Befunden und deren Erklärungen. Die empirischen Befunde sind die Ergebnisse einer statistischen Analyse zweier Datensammlungen, die über Internetschnittstellen von eBay gewonnen wurden. Die Screen-Scraping- und HTTP-Parsing-Methoden ermöglichten einen formatierten Datengewinn aus über 200,000 Auktionen.
Die Auswahl der Auktionen wurde so getroffen, dass die Datensammlung ein möglichst weites Spektrum der Endpreise und anderer dynamischer Variablen abdeckt.
Bei einigen Studien müssen Agentenaktivitäten über einen längeren Zeitraum beobachtet werden. Das ist der Grund, weshalb zwei unterschiedliche Datenmengen gesammelt wurden. Die eine beinhaltet zufällig ausgewählte Auktionen aus allen eBay-Kategorien, die gleichzeitig an einem bestimmten Datum liefen. Die andere beinhaltet alle Auktionen zweier ausgewählter Kategorien über den Zeitraum von 10 Monaten. Die Methoden und eine detaillierte Beschreibung der Datenmengen sind in Abschnitt 5 gegeben.
Die Ergebnisse der statistischen Analyse können in zwei Hauptgruppen von Wahrscheinlichkeitsverteilungen von Messgrößen und Beziehungen zwischen unterschiedlichen Messgrößen geteilt werden. Die Wahrscheinlichkeitsverteilungen diverser Größen wie Endpreis, Startpreis, Auktionsdauer, Anzahl der positiven Bewertungen, Gesamtanzahl der Gebote eines Käufers, Gesamtanzahl der Angebote eines Verkäufers, Gebotszeit (Zeit bis zum Auktionsende), Return (definiert als die relative Differenz zwischen Gebot und aktuellem Preis) sind untersucht worden.
Wir haben auch die Beziehungen zwischen korrelierten Variablen untersucht. Beispiele sind Beziehungen zwischen Endpreis und Anzahl der Gebote, Endpreis und Anzahl der verschiedenen Bieter und Return und der relativen Gebotzeit.

Unsere Studie zeigt, dass die Wahrscheinlichkeitsdichten vieler Messgrößen einfache Formen annehmen, wie Potenzgesetze oder Exponentialfunktionen.
Potenzgesetze sind als untrennbare Eigenschaft mit komplexen Systemen verbunden. Sie tauchen überall auf, unabhängig davon, ob man Natur, Menschen, Evolution der Organismen oder Börsenmärkte beobachtet.
Sie sind seit mehr als einem Jahrhundert durch sozial- und naturwissenschaftliche Beobachtungen zu einer der wichtigsten Fragen der Wissenschaft geworden. Gegenstand früher Betrachtungen waren die Verteilung des Volksvermögens, bekannt als Pareto-Verteilung [1] und die Verteilung von Wörtern in Sprachen, bekannt als Zipfsches Gesetz [2].
Ähnliche Potenzgesetze wurden in der Physik bei der Untersuchung der Systeme in deren kritischen Punkten in Zusammenhang mit Phasenübergängen zweiter Ordnung beobachtet. Solche Systeme weisen eine skalenfreie Struktur auf, wobei Fluktuationen auf allen Längenskalen eine wichtige Rolle spielen.
Erst nach Arbeiten von Wilson [4] und der Einführung der Theorie der Renormierungsgruppe haben Physiker die Möglichkeit solche Probleme systematisch zu studieren. Viel Arbeit wurde in den letzten drei Jahrzehnten investiert und eine hilfreiche Klassifikation, die kritische Systeme in Universalitätsklassen zusammenfasst, wurde gefunden. Diese ermöglichte Physikern eine große Anzahl von unterschiedlichen Systemen durch Identifizierung ihrer zugehörigen Universalitätsklassen zu verstehen.
Dieses Paradigma der statistischen Physik war immer eine Motivation für Physiker, sich mit nichtphysikalische Systeme, die aus vielen wechselwirkenden Agenten bestehen, zu beschäftigen.
Die Wechselwirkungen der Agenten in nichtphysikalischen Systemen sind allerdings entweder nicht bekannt oder mathematisch nicht definierbar. Sogar mit Hilfe von Annahmen und Approximationen bleiben solche Systeme wegen der Nichtlinearität der Wechselwirkungen analytisch unlösbar.
Neue numerische Methoden wie Monte Carlo-Simulationen bieten die Möglichkeit einer numerischen Behandlung. Die übliche Vorgehensweise ist die Definition von Modellen und deren minimalen mikroskopischen Regeln. Diese Modelle werden für Simulationen verwendet. Die Ergebnisse und Vorhersagen werden dann mit den realen Systemen und bekannten empirischen Daten verglichen.
Eine solche Methode wurde von Physikern für die Modellierung des Straßenverkehrs und der Fußgängerdynamik, biologischer und ökonomischer Systeme usw. bereits erfolgreich verwendet.
Potenzgesetze werden als eine der emergenten Eigenschaften komplexer Systeme beobachtet. Komplexe Systeme teilen eine Reihe von Eigenschaften. Die wichtigsten sind Emergenz, Nichtlinearität, Offenheit und Selbstorganisation. Trotz vieler Bemühungen solche Eigenschaften zu verstehen, gilt die Komplexitätstheorie als eine neue Wissenschaft, die noch am Anfang ihrer Entwicklung steht.
Eines der wichtigsten Konzepte befasst sich mit der Ähnlichkeit vom Verhalten dieser Systeme mit Gleichgewichtssystemen in ihren kritischen Punkten. Dieses Konzept, bekannt als selbstorganisierte Kritikalität, besagt, dass Systeme im Nichtgleichgewicht dazu neigen, sich von selbst in einen kritischen Zustand zu entwickeln.

In Abschnitt 2 werden sowohl wichtige Eigenschaften komplexer Systeme als auch Prozesse für die Erzeugung von Potenzgesetzen diskutiert.
Unsere Untersuchung zeigt, dass manche empirisch gefundenen Potenzgesetze mittels einfacher Wachstumsprozesse zu verstehen sind. Zum Beispiel lässt sich die Verteilung der Anzahl der Angebote eines Verkäufers durch Anwendung des so genannten Yule-Prozesses quantitativ verstehen (siehe Abschnitt 6.1).
Eine der interessantesten Ergebnisse unserer Untersuchung ist die skalenfreie Beziehung zwischen mittlerem Preis und Anzahl der Gebote. Ähnliche Beziehungen sind auch zwischen anderen Messgrößen gefunden worden. Diese skalenfreien Beziehungen sind Hinweise darauf, dass das System sich im kritischen Zustand befindet.
Skalenfreiheit taucht häufig in Studien zu menschbasierten Systemen auf und ist direkt mit der fraktalen Struktur des Systems, welches selbstähnliche Eigenschaften besitzt, verbunden.
Ein anderes interessantes Ergebnis dieser Studie zeigt, dass die Bieter es vorziehen, ihre Gebote erst gegen Ende der Auktion abzugeben. Dieses Ergebnis steht nicht im Einklang mit den Vorhersagen der Spieltheorie. Diese besagt, dass in Zweitpreisauktionen (Vickrey-Auktionen) mit rationalen Agenten die Bietzeit keine Rolle spielt und es keine Tendenz geben sollte, weniger zu bieten als man glaubt, dass die Ware wert sei.
Viele spieltheoretische Studien haben das Ziel, Strategien, die zum Spätbieten führen, als rationale Strategien plausibel zu machen. Eine kurze Zusammenfassung dieser Studien ist in Abschnitt 4 gegeben.
Unsere Studie zeigt, dass die Wahl der Bietstrategie nicht der einzige Parameter ist, der für das exponentielle Verhalten der Gebotzeitverteilung verantwortlich ist, sondern dabei andere Mechanismen wie Sortierungskriterien von eBay eine sehr wichtige Rolle spielen.
In der Arbeit gefundene einfache robuste Beziehungen wie die zwischen Endpreis und Anzahl der Gebote bieten die Möglichkeit, normale und manipulierte Auktionen automatisch zu unterscheiden.
Eines der häufigsten Internetverbrechen ist das verbotene Shill-bidding, wobei ein Verkäufer versucht beim Bieten auf eigenen Auktionen den Preis eines Artikels künstlich hochzutreiben. Wir fanden heraus, dass erfolgreiches Shill-bidding zu einer statistisch messbaren Abweichung vom Durchschnittsverhalten führt.
Die Kritikalität dieses Systems kann durch die Wahrscheinlichkeitsverteilung von Returns wiedergegeben werden. Die Verwendung dieser Verteilung liefert Erklärungen für andere gemessene Verteilungen und Verhalten. Dieses wurde in einem einfachen Model getestet.
Trotz qualitativer und quantitativer Erklärungen für viele beobachtete Eigenschaften können wir zurzeit kein vereinheitlichtes Modell für dieses System vorschlagen.
Diese Studie hat herausgefunden, dass viele Eigenschaften als einfache Ergebnisse stochastischer Prozesse zu verstehen sind. Einige andere sind ohne Einbeziehung der Rationalität der Agenten nicht zu erklären.

English Abstract

The present work studies eBay online auctions as a complex system, where agents in form of sellers and bidders interact in a large number. In contrast to unintelligent agents of physical systems, humans have the choice between different possible rational strategies.
The empirical findings in this work are the result of statistical analysis of two major sets of data with more than 200,000 auctions. Probability distribution functions and relations between different variables are studied. Statistical analysis of the eBay data shows that the probability density functions for a wide range of quantities follow rather simple functionalities like exponential and power laws. Similar power-law distributions were observed in physics when studying the behavior of systems at their critical points, related to second order phase transitions. Although lots of relations and distributions could be understood qualitatively and quantitatively, there exists, up to now, no unifying model, which describes all of the observations.
As an application we have found that a kind of fraud known as shill bidding leads to significant deviations from the average behavior.
eBay regarded as a complex system seems to have features of totally stochastic processes, irrational agents and fully rational agents altogether.

Deutsche Kurzzusammenfassung

Die vorliegende Arbeit beschäftigt sich mit eBay-Auktionen als einem komplexen System. Das Onlineauktionshaus eBay bildet ein System aus vielen Agenten, die als Käufer und Verkäufer durch Auktionen in Wechselwirkung zu einander treten. Anders als in Systemen unintelligenter (unbelebter) Agenten können Menschen in ihrem sozialen System zwischen mehreren existierenden rationalen Strategien wählen.
Die empirischen Befunde dieser Arbeit sind die Ergebnisse einer statistischen Analyse von über 200,000 Auktionen. Die Wahrscheinlichkeitsverteilungen und die Beziehungen zwischen korrelierten Variablen wurden untersucht. Unsere Studie zeigt, dass die Wahrscheinlichkeitsdichten vieler Messgrößen einfache Formen annehmen, wie Potenzgesetze oder Exponentialfunktionen. Ähnliche Potenzgesetze wurden in der Physik bei der Untersuchung der Systeme in deren kritischen Punkten in Zusammenhang mit Phasenübergängen zweiter Ordnung beobachtet.
Als Anwendung konnte gezeigt werden, dass betrügerisches Shill-bidding zu einer deutlichen Abweichung vom mittleren Verhalten führt.
Diese Studie hat herausgefunden, dass viele Eigenschaften als einfache Ergebnisse stochastischer Prozesse zu verstehen sind. Einige andere sind ohne Einbeziehung der Rationalität der Agenten nicht zu erklären.

Danksagung

Ein ganz besonderes Dankeschön geht an Herrn Priv-Doz. Dr. A. Schadschneider. Seine Diskussions- und Hilfsbereitschaft trugen wesentlich zum Gelingen dieser Arbeit bei. Insbesondere seine Aufmerksamkeit und Unterstützung in den letzten Jahren lernte ich sehr zu schätzen. Eine solche angenehme Zusammenarbeit wünsche ich mir auch für die Zukunft.

Ein weiteres Dankeschön geht an die Herren Prof. J. Zittartz und Prof. A. Rosch, die immer für eine gute Atmosphäre in ihrer Arbeitsgruppe sorgten.

Ein schöner Dank geht ebenfalls an Herrn Prof. Dr. J. Krug für interessante Diskussionen und hilfreiche Unterstützung.

Frau Prof. Dr. M. Paczuski und Herrn Prof. Dr. A. Ockenfels möchte ich für interessante Gespräche danken.

Die sowohl entspannte als auch produktive Stimmung in Raum 111 mit Marc Andre Ahrens, Erik Bartel, Alexander John, Clemens Müller-Gugenberger und Dr. Christian Dziurzik schufen eine sehr nette Arbeitsatmosphäre.

Desweiteren hatten Dr. Ansgar Kirchner, Dr. Andreas Kemper, Dr. Frank Zielen, Dr. Kavita Jain, Prof. Dr. Katsuhiro Nishinari und Dr. Hossein Yavartanoo immer ein offenes Ohr für meine Fragen, dafür ein Dankeschön.

Danken möche ich auch Dr. Andreas Sindermann. So manches Computerprogramm wäre ohne sein Hilfe sicherlich nicht so schnell zustande gekommen.

Für das sehr aufmerksame Korrekturlesen der vorliegenden Arbeit bin ich Andreas Schadschneider, Alexander John und Erik Bartel sehr dankbar.

Mein besonderer Dank gilt meiner Familie, meinem Sohn Samuel und Julia Heygster für ihre stete Unterstützung.

Acknowledgments

I would like to thank Priv-Doz. Dr. A. Schadschneider for supervising this thesis and interesting discussions.
I am grateful to Prof. J. Zittartz and Prof. A. Rosch for providing an enjoyable working environment at the institute für Theoretische Physik der Universität zu Köln.
I wish to thank Prof. Dr. J. Krug, Prof. Dr. M. Paczuski, Prof. Dr. A. Ockenfels, Dr. Ansgar Kirchner, Marc Andre Ahrens, Erik Bartel, Alexander John, Clemens Müller-Gugenberger, Dr. Andreas Kemper, Dr. Frank Zielen, Dr. Kavita Jain, Prof. Dr. Katsuhiro Nishinari, Dr. Andreas Sindermann, Dr. Christian Dziurzik and Dr. Hossein Yavartanoo for fruitful discussions and hints.
Finally, I would like to thank my parents, brothers, my son Samuel and Julia Heygster for their support. Especially I wish to thank Ali Namazi for awaking my interest for physics and for his scientific support during my whole education.
My love goes to Samuel Namazi.

Erklärung

Ich versichere, dass ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfmittel vollständig angegeben und die Stellen der Arbeit – einschliesslich Tabellen, Karten, und Abbildungen –, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie noch nicht veröffentlicht worden ist sowie, dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen dieser Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Herrn Privat-Dozent Dr. A. Schadschneider betreut worden.

Köln, den 27. Mai 2005

Lebenslauf

Persönliche Daten:

Name	Alireza Namazi
Geburtstag	23.04.1972
Geburtsort	Teheran
Familienstand	geschieden, ein Kind (Samuel H. A.)
Staatsangehörigkeit	Iraner

Studium:

1990-1996	Physikstudium an der Sharif University of Technology in Teheran (Iran).
Sep. 1996	Abschluss: B.S. (Bachelor of Science).
1997-04/2000	Physikstudium an der Universität zu Köln.
04/2000 - 04/2001	Diplomarbeit am Zentrum für angewandte Informatik der Universität zu Köln (ZAIK) mit dem Thema "Boundary induced phase transitions in a traffic flow model with metastable states."
April 2001	Abschluss: Diplom Physik
seit Februar 2002	Promotion in theoretischer Physik bei Priv.-Doz. Andreas Schadschneider.

Wissenschaftliche Arbeitsverhältnisse:

1998 -1999	Studentische Hilfskraft am Seminar für Wirtschaftsinformatik und Operational Research der Universität zu Köln. Mitarbeit am FpEdit-Projekt.
1999 - 2000	Studentische Hilfskraft am Mathematischen Institut der Universität zu Köln. Entwicklung von MathNet.
seit Feb. 2002	Wissenschaftliche Hilfskraft bzw. wissenschaftlicher Mitarbeiter am Institut für Theoretische Physik der Universität zu Köln.

Stipendien und Preise:

03/1999 - 03/2000	Akademisches Auslandsamt der Universität zu Köln.
05/2000 - 05/2001	Living System AG. Dritter Platz bei dem Internetideen-Wettbewerb.
2004	Kölner Gymnasial- und Stiftungsfonds.